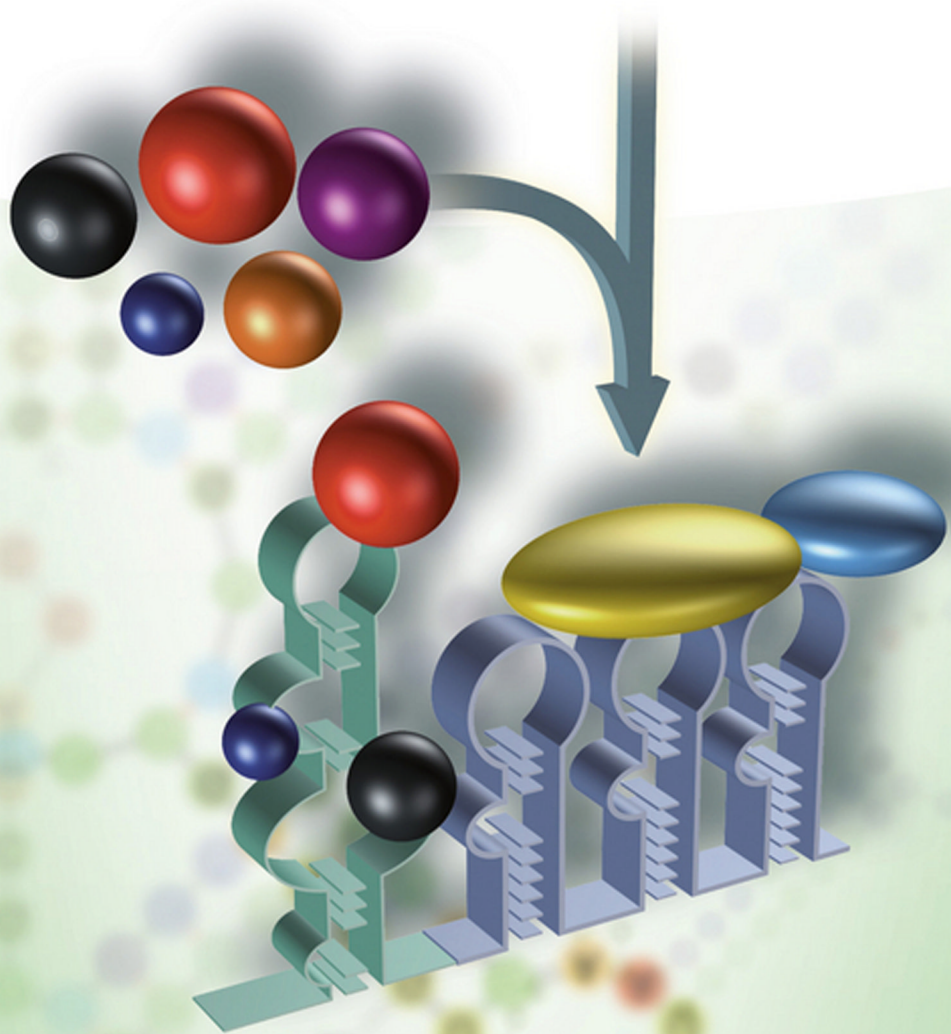


Edited by Roland K. Hartmann, Albrecht Bindereif,
Astrid Schön, and Eric Westhof

Handbook of RNA Biochemistry

Second, Completely Revised and Enlarged Edition



*Edited by Roland K. Hartmann,
Albrecht Bindereif, Astrid Schön,
and Eric Westhof*

Handbook of RNA Biochemistry

Related Titles

Meister, G.

RNA Biology

An Introduction

2011

ISBN: 978-3-527-32278-7

Gu, J., Bourne, P. E. (eds.)

Structural Bioinformatics

2009

ISBN: 978-0-470-18105-8

Gjerde, D. T., Hoang, L., Hornby, D.

RNA Purification and Analysis

**Sample Preparation, Extraction,
Chromatography**

2009

ISBN: 978-3-527-32116-2

*Edited by Roland K. Hartmann, Albrecht Bindereif,
Astrid Schön, and Eric Westhof*

Handbook of RNA Biochemistry

Second, Completely Revised and Enlarged Edition

WILEY-VCH
Verlag GmbH & Co. KGaA

The Editors

Prof. Dr. Roland K. Hartmann

Philipps-Universität Marburg
Institut für Pharma. Chemie
Marbacher Weg 6
35037 Marburg
Germany

Prof. Dr. Albrecht Bindereif

Justus-Liebig-Universität
Institut für Biochemie
Heinrich-Buff-Ring 58
35392 Gießen
Germany

Dr. Astrid Schön

Universität Leipzig
Molecular Cell Therapy
Deutscher Platz 5
04103 Leipzig
Germany

Prof. Dr. Eric Westhof

CNRS - UPR 9002, Inst. de
Biol. Mol. et Cellulaire
15 rue René Descartes
06708 Strasbourg
France

All books published by **Wiley-VCH** are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.:

applied for

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <<http://dnb.d-nb.de>>.

© 2014 Wiley-VCH Verlag GmbH & Co.
KGaA, Boschstr. 12, 69469 Weinheim,
Germany

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Composition Laserwords Private Limited,
Chennai, India

Printing and Binding Markono Print Media
Pte Ltd, Singapore

Cover Design Schulz Grafik-Design,
Fußgönheim

Print ISBN: 978-3-527-32776-8

ePDF ISBN: 978-3-527-65055-2

ePub ISBN: 978-3-527-65054-5

mobi ISBN: 978-3-527-65053-8

oBook ISBN: 978-3-527-64706-4

Printed in Singapore
Printed on acid-free paper

Contents

Preface XXXV

List of Contributors XXXVII

Part I RNA Synthesis and Detection 1

1	Enzymatic RNA Synthesis Using Bacteriophage T7 RNA Polymerase	3
	<i>Markus Gößringer, Dominik Helmecke, Karen Köhler, Astrid Schön, Leif A. Kirsebom, Albrecht Bindereif, and Roland K. Hartmann</i>	
1.1	Introduction	3
1.2	Description of Method – T7 Transcription <i>In vitro</i>	4
1.2.1	Templates	4
1.2.1.1	Strategy (i): Insertion into a Plasmid	4
1.2.1.2	Strategy (ii): Direct Use of Templates Generated by PCR	5
1.2.1.3	Strategy (iii): Annealing of a T7 Promoter DNA Oligonucleotide to a Single-Stranded Template	5
1.2.2	Special Demands on the RNA Product	5
1.2.2.1	Homogeneous 5' and 3' Ends, Small RNAs, Functional Groups at the 5' End	5
1.2.2.2	Modified Substrates	6
1.3	Transcription Protocols	8
1.3.1	Transcription with Unmodified Nucleotides	9
1.3.2	Transcription with 2'-Fluoro-Modified Nucleotides	16
1.3.3	T7 Transcripts with 5'-Cap Structures	17
1.3.4	Purification	18
1.4	Troubleshooting	20
1.4.1	Low or No Product Yield	20
1.5	Rapid Preparation of T7 RNA Polymerase	21
1.5.1	Required Material	21
1.5.1.1	Medium	21
1.5.1.2	Buffers and Solutions	21
1.5.1.3	Electrophoresis and Chromatography	22
1.5.2	Procedure	22
1.5.2.1	Cell Growth, Induction, and Test for Expression of T7 RNAP	22

1.5.2.2	Purification of T7 RNAP	23
1.5.3	Notes and Troubleshooting	24
	References	25
2	Production of RNAs with Homogeneous 5'- and 3'-Ends	29
	<i>Mario Mörl and Roland K. Hartmann</i>	
2.1	Introduction	29
2.2	Description of Approach	30
2.2.1	Cis-Cleaving Autocatalytic Ribozyme Cassettes	30
2.2.1.1	The 5'-Cassette	30
2.2.1.2	The 3'-Cassette	30
2.2.1.3	Purification of Released RNA Product and Conversion of End Groups	31
2.2.2	Trans-Cleaving Ribozymes for the Generation of Homogeneous 3' Ends	33
2.2.3	Further Strategies toward Homogeneous Ends	35
2.3	Critical Experimental Steps, Changeable Parameters, Troubleshooting	36
2.3.1	Construction of Cis-Cleaving 5'- and 3'-Cassettes	36
2.4	PCR Protocols	37
2.5	Potential Problems	42
	References	42
3	RNA Ligation	45
	<i>Janne J. Turunen, Liudmila V. Pavlova, Martin Hengesbach, Mark Helm, Sabine Müller, Roland K. Hartmann, and Mikko J. Frilander</i>	
3.1	General Introduction	45
3.1.1	T4 Polynucleotide Ligases	46
3.1.2	Reaction Mechanism	46
3.1.3	Advantages of T4 DNA Ligase for RNA Ligation	49
3.1.4	Chapter Structure	49
3.2	RNA Ligation Using T4 DNA Ligase (T4 Dnl)	50
3.2.1	Overview of the RNA Ligation Method Using the T4 DNA Ligase (T4 Dnl)	51
3.2.2	Large-Scale Transcription and Purification of RNAs	53
3.2.3	Generating Homogeneous Acceptor 3'-Ends for Ligation	53
3.2.4	Site-Directed Cleavage with RNase H	54
3.2.5	Dephosphorylation and Phosphorylation of RNAs	56
3.2.6	RNA Ligation	57
3.2.7	Troubleshooting	58
3.3	Simultaneous Splint Ligation of Five RNA Fragments to Generate RNAs for FRET Experiments	66
3.3.1	Introduction	66
3.3.2	Construct Design	68
3.3.3	Troubleshooting	70

3.3.3.1	Low Overall Ligation Efficiency	70
3.3.3.2	Undesired Ligation By-products	70
3.3.3.3	RNA Degradation	70
3.4	T4 RNA Ligase(s)	70
3.4.1	Introduction	70
3.4.2	Mechanism and Substrate Specificity	71
3.4.2.1	Early Studies	71
3.4.2.2	Substrate Specificity and Reaction Conditions	72
3.4.3	Applications of T4 RNA Ligase	73
3.4.3.1	End-Labeling	73
3.4.3.2	Circularization	75
3.4.3.3	Intermolecular Ligation of Polynucleotides	75
3.4.4	T4 RNA Ligation of Large RNA Molecules	76
3.4.5	Application Examples and Protocols	79
3.4.5.1	Production of Full-Length tRNAs	79
3.4.6	Troubleshooting	84
	References	84
4	Northern Blot Detection of Small RNAs	89
	<i>Benedikt M. Beckmann, Arnold Grünweller, and Roland K. Hartmann</i>	
4.1	Introduction	89
4.1.1	Isolation of RNA	89
4.1.1.1	Kits	90
4.1.1.2	Do it Yourself	90
4.1.1.3	Quality Control	90
4.1.2	Native <i>versus</i> Denaturing Gels	90
4.1.3	Transfer of RNA and Fixation to Membranes	91
4.1.4	Hybridization with a Complementary Probe	92
4.1.4.1	Design of DNA/LNA Mixmer Probes	92
4.1.5	Detection of DIG-Labeled Probes	95
4.1.6	Troubleshooting	95
4.1.7	Application Example	96
4.1.8	Limitations of the Method	96
4.2	Northern Hybridization Protocols	98
	References	102
5	Rapid, Non-Denaturing, Large-Scale Purification of <i>In Vitro</i> Transcribed RNA Using Weak Anion-Exchange Chromatography	105
	<i>Laura E. Easton, Yoko Shibata, and Peter J. Lukavsky</i>	
5.1	Introduction	105
5.2	Materials	106
5.2.1	Cloning and Plasmid Purification	106
5.2.2	<i>In Vitro</i> Transcription	106
5.2.3	Weak Anion-Exchange FPLC	107

5.3	Protocols for Plasmid Design and Preparation, RNA Transcription, and Weak Anion-Exchange Purification	107
5.4	Troubleshooting	115
	Acknowledgments	115
	References	116
6	3'-Terminal Attachment of Fluorescent Dyes and Biotin	117
	<i>Dagmar K. Willkomm and Roland K. Hartmann</i>	
6.1	Introduction	117
6.2	Description of Method	118
6.3	History of the Method	118
6.4	Troubleshooting	124
6.4.1	Problems Caused Before the Labeling Reaction	124
6.4.1.1	Quality of the RNA 3' Ends	124
6.4.1.2	Purity of the RNA to Be Labeled	124
6.4.2	Problems with the Labeling Reaction Itself	124
6.4.2.1	pH of Reagents	124
6.4.2.2	Stability of Reagents	124
6.4.3	Postlabeling Problems	125
6.4.3.1	Removal of Labeling Reagents	125
6.4.3.2	Loss of RNA Material during Downstream Purification	125
6.4.3.3	Stability of Labeled RNA	125
	Acknowledgment	125
	References	125
7	Chemical RNA Synthesis, Purification, and Analysis	129
	<i>Brian S. Sproat</i>	
7.1	Introduction	129
7.2	Description	132
7.2.1	The Solid-Phase Synthesis of RNA	132
7.2.2	Deprotection	136
7.2.3	Purification	138
7.2.3.1	Anion-Exchange HPLC Purification	139
7.2.3.2	Reversed-Phase HPLC Purification of Trityl-On RNA	140
7.2.3.3	Detritylation of Trityl-On RNA	142
7.2.3.4	Desalting by HPLC	142
7.2.4	Analysis of the Purified RNA	143
7.3	Troubleshooting	144
	References	147
8	Modified RNAs as Tools in RNA Biochemistry	151
	<i>Thomas E. Edwards and Snorri Th. Sigurdsson</i>	
8.1	Introduction	151
8.1.1	Modification Strategy: the Phosphoramidite Method	152
8.1.2	Modification Strategy: Postsynthetic Labeling	154

8.2	Description of Methods	156
8.2.1	Postsynthetic Modification: the 2'-Amino Approach	156
8.2.2	Reaction of 2'-Amino Groups with Succinimidyl Esters	158
8.2.3	Reaction of 2'-Amino Groups with Aromatic Isothiocyanates	158
8.2.4	Reaction of 2'-Amino Groups with Aliphatic Isocyanates	159
8.3	Experimental Protocols	159
8.3.1	Synthesis of Aromatic Isothiocyanates and Aliphatic Isocyanates	160
8.3.2	Postsynthetic Labeling of 2'-Amino-Modified RNA	161
8.3.3	Postsynthetic Labeling of 4-Thiouridine-Modified RNA	164
8.3.4	Verification of Label Incorporation	164
8.3.5	Potential Problems and Troubleshooting	165
	References	166

Part II Structure Determination 173

9	Direct Determination of RNA Sequence and Modification by Radiolabeling Methods	175
	<i>Olaf Gimple and Astrid Schön</i>	
9.1	Introduction	175
9.2	General Methods	175
9.3	Isolation of Pure RNA Species from Biological Material	176
9.3.1	Preparation of Size-Fractionated RNA	176
9.3.2	Isolation of a Single Unknown RNA Species Following a Functional Assay	176
9.3.2.1	Solutions for Electrophoresis, Staining, and Elution of RNAs from Gels	176
9.3.2.2	Two-Dimensional Electrophoresis of RNA	177
9.3.2.3	Comments on the Electrophoretic Purification and Elution of RNA Species	178
9.3.3	Isolation of Single RNA Species with Partially Known Sequence	178
9.3.3.1	Materials for Hybrid Selection of Single RNA Species	178
9.4	Radioactive Labeling of RNA Termini	180
9.4.1	Materials for 5'-End Labeling of RNAs	180
9.4.2	3'-Labeling of RNAs	181
9.4.2.1	Materials for 3'-End Labeling of RNAs	182
9.5	Sequencing of End-Labeled RNA	183
9.5.1	Sequencing by Base-Specific Enzymatic Hydrolysis of End-Labeled RNA	184
9.5.1.1	Materials Required for Enzymatic Sequencing	185
9.5.1.2	Interpretation and Troubleshooting	186
9.5.2	Sequencing by Base-Specific Chemical Modification and Cleavage	187
9.5.2.1	Materials Required for Chemical Sequencing	188

9.5.2.2	Interpretation and Troubleshooting	189
9.6	Determination of Terminal RNA Sequences by Two-dimensional Mobility Shift	190
9.6.1	Materials Required for Mobility Shift Analysis	190
9.7	Determination of Modified Nucleotides by Postlabeling Methods	194
9.7.1	Analysis of Total Nucleotide Content	195
9.7.1.1	Materials Required for RNA Nucleotide Analysis	195
9.7.1.2	Interpretation and Troubleshooting	197
9.7.2	Determination of Position and Identity of Modified Nucleotides	198
9.7.2.1	Interpretation and Troubleshooting	199
9.8	Conclusions and Outlook	201
	Acknowledgments	202
	References	202
10	Probing RNA Structure <i>In Vitro</i> with Enzymes and Chemicals	205
	<i>Anne-Catherine Helfer, Cédric Romilly, Clément Chevalier, Efthimia Lioliou, Stefano Marzi, and Pascale Romby</i>	
10.1	Introduction	205
10.2	Enzymatic and Chemical Probes	207
10.2.1	Enzymes	207
10.2.2	Base-Specific Chemical Probes	210
10.2.3	Backbone-Specific Chemical Probes	211
10.3	<i>In Vivo</i> DMS Modification	222
10.3.1	Generalities	222
10.3.2	<i>In Vivo</i> Probing	222
10.4	Commentary	223
10.4.1	Critical Parameters	223
10.4.1.1	RNA Preparation	223
10.4.1.2	Homogeneous RNA Conformation	224
10.4.1.3	Chemical and Enzymatic Probing	224
10.4.1.4	<i>In Vivo</i> DMS Mapping	225
10.5	Troubleshooting	225
	Acknowledgments	227
	References	227
11	Probing RNA Solution Structure by Photocrosslinking: Incorporation of Photoreactive Groups at RNA Termini and Determination of Crosslinked Sites by Primer Extension	231
	<i>Michael E. Harris</i>	
11.1	Introduction	231
11.1.1	Applications of RNA Modifications	231
11.1.2	Techniques for the Incorporation of Modified Nucleotides	232
11.2	Description	233

11.2.1	5'-End Modification by Transcription Priming	233
11.2.2	Chemical Phosphorylation of Nucleosides to Generate 5'-Monophosphate or 5'-Monophosphorothioate Derivatives	234
11.2.3	Attachment of an Aryl Azide Photocrosslinking Agent to a 5'-Terminal Phosphorothioate	236
11.2.4	3'-Addition of an Aryl Azide Photocrosslinking Agent	238
11.3	Troubleshooting	240
11.4	Probing RNA Structure by Photoaffinity Crosslinking with 4-Thiouridine and 6-Thioguanosine	240
11.4.1	Introduction	240
11.4.2	Description	243
11.4.2.1	General Considerations: Reaction Conditions and Concentrations of Interacting Species	243
11.4.2.2	Application Example – RNase P RNA and s ⁶ G-Modified Precursor tRNA	244
11.4.2.3	Generation and Isolation of Crosslinked RNAs	246
11.4.2.4	Primer Extension Mapping of crosslinked Nucleotides	247
11.4.3	Troubleshooting	249
	References	250
12	Terbium(III) Footprinting as a Probe of RNA Structure and Metal Binding Sites	255
	<i>Dinari A. Harris, Gabrielle C. Todd, and Nils G. Walter</i>	
12.1	Introduction	255
12.2	Application Example	261
12.3	Troubleshooting	265
12.4	Frontiers in Footprinting Data Analysis	265
	References	266
13	Pb²⁺-Induced Cleavage of RNA	269
	<i>Leif A. Kirsebom and Jerzy Ciesiolka</i>	
13.1	Introduction	269
13.2	Pb ²⁺ -Induced Cleavage to Probe Metal Ion Binding Sites, RNA Structure, and RNA–Ligand Interactions	271
13.2.1	Probing High-Affinity Metal Ion Binding Sites	271
13.2.2	Pb ²⁺ -Induced Cleavage and RNA Structure	273
13.2.3	Pb ²⁺ -Induced Cleavage to Study RNA–Ligand Interactions	274
13.2.4	Pb ²⁺ -Induced Cleavage of RNA <i>In Vivo</i>	275
13.3	Troubleshooting	279
13.3.1	No Pb ²⁺ -Induced Cleavage Detected	279
13.3.2	Complete Degradation of the RNA	280
13.3.3	<i>In Vivo</i>	280
	Acknowledgments	280
	References	281

14	Identification and Characterization of Metal Ion Coordination Interactions with RNA by Quantitative Analysis of Thiophilic Metal Ion Rescue of Site-Specific Phosphorothioate Modifications	285
	<i>Michael E. Harris</i>	
14.1	Introduction	285
14.1.1	Thiophilic Metal Ion Rescue of RNA Phosphorothioate Modifications	286
14.2	Purification of Phosphorothioate Stereoisomers by RP-HPLC	290
14.3	Techniques for Incorporation of Phosphorothioates into RNA	291
14.4	Kinetic Analysis of Thiophilic Metal Ion Rescue	293
14.5	Data Analysis by Fitting to Simple Equilibrium Models	295
	References	297
15	Probing RNA Structure and Ligand Binding Sites on RNA by Fenton Cleavage	301
	<i>Corina G. Heidrich and Christian Berens</i>	
15.1	Introduction	301
15.2	Comments and Troubleshooting	312
	References	314
16	Measuring the Stoichiometry of Magnesium Ions Bound to RNA	319
	<i>Andrew J. Andrews and Carol A. Fierke</i>	
16.1	Introduction	319
16.2	Separation of Free Mg^{2+} from RNA-bound Mg^{2+}	320
16.3	Forced Dialysis Is the Preferred Method for Separating Bound and Free Mg^{2+}	321
16.4	Alternative Methods for Separating Free and Bound Mg^{2+} Ions	323
16.5	Determining the Concentration of Free Mg^{2+} in the Flow-Through	324
16.6	How to Determine the Concentration of Mg^{2+} Bound to the RNA and the Number of Binding Sites on the RNA	324
16.7	Conclusion	327
16.8	Troubleshooting	327
	References	327
17	Nucleotide Analog Interference Mapping and Suppression (NAIM/NAIS): a Combinatorial Approach to Study RNA Structure, Folding, and Interaction with Proteins	329
	<i>Olga Fedorova, Marc Boudvillain, and Christina Waldsich</i>	
17.1	Introduction	329
17.1.1	NAIM: a Combinatorial Approach for RNA Structure–Function Analysis	329
17.1.1.1	Description of the Method	330
17.1.2	NAIS: a Chemogenetic Tool for Identifying RNA Tertiary Contacts and Interaction Interfaces	332

17.1.2.1	General Concepts	332
17.1.2.2	Applications: Elucidating Tertiary Contacts in Group I and Group II Ribozymes	332
17.2	Experimental Protocols for NAIM	333
17.2.1	Nucleoside Analog Thiotriphosphates	333
17.2.2	Preparation of Transcripts Containing Phosphorothioate Analogs	335
17.2.2.1	Tips and Troubleshooting	336
17.2.3	Radioactive Labeling of the RNA Pool	337
17.2.4	The Selection Step of NAIM: Three Applications to Studies of RNA Function	339
17.2.4.1	Group II Intron Ribozyme Activity: Selection through Transesterification	339
17.2.4.2	Group II Ribozyme Folding: Selection through Mg^{2+} -Induced Compaction of RNA	344
17.2.4.3	RNA–Protein Interactions: a One-Pot Reaction for Studying Rho-Independent Transcription Termination	347
17.2.4.4	RNA–Protein Interactions: Elucidation of the Rho Helicase Activation Mechanism via Unwinding Activity	351
17.2.5	Iodine Cleavage of RNA Pools	354
17.2.5.1	Experimental Procedure	355
17.2.5.2	Tips and Troubleshooting	355
17.2.6	Analysis and Interpretation of NAIM Results	355
17.2.6.1	Quantification of Interference Effects	355
17.3	Experimental Protocols for NAIS	358
17.3.1	Design and Construction of RNA Mutants	358
17.3.1.1	General Considerations	358
17.3.1.2	Preparation of RNA Molecules Containing Single-Atom Substitutions	359
17.3.2	Functional Analysis of Mutants for NAIS Experiments	362
17.3.3	The Selection Step for NAIS	362
17.3.4	Data Analysis and Presentation	363
	Acknowledgments	364
	References	364
18	Nucleotide Analog Interference Mapping (NAIM): Application to the RNase P System	369
	<i>Simona Cuzic-Feltens and Roland K. Hartmann</i>	
18.1	Introduction	369
18.1.1	Nucleotide Analog Interference Mapping (NAIM) – the Approach	369
18.1.2	Critical Aspects of the Method	371
18.1.2.1	Analog Incorporation	371
18.1.2.2	Functional Assays	372
18.1.2.3	Factors Influencing the Outcome of NAIM Studies	372

18.1.3	Interpretation of Results	373
18.2	NAIM Analysis of <i>cis</i> -Cleaving RNase P RNA-tRNA Conjugates	375
18.2.1	Biochemical and kinetic characterization of a <i>cis</i> -Cleaving <i>E. coli</i> RNase P RNA-tRNA Conjugate	375
18.2.2	Application Example	378
18.2.3	Data Evaluation	386
18.3	Troubleshooting	387
18.3.1	RNA Transcription Reaction Did Not Work	387
18.3.2	RNA Degradation	389
18.3.3	Inefficient RNA Elution from Denaturing PAA Gels	389
18.3.4	RNA Is Degraded after Elution	389
18.3.5	Inefficient 3'- or 5'-End-Labeling	389
18.3.6	Iodine-Induced Hydrolysis Failed or Was Inefficient	391
18.3.7	Unsatisfactory Gel Performance after Iodine Cleavage (Band Smearing, Curved Bands, Irregular Shape of Bands, Unequal Band Migration in Different Lanes, and Insufficient Band Separation)	392
	References	393
19	Identification of Divalent Metal Ion Binding Sites in RNA/DNA-Metabolizing Enzymes by Fe(II)-Mediated Hydroxyl Radical Cleavage	397
	<i>Yan-Guo Ren, Niklas Henriksson, and Anders Virtanen</i>	
19.1	Introduction	397
19.2	Probing Divalent Metal Ion Binding Sites	398
19.2.1	Fe(II)-Mediated Hydroxyl Radical Cleavage	398
19.2.2	How to Map Divalent Metal Ion Binding Sites	399
19.2.3	How to Use Aminoglycosides as Functional and Structural Probes	401
19.3	Notes and Troubleshooting	403
	References	404
20	RNA Structure and Folding Analyzed Using Small-Angle X-Ray Scattering	407
	<i>Nathan J. Baird, Jeremy West, and Tobin R. Sosnick</i>	
20.1	Introduction	407
20.2	Description of Method	410
20.2.1	General Requirements	410
20.2.2	SAXS Application Example	411
20.2.3	General Information	412
20.2.4	Question 1: The Global Conformation of the S-Domain Folding Intermediate	412
20.2.5	Question 2: The Stable, Extended Conformation of the S-Domain Folding Intermediate	414
20.2.6	Question 3: The Utility of Low-Resolution Real-Space Reconstructions in RNA Modeling	416

20.3	Troubleshooting	421
20.3.1	Problem 1: Radiation Damage and Aggregation	421
20.3.2	Problem 2: High Scattering Background	422
20.3.3	Problem 3: Scattering Results Cannot Be Fit to Simple Models	422
20.4	Conclusions – Outlook	422
	Acknowledgments	423
	Abbreviations	423
	References	423
21	Temperature-Gradient Gel Electrophoresis of RNA	427
	<i>Detlev Riesner and Gerhard Steger</i>	
21.1	Introduction	427
21.2	Method	428
21.2.1	Principle	428
21.2.2	Instruments	429
21.2.3	Handling	429
21.3	Optimization of Experimental Conditions	430
21.3.1	Pore Size of the Gel Matrix	430
21.3.2	Electric Field	430
21.3.3	Ionic Strength and Urea	431
21.4	TGGE – General Interpretation Rules	431
21.5	Examples of TGGE Applications	433
21.5.1	Example 1: Analysis of Different RNA Molecules in a Single TGGE	434
21.5.2	Example 2: Analysis of Structure Transitions in a Single RNA – Detection of Specific Structures by Oligonucleotide Hybridization	435
21.5.3	Example 3: Analysis of Mutants	438
21.5.4	Example 4: Detection of Protein–RNA Complexes by TGGE	439
21.5.5	Outlook	442
	References	443
22	UV Melting Studies with RNA	445
	<i>Philippe Dumas, Eric Ennifar, Francois Disdier, and Philippe Walter</i>	
22.1	Introduction	445
22.2	A Simplified Account of the Physical Basis of UV Absorption	445
22.3	Definitions and Nomenclature	446
22.4	Well-Known and Less Well-Known Characteristics of UV Absorption by Nucleic Acids Bases	447
22.5	The Basis of UV Melting Experiments for Thermodynamic Studies	449
22.5.1	The Only Valid Definition of a Melting Temperature	450
22.5.2	Reminders	450
22.5.3	Unimolecular Transitions	451
22.5.4	Bimolecular Transitions	452

22.5.4.1	Entropic Considerations	452
22.5.4.2	Basic and Less Basic Equations about Melting Curves Involving Bimolecular Transitions	454
22.5.4.3	Higher Order Transitions	455
22.5.4.4	Influence of the Temperature Dependence of the Absorbance Parameters	455
22.5.4.5	The Different Ways of Obtaining T_m , ΔH , and ΔS	455
22.6	The Two-State Approximation and Its Limitations	459
22.7	Equilibrium and Non-equilibrium	459
22.8	A Common Pitfall with Self-Complementary Sequences	460
22.9	Extracting Thermodynamic Information from Melting Curves of Large RNAs	461
22.10	Parameters Influencing the Melting Temperature	462
22.11	Practical Problems	463
22.11.1	Evaporation during Heating: an Important Improvement	463
22.11.2	Sloping Baseline	464
22.12	A Neat Experimental Solution to the Sloping Baseline	468
22.12.1	pH Variation and Buffers	468
22.12.2	RNA Degradation	470
22.12.3	Heating Rate and Data Sampling	471
22.12.4	Experimental Data Processing	472
22.12.5	Softwares	473
	Acknowledgment	473
	Appendix A: Difference between T_m and T_{max} and DMC Normalization	473
	Appendix B: Experimental Setup against Evaporation	475
	Appendix C: The Subtleties with Partial Derivatives for ΔC_p Determination	475
	Appendix D: Buffer pK_a Variation with the Temperature	476
	References	476
23	RNA Crystallization	481
	<i>Jiro Kondo, Claude Sauter, and Benoît Masquida</i>	
23.1	Introduction	481
23.2	RNA Purification	482
23.2.1	HPLC Purification	482
23.2.2	Gel Electrophoresis	483
23.2.3	RNA Recovery	484
23.2.3.1	Elution of the RNA from the Gel	484
23.2.3.2	Concentrating and Desalting	484
23.3	RNA Crystallization	485
23.3.1	Renaturing the RNA	485
23.3.2	Search for Crystallization Conditions	485
23.3.3	Evaluation of Crystallization Assays	488
23.3.4	The Optimization Process	489

23.3.5	Designing RNA Constructs with Improved Crystallization Capabilities	491
23.3.6	Crystallizing Complexes with Organic Ligands: the Example of Aminoglycosides	493
23.4	Conclusions	494
	References	495
24	Studying RNA Using Single Molecule Fluorescence Resonance Energy Transfer	499
	<i>Felix Spenkuch, Olwen Domingo, Gerald Hinze, Thomas Basché, and Mark Helm</i>	
24.1	Introduction	499
24.1.1	The Advantages of Single Molecule Fluorescence Resonance Energy Transfer	499
24.1.2	Chapter Scope	500
24.1.3	Typical Topics of RNA Dynamics Addressed by Single Molecule FRET	500
24.2	Theory of Fluorescence Resonance Energy Transfer	502
24.3	Experimental Design	503
24.3.1	Considerations for Construct Design	503
24.4	smFRET Experiments Using Immobilized Molecules	505
24.4.1	Instrumental Setup	505
24.4.2	Means of Signal Correction and Data Analysis	505
24.4.3	The Choice of Dye Pairs for FRET	507
24.4.4	Buffer Handling in Single Molecule Experiments	508
24.4.5	Strategies for Dye Labeling of RNA Constructs	508
24.4.6	Postsynthetic Labeling of Alkyne-Containing RNA Oligonucleotides	509
24.4.7	Tuning Dye Endurance: Antifading Agents	510
24.5	Troubleshooting	520
24.5.1	RNase Contamination	520
24.5.2	Removal of Unbound Fluorophores	521
24.5.3	Drying of Samples	521
24.5.4	Donor-Only Populations	521
24.5.5	Too Dense or Too Sparse Surface Coverage	521
	References	522
25	Atomic Force Microscopy Imaging and Force Spectroscopy of RNA	527
	<i>Malte Bussiek, Antonie Schöne, and Wolfgang Nellen</i>	
25.1	Introduction	527
25.2	AFM Imaging of RNA Structures	528
25.2.1	General Preconditions: Mode of Operation, Data Analysis, and Resolution	528
25.2.2	Surface Preparation Conditions	531

25.2.3	Imaging in Liquid	535
25.2.4	Experimental Example of Salt-Dependent RNA Folding Using a Designed RNA Construct	535
25.3	Example Protocol: RNA Preparation for AFM Imaging in Air Using PL-Coated Mica	537
25.4	Troubleshooting	538
25.5	Force Spectroscopy AFM	540
25.6	Outlook	544
	Acknowledgments	544
	References	544

Part III RNA Genomics & Bioinformatics, Global Approaches 547

26	Secondary Structure Prediction	549
	<i>Gerhard Steger</i>	
26.1	Introduction	549
26.2	Thermodynamics	550
26.3	Formal Background	552
26.4	mFold and UNAFold	555
26.4.1	Input to the mFold Server	556
26.4.1.1	Sequence Name	556
26.4.1.2	Sequence	556
26.4.1.3	Constraints	556
26.4.1.4	Further Parameters	558
26.4.1.5	Immediate <i>versus</i> Batch Jobs	561
26.4.2	Output from the mFold Server	561
26.4.2.1	Energy Dot Plot	561
26.4.2.2	Extra Files	563
26.4.2.3	Download All Foldings	563
26.4.2.4	View ss-Count Information	564
26.4.2.5	View Individual Structures	564
26.4.2.6	Dot Plot Folding Comparisons	565
26.5	RNAfold	565
26.5.1	Input to the RNAfold Server	566
26.5.1.1	Sequence and Constraints	566
26.5.1.2	Further Parameters	567
26.5.1.3	Immediate <i>versus</i> Batch Jobs	568
26.5.2	Output from the RNAfold Server	570
26.5.2.1	Text Output of Secondary Structure	570
26.5.2.2	Probability Dot Plot	570
26.5.2.3	Graphical Output of Secondary Structure	570
26.5.2.4	Mountain Plot	571
26.6	Troubleshooting	571
	Acknowledgment	573
	References	573

27	RNA Secondary Structure Analysis Using Abstract Shapes	579
	<i>Robert Giegerich and Björn Voß</i>	
27.1	Introduction to Abstract Shape Analysis	579
27.1.1	Looking Deeper into the RNA Folding Space	579
27.1.2	Overview of Functions of Abstract Shape Analysis	580
27.1.3	Definition of Shape Abstraction	580
27.1.3.1	Shapes	580
27.1.3.2	Shape Abstraction Function	581
27.1.3.3	Shape Representative Structures (shreps)	581
27.1.3.4	Levels of Abstraction	581
27.1.3.5	Shape Probabilities	582
27.1.3.6	Consensus Shape	582
27.1.4	General Caveats when Working with Abstract Shapes	582
27.1.5	Applications of Abstract Shape Analysis	583
27.2	Protocol 1: Computing Shape Representative Structures	584
27.2.1	Useful Parameters for <i>RNAshapes</i>	585
27.3	Protocol 2: Probabilistic Shape Analysis	585
27.3.1	Useful Parameters	587
27.4	Protocol 3: Comparative Shape Analysis from Aligned Sequences	587
27.4.1	Useful Parameters for <i>RNALishapes</i>	588
27.5	Protocol 4: Comparative Shape Analysis from Unaligned Sequences	588
27.5.1	Useful Parameters for <i>RNAshapes</i>	592
27.6	<i>RNAshapes</i> Parameter Overview	592
27.7	<i>RNALishapes</i> Parameter Overview	593
	References	594
28	Screening Genome Sequences for known RNA Genes or Motifs	595
	<i>Daniel Gautheret</i>	
28.1	Introduction	595
28.2	Choosing the Right Search Program	596
28.3	Overview of the RNA Search Procedure	597
28.4	Assessing Search Specificity	598
28.5	A Test Case: Looking for Homologs of a Bacterial sRNA	600
28.5.1	Building a First Training Set with BLASTN	600
28.5.2	Alignment and Structure Prediction	602
28.5.3	Searching with HMMER	604
28.5.4	Searching with RNAMOTIF	606
28.5.5	Searching with ERPIN	609
28.5.6	Searching with INFERNAL	614
28.6	Conclusion	615
28.7	Supplemental Data	615

28.8	Program Versions and Download Sites	616
	Acknowledgments	616
	References	616
29	Homology Search for Small Structured Non-coding RNAs	619
	<i>Manja Marz, Stefanie Wehner, and Peter F. Stadler</i>	
29.1	Introduction	619
29.2	Materials	619
29.2.1	Sequence Data	619
29.2.2	Web Services	620
29.2.3	Web Service-Independent Software	621
29.3	Protocol: mascRNAs	621
29.3.1	The Seed	622
29.3.2	Low-Hanging Fruits: Initial BLAST Search	623
29.3.3	Initial Secondary Structure Model	624
29.3.4	Drilling Deep – Structure-Based Searches	625
29.4	Concluding Remarks	629
	Acknowledgments	630
	References	630
30	Predict RNA 2D and 3D Structure over the Internet Using MC-Tools	633
	<i>Stephen Leong Koan, Jonathan Roy, Marc Parisien, and François Major</i>	
30.1	Introduction	633
30.2	Materials	634
30.2.1	Equipment	634
30.2.2	Data	634
30.3	MC-Tools	635
30.3.1	MC-Fold	635
30.3.2	MC-Cons	636
30.3.3	MC-Sym	636
30.4	Troubleshooting	663
	Acknowledgments	663
	References	663
31	S2S-Assemble2: a Semi-Automatic Bioinformatics Framework to Study and Model RNA 3D Architectures	667
	<i>Fabrice Jossinet and Eric Westhof</i>	
31.1	Introduction	667
31.2	S2S: an Interactive RNA Alignment Viewer and Editor	668
31.3	Assemble2: an Interactive RNA 3D Modeler	671
31.4	The Semi-Automatic Architecture of S2S and Assemble2	672
31.5	Installation of S2S and Assemble2	673
	References	685

32	Molecular Dynamics Simulations of RNA Systems	687
	<i>Pascal Auffinger</i>	
32.1	Introduction	687
32.2	MD Methods	689
32.3	Simulation Setups	689
32.3.1	Selecting an Appropriate Starting Structure	689
32.3.1.1	Model-Built Structures	689
32.3.1.2	X-Ray and Neutron Diffraction Structures	689
32.3.1.3	Cryo-Electron Microscopy (Cryo-EM) Structures	690
32.3.1.4	NMR Structures	690
32.3.2	Checking the Starting Structure	690
32.3.2.1	Conformational Checks	690
32.3.2.2	Rare Non-covalent Interactions	691
32.3.2.3	Protonation Issues	692
32.3.2.4	Solvent	692
32.3.3	Adding Hydrogen Atoms	693
32.3.4	Choosing the Environment (Crystal, Liquid) and Ion Types	693
32.3.5	Setting the Box Size and Placing the Ions and Water	693
32.3.5.1	Box Size	693
32.3.5.2	Monovalent Ions	693
32.3.5.3	Divalent Ions	694
32.3.5.4	Minimal Salt Conditions	694
32.3.5.5	Water Molecules	694
32.3.5.6	Building Initial Solute and Solvent Configurations	694
32.3.6	Choosing the Program and Force Field	695
32.3.6.1	Programs	695
32.3.6.2	Force Fields	695
32.3.6.3	Parameterization of Modified Nucleotides, Ligands, and Ions	696
32.3.6.4	Clustering Artifacts and Ion Parameters	696
32.3.6.5	Water Models	696
32.3.7	Treatment of Electrostatic Interactions	697
32.3.8	Other Simulation Parameters	697
32.3.8.1	Thermodynamic Ensemble	697
32.3.8.2	Temperature and Pressure	698
32.3.8.3	Shake, Time Steps, and Update of the Non-bonded Pair List	698
32.3.8.4	The “Flying Ice Cube Problem”	698
32.3.9	Equilibration	699
32.3.10	Sampling	699
32.3.10.1	How Long Should a Simulation Be?	699
32.3.10.2	When to Stop a Simulation	700
32.3.10.3	Multiple Molecular Dynamics (MMD) Simulations	701
32.3.10.4	Simulations of Large Systems	701
32.4	Analysis	701
32.4.1	Evaluating the Quality of the Trajectories	701
32.4.1.1	Consistency Checks	702

32.4.1.2	Comparison with Experimental Data	702
32.4.1.3	Visualization	702
32.4.1.4	Validation through Statistical Survey of Structural Databases	703
32.4.2	Convergence Issues	703
32.4.3	Conformational Parameters	703
32.4.4	Data Analysis	704
32.4.4.1	Clustering	704
32.4.4.2	Analysis Packages	704
32.4.4.3	Solvent Analysis	704
32.5	Perspectives	704
	Acknowledgments	705
	References	705
33	Identification and Characterization of Small Non-coding RNAs in Bacteria	719
	<i>Dimitri Podkaminski, Marie Bouvier, and Jörg Vogel</i>	
33.1	Introduction	719
33.2	Expression-Based Discovery of sRNAs	720
33.2.1	Microarray	720
33.2.2	High-Throughput Sequencing and RNA-Seq	721
33.2.3	Hfq Coimmunoprecipitation	724
33.3	Expression-Independent Searches	726
33.3.1	Biocomputational Approaches	726
33.3.2	Genomic SELEX	728
33.4	Deciphering the Biological Role of an sRNA	728
33.4.1	sRNA Expression Profile	729
33.4.2	sRNA Deletion	729
33.4.3	sRNA Overexpression	731
33.4.4	sRNA Pulse Expression Combined with Transcriptome Analysis	733
33.4.5	sRNA Libraries	734
33.4.6	Finding sRNA-Associated Proteins	735
33.4.7	Biocomputational Approaches to Find Targets	736
33.5	Experimental Target Validation	737
33.5.1	Reporter Gene Fusions and sRNA Chimera	738
33.5.2	<i>In vitro</i> RNA–RNA Footprinting	739
33.5.3	<i>In vitro</i> Characterization of sRNA Function	741
33.6	Conclusions	742
	Acknowledgments	776
	References	776
34	The Identification of Bacterial Non-coding RNAs through Complementary Approaches	787
	<i>Björn Voß and Wolfgang R. Hess</i>	
34.1	Introduction	787

34.2	Computational Prediction	787
34.2.1	Workflow	788
34.2.2	Results and Interpretation	789
34.2.3	Alternative Approaches	790
34.2.4	Troubleshooting	791
34.2.4.1	Choice of Genomes	791
34.2.4.2	Short mRNAs and Dual-Function RNAs	794
34.3	Experimental Approaches for High-Throughput RNomics in Bacteria	794
34.3.1	Microarray Analysis	794
34.3.1.1	Considerations for the Design of Tiling Microarrays	795
34.3.1.2	Considerations for the Design of Expression Microarrays	796
34.3.1.3	Direct Labeling of RNA for Microarray Hybridization	796
34.4	Troubleshooting	799
	Acknowledgments	799
	References	800
35	Experimental RNomics, a Global Approach to Identify Non-coding RNAs in Model Organisms, and RNPomics to Analyze the Non-coding RNP Transcriptome	801
	<i>Mathieu Rederstorff and Alexander Hüttenhofer</i>	
35.1	Introduction	801
35.2	Computational Analysis of ncRNA Sequences	811
35.3	Notes	812
35.4	Computational Analysis of ncRNA Sequences	816
35.5	Notes	816
	Acknowledgments	817
	References	817
36	Computational Methods for Gene Expression Profiling Using Next-Generation Sequencing (RNA-Seq)	821
	<i>John C. Castle</i>	
36.1	Introduction	821
36.2	Procedure Overview	822
36.2.1	Understand the Experiment and the Molecular Biology Protocol	823
36.2.1.1	Library Generation	823
36.2.1.2	Sequencing	825
36.2.2	Align Reads	826
36.2.3	Associate Reads with Transcripts	827
36.2.4	Determine Expression and Uncertainty	828
36.2.5	Normalization	828
36.2.6	Output and Viewing	828
36.2.7	Troubleshooting	829
36.2.8	The Future Is Bright!	830

36.3	Protocols: Useful Algorithms, Formats, and Tools	830
	References	830
37	Characterization and Prediction of miRNA Targets	833
	<i>Jean Hausser and Mihaela Zavolan</i>	
37.1	Introduction	833
37.2	Description	834
37.2.1	Building a Set of “Positives” and “Negatives”; Obtaining Examples of Functional and Non-functional miRNA Binding Sites	835
37.2.1.1	Comparative genomics	836
37.2.1.2	miRNA perturbation and omics	837
37.2.1.3	Immunoprecipitation of RISC components	838
37.2.1.4	Measuring translation repression directly with polysome profiles	839
37.2.1.5	Which data set should one use for inferring properties that characterize functional miRNA binding sites?	839
37.2.2	Properties of Functional miRNA Binding Sites	840
37.2.2.1	The “seed” binding criterion	840
37.2.2.2	Evolutionary conservation	841
37.2.2.3	Stability of the miRNA–mRNA duplex	841
37.2.2.4	Structural accessibility	841
37.2.2.5	Sequence composition	842
37.2.2.6	Spatial effects	842
37.2.3	Combining Properties and Examples into a Predictive Model	843
37.2.3.1	Inferring properties that consistently predict miRNA targeting across data sets	843
37.2.3.2	Training a miRNA target prediction model	846
37.3	Troubleshooting	847
37.3.1	Using miRNA target predictions in an experimental setting	847
37.3.1.1	How accurate are miRNA target predictions?	848
37.3.1.2	Which miRNA target prediction method should I use?	849
37.3.1.3	How many targets does a miRNA have?	850
37.3.1.4	Why does a particular high-confidence predicted target not change in response to miRNA overexpression?	850
37.3.1.5	Transcript x is a target of miRNA y according to method z, yet it does not have an “miRNA y seed match” in the 3' UTR	850
37.3.1.6	The list of targets predicted by method x has a different type of identifiers (Entrez Gene ID/RefSeq ID/Ensembl transcript/ . . .) than the list predicted by method y or the list that one obtains in a large-scale validation experiment (e.g., microarray measurement)	851
37.3.2	The Complexity of Gene Regulation and its Impact on Designing Accurate miRNA Target Prediction Methods	851
	References	853

38	Barcoded cDNA Libraries for miRNA Profiling by Next-Generation Sequencing 861
	<i>Markus Hafner, Neil Renwick, John Pena, Aleksandra Mihailovic, and Thomas Tuschl</i>
38.1	Introduction 861
38.2	Overview of the Method 862
38.3	Troubleshooting 872
	Acknowledgments 872
	References 872
39	Transcriptome-Wide Identification of Protein Binding Sites on RNA by PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) 877
	<i>Jessica I. Hoell, Markus Hafner, Markus Landthaler, Manuel Ascano, Thalia A. Farazi, Greg Wardle, Jeff Nusbaum, Pavol Cekan, Mohsen Khorshid, Lukas Burger, Mihaela Zavolan, and Thomas Tuschl</i>
39.1	Introduction 877
39.2	Troubleshooting 897
	Acknowledgments 897
	References 897
40	Global Analysis of Protein–RNA Interactions with Single-Nucleotide Resolution Using iCLIP 899
	<i>Julian König, Nicholas J. Mc Glinicy, and Jernej Ule</i>
40.1	Introduction 899
40.2	Procedure 900
40.2.1	Overview 900
40.2.2	Antibody and Library Preparation Quality Control 902
40.2.3	Oligonucleotide Design 903
40.2.4	Troubleshooting 904
	Acknowledgments 917
	References 917
	Part IV RNA Function, RNP Analysis, SELEX, RNAi 919
41	Use of RNA Affinity Matrices for the Isolation of RNA Binding Proteins 921
	<i>Markus Englert, Bettina Späth, Steffen Schiffer, Sylvia Rösch, Hildburg Beier, and Anita Marchfelder</i>
41.1	Introduction 921
41.2	Applications 927
41.2.1	Purification of the Nuclear tRNase Z from Wheat Germ 927
41.2.2	Purification of the tRNA-Splicing Ligase from Wheat Germ 930

41.3	Notes	932
	References	932
42	Biotin-Based Affinity Purification of RNA–Protein Complexes	935
	<i>Marco Preußner, Silke Schreiner, Inna Grishina, Zsolia Palfi, Jingyi Hui, and Albrecht Bindereif</i>	
42.1	Introduction	935
42.2	Materials	937
42.2.1	Biotinylated Probes	937
42.2.2	Affinity Matrices	937
42.2.3	Cell Extracts	938
42.2.4	Buffers and Solutions	938
42.2.5	Additional Materials	939
42.3	Methods	939
42.3.1	Affinity Purification of RNA–Protein Complexes (RNPs)	939
42.3.1.1	Depletion of Total Cell Lysate from SAg-Binding Material (Preclearing)	940
42.3.1.2	Preblocking Streptavidin Agarose Beads	941
42.3.1.3	Affinity Selection of RNPs for Biochemical Studies	941
42.3.1.4	Elution of Affinity-Selected RNPs for Functional Studies by a Displacement Oligonucleotide	945
42.3.2	Affinity Purification of Specific RNA Binding Proteins by Biotinylated RNAs	948
42.3.3	Depletion of Nuclear Extract with Biotinylated RNA	951
42.4	Troubleshooting	952
42.4.1	Biotinylated 2'OMe RNA Oligonucleotides	952
42.4.2	Extracts and Buffers	952
42.4.3	Optimization of the Experimental Conditions, When Yields Are Low	952
42.4.4	Optimization of the Experimental Conditions in the Case of High Background	953
	References	953
43	Affinity Purification of Spliceosomal and Small Nuclear Ribonucleoprotein Complexes	957
	<i>Julia Dannenberg, Patrizia Fabrizio, Cindy L. Will, and Reinhard Lührmann</i>	
43.1	Introduction	957
43.2	Immunoaffinity Purification	958
43.2.1	Generation of Antipeptide Antibodies: Peptide Selection Criteria	958
43.3	RNA Aptamer-Based Affinity Purification	963
43.3.1	Approaches for the Isolation of Native Spliceosomal Complexes	963
	Acknowledgments	971
	References	972

44	Study of RNA–Protein Interactions and RNA Structure in Ribonucleoprotein Particles (RNPs)	975
	<i>Virginie Marchand, Annie Mougin, Agnès Méreau, Isabelle Behm-Ansmant, Yuri Motorin, and Christiane Branlant</i>	
44.1	Introduction	975
44.2	Methods	978
44.2.1	RNP Reconstitution	978
44.2.1.1	Equipment, Materials, and Reagents	978
44.2.1.2	RNA Preparation and Renaturation Step	980
44.2.2	EMSA	981
44.2.2.1	EMSA Method	981
44.2.2.2	Supershift Method	983
44.2.2.3	Identification of Proteins Contained in RNP by EMSA Experiments Coupled to a Second Gel Electrophoresis and Western Blot Analysis	984
44.2.3	Purification of RNPs Reconstituted in Complex Cellular Extracts	986
44.2.4	Methods for RNP Purification Using Tobramycin–Sepharose or MS2-MBP Affinity Chromatography	987
44.2.4.1	Equipment and Materials Common to the Two Approaches	987
44.2.4.2	RNP Purification Using Tobramycin–Sepharose	987
44.2.4.3	Formation of RNPs in the Cellular Extract	989
44.2.4.4	Elution of Purified RNPs under Native Conditions	989
44.2.4.5	MS2-MBP Affinity Chromatography	989
44.2.4.6	Elution and Analysis of Purified RNPs	990
44.2.4.7	Analysis of the Purified RNP Protein Content	990
44.2.5	Probing of RNA Structure	991
44.2.5.1	Properties of the Probes Used	991
44.2.5.2	Equipment, Material, and Reagents	993
44.2.5.3	Probing Method	994
44.2.6	UV Crosslinking and Immunoselection	999
44.2.6.1	Equipment, Materials, and Reagents	1000
44.2.6.2	UV-Crosslinking Method	1003
44.3	Commentaries and Pitfalls	1005
44.3.1	RNP Purification and Reconstitution	1005
44.3.1.1	RNA Purification and Renaturation	1005
44.3.1.2	EMSA	1005
44.3.1.3	Tobramycin–Sepharose Affinity Chromatography	1006
44.3.2	Probing Conditions	1006
44.3.2.1	Choice of the Probes Used	1006
44.3.2.2	Ratio of RNA/Probes	1007
44.3.3	UV Crosslinking	1008
44.3.3.1	Photoreactivity of Individual Amino Acids and Nucleotide Bases	1008
44.3.3.2	Labeled Nucleotide in RNA	1008
44.3.4	Immunoprecipitations	1008

44.3.4.1	Efficiency of Immunoabsorbents for Antibody Binding	1008
44.4	Troubleshooting	1008
44.4.1	RNP Purification by Tobramycin–Sepharose or MS2-MBP Affinity Chromatography	1008
44.4.2	RNP Reconstitution	1009
44.4.3	RNA Probing	1009
44.4.4	UV Crosslinking	1009
44.4.5	Immunoprecipitations	1009
	Acknowledgments	1010
	References	1010
45	Immunopurification of Endogenous RNAs Associated with RNA Binding Proteins <i>In vivo</i>	1017
	<i>Minna-Liisa Änkö and Karla M. Neugebauer</i>	
45.1	Introduction	1017
45.2	Description of Methods	1017
45.2.1	Overview	1017
45.2.2	Analysis of Coimmunoprecipitated RNA	1022
45.2.2.1	Microarray Analysis of Immunopurified RNA	1022
45.2.2.2	RT-PCR Analysis of Immunopurified RNA	1024
45.2.2.3	Next-Generation Sequencing of Immunopurified RNA	1025
45.3	Troubleshooting	1025
45.3.1	Critical Points and Common Problems	1025
45.3.2	Uncrosslinked or Crosslinked RNA Immunoprecipitation	1026
45.3.3	Microarray Data Analysis	1026
45.4	Conclusions	1027
	Acknowledgments	1027
	References	1027
46	Protein–RNA Crosslinking in Native Ribonucleoprotein Particles	1029
	<i>Olexandr Dybkov, Henning Urlaub, and Reinhard Lührmann</i>	
46.1	Introduction	1029
46.2	Overall Strategy	1030
46.3	UV Crosslinking	1031
46.4	Identification of UV-Induced Protein–RNA Crosslinking Sites by Primer Extension Analysis	1033
46.5	Identification of Crosslinked Proteins	1037
46.6	Troubleshooting	1040
	Acknowledgments	1050
	References	1050
47	Sedimentation Analysis of Ribonucleoprotein Complexes	1055
	<i>Tanja Rösel, Jan Medenbach, Andrey Damianov, Silke Schreiner, and Albrecht Bindereif</i>	
47.1	Introduction	1055

47.2	Glycerol Gradient Centrifugation	1056
47.3	Fractionation of Ribonucleoproteins (RNPs) by Cesium Chloride Density Gradient Centrifugation	1061
	References	1065
48	Identification and Characterization of RNA Binding Proteins through Three-Hybrid Analysis	1067
	<i>Felicia Scott and David R. Engelke</i>	
48.1	Introduction	1067
48.2	Basic Strategy of the Method	1068
48.3	Detailed Components	1070
48.3.1	Yeast Reporter Strain	1070
48.3.2	Plasmids	1070
48.3.3	Hybrid RNA	1071
48.3.3.1	Technical Considerations for the Hybrid RNA	1071
48.3.4	Activation Domain FP2	1073
48.3.4.1	Technical Considerations for the Activation Domain of FP2	1074
48.3.5	Positive Controls	1075
48.4	Troubleshooting	1079
48.5	Additional Applications	1081
48.6	Summary	1082
	Acknowledgments	1083
	References	1083
49	Experimental Identification of MicroRNA Targets	1087
	<i>Michaela Beitzinger and Gunter Meister</i>	
49.1	Introduction	1087
49.2	Troubleshooting and Notes	1093
49.3	Buffers and Solutions	1094
	References	1095
50	Aptamer Selection against Biological Macromolecules: Proteins and Carbohydrates	1097
	<i>Franziska Peter and C. Stefan Voertler</i>	
50.1	Introduction	1097
50.2	General Strategy	1098
50.2.1	Choosing a Suitable Target	1100
50.2.1.1	Protein Targets	1100
50.2.1.2	Carbohydrate Targets	1101
50.2.2	Immobilization of the Target	1102
50.2.3	Selection Assays	1103
50.2.4	Design and Preparation of the Library	1103
50.3	Running the <i>In vitro</i> Selection Cycle	1104
50.4	Analysis of the Selection Outcome	1106
50.5	Troubleshooting	1107

Acknowledgments 1131

References 1131

51 In Vitro Selection against Small Targets 1139

*Dirk Eulberg, Christian Maasch, Werner G. Purschke, and
Sven Klussmann*

51.1 Introduction 1139

51.2 Target Immobilization 1142

51.2.1 Covalent Immobilization 1143

51.2.1.1 Epoxy-Activated Matrices 1143

51.2.1.2 NHS-Activated Matrices 1145

51.2.1.3 Pyridyl Disulfide-Activated Matrices 1146

51.2.2 Non-covalent Immobilization 1147

51.3 Nucleic Acid Libraries 1148

51.3.1 Library Design 1148

51.3.2 Starting Pool Preparation 1149

51.4 Enzymatics 1150

51.4.1 Reverse Transcription 1151

51.4.2 Polymerase Chain Reaction 1152

51.4.3 In Vitro Transcription 1153

51.5 Partitioning 1154

51.6 Binding Assays 1159

51.6.1 Equilibrium Dialysis 1159

51.6.2 Equilibrium Filtration Analysis 1160

51.6.3 Isocratic Competitive Affinity Chromatography 1161

References 1162

52 SELEX Strategies to Identify Antisense and Protein Target Sites in RNA or hnRNP Complexes 1165

Martin Lützelberger, Martin R. Jakobsen, and Jørgen Kjems

52.1 Introduction 1165

52.1.1 Applications for Antisense 1166

52.1.2 Selecting Protein Binding Sites 1166

52.2 Construction of the Library 1166

52.2.1 Generation of Random DNA Fragments from Genomic or Plasmid
DNA 1168

52.2.2 Preparing RNA Libraries from Plasmid, cDNA, or Genomic
DNA 1168

52.3 Identification of Optimal Antisense Annealing Sites in RNAs 1169

52.4 Identification of Natural RNA Substrates for Proteins and Other
Ligands 1171

52.5 Cloning, Sequencing, and Validating the Selected Inserts 1171

52.6 Troubleshooting 1172

52.6.1 Sonication of Plasmid DNA does not Yield Shorter Fragments 1172

52.6.2 Inefficient Ligation 1172

52.6.3	Inefficient <i>Mme</i> I Digestion	1172
52.6.4	The Amplification of the Unselected Library is Inefficient	1173
52.6.5	The Library Appears to be Non-Random in the Unselected Pool	1173
52.6.6	The Selected RNAs do not Bind to Native Protein	1173
	References	1182
53	Genomic SELEX	1185
	<i>Jennifer L. Boots, Katarzyna Matylla-Kulinska, Marek Zywicki, Bob Zimmermann, and Renée Schroeder</i>	
53.1	Introduction	1185
53.2	Description of the Methods	1186
53.2.1	Library Construction	1186
53.2.2	Choice of Bait	1188
53.2.3	SELEX Procedure	1188
53.2.3.1	Transcription of Genomic Library into RNA Library	1190
53.2.3.2	Counter Selection	1190
53.2.3.3	Positive Selection	1190
53.2.3.4	Recovery and Amplification of Selected Sequences	1191
53.2.3.5	Neutral SELEX	1192
53.2.3.6	Cloning and Sequencing	1194
53.2.4	Troubleshooting	1194
53.3	Evaluation of Obtained Sequences	1194
53.3.1	Computational Analysis of SELEX-Derived Sequences	1194
53.3.1.1	Read Filtering and Cleaning	1196
53.3.1.2	Genome Mapping	1196
53.3.1.3	Assembly and Annotation	1197
53.3.1.4	Enrichment Analysis	1197
53.3.1.5	Benefits of Sequencing the Initial Library	1198
53.3.1.6	Identification of the Binding Motif	1198
53.3.2	Biochemical Analysis of the Genomic Aptamers	1199
53.3.2.1	Validation of the RNA–Protein Interaction	1199
53.3.2.2	Expression Analysis of Genomic Aptamers	1199
53.3.2.3	Reconstruction of the Whole-Transcript-Comprising Genomic Aptamer	1200
53.3.2.4	Determining the Function of the RNA–Protein Interaction	1200
53.4	Conclusions and Outlook	1202
	Acknowledgments	1202
	References	1202
54	<i>In vivo</i> SELEX Strategies	1207
	<i>Thomas A. Cooper</i>	
54.1	Introduction	1207
54.2	Procedure Overview	1208
54.2.1	Design of the Randomized Exon Cassette	1210
54.2.2	Design of the Minigene	1212

54.2.3	RT-PCR Amplification	1213
54.2.4	Monitoring for Enrichment of Exon Sequences That Function as Splicing Enhancers	1213
54.2.5	Troubleshooting	1214
	Acknowledgments	1218
	References	1219
55	Gene Silencing Methods Using Vector-Encoded siRNAs or miRNAs	1221
	<i>Ying Poi Liu and Ben Berkhout</i>	
55.1	Introduction	1221
55.2	Background Information	1221
55.3	Construction of shRNA Vectors	1223
55.4	Construction of miRNA Vectors	1228
55.5	Construction of Extended shRNAs and lhrRNAs	1229
55.6	Production of Lentiviral Vectors Encoding Anti-HIV-1 shRNAs or e-shRNAs	1230
55.6.1	Troubleshooting	1234
	References	1235
56	Using Chemical Modification to Enhance siRNA Performance	1243
	<i>Jesper B. Bramsen, Arnold Grünweller, Roland K. Hartmann, and Jørgen Kjems</i>	
56.1	Introduction	1243
56.2	Numerous siRNA Designs: What siRNA Architecture to Choose?	1243
56.3	siRNA Tolerance Toward Modification	1244
56.4	Tools for Chemical Modification of siRNAs	1245
56.4.1	siRNA Backbone Modifications	1246
56.4.2	Ribose 2'-OH Substitutions	1248
56.4.3	Alteration of the Ribose Backbone	1251
56.4.4	Base Modifications	1252
56.5	Improving siRNA Potency	1252
56.6	Enhancing siRNA Nuclease Resistance	1253
56.6.1	siRNA Stability and Ribonucleases	1253
56.6.2	Strategies for siRNA Stabilization	1254
56.7	Enhancing siRNA Silencing Duration	1255
56.8	siRNA Immunogenicity	1256
56.8.1	Cellular Response to siRNA	1256
56.8.2	Chemical Modification Can Abrogate siRNA Immunogenicity	1257
56.9	Reducing siRNA Off-Target Effects by Chemical Modification	1258
56.9.1	Off-Target Effects Caused by miRNA-Like Activities	1258
56.9.2	Reducing Off-Targeting by Chemical Modification of the siRNA Guide Strand Seed Region	1258
56.9.3	Avoiding Passenger Strand Off-Targeting	1259

56.10	Chemical Modifications Can Improve siRNA Pharmacokinetics	1259
56.10.1	Enhancing Cellular Delivery by siRNA Conjugation	1260
56.10.2	Altering Biodistribution by siRNA Conjugation	1261
56.11	Chemical Modification of siRNAs – State of the Art	1261
56.12	A Guide for <i>In vivo</i> Studies	1261
	References	1265

Appendix: UV Spectroscopy for the Quantitation of RNA 1279

Index 1283

Preface

Why a second edition of the Handbook of RNA Biochemistry about eight years after release of the first edition? We see several profound reasons, the most fundamental one being that new biological and biochemical questions induce new technological advances, which in turn drives our research capabilities and opens up insights into novel RNA functions and mechanistic principles. For example, a multitude of novel non-coding RNAs (ncRNAs) have been uncovered, which entails a need not only for bioinformatic tools to predict their structure and to search for homologs, but also for further developments in biochemical tools for their functional analysis.

In the last decade, research in RNA biology, and here most notably global approaches, experienced an incredible boom, largely driven by new genome-wide and high-throughput technology, as well as RNA bioinformatics. Therefore, high-throughput and deep-sequencing approaches are covered by new chapters in this second edition (Chapters 34, 36–40), and contributions present already in the first edition have been thoroughly updated (Chapters 33 and 35) to keep pace with the fast evolution of these powerful methods.

Although unmodified RNA contains only four different nucleotides, the prediction of RNA secondary and tertiary structures and RNA homology searches are inherently sophisticated tasks of pivotal importance for RNA researchers. Chapters 26–32 are dedicated to these demands. There is also a need for protocols that enable experimental scientists to competently utilize bioinformatic, preferably web-based, tools. This aspect has been taken into account throughout this second edition.

All the chapters already present in the first edition have been updated, which concerns practical details (such as on company names, providers of enzymes and materials, web addresses), methodological details, and protocol streamlining. For example, the topics of RNA ligation or photoaffinity crosslinking to probe RNA structure, each previously represented by two separate chapters, are now consolidated in single chapters (Chapters 3 and 11). In addition, “old,” but very informative, RNA techniques such as gel- or TLC-based approaches to identify modified nucleosides or temperature-gradient gel electrophoresis of RNA, currently out of fashion, have been described in even more detail (Chapters 9 and 21) in order to preserve this kind of more traditional knowledge, which may experience an unforeseen revival at some point in the future.

Methodology in the area of RNA interference is crucial for so many researchers in various fields and is not restricted to RNA specialists, but also essential for RNA-based biotechnology and application in molecular medicine. We have therefore included new chapters on vector-encoded siRNA or miRNA techniques (Chapter 55), miRNA analysis (Chapter 49), and the application of chemically modified siRNAs (Chapter 56).

In summary, we have expanded the number of chapters and protocols, all written by experts in their fields, included new methods and approaches, strengthened the ready-to-use-lab-protocol format, and eliminated some redundancies.

We wish all readers scientific success in the application of our protocols, as well as several *Eureka!* experiences when, or after, consulting this Handbook for experimental strategies to tackle their specific biological questions or problems in RNA research.

March 2013

*Roland K. Hartmann
Albrecht Bindereif
Astrid Schön
Eric Westhof*

List of Contributors

Andrew J. Andrews

Fox Chase Cancer Center
333 Cottman ave.
Philadelphia,
PA 19111-2497
USA

Minna-Liisa Änkö

Max Planck Institute of
Cell Biology and Genetics
Pfotenhauerstrasse 108
01307 Dresden
Germany

and

Walter and Eliza Hall Institute of
Medical Research
Chemical Biology Division
Parkville, Melbourne
Australia

Manuel Ascano

The Rockefeller University
Laboratory of RNA Molecular
Biology
Howard Hughes Medical
Institute
1230 York Avenue
New York, NY 10065
USA

Pascal Auffinger

Modélisations et Simulations des
Acides Nucléiques
UPR 9002
Institut de Biologie Moléculaire et
Cellulaire du CNRS
15, rue René Descartes
67084 Strasbourg Cedex
France

Nathan J. Baird

National Heart
Lung and Blood Institute
National Institutes of Health
50 South Dr.
Bethesda, MD 20892
USA

Thomas Basché

Johannes Gutenberg-Universität
Institute of Physical Chemistry
Duesbergweg 10-14
55099 Mainz
Germany

Benedikt M. Beckmann

European Molecular Biology
Laboratory (EMBL)
Meyerhofstrasse 1
69117 Heidelberg
Germany

Isabelle Behm-Ansmant

Nancy Université
Laboratoire ARN-RNP
Maturation-Structure-Fonction
Enzymologie Moléculaire et
Structurale (AREMS)
UMR 7214 CNRS-UL
Batiment Biopôle
9, avenue de la Forêt de Haye
54506 Vandoeuvre-les-Nancy
France

Hildburg Beier

Universität Würzburg
BEEgroup
Biozentrum
Am Hubland
97074 Würzburg
Germany

Michaela Beitzinger

Universität Regensburg
Lehrstuhl Biochemie I
Universitätsstrasse 31
93053 Regensburg
Germany

Christian Berens

Friedrich-Alexander-Universität
Erlangen-Nürnberg
Lehrstuhl für Mikrobiologie
Department Biologie
Staudtstr. 5
91058 Erlangen
Germany

Ben Berkhout

University of Amsterdam
Laboratory of Experimental
Virology
Department of Medical
Microbiology
Center for Infection and
Immunity Amsterdam (CINIMA)
Academic Medical Center
Meibergdreef 15, K3-113D
1105 AZ Amsterdam
The Netherlands

Albrecht Bindereif

Justus-Liebig-Universität Gießen
Fachbereich Biologie und Chemie
Institut für Biochemie
Heinrich-Buff-Ring 58
35392 Gießen
Germany

Jennifer L. Boots

University of Vienna
Department of Biochemistry and
Molecular Cell Biology
Max F. Perutz Laboratories
Doktor-Bohr-Gasse 9/5
1030 Vienna
Austria

Marc Boudvillain

CNRS
Centre de Biophysique
Moléculaire
rue Charles Sadron
45071 Orleans
France

Marie Bouvier

University of Würzburg
 Institute for Molecular Infection
 Biology
 RNA Biology Group
 Josef-Schneider-Strasse 2
 97080 Würzburg
 Germany

Jesper B. Bramsen

University of Aarhus
 Interdisciplinary Nanoscience
 Center (iNANO)
 Ny Munkegade 118
 8000 Aarhus C
 Denmark

and

University of Aarhus
 Department of Molecular Biology
 and Genetics
 C. F. Møllers Allé 3
 8000 Aarhus C
 Denmark

Christiane Branlant

Nancy Université
 Laboratoire ARN-RNP
 Maturation-Structure-Fonction
 Enzymologie Moléculaire et
 Structurale (AREMS)
 UMR 7214 CNRS-UL
 Batiment Biopôle
 9, avenue de la Forêt de Haye
 54506 Vandoeuvre-les-Nancy
 France

Lukas Burger

Biozentrum der Universität Basel
 and Swiss Institute of
 Bioinformatics (SIB)
 Klingelbergstr. 50–70
 4056 Basel
 Switzerland

Malte Bussiek

University of Kassel
 Abt. Genetik and CINSaT
 Heinrich-Plett-Str. 40
 34132 Kassel
 Germany

John C. Castle

Johannes Gutenberg Medical
 University of Mainz
 TRON –Translational Oncology
 Saarstr. 21
 55122 Mainz
 Germany

Pavol Cekan

The Rockefeller University
 Laboratory of RNA Molecular
 Biology
 Howard Hughes Medical
 Institute
 1230 York Avenue, Box 186
 New York, NY 10065
 USA

Clément Chevalier

Université de Strasbourg
 Architecture et Réactivité de
 l'ARN
 UPR 9002 CNRS
 IBMC
 15, rue René Descartes
 67084 Strasbourg
 France

Jerzy Ciesiolka

Polish Academy of Sciences
 Institute of Bioorganic Chemistry
 Laboratory of RNA Biochemistry
 Noskowskiego 12/14
 61-704 Poznan
 Poland

Thomas A. Cooper

Baylor College of Medicine
Department of Pathology and
Immunology
One Baylor Plaza
Houston, TX 77030
USA

Simona Cuzic-Feltens

Martin-Luther Universität Halle
Naturwissenschaftliche Fakultät I
– Biowissenschaften
Institut für Biochemie und
Biotechnologie
Kurt-Mothes Str.3
06120 Halle (Saale)
Germany

Andrey Damianov

Justus-Liebig-Universität Giessen
Institut für Biochemie
Heinrich-Buff-Ring 58
35392 Giessen
Germany

and

University of California
at Los Angeles
Howard Hughes Medical
Institute
Los Angeles, CA 90095
USA

Julia Dannenberg

Max Planck Institute of
Biophysical Chemistry
Department of Cellular
Biochemistry
Am Fassberg 11
37077 Göttingen
Germany

François Disdier

Université de Strasbourg
Equipe de Biophysique et
Biologie Structurale
Unité 'Architecture et Réactivité
de l'ARN'
Institut de Biologie Moléculaire et
Cellulaire du CNRS
15, rue René Descartes
67084 Strasbourg
France

Olwen Domingo

Johannes Gutenberg-Universität
Institute of Pharmacy and
Biochemistry
Staudinger Weg 5
55128 Mainz
Germany

Philippe Dumas

Université de Strasbourg
Equipe de Biophysique et
Biologie Structurale
Unité 'Architecture et Réactivité
de l'ARN'
Institut de Biologie Moléculaire et
Cellulaire du CNRS
15, rue René Descartes
67084 Strasbourg
France

Olexandr Dybkov

Max-Planck-Institute for
Biophysical Chemistry
Department of Cellular
Biochemistry
Am Fassberg 11
37077 Göttingen
Germany

Laura E. Easton

MRC Laboratory of Molecular
Biology
Structural Studies Division
Hills Road
Cambridge CB2 0QH
UK

Thomas E. Edwards

University of Iceland
Department of Chemistry
Science Institute
Dunhaga 3
107 Reykjavik
Iceland

David R. Engelke

University of Michigan
Department of Biological
Chemistry
1150 W. Medical Center Drive
Ann Arbor, MI 48109-0600
USA

Markus Englert

Yale University
Department of Molecular
Biophysics and Biochemistry
Howard Hughes Medical
Institute
New Haven, CT 06520-8114
USA

Eric Ennifar

Université de Strasbourg
Equipe de Biophysique et
Biologie Structurale
Unité 'Architecture et Réactivité
de l'ARN'
Institut de Biologie Moléculaire et
Cellulaire du CNRS
15, rue René Descartes
67084 Strasbourg
France

Dirk Eulberg

Neue Welt 14
10247 Berlin
Germany

and

NOXXON Pharma AG
Max-Dohrn-Str. 8-10
10589 Berlin
Germany

Patrizia Fabrizio

Max Planck Institute of
Biophysical Chemistry
Department of Cellular
Biochemistry
37077 Göttingen
Germany

Thalia A. Farazi

The Rockefeller University
Laboratory of RNA Molecular
Biology
Howard Hughes Medical
Institute
1230 York Avenue
New York, NY 10065
USA

Olga Fedorova

Yale University and Howard
Hughes Medical Institute
Department of Molecular Cellular
and Developmental Biology
266 Whitney Avenue
New Haven, CT 06520
USA

Carol A. Fierke

University of Michigan
Department of Chemistry
930 North University
Ann Arbor, MI 48189-1055
USA

Mikko J. Frilander

University of Helsinki
Institute of Biotechnology
PL56 (Viikinkaari 9)
00014 Helsinki
Finland

Daniel Gautheret

Université Paris-Sud
CNRS-UMR8621
Institut de Génétique et
Microbiologie
Bâtiment 400
91405 Orsay Cedex
France

Robert Giegerich

Bielefeld University
Faculty of Technology and Center
of Biotechnology
Universitätsstrasse
33501 Bielefeld
Germany

Olaf Gimpl

Bayerische
Julius-Maximilians-Universität
Institut für Biochemie
Biozentrum
Am Hubland
97074 Würzburg
Germany

Markus Gößringer

Philipps-Universität Marburg
Institut für Pharmazeutische
Chemie
Marbacher Weg 6
35037 Marburg
Germany

Arnold Grünweller

Philipps-Universität Marburg
Institute of Pharmaceutical
Chemistry
Marbacher Weg 6
35037 Marburg
Germany

Inna Grishina

Justus-Liebig-Universität Giessen
Institut für Biochemie
Heinrich-Buff-Ring 58
35392 Giessen
Germany

and

Justus-Liebig-Universität Giessen
Biochemisches Institut
Friedrichstrasse 24
35392 Giessen
Germany

Markus Hafner

The Rockefeller University
Howard Hughes Medical
Institute
Laboratory of RNA Molecular
Biology
1230 York Avenue
New York, NY 10065
USA

Dinari A. Harris

University of Michigan
Department of Chemistry
930 N. University
Ann Arbor, MI 48109-1055
USA

Michael E. Harris

Case Western Reserve University
School of Medicine
Department of Biochemistry
10900 Euclid Avenue
Cleveland, OH 44106-4973
USA

Roland K. Hartmann

Philipps-Universität Marburg
Fachbereich Pharmazie
Institut für Pharmazeutische
Chemie
Marbacher Weg 6
Building C
35037 Marburg
Germany

Jean Hausser

University of Basel
Biozentrum
Klingelbergstrasse 50-70
4056 Basel
Switzerland

Corina G. Heidrich

Friedrich-Alexander-Universität
Erlangen-Nürnberg
Lehrstuhl für Mikrobiologie
Department Biologie
Staudtstr. 5
91058 Erlangen
Germany

Anne-Catherine Helfer

Université de Strasbourg
Architecture et Réactivité de
l'ARN
UPR 9002 CNRS
IBMC
15, rue René Descartes
67084 Strasbourg
France

Mark Helm

Johannes Gutenberg-Universität
Mainz
Institut für Pharmazie und
Biochemie
Staudinger Weg 5
55128 Mainz
Germany

Dominik Helmecke

Philipps-Universität Marburg
Institut für Pharmazeutische
Chemie
Marbacher Weg 6
35037 Marburg
Germany

Martin Hengesbach

University of California
Santa Cruz
Department of Chemistry and
Biochemistry
1156 High St
Santa Cruz, CA 95060
USA

Niklas Henriksson

Uppsala University
Department of Cell and
Molecular Biology
BMC
Husargatan 3
751 24 Uppsala
Sweden

Wolfgang R. Hess

University of Freiburg
Faculty of Biology
Genetics & Experimental
Bioinformatics
Institute of Biology III
Schänzlestr. 1
79104 Freiburg
Germany

Gerald Hinze

Johannes Gutenberg-Universität
Institute of Physical Chemistry
Duesbergweg 10-14
55099 Mainz
Germany

Jessica I. Hoell

The Rockefeller University
Laboratory of RNA Molecular
Biology
Howard Hughes Medical
Institute
1230 York Avenue
New York, NY 10065
USA

Alexander Hüttenhofer

Innsbruck Medical University
Section for Genomics and
RNomics
Innsbruck Biocenter
Fritz Pregl Strasse 3
6020 Innsbruck
Austria

Jingyi Hui

Justus-Liebig-Universität Giessen
Institut für Biochemie
Heinrich-Buff-Ring 58
35392 Giessen
Germany

and

Institute of Biochemistry and
Cell Biology
Chinese Academy of Sciences
200031 Shanghai
China

Martin R. Jakobsen

Aarhus University
Department of Molecular Biology
and Genetics
C.F.Møllers Alle
8000 Aarhus C
Denmark

Fabrice Jossinet

Université de Strasbourg
Architecture et Réactivité de
l'ARN
Institut de Biologie Moléculaire et
Cellulaire du CNRS
67084 Strasbourg
France

Mohsen Khorshid

Biozentrum der Universität Basel
and Swiss Institute of
Bioinformatics (SIB)
Klingelbergstr. 50–70
4056 Basel
Switzerland

Leif A. Kirsebom

Uppsala University
Department of Cell and
Molecular Biology
Biomedical Center
Box 596
751 24 Uppsala
Sweden

Jørgen Kjems

University of Aarhus
Interdisciplinary Nanoscience
Center (iNANO)
Ny Munkegade 118
8000 Aarhus C
Denmark

and

University of Aarhus
Department of Molecular Biology
and Genetics
C. F. Møllers Allé 3
8000 Aarhus C
Denmark

Sven Klussmann

Thrasoltstr. 1
10585 Berlin
Germany

and

NOXXON Pharma AG
Max-Dohrn-Str. 8-10
10589 Berlin
Germany

Karen Köhler

Philipps-Universität Marburg
Institut für Pharmazeutische
Chemie
Marbacher Weg 6
35037 Marburg
Germany

Julian König

MRC Laboratory of Molecular
Biology
Department of Structural Studies
Hills Road
Cambridge CB2 0QH
UK

Jiro Kondo

Sophia University
Department of Materials and
Life Sciences
Faculty of Science and
Technology
7-1 Kioi-cho, Chiyoda-ku
102-8554, Tokyo
Japan

Markus Landthaler

Max-Delbrück-Center for
Molecular Medicine
Berlin Institute for Medical
Systems Biology
Robert-Rössle-Str. 10
13125 Berlin
Germany

Stephen Leong Koan

Université de Montréal
Institute for Research in
Immunology and Cancer (IRIC)
Department of Computer Science
Montréal QC H3C 3J7
Canada

Efthimia Lioliou

Université de Strasbourg
Architecture et Réactivité de
l'ARN
UPR 9002 CNRS
IBMC
15, rue René Descartes
67084 Strasbourg
France

Reinhard Lührmann

Department of Cellular
Biochemistry
Max-Planck-Institute for
Biophysical Chemistry
Am Fassberg 11
37077 Göttingen
Germany

Martin Lützelberger

Technical University of
Braunschweig
Institute of Genetics
Spielmannstr. 7
38 106 Braunschweig
Germany

Peter J. Lukavsky

MRC Laboratory of
Molecular Biology
Structural Studies Division
Hills Road
Cambridge CB2 0QH
UK

and

Masaryk University
CEITEC - Central European
Institute of Technology
Kamenice 5/A4/2.33
62500 Brno
Czech Republic

Christian Maasch

Ernststr. 27
13509 Berlin
Germany

and

NOXXON Pharma AG
Max-Dohrn-Str. 8-10
10589 Berlin
Germany

François Major

Université de Montréal
Institute for Research in
Immunology and Cancer (IRIC)
Department of Computer Science
Montréal QC H3C 3J7
Canada

Virginie Marchand

Nancy Université
Laboratoire ARN-RNP
Maturation-Structure-Fonction
Enzymologie Moléculaire et
Structurale (AREMS)
UMR 7214 CNRS-UL
Batiment Biopôle
9, avenue de la Forêt de Haye
54506 Vandoeuvre-les-Nancy
France

and

European Molecular Biology
Laboratory (EMBL)
Meyerhofstrasse 1
69117 Heidelberg
Germany

Anita Marchfelder

Universität Ulm
Biologie II
Albert-Einstein-Allee 11
89069 Ulm
Germany

Manja Marz

Friedrich-Schiller-University Jena
Faculty of Mathematics and
Computer Science
Leutragraben 1
07743 Jena
Germany

Stefano Marzi

Université de Strasbourg
Architecture et Réactivité de
l'ARN
UPR 9002 CNRS
IBMC
15, rue René Descartes
67084 Strasbourg
France

Benoît Masquida

Chargé de Recherche CNRS
GMGM
UMR 7156 GMGM, IPCB
21, rue René Descartes
67084 Strasbourg
France

Katarzyna Matylla-Kulinska

University of Vienna
Department of Biochemistry and
Molecular Cell Biology
Max F. Perutz Laboratories
Doktor-Bohr-Gasse 9/5
1030 Vienna
Austria

Agnès Mèreau

Nancy Université
Laboratoire ARN-RNP
Maturation-Structure-Fonction
Enzymologie Moléculaire et
Structurale (AREMS)
UMR 7214 CNRS-UL
Batiment Biopôle
9, avenue de la Forêt de Haye
54506 Vandoeuvre-les-Nancy
France

and

Université de Lorraine
UMR CNRS AREMS team
ARN-RNP
Structure-Function-Maturation
Biople, Faculté de médecine
54505 Vandoeuvre-les-Nancy
France

Nicholas J. McGlincy

MRC Laboratory of
Molecular Biology
Department of Structural Studies
Hills Road
Cambridge CB2 0QH
UK

Jan Medenbach

Justus-Liebig-Universität Giessen
Institut für Biochemie
Heinrich-Buff-Ring 58
35392 Giessen
Germany

and

University of Regensburg
Universitätsstrasse 31
93053 Regensburg
Germany

Gunter Meister

Universität Regensburg
Lehrstuhl Biochemie I
Universitätsstrasse 31
93053 Regensburg
Germany

and

Max Planck Institute of
Biochemistry
Center for Integrated Protein
Science Munich
Am Klopferspitz 18
82152 Martinsried
Germany

Aleksandra Mihailovic

The Rockefeller University
Howard Hughes Medical
Institute
Laboratory of RNA Molecular
Biology
1230 York Avenue
New York, NY 10065
USA

Mario Mörl

Universität Leipzig
Fakultät für Biowissenschaften
Pharmazie und Psychologie
Institut für Biochemie
Brüderstraße 34
04103 Leipzig
Germany

Yuri Motorin

Nancy Université
Laboratoire ARN-RNP
Maturation-Structure-Fonction
Enzymologie Moléculaire et
Structurale (AREMS)
UMR 7214 CNRS-UL
Batiment Biopôle
9, avenue de la Forêt de Haye
54506 Vandoeuvre-les-Nancy
France

Annie Mougin

Nancy Université
Laboratoire ARN-RNP
Maturation-Structure-Fonction
Enzymologie Moléculaire et
Structurale (AREMS)
UMR 7214 CNRS-UL
Batiment Biopôle
9, avenue de la Forêt de Haye
54506 Vandoeuvre-les-Nancy
France

and

Université Paul Sabatier
Laboratoire de Biologie
Moléculaire Eucaryote
UMR 5099 CNRS-UPS
118, route de Narbonne
31062 Toulouse cedex 4
France

Sabine Müller

Ernst-Moritz-Arndt-Universität
Greifswald
Mathematisch-
Naturwissenschaftliche Fakultät
Institut für Biochemie
Felix-Hausdorff-Str. 4
17487 Greifswald
Germany

Wolfgang Nellen

University of Kassel
Abt. Genetik and CINSaT
Heinrich-Plett-Str. 40
34132 Kassel
Germany

Karla M. Neugebauer

Max Planck Institute of Cell
Biology and Genetics
Pfotenhauerstrasse 108
01307 Dresden
Germany

Jeff Nusbaum

The Rockefeller University
Laboratory of RNA Molecular
Biology
Howard Hughes Medical
Institute
1230 York Avenue
New York, NY 10065
USA

Zsöfia Palfi

Justus-Liebig-Universität Giessen
Institut für Biochemie
Heinrich-Buff-Ring 58
35392 Giessen
Germany

and

Medical University of Vienna
Department of Neuronal Cell
Biology
Spitalgasse 4
1090 Wien
Austria

Marc Parisien

University of Chicago
Biochemistry Department
Chicago, IL 60637
USA

Liudmila V. Pavlova

Philipps-Universität Marburg
Faculty of Pharmacy
Marbacher Weg 6
35037 Marburg
Germany

John Pena

The Rockefeller University
Howard Hughes Medical
Institute
Laboratory of RNA Molecular
Biology
1230 York Avenue
New York, NY 10065
USA

Franziska Peter

University of Leipzig
Institute of Biochemistry
Institute of Biochemistry
Brüderstrasse 34
04103 Leipzig
Germany

Dimitri Podkaminski

University of Würzburg
Institute for Molecular Infection
Biology
RNA Biology Group
Josef-Schneider-Strasse 2
97080 Würzburg
Germany

Ying Poi Liu

University of Amsterdam
Laboratory of Experimental
Virology
Department of Medical
Microbiology
Center for Infection and
Immunity Amsterdam (CINIMA)
Academic Medical Center
Meibergdreef 15, K3-113D
1105 AZ Amsterdam
The Netherlands

Marco Preußner

Justus-Liebig-Universität Giessen
Institut für Biochemie
Heinrich-Buff-Ring 58
35392 Giessen
Germany

and

Institute of Molecular Biology and
Tumor Research
Philipps-Universität Marburg
Emil-Mannkopff-Strasse 2
35032 Marburg
Germany

Werner G. Purschke

Wriezener Str. 30
13359 Berlin
Germany

and

NOXXON Pharma AG
Max-Dohrn-Str. 8-10
10589 Berlin
Germany

Mathieu Rederstorff

Université de Lorraine
Biopôle
UMR 7365 IMOPA
9, avenue de la Forêt de Haye
54506 Vandoeuvre-lès-Nancy
France

Yan-Guo Ren

Uppsala University
Department of Cell and
Molecular Biology
BMC
Husargatan 3
751 24 Uppsala
Sweden

Neil Renwick

The Rockefeller University
Howard Hughes Medical
Institute
Laboratory of RNA Molecular
Biology
1230 York Avenue
New York, NY 10065
USA

Detlev Riesner

Heinrich-Heine-Universität
Düsseldorf
Department of Biology
Institut für Physikalische Biologie
Universitätsstraße 1
40225 Düsseldorf
Germany

Sylvia Rösch

Universität Ulm
Biologie II
Albert-Einstein-Allee 11
89069 Ulm
Germany

Tanja Rösel

Justus-Liebig-Universität Giessen
Institut für Biochemie
Heinrich-Buff-Ring 58
35392 Giessen
Germany

Pascale Romby

Université de Strasbourg
Architecture et Réactivité de
l'ARN
UPR 9002 CNRS
IBMC
15, rue René Descartes
67084 Strasbourg
France

Cédric Romilly

Université de Strasbourg
Architecture et Réactivité de
l'ARN
UPR 9002 CNRS
IBMC
15, rue René Descartes
67084 Strasbourg
France

Jonathan Roy

Université de Montréal
Institute for Research in
Immunology and Cancer (IRIC)
Department of Computer Science
Montréal QC H3C 3J7
Canada

Claude Sauter

Université de Strasbourg
Architecture et Réactivité de
l'ARN
Institut de biologie moléculaire et
cellulaire du CNRS
15, rue René Descartes
67084 Strasbourg
France

Steffen Schiffer

Universität Ulm
Biologie II
Albert-Einstein-Allee 11
89069 Ulm
Germany

Astrid Schön

Universität Leipzig
Molekulare Zelltherapie
Biotechnologisch-
Biomedizinisches Zentrum
Deutscher Platz 5
04105 Leipzig
Germany

Antonie Schöne

University of Kassel
Abt. Genetik und CINSaT
Heinrich-Plett-Str. 40
34132 Kassel
Germany

Silke Schreiner

Justus-Liebig-Universität Giessen
Institut für Biochemie
Heinrich-Buff-Ring 58
35392 Giessen
Germany

Renée Schroeder

University of Vienna
Department of Biochemistry and
Molecular Cell Biology
Max F. Perutz Laboratories
Doktor-Bohr-Gasse 9/5
1030 Vienna
Austria

Felicia Scott

Macomb Community College
45575 Garfield Road
Clinton Township, MI 48038
USA

Yoko Shibata

MRC Laboratory of
Molecular Biology
Structural Studies Division
Hills Road
Cambridge CB2 0QH
UK

and

MedImmune
Milstein Building
Granata Park
Cambridge CB21 6GH
UK

Snorri Th. Sigurdsson

University of Iceland
Department of Chemistry
Science Institute
Dunhaga 3
107 Reykjavik
Iceland

Tobin R. Sosnick

The University of Chicago
Department of Biochemistry and
Molecular Biology
Institute for Biophysical
Dynamics
920 East 58th Street
Chicago, IL 60637
USA

Bettina Späth

Universität Ulm
Biologie II
Albert-Einstein-Allee 11
89069 Ulm
Germany

Felix Spenkuch

Johannes Gutenberg-Universität
Institute of Pharmacy and
Biochemistry
Staudinger Weg 5
55128 Mainz
Germany

Brian S. Sproat

Chemconsilium GCV
Jaarmarktstraat 48
2221 Heist-op-den-Berg
Belgium

Peter F. Stadler

University of Leipzig
Computer Science
Bioinformatics Group
Haertelstrasse 16-18
04107 Leipzig
Germany

Gerhard Steger

Heinrich-Heine-Universität
Düsseldorf
Department of Biology
Institut für Physikalische Biologie
Geb. 26.12.
Universitätsstraße 1
40225 Düsseldorf
Germany

Gabrielle C. Todd

University of Michigan
Department of Chemistry
930 N. University
Ann Arbor, MI 48109-1055
USA

Janne J. Turunen

University of Helsinki
Institute of Biotechnology
PL56 (Viikinkaari 9)
00014 Helsinki
Finland

and

Karolinska Institutet
Department of Laboratory
Medicine Hälsovägen 7
SE 14157 Huddinge
Sweden

Thomas Tuschl

The Rockefeller University
Howard Hughes Medical
Institute
Laboratory of RNA Molecular
Biology
1230 York Avenue
New York, NY 10065
USA

Jernej Ule

MRC Laboratory of Molecular
Biology
Department of Structural Studies
Hills Road
Cambridge CB2 0QH
UK

Henning Urlaub

Max-Planck-Institute for
Biophysical Chemistry
Bioanalytical Mass Spectrometry
Group
Am Fassberg 11
37077 Göttingen
Germany

and

University Medical Center
Göttingen Bioanalytics
Department of Clinical Chemistry
Robert-Koch-Strasse 40
37075 Göttingen
Germany

Anders Virtanen

Uppsala University
Department of Cell and
Molecular Biology
BMC
Husargatan 3
751 24 Uppsala
Sweden

C. Stefan Voertler

University of Leipzig
Institute of Biochemistry
Institute of Biochemistry
Brüderstrasse 34
04103 Leipzig
Germany

Jörg Vogel

University of Würzburg
Institute for Molecular Infection
Biology
RNA Biology Group
Josef-Schneider-Strasse 2
97080 Würzburg
Germany

Björn Voß

University of Freiburg
Faculty of Biology
Genetics & Experimental
Bioinformatics
Institute of Biology III
Schänzlestr. 1
79104 Freiburg
Germany

Christina Waldsich

University of Vienna
Department of Biochemistry and
Cell Biology
Max F. Perutz Laboratories
Dr. Bohrgasse 9/5
1030 Vienna
Austria

Nils G. Walter

University of Michigan
Department of Chemistry
930 N. University
Ann Arbor, MI 48109-1055
USA

Philippe Walter

Université de Strasbourg
Equipe de Biophysique et
Biologie Structurale
Unité 'Architecture et Réactivité
de l'ARN'
Institut de Biologie Moléculaire et
Cellulaire du CNRS
15, rue René Descartes
67084 Strasbourg
France

Greg Wardle

The Rockefeller University
Laboratory of RNA Molecular
Biology
Howard Hughes Medical
Institute
1230 York Avenue
New York, NY 10065
USA

Stefanie Wehner

Friedrich-Schiller-University Jena
Faculty of Mathematics and
Computer Science
Leutragraben 1
07743 Jena
Germany

Jeremey West

The University of Chicago
Department of Chemistry
920 East 58th Street
Chicago, IL 60637
USA

Eric Westhof

Université de Strasbourg
Architecture et Réactivité de
l'ARN
Institut de Biologie Moléculaire et
Cellulaire du CNRS
67084 Strasbourg
France

Cindy L. Will

Max Planck Institute of
Biophysical Chemistry
Department of Cellular
Biochemistry
37077 Göttingen
Germany

Dagmar K. Willkomm

Universitätsklinikum
Schleswig-Holstein
Campus Lübeck
Institut für Medizinische
Mikrobiologie und Hygiene
Ratzeburger Allee 160
23538 Lübeck
Germany

Mihaela Zavolan

Biozentrum der Universität Basel
and Swiss Institute of
Bioinformatics (SIB)
Klingelbergstr. 50-70
4056 Basel
Switzerland

Bob Zimmermann

University of Vienna
Department of Biochemistry and
Molecular Cell Biology
Max F. Perutz Laboratories
Doktor-Bohr-Gasse 9/5
1030 Vienna
Austria

Marek Zywicki

Medical University Innsbruck
Division of Genomics and
RNomics
Innsbruck Biocenter
Fritz-Pregl-Str. 3
6020 Innsbruck
Austria

Part I

RNA Synthesis and Detection

1

Enzymatic RNA Synthesis Using Bacteriophage T7 RNA Polymerase

Markus Gößringer, Dominik Helmecke, Karen Köhler, Astrid Schön,
Leif A. Kirseborn, Albrecht Bindereif, and Roland K. Hartmann

1.1

Introduction

Bacteriophage T7 RNA polymerase (T7 RNAP) was first cloned and overexpressed from bacteriophage T7-infected *Escherichia coli* cells in the early 1980s [1]. In contrast to multisubunit DNA-dependent RNAPs from eukaryotes and prokaryotes, T7 RNAP consists of a single subunit of about 100 kDa [2]. The subdomains adopt a hand-like shape with palm, thumb, and fingers around a central cleft where the active site containing the functionally essential amino acid residues is located, creating a binding cavity for magnesium ions and ribonucleotide substrates. For RNA synthesis, the unwound template strand is positioned such that the template base -1 is anchored in a hydrophobic pocket in direct vicinity of the active site [3].

T7 RNAP is highly specific to its own promoters and exhibits no affinity even to closely related phage T3 promoters, although the 23 bp consensus sequences are very similar (Figure 1.1a). During the initiation process, the polymerase goes through several elongation attempts, generating short abortive oligoribonucleotides. Only when the nascent RNA transcript exceeds 9–12 nt do initiation complexes convert to stable elongation complexes. Transcription proceeds with an average rate of $200\text{--}260\text{ nt s}^{-1}$ until the elongation complex encounters a termination signal or falls off the template end during *in vitro* run-off transcription [4, 5]. The error frequency in transcripts of wild-type (wt) T7 RNAP is about 6×10^{-5} [6].

In the following sections, we describe protocols that have been used routinely for T7 transcriptions. Further, a robust and simple protocol for the partial purification of T7 RNAP is included, which yields an enzyme preparation that fully satisfies all *in vitro* transcription demands. The given transcription protocols suffice for most purposes. However, in special cases, such as the synthesis of milligram quantities, modified RNAs, or very A + U-rich RNAs, it may be worthwhile to further optimize the transcription conditions. We would also like to draw the reader's attention to the paper by Milligan and Uhlenbeck [10], which briefly discusses many fundamental aspects of T7 transcription.



1.2

(c) T7 class III promoter region with the recommended G identities at positions +1 and +2 of the RNA transcript shown in gray.

Description of Method – T7 Transcription *In vitro*

influence transcription yields significantly (Figure 1.1b,c; [10]).

Templates

template DNA oligonucleotide.

1.2.1.1 Strategy (i): Insertion into a Plasmid

sequence of a plasmid clone has been confirmed, the DNA can be conveniently

amplified by *in vivo* plasmid replication exploiting the high fidelity of bacterial DNA polymerases. The RNA expression cassette (either with or without the T7 promoter sequence) is usually obtained by PCR and cloned into a bacterial plasmid. Since PCR amplification is error-prone, plasmid inserts ought to be sequenced. When the T7 RNAP promoter region from –17 to –1 is not encoded in the PCR fragment, one can use commercially available T7 transcription vectors (e.g., pGEM[®] 3Z and derivatives from Promega, or the pPCR-Script series from Agilent Technologies/Stratagene) containing the T7 promoter and a multiple cloning site for insertion of the RNA expression cassette. If there are no sequence constraints at the transcript 5' end, we routinely design templates encoding 5'–GGA at positions +1 to +3 of the RNA transcript, which usually results in high transcription yields. Whenever possible, at least the nucleotide preferences at positions +1 and +2 should be taken into account (Figure 1.1b,c). Directly downstream of the expression cassette, a restriction site is required for template linearization to terminate RNA synthesis (run-off transcription); restriction enzymes producing 5' overhangs are preferred over those producing blunt ends or 3' overhangs [10]. Beyond common type II restriction enzymes generating 5' overhangs (e.g., *Bam* HI, *Eco* RI), “type IIS” enzymes (e.g., *Fok*I) are of interest because they cleave sequence-independently outside of their recognition sequence and thus permit to design RNA transcript 3' ends of complete identity to natural counterparts. Individual steps of template preparation are (i) ligation of (PCR) insert into plasmid, (ii) cloning in *E. coli*, purification and sequencing of plasmid, (iii) linearization of plasmid DNA for run-off transcription, (iv) phenol/chloroform extraction and ethanol precipitation of template DNA before (v) use in T7 transcription assays.

1.2.1.2 Strategy (ii): Direct Use of Templates Generated by PCR

Direct use of PCR fragments as templates is faster than insertion into a plasmid and preferred if only minor amounts of RNA are required. In this case, the T7 promoter sequence is encoded by the 5' primer used in the PCR reaction. A downstream restriction site producing 5' overhangs may be conveniently included in the 3' primer.

1.2.1.3 Strategy (iii): Annealing of a T7 Promoter DNA Oligonucleotide to a Single-Stranded Template

This strategy is the fastest, and we have used it to synthesize small amounts of an RNA 30-mer for 5'–labeling purposes (Protocol 4). Here, the shorter T7 promoter DNA oligonucleotide is annealed to the complementary single-stranded DNA template oligonucleotide. The complementary double-stranded region is sufficient to initiate transcription by T7 RNAP.

1.2.2

Special Demands on the RNA Product

1.2.2.1 Homogeneous 5' and 3' Ends, Small RNAs, Functional Groups at the 5' End

While T7 RNAP usually initiates transcription at a defined position, it tends to append one or occasionally even a few more non-templated nucleotides to the

product 3' terminus [10, 12]. 5' End heterogeneity may become a problem when the template encodes unusual 5'-terminal sequences, such as 5'-CACUGU, 5'-CAGAGA, or 5'-GAAAAA [13], or when transcripts are initiated with multiple guanosines [14]. For example, in the case of transcripts starting with 5'-GGGGG, 75% had canonical 5' ends, relative to >99% for 5'-GCGGA, 87% for 5'-GGGCC, 97% for 5'-GGGAG, and only 66% for 5'-GGGGC [14]. Thus, it is recommended that more than two consecutive G residues at the 5' end be avoided. 5' End heterogeneity seems to be a problem associated with T7 class III promoters (Figure 1.1a) because almost complete 5' end homogeneity of T7 transcripts has been achieved with templates directing transcription from the more rarely used T7 ϕ 2.5 class II promoter (Figure 1.1a), at which T7 RNAP initiates synthesis with an A instead of a G residue. Transcription yields from this promoter were reported to equal those of the commonly used T7 class III promoter [15].

For the production of RNAs with 100% 5' and 3' end homogeneity, several methods are available. In one approach (Chapter 3), the downstream PCR primer introduces two 2'-OCH₃-modified RNA nucleotides at the 5'-terminal positions of the template strand, which suppresses the addition of 3'-terminal non-templated residues during transcription. Alternatively, hammerhead or hepatitis delta virus (HDV) ribozymes may be tethered to the RNA of interest on one or both sides (Chapter 2). The ribozyme(s) will release the RNA product by self-cleavage during transcription. Such a *cis*-acting ribozyme placed upstream releases the RNA of interest with a 5'-OH terminus directly accessible to 5' endlabeling (Chapter 9), and simultaneously eliminates the problem of 5' end heterogeneity as well as constraints on the identity of the 5' terminal nucleotide of the RNA of interest (Chapter 2). The same strategy may also be considered for the synthesis of large amounts of smaller RNAs. Chemical synthesis and purification of 10 mg of, for example, an RNA 15-mer by a commercial supplier can be quite expensive. In such a case, a cheaper alternative would be to transcribe the 15-mer sandwiched between two *cis*-cleaving ribozymes, resulting in posttranscriptional release of the 15-mer with uniform 5' and 3' ends. Purification of the 15-mer (and separation from the released ribozyme fragments) can then be achieved by preparative denaturing polyacrylamide gel electrophoresis (PAGE; Section 1.3.4). If T7 RNAP is self-prepared according to the protocol described in this chapter, synthesis of 10 mg of a 15-mer will become quite affordable.

Normally, transcription by T7 RNAP is initiated with GTP, resulting in 5'-triphosphate ends. If, however, 5'-OH ends or 5'-monophosphate termini are preferred, T7 RNAP can be prompted to initiate transcripts with guanosine or 5'-ApG (to generate 5'-OH ends for direct endlabeling with ³²P), or 5'-GMP (to generate 5'-monophosphates), when these components are added to reaction mixtures in excess of GTP [16]. RNA transcripts with 5'-GMP ends are preferred when the RNA is used for ligation with other RNA molecules.

1.2.2.2 Modified Substrates

There are a number of modified nucleoside-5'-triphosphates known to be substrates for T7 RNAP. Table 1.1 has been adopted from Milligan and Uhlenbeck [10] and expanded by addition of more recent information.

Table 1.1 Nucleotide analogs for internal or 5'-terminal incorporation into T7 transcripts.

NTP	wt T7 RNAP	Y639F T7 RNAP	Y639F/H784A T7 RNAP	References
NTP α S (Sp)	+			[17]
NTP α S (Rp)	–			[17]
5-Br-UTP	+			[10]
5-F-UTP	+			[10]
5-Hexamethyleneamino-UTP	+			[10]
6-Aza-UTP	+			[10]
4-Thio-UTP	+			[10]
Pseudo-UTP	+			[10]
8-Br-ATP	+			[10]
7-Me-GTP	–			[10]
ITP (with initiator) ^a	+			[18]
2'-dNTP	+/-, ^{-b}	+		[10, 19, 20]
2'-dNTP α S	+/-	+		[21–23]
2'-O-Me-NTP or -NTP α S	+/-		+	[21, 24]
2'-O-Me-NTP		+		[20]
2'-N ₃ -NTP	– ^c		+	[19, 24]
2'-F-(A,C,U)TP	(+), +/- ^d	+		[19, 20]
2'-Amino-UTP	(+), +/- ^d			[19]
2'-Amino-(A,C,U)TP		+		[20]
LNA-ATP	+			[24]
LNA-UTP	+			[25]
tCTP ^e	+			[26]
GTP γ S	+			[27]
5'-Biotin-GMP		+		[28]
6-Thio-GMP	+			[29]
GMPS ^f	+			[30]

+/-: Low incorporation efficiency.

(+: Reduced incorporation efficiency.

^aInosine triphosphate (ITP) cannot be used to start transcription, but can substitute for GTP during elongation if a primer, such as 5'-ApG or 5'-GMP, is present as initiator of transcription.^bTested for 2'-dCTP fully replacing CTP during transcription [19].^cTested for 2'-N₃-CTP fully replacing CTP during transcription [19].^dTranscription efficiency strongly dependent on the specific RNA transcript [19].^e1,3-Diaza-2-oxophenothiazine-ribose-5'-triphosphate (fluorescent ribonucleotide).^fGuanosine 5'-monophosphorothioate.

Owing to discrimination of NTPs and dNTPs by wt T7 RNAP, the polymerase incorporates NTPs 70- to 80-fold more efficiently than dNTPs in the presence of Mg²⁺ as the metal ion cofactor. However, a T7 RNAP mutant (Y639F) carrying a tyrosine to phenylalanine exchange at position 639 [31] was shown to have only about fourfold higher preference for NTPs than dNTPs [31, 32], and thus permits more efficient incorporation of substrates lacking the ribose 2'-hydroxyl, such as 2'-fluoro or 2'-amino nucleotides [32]. Incorporation of substrate analogs with 2'-ribose modifications can also be stimulated to some extent in reactions catalyzed by wt T7

RNAP upon addition of Mn^{2+} [21]. Similarly, dNTPαS analogs were partially incorporated into RNase P RNA in a sequence-specific manner under mixed metal ion conditions (Mg^{2+}/Mn^{2+} ; [23]). Despite these achievements, the Y639F mutant T7 RNAP is nowadays the enzyme of choice for the incorporation of all nucleotides with 2'-ribose modifications. For detailed protocols tailored to the specific nucleotide analog, the reader is referred to [20, 32, 33]. The Y639F T7 RNAP can, of course, also be used for the synthesis of RNA transcripts with standard nucleotides. This mutant T7 RNAP can be purchased from Epicentre Biotechnologies (Madison, WI, USA) under the product name T7 R&DNA™ Polymerase, and a strain overexpressing the Y639F enzyme can be obtained from the Sousa Laboratory (Section 1.5.1).

A double mutant of T7 RNAP (Y639F/H784A) was shown to increase the efficiency of full-length RNA synthesis using modified NTPs with bulky 2'-substituents, such as 2'-OCH₃ and 2'-N₃ ([34]; 2'-OCH₃– and 2'-N₃–modified 5'-NTPs are available from TriLink BioTechnologies, San Diego, CA, USA). Analog incorporation was efficient during the elongation phase of transcription, but still inefficient during the poorly processive initiation phase (<9 nt), particularly for the 2'-OCH₃ analog [34].

Further modifications can be introduced into transcripts by initiator (oligo)nucleotides. Di- to hexanucleotides with a 3' terminal guanine base, including di- to tetranucleotides with internal or terminal 2'-deoxy- or 2'-O-methylated residues, were tested as initiators of transcription by wt T7 RNAP [35]. 5'-Terminal incorporation varied between 20% (hexamer) and 80–95% in the case of 5'-ApG or a 5'-biotinylated ApG. ApG, 5'-biotin-ApG, 5'-fluorescein-ApG, and 5'-TAMRA-ApG are available from IBA, Göttingen, Germany (furthermore, a series of other ApG derivatives on request). For 5'-terminal labeling with biotin, 5'-biotin-GMP (TriLink BioTechnologies) can be added as starter nucleotide to transcription reactions in excess of GTP (e.g., 2 mM GTP, 3 mM 5'-biotin-GMP [28]). Similarly, 6-thio-GMP and guanosine 5'-monophosphorothioate (GMPS) were used as starter nucleotides to directly introduce a photoagent for site-specific crosslinking (6-thio-GMP [29]) or for derivatization with a photoagent (GMPS [30]) (Chapter 11). Also, transcription by T7 RNAP, in this case from the T7 φ2.5 class II promoter, was initiated with coenzymes containing an adenosine moiety, such as CoA (3'-dephospho-coenzyme A), NAD, or FAD. Reduced NADH and oxidized FAD are highly fluorescent, which opens up the perspective to employ coenzyme-linked RNAs for the study of RNA–RNA or RNA–protein interactions by fluorescence techniques [36].

1.3

Transcription Protocols

Generally, all solutions are prepared using double-distilled RNase-free water (dd H₂O) (exception: deionized H₂O to prepare and dilute 5 x TBE (Tris/Borate/EDTA) to 1 x TBE electrophoresis buffer) and analytical grade reagents. All stock solutions, except for electrophoresis buffers at their final dilution, are subjected to sterile filtration (Filtropur S 0.2; Sarstedt) and stored at –20 °C if not stated otherwise.

1.3.1

Transcription with Unmodified Nucleotides

The protocols given below have been applied to template DNAs directing transcription from the T7 class III promoter. Transcription yields can differ substantially, depending on the individual DNA template and the origin of T7 RNAP. In Protocols 1–5 (Hartmann Laboratory), T7 RNAP from Fermentas or home-made T7 RNAP (Section 1.5.2) has been used. These protocols may be suboptimal for some RNA transcripts and when using T7 RNAP from other sources. This has been accounted for by including protocols from the Kirsebom (Protocols 6 and 7) and Schön (Protocol 8) Laboratories. Independent of the type of protocol, it is advisable to put some effort into the optimization of the transcription protocol if large amounts of RNA are to be produced or if the transcript represents a “standard RNA” in the laboratory, used over longer periods and requiring repeated synthesis. In addition to commercially available T7 RNAP, protocols for the partial purification of T7 RNAP from bacterial overexpression strains have been reported [1, 37–39]. We have used and refined the protocol detailed in Section 1.5.

When a new DNA template is used for the first time, one approach is to perform small test transcriptions on a 50 μl scale according to a basic protocol (e.g., Protocol 1). Reaction mixtures should be prepared at room temperature, since DNA may precipitate in the presence of spermidine at low temperatures. In the case of plasmid DNA, template amounts of 40–80 $\mu\text{g ml}^{-1}$ (final assay concentration) are used as a rule of thumb, whereas a PCR template of 400 bp is adjusted to about 5 $\mu\text{g ml}^{-1}$ final assay concentration. In the Hartmann Laboratory, we usually incubate transcription mixes for 4–6 h at 37°C, although overnight incubations have been used as well. A variation is to add another aliquot (e.g., 2 U μl^{-1} of reaction mix) of T7 RNAP after 2 h at 37°C, followed by a further 2 h incubation period at 37°C. In transcriptions according to, for example, Protocol 6, a white precipitate may appear because of pyrophosphate accumulation. This is avoided, for example, in Protocol 1, where pyrophosphate is hydrolyzed because of the presence of pyrophosphatase. Extension of incubation periods beyond 4–6 h did not prove to be advantageous in our hands and may be associated with some product degradation since T7 RNAP has a DNase and RNase function, which is normally inhibited by NTP substrates added in excess to *in vitro* transcription assays. However, after extended incubation periods, the NTP concentration may drop below a critical limit, thereby favoring RNA degradation [40]. Protocol 2 represents an inexpensive strategy to incorporate a 5'-terminal guanosine. Since guanosine has a low solubility, a 30 mM solution is prepared and kept at 75°C; the reaction mixture – except for guanosine and T7 RNAP – is prepared at room temperature and prewarmed to 37°C before addition of guanosine and T7 RNAP.

We would like to note here that transcription from the T7 $\phi 2.5$ class II promoter is initiated with A instead of G (Figure 1.1a), opening up the perspective to incorporate an adenosine at the 5' end. Adenosine would fulfill the same purpose as the aforementioned guanosine used in the case of the class III promoter, but is advantageous because of its better water solubility [15, 36].

As an alternative to starting class III-promoter-directed transcripts with guanosine, the dinucleotide 5'-ApG (see above) may be employed. 5'-Terminal incorporation of the dinucleotide leads to a -1 adenosine extension of the transcript (Protocol 3). ApG, which is more convenient to use than guanosine, is available from TriLink BioTechnologies and IBA, but much more expensive than guanosine. Transcripts can further be initiated with 5'-GMP (Protocol 3) if present in excess of 5'-GTP (15 mM vs 3.75 mM). Although the majority of RNA products should possess a 5'-terminal monophosphate, a dephosphorylation/phosphorylation strategy (Chapter 3) may be preferred to obtain RNA products with 100% 5'-monophosphate ends.

Protocol 4 is a quick protocol for the synthesis of small amounts of shorter RNAs for 5'-endlabeling purposes. The promoter and template DNA oligonucleotides are simply added to the reaction mixture, which is then preincubated for 1 h at 37 °C before starting transcription by adding T7 RNAP.

Protocol 5 is tailored to the enzymatic synthesis of large amounts of RNA (e.g. for structural studies).

Protocol 6 (from the Kirsebom Laboratory) has the characteristics of high T7 RNAP concentrations and the presence of RNase inhibitor, suitable for large-scale transcriptions. Protocol 7 is used in the Kirsebom Laboratory for the production of internally labeled RNA.

Protocol 8 (from the Schön Laboratory) has been used for standard transcriptions as well as for the synthesis of extremely A+U-rich RNAs, with the ratio of individual NTPs adapted to their proportion in the final transcript.

Protocol 1: Standard Protocol (Hartmann Laboratory)

	Final concentration	1000 µl
HEPES pH 7.5, 1 M	80 mM	80 µl
DTT 100 mM	5 mM	50 µl
MgCl ₂ 3 M	20 mM	6.7 µl
Spermidine 100 mM	1 mM	10 µl
NTP mix (25 mM each)	4 mM (each)	160 µl
Template (linearized plasmid 3.2 kb) 1 µg µl ⁻¹	40 µg ml ⁻¹	40 µl
Pyrophosphatase ^a 200 U ml ⁻¹	1 U ml ⁻¹	5 µl
T7 RNAP 200 U µl ⁻¹	1000–2000 U ml ⁻¹	5–10 µl
RNase-free water	—	to 1000 µl

For small scale transcriptions (50 µl final volume), reaction mixes are incubated for 4–6 h at 37 °C. For preparative transcription, usually 1 ml reaction mixtures are prepared and then incubated in 200 µl aliquots (for better thermal equilibration) for 2 h at 37 °C; then a second aliquot of T7 RNAP is added (400 U per 200 µl reaction mix), followed by another 2 h of incubation at 37 °C. Efficient transcription reactions in the 1 ml scale result in a product yield of about 3 nmol.

^aPyrophosphatase from yeast (Roche Cat.No. 10108987001, EC 3.6.1.1, 200 U mg⁻¹, < 0.01% ATPase and phosphatases each).

Transcription yields depend on the particular template DNA and RNA transcript. If the yield and quality (fraction of full-length transcript) of the transcription is unsatisfactory, we vary the dithiothreitol (DTT) concentration between 5 and

15 mM, and that of Mg^{2+} between 10 and 40 mM (in 50 μl test reactions). Furthermore, the DNA template, pyrophosphatase, and T7 RNAP concentrations may be increased to 80 $\mu\text{g ml}^{-1}$, 2–5 U ml^{-1} , and 3000 U ml^{-1} , respectively. We further observed that addition of 3% dimethyl sulfoxide (DMSO) can increase transcription yields severalfold (test the range between 0.5 and 10% DMSO). Another additive worth testing is betaine (Chapter 51).

Protocol 2: Transcripts Initiated with Guanosine for 5' Endlabeling (Hartmann Lab)

	Final concentration	1000 μl
HEPES pH 7.5, 1 M	80 mM	80 μl
DTT 100 mM	5 mM ^a	50 μl
MgCl_2 3 M	20 mM ^a	6.7 μl
Spermidine 100 mM	1 mM	10 μl
NTP mix (25 mM each)	3.75 mM (each)	150 μl
Template (linearized plasmid 3.2 kb) 1 $\mu\text{g ml}^{-1}$	40 $\mu\text{g ml}^{-1}$	40 μl
Pyrophosphatase ^b 200 U ml^{-1}	1 U ml^{-1}	5 μl
RNase-free water	—	348.3 μl

^aMay require optimization, as described for Protocol 1.

^bSee Protocol 1.

- 1) Prewarm mixture to 37 °C.
- 2) Then add 300 μl 30 mM guanosine (kept at 75 °C) to a final concentration of 9 mM.
- 3) Finally add 10 μl T7 RNAP (200 U ml^{-1}) to a final concentration of 2000 U ml^{-1} .
- 4) For incubation, see Protocol 1.

Protocol 3: Initiation with 5'-GMP or 5'-ApG, to Provide Transcripts with 5'-Monophosphates (5'-GMP) or with 5'-Hydroxyls for 5'-Endlabeling Purposes (5'-ApG) (Hartmann Lab)

	Final concentration	1000 μl
HEPES pH 7.5, 1 M	80 mM	80 μl
DTT 100 mM	5 mM ^a	50 μl
MgCl_2 3 M	20 mM ^a	6.7 μl
Spermidine 100 mM	1 mM	10 μl
NTP mix (25 mM each)	3.75 mM (each) ^b	150 μl
5'-GMP 100 mM (initiator) ^b	15 mM	150 μl
Template (linearized plasmid 3.2 kb) 1 $\mu\text{g ml}^{-1}$	40 $\mu\text{g ml}^{-1}$	40 μl
Pyrophosphatase ^c 200 U ml^{-1}	5 U ml^{-1}	25 μl
T7 RNAP 200 U ml^{-1}	2000 U ml^{-1}	10 μl
RNase-free water	—	ad 1000 μl

^aMay require optimization, as described for Protocol 1.

^bWhen 5'-GMP is replaced with the dinucleotide 5'-ApG for transcription initiation, adjust 5'-ApG to 7.5 mM and NTPs to 2.5 mM each (final concentrations). For incubation, see Protocol 1.

^cSee Protocol 1.

Protocol 4: DNA Template Assembled from Oligonucleotides, Transcripts Initiated with Guanosine (Hartmann Lab)

	Final concentration	500 µl
HEPES pH 8.0, 1 M	160 mM	80 µl
DTT 100 mM	15 mM	75 µl
MgCl ₂ 3 M	33 mM	5.5 µl
Spermidine 100 mM	1 mM	5 µl
Promoter DNA oligonucleotide 3.3 µg µl ^{-1a}	132 µg ml ⁻¹	20 µl
Template DNA oligonucleotide 2.1 µg µl ^{-1b}	84 µg ml ⁻¹	20 µl
BSA (bovine serum albumin) 20 mg ml ⁻¹	0.12 mg ml ⁻¹	3 µl
RNase-free water	—	51.5 µl

^aPromoter DNA oligonucleotide: 5'-TAA TAC GAC TCA CTA TAG.
^bIn this example, the template DNA oligonucleotide had the sequence 5'-GGT CAT AGG TAT TCC CCC TCT CTC CAT TCC TAT AGT GAG TCG TAT TAA, resulting in an RNA product with the sequence 5'-GGA AUG GAG AGA GGG GGA AUA CCU AUG ACC; the design of DNA oligonucleotides was adapted from [12, 41].

- 1) Mix and preincubate for 1 h at 37 °C.
- 2) Add 75 µl NTP mix (25 mM each) to a final concentration of 3.75 mM (each)^c
- 3) Add 5 µl Pyrophosphatase (200 U ml⁻¹)^d to a final concentration of 2 U ml⁻¹.
- 4) Prewarm mixture to 37 °C.
- 5) Then add 150 µl 30 mM guanosine (kept at 75 °C) to a final concentration of 9 mM.
- 6) Add 10 µl T7 RNAP (200 U µl⁻¹) to a final concentration of 4000 U ml⁻¹.
- 7) Incubate at 37 °C for 4 h.

^cTo increase the percentage of transcripts initiated with guanosine, the ratio of guanosine to NTPs may be increased; for example, by reducing the NTP concentration to 1.5 mM each.
^dSee Protocol 1.

Protocol 5: T7 Synthesis of Large Amounts of RNA for Structural Studies (Hartmann Lab)

For such large-scale transcriptions, we first determine the optimal Mg²⁺ concentration (variation between 10 and 40 mM) in 50 µl test reactions.

	Final concentration	~15 ml
Tris-HCl pH 7.9, 1 M	200 mM	3 ml
DTT 1 M	20 mM	0.3 ml
MgCl ₂ 150 mM	15 mM	1.5 ml
Spermidine 100 mM	2 mM	0.3 ml
Template (linearized plasmid 3.2 kb) 200 µg ml ⁻¹	100 µg ml ⁻¹	7.5 ml
NTP mix (25 mM each)	4 mM (each)	2.4 ml
T7 RNAP (in-house preparation)	—	37.5 µl
RNase-free water	—	—

- 1) Incubate at 37 °C overnight.
- 2) *Optional:* add 0.5 M Na₂EDTA (pH 7.5 or 8.0) until all insoluble Mg²⁺-pyrophosphates (Section 1.3.4) have been dissolved.
- 3) Precipitate the RNA with ethanol (or isopropanol).

- 4) After ethanol precipitation, centrifuge, wash the RNA pellet with 70% ethanol, briefly centrifuge again, air-dry the pellet, dissolve it in 4–6 ml ddH₂O, and add an equal volume of denaturing loading buffer (for more details, see Section 1.3.4).
- 5) Load the sample (e.g., 12 ml) onto five preparative (e.g., 10%) polyacrylamide (PAA) 8 M urea gels (30 cm (width) × 20 cm (height) × 1 mm, (gel thickness), gel pocket 23 cm wide). For each gel, one may load 2 µg of a reference RNA into a small pocket (0.7 cm wide) next to the preparative sample, which helps to locate the RNA of interest during UV shadowing. Run the gels at 40 mA for 4–5 h in 1 x TBE at room temperature.
- 6) Localize the RNA of interest by UV shadowing (Chapter 3) and excise the corresponding gel strip. Combine the gel strips from all five gels, chop into smaller gel slices and transfer into a 50 ml Falcon tube containing 20 ml 0.3 M NaOAc pH 5; shake in an Orbital Shaker overnight at 4–10 °C, withdraw the solution using a syringe and filter over a sterile filter to get rid of residual gel pieces; precipitate with ethanol and redissolve in ~100 µl ddH₂O. Perform a second elution in the same manner as the first.

Using this protocol (without Na₂EDTA addition), we were able to purify 7.5 mg of the catalytic domain of *E. coli* RNase P RNA (~220 nt) involving two consecutive gel elutions.

Protocols 6 and 7: Kirsebom Laboratory

- 10 x Transcription buffer (TRX), for transcription Protocols 6 and 7:

	Final concentration	1000 µl
Tris-HCl pH 7.7, 1 M	400 mM	400 µl
MgCl ₂ 3 M	240 mM	80 µl
Spermidine 100 mM	20 mM	200 µl
RNase-free water	—	320 µl

Protocol 6: Nonradioactive Transcription, Volume Sufficient for 4 Reactions (Mix Prepared for 4.5 Reactions)

	Final concentration ^a	432 µl
10 x TRX	1 x	45 µl
DTT 0.5 M	10 mM	9 µl
0.2% Triton X100	0.01%	22.5 µl
ATP (100 mM)	2 mM	9 µl
GTP (100 mM)	2 mM	9 µl
CTP (100 mM)	2 mM	9 µl
UTP (100 mM)	2 mM	9 µl
RNase inhibitor 40 U µl ⁻¹ (Fermentas)	32 U ml ⁻¹	0.6 µl
T7 RNAP 200 U µl ⁻¹ (Ambion or prepared in-house)	10 000 U ml ⁻¹	22.5 µl
RNase-free water	—	296 µl

^aFinal concentrations after addition of template.

- 1) To 96 μl of this mix, add 4 μl template DNA (linearized plasmid ≈ 3.2 kb; 1 $\mu\text{g } \mu\text{l}^{-1}$; final concentration 40 $\mu\text{g ml}^{-1}$).
- 2) Incubate at 37 °C for ≤ 10 h.

Protocol 7: Internal Radioactive Labeling Mix, Volume Sufficient for 9 Reactions (Mix Prepared for 10 Reactions)

	Final concentration ^a	230 μl
10 x TRX	1 x	25 μl
DTT 0.5 M	10 mM	5 μl
0.2% Triton X100	0.01%	12.5 μl
ATP (100 mM)	2 mM	5 μl
GTP (100 mM)	2 mM	5 μl
CTP (100 mM)	2 mM	5 μl
UTP 1 mM	0.2 mM	50 μl
$[\alpha\text{-}^{32}\text{P}]\text{UTP}$ 800 Ci mmol^{-1} (20 mCi ml^{-1}) ^b	2 mCi ml^{-1}	25 μl
RNase inhibitor 24 U μl^{-1}	31.7 U ml^{-1}	0.33 μl
T7 RNAP 200 U μl^{-1}	10 000 U ml^{-1}	12.5 μl
RNase-free water	—	84.7 μl

^aFinal concentrations after addition of template.

^bOther ^{32}P -labeled nucleotides may be used as well, also in combinations and with different specific activities.

- 1) To 23 μl of this mix, add 2 μl template DNA (linearized plasmid ≈ 3.2 kb; 1 $\mu\text{g } \mu\text{l}^{-1}$; final concentration 80 $\mu\text{g ml}^{-1}$).
- 2) Incubate at 37 °C for ≤ 10 h.

Protocol 8: Preparative Transcription of RNAs with Biased Nucleotide Composition (Schön Lab)

To account for unbalanced nucleotide composition, such as in RNase P RNAs from the *Cyanophora paradoxa* cyanelle [42] or from a plant-pathogenic phytoplasma [43], the relative concentrations of NTPs are adjusted accordingly. For phytoplasma RNase P RNA (about 73% A+U), the composition of the nucleotide mix was calculated as follows:

	Calculated mol% of each nucleotide in the transcript	Concentration of each NTP in the nucleotide mix (mM)	Final concentration of each NTP in the reaction mix (mM)
ATP	41.08	33	3.3
CTP	11.06	9	0.9
GTP	16.03	12.5	1.25
UTP	31.83	25.5	2.55
Total	100	80	8.0

The following sample protocol is routinely used for the preparation of large amounts of RNA and can be easily adjusted to the transcription of templates with a biased nucleotide composition. In such cases, the “standard” NTP mix (20 mM each NTP) is replaced with the template-specific NTP mix with adjusted nucleotide concentrations.

- 10 x Transcription buffer: 400 mM Tris-HCl (pH 7.9 at 25 °C), 120 mM MgCl₂, 50 mM DTT, 50 mM NaCl and 10 mM spermidine.

	Final concentration	Volume added
10 x Transcription buffer	1 x buffer	25 µl
NTP mix (here: 33 mM ATP/9 mM CTP/12.5 mM GTP/25.5 mM UTP)	0.1 x NTP mix	25 µl
Template (linearized plasmid 3.2 kb) 0.1 µg µl ⁻¹	50 µg ml ⁻¹	125 µl
T7 RNAP (own preparation, 5–10 µg µl ⁻¹ total protein) ^a	10–50 µg ml ^{-1a}	2–5 µl ^a
RNase-free water	—	ad 250 µl

^aDepending on the specific activity of the individual T7 RNAP preparation; see Section 1.5.2.2.

Note that because of the high concentrations of NTPs and Mg²⁺ in the reaction, insoluble precipitates may form if the complete mix is kept on ice. It is thus advisable to start with water before adding the other components, and to prewarm the mix to 37 °C before adding the template and polymerase. We let the reactions proceed for at least 2 h (up to overnight) at 37 °C and quench the excess Mg²⁺ by adding Na₂EDTA to a final concentration of 25 mM before phenol extraction and EtOH precipitation (Section 1.3.4, step 1).

After transcription, product yield and quality (5–10 µl aliquot plus an equal volume of gel loading buffer) are checked by PAGE in the presence of 8 M urea and staining with toluidine blue or ethidium bromide. Load at least one reference RNA on the gel to identify the genuine product, since sometimes a complex mixture of bands is observed. A small number of bands in addition to the product band points to good transcription performance, and high yields of transcription correlate with the observation that the RNA product appears as a prominent band, while the DNA template is faintly visible. However, with some templates one has to be satisfied with product amounts exceeding that of the template less than tenfold. Aberrant transcripts of similar size and abundance as the desired product, sometimes even appearing as a smear, can make identification and gel purification of the RNA product of interest impossible. In view of such potential problems, the best transcription protocol will be the one generating the highest amount of specific product at the lowest cost of incorrect products. If RNA yields are not satisfactory, vary concentrations of template, MgCl₂, T7 RNAP, or DTT for further optimization.

1.3.2

Transcription with 2'-Fluoro-Modified Nucleotides

To produce nuclease-resistant 2'-fluoro-modified RNAs with the Y639F mutant T7 RNAP, we replaced UTP and CTP with the corresponding 2'-fluoro analogs (Protocol 9: 2'-fluoro analogs supplied by IBA, Protocol 10: 2'-fluoro analogs supplied by Epicentre Biotechnologies, or TriLink BioTechnologies). Protocol 9 has been employed to transcribe an internally ^{32}P -labeled RNA library for *in vitro* selection using a 117 bp double-stranded PCR template including an internal segment of 60 randomized positions. Transcription assays were incubated for 3–4 h at 37 °C. Protocol 10 was used to enzymatically synthesize one individual aptamer (~100 nt; without ^{32}P -labeling) isolated in the above *in vitro* selection study. In this particular case, the optimal Mg^{2+} concentration was determined as 6 mM.

2'-F-ATP and 2'-F-GTP analogs have become available as well (TriLink BioTechnologies), opening the perspective to also replace purines with 2'-fluoro analogs in T7 transcripts.

Epicentre Biotechnologies also offers the DuraScribe® T7 and SP6 Transcription Kits for the synthesis of transcripts with 2'-F-pyrimidine substitutions. Such transcripts, additionally started with 5'-biotin-GMP, have been synthesized in an aptamer study [28].

Protocol 9: Transcription of a 2'-Fluoro- Modified and ^{32}P -Labeled RNA Library (Hartmann Lab)

- 10 x Transcription buffer: 400 mM Tris-HCl (pH 8.0), 200 mM MgCl_2 , 10 mM spermidine, 0.1% Triton X100.

	Final concentration	150 μl
10 x Transcription buffer	1 x buffer	15 μl
DTT 100 mM	5 mM	7.5 μl
2'-F-CTP, 2'-F-UTP (10 mM each)	1.25 mM (each)	18.75 μl
ATP, GTP (10 mM each)	1.25 mM (each)	18.75 μl
$[\alpha\text{-}^{32}\text{P}]\text{ATP}$ 800 Ci mmol^{-1} (10 mCi ml^{-1})	0.2 mCi ml^{-1}	3 μl
PCR template	1–3.33 nmol ml^{-1}	0.15–0.5 nmol
Y639F mutant T7 RNAP (~4 μg μl^{-1})	100 μg ml^{-1}	3.8 μl
RNase-free water	—	ad 150 μl

Protocol 10: Transcription of an Individual 2'-Fluoro-Modified RNA Aptamer (Hartmann Lab)

- 2 x Transcription buffer: 160 mM HEPES (pH 7.5), 4 mM spermidine, 20 mM DTT, 0.24 mg ml^{-1} BSA (Sigma–Aldrich A7906).

	Final concentration	50 μ l
2 x Transcription buffer	1 x buffer	25 μ l
MgCl ₂ 0.5 M	6 mM	0.6 μ l
Pyrophosphatase 0.1 mg/ml ^a	5 μ g ml ⁻¹	2.5 μ l
Template (BamH1-linearized plasmid 3.1 kb) 1 μ g μ l ⁻¹	80 μ g ml ⁻¹	4.0 μ l
2'-F-CTP, 2'-F-UTP (25 mM each)	1.88 mM (each)	3.75 μ l
ATP, GTP (25 mM each)	1.88 mM (each)	3.75 μ l
Y639F mutant T7 RNAP (prepared according to Section 1.5.2.2)	—	1 μ l
RNase-free water	—	ad 50 μ l

^aPyrophosphatase (Roche) diluted from stock (1 mg ml⁻¹) with 50% glycerol.

1.3.3

T7 Transcripts with 5'-Cap Structures

Eukaryotic mRNAs, snRNAs, and many viral RNAs carry 5'-cap structures that are important for RNA stability, splicing, intracellular transport, initiation of protein synthesis, or infectivity (viral RNAs). 5'-Capped RNAs for biochemical studies can be synthesized by T7 RNAP, with four different cap structures available from Epicentre Biotechnologies CellScript: unmethylated (G[5']ppp[5']G), monomethylated (m⁷G[5']ppp[5']G), trimethylated (m₃^{2,2,7}G[5']ppp[5']G), and antireverse cap analog (ARCA; m₂^{7,3'-O}G[5']ppp[5']G). With the m⁷G[5']ppp[5']G analog, 50% of the cap is incorporated in the reverse orientation, which compromises translation efficiency. In contrast, ARCA is incorporated exclusively in the correct orientation, but the priming of transcripts with ARCA is less efficient than with the standard monomethylated cap analog [44]. Protocol 11 (see below) has been routinely applied to transcription of RNAs with unmethylated or monomethylated cap analogs. If the outcome with this protocol is unsatisfactory for ARCA-initiated T7 transcription, Epicentre offers the MessageMAX™ T7 ARCA-Capped Message Transcription Kit for high-yield transcription.

Protocol 11: Transcription with 5'-Cap Analogs (Bindereif Laboratory)

- 5 x Cap transcription buffer: 200 mM Tris-HCl (pH 7.9), 30 mM MgCl₂, 50 mM DTT, 50 mM NaCl, 10 mM spermidine
- T7 RNAP (Fermentas)
- RNaseOUT™ Recombinant Ribonuclease Inhibitor (Life Technologies Invitrogen).

	Final concentration	25 μ l
5 x Cap transcription buffer	1 x	5 μ l
ATP, CTP, UTP (5 mM each)	0.5 mM (each)	2.5 μ l
GTP 1 mM	0.1 mM	2.5 μ l
Cap analog (10 mM)	1 mM	2.5 μ l
Template (linearized plasmid 3.2 kb) 1 μ g μ l ⁻¹	40 μ g ml ⁻¹	1 μ l
T7 RNAP 20 U μ l ⁻¹	800 U ml ⁻¹	1 μ l
RNaseOUT 40 U μ l ⁻¹	1600 U ml ⁻¹	1 μ l
RNase-free water	—	ad 25 μ l

Incubate for 1–2 h at 37 °C.

1.3.4

Purification

Denaturing PAGE is most commonly used to purify RNA transcripts. However, the denaturing conditions may lead to aberrant folding or aggregation of the RNA sample, which may persist despite a refolding procedure applied before RNA use. For such cases, Chapter 5 describes the method of non-denaturing RNA purification by weak anion-exchange chromatography. Alternative methods are preparative high-performance liquid chromatography (HPLC) if available (see Chapters 7 and 23) or native PAGE to purify a homogeneously folded RNA conformer. If the full-length RNA of interest is the main product of the *in vitro* transcription, one may simply use Amicon[®] Ultra centrifugal filter devices provided with different molecular weight cutoffs (e.g., 50 kDa; Millipore) to concentrate the RNA, to remove salts and exchange buffer.

The purification steps outlined below involve denaturing PAGE. Some purification steps are optional and depend on transcription quality and the demands on product purity. Often, a purification procedure only including steps 5–8 is sufficient. Another protocol for purification of RNA by denaturing PAGE is described in Chapter 3, Section 3.4, Protocol 3.

- 1) *Insoluble pyrophosphate complexes.* In transcription reactions without pyrophosphatase or when pyrophosphatase activity is low, a white pyrophosphate precipitate may form. In such cases, remove the precipitate before steps 2 (see below) by centrifugation at 14000 g for about 5 min directly after transcription. Carefully remove the clear supernatant and transfer to a new Eppendorf tube for further sample processing. Also, Na₂EDTA (500 mM, pH 7.5) may be added immediately after transcription to give a final concentration of 50–100 mM. By chelating Mg²⁺, the formation of insoluble precipitates is substantially reduced.
- 2) *DNase I digestion to remove template DNA.* Add 10 U DNase I (RNase-free, Roche) per 200 μ l and incubate for 20 min at 37 °C.
- 3) *Phenol and chloroform extractions to remove the enzyme(s).* For extraction with phenol (Roti[®] Aqua-Phenol, water-saturated, pH 4.5–5.0; Roth #A980.3), mix

- the sample with 1 vol. phenol, vortex for 30 s, centrifuge 1–5 min (until phases have cleared) at 12 000 g, withdraw the aqueous upper phase and mix it with 1 vol. chloroform, vortex for 30 s, centrifuge 3 min at 12 000 g, and transfer the aqueous upper phase to a new tube, avoiding to withdraw any chloroform.
- 4) *Removal of salt for better gel resolution.* For example, use illustra NAP-10 Columns (GE Healthcare); column material: Sephadex G-25.
 - 5) *Ethanol precipitation.* Ethanol precipitation is performed to remove residual chloroform and salts, and to concentrate the RNA. Mix sample with 2.5 vol. ethanol, 0.1 vol. 3 M NaOAc (pH 4.7), and 1 μl glycogen ($20 \mu\text{g} \mu\text{l}^{-1}$); leave for 10–20 min at -70°C or at least 2 h at -20°C . Centrifuge for 30–45 min at 4°C and 16 000 g. Wash the pellet with 70% ethanol and centrifuge again for 10 min. After ethanol precipitation and air drying of the pellet, redissolve it in a small volume of RNase-free water and add an equal volume of gel loading buffer (2 x TBE (see step 6), pH 8.0, 2.7 M urea, 67% formamide, 0.02% (w/v) each bromophenol blue (BPB) and xylene cyanol blue (XCB)).
 - 6) *Preparative denaturing PAGE.* A gel well, 6–7 cm wide and 1 mm thick, is appropriate for loading the product RNA from an efficient 1 ml transcription reaction. The pocket size is of some importance, as an overloaded gel may cause separation problems; on the other hand, if the pocket is too large, RNA bands may be barely visible and elution efficiency may decrease. After electrophoresis, the desired RNA band is visualized at 254 nm by UV shadowing and marked for gel excision (for details, see Chapter 3); gel running buffer: 1 x TBE (89 mM Tris base, 89 mM boric acid, 2 mM EDTA).
 - 7) *Elution of RNA product.*
 - a. **Diffusion elution.** Cover the excised gel pieces with elution buffer and shake overnight at 4°C ; in the case of efficient transcriptions, a second gel elution step in fresh elution buffer may substantially increase the yield. Different elution buffers can be used: buffer A: 1 mM EDTA, 200 mM Tris-HCl (pH 7.0)/buffer B: 0.3–1.0 M NaOAc (pH 4.7–5.0)/buffer C (successfully used for the elution of phosphorothioate-modified RNAs): 1 M NH_4OAc (pH 7.0). Usually, buffer A is used; buffer B was found to be advantageous in cases where elution efficiency with buffer A was low. After elution, RNA is concentrated by ethanol precipitation. A lower NaOAc concentration of buffer B (e.g., 0.3 M) lowers the risk of precipitating NaOAc by ethanol precipitation, which may compromise RNA solubility.
 - b. **Alternatively – electroelution.** Excised gel pieces containing the RNA are placed in an Elutrap[®] chamber (Schleicher & Schuell BioScience, VWR) following the manufacturer's protocol. The RNA is eluted in 0.5 x TBE buffer (see step 6). The final volume of RNA solution after elution is approximately 600 μl , depending on the extent of evaporation during the elution process. The elution is permitted to proceed overnight at 150 V/20 mA or for 4–6 h at 200–300 V/30 mA. During the elution process there is evaporation, resulting in condensation on the lid of the Elutrap chamber. This lid has to be closed when the Elutrap is running. To minimize evaporation/condensation, the Elutrap should not be run at a voltage higher than 150 V overnight. After

elution, the RNA is extracted once with phenol and twice with chloroform, followed by ethanol precipitation.

- 8) *Quantification (UV spectroscopy, see Chapter 3 and Appendix) and quality check (analytical denaturing PAGE).*

1.4

Troubleshooting

1.4.1

Low or No Product Yield

- Evidence for RNase contamination: check electrophoresis equipment; the use of a single graduated cylinder for diluting 5 x TBE stocks to 1 x TBE buffer, irrespective of DNA or RNA work, is an obvious source of RNase contamination. Decontaminate electrophoresis equipment by soaking in 3% hydrogen peroxide for 10 min. Remove peroxide by extensively rinsing with RNase-free water (diethyl pyrocarbonate (DEPC)-treated or double-distilled). Also, freshly prepare 1 x TBE buffer. Check quality of deionized water used for preparing the TBE electrophoresis buffer, which may contain traces of transition metal ions. If necessary, switch to double-distilled or DEPC-treated water as used for all other aqueous solutions in RNA work. Several companies offer RNase inhibitors, but this is not a solution in case of general RNase degradation problems. For avoidance of RNase contamination, see also web pages of, for example, New England Biolabs (NEB) or Ambion (NEB web page, Technical Reference, General Molecular Biology Data “Avoiding Ribonuclease Contamination”; Ambion Life Technologies, RNA Technical Resources page, “The Basics: RNase Control”).
- If product yields are low with a protocol that had already been successfully used for the same template, repeat transcription assay once without any alteration on a 50 µl scale; if unsuccessful, test different enzyme batches or enzymes from alternative suppliers as well. Differences between enzyme preparations can be considerable.
- Be sure that all components (except enzymes) have been warmed up to ambient temperature before preparation of reaction mixtures.
- Check that thawed stock solutions, particularly concentrated transcription buffers, do not contain precipitated ingredients. For nucleotide solutions, limit freeze/thawing cycles, store in aliquots at -20°C , and adjust stock solutions (in H_2O) with NaOH to pH 7.0 or consider to buffer with 10–40 mM Tris-HCl adjusted to the pH used in transcription reactions; be aware that diluted working solutions may degrade rapidly.
- Check integrity of the DNA template (prepare new template if quality is equivocal), do not use linearized plasmids with 3' overhangs and consider transcribing a control template for comparison. Take particular care to effectively remove traces of SDS, EDTA, RNases, salts (~50% inhibition of T7 RNAP by NaCl or KCl above 150 mM) as well as traces of phenol and chloroform.

- For templates with a highly biased nucleotide composition (e.g., coding for RNAs with extremely high A+U content [42, 43]), adjust the NTP solution according to the nucleotide composition of the transcript (Protocol 8). However, do not alter the total nucleotide concentration of the reaction mix. For G+C – rich template DNAs, increased GTP and CTP levels, an elevated assay temperature (42 °C), addition of *E. coli* single-stranded DNA-binding protein (SSB; [45]) or addition of 3% DMSO may be considered to increase transcript yields.
- In transcription reactions without pyrophosphatase, pyrophosphate precipitates may form. Such precipitates can be removed as described in step 1 of Section 1.3.4.
- Gel entry problems or smear on gel: consider to apply steps 1–5 of Section 1.3.4 before proceeding to gel purification (step 6).

1.5

Rapid Preparation of T7 RNA Polymerase

This protocol is based on the publications of Grodberg and Dunn [37] and Zawadzki and Gross [38] and provides a fast and efficient procedure for the preparation of a highly stable T7 RNAP, which is sufficiently pure for most purposes. The chromatography is described for fast-performance liquid chromatography (FPLC), but any standard low-pressure equipment will give satisfactory results if the procedure is adapted accordingly.

1.5.1

Required Material

- *E. coli* BL21 pAR1219 ([1, 46], provided by F.W. Studier, Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA) or *E. coli* BL21 pDPT7Y639F ([47], provided by R. Sousa, Department of Biochemistry, University of Texas Health Science Center, San Antonio, Texas 78229, USA).
- Branson Sonifier 250 with Tip 1/2" (end diameter 13 mm) (Heinemann, Schwäbisch Gmünd, Germany, Cat. No. 101-063-674 and 101-148-013).

1.5.1.1 Medium

LB (Luria-Bertani) medium [48] supplemented with 50–100 $\mu\text{g ml}^{-1}$ ampicillin.

1.5.1.2 Buffers and Solutions

- 1 M IPTG;
- TEN buffer (50 mM Tris HCl pH 8.1; 2 mM Na_2EDTA ; 20 mM NaCl);
- Phenylmethyl sulfonyl fluoride (PMSF), 20 mg ml^{-1} in isopropanol;
- Leupeptin, 5 mg ml^{-1} ;
- 0.8% Na-deoxycholate solution;
- 2 M ammonium sulfate (enzyme grade);
- Polymin P: 10% solution, adjusted to pH 8 with HCl;
- Saturated ammonium sulfate solution (4.1 M; adjust pH to 7 with some drops of concentrated Tris base, keep at 4 °C where a precipitate will form);

- Buffer C-10, C-100, C-300, C-500, C-1000: buffer C (20 mM Na-phosphate pH 7.7; 1 mM Na₂EDTA; 1 mM DTT; 5% (w/v) glycerol) adjusted to 10, 100, 300, 500, or 1000 mM NaCl. The basis for all C buffers is a 1 M stock of phosphate buffer pH 7.7 (diluted from a mixture of 84.5 vol. of 1 M Na₂HPO₄ and 15.5 vol. of 1 M NaH₂PO₄; if crystals have formed, they need to be dissolved before this solution is used). Directly before use, supplement with NaCl (stock 5 M), PMSF, DTT (stock 1 M), and glycerol to the desired final concentrations; adjust the final volume by addition of double-distilled water precooled to 4 °C.
- 4 x Laemmli gel loading buffer: 100 mM Tris/HCl pH 6.8, 8% (w/v) SDS, 30% glycerol (w/v), 8% (v/v) β-mercaptoethanol, 0.04% (w/v) BPB; adjust pH before addition of BPB.
- sterile glycerol (autoclaved) 99.5%.

1.5.1.3 Electrophoresis and Chromatography

- Laemmli-type SDS gel for protein separation under denaturing conditions (10% PAA);
- Merck EMD Fractogel SO₃⁻, equilibrated in buffer C-100 in a 2 × 10 cm column.

1.5.2

Procedure

1.5.2.1 Cell Growth, Induction, and Test for Expression of T7 RNAP

- 1) Inoculate 25 ml LB containing 50–100 µg ml⁻¹ ampicillin with a colony from a fresh plate culture and grow overnight at 37 °C.
- 2) Two 2 l flasks with 500 ml of the same medium are inoculated 1 : 50 from this culture; grow at 37 °C under vigorous shaking.
- 3) When the cultures have reached an OD₆₀₀ of about 0.6 (which should take not longer than 3 h), transfer 1 ml to an Eppendorf tube, centrifuge for 5 min at 5000 rpm in a desktop centrifuge, and keep the sediment as a control.
- 4) Then induce T7 RNAP expression in the remaining culture by addition of IPTG to a final concentration of 1 mM and shake the culture for 3 h at 37 °C.
- 5) Take a 1 ml sample as in step 3; harvest the bulk of the culture by centrifugation (10 min, 3000–5000 g), wash once with TEN buffer, shock-freeze in liquid N₂ or dry ice, and keep at –80 °C until needed.
- 6) Analyze the 1 ml samples from steps 3 and 5 for expression of T7 RNAP as follows: resuspend the cell sediment in an appropriate volume (example: if OD₆₀₀ before induction was 0.6, suspend pellet in 60 µl and if OD₆₀₀ after induction was 2.5, suspend pellet in 250 µl) of 1 x Laemmli gel loading buffer (note that this buffer is usually prepared as a more concentrated stock solution; see Section 1.5.1.2) and denature for 2 min at 95 °C. Then load 10 µl of each sample onto an SDS-10% polyacrylamide gel with appropriate size markers. If expression induction has been successful (a strong band of about 100 kDa should appear 2–3 h after induction), proceed with enzyme purification.

1.5.2.2 Purification of T7 RNAP

Generally, all steps are performed on ice or at 4 °C, and all buffers are supplemented with the protease inhibitor PMSF (20 µg ml⁻¹, if not stated otherwise). From each purification step, a small sample should be retained for SDS-PAGE monitoring of the purification process. All buffers used for FPLC must be filtered through a 0.2 µm filter.

- 1) Resuspend cells in 24 ml of TEN buffer supplemented with 50 µl of PMSF and 20 µl of leupeptin stock solutions (Section 1.5.1.2).
- 2) Add 2.5 ml of 0.8% Na-deoxycholate solution and incubate for 20 min.
- 3) Sonicate the sample to further promote cell lysis and to shear the DNA (15 min; duty cycle 50%; output 4; in an ice-water bath).
- 4) Centrifuge to clear the supernatant (15 min, 18 000 g, 4 °C).
- 5) Add 5 ml 2 M ammonium sulfate and adjust the total volume to 50 ml with TEN buffer.
- 6) Precipitate DNA by slow addition of 5 ml Polymix P and stirring for 20 min.
- 7) After centrifugation (15 min, 18 000 g, 4 °C), keep the supernatant and determine its volume.
- 8) Precipitate the enzyme from the supernatant by slow addition of 0.82 vol. of saturated ammonium sulfate and stirring for another 15 min.
- 9) After centrifugation for 15 min at 18 000 g, 4 °C, resuspend the sediment in 10 ml buffer C-100, and dialyze 3 × against 1 l of the same buffer (1 × 1 h, 1 × 1 h, 1 × overnight). Dialysis should be extensive in order to completely remove the ammonium sulfate contained in the sediment; otherwise, T7 RNAP will not bind to the cation exchange column.
- 10) Remove insoluble material by centrifugation as in step 9.
- 11) Filtrate the supernatant using a syringe filter (Sarstedt, Filtropur S 0.45).
- 12) Apply the filtrate to the EMD-SO₃⁻-column at a flow rate of 1–2 ml min⁻¹. Wash the column with 10 vol. of buffer C-100 or until protein is no longer detectable in the flow-through. Then apply a 500 ml gradient from 100 to 1000 mM NaCl in buffer C and collect 3 ml fractions. T7 RNAP elutes between 150 and 250 mM NaCl from the EMD-SO₃⁻-column, as visible by the high protein content in these fractions. Finally, wash the column with buffer C-1000; the resin can be reused after equilibration in buffer C-100.
- 13) Five microliters of each fraction and the flow-through, and 2–5 µl of the input material loaded in step 12, are then analyzed by SDS-PAGE.
- 14) The fractions containing the highest amounts of T7 RNAP (around 13 fractions each 3 ml) are pooled and dialyzed 3 × against 1 l of buffer C-10 (1 × 1 h, 1 × 1 h, 1 × overnight).
- 15) After dialysis, centrifuge (15 min at 18 000 g, 4 °C). It depends on the T7 RNAP concentration in the dialyzed sample if the enzyme will precipitate at this point or not. If there is massive precipitation, proceed to step 17; if no pellet is visible, concentrate the supernatant as described in step 16.
- 16) The supernatant is concentrated with centrifugal filter devices (Amicon Ultra-15, Millipore) to about 0.5–1 ml, followed by addition of an equal volume of

sterile glycerol; store in aliquots at -20°C . Yet, if the enzyme precipitates during this concentration process (and clogs the filter), then disperse the supernatant in the filter unit by pipetting up and down and transfer it to a centrifugation tube, followed by centrifugation (15 min at 18 000 g, 4°C). The pellet is handled as described in step 17. The supernatant is concentrated in a new centrifugal filter to about 0.3–0.5 ml, followed by addition of an equal volume of sterile glycerol and storage in aliquots at -20°C .

- 17) The pellet containing T7 RNAP is dissolved in 0.3–0.5 ml C-300 buffer; then add an equal volume of sterile glycerol and store the enzyme preparation in aliquots at -20°C .

Specific activity of T7 RNAP may be determined by incorporation of ^{32}P -labeled nucleotides into acid-precipitable material and can reach $400\,000\text{ U mg}^{-1}$ (one U is defined as the incorporation of 1 nmol AMP into acid-precipitable material in 1 h at 37°C [12]). For most purposes, it is sufficient to titrate the amount of enzyme preparation needed to give good transcription yields without too many side products. To analyze a typical preparation of T7 RNAP ($5\text{--}10\text{ }\mu\text{g }\mu\text{l}^{-1}$ total protein content), we usually test $1\text{ }\mu\text{l}$ and $1\text{ }\mu\text{l}$ of a 1 : 10 dilution according to Protocol 1 scaled down to $50\text{ }\mu\text{l}$ and containing a standard template DNA (about 50% G + C). Transcription efficiency and quality are then evaluated by analyzing $5\text{ }\mu\text{l}$ (1/10th) of the test transcription on a denaturing PAA gel followed by gel staining with ethidium bromide. Alternatively, for easier detection of abortive transcripts and degradation products, a radioactive tracer (e.g., $10\text{ }\mu\text{Ci}$ of $[\alpha\text{-}^{32}\text{P}]\text{UTP}$ or $[\alpha\text{-}^{32}\text{P}]\text{GTP}$) can be added to the transcription reaction and the products visualized on a Phosphorimager or by autoradiography (Protocol 7).

1.5.3

Notes and Troubleshooting

- 1) We advise not to use lysozyme for cell disruption (Section 1.5.2.2, steps 2 and 3). In our hands, lysozyme treatment caused T7 RNAP to appear entirely in the insoluble fraction, for reasons not understood at present.
- 2) If protein gel electrophoresis of crude cell samples (Section 1.5.2.1, step 6) yields smeared bands, try to shear the DNA by sonication as described for the enzyme purification (Section 1.5.2.2, step 3). Alternatively, samples can be squeezed with a syringe through a thin (0.7 mm, or 22-gauge) needle.
- 3) If expression of T7 RNAP is insufficient, start again with a fresh cell clone picked from an LB-ampicillin agar plate; if this is unsuccessful, prepare a new IPTG stock solution, vary the IPTG concentration used for induction, change growth times before and after induction, or switch to other growth media, such as TB or 2 x YT [48], or M9 Minimal Medium supplemented with trace elements [49].
- 4) The Fractogel EMD- SO_3^- -column can be substituted by any strong cation exchanger of the SO_3^- -type. However, with most other column matrices, T7 RNAP elutes at different NaCl concentrations.

- 5) High enzyme concentrations and low salt favor the precipitation of T7 RNAP during dialysis. For this reason, the dialysis step is performed in buffer C-10 (Section 1.5.2.2, step 14). When performing the dialysis step in C-100, we did not observe any precipitation of T7 RNAP, even not during the volume reduction procedure with centrifugal filter devices (step 16), indicating that the enzyme concentration was not high enough for precipitation at 100 mM NaCl.

References

1. Davanloo, P., Rosenberg, A.H., Dunn, J.J., and Studier, F.W. (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 2035–2039.
2. Ebright, R.H. (2000) RNA polymerase: structural similarities between bacteria RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.*, **304**, 687–698.
3. Cheetham, G.M.T. and Steitz, T.A. (2000) Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.*, **10**, 117–123.
4. Diaz, G.A., Rong, M., McAllister, W.T., and Durbin, R.K. (1996) The stability of abortively cycling T7 RNA polymerase complexes depends upon template conformation. *Biochemistry*, **35**, 10837–10843.
5. Kochetkov, S.N., Rusakova, E.E., and Tunitskaya, V.L. (1998) Recent studies of T7 RNA polymerase mechanism. *FEBS Lett.*, **440**, 264–267.
6. Brakmann, S. and Grzeszik, S. (2001) An error-prone T7 RNA polymerase mutant generated by directed evolution. *ChemBioChem*, **2**, 212–219.
7. Huang, F., Bugg, C.W., and Yarus, M. (2000) RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. *Biochemistry*, **39**, 15548–15555.
8. Dunn, J.J. and Studier, F.W. (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.*, **166**, 477–535.
9. Lee, S.S. and Kang, C. (1993) Two base pairs at -9 and -8 distinguish between the bacteriophage T7 and SP6 promoters. *J. Biol. Chem.*, **268**, 19299–19304.
10. Milligan, J.F. and Uhlenbeck, O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Meth. Enzymol.*, **180**, 51–63.
11. Pokrovskaya, I.D. and Gurevich, V.V. (1994) *In vitro* transcription: preparative RNA yields in analytical scale reactions. *Anal. Biochem.*, **220**, 420–423.
12. Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.
13. Helm, M., Brulé, H., Giegé, R., and Florentz, C. (1999) More mistakes by T7 RNA polymerase at the 5' ends of *in vitro*-transcribed RNAs. *RNA*, **5**, 618–621.
14. Pleiss, J.A., Derrick, M.L., and Uhlenbeck, O.C. (1998) T7 RNA polymerase produces 5' end heterogeneity during *in vitro* transcription from certain templates. *RNA*, **4**, 1313–1317.
15. Coleman, T.M., Wang, G., and Huang, F. (2004) Superior 5' homogeneity of RNA from ATP-initiated transcription under the T7 phi 2.5 promoter. *Nucleic Acids Res.*, **32**, e14.
16. Sampson, J.R. and Uhlenbeck, O.C. (1988) Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 1033–1037.
17. Griffiths, A.D., Potter, B.V., and Eperon, I.C. (1987) Stereospecificity of nucleases towards phosphorothioate-substituted RNA: stereochemistry of transcription by

- T7 RNA polymerase. *Nucleic Acids Res.*, **15**, 4145–4162.
18. Axelrod, V.D. and Kramer, F.D. (1985) Transcription from bacteriophage T7 and SP6 RNA polymerase promoters in the presence of 3'-deoxyribonucleoside 5'-triphosphate chain terminators. *Biochemistry*, **24**, 5716–5723.
 19. Aurup, H., Williams, D.M., and Eckstein, F. (1992) 2'-Fluoro- and 2'-amino-2'-deoxynucleoside 5'-triphosphates as substrates for T7 RNA polymerase. *Biochemistry*, **31**, 9636–9641.
 20. Padilla, R. and Sousa, R. (1999) Efficient synthesis of nucleic acids heavily modified with non-canonical ribose 2'-groups using a mutant T7 RNA polymerase. *Nucleic Acids Res.*, **27**, 1561–1563.
 21. Conrad, F., Hanne, A., Gaur, R.K., and Krupp, G. (1995) Enzymatic synthesis of 2'-modified nucleic acids: identification of important phosphate and ribose moieties in RNase P substrates. *Nucleic Acids Res.*, **23**, 1845–1853.
 22. Gaur, R.K. and Krupp, G. (1993) Enzymatic RNA synthesis with deoxynucleoside 5'-O-(1-thiotriphosphates). *FEBS Lett.*, **315**, 56–60.
 23. Hardt, W.D., Erdmann, V.A., and Hartmann, R.K. (1996) Rp-deoxy-phosphorothioate modification interference experiments identify 2'-OH groups in RNase P RNA that are crucial to tRNA binding. *RNA*, **2**, 1189–1198.
 24. Veedu, R.N., Vester, B., and Wengel, J. (2008) Polymerase chain reaction and transcription using locked nucleic acid nucleotide triphosphates. *J. Am. Chem. Soc.*, **130**, 8124–8125.
 25. Veedu, R.N., Burri, H.V., Kumar, P., Sharma, P.K., Hrdlicka, P.J., Vester, B., and Wengel, J. (2010) Polymerase-directed synthesis of C₅-ethynyl locked nucleic acids. *Bioorg. Med. Chem. Lett.*, **20**, 6565–6568.
 26. Stengel, G., Urban, M., Purse, B.W., and Kuchta, R.D. (2010) Incorporation of the fluorescent ribonucleotide analogue tCTP by T7 RNA polymerase. *Anal. Chem.*, **82**, 1082–1089.
 27. Logsdon, N., Lee, C.G.L., and Harper, J.W. (1992) Selective 5' modification of T7 RNA polymerase transcripts. *Anal. Biochem.*, **205**, 36–41.
 28. Kim, Y.-H., Sung, H.J., Kim, S., Lee, J.W., Moon, J.Y., Choi, K., Jung, J.E., Lee, Y., Koh, S.S., Rhee, S.G., Heo, K., and Kim, I.H. (2011) An RNA aptamer that specifically binds pancreatic adenocarcinoma up-regulated factor inhibits migration and growth of pancreatic cancer cells. *Cancer Lett.*, **313**, 76–83, doi: 10.1016/j.canlet.2011.08.027.
 29. Christian, E.L., McPheeters, D.S., and Harris, M.E. (1998) Identification of individual nucleotides in the bacterial ribonuclease P ribozyme adjacent to the pre-tRNA cleavage site by short-range photo-cross-linking. *Biochemistry*, **37**, 17618–17628.
 30. Harris, M.E. and Christian, E.L. (1999) Use of circular permutation and end modification to position photoaffinity probes for analysis of RNA structure. *Methods*, **18**, 51–59.
 31. Sousa, R. and Padilla, R. (1995) A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J.*, **14**, 4609–4621.
 32. Huang, Y., Eckstein, F., Padilla, R., and Sousa, R. (1997) Mechanism of ribose 2'-group discrimination by an RNA polymerase. *Biochemistry*, **36**, 8231–8242.
 33. Ryder, S.P. and Strobel, S.A. (1999) Nucleotide analog interference mapping. *Methods*, **18**, 38–50.
 34. Padilla, R. and Sousa, R. (2002) Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Res.*, **30**, e138.
 35. Pitulle, C., Kleineidam, R.G., Sproat, B., and Krupp, G. (1992) Initiator oligonucleotides for the combination of chemical and enzymatic RNA synthesis. *Gene*, **112**, 101–105.
 36. Huang, F. (2003) Efficient incorporation of CoA, NAD and FAD into RNA by *in vitro* transcription. *Nucleic Acids Res.*, **31**, e8.
 37. Grodberg, J. and Dunn, J.J. (1988) *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol.*, **170**, 1245–1253.

38. Zawadzki, V. and Gross, H.J. (1991) Rapid and simple purification of T7 RNA polymerase. *Nucleic Acids Res.*, **19**, 1948.
39. Ellinger, T. and Ehricht, R. (1998) Single-step purification of T7 RNA polymerase with a 6-histidine tag. *BioTechniques*, **24**, 718–720.
40. Sastry, S.S. and Ross, B.M. (1997) Nuclease activity of T7 RNA polymerase and the heterogeneity of transcription elongation complexes. *J. Biol. Chem.*, **272**, 8644–8652.
41. Liu, F. and Altman, S. (1996) Requirements for cleavage by a modified RNase P of a small model substrate. *Nucleic Acids Res.*, **24**, 2690–2696.
42. Baum, M., Cordier, A., and Schön, A. (1996) RNase P from a photosynthetic organelle contains an RNA homologous to the cyanobacterial counterpart. *J. Mol. Biol.*, **257**, 43–52.
43. Wagner, M., Fingerhut, C., Gross, H.J., and Schön, A. (2001) The first phytoplasmic RNase P RNA provides new insights into the sequence requirements of this ribozyme. *Nucleic Acids Res.*, **29**, 2661–2665.
44. Peng, Z.H., Sharma, V., Singleton, S.F., and Gershon, P.D. (2002) Synthesis and application of a chain-terminating dinucleotide mRNA cap analog. *Org. Lett.*, **4**, 161–164.
45. Aziz, R.B. and Soreq, H. (1990) Improving poor *in vitro* transcription from G,C-rich genes. *Nucleic Acids Res.*, **18**, 3418.
46. Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dudendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Meth. Enzymol.*, **185**, 60–89.
47. Bonner, G., Patra, D., Lafer, E.M., and Sousa, R. (1992) Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure. *EMBO J.*, **11**, 3767–3775.
48. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning*, Cold Spring Harbor Laboratory Press.
49. Mossakawska, D.E. and Smith, R.A.G. (1997) in *Protein NMR Techniques* (ed. D.G. Reid), Humana Press Inc., Totowa, NJ, 325–335.

2

Production of RNAs with Homogeneous 5'- and 3'-Ends

Mario Mörl and Roland K. Hartmann

2.1

Introduction

Synthesis of RNA molecules by *in vitro* transcription, primarily using T7 RNA polymerase (T7 RNAP) for reasons of price and efficiency, is a widely used method in biochemistry and molecular biology. For most purposes, it is sufficient to purify transcripts by electrophoretic separation on denaturing polyacrylamide (PAA) gels. Thereby, template DNA, enzyme, and nucleotides, as well as incomplete transcripts (resulting from premature transcription termination), can be removed efficiently. However, the seemingly uniform population of full-length transcripts is, in most cases, heterogeneous because T7 RNAP tends to add one or occasionally even a few additional, non-encoded nucleotides to the 3'-end [1, 2]. Also, 5'-end heterogeneity of T7 transcripts has been observed with certain templates [3, 4] (Chapter 1). Such microheterogeneous full-length transcripts can rarely be separated efficiently by denaturing polyacrylamide gel electrophoresis (PAGE). Thus, if applications require complete homogeneity at the transcript 5'- and 3'-ends, for example, in the context of *in vitro* aminoacylation of tRNAs, ligation of RNA fragments (Chapter 3), RNA crystallization, or NMR studies, refined approaches are needed.

Apart from product end heterogeneity, another problem associated with *in vitro* transcription stems from the fact that the sequence requirements of the core promoter regions recognized by phage RNAPs also put some constraints on the 5'-terminal nucleotide identities of the transcript. For example, transcription by T7 RNAP from the predominantly used (class III) T7 promoter only proceeds with reasonable efficiency if the transcript is initiated with a G residue (for more details, see Chapter 1). As a consequence, synthesis yields will be low for transcripts with a different nucleotide identity at position +1.

An approach toward eliminating 5'- and 3'-end heterogeneity as well as sequence constraints at the 5'-end is the use of catalytic RNA entities acting in *cis* or in *trans*, which precisely release the RNA of interest from a primary transcript with extra 5'- and 3'-flanking sequences. These strategies permit the production of RNA without any sequence restrictions and with essentially complete 5'- and 3'-end homogeneity.

2.2

Description of Approach

2.2.1

Cis-Cleaving Autocatalytic Ribozyme Cassettes

2.2.1.1 The 5'-Cassette

An elegant and efficient way to produce transcripts with uniform 5'-ends is to construct a DNA template encoding a self-cleaving hammerhead ribozyme immediately upstream of the RNA sequence of interest. The hammerhead ribozyme is a small structural element, originally identified in pathogenic plant viroids and virusoids, which catalyzes phosphodiester bond hydrolysis at a single defined position. With a length of about 50 nt, the hammerhead sequence can easily be inserted as a DNA cassette at the 5'-end of the DNA sequence to be transcribed (Figure 2.1). In the primary transcript (Figure 2.1, middle), the 5'-part of the RNA of interest is integrated into the secondary structure of the hammerhead ribozyme. Hammerhead self-cleavage occurs already during transcription – provided that Mg^{2+} ions are present at sufficiently high concentrations (≥ 5 mM), as is the case under standard transcription conditions. In addition to 5'-end uniformity, another advantage of a 5'-hammerhead is that optimal transcription start sequences can be used upstream of the ribozyme cassette in favor of high transcription yields.

Within a 5'-flanking hammerhead cassette, the sequence of helix P1 is dictated by the 5'-terminal nucleotides of the RNA of interest (Figure 2.1, middle). As the only requirement for P1 is the formation of an A-helix, there are basically no sequence constraints for this structural element. Thus, any efficient minimal hammerhead variant matching the consensus sequence shown in Figure 2.1 can be used [5–8]. Two examples of 5'-hammerhead constructs used successfully in our hands are shown in Figure 2.2a,b.

2.2.1.2 The 3'-Cassette

For the production of RNAs with homogeneous 3'-ends, hammerhead cassettes (Figure 2.2c) can be again tethered to the RNA of interest. A disadvantage of these structures are some sequence constraints imposed on the product RNA 3'-end, such as the requirement for a 3'-terminal NUH sequence motif (N = any nucleotide; H = C, U, or A), preferentially GUC (Figure 2.2c; [5]).

The autocatalytic domain of the hepatitis delta virus (HDV ribozyme), however, has no such sequence requirements upstream of its cleavage site and can therefore be used for any transcript sequence of interest (Figures 2.1 and 2.2a). This ribozyme folds into four helical domains that form a pseudoknotted structure, with the cleavage site located immediately upstream of helix P1 at the 5'-end of the ribozyme [9, 10]. Using an optimized version of the HDV ribozyme, homogeneous RNA 3'-ends can easily be generated [11]. The HDV ribozyme is already sufficiently active at Mg^{2+} concentrations of about 1 mM, usually resulting in efficient self-cleavage during transcription. A frequently observed problem is that sequences of the RNA of interest interfere with proper folding of the HDV

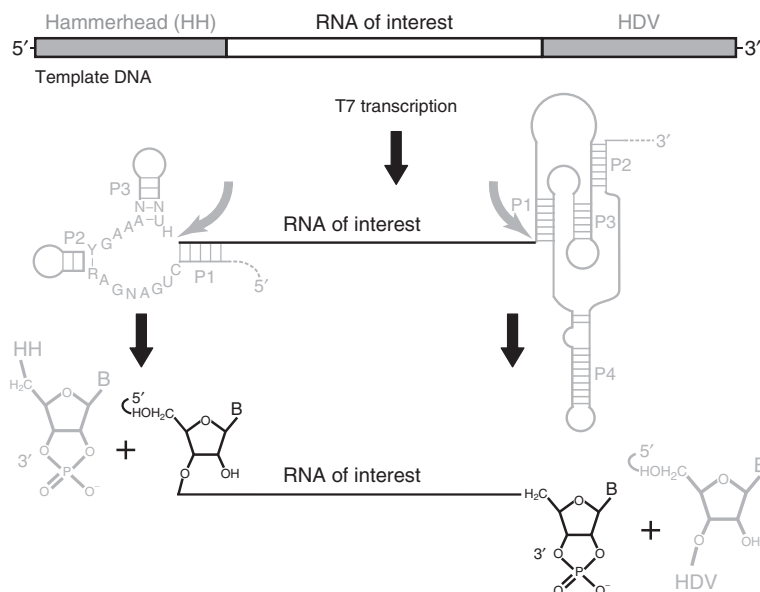


Figure 2.1 5'- and 3'-Cassettes for the generation of homogeneous RNA ends. The DNA molecule to be transcribed (linearized plasmid or PCR product) includes extra sequence cassettes upstream and downstream of the sequence encoding the RNA of interest. These cassettes (indicated in gray at the top) are transcribed into self-cleaving ribozyme structures (hammerhead at the 5'-end and HDV ribozyme at the 3'-end) flanking the RNA of interest. Conserved nucleotides in the hammerhead core structure are indicated; Y = pyrimidine; R = purine; NUH (N = any

nucleotide; H = C, U, or A), preferentially GUC. Both cassettes are designed to permit insertion of virtually any sequence between the ribozymes, rendering this strategy applicable to essentially any RNA of interest. The resulting cleavage product carries a 5'-hydroxyl group, which can be converted to a (radioactively labeled) 5'-phosphate. At the 3'-terminus, a 2',3'-cyclic phosphate group is generated, which interferes with some RNA functions. Protocols for the removal of the 3'-terminal phosphate group are described in the section "3' Dephosphorylation Protocols."

ribozyme, leading to reduced cleavage efficiency. This problem has been solved by incubating the primary transcript repeatedly (10 cycles) for 3 min at 60 °C and 3 min at 25 °C. This procedure should allow the ribozyme cassette of the transcript to adopt, at least transiently, its active structure and has in our hands led to quantitative cleavage in many cases. However, if not successful, one may adjust the transcription sample (optionally after phenol/chloroform extraction and ethanol precipitation) to 30-40 mM MgCl₂, followed by up to 13 thermal cycles of 1 min 72 °C / 5 min 65 °C / 10 min 37 °C [12].

2.2.1.3 Purification of Released RNA Product and Conversion of End Groups

After cotranscriptional self-cleavage of the terminal ribozyme cassette(s), it is advisable to purify the released RNA of interest by denaturing or non-denaturing PAGE and subsequent gel elution in, for example, 500 mM ammonium acetate,

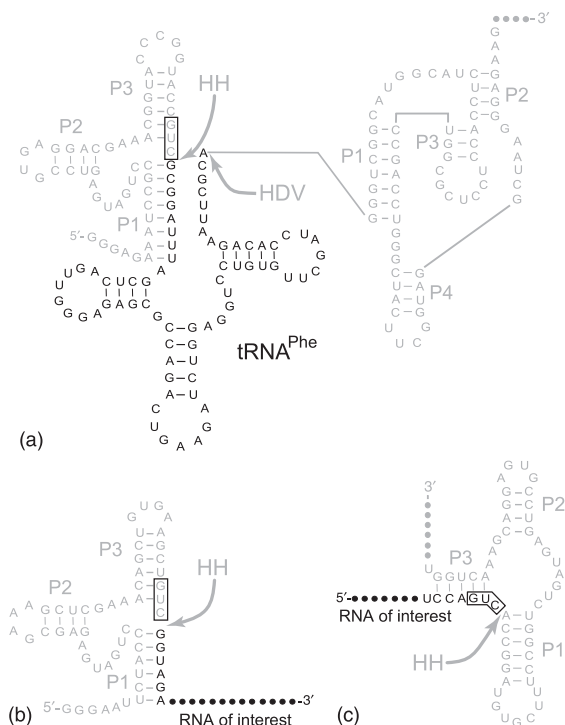


Figure 2.2 Examples of terminal ribozyme cassettes in primary transcripts. The RNA of interest is shown in black, the ribozyme modules are depicted in gray. Ribozyme self-cleavage sites are indicated by arrows. The preferred GUC triad of hammerhead

have displayed efficient and precise autocatalytic cleavage in the authors' laboratories. (a) Sequence of a primary transcript consisting of a 5'-hammerhead, a central tRNA^{Phe} as the RNA of interest, and a 3'-terminal HDV ribozyme. (b,c) Examples of a 5'- and a 3'-terminal hammerhead, respectively.

pH 5.7, 0.1 mM EDTA, and 0.1% SDS at 4 °C overnight, followed by ethanol precipitation ([13]; for detailed protocols, see Chapters 1 and 3; for non-denaturing RNA purification using weak anion-exchange chromatography, see Chapter 5). If, for example, a tRNA product is released from the primary transcript, the ribozyme cassettes will have similar lengths of about 50–70 nt, complicating purification of the tRNA. A practical solution to the problem is to extend the 5'- and/or 3'-termini of the primary transcripts, thereby increasing the size of the released terminal fragments containing the ribozyme cassettes.

Another important point to be considered is the nature of the 5'- and 3'-ends generated by these ribozymes. Both hammerhead and HDV ribozyme cassettes catalyze transesterification reactions induced by a nucleophilic attack of the 2'-OH group of the neighboring ribose, resulting in 5'-hydroxyl and 2', 3'-cyclic phosphate termini instead of 5'-phosphates and 3'-hydroxyls. A terminal 5'-OH group is for most purposes neutral to RNA function and may even be advantageous, because it is directly accessible to 5'-endlabeling with [γ -^{32/33}P]ATP by T4 polynucleotide

kinase (T4 PNK). The 3'-terminal cyclophosphate, however, interferes with a variety of downstream experiments, such as aminoacylation of tRNAs, ligation of RNA molecules, or 3'-endlabeling using [5',³²P]pCp. To remove the 2', 3'-cyclic phosphate, the phosphatase activity of T4 PNK can be employed with excellent efficiency (see section "3' Dephosphorylation Protocols" and [14, 15]).

2.2.2

Trans-Cleaving Ribozymes for the Generation of Homogeneous 3' Ends

As an alternative to the *cis*-cleaving ribozyme cassettes, homogeneous 3'-ends can be obtained by *trans*-cleavage, for example, using a *trans*-acting hammerhead ribozyme [16], the *Neurospora* Varkud satellite (VS) ribozyme [16], or the catalytic RNA subunit of a bacterial RNase P [17] as detailed in the following sections. A *trans*-acting VS ribozyme has the advantage that the RNA of interest requires only a short stem-loop 3'-extension of 24 nt which serves as the ribozyme substrate recognized by tertiary interactions [16]. As the HDV ribozyme, the VS ribozyme has no specific sequence requirements upstream of the cleavage site and generates 2', 3'-cyclic phosphate ends [18]. Similar to the VS ribozyme strategy, it is possible to use the HDV ribozyme in a *trans*-cleavage reaction. Here, the transcripts are extended by as few as seven nucleotides designed to provide the 5'-strand of helix P1 (Figure 2.2a, structure on the right). A correspondingly 5'-truncated HDV RNA can base-pair with this extension, thereby restoring the P1 helix and a functional HDV ribozyme structure. Cleavage of the target RNA then occurs seven nucleotides upstream from its 3'-end [19].

Bacterial RNase P RNA (~380 nt) can accurately and efficiently remove 5'-flanking sequences of tRNA precursor transcripts by endonucleolytic cleavage. To exploit this system for the production of RNAs with homogeneous 3'-ends, the RNA of interest is fused to a downstream tRNA sequence. As a result, the construct mimics a tRNA precursor molecule, whose 5'-leader sequence represents the RNA of interest (Figure 2.3). Precise endonucleolytic cleavage by the RNase P ribozyme then releases the RNA product with uniform 3'-OH termini (Figure 2.3) that, in contrast to the 2', 3'-cyclic phosphates generated by hammerhead and HDV ribozymes, are suitable for many downstream applications.

As the substrate recognition elements for RNase P RNA are located predominantly in the tRNA structure, there are little upstream sequence requirements to be fulfilled. One restriction is to avoid a G residue at position -1 relative to the tRNA, which otherwise may cause some aberrant cleavage between positions -1 and -2. In addition, for reasons of cleavage fidelity, the use of a bacterial class I tRNA moiety with a 7-bp acceptor stem and short variable arm is recommended (such as *Bacillus subtilis* tRNA^{Asp} [17, 21] or *Thermus thermophilus* tRNA^{Gly} [22]). Furthermore, any base-pairing potential between position -1 and the discriminator base of the tRNA (position +73) should be avoided, the discriminator should be D (= G, A, or U) and not C when using *Escherichia coli* RNase P RNA [20], and the first acceptor stem base pair should be G₊₁-C₊₇₂. However, note that some class I tRNA transcripts, despite meeting all the aforementioned requirements,

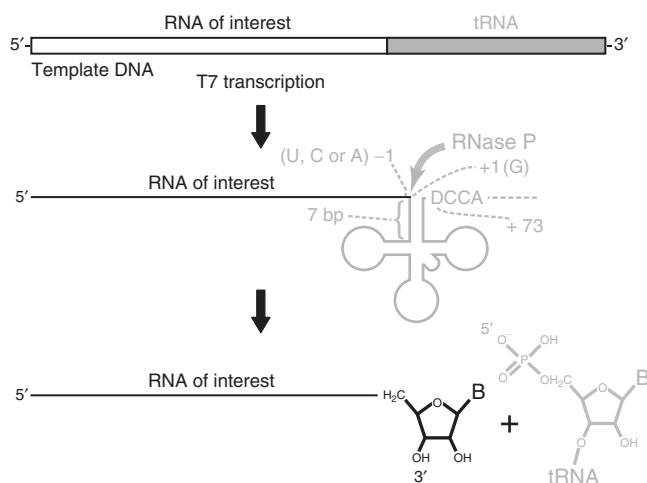


Figure 2.3 Cleavage by bacterial RNase P RNA to generate homogeneous 3'-ends. The DNA template for transcription is designed in a way that a tRNA gene (in gray) is positioned immediately downstream of the sequence encoding the RNA molecule of interest. By incubating the primary transcript, derived from run-off transcription, with bacterial RNase P RNA, the tRNA molecule is cleaved off at its 5'-end and the upstream portion of the primary transcript is released. Aspects of importance: avoid a G residue at position -1 ; avoid base pairing between nt

-1 and discriminator ($+73$); the discriminator should be D (= G, A, or U) and not C when using *E. coli* RNase P RNA [20]; the first acceptor stem base pair should be $G_{+1} - C_{+72}$; choose a class I tRNA with a 7-bp acceptor stem, short variable arm and stable fold (G–C base pairs prevailing in stems); encode CCA at the tRNA 3'-terminus; a 3'-extension of up to six nucleotides beyond CCA does not interfere with RNase P cleavage, but a second CCA trinucleotide should be avoided within this extension.

may give rise to a small percentage of miscleavage [23]; as this is related to tRNA folding stability, avoid tRNA moieties that contain non-canonical base pairs or clusters of A–U base pairs in their stem regions. We have successfully used RNase P RNAs from *E. coli* and *T. thermophilus*, which are available from the authors as T7 expression plasmids. Yet, since achieving 100% 3'-end homogeneity – and not 98% – by use of this approach may require some optimization, we suggest to first consider pursuing an HDV-based strategy as a quicker alternative if absolute homogeneity is critical for downstream applications.

In applications where a short 3'-appendage to the RNA of interest is preferable (e.g., when incorporating costly isotopically labeled nucleotides for NMR studies), the tethered RNase P substrate can be reduced to a hairpin structure of less than half the tRNA size. One such substrate is pATSerU $_{-1}$ G $_{+73}$ and its variants pATSerU $_{-1}$ U $_{+73}$, pATSerC $_{-1}$ A $_{+73}$, and pATSerA $_{-1}$ U $_{+73}$, which release an RNA of interest with a 3'-terminal U, C, or A residue, respectively [24]. These substrates showed >99% cleavage at the canonical RNase P cleavage site under conditions of 50 mM Tris/HCl, pH 7.2, 5% (w/v) PEG 6000, 100 mM NH₄Cl, and 40 mM MgCl₂ [24] (Leif A. Kirsebom, personal communication).

The RNase P RNA-based approach could have advantages over a *cis*-cleaving HDV cassette in cases where interactions between the RNA of interest and the HDV cassette prevent the ribozyme from adopting an active conformation. Such folding interference is less likely for a tRNA cassette because tRNAs are among the most stable autonomous RNA folding units. If folding of the downstream tRNA may nevertheless be impeded by the RNA of interest at 37 °C, thermostable RNase P RNA from *T. thermophilus* may be used for cleavage at elevated temperatures, assuming that folding interference is abolished under such conditions. This ribozyme, when acting on a transcript with, for example, a G/C-rich tRNA^{Gly} cassette from the same organism [25], will cleave off the tRNA moiety with high precision at temperatures of up to ~70 °C [26].

The *trans*-cleaving approach may be somewhat laborious if the ribozyme is prepared independently of the transcript of interest and if *in vitro* transcription is followed by a second, independent incubation step for the cleavage reaction. However, the procedure can be simplified by transcribing both RNAs simultaneously from two different templates added to the same reaction mix [16]. For the RNase P-based approach, however, we recommend to transcribe the substrate RNA separately, followed by phenol/chloroform extraction and ethanol precipitation before incubation with RNase P RNA in cleavage buffer (see section “Protocol for RNase P Cleavage”) that differs from T7 transcription buffers.

2.2.3

Further Strategies toward Homogeneous Ends

In addition to those described above, other approaches for generating homogeneous ends are available as well. Two methods, an RNase H-based strategy and another making use of T7 transcription templates with two consecutive 2'-O-methyl nucleotides at the 5'-end of the template strand, are detailed in Chapter 3. The RNase H-based strategy can be employed to generate homogeneous 5'- and 3'-ends, while the 2'-O-methyl approach is suited for the production of homogeneous 3'-ends only.

Recently, almost complete 5'-end homogeneity of T7 transcripts was demonstrated with templates directing transcription from the less frequently used T7 class II promoter, at which T7 RNAP initiates synthesis with an A instead of a G residue. Transcription yields from this promoter were reported to equal those of the commonly used T7 class III promoter (Chapter 1; [27]).

An elegant variation of the *trans*-cleavage concept, usually relying on ribozymes, involves the 10–23 DNA enzyme [27–29]. This type of DNAzyme has a 15-nt core DNA sequence flanked by two arms that form 8- to 10-bp long hybrid helices with the substrate RNA. Cleavage occurs within a 5'-RY motif (R = A, G; Y = C, U; for details, see [30]) at the junction of the two hybrid helices, resulting in 5'-OH and 2', 3'-cyclic phosphate termini. Sequence-tailored versions of the 10–23 DNAzyme, representing simple DNA oligonucleotides (~30 nt), can be easily obtained from commercial suppliers and provide an inexpensive and effortless, although less explored, alternative to ribozymes in the production of RNAs with homogeneous ends.

2.3

Critical Experimental Steps, Changeable Parameters, Troubleshooting

2.3.1

Construction of *Cis*-Cleaving 5'- and 3'-Cassettes

A critical step in the construction of hammerhead and/or HDV ribozyme cassettes is to establish an efficient overlap extension PCR (Figure 2.4). Usually, the 5'-hammerhead cassette is created by two overlapping oligonucleotides that cover the complete hammerhead domain (Figure 2.4a). In addition, the upstream primer can carry the sequence for the T7 RNAP promoter at the 5'-end (and/or a terminal restriction enzyme cleavage site if plasmid cloning of the PCR fragment

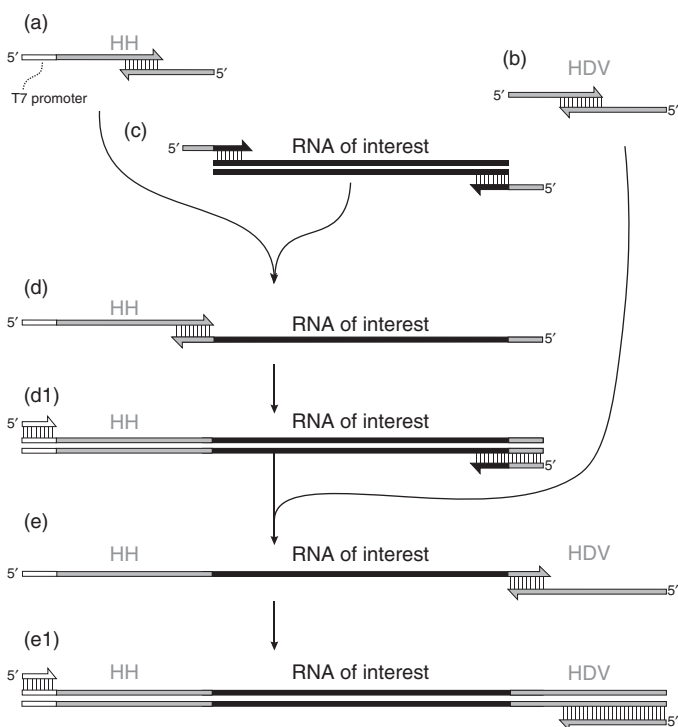


Figure 2.4 Construction of transcription templates carrying autocatalytic ribozyme cassettes at the 5'- and 3'-termini. (a,b) By PCR extension of overlapping primer pairs, the initial cassettes (HH, hammerhead; HDV, HDV ribozyme) are generated. (c) Using primers with 5'-extensions overlapping the ribozyme cassettes, the sequence of interest is amplified in a third PCR. (d) Subsequently, PCR products from reactions

a and c are combined (overlap extension PCR without addition of primers) to fuse the hammerhead sequence to the sequence encoding the RNA of interest. (d1) The resulting product is further amplified using the indicated primers. (e) Eventually, the same strategy is applied to append the HDV sequence. (e1) The final product carrying both cassettes at the corresponding ends is further amplified with terminal primers.

is intended). A minor disadvantage is that such an oligonucleotide will be extended by at least 17 nt, which is associated with lower yields of chemical synthesis. Alternatively, an upstream primer without T7 promoter sequence may be used. The PCR product can then be inserted into a cloning vector that encodes the T7 promoter immediately upstream of the cloning site, or the T7 promoter is introduced in a follow-up PCR step with a shorter upstream primer.

In a second PCR, the 3'-cassette representing the HDV ribozyme domain is synthesized, again by the use of overlapping primers (Figure 2.4b). In a third step, the sequence encoding the RNA of interest is amplified (Figure 2.4c). Here, the upstream primer includes a region overlapping the hammerhead sequence, such that the resulting product can be used for an overlap extension in combination with the hammerhead PCR product (Figure 2.4d). Similarly, the downstream primer introduces an extension corresponding to the 5'-part of the HDV ribozyme cassette. In a final overlap extension (Figure 2.4e), the amplified product from Figure 2.4d1 is combined with the HDV PCR product in order to generate the full-length construct. Before cloning this product into an appropriate plasmid, it is further amplified using terminal primers (Figure 2.4e1). After cloning, inserted sequences need to be verified, because the numerous PCR steps involved may lead to aberrant products or sequence deviations.

A complete protocol representing an example of the PCR strategy shown in Figure 2.4 is detailed below. The resulting final primary transcript with upstream hammerhead and downstream HDV ribozyme cassettes is illustrated in Figure 2.2a. PCRs outlined below have been successfully performed with *Taq* DNA polymerase. However, one may consider to use a thermostable DNA polymerase with 3'–5' proofreading activity, such as *Pfu* polymerase, for the generation of all PCR products that are subsequently used in overlap extension reactions. The reason is that *Taq* polymerase (which has no proofreading activity) tends to add a single nontemplated A residue to the 3'-end of PCR products. While this activity is exploited in some cloning strategies (TA-cloning kits), it potentially interferes with overlap extensions, because the additional 3'-terminal A does not base pair with the complementary strand. As a consequence, the fraction of strands carrying this extra A residue may not be extended, thus decreasing the overall yield of extension product [31].

2.4

PCR Protocols

PCRs were performed with 2.5 U *Taq* DNA polymerase per 50 µl standard reaction volume in 10 mM Tris/HCl, pH 8.3, 1.5 mM MgCl₂, and 50 mM KCl. For overlap extensions, the complementary stretches of primer pairs are underlined. The core T7 promoter is given in *italics*, and a terminal *Bgl* II site is in lowercase letters. Reactions a to e1 correspond to the steps shown in Figure 2.4. Note that when including terminal restriction enzyme recognition sites in the PCR product for cloning purposes, a few flanking nucleotides beyond the restriction sites have to be added for efficient restriction enzyme cleavage after PCR amplification (for details,

see New England Biolabs (NEB) Web site, Technical Reference, and Restriction Endonucleases).

(a) Overlap extension of regions P2, P3, and the 5' part of P1 of the hammerhead cassette (Figure 2.1), including an upstream T7 promoter and a *Bgl* II site

Primer 1 *sense*: 5'-GGa gat ctA ATA CGA CTC ACT ATA GGG AGA AAT CCG CCT GAT GAG-3'

Primer 2 *antisense*: 5'-GAC GGT ACC GGG TAC CGT TTC GTC CTC ACG GAC TCA TCA GGC GGA-3'

PCR profile: 20 cycles: 1 min 94 °C/1 min 40 °C/30 s 72 °C

Resulting sequence: 5'-GGa gat ctA ATA CGA CTC ACT ATA GGG AGA AAT CCG CCT GAT GAG TCC GTG AGG ACG AAA CGG TAC CCG GTA CCG TC-3'; 77 bp

(b) Overlap extension of the HDV ribozyme cassette

Primer 3 *sense*: 5'-GGG TCG GCA TGG CAT CTC CAC CTC CTC GCG GTC CG A CCT GGG CTA-3'

Primer 4 *antisense*: 5'-CTT CTC CCT TAG CCT ACC GAA GTA GCC CAG GTC GGA CCG CGA GGA-3'

PCR profile: 20 cycles: 1 min 94 °C/1 min 60 °C/30 s 72 °C

Resulting sequence: 5'-GGG TCG GCA TGG CAT CTC CAC CTC CTC GCG GTC CGA CCT GGG CTA CTT CGG TAG GCT AAG GGA GAA G-3'; 67 bp

(c) Amplification of the template encoding the RNA of interest (here: yeast tRNA^{Phe}; Figure 2.2a) using primers overlapping with hammerhead and HDV sequence

Primer 5 *sense* (5'-extension into the hammerhead sequence underlined):

5'-GTA CCC GGT ACC GTC GCG GAT TTA GCT CAG-3'

Primer 6 *antisense* (5'-extension into the HDV ribozyme sequence underlined):

5'-TGG AGA TGC CAT GCC GAC CCT GCG AAT TCT GTG G-3'

PCR profile: once 2 min 94 °C, then 30 cycles: 1 min 94 °C/1 min 42 °C/30 s 72 °C

Resulting sequence (regions overlapping with hammerhead and HDV ribozyme sequences are underlined): 5'-GTA CCC GGT ACC GTC GCG GAT TTA GCT CAG TTG GGA GAG CGC CAG ACT GAA GAT CTG GAG GTC CTG TGT TCG ATC CAC AGA ATT CGC AGG GTC GGC ATG GCA TCT CCA-3'; 108 bp

(d) Overlap extension of products from a and c

PCR profile: once 4 min 94 °C, then 10 cycles: 1 min 94 °C/2 min 40 °C/45 s 72 °C

(d1) Addition of primers, product amplification

Primer 7 *sense*: 5'-GGa gat ctA ATA CGA CTC ACT ATA GGG-3'

Primer 6 *antisense*: 5'-TGG AGA TGC CAT GCC GAC CCT GCG AAT TCT GTG G-3'

PCR profile: 30 cycles: 1 min 94 °C/2 min 55 °C/45 s 72 °C

Resulting sequence (hammerhead region and overlap with HDV ribozyme cassette underlined): 5'-GGa gat ct A ATA CGA CTC ACT ATA GGG AGA

AAT CCG CCT GAT GAG TCC GTG AGG ACG AAA CGG TAC C
CG GTA CCG TCG CGG ATT TAG CTC AGT TGG GAG AGC GCC AGA
CTG AAG ATC TGG AGG TCC TGT GTT CGA TCC ACA GAA TTC GCA
GGG TCG GCA TGG CAT CTC CA-3'; 170 bp

- (e) **Overlap extension using the product obtained in d1 (carrying a 3'-extension into the HDV-coding sequence) and the PCR product for the HDV ribozyme cassette (b)**

PCR profile: once 4 min 94 °C, then 10 cycles: 1 min 94 °C/2 min 60 °C/45 s 72 °C

(e1) Addition of primers, product amplification

Primer 7 *sense*: 5'-GGa gat ctA ATA CGA CTC ACT ATA GGG-3'

Primer 4 *antisense*: 5'-CTT CTC CCT TAG CCT ACC GAA GTA GCC CAG GTC GGA CCG CGA GGA-3'

PCR profile: 30 cycles: 1 min 94 °C/2 min 60 °C/45 s 72 °C

Resulting sequence (hammerhead and HDV ribozyme regions underlined):

5'-GGa gat ctA ATA CGA CTC ACT ATA GGG AGA AAT CCG CCT GAT GAG
TCC GTG AGG ACG AAA CGG TAC CCG GTA CCG TCG CGG ATT TAG
CTC AGT TGG GAG AGC GCC AGA CTG AAG ATC TGG AGG TCC TGT
GTT CGA TCC ACA GAA TTC GCA GGG TCG GCA TGG CAT CTC CAC
CTC CTC GCG GTC CGA CCT GGG CTA CTT CGG TAG GCT AAG
GGA GAA G-3'; 217 bp.

3' Dephosphorylation Protocols

As discussed earlier, the activities of both hammerhead and HDV ribozymes lead to the release of RNA molecules that carry 5'-OH and 2', 3'-cyclic phosphate groups at their termini. While the 5'-ends can be phosphorylated by standard T4 PNK procedures, several efficient and robust protocols can be used to remove the 2', 3'-cyclic phosphate group [11]. However, the efficiency of these protocols may vary with the RNA substrate to be dephosphorylated. Hence, it is recommended to test different dephosphorylation procedures if one method does not give satisfying results. An easy test for the removal of the terminal 2', 3'-cyclic phosphate group is to analyze aliquots of the RNA before and after treatment with T4 PNK by denaturing PAGE: the removal of the phosphate group leads to a reduced net charge of the transcript, which, for small RNAs (<100 nt), can be monitored by a lower electrophoretic mobility of the RNA in comparison to the untreated RNA (Chapter 6).

3' Dephosphorylation Protocol 1

Up to 50 pmol of RNA are incubated with 6 U T4 PNK (NEB) in a final volume of 50 µl for 6 h at 37 °C in the following buffer:

- 100 mM Tris/HCl, pH 6.5
- 100 mM magnesium acetate
- 5 mM β-mercaptoethanol.

3' Dephosphorylation Protocol 2

A total of 100 pmol of RNA are incubated with 1 U T4 PNK (NEB) in a final volume of 50 μ l for 6 h at 37 °C in a buffer containing:

- 100 mM imidazole/HCl, pH 6.0
- 10 mM MgCl₂
- 0.1 mM ATP
- 10 mM β -mercaptoethanol
- 20 μ g/ml BSA (RNase-free).

This protocol is also applied as an upscaled version using up to 500 pmol of RNA and 10 U T4 PNK.

3' Dephosphorylation Protocol 3

Up to 300 pmol of RNA are incubated with 10 U T4 PNK (NEB) in a final volume of 20 μ l for 6 h at 37 °C in the following buffer:

- 100 mM 2-(*N*-morpholino)ethanesulfonic acid (MES)/NaOH, pH 5.5
- 300 mM NaCl
- 10 mM MgCl₂
- 10 mM β -mercaptoethanol.

Protocol for RNase P Cleavage

- 5 x RNase P cleavage buffer: 250 mM Tris-HCl, pH 7.5, 0.5 mM EDTA (disodium salt), 0.5 M Mg(OAc)₂, 0.5 M NH₄OAc, and 25% PEG 6000.
- 1) 20 pmol of *E. coli* RNase P RNA (produced by *in vitro* transcription) are preincubated for 5 min at 55 °C and 55 min at 37 °C in 15 μ l of 1 x RNase P cleavage buffer. This preincubation step is not essential when using *E. coli* RNase P RNA, but it increases the proportion of ribozyme molecules competent for substrate binding [32].
 - 2) 10 pmol of an RNA-tRNA primary transcript of the type shown in Figure 2.3 (in 15 μ l 1 x RNase P cleavage buffer) are added to the RNase P RNA preincubation mixture (final volume 30 μ l), followed by incubation for 1.5 h at 37 °C. A 2 : 1 ratio of ribozyme to substrate favors efficient cleavage. However, other ratios, such as 1 : 2 or 1 : 5 (ribozyme : substrate), will work as well and help to save ribozyme material. As an alternative to *E. coli* RNase P RNA, the thermostable counterpart from *T. thermophilus* may be used (see Notes 1 and 2).
 - 3) Load ~3 μ l (50–100 ng substrate RNA) of the processing reaction onto a PAA/8 M urea gel (5–20%, depending on the size of the RNA of interest released by RNase P RNA cleavage) to separate ribozyme and processing products. Also load an equivalent amount of the unprocessed substrate for comparison. Stain with ethidium bromide (EtBr) or SYBR[®] Gold (see Note 4).

- 4) Conduct multiple assays of the type described in steps 1 and 2 or scale up correspondingly. Run a preparative denaturing PAA gel (5–10%), localize the RNA of interest by UV shadowing (detection limit: $\sim 0.3 \mu\text{g}$; for methodological details on UV shadowing, see Chapter 3), and elute from gel slices by passive or electroelution (Chapters 1 and 3). We usually load 20–150 μg of the RNA of interest (between 75 and 400 nt in length) onto the preparative gel (30 cm (width) $\times$ 20 cm (height) $\times$ 1 mm (gel thickness), gel pocket 24 cm wide); 2 μg of a reference RNA loaded into a small pocket (0.7 cm wide) next to the preparative sample are helpful for unambiguous identification of the RNA of interest during UV shadowing.

Notes

- 1) Cleavage by the thermostable RNase P RNA from *T. thermophilus* can be performed in the same buffer as used for the *E. coli* RNA (see above). However, a typical incubation temperature is 55 °C [32]. If cleavage assays with *T. thermophilus* RNase P RNA are preferred to be performed at 37 °C, this ribozyme has to be preincubated at 55 °C (10–20 min), because it is misfolded and inactive after preparation by denaturing PAGE; refolding into its active structure has an activation barrier that is only very slowly overcome at 37 °C but rapidly at 55 °C. Thus, the 55 °C preincubation step is not necessary when cleavage assays are conducted at 55–65 °C.
- 2) For the design of the ribozyme/substrate ratio, it is helpful to consider that the cleavage rate by *E. coli* RNase P RNA under multiple turnover conditions will be limited by product release [25], not exceeding 1 min⁻¹ at 37 °C under the conditions outlined above. However, cleavage rates of up to 25 min⁻¹ have been obtained with *T. thermophilus* RNase P RNA [25] at 55 °C using essentially the same buffer as for *E. coli* RNase P RNA above. Use of a thermostable tRNA moiety as part of the RNA-tRNA substrate, such as the tRNA^{Gly} of *T. thermophilus* [25], will reduce partial tRNA melting during processing at elevated temperature.
- 3) The use of *T. thermophilus* instead of *E. coli* RNase P RNA may also be advantageous in cases where the primary transcript is suspected to form stable structures that impair access of RNase P RNA to the cleavage site. Processing by the thermostable P RNA can be conducted at 55–65 °C, conditions that may ameliorate cleavage site accessibility. Above 70–75 °C, cleavage efficiency is expected to sharply drop because partial melting of thermostable RNase P RNA and tRNA moiety as well as the temperature-dependent loss of ribozyme-substrate affinity will become substantial [25, 26].
- 4) Staining with SYBR Gold (Life Technologies, Invitrogen): note that degradation of the dye after repeated freeze–thaw cycles has been reported; with SYBR Gold, we were able to detect 5–10 ng (~ 1 –2 pmol) of an RNA 14-mer (5'-GUU CGG UCA CGA CU) using a gel documentation system (Biostep DH-50) with EtBr filter (575 nm); for staining with EtBr, at least 100 ng of a short unstructured RNA (such as the aforementioned RNA 14-mer) or 20–50 ng of a longer and more structured RNA of interest should be loaded.

2.5

Potential Problems

Although the described systems do not have any known restrictions concerning the sequence of the RNA of interest, one should keep in mind that some RNA sequences might interfere with the correct folding of the ribozyme cassettes or the linked tRNA molecule. Such misfolding may result in a low reaction efficiency or even in no cleavage at all. As a first approach, a temperature cycling procedure as the one described in Section 2.2.1.2 should be attempted. If unsuccessful, it is advisable to switch from *cis*-cleaving ribozyme cassettes to the *trans*-cleaving RNase P strategy or vice versa. To avoid such failures from the beginning, we advise the reader to scrutinize the structure of the primary transcript by Mfold [33] or RNAfold [34] before experimental work.

References

1. Draper, D.E., White, S.A., and Kean, J.M. (1988) Preparation of specific ribosomal RNA fragments. *Methods Enzymol.*, **164**, 221–237.
2. Milligan, J.F. and Uhlenbeck, O.C. (1989) Synthesis of small RNAs using T₇ RNA polymerase. *Methods Enzymol.*, **180**, 51–62.
3. Pleiss, J.A., Derrick, M.L., and Uhlenbeck, O.C. (1998) T₇ RNA polymerase produces 5' end heterogeneity during *in vitro* transcription from certain templates. *RNA*, **4**, 1313–1317.
4. Helm, M., Brulé, H., Giegé, R., and Florentz, C. (1999) More mistakes by T₇ RNA polymerase at the 5' ends of *in vitro*-transcribed RNAs. *RNA*, **5**, 618–621.
5. Birikh, K.R., Heaton, P.A., and Eckstein, F. (1997) The structure, function and application of the hammerhead ribozyme. *Eur. J. Biochem.*, **245**, 1–16.
6. Clouet-d'Orval, B. and Uhlenbeck, O.C. (1997) Hammerhead ribozymes with a faster cleavage rate. *Biochemistry*, **36**, 9087–9092.
7. Stage-Zimmermann, T.K. and Uhlenbeck, O.C. (1998) Hammerhead ribozyme kinetics. *RNA*, **4**, 875–889.
8. Persson, T., Hartmann, R.K., and Eckstein, F. (2002) Selection of variants with low Mg²⁺ requirement: importance of stem-loop II. *Chembiochem*, **3**, 1066–1071.
9. Been, M.D., Perrotta, A.T., and Rosenstein, S.P. (1992) Secondary structure of the self-cleaving RNA of hepatitis delta virus: applications to catalytic RNA design. *Biochemistry*, **31**, 11843–11852.
10. Chowrira, B.M., Pavco, P.A., and McSwiggen, J.A. (1994) *In vitro* and *in vivo* comparison of hammerhead, hairpin, and hepatitis delta virus self-processing ribozyme cassettes. *J. Biol. Chem.*, **269**, 25856–25864.
11. Schürer, H., Lang, K., Schuster, J., and Mörl, M. (2002) A universal method to produce *in vitro* transcripts with homogeneous 3' ends. *Nucleic Acids Res.*, **30**, e56.
12. Wagner, S.C., Avis, J.M., and Conn, G.L. (2003) General plasmids for producing run-off *in vitro* transcripts with homogeneous ends. *Nucleic Acids Res.*, **31**, e82.
13. Peattie, D.A. (1979) Direct chemical methods for sequencing RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 1760–1764.
14. Cameron, V. and Uhlenbeck, O.C. (1977) 3'-Phosphatase activity in T₄ polynucleotide kinase. *Biochemistry*, **16**, 5120–5126.
15. Povirk, L.F. and Steighner, R.J. (1990) High ionic strength promotes selective 3'-phosphatase activity of T₄ polynucleotide kinase. *Biotechniques*, **9**, 562.
16. Ferré-D'Amaré, A.R. and Doudna, J.A. (1996) Use of *cis*- and *trans*-ribozymes to

- remove 5' and 3' heterogeneities from milligrams of *in vitro* transcribed RNA. *Nucleic Acids Res.*, **24**, 977–978.
17. Ziehler, W.A. and Engelke, D.R. (1996) Synthesis of small RNA transcripts with discrete 5' and 3' ends. *Biotechniques*, **20**, 622–624.
 18. Collins, R.A. (2002) The *Neurospora* Varkud satellite ribozyme. *Biochem. Soc. Trans.*, **30**, 1122–1126.
 19. Wichlacz, A., Legiewicz, M., and Ciesiolka, J. (2004) Generating *in vitro* transcripts with homogenous 3' ends using *trans*-acting antigenomic delta ribozyme. *Nucleic Acids Res.*, **32**, E39.
 20. Kirsebom, L.A. (2002) RNase P RNA-mediated catalysis. *Biochem. Soc. Trans.*, **30**, 1153–1158.
 21. Oh, B.K., Frank, D.N., and Pace, N.R. (1998) Participation of the 3'-CCA of tRNA in the binding of catalytic Mg²⁺ ions by ribonuclease P. *Biochemistry*, **37**, 7277–7283.
 22. Busch, S., Kirsebom, L.A., Notbohm, H., and Hartmann, R.K. (2000) Differential role of the intermolecular base-pairs G292-C(75) and G293-C(74) in the reaction catalyzed by *Escherichia coli* RNase P RNA. *J. Mol. Biol.*, **299**, 941–951.
 23. Loria, A. and Pan, T. (1998) Recognition of the 5' leader and the acceptor stem of a pre-tRNA substrate by the ribozyme from *Bacillus subtilis* RNase P. *Biochemistry*, **37**, 10126–10133.
 24. Brännvall, M., Pettersson, B.M., and Kirsebom, L.A. (2003) Importance of the +73/294 interaction in *Escherichia coli* RNase P RNA substrate complexes for cleavage and metal ion coordination. *J. Mol. Biol.*, **325**, 697–709.
 25. Hardt, W.D., Schlegl, J., Erdmann, V.A., and Hartmann, R.K. (1995) Kinetics and thermodynamics of the RNase P RNA cleavage reaction: analysis of tRNA 3'-end variants. *J. Mol. Biol.*, **247**, 161–172.
 26. Hartmann, R.K. and Erdmann, V. (1991) Analysis of the gene encoding the RNA subunit of ribonuclease P from *T. thermophilus* HB8. *Nucleic Acids Res.*, **19**, 5957–5964.
 27. Coleman, T.M., Wang, G., and Huang, F. (2004) Superior 5' homogeneity of RNA from ATP-initiated transcription under the T₇ phi 2.5 promoter. *Nucleic Acids Res.*, **32**, e14.
 28. Santoro, S.W. and Joyce, G.F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 4262–4266.
 29. Santoro, S.W. and Joyce, G.F. (1998) Mechanisms and utility of an RNA-cleaving DNA enzyme. *Biochemistry*, **37**, 13330–13342.
 30. Cairns, M.J., King, A., and Sun, L.Q. (2003) Optimisation of the 10-23 DNAzyme-substrate pairing interactions enhanced RNA cleavage activity at purine–cytosine target sites. *Nucleic Acids Res.*, **31**, 2883–2889.
 31. Shevchuk, N.A., Bryksin, A.V., Nusinovich, Y.A., Cabello, F.C., Sutherland, M., and Ladisch, S. (2004) Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. *Nucleic Acids Res.*, **32**, E19.
 32. Hardt, W.D., Schlegl, J., Erdmann, V.A., and Hartmann, R.K. (1993) Gel retardation analysis of *E. coli* M1 RNA-tRNA complexes. *Nucleic Acids Res.*, **21**, 3521–3527.
 33. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
 34. Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., and Schuster, P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, **125**, 167–188.

3

RNA Ligation

Janne J. Turunen, Liudmila V. Pavlova, Martin Hengesbach, Mark Helm,
Sabine Müller, Roland K. Hartmann, and Mikko J. Frilander

3.1

General Introduction

The most efficient RNA synthesis method used today continues to be *in vitro* run-off transcription by T7 RNA polymerase (Chapter 1). Although this technique produces large amounts of RNA at relatively low cost with reasonable effort, it suffers from essentially two drawbacks: one is the potential 5'- and/or 3'-end heterogeneity of the product and the other is the inability to introduce internal modifications at specific sites. End heterogeneity can be overcome by several approaches described in detail in this chapter and Chapter 2, but for site-specific incorporation of nucleotide modifications, chemical RNA synthesis is, in most instances, inevitable [1].

Development of new commercially available phosphoramidites of natural and unnatural nucleosides, improvement of 2'-OH protecting groups, and the use of more efficient activators are innovations that have paved the way for chemical synthesis of RNA molecules in better and more reproducible yields ([2–5]; see also Chapters 7 and 8). Despite these advances, substantial amounts of longer RNA molecules are still very difficult and expensive to synthesize. At present, the efficiency of chemical RNA synthesis is such that high-quality RNA molecules up to 80 nt long can be obtained from commercial sources. As the repertoire of various phosphoramidite monomers used in chemical RNA synthesis is constantly growing, RNA oligonucleotides containing modified nucleotides are frequently used in the construction of long chimeric RNA molecules. In such cases, the RNA oligonucleotides containing the desired modification are ligated to other RNA molecules, which are often produced by *in vitro* transcription reactions. This combination provides a way to modify functional groups in virtually any position in a given RNA molecule. As the product yield of RNA ligation reactions is usually far below 100%, the approach should be carried out only if the RNA of interest cannot be obtained as one strand by chemical synthesis. Such cases include (i) the requirement for site-specifically modified RNA molecules that are longer than what can be synthesized chemically; (ii) the introduction of internal radioactive groups; (iii) the need for, for example, mRNAs with internal site-specific modifications,

whose 5'-portion is prepared by T7 transcription to introduce a 5'-cap; or (iv) the introduction of several modifications in a combinatorial manner to produce differentially modified variants of the same RNA.

3.1.1

T4 Polynucleotide Ligases

The genome of bacteriophage T4 encodes three polynucleotide ligases that seal single and double strand breaks of nucleic acids during infection of host bacteria: T4 DNA ligase (T4 Dnl), T4 RNA ligase 1 (T4 Rnl1), and T4 RNA ligase 2 (T4 Rnl2). All three ligases use 5'-phosphate and 3'-hydroxyl end groups for ligation and require ATP as a cofactor. However, they differ in their nucleic acid substrate preferences. For example, T4 Rnl2, in contrast to T4 Rnl1, is much more active in joining nicks in dsRNA *versus* ssRNA [6]. However, both T4 Rnl1 and T4 Rnl2 also efficiently circularize linear ssRNA [7]. Figure 3.1 summarizes the substrate type and ligation efficiency for the three T4 ligases, including engineered variants of T4 Rnl2 (all available from NEB). T4 Rnl2 “Truncated K227Q” is a point mutant of T4 Rnl2 “Truncated” that further reduces the formation of unwanted ligation by-products in reactions using preadenylated DNA adaptor oligonucleotides for ligation to RNA 3'-ends in directional cDNA library constructions (see Chapters 38 and 39 for application). The preadenylated adapter mimics the adenylated donor RNA in Figure 3.2A (step 2).

An interesting alternative to the T4 RNA ligases is the thermostable RNA ligase 1 from *Thermus scotoductus* bacteriophage TS2126, which has a temperature optimum of 60–70 °C [12] and is commercially available from Epicentre. In some cases, strong RNA secondary structures can reduce the accessibility of the RNA termini, thus inhibiting the ligation reaction. Use of the thermostable RNA ligase 1 at elevated ligation temperature may help overcome such problems. However, the suitability of this enzyme for preparative scale work is not yet known.

3.1.2

Reaction Mechanism

All three phage T4 ligases follow the same basic enzyme mechanism [6], which is illustrated for T4 Rnl1 in Figure 3.2. The reaction consists of three distinct and reversible steps (Figure 3.2A; [8]). In the first step, the ligase reacts with ATP to form an adenylated enzyme intermediate. In the second step, a donor substrate with a 5'-terminal monophosphate is bound by the enzyme, and the enzyme-linked 5'-AMP moiety is now transferred to the 5'-phosphate of the donor substrate, yielding the adenylated donor product in which the terminal adenosine moiety is attached *via* a 5', 5'-diphosphate bridge $[A(5')pp(5')Np(Np)_n]$. In the final transesterification step, the phosphodiester bond connecting the two 5', 5'-linked phosphates is cleaved and a phosphodiester bond is formed between the donor and acceptor substrate, with concomitant release of AMP.

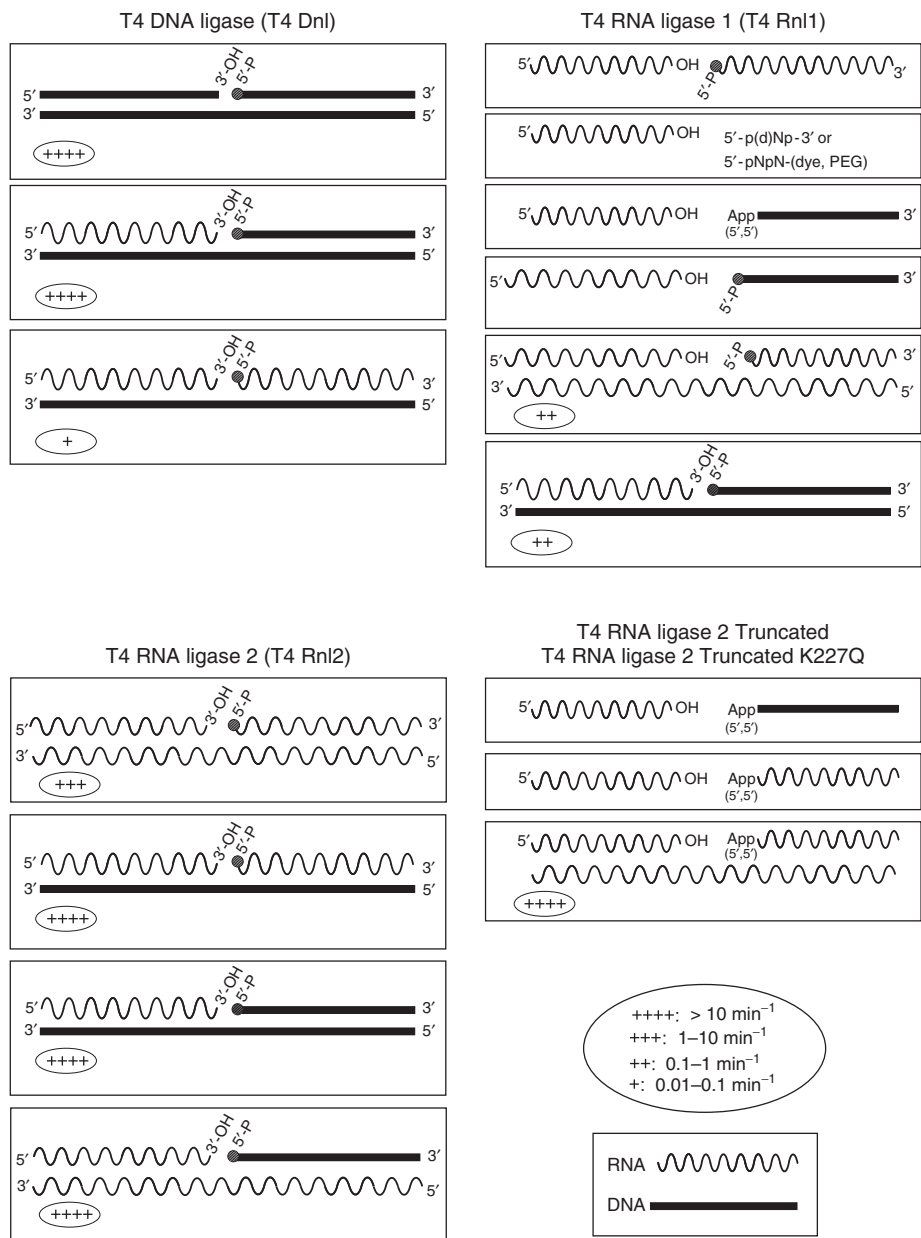
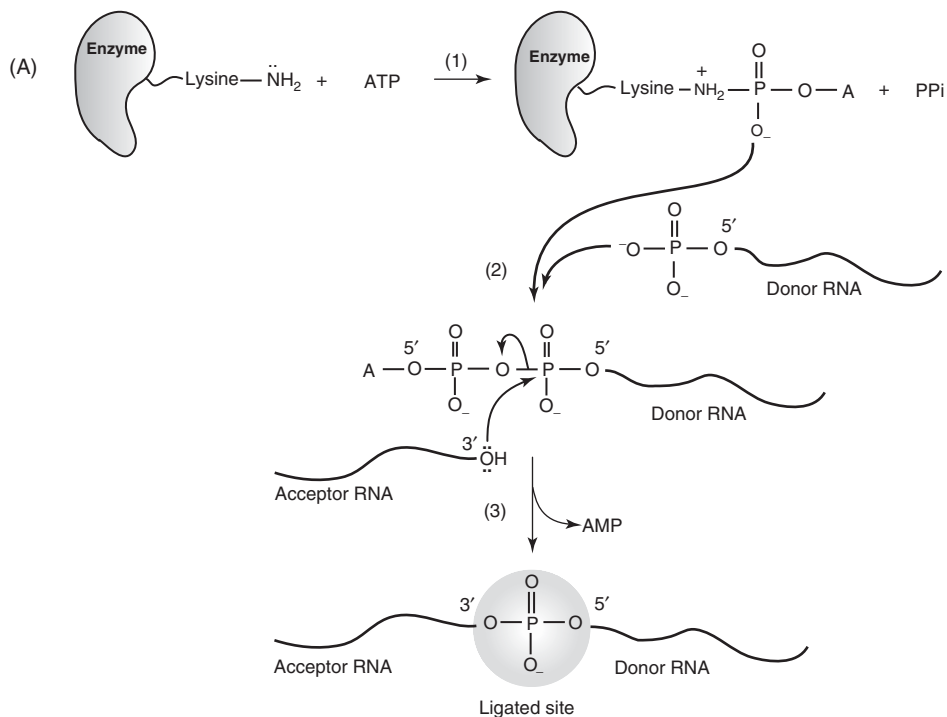
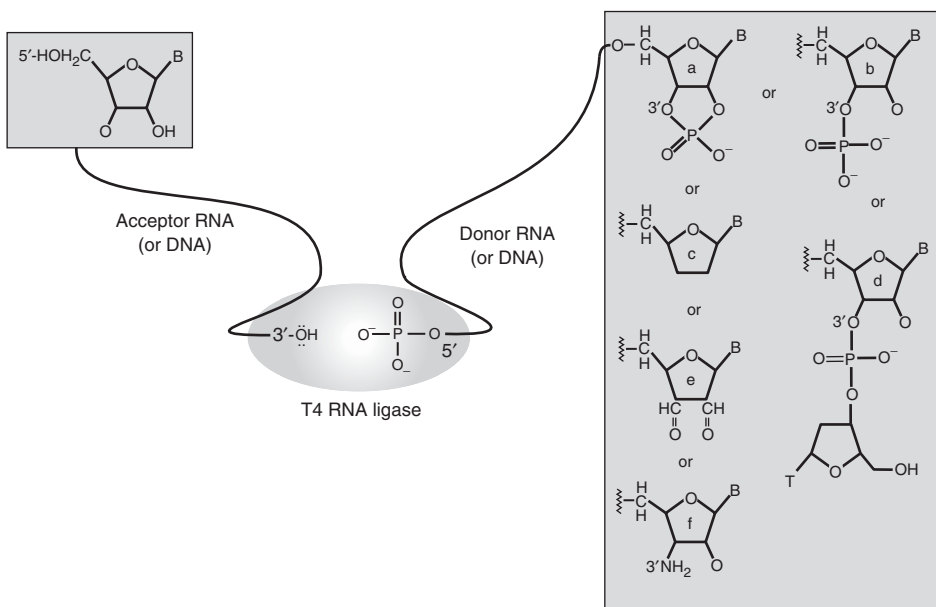


Figure 3.1 Substrate specificity of T4 DNA ligase (T4 Dnl), T4 RNA ligases 1 and 2 (T4 Rnl1, T4 Rnl2), and T4 RNA ligases 2 Truncated and Truncated K227Q (modified based on information provided on the NEB web page). The efficiencies (ligation rates) for

double-stranded substrates (Source: Taken from Ref. [6]) are indicated by plus signs within ovals; the correlation with ligation rates is specified in the lower right part of the table.



(B)



3.1.3

Advantages of T4 DNA Ligase for RNA Ligation

Several features of T4 Dnl have made it a popular enzyme for the construction of long RNA molecules. As a consequence of the strict requirement for an uninterrupted and unbulged double-helical structure, end-joining by T4 Dnl only occurs if the two substrate RNA molecules have the correct termini at the junction, thus avoiding the formation of aberrant ligation products with nucleotide deletions or insertions (Figure 3.3c). It also does not display any sequence specificity on either donor or acceptor molecules. Furthermore, the use of T4 Dnl does not lead to an unwanted formation of circular RNA structures that can be a problem when using T4 RNA ligases with RNA molecules containing unprotected 3'- or 5'-ends (Section 3.4). Interestingly, the relatively recently discovered T4 Rnl2 is more efficient in ligating short RNA substrates annealed to a DNA splint than T4 Dnl ([6]; see also Figure 3.1). Although our current knowledge of the function of T4 Rnl2 on long RNA substrates is limited, it is possible that it could substitute T4 Dnl as the enzyme of choice in certain procedures (see also Section 3.3.2).

3.1.4

Chapter Structure

After a general methodological part, we provide the basic protocols for RNA ligation by T4 Dnl, followed by specialized protocols for the simultaneous ligation of five RNA oligonucleotides. The latter type of protocol has been developed for the

Figure 3.2 (A) Mechanism of the ligation reaction catalyzed by T4 Rnl1. In the ligation reaction, an acceptor substrate with a 3'-hydroxyl group reacts with a donor substrate carrying a 5'-phosphate group, resulting in a 3'-5'-phosphodiester bond. The reaction cycle consists of three distinct and reversible steps [8]: initially, the ligase reacts with ATP to form an adenylated enzyme intermediate (at lysine 99; [9, 10]), with concomitant release of pyrophosphate (1). Then a donor substrate with a 5'-terminal monophosphate is bound by the enzyme and converted to the adenylated donor $A(5')pp(5')Np(Np)_n$, an intermediate in which the terminal adenosine moiety is attached *via* a 5',5'-diphosphate bridge to the donor RNA (2). In a final transesterification step, the phosphodiester bond connecting the two 5',5'-linked phosphates is cleaved and a phosphodiester bond is formed between the donor and acceptor

substrate, with concomitant release of AMP (3). (B) End groups of donor and acceptor substrates that prevent the formation of alternative products (intramolecular donor or acceptor cyclization, or formation of donor or acceptor tandems) in intermolecular ligation reactions of two oligo(ribo)nucleotides. In the case of the acceptor substrate, a 5'-OH terminus prevents this substrate from acting as a donor. A 3'-OH terminus is mandatory for acceptor function, whereas blockage of this end group in the case of donor substrates precludes their function as an acceptor; 3'-end blockage is achieved by introducing a terminal 2',3'-cyclic phosphate (a), 3'-phosphate (b), a dideoxy residue (c; [11]), a 3'-inverted deoxythymidine (d; modification available from Thermo Scientific Dharmacon, Colorado, USA), a periodate-oxidized 3'-terminal ribose (e; for details, see Chapter 6), or a 3'-amino group (f; see NEB web page).

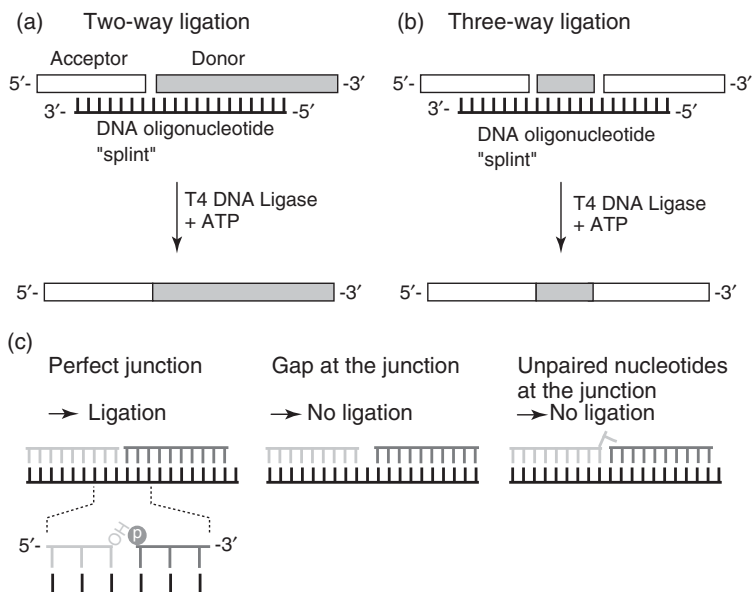


Figure 3.3 The principle of the RNA ligation with T4 DNA ligase (T4 Dnl). A DNA splint oligonucleotide (black) hybridizes with two (a – two-way ligation) or three (b – three-way ligation) RNA molecules (gray and white) and forms a double-helical structure. (c) Ligation requires 3'-OH

and 5'-monophosphate on acceptor and donor molecules, respectively. A gap in the double-stranded helix structure or unpaired nucleotides at the junction (such as $n + 1$ products resulting from T7 transcription) will inhibit the ligation.

assembly of 75–90 nt long RNAs, such as tRNAs, carrying two or more site-specific modifications, for example, for fluorescence resonance energy transfer (FRET) analyses. The last part of the chapter focuses on the widely applied T4 Rnl1, including its enzymatic features, scope of substrates, and ligation protocols.

3.2

RNA Ligation Using T4 DNA Ligase (T4 Dnl)

RNA ligation methods for generating site-specifically modified long RNA molecules using the T4 Dnl were initially described some 20 years ago [13]. More recently, this method has been widely used to provide chimeric RNAs to study various RNA–RNA and RNA–protein interactions in diverse biochemical reconstitution systems and also in live cells (for example, in *Xenopus* oocytes). The modifications introduced include simple insertion of a single radioactive group at a specific location of an RNA molecule, and more complicated alterations in which different nucleotide analogs, crosslinking groups, or RNA backbone modifiers have been inserted into long RNA molecules at specific positions [14–22].

3.2.1

Overview of the RNA Ligation Method Using the T4 DNA Ligase (T4 Dnl)

Despite the seemingly simple overall reaction, that is, the joining of two or three RNA molecules together with the aid of T4 Dnl, the execution of a ligation experiment can be a laborious task if large amounts of the ligated products are needed. As several excellent reviews on this subject have already been published [23–25], we only give a brief overview on the basic principles of the approach before detailing the practical steps involved in RNA ligation by T4 Dnl. In addition, we describe special methods to generate high-quality *in vitro* transcripts with homogeneous 5'- and 3'-termini, which are required for efficient large-scale production of chimeric RNAs. Other widely used strategies for the generation of homogeneous 5'- and 3'-ends are detailed in Chapter 2.

Although T4 Dnl is ordinarily used to ligate DNA molecules, it can also catalyze the formation of a phosphodiester bond between two RNA molecules or between RNA and DNA molecules, as its original name “T4 polynucleotide ligase” indicates [26]. The principle of RNA ligation by T4 Dnl is depicted in Figure 3.3. Typical applications are the so-called two-way (Figure 3.3a) and three-way (Figure 3.3b) ligations, in which either two or three RNA pieces, respectively, are joined together. Normally, one of the pieces contains the desired modification(s), while the other(s) are used to reconstitute the full-length RNA molecule under study. The RNA pieces to be ligated are aligned and held together with a complementary bridging DNA oligonucleotide, also known as *DNA splint* or *cDNA template*. The ligase catalyzes phosphodiester bond formation between the 5'-phosphate of the donor (3'-substrate RNA) and the 3'-hydroxyl of the acceptor (5'-substrate RNA). Therefore, RNAs containing, for example, a 3'-phosphate or a 5'-triphosphate are not ligated. Furthermore, the RNA/DNA double helix formed by the two RNA pieces and the DNA splint has to be consecutive without any bulges or gaps, especially at the point of the junction of the two RNA molecules.

A simplified flowchart of a three-way ligation experiment is presented in Figure 3.4 to illustrate the typical steps needed for the preparation of individual RNA pieces for the RNA ligation. In this example, the modified nucleotides were incorporated into the central piece obtained by chemical synthesis. The 5'-piece, initiated with a 5'-terminal GpppG cap, and the 3'-piece were produced by *in vitro* transcription using T7 RNA polymerase. In this scheme, the preparation of the 5'-piece is relatively simple, while the 3'-piece requires either T7 transcription in the presence of an excess of GMP over GTP (Chapter 1), or further transcript processing to generate a 5'-monophosphate and/or correct 5'-terminal sequence (dephosphorylation followed by phosphorylation to generate 5'-monophosphate termini, or alternatively site-specific cleavage with RNase H). As the quality of the chemically synthesized pieces is typically sufficient for use without further purification, we concentrate on the production of high-quality transcripts by T7 RNA polymerases in the following sections.

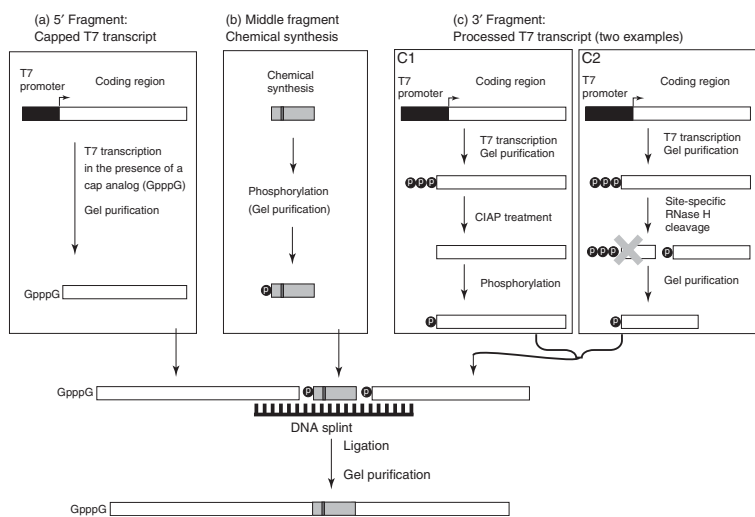


Figure 3.4 An example of steps needed for the production of RNA pieces for a three-way ligation with T4 DNA ligase (T4 Dnl). (a) 5'-Fragment: capped RNA is produced by transcription with T7 RNA polymerase. (b) Middle fragment: RNA oligonucleotide, containing a modified nucleotide (black stripe) is produced by chemical synthesis. (c) 3'-Fragment: two alternative examples are presented. (C1) The 3'-fragments are produced by T7 transcription, followed by dephosphorylation (to remove the 5'-triphosphate) and phosphorylation (to add a single phosphate to the 5'-terminus). (C2) The 3'-fragment is initially produced as a longer precursor, which is subsequently cleaved and gel-purified.

3.2.2

Large-Scale Transcription and Purification of RNAs

When preparing RNA for the ligation, it is necessary to start with a relatively large initial amount of RNA to account for the unavoidable losses at the various stages of the procedure (substrate manipulations, RNA ligation itself, and the subsequent purification steps). A practical rule of thumb used in the Frilander laboratory is to start with at least a 10-fold excess of each individual RNA piece compared to the amount of the ligated product needed in the final experiments. As described in Chapter 1, transcription by T7 phage RNA polymerase can be used to generate large quantities of RNAs from defined DNA templates containing phage-specific promoters. The templates can be linearized plasmids or PCR products or even annealed oligonucleotides [27, 28]. In ligation experiments, the use of PCR products is preferred because they provide an easy way to specify nucleotides at the point of junction of the two RNAs to be ligated. Here we describe conditions for generating large amounts of RNA (several nanomoles) from a single transcription reaction (Protocol 1). This reaction is suitable for RNA molecules that fit the T7 promoter consensus, where at least the first transcribed nucleotide, or if possible, both the first and the second nucleotides should be G residues. If other than a G residue is required as the initial nucleotide, the transcript can be initially produced as a longer precursor, which is subsequently cleaved at a specific site by a ribozyme (Chapter 2) or RNase H (Section 3.2.4). Alternatively, the transcription can be primed with a suitable dinucleotide [13, 29, 30].

Using Protocol 1, we have obtained yields of up to 4–5 nmol of gel-purified RNA based on a 200 μ l T7 transcription reaction. The reaction conditions described in Protocol 1 are not recommended for the production of capped RNAs, as the high concentrations of divalent cations and spermidine tend to precipitate the cap analogs. Instead, modified conditions described in Protocol 2 should be used for the transcription of capped RNAs.

After transcription, the DNA template is degraded with DNase to ensure that the contaminating template DNA (which will have almost the same mobility in the gel as the transcript itself) will not interfere with further steps of the ligation procedure. Subsequently, the full-length transcript RNA will be purified from prematurely terminated products by denaturing gel electrophoresis. Following electrophoresis, the bands are visualized using the “UV shadow” technique, excised, and eluted by passive diffusion (Protocol 3).

3.2.3

Generating Homogeneous Acceptor 3'-Ends for Ligation

PCR-based template generation followed by transcription with phage polymerases is a simple and efficient method for generating large amounts of RNA fragments for ligation purposes. However, a problem with the phage polymerases is that they can add nontemplated nucleotides to the 3'-end of the synthesized RNA molecule. In the worst cases, more than 50% of the synthesized RNA molecules can contain

these so-called $n + 1$ and $n + 2$ nt [27, 31]. As T4 Dnl requires an absolute match between the DNA splint and the two RNA molecules to be ligated, the nontemplated addition of extra nucleotides can significantly reduce the efficiency of the RNA ligation. This is not necessarily a severe problem in small-scale or initial screening experiments, in which limited amounts of the ligated products are often sufficient. However, if large amounts of the ligated products are needed, inefficient ligation may become a major limitation. In such cases, methods producing RNAs with homogeneous 3'-termini can often lead to a severalfold increase in the quantity of the ligated products.

A simple method for reducing nontemplated nucleotide addition has been described by Kao *et al.* [32]. In this method, the very 5'-end of the downstream PCR primer used in the synthesis of the template DNA for T7 transcription is modified: instead of standard deoxyribonucleotides, it contains two 2'-O-methyl RNA residues (Figure 3.5a). During the PCR reaction, the primers are incorporated into the synthesized DNA fragments, which subsequently serve as templates in transcription by T7 RNA polymerase. During transcription, the modified nucleotides at the 5'-end of the template strand will lead to a significant reduction of the nontemplated addition of extra nucleotides to the 3'-end of transcripts [32]. We have used this method successfully and have been able to raise the ligation efficiency from 20 to 70% with no other alterations in the procedure (Figure 3.5b). We have purchased our oligonucleotides from Sigma-Aldrich (www.sigmaaldrich.com), from Dharmacon Research (www.dharmacon.com), or from the Keck oligonucleotide synthesis facility at Yale University (<http://medicine.yale.edu/keck/oligo/>).

Another way of creating homogenous 3'-ends is to synthesize the RNA as a longer precursor and to cut it into the desired length using site-directed cleavage by RNase H or a *cis*- or *trans*-acting ribozyme. Ribozyme cleavage is described in Chapter 2, while the site-specific RNase H cleavage, which can be used to trim both 3'- and 5'-end of the RNA molecule, is described in detail in Section 3.2.4.

3.2.4

Site-Directed Cleavage with RNase H

At times, it may not be possible to transcribe the desired RNA directly, for example, when the first nucleotide is not a guanosine. One possible solution for this is the site-specific cleavage of a longer precursor RNA molecule using RNase H and chimeric 2'-O-methyl RNA/DNA oligonucleotides. This method may also be used to solve problems with 5'- or 3'-end heterogeneity.

RNase H recognizes and binds nucleic acids that are duplexes of DNA and RNA and cleaves the backbone of the RNA strand leaving a 5'-phosphate and a 3'-hydroxyl [34]. The site of cleavage may be specified when using oligonucleotides containing a short DNA stretch (3 or 4 nt), which is flanked by 2'-O-methyl-RNA sequences [35, 36]. We have successfully used hybrid 20-mer oligonucleotides that contain three 2'-O-methyl RNA residues at the 5'-end, followed by four DNA residues and thirteen 2'-O-methyl residues (Figure 3.5c). An important observation to be noted [33] is that the exact position of cleavage is, for unknown reasons, dependent on the

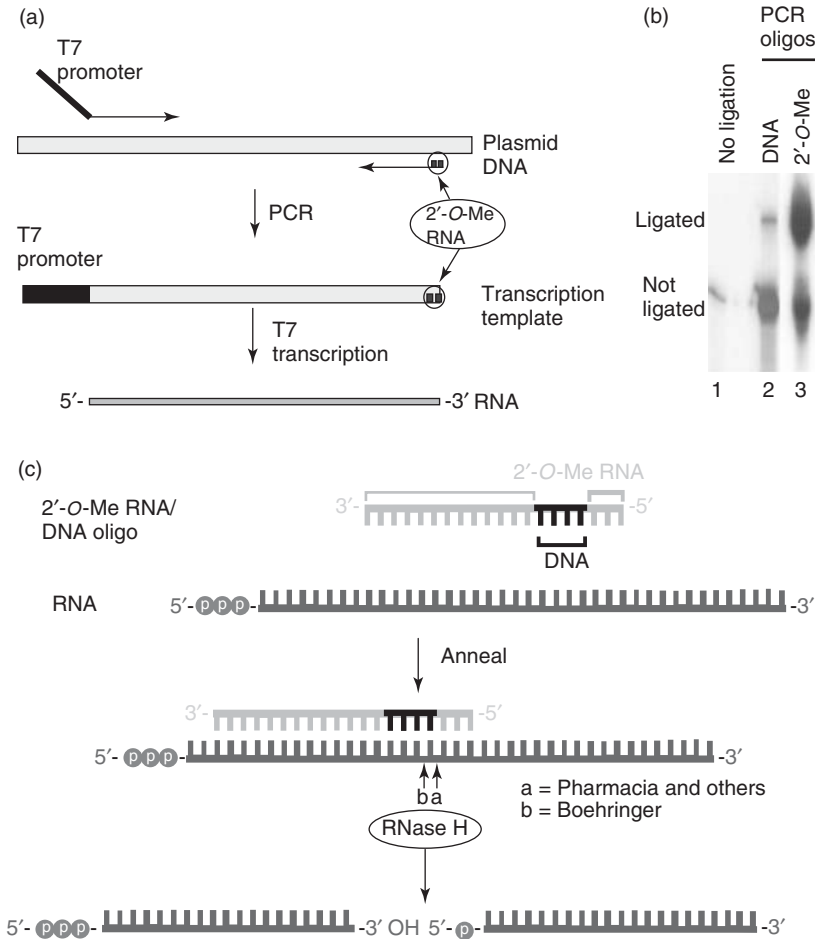


Figure 3.5 (a) Production of T7 transcripts with homogenous 3'-ends. The template for T7 transcription is produced by PCR. The upstream primer contains a promoter for T7 RNA polymerase, while the downstream primer contains two 2'-O-methyl RNA residues at the 5'-end of the oligonucleotide. Both primers are incorporated into the PCR product, and during transcription with T7 RNA polymerase, the 2'-O-methyl RNA residues prevent the addition of nontemplated nucleotides to the 3'-end of the RNA molecules. (b) Comparison of the ligation efficiencies when a standard DNA oligonucleotide and a hybrid DNA/2'-O-Methyl RNA oligonucleotide (depicted in panel A) were used as the

downstream primer in the PCR reaction to produce a template for T7 transcription. Lane 1: control lane containing a ^{32}P -labeled donor molecule; lane 2: ligation with an RNA fragment derived from transcription using an all-DNA template; lane 3: ligation with an RNA fragment derived from transcription using a template with two 5'-terminal 2'-O-methyl modifications. In the ligation reactions (lanes 2 and 3), the donor RNA fragments were radioactively labeled, while the acceptor RNAs were unlabeled. (c) The principle of site-specific RNase H cleavage. The cleavage sites of Pharmacia/Sigma/Takarashuzo/USB/GE Healthcare (a) or Boehringer Mannheim (b) RNase H [33] are indicated.

commercial source of the RNase H. The enzymes previously supplied by Pharmacia, Sigma, and Takarashuzo were reported to cleave the phosphodiester bond that is located 3' to the ribonucleotide base-paired with the 5'-most deoxyribonucleotide of the oligonucleotide, whereas the enzyme from Boehringer Mannheim cleaved the bond located one nucleotide upstream (5'-direction) in the RNA molecule (Figure 3.5c). After the report was made, however, there have been mergers of the aforementioned companies, and it is unclear which enzyme sources and purification protocols are now used by the merged companies. In our studies, we have used RNase H supplied by USB or GE Healthcare (formerly Amersham Biosciences) and found that they function as the one supplied earlier by Pharmacia. Alternatively, we have also used a His-tagged RNase H expressed in a standard *Escherichia coli* strain and purified in our laboratory. If other enzyme sources are used, it is advisable to map the exact cleavage site by primer extension analysis. If available, high-concentration RNase H should preferably be used, as large amounts of glycerol may inhibit the reaction. It should be noted, however, that a large excess of RNase H should be avoided, as the enzyme has some unspecific RNase activity.

Protocol 4 describes a general cleavage strategy using RNase H and hybrid DNA/2'-O-methyl RNA oligonucleotides. In the first step, the RNA and the hybrid oligonucleotide are allowed to anneal, after which the appropriate buffers and enzymes are added. For efficient annealing, the RNA and hybrid oligonucleotide should be initially denatured completely by heating them to 95 °C and then, by lowering the temperature slowly, allowed to anneal. Addition of a monovalent salt, such as KCl, further enhances the annealing, but the concentration in the final cleavage reaction should not exceed 50 mM. Small amounts of EDTA are included to chelate any traces of divalent cations, for example, Mg^{2+} , to reduce chemical degradation of the RNA at high temperatures. The amounts of RNA and hybrid oligonucleotide can be adjusted to suit the particular experiment, but the reaction should always contain close to equimolar amounts of the oligonucleotide relative to the RNA to be cleaved. Large excess of the hybrid oligonucleotide may lead to aberrant cleavage at additional sites. We typically use a 10–15% excess of the oligonucleotide to ensure efficient annealing. If the 5'-end of the fragment is to be dephosphorylated for a subsequent labeling with a radioactive phosphate, the final yield of the cleaved RNA product may be increased by carrying out the dephosphorylation before gel purification.

3.2.5

Dephosphorylation and Phosphorylation of RNAs

The 5'-triphosphate resulting from the transcription reaction (Figure 3.4c) must be converted to a 5'-monophosphate if the RNA is to be used as a donor RNA in the ligation reaction. This is achieved by first dephosphorylating the RNA and then phosphorylating the resulting 5'-hydroxyl with an unlabeled or radioactively labeled phosphate. The dephosphorylation catalyzed by calf intestinal alkaline phosphatase (CIAP) is carried out at 50 °C (Protocol 5). The elevated temperature is used to reduce the effect of RNA secondary structure on dephosphorylation. Following

the dephosphorylation, it is necessary to completely remove the CIAP, as it could seriously inhibit further downstream steps. As very large amounts of CIAP are used, we typically remove it by proteinase K digestion. It is also possible to use other phosphatases (such as SAP – shrimp alkaline phosphatase), which are easier to inactivate compared to CIAP. However, our experience with SAP is that at least with some substrates it does not work as efficiently as CIAP.

As mentioned earlier, the donor RNA must have a 5'-phosphate for successful ligation. Dephosphorylated transcripts or chemically synthesized oligonucleotides can be phosphorylated by T4 polynucleotide kinase (PNK) in the presence of ATP (note that RNA oligonucleotides can also be phosphorylated during the chemical synthesis). T4 PNK can also be used in site-specific labeling to insert a single radioactive phosphorus at the junction between the acceptor and the donor. Protocol 6 describes phosphorylation with [γ - ^{32}P]ATP resulting in a specific activity of approximately $0.5 \times 10^6 \text{cpm pmol}^{-1}$. The protocol can also be used for nonradioactive phosphorylation to create the 5'-phosphate required for ligation, in which case 500 μM of unlabeled ATP should be used.

3.2.6

RNA Ligation

In the first step, the RNA fragments are aligned together with a bridging oligonucleotide, also known as the *DNA splint* or *cDNA template*. In the second step, the reaction mix is added and the RNA ends at the junction are joined by T4 Dnl in the presence of ATP. The major consideration when planning both steps is to know the molar concentrations of each RNA piece and the DNA splint oligonucleotide. Furthermore, the ligation volume should be kept as small as possible, as the ligation proceeds more efficiently when the reactants are concentrated. The final volume of the reaction should therefore be kept at approximately 10–20 μl . Finally, to help the detection of the ligated products in purification gels and their subsequent quantification one should always include trace amounts of radioactively labeled RNAs in the ligation reaction. This should be done even when the aim is to produce unlabeled chimeric RNAs.

Splints spanning 20 nt on both sides of the junction align the substrates efficiently, although splints down to about 20 nt (10 on each side) can be used [23]. For efficient ligation, the three polynucleotides should be present in equal molar amounts. If one of the RNA fragments is scarce, the splint and the other fragment may be added in excess to drive the reaction. Adding splint and donor RNA in excess can also be used to increase the yield if the 3'-end of the acceptor RNA is heterogeneous. However, it is important that the concentration of at least one of the RNA fragments is greater than that of the DNA splint. If excess amounts of the splint are used, the individual RNA fragments may hybridize to different splint molecules and thus be separated from their ligation partners.

Some monovalent salt and EDTA may be added to the annealing reaction, for the same reasons as previously described for annealing before RNase H cleavage (Section 3.2.4). Performing the annealing in a thermal cycler with a heated lid

prevents condensation of water on the lid of the tube, which may even lead to sample drying in the case of small volumes.

The buffer used for the ligation reaction may be made by the researcher, as the one described in Protocol 7, or a commercial one supplied with the enzyme can be used. Macromolecular crowding agents, such as polyvinyl alcohol (PVA), polyethylene glycol (PEG), or polyvinyl pyrrolidone (PVP) should be added to increase the effective concentration of the reactants. Some thought should be lent to the amount of ligase used, as it has been reported that T4 Dnl does not turn over efficiently on RNA-containing duplexes [23]. Therefore, a stoichiometric amount of T4 Dnl should be used, with one Weiss unit corresponding to approximately 1 pmol.

Finally, the incubation time and temperature should be considered. Incubating the reaction at 30 °C for 4 h is a widely used approach. We have also performed the incubation overnight at room temperature (~20–25 °C), and this seems to result in somewhat higher yields, probably because of both increased incubation time and the reduced temperature, which stabilizes the double-stranded structures at the junction.

3.2.7

Troubleshooting

- A typical problem with ligations is a low yield of the final ligated product. For two-way ligation (two RNA fragments, one junction), the efficiency should be at least 20%, but even nearly stoichiometric ligations are possible with high-quality RNA. For three-way ligations (three RNA fragments, two junctions), the efficiencies can sometimes be less than 10%, but one should be able to increase the efficiency to approximately 30% relatively easily. If the ligation yield is very low, the first thing to do is to determine which one of the RNA or DNA fragments is responsible for the low efficiency. This can be done easily by setting up small-scale test ligations that contain only about 1 pmol of each fragment. Here, short DNA oligonucleotides can be used as acceptors or donors during troubleshooting instead of the actual RNA fragments.
- If the problem with the ligation can be pinpointed to the acceptor RNA, the most obvious question would be the quality of the 3'-end of the RNA molecule: is it homogeneous or does it contain nontemplated nucleotides? Nontemplated nucleotide additions can often be avoided using the techniques described in Sections 3.2.3 and 3.2.4. Another possible problem with the acceptor is the presence of a stable RNA secondary structure that could prevent the hybridization of the DNA splint oligonucleotide. One way to resolve this is destabilization of the structure by site-specific mutations. The mutations near the junction are often easy to incorporate by means of the PCR oligonucleotides that are used in the generation of templates for T7 transcription. If the strategy does not allow mutations, alternatives are the use of DNA splints that cover the entire RNA or to involve an additional “disrupter” oligonucleotide that binds adjacent to the splint oligonucleotide and prevents the formation of stable RNA secondary structures.

- A typical problem with the donor RNA, in addition to inefficient annealing to the splint DNA owing to intramolecular RNA structure formation, is inefficient dephosphorylation (or phosphorylation) at the 5'-end of the RNA. This can be caused, for example, by a stable secondary structure at the 5'-end of the molecule. A larger amount of phosphatase (or kinase) may be used to overcome this problem, or, alternatively, an analogous disrupter oligonucleotide strategy could be designed.
- Apart from the problems with ligations, one might also experience difficulties at stages that are further downstream of the actual RNA ligation reaction, but which are the results of the ligation procedure. One such, at least theoretical, possibility is the cleavage of the ligated RNA during incubation with cellular extract due to endogenous RNase H activity and residual amounts of the DNA splint oligonucleotide in the ligated RNA sample. We have not observed any RNase H activity resulting from a contaminating DNA splint in any of our studies. However, if this is of concern, it can be avoided by treating the ligation reactions with DNase before purification.

General Buffers and Procedures (GBP)

Preparation of RNase-Free Water

A major cause of RNA degradation is ribonuclease contamination of the water used for the preparation of buffers and solutions. It is therefore recommended to use double-distilled water, which might be further treated with diethylpyrocarbonate (DEPC; add 1/1000 vol. DEPC and stir vigorously for 2 h followed by autoclaving for DEPC decomposition). For longer storage, the water may be additionally filtered through a 0.6 μm filter and kept in 1 ml aliquots in Eppendorf tubes, preferably at -20°C . It should be noted that DEPC remnants may interfere with enzymatic reactions. Thus, we prefer double-distilled water, frozen at -20°C for storage; store water at ambient temperature only when freshly autoclaved, but freeze in aliquots once the bottles have been opened.

Gel Running Buffer (5 x TBE)

54 g of Tris base (446 mM), 27.5 g of boric acid (445 mM), and 20 ml of 0.5 M EDTA (pH 8.0) per liter; no pH adjustment required; store at room temperature.

2 x Gel Loading Buffer for Denaturing PAGE

Weigh 48 g urea, 50 mg bromophenol blue, 50 mg xylene cyanol, and 3.72 g EDTA; add 20 ml 5 x TBE (see above), adjust the volume to 80 ml with double-distilled water, and finally to 100 ml after complete dissolving [37]. Store the solution at room temperature.

Staining Buffer (Ethidium Bromide Solution)

PAA/8 M urea gels are stained in 1 x TBE containing $0.5\text{--}1.0\ \mu\text{g ml}^{-1}$ ethidium bromide (EtBr). An EtBr stock solution at a concentration of $10\ \text{mg ml}^{-1}$ in water is preferred. The solution should be stored in dark bottles (e.g., bottles covered with aluminum foil) at room temperature or 4°C .

Calculation of RNA Concentration

Two alternative formulas can be used for the calculation of RNA concentrations based on UV spectrometry measurements:

- 1) $c = A_{260} \times \text{dilution factor of cuvette solution} / (\alpha_{m260} \times \text{number of nucleotides} \times l)$, where A_{260} is the absorbance at 260 nm, α_{m260} is the average molar absorption coefficient of the 4 nt at 260 nm (for single-stranded DNA and RNA, an average α_{m260} of $10\,800\text{ M}^{-1}\text{ cm}^{-1}$ is appropriate), and l is the pathlength of cuvette (normally 1 cm).
- 2) A_{260} unit (absorbance of 1 measured at 260 nm in a 1 cm cuvette) corresponds to approximately $40\text{ }\mu\text{g ml}^{-1}$ single-stranded RNA (ssRNA). Total amount of RNA (μg) = $40 \times A_{260} \text{ units} \times \text{dilution factor of cuvette solution} \times \text{total volume of RNA stock solution in milliliters}$.

For more precise determination of RNA concentrations, it is possible to calculate the molar absorptivity of each RNA molecule on the basis of their sequence composition using the nearest neighbor formula. To do this, there are several online tools available, such as OligoCalc (<http://www.basic.northwestern.edu/biotools/oligocalc.html>). For further details on determination of RNA concentration by UV spectroscopy, see Appendix.

Protocols for Ligation by T4 DNA Ligase (T4 Dnl)

Protocol 1: Transcription

- 5 x T7 transcription buffer: 600 mM Hepes-KOH, pH 7.5; 120 mM MgCl_2 ; 100 mM DTT; 5 mM spermidine.

T7 transcription, 200 μl		Final concentration
40 μl	5 x T7 transcription buffer	1 x
40 μl	25 mM each rNTP	5 mM
5 μl	40 $\text{U }\mu\text{l}^{-1}$ RNase inhibitor (Promega)	1 $\text{U }\mu\text{l}^{-1}$
2–10 μl	100 $\text{U }\mu\text{l}^{-1}$ T7 RNA polymerase	1–5 $\text{U }\mu\text{l}^{-1}$
40 μl	50–250 $\text{ng }\mu\text{l}^{-1}$ PCR product	10–50 ng
To 200 μl	With RNase-free water	

- 1) Combine reaction components, add the transcription buffer last to avoid precipitation of nucleotides by high concentrations of spermidine present in the transcription buffer.
- 2) Incubate at 37°C for 1.5 h, and then add more enzyme ($0.5\text{--}1 \times$ the original amount) and incubate for an additional 1.5 h. Typically, a white pyrophosphate precipitate will start to accumulate at the later stages of the transcription reaction.
- 3) Add 1 U of RNase-free DNase, such as RQ1 DNase (Promega) for each microgram of template DNA, and continue the incubation for an additional 15–30 min.

- 4) Extract RNA once with phenol (pH 7). Phenol is used here instead of phenol-chloroform-isoamyl alcohol, as the pyrophosphate precipitate will form a pellet with the former, but will remain in the interphase if the latter is used.
- 5) Precipitate RNA by adding 20 μl 3 M NaOAc (pH 5.2) or NaCl and 2.5 vol. of ethanol.
- 6) Dissolve the pellet in 5 μl RNase-free water and gel-purify the RNA by denaturing PAGE as described in Protocol 3.

Protocol 2: Transcription of Capped RNAs

- 5 x transcription buffer for capped RNA transcription (Promega): 200 mM Tris-HCl, pH 7.9; 30 mM MgCl_2 ; 10 mM spermidine; 50 mM NaCl

T7 transcription, 200 μl		Final concentration
40 μl	5 x transcription buffer	1 x
2 μl	100 mM ATP	1 mM
2 μl	100 mM CTP	1 mM
2 μl	100 mM UTP	1 mM
1 μl	100 mM GTP	0.5 mM
40 μl	10 mM G(5')ppp(5')G	2 mM ^a
40 μl	50–250 ng μl^{-1} PCR product	10–50 ng
5 μl	40 U μl^{-1} RNase inhibitor (Promega)	1 U μl^{-1}
2–10 μl	100 U μl^{-1} T7 RNA polymerase	1–5 U μl^{-1}
To 200 μl	With RNase-free water	

Carry out the reaction as described in Protocol 1 and gel-purify according to Protocol 3.

^aCap analogs should be at least in fourfold molar excess relative to GTP to ensure efficient initiation with the cap analog. Similarly, 5'-hydroxyl- or 5'-monophosphate-containing RNAs can be produced by priming the transcription reaction with guanidine or GMP, respectively, under reaction conditions comparable to those used with the cap analogs (Chapter 1).

Protocol 3: Purification of RNA by Denaturing PAGE

- Gel loading buffer: 0.01% bromophenol blue; 0.006% xylene cyanol in 7.5 M urea/1 x TBE.
- RNA elution buffer: 50 mM Tris-HCl, pH 7.5; 10 mM EDTA; 0.1% SDS; 0.3 M NaCl.

Before electrophoresis, the reaction mixture should be extracted once with an equal volume of phenol (pH 7) followed by ethanol precipitation, as proteins in the sample may cause smearing of the bands and retain some of the RNA in the wells. With large amounts (several nanomoles) of RNA, it is advisable to visually verify that the pellet has fully dissolved, by first dissolving it in a small volume of water (5–10 μl) before applying the gel loading buffer.

Prepare the polyacrylamide gel (19 : 1 acrylamide: *N,N'*-methylene bisacrylamide) in 7.5 M urea/1 x TBE. The percentage of acrylamide should be adjusted to the size

of the RNA fragment to be purified. The thickness of the gel should be chosen according to the amount of RNA to be loaded. As the transcription reactions usually contain several nanomoles of RNA, a relatively thick gel (approximately 1 mm) should be used to avoid smearing of the bands. In addition, RNA samples should be distributed between several wells, even though this reduces the RNA elution yields. In later steps, when dealing with smaller amounts of RNA, a thinner gel (0.3–0.5 mm) may be used to increase the elution yields.

- 1) Prerun the denaturing gel at least 20–30 min at 60 W.
- 2) Dissolve the RNA pellet in a small amount of water (5–10 μ l) and add 1 vol. of gel loading buffer. The final urea concentration should be at least 3.5 M but loading buffers containing up to 7.5 M urea can be used.
- 3) Heat the RNA samples at 95 °C for 3–5 min. Immediately put the samples on ice. Centrifuge the samples if there is water condensation on the lid. This is especially important if a high-concentration urea loading buffer is used, as otherwise the loading buffer (and the sample) can precipitate in the pipette tip during loading.
- 4) Load the samples and run the gel at approximately 60 W for at least 30–60 min, or longer, if necessary.
- 5) Separate the gel plates. Place the gel between two Saran wrap (or alike) sheets.
- 6) Visualize the RNA bands using UV shadowing: the gel (between two sheets of Saran wrap) is placed on a fluorescent TLC plate or intensifying screen (a sheet of white paper will do as well) and illuminated briefly with UV light (254 nm). RNA bands are visualized as dark bands on a fluorescent background. Use a pen or marker to indicate the location of each band. The exposure time should be minimized to avoid damage to RNA by the UV light.
- 7) Excise the bands and add approximately 5 vol. of elution buffer to the gel slices. Siliconized microcentrifuge tubes should be used here, as we have noticed that this ensures that the gel slices remain submerged in the elution buffer. Carry out the elution overnight at room temperature using a tube rotator, “rocking table,” or a similar device to provide gentle shaking during elution.
- 8) Collect the supernatant. For increased yield, replace the buffer and continue elution for a further 4–6 h. Extract the eluates once with phenol/chloroform/isoamyl alcohol (25 : 24 : 1) and once with chloroform/isoamyl alcohol (24 : 1). This helps remove any impurities present in the gel, as well as any remaining gel fragments. The RNA is concentrated by ethanol precipitation. Note that there is no need to add salt, as the elution buffer already contains a sufficient amount of salt for the precipitation. Wash the pellet at least three times with 70% ethanol to remove any traces of SDS and dissolve it in RNase-free water. After purification the concentration of the transcript should be measured by UV spectrometry.
- 9) If purifying radiolabeled RNAs, the bands can be visualized by autoradiography instead of UV shadowing. In this case, the gel should be left on one of the glass plates and covered with Saran wrap. Pieces of fluorescent tape (such as RadTape from Diversified Biotech; <http://divbio.com/>) serving as alignment

marks are attached to the wrapped gel and illuminated briefly with light. In the darkroom, place an X-ray film on top of the gel. Expose the film (depending on the activity of the RNA to be purified, this can be anything from 10 s to several hours). Use the markings from the fluorescent tape to align the gel and the film, mark the bands, cut them out, and elute as described in step 7.

Protocol 4: Site-Directed Cleavage with RNase H

- 5 x RNase H buffer (GE Healthcare): 100 mM Tris-HCl, pH 7.5; 100 mM KCl; 50 mM MgCl₂; 0.5 mM EDTA; 0.5 mM DTT.

Annealing reaction, 13 μ l		Final concentrations
5 μ l	400 μ M RNA (2 nmol)	154 μ M
5.5 μ l	400 μ M 2'-O-methyl RNA/DNA oligo (2.2 nmol)	169 μ M
1.3 μ l	1 M KCl	100 mM
1.2 μ l	1 mM EDTA	92 μ M

Run the following program in a thermal cycler: 95 °C 5 min, 85 °C 10 s; decrease the temperature by $-0.1^{\circ}\text{C s}^{-1}$ until the temperature of 35 °C is reached. Alternatively, the annealing may be carried out in a heating block. In this case, place the samples first in a hot heating block (95 °C) for 3–5 min, then remove the block from the heating unit, and allow slow cooling to room temperature. In each case, the tubes should be checked for any condensed water on the lid after the annealing and centrifuged, if necessary.

Cleavage reaction, 50 μ l		Final concentrations
13 μ l	Annealing mix	
10 μ l	5 x RNase H buffer	1 x
1.5 μ l	40 U μ l ⁻¹ RNase inhibitor (Promega)	~ 1 U μ l ⁻¹
0.5 μ l	100 mM DTT	1 mM
15 μ l	5 U μ l ⁻¹ RNase H	1.5 U μ l ⁻¹
10 μ l	RNase-free water	

- 1) Incubate at 37 °C for 3–4 h, and extract once with phenol/chloroform/isoamyl alcohol and ethanol-precipitate. If dephosphorylation is to be performed before electrophoresis, carry out chloroform/isoamyl alcohol extraction after the phenol extraction to remove any traces of phenol.
- 2) Separate the cleavage products in a denaturing polyacrylamide gel and visualize the bands by UV shadowing. Use approximately 200 pmol of uncut RNA as a control to distinguish the full-length RNA. Elute and purify as described in Protocol 3.

Protocol 5: Dephosphorylation

- 10 x CIAP buffer (Finnzymes): 100 mM Tris-HCl, pH 7.9; 100 mM MgCl₂; 10 mM DTT; 500 mM NaCl

Dephosphorylation reaction, 100 μ l		Final concentrations
10 μ l	10 x CIAP buffer	1 x
10 μ l	200 μ M RNA (2 nmol)	20 μ M
2.5 μ l	40 U μ l ⁻¹ RNase inhibitor (Promega)	1 U μ l ⁻¹
10 μ l	10 U μ l ⁻¹ CIAP (Finnzymes)	1 U μ l ⁻¹
68 μ l	RNase-free water	

- 1) Incubate at 50 °C for 1 h.
- 2) Add at least 1 vol. of an appropriate 1 x buffer for proteinase K (the RNA elution buffer described earlier also works very well) and 100 μ g of proteinase K to the reaction and continue incubation at 50 °C for 1 h.
- 3) Extract RNA with phenol/chloroform/isoamyl alcohol followed by extraction with chloroform/isoamyl alcohol and ethanol precipitation. If the RNA fragment has been cleaved with RNase H before the dephosphorylation and has not been purified by gel electrophoresis, this should be performed at this stage.

Protocol 6: Site-Specific Labeling with Radioactive Phosphorus at the Donor 5'-End

- 10 x T4 PNK buffer: 700 mM Tris-HCl, pH 7.5; 100 mM MgCl₂; 50 mM DTT.

Phosphorylation reaction, 40 μ l		Final concentrations
4 μ l	10 x T4 PNK buffer	1 x
5 μ l	20 μ M RNA (100 pmol)	2.5 μ M
6 μ l	200 μ M ATP ^a	30 μ M
1 μ l	40 U μ l ⁻¹ RNase inhibitor (Promega)	1 U μ l ⁻¹
20 μ l	γ - ³² P-ATP (10 μ Ci μ l ⁻¹ , 6000 Ci mmol ⁻¹) ^a	1.7 μ M
1 μ l	10 U μ l ⁻¹ T4 PNK	0.25 U μ l ⁻¹
3 μ l	RNase-free water	

^aThe ratio of radioactively labeled *versus* cold ATP can be adjusted depending on the particular experiment. However, the total concentration of ATP in the reaction should be kept at least 2 x higher than the concentration of the RNA. Note that for efficient and quantitative labeling, the concentration of ATP should be relatively high. The K_m of T4 PNK for ATP is approximately 14–140 μ M (depending on the substrate), and the phosphorylation may thus not proceed efficiently at low ATP concentrations. If large amounts of highly active RNA are needed, we use a more crude, but concentrated, γ -³²P-ATP (such as NEG 035C from Perkin Elmer) with an activity of about 150 μ Ci μ l⁻¹ at a concentration of 15 μ M. If RNA is not to be radioactively labeled, the ³²P-labeled ATP can be replaced with 500 μ M unlabeled ATP.

- 1) Incubate at 37 °C for 30–60 min. Subsequently, raise the concentration of cold ATP to 45 μM and continue the incubation for 15 min.
- 2) Extract once with phenol/chloroform/isoamyl alcohol and once with chloroform, and ethanol-precipitate.
- 3) Dissolve the pellet in a few microliters of water. Alternatively, the dry pellet can be directly used in the ligation reaction if dividing the RNA donor into smaller aliquots is not needed.

Protocol 7: Ligation

- 10 x ligation buffer: 500 mM Tris-HCl, pH 7.5; 100 mM MgCl_2 ; 200 mM DTT; 10 mM ATP

Annealing mix for a two-way ligation, 5 μl ^a		Final concentration
1.2 μl	50 μM donor RNA (50 pmol) ^b	10 μM
1.4 μl	50 μM acceptor RNA (70 pmol)	14 μM
1.2 μl	50 μM splint DNA oligo (60 pmol)	12 μM
0.5 μl	1 mM EDTA	0.1 mM
0.5 μl	500 mM KCl	50 mM

^aEven if the aim is to produce an unlabeled product, we recommend to add in trace amounts one of the RNA fragments radioactively labeled. This provides an easy way to quantify the total yield of the ligation reaction. Simply compare the amount of radioactivity included in the reaction with the amount of radioactivity after the gel purification (by measuring small samples with liquid scintillation counting) and multiply this percentage with the total amount of this particular RNA fragment in the reaction.

^bWe have successfully used RNA concentrations ranging from 100 nM to 10 μM in the ligation reaction.

- 1) Run the following program in a thermal cycler: 95 °C 5 min, 85 °C 10 s, followed by cooling to 35 °C with a slope of $-0.1\text{ }^\circ\text{C s}^{-1}$.
- 2) Check for condensation at the lid and centrifuge the tube(s), if necessary. Compose ligation reaction as follows:

Ligation reaction, 10 μl		Final concentration
5 μl	Annealing reaction	
1 μl	10 x ligation buffer	1 x
1.5 μl	13% polyvinyl alcohol (PVA)	~2%
0.5 μl	40 U μl^{-1} RNase inhibitor	2 U μl^{-1}
2 μl	30 Weiss U μl^{-1} T4 Dnl (Fermentas)	6 U μl^{-1}

- 3) Incubate at room temperature overnight or at 30 °C for at least 4 h.

- 4) Extract once with phenol, and ethanol-precipitate. Separate the ligated products from the unligated donor and acceptor in a denaturing polyacrylamide gel. Use radioactively labeled donor, acceptor, or a full-length transcript as control. Owing to the small amount of RNA present, the bands must be visualized using autoradiography. Elute and purify the excised samples as described in Protocol 3. As the amount of RNA can be relatively small after the ligation, care must be taken to maximize the yields of the gel-purified product. For this purpose, we use siliconized microcentrifuge tubes in the elution step and carry out the RNA precipitation in the presence of carrier (10 μ g glycogen).

3.3

Simultaneous Splint Ligation of Five RNA Fragments to Generate RNAs for FRET Experiments

3.3.1

Introduction

The following protocols for a one-pot reaction involving phosphorylation, ligation, and splint digestion were specifically developed for the synthesis of 75–90 nt long RNAs carrying two or more site-specific modifications, including two fluorescent dyes forming a FRET pair that permits monitoring dynamic changes in RNA structure (Figure 3.6). Beyond fluorescent dyes, further modifications may be introduced, such as naturally occurring nucleotide modifications, locked nucleic acids, or biotin residues. When planning the incorporation of several modifications, it is helpful to compose the RNA of several fragments, ideally each one carrying a single modification. In this way, the construction of RNA variants carrying different subsets of modifications (e.g., dyes plus or minus a 5'-terminal biotin label) is easily achievable *via* a “construction kit” approach. We have successfully ligated up to five fragments in a single reaction, illustrating the suitability of the cDNA splint to arrange the RNA fragments in correct order by virtue of sequence complementarity. Typical ligation yields are around 60% per ligation site, resulting in about 10–15% final product yield in a five-piece ligation.

Several companies offer a variety of modified building blocks, ranging from degenerate nucleotides to modifiable bases or sugars. Several of these carry moieties suitable for covalently linking tags or dyes to the RNA of interest, using different attachment chemistries. A variety of tags and dyes are available that can be coupled to the RNA *via* these moieties. Owing to the absolute requirement for canonical Watson–Crick base pairing to the DNA splint during ligation, the possible sites of attachment are limited. Usually, both the 5'- and the 3'-end of the final construct are available for labeling. The main strategy for label attachment at internal positions is to use linkage sites that are not involved in base pairing, such as the C5 of the uracil base, or the C8 of a guanosine moiety. Custom-modified oligonucleotides to generate such base-linked derivatives are available (e.g., a C5-amino-dT derivative) from several companies (e.g., IBA, Göttingen, Germany),

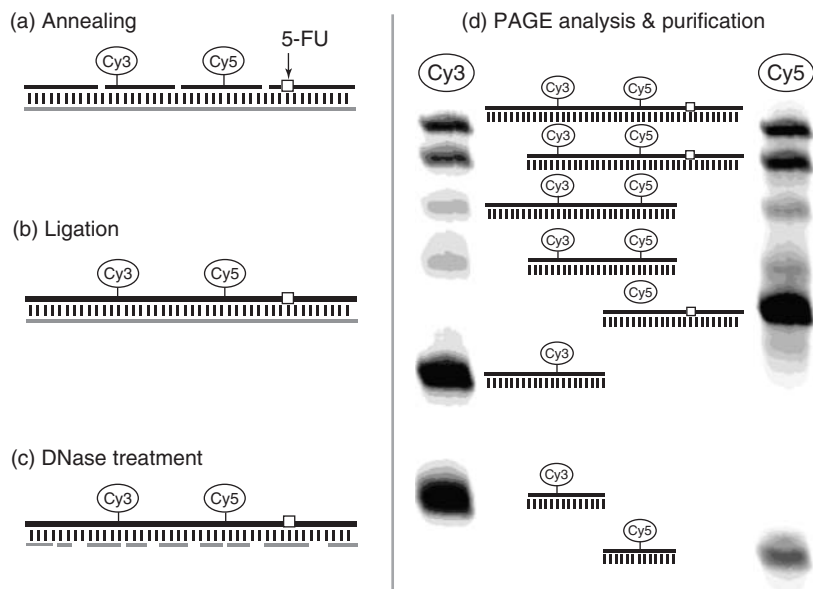


Figure 3.6 Splint ligation of multiple RNA fragments including fluorescent dye modifications. (a) The RNA fragments containing, for example, fluorescent Cy3, Cy5 modifications, or 5-fluoro-uridine (5-FU) are aligned by hybridization on a complementary DNA splint. (b) T4 Dnl closes the “nicks” in the RNA by joining the adjacent 3′-hydroxyl and

5′-phosphate ends. (c) The DNA splint is removed by incubation with DNase I. (d) The resulting reaction mixture is subjected to denaturing PAGE. Fluorescence scanning reveals various fragments and partial ligation products carrying one or both dyes, respectively, in addition to the full-length product. The latter is excised and eluted from the gel.

and there are options to obtain oligonucleotides that are modifiable by NHS chemistry (e.g., Thermo Scientific Dharmacon, Lafayette, Colorado, USA). These comprise internal (5-amino-allyl-uridine) as well as terminal (5′-amino, 6- or 12-carbon linker) options. Such oligonucleotides have been used in a number of studies on RNA structure and function [38–40]. For naturally occurring modified nucleotides that are commercially available, see references [41, 42].

In general, the ligation scheme should be designed such that these modifications do not come to lie close to a ligation site. In the case of modifications that alter or hamper canonical base pairing in proximity of the ligation site, the resulting hybridized complex will undergo ligation with only limited efficiency. This also holds true for modifications that interfere with ligase binding to the complex (e.g., linker modifications of nucleobases). In general, at least two unmodified canonical base pairs should be present 3′ and 5′ of the ligation site to avoid steric interferences.

The experimental approach can be grouped into three steps, starting from modified oligonucleotides and resulting in the full-length RNA construct: (i) design of construct, (ii) phosphorylation and ligation, and (iii) purification.

3.3.2

Construct Design

Construct design follows a reverse-engineering approach: on the basis of experimental knowledge and/or available structures, there are usually regions within each molecule of interest that are amenable to labeling. In the case of FRET-based studies, two attachment sites have to be identified that appear suitable to report alterations in their spatial distance. It has been proved useful to use attachment sites that come to lie within helical regions of the structured RNA. In this case, fluorescent dyes attached to the uracil C5 or guanosine C8 protrude from the major groove. This minimizes the risk of the structure being affected by the presence of bulky dyes. The distance of the dyes to either end of a fragment should be two or more nucleotides that are not modified. This ensures proper hybridization of the fragment ends to the DNA splint and minimizes the risk of terminal stacking interactions between dyes and helix ends.

According to the attachment sites chosen, junction sites for the singly modified fragments have to be identified. In order to assemble the full-length RNA with several modifications, it has to be split into parts that carry only one labeling moiety each. Since the minimal length of fragments is dictated by their melting temperature in the DNA–RNA heteroduplex, we do not recommend using less than 8 nt at a ligation temperature of 16 °C. In the case of single-molecule studies, the requirement of only low amounts of labeled RNA constructs permits increasing the number of RNA fragments to be ligated to up to five or even six. This opens possibilities to synthesize a variety of constructs from a limited number of oligonucleotides in a combinatorial manner. It has to be pointed out that replacing T4 Dnl with T4 Rnl2 (which also ligates RNA oligonucleotides in an RNA–DNA duplex) is strongly discouraged in ligation experiments of more than two fragments. The unspecific “background” ligation of RNA oligonucleotides (thought to arise from the formation of transient dsRNA structures) will complicate identification and isolation of the desired specific RNA ligation product. In other cases, T4 Rnl2 might be used for RNA oligonucleotides that have been proved to be difficult to ligate by T4 Dnl because of their length or modifications.

As longer DNA molecules (up to 200 nt) are easily available and comparably cheap, we suggest covering the entire RNA construct, including all fragments from the 5'- to the 3'-end, with the hybridizing DNA splint. In the case of longer RNA constructs, several splints can be used. In this case, the reaction conditions might need further optimization.

Protocols for the Simultaneous Splint Ligation of Five RNA Fragments**Protocol 8: Phosphorylation**

- 5 x KL buffer: 250 mM Tris-HCl, pH 7.4; 50 mM MgCl₂; 25 mM DTT; 10 mM ATP

- 1) Adjust the solution of RNA fragments (final concentration: 40 μM) to 1 x KL buffer.
- 2) Add 0.75 U μl^{-1} T4 PNK (Fermentas) and incubate at 37 °C for 60 min.

The phosphorylated RNA fragments can be directly used in Protocol 9. Note that the buffer composition of the phosphorylation reaction has to be taken into account when calculating the buffer composition of the ligation reaction. If higher oligonucleotide concentrations are required, phenol/diethylether extraction and subsequent ethanol precipitation may be conducted before ligation.

We have successfully applied Protocol 9 in conjunction with Protocol 8 for large-scale ligations (nmol scale, reaction volumes up to 1 ml) using T4 Dnl prepared in-house. Protocol 10 is recommended when ligations are not routinely performed in the laboratory, and when the RNA oligonucleotides are already 5'-phosphorylated, making Protocol 8 dispensable. In addition, ATP stability is more guaranteed in Protocol 10 relative to Protocol 9, because in the former half of the ATP is only added after the 75 °C denaturation step to reduce the potential of thermal ATP decomposition.

Protocol 9: Ligation

- 1) Combine equimolar amounts of RNA fragments and DNA splint, at concentrations of 1–10 μM each.
- 2) Adjust to 1 x KL buffer and incubate at 75 °C for 3 min (denaturation).
- 3) Incubate at room temperature for 15 min (annealing).
- 4) Add T4 Dnl (Thermo Scientific Fermentas, approximately 2 U (Weiss units) μl^{-1} final concentration, or equivalent amount of T4 Dnl prepared in-house) and incubate for at least 4 h or overnight at 16 °C.

Protocol 10: Ligation

- T4 Dnl (high conc., 2000 cohesive end units μl^{-1} ; NEB)
 - 10 x T4 Dnl buffer (NEB): 500 mM Tris-HCl; 100 mM MgCl_2 ; 10 mM ATP; 10 mM DTT, pH 7.5 (25 °C)
- 1) Combine equimolar amounts of RNA fragments and DNA splint, at final concentrations of 1–10 μM each.
 - 2) Adjust the solution to 0.5 x T4 Dnl buffer and incubate at 75 °C for 3 min (denaturation).
 - 3) Incubate at room temperature for 15 min (annealing).
 - 4) Add T4 Dnl, approximately 2 μl (4000 U) per 100 μl reaction volume.
 - 5) Adjust the solution to 1 x T4 Dnl buffer and incubate at 16 °C or room temperature overnight.

All ligation reactions can be purified directly using denaturing PAGE. In the case of large reaction volumes, phenol/diethylether extraction and ethanol precipitation should precede denaturing PAGE.

3.3.3

Troubleshooting3.3.3.1 **Low Overall Ligation Efficiency**

- Perform ligations with individual pairs of adjacent RNA oligonucleotides and identify junctions that ligate poorly.
- Optimize ligation conditions (annealing temperature, Mg^{2+} concentration, ATP concentration, and ligation temperature).
- Check phosphorylation efficiency by spiking the reaction with radiolabeled ATP.

3.3.3.2 **Undesired Ligation By-products**

- Optimize the annealing step (possibly using non-denaturing PAGE to analyze annealing efficiency) and Mg^{2+} concentration.
- Try using longer DNA splints, up to full coverage of the ligated RNA oligonucleotides.

3.3.3.3 **RNA Degradation**

- Check for RNase contaminations in oligonucleotide stock solutions.
- Use RNase inhibitor.
- Replace reaction buffers.
- Use DNA splints that cover the entire RNA sequence.

3.4

T4 RNA Ligase(s)

3.4.1

Introduction

This section focuses on T4 Rnl1 (EC 6.5.1.3; also named RnlA RNA ligase 1) that belongs to a family of oligonucleotide end-joining enzymes involved in RNA repair, splicing, and editing pathways [43–45]. The enzyme, introduced into molecular biology laboratories about 40 years ago [46–48], catalyzes the formation of phosphodiester bonds between 5'-phosphate and 3'-hydroxyl ends of preferentially ssRNA and, less efficiently, ssDNA. The oligo(ribo)nucleotide carrying the terminal 3'-hydroxyl group is termed *acceptor substrate*, and the one providing the terminal 5'-monophosphate is described as the donor substrate (Figure 3.2A). In intramolecular circularization reactions, both end groups (3'-OH, 5'-phosphate) are located on the same oligo(ribo)nucleotide molecule. The biological role of T4 RNA ligase seems to have its seeds in the intricate antagonisms of T4 phage and *E. coli* host strains: T4 infection was shown to induce activation of a nuclease that cleaves bacterial tRNA^{Lys} in the anticodon loop. T4 RNA ligase in concert with T4 PNK catalyzes the repair of the damaged tRNA^{Lys} [49, 50].

3.4.2

Mechanism and Substrate Specificity3.4.2.1 **Early Studies**

Toward understanding the substrate specificity of T4 Rnl1, it is instructive to briefly review results of early studies. The enzyme was first described by its ability to circularize 5'-³²P-labeled tRNA and polyhomoribonucleotides, such as poly(A), in a reaction requiring ATP and Mg²⁺ [46]. Relative to poly(A), the reaction occurred about fourfold less efficiently with poly(I), circa 100-fold less efficiently with poly(C) and poly(U), and at least 800-fold less efficiently with poly(dA). Circularization of poly(A) molecules with an average chain length between 34 and 40 nt was about twice as efficient as for those 70–100 nt in length, and no intermolecular ligation products were observed in this early study [46]. Subsequently, the shortest circularizing polyadenylate was shown to be (pA)₈, the optimal chain length for this reaction being 10–30 A residues [51]. Cyclization was generally found to be the preferred reaction by orders of magnitude over intermolecular joining [52, 53], with four exceptions representing conditions that favor intermolecular ligation: (i) when the donor is too short to cyclize, (ii) when a DNA donor is combined with a 5'-dephosphorylated RNA acceptor, (iii) when the acceptor carries 5'- and 3'-hydroxyls and the donor a 5'-phosphate and a blocked 3'-terminus (Figure 3.2B; [54]), and (iv) when the donor 5'-phosphate of one nucleic acid molecule is juxtaposed to the 3'-hydroxyl of a second (acceptor) nucleic acid molecule by base-pairing interactions, resulting in a *quasi*-intramolecular reaction (see below).

For ligation of DNA, studies with [5'-³²P]-labeled oligodeoxythymidylates of various length ([5'-³²P]dT_n) revealed that ligase-catalyzed cyclization requires a minimal chain length of 6 dT residues, and the best efficiency was obtained with chains of 20 [53]. As DNA is a less efficient acceptor than RNA, adenylated A(5')pp(5')dT_n intermediate accumulated to some extent. Interestingly, addition of the ribotrinucleotide ApApA, beyond serving as acceptor substrate, also stimulated cyclization of [5'-³²P]dT_n, suggesting that acceptors function not only as substrates for ligation but also as cofactors for adenylation of the donor. ApA instead of ApApA neither stimulated donor adenylation nor was it joined [52, 53]. Similarly, ApA, IpI, or UpU were found to be inactive as acceptors in the overall reaction with pAp as donor [54].

Analysis of minimal donor substrates of the pNp type (nucleoside-3',5'-bisphosphate) in the adenylation partial reaction revealed the highest efficiency of A(5')pp(5')Np formation for pCp, whereas the reaction with pUp and pAp was threefold and with pGp tenfold less efficient [8]. A similar hierarchy was seen in the overall reaction [54, 55]. On the acceptor side, ApApA was a much better substrate than UpUpU (for acceptor DNA/RNA mixmers with a 3'-terminal stretch of RNA purines, see Chapter 33). Ligation of the “poor” pGp donor to the “good” ApApA acceptor could be stimulated to proceed to almost completion by increasing the pGp concentration from 1 to 10 mM [8]. Furthermore, this study led to the conclusion that the enzyme exhibits more specificity in the donor adenylation reaction than in the subsequent joining of adenylated donor to acceptor. Formation

of the UpUpUpGp product was much more efficient in the reaction of UpUpU with the adenylated donor A(5')pp(5')Gp than in the overall reaction starting from UpUpU, pGp, and the cofactor ATP. Thus, preadenylation of the donor substrate may serve as a strategy to improve product yields when dealing with “poor” substrates ([8]; see Chapters 38 and 39). However, this principle may not be generally valid, as the results of two other studies investigating the synthesis of UpUpUpAp from UpUpU and pAp suggested rate limitation at the level of transfer of adenylated donor A(5')pp(5')Ap to acceptor [54, 55].

3.4.2.2 Substrate Specificity and Reaction Conditions

Substrate specificity of T4 Rnl1 can be summarized as follows. The minimal donor substrate is a nucleoside-3', 5'-bisphosphate, with efficiency decreasing in the order pCp > pUp $\approx$ pAp > pGp. Isocytidine-3', 5'-bisphosphate was reactive as well, indicating that the modified bases are also tolerated [56]. Except for p(dCp), the deoxyribonucleoside-3', 5'-bisphosphates were found to be poorer donor substrates than the corresponding pNp ribonucleosides in the overall reaction [54]. 5'-AMP (pA) is not a donor, and also nucleoside-2', 5'-bisphosphates are neither donor substrates nor effective inhibitors. Thus, in the donor substrate, the enzyme specifically recognizes the 5'-terminal phosphate and ribonucleoside plus the next 3'-linked phosphate; the chain length of the donor exerts only marginal effects on reaction extent [54].

The smallest reactive acceptors are trinucleoside diphosphates (NpNpN) with a 3'-terminal hydroxyl [52]. Recently, dinucleoside polyphosphates with at least four bridging phosphates, such as Gp₄G, were also identified as acceptor substrates [57]. The 3'-terminal ribose moiety is important for acceptor recognition, and a 3'-terminal adenosine is preferred over cytidine and guanosine, showing intermediate reactivity, while a uridine residue is a relatively poor substrate ($A > C \geq G > U$; [54, 55]). However, intermolecular ligation yields are not simply dependent on the identity of the 3'-terminal base. For example, lower yields were observed with trimeric NpNpN acceptors containing a U residue at any of the three positions [54]. In another study comparing trimeric acceptors equal in base composition, two consecutive purines enhanced ligation yields with pCp as donor (e.g., GpApU > UpApG > ApUpG; [55]).

Regarding DNA oligonucleotides as substrates for T4 Rnl1, combined appreciation of several studies [53, 58, 59] revealed the following rules of thumb: DNAs are less efficient substrates than RNAs, but discrimination against DNA occurs mainly at the acceptor substrate level. To improve ligation efficiency, DNA/RNA mixmers with a 3'-terminal stretch of RNA purine residues may be used as acceptor substrates (Chapter 33).

Despite these substrate preferences, several strategies have been successfully used to optimize reaction conditions for poorer substrates. Reductions in Mg²⁺ concentration or addition of dimethyl sulfoxide (DMSO) improved product (UpUpUpAp) yields for a UpUpU acceptor and a pAp donor [54]. Stimulation by DMSO, however, was not observed in ligation of UpUpU with pCp [55], suggesting that the benefit of such additives may not be taken for granted. Also, the joining

of UpUpU and pCp to pUpUpUpCp at high enzyme concentration (350 U ml^{-1}) was stimulated from 6 to 60% product yield when ligation mixtures were incubated for 18 h at 15°C instead of 1 h at 37°C [55]. Ligation efficiencies with DNA were increased to some extent by elevated enzyme concentration, (partial) replacement of Mg^{2+} with Mn^{2+} , reduction of incubation temperature to circa 17°C , variation of substrate : ATP ratio as well as their individual concentrations, or low ATP concentration plus an ATP regeneration system [58, 59].

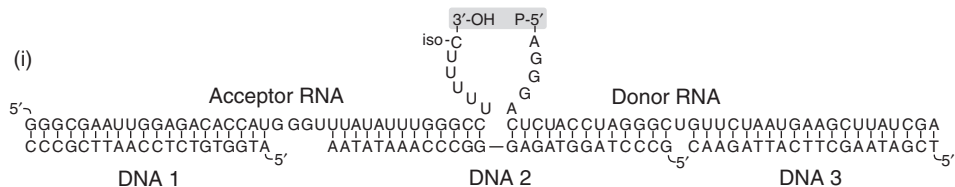
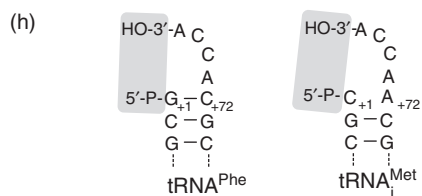
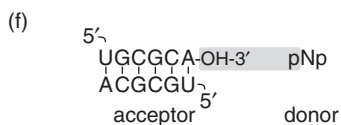
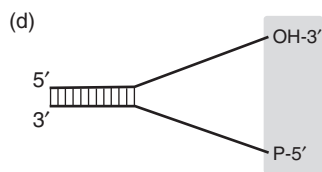
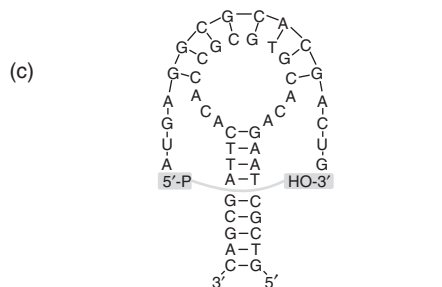
In summary, the substrate specificity of T4 Rnl1 is rather broad, permitting the ligation of essentially any RNA or DNA sequence, mostly with satisfactory efficiency. Best reaction yields are commonly obtained at pH 7.2–7.8, 10–20 mM Mg^{2+} , 10–20% DMSO, and often at temperatures as low as 5°C , with substantial activity even exerted at 0°C [60]. Several of the above-mentioned parameters may be varied to optimize ligation yields. Furthermore, additives beyond DMSO, such as PEG 8000 or hexammine cobalt chloride, have been shown to improve product formation [11]. Another aspect is to prevent formation of unwanted by-products. Thus, for intermolecular joining of two oligo(ribo)nucleotides, one should generally bear the following aspects in mind, as illustrated in Figure 3.6b: the acceptor substrate ought to carry a 5'-hydroxyl terminus to prevent acceptor cyclization or joining of two acceptor substrates; similarly, the donor substrate should be blocked at its 3'-terminus by a 2', 3'-cyclic phosphate, a 3'-phosphate, a dideoxy residue [11], a 3'-inverted deoxythymidine, a periodate-oxidized terminal ribose (Chapter 6), or a 3'-amino group (Figure 3.2B), to avoid donor cyclization or formation of donor tandems.

3.4.3

Applications of T4 RNA Ligase

3.4.3.1 End-Labeling

A common application of T4 Rnl1 is 3'-endlabeling with $[5'-^{32}\text{P}]\text{pCp}$. A ribocytidine dinucleotide bearing a 5'-phosphate and a 3'-terminal nonradioactive label, such as a fluorescein group ($\text{pCpC}_{3'\text{-fluorescein}}$), was also shown to be efficiently attached to the 3'-end of RNA substrates by the enzyme [61]. In a related application, a 5'-phosphorylated pCpC dinucleotide with a 3'-terminal PEG linker including an internal photocleavage site and a terminal primary aliphatic amino group for coupling purposes was attached to RNA acceptor 3'-ends [62]. Kinoshita *et al.* [63] have made use of the fact that the ligase attaches an AMP residue *via* a 5', 5'-pyrophosphate linkage to the 5'-phosphate end of the donor RNA or DNA, representing the intermediate donor activation step on the reaction pathway (Figure 3.2A). In the absence of an acceptor nucleic acid, this intermediate was shown to be stable, and the authors demonstrated that fluorescent 2-aminopurine riboside triphosphate or 3'-amino ATP (for subsequent biotinylation) could replace the normal ATP in this 5'-endlabeling reaction catalyzed by T4 Rnl1.



3.4.3.2 Circularization

The enzyme has further been exploited for intramolecular circularization of linear RNA to produce an authentic infectious circular RNA (371 nt) of the citrus exocortis viroid strain A (CEV-A; [64]). One site of ligation was located within a dinucleotide internal loop of the viroid's predicted rod-like structure [64, 65], juxtaposing the reacting end groups (Figure 3.7a). T4 Rnl1 was further employed to produce circular versions of hammerhead ribozyme strands as small as 15 nt, which exhibited increased activity, a reduced requirement for divalent metal ions, as well as increased resistance against nucleolytic degradation [66]. Interestingly, efficient T4-Rnl1-catalyzed circularization of such hammerhead ribozyme oligoribonucleotides was achieved after their internal 7–8 nt had been annealed to a complementary DNA oligonucleotide, either linear or presented within a DNA hairpin loop (Figure 3.7b,c). This setup favored circularization over formation of linear dimers, and the short central RNA–DNA duplex constrained the overall flexibility of the RNA oligonucleotide, while simultaneously juxtaposing the single-stranded 5'- and 3'-ends to be ligated.

3.4.3.3 Intermolecular Ligation of Polynucleotides

Ligation of RNA oligonucleotides to the 5'-end of mRNAs or other RNAs is used in 5'-RACE (rapid amplification of cDNA ends) strategies to map RNA 5'-ends [69, 70]. Likewise, ligation of DNA or RNA oligonucleotides to RNA 3'-ends is the initial step before reverse transcription and PCR in approaches to map RNA 3'-ends (3'-RACE), or to determine the length of poly(A) tails and to identify polyadenylation sites [69, 70]. Nowadays, adapter oligonucleotide ligation to RNA 5'-ends is also widely applied in deep sequencing studies (Chapters 38 and 39).

Nishigaki *et al.* [67] used T4 Rnl1 to tie two DNA single strands together (the 3'-terminal nucleotide of the acceptor oligonucleotide was a riboC in these experiments to increase ligation efficiency). To bring the reacting ends in proximity to each other, they equipped the two oligonucleotides with 12 nt long complementary

Figure 3.7 Substrates that have been successfully used in the reaction catalyzed by T4 Rnl1. (a) Intramolecular circularization of linear RNA to produce an authentic infectious circular RNA (371 nt) of the citrus exocortis viroid strain A (CEV-A; [64]). (b,c) Setups to produce circular versions of hammerhead ribozyme strands, using complementary DNA oligonucleotides that are either linear (b) or that present the complementary sequence within a hairpin loop (c) [66]. (d) “Y”-shape design for intermolecular ligation [67]. (e) Documented strategies used for the ligation of broken tRNA structures. (f,g) Double-stranded acceptor substrates [68] with blunt ends (f) or with a single nucleotide 5'-overhang (g). (h) Donor

termini in the context of tRNA structures. The 5'-terminal phosphate of tRNA^{Phe} is an inefficient donor; in contrast, the 5'-terminal phosphate of tRNA^{Met} has excellent donor substrate quality owing to the mispairing between C₊₁ and A₊₇₂ [60]. (i) Adaptation of the “DNA splint oligonucleotide” principle to the T4 RNA ligase reaction [56]. DNA oligonucleotides DNA 1 and DNA 3, used to prevent formation of unwanted ligation by-products, are likely to be dispensable when the acceptor RNA carries a 5'-OH terminus and the donor RNA is blocked at its 3'-end, as specified in Figure 3.2b. IsoC: isocytidine used instead of a natural nucleoside as the 3'-terminal nucleotide of the acceptor RNA in this particular experiment.

sequences at one terminus, such that they formed “Y”-like hybrid structures, with the blunt-ended helix representing the stem and the two single-stranded arms presenting the donor and acceptor groups at their tips (Figure 3.7d). This setup converted the intermolecular reaction into a *quasi*-intramolecular reaction, and the blunt-ended stem of the “Y” ensured that only the unpaired end of each oligonucleotide reacted. Even dangling “Y” arms of up to circa 50 nt each still gave ligation yields of circa 20%.

Tessier *et al.* [11] optimized the T4-Rnl1-catalyzed reaction for the joining of pure DNA oligonucleotides, in this case a 25-mer and a 23-mer. Ligation yields of more than 50% were achieved by including PEG 8000 and hexamine cobalt chloride in the reaction. Oligonucleotide joining was favored over accumulation of the adenylated donor intermediate by restricting the ATP cofactor concentration to 20 μ M. In this setup, the acceptor carried 5'- and 3'-OH end groups, whereas the donor oligonucleotide carried the 5'-phosphoryl group, but its 3'-end was blocked *via* a single dideoxy analog (Figure 3.2B), added by terminal transferase, to avoid ligation of two donor oligonucleotides.

3.4.4

T4 RNA Ligation of Large RNA Molecules

For the ligation of larger RNAs, several aspects should be kept in mind:

- 1) **Proximity of ends:** To increase the probability for an enzyme molecule to simultaneously bind both reacting end groups, the ssRNA acceptor and donor ought to be brought into close proximity to each other. Generally, this is accomplished by letting the ends protrude from a helical region, resulting in a hairpin loop as the reaction product. Such a design converts intermolecular reactions into *quasi*-intramolecular reactions. The importance of the structural context of the ligation site is easily illustrated for tRNA molecules. The secondary structure of tRNA is built from three stem-loop structures and one stem, which together form what is known as a *cloverleaf structure* (Figure 3.8). On the basis of the enzyme's preference for ssRNA termini, the three loops are expected to be favorable ligation areas in a tRNA molecule. Indeed, corresponding ligation strategies have been established for the anticodon loop (Figures 3.7e and 3.8a; [71–73]) and the D loop (Figures 3.7e and 3.8b, [74]), while so far T loop ligation in the context of a full-length tRNA structure has not been described. The size of the product hairpin loop will influence ligation efficiency: increased loop size means more flexibility of the dangling single strands and increased average distance of the reactive groups before ligation. This is expected to reduce the efficiency of product formation. On the other hand, ligation within a tetraloop may occur less efficiently because the acceptor and donor ends may already be conformationally restricted before end joining.
- 2) **Acceptor substrates:** Earlier studies with short acceptor oligoribonucleotide duplexes (6 bp) and pNp donors have indicated that acceptor substrates can be double-stranded or even blunt-ended (Figure 3.7f; [68]). Moreover, duplexes

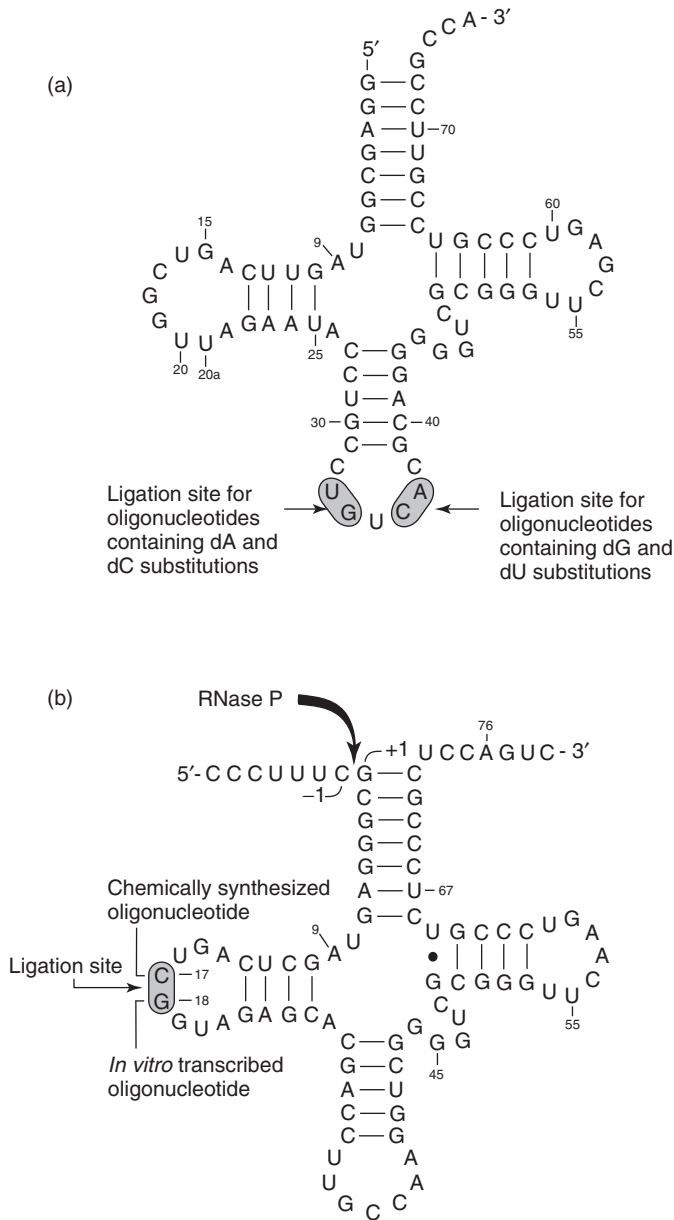


Figure 3.8 (a) Schematic illustration of the T4 RNA ligation strategy for *E. coli* tRNA^{Asp}. The two alternative ligation sites (highlighted in gray) were placed in the anticodon loop, between U33 and G34 for oligonucleotides containing 2'-deoxyA or 2'-deoxyC modifications, and between C36 and A37 for those

containing 2'-deoxyG or 2'-deoxyU modifications. (b) Schematic representation of the T4 RNA ligation strategy for *Thermus thermophilus* ptRNA^{Gly}. The arrow between positions -1 and +1 marks the canonical RNase P cleavage site; the ligation site between C17 and G18 is highlighted in gray.

with single nucleotide 5'-overhangs reacted most efficiently with the pNp donor that can base-pair with the duplex overhang (Figure 3.7g). For example, with single cytidine 5'-overhangs on each end of the acceptor duplex, best yields were obtained for the pGp donor [68]. Regarding blunt-ended acceptor duplexes, it should, however, be mentioned that a dA₈ DNA acceptor oligonucleotide was joined less efficiently to a donor substrate in the presence of a complementary dT₈ DNA oligonucleotide [59], suggesting that a blunt-ended duplex, which is expected to form from the two oligonucleotides, is a less efficient type of acceptor substrate.

- 3) **Donor substrates:** Donor termini that are part of a helix are inefficient substrates [60]. For example, the 5'-terminal phosphate of G₊₁ of yeast tRNA^{Phe} was a relatively inefficient donor because G₊₁ is base-paired to C₊₇₂, forming the terminal acceptor stem base pair (Figure 3.7h). In comparison, the 5'-terminal monophosphate at C₊₁ of *E. coli* initiator tRNA^{Met}, carrying a single mismatch at the acceptor stem terminus (C₊₁ and A₊₇₂, Figure 3.7h), was a much better donor in the T4-Rnl1-catalyzed ligation reaction. Here, intramolecular cyclization between the 5'-terminal phosphate and the 3'-hydroxyl at the 3'-end (A₊₇₆) was the favored reaction, already occurring with high efficiency at very low enzyme concentrations [60].
- 4) **Accessibility of ligation sites:** In the case of large RNAs, the helix structure that clamps the preferably single-stranded acceptor and donor substrates should be positioned at the surface of the RNA complex to ensure enzyme access. In the case of RNAs for which the ligation site is embedded in higher-order structures, experimenters usually switch to the “DNA splint” ligation technique using T4 Dnl, as it involves disruption of RNA structure (Section 3.2 and 3.3, [13]). However, it may sometimes have escaped attention that the “splint principle” has also been successfully adapted to the T4 Rnl1 reaction by annealing donor and acceptor RNAs to a bridging DNA oligonucleotide for juxtaposition of reacting end groups [56]. Yet, in this setup, the bridging oligonucleotide design excluded 5–6 nt of each, the acceptor 3'-end and the donor 5'-end, from the RNA–DNA hybrid, creating a broken bulge loop structure in the RNA strand, while the DNA strand was entirely engaged in base pairing (Figure 3.7i).
- 5) **RNA end homogeneity:** A problem connected with *in vitro* transcription is 3'- and/or 5'-heterogeneity of RNA products, which may reduce ligation efficiency and compromise product homogeneity (for details, see Chapter 1). This problem can be solved either by sandwiching the RNA of interest between terminal *cis*-cleaving ribozymes (Figure 3.9 and Chapter 2), or by use of alternative approaches described in Section 3.2.3 and 3.2.4. When involving *cis*-cleaving ribozymes, however, it is important to keep in mind that the cleavage reaction produces a 2', 3'-cyclic phosphate at the 3'-end and a hydroxyl group at the 5'-terminus. A 5'-hydroxyl is optimal for acceptor substrates, but requires phosphorylation when present on the donor substrate. Conversely, a 2', 3'-cyclic phosphate nicely blocks the donor 3'-end, but has to be removed when present on an acceptor substrate. Protocols to remove 2', 3'-cyclic phosphates are described in Chapters 2 and 6.

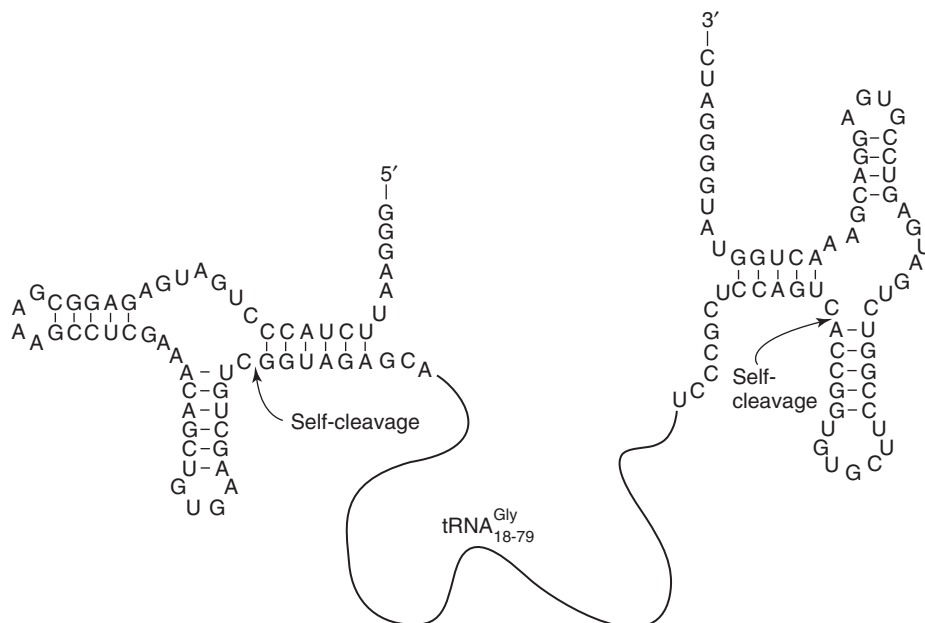


Figure 3.9 A primary transcript with self-cleaving hammerhead ribozyme structures at its 5'- and 3'-termini. During transcription, the flanking regions are removed

by ribozyme self-cleavage, releasing the internal RNA molecule of interest, in this case the donor substrate for the ligation illustrated in Figure 3.8b (nt 18–79 of tRNA^{Gly}).

3.4.5

Application Examples and Protocols

3.4.5.1 Production of Full-Length tRNAs

The T4-Rnl1-catalyzed ligation procedure was successfully used for the production of circa 80 nt long tRNA derivatives with site-specific modifications. In the first case, a 77 nt long *E. coli* tRNA^{Asp} was prepared by enzymatic ligation of two chemically synthesized oligonucleotides, each between 34 and 43 nt in length. The ligation sites were placed in the anticodon loop, the most explored region of tRNA for ligation by T4 Rnl1. In this study [72], analyzing the effect of 2'-deoxy modifications on aminoacylation, tRNA variants were prepared that either contained single-site 2'-deoxy modifications or had all, for example, A residues in the 5'- or 3'-half or even in the entire tRNA replaced with the 2'-deoxyA analog. The ligation position of tRNA halves containing 2'-deoxyA- and 2'-deoxyC modifications at every A and C position, respectively, was placed between anticodon nucleotides U33 and G34, while nucleotides C36 and A37 were selected as ligation site when the substrate halves contained 2'-deoxyG- and 2'-deoxyU modifications (Figure 3.8a). Ligation yields were between 30 and 50%. These findings illustrate that satisfactory product yields can be obtained despite seemingly unfavorable identities (Section 3.4.2.1) of the acceptor 3'-terminus (U33) and the donor 5'-terminus (G34).

In the second application, a bacterial precursor tRNA^{Gly} (ptRNA^{Gly} from *Thermus thermophilus*) was prepared from a 24 nt acceptor substrate obtained by chemical synthesis (representing the 5'-portion of the ptRNA) and a 62 nt donor substrate representing the 3'-proximal portion of the ptRNA and generated by T7 RNA transcription (Figure 3.8b). The 24 nt acceptor oligonucleotide carried 5'- and 3'-terminal hydroxyl groups, as routinely present in chemically synthesized RNAs. The 62 nt donor RNA was released from a primary transcript with terminal *cis*-hammerheads (Figure 3.9), generating the aforementioned 5'-hydroxyl and 2', 3'-cyclic phosphate end groups. Before ligation, the 5'-end was phosphorylated (Protocol 12; see also Section "Protocols for Ligation by T4 DNA ligase (T4 Dnl)", Protocol 6) using T4 PNK. Here, the purpose was to study the effect of ribose modifications at nt -1 of ptRNA^{Gly} on catalysis by *E. coli* RNase P RNA [74]. The ligation site was placed in the D loop, between positions C17 and G18 (Figure 3.8b), to minimize the length of the chemically synthesized RNA oligonucleotides carrying the single-site modification and thus to reduce the costs of chemical synthesis. The ligation yield was about 50% (Protocol 13), again despite unfavorable identity of the donor 5'-terminus (G18).

Protocol 14 describes ligation of two RNA oligonucleotides using T4 Rnl1 and a DNA splint, which was the most efficient approach towards the synthesis of a 129 nt long aptamer.

Specific Protocols

Some of the routine buffers and procedures used are detailed in Section "General Buffers and Procedures (GBP)." For detailed protocols on preparative *in vitro* transcription (0.2–1 ml scale) by T7 RNA polymerase, posttranscriptional DNA template digestion with RNase-free DNase I, phenol/chloroform extractions, ethanol precipitation, and gel elution of RNA, the reader is referred to Section "Protocols for Ligation by T4 DNA ligase (T4 Dnl)" and Chapter 1.

Protocol 11: RNA Preparation

- 1) *In vitro* transcription (1 ml scale) by T7 RNA polymerase.
- 2) DNA template digestion with RNase-free DNase I.
- 3) Phenol extraction followed by chloroform extraction (see Chapter 1, Section 1.3.4, step 3, therein). It may further be advisable to remove excess NTPs and/or salt components before EtOH precipitation (step 4) by use of NAP 10 columns (GE Healthcare) or equivalent matrices. Removal of excess NTPs is, in general, necessary only when using ³²P-radiolabeled nucleotides to avoid background radioactivity in purification gels. Salt components may impair band separation by denaturing PAGE, particularly when optimal gel resolution is essential, such as in separations of full-length RNA product from *n* + 1 and *n* - 1 species.
- 4) Precipitate the RNA with ethanol.

- 5) Dissolve the RNA pellet in 20–40 μl RNase-free water and purify the RNA of interest by denaturing PAGE according to Protocol 15.
- 6) Finally, determine the RNA concentration by UV spectroscopy.

Protocol 12: 5'-Phosphorylation of Donor Oligonucleotide

Note that T4 PNK requires a free hydroxyl group at the 5'-terminus. The protocol is given for a total reaction volume of 50 μl .

- T4 PNK (10 U μl^{-1} , Fermentas #EK0031)
 - 10 x T4 PNK buffer (Fermentas): 500 mM Tris-Cl (pH 7.6 at 25 °C), 100 mM MgCl_2 , 50 mM DTT, 1 mM spermidine, 1 mM EDTA
 - 100 mM ATP, adjusted to pH 7.0 with NaOH.
- 1) Combine 3 nmol of donor RNA with 5 μl 10 x T4 PNK buffer, 2.5 μl 100 mM ATP, and RNase-free water to a final volume of 47.5 μl ; mix gently.
 - 2) Add 2.5 μl T4 PNK (final concentration 0.5 U μl^{-1}) and incubate the mixture at 37 °C for 1 h.
 - 3) To increase the 5'-phosphorylation efficiency, add another 2.5 μl 100 mM ATP and 2.5 μl T4 PNK, and incubate the mixture for another 1 h at 37 °C. Either directly add 55 μl 2 x gel loading buffer for PAGE purification (Protocol 15) or proceed to step 4.
 - 4) Precipitate the RNA with ethanol.
 - 5) Dissolve the RNA pellet in about 20 μl RNase-free water and purify the RNA by denaturing PAGE according to Protocol 15.
 - 6) Determine the RNA concentration by UV spectroscopy.

5'-Phosphorylation protocols often include the addition of an RNase inhibitor in step 1 (final concentration 0.5–1.0 U μl^{-1} ; see Section “Protocols for Ligation by T4 DNA ligase (T4 Dnl),” Protocol 6). According to our experience, this is not necessary but may be considered in the case of problems with RNA degradation.

Protocol 13: Ligation with T4 RnlI

The protocol described below was used to generate full-length *E. coli* tRNA^{Asp} and *T. thermophilus* ptRNA^{Gly}. The tRNA^{Asp} was generated by ligation of pairs of chemically synthesized oligoribonucleotides, 34–43 nt in length (Figure 3.8a); ptRNA^{Gly} was generated by ligation of its 3'-portion (nucleotides 18–79), transcribed *in vitro*, to a chemically synthesized 24-meric oligoribonucleotide contributing the ptRNA^{Gly} 5'-portion (Figure 3.8b). For the chemical synthesis and purification of oligoribonucleotides, see Chapter 7 of this edition and Chapter 4 of the first edition of the *Handbook of RNA Biochemistry* [75].

- 1) If the donor 3'-oligonucleotide does not contain a 5'-phosphate, 5'-phosphorylate according to Protocol 12 before ligation. It is recommended to check the purity of the oligoribonucleotides to be used in the ligation reaction by analytical denaturing PAGE. If necessary, also gel-purify the freshly phosphorylated donor oligoribonucleotide as described in Protocol 15.

- 2) Ligation reactions are performed in a total volume of 200 μl (reached in step 5). As the first step, combine 3 nmol of 5'-phosphorylated donor oligoribonucleotide and 4.5 nmol of the corresponding acceptor oligoribonucleotide with:

Final concentration in 200 μl		
20 μl	1 M HEPES-KOH, pH 7.5	100 mM
20 μl	100 mM DTT	10 mM
To 156 μl	With RNase-free water	

- 3) Denature for 3 min at 90 $^{\circ}\text{C}$, followed by incubation for 10 min at 65 $^{\circ}\text{C}$ and slow cooling (45 min) to ambient temperature in a metal block removed from the heating apparatus to anneal the oligoribonucleotides.
- 4) After the cooling step, add:

Final concentration in 200 μl		
12 μl	250 mM MgCl_2	15 mM
1 μl	100 mM ATP	0.5 mM
20 μl	100% DMSO	10% (v/v)
3.75 μl	25–40 $\text{U } \mu\text{l}^{-1}$ RNase inhibitor (Fermentas or Promega)	0.5–0.75 $\text{U } \mu\text{l}^{-1}$

Mix gently.

- 5) Add 7 μl of T4 Rnl1 (stock solution 20 000 U ml^{-1} ; final concentration 0.7 $\text{U } \mu\text{l}^{-1}$, NEB), resulting in the final volume of 200 μl . Incubate at 16 $^{\circ}\text{C}$ for about 12–15 h.
- 6) To analyze ligation efficiency, withdraw 1.5 μl from the ligation reaction and mix with 10 μl loading buffer, load onto a denaturing 8–12% PAA gel, stain the gel with ethidium bromide, and visualize RNA bands by exposure to UV light.
- 7) Ethanol-precipitate the bulk of the ligation reaction.
- 8) Dissolve the resulting RNA pellet in 15–30 μl RNase-free water and purify the RNA by denaturing PAGE according to Protocol 15.
- 9) Determine the RNA concentration by UV spectroscopy.

Protocol 14: Ligation with T4 Rnl1 and DNA Splint

The protocol described below was used to generate the full-length *aptamer* RNA of the flavine mononucleotide-dependent riboswitch from *Bacillus subtilis* by ligation

of pairs of chemically synthesized oligoribonucleotides, 64 and 65 nt in length, and a 51 nt DNA splint, forming 23 base pairs with the acceptor oligonucleotide and 24 base pairs with the donor nucleotide, such that 2 nucleotides of both donor and acceptor strand remained single-stranded at the ligation site. This setup is similar to the one illustrated in Figure 3.7i. The yield of ligation product was 42%.

For the chemical synthesis and purification of oligoribonucleotides, see Chapter 7 of this edition and Chapter 4 of the first edition of the *Handbook of RNA Biochemistry* [75].

- 10 x T4 RNA Ligase buffer (500 mM Tris-HCl, 100 mM MgCl₂, 10 mM DTT, pH 7.5, NEB)
 - T4 RNA Ligase, 10 000 U ml⁻¹, NEB cat. no. M0204L
 - DNase I (Fermentas, 1 U µl⁻¹)
- 1) 5'-phosphorylate the donor oligonucleotide according to Protocol 12. It is recommended to 5'-phosphorylate the donor oligonucleotide directly prior to use in the ligation reaction. If necessary, gel-purify the freshly phosphorylated donor oligoribonucleotide as described in Protocol 15. The purity of all oligoribonucleotides to be used in the ligation reaction should be checked by analytical denaturing PAGE.
 - 2) Ligation reactions are performed in a total volume of 20 µl (reached in step 5). At first, combine the 5'-phosphorylated donor oligoribonucleotide (final concentration 20 µM), the corresponding acceptor oligoribonucleotide (final concentration 10 µM) and the DNA splint (final concentration 15 µM) with:

Final concentration in 20 µl		
2 µl	10 x T4 RNA Ligase Puffer	1 x
2 µl	100% DMSO	10% (v/v)
2 µl	10 mM ATP	1 mM
To 18 µl	With RNase-free water	

- 3) Denature for 5 min at 90 °C, followed by cooling over 15 min to 22 °C.
- 4) Add 2 µl of T4 RNA Ligase (final concentration 1 U µl⁻¹), resulting in the final volume of 20 µl. Incubate at 16 °C overnight.
- 5) Add 1 µl DNase I (final concentration 0.05 U µl⁻¹) and incubate 30 min at 37 °C.
- 6) Ethanol-precipitate the bulk of the ligation reaction.
- 7) Dissolve the resulting RNA pellet in 5 – 10 µl RNase-free water and purify the RNA by denaturing PAGE according to Protocol 15 below.

Protocol 15: Preparative Purification of RNA by Denaturing PAGE

The appropriate gel concentration depends, as usual, on the size of the RNA; a 12% PAA/8M urea sequencing gel (40 cm long, 1 mm thick) was used for the

purification of ligated tRNA^{Asp} (77 nt), and a corresponding 10% gel for the purification of ligated ptRNA^{Gly} (86 nt). To facilitate localizing the RNA product on the gel, it is advisable to increase the concentration of the product band by pooling three separate ligation reactions ($3 \times 15\text{--}30\ \mu\text{l}$), each based on 3 nmol of input donor RNA (see Protocol 13, step 8).

- 1) Mix the 45–90 μl of pooled ligation product with 45–90 μl 2 x gel loading buffer and pipette into a 2 cm broad well of the PAA/8 M urea gel. As an example, the above-mentioned 86 nt ligation product was run on a circa 40 cm long, 1 mm thick 10% PAA/8 M urea gel until the xylene cyanol marker had migrated 22–25 cm from the top.
- 2) Detect the ligated product by UV shadowing, excise from the gel and elute from crushed gel slices in elution buffer overnight at 4 °C. A detailed description of this procedure is given in Section “Protocols for Ligation by T4 DNA ligase (T4 Dnl),” Protocol 3.
- 3) Concentrate the eluted RNA by ethanol precipitation and redissolve in 20–40 μl RNase-free water.
- 4) Determine the RNA concentration by UV spectroscopy.

3.4.6

Troubleshooting

Low yields of the ligation reaction may have the following reasons:

- 1) Check for unfavorable secondary structure formation of the RNA fragments, particularly at the ligation joint. For this purpose, software such as OLIGO version 4.0 (National Bioscience), Mfold [76], or RNAfold [77] can be employed.
- 2) Heterogeneous 3'-ends of RNA transcripts: for RNA fragments of up to about 40 nt, it is usually sufficient to purify the RNA by preparative PAGE in the presence of 7.5–8.0 M urea before the ligation reaction. However, for RNA transcripts longer than approximately 50 nt, RNAs slightly differing in length from the main product are hard to get rid of by preparative gel purification. Methods to eliminate the problem of 5'- and 3'-end heterogeneities are described in this chapter (Sections 3.2.3 and 3.2.4) and Chapter 2.
- 3) RNA degradation: this is usually due to RNase contamination in water or solutions. Prepare and store RNase-free water as described in Section “General Buffers and Procedures (GBP).” Check individual solutions for RNase activity, and prepare fresh buffers and solutions.

References

1. Verma, S. and Eckstein, F. (1998) Modified oligonucleotides: synthesis and strategy for users. *Annu. Rev. Biochem.*, **67**, 99–134.
2. Persson, T., Kutzke, U., Busch, S., Held, R., and Hartmann, R.K. (2001) Chemical synthesis and biological investigation of a 77-mer oligoribonucleotide

- with a sequence corresponding to *E. coli* tRNA(Asp). *Bioorg. Med. Chem.*, **9**, 51–56.
3. Vargeese, C., Carter, J., Yegge, J., Krivjansky, S., Settle, A., Kropp, E., Peterson, K., and Pieken, W. (1998) Efficient activation of nucleoside phosphoramidites with 4,5-dicyanoimidazole during oligonucleotide synthesis. *Nucleic Acids Res.*, **26**, 1046–1050.
 4. Hogrefe, R.I., McCaffrey, A.P., Borozdina, L.U., McCampbell, E.S., and Vaghefi, M.M. (1993) Effect of excess water on the desilylation of oligoribonucleotides using tetrabutylammonium fluoride. *Nucleic Acids Res.*, **21**, 4739–4741.
 5. Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzales, C., Scaringe, S., and Usman, N. (1995) Synthesis, deprotection, analysis and purification of RNA and ribozymes. *Nucleic Acids Res.*, **23**, 2677–2684.
 6. Bullard, D.R. and Bowater, R.P. (2006) Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4. *Biochem. J.*, **398**, 135–144.
 7. Ho, C.K. and Shuman, S. (2002) Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12709–12714.
 8. McLaughlin, L.W., Piel, N., and Graeser, E. (1985) Donor activation in the T4 RNA ligase reaction. *Biochemistry*, **24**, 267–273.
 9. Heaphy, S., Singh, M., and Gait, M.J. (1987) Effect of single amino acid changes in the region of the adenyllylation site of T4 RNA ligase. *Biochemistry*, **26**, 1688–1696.
 10. Wang, L.K., Ho, C.K., Pei, Y., and Shuman, S. (2003) Mutational analysis of bacteriophage T4 RNA ligase 1. Different functional groups are required for the nucleotidyl transfer and phosphodiester bond formation steps of the ligation reaction. *J. Biol. Chem.*, **278**, 29454–29462.
 11. Tessier, D.C., Brousseau, R., and Vernet, T. (1986) Ligation of single-stranded oligodeoxyribonucleotides by T4 RNA ligase. *Anal. Biochem.*, **158**, 171–178.
 12. Blondal, T., Thorisdottir, A., Unnsteinsdottir, U., Hjorleifsdottir, S., Avarsson, A., Ernstsson, S., Fridjonsson, O.H., Skirnisdottir, S., Wheat, J.O., Hermannsdottir, A.G. *et al.* (2005) Isolation and characterization of a thermostable RNA ligase 1 from a *Thermus scotoductus* bacteriophage TS2126 with good single-stranded DNA ligation properties. *Nucleic Acids Res.*, **33**, 135–142.
 13. Moore, M.J. and Sharp, P.A. (1992) Site-specific modification of pre-mRNA: The 2'-hydroxyl groups at the splice sites. *Science*, **256**, 992–997.
 14. Frilander, M.J. and Steitz, J.A. (2001) Dynamic Exchanges of RNA Interactions Leading to Catalytic Core Formation in the U12-Dependent Spliceosome. *Mol. Cell*, **7**, 217–226.
 15. Gozani, O., Feld, R., and Reed, R. (1996) Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev.*, **10**, 233–243.
 16. Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000a) The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.*, **19**, 6860–6869.
 17. Le Hir, H., Moore, M.J., and Maquat, L.E. (2000b) Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev.*, **14**, 1098–1108.
 18. MacMillan, A.M., Query, C.C., Allerson, C.R., Chen, S., Verdine, G.L., and Sharp, P.A. (1994) Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev.*, **8**, 3008–3020.
 19. Maroney, P.A., Romfo, C.M., and Nilsen, T.W. (2000) Functional recognition of the 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol. Cell*, **6**, 317–328.
 20. McConnell, T.S. and Steitz, J.A. (2001) Proximity of the invariant loop of U5

- snRNA to the second intron residue during pre-mRNA splicing. *EMBO J.*, **20**, 3577–3586.
21. Pasmán, Z. and García-Blanco, M.A. (1996) The 5' and 3' splice sites come together via a three dimensional diffusion mechanism. *Nucleic Acids Res.*, **24**, 1638–1645.
 22. Query, C.C., Moore, M.J., and Sharp, P.A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.*, **8**, 587–597.
 23. Moore, M.J. and Query, C.C. (1998), Uses of site-specifically modified RNAs constructed by RNA ligation in *RNA:Protein Interactions A Practical Approach* (ed. C.W.J. Smith), Oxford University Press, Oxford, pp. 75–108.
 24. Reed, R. and Chiara, M.D. (1999) Identification of RNA-protein contacts within functional ribonucleoprotein complexes by RNA site-specific labeling and UV crosslinking. *Meth. Enzymol.*, **18**, 3–12.
 25. Yu, Y.T. (1999) Construction of 4-thiouridine site-specifically substituted RNAs for cross-linking studies. *Meth. Enzymol.*, **18**, 13–21.
 26. Kleppe, K., van de Sande, J.H., and Khorana, H.G. (1970) Polynucleotide Ligase-Catalyzed Joining of Deoxyribo-oligonucleotides on Ribopolynucleotide Templates and of Ribo-oligonucleotides on Deoxyribopolynucleotide Templates. *Proc. Natl. Acad. Sci. U.S.A.*, **67**, 68–73.
 27. Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.
 28. Milligan, J.F. and Uhlenbeck, O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Meth. Enzymol.*, **180**, 51–62.
 29. Sontheimer, E.J. and Steitz, J.A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262**, 1989–1996.
 30. Wyatt, J.R., Sontheimer, E.J., and Steitz, J.A. (1992) Site-specific crosslinking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev.*, **6**, 2542–2553.
 31. Krupp, G. (1988) RNA synthesis: strategies for the use of bacteriophage RNA polymerases. *Gene*, **72**, 75–89.
 32. Kao, C., Zheng, M., and Rüdiger, S. (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA*, **5**, 1268–1272.
 33. Lapham, J., Yu, Y.T., Shu, M.D., Steitz, J.A., and Crothers, D.M. (1997) The position of site-directed cleavage of RNA using RNase H and 2'-O-methyl oligonucleotides is dependent on the enzyme source. *RNA*, **3**, 950–951.
 34. Berkower, I., Leis, J., and Hurwitz, J. (1973) Isolation and Characterization of an Endonuclease from *Escherichia coli* Specific for Ribonucleic Acid in Ribonucleic Acid Deoxyribonucleic Acid Hybrid Structures. *J. Biol. Chem.*, **248**, 5914–5921.
 35. Inoue, H., Hayase, Y., Iwai, S., and Ohtsuka, E. (1987) Sequence-dependent hydrolysis of RNA using modified oligonucleotide splints and RNase H. *FEBS Lett.*, **215**, 327–330.
 36. Lapham, J. and Crothers, D.M. (1996) RNase H cleavage for processing of *in vitro* transcribed RNA for NMR studies and RNA ligation. *RNA*, **2**, 289–296.
 37. Dawson, R.M.C. Elliott, D.C. Elliott, W.H. and Jones, K.M. (1989) *Data for Biochemical Research*, Oxford University Press, Oxford, UK.
 38. Kobitski, A.Y., Hengesbach, M., Helm, M. and Nienhaus, G.U. (2008) Sculpting an RNA conformational energy landscape by a methyl group modification—a single-molecule FRET study. *Angew. Chem. Int. Ed.*, **47**, 4326–4330.
 39. Stone, M.D. Mihalusova, M., O'connor, C.M., Prathapam, R., Collins, K. and Zhuang, X. (2007) Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature*, **446**, 458–461.
 40. Voigts-Hoffmann, F., Hengesbach, M., Kobitski, A.Y., van Aerschot, A., Herdewijn, P., Nienhaus, G.U., and Helm, M. (2007) A methyl group controls conformational equilibrium in

- human mitochondrial tRNA(Lys). *J. Am. Chem. Soc.*, **129**, 13382–13383.
41. Hengesbach, M., Kobitski, A., Voigts-Hoffmann, F., Frauer, C., Nienhaus, G.U., and Helm, M. (2008) RNA intramolecular dynamics by single-molecule FRET. *Curr. Protoc. Nucleic Acid Chem.*, **11**, Unit 11.12.
 42. Hayrapetyan, A., Seidu-Larry, S. and Helm, M. (2009) in *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution* (ed. H. Grosjean), Landes Bioscience, Austin, TX, pp. 629–635.
 43. Arn, E.A. and Abelson, J. (1998), *J. RNA Ligases: Function, Mechanism, and Sequence Conservation in RNA Structure and Function* (eds R.W. Simons and M. Grunberg-Manago), Cold Spring Harbor, New York, pp. 695–726.
 44. Abelson, J., Trotta, C.R., and Li, H. (1998) tRNA splicing. *J. Biol. Chem.*, **273**, 12685–12688.
 45. Schnauffer, A., Panigrahi, A.K., Panicucci, B., Igo, R.P. Jr., Wirtz, E., Salavati, R., and Stuart, K. (2001) An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei*. *Science*, **291**, 2159–2162.
 46. Silber, R., Malathi, V.G., and Hurwitz, J. (1972) Purification and properties of bacteriophage T4-induced RNA ligase. *Proc. Natl. Acad. Sci. U.S.A.*, **69**, 3009–3013.
 47. Kaufmann, G. and Littauer, U.Z. (1974) Covalent joining of phenylalanine transfer ribonucleic acid half-molecules by T4 RNA ligase. *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 3741–3745.
 48. Walker, G.C., Uhlenbeck, O.C., Bedows, E., and Gumpert, R.I. (1975) T4-induced RNA ligase joins single-stranded oligoribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 122–126.
 49. Amitsur, M., Levitz, R., and Kaufmann, G. (1987) Bacteriophage T4 anticodon nuclease, polynucleotide kinase and RNA ligase reprocess the host lysine tRNA. *EMBO J.*, **6**, 2499–2503.
 50. Tyndall, C., Meister, J., and Bickle, T.A. (1994) The *Escherichia coli* prr region encodes a functional type IC DNA restriction system closely integrated with an anticodon nuclease gene. *J. Mol. Biol.*, **237**, 266–274.
 51. Kaufmann, G., Klein, T., and Littauer, U.Z. (1974) T4 RNA ligase: substrate chain length requirements. *FEBS Lett.*, **46**, 271–275.
 52. Kaufmann, G. and Kallenbach, N.R. (1975) Determination of recognition sites of T4 RNA ligase on the 3' -OH and 5' -P termini of polyribonucleotide chains. *Nature*, **254**, 452–454.
 53. Sugino, A., Snopek, T.J., and Cozzarelli, N.R. (1977) Bacteriophage T4 RNA ligase. Reaction intermediates and interaction of substrates. *J. Biol. Chem.*, **252**, 1732–1738.
 54. England, T.E. and Uhlenbeck, O.C. (1978) Enzymatic oligoribonucleotide synthesis with T4 RNA ligase. *Biochemistry*, **17**, 2069–2076.
 55. Romaniuk, E., McLaughlin, L.W., Neilson, T., and Romaniuk, P.J. (1982) The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction. *Eur. J. Biochem.*, **125**, 639–643.
 56. Bain, J.D. and Switzer, C. (1992) Regioselective ligation of oligoribonucleotides using DNA splints. *Nucleic Acids Res.*, **20**, 4372.
 57. Atencia, E.A., Montes, M., Sillero, M.A.G., and Sillero, A. (2000) Several dinucleoside polyphosphates are acceptor substrates in the T4 RNA ligase catalyzed reaction. *Eur. J. Biochem.*, **267**, 1707–1714.
 58. Hinton, D.M., Baez, J.A., and Gumpert, R.I. (1978) T4 RNA Ligase joins 2'-deoxyribonucleoside 3',5'-bisphosphates to oligodeoxyribonucleotides. *Biochemistry*, **17**, 5091–5097.
 59. McCoy, M.I. and Gumpert, R.I. (1980) T4 ribonucleic acid ligase joins single-strand oligo(deoxyribonucleotides). *Biochemistry*, **19**, 635–642.
 60. Bruce, A.G. and Uhlenbeck, O.C. (1978) Reactions at the termini of tRNA with T4 RNA ligase. *Nucleic Acids Res.*, **5**, 3665–3677.
 61. Igloi, G.L. (1996) Nonradioactive labeling of RNA. *Anal. Biochem.*, **233**, 124–129.

62. Hausch, F. and Jäschke, A. (1997) Libraries of multifunctional RNA conjugates for the selection of new RNA catalysts. *Bioconjug. Chem.*, **8**, 885–890.
63. Kinoshita, Y., Nishigaki, K., and Husimi, Y. (1997) Fluorescence-, isotope- or biotin-labeling of the 5'-end of single-stranded DNA/RNA using T4 RNA ligase. *Nucleic Acids Res.*, **25**, 3747–3748.
64. Rigden, J.E. and Rezaian, M.A. (1992) *In vitro* synthesis of an infectious viroid: analysis of the infectivity of monomeric linear CEV. *Virology*, **186**, 201–206.
65. Skoric, D., Conerly, M., Szychowski, J.A., and Semancik, J.S. (2001) CEVd-induced symptom modification as a response to a host-specific temperature-sensitive reaction. *Virology*, **280**, 115–123.
66. Wang, L. and Ruffner, D.E. (1998) Oligoribonucleotide circularization by 'template-mediated' ligation with T4 RNA ligase: synthesis of circular hammerhead ribozymes. *Nucleic Acids Res.*, **26**, 2502–2504.
67. Nishigaki, K., Taguchi, K., Kinoshita, Y., Aita, T., and Husimi, Y. (1998) Y-ligation: an efficient method for ligating single-stranded DNAs and RNAs with T4 RNA ligase. *Mol. Divers.*, **4**, 187–190.
68. Sugimoto, N., Matsumura, A., Hasegawa, K., and Sasaki, M. (1991) Effect of Unpaired Terminal Nucleotides of Substrate RNAs on the RNA Ligation with T4 RNA Ligase. *Bull. Chem. Soc. Jpn.*, **64**, 2978–2982.
69. Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E.G.H., Margalit, H., and Altuvia, S. (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.*, **11**, 941–950.
70. Liu, X. and Gorovsky, M.A. (1993) Mapping the 5' and 3' ends of *Tetrahymena thermophila* mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE). *Nucleic Acids Res.*, **21**, 4954–4960.
71. Ohtsuki, T., Kawai, G., and Watanabe, K. (1998) Stable isotope-edited NMR analysis of *Ascaris suum* mitochondrial tRNAMet having a TV-replacement loop. *J. Biochem.*, **124**, 28–34.
72. Vörtler, C.S., Fedorova, O., Persson, T., Kutzke, U., and Eckstein, F. (1998) Determination of 2'-hydroxyl and phosphate groups important for aminoacylation of *Escherichia coli* tRNAAsp: a nucleotide analogue interference study. *RNA*, **4**, 1444–1454.
73. Sherlin, L.D., Bullock, T.L., Nissan, T.A., Perona, J.J., Lariviere, F.J., Uhlenbeck, O.C., and Scaringe, S.A. (2001) Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization. *RNA*, **7**, 1671–1678.
74. Persson, T., Cuzic, S., and Hartmann, R.K. (2003) Catalysis by RNase P RNA: unique features and unprecedented active site plasticity. *J. Biol. Chem.*, **278**, 43394–43401.
75. Persson, T., Willkomm, D.K., and Hartmann, R.K. (2005) Handbook of RNA Biochemistry, Wiley-VCH Verlag, Weinheim, Germany, pp. 53–74.
76. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
77. Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.

4

Northern Blot Detection of Small RNAs

Benedikt M. Beckmann, Arnold Grünweller, and Roland K. Hartmann

4.1

Introduction

During the past years, new classes of small RNAs (sRNA) have been discovered [1, 2], also owing to advances in high-throughput sequencing. Characterization of these short RNA species often requires modifications of established methods, primarily, RT-PCR and Northern hybridization, to ensure sensitive and specific detection and quantification. Although quantification of sRNAs can be achieved by qRT-PCR (such as the stem-loop primer technique developed for miRNAs [3]), the Northern blot has remained indispensable, as it permits visualizing and roughly quantifying cellular levels of RNAs and their processing intermediates relative to endogenous RNA standards, such as 5S rRNA. Analysis of sRNAs <30 nt (e.g., miRNAs) using Northern hybridization is challenging because of their limited interaction surface that increases the sensitivity to steric impediments caused by immobilization on hybridization membranes. Different adaptations to overcome these limitations have been described for mature miRNAs [4, 5], ~22 nt in length, but were insufficient for the detection of even smaller cellular RNAs in our hands. We have developed an approach combining 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) crosslinking with non-radioactive 5'-digoxigenin (DIG)-end-labeled mixer probes and native polyacrylamide gels to detect RNAs as short as ~14 nt with high sensitivity and specificity [6]. Apart from such very short cellular RNAs, the described Northern blot procedure can be easily adapted to larger sRNAs, mainly by using an *in vitro* transcribed RNA probe internally labeled with DIG-11-UTP.

4.1.1

Isolation of RNA

Purified cellular RNA of highest quality is usually the starting point of a Northern hybridization. Different purification methods and protocols have been described, and several commercial kits are available for different purposes and cell types.

4.1.1.1 Kits

Using a kit is a convenient way to obtain RNA extracts from cells. RNA preparation by this route is usually fast. Other advantages are quality-controlled enclosed buffers and solutions. A feature of kit-based procedures can be the exclusion of RNAs falling below or exceeding certain length limits, especially if spin columns are used for nucleic acid purification. As a potential drawback, RNA extracted from kits designed for a special purpose (e.g., extraction of sRNAs) do not fully represent the biological RNA distribution [7]. For this reason, we prefer total RNA extracts even for Northern hybridization of short RNAs.

4.1.1.2 Do it Yourself

We routinely use two robust methods to extract total RNA from bacterial cells and from eukaryotic cell lines:

- single-step method (Trizol)
- hot phenol method

The “single-step method” provides in general high-quality RNA because of the chaotropic reagent guanidinium isothiocyanate that inactivates cellular nucleases and therefore protects the RNA during the purification process [8, 9].

In our hands, the “hot phenol method” [10] usually results in higher amounts of purified RNA compared to the single-step method. It should, however, be taken into consideration that RNA duplexes will be (partially) denatured owing to the 65 °C heating step. Therefore, the hot phenol method is not recommended if the native fold of RNA(s) of interest or intermolecular RNA–RNA interactions are at risk to be irreversibly disrupted at 65 °C.

4.1.1.3 Quality Control

An important step after purification is the quality control of the cellular RNA. This is usually achieved by analyzing a small amount of RNA on a denaturing gel and subsequent ethidium bromide staining. The abundance of the two major cellular RNA species (16–18S and 23–28S ribosomal RNA) is then used as a marker for the overall integrity of the purified RNA. For details, see the troubleshooting section.

4.1.2

Native *versus* Denaturing Gels

Electrophoretic separation of RNAs is usually achieved by denaturing gels. For long RNAs (>200 nt), agarose gels containing glyoxal or formaldehyde [11] may be used. For shorter RNAs, polyacrylamide (PAA) gels containing up to 8 M urea are commonly utilized. As our focus is on short RNAs, only PAA gels will be discussed.

Native PAA gels provide the possibility to detect intermolecular RNA–RNA interactions, to identify conformational differences between RNA molecules of identical length and to evaluate the conformational homogeneity of single RNA

species. We have reported evidence that native compared to denaturing gels increase the detection sensitivity for RNAs shorter than 20 nt in protocols involving EDC crosslinking [6]. Apparently, urea interferes with EDC chemistry at the membrane surface, resulting in less efficient RNA immobilization and thus weaker Northern blot signals [6]. Nonetheless, at least for miRNAs, robust Northern blot protocols involving denaturing (8 M urea) PAA gels have been developed [12].

4.1.3

Transfer of RNA and Fixation to Membranes

To transfer short RNAs from PAA gels to nylon membranes, we use semidry blotting in 0.5 x TBE as transfer buffer for at least 6 h (for details, see Section 4.2).

For subsequent RNA fixation on the membrane, several methods are in use, the most common techniques being UV crosslinking or baking of the membrane at 80–120 °C. These methods, however, entail the disadvantage that the nitrogenous bases, for example uracil, form covalent bonds to free amine groups of the membrane surface [4, 5]. These bases are then no longer available for Watson–Crick base pairing to the complementary probe (Figure 4.1), leading to a decrease of signal strength. To circumvent this problem, which deteriorates for short RNAs, chemical crosslinking with EDC can be applied [4, 5]. Here, primarily the 5′-phosphate group of RNA is covalently linked to a membrane amine group, leaving the bases accessible to base pairing with the probe. Chemical EDC crosslinking is thus most efficient for 5′-phosphorylated RNA, but also works, with reduced efficiency, for RNAs carrying a 5′-OH end, apparently through reaction of internal phosphodiesterases [6].

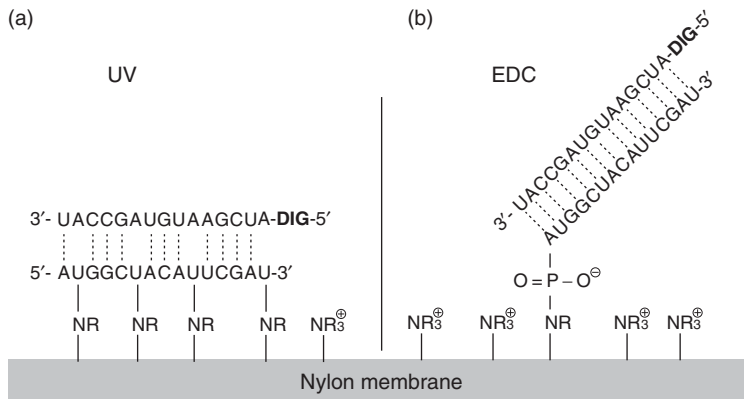


Figure 4.1 Schematic presentation of differences between UV- and EDC crosslinking procedures. (a) Treatment of RNA with UV light leads to covalent attachment of bases to amine groups of the nylon membrane. Probe hybridization to these

and likely also neighboring bases is then precluded. (b) EDC crosslinking primarily occurs via 5′-phosphate groups, thus leaving the nucleobases unaffected and consequently accessible to pairing with the probe.

4.1.4

Hybridization with a Complementary Probe

For hybridization to RNA, RNA or DNA probes can be used. For routine Northern hybridization, we use RNA probes generated by T7 run-off transcription from PCR fragments or linearized plasmid DNAs. These RNA probes can either be labeled post-transcriptionally (5'-endlabeling with $\gamma^{32}\text{P}$ -ATP) or during transcription by the incorporation of radioactive or modified nucleotides such as $\alpha^{32}\text{P}$ -NTPs or DIG-11-UTP.

An effective strategy to improve binding of Northern probes to short RNAs is incorporation of LNA (locked nucleic acid) residues into DNA oligonucleotides [13, 14]. Such DNA/LNA mixmers increase duplex stability and thus permit applying elevated hybridization temperatures [15, 16], such as 72 °C. We have used 5'-DIG end-labeled DNA/LNA mixmer probes (custom-synthesized by Exiqon¹⁾) in combination with EDC crosslinking of short RNAs (<20 nt) to nylon membranes. This approach, but using denaturing PAGE, has also been applied to the detection of mature (~22 nt) miRNAs [12].

4.1.4.1 **Design of DNA/LNA Mixmer Probes**

Design of DNA/LNA mixmers as probes can be challenging, especially for short oligonucleotides. One of the main features of LNA is the fixation of the ribose moiety in its C3'-*endo* conformation leading to a significant increase in melting temperature [15, 16]. A single LNA residue stabilizes not only base pairing with the complementary RNA nucleotide but also that of immediate DNA neighbors with their complementary RNA bases. Thus, DNA/LNA mixmers confer substantial duplex stabilization, making the use of all-LNA probes unnecessary. In addition, for all-LNA probes, the risk of self-complementarity and dimerization increases. We apply several rules for our design of probes that are usually 12–16 nt in length (see also Figure 4.2):

- Run `mfold`²⁾ [17] and/or `RNAfold`³⁾ [18] with your probe sequence to check for secondary structures. Nucleotides predicted to form internal duplexes should not be considered for LNA positioning.
- To check for potential dimerization, run `mfold` or `RNAfold` with a tandem repeat of your sequence, the two identical sequence stretches connected by a “NNNNN” linker; alternatively, use `RNAcofold`⁴⁾ [19] (paste identical sequence in 1st and 2nd window). Avoid LNA modifications in regions predicted to be base-paired.
- Use 2–5 LNA modifications for a 9–16 nt probe.

1) <http://www.exiqon.com>.

2) <http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi>.

3) <http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>.

4) <http://rna.tbi.univie.ac.at/cgi-bin/RNAcofold.cgi>.

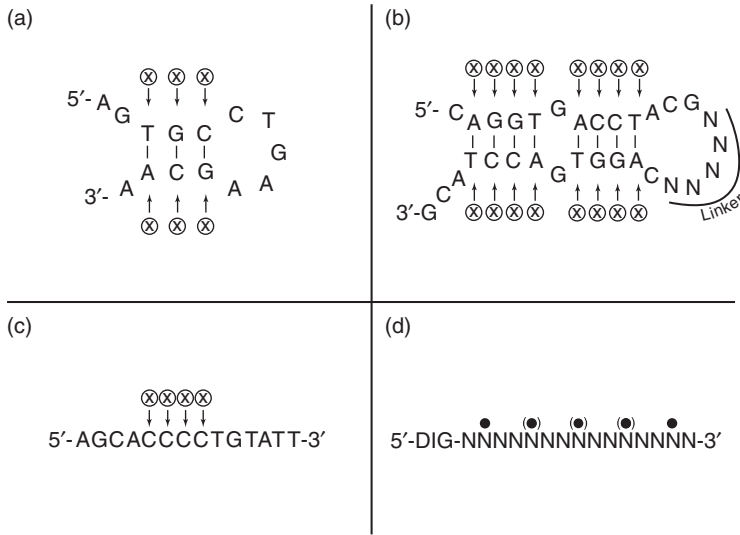


Figure 4.2 Rules for LNA/DNA mixer design of short (<20 nt) probes. Use a folding program to detect either (a) self-complementarity or (b) dimerization potential. Avoid LNA residues in such paired regions. (c) LNA modifications in C- or

G-stretches increase the danger of interactions with such motifs in other RNAs, for example, ribosomal RNAs; (X) omit LNA modification. (d) General scheme of placing LNA modifications; (X) place LNA modification, (●) optional LNA modification.

- Do not position LNA residues within homo-G or -C stretches (e.g., 5'-NNN GGG GNN NNN-3'; Figure 4.2c), as this will increase non-specific binding and will compromise sensitivity (Figure 4.3).
- Use BLAST⁵⁾ to screen for probe complementarity to other regions of the genome. Keep in mind, however, that only expressed RNAs may cause non-specific signals. If you find hits in annotated genes, place LNAs in positions that lack complementarity to this potentially cross-hybridizing cellular RNA.
- If the probe's length is too short (<18 nt) for BLAST, try other bioinformatic tools such as RNAbob⁶⁾ or fragrep [20].
- If possible, place LNA modifications at the second and the penultimate position of the probe to increase stability (Figure 4.2d).
- Keep in mind that LNA favors the C3'-endo conformation of its two neighboring DNA residues. Therefore, one LNA modification should be followed by two unmodified DNA nucleotides.
- A useful tool for LNA/DNA mixer design can also be found at Exiqon's homepage.

5) <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.

6) rnabob (<http://selab.janelia.org/software.html>) as well as fragrep (<http://www.bierinformatik.de/Software/fragrep/>) are not available as web-based services but have to be installed locally and run as console programs.

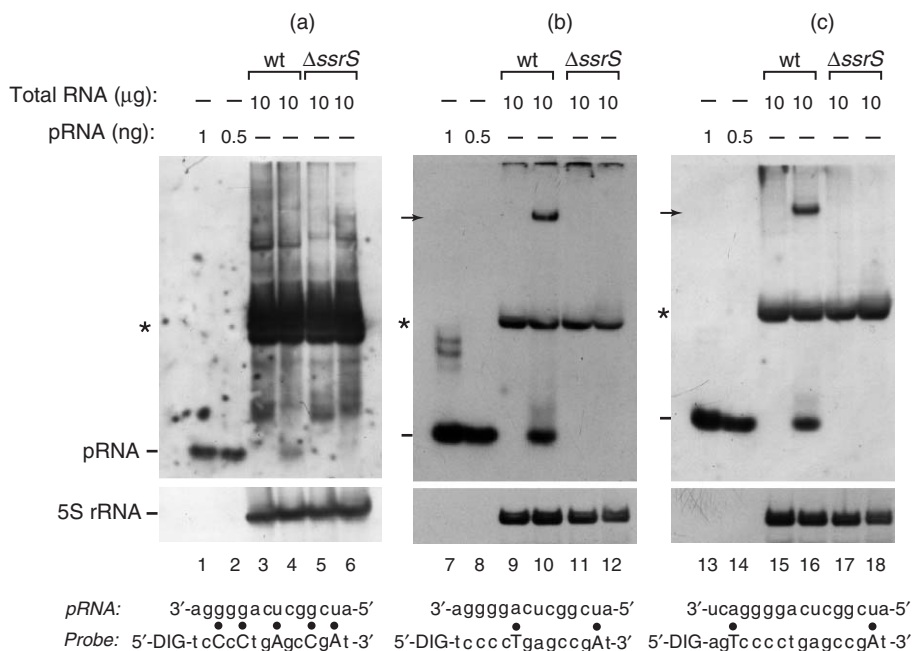


Figure 4.3 Specificity of different probes for the detection of pRNAs in RNA extracts from *E. coli*. (a–c) Detection of pRNAs derived from *E. coli* 6S RNA in total RNA extracts from the wild-type (wt) strain MG1655 (lanes 3, 4, 9, 10, 15, and 16) versus RNA extracts derived from a MG1655 6S RNA knockout ($\Delta ssrS$) strain (lanes 5, 6, 11, 12, 17, and 18) using three different probes shown below the corresponding panel; uppercase letters indicate LNA and lowercase letters DNA residues, with filled circles indicating to which *E. coli* pRNA residues the LNA residues base pair; the sequence of a chemically synthesized pRNA 14-mer is shown above the probes (panels a and b); in panel (c), a 16-mer probe was used to extend base pairing to longer endogenous pRNA by two nucleotides; these two nucleotides (lowercase letters) are not present in the chemically synthesized pRNA 14-mer (5'-AUC GGC UCA GGG GA-3'). Deviating from the standard procedure, RNA samples were heated in denaturing loading buffer for 5 min at 95 °C immediately before gel loading. Total RNA was extracted from cells in stationary phase (lanes 3, 5, 9, 11, 15, and 17) or during outgrowth (lanes 4, 6, 10, 12, 16, and

18) by the hot phenol method; for the definition of stationary phase and outgrowth, see Ref. [6]. The chemically synthesized *E. coli* pRNA 14-mer was loaded as size control in lanes 1, 2, 7, 8, 13, and 14 (pRNA positions marked by dashes at the left margin of each panel). Membranes were stripped after hybridization with pRNA probes and hybridized with a 5S rRNA-specific probe (as loading control) shown at the bottom of each panel. A pRNA signal was seen under outgrowth conditions for wt (wild-type) bacteria (lanes 4, 10, and 16) but not for $\Delta ssrS$ bacteria (lanes 6, 12, and 18). Probe 1 (panel a) was the least specific and yielded the highest background among the three different probes, which we attribute to the high number of 5 LNA residues, two of which are in the tetraacytidylate stretch where they favor base pairing with tetraguanlylates in other RNAs. Signals indicated by arrows in lanes 10 and 16 are 6S RNA-pRNA hybrids [6]; asterisks mark a cross-hybridizing RNA of unknown identity. The hybridization temperature was 72 °C in panel (a) and 68 °C in panels (b) and (c). Source: Adapted from Ref. [6].

4.1.5

Detection of DIG-Labeled Probes

Visualization of DIG-labeled probes is achieved by immunological detection using an antibody (Anti-Digoxigenin AP Fab Fragments) specific for the DIG-modified nucleotides. The antibody is coupled to an alkaline phosphatase that dephosphorylates a dioxetane phenyl phosphate (CDP-Star). The resulting chemoluminescent signal is then detected by exposing an X-ray film to the membrane. As DIG technology is owned by Roche,⁷⁾ we obtained all reagents and antibodies from this source.

4.1.6

Troubleshooting

Several problems may arise during this multistep procedure. Consider the following hints:

- **Quality of RNA:** When total RNA has been prepared, check its integrity before conducting a Northern blot. Prepare a denaturing 1% agarose gel [11] or a denaturing 5% polyacrylamide gel and load 2–3 µg total RNA using a denaturing loading buffer (see below). Run the gel until the bromophenol blue is close to the bottom and stain with ethidium bromide. If the RNA is intact, 28S and 18S rRNA species in mammals (23S and 16S rRNA in bacteria, respectively) should constitute the main bands, with the larger rRNA yielding roughly twice the staining intensity of the smaller rRNA. Both bands should be discrete; smearing indicates degradation of the RNA.
- **Loading control:** Loading the same amounts of total cellular RNA is not sufficient to ensure comparability between different samples. One possibility is to stain the gel with ethidium bromide before blotting and check for equal amounts of ribosomal or tRNA species. A commonly used strategy is the additional detection of a housekeeping RNA such as 5S rRNA. After detection of the RNA of interest, the membrane is stripped from its first probe (see below) and 5S rRNA levels are determined and compared with a 5S rRNA-specific probe.
- **Background signals:** There are many reasons for this problem; they also differ depending on membrane material, probe, and so on. The reader is referred to [11], Section 7.45, for a first troubleshooting. When using EDC crosslinking and/or DIG-labeled probes for short RNA detection, the following additional problems might arise:
 - Longer exposure of X-ray films to the AP Fab Fragments/CDP-Star solution leads to high background: as the CDP-Star solution is highly concentrated, dilute it 1 : 20 in detection buffer and increase exposure time by a factor of 2–3 [21].

7) <http://www.roche-applied-science.com>.

- Signal speckles may appear all over the membrane, likely caused by aggregation of DIG/LNA probes. Increase hybridization temperature up to 72 °C or design a probe with fewer LNA modifications.
- Shadow-like background signals: Centrifuge the anti-digoxigenin antibody before usage (5 min, 11000 g).
- Weak signal: Adjust hybridization temperature. Do not forget to heat the probe to 95 °C for 3 min directly before hybridization.
- RNA aggregation: Owing to the lack of denaturing reagents in the gel, RNA may form multimers. As low amounts of urea do not substantially disturb EDC crosslinking, the usage of a denaturing gel loading buffer in combination with heating the samples for up to 5 min at 95 °C was shown to resolve such multimers [6].

4.1.7

Application Example

We applied the described method to detect short RNA species in bacterial cells. In *Escherichia coli*, RNA polymerase synthesizes the so-called pRNAs, ~10–20 nt in length, using the abundant riboregulator 6S RNA as template [22]. Synthesis of pRNAs peaks during outgrowth from stationary phase (Figure 4.3).

4.1.8

Limitations of the Method

We further explored the performance of our method with respect to the detection of RNAs even shorter than 14 nt. In an example experiment, we spotted an RNA 14-mer, fully complementary to the 14-meric DNA/LNA probe, and an RNA 8-mer, complementary to the distal eight nucleotides of the DNA/LNA probe, on a nylon membrane (Figure 4.4a). Detection sensitivity was largely decreased for the RNA 8-mer compared to the “full length” 14-mer. However, the sensitivity of detection could be partially restored using an isosequential all-LNA 8-mer.

We then tested different temperatures and buffer conditions in order to improve detection of the RNA 8-mer. While lowering the hybridization temperature from 50 to 25 °C tended to slightly improve the lower detection limit, reductions in the concentration of the chaotropic urea did not improve detection sensitivity (Figure 4.4b). While the Roche buffer tends to minimize background signals, we see a slightly increased detection sensitivity for our simple phosphate buffer containing 4 M urea [23] in the case of 8-mer detection at both temperatures.

On the basis of these observations, we advice researchers trying to detect RNAs of such short length (<14 nt) to decrease the hybridization temperature to 25 °C, to use a hybridization buffer containing 4 M urea, and, optionally, to include (as positive control) an LNA-based mimic of the short RNA to be detected (see Figure 4.4a). However, be aware that small all-LNA probes tend to have a somewhat slower migration velocity in native PAA gels relative to their RNA counterparts (own observation).

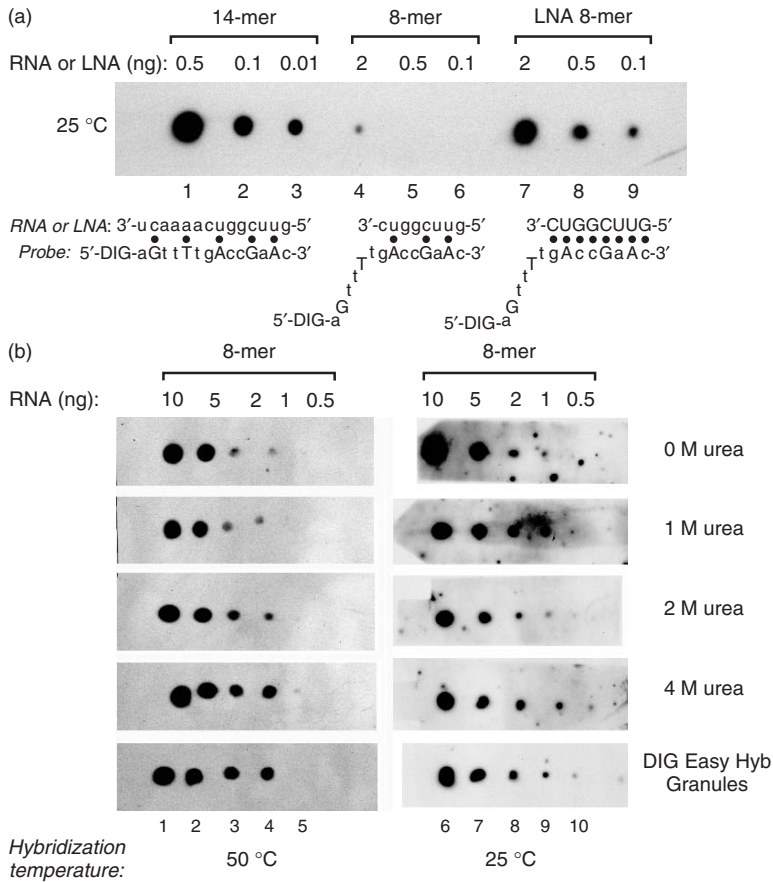


Figure 4.4 Detection sensitivity limitations. (a) Dot blot of chemically synthesized 5'-phosphorylated RNA 14- or 8-mers spotted directly onto the nylon membrane before EDC crosslinking; RNA and probe sequences are given at the bottom: the LNA 8-mer carried LNA instead of RNA residues indicated by uppercase letters; for the probes, uppercase letters indicate LNA and lowercase letters DNA residues; filled circles between target RNA (or LNA) and probe indicate base pairs involving stabilizing LNA residues. Detection sensitivity of the RNA 8-mer was estimated to be ≥ 40 -fold reduced compared to the 14-mer (cf. lanes 1–3 versus 4–6). The LNA 8-mer improved detection sensitivity relative to the isosequential RNA 8-mer; however, sensitivity for the LNA 8-mer was still ~ 20 -fold lower than for the RNA 14-mer (cf. lane 9 versus 3). (b)

Influence of hybridization temperature and buffer on detection sensitivity. “DIG Easy Hyb Granules” (bottom panels) describes the hybridization buffer provided by Roche, for which the composition is undisclosed, but said to correspond to 50% formamide in terms of nucleic acid denaturation efficacy. The other buffers contained 50 mM $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ pH 7.0, 0.02% (w/v) SDS, and indicated concentrations of urea. At the lower hybridization temperature of 25 °C, background speckles increased relative to 50 °C, particularly at 1 and 0 M urea (cf. lanes 1–5 versus 6–10). Reduction of the urea concentration below 4 M tended to impair detection sensitivity. Sensitivity appeared best for the simple phosphate-based buffer containing 4 M urea combined with a hybridization temperature of 25 °C.

4.2

Northern Hybridization Protocols

Protocol: RNA Extraction from Cells using Hot Phenol

- Extraction buffer: 10 mM NaOAc, 150 mM sucrose, adjust to pH 4.8 with acetic acid; pass solution through a sterile filter.
 - 10 x TE: 100 mM Tris-HCl pH 7.4, 10 mM EDTA.
 - Acid phenol: Use 1 x TE-saturated phenol and discard upper aqueous phase. Add 1/10 vol. of 300 mM NaOAc pH 4.9, vortex, centrifuge for 10 min at 3000 g, and discard upper aqueous phase. Repeat three times.
- 1) Resuspend 100 mg tissue or cell pellet (e.g., from 50 ml *E. coli* cell suspension at OD₆₀₀ 1.0 or up to 10⁶ HeLa cells) in 3 ml of precooled (4 °C) extraction buffer by vigorous vortexing in 15-ml Falcon tubes.
 - 2) (optional) If Gram-positive bacteria are used, add 100 µl lysozyme (20 mg ml⁻¹ in 1 x TE) and incubate at room temperature (RT) for 10 min or
 - 3) (optional) if RNA degradation occurs during lysozyme treatment, one may alternatively resuspend the frozen pellet (step 1) directly in phenol and use glass beads for cell disruption (see Ref. [24] for details).
 - 4) Add 333 µl 10% SDS to a final concentration of 1% SDS.
 - 5) Add 3 ml of preheated (65 °C) acid phenol, vortex.
 - 6) Incubate at 65 °C for 5 min.
 - 7) Incubate on ice for 5 min.
 - 8) Centrifuge for 20 min at 8200 g; transfer upper aqueous phase in 400-µl aliquots to Eppendorf tubes.
 - 9) Perform phenol and chloroform extractions: add 1 vol. of acid phenol, vortex and centrifuge for 5 min at 11000 g; keep upper aqueous phase and repeat with 1 vol. of chloroform.
 - 10) Precipitate RNA by adding 0.1 vol. of 3 M NaOAc pH 4.9 and 2.5 vol. of ethanol. Cool solution at -20 °C for 20 min and centrifuge for 30 min at 11000 g and 4 °C, discard supernatant.
 - 11) (optional) Wash pellet with 70% ethanol, centrifuge for 10 min at 11000 g and 4 °C, discard supernatant.
 - 12) Air-dry the pellets for 5 min before redissolving the RNA in a total volume of 20 µl of RNase-free double-distilled H₂O (ddH₂O) or 1 x TE.

Protocol: RNA Extraction by the "Single-Step Method"

- Denaturation solution: 4 M guanidinium isothiocyanate, 25 mM sodium citrate, 0.5% (w/v) *N*-lauroylsarcosyl, 0.1 M β-mercaptoethanol, pH 7.0.
- 10 x TE: 100 mM Tris-HCl pH 7.4, 10 mM EDTA.
- Acid phenol: Use 1 x TE-saturated phenol and discard upper aqueous phase. Add 1/10 vol of 300 mM NaOAc pH 4.9, vortex, centrifuge for 10 min at 3000 g, and discard upper aqueous phase. Repeat three times.

- 1) Resuspend 100 mg tissue or cell pellet (e.g., from 50 ml *E. coli* cell suspension at OD₆₀₀ 1.0 or up to 10⁶ HeLa cells) in 1 ml denaturation solution, vortex (in 15-ml Falcon tubes).
- 2) (optional) If Gram-positive bacteria are used, add 100 µl lysozyme (20 mg ml⁻¹ in 1 x TE) and incubate at RT for 10 min or
- 3) (optional) if RNA degradation occurs during lysozyme treatment, one may alternatively resuspend the frozen pellet (step 1) directly in phenol and use glass beads for cell disruption (see Ref. [24] for details).
- 4) Add 100 µl 2 M NaOAc pH 4.0, vortex.
- 5) Add 1 ml acid phenol and 200 µl chloroform, vortex.
- 6) Incubate on ice for 15 min.
- 7) Centrifuge for 20 min at 8200 g; transfer upper aqueous phase in 600-µl aliquots to Eppendorf tubes.
- 8) Precipitate by adding 1 vol. isopropanol, incubate at -20 °C for 20 min, centrifuge for 30 min at 11000 g and 4 °C and discard supernatant.
- 9) Dissolve pellet in 300 µl denaturation solution.
- 10) Precipitate again with isopropanol.
- 11) Wash pellet with 70% ethanol, centrifuge for 10 min at 11000 g and 4 °C, discard supernatant.
- 12) Redissolve RNA in a total volume of 20 µl RNase-free ddH₂O or 1 x TE.

Protocol: Northern Hybridization of Small RNA

PAGE and Transfer of RNA

- 5 x TBE: 445 mM Tris, 445 mM borate, 10 mM EDTA.
 - Native polyacrylamide gel: 1 x TBE, acrylamide/bisacrylamide (24:1), 10% APS (ammonium persulfate), TEMED (*N, N, N', N'*-tetramethylethylenediamine).
 - Denaturing polyacrylamide gel: 1 x TBE, 8 M urea, acrylamide/bisacrylamide (24:1), 10% APS, TEMED.
 - 6 x native loading buffer: 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol blue, 30% (v/v) glycerol.
 - 2 x denaturing loading buffer: 0.02% (w/v) bromophenol blue, 0.02% (w/v) xylene cyanol blue, 2.6 M urea, 66% (v/v) deionized formamide, 2 x TBE.
 - Whatman paper (>1 mm thickness).
 - Nylon membranes, positively charged (Roche, No. 11209299001); microporous nylon 66 membrane on a polyester support, carrying positively quaternary ammonium groups; surface properties: 0.45 µm pore size.
 - Semidry blotter (C.B.S. Scientific Co., model EBU-4000).
- 1) Redissolve 10 µg of ethanol-precipitated total RNA in 1 vol. of 2 x denaturing loading buffer or 0.2 vol. of native loading buffer, respectively.
 - 2) Incubate mixture at 95 °C for 3 min and put on ice immediately.
 - 3) Rinse gel pockets with a syringe, load the samples and run the gel using 1 x TBE as buffer (15 mA for a 15 × 20 cm sized gel with 1 mm thickness).

- 4) After the run, wash gel in 0.5 x TBE for 5 min at RT.
- 5) Transfer RNA to a positively charged nylon membrane using a semidry blotting apparatus (Figure 4.5): Place 1 piece of Whatman paper (gel size) soaked with 0.5 x TBE on the anode plate. Layer on top in the following order: membrane (soaked with 0.5 x TBE, gel size), the gel to be transferred, and a piece of Whatman paper (soaked with 0.5 x TBE, gel size). Put the cathode plate on top and run transfer overnight at 0.27 mA cm^{-2} . Take care to avoid any air bubbles between membrane and gel.

Fixation of Small RNAs by EDC CrossLinking

- EDC solution: 130 mM 1-methylimidazole, 160 mM 1-ethyl-3-(3-dimethylamino-propyl) carbodiimide, pH 8.0.
- 1) Soak a Whatman paper with EDC solution and place the membrane on it (membrane side with blotted RNA should not get in direct contact with the EDC solution to avoid detachment of RNA).
 - 2) Wrap in kitchen wrapping film and incubate for 2 h at 60°C .
 - 3) Wash membrane carefully with ddH_2O , wrap in kitchen wrapping film and store at 4°C until use.

Probe Generation by T7 Transcription

- 5 x T7 transcription buffer: 200 mM Tris-HCl, 30 mM MgCl_2 , 50 mM DTT, 50 mM NaCl, 10 mM spermidine, pH 7.9.
 - 10 x DIG labeling mix (Roche).
 - T7 RNA polymerase, for example, purchased from Fermentas⁸⁾ or NEB.⁹⁾
- 1) Mix T7 transcription reaction and incubate for 2 h at 37°C .
 - 2) Stop transcription reaction by adding $2 \mu\text{l}$ 10 mM EDTA and store at -20°C (Table 4.1).

Hybridization

Hybridization temperature depends on the probe's properties (length, G/C content, LNA modification, etc.) and on the hybridization solution. With the hybridization

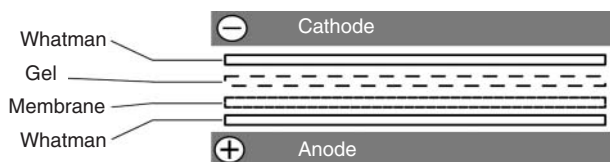


Figure 4.5 Northern blot assembly.

8) <http://www.fermentas.com>.

9) <http://www.neb.com>.

Table 4.1 T7 transcription reaction.

PCR template (200 ng μl^{-1})	5 μl
5 x T7 transcription buffer	4 μl
10 x DIG labeling mix	2 μl
T7 polymerase (10 U μl^{-1})	2 μl
ddH ₂ O	7 μl

buffers described here, we usually perform the first hybridization experiment at 68 °C in the case of probes transcribed by T7 RNA polymerase in the presence of DIG-11-UTP (see above), or at 50 °C in the case of 12–16 nt LNA/DNA mixmer probes. The example protocol is given for 50 °C.

- Hybridization solution: DIG Easy Hyb Granules (Roche) or
 - 50 mM Na₂HPO₄/NaH₂PO₄ pH 7.0, 0.02% (w/v) SDS, 4 M urea.
- 1) Preheat 7 ml of hybridization solution to 50 °C.
 - 2) Place the membrane in a hybridization tube (or a 50 ml Falcon tube for small membranes). Make sure that the “RNA side” does not touch the glass wall but is directed inward.
 - 3) Add hybridization solution and perform prehybridization at 50 °C for 2 h with slow rotation in a hybridization oven.
 - 4) Denature the probe (10 μl T7 transcript or 300 pmol DIG-LNA/DNA mixmer) for 3 min at 95 °C, put on ice immediately.
 - 5) Add the probe to 7 ml preheated (50 °C) hybridization solution, discard the prehybridization solution, and add solution containing the probe to the membrane.
 - 6) Perform hybridization overnight (or at least for 6 h) at 50 °C with slow rotation in a hybridization oven.

Detection of DIG-Labeled Probes

- 20 x SSC: 3 M NaCl, 0.3 M sodium citrate, pH 7.0.
- Stringent wash buffer 1: 2 x SSC, 0.1% (w/v) SDS.
- Stringent wash buffer 2: 0.1 x SSC, 0.1% (w/v) SDS.
- Maleic acid buffer: 0.1 M maleic acid, 0.15 M NaCl, pH 7.5.
- Blocking solution: 10 x Blocking solution (Roche) diluted 1 : 10 in maleic acid buffer.
- Antibody solution: Blocking solution with 1 : 10,000 diluted Anti-Digoxigenin-AP Fab Fragments (Roche).
- Wash buffer: 0.1 M maleic acid, 0.15 M NaCl, 0.3% (v/v) Tween 20, pH 7.5.
- Detection buffer: 0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5.
- CDP-Star (Roche).
- Autoradiography film: Kodak BioMax Light Film.

For detection, all steps should be carried out using clean glass/plastic ware. Take care that membranes are covered with buffer and that shaking takes place at low speed.

- 1) Wash membrane twice in stringent wash buffer 1 for 5 min at RT.
- 2) Wash membrane twice in stringent wash buffer 2 for 15 min at RT.
- 3) Wash membrane for 2 min in ddH₂O.
- 4) Incubate in blocking solution for 30 min at RT.
- 5) Incubate in antibody solution for 30 min at RT.
- 6) Wash twice in wash buffer for 15 min at RT.
- 7) Incubate in detection buffer for 5 min at RT.
- 8) Place membrane on a kitchen wrapping film, add some drops of CDP-Star until the membrane is covered with the solution, wrap it up completely in the kitchen wrapping film, and incubate for 10 min at RT to allow development of a robust chemical signal. Avoid any air bubbles.
- 9) Place an autoradiography film of identical size on the gel and expose it for 5 min to 1 h.
- 10) Develop the film according to the manufacturer's instructions.
- 11) Take care that the membrane does not run dry after detection if subsequent stripping is intended!

Stripping of Membranes

- Stripping buffer: 50% deionized formamide, 50 mM Tris-HCl, pH 7.5, 5% (w/v) SDS, always freshly prepared.
 - 2 x SSC: 300 mM NaCl, 30 mM sodium citrate, pH 7.0.
- 1) Wash membrane with ddH₂O.
 - 2) Incubate twice in stripping buffer for 60 min at 80 °C while shaking gently.
 - 3) Wash twice in 2 x SSC for 5 min at RT.
 - 4) Start with prehybridization and prepare for hybridization with new probe (see above).

References

1. Waters, L.S. and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell*, **136**, 615–628.
2. Ghildiyal, M. and Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.*, **10**, 94–108.
3. Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R., Lao, K.Q., Livak, K.J., and Guegler, K.J. (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.*, **33**, e179.
4. Pall, G.S., Codony-Servat, C., Byrne, J., Ritchie, L., and Hamilton, A. (2007) Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res.*, **35**, e60.
5. Pall, G.S. and Hamilton, A.J. (2008) Improved northern blot method for enhanced detection of small RNA. *Nat. Protoc.*, **3**, 1077–1084.
6. Beckmann, B.M., Grünweller, A., Weber, M.H., and Hartmann, R.K.

- (2010) Northern blot detection of endogenous small RNAs (approximately 14 nt) in bacterial total RNA extracts. *Nucleic Acids Res.*, **38**, e147.
7. Mraz, M., Malinova, K., Mayer, J., and Pospisilova, S. (2009) MicroRNA isolation and stability in stored RNA samples. *Biochem. Biophys. Res. Commun.*, **390**, 1–4.
 8. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156–159.
 9. Chomczynski, P. and Sacchi, N. (2006) The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat. Protoc.*, **1**, 581–585.
 10. Mattatall, N.R. and Sanderson, K.E. (1996) *Salmonella typhimurium* LT2 possesses three distinct 23S rRNA intervening sequences. *J. Bacteriol.*, **178**, 2272–2278.
 11. Sambrook, J. and Russel, D.W. (2001) *Molecular Cloning*, 3rd edn, Cold Spring Harbor Laboratory Press, Chapter 7.
 12. Kim, S.W., Li, Z., Moore, P.S., Monaghan, A.P., Chang, Y., Nichols, M., and John, B. (2010) A sensitive non-radioactive northern blot method to detect small RNAs. *Nucleic Acids Res.*, **38**, e98.
 13. Várallyay, E., Burgyán, J., and Havelda, Z. (2008) MicroRNA detection by northern blotting using locked nucleic acid probes. *Nat. Protoc.*, **3**, 190–196.
 14. Válczi, A., Hornyik, C., Varga, N., Burgyán, J., Kauppinen, S., and Havelda, Z. (2004) Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. *Nucleic Acids Res.*, **32**, e157.
 15. Vester, B. and Wengel, J. (2004) LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. *Biochemistry*, **43**, 13, 233–13, 241.
 16. Grünweller, A. and Hartmann, R.K. (2007) Locked nucleic acid oligonucleotides: the next generation of antisense agents? *BioDrugs*, **21**, 235–243.
 17. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
 18. Hofacker, I.L. and Stadler, P.F. (2006) Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics*, **22**, 1172–1176.
 19. Bernhart, S.H., Tafer, H., Mückstein, U., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2006) Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol. Biol.*, **1**, 3.
 20. Mosig, A., Sameith, K., and Stadler, P. (2006) Fragrep: an efficient search tool for fragmented patterns in genomic sequences. *Geno. Prot. Bioinform.*, **4**, 56–60.
 21. Grünweller, A., Purschke, W.G., Kügler, S., Kruse, C., and Müller, P.K. (1997) Chicken vigilin gene: a distinctive pattern of hypersensitive sites is characteristic for its transcriptional activity. *Biochem. J.*, **326**, 601–607.
 22. Wassarman, K.M. and Saecker, R.M. (2006) Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science*, **314**, 1601–1603.
 23. Simard, C., Lemieux, R., and Côté, S. (2001) Urea substitutes toxic formamide as destabilizing agent in nucleic acid hybridizations with RNA probes. *Electrophoresis*, **22**, 2679–2683.
 24. Madhugiri, R. and Evguenieva-Hackenberg, E. (2009) RNase J is involved in the 5'-end maturation of 16S rRNA and 23S rRNA in *Sinorhizobium meliloti*. *FEBS Lett.*, **583**, 2339–2342.

5

Rapid, Non-Denaturing, Large-Scale Purification of *In Vitro* Transcribed RNA Using Weak Anion-Exchange Chromatography

Laura E. Easton, Yoko Shibata, and Peter J. Lukavsky

5.1

Introduction

Over the past decade, new roles for RNA in biology have emerged and established a central role for RNA in the posttranscriptional regulation of gene expression. Many large and small non-protein-coding RNAs have been identified [1–5], which are involved in various processes of gene regulation such as translational silencing, mRNA degradation, and possibly chromatin silencing [2–6]. Extensive biochemical and structural studies of these RNAs are now under way to define their interacting protein and RNA partners, the specificity of their interactions, and the pathways in which such interactions occur in order to gain insights into their functions. To support such studies, simple and effective methods for the preparation and purification of large amounts of structurally homogeneous RNA are required.

To date, there are several established methods of purifying *in vitro* transcribed RNA; denaturing polyacrylamide gel electrophoresis (PAGE) [7], size-exclusion chromatography [8, 9] as well as affinity chromatography and weak anion-exchange chromatography [10–14]. Denaturing PAGE purification is the most commonly used technique that can achieve RNAs highly homogeneous in their molecular sizes. However, this can be time-consuming, involving several steps to run the gel, excise the RNA band, elute the RNA from the gel, and then dialyze and concentrate the RNA sample. Moreover, denaturing RNAs can lead to subsequent misfolding or aggregation of the RNA sample. In addition, it is difficult to fully remove water-soluble acrylamide from the RNA sample, which can interfere with structural and biophysical studies [8]. Chromatographic techniques are often employed to purify an RNA without denaturing it in the process. Size-exclusion chromatography allows not only separation of full-length RNA from aborted RNA transcripts, unincorporated nucleoside triphosphate (NTPs) and the DNA template, but also resolves monomeric and multimeric RNA species [8, 9]. However, this method still requires initial removal of T7 RNA polymerase by phenol extraction, and a subsequent concentrating step of the crude transcription reaction in order to achieve maximum resolution during size exclusion chromatography, which again can cause RNA aggregation. Affinity chromatography can be used when an affinity

tag [10, 11] is incorporated into the nascent transcript. Release of the RNA of interest from the affinity matrix may be achieved *via* a *trans*-acting ribozyme [15] or an inducible *cis*-acting ribozyme separating the affinity tag and the RNA of interest [10, 11]. This technology produces RNAs with homogeneous 3' ends and can be applied to RNA of any size. However, *in vitro* transcription of large-sized RNAs can become inefficient and preparation of isotopically labeled RNA samples commonly used for NMR structural studies can become expensive. If homogeneous 3' ends are not a strict requirement for subsequent biochemical or structural studies, weak anion-exchange chromatography is a fast and reliable alternative to affinity-based RNA purification methods [12–14].

5.2

Materials

5.2.1

Cloning and Plasmid Purification

- *Hind* III, *Bbs* I, and *Eco* RI (New England Biolabs)
- Polymerase chain reaction (PCR) Supermix (Life Technologies Invitrogen) or similar PCR reagents
- Calf intestine alkaline phosphatase (CIAP, 1 U μl^{-1} , Roche Applied Science)
- 10 x CIAP buffer (Roche Applied Science): 0.5 M Tris-HCl, 1 mM EDTA, pH 8.5 (20 °C)
- T4 DNA ligase (New England, NEB, # M0202L) 400 cohesive end units μl^{-1}
- 10 x T4 DNA ligase buffer (NEB): 500 mM Tris-HCl, 100 mM MgCl_2 , 10 mM ATP, 100 mM Dithiothreitol, pH 7.5 at 25 °C
- pUC18 plasmid (or similar high copy number plasmid)
- DH5 α competent cells (Life Technologies Invitrogen)
- QIAquick PCR purification kit (Qiagen)
- QIAquick gel extraction kit (Qiagen)
- QIAprep spin Miniprep kit (Qiagen)
- QIAfilter plasmid mega kit (Qiagen)
- Ampicillin or Carbenicillin (Melford Laboratories)
- 1 x TAE electrophoresis buffer (1 l): 40 mM Tris base, 40 mM glacial acetic acid, 5 mM EDTA at pH 8.0
- SOB media [16]
- TYE agar plates [16]
- 2 x TY broth [16]

5.2.2

In Vitro Transcription

- T7 RNA polymerase (prepared in-house)
- rNTPs (Sigma–Aldrich) prepared at 100 mM total rNTP concentration
- Inorganic pyrophosphatase (Sigma–Aldrich)
- Magnesium chloride (Sigma–Aldrich)

- 2 x Urea loading dye [16]
- 10 x Transcription buffer: 400 mM Tris–HCl, pH 8.1 at 37 °C, 10 mM spermidine, 50 mM dithiothreitol (DTT), 0.1% Triton X-100
- 1 x Tris/Borate/EDTA (TBE) electrophoresis buffer: 89 mM Tris base, 89 mM boric acid, and 2 mM EDTA pH 8.3
- Eight percent denaturing polyacrylamide gel: 1 x TBE buffer, 8% acrylamide (19 : 1 acrylamide: *N,N'*-methylene bisacrylamide), 8 M urea, 1% w/v ammonium persulfate (APS), and 0.5% v/v TEMED (*N, N, N', N'*-tetramethylethylenediamine)
- 0.1% Toluidine blue stain

5.2.3

Weak Anion-Exchange FPLC

- AKTA prime fast-performance liquid chromatography (FPLC) system with a 50 ml superloop or an equivalent system
- Five milliliters of diethylaminoethyl (DEAE) sepharose Fastflow columns (GE Healthcare)
- Buffer A: 50 mM sodium phosphate, pH 6.5, 150 mM sodium chloride, and 0.2 mM EDTA
- Buffer B: 50 mM sodium phosphate, pH 6.5, 2 M sodium chloride, and 0.2 mM EDTA

5.3

Protocols for Plasmid Design and Preparation, RNA Transcription, and Weak Anion-Exchange Purification

Here, we present protocols routinely used in our laboratory for large-scale preparation of highly pure RNA samples. We describe the generation of DNA templates for *in vitro* RNA transcription, preparation of milligram quantities of linearized plasmid DNA template for “runoff” RNA transcription (Protocol 1), the production of milligram quantities of RNA and a rapid, large-scale purification method for native, structurally homogeneous RNAs suitable for biochemical and structural studies (Protocol 2).

RNAs are prepared by *in vitro* transcription using T7 RNA polymerase and a linearized plasmid DNA template, and after addition of EDTA, the crude transcription reaction is applied directly to weak anion-exchange chromatography using a DEAE-sepharose column without further manipulation. T7 RNA polymerase and unincorporated rNTPs, which do not bind to the column matrix, are found in the flow-through. Small abortive transcripts, the desired RNA product, and the plasmid DNA template are separated on the DEAE column over a shallow salt gradient. This easy-to-implement, native RNA purification method eliminates the need for large affinity tags and *cis*-acting ribozymes to be incorporated into the RNA transcript and omits the tedious removal of T7 RNA polymerase by phenol/chloroform extraction. In addition, unincorporated isotopically labeled rNTPs can be easily recycled

from the column flow-through and oligomeric RNA species can be efficiently separated from monomeric RNA based on their overall difference in charge.

Protocol 1: Design and Preparation of the Plasmid DNA Template for “Runoff” RNA Transcription

This section describes the cloning strategy employed to obtain a plasmid DNA template that can be linearized exactly at the nucleotide position corresponding to the desired terminal nucleotide of the RNA of interest. Overlapping primers are used to prepare a construct containing the T7 promoter sequence, directly followed by the compulsory guanosines in the first and second transcript positions, the DNA template sequence for the RNA of interest, and an appropriate 3' end containing a *Bbs* I restriction site. The *Bbs* I restriction enzyme cleaves at position 2 and 6 upstream of the recognition sequence, independent of the nucleotide sequence at the site of cleavage (Figure 5.1). (If the RNA sequence contains an internal *Bbs* I restriction site, *Bsa* I can be used as an alternative allowing cleavage at upstream positions 1 and 5.) The construct should be flanked at either end with restriction sites; we usually use *Hind* III and *Eco* RI, in order to subclone into a high copy number vector such as pUC18. The plasmid can then be amplified in *Escherichia coli* cells, such as DH5 α so that milligram quantities can be purified and linearized for *in vitro* RNA transcription (Figure 5.1).

Protocol 1.1: Cloning the Plasmid DNA-Template-Encoding RNA

- 1) DNA template synthesis (Figure 5.1). In order to synthesize a template for RNA sequences less than 90 nt, a two-primer (up to 80 nt each) approach can be employed. The forward primer should contain the *Hind* III restriction site (AAGCTT), the T7 promoter sequence (AATACGACTCACTATA), and the compulsory first two G residues, followed by the 5' sequence of the required RNA. The reverse primer should contain the 3' end of the RNA sequence followed by two linker residues (TT), the *Bbs* I (GTCTTC) and *Eco* RI (GAATTC) restriction sites (if using *Bsa* I, a single linker residue is required). The 3' end of each primer should overlap in an 18–22 nt complementary sequence with a melting temperature of $\sim 60^{\circ}\text{C}$ to allow for primer annealing. Standard PCR reactions are used to extend the primers and amplify the DNA product. DNA templates for RNA sequences greater than 90 nt can be prepared using overlapping primers and consecutive PCR reactions [8].
- 2) Analyze the size of the PCR product by running 5 μl of the PCR reaction on a 1% agarose gel (1 x TAE buffer) run at 100 V for 30 min. If correct, purify the rest of the PCR product using the QIAquick PCR purification kit following the manufacturer's instructions (similar commercially available kits can be used in steps 2 and 4).
- 3) Digest the PCR product with *Hind* III and *Eco* RI using the manufacturer's recommended reaction conditions; typically, we use 10 U of each enzyme in a 50 μl reaction mixture containing the purified DNA from step 2 and

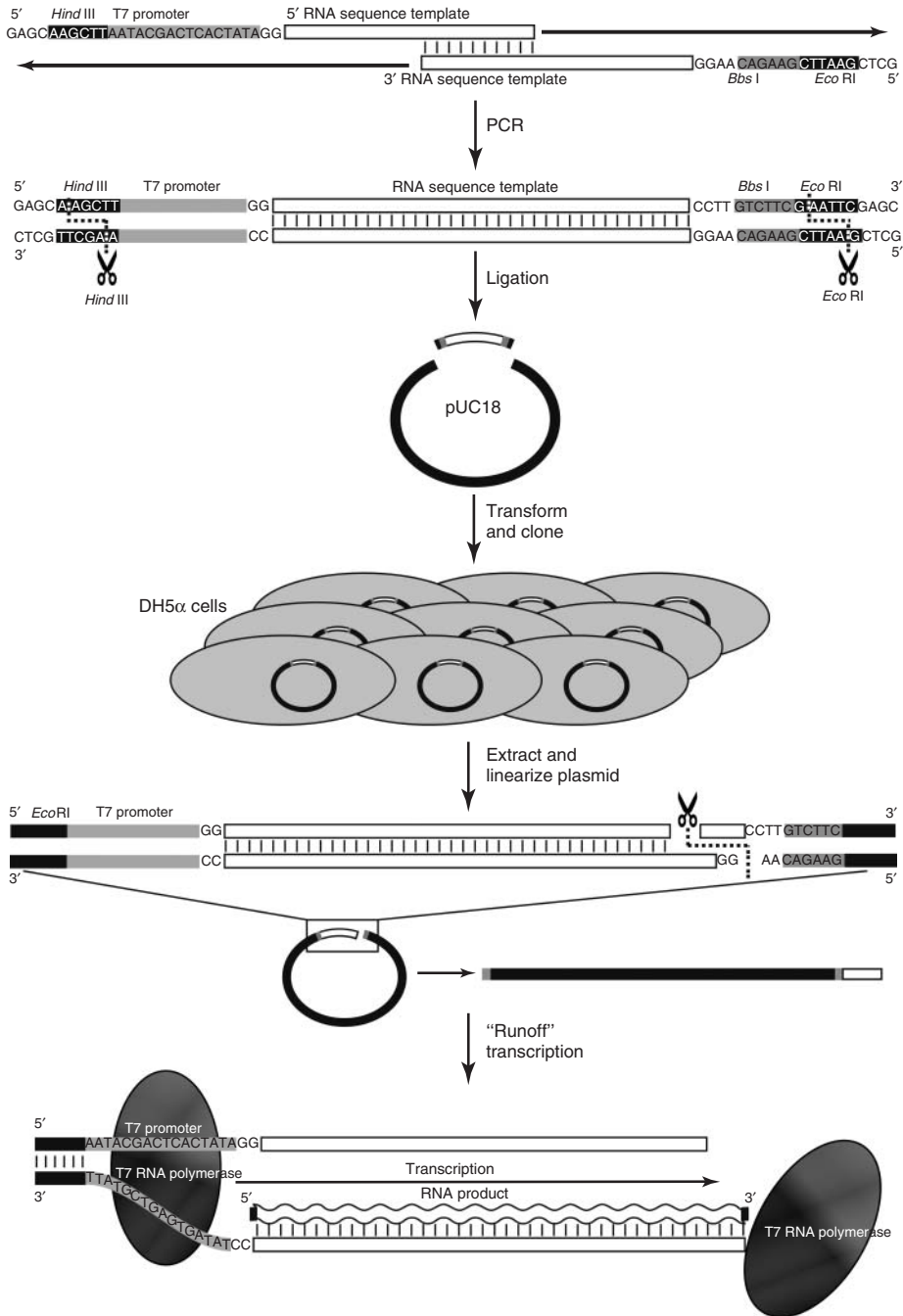


Figure 5.1 Schematic overview of the techniques involved in preparing an *in vitro* transcribed RNA. For details, see text.

incubate in a 37 °C water bath for 2 h. Arrest the reaction by heat-inactivating the enzymes following the manufacturer's instructions. (In steps 3, 4, and 6, reaction conditions are given, but the manufacturer's guidelines should be followed if using alternative suppliers or fast digest and rapid ligation enzymes.)

- 4) Digest the pUC18 plasmid with *Hind* III and *Eco* RI. To prevent self-ligation of the pUC18 backbone, the ends should be dephosphorylated with CIAP: 45 µl digested pUC18 vector, 5 µl 10 x CIAP buffer, 1 µl CIAP (1 U), incubate for 1 h at 37 °C.
- 5) Load digestion mixture onto an agarose gel (1 x TAE), run at 100 V for 30 min, then excise the desired band and purify the pUC18 backbone using the QIAquick gel extraction kit.
- 6) Ligate the *Hind* III/*Eco* RI-digested PCR product into the *Hind* III/*Eco* RI pUC18 backbone using T4 DNA ligase: 10 µl purified PCR product, 1 µl purified pUC18 backbone, 2 µl 10 x T4 DNA ligase buffer, 6 µl Millipore water, and 1 µl T4 DNA ligase (400 U); incubate at 16 °C for 4–6 h or overnight.
- 7) Transform DH5α cells with 10 µl of the ligation reaction (30 min incubation on ice, 45 s heat shock at 42 °C, recover in 400 µl SOB medium at 37 °C with shaking for 2 h).
- 8) Spread 200 µl of the recovered culture onto a TYE + ampicillin (50 µg ml⁻¹) agar plate and allow to grow at 37 °C overnight.
- 9) Pick several single colonies; for each colony, restreak a TYE + ampicillin agar plate and inoculate 5 ml 2 x TY media. Incubate the agar plates at 37 °C and grow the 5 ml cultures in a 37 °C shaker incubator overnight.
- 10) Purify the plasmids from the 5 ml cultures using a mini prep kit and confirm the sequence.

Protocol 1.2: Preparation of Milligram Quantities of Linearized Plasmid DNA

- 1) Inoculate 5 ml 2 x TY + carbenicillin (50 µg ml⁻¹) with a single agar plate colony (derived from a clone with confirmed DNA sequence of the plasmid insert) and grow to 0.5 OD₆₀₀ in a 37 °C shaker incubator. Inoculate four 2 l Erlenmeyer flasks each containing 0.5 l of 2 x TY medium/carbenicillin (50 µg ml⁻¹) with 1 ml each of the starter culture, grow in a 37 °C shaker incubator overnight, and harvest the cells next day by centrifugation.
- 2) Plasmid extraction and purification. Using the Qiagen QIAfilter plasmid mega kit, we typically obtain 3–6 mg plasmid DNA from the 2 l culture. Briefly, resuspend the cells in 125 ml buffer P1, chemically lyse the cells by adding 125 ml buffer P2 and incubate for 5–7 min at room temperature; arrest cell lysis by adding 125 ml buffer P3. Clear the lysate by filtering through a QIAfilter mega-giga cartridge. Equilibrate a QIAGEN-tip 2500 with 35 ml buffer QBT before loading the clarified lysate by gravitational flow. Wash the QIAGEN-tip 2500 with 6–8 tip volumes of buffer QC. Elute the plasmid into a clean centrifuge tube with 35 ml buffer QF, once eluted precipitate by adding

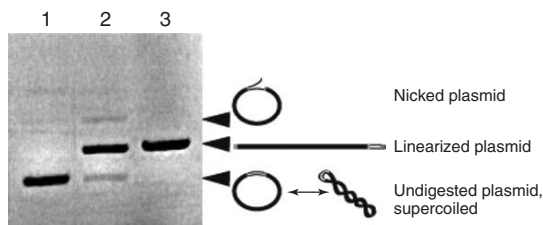


Figure 5.2 Agarose gel analysis of plasmid linearization; lane 1, undigested plasmid; lane 2, partially digested – “nicked” plasmid; lane 3, fully digested, linear plasmid.

24.5 ml of room temperature isopropanol and pellet by centrifugation at 15 000 g for 15 min at 4 °C.

- 3) Wash the plasmid pellet twice with 7 ml 70% ethanol and pellet it, as before, by centrifugation after each wash. Carefully remove excess ethanol with a 1 ml pipette before air-drying at room temperature for 1 h. The pellet becomes glassy when dry and difficult to see; therefore, it is advisable to mark the position of the pellet on the outside of the centrifuge tube.
- 4) Gently redissolve the pellet with 3 ml RNase-free water: use 2 ml to redissolve and a further 1 ml to wash the centrifuge tube. Measure the absorbance at 260 nm to estimate the concentration (pUC18: 1 OD ml^{-1} = 50 μg). Dilute the sample with RNase-free water and 10 x *Bbs* I reaction buffer so that the final concentration of plasmid DNA is 700 $\mu\text{g ml}^{-1}$ in 1 x *Bbs* I reaction buffer; for example, to 3 ml DNA solution (1633 $\mu\text{g ml}^{-1}$) add 3.3 ml RNase-free water and 0.7 ml 10 x *Bbs* I reaction buffer.
- 5) Take a 50 μl aliquot of this diluted plasmid solution and store it at –20 °C. Add 50 U ml^{-1} *Bbs* I to the remaining DNA sample and incubate in a 37 °C water bath for 16 h.
- 6) Load 10 μl of a 1/100 dilution of both, the undigested aliquot and the *Bbs* I-digested plasmid onto a 1% agarose gel (1 x TAE buffer) and run at 100 V for 30 min. Fully digested, linearized plasmid migrates slower than the supercoiled uncut DNA (Figure 5.2). Incomplete, single-strand digestion can result in nicked DNA (only one DNA strand is cleaved), which migrates slower than the linearized plasmid (Figure 5.2). In such a case, add additional 25 U ml^{-1} *Bbs* I and incubate for a further 8 h at 37 °C. Once the plasmid is fully linearized, it can be stored at –20 °C until required.

Protocol 2: *In Vitro* Transcription and Purification of Milligram Quantities of RNA

This section describes how to optimize the magnesium concentration for the transcription reaction before performing large-scale transcription to yield milligram quantities of RNA, and the anion-exchange chromatography method to purify the final RNA product from the enzymes and plasmid DNA. It is important to optimize the magnesium concentration for each RNA sequence as the optimal condition

can only be determined empirically and can drastically affect the final RNA yield. This is done by performing a series of small 25 μl transcriptions, each containing an increasing concentration of magnesium.

As the transcription reaction progresses, pyrophosphate is released. This anion sequesters the Mg^{2+} ions, and, therefore, accumulation of pyrophosphate can lower the final RNA yield. By adding inorganic pyrophosphatase to the reaction mixture, the pyrophosphate is converted into inert phosphate ions. This is especially important for a large-scale transcription reaction containing isotope-labeled nucleotides. The transcription reaction will contain a mixture of plasmid DNA, full-length RNA product, aborted RNA products, and unincorporated rNTPs. Each component of this mixture differs greatly in size as well as in the number of phosphate ions and therefore overall negative charge, which can be exploited to isolate the full-length RNA from the reaction mixture on a weak anion-exchange column. This method is non-denaturing similar to size exclusion or affinity chromatography. Similar to size exclusion, chromatography based on overall charge also allows the separation of any multimeric species from the monomeric RNA. In addition, the phenol extraction step to remove the T7 RNA polymerase is unnecessary, as the enzyme does not bind to the column.

The weak anion-exchange chromatographic purification utilizes three 5 ml DEAE-sepharose Fastflow columns; alternatively, a 20 ml DEAE Fastflow column can be used with the same results. All FPLC steps are performed at room temperature. Once loaded onto the DEAE columns, the transcription mixture can be separated by a gradient of increasing salt concentration (see below). The T7 RNA polymerase washes off the column quickly in the near void volume together with the unincorporated rNTPs that also bind very weakly. The small aborted RNA transcripts elute at low salt concentration at the beginning of the gradient. The full-length RNA, followed by any multimer species of the RNA, elute at higher salt concentrations. Finally, the linearized plasmid is eluted at the end of the gradient at the highest salt concentration (Figure 5.3a).

Protocol 2.1: Preparation of Milligram Quantities of RNA

- 1) Prepare a total of thirteen 25 μl transcriptions. Each reaction should contain 4 mM each rNTP, 70 $\mu\text{g ml}^{-1}$ *Bbs* I-linearized plasmid, 1 x transcription buffer (see above), 1200 U ml^{-1} T7 RNA polymerase, 1 U ml^{-1} inorganic pyrophosphatase, 4–52 mM MgCl_2 in 4 mM increments. Incubate the reaction mixtures in a 37 °C water bath for 1 h.
- 2) Mix 2 μl of each transcription reaction with 2 μl 2 x urea loading dye to denature the sample. Pre-run an 8% denaturing polyacrylamide gel at 200 V for 20 min with 1 x TBE buffer. It is important to pre-run the gel to remove any non-polymerized acrylamide and warm the gel for efficient denaturation of the RNA. Load samples and run the 8% denaturing polyacrylamide gel at 200 V for 20 min in 1 x TBE buffer.
- 3) Visualize the RNA bands, either by UV shadowing (see Chapter 3 for details) or by staining with 0.1% toluidine blue. Identify the reaction condition with the highest RNA yield and thus the optimal magnesium concentration.

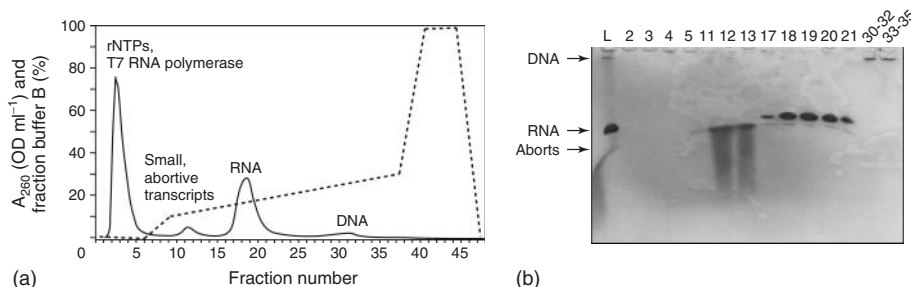


Figure 5.3 RNA purification by weak anion-exchange chromatography. (a) Typical elution profile of a 20 ml *in vitro* transcription reaction of the 48 nt K10 TLS RNA, represented by the black line. Unincorporated NTPs, aborted RNAs, full-length RNA and linear plasmid are all well separated. The gradient trace (fraction buffer B in percentage) is represented by the dashed

line. (b) Denaturing 8% PAGE analysis of the eluted fractions. The transcription reaction is shown in lane L, other lanes contain fractions as numbered in (a). Lanes 30–32 and 33–35 show DNA purified from 1 ml of pooled fractions 30–32 and 33–35, respectively, using the QIAGEN PCR purification kit.

- 4) Prepare a 20 ml transcription reaction mixture as above using the optimal magnesium concentration and incubate in a 37 °C water bath for 2–4 h. Incubation for more than 4 h is not recommended as it also increases the likelihood of degradation. Analyze the transcription yield as before by comparing against the optimal trial reaction.
- 5) Add 0.5 M EDTA (pH 8.1 at 37 °C) to a final concentration of 50 mM to arrest the reaction. Retain a 20 µl aliquot of the reaction as a control sample and store at –20 °C. The transcription mixture can be purified immediately or stored at –20 °C until needed.

Protocol 2.2: Weak Anion-Exchange Chromatographic Purification of RNA

- 1) Equilibrate the DEAE columns with three-column volumes buffer A. Load the transcription reaction mixture into the 50 ml superloop. Program the salt gradient as follows:

0–70 ml	0% Buffer B	1 ml min ^{–1}
70–100 ml	0–10% Buffer B	2 ml min ^{–1}
100–380 ml	10–30% Buffer B	2 ml min ^{–1}
380–410 ml	30–100% Buffer B	4 ml min ^{–1}
410–455 ml	100% Buffer B	4 ml min ^{–1}
455–485 ml	100 to 0% Buffer B	4 ml min ^{–1}

Collect 10 ml fractions throughout the gradient.

- 2) Analyze the fractions by 8% denaturing PAGE. Load 5 µl of each fraction and 2 µl of the crude transcription reaction mixture, each combined with 5 µl 2 x urea loading dye, onto a pre-run 8% denaturing polyacrylamide gel. Run the gel at 200 V for 20 min in 1 x TBE buffer.

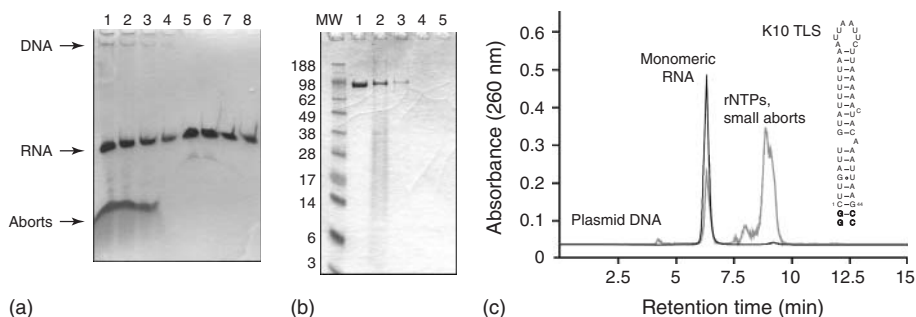


Figure 5.4 Analysis of RNA purified by weak anion-exchange chromatography. (a) Denaturing 8% PAGE analysis of the crude transcription reaction in comparison to a pool (50 ml) of purified RNA fractions, visualized by 0.1% toluidine blue staining. Lanes 1-4: 2, 1, 0.5, and 0.25 μ l of the crude transcription reaction; lanes 5-8: 5, 2.5, 1.25, and 0.625 μ l of the purified RNA pool. (b). Denaturing SDS-PAGE analysis of the peak fractions shown in Figure 5.3a, visualized by Coomassie staining, to confirm that the purified RNA is free of T7 RNA polymerase. The crude transcription reaction (lane 2) and pooled fractions from each peak: rNTPs and T7 RNA polymerase (pooled flow-through, lane 3), abortive transcripts (lane 4) and RNA (lane 5), were precipitated with TCA, and the redissolved pellets loaded onto the gel. T7 RNA polymerase was loaded in lane 1 as a control; MW, molecular weight marker with sizes (in kDa) indicated on the left (c). Gel filtration of purified K10 RNA. The gel filtration profile of the 20 ml crude transcription reaction is represented by a gray line; monomeric RNA and contaminating DNA, rNTPs, and RNA aborts are indicated. Gel filtration of pooled DEAE-purified RNA fractions (17–21) is represented by a black line. The secondary structure of K10 TLS RNA is shown on the right.

- 3) Visualize the RNA bands on the gel and pool the RNA-containing fractions (Figure 5.3b). At this point, the collected RNA fractions should be free of T7 RNA polymerase, linear plasmid, and most abortive RNA transcripts (Figure 5.4b,c). However, for RNAs from 25 to 35 nt, there might be some contamination with small abortive RNA transcripts (Figure 5.4a), but these are removed using either a 3 or a 10 kDa molecular weight cutoff (MWCO) 15 ml Centrprep Centrifugal Concentrator (Millipore, various centrifuges can be used, in this instance Kendro Legend RT, rotor 75003057).
- 4) Equilibrate a 15 ml Centrprep with 15 ml buffer A at 3000 g and 4 °C before loading the pooled RNA sample. Centrifuge at 3000 g and 4 °C until the entire sample has been loaded and the volume reduced to 5 ml. Exchange the buffer of the RNA sample by concentrating to 1 ml and then diluting it again to 15 ml with the desired buffer. Repeat this process further two to four times (depending on the buffer and salt purity required) and then continue until the final required concentration of the RNA sample.
- 5) Dilute 1 μ l of the final RNA sample to the equivalent concentration of the stored aliquot from the crude 20 ml transcription reaction mixture. Mix 2 μ l diluted sample and 2 μ l crude transcription mixture with 5 μ l 2 x urea loading dye and load onto a pre-run 8% denaturing polyacrylamide gel and run at 200 V for 20 min in 1 x TBE buffer. Visualize the RNA bands on the gel to

confirm that the RNA is pure and has not undergone any degradation. Store the concentrated RNA sample at -20°C .

5.4

Troubleshooting

Low plasmid yield following the QIAfilter plasmid mega prep has been observed when purification is performed at temperatures exceeding 25°C , or if the buffers have been stored at temperatures above 25°C or for an extended period of time. In such cases, all buffers for the QIAfilter plasmid mega prep should be prepared freshly according to the manufacturer's instructions. Higher temperatures ($>25^{\circ}\text{C}$) also yield less pure and more salty plasmid DNA. Digestion of plasmid DNA with *Bbs* I is sensitive to higher salt concentration. Therefore, it is crucial to wash the plasmid pellet well with 70% ethanol to remove any excess salt. If plasmid linearization still poses problems leading to only partial digestion or even plasmid degradation, then halve the salt concentration in the reaction mixture relative to that recommended by the manufacturer. Store *Bbs* I at -80°C ; otherwise the enzyme quickly loses activity.

In our hands, the optimal MgCl_2 concentration for RNA transcription reactions ranges from 15 to 30 mM when using rNTPs from Sigma–Aldrich. The rNTPs from other manufacturers may contain different salts and require different concentrations of MgCl_2 for optimal transcription yield. Indeed, isotopically labeled rNTPs prepared in-house [17] often require the addition of much lower MgCl_2 amounts, as they are typically obtained as magnesium salts.

RNA purification using the weak anion-exchange chromatography described here can be used to purify RNAs ranging in size from 30 to 600 nt in length. However, when the size of the RNA is small (25–35 nt), some contamination with aborted RNA transcripts has been observed. These small abortive RNA contaminations can be removed with the 15 ml Centriprep 3 or 10 kDa MWCO concentrator (Protocol 2.2). In order to avoid contamination of very large RNAs (500–600 nt) with the linearized plasmid DNA, only the utmost peak fractions should be pooled. When dimeric or multimeric RNA species are observed, it is advisable to test the eluted RNA fraction by analytical size-exclusion chromatography so that only fractions containing the desired, monomeric RNA species are pooled as described before [14]. In our experience, dimeric and multimeric RNA species can often, at least partially, be refolded into monomeric species by simply diluting 10-fold with water, to reduce salt concentration, before boiling for 10 min and snap cooling on ice. Again, the resulting monomeric species can be separated easily from the oligomers using either size-exclusion or weak anion-exchange chromatography.

Acknowledgments

This work was supported by a grant from HFSP (RGP0024/2008) to P.J.L. We kindly thank all members of the Lukavsky group for their helpful suggestions and discussions.

References

1. Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J. *et al.* (2005) RIKEN Genome Exploration Research Group; Genome Science Group (Genome Network Project Core Group); FANTOM Consortium, Antisense transcription in the mammalian transcriptome. *Science*, **309**, 1564–1566.
2. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, **75**, 843–854.
3. Wightman, B., Ha, I., and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*, **75**, 855–862.
4. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.
5. Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
6. Stevenson, D.S. and Jarvis, P. (2003) Chromatin silencing: RNA in the driving seat. *Curr. Biol.*, **13**, R13–R15.
7. Wyatt, J.R., Chastain, M., and Puglisi, J.D. (1991) Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques*, **11**, 764–769.
8. Lukavsky, P.J. and Puglisi, J.D. (2004) Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides. *RNA*, **10**, 889–893.
9. Kim, I., Mckenna, S.A., Viani Puglisi, E., and Puglisi, J.D. (2007) Rapid purification of RNAs using fast performance liquid chromatography (FPLC). *RNA*, **13**, 289–294.
10. Batey, R.T. and Kieft, J.S. (2007) Improved native affinity purification of RNA. *RNA*, **13**, 1384–1389.
11. Kieft, J.S. and Batey, R.T. (2004) A general method for rapid and nondenaturing purification of RNAs. *RNA*, **10**, 988–995.
12. Wincott, F., Drenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., and Usman, N. (1995) Synthesis, deprotection, analysis and purification of RNA and ribozymes. *Nucleic Acids Res.*, **23**, 2677–2684.
13. Anderson, A.C., Scaringe, S.A., Earp, B.E., and Frederick, C.A. (1996) HPLC purification of RNA for crystallography and NMR. *RNA*, **2**, 110–117.
14. Easton, L.E., Shibata, Y., and Lukavsky, P.J. (2010) Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA*, **16**, 647–653.
15. Ferre-D'amare, A.R. and Doudna, J.A. (1996) Use of *cis*- and *trans*-ribozymes to remove 5' and 3' heterogeneities from milligrams of *in vitro* transcribed RNA. *Nucleic Acids Res.*, **24**, 977–978.
16. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, 2nd edn, vol. 3, Cold Spring Harbor Press, New York.
17. Batey, R.T., Inada, M., Kujawinski, E., Puglisi, J.D., and Williamson, J.R. (1992) Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. *Nucleic Acids Res.*, **20**, 4515–4523.

6

3'-Terminal Attachment of Fluorescent Dyes and Biotin

Dagmar K. Willkomm and Roland K. Hartmann

6.1

Introduction

A large number of experimental approaches in RNA biochemistry are based on some sort of label or tag within RNA molecules. Several of the methods at hand for RNA labeling are widely used, such as incorporation of modified nucleotides during solid-phase synthesis (Chapters 7 and 8) or transcription (Chapter 1), as well as posttranscriptional random chemical labeling by commercially available kits (e.g., the ULYSIS[®] Nucleic Acid Labeling Kits from Invitrogen Life Technologies). In contrast, the selective chemical attachment of a label to periodate-oxidized RNA 3' ends described here is less common, despite some advantages over alternative labeling techniques: primarily, the procedure is efficient, inexpensive, and suitable for large-scale preparations. Further, it is not restricted to newly synthesized RNA, and can therefore also be applied to commercially available RNAs or molecules isolated from cells and tissue. Moreover, the tag is incorporated at the RNA end, where it is less likely to impede proper folding than elsewhere in the molecule. Finally, when a single labeling group is added to each RNA molecule's 3' end, the technique yields a rather uniform RNA population in terms of structure and RNA/label ratio.

Both fluorescent and biotin RNA tags are versatile tools in a broad array of techniques. One of the traditional domains of RNA 3'-end-labeling with fluorescent dyes is fluorescence resonance energy transfer (see also Chapter 24) and related methods, which were initially used to probe RNA positioning within the ribosome [1, 2] and to analyze RNA–protein interactions within the signal recognition particle [3, 4]; further applications range from RNA–RNA binding measurements [5] to the use of fluorescent RNAs as hybridization probes [6, 7] and for microinjection experiments [8, 9].

Biotin as a tag is widely used because of its exceptionally tight binding to streptavidin, a 60 kDa bacterial protein, and avidin, the related protein from egg white. With an estimated dissociation constant of about 10^{-14} M, the complex formation between biotin and streptavidin is essentially irreversible under a wide variety of conditions ([10] and references therein, [11]). More recently, modified

avidins, which allow reversible binding, have also become available [6]. Accordingly, biotin tags have their uses in immobilization of RNA to solid supports as a prerequisite for capture assays and affinity chromatography (Chapter 42; [12, 13]) or surface plasmon resonance measurements [14, 15]. Finally, among many further applications (reviewed in [16]), biotin-labeled RNAs are also employed as probes, with a subsequent detection mostly based on streptavidin–alkaline phosphatase conjugates [17].

6.2

Description of Method

The underlying reaction mechanism of 3'-label chemical attachment is based on the selective periodate-mediated oxidation of the RNA's 3'-terminal ribose *cis*-diol (Figure 6.1a; [18] and references therein). Oxidation results in a dialdehyde that is highly susceptible to nucleophilic attack, and will therefore readily react with nucleophilic amino components such as hydrazine derivatives. As a final step, the reaction product can be stabilized by borohydride reduction (Figure 6.1a).

6.3

History of the Method

The reaction of dialdehydes, derived from periodate oxidation of riboses, with amines was originally explored in the context of nucleosides and nucleoside-5'-phosphates (5'-NMPs). For oxidized 5'-AMP, it was shown that alkaline conditions cause phosphate release and cleavage of the N-glycosidic bond [20]. With phenylhydrazine, nucleoside bisphenylhydrazones were reported to form at the C2'- and C3'-carbon, which favored nucleoside degradation under acidic conditions [21]. Schiff base formation at the C2'- and C3'-carbons was observed for reaction of methylamine with oxidized 5'-AMP; dependent on the methylamine concentration, phosphate release *via* β -elimination occurred predominantly in the neutral pH range [20, 22]. Using isonicotinic acid hydrazide as the nucleophile in reactions with periodate-oxidized nucleosides and nucleotide-5'-phosphates, it was demonstrated by elemental and spectroscopic analyses that the reaction products are morpholine derivatives (III, Figure 6.1a; [23]). Importantly, in contrast to reactions with methylamine, β -elimination of the 5'-phosphate in reaction products with oxidized 5'-NMP was suppressed in the pH range of 4–6, providing first evidence that carboxylic acid (acyl) hydrazides inhibit this unwanted reaction [23]. The same study [23] also demonstrated the formation of stable morpholine derivatives with the oxidized 3'-terminal ribose of tRNA and rRNA. Very similar results were later obtained with isonicotinic acid hydrazide and the related 1-(carboxymethyl)pyridinium chloride hydrazide or benzoic acid hydrazide [19]. Interestingly, the morpholine derivative products (III, Figure 6.1a) were found to be stable without reduction by NaBH₄ [19, 23]. The latter reduction step was required to stabilize the Schiff base products resulting from the reaction of oxidized 5'-AMP with methylamine, in order to prevent degradation by phosphate elimination [20].

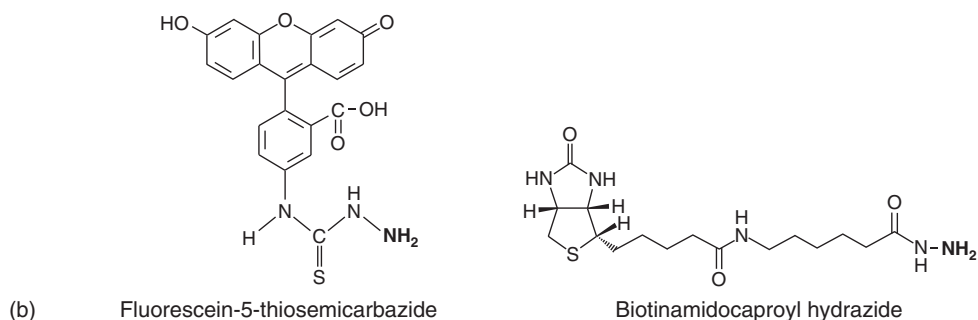
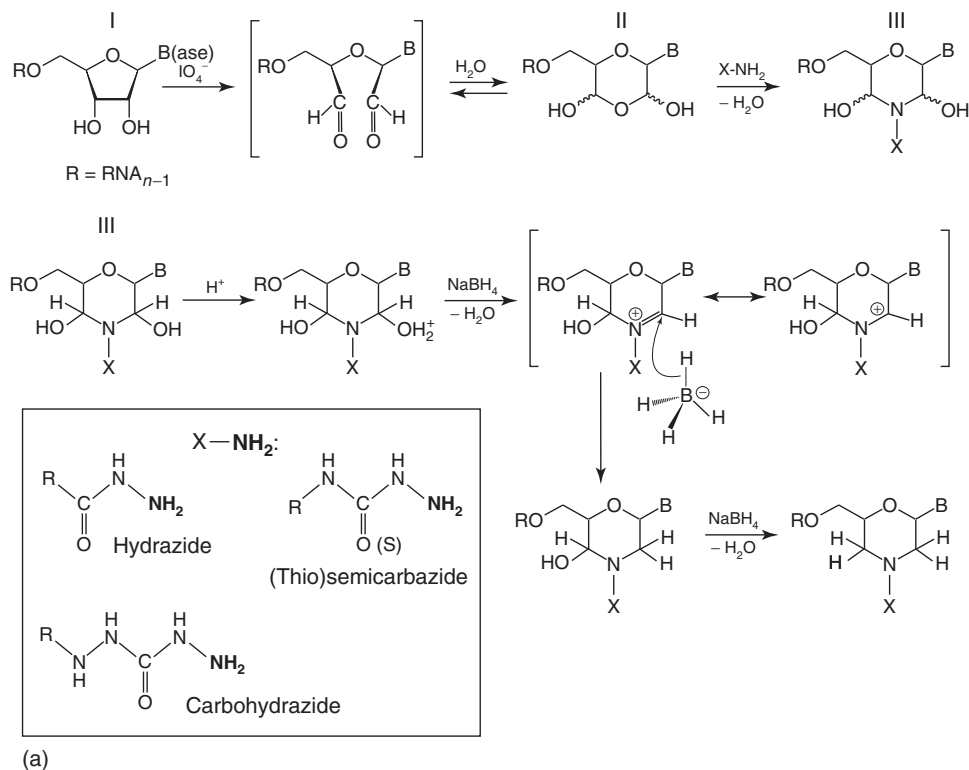


Figure 6.1 (a) Reaction mechanism for modifications at the 3'-terminal ribose of RNAs. Oxidation of the ribose *cis*-diol with periodate results in a reactive dialdehyde that is rapidly hydrated to II, as inferred from IR and NMR spectroscopic analyses [19]; II is then attacked by the nucleophilic amino group of hydrazides, (thio)semicarbazides, or

carbohydrazides (see box; adapted from [18]). The morpholine derivatives (III) can be further stabilized by reduction with NaBH_4 . (b) Examples of coupling reagents: fluorescein-5-thiosemicarbazide, the hydrazide most frequently used as an RNA fluorescence tag, and biotinamidocaproyl hydrazide, to be used for biotin labeling.

With regard to the requirements of the labeling chemistry, a variety of hydrazine derivatives have been developed and are commercially available. Biotin can be purchased as biotin hydrazide (where the carboxylic acid moiety of biotin is derivatized to an acyl hydrazide) from a number of sources. In two other biotin derivatives in the market, biotinamidocaproyl hydrazide (= biotin-X-hydrazide = biotin-6-aminohexanoic hydrazide, Figure 6.1b; from Sigma-Aldrich) and biotin-XX-hydrazide (Calbiochem Merck, Molecular Probes Invitrogen Life Technologies), the acyl hydrazide moiety providing the nucleophilic nitrogen introduces a 7- or 14-atom spacer, respectively, between the biotin moiety and the oxidized ribose to minimize steric interference between the RNA and the tag. Biotin-X-hydrazide is the more commonly used reagent because of its better water solubility.

Regarding fluorescent dyes, the fluorophore is, in most cases, directly conjugated to the semicarbazide or hydrazide moiety, as, for example, in fluorescein-5-thiosemicarbazide (Figure 6.1b) and Alexa Fluor® hydrazides (both from Molecular Probes Invitrogen Life Technologies), as well as eosin-5-thiosemicarbazide (Sigma-Aldrich has withdrawn it from the market). However, reagents with a short extra spacer between the fluorophore and the reactive group, advantageous to some applications [24], are also available. For a compilation of the diverse fluorophore structures and fluorescence properties, see [6].

Protocol 1

3' Labeling

Biotin Attachment

- 1) Incubate up to about 3–4 nmol RNA in a total volume of 100 μ l 40 mM KIO_4 for 1 h at room temperature in the dark [12].
- 2) Stop the reaction with 100 μ l of 50% ethylene glycol, then add 1/10th volume of 3 M NaOAc (pH 5.2) and 2.5 volumes of 96% ethanol to precipitate the RNA. After centrifugation, wash the pellet with 70% ethanol and air-dry.
- 3) Dissolve the dried pellet in 100 μ l of 10 mM biotinamidocaproyl hydrazide and incubate for 2 h at 37 °C.
- 4) Add 100 μ l of 0.2 M NaBH_4 and 200 μ l of 1 M Tris–HCl pH 8.2. Incubate for 30 min on ice in the dark.
- 5) Purify the RNA (Section RNA Downstream Purifications) to remove the salt and coupling reagent.

Special care has to be taken with some of the solutions:

- KIO_4 needs to be prepared as a 50 mM aqueous solution adjusted to pH 7.0 with 10 N NaOH. The KIO_4 will dissolve only on addition of NaOH; at pH 7.0, 50 mM is at the limit of solubility. A more alkaline pH, although helpful for dissolving the KIO_4 , reduces the yield of the overall reaction [19].

- Biotinamidocaproyl hydrazide has low solubility in water; to prepare the required 10 mM solution, incubate for 2 min at 95 °C.
- 0.2 M NaBH₄: the solution sets free hydrogen. Therefore, prepare small portions (Eppendorf tube scale is fine).

For recent applications of this protocol, see [25, 26].

Fluorescence Labeling

Although similar in chemistry, a slightly different protocol has been used in our laboratory for fluorescence labeling. In order to minimize photobleaching of the fluorescent dye, all reactions need to be carried out in the dark [5].

- 1) Incubate 20 nmol of RNA in a volume of 400 µl 2.5 mM NaIO₄, 100 mM NaOAc (pH 5.0) for 50 min on ice.
- 2) Precipitate with ethanol, wash the pellet with 70% ethanol and air-dry.
- 3) For the coupling reaction, dissolve the RNA in 400 µl of 100 mM NaOAc (pH 5.0) and 1 mM fluorescein-5-thiosemicarbazide (200 mM stock solution in dimethyl formamide) and incubate on ice overnight.
- 4) Ethanol-precipitate and redissolve in 50 µl double-distilled water.
- 5) Purify (Section RNA Downstream Purifications).

Preparatory Procedures: Dephosphorylation of RNA Produced with 3' Hammerheads

Transcription of RNA with 3' *cis*-hammerheads is particularly attractive because it gives defined homogeneous 3' ends as opposed to the heterogeneous 3' ends that result from run-off transcription (Chapter 1). 3' Labeling of these RNAs released by hammerhead self-cleavage, however, poses a problem because their 3' end is masked by a 2', 3' cyclic phosphate. Before the labeling reaction, these RNAs therefore require 3' dephosphorylation, which can be done in a kinase reaction at low concentration of ATP, thus making use of the T4 polynucleotide kinase phosphatase activity [27]. The method can also be applied to other RNA substrates with a 3'-phosphate.

- 1) Incubate 300 pmol RNA with 2', 3' cyclic phosphate ends for 6 h at 37 °C in a total volume of 100 µl containing:
 - a. 0.1 mM ATP
 - b. 100 mM imidazole/HCl (pH 6.0)
 - c. 10 mM MgCl₂
 - d. 10 mM β-mercaptoethanol
 - e. 2 µg bovine serum albumin (BSA)
 - f. 20 U T4 polynucleotide kinase
- 2) Extract first with an equal volume of phenol/chloroform (1 + 1 by volume) and then with chloroform.
- 3) Desalt by gel chromatography (Section Gel Chromatography) and precipitate with ethanol.

RNA Downstream Purifications

Gel Chromatography

Among the diverse options, Sephadex G-50 gel exclusion chromatography with self-made spin columns [28] has, in our hands, proved to be a cheap and efficient means to remove unincorporated label, particularly biotin, with almost no loss of RNA material.

- 1) Prepare Sephadex G-50 slurry: add an equal volume of double-distilled water to the Sephadex, let it swell for several hours with occasional gentle shaking. Wash twice with double-distilled water: let it sediment (or spin briefly at low speed), exchange the water, resuspend, and repeat once more. Perform a third wash with 0.1 x TE (1 mM Tris, 0.1 mM EDTA) pH 7.4, let it sediment and adjust the final volume of the supernatant liquid phase to 30% of the total volume (Sephadex plus supernatant). Resuspend before use.
- 2) Remove the plunger from a 2.5 ml syringe, plug the syringe with a small amount of siliconized glass wool (or a fitting piece of filter paper, e.g., from Whatman-GE Healthcare), and fill it with 1.5 ml Sephadex G-50 slurry.
- 3) Place it in a 15 ml disposable tube and centrifuge in a swinging-bucket rotor for 2 min at 550 g. Discard flow-through.
- 4) Place a decapped Eppendorf tube at the bottom of the 15 ml tube. Apply 200 μ l of sample to the column and spin exactly as before. The flow-through collected in the Eppendorf tube contains the purified RNA.

Depending on the centrifuge, the centrifugation step might need optimization. The aim is that the volume of the flow-through finally collected will be identical to the sample volume applied to the column. Alternatively, ready-made columns are available from a number of suppliers. We use, for example, Micro Bio-Spin[®] Chromatography columns (Bio-Rad):

- 1) Add 800 μ l preswelled Sephadex G-50 or G-75 slurry (2.5 g/50 ml; see above for the preparation) to the 1-ml Micro Bio-Spin[®] column.
- 2) Pack the column by centrifugation for 1 min at 4000 rpm in a tabletop centrifuge.
- 3) Put the column on top of an Eppendorf tube, load 50 μ l of sample onto the matrix center and centrifuge again for 1 min at 4000 rpm to collect the flow-through containing the purified RNA.

Purification on Denaturing Polyacrylamide Gels

With chromatographic purifications, there is the risk of low amounts of coupling reagents not being removed. For many applications, it might therefore prove useful to check the purification by running a sample on a denaturing polyacrylamide gel, where unincorporated fluorescent dye can be seen to migrate below the bromophenol blue band, or to do a gel purification in the first place.

The method of gel purification, described extensively elsewhere in this book (Chapter 3), is convenient and efficient for RNAs of 400 nt or less. In addition to

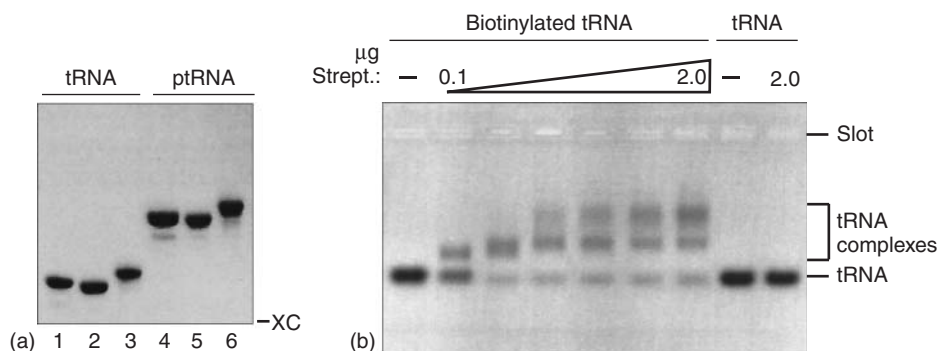


Figure 6.2 (a) Denaturing 9% polyacrylamide gel showing a tRNA^{Gly} (79 nt, lanes 1–3) and its 93 nt precursor (lanes 4–6) – both with homogeneous 3′ ends due to release from a primary transcript carrying a 3′-terminal *cis*-hammerhead – after hammerhead self-removal and gel purification (lanes 2 and 5), after subsequent 3′ dephosphorylation (lanes 1 and 4), and finally after 3′ labeling with fluorescein-5-thiosemicarbazide (lanes 3 and 6). Both the dephosphorylated molecules (lacking the negative charge of one phosphate group) and the tagged molecules (of increased molecular weight) run somewhat slower than the unmodified RNAs. XC: position of xylene cyanol blue, at 20 cm from the slot.

(b) Streptavidin retardation experiment to monitor the efficiency of biotinylation of yeast total tRNA (Roche). About 140 ng of biotinylated tRNA were incubated with up to 2 μg of streptavidin (Strept.) for 15 min in 10 mM Tris pH 7.4, 2.5 mM MgCl₂, and 100 mM NaCl. Samples were run on a 0.8% ethidium bromide-stained agarose gel. Different streptavidin–tRNA complexes were resolved on the gel, which may be related to the fact that streptavidin forms tetramers. At saturating amounts of streptavidin, over 80% of the tRNA showed reduced gel mobility. Unbiotinylated tRNA incubated with 2 μg streptavidin is shown as a control in the last lane on the right.

complete removal of the coupling reagent, for shorter RNAs, such as tRNAs, it also allows elimination of unmodified RNA from the RNA pool because the attached dye slows down RNA migration (Figure 6.2a). For larger RNAs, the difference in electrophoretic mobility caused by the 3′ label might well be too small to allow discrimination of modified and unmodified molecules.

Quality Control

Particularly while setting up the procedure, it is advisable to run controls for labeling efficiency. For fluorescence labeling, the expected efficiency of the reaction is over 90%, and for homogeneous populations of small RNAs this can be monitored by denaturing polyacrylamide gel electrophoresis (Figure 6.2a). Biotin attachment should also proceed almost quantitatively. Biotinylation efficiency can be analyzed by gel shift assays with saturating amounts of streptavidin (commercially available, e.g., from Sigma–Aldrich) on native agarose gels (Figure 6.2b). As opposed to fluorescence labeling analysis, the shift caused by streptavidin is fairly large, so that the labeling of slightly heterogeneous RNA populations and larger RNAs can also be monitored.

6.4

Troubleshooting

When labeling and/or downstream application efficiencies are low, check the following aspects.

6.4.1

Problems Caused Before the Labeling Reaction

6.4.1.1 Quality of the RNA 3' Ends

While the RNA 2'- and 3'-hydroxyls are a prerequisite for the reaction chemistry, a number of cleavage activities generate RNA 3'-phosphate or 2', 3'-cyclic phosphate ends, including several RNases (e.g., RNase A and RNase T1 [29]), metal ions [30], and small ribozymes (reviewed in [31]). The respective RNAs will need to be dephosphorylated before the labeling reaction (Section Preparatory Procedures: Dephosphorylation of RNA Produced with 3' Hammerheads). If the labeling efficiency is low, even after a dephosphorylation step, the extent of dephosphorylation can be checked by denaturing polyacrylamide gel electrophoresis (PAGE) (Figure 6.2a). However, PAGE resolution limits such analysis to RNAs smaller than 100 nt.

6.4.1.2 Purity of the RNA to Be Labeled

Nucleotides also react with the labeling reagents when subjected to periodate oxidation. Thus, after transcription of the RNA to be labeled, thoroughly purify by either Sephadex columns or, preferably, polyacrylamide gel electrophoresis. Impurities such as proteins or salts might impede the labeling reaction as well.

When the labeling efficiency for the RNA of interest is low, even though 2', 3'-hydroxyl ends should be present and the RNA has been thoroughly purified and desalted before the labeling, the most likely cause is poor RNA quality because of degradation. In this case, an entirely new preparation of RNA should be used.

6.4.2

Problems with the Labeling Reaction Itself

6.4.2.1 pH of Reagents

A crucial and limiting aspect to the overall yield of the labeling reaction is the lability of the dialdehyde reaction intermediate [18]. Since it is destabilized at basic pH, the reaction conditions need to be kept non-alkaline. In particular, for the periodate oxidation according to the biotin-labeling protocol, neutral pH of the 50 mM KIO_4 solution is essential. At higher pH, the reaction will be markedly less efficient.

6.4.2.2 Stability of Reagents

In general, it is advisable to store all labeling reagents at -20°C . Nevertheless, especially the sodium borohydride solution will suffer from repeated freeze-thaw cycles and should be stored in aliquots. We have further prepared the periodate

solutions freshly after three freeze–thaw cycles and have kept the borohydride and periodate solutions for no longer than a few months.

6.4.3

Postlabeling Problems

6.4.3.1 Removal of Labeling Reagents

For most downstream applications, special care has to be taken to efficiently remove excess labeling reagents. Ethanol precipitation, even repeatedly, will generally not suffice. Substantial quantities of (anionic) contaminant fluorescent dyes will be visible on polyacrylamide gels, utilizing an appropriate excitation light source. Such gels need to be run very shortly in order to prevent the low-molecular-weight dye from exiting the gel and to minimize lateral diffusion. Unincorporated biotin derivatives can be detected by competition with the biotinylated RNA for binding to streptavidin, resulting in a considerable increase in the amount of streptavidin needed for saturation in a control shift assay.

6.4.3.2 Loss of RNA Material during Downstream Purification

The risk of losing the labeled RNA during purification is not significantly altered when compared to unmodified RNA. A very important exception is phenol extraction of biotinylated material: since biotin increases hydrophobicity, the biotinylated RNA may be retained to some extent at the water/phenol interphase. Phenol extraction of biotinylated RNA should therefore be avoided.

6.4.3.3 Stability of Labeled RNA

Thiosemicarbazide adducts tend to degrade above pH 8 and at elevated temperatures [32], and thus should strictly be kept cold and in the dark, as well as below pH 8.

Acknowledgment

We thank Martin Schlitzer for insightful comments on the chemistry of oxidized riboses.

References

1. Odom, O.W., Robbins, D.J., Lynch, J., Dottavio-Martin, D., Kramer, G., and Hardesty, B. (1980) Distances between 3' ends of ribosomal ribonucleic acids reassembled into *Escherichia coli* ribosomes. *Biochemistry*, **19**, 5947–5954.
2. Stöffler-Meilicke, M., Stöffler, G., Odom, O.W., Zinn, A., Kramer, G., and Hardesty, B. (1981) Localization of 3' ends of 5S and 23S rRNAs in reconstituted subunits of *Escherichia coli* ribosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **78**, 5538–5542.
3. Janiak, F., Walter, P., and Johnson, A.E. (1992) Fluorescence-detected assembly of the signal recognition particle: binding of the two SRP protein heterodimers to SRP RNA

- is noncooperative. *Biochemistry*, **31**, 5830–5840.
4. Lentzen, G., Dobberstein, B., and Wintermeyer, W. (1994) Formation of SRP-like particle induces a conformational change in *E. coli* 4.5S RNA. *FEBS Lett.*, **348**, 233–238.
 5. Busch, S., Kirsebom, L.A., Notbohm, H., and Hartmann, R.K. (2000) Differential role of the intermolecular base-pairs G292-C(75) and G293-C(74) in the reaction catalyzed by *Escherichia coli* RNase P RNA. *J. Mol. Biol.*, **299**, 941–951.
 6. Johnson, I. and Spence, M.T.Z. (2010) *Molecular Probes® Handbook, A Guide to Fluorescent Probes and Labeling Technologies*, 11th edn. This handbook provides comprehensive information on labeling and detection reagents and techniques. For modification of aldehydes, see Chapter 3 (Click Chemistry and Other Functional Group Modifications), Section 3.3 (Reagents for Modifying Aldehydes and Ketones) therein.
 7. Egger, D., Bolten, R., Rahner, C., and Bienz, K. (1999) Fluorochrome-labeled RNA as a sensitive, strand-specific probe for direct fluorescence in situ hybridization. *Histochem. Cell Biol.*, **111**, 319–324.
 8. Kruse, C., Willkomm, D.K., Grünweller, A., Vollbrandt, T., Sommer, S., Busch, S., Pfeiffer, T., Brinkmann, J., Hartmann, R.K., and Müller, P.K. (2000) Export and transport of tRNA are coupled to a multi-protein complex. *Biochem. J.*, **346**, 107–115.
 9. Pederson, T. (2001) Fluorescent RNA cytochemistry: tracking gene transcripts in living cells. *Nucleic Acids Res.*, **29**, 1013–1016.
 10. Jones, M.L. and Kurzban, G.P. (1995) Noncooperativity of biotin binding to tetrameric streptavidin. *Biochemistry*, **34**, 11750–11756.
 11. Wilchek, M. and Bayer, E.A. (1990) Introduction to avidin-biotin technology. *Meth. Enzymol.*, **184**, 5–45.
 12. Von Ahsen, U. and Noller, H.F. (1995) Identification of bases in 16S rRNA essential for tRNA binding at the 30S ribosomal P site. *Science*, **267**, 234–237.
 13. Jestin, J.-L., Dème, E., and Jacquier, A. (1997) Identification of structural elements critical for inter-domain interactions in a group II self-splicing intron. *EMBO J.*, **16**, 2945–2954.
 14. Buckle, M., Williams, R.M., Negroni, M., and Buc, H. (1996) Real time measurements of elongation by a reverse transcriptase using surface plasmon resonance. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 889–894.
 15. Hendrix, M., Priestley, E.S., Joyce, G.F., and Wong, C.-H. (1997) Direct observation of aminoglycoside-RNA interactions by surface plasmon resonance. *J. Am. Chem. Soc.*, **119**, 3641–3648.
 16. Wilchek, M. and Bayer, E.A. (1999) Foreword and introduction to the book (strept)avidin-biotin system. *Biomol. Eng.*, **16**, 1–4.
 17. Boyle, A. and Perry-O'Keefe, H. (2001) Labeling and colorimetric detection of nonisotopic probes. *Current Protocols in Molecular Biology*, **20**, 3.18.1–3.18.9.
 18. Hansske, F. and Cramer, F. (1979) Modification of the 3' terminus of tRNA by periodate oxidation and subsequent reaction with hydrazides. *Methods Enzymol.*, **59**, 172–181.
 19. Hansske, F., Sprinzl, M., and Cramer, F. (1974) Reaction of the ribose moiety of adenosine and AMP with periodate and carboxylic acid hydrazides. *Bioorg. Chem.*, **3**, 367–376.
 20. Khym, J.X. and Cohn, W.E. (1961) Amine-induced cleavage of periodate-oxidized nucleotide residues. *J. Biol. Chem.*, **236**, PC9–PC10.
 21. Khym, J.X. and Cohn, W.E. (1960) Characterizations and some chemical reactions of periodate-oxidized nucleosides. *J. Am. Chem. Soc.*, **82**, 6380–6386.
 22. Steinschneider, A. (1971) Effect of methylamine on periodate-oxidized adenosine 5'-phosphate. *Biochemistry*, **10**, 173–178.
 23. Hunt, J.A. (1965) Terminal-sequence studies of high-molecular-weight ribonucleic acid. The reaction of periodate-oxidized ribonucleosides, 5'-ribonucleotides and ribonucleic acid with isoniazid. *Biochem. J.*, **95**, 541–551.
 24. Klostermeier, D. and Millar, D.P. (2001) RNA conformation and folding studied with fluorescence resonance energy transfer. *Methods*, **23**, 240–254.

25. Rossbach, O., Hung, L.H., Schreiner, S., Grishina, I., Heiner, M., Hui, J., and Bindereif, A. (2009) Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. *Mol. Cell Biol.*, **29**, 1442–1451.
26. Li, D., Meyer, M.H., Willkomm, D.K., Keusgen, M., and Hartmann, R.K. (2010) Analysis of bacterial RNase P RNA and protein interaction by a magnetic biosensor technique. *Biochimie*, **92**, 772–778.
27. Cameron, V. and Uhlenbeck, O.C. (1977) 3'-Phosphatase activity in T₄ polynucleotide kinase. *Biochemistry*, **16**, 5120–5126.
28. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, New York.
29. Davidson, J.N. (1972) *The Biochemistry of the Nucleic Acids*, 7th edn., Academic Press, New York.
30. Pan, T., Long, D.M., and Uhlenbeck, O.C. (1993) Divalent metal ions in RNA folding and catalysis, in *The RNA World* (eds R.F. Gesteland and J.F. Atkins), Cold Spring Harbor Laboratory Press, New York, 271–302.
31. Takagi, Y., Warashina, M., Stec, W.J., Yoshinari, K., and Taira, K. (2001) Recent advances in the elucidation of the mechanisms of action of ribozymes. *Nucleic Acids Res.*, **29**, 1815–1834.
32. Dulbecco, R. and Smith, J.D. (1960) The detection of terminal nucleoside residues in polyribonucleotides. *Biochim. Biophys. Acta*, **39**, 358–361.

7

Chemical RNA Synthesis, Purification, and Analysis*Brian S. Sproat*

7.1

Introduction

The interest in chemically synthesized RNA took a dramatic leap forward with the discovery and application of the small interfering RNA (siRNA) technology [1, 2] and the subsequent discovery of a multitude of different types of non-coding RNAs, including microRNAs. The siRNA technology, in particular, has revolutionized functional genomics and target validation during the past 10 years and surpasses antisense technology in its applicability. Progress has been so rapid that several siRNA-based therapeutics are already in clinical trials and many more compounds are in preclinical studies. Until some 10 years or so ago, the chemical synthesis of RNA lagged a long way behind the well-established DNA synthesis technology. Several pioneers in the field have been responsible for the development of the three solid-phase RNA synthesis chemistry variants that are currently used by almost all commercial suppliers of RNA; these are the 2'-O-TBDMS method [3], the TOM method [4], which is a variant of the TBDMS method, and the 2'-ACE method [5, 6]. The abbreviations refer to the protecting groups used for the ribose 2'-hydroxyl group: TBDMS is *tert*-butyldimethylsilyl, TOM is triisopropylsilyloxymethyl, and ACE is *bis*(acetoxyethoxy)methyl. In the past, problems with RNA synthesis were largely caused by poor quality RNA phosphoramidites (the building blocks for solid-phase synthesis), inappropriate protecting groups combined with suboptimal deprotection protocols, and poor activating agents. This combination of largely unavoidable obstacles combined with the intrinsic chemical and biological instability of RNA led, in most cases, to failed syntheses. However, the boom in the usage of synthetic RNA, both for siRNA and other applications such as ribozymes and aptamers, has had a very positive effect in that the speciality reagent suppliers have improved the quality of the building blocks, leading to healthy competition and affordable prices. Moreover, the use of optimized protecting groups, coupling agents, and finely tuned deprotection protocols has finally made chemical RNA synthesis robust and reliable. Now that siRNAs are being developed as therapeutics, interest in finding still better 2'-hydroxyl protecting groups and simplified deprotection protocols has intensified dramatically and the past years have seen the development of several

highly promising candidates, a couple of which may eventually replace the widely used TBDMS chemistry that is currently in use for all siRNAs that are in or about to enter clinical trials. Worthy of mention here is the 2'-O-(2-cyanoethoxymethyl) protecting group, abbreviated to CEM [7–10], whereby a 110-mer oligoribonucleotide has been synthesized on a small scale using this blocking group and fully characterized [9]. Commercialization of these building blocks is currently being done by Nippon Shinyaku Co. Ltd., Ibaraki, Japan. A closely related blocking group, the 2'-O-[2-(4-tolylsulfonyl)ethoxymethyl], or TEM [11, 12], has been proposed but has some special issues during the deprotection that will likely prevent its more widespread use. The 2'-O-[4-(*N*-dichloroacetyl-*N*-methylamino)benzyloxymethyl] group [13] enables efficient synthesis of RNA, but after removal of the base-sensitive protecting groups the resultant 2'-O-[4-(*N*-methylamino)benzyloxymethyl] group requires cleavage at pH 3.8 for 30 min at 90 °C in buffered acetic acid. Most recently, a couple of research groups have proposed new blocking groups that may well revolutionize the chemical synthesis of RNA by making it almost as simple as the DNA synthesis. The first of these is the 2'-O-(levulinylloxymethyl) protecting group, otherwise referred to as acetal levulinyl ester (ALE), in which the nucleobases are also protected with levulinyl groups, thus enabling a very mild cleavage of the nucleobase and 2'-hydroxyl protecting groups with hydrazine in pyridine/acetic acid [14, 15]. However, cleavage from the support still requires an additional reagent. The other two new RNA synthesis routes require only single-step deprotections and completely avoid the need for toxic fluoride reagents that are an essential feature of the 2'-O-TBDMS- and 2'-O-TOM-based RNA synthesis chemistries. The use of 2'-O-acyloxymethyl protecting groups, in particular the 2'-O-(pivaloyloxymethyl) (POM) group, enables fast and efficient oligoribonucleotide synthesis, whereby deprotection is achieved simply with aqueous ammonia [16–18], since under these conditions the 2'-OH group remains protected as a hemiacetal of formaldehyde, which only breaks down on removal of ammonia as the pH drops. A minor hitch is the requirement of a trapping agent for the formaldehyde, which otherwise adds onto the guanine base to form an adduct in the presence of ammonia. Preliminary results look very promising, although until now no large-scale syntheses have been reported. The 2'-O-POM-protected RNA monomers are available for testing from the group of Debart in Montpellier, France. A completely different strategy has been adopted by Dellinger's group at Agilent Technologies and has led to the development of the 2'-O-(1,1-dioxo-thiomorpholine-4-thionocarbonyl) protecting group, a 2'-thionocarbamate (TC)-based protecting group [19], which can be cleaved along with the phosphate protecting groups, the nucleobase protecting groups, and the linkage to the solid support using ethylene diamine/toluene (1 : 1 v/v) for 2 h at room temperature. Under these conditions, the deprotected RNA stays adsorbed on the support and, after removal of the reagent, washing and drying, is eluted in sterile buffer ready for purification. This so-called TC chemistry method should gain widespread use in high throughput syntheses for siRNA screenings as well as for large-scale syntheses. Building blocks are available from Link Technologies, Inc. and SAFC Prologo under license from Agilent. Thus far, synthesis results look

very promising indeed and the industrial scale-up of the monomers is straightforward since almost all reaction intermediates are crystalline and yields are close to quantitative.

Despite the recent progress as outlined above, the methods and procedures described here have been restricted to the industry standard TBDMS chemistry but also apply to the closely related TOM chemistry, whereas most commercially available synthesizers are not compatible with the highly specialized 2'-ACE chemistry. The synthesis method described here is of course one of many variant methods, but all methods are in the end a variation of the basic methods described here. Since synthesis is performed on solid phase using the well-established phosphoramidite chemistry [20, 21], it can be performed manually or on any of the commercially available synthesizers. Synthesis is initiated from the 3' end starting with the 3'-terminal nucleoside anchored most commonly via a succinyl linkage to an insoluble matrix, generally aminopropyl or long-chain alkylamine-functionalized controlled pore glass, or aminomethyl polystyrene, contained in an appropriate reaction vessel. The nucleobases of the phosphoramidites and functionalized supports are preferably protected with *N*-phenoxyacetyl (pac) [22] or *N*-*tert*-butylphenoxyacetyl (tac) [23] groups, enabling very mild deprotection of the RNA at the end of the synthesis. However, the use of *N*⁶-benzoyl A, *N*⁴-acetyl C, and *N*²-isobutyryl G phosphoramidites leads to similar results regarding the yield and purity of the RNA. The structure of the tac-protected cytidine building block is illustrated in Figure 7.1. The advantage of the solid-phase method is that reagents are introduced into the vessel for removing protecting groups and enabling chain extension of the RNA one nucleotide at a time and excess reagents are simply flushed with a suitable solvent, in this case acetonitrile. The cyclical process is repeated until the desired length of RNA is obtained. Since no intermediate purification steps are possible, the purification is done at the end of the assembly when most of the protecting groups have been removed. In practice, all reactions proceed

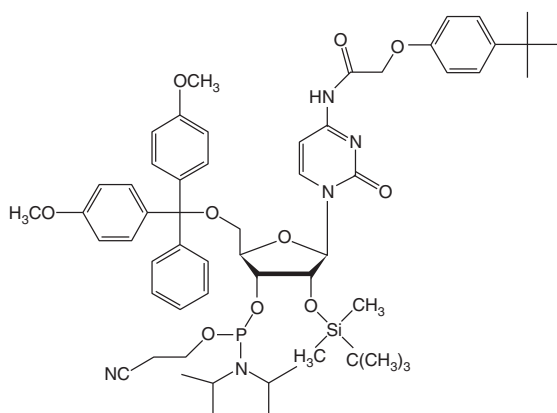


Figure 7.1 Structure of a standard cytidine phosphoramidite building block, carrying 4-*t*-butylphenoxyacetyl protection of the exocyclic amino group.

close to 100% yield and the chain-extension reaction has a yield in the range of 98.5–99%, thus enabling in most cases a straightforward purification of the crude product.

On completion of the synthesis, the fully protected support-bound RNA is deprotected in a stepwise manner. In the first step, the linkage to the solid phase and the nucleobase and phosphate-protecting groups are cleaved. In the second step, the TBDMS groups are cleaved using triethylamine *tris*(hydrofluoride) [24, 25]. When RNAs longer than about 25 nt are synthesized, it is best to leave the 5'-terminal dimethoxytrityl group attached as it is lipophilic and can be used as a purification aid. The crude RNA is then purified by anion-exchange and/or reversed-phase high-performance liquid chromatography (HPLC) according to the length of the RNA and the purity required. For applications such as nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography, purities of greater than 98% are desirable.

7.2

Description

7.2.1

The Solid-Phase Synthesis of RNA

This section is devoted to the synthesis of the fully protected RNA on solid phase. The various steps involved in each cycle of the synthesis are illustrated in Figure 7.2. Each cycle comprises a detritylation step that unmasks the 5'-hydroxyl group for chain extension, a coupling step in which the desired nucleotide as a phosphoramidite building block activated with 5-(benzylthio)-1*H*-tetrazole [26] is added, a capping step that acylates any unreacted 5'-hydroxyl group, an oxidation step that converts the phosphite triester to a phosphate triester, a second capping step that removes any occluded iodine and, of course, in between washing steps with acetonitrile to remove excess reagents. 5-(Benzylthio)-1*H*-tetrazole (BTT) for activation of the sterically hindered 2'-*O*-TBDMS-protected phosphoramidites is strongly preferred over conventional 1*H*-tetrazole with regard to both speed and coupling efficiency [26]; however 4,5-dicyanoimidazole (DCI) [27] and 5-(ethylthio)-1*H*-tetrazole (ETT) [28] can also be used with similar efficiency. Moreover, syntheses can be performed manually or be machine assisted using the following reagents and equipment:

- RNA monomers: 5'-*O*-Dimethoxytrityl-*N*(pac or tac)-2'-*O*-TBDMS-3'-*O*-(β -cyanoethylphosphoramidites) of adenosine (A), uridine (U), cytidine (C), and guanosine (G). These compounds are available, for instance, from Thermo Scientific (Milwaukee, USA) or SAFC Proligo (Hamburg, Germany) and should be stored dry under argon at -20°C . Alternative suppliers of fast deprotecting RNA phosphoramidites are Glen Research (Sterling, VA, USA) and ChemGenes (Wilmington, MA, USA).

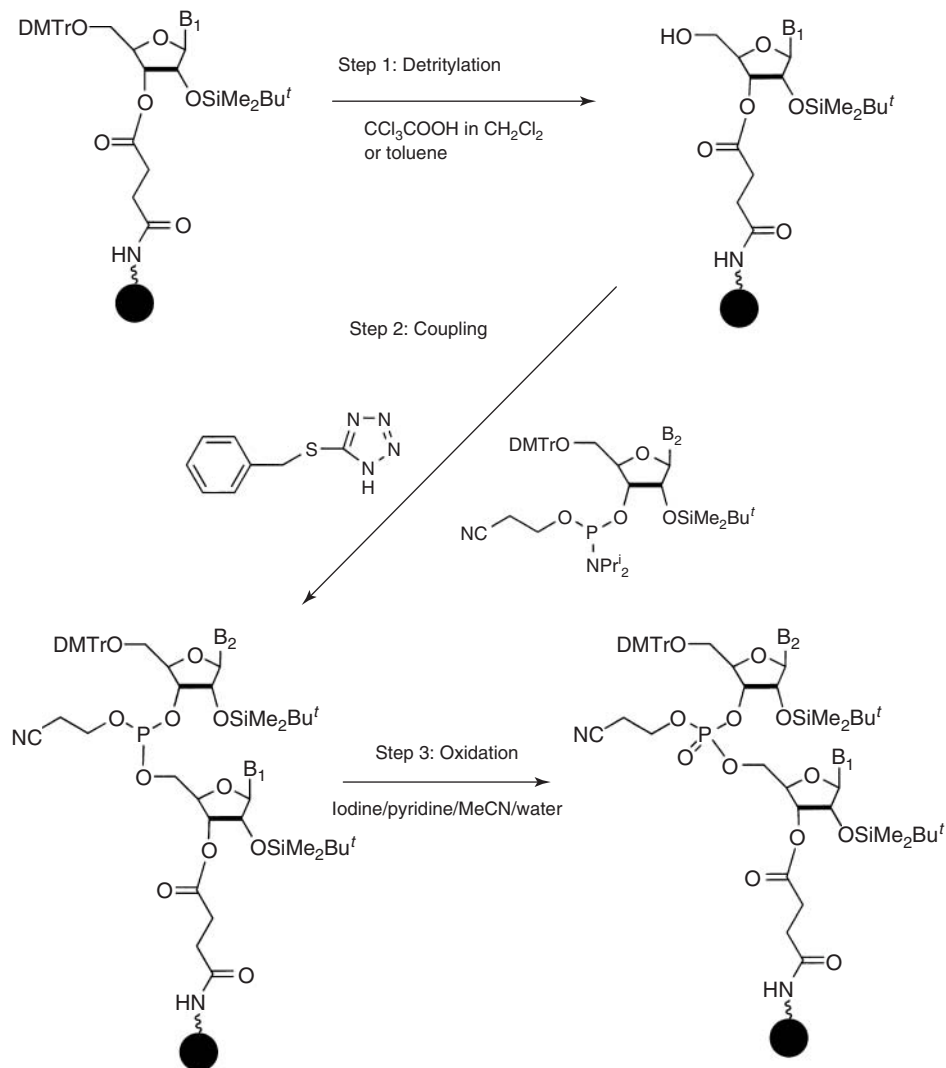


Figure 7.2 Scheme illustrating a single cycle of solid-phase RNA synthesis *via* the phosphoramidite method. The filled-in black circle represents the controlled pore glass support. B₁ and B₂ represent

protected nucleobases, for example, uracil-1-yl, *N*⁴-(4-*t*-butylphenoxyacetyl) cytosine-1-yl, *N*²-(4-*t*-butylphenoxyacetyl) guanine-9-yl or *N*⁶-(4-*t*-butylphenoxyacetyl) adenine-9-yl.

- Solid-phase supports, either CPG from, for example, SAFC Proligo, Thermo Scientific, ChemGenes, Glen Research, Link Technologies (Glasgow, UK), and Prime Synthesis (Aston, PA, USA) or polystyrene from, for example, GE Healthcare, Kinovate Life Sciences (Oceanside, CA, USA), ChemGenes, Glen Research, and Prime Synthesis, functionalized with A, U, C, and G.

- BTT, the activating agent that is available with a very low residual water content from, for example, emp Biotech (Berlin, Germany), CMS (Oxford, UK), Link Technologies, and American International Chemical, Inc. (Framingham, MA, USA).
- Capping solutions A (fast deprotection since it is based on 4-*tert*-butylphenoxy-acetic anhydride) and B from SAFC Proligo. Capping solutions are also available from, for example, Merck, Riedel-de-Haen, Biosolve, and Malinckrodt-Baker.
- Oxidation solution containing iodine from, for example, SAFC Proligo, Merck, Riedel-de-Haen, Biosolve, or Malinckrodt-Baker. Since the iodine concentration is often only 17 mM, the concentration should be adjusted to 50 mM when performing large-scale syntheses, by adding the correct amount of high-purity iodine. The composition of this solution varies considerably between suppliers but the basic composition is a solution of iodine in tetrahydrofuran (THF) or acetonitrile containing about 10 volume% of pyridine plus a few percentage of water. *Warning: Do not use THF-containing solutions on ÄKTA synthesizers.*
- Deblock solution comprising 3% trichloroacetic acid in dichloromethane from SAFC Proligo or another supplier. For large-scale syntheses, an alternative deblock solution containing up to 6% dichloroacetic acid in toluene is usually used, in particular on the ÄKTA oligopilot plus 10 (GE Healthcare).
- DNA synthesis grade acetonitrile containing less than 30 ppm water from, for example, Malinckrodt-Baker, Merck, Riedel-de-Haen, Romil (Waterbeach, Cambridge, UK), and Biosolve.
- Assorted 1000 series gastight syringes with volumes of 0.5, 1, and 2.5 ml, which can be purchased from Hamilton Company (Reno, NV, USA).
- DNA/RNA synthesizer from, for example, Applied Biosystems, GE Healthcare, BioAutomation (Plano, TX, USA), or other manufacturers or, for manual synthesis, a glass reaction vessel fitted with a B14 ground glass joint at the top and a fine porosity glass frit and a tap at the bottom [29]. A set of suitable vessels can be made by any laboratory glass blower.
- Dry argon supply, purity grade 4.8 or higher.

Manual RNA Synthesis

- 1) Weigh out the requisite amounts of the tac- or pac-protected monomers in small vials that can be closed with a septum and dry them overnight *in vacuo* over separate containers of phosphorus pentaoxide and potassium hydroxide pellets to remove traces of water. Suitable vials for this purpose are the amber glass bottles used by suppliers of DNA and RNA phosphoramidites. To perform syntheses in the range of 1–3 μmol , it is recommended that 8–10 equiv. of monomer per coupling be used relative to the amount of support used. For synthesis scales above 5 μmol , the excess monomer should be reduced to 5-fold or even to 2.5-fold when you have gained some experience, in order to reduce costs.
- 2) Carefully release the vacuum with dry argon and seal the bottles with tight-fitting rubber septa.

- 3) Using a gastight syringe, dissolve each of the monomers in the requisite volume of dry acetonitrile to give a 0.1 M solution and seal with Parafilm. It is not recommended to store the monomer solutions for more than two to three days at room temperature.
- 4) Prepare an adequate volume of a 0.3 M solution of BTT in very dry acetonitrile in a tightly closed bottle under argon.
- 5) Prepare 100 ml of capping mixture comprising 1 volume of fast deprotection capping solution A and 1.1 volumes of capping solution B in a tightly stoppered flask. Fresh capping mixture should be prepared every 4–5 h.
- 6) Weigh out the requisite amount of CPG or polystyrene carrying the desired 3'-terminal ribonucleoside into the glass reaction vessel. For a 1 μ mol scale synthesis, the vessel should have a volume of about 5 ml; whereas for 10 μ mol scale, a volume of 20 ml is more appropriate to allow efficient washing.
- 7) Using a Pasteur pipette, add 3% TCA in dichloromethane (deblock solution) to the support and let it percolate through. Immediately, a deep orange color is produced, characteristic of the released dimethoxytrityl cation. Continue to add acid until the effluent is colorless.
- 8) Now drain the support using a slight pressure of dry nitrogen or, better still, argon since it is much heavier and diffuses much more slowly when the vessel is left briefly unstoppered.
- 9) Wash the support 8–10 times in a batchwise manner with acetonitrile using a Teflon wash bottle, removing the supernatant each time using argon pressure.
- 10) Just before the coupling step, wash the support once with very dry acetonitrile (less than 30 ppm water), flush using argon pressure, close the tap, and stopper the vessel.
- 11) To the support, add with a gastight syringe the desired monomer as a 0.1 M solution in acetonitrile, and add with a second gastight syringe an equal volume of 0.3 M BTT solution in acetonitrile, stopper the vessel immediately, and agitate several times for a period of 6 min.
- 12) While the coupling step is in progress, clean both syringes thoroughly with acetonitrile and store them in a desiccator.
- 13) Flush the coupling mixture, wash the support once with acetonitrile, and flush using argon pressure.
- 14) Add a few milliliters of capping mixture to the reaction vessel, stopper and agitate for 1 min, and then drain.
- 15) Wash the support once with acetonitrile and flush using argon pressure.
- 16) Add a few milliliters of oxidation mixture and allow it to slowly percolate through the support for 2 min. This step oxidizes the phosphite triester to a phosphate triester.
- 17) Drain the support and wash once with acetonitrile and drain using argon pressure.
- 18) Once again add a few milliliters of capping mixture, agitate for 30 s, and then drain using argon pressure.

- 19) Now wash the support thoroughly (six times) with acetonitrile, draining each time using argon pressure.
- 20) Repeat steps 7–19 for every addition of nucleotide until the 5' end of the oligoribonucleotide has been reached.
- 21) If the RNA is longer than about 25 nt, the final trityl group should be left on as a lipophilic purification handle. For RNAs less than 25 nt in length, remove the final trityl group as in step 7 and wash the support very thoroughly with acetonitrile.
- 22) Finally dry the support carefully using a stream of argon.

Automated RNA Synthesis

In order to perform automated RNA synthesis, follow the instructions for the particular synthesizer plus the program for the RNA synthesis scale you intend to use. Now the CPG or polystyrene support is placed inside a small plastic cartridge or a steel column; prepacked columns are also available commercially for several different synthesis scales from the support suppliers listed in Section 7.2.1. All the reagents that you will need are commercially available in the correct bottles to fit the various instruments in the market. To ensure that reagents stay dry during the synthesis, activated molecular sieves or trap bags can be added, but be sure to use very good quality sieves for this purpose.

7.2.2

Deprotection

In the first deprotection step, the succinate linkage connecting the 3' terminus of the protected RNA to the solid support is cleaved, the β -cyanoethyl phosphate protecting groups are removed by β -elimination and, in addition, the exocyclic nucleobase-protecting groups are cleaved. In our hands, this step is best performed with a 1 : 1 mixture of concentrated aqueous ammonia and 8 M ethanolic methylamine, which prevents premature loss of the TBDMS groups that would otherwise lead to degradation of the RNA under basic conditions. *Warning: methylamine is not compatible with N^4 -benzoyl protected C.* It is only compatible with N^4 -acetyl-, pac-, or tac-protected C. In the second deprotection step, the TBDMS groups are removed using triethylamine *tris*(hydrofluoride) in an appropriate solvent. At this point, you will need the following reagents:

- High-purity concentrated aqueous ammonium hydroxide. This solution is highly irritating to the eyes and respiratory system and must only be used in a well-ventilated fume cupboard.
- Anhydrous 8 M methylamine in ethanol. This compound is also highly irritating to the eyes and respiratory system and must only be used in a well-ventilated fume cupboard.
- Anhydrous dimethylsulfoxide (e.g., Fluka, Biotech. Grade).
- Triethylamine *tris*(hydrofluoride) (e.g., Sigma–Aldrich, Acros). This compound is hazardous and toxic and should only be handled wearing full protection and used only in a well-ventilated fume cupboard.

- Anhydrous triethylamine.
- *N*-Methylpyrrolidone, peptide synthesis grade (e.g., Acros).
- Ethoxytrimethylsilane (e.g., TCI Europe, Fluka).
- *Tert*-butyl methyl ether (e.g., Fluka).

Deprotection of Base-Labile Protecting Groups

- 1) Transfer the support obtained from Section 7.2.1 into a small screw top vial or Duran bottle equipped with a tight-fitting screw top.
- 2) Add a mixture of concentrated aqueous ammonia and 8 M ethanolic methylamine, 1 : 1 by volume. A volume of 2 ml is adequate for a 0.2–1 μmol scale synthesis. Otherwise, use 2 ml μmol^{-1} of support.
- 3) Close the vial or bottle tightly and seal further with Parafilm.
- 4) Place the vial or bottle in a preheated oven at 65 °C for 20 min for small vials but 40 min for larger bottles, which take longer to equilibrate thermally.
- 5) Allow the vial/bottle to cool completely to room temperature before opening in a fume cupboard.
- 6) Carefully remove the supernatant and wash the support several times with a few milliliters of ethanol/sterile water (1 : 1 by volume).
- 7) Combine the supernatant and washings in a Falcon tube and dry in a SpeedVac or, for bigger volumes, evaporate to dryness on a rotary evaporator. Do not use water bath temperatures above 30 °C for trityl-on material.
- 8) Dry the residue once by evaporation of absolute ethanol.

Desilylation of Trityl-Off RNA

- 1) Add a 1 : 1 mixture of dry DMSO and triethylamine *tris*(hydrofluoride) [30], using 600 μl μmol^{-1} , to the trityl-off residue in the Falcon tube as obtained in the Section “Deprotection of Base-Labile Protecting Groups” and sonicate briefly. If you have dried down the oligoribonucleotide in a glass flask, dissolve it in the required volume of dry DMSO using gentle warming of the flask using a hair dryer, transfer the solution to a Falcon tube, and add an equal volume of the fluoride reagent.
- 2) Close the tube, seal with Parafilm, and place it in a preheated oven at 65 °C for 2.5 h.
- 3) Cool the tube to room temperature.
- 4) Add 2 volumes of ethoxytrimethylsilane to destroy the excess fluoride reagent, close the tube, and shake vigorously at intervals for 10 min. At this point, a white precipitate appears.
- 5) Open the tube carefully and add 5 volumes of *tert*-butyl methyl ether, close, and agitate vigorously.
- 6) Collect the precipitate by centrifugation at 4000 rpm at 4 °C for 5 min.
- 7) Remove the supernatant by careful decantation.
- 8) Resuspend the pellet in *tert*-butyl methyl ether, close the tube, agitate, and again collect the precipitate by centrifugation.

- 9) Repeat steps 7 and 8 twice more.
- 10) Finally dry the RNA pellet carefully *in vacuo*.

Desilylation of Trityl-On RNA

- 1) Prepare a solution of *N*-methylpyrrolidone/triethylamine/triethylamine *tris*(hydrofluoride) (6 : 3 : 4 by volume) [31] immediately before use and add 600 μl μmol^{-1} of the trityl-on residue as obtained in the Section “Deprotection of Base-Labile Protecting Groups”. For material that has been dried down in a glass flask, dissolve the residue in the minimum volume of dry DMSO without warming, transfer the solution to a Falcon tube, and add 600 μl of freshly prepared desilylation solution per micromole.
- 2) Perform steps 2–8 as described in Section “Desilylation of Trityl-Off RNA.”
- 3) Finally dry the RNA pellet very briefly using a stream of argon gas, then dissolve it immediately in sterile 0.1 M aqueous ammonium bicarbonate, and purify immediately by reversed-phase HPLC as the DMTr-group has a limited half-life under these conditions.

7.2.3

Purification

This entire section is devoted to the anion-exchange HPLC purification of fully deprotected RNA using gradients of sodium perchlorate [31] or lithium perchlorate [28] as chaotropes, the reversed-phase HPLC purification of trityl-on RNA, detritylation, and desalting. In this section, you will need the following items:

- 1) A biocompatible HPLC system from, for example, GE Healthcare, Waters Corporation (Milford, MA, USA), Dionex Corporation (Sunnyvale, CA, USA), or other.
- 2) A set of anion-exchange HPLC columns, for instance, MonoQ 5/5, Source 15Q 16/10, and/or a FineLINE 35 pilot column packed with Source 15Q (GE Healthcare) or TSK gel SuperQ-5PW resin (Tosoh Bioscience LLC, King of Prussia, PA, USA) or Dionex DNAPac PA100, PA200, or DNASwift SAX-1S monolith columns (Dionex Corporation). Alternatives to the FineLINE 35 column are the AP-2 and AP-5 columns from Waters and the Kronlab ECO glass columns from YMC Europe GmbH (Dinslaken, Germany), which are available in a variety of internal diameters of which the most useful are 15, 20, 25, 32, and 50 mm.
- 3) Sodium perchlorate (very high purity). *Note that this salt is toxic and corrosive.* For larger scale purifications, it is more convenient to work with sodium bromide (Fluka BioChemika Ultra grade or equivalent).
- 4) Disodium EDTA.
- 5) 1 M sterile Tris–HCl buffer, pH 7.4.
- 6) A Hi-Prep 26/10 desalting column (GE Healthcare).
- 7) Reversed-phase HPLC columns, for instance, Hamilton PRP-1, 7 $\times$ 305 mm, XTerraTM RP₈, 4.6 $\times$ 250 mm (Waters), XTerraTM RP₈, 19 $\times$ 300 mm, and/or

a FineLINE 35 pilot column packed with Source 15RPC (GE Healthcare). Equivalent columns are also available from Phenomenex and other suppliers.

- 8) HPLC grade acetonitrile.
- 9) High-purity ammonium bicarbonate.
- 10) Glacial acetic acid.

7.2.3.1 Anion-Exchange HPLC Purification

The purity of oligoribonucleotides less than 25 nt in length, obtained by anion-exchange HPLC as the only purification step, is perfectly adequate for most biological applications. It generally results in a purity in the range of 95–98%. Longer RNAs must be purified in the trityl-on mode (Section 7.2.3.2). For anion-exchange HPLC, it is recommended that a gradient of sodium perchlorate or sodium bromide in sterile water/acetonitrile (9 : 1 v/v) containing 50 mM Tris–HCl buffer pH 7.6 and 50 μ M EDTA be used. The reason for adding EDTA is to complex traces of heavy metals that could otherwise lead to degradation of the RNA. The presence of acetonitrile in the buffers minimizes non-specific interactions between the RNA and the separation medium, which can be quite pronounced if lipophilic tags such as fluorescent dyes or cholesterol are present. The 4 $\times$ 250 mm DNAPac columns are fine for small-scale purifications of up to about 100 μ g of RNA, but are generally reserved for analytical work. The recently introduced DNASwift™ SAX-1S 5 $\times$ 150 mm monolith column is an excellent choice for the purification of up to several milligrams of RNA with flow rates up to 3 ml min⁻¹ and short gradient times (around 15–20 min). Recommended are the Source 15Q 16/10 columns for purification of 1 μ mol scale syntheses with a flow rate of 5 ml min⁻¹, whereas for syntheses at the 10–100 μ mol scale, purification is best achieved using a FineLINE Pilot 35 column packed with Source 15Q or TSK gel SuperQ-5PW and eluted at 20 ml min⁻¹. The low salt or A buffer preferably contains 10 mM sodium perchlorate (or sodium bromide) and the high salt or B buffer contains 600 mM sodium perchlorate (or 1 M sodium bromide). It has been found that a gradient from 10 to 60% B for 40 min gives good resolution. When not in use, the columns should be stored in 20% ethanol in sterile water to prevent microbial growth. For long-term storage of columns, it is advisable to add 0.2% sodium azide as an antimicrobial. Before using a column that has been stored, wash it with several column volumes of sterile water. The column is then equilibrated by washing it with several column volumes of buffer B followed by several column volumes of 90% buffer A plus 10% buffer B before injecting the sample of fully deprotected RNA as obtained in the Section “Desilylation of Trityl-Off RNA” dissolved in buffer A and running the salt gradient. The desired product peak is the late-eluting major component. This material is then desalted as described in Section 7.2.3.4. A typical trace of an anion-exchange HPLC purification is shown in Figure 7.3. In this example, the oligomer is a 21-mer synthesized manually on a 20 μ mol scale and purified on Source 15Q packed in a FineLINE Pilot 35 column. As can be seen, the failure peaks are very small compared to the product peak, which elutes at about 25 min. These separation

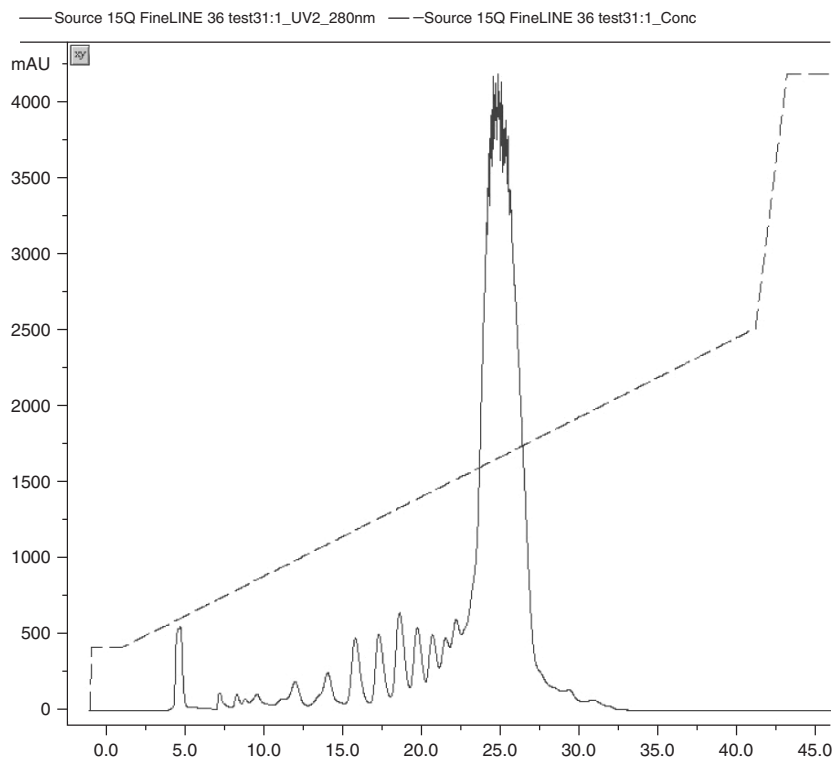


Figure 7.3 Preparative anion-exchange HPLC trace of a 21-mer oligoribonucleotide synthesized manually on 20 μmol scale and purified on Source 15Q packed in a Fine-*LINE* Pilot 35 column. The column was eluted with a linear gradient from 10 to 60% B for 40 min at a flow rate of 20 ml min^{-1} .

Buffer A was 10 mM sodium perchlorate, 50 μM EDTA and 50 mM Tris-HCl pH 7.6 in sterile water/acetonitrile (9 : 1 v/v), and buffer B was the same as buffer A except that the sodium perchlorate concentration was 600 mM. Absorbance was monitored at 280 nm. The x-axis is in minutes.

media should be cleaned in place every now and then using the recommended protocols from the manufacturer to ensure optimal performance and to avoid potential cross-contamination.

7.2.3.2 Reversed-Phase HPLC Purification of Trityl-On RNA

The highly lipophilic dimethoxytrityl group profoundly retards the full-length trityl-on RNA when purification is performed on a reversed-phase HPLC column. Although a better separation of failure peaks from the desired trityl-on product peak is obtained using aqueous triethylammonium acetate/acetonitrile buffers, for ease of salt removal and minimal damage to the RNA the use of ammonium bicarbonate is strongly preferred instead of triethylammonium acetate. However, make up the ammonium bicarbonate buffer fresh; otherwise, store it cold, since it has a limited stability at room temperature in contrast to triethylammonium acetate.

Columns recommended for trityl-on purification are the Hamilton PRP-1, 7×305 mm, for purifications on a few micromoles scale, or a 21.5×250 mm column for 10–20 μmol scale purifications. As an alternative for larger scale purifications, a FineLINE Pilot 35 column packed with Source 15RPC can be used. The buffers required are 0.1 M ammonium bicarbonate prepared in sterile water (buffer A) and 0.1 M aqueous ammonium bicarbonate/acetonitrile (1 : 1 by volume), which is buffer B. A useful gradient to use is 0–90% B for 40 min. The failure peaks elute early and are well separated from the desired trityl-on product peak, which elutes last. The pure product fraction should be collected in a polypropylene Falcon tube and dried down in a SpeedVac. Residual ammonium bicarbonate is then removed by lyophilization of the product, which is then ready for detritylation. A typical trace of a trityl-on RNA purification by reversed-phase HPLC is shown in Figure 7.4. The example shows a trityl-on 34-mer oligoribonucleotide synthesized by machine at a 20 μmol scale and purified on a FineLINE Pilot 35 column packed with Source 15RPC. As can be seen, the desired product peak elutes at 20–25 min well separated from the trityl-off failure sequences, which elute between 7 and 12 min.

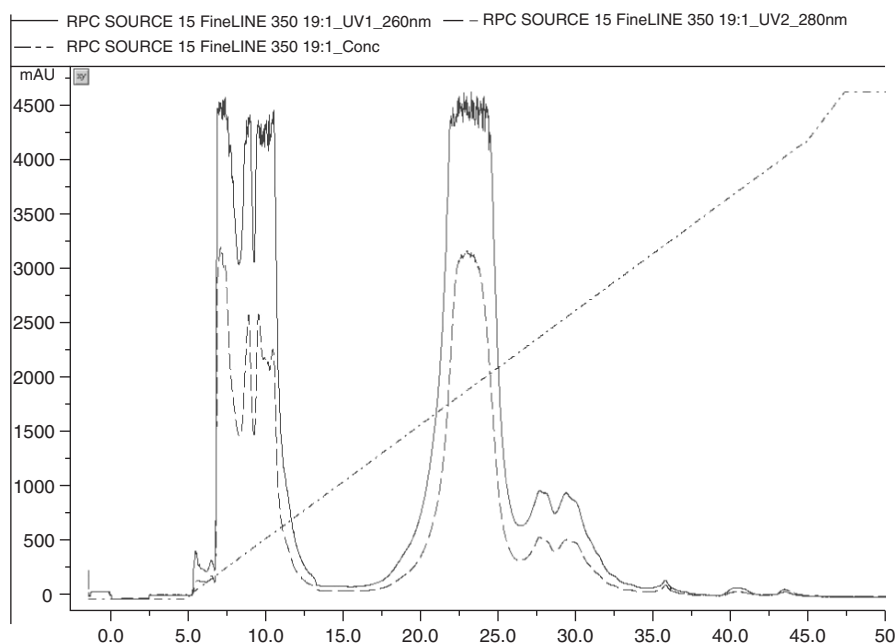


Figure 7.4 Preparative reversed-phase HPLC trace of a trityl-on 34-mer oligoribonucleotide synthesized by machine on a 20 μmol scale and purified on Source 15RPC packed in a FineLINE Pilot 35 column. The column was eluted with a linear gradient from 0 to 90% B for 40 min at

a flow rate of 25 ml min^{-1} . Buffer A was 0.1 M sterile aqueous ammonium bicarbonate and B was 0.1 M aqueous ammonium bicarbonate/acetonitrile (1 : 1 v/v). The solid line traces absorption at 260 nm, the dashed line that at 280 nm. The x-axis is in minutes.

Very recently, ion-pair reversed-phase HPLC has been proposed as a method for purification of single-stranded RNA as well as siRNA duplexes, whereby on-column annealing of the guide and passenger strands was performed leading to very high-quality duplexes [32].

7.2.3.3 Detritylation of Trityl-On RNA

- 1) Dissolve the HPLC-purified trityl-on RNA, as obtained in Section 7.2.3.2, in 3% sterile aqueous acetic acid ($200\ \mu\text{l}\ \mu\text{mol}^{-1}$) and keep for 45 min at room temperature. The pH should be about 3.5.
- 2) Neutralize the solution by careful addition of solid ammonium bicarbonate until evolution of carbon dioxide ceases. The pH will now be about 7.8.
- 3) Dry the sample in a SpeedVac.
- 4) Repurify the product by anion-exchange HPLC, as described in Section 7.2.3.1, which, in addition, converts the RNA from the ammonium form into the biologically preferred sodium form.
- 5) Desalt according to Section 7.2.3.4.

As an alternative to steps 4 and 5, the salt exchange can be performed in a reliable and high-yielding manner by dissolving the residue from step 3 in sterile 0.3 M aqueous sodium acetate ($400\ \mu\text{l}\ \mu\text{mol}^{-1}$ synthesis scale) and adding 2.5 volumes of absolute ethanol. After mixing and storage at -70°C for 20 min, the precipitated RNA is recovered by centrifugation, washed once with absolute ethanol, and then dried carefully *in vacuo*. To ensure a complete exchange of cation from ammonium to sodium, the precipitation procedure should be repeated once more. This protocol is convenient if the detritylated RNA is of sufficient purity as determined by analytical anion-exchange HPLC or polyacrylamide gel electrophoresis (PAGE).

7.2.3.4 Desalting by HPLC

Oligoribonucleotides purified by anion-exchange HPLC must be free of excess salt, buffer, and EDTA regardless of the intended application. This is readily achieved using a desalting column such as the Hi-Prep 26/10, which is filled with Sephadex G-25. The sample should be loaded on in a volume not greater than 15 ml, but preferably less than 10 ml, to obtain a complete separation between the RNA that elutes first and the salt peak that follows it. The column should be stored in and eluted with 20% ethanol in sterile water, which prevents microbial growth. For long-term storage, it is recommended that 0.2% sodium azide be added as an antimicrobial agent. Careful monitoring of the column effluent by UV and conductivity avoids contamination of the RNA by incompletely removed salt, which does not absorb in the UV spectral range. The desalted RNA in aqueous ethanol is first concentrated in a SpeedVac and finally lyophilized to give the pure RNA in its sodium form as a fluffy white solid.

As an alternative to the HPLC method, fast desalting of micromole-scale RNA samples can be achieved using the NAP columns from Amersham. For very large-scale desalting, ultrafiltration is the method of choice, since, in addition, it

results in sample concentration, whereas other desalting methods such as dialysis and gel filtration lead to sample dilution. Suitable laboratory-scale tangential flow filtration (TFF) systems useful for ultrafiltration and diafiltration are the Minim™ II (Pall), the Cogent μ Scale (Millipore) and the AccuTec™ (SciLog) for instance. It is recommended that 1 or 3 kDa cutoff cassettes be used when performing ultrafiltration/diafiltration of oligoribonucleotides. The desalted RNA is then isolated by freeze-drying.

7.2.4

Analysis of the Purified RNA

The purity of the RNA purified by the HPLC methods described above should be checked by analytical anion-exchange HPLC using a Dionex DNAPac PA200 column and buffers described in Section 7.2.3.1. Suitable alternatives to anion-exchange HPLC are capillary gel electrophoresis and PAGE, both of which operate under denaturing conditions. These methods are not described here. Nowadays, it is also recommended that the purity be analyzed by ion-pair reversed-phase HPLC on, for instance, an XBridge™ OST C18 4.6 $\times$ 50 mm, 2.5 μ m column (Waters) using a gradient of acetonitrile in aqueous triethylammonium acetate buffer pH 7.

Moreover, RNA destined for structural studies such as NMR spectroscopy or X-ray crystallography should certainly be analyzed in addition by analytical reversed-phase HPLC on a high-resolution column such as the 5 μ m XTerra™ RP₈, 4.6 $\times$ 250 mm column from Waters using a gradient from 0 to 20% acetonitrile in aqueous ammonium bicarbonate, combined with a flow rate of 1 ml min⁻¹. RNA that is deemed not to be of sufficient purity should be further purified by preparative reversed-phase HPLC on the 7 μ m XTerra™ RP₈, 19 $\times$ 300 mm column using a gradient similar to that used for the analytical reversed-phase analysis but with a flow rate of 12 ml min⁻¹. The material is desalted and exchanged to the sodium form using the procedures described above. As an example, the preparative reversed-phase HPLC trace of an anion-exchange HPLC-purified 27-mer oligoribonucleotide, destined for NMR spectroscopy, is illustrated in Figure 7.5 and an analytical reversed-phase HPLC of the now double-purified material is illustrated in Figure 7.6.

The recent developments in ultra performance liquid chromatography (UPLC) have dramatically improved the analytical sensitivity and resolution of synthetic oligonucleotides using sub-2 μ m particles for the separations. Generally, such an ion-pair reversed-phase UPLC system (Waters Acquity or nanoAcquity, or Agilent 1200) is directly coupled to an electrospray ionization (ESI) mass spectrometer. The separation is carried out on, for instance, an Acquity BEH OST C18, 50 $\times$ 2.1 mm, 1.7 μ m column (Waters) or Zorbax Eclipse column (Agilent) eluting with a gradient of methanol in an aqueous buffer pH 7.9 containing 400 mM hexafluoropropan-2-ol (HFIP) and 16.3 mM triethylamine at 60 °C. The flow rate is generally around 0.2 ml min⁻¹, resulting in gradient runs of approximately 10 min. This HFIP-containing buffer was first developed by Apffel *et al.* [33], and separation

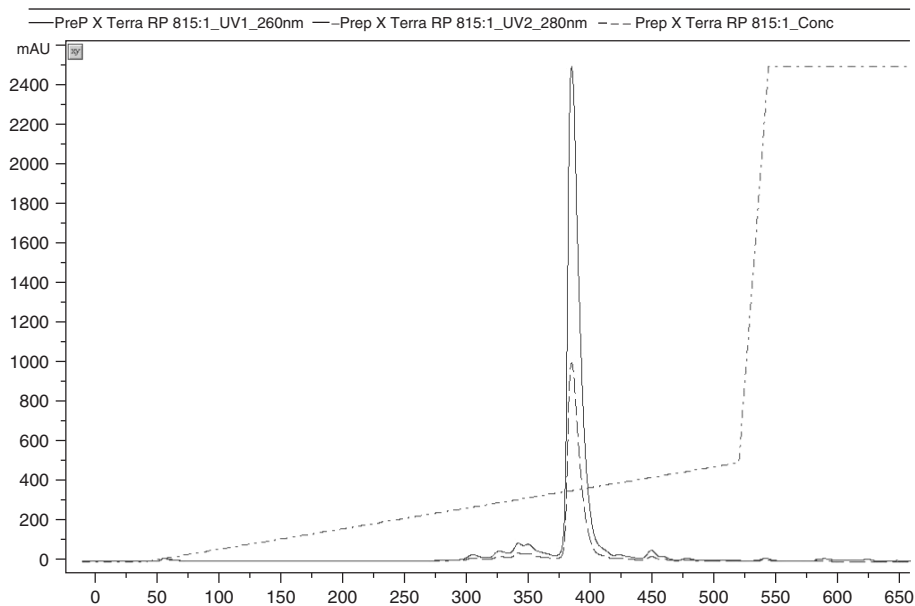


Figure 7.5 Preparative reversed-phase HPLC trace of a 27-mer oligoribonucleotide on a 7 μm XTerraTM RP₈, 19 $\times$ 300 mm column. The compound was initially purified trityl-on by anion-exchange HPLC. The column was eluted with a linear gradient from 0 to 20%

B for 40 min at a flow rate of 12 ml min⁻¹. Buffer A was 0.1 M sterile aqueous ammonium bicarbonate and B was acetonitrile. The solid line traces absorption at 260 nm, the dashed line that at 280 nm. The x-axis is in milliliters.

is according to length and there are no base composition dependencies [34]. To check product authenticity, the molecular weight of the RNA should be determined by mass spectroscopy, either ESI coupled with UPLC as above or by matrix-assisted laser desorption ionization time of flight (MALDI-TOF) [35]. It should be mentioned here that the latter method has a lower mass accuracy. A mass determination is absolutely essential when modified nucleotides that contain, for instance, unusual protecting groups are incorporated. As a final proof of authenticity, the RNA can also be sequenced by standard methods to check the absolute order of the monomeric units within the sequence, or more conveniently by a combination of mass spectroscopy methods and chemical/enzymatic degradation [36–38]. These methods are outside the scope of this chapter and the reader is advised to consult the literature references given above.

7.3

Troubleshooting

Many potential sources of problems can be eliminated by taking some simple precautions during the synthesis and purification steps. The following points

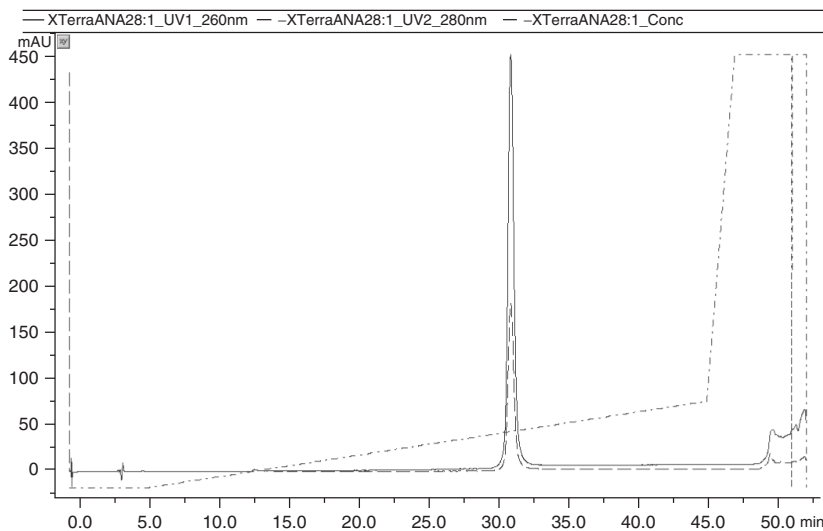


Figure 7.6 Analytical reversed-phase HPLC trace of a double HPLC-purified 27-mer oligoribonucleotide run on a 5 μm XTerraTM RP₈, 4.6 $\times$ 250 mm column. The column was eluted with a linear gradient from 0 to 20% B for 40 min at a flow

rate of 1 ml min^{-1} . Buffer A was 0.1 M sterile aqueous ammonium bicarbonate and B was acetonitrile. The solid line traces absorption at 260 nm, the dashed line that at 280 nm. The x-axis is in minutes.

should be heeded during the solid-phase synthesis of RNA. The coupling step in the solid-phase synthesis is very sensitive to traces of water and it is essential to use very dry acetonitrile for monomer dissolution. It is also imperative to allow monomers to reach room temperature before opening and weighing out material; otherwise, condensation of atmospheric water will occur leading to eventual degradation. The bottle contents should also be put back under argon before sealing and storing again at -20°C . Addition of activated molecular sieves or trap bags to the monomer and activator solutions ensures that they stay dry during the synthesis. It is also critical to wash away the acid from the detritylation step with copious acetonitrile washes; otherwise, residual acid will cause serious problems with the coupling step.

Incomplete oxidation will cause serious problems with the overall synthesis yield and quality as any residual phosphite triester is cleaved at the internucleotide linkage by the acid used in the detritylation step. As mentioned, the standard oxidation mixture from several commercial suppliers is only 17 mM in iodine, that is, 17 $\mu\text{mol ml}^{-1}$; so use enough solution for larger scale syntheses to ensure that there is an ample excess of reagent and/or increase the iodine concentration to 50 mM. For a 20 μmol scale synthesis, use 5 ml of the oxidation mixture. At the end of the solid-phase assembly, the CPG or polystyrene bearing trityl-off protected RNA can be stored cold and dry ready for deprotection at an appropriate time. However, CPG-bearing trityl-on RNA must be deprotected and purified immediately on

completion of the synthesis; otherwise, there will be partial or complete loss of the trityl group during storage. This is particularly bad for RNAs that terminate at the 5' end with one or more G residues.

For safety reasons, handle all chemicals in a well-ventilated fume cupboard and wear suitable resistant disposable gloves, particularly when handling toxic materials such as triethylamine *tris*(hydrofluoride). In addition, make sure that you have read the handling protocols for all chemicals, in particular those with which you are not familiar. Concerning the deprotection step with ammonia/methylamine, avoid too great an air space in the vial or bottle; otherwise, most of the ammonia and methylamine will end up in the vapor phase. In the worst case, this could lead to an incomplete deprotection. Take care when handling trityl-on RNA; do not overdry or let it get too warm, and purify immediately to avoid partial detritylation that will result in an unnecessary loss of product. This problem seems to be particularly serious with sequences that have one or more G residues at the 5' end.

To avoid inadvertent degradation of unprotected RNA by RNases, use RNase-free salts, sterilize all buffers by autoclaving, and sterilize all glassware in a 180 °C oven. In addition, wear disposable gloves and use sterile plasticware such as Eppendorf and Falcon tubes as much as possible. Concerning anion-exchange HPLC purification, oligoribonucleotides that contain four or more consecutive G residues are notoriously difficult to purify since they form tetraplexes and higher aggregates in solution. Such RNAs are best purified using a lithium perchlorate gradient since these structures do not form if lithium ions are present instead of sodium or potassium ions. Of course, before use in biological experiments, the lithium ions must be replaced with sodium ions since lithium ions are toxic in many biological systems. Denaturants such as formamide (not really recommended because of its toxicity) or urea (up to 8 M) can also be added to the salt gradient to reduce problems caused by strong secondary structures. An alternative is to perform the chromatography at high temperature, typically around 60 °C; however, this necessitates the use of a thermostatted column and preheated buffers. For all HPLC work and counterion changes, it is essential to use the highest quality reagents commercially available, such as those used for molecular biology work, and to ensure that they are RNase- and DNase-free and have very low (a few parts per million maximum) residual heavy metal ion content.

In large-scale purifications, to avoid product shoot through due to the ionic strength of the applied sample solution being too high, it is advisable to apply the crude RNA sample to the FineLINE Pilot 35 column dissolved in a volume of 10–50 ml of 50 mM Tris–HCl buffer pH 7.4, using a 10 or 50 ml superloop. As an alternative, desalt the sample before purification.

It is important to note that RNA samples as sodium salts are not suitable for mass spectrometry. Mass spectrometry samples of RNA are best prepared as ammonium salts. This can be done in several ways. One way is to perform anion-exchange HPLC using ammonium sulfate for elution, followed by desalting on a small NAP cartridge. In this case, store all buffers in plastic bottles and collect the product in an Eppendorf tube. A second way is to take a small aliquot of the RNA in its sodium form and exchange the sodium ions with ammonium

ions by using ammonium form Dowex 50 cation exchange resin. A third way is to purify a small sample of RNA by reversed-phase HPLC using the aqueous ammonium bicarbonate/acetonitrile system followed by lyophilization to remove the residual salt. Once in the ammonium form, the RNA should not be in contact with glass surfaces; otherwise, sodium and potassium ions will be picked up, which will severely impair the quality of the mass spectra. This latter point is of great importance when you try to analyze very long RNAs, for instance, in the 40- to 70-mer range.

Following the protocols and troubleshooting hints given above, the reader should be in a position to synthesize and purify RNA with success.

References

1. Ramaswamy, G. and Slack, F.J. (2002) siRNA. A guide for RNA silencing. *Chem. Biol.*, **9**, 1053–1055.
2. Couzin, J. (2002) Breakthrough of the year. Small RNAs make big splash. *Science*, **298**, 2296–2297.
3. Usman, N., Ogilvie, K.K., Jiang, M.-Y., and Cedergren, R.J. (1987) Automated chemical synthesis of long oligoribonucleotides using 2'-O-silylated ribonucleoside 3'-O-phosphoramidites on a controlled-pore glass support: synthesis of a 43-nucleotide sequence similar to the 3'-half molecule of an *Escherichia coli* formylmethionine tRNA. *J. Am. Chem. Soc.*, **109**, 7845–7854.
4. Pitsch, S., Weiss, P.A., Jenny, L., Stutz, A., and Wu, X. (2001) Reliable chemical synthesis of oligoribonucleotides (RNA) with 2'-O-[(triisopropylsilyl)oxy]methyl(2'-O-tom)-protected phosphoramidites. *Helv. Chim. Acta*, **80**, 3773–3795.
5. Scaringe, S.A., Wincott, F.E., and Caruthers, M.H. (1998) Novel RNA synthesis method using 5'-O-silyl-2'-O-orthoester protecting groups. *J. Am. Chem. Soc.*, **120**, 11820–11821.
6. Scaringe, S.A. (2000) Advanced 5'-silyl-2'-orthoester approach to RNA oligonucleotide synthesis, in *Methods in Enzymology*, vol. 317. Academic Press, pp. 3–18.
7. Ohgi, T., Masutomi, Y., Ishiyama, K., Kitagawa, H., Shiba, Y., and Yano, J. (2005) A new RNA synthetic method with a 2'-O-(2-cyanoethoxymethyl) protecting group. *Org. Lett.*, **7**, 3477–3480.
8. Shiba, Y., Kitagawa, H., Masutomi, Y., Ishiyama, K., Ohgi, T., and Yano, J. (2006) A novel RNA synthetic method with a 2'-O-(2-cyanoethoxymethyl) protecting group. *Nucleic Acids Symp. Ser.*, **50**, 11–12.
9. Shiba, Y., Masuda, H., Watanabe, N., Ego, T., Takagaki, K., Ishiyama, K., Ohgi, T., and Yano, J. (2007) Chemical synthesis of a very long oligoribonucleotide with 2-cyanoethoxymethyl (CEM) as the 2'-O-protecting group: structural identification and biological activity of a synthetic 110 mer precursor-microRNA candidate. *Nucleic Acids Res.*, **35**, 3287–3296.
10. Ohgi, T., Kitagawa, H., and Yano, J. (2008) Chemical synthesis of oligoribonucleotides with 2'-O-(2-cyanoethoxymethyl)-protected phosphoramidites. *Curr. Protoc. Nucleic Acid Chem.*, **34**, 2.15.1–2.15.19.
11. Zhou, C., Honcharenko, D., and Chattopadhyaya, J. (2007) 2-(4-Tolylsulfonyl)ethoxymethyl (TEM)-a new 2'-OH protecting group for solid-supported RNA synthesis. *Org. Biomol. Chem.*, **5**, 333–343.
12. Zhou, C., Pathmasiri, W., Honcharenko, D., Chatterjee, S., Barman, J., and Chattopadhyaya, J. (2007) High-quality oligo-RNA synthesis using the new 2'-O-TEM protecting group by selectively quenching the addition of p-tolyl vinyl sulphone to exocyclic amino functions. *Can. J. Chem.*, **85**, 293–301.

13. Cieślak, J., Grajkowski, A., Kauffman, J.S., Ruff, R.J., and Beaucauge, S.L. (2008) The 4-(*N*-dichloroacetyl-*N*-methylamino) benzyloxymethyl group for 2'-hydroxyl protection of ribonucleosides in the solid-phase synthesis of oligoribonucleotides. *J. Org. Chem.*, **73**, 2774–2783.
14. Lackey, J.G. and Damha, M.J. (2008) The acetal levulinyl ester (ALE) group for the 2'-hydroxyl protection of ribonucleosides and the synthesis of oligoribonucleotides. *Nucleic Acids Symp. Ser.*, **52**, 35–36.
15. Lackey, J.G., Mitra, D., Somoza, M.M., Cerrina, F., and Damha, M.J. (2009) Acetal levulinyl ester (ALE) groups for 2'-hydroxyl protection of ribonucleosides in the synthesis of oligoribonucleotides on glass and microarrays. *J. Am. Chem. Soc.*, **131**, 8496–8502.
16. Lavergne, T., Martin, A., Debart, F., and Vasseur, J.J. (2008) A whole base-labile strategy for RNA synthesis with 2'-O-acetalester protections. *Nucleic Acids Symp. Ser.*, **52**, 51–52.
17. Lavergne, T., Bertrand, J.R., Vasseur, J.J., and Debart, F. (2008) A base-labile group for 2'-OH protection of ribonucleosides: a major challenge for RNA synthesis. *Chem. Eur. J.*, **14**, 9135–9138.
18. Martin, A.R., Lavergne, T., Vasseur, J.J., and Debart, F. (2009) Assessment of new 2'-O-acetalester protecting groups for regular RNA synthesis and original 2'-modified proRNA. *Bioorg. Med. Chem. Lett.*, **19**, 4046–4049.
19. Dellinger, D.J., Timär, Z., Myerson, J., Sierzchala, A.B., Turner, J., Ferreira, F., Kupihâr, Z., Dellinger, G., Hill, K.W., Powell, J.A., Sampson, J.R., and Caruthers, M.H. (2011) Streamlined process for the chemical synthesis of RNA using 2'-O-thionocarbamate-protected nucleoside phosphoramidites in the solid phase. *J. Am. Chem. Soc.*, **133**, 11540–11556.
20. Matteucci, M.D. and Caruthers, M.H. (1981) Synthesis of deoxyoligonucleotides on a polymer support. *J. Am. Chem. Soc.*, **103**, 3185–3191.
21. Sinha, N.D., Biernat, J., McManus, J., and Köster, H. (1984) Polymer support oligonucleotide synthesis XVIII: use of beta-cyanoethyl-*N,N*-dialkylamino-*N*-morpholino phosphoramidite of deoxynucleosides for the synthesis of DNA fragments simplifying deprotection and isolation of the final product. *Nucleic Acids Res.*, **12**, 4539–4557.
22. Chaix, C., Molko, D., and Téoule, R. (1989) The use of labile base protecting groups in oligoribonucleotide synthesis. *Tetrahedron Lett.*, **30**, 71–74.
23. Sinha, N.D., Davis, P., Usman, N., Pérez, J., Hodge, R., Kremsky, J., and Casale, R. (1993) Labile exocyclic amine protection of nucleosides in DNA, RNA and oligonucleotide analog synthesis facilitating N-deacylation, minimizing depurination and chain degradation. *Biochimie*, **75**, 13–23.
24. Gasparutto, D., Livache, T., Bazin, H., Duplaa, A.-M., Guy, A., Khorlin, A., Molko, D., Roget, A., and Téoule, R. (1992) Chemical synthesis of a biologically active natural tRNA with its minor bases. *Nucleic Acids Res.*, **20**, 5159–5166.
25. Westman, E. and Strömberg, R. (1994) Removal of *t*-butyldimethylsilyl protection in RNA-synthesis. Triethylamine trihydrofluoride (TEA, 3HF) is a more reliable alternative to tetrabutylammonium fluoride (TBAF). *Nucleic Acids Res.*, **22**, 2430–2431.
26. Welz, R. and Müller, S. (2002) 5-(Benzylmercapto)-1H-tetrazole as activator for 2'-O-TBDMS phosphoramidite building blocks in RNA synthesis. *Tetrahedron Lett.*, **43**, 795–797.
27. Vargeese, C., Carter, J., Yegge, J., Krivjansky, S., Settle, A., Kropp, E., Peterson, K., and Pieken, W. (1998) Efficient activation of nucleoside phosphoramidites with 4,5-dicyanoimidazole during oligonucleotide synthesis. *Nucleic Acids Res.*, **26**, 1046–1050.
28. Sproat, B., Colonna, F., Mullah, B., Tsou, D., Andrus, A., Hampel, A., and Vinayak, R. (1995) An efficient method for the isolation and purification of oligoribonucleotides. *Nucleosides Nucleotides*, **14**, 255–273.
29. Sproat, B.S. and Gait, M.J. (1984) in *Oligonucleotide Synthesis: A Practical Approach* (ed. M.J. Gait), IRL Press, Ltd, Oxford, pp. 83–115.

30. Vinayak, R., Andrus, A., and Hampel, A. (1995) Rapid desilylation of oligoribonucleotides at elevated temperatures: cleavage activity in ribozyme-substrate assays. *Biomed. Peptides, Proteins Nucleic Acids*, **1**, 227–230.
31. Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., and Usman, N. (1995) Synthesis, deprotection, analysis and purification of RNA and ribozymes. *Nucleic Acids Res.*, **23**, 2677–2684.
32. McCarthy, S.M., Gilar, M., and Gebler, J. (2009) Reversed-phase ion-pair liquid chromatography analysis and purification of small interfering RNA. *Anal. Biochem.*, **390**, 181–188.
33. Apffel, A., Chakel, J.A., Fischer, S., Lichtenwalter, K., and Hancock, W.S. (1997) Analysis of oligonucleotides by HPLC-electrospray ionization mass spectrometry. *Anal. Chem.*, **69**, 1320–1325.
34. Fountain, K.J., Gilar, M., and Gebler, J.C. (2003) Analysis of native and chemically modified oligonucleotides by tandem ion-pair reversed-phase high-performance liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.*, **17**, 646–653.
35. Pieves, U., Zürcher, W., Schär, M., and Moser, H.E. (1993) Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic Acids Res.*, **21**, 3191–3196.
36. Huang, T.Y., Liu, J., Liang, X., Hodges, B.D., and McLuckey, S.A. (2008) Collision-induced dissociation of intact duplex and single-stranded siRNA anions. *Anal. Chem.*, **80**, 8501–8508.
37. Bahr, U., Aygün, H., and Karas, M. (2009) Sequencing of single and double stranded RNA oligonucleotides by acid hydrolysis and MALDI mass spectrometry. *Anal. Chem.*, **81**, 3173–3179.
38. Farand, J. and Gosselin, F. (2009) De novo sequence determination of modified oligonucleotides. *Anal. Chem.*, **81**, 3723–3730.

8

Modified RNAs as Tools in RNA Biochemistry

Thomas E. Edwards and Snorri Th. Sigurdsson

8.1

Introduction

RNA displays a vast variety of functions in that it carries genetic information, regulates gene expression, catalyzes reactions, and participates in all facets of protein expression [1]. In addition to the four basic nucleosides (adenosine, guanosine, cytidine, and uridine), many RNA molecules contain modified nucleosides essential for function. The fact that these modifications are essential for function in some RNAs but entirely absent in others indicates a significant layer of complexity in the hierarchy of RNA structure. With the progress in the chemical synthesis of RNA over the past 20 years, modified nucleosides can now be readily incorporated at specific positions in RNA. These advances in solid-phase synthesis have promoted a cornucopia of experiments examining the influence of single functional group modification on the biological function of RNA.

Modified nucleosides have also been site-specifically incorporated into RNA as reporter groups for biochemical and biophysical structure–function analysis. There is a large diversity in these approaches. For example, fluorescent probes have been used to report internal changes during RNA folding [2] as well as to measure interhelical distances for determining the global structure of RNA [3–5]. Disulfide crosslinks have been used to restrict RNA helical elements to validate structural models based on other techniques [6]. These are but a few examples of site-specific incorporation of RNA structure–function probes.

The major goals of this chapter are to review the various types of modifications that can be incorporated site-specifically into RNA by chemical synthesis and to provide a general method for the incorporation of reporter groups into RNA for biochemical and biophysical analysis. This includes comparison of the two central strategies for the incorporation of modified nucleosides into RNA, the phosphoramidite strategy and postsynthetic labeling. The phosphoramidite strategy utilizes chemical synthesis of a modified nucleoside phosphoramidite in conjunction with solid-phase synthesis, whereas postsynthetic labeling utilizes incorporation by the phosphoramidite method of a convertible nucleoside containing a reactive group, which is selectively modified after oligonucleotide synthesis with a labeling reagent.

The advantages and disadvantages of each modification strategy are described. Finally, two general and efficient modification strategies are presented, postsynthetic labeling of 2'-amino groups and 4-thiouridine.

8.1.1

Modification Strategy: the Phosphoramidite Method

While enzymatic synthesis can be used to prepare uniformly labeled RNA, modified nucleosides can be incorporated site-specifically into RNA by solid-phase chemical synthesis using modified nucleoside phosphoramidites [7]. The main advantage of this method is that it allows for the incorporation of a desired modification or reporter group at a specific position in the RNA. While this is a highly effective and powerful method, it has several disadvantages. In most cases, the synthesis of a modified phosphoramidite requires a lengthy and costly synthetic route. Furthermore, the reporter group must be stable under the conditions used in solid-phase oligonucleotide synthesis (e.g., incubation with acid, base, and oxidizing solutions) as well as the deprotection conditions. However, phosphoramidites of many desirable modified nucleosides are commercially available, providing rapid, cost-effective access to a variety of modified RNAs (Table 8.1).

There are four basic categories of RNA modifications that can be incorporated into RNA *via* the phosphoramidite method: end-(5' and 3'), base, phosphate, and sugar modifications. Figure 8.1 shows selected examples of such modifications and Table 8.1 lists several of the RNA modifications that are commercially available as phosphoramidites and/or modified RNAs. Notable examples of RNA modification that utilize phosphoramidites that are not commercially available (many of which are reviewed in [17, 18, 25–28]) include the base modifications 2-deoxyribonolactone [29] and 5-ketone pyrimidine derivatives [30] as well as nitroxide spin labels in purines [27, 31] or at the 5 position of pyrimidines [27, 32]. Internucleotide linkages include boranophosphates [33] and phosphoroselenoates [34]. Sugar modifications include 1'-deutero [35], 2'-modifications (O-(2-thioethyl) [36] and O-(2-aminoethyl) [37]), 5'-modifications (tallo or C-methyl [38, 39], chloro [40], and amino [40]), and perdeuterated ribose [41]. Fluorescent labels have also been synthetically attached to the 2' position *via* an ether linkage [42], a carbamoyl linker [43], an arabino carbamoyl linker [44], and an amido linkage [45].

The phosphoramidite method has been particularly useful in the investigation of ribozyme cleavage mechanisms. For example, incorporation of a 5'-C-methyl-modified nucleoside near the cleavage site of the hammerhead ribozyme resulted in a kinetically trapped intermediate in a crystal [39]. A crystal structure of the hairpin ribozyme containing a 5'-chloro group at the cleavage site provided structural information for comparison with the non-cleaved state (all RNA) and a vanadyl transition state mimic, providing valuable information about the entire mechanistic pathway [46]. The 2'–5' phosphodiester linkage has also been incorporated *via* the phosphoramidite method for X-ray crystallographic analyses of ligation chemistry of ribozymes [47, 48].

Table 8.1 Commercially available modifications that can be incorporated into RNA by the phosphoramidite method.

Site	Modification	Commercial source ^a	References ^b
End labeling			Reviewed in [8–10]
5' End	Fluorescent dyes	CG, TS, GR	[3–5]
	Amino groups	CG, TS, GR	[11, 12]
	Biotin	CG, TS, GR	—
	Photo-cleavable biotin	TS, GR	—
	5'-Thiol	TS, GR	—
	Acridine	GR	—
3' End	Cholesterol	TS	—
	Fluorescent dyes	CG, TS, GR	—
	Amino groups	CG, TS, GR	—
	Inverted abasic	TS	—
	Puromycin	CG, TS	[13]
	Dideoxy G,C	TS	—
	Biotin	CG, GR	—
	Acridine	GR	—
	Psoralen	CG, GR	—
	Cholesterol	CG, GR	—
Internucleotide	DNP	CG, GR	—
	S (non-bridging)	CG, TS, GR	[14–16]
	3, 9, 18 Atom spacers	CG, TS, GR	—
2'–5'		GR	—
Sugar			Reviewed in [17, 18]
1'	Abasic	TS, GR	—
2'	NH ₂ U,C	TS, CG	[19]
	F U,C	CG, TS	—
	OCH ₃	CG, TS, GR	—
	SCH ₃ U	GR	[20]
	OCH ₂ CH ₂ CH ₂ NH ₂	CG	[21]
	NHCOCH ₂ CH ₂ CH ₂ pyrU	TS	[22]
	LNA	RT, EX, GR	[23]
Purine			Reviewed in [17, 18]
	N6,N6-Dimethyl A	TS	—
	Inosine	CG, TS, GR	—
	Purine ribonucleoside	CG, TS	—
	Ribavirin	TS	—
	7-Deaza A,G	CG	—
	2-Aminopurine	CG, TS, GR	[2]
	2,6-Diaminopurine	TS	—
Pyrimidine	8-Bromo A	CG	—
			Reviewed in [17, 18]
	N3-Methyl U,rT	TS, CG	—
	N3-Thiobenzoyl ethyl U	CG	—
	4-Triazoyl U	CG	—

(continued overleaf)

Table 8.1 (Continued)

Site	Modification	Commercial source ^a	References ^b
	N4-Ethyl C	CG	—
	Pyridine-2-one	CG	—
	Pyrrolo-C	GR	—
	2,2'-Anhydro U	CG	—
	5-Methyl U,C	TS	—
	4-Thio uridine	CG, TS, GR	[24]
	5-Fluoro U,C	CG, TS, GR	—
	5-Bromo U,C	CG, TS, GR	—
	5-Iodo U	CG, TS, GR	—
	5-Methyl C	TS	—
	Pseudouridine	CG, TS, GR	—
	5-CHCHCH ₂ NH ₂ U	TS	—
	2'-N-Butyryl pyrene U	TS	—

^aFor commercial sources: CG, ChemGenes; GR, Glen Research; TS, ThermoScientific; RT, RiboTask; EX, Exiqon. Note that chemical suppliers are subject to change and this list is a representative example at the time of publication. Several other companies that sell modified RNA and modified RNA phosphoramidites exist.

^bReferences are select examples and may be reviews, applications, or synthetic procedures.

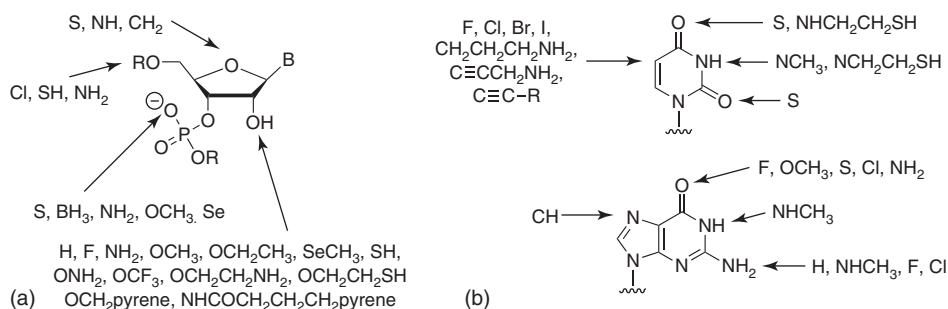


Figure 8.1 Selected examples of modifications that can be incorporated at the sugar (a), phosphodiester backbone (a) or base (b) using the phosphoramidite method.

8.1.2

Modification Strategy: Postsynthetic Labeling

Postsynthetic modification of convertible nucleosides enables the site-specific incorporation of a wide variety of reporter groups into RNA. The main advantage of this approach is that once the RNA has been prepared, it enables the rapid and efficient production of a wide variety of modified RNAs. Another advantage is that sensitive reporter groups that would otherwise be unstable to the conditions

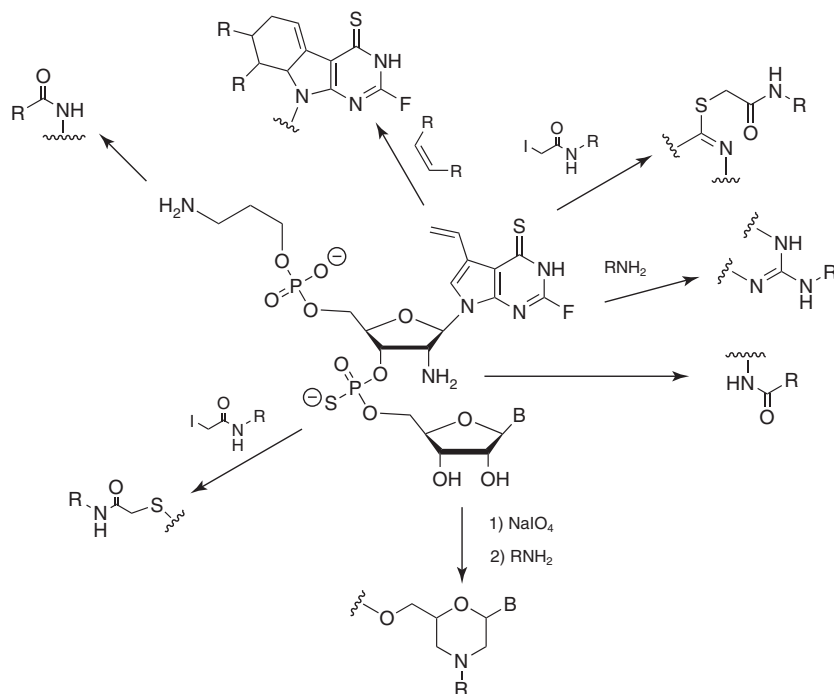


Figure 8.2 Selected examples of RNA postsynthetic labeling.

of solid-phase oligonucleotide synthesis can be incorporated into RNA. Possible disadvantages of this strategy are that in some cases additional purification steps are necessary and that it may be necessary to synthesize the convertible nucleoside phosphoramidite, if it is not commercially available.

Postsynthetic modification strategies have been developed for attachment of reporter groups to the 5' and 3' ends and at internal sites on the base, phosphate, and sugar (Figure 8.2, Table 8.2). The main focus of this chapter is the general strategy of postsynthetic labeling of 2'-amino-containing RNA and this approach is described in detail in the next section. In addition to the modifications shown in Figure 8.2 and Table 8.2, various groups can be attached to the 5-position of pyrimidines *via* on-column Pd-catalyzed coupling reactions [27, 49]. A variety of molecules have also been attached to the phosphodiester backbone through phosphorothioate [14–16] or phosphoramidate linkages [50]. However, these labeling strategies are problematic for RNA because of the inherent instability of these linkages in the presence of 2'-hydroxyl groups; consequently, this problem is overcome by incorporation of a 2'-deoxy or 2'-methoxy group at the nucleotide 5' of the modification. Another disadvantage of phosphorothioate modifications is the generation of a diastereomeric mixture. However, the diastereomers can sometimes be physically separated [51].

Table 8.2 Select examples of modifications for postsynthetic RNA derivatization.

Modification	Molecular handle	Commercially available? ^a	Labeling reactants	References ^b
1	3-Amino modifiers	CG, TS, GR	Activated esters	[11, 12]
2 ^c	Diene	No	Dienophile	[52]
3	Sulfur-containing bases	CG, TS, GR	Iodoacetamides, disulfides	[53, 54] [24]
4	Convertible F or ClΦ ^d nucleosides	No	Amines	[55]
5	2'-Amino	CG, TS	Isothiocyanates, NHS esters, isocyanates	[19]
6	2', 3'-Diols	All (RNA)	NaIO ₄ , amines	[56]
7	Non-bridging phosphorothioates	CG, TS, GR	Iodoacetamides	[14–16, 57]

^aFor commercial sources: CG, ChemGenes; GR, Glen Research; TS, ThermoScientific. Please note that chemical suppliers are subject to change and this list is a representative example at the time of publication. Several other companies that sell modified RNA and modified RNA phosphoramidites exist.

^bReferences are select examples and may be reviews, applications, or synthetic procedures.

^cThis modification strategy has only been applied to DNA thus far, but is of select interest.

^d4-Chlorophenoxy.

8.2

Description of Methods

8.2.1

Postsynthetic Modification: the 2'-Amino Approach

Postsynthetic labeling of the 2'-amino group (Figure 8.3a,b) has emerged as an effective approach for the site-specific incorporation of reporter groups into RNA. Several notable examples include the incorporation of disulfide crosslinking reagents for the evaluation of RNA helical orientation [6, 58, 59], the incorporation of fluorescent probes for the study of RNA folding and ligand binding [22, 60], and the incorporation of Electron Paramagnetic Resonance (EPR)-active probes [19, 61] for the study of RNA internal dynamics [62–67] and for distance measurements [68, 69]. RNAs containing 2'-amino groups at specific pyrimidine nucleotides (Thermo Scientific) and 2'-amino-modified pyrimidine phosphoramidites (ChemGenes) are now commercially available. Because the 2'-amino group is an aliphatic amine, it is more reactive (i.e., nucleophilic) than the aromatic amines or hydroxyl groups native to RNA, making this method of postsynthetic labeling highly selective. The major advantage of the 2'-amino group over other amino-based modifiers (e.g., 5'- and 3'-amino modifiers, 5-alkylamino-modified pyrimidines) is that it offers a minimal linker length. Several chemical conjugation approaches exist, including reaction with succinimidyl esters (often referred to as *N*-hydroxysuccinimide (NHS)

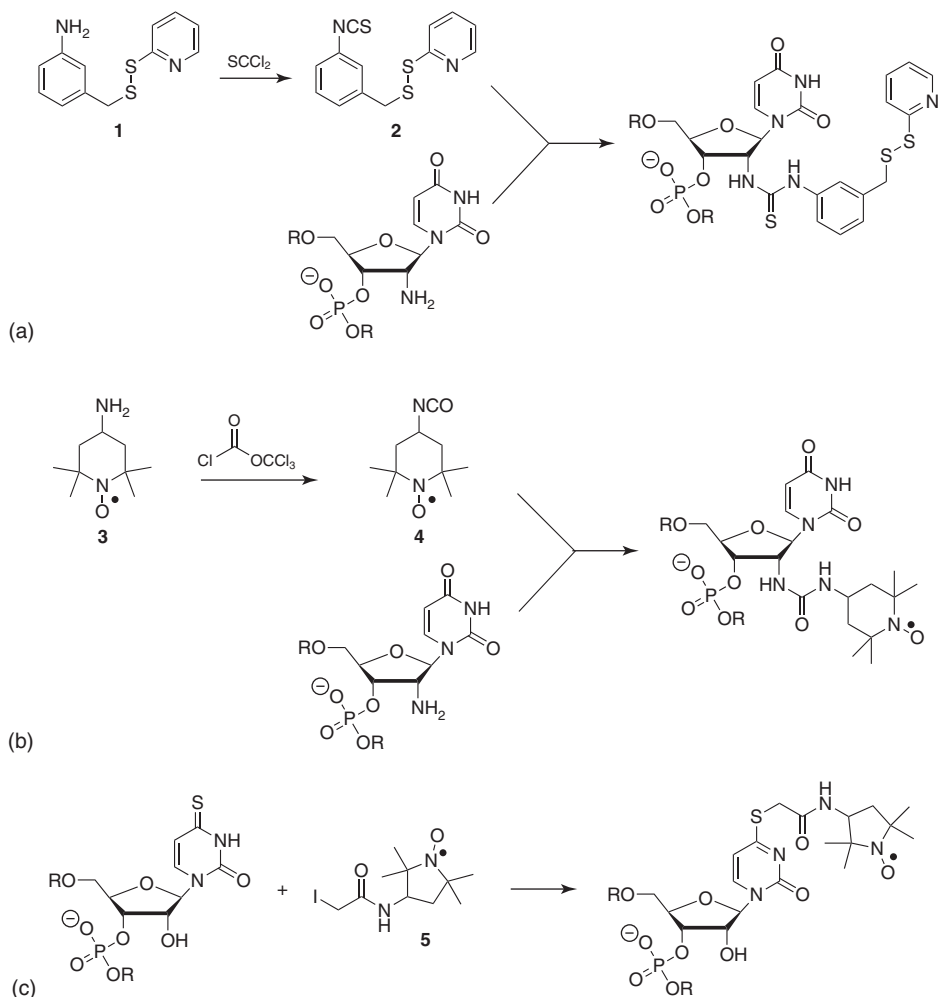


Figure 8.3 Preparation of an isothiocyanate crosslinking reagent **2** (a) and an EPR spin-labeling reagent 4-isocyanato TEMPO **4** (b) from the corresponding amines using thiophosgene and

diphosgene, respectively, and their subsequent incorporation into 2'-amino-modified RNA. (c) Postsynthetic modification of 4-thiouridine by alkylation with spin-label **5**.

esters) to produce amide-modified RNA [58], reaction with aromatic isothiocyanates to form thiourea-linked RNA [6, 70], and reaction with aliphatic isocyanates to prepare urea-tethered RNA [70, 71]. These three methods are described in detail below and examples employing these methods to address biochemical and biophysical questions are provided. Of notable importance for this modification strategy is an alternative 2' protection approach that has been developed based on a photocleavable protecting group in place of the standard 2'-trifluoroacetyl group; after removal

of the protecting group, the 2'-amino group may be derivatized on-column, providing many advantages over solution-based postsynthetic modification of deprotected oligonucleotides [72]. Other approaches for the attachment of reporter groups to the 2' position of oligonucleotides that will not be addressed in depth include chelation of metal ions such as ruthenium to 2'-amino-modified oligonucleotides [73, 74], incorporation of fluorescamine at a 2'-amino group using a Michael addition and rearrangement reaction [75], reaction of amines with a 2'-O-(acetaldehyde) group [76], and reduction of thiol-containing compounds with 2'-O-(2-thioethyl) to form disulfide-linked modified RNAs [36]. The attachment of sterically hindered compounds to the 2'-amino group may be difficult and can be overcome by use of the 2'-O-(2-aminoethyl) modification [37].

8.2.2

Reaction of 2'-Amino Groups with Succinimidyl Esters

The reaction of 2'-amino groups with succinimidyl esters to produce amido-linked modified RNAs has been used to incorporate thiol groups into RNA for crosslinking [58, 77–79]. Succinimidyl ester modifications of 2'-amino groups have also been used to identify base-pair mismatches [80, 81], to incorporate fluorescent pyrene labels [22, 82–84], and to probe steric interference in the hammerhead ribozyme [85]. Catalysis of this chemistry by the phosphodiester on the 3' position adjacent to the 2'-amino-containing nucleoside and/or the 3'-oxygen, by stabilizing the developing positive charge on the nitrogen by electrostatic or hydrogen bonding interactions, has been reported [86]. One advantage of this method is that many succinimidyl esters are commercially available (e.g., Invitrogen and ChemGenes have a wide variety of amine-reactive succinimidyl ester dyes available). The major drawback of this method is that this chemistry often suffers from low labeling efficiency, for example, the pyrene succinimidyl esters typically coupled with only 20–26% yield after purification [22]. However, in some cases, it is possible to overcome this low coupling efficiency by the use of the corresponding carboxylic acid with an activating agent, such as a carbodiimide, which may result in nearly quantitative coupling [37, 72, 77]. Another disadvantage of this modification approach is that 2'-amido modifications destabilize RNA when incorporated at internal positions ($\Delta T_m \sim -5$ to 12°C per modification) [45, 69, 87, 88]. However, some 2'-amido-linked modifications located at end positions increase RNA stability [45, 89], likely a result of these particular modifications that contain large aromatic groups (e.g., fluorescent labels), which may stack onto the end of the helix. Nevertheless, this destabilizing effect of 2'-amido groups at internal sites should be kept in mind when incorporating reporter groups into RNA.

8.2.3

Reaction of 2'-Amino Groups with Aromatic Isothiocyanates

The reaction of 2'-amino groups with aromatic isothiocyanates (Figure 8.3a) has been used to incorporate fluorescent probes [45, 60], disulfide crosslinks [6, 59],

and photocrosslinking agents [90] into RNA. The main advantage of this method is the highly efficient chemistry, which has resulted in reported conversion yields in excess of 90% in all cases. In addition, fewer equivalents of the isothiocyanate labeling reagent are required than for the succinimidyl ester chemistry. A drawback is that although some isothiocyanates are commercially available, most must be prepared from the corresponding amine and thiophosgene; however, this synthetic conversion is relatively straightforward [70, 91]. There is limited available UV thermal stability data for 2'-thioureido modifications, and all of the data involves incorporation of large fluorescent probes (fluorescein and rhodamine). Following a pattern similar to that observed for the 2'-amido modifications, these modifications are rather destabilizing at internal positions, but can have a stabilizing effect at end positions [45, 60].

8.2.4

Reaction of 2'-Amino Groups with Aliphatic Isocyanates

The reaction of 2'-amino groups with aliphatic isocyanates (Figure 8.3b) is a versatile platform for the incorporation of biochemical and biophysical reporter groups into RNA. This method has been used to incorporate disulfide crosslinks [59, 71, 92]; an activated disulfide that can be used to conjugate a wide variety of groups such as cholesterol [93], glutathione and biman [71], nitrophenol [72], pyrene [37], and nitroxide spin labels [19]. Similar to the isothiocyanate coupling, this chemistry is highly efficient, typically displaying quantitative yields for unhindered isocyanates [71, 72] and good yields for structurally hindered isocyanates (e.g., 90% yields are routinely observed for the secondary isocyanate 4-isocyanato-TEMPO [19, 62]). Owing to the high selectivity and efficiency of this reaction, the crude deprotected RNA can be labeled directly, allowing for only one purification and therefore high yields. Another advantage is the relatively fast coupling time (15–60 min). In addition, 2'-ureido modifications are not as destabilizing as 2'-amido modifications [71, 92]; for example, incorporation of the EPR spin probe TEMPO through a 2'-urea linkage at internal base-pairing nucleotides resulted in a minor decrease in stability as measured by a small decrease in melting temperature of 1–3 °C [19, 69]. The main drawback is that usually the isocyanate labeling reagent must be prepared from the corresponding amine; however, this chemistry is straightforward and pure isocyanates are obtained in high yields after purification by extraction [70, 71, 94].

8.3

Experimental Protocols

The general experimental protocols for two representative examples of RNA labeling by the 2'-amino approach are detailed: incorporation of a crosslinking reagent (Figure 8.3a) for validation of existing structural models [6, 70] and incorporation of an EPR spin probe (Figure 8.3b) for biophysical analysis of structure [68, 69] and dynamics of RNA molecular recognition [19, 62–67]. We have also included

an example of base-labeling using 4-thiouridine for RNAs that cannot be modified in the 2' position because of loss of function (Figure 8.3c).

8.3.1

Synthesis of Aromatic Isothiocyanates and Aliphatic Isocyanates

The 2'-amino postsynthetic labeling approach often requires the synthesis of the desired labeling reagent from the corresponding amine, which may be commercially available. The isothiocyanate crosslinking reagent **2** was prepared from the corresponding amine **1** and thiophosgene (Figure 8.3a) [6, 70], as reviewed in [61].

- 1) Add a solution of amine **1** (for synthesis, see [6]; 8.20 g, 33 mmol) in chloroform (250 ml) dropwise to a solution of thiophosgene (4.17 g, 36.3 mmol) in chloroform (50 ml) over 10 min at room temperature.
- 2) Stir for 1 h at room temperature.
- 3) Dilute the mixture with methylene chloride (330 ml).
- 4) Wash with NaOH (1 M aq., 165 ml).
- 5) Extract the aqueous phase with additional methylene chloride (40 ml).
- 6) Combine the organic phases.
- 7) Dry the combined organic phases (Na_2SO_4) and filter off the salt.
- 8) Remove the solvent *in vacuo*.
- 9) Purify the crude product by flash column chromatography (CH_2Cl_2).
- 10) This procedure produces an oil (in our hands 8.80 g, 92% yield).

The isocyanate spin-labeling reagent, 4-isocyanato TEMPO **4**, was prepared from 4-amino TEMPO **3** (Acros and Sigma–Aldrich) and diphosgene (Figure 8.3b) [19].

- 1) Precool a solution of **3** (198 mg, 1.15 mmol) in anhydrous CH_2Cl_2 (1.5 ml) in a rock salt/ice water bath at -8°C .
- 2) Separately, precool in the same bath a solution of trichloromethyl chloroformate (diphosgene, 25 μl , 0.21 mmol) in CH_2Cl_2 (1.5 ml).
- 3) Rapidly (~ 8 s) add the solution of amine under a positive pressure of nitrogen to the stirred solution of trichloromethyl chloroformate at -8°C .
- 4) Remove the cooling bath and allow the reaction to stir for 2 min.
- 5) Dilute the crude reaction mixture to 20 ml with CH_2Cl_2 .
- 6) Wash the organic layer successively with NH_4Cl (1 M aq., 4×20 ml) and NaOH (1 M aq., 20 ml).
- 7) Dry the organic layer with Na_2SO_4 and filter off the salt.
- 8) Remove the solvent *in vacuo*.
- 9) This protocol typically yields a peach-colored solid (66 mg, 29% based on starting amine or 87% maximum of theoretical yield).
- 10) Store the isocyanate desiccated at -20°C in CH_2Cl_2 (0.5 mg per 50 μl). Small quantities of isocyanates hydrolyze slowly when stored concentrated at -20°C ($\sim 30\%$ after four weeks) and rapidly when stored in dimethylformamide (DMF) at -20°C [71]. However, isocyanates can be stored in CH_2Cl_2 solutions as described above for several months after preparation.

The syntheses of the isothiocyanate and isocyanate can be readily performed on a scale ranging from 25 mg to several grams. Preparation of isothiocyanates and isocyanates produces acid (HCl), which combines with the starting amine to produce an unreactive ammonium salt, thereby decreasing the yield of the reaction. This is particularly problematic for isocyanates since it is only possible to convert one-third of the amine to the corresponding isocyanate using this protocol. Alternatively, the non-nucleophilic base Proton Sponge® (1,8-bis((dimethyl)amino)naphthylene, 2.5 equiv.; Sigma–Aldrich) can be included in the reaction, and thereby increase the yield of the reaction. This is especially advantageous if the starting amine is expensive or only available in small quantities. If the light-sensitive Proton Sponge® is used, the reaction should be performed in the dark. After the reaction, Proton Sponge® is removed by extraction using the protocol described above (step 6).

Similar to most chemicals commonly used in any chemistry laboratory, thiophosgene and diphosgene are harmful, but since they are liquids and used in small quantities, they are relatively simple to handle. However, these reactions should be carried out in a well-ventilated area. Likewise, isothiocyanates and isocyanates are toxic chemicals, but they are simple to use and require only standard laboratory safety equipment (e.g., nitrile gloves).

8.3.2

Postsynthetic Labeling of 2'-Amino-Modified RNA

RNA modified with 2'-amino groups can be purchased from several companies. The standard 2'-trifluoroacetyl protecting group is readily cleaved under standard RNA deprotection conditions, and thus no additional deprotection step is necessary. Reaction of 2'-amino-modified RNA with isothiocyanates or isocyanates is typically done under conditions where the RNA is denatured, in aqueous DMF and/or formamide. The organic solvents also act as cosolvents for dissolving the isothiocyanates or isocyanates. It is pertinent that highly pure amine-free anhydrous DMF be used in these reactions because of the high reactivity of succinimidyl esters, isothiocyanates and isocyanates toward amines. Furthermore, we recommend ethanol precipitation of 2'-amino-modified RNAs, effectively converting ammonium salts of RNA from chemical synthesis into sodium salts, before the reaction with these amine-reactive compounds as a precaution against unwanted side reactions.

For labeling of 2'-amino-modified RNA with aromatic isothiocyanates:

- 1) Dissolve the 2'-amino-containing RNA in 5 μ l of 50 mM borate buffer, pH 8.6 (RNA concentration $\sim$ 2 mM, usually performed on 10–40 nt long RNAs).
- 2) Add **2** (100 mM in DMF, 5 μ l).
- 3) Incubate at 37 °C for 28 h (final concentrations: 1 mM 2'-amino RNA, 50 mM isothiocyanate **2**; 50% aqueous DMF, v/v). This reaction proceeds more slowly at room temperature.

Labeling reactions with the aliphatic isocyanates were carried out in a rock salt/ice water bath (-8°C) in a cold room (5°C). If performed at higher temperatures, increased rates of isocyanate hydrolysis result in lower yields. Furthermore,

non-specific labeling has been observed at 37 °C [71]. Analytical scale reactions can be performed using the following procedure, provided reaction amounts are scaled down in such a way that all concentrations of reactants and buffer remain constant.

For preparative scale reactions of 2'-amino-modified RNA with aliphatic isocyanates:

- 1) Dissolve the crude (i.e., not yet gel- or high-performance liquid chromatography (HPLC)-purified), deprotected 2'-amino-containing RNA (one-quarter of a 1 μ mol synthesis) in 100 μ l 70 mM boric acid buffer, pH 8.6.
- 2) Cool the solution in a rock salt/ice water bath (-8°C) in a cold room (5°C).
- 3) Treat the solution sequentially with precooled solutions of formamide (60 μ l, 0°C) and freshly prepared isocyanate in anhydrous DMF (75 mM, 40 μ l, -8°C). Final concentrations: 1 mM 2'-amino RNA, 15 mM isocyanate 4; 50% aqueous borate buffer, 30% formamide, and 20% DMF, v/v/v.
- 4) Incubate for 1 h at -8°C .
- 5) Treat the oligoribonucleotide solution with a second aliquot of freshly prepared isocyanate (40 μ l, 75 mM in DMF).
- 6) Incubate for 1 h at -8°C .
- 7) Wash the solution with CHCl_3 (2×300 μ l) at room temperature.
- 8) Add sodium acetate (3.0 M, 20 μ l, pH 5.3).
- 9) Add absolute ethanol (-20°C , 1.3 ml).
- 10) Precipitate the RNA by storage at -20°C for 4 h.
- 11) Centrifuge the sample ($\sim 13\,000$ g, 15 min, 5°C).
- 12) Remove the supernatant.
- 13) Wash the pellet with cold absolute ethanol (2×50 μ l).
- 14) Dry the pellet *in vacuo*.
- 15) Dissolve the pellet in RNase-free water (50 μ l).
- 16) Dilute with aqueous urea (8 M, 150 μ l).
- 17) Purify the RNA by 20% denaturing polyacrylamide gel electrophoresis (PAGE) (20 cm gel for short oligos up to 20 nt in length, 20 h at 400–600 V or three-quarters the length of the gel; 40 cm gel for longer oligos up to 50 nt in length, up to 72 h at 600 V or less time if higher voltage, e.g., 1200–1800 V, is used).
- 18) Yields typically range from 100 to 170 nmol for a quarter of a 1.0 μ mol synthesis, depending on the length and quality of the RNA synthesis.

To monitor the extent of labeling:

- 1) Remove an aliquot (1.0 μ l) of the reaction mixture from step 15 of the above protocol.
- 2) Dilute with RNase-free water (19 μ l).
- 3) Wash with chloroform (2×75 μ l) to remove excess labeling reagent.
- 4) Analyze the reaction by one of the following three methods:
 - a. Reversed-phase (RP)-HPLC on an analytical column (C18, 4.6×250 mm, 5 μ m column) at 1.5 ml min^{-1} using the following protocol: solvent A, 50 mM Et_3NHOAc (pH 7.0); solvent B, 70% $\text{CH}_3\text{CN}/30\%$ of 50 mM

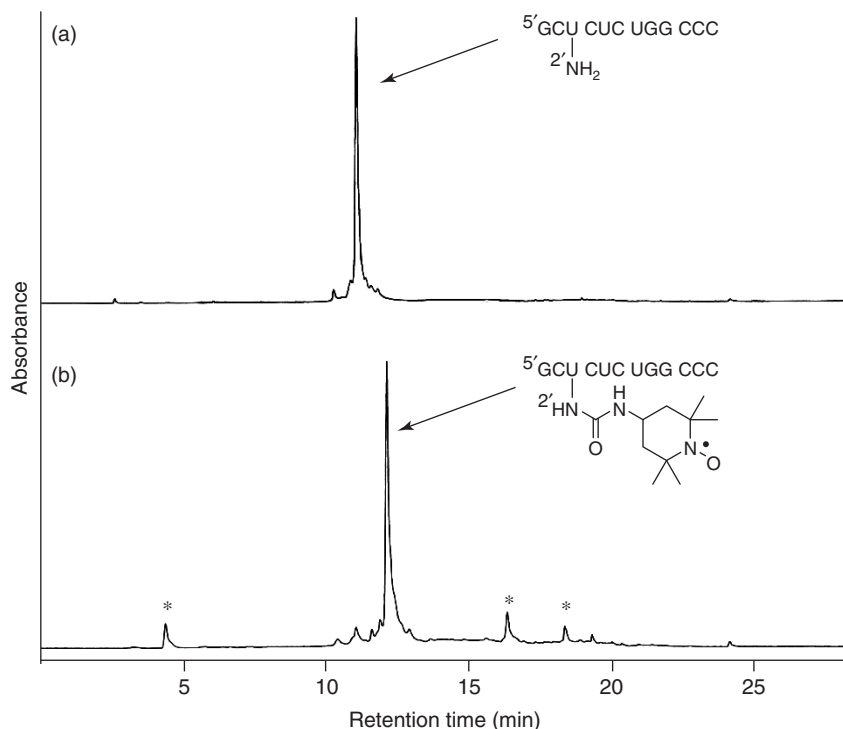


Figure 8.4 HPLC analysis of 5'-GC(2'-NH₂ U) CUC UGG CCC before (a) and after (b) reaction with **4**, which shows 95% conversion to the labeled RNA with increased retention time. The asterisks correspond to **4** and its hydrolysis products. HPLC chromatograms were obtained at 260 nm using an analytical column (C18, 4.6 × 250 mm, 5 μm column) run at 1.5 ml min⁻¹

according to the following protocol: solvent A, 50 mM Et₃NHOAc (pH 7.0) and solvent B, 70% CH₃CN/30% of 50 mM Et₃NHOAc (pH 7.0); 15 min linear gradient from 0 to 23% B, 5 min linear gradient to 100% B, isocratic for 10 min, 3 min linear gradient to initial conditions, and 15 min equilibrium time between runs.

Et₃NHOAc (pH 7.0); 15 min linear gradient from 0 to 23% B, 5 min linear gradient to 100% B, isocratic for 10 min, 3 min linear gradient to initial conditions, and 15 min equilibrium time between runs. A representative example is given in Figure 8.4, which shows the reaction of **4** with 5'-GC(2'-NH₂ U) CUC UGG CCC.

- b. Twenty percent denaturing PAGE (20 cm gel, 400 V for 3.5 h) followed by UV shadow visualization (for details, see Chapter 3).
- c. Analytical ion exchange (IE)-HPLC on a Dionex DNAPac PA100 (4 × 250 mm) analytical column heated at 50 °C by a column warmer. Separation will not be achieved without heating the column. Solvent gradients for analytical IE-HPLC were run at 1.0 ml min⁻¹ as follows: solvent A, 25 mM Tris · HCl, pH 8.0; solvent B, 1.0 M NaCl, 25 mM Tris · HCl, pH 8.0; 35 min linear gradient from 10% B to 80% B, 5 min linear gradient to 10% B.

Short-labeled RNAs (up to 20 nt in length) can also be purified utilizing these HPLC protocols, although we recommend 20% denaturing PAGE purification, since the hydrolysis products of some isocyanates may coelute with the labeled RNA using RP-HPLC.

8.3.3

Postsynthetic Labeling of 4-Thiouridine-Modified RNA

If one knows *a priori* that modification of the 2'-hydroxyl group will likely interfere with biological function (e.g., 2'-OH is involved in an essential hydrogen bond), it may be necessary to label using an alternative postsynthetic labeling strategy. In this case, another simple, straightforward method is the labeling of 4-thiouridine residues with iodoacetamides [53, 54] or sulfur-based compounds [24, 57]. One of the advantages of this method is that the labeling reaction can be followed by monitoring the consumption of UV signal at 320 nm, which corresponds to the thiocarbonyl. This labeling strategy changes the base-pairing properties of this residue. However, UV thermal denaturation data as well as NMR structural data indicate that 4-thiouridine residues can be labeled in this manner without disruption of helical stacking [24]. Labeling of the 4-amine group of cytidine with a crudely analogous modification, however, resulted in severe thermal instability of DNA [95]. Therefore, caution should be exercised when choosing such a labeling strategy for base-pairing residues.

Labeling of 4-thiouridine with the iodoacetamide spin-labeling reagent 3-(2-iodoacetamido)-proxyl 5 (modified procedure based on that reported in [54]):

- 1) Dissolve 4-thiouridine-modified RNA (1/4 of a 1 μ mol synthesis) in 166 μ l of buffer (100 mM sodium phosphate, pH 8).
- 2) Acquire UV spectrum of an aliquot of the above mixture, monitoring at 260 and 320 nm.
- 3) Dissolve 6 mg of 3-(2-iodoacetamido)-proxyl 5 (Figure 8.3c, Sigma-Aldrich) into 14 μ l of ethanol and 20 μ l of anhydrous DMF (0.5 M labeling reagent).
- 4) Mix 5 and 4-thiouridine-modified RNA; final concentrations: $\sim$ 1 mM RNA, 85 mM 5, 83% phosphate buffer/7% ethanol/10% DMF (v/v/v).
- 5) Owing to light sensitivity of 4-thiouridine residues, cover samples with aluminum foil.
- 6) Vortex vigorously until absorbance at 320 nm disappears (typically 18–28 h).
- 7) Once the reaction is complete, precipitate and purify RNA as described above.

8.3.4

Verification of Label Incorporation

Whenever a modification is introduced into RNA, either by solid-phase chemical synthesis using a phosphoramidite or by postsynthetic modification, several steps are necessary to verify that the incorporation was successful. Not all modifications

are incorporated as intended. For example, the 5-trifluoromethyl-2'-deoxyuridine phosphoramidite was prepared for the purpose of ^{19}F NMR spectroscopy of nucleic acids; however, standard oligonucleotide deprotection conditions converted the 5-trifluoromethyl group to a 5-cyano group, prompting the use of alternative mild deprotection conditions [96]. Incorporation of the modified nucleoside should be verified by mass spectrometry as well as enzymatic digestion in conjunction with RP-HPLC analysis. For the latter analysis, the RNAs (2 nmol or ~ 0.2 OD₂₆₀) should be digested with snake venom phosphodiesterase (0.5 U) and calf intestinal alkaline phosphatase (8 U) at 37 °C for 5 h in 5 mM Tris·HCl, pH 7.4 (20 μl) and then analyzed by analytical RP-HPLC using the same protocol as that detailed above for monitoring the extent of labeling. For example, HPLC analysis of the enzymatic digestion of 5'-GCU C(2'-NH₂ U)C UGG CCC resulted in peaks corresponding to C, 2'-NH₂ U, U, and G (Figure 8.5a), whereas after labeling with the spin-label isocyanate **4** HPLC analysis revealed the absence of the 2'-NH₂ U peak and the presence of a new peak (Figure 8.5b) that coeluted with the modified spin-labeled nucleoside prepared by chemical synthesis [19].

8.3.5

Potential Problems and Troubleshooting

It is always important to determine if the modification interferes (intentionally or unintentionally) with the structure and function of the molecule using a standard

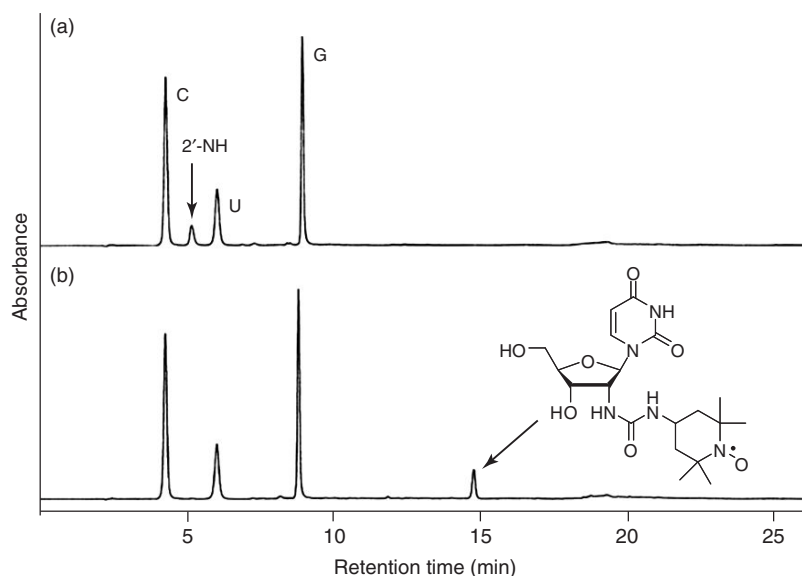


Figure 8.5 HPLC analysis of enzymatically digested RNA. (a) Enzymatic digestion of 5'-GCU C(2'-NH₂ U)C UGG CCC and (b) enzymatic digestion of the product of the 2'-NH₂-modified oligonucleotide from a after reaction with isocyanate **4**. HPLC chromatograms were obtained as in Figure 8.4.

structural (e.g., UV thermal denaturation and/or other biophysical spectroscopy or crystallography) and functional (binding or enzymatic) assay. For example, the effect of incorporation of nitroxide spin labels at the 2' position on RNA has been investigated by UV thermal denaturation [19, 69], whereas their effect on RNA–protein complex formation was investigated by electrophoretic mobility shift analysis [62].

If the labeling reaction does not work or its yields are low, this is generally the result of one of four problems. First, the isocyanate may be hydrolyzed or not prepared properly. The quality of the isocyanate can be determined by spectroscopic methods (e.g., NMR) and/or tested by reaction with a simple aliphatic amine such as benzylamine (30 min in CH_2Cl_2 at room temperature). Second, the 2'-trifluoroacetyl protecting group may not have been fully removed, which may not be readily apparent because 2'-trifluoroacetamido and 2'-amino-modified RNAs often have similar mobility on HPLC or in gels. However, this can be readily investigated by enzymatic digestion of the RNA, followed by HPLC analysis as described above. For example, if the 2'-trifluoroacetyl group is not fully removed, a new peak is observed by HPLC analysis with a retention time of ~ 5 min corresponding to the 2'-trifluoroacetamido uridine nucleoside (e.g., in the order of C, 2'- NH_2 U, U, 2'- NHCOCF_3 U, G, A). Third, if the temperature of the reaction is not low enough, the yields are lower, presumably because of the competing hydrolysis of the isocyanate. Therefore, it is important to monitor the temperature of the ice–salt bath with a thermometer. Fourth, lower yields are obtained if the RNA is not completely dissolved at the beginning of the reaction.

References

1. Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds) (1999) *The RNA World*, Cold Springs Harbor Laboratory Press, Cold Springs Harbor.
2. Menger, M., Eckstein, F., and Porschke, D. (2000) Multiple conformational states of the hammerhead ribozyme, broad time range of relaxation and topology of dynamics. *Nucleic Acids Res.*, **29**, 4428–4434.
3. Tuschl, T., Gohlke, C., Jovin, T.M., Westhof, E., and Eckstein, F. (1994) A three-dimensional model for the hammerhead ribozyme based on fluorescence measurements. *Science*, **266**, 785–788.
4. Bassi, G.S., Murchie, A.I.H., Walter, F., Clegg, R.M., and Lilley, D.M.J. (1997) Ion-induced folding of the hammerhead ribozyme: a fluorescence resonance energy transfer study. *EMBO J.*, **16**, 7481–7489.
5. Lafontaine, D.A., Norman, D.G., and Lilley, D.M.J. (2002) The global structure of the VS ribozyme. *EMBO J.*, **21**, 2461–2471.
6. Sigurdsson, S.T., Tuschl, T., and Eckstein, F. (1995) Probing RNA tertiary structure: interhelical crosslinking of the hammerhead ribozyme. *RNA*, **1**, 575–583.
7. Venkatesan, N., Kim, S.J., and Kim, B.H. (2003) Novel phosphoramidite building blocks in synthesis and applications toward modified oligonucleotides. *Curr. Med. Chem.*, **10**, 1973–1991.
8. Kinoshita, Y., Nishigaki, K., and Husimi, Y. (1997) Fluorescence-isotope- or biotin-labeling of the 5'-end of single-stranded DNA/RNA using T_4 RNA ligase. *Nucleic Acids Res.*, **25**, 3747–3748.
9. Rodgers, J.T., Patel, P., Hennes, J.L., Bolognia, S.L., and Mascotti, D.P. (2000)

- Use of biotin-labeled nucleic acids for protein purification and agarose-based chemiluminescent electromobility shift assays. *Anal. Biochem.*, **277**, 254–259.
10. Harwood, A.J. (2002) End-labeling of DNA probes. *Methods Mol. Biol.*, **187**, 17–22.
 11. McMin, D.L. and Greenberg, M.M. (1998) Postsynthetic conjugation of protected oligonucleotides containing 3'-alkylamines. *J. Am. Chem. Soc.*, **120**, 3289–3294.
 12. Chockalingam, P.S., Jurado, L.A., and Jarrett, H.W. (2001) DNA affinity chromatography. *Mol. Biotechnol.*, **19**, 189–199.
 13. Roberts, R.W. and Szostak, J.W. (1997) RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 12297–12302.
 14. Musier-Forsyth, K. and Schimmel, P. (1994) Acceptor helix interactions in a class II tRNA synthetase: photoaffinity cross-linking of an RNA miniduplex substrate. *Biochemistry*, **33**, 773–779.
 15. Konarska, M.M. (1999) Site-specific derivatization of RNA with photocrosslinkable groups. *Methods*, **18**, 22–28.
 16. Qin, P.Z., Butcher, S.E., Feigon, J., and Hubbell, W.L. (2001) Quantitative analysis of the isolated GAAA tetraloop/receptor interaction in solution: a site-directed spin labeling study. *Biochemistry*, **40**, 6929–6936.
 17. Earnshaw, D.J. and Gait, M.J. (1998) Modified oligoribonucleotides as site-specific probes of RNA structure and function. *Biopolymers (Nucleic Acid Sci.)*, **48**, 39–55.
 18. Verma, S. and Eckstein, F. (1998) Modified oligonucleotides: synthesis and strategy for users. *Annu. Rev. Biochem.*, **67**, 99–134.
 19. Edwards, T.E., Okonogi, T.M., Robinson, B.H., and Sigurdsson, S.T. (2001) Site-specific incorporation of nitroxide spin-labels into internal sites of the TAR RNA; structure-dependent dynamics of RNA by EPR spectroscopy. *J. Am. Chem. Soc.*, **123**, 1527–1528.
 20. Teplova, M., Wilds, C.J., Wawrzak, Z., Tereshko, V., Du, Q., Carrasco, N., Huang, Z., and Egli, M. (2002) Covalent incorporation of selenium into oligonucleotides for X-ray crystal structure determination via MAD: proof of principle. Multiwavelength anomalous dispersion. *Biochemie*, **84**, 849–858.
 21. Huang, L., Chappell, L.L., Iranzo, O., Baker, B.F., and Morrow, J.R. (2000) Oligonucleotide conjugates of Eu(III) tetraazamacrocycles with pendent alcohol and amide groups promote sequence-specific RNA cleavage. *J. Biol. Inorg. Chem.*, **5**, 85–92.
 22. Silverman, S.K. and Cech, T.R. (1999) RNA tertiary folding monitored by fluorescence of covalently attached pyrene. *Biochemistry*, **38**, 14224–14237.
 23. Koshkin, A.A., Fensholdt, J., Pfundheller, H.M., and Lamholt, C. (2001) Simplified and efficient route to 2'-O, 4'-C-methylene-linked bicyclic ribonucleosides (locked nucleic acid). *J. Org. Chem.*, **66**, 8504–8512.
 24. Qin, P.Z., Hideg, K., Feigon, J., and Hubbell, W.L. (2003) Monitoring RNA base structure and dynamics using site-directed spin labeling. *Biochemistry*, **42**, 6772–6783.
 25. Sproat, B.S. (1995) Chemistry and applications of oligonucleotide analogues. *J. Biotech.*, **41**, 221–238.
 26. Eckstein, F. (2002) Developments in RNA chemistry, a personal view. *Biochemie*, **84**, 841–848.
 27. Schiemann, O., Piton, N., Plackmeyer, J., Bode, B.E., Prisner, T.F., and Engels, J.W. (2007) Spin labeling of oligonucleotides with the nitroxide TPA and use of PELDOR, a pulse EPR method, to measure intramolecular distances. *Nat. Protoc.*, **2**, 904–923.
 28. Sowa, G.Z. and Qin, P.Z. (2008) Site-directed spin labeling studies on nucleic acid structure and dynamics. *Prog. Nucleic Acids Res. Mol. Biol.*, **82**, 147–197.
 29. Lenox, H.J., McCoy, C.P., and Sheppard, T.L. (2001) Site-specific generation of deoxyribonolactone lesions in DNA oligonucleotides. *Org. Lett.*, **3**, 2415–2418.
 30. Dey, S. and Sheppard, T.L. (2001) Ketone-DNA: A versatile postsynthetic

- DNA decoration platform. *Org. Lett.*, **3**, 3983–3986.
31. Piton, N., Mu, Y., Stock, G., Prisner, T.F., Schiemann, O., and Engels, J.W. (2007) Base-specific spin-labeling of RNA for structure determination. *Nucleic Acids Res.*, **35**, 3128–3143.
 32. Gannett, P.M., Darian, E., Powell, J., Johnson, E.M. II, Mundoma, C., Greenbaum, N.L., Ramsey, C.M., Dalal, N.S., and Budil, D.E. (2002) Probing triplex formation by EPR spectroscopy using a newly synthesized spin label for oligonucleotides. *Nucleic Acids Res.*, **30**, 5328–5337.
 33. Summers, J.S. and Shaw, B.R. (2001) Boranophosphates as mimics of natural phosphodiester in DNA. *Curr. Med. Chem.*, **8**, 1147–1155.
 34. Holloway, G.A., Pavot, C., Scaringe, S.A., Lu, Y., and Rauchfuss, T.B. (2002) An organometallic route to oligonucleotides containing phosphoroselenoate. *ChemBioChem*, **3**, 1061–1065.
 35. Chen, B., Jamieson, E.R., and Tullius, T.D. (2002) A general synthesis of specifically deuterated nucleotides for studies of DNA and RNA. *Bioorg. Med. Chem. Lett.*, **12**, 3093–3096.
 36. Douglas, M.E., Beijer, B., and Sproat, B.S. (1994) An approach towards thiol mediated labelling in the minor groove of oligonucleotides. *Bioorg. Med. Chem. Lett.*, **4**, 995–1000.
 37. Hwang, J.T. and Greenberg, M.M. (2001) Synthesis of 2'-modified oligodeoxynucleotides via on-column conjugation. *J. Org. Chem.*, **66**, 363–369.
 38. Beigelman, L., Karpeisky, A., and Usman, N. (1995) Synthesis of 5'-C-methyl-D-allo- and L-talo-ribonucleoside 3'-O-phosphoramidites and their incorporation into hammerhead ribozymes. *Nucleosides Nucleotides*, **14**, 901–905.
 39. Murray, J.B., Terwey, D.P., Maloney, L., Karpeisky, A., Usman, N., Beigelman, L., and Scott, W.G. (1998) The structural basis of hammerhead ribozyme self-cleavage. *Cell*, **92**, 665–673.
 40. Massey, A.P. and Sigurdsson, S.T. (2004) Chemical syntheses of inhibitory substrates of the RNA-RNA ligation reaction catalyzed by the hairpin ribozyme. *Nucleic Acids Res.*, **32**, 2017–2022.
 41. Földesi, A., Trifonova, A., Dinya, Z., and Chattopadhyaya, J. (2001) Total synthesis of 2',3',4',5',5''-(2)h(5)-ribonucleosides: the key building blocks for NMR structure elucidation of large RNA. *J. Org. Chem.*, **66**, 6560–6570.
 42. Yamana, K., Iwase, R., Furutani, S., Tschida, H., Zako, H., Yamaoka, T., and Murakami, A. (1999) 2'-Pyrene modified oligonucleotide provides a highly sensitive fluorescent probe of RNA. *Nucleic Acids Res.*, **27**, 2387–2392.
 43. Korshun, V.A., Stetsenko, D.A., and Gait, M.J. (2002) Novel uridin-2'-yl carbamates: synthesis, incorporation into oligodeoxyribonucleotides, and remarkable fluorescence properties of 2'-pyren-1-ylmethylcarbamate. *J. Chem. Soc., Perkin Trans. 1*, **8**, 1092–1104.
 44. Dioubankova, N.N., Malakhov, A.D., Stetsenko, D.A., Gait, M.J., Volynsky, P.E., Efremov, R.G., and Korshun, V.A. (2003) Pyrenemethyl ara-uridine-2'-carbamate: a strong interstrand excimer in the major groove of a DNA duplex. *ChemBioChem*, **4**, 841–847.
 45. Yamana, K., Mitsui, T., and Nakano, H. (1999) Synthesis and properties of oligonucleotide duplexes containing donor and acceptor fluorophores at 2'-positions. *Tetrahedron*, **55**, 9143–9150.
 46. Rupert, P.B., Massey, A.P., Sigurdsson, S.T., and Ferré-D'Amaré, A.R. (2002) Transition state stabilization by a catalytic RNA. *Science*, **298**, 1421–1424.
 47. Torelli, A.T., Krucinska, J., and Wedekind, J.E. (2007) A comparison of vanadate to a 2'-5' linkage at the active site of a small ribozyme suggests a role for water in transition-state stabilization. *RNA*, **13**, 1052–1070.
 48. Pitt, J.N. and Ferré-D'Amare, A.R. (2009) Structure-Guided Engineering of the Regioselectivity of RNA Ligase Ribozymes. *J. Am. Chem. Soc.*, **131**, 3532–3540.
 49. Khan, S.I. and Grinstaff, M.W. (1999) Structure-guided engineering of the

- regioselectivity of RNA ligase ribozymes. *J. Am. Chem. Soc.*, **121**, 4704–4705.
50. Fidanza, J.A. and McLaughlin, L.W. (1992) Use of a thiol tether for the site-specific attachment of reporter groups to DNA. *J. Org. Chem.*, **57**, 2340–2346.
 51. Grant, G.P., Popova, A., and Qin, P.Z. (2008) Diastereomer characterizations of nitroxide-labeled nucleic acids. *Biochem. Biophys. Res. Commun.*, **371**, 451–455.
 52. Okamoto, A., Taiji, T., Tainaka, K., and Saito, I. (2002) Oligonucleotides containing 7-vinyl-7-deazaguanine as a facile strategy for expanding the functional diversity of DNA. *Bioorg. Med. Chem. Lett.*, **12**, 1895–1896.
 53. Hara, H., Horiuchi, T., Saneyoshi, M., and Nishimura, S. (1970) 4-Thiouridine-specific spin-labeling of *E. coli* transfer RNA. *Biochem. Biophys. Res. Commun.*, **38**, 305–311.
 54. Ramos, A. and Varani, G. (1998) A new method to detect long-range Protein-RNA contacts: NMR detection of electron-proton relaxation induced by nitroxide spin-labelled RNA. *J. Am. Chem. Soc.*, **120**, 10992–10993.
 55. Allerson, C.R., Chen, S.L., and Verdine, G.L. (1997) A chemical method for site-specific modification of rna: the convertible nucleoside approach. *J. Am. Chem. Soc.*, **119**, 7423–7433.
 56. Bellon, L., Workman, C., Scherrer, J., Usman, N., and Wincott, F. (1996) Morpholino-linked ribozymes: a convergent synthetic approach. *J. Am. Chem. Soc.*, **118**, 3771–3772.
 57. Grant, G.P.G., Boyd, N., Herschlag, D., and Qin, P.Z. (2009) Motions of the substrate recognition duplex in a group I intron assessed by site-directed spin labeling. *J. Am. Chem. Soc.*, **131**, 3136–3137.
 58. Cohen, S.B. and Cech, T.R. (1997) Dynamics of thermal motions within a large catalytic RNA investigated by cross-linking with thiol-disulfide interchange. *J. Am. Chem. Soc.*, **119**, 6259–6268.
 59. Earnshaw, D.J., Masquida, B., Muller, S., Sigurdsson, S.T., Eckstein, F., Westhof, E., and Gait, M.J. (1997) Inter-domain cross-linking and molecular modelling of the hairpin ribozyme. *J. Mol. Biol.*, **274**, 197–212.
 60. Aurup, H., Tuschl, T., Benseler, F., Ludwig, J., and Eckstein, F. (1994) Oligonucleotide duplexes containing 2'-amino-2'-deoxycytidines: thermal stability and chemical reactivity. *Nucleic Acids Res.*, **22**, 20–24.
 61. Edwards, T.E. and Sigurdsson, S.T. (2007) Site-specific incorporation of nitroxide spin-labels into 2'-positions of nucleic acids. *Nat. Protoc.*, **2**, 1954–1962.
 62. Edwards, T.E., Okonogi, T.M., and Sigurdsson, S.T. (2002) Investigation of RNA-protein and RNA-metal ion interactions by electron paramagnetic resonance spectroscopy. The HIV TAR-Tat motif. *Chem. Biol.*, **9**, 699–706.
 63. Edwards, T.E. and Sigurdsson, S.T. (2002) Electron paramagnetic resonance dynamic signatures of TAR RNA-small molecule complexes provide insight into RNA structure and recognition. *Biochemistry*, **41**, 14843–14847.
 64. Edwards, T.E. and Sigurdsson, S.T. (2003) EPR spectroscopic analysis of TAR RNA-metal ion interactions. *Biochem. Biophys. Res. Commun.*, **303**, 721–725.
 65. Edwards, T.E., Robinson, B.H., and Sigurdsson, S.T. (2005) Identification of amino acids that promote specific and rigid TAR RNA-Tat protein complex formation. *Chem. Biol.*, **12**, 329–337.
 66. Edwards, T.E. and Sigurdsson, S.T. (2005) EPR Spectroscopic Analysis of U7 Hammerhead Ribozyme Dynamics during Metal Ion Induced Folding. *Biochemistry*, **44**, 12870–12878.
 67. Kim, N.K., Murali, A., and DeRose, V.J. (2005) Separate metal requirements for loop interactions and catalysis in the extended hammerhead ribozyme. *J. Am. Chem. Soc.*, **127**, 14134–14135.
 68. Schiemann, O., Weber, A., Edwards, T.E., Prisner, T.F., and Sigurdsson, S.T. (2003) Nanometer Distance Measurements on RNA Using PELDOR. *J. Am. Chem. Soc.*, **125**, 3434–3435.
 69. Kim, N.K., Murali, A., and DeRose, V.J. (2004) A Distance Ruler for RNA Using EPR and Site-Directed Spin Labeling. *Chem. Biol.*, **11**, 939–948.

70. Sigurdsson, S.T. (1999) Thiol-containing RNA for the study of structure and function of ribozymes. *Methods*, **18**, 71–77.
71. Sigurdsson, S.T. and Eckstein, F. (1996) Site specific labelling of sugar residues in oligoribonucleotides: reactions of aliphatic isocyanates with 2' amino groups. *Nucleic Acids Res.*, **24**, 3129–3133.
72. Hwang, J.T. and Greenberg, M.M. (1999) Synthesis of modified oligodeoxyribonucleotides on a solid-phase support via derivatization of a selectively revealed 2'-Amino-2'-Deoxyuridine. *Org. Lett.*, **1**, 2021–2024.
73. Meade, T.J. and Kayyem, J.F. (1995) Electron transfer through DNA: site-specific modification of duplex DNA with ruthenium donors and acceptors. *Angew. Chem. Int. Ed.*, **34**, 352–353.
74. Krider, E.S., Miller, J.E., and Meade, T.J. (2002) 2'-modified nucleosides for site-specific labeling of oligonucleotides. *Bioconjug. Chem.*, **13**, 155–162.
75. Merino, E.J. and Weeks, K.M. (2003) Fluorogenic resolution of ligand binding by a nucleic acid aptamer. *J. Am. Chem. Soc.*, **125**, 12370–12371.
76. Zatsepin, T.S., Stetsenko, D.A., Arzumanov, A., Romanova, E., Gait, M.J., and Oretskaya, T.S. (2002) Synthesis of peptide-oligonucleotide conjugates with single and multiple peptides attached to 2'-aldehydes through thiazolidine, oxime, and hydrazine linkages. *Bioconjug. Chem.*, **13**, 822–830.
77. Blount, K.F. and Uhlenbeck, O.C. (2002) Internal equilibrium of the hammerhead ribozyme is altered by the length of certain covalent cross-links. *Biochemistry*, **41**, 6834–6841.
78. Stage-Zimmerman, T.K. and Uhlenbeck, O.C. (2001) A covalent crosslink converts the hammerhead ribozyme from a ribonuclease to an RNA ligase. *Nat. Struct. Biol.*, **8**, 863–867.
79. Buchmueller, K.L., Hill, B.T., Platz, M.S., and Weeks, K.M. (2003) RNA-tethered phenyl azide photocrosslinking via a short-lived indiscriminant electrophile. *J. Am. Chem. Soc.*, **125**, 10850–10861.
80. John, D.M. and Weeks, K.M. (2000) Tagging DNA mismatches by selective 2'-amine acylation. *Chem. Biol.*, **7**, 405–410.
81. John, D.M. and Weeks, K.M. (2002) Chemical interrogation of mismatches in DNA-DNA and DNA-RNA duplexes under non-stringent conditions by selective 2'-amine acylation. *Biochemistry*, **41**, 6866–6874.
82. Silverman, S.K., Deras, M.L., Woodson, S.A., Scaringe, S.A., and Cech, T.R. (2000) Multiple folding pathways for the P4-P6 RNA domain. *Biochemistry*, **39**, 12465–12475.
83. Silverman, S.K. and Cech, T.R. (2001) An early transition state for folding of the P4-P6 RNA domain. *RNA*, **7**, 161–166.
84. Blount, K.F. and Tor, Y. (2003) Using pyrene-labeled HIV-1 TAR to measure RNA-small molecule binding. *Nucleic Acids Res.*, **31**, 5490–5500.
85. Blount, K.F., Grover, N.L., Mokler, V., Beigelman, L., and Uhlenbeck, O.C. (2002) Steric interference modification of the hammerhead ribozyme. *Chem. Biol.*, **9**, 1009–1016.
86. Chamberlin, S.I., Merino, E.J., and Weeks, K.M. (2002) Catalysis of amide synthesis by RNA phosphodiester and hydroxyl groups. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 14688–14693.
87. Hendrix, C., Devreese, B., Rozenski, J., van Aerschot, A., De Bruyn, A., Van Beeumen, J., and Herdewijn, P. (1995) Incorporation of 2'-amido-nucleosides in oligodeoxynucleotides and oligoribonucleotides as a model for 2'-linked conjugates. *Nucleic Acids Res.*, **23**, 51–57.
88. Hendrix, C., Mahieu, M., Anne, J., Calenbergh, S.V., Van Aerschot, A., Content, J., and Herdewijn, P. (1995) Catalytic activity and stability of hammerhead ribozymes containing 2'-acetamido-2'-deoxyribonucleosides. *Biochem. Biophys. Res. Commun.*, **210**, 67–73.
89. Kryatova, O.P., Connors, W.H., Bleczinski, C.F., Mokhir, A.A., and Richert, C. (2001) A 2'-acylamido cap that increases the stability of

- oligonucleotide duplexes. *Org. Lett.*, **3**, 987–990.
90. Sergiev, P., Dokudovskaya, S., Romanova, E., Topin, A., Bogdanov, A., Brimacombe, R., and Dontsova, O. (1998) The environment of 5S rRNA in the ribosome: cross-links to the GTPase-associated area of 23S rRNA. *Nucleic Acids Res.*, **26**, 2519–2525.
 91. Sigurdsson, S.T. and Eckstein, F. (1995) Structure-function relationships of hammerhead ribozymes: from understanding to applications. *Trends Biotech.*, **13**, 286–289.
 92. Alefelder, S. and Sigurdsson, S.T. (2000) Interstrand disulfide cross-linking of internal sugar residues in duplex RNA. *Bioorg. Med. Chem.*, **8**, 269–273.
 93. Bramlage, B., Alefelder, S., Marschall, P., and Eckstein, F. (1999) Inhibition of luciferase expression by synthetic hammerhead ribozymes and their cellular uptake. *Nucleic Acids Res.*, **27**, 3159–3167.
 94. Sigurdsson, S.T., Seeger, B., Kutzke, U., and Eckstein, F. (1996) A mild and simple method for the preparation of isocyanates from aliphatic amines using trichloromethyl chloroformate. Synthesis of an isocyanate containing an activated disulfide. *J. Org. Chem.*, **61**, 3883–3884.
 95. Bannwarth, W. and Schmidt, D. (1994) Oligonucleotides containing spin-labeled 2'-deoxycytidine and 5-methyl-2'-deoxycytidine as probes for structural motifs of DNA. *Bioorg. Med. Chem. Lett.*, **4**, 977–980.
 96. Markley, J.C., Chirakul, P., Sologub, D., and Sigurdsson, S.T. (2001) Incorporation of 2'-deoxy-5-(trifluoromethyl)uridine and 5-cyano-2'-deoxyuridine into DNA. *Bioorg. Med. Chem. Lett.*, **11**, 2453–2455.

Part II

Structure Determination

9

Direct Determination of RNA Sequence and Modification by Radiolabeling Methods

Olaf Gimple and Astrid Schön

9.1

Introduction

The large number of genome sequences now available has allowed the identification of many novel RNA species, simply by deduction from the published DNA sequences [1]. Nevertheless, direct sequencing of RNA molecules is still indispensable in many cases. The most important reason is that even nowadays, novel RNA species may only be discovered following a functional assay. If no hint to the sequence can be obtained by genomic data mining, if RNA editing may occur in this organism, or simply, if the RNA is derived from an organism where no genomic sequences are available, then the RNA has to be purified before direct sequence determination. The second, and probably even more intriguing rationale is the observation that a large number of RNAs, such as rRNAs, snRNAs, snoRNAs, and specifically tRNAs, contain modified nucleobases which often play crucial roles in the function of these RNAs [2–9]. Although in many instances, plain RNA sequences can be obtained by novel, massively parallel DNA or cDNA sequencing methods [10], the type and position of modified bases has to be determined by direct analysis of the purified RNA. In this chapter, we will describe methods for the rigorous purification of single small RNA species from the bulk of cellular RNA, their sequence determination, and the identification of modified nucleotides by *in vitro* labeling methods.

9.2

General Methods

It is anticipated that the reader is familiar with standard biochemical and molecular biology procedures, such as gel electrophoresis, column- and thin-layer chromatography (TLC), and handling of radioactive materials. In order to avoid RNase contamination, all aqueous solutions and plasticware should be sterilized by autoclaving, or prepared from RNase-free (diethylpyrocarbonate-treated) water. Glassware should be baked at 150 °C for 4 h. Centrifugations are performed in a microcentrifuge at 10 000 g, if not stated otherwise.

9.3

Isolation of Pure RNA Species from Biological Material

9.3.1

Preparation of Size-Fractionated RNA

Numerous methods for the preparation of crude RNA from cells and tissues can be found in handbooks describing standard molecular biology techniques, for example, [11]. All these methods consist of a cell disruption step under denaturing conditions, followed by separation of the nucleic acids from protein and cell debris by extraction with an organic solvent and precipitation with alcohol. For purification of a single RNA species from this total RNA population, it is preferable to perform a size selection before further manipulations. The simplest procedure consists of fractionated precipitation with NaCl, where large RNAs and polysaccharides are separated from the “soluble” (i.e., small) RNAs by subsequent centrifugation [12]. A more elaborate scheme for the large scale purification of tRNAs (and other small RNAs) from human and animal tissue has been described by Roe [13]. This procedure can be easily scaled down and adapted to other tissues. The DEAE anion exchange chromatography described in that paper may be conveniently replaced with ready-to-use columns for small scale nucleic acid preparations, following the manufacturer’s instructions for RNA preparation (e.g., Macherey-Nagel Nucleobond AX). Alternatively, ribosome-free cell extracts (S100) or partially purified enzyme preparations can be used for the isolation of specific RNA populations [14, 15]. The RNAs obtained by any of these procedures are ready for further purification and functional assays.

9.3.2

Isolation of a Single Unknown RNA Species Following a Functional Assay

If the RNA of interest can be traced by a functional assay such as aminoacylation, ribozyme activity or similar, it can be purified and identified, regardless of prior sequence information. Although chromatographic procedures such as anion exchange, gel filtration, or reverse phase chromatography can be used in any combination [2, 3], isolation of a single species usually requires preparative separation by one- or two-dimensional gel electrophoresis as a final step. To achieve the best possible separation based on differences in RNA length, base composition, nucleotide modification, and structure, the first dimension gel should be run under semi-denaturing and slightly acidic conditions, and the second dimension under fully denaturing and slightly alkaline conditions, as described in [16, 17] and below.

9.3.2.1 Solutions for Electrophoresis, Staining, and Elution of RNAs from Gels

- Urea gel loading buffer: 8 M urea, 0.03% (w/v) bromophenol blue (BPB), 0.03% (w/v) xylene cyanol FF (XC), 0.06% (w/v) Sigma brilliant blue (SBB) in TE buffer (10 mM Tris-Cl pH 7.5, 1 mM Na₂EDTA). It is most convenient to prepare this

solution from dye stock solutions (see Section 9.5.1.1) and 9 M urea prepared in RNase-free water. For semi-denaturing gels, the urea concentration is reduced accordingly.

- First dimension gel: 10% Polyacrylamide (PAA; 19 : 1, i.e., acrylamide : N,N'-methylene bisacrylamide), 25 mM Na-Citrate pH 3.5; 3.5 M urea; electrophoresis buffer: 25 mM Na-Citrate pH 3.5, 4 mM Na₂EDTA.
- Second dimension gel: 15% PAA (29 : 1), 8 M urea; TBD buffer for gel and electrophoresis: 90 mM Tris, 90 mM H₃BO₄, 2.5 mM EDTA. This buffer is most conveniently made from a 10-fold concentrated stock solution and will have a pH of 8.3 without further adjustment.
- Staining solution: 0.4% Toluidine blue O (w/v) in 50% MeOH (v/v), 10% glacial acetic acid (v/v).
- Destaining solution: 50% MeOH (v/v), 10% glacial acetic acid (v/v).
- Elution buffer: 0.5 M Tris-Cl pH 7, 0.1% SDS (w/v), 0.1 mM Na₂EDTA, 1 mM MgCl₂.

9.3.2.2 Two-Dimensional Electrophoresis of RNA

First and second dimension gels should be 40 cm long, 20 cm wide, and 0.5–0.7 mm thick. The first dimension is run under semi-denaturing conditions, that is, at a power of not more than 20 W. To obtain a homogeneous temperature distribution during electrophoresis, a thermostatic apparatus is useful. After electrophoresis in the first dimension, a 16 cm long region of the lane is cut out. The upper and lower excision sites of this strip depend on the anticipated migration distance of the desired RNA. This may be estimated from the migration of marker dyes: BPB comigrates with approximately 16, XC with 60, and SBB with about 100 nt in 10% PAA gels. The gel piece is then polymerized into the fully denaturing second dimension gel such that the direction of electrophoresis is turned by 90° compared to the first dimension. This is easily accomplished by positioning the gel strip at the lower end of a gel plate directly above the lower sealing, assembling and pouring the gel as usual. The glass plate sandwich with the polymerized gel is then fixed in the electrophoresis apparatus such that the strip region from the first dimension is at the top (cathode). The second dimension is run at 40 W to obtain a higher gel temperature; again, a thermostatic apparatus will improve the results. After electrophoresis, the RNAs are visualized by Toluidine Blue staining for 20 min and destaining until a clear background is achieved. The isolated spots are cut out with a scalpel, and 150 µl elution buffer is added to each reaction tube containing a gel piece of 4 mm maximum diameter (if pieces are larger, adjust buffer volume accordingly). After freezing the tubes to disrupt the gel structure, the RNA is eluted by vigorous shaking at room temperature for at least 8 h. The buffer is collected, another 150 µl is added to each tube to wash the gel piece, and the RNA is precipitated from the combined buffer fractions by addition of 750 µl EtOH, overnight precipitation at –20 °C, and centrifugation for 20 min at 10 000 g. After washing with ice-cold 70% ethanol and vacuum drying for 10 min, the RNAs can be dissolved in H₂O or the desired buffer for functional analysis or sequencing.

9.3.2.3 Comments on the Electrophoretic Purification and Elution of RNA Species

One parameter influencing gel resolution is the amount of RNA loaded. To obtain optimal resolution, not more than 50 μg of a prefractionated “small RNA” preparation should be loaded per 1 cm of lane width of a first dimension gel (0.5 mm thick, 20 cm wide, and 40 cm long). For nearly homogeneous preparations obtained as described in Section 9.3.3, 5–10 μg may be loaded onto each 1 cm wide lane of a preparative gel; as little as 0.1 μg (Toluidine staining) or 1 ng (silver staining) RNA can be detected on analytical scale.

The second parameter important for good separation is the polymer concentration, which should be adjusted to the expected size of the RNA of interest; for two-dimensional separation, the first gel should be of lower concentration than the second. For example, 10 and 15% PAA should be used for RNA sizes between 75 and 90 nt, and electrophoresis should proceed until XC has reached the bottom of the gel. For smaller molecules, up to 25% PAA (37.5 : 1) may be used. Note that the efficiency of gel elution is dependent on the gel concentration: while a 350 nt long RNA is easily recovered from an 8 or 10% PAA gel, yields are lower from a 15% gel. If very small amounts of RNA have to be recovered from large volumes of elution buffer, 1–10 μg glycogen may be added as a coprecipitant if it does not interfere with subsequent analysis.

An additional important point is to keep the RNA fully denatured and intact during electrophoresis. We use TBD instead of the standard TBE buffer for preparative and analytical (sequencing) gels. Empirical analysis showed that the concentration of 2.5 mM EDTA in the TBD buffer is optimal for the reduction of secondary structure effects during electrophoresis.

9.3.3

Isolation of Single RNA Species with Partially Known Sequence

If part or all of the primary sequence of an RNA species is known, for example from genomic analysis or direct sequencing, a hybrid selection method can be applied to obtain the desired species in sufficient quantities for further studies. This protocol has been optimized following published procedures [18, 19].

9.3.3.1 Materials for Hybrid Selection of Single RNA Species

- 1) **Buffers:** 20 $\times$ SSC, 6 $\times$ SSC, 1 $\times$ SSC; TE: 10 mM Tris-Cl pH 7.5, 1 mM Na_2EDTA ; 2 M NaOAc pH 5.
- 2) **Affinity matrix:** Ultra-Link Streptavidin Plus-Beads (Pierce, No. 53117; capacity 66.7 pmol Biotin/ μl according to the manufacturer) are used in this protocol; equivalent products from other manufacturers may give similar results.
- 3) **Nucleic acids:**
 - a. **Bait oligonucleotide:** 3'-biotinylated deoxyoligonucleotide, complementary to the desired RNA. Most manufacturers offer the biotin-coupled oligonucleotides at excellent quality. Note that the calculated T_m of oligonucleotide-target RNA complexes should be about 70 $^\circ\text{C}$ under standard conditions

(1 M NaCl at 25 °C at pH 7). If possible, choose a single-stranded variable RNA region as target. The working solution of the oligonucleotide is adjusted to 1 nmol/μl H₂O.

- b. **Prey (target) RNA:** crude RNA preparation (see Section 9.3.1) in TE or H₂O. The best efficiency in hybrid selection is achieved with size-fractionated RNA preparations.
- 4) **Equipment:** 2–3 thermostated shakers holding 1.5 ml reaction tubes. Speed should be adjusted such that the beads are just kept from sedimenting, but do not move vigorously.

Protocol 1: Procedure for the Purification of a Single RNA Species from 1 mg Crude Small RNAs

Protocol 1.1: Coupling of Biotinylated Oligonucleotide to Streptavidin Beads

- 1) The Streptavidin beads are supplied as a suspension and tend to sediment fast. Before removing the required amount, shake the suspension until homogeneous, and use a pipet tip with enlarged opening (cut apical end with a scalpel) to avoid clogging. Take 15 μl of bead suspension per 1 mg prefractionated RNA.
- 2) Wash beads twice in 200 μl TE and resuspend in 200 μl TE (“Washing” of beads means suspend *thoroughly but carefully* by gentle vortexing, collect by centrifugation at 2000 r.p.m. for 5 min in a microcentrifuge, remove supernatant).
- 3) Add 1 μl Biotin-oligonucleotide solution; shake for 15 min at room temperature and wash twice in 6 × SSC.

Protocol 1.2: Hybridization of RNA to Coupled Oligonucleotides

- 1) Preheat one shaker at 65 °C and a second one at 90 °C.
- 2) Adjust RNA volume with TE to 100 μl (final concentration: 10 mg/ml).
- 3) Denature by heating for 2 min at 90 °C and snap-cool in ice water.
- 4) Adjust RNA to 6 × SSC by adding 43 μl of 20 × SSC and add to beads.
- 5) Shake for 10 min at 65 °C, turn off heater and continue shaking while the block cools down to room temperature.
- 6) Immediately after the block has reached room temperature, collect beads by centrifugation at 2000 r.p.m. for 5 min and wash three times at room temperature with 6 × SSC followed by three times with 1 × SSC. Save the beads for elution and keep the supernatants in case the hybridization needs further optimization.

Protocol 1.3: Elution and Electrophoresis of Affinity-Purified RNA

Preheat two shakers at 60 °C and 75 °C, respectively.

Elution of the desired RNA from the beads is achieved in three steps; retain the supernatant of each step for further analysis.

- 1) First, the beads are resuspended in 100 μ l TE and shaken at room temperature for 5 min.
- 2) After centrifugation as above (Protocol 1.2, step 6), resuspend the beads in 100 μ l TE prewarmed at 60 °C, and incubate at 60 °C for 5 min.
- 3) Repeat step 2, but use 75 °C for prewarming and incubation.
- 4) If no RNA is recovered, elution may be repeated at 90 °C.
- 5) Precipitate the RNA from the eluted fractions by addition of 0.1 volume of 2 M NaOAc (pH 5) and 2.5 volumes of EtOH, followed by storage at –20 °C for at least 30 min and centrifugation.
- 6) Wash the precipitate with 100 μ l ice-cold 70% EtOH, vacuum-dry, and dissolve in 10–20 μ l urea gel loading buffer (see Section 9.3.2.1).
- 7) After denaturation for 2 min at 95 °C, load samples onto a denaturing PAA gel for electrophoretic separation; gel composition and running conditions are as described for the second dimension (Sections 9.3.2.1 and 9.3.2.2).
- 8) After electrophoresis, stain the gel with Toluidine Blue, and excise and elute the RNA band(s) of interest. This additional electrophoretic step may be required to ensure absolute purity of the desired RNA because in some cases, co-purification of RNA species with similar sequence cannot be totally avoided. This step may be omitted if unnecessary. In such cases, determine the purity of bead eluates collected in steps 1–4 on an analytical scale. For this purpose, mix 2–5 μ l of the bead eluates with 4 volumes of loading buffer, electrophorese as in step 7 above, and detect the RNA by silver staining [20].

9.4

Radioactive Labeling of RNA Termini

End labeling of RNA with ^{32}P is a prerequisite for direct sequence analysis, and free 5'- or 3'-OH-groups are required for most labeling reactions. The removal of 5'-cap structures using tobacco pyrophosphatase has been described [21]. 5'- and 3'- phosphate residues can be easily removed by calf intestine alkaline phosphatase (CIP) before the labeling reaction. Be aware that not all end-modifying reactions work equally well with all types of RNA. For example, the "hidden" 5'-end of tRNAs is often hard to dephosphorylate; thus, labeling by phosphate exchange (see Protocol 2.2) is the preferred method in these cases.

9.4.1

Materials for 5'-End Labeling of RNAs

- **Enzymes:** Calf intestine alkaline phosphatase (CIP; 10 mU/ μ l) and T4 polynucleotide kinase (PNK; 5 U/ μ l) are available from Life Technologies, Roche Applied Science, New England Biolabs, or Promega.
- **Buffers and reagents:** 50 mM nitrilo-tri-acetic acid (NTA), pH 8; IMID-mix: 250 mM imidazole, 25 mM DTT, 0.5 mM spermine, 0.5 mM Na_2EDTA , 50 mM MgCl_2 , pH 6.6 with HCl; MgCl_2 /spermine solution: 0.2 M MgCl_2 , 32 mM spermine; 100 mM DTT.

- **Radioactive materials:** γ - ^{32}P -ATP (10 $\mu\text{Ci}/\mu\text{l}$; 3000 Ci/mmol).

Protocol 2.1: Dephosphorylation and 5'-Labeling of RNA

- 1) For dephosphorylation, vacuum-dry 10–100 pmol of purified RNA, dissolve in 8 μl H_2O , denature for 2 min at 90°C and snap-cool in ice water.
- 2) Add 1 μl 1 M Tris-Cl pH 8 and 10 mU CIP (1 μl) and incubate at 50°C for 1 h.
- 3) Inactivate the enzyme by addition of 3.3 μl NTA solution followed by incubation at 50°C for 20 min.
- 4) For radioactive labeling, prepare one 1.5 ml reaction tube containing 200 μCi γ - ^{32}P -ATP and one with 10 μl of urea gel loading buffer (Section 9.3.2.1). Dry both solutions in a desiccator.
- 5) Dissolve the dried γ - ^{32}P -ATP by adding 6.5 μl of the RNA solution from step 3.
- 6) Add 1 μl of the MgCl_2 /spermine solution, 1 μl of 100 mM DTT, and 1 μl PNK, and incubate at 37°C for 30 min.
- 7) Terminate the reaction by transferring the whole mixture into the tube containing the dry loading buffer.
- 8) After denaturation (2 min, 90°C), separate the labeled RNA on a denaturing PAA gel, localize it by autoradiography or phosphorimaging, and recover the RNA by elution (Section 9.3.2.2). To increase recovery of the labeled RNA during precipitation, 10 μg yeast tRNA per 300 μl elution buffer may be added as a coprecipitant.

Protocol 2.2: 5'-Labeling by Phosphate Exchange

Many small RNAs are highly structured and have a recessed 5'-end difficult to access by the phosphatase. Thus, the exchange reaction described in [22] is the labeling method of choice for these molecules.

- 1) Dissolve 10–100 pmol RNA in 2 μl IMID-mix, then add 1.25 μl 0.5 mM ADP and 5.75 μl H_2O .
- 2) Transfer mixture into a reaction tube containing 200 μCi dry γ - ^{32}P -ATP. Initiate the reaction by addition of 1 μl PNK (5 U), and incubate for 30 min at 37°C .
- 3) Terminate reaction by pipetting the mixture onto dried loading buffer, and separate by electrophoresis as described above.

9.4.2

3'-Labeling of RNAs

Most natural RNAs possess a 3'-OH end and can thus be directly labeled at this terminus. The most popular method is the ligation of radioactive pCp to the RNA [23]. This procedure has been described in detail [16, 24]; an abbreviated and efficient variation including the synthesis of pCp is presented below.

3'-labeling of RNAs is usually more efficient than 5'-labeling and requires less material. However, larger RNAs are often poor substrates for the ligation reaction and may be preferably labeled by poly(A) polymerase using α - ^{32}P -labeled 3'-deoxyadenosine (Cordycepin) to prevent chain elongation [25]. As an exceptional case owing to their 3'-terminal sequence, tRNAs can be specifically labeled with α - ^{32}P -ATP at their CCA end by tRNA-nucleotidyltransferase [4, 26], eliminating the problem of sequence heterogeneity. A special method to label RNAs with (at least partially) known sequence is "*splint labeling*" of the 3'-end with DNA polymerase [27].

9.4.2.1 Materials for 3'-End Labeling of RNAs

- **Enzymes:** T4 polynucleotide kinase (PNK) and T4 RNA ligase (both at 5 U/ μl) are available from Life Technologies, Roche Applied Science, New England Biolabs, or Promega. Yeast Poly(A) polymerase (about 500 U/ μl) and T7 DNA polymerase (10–25 U/ μl) are available from USB (Affymetrix).
- **Buffers and reagents:** 1 mM 3'-cytidine monophosphate (3'-Cp);
pCp-mix: 175 mM Tris-Cl pH 8, 15 mM MgCl_2 , 12 mM DTT, 2.4 mM spermine;
D-mix: 10 mM Hepes-KOH pH 8.3 in 33 % (v/v) DMSO;
Ligase-mix: 120 mM Hepes-KOH pH 8.3, 10 mM DTT, 30 mM MgCl_2 , 30 $\mu\text{g}/\text{ml}$ RNase-free BSA or gelatin, 3 mM ATP in 25 % (v/v) DMSO. Please note that the DMSO should be deionized freshly before preparation of these solutions; aliquots of DMSO or the reaction mixes can be stored frozen for several months.
5 $\times$ Poly(A) polymerase reaction buffer: 100 mM Tris-HCl, pH 7.0, 3 mM MnCl_2 , 0.1 mM EDTA, 1 mM DTT, 500 $\mu\text{g}/\text{ml}$ acetylated BSA, 50% glycerol.
5 $\times$ T7 DNA polymerase reaction buffer: 200 mM Tris-HCl (pH 7.5), 100 mM MgCl_2 and 250 mM NaCl.
- **Radioactive materials:** γ - ^{32}P -ATP (10 $\mu\text{Ci}/\mu\text{l}$; >3000 Ci/mmol), α - ^{32}P -Cordycepin-triphosphate, and α - ^{32}P -dATP (each of highest specific activity available) can be obtained from Hartmann Analytic or NEN/Perkin Elmer.

Protocol 3: 3'-Labeling of RNA by Ligation of 5'- ^{32}P -pCp

Protocol 3.1.1: Preparation of 5'- ^{32}P -pCp

- 1) For one labeling reaction, dry down 100 μCi (3.7 MBq) γ - ^{32}P -ATP and dissolve in 2 μl pCp-mix.
- 2) Add 1 μl 1 mM 3'-Cp and 5 U PNK, incubate for 1 h at 37 °C and dry under vacuum.

Alternatively, 5'- ^{32}P -pCp can be purchased from several suppliers and used directly for ligation.

Protocol 3.1.2: Ligation of 5'- ^{32}P -pCp to RNA

- 1) Dissolve the dry RNA (4–10 pmol) in 2 μl D-mix.
- 2) Denature for 2 min at 90 °C and snap-cool on ice.

- 3) Add 2 μ l Ligase-mix and transfer the mixture into the tube containing the dry ^{32}P -pCp.
- 4) Start the reaction by addition of 2 μ l T4 RNA ligase (5 U/ μ l), and incubate at 4 °C for 16–30 h.
- 5) Purify by one-dimensional denaturing PAGE (see Section 9.3.2.1, second dimension gel).

Protocol 3.2: 3'-Labeling of RNA with Poly(A) Polymerase and Cordycepin

For this highly sensitive labeling method, 2–10 nmol of RNA 3'-ends are sufficient.

- 1) Mix the purified RNA with 2 μ l 5 $\times$ Poly(A) polymerase reaction buffer (supplied by the manufacturer), 2 μ l (20 μ Ci) α - ^{32}P -Cordycepin-triphosphate and H_2O to a final volume of 9 μ l.
- 2) Add 1 μ l Poly(A) polymerase and incubate the reaction at 30 °C for 20 min.
- 3) Because the enzyme tends to bind to the RNA, a phenol extraction and precipitation step is advisable before proceeding to one-dimensional denaturing PAGE.

Protocol 3.3: 3'-Labeling of RNA with T7 DNA Polymerase

This “splint labeling” method is a special case because a single RNA species may be labeled specifically within a mixture of other nucleic acids, provided that at least the immediate 3'-sequence of that RNA is known. A DNA oligonucleotide should be obtained that is complementary to the immediate 3'-RNA end (T_m of hybrid duplex about 40 °C) and has one extra T residue at its 5'-end, providing a “5'-T-overhang” after annealing.

- 1) For annealing, combine an RNA preparation containing approximately 10–50 pmol of the desired species with 50 pmol of this oligonucleotide in a total volume of 17 μ l H_2O .
- 2) Incubate for 5 min at 70 °C and slowly cool down to 50 °C; this takes about 30 min.
- 3) Terminate the annealing process on ice.
- 4) Add 5 μ l of 5 $\times$ T7 DNA polymerase reaction buffer (supplied by the manufacturer), 2 μ l (20 μ Ci) α - ^{32}P -dATP, and 1 μ l T7 DNA polymerase (about 20 U/ μ l) and incubate for 30 min at 37 °C.
- 5) Analyze and purify the reaction products by denaturing PAGE.

9.5

Sequencing of End-Labeled RNA

Genomic sequencing has opened the view on a large number of putative novel RNA species. Their existence and primary sequence can easily be verified by a number of indirect techniques, including RT-PCR for the known part of the RNA sequence

and variations of the RACE method to determine the initiation and termination points of transcription, or of processing sites [10, 11]. However, if no hint to the sequence of an interesting functional RNA can be obtained by data mining, or if nucleotide modifications are suspected to play a role in RNA function, direct sequence analysis of the RNA is the method of choice. Base-specific enzymatic and chemical cleavage of end-labeled RNA can provide hints on the existence and identity of many modified nucleotides [17, 28–32]; detailed working protocols for both sequencing methods will be presented in Sections 9.5.1 and 9.5.2. The enzymatic as well as the chemical sequencing methods rely on cleavage reactions that are not completely specific for only one of the four major nucleobases; thus, the sequence has to be deduced from a partly ambiguous cleavage pattern in both cases. The advantage of the enzymatic over the chemical method is that sequence can be obtained from either labeled end, that more information on the nature of modifications can be deduced, and that the procedure is straightforward and fast. The main disadvantage is that strong secondary structure of an RNA may inhibit cleavage by certain enzymes, and that it may be difficult to obtain the required enzymes at sufficient quality. In contrast, the chemical modifications and subsequent cleavage reactions require only a small number of chemicals, and are mostly insensitive to secondary structure under the conditions used.

A major disadvantage of the enzymatic and chemical RNA sequencing approaches discussed here is that none of them will allow unambiguous reading of the terminal few nucleotides. This can be overcome by the two-dimensional mobility shift method [16, 24, 33] outlined in Section 9.6. If desired, the labeled 5'- or 3'-terminal nucleotide of any RNA can be determined by cleavage with nuclease P1 or RNase T2, respectively, and subsequent TLC analysis, as described below (Section 9.7.1). If the exact nature and position of nucleoside modifications is to be determined, a position-specific nucleotide analysis by TLC has to be performed [24, 34]. This method, described in Section 9.7.2, avoids labeling of the original RNA termini and gives the best information regarding modified nucleotides, but is quite laborious and time-consuming to perform. Also, it is quite often prone to secondary cleavages and thus not absolutely reliable for primary sequence determination.

In conclusion, the inherent advantages and disadvantages of the aforementioned methods demand a careful evaluation of the specific goals of each sequencing project, in order to determine in which combination they should be used.

9.5.1

Sequencing by Base-Specific Enzymatic Hydrolysis of End-Labeled RNA

For enzymatic sequencing, either 5'- or 3'-labeled RNA may be used. Although 10 000 c.p.m. per reaction are optimal, as little as 1000 c.p.m. may be used if sufficient exposure time is allowed. To resolve problems arising from secondary structure, all reactions are performed under denaturing conditions (8 M urea, 50 °C).

9.5.1.1 Materials Required for Enzymatic Sequencing

- **RNA:** 5'- or 3'-³²P-labeled, gel-purified RNA (see Sections 9.4.1 and 9.4.2), minimum total amount 10 000 c.p.m.; 1 µg/µl yeast tRNA (USB/Affymetrix, Sigma-Aldrich or Life Technologies).
- **Enzymes:** RNase T1 and *Staphylococcus aureus* nuclease are available from Calbiochem (Merck), Life Technologies (Ambion), Roche Biochemicals and Worthington, respectively. The adenosine-specific RNase U2 is included in the protocols, although it is not commercially available anymore. The same is true for the cytidine-specific RNase CL3, but this enzyme can be prepared with little effort [35]. Because the quality differs between production batches, each lot should be tested separately, using the protocol below. Working solutions of 100 mU/µl (RNases T1 and U2), 3 U/µl for *Staphylococcus* nuclease, and 13 mU/µl for RNase CL3 should be freshly prepared in H₂O before use.
- **Reaction mixes:** T1 mix and H⁺ mix should be freshly prepared from appropriate stock solutions; all other solutions may be prepared in advance and stored at -20 °C.

Dye stock: 1% (w/v) BPB, 1% (w/v) XC, 2% (w/v) SBB in 10 mM Tris pH 7.5.

T1 mix: 20 mM Na-Citrate pH 3.5, 1 mM Na₂EDTA, 0.03% (v/v) dye stock, 8.1 M urea;

S7 mix: 20 mM Tris-Cl pH 7.5, 10 mM CaCl₂, 0.03 % (v/v) dye stock, 8.1 M urea;

CL3 mix: 20 mM Tris-Cl pH 7.5, 0.03 % (v/v) dye stock, 8.1 M urea;

H⁺ mix: 0.22 N H₂SO₄, 0.03% (v/v) dye stock, 6.8 M urea.

Protocol 4: Working Procedure for Enzymatic Sequencing

- 1) Prepare 10 Eppendorf-type reaction tubes with equal aliquots of your end-labelled, gel-purified RNA (between 1000 and 10 000 c.p.m. per reaction). Label one tube as control without enzyme (-E), one for the ladder (H⁺), and two for each of the 4 enzymes (see Table 9.1). Calculate the amount of carrier RNA (from the elution and precipitation) in each aliquot, and adjust to 1 µg/tube for the control and enzymatic reactions. For the acid ladder, adjust

Table 9.1 Cleavage reactions for enzymatic sequencing.

Reaction	Yeast tRNA (µg)	Mix	Enzyme	Incubation	Specificity
-E	1	T1 mix	–	15 min, 50 °C	–
T1	1	T1 mix	RNase T1	15 min, 50 °C	G
U2	1	T1 mix	RNase U2	15 min, 50 °C	A >> ms ² i ⁶ A, G
S7	1	S7 mix	<i>Staphylococcus</i> nuclease	15 min, 50 °C	A, U > T, s ⁴ U, ms ² i ⁶ A, m ² A
CL3	1	CL3 mix	RNase CL3	15 min, 50 °C	C >> A, T
H ⁺	5	H ⁺ mix	–	3 min, 100 °C	All except 2'-O-methyl

carrier to 5 µg. Dry down the contents of all tubes at room temperature (this takes about 30 min in a Speedvac, or 2 h in an evacuated desiccator over fresh desiccant). Prepare an ice box with wet ice, and one with finely crushed dry ice, both large enough to hold the 10 tubes deeply immersed. Preheat three water baths, one at 50 °C and the other two at 65 °C and 95–100 °C, respectively. Alternatively, metal-based heating blocks may be used, but heat transfer is faster and more efficient in water.

- 2) Add 4 µl of the respective enzyme reaction mix to each labeled tube (except H⁺; see Table 9.1), spin down shortly, denature for 5 min at 65 °C, and quickly chill on ice.
- 3) Add 1 µl of the respective enzyme working solution to the first of your two tubes for the same enzyme (for example, T1). Mix by pipetting in and out, and transfer exactly 1 µl to the second tube. Immediately, put the two tubes in the 50 °C bath and incubate for exactly 15 min; stop the reactions in dry ice. The remaining tubes are treated in the same way.
- 4) For the acid ladder, add 4 µl of H⁺ mix to the respective tube, followed by incubation in a boiling water bath for exactly 3 min and quenching in dry ice.
- 5) All samples may be stored overnight at –80 °C at this stage.
- 6) After a short spin at room temperature (be aware that the urea may precipitate at lower temperatures), load samples immediately onto a denaturing PAA sequencing gel (40 cm long, 20 cm wide, 0.4 mm thick, 12 lanes per RNA; composition as in Section 9.3.2.1, second dimension). For RNAs 70–90 nt long, use 20% PAA; for longer RNAs, use 15% PAA. To read over the entire sequence length, it is advisable to prepare enough material for two runs: a short run (BPB just leaving the 20% PAA gel), and a long run (XC migrating to the lower edge of the gel). If the sequence should be read up to the labeled end, precipitate the cleavage reactions with EtOH, omit BPB from the loading mix and use a 25% PAA gel (see Figure 9.1).

9.5.1.2 Interpretation and Troubleshooting

From the counting ladder (H⁺) and the highly base-specific RNases T1 and U2 (and CL3 if available in good quality), it is straightforward to deduce a large part of the sequence. To read the band pattern created by *Staphylococcus* nuclease, recall that this enzyme cleaves 5' of the respective nucleotide, leading to a band shift (see Figure 9.1). A gap in the counting ladder (H⁺) indicates a 2'-ribose methylation of the corresponding nucleotide. Single weak bands may result from base modifications; see Table 9.1 for an incomplete overview and [16, 31] for a full discussion of this issue.

If parts of the gel show weak bands in all lanes, strong secondary structure may hinder efficient cleavage; in this case, denature the RNA at 90 °C, and run the reactions at 65 °C (you may have to use a higher enzyme concentration owing to partial enzyme denaturation). If bands are compressed on the gel, insufficient denaturation during electrophoresis is the reason. Make sure that the gel is run at 40 W (for a gel 40 cm long, 20 cm wide and 0.4 mm thick), or use a thermostated electrophoresis apparatus at 60–65 °C.



reading from the second nucleotide (bottom of the gel). The position of the dyes (SBB, XC, BPB) is given on the right, and the sequence of the first 18 nt on the left side of the panel (the terminal U was determined by end-group analysis). Note the gap in the ladder (H^+) at the position of 2'-O-methyl-guanosine (G_m). For details of the sequence and its interpretation, see [5].

Sequencing by Base-Specific Chemical Modification and Cleavage

Chemical sequencing gives clear results only for 3'-labeled RNA, because of the inhomogeneous cleavage products 5' of the attacked nucleotide [29]. The precipitations required to stop the reactions and to remove the aniline before electrophoresis lead to some loss of material; thus, a higher amount of radioactive starting material should be used relative to enzymatic sequencing. The following protocol is a simplified and slightly modified version of the original protocol, which should be consulted for full details [29, 30].

9.5.2.1 Materials Required for Chemical Sequencing

- **RNA:** 3'-³²P-labeled, gel-purified RNA (Section 9.4.2), minimum total amount 25 000 c.p.m.; 10 µg/µl yeast tRNA (Roche Biochemicals).
- **Chemicals:** hydrazine, diethylpyrocarbonate (DEPC), aniline, NaBH₄, dimethylsulfate (DMS), and EtOH should be of the highest purity available and stored dry at 4 °C.
- **Buffers and reagents:** 50 mM NaOAc, pH 4;
1 M aniline acetate, pH 4.5 (mix H₂O : acetic acid : aniline at a volume ratio of 7 : 3 : 1; spin down precipitate, check pH of an aliquot, store frozen in aliquots);
0.5 M NaBH₄ (prepare fresh before use);
0.3% (w/v) DMS in 50 mM NaOAc, pH 4 (mix directly before use);
hydrazine/H₂O: mix equal volumes and keep on ice until use;
3 M NaCl in hydrazine: dry NaCl in an oven at 120 °C for 2 h, store in a desiccator; dissolve in hydrazine and keep on ice until use.
- **Waste disposal:** DMS is a carcinogen; all solutions containing it (for example, reagents and EtOH supernatants from the first precipitation) should be disposed into a bottle containing 5 M NaOH. Hydrazine waste is inactivated with 3 M FeCl₃. Aniline, DMS, and hydrazine bottles should be opened only under a flow hood.

Protocol 5: Working Procedure for Chemical RNA Sequencing

- 1) Prepare five reaction tubes containing equal aliquots of your end-labeled, gel-purified RNA (between 5000 and 20 000 c.p.m. per reaction) plus 20 µg yeast tRNA and label them according to the four sequencing reactions and control (A, C, G, U, -E; see Table 9.2).
- 2) Dry down the contents of all tubes at room temperature (about 30 min in a Speedvac, or 2 h in an evacuated desiccator over fresh desiccant). Prepare an ice box with wet ice, and preheat one water bath at 60 °C and another at 90 °C.
- 3) All chemical modification reactions are carried out according to the flow sheet in Table 9.2; note that all precipitations are on ice (or at -20 °C) for 5 min, and all centrifugations are at 4 °C and 10 000 g for 5 min (precipitations) or 1 min (wash), respectively. Be careful to remove all of the supernatants to avoid carryover of reagents, which will cause a smear on the sequencing gel.
- 4) For aniline cleavage, add 10 µl of aniline acetate to all tubes including the control tube. Incubate for 20 min at 60 °C, stop on ice and precipitate with 150 µl NaOAc and 650 µl EtOH. After two washes with 800 µl EtOH to completely remove residual aniline, dry the RNA, dissolve in urea gel loading buffer, denature for 3 min at 95 °C and analyze on a sequencing gel (see Protocol 4, step 6).
- 5) If desired, a counting ladder may be prepared by acid hydrolysis (Protocol 4, step 4) and run in parallel. Note that owing to the different cleavage mechanisms, the bands of this ladder migrate by about 1 nt slower than the corresponding bands obtained by chemical cleavage.

Table 9.2 Working table for chemical sequence analysis of RNA

Specificity	G	A	U	C
Modification:				
<i>Reagent</i>	10 μ l 0.3% DMS (w/v) in 50 mM NaOAc, pH 4	150 μ l 50 mM NaOAc (pH 4) + 1 μ l DEPC	10 μ l hydrazine/H ₂ O (equal volumes)	10 μ l 3 M NaCl in hydrazine
<i>Incubation</i>	40 s, 90 °C	10 min, 90 °C	10 min, 0 °C	10 min, 90 °C
<i>First precipitation</i>	150 μ l NaOAc, 650 μ l EtOH	400 μ l EtOH	150 μ l NaOAc, 550 μ l EtOH	500 μ l 80% EtOH
<i>Second precipitation</i>	—	150 μ l NaOAc, 450 μ l EtOH	150 μ l NaOAc, 450 μ l EtOH	150 μ l NaOAc, 450 μ l EtOH
<i>Wash</i>	800 μ l EtOH	800 μ l EtOH	800 μ l EtOH	800 μ l EtOH
Reduction:				
<i>Reagent</i>	10 μ l 0.5 M NaBH ₄	—	—	—
<i>Incubation</i>	10 min, 0 °C (dark)	—	—	—
<i>Precipitation</i>	150 μ l NaOAc, 650 μ l EtOH	—	—	—
<i>Wash</i>	800 μ l EtOH	—	—	—

Starting material is dried 3'-end labeled RNA containing 20 μ g yeast tRNA per reaction tube. All centrifugations are performed in a microcentrifuge at 4 °C (10 000 g); all precipitations are done in crushed dry ice for 10 min. NaOAc stands for 50 mM NaOAc, pH 4.

9.5.2.2 Interpretation and Troubleshooting

From the counting ladder (H⁺), and the highly base-specific cleavages at G, A, and U residues, a large part of the sequence can be easily deduced. Because the C reaction modifies both pyrimidines (although with lower efficiency for U residues), a “subtractive reading” of the U and C lanes is required for unambiguous identification of both bases. Some modified bases can be identified very easily: m⁷G is sensitive toward aniline without any further modification and thus appears as an extremely strong band in all lanes including the control; this reaction may even prevent reading of the subsequent bases. All uridine derivatives except pseudouridine are weakly reactive toward the U reaction; ac⁴C can be recognized as a band in all lanes, but weaker than the appearance of m⁷G. For a more complete overview and details of the procedure, see [16, 31]. Band compression due to strong secondary structure of the RNA can be avoided by running the gel at higher temperature (see Section 9.5.1.2).

9.6

Determination of Terminal RNA Sequences by Two-dimensional Mobility Shift

This method is suitable for obtaining sequence information close to the labeled RNA end from statistical RNA hydrolysates. Before two-dimensional analysis (see application example in Figure 9.2), an end-group analysis of the labeled terminal nucleotide is performed. This is necessary because separation in the first dimension is mostly depending on the net charge of the bases at pH 3.5, which depends on their pK_a values. Although the negatively charged phosphate groups promote migration of the RNA fragments strictly according to length, the different charges of the bases lead to retardation of C- and A-rich fragments relative to U- and G-rich fragments. This influence is greater for short fragments; thus, the identity of the end group determines the position where the first dimension strip should be blotted onto the second dimension TLC plate to avoid loss of sequence information (see Protocol 6.3 and Figures 9.2 and 9.3). The terminal nucleotide is identified from an aliquot of the labeled RNA, as described in Protocol 6.1 for 5'- or 3'-labeled molecules. For the two-dimensional separation, a statistical hydrolysate is obtained either by cleavage with nuclease P1, or by alkaline hydrolysis in NaHCO_3 followed by neutralization with HCl [24, 33, 36]. Either reaction product is dried and dissolved in 1 μl H_2O .

9.6.1

Materials Required for Mobility Shift Analysis

Except for materials unique to mobility shift analysis, enzymes and buffers are listed in Section 9.7.1.1. Cellulose acetate (CA) strips in different formats are available from Sigma-Aldrich and from Alpha Laboratories; for example, the SAVANT Instruments apparatus requires strips of 3×55 cm. DEAE Cellulose TLC plates (20×20 or 20×40 cm) can be obtained from Sigma-Aldrich, Merck Millipore, and Macherey-Nagel. In our hands, TLC plates from the latter supplier (CEL DEAE/HR mix 20) stand out in terms of resolution and handling stability. The buffer for equilibration of the first dimension strips is 5% (v/v) acetic acid, 1 mM Na_2EDTA , 7 M urea (pH 3.5); electrophoresis buffer is 5% (v/v) acetic acid, 1 mM Na_2EDTA , 0.5% pyridine (pH 3.5). The second dimension solvent (30 mM homomix) is prepared as follows: 30 g yeast RNA (Sigma-Aldrich, Type VI), 420 g urea and 1.7 g KOH are dissolved in 900 ml H_2O and stirred for 20 h at 65°C . After cooling to room temperature, the pH is adjusted to 4.7 with acetic acid, and the volume brought to 1 l.

Protocol 6.1: End-Group Analysis of RNA

- 1) For end group analysis of 5'-labeled RNA, dry a small aliquot (1000–5000 c.p.m) together with 5 μg of yeast tRNA. Redissolve the RNA in a solution of 0.1 μg nuclease P1 in 10 μl 50 mM ammonium acetate pH 5.3 and incubate for 2 h at 50°C .

- 2) The end group of 3'-labeled RNA is determined in a similar manner: dry the sample including 5 µg of yeast tRNA, redissolve in 10 µl 10 mM ammonium acetate pH 4.6 containing 0.5 U RNase T2, and incubate for 5 h at 37 °C.
- 3) Either sample is dried in a desiccator and analyzed by two-dimensional nucleotide analysis, as described in Protocol 7.2.

Protocol 6.2.1: Preparation of a Statistical RNA Hydrolysate by Alkaline Hydrolysis

- 1) Vacuum-dry the radiolabeled RNA sample plus 25 µg yeast tRNA in a reaction tube; preheat a thermostated water bath or heating block to 95 °C.
- 2) Thoroughly resuspend the dried RNA in 4 µl of hydrolysis buffer (50 mM NaHCO₃, 0.5 mM EDTA pH 9.1); hydrolyze at 95 °C for 15 min, snap-cool on ice.
- 3) Spin down briefly to recover condensed water, add 1.5 µl 2 M HCl and incubate for 2 min at 37 °C to open cyclic phosphates. Dry the samples (if necessary, they may be kept at -20 °C at this stage). Directly before the first dimension run, dissolve the sample in 1–2 µl H₂O. This hydrolysis method can be used for all 5'- or 3'-labeled RNAs.

Protocol 6.2.2: Preparation of a Statistical RNA Hydrolysate by Enzymatic Hydrolysis with Nuclease P1

- 1) Vacuum-dry the radiolabeled RNA sample plus 100 µg yeast tRNA.

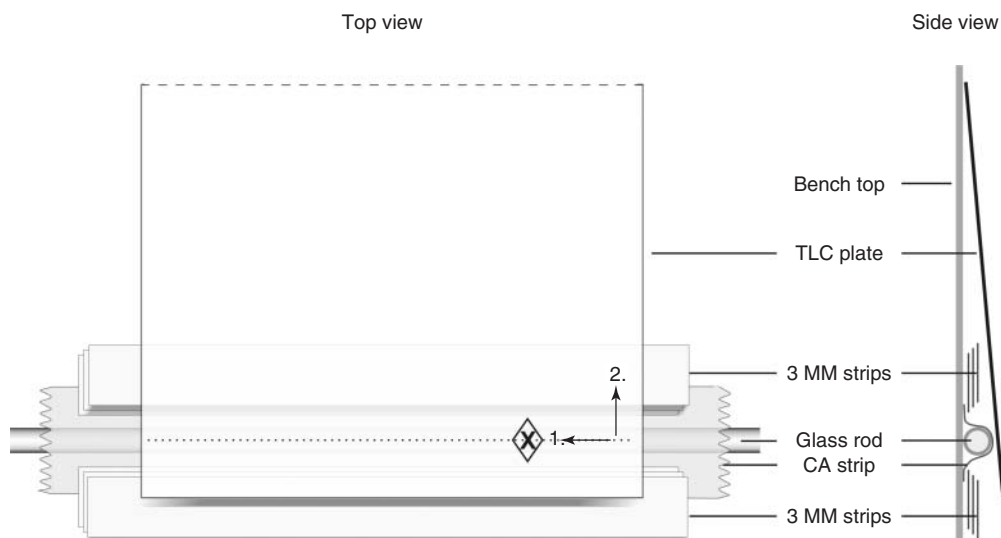


Figure 9.2 Schematic depiction of the assembly used for the transfer of RNA fragments from first dimension (CA strip) to second dimension (DEAE plate). For details, see Protocol 6.3, step 2. The fringed ends of the CA strip indicate that they

extend beyond the figure. The broken line at the top edge of the TLC plate illustrates that the TLC plate may be 20 or 40 cm long; the larger plates are required to resolve sequences of 5 or more nucleotides.

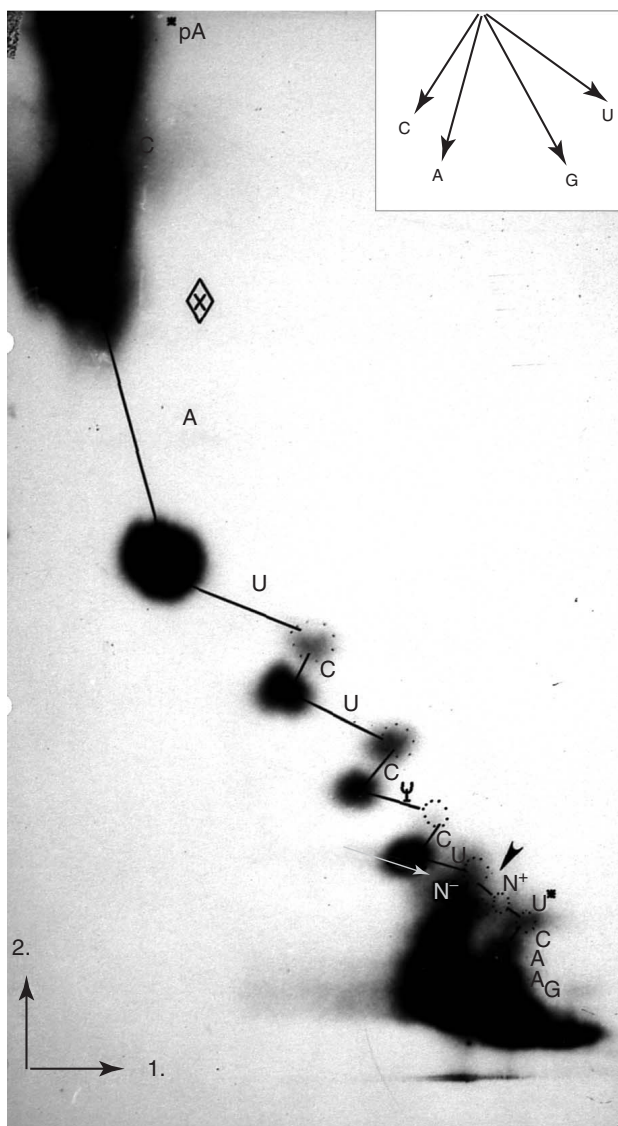
- 2) For each RNA aliquot to be withdrawn in step 5 below, prepare one reaction tube containing 3 μ l 250 mM Na₂EDTA, positioned in a small insulated vessel filled with crushed dry ice.
- 3) Dissolve the RNA sample in 10 μ l 50 mM ammonium acetate pH 5.3 and add 1.5 μ l nuclease P1 (5 ng/ μ l). The large amount of RNA added will increase the volume by 1–2 μ l.
- 4) Incubate in a water bath or heating block at 50 °C.
- 5) Transfer 3 μ l aliquots at 0, 2, 5, and 10 min into one of the tubes on dry ice (see step 2). Denature the samples for 4 min at 95 °C and vacuum-dry. Dissolve in 1–2 μ l H₂O. Note that this hydrolysis method cannot be used for pCp-labeled RNA, because nuclease P1 can also remove the unlabeled 3'-phosphate introduced by 5'-³²P-pCp, generating a mixture of fragments that include or lack this 3'-terminal phosphate.

Protocol 6.3: Mobility Shift Analysis

- 1) The oligonucleotide mixture obtained from either alkaline hydrolysis or nuclease P1 digestion (Protocol 6.2.1 or 6.2.2) is first separated by net charge (i.e., base composition), using high-voltage electrophoresis on CA strips under denaturing conditions at pH 3.5. 1 μ l of RNA hydrolysate is applied as a small spot 18 cm from one end. As a tracer, 1 μ l of 1% (w/v) XC is applied at both sides of the sample; it should migrate a distance of about 8–9 cm, which is conveniently marked with a pencil on the strip. The original apparatus is a solvent-cooled tank (Savant Instruments) run at 5000 V; however, any paper electrophoresis unit with a good cooling device (e.g., from Alpha Laboratories) should give comparable results.
- 2) In the meantime, prepare the plate for the second dimension: label the starting line 2 cm from the short edge of a 20 $\times$ 40 cm DEAE Cellulose TLC plate, either with a soft pencil on the DEAE side, or with a water-resistant marker pen on the glass back side. On the glass back side, also label the position

Figure 9.3 Sequence analysis of a tRNA fragment by two-dimensional mobility shift. Barley chloroplast tRNA^{Glu} [4] was subjected to total fragmentation by RNase T1. The cleavage products were labeled at their 5'-OH termini (Section 9.4.1) and separated by denaturing gel electrophoresis. A 16 nt long fragment spanning the anticodon region was recovered and sequenced as described in Protocol 6.3. Alkaline hydrolysis led to complete cleavage at position 34 of the tRNA (not shown). Statistical cleavage by nuclease P1 (Protocol 6.2.2) resulted in branching of the chromatogram at position 34 (black arrowhead): the alkali-labile nucleotide at this position behaves either C-like (left branch; N⁻ indicated by the grey arrow) or U-like (right branch, N⁺). Sequencing by

position-specific nucleotide analysis followed by CNBr modification (Section 9.7.2; not shown) identified this anticodon nucleotide as m⁵s²U, which is also present in *E. coli* tRNA^{Glu} [4]. The sequence is written in the gaps between fragment signals. Note that the chromatogram is read from the top (labeled adenosine terminus) to the bottom. The chromatographic properties of the major nucleoside monophosphates in both dimensions determine the directional distances between the oligonucleotide fragments, as is schematically depicted in the inset (top right). The orientation of first and second dimension runs is given by the arrows in the lower left corner; the position of the Xylene cyanol marker after the run is indicated by a diamond.



(for criteria, see below) where the XC marker indicated by the crossed diamond in Figure 9.2 will be blotted. Preparation of the plate, paper strips, and blotting assembly (see below) in advance is advisable to prevent drying out of the CA strip. After the first dimension run, the oligonucleotides are blotted onto the prepared TLC plate as follows (Figure 9.2): mount a glass rod or 2 ml glass pipette (diameter ca. 5 mm) on a water-resistant table and fix at both ends with adhesive tape. Cut Whatman 3 MM paper into strips of 3×30 cm (6 strips per plate); soak in water. Position the wet CA strip lengthwise on the rod; carefully fold both ends of the CA strip (55 cm long) to direct them away from the rod such that the XC dye marker is positioned about 10 cm from the folded edge, and the CA strip is covering the rod for a length of about 30 cm (see Figure 9.2). Place the wet 3MM paper along both long sides of the CA strip with about 5 mm overlap; this serves as a reservoir for blotting and prevents curling of the CA edges. If necessary, cover the outer half of the strips with Saran wrap to prevent drying. Put the plate (with the DEAE side facing down) on top of the CA strip covering the rod such that the labeled starting line is touching the CA strip and the XC dye is at the proper position in the lower left corner of the DEAE side (note that the plate is inverted!). The XC marker is positioned at the following distance from the right edge of the inverted TLC plate: at 1, 3, 5, or 6 cm when the terminal nucleotide is U, G, A, or C, respectively. This careful positioning is important for inclusion of all RNA fragments within the width of the plate, because uridine residues have the largest influence on mobility in the first dimension, particularly in small oligonucleotides (see Figure 9.3). Thus, terminal U residues shift the cohort of labeled fragments to the right relative to the XC marker. An evenly distributed weight (e.g., several small lead containers or water-filled beakers) put on top of the contact line will increase transfer efficiency. The transfer is finished when the dye is visible on the plate (after about 1 h).

- 3) For separation in the second dimension, three Whatman 3MM strips (8×30 cm) serving as wicks are fixed at the top end of the plate with clips. In a 65°C incubator, prewarm two chromatography tanks with tight fitting lids; fill one with water and the other with 30 mM homomix, about 1.5 cm high. Run the plate in water until the front has migrated about 20 cm. Transfer the plate to the tank containing the homomix and continue chromatography in the same direction for about 10 h, again using XC as a migration marker, which should reach the top of the plate. After chromatography, the wicks are carefully removed, the plates are thoroughly dried, and an X-ray film or imager plate is exposed. Sequences up to 12 nucleotides can be resolved on 20×20 cm plates; for longer sequences, 20×40 cm plates are required.

9.7

Determination of Modified Nucleotides by Postlabeling Methods

In many cases, it is desirable to obtain an overview of the modified nucleotides present in a purified RNA species, or in an RNA population obtained from a

certain organism. If the pure RNA species can be easily obtained in microgram amounts HPLC analysis is the method of choice because UV spectra provide additional information on the nature of the nucleobase. Coupled HPLC-MS will even identify unknown or novel nucleotides [37, 38]. However, the required apparatus infrastructure is not readily available for most laboratories, and even though the sensitivity of the methods is impressive, availability of the biological samples may be limiting. A reliable alternative to determine the nucleotide content of sub-picomolar amounts of RNA is the postlabeling of an RNA hydrolysate, followed by two-dimensional TLC analysis of the products [24, 39–44]. An overview of these diverse methods is found in [45]. The determination of sequence and base modification at the same time has been made possible by the coupling of limited RNA fragmentation and end-group identification of the terminally labeled, separated fragments [34].

9.7.1

Analysis of Total Nucleotide Content

The first step of this procedure is the hydrolysis of the purified RNA (Section 9.3) by a mixture of RNases T2 and A. The resulting nucleoside 3'-phosphates are then radioactively labeled at the 5'-end by PNK; these 5'-³²P-labeled 3',5'-nucleoside diphosphates are then converted to the corresponding nucleoside 5'-phosphates by nuclease P1. After elimination of residual ATP by Apyrase, the mixture is subjected to two-dimensional TLC, with an excess of unlabeled nucleoside 5'-phosphates comigrating as standards (see application example in Figure 9.4).

9.7.1.1 Materials Required for RNA Nucleotide Analysis

- **Enzymes:** Pancreatic ribonuclease (RNase A) and T4 PNK (3'-phosphatase free) are from Roche Biochemicals; recombinant RNase T2 is available from Mobitec. Prepare a working solution (T2/A mix) containing 50 mU/μl RNase T2 and 0.1 μg/μl RNase A in H₂O (can be stored frozen); Apyrase (5 U/ml; Sigma-Aldrich); Nuclease P1 (Roche or Sigma-Aldrich); prepare a working solution of 5 ng/μl (Protocol 6.2.2) or 10 ng/μl in 50 mM ammonium acetate pH 5.3.
- **Materials for TLC:** Two glass tanks with tightly closing lids, preferably with fittings for several plates; Cellulose TLC plates (plastic or glass backed, non-fluorescent, analytical scale), 20 × 20 cm (Macherey-Nagel or Merck).
Solvent A: Isobutyric acid : conc. Ammonia : H₂O (57.7 : 3.8 : 38.5, (v/v)), pH 4.3;
Solvent B: Isopropanol : conc. HCl : H₂O (70 : 15 : 15).
pN marker mix: 5 mg/ml each of pA, pG, pC, and pU; store frozen in small aliquots.
- **Radioactive materials:** γ-³²P-ATP (10 μCi/μl; >3000 Ci/mmol).

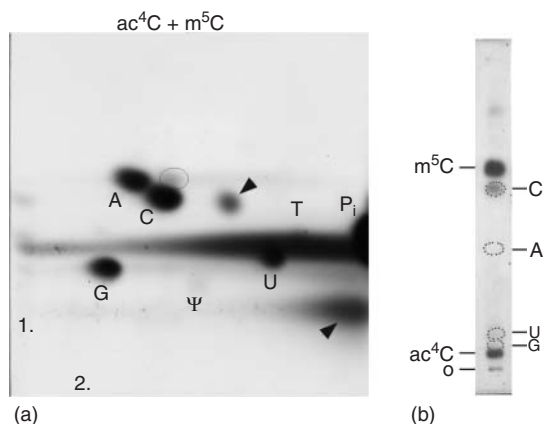


Figure 9.4 Identification of modified nucleotides in two different chromatographic systems. (a) Total nucleotide analysis of barley chloroplast tRNA^{Glu} by two-dimensional cellulose TLC was performed as described in Section 9.7.1. The positions of the four major nucleotides, pseudouridine (Ψ), ribothymidine (T), and of the inorganic phosphate derived from the labeling reaction are indicated by letters. Arrowheads show unidentified reaction products. The long arrow points to the position where ac⁴C and m⁵C comigrate. This spot is only detectable after long exposures (not shown) and was recovered for secondary analysis. Starting point and directions of both chromatographic runs are indicated by small

arrows in the left bottom corner; weak spots are circled. (b) Here, the two nucleoside monophosphates ac⁴C and m⁵C, which comigrate in the cellulose TLC system in both dimensions, were separated according to net charge by chromatography on PEI plates using the first dimension solvent described in [39]. The cytidine spot is due to carryover resulting from the slight overlap with these modified nucleotides in the two-dimensional primary analysis (see (a)). Positions of non-radioactive marker nucleotides (dotted circles) are shown on the right side of the panel, and the starting point (o) is indicated.

Protocol 7.1: Preparation of 5'-³²P-Labeled Nucleoside Monophosphates

- 1) Dry purified RNA (2–20 pmol) under vacuum and dissolve in 2 μl 50 mM ammonium acetate pH 4.5 and 6 μl H₂O (include a control sample without RNA).
- 2) Add 1 μl of RNase T2/A mixture and incubate the sample for 5 h at 37 °C.
- 3) To the resulting hydrolysate, add 1 μl 10 × concentrated PNK buffer (provided by the manufacturer), 0.5 μl 0.1 mM ATP, 25 μCi γ-³²P-ATP and 5 U PNK.
- 4) Incubate for 30 min at 37 °C.
- 5) Add 1 μl Apyrase, incubate for another 30 min at 37 °C, and proceed with half of the mixture (save the other half at –20 °C). Vacuum-dry this aliquot, add 10 μl of nuclease P1 solution and incubate for 3 h at 37 °C.

Protocol 7.2: Two-Dimensional TLC of Nucleoside Monophosphates

- 1) For analytical TLC, combine 1 μl of the labeled RNA hydrolysate from step 5 above with 1 μl pN marker mix.

- 2) Mark the start point with a soft pencil in the lower left corner of a cellulose plate, 1.5 cm from each edge.
- 3) Load the sample in a small spot (preferably with a drawn-out capillary) and let dry; if necessary, use a hair dryer at medium heat.
- 4) Develop the first dimension in solvent A until the front has reached the top edge; carefully remove excess solvent from the back and edges, then dry the plate thoroughly under a fume hood.
- 5) For chromatography in the second dimension, turn the plate by 90° compared to the first dimension and develop in solvent B.
- 6) After drying, expose an X-ray film or Phosphorimager screen to the plates. The marker nucleotides are visualized as dark blue spots under a UV lamp at 254 nm; their position is marked as an aid in the identification of unknown nucleotides.

If a specific nucleotide has to be subjected to secondary analysis, the whole reaction mix (Protocol 7.1, step 5) should be loaded onto the TLC plate. The corresponding spot is then localized and recovered as follows: attach fluorescent markers to the corners of the TLC plate and illuminate briefly; in the darkroom, fix an X-ray film on the TLC plate, expose for several hours (if necessary, take several exposures); alternatively, visualize the exposure in a Phosphorimager and print the picture at actual size, preferably on a transparent medium. Fix the image (X-ray film or printout) on a light box, put the TLC plate (with its top side up) on the image, align the illuminated TLC plate and image *via* their markers. Label the region(s) of interest with a pencil, and then outline the window to be scraped off with a scalpel before scraping off the cellulose layer within the area containing the nucleotide of interest. For each nucleotide to be recovered, prepare a pipet tip (20–200 µl size) with a small cotton plug fixed inside the tip close to its apex (narrow end). Using a flexible, small diameter tubing (tightly fitting this end), connect the tip to an aspirator. Using only partial vacuum, scrape off the region of interest with the wide end of the tip and collect the cellulose by vacuum. Put each tip into a 1.5 ml reaction tube (wide tip opening at the top, preferably with a hole in the lid to stabilize the tip during centrifugation), fill 50 µl H₂O into the tip, let stand for 5 min to hydrate the cellulose, and elute the radioactive nucleotide(s) by brief centrifugation [16]. Recovery is easily tested with a scintillation counter; if necessary, repeat the elution step. The recovered nucleotides are centrifuged for 10 min at 10 000 g to remove residual cellulose particles. After drying *in vacuo*, the supernatant is ready for secondary analysis.

9.7.1.2 Interpretation and Troubleshooting

The four marker nucleotides should appear under UV as clearly separated spots (see reference patterns in [17, 40–44]; the ³²P_i (resulting from hydrolysis of unincorporated ATP) should be visible as a prominent spot on the X-ray film in the center of the right edge. If separation of nucleotides is unacceptable, check the pH of the solvents and replace them if necessary; make sure that the sample was dried completely after application (use an infrared lamp or hot air if necessary), and check that the lids of the tanks close tightly.

If the starting point looks streaky in the first dimension, the problem might be the amount of protein in your sample; try to reduce the amount of enzyme used and extend the incubation times instead. If the problem occurs in a secondary analysis, it might be due to cellulose particles. In this case, repeat the centrifugation step (see previous section) at higher speed and take care in avoiding carryover of cellulose particles.

If comparison to the standard pattern reveals that many dinucleotides are present in your sample, you should analyze an aliquot of your sample before nuclease P1 digestion. This reveals whether you should increase the amount of RNase T2/A mix and/or nuclease P1, or extend the respective incubation time.

9.7.2

Determination of Position and Identity of Modified Nucleotides

This sequencing method allows the simultaneous elucidation of sequence and base modifications. Limited RNA hydrolysis (ideally, one cut per RNA molecule) is performed non-enzymatically, and the resulting fragments are radioactively labeled [24, 34]. After electrophoretic separation, the 5'-terminal nucleotide of each isolated fragment is determined by TLC. In the original protocol, each band corresponding to an RNA fragment was analyzed by two-dimensional TLC using one plate per fragment [34]. A less time-consuming and laborious approach uses one-dimensional TLC in two separate solvent systems. Here, up to 20 different fragments are loaded simultaneously onto one TLC plate, and onto a second replica plate to be developed with another solvent. The use of two different solvent systems is often necessary because many of the modified nucleotides comigrate with one of the four standard nucleotides, and can only be identified unambiguously by the combination of both solvent systems. The material required is mostly identical to that specified in Section 9.7.1.1; in addition, sterile glass capillaries (5 or 10 μ l size) and a gas burner are needed.

Protocol 8.1: Generation and Separation of 5'-Labeled Random RNA Fragments

- 1) In separate reaction tubes, dry down 20–40 pmol purified RNA, 50–100 μ Ci γ -³²P-ATP and 10 μ l of urea gel loading buffer (see Section 9.3.2.1).
- 2) Preheat a water bath to 95 °C.
- 3) Dissolve the RNA in 1.5 μ l H₂O, transfer it into the center of the capillary by aspiration, and seal the ends with the flame. Perform hydrolysis for exactly 30 s in the boiling water bath and stop in ice water.
- 4) Cut open the ends of the capillary with a glass cutter, transfer the contents back to the original tube; recover residual RNA by rinsing the capillary with 5 μ l H₂O.
- 5) Transfer the combined RNA samples to the tube containing the dry γ -³²P-ATP and proceed with 5'-labeling (see Section 9.4, Protocol 2.1). Note that the fragments generated by alkaline hydrolysis have 5'-OH termini, and so the

kinase reaction is carried out without dephosphorylation. The RNA fragments are separated by electrophoresis on a 15% PAA gel. Best separation results are achieved on denaturing gels of 60 or 80 cm length, and running at 60 or 80 W until BPB has reached the bottom; if such glass plates are not available, a short and long run should be performed on a 40 cm long gel.

- 6) Localize fragments by autoradiography, cut out, elute and ethanol-precipitate as described in Section 9.3.2.2, including 10 µg yeast tRNA per band.

Protocol 8.2: Identification of the 5'-End Group of the RNA Fragments

- 1) After ethanol precipitation, dissolve each sample in 10 µl of nuclease P1 solution (Section 9.7.1.1; 10 ng/µl in 50 mM ammonium acetate pH 5.3).
- 2) Incubate for 2 h at 50 °C.
- 3) Remove an aliquot (1–5 µl, depending on labeling efficiency) and dry (the rest may be stored frozen).
- 4) Dissolve each dry sample in 2 µl of pN marker mix, and apply equal amounts to each of two cellulose TLC plates (see Section 9.7.1.1). The TLC plates are prepared such that 16–20 samples can be applied as thin horizontal streaks, about 0.75 cm wide and with 0.5 cm spacing, 1.5 cm from the bottom edge.
- 5) Develop one of the TLC plates with solvent A and the other with solvent B (see Protocol 7.2). Using this parallel analysis in two different solvent systems, the large number of samples can be analyzed fast and conveniently, and the RNA sequence can be read directly from the TLC plate if the 5'-terminal nucleotides (5'-end groups) derive from RNA fragments successively shortened by 1 nt from the 5'-end (see Figure 9.5). If a 2'-O-methylated dinucleotide is detected, an aliquot of the corresponding sample is digested with 1–10 µg of nuclease P1 (5 h at 65 °C) and analyzed as before, or by two-dimensional TLC (see Protocol 7.2).

An example for electrophoretic separation of the 5'-labeled RNA fragments derived from statistical hydrolysis, chromatography of the fragment end groups in both solvents, and sequence readout is given in Figure 9.5.

9.7.2.1 Interpretation and Troubleshooting

Ideally, the band pattern visible after electrophoresis should show an even distribution, with a resolution up to the penultimate nucleotide. If the ladder is shifted significantly toward the smaller fragments, try to increase the amount of RNA, or reduce the hydrolysis time. A specific problem may arise if a labile modified nucleotide, or a C–A bond in a single-stranded region, is present in the RNA. In this case, near-quantitative hydrolysis of the corresponding phosphodiester bond may occur and lead to a complete lack of bands above this point; in this case, the nucleotide 3' of the cleaved bond will be visible in all other samples [4]. Most problems arising from the TLC systems have been discussed in Section 9.7.1. In some cases, it may be necessary to reanalyze specific modified nucleotides in a

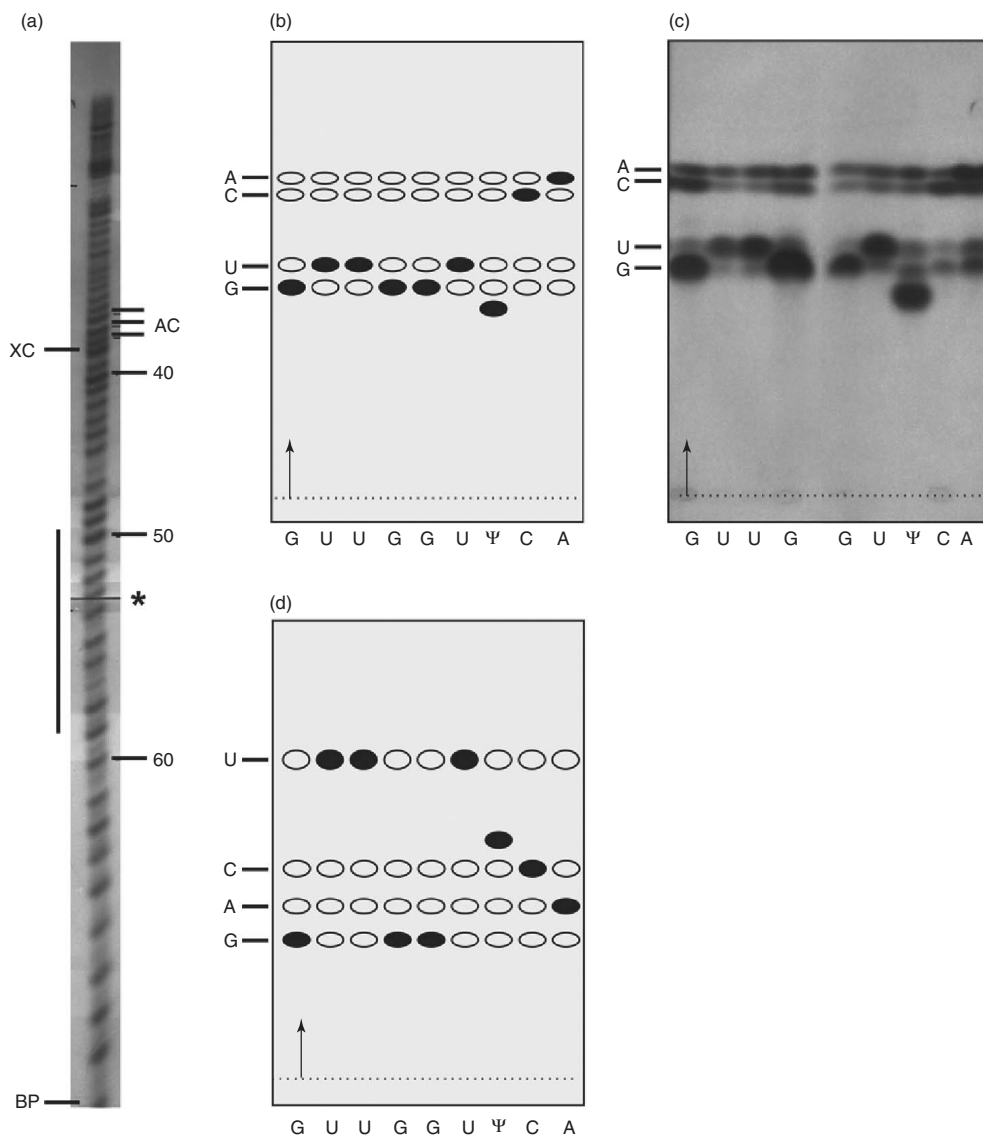


Figure 9.5 Sequencing by position-specific nucleotide analysis. The figure shows part of the sequence analysis of a *Mycoplasma* tRNA^{Ile} [46] by limited RNA hydrolysis and determination of the 5'-end groups of the resulting fragments (Section 9.7.2). (a) Separation of the 5'-labeled fragments obtained by statistical hydrolysis of the tRNA. Electrophoresis was performed on a denaturing 15% PAA gel (80 × 20 × 0.035 cm) for 9 h at 2500 V/80 W; only the lane containing the limited hydrolysis products derived from tRNA^{Ile} is shown. The positions of XC and BPB marker dyes and the anticodon nucleotides (AC) are indicated with horizontal bars, and the numbering of fragments from the top (5'-end of the RNA) is given on the right side of the gel. The T-stem/loop region analyzed on the TLC plates is indicated by a vertical bar at the left side of the gel. Two X-ray films (20 × 40 cm each) were required to cover the whole length of this gel; the position where they overlap is indicated by an asterisk. (b–d) TLC analysis of nucleotides 49–57 covering the T-stem and part of the T-loop. The chromatographic

patterns obtained in solvent A (Section 9.7.1; b,c) and solvent B (d) are shown. (b,d) Schematic drawings of the chromatographic pattern; (c) genuine experimental results corresponding to those sketched in (b). The positions of the internal nucleotide standards are indicated at the left edges of the plates (A, C, G, U stands for the respective 5'-phosphate) and depicted as ovals in the schematic drawings; the sequence is read from left to right. The background of adenosine which is evident in (c) is due to hydrolysis of several pyrimidine-A bonds in single-stranded regions. In general, the presence of these labile bonds leads to accumulation of adenosine end groups in the shorter range of the fragment population due to secondary cleavage. The variable intensities of the sequence signals are due to different labeling efficiencies for purines and pyrimidines. The sequence derived from the TLC analysis shows that the ubiquitous pseudouridine (ψ) is present in the T-loop, but the adjacent ribothymidine is replaced with uridine in this hypomodified tRNA.

different solvent system. A two-dimensional chromatography system with slightly different separation properties has been described in [42]. The closely related nucleotides ac⁴C and m⁵C cannot be distinguished in the solvents described above (Section 9.7.1.1), but can be readily identified by chromatography on PEI plates [39] (Figure 9.4b). Thionucleotides can be identified after modification with CNBr and separation of the products on cellulose plates [47].

9.8

Conclusions and Outlook

The increasing number of genomic sequences has led to the detection of numerous novel RNAs with mostly unknown functions. In many cases, modified nucleotides may play a role in increasing their structural stability, or facilitating specific interactions with proteins or other RNAs; in some cases, editing may even change the primary sequence and coding potential of an RNA. The methods presented here allow the rigorous purification of any desired RNA from biological samples and provide tools for the identification of type and position of modified nucleotides. They may thus help in elucidating the function of these RNAs by identifying novel interaction points with other macromolecules. We anticipate an increasing

application of direct RNA sequencing methods, specifically in the context of future investigations on novel RNA species.

Acknowledgments

Research in our laboratory is supported by grants from the Deutsche Forschungsgemeinschaft, by the European Union and the Free State of Saxony.

References

1. Filipowicz, W. (2000) Imprinted expression of small nucleolar RNAs in brain: time for RNomics. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 14035–14037.
2. Beier, H., Barciszewska, M., Krupp, G.R., Mitnacht, R., and Gross, H.J. (1984) UAG readthrough during TMV RNA translation: isolation and sequence of two tRNAs^{Tyr} with suppressor activity from tobacco plants. *EMBO J.*, **3**, 351–356.
3. Beier, H., Barciszewska, M., and Sickinger, H.D. (1984) The molecular basis for the differential translation of TMV RNA in tobacco protoplasts and wheat germ extracts. *EMBO J.*, **3**, 1091–1096.
4. Schön, A., Krupp, G., Gough, S., Berry-Lowe, S., Kannangara, C.G., and Söll, D. (1986) The RNA required in the first step of chlorophyll synthesis is a chloroplast glutamate tRNA. *Nature*, **322**, 281–284.
5. Schön, A., Kannangara, C.G., Gough, S., and Söll, D. (1988) Protein biosynthesis in organelles requires misaminoacylation of transfer RNA. *Nature*, **331**, 187–190.
6. Yasukawa, T., Suzuki, T., Ishii, N., Ohta, S., and Watanabe, K. (2001) Wobble modification defect in tRNA disturbs codon-anticodon interaction in a mitochondrial disease. *EMBO J.*, **20**, 4794–4802.
7. Behm-Ansmant, I., Urban, A., Ma, X., Yu, Y., Motorin, I., and Branlant, C. (2003) The *Saccharomyces cerevisiae* U2 snRNA: pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA: Ψ-synthase also acting on tRNAs. *RNA*, **9**, 1371–1382.
8. Agris, P.F. (2008) Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep.*, **2**, 629–635.
9. Higa-Nakamine, S., Suzuki, T., Uechi, T., Chakraborty, A., Nakajima, Y., Nakamura, M., Hirano, N., Suzuki, T., and Kenmochi, N. (2012) Loss of ribosomal RNA modification causes developmental defects in zebrafish. *Nucleic Acids Res.*, **40**, 391–398.
10. Oszolak, F. and Milos, P.M. (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet.*, **12**, 87–98.
11. Green, M.R. and Sambrook, J. (2012) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press.
12. Zubay, G. (1962) The isolation and fractionation of soluble ribonucleic acid. *J. Mol. Biol.*, **4**, 347–356.
13. Roe, B.A. (1975) Studies on human tRNA I. The rapid, large scale isolation and partial fractionation of placenta and liver tRNA. *Nucleic Acids Res.*, **2**, 21–42.
14. Heubeck, C.H. and Schön, A. (2001) Cyanelle ribonuclease P: isolation and structure-function studies of an organellar ribonucleoprotein enzyme. *Methods Enzymol.*, **342**, 118–134.
15. Krehan, M., Heubeck, C.H., Menzel, N., Seibel, P., and Schön, A. (2012) RNase MRP RNA and RNase P activity in plants are associated with

- a Pop1p containing complex. *Nucleic Acids Res.*, **40**, 7956–7966. doi: 10.1093/nar/gks476
16. Krupp, G. and Gross, H.J. (1983) Sequence analysis of *in vitro* ³²P-labeled RNA, in *The Modified Nucleotides of Transfer RNA II* (eds P.F. Agris and R.A. Kopper), Alan R. Liss, New York, pp. 11–59.
 17. Kuchino, Y., Hanyu, N., and Nishimura, S. (1987) Analysis of modified nucleosides and nucleotide sequence of tRNA. *Methods Enzymol.*, **155**, 379–396.
 18. Wakita, K., Watanabe, Y.I., Yokogawa, T., Kumazawa, Y., Nakamura, S., Ueda, T., Watanabe, K., and Nishikawa, K. (1994) Higher-order structure of bovine mitochondrial tRNA^{Phe} lacking the 'conserved' GG and TΨCG sequences as inferred by enzymatic and chemical probing. *Nucleic Acids Res.*, **22**, 347–353.
 19. Suzuki, T., Ueda, T., and Watanabe, K. (1996) A new method for identifying the amino acid attached to a particular RNA in the cell. *FEBS Lett.*, **381**, 195–198.
 20. Blum, H., Beier, H., and Gross, H.J. (1987) Improved silver staining of plant proteins. *Electrophoresis*, **8**, 93–99.
 21. Efstratiadis, A., Vournakis, J.N., Donis-Keller, H., Chaconas, G., Dougall, D.K., and Kafatos, F.C., (1977) in End labeling of enzymatically de-capped mRNA. *Nucleic Acids Res.*, **4**, 4165–4174.
 22. Berkner, K.L. and Folk, W.R. (1977) Polynucleotide kinase exchange reaction: quantitative assay for restriction endonuclease-generated 5'-phosphoroyl termini in DNA. *J. Biol. Chem.*, **252**, 3176–3184.
 23. England, T.E., Bruce, A.G., and Uhlenbeck, O.C. (1980) Specific labelling of 3' termini of RNA with T4 RNA ligase. *Methods Enzymol.*, **65**, 65–74.
 24. Beier, H. and Gross, H.J. (1991) Sequence analysis of RNA, in *Essential Molecular Biology II* (ed. T.A. Brown), IRL Press, Oxford.
 25. Lingner, J. and Keller, W. (1993) 3'-end labelling of RNA with recombinant yeast poly(A) polymerase. *Nucleic Acids Res.*, **21**, 2917–2920.
 26. Ledoux, S. and Uhlenbeck, O.C. (2008) 3'-³²P labeling tRNA with nucleotidyl-transferase for assaying aminoacylation and peptide bond formation. *Methods*, **44**, 74–80.
 27. Hausner, T.P., Giglio, L.M., and Weiner, A.M. (1990) Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. *Genes Dev.*, **4**, 2146–2156.
 28. Donis-Keller, H., Maxam, A.M., and Gilbert, W. (1977) Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res.*, **4**, 2527–2538.
 29. Peattie, D.A. (1979) Direct chemical method for sequencing RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 1760–1764.
 30. Waldmann, R., Gross, H.J., and Krupp, G. (1987) Protocol for rapid chemical RNA sequencing. *Nucleic Acids Res.*, **15**, 7209.
 31. Lankat-Buttgereit, B., Gross, H.J., and Krupp, G. (1987) Detection of modified nucleosides by rapid RNA sequencing methods. *Nucleic Acids Res.*, **15**, 7649.
 32. Kuchino, Y. and Nishimura, S. (1989) Enzymatic RNA sequencing. *Methods Enzymol.*, **180**, 154–163.
 33. Silberklang, M., Gillum, A.M., and RajBhandary, U.L. (1979) Use of *in vitro* ³²P-labeling in the sequence analysis of nonradioactive tRNAs. *Methods Enzymol.*, **59**, 58–109.
 34. Stanley, J. and Vassilenko, S. (1978) A different approach to RNA sequencing. *Nature*, **274**, 87–89.
 35. Levy, C.C. and Karpetsky, T.P. (1980) The purification and properties of chicken liver RNase. *J. Biol. Chem.*, **255**, 2153–2159.
 36. Wengler, G., Wengler, G., and Gross, H.J. (1979) Replicative form of Semliki Forest virus RNA contains an unpaired guanosine. *Nature*, **282**, 754–756.
 37. McCloskey, J.A., Whitehill, A.B., Rozenski, J., Qiu, F., and Crain, P.F. (1999) New techniques for the rapid characterization of oligonucleotides by mass spectrometry. *Nucleosides Nucleotides*, **18**, 1549–1553.
 38. Sauerwald, A., Sitaramaiah, D., McCloskey, J.A., Söll, D., and

- Crain, P.F. (2005) Acetyladenosine: a new modified nucleoside from *Methanopyrus kandleri*. *FEBS Lett.*, **579**, 2807–2810.
39. Gupta, R.C., Randerath, E., and Randerath, K. (1976) An improved separation procedure for nucleoside monophosphates on polyethyleneimine-(PEI)-cellulose thin layers. *Nucleic Acids Res.*, **3**, 2915–2921.
 40. Nishimura, S. (1979) in *Transfer RNA: Structure, Properties and Recognition* (eds P. Schimmel, D. Söll, and J. Abelson), Cold Spring Harbor Laboratory Press, pp. 551–552.
 41. Nishimura, S. and Kuchino, Y. (1983) *Methods of DNA and RNA Sequencing*, Praeger, New York, pp. 235–255.
 42. Keith, G. (1995) Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie*, **77**, 142–144.
 43. Grosjean, H., Keith, G., and Droogmans, L. (2004) Detection and Quantification of Modified Nucleotides in RNA Using Thin-Layer Chromatography, in *RNA Interference, Editing, and Modification*, Humana Press, pp. 357–391.
 44. Grosjean, H., Droogmans, L., Roovers, M., and Keith, G. (2007) Detection of enzymatic activity of transfer RNA modification enzymes using radiolabeled tRNA substrates. *Methods Enzymol.*, **425**, 55–101.
 45. Kellner, S., Burhenne, J., and Helm, M. (2010) Detection of RNA modifications. *RNA Biol.*, **7**, 237–247.
 46. Schön, A. (1987) Sequence of an isoleucine tRNA from *Mycoplasma mycoides* sp. capri. *Nucleic Acids Res.*, **15**, 5488.
 47. Saneyoshi, M. and Nishimura, S. (1970) Selective modification of 4-thiouridylylate residue in *Escherichia coli* transfer RNA with cyanogen bromide. *Biochim. Biophys. Acta*, **204**, 389–399.

10

Probing RNA Structure *In Vitro* with Enzymes and Chemicals

Anne-Catherine Helfer, Cédric Romilly, Clément Chevalier, Efthimia Lioliou, Stefano Marzi, and Pascale Romby

10.1

Introduction

In all organisms, RNAs fulfill a diverse range of essential functions, including their role as major regulators of gene expression [1–3]. These functions result from specific RNA tertiary structures, and RNA structural transitions or conformational rearrangements often occur during regulation, leading to different genetic responses [4, 5]. Indeed, the regulatory regions of bacterial mRNAs can sense physical and metabolic changes. These *cis*-acting regions respond to temperature [6, 7], pH [8], divalent ions [9], and the intracellular concentration of a wide variety of *trans*-acting factors, including metabolites [5], uncharged tRNAs [10], proteins [11], and non-coding RNAs [3, 12, 13]. Hence, the biological functions of RNA are linked to the dynamics and versatility of its structure, which promotes the formation of stable or transient intra- or intermolecular interactions. Consequently, there is increasing interest in studies on the structure of both free RNAs and RNAs interacting with ligands.

Although the structure of the whole ribosome, representing one of the largest RNA assemblies in nature, is basically solved [14], analysis of the architecture of long RNA molecules is still a challenging experimental problem. Among the wide array of methods applied to this problem, chemical and enzymatic probing has the obvious advantage that it permits to probe RNA molecules of any size under a wide range of experimental conditions (i.e., varying temperature, pH, or different mono- and divalent ion concentrations). These methods map the reactivity and/or the accessibility of each nucleotide toward enzymes and chemicals and unambiguously define unpaired regions of the RNA. These data are often used as constraints in computer folding programs, facilitating the prediction of RNA secondary structure inferred from the sequence [15, 16]. The resulting secondary structure model can be further validated by site-directed mutagenesis studies coupled with a probing approach to evaluate the structural consequences of each mutation. Chemical probing also provides information regarding the tertiary folding of RNA. Large RNAs are usually composed of stable secondary structural

elements that are brought together by long-distance interactions to form compact domains [17]. Crystallographic studies performed on the ribosome and various regulatory elements of mRNAs have revealed the presence of numerous RNA motifs that either mediate specific interactions to induce compact folding of rRNAs or constitute specific ligand binding sites [18, 19]. These motifs are composed of stacked arrays of non-Watson-Crick base pairs and are characterized by a specific pattern of nucleobase reactivities. Hence, the correlation between X-ray structure and chemical modification patterns can be used to determine the existence of structural motifs in RNAs of unknown tertiary structure. Other chemicals, such as hydroxyl radicals generated by Fe-EDTA, which cleave the ribose [20], and *N*-ethyl-*N*-nitrosourea, which alkylates phosphates [21], are useful to probe the solvent-accessible surface and divalent ion binding pockets of large, structured RNAs. Other methods that probe RNA in a sequence-independent manner have been shown to be powerful in predicting RNA secondary structure and aid in visualizing conformational changes [22–25]. These methods, such as in-line probing [22] or selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) [26, 27], are based on differences in the flexibility of individual internucleotide linkages.

To define the rules that govern the RNA folding process, time-resolved probing techniques have been developed using either base-specific reagents [28] or hydroxyl radicals [29]. With these tools, it is now possible to monitor conformational rearrangements and transient interactions that occur during the RNA folding process or during the assembly of RNA–ligand complexes. As a new development, time-resolved chemical probing in single RNA molecule experiments would be of interest in analyzing the intermediates that form along the folding pathway toward the native and functional RNA state. Indeed, recent single-molecule fluorescence resonance energy transfer experiments have revealed that RNA can fold into multiple distinct states that interconvert [30]. This ability of RNA to adopt different conformations may have an impact on cell biology, perhaps by mediating different responses to environmental changes [31].

Structure-specific chemical probes are also unique tools to map the structure of RNA *in vivo* under different cell growth conditions. The usefulness of probes, however, depends on their ability to penetrate the cell wall of bacteria or membranes. To date, the reagents that have been successfully used for *in vivo* RNA probing are dimethylsulfate (DMS) (e.g., [7]), kethoxal [32], and lead(II) ions [33]. The novel SHAPE approach based on selective 2'-hydroxyl acylation of the ribose units might also be applicable to *in vivo* probing [34]. With the development of high-throughput sequencing analysis combined with lead(II)-induced cleavage *in vivo*, it may now be feasible to obtain structural information on a pool of RNA molecules expressed under specific conditions. Such an approach might also permit to map the binding sites of regulators that target multiple mRNAs. Other technical improvements using capillary electrophoresis combined with fluorescent DNA probes for primer extension allow quantitative analysis of the chemical and enzymatic probing data [35, 36]. This technique, based on DNA sequencing technology, makes it possible to probe very long RNA molecules in one step, such as the entire HIV-1 RNA genome [37].

Other methods such as chemical interference approaches have been derived from footprinting experiments. Chemical interference defines a set of nucleotides that have lost the capability to interact with a ligand when they are modified by a chemical probe (e.g., [38, 39]). In this respect, nucleotide analog interference mapping (NAIM) (Chapters 17 and 18) is a powerful method to investigate the effects of substituting specific nucleotide functional groups on RNA structure or ligand binding [40, 41]. Chemical probes tethered to proteins or RNAs can provide topographical information on ligand–RNA complexes by inducing site-specific cleavages in RNA regions close to the binding interface (e.g., [42, 43]).

In this chapter, an experimental guide is given for several probes commonly used to map the structure of RNAs and RNA–ligand interactions.

10.2

Enzymatic and Chemical Probes

These probes are used under limited conditions in which less than one cleavage or modification occurs per RNA molecule with a statistical distribution. The identification of the cleavages or modifications can be achieved by various methodologies depending on the length of the RNA molecule and the nature of the modification. The first approach uses end-labeled RNA, but it only detects scissions and is limited to RNAs of less than 300 nt. The second approach uses primer extension to detect stops of reverse transcriptase at modified nucleotides or cleavage sites and can be applied to RNAs of any size. Primer extension with reverse transcriptase is usually done by use of a 5′-³²P-end-labeled DNA oligonucleotide, followed by separation of the labeled cDNA fragments with conventional denaturing polyacrylamide gel electrophoresis (PAGE), allowing the identification of modifications/cleavages for 150–200 nt per experiment. A more recent approach uses primer extension with a 5′-fluorescently tagged DNA primer; the cDNA products are then resolved by capillary electrophoresis. Although the classical approach using a radiolabeled primer is sensitive, better quality of the quantification of the nucleotide reactivities/cleavages has been obtained using capillary electrophoresis of fluorescently labeled cDNAs [26, 35, 36]. In addition, modifications or cleavages can be identified over a range of 650–1000 nt per experiment.

The proposed experimental guide has been inferred from structure probing of the regulatory RNAIII from *Staphylococcus aureus* [44, 45] and is confined here to several commonly used probes (Table 10.1). For RNA structure probing *in vivo* in bacterial cells, protocols are also given in Chapter 13, and in [46] for probing RNA structure in eukaryotic cells.

10.2.1

Enzymes

RNase T1 (specific for unpaired guanines), RNase T2 (with preference for unpaired adenines), and RNase V1 (specific for double-stranded regions) are easy to use and

Table 10.1 Structure-specific probes for RNA.

Probes	Molecular weight	Specificity, product	Detection		RNA structure, footprint	In vivo mapping	Considerations of buffers, pH, temperature, etc.
			Direct	Indirect			
Ribonucleases							
RNase T1	11 000	Unpaired G ... Gp3'	+	+	II, III, F	–	Active under a wide range of conditions (temperature range: 4–55 °C, works with or without Mg ²⁺ and salt, and active in the presence of urea)
RNase T2	36 000	Unpaired A >> U,C,G Ap3'	+	+	II, III, F	–	Active under various conditions (temperature range: 4–55 °C and works with or without Mg ²⁺)
RNase V1	15 900	Paired or stacked nucleotide 5'pN ...	+	+	II, III, F	–	Requires magnesium ions and temperature range: 4–50 °C
Chemicals							
DMS	126	A(N1) N1–CH ₃	–	+	II, III, F	+	Reactive at pH 4.5–9 and temperature range: 4–90 °C; tris buffer should be avoided as DMS reacts with amine groups
		C(N3) N3–CH ₃	s	+	II, III, F	+	Reactive at pH 4.5–9 and temperature range: 4–90 °C; tris buffer should be avoided as DMS reacts with amine groups
		G(N7) N7–CH ₃	s	s	II, III, F	+	Reactive at pH 4.5–9 and temperature range: 4–90 °C; tris buffer should be avoided as DMS reacts with amine groups
DEPC	174	A(N7) N7–CO ₂ C ₂ H ₅	s	+	II, III	?	Same comment as for DMS
CMCT	424	G(N1) N1-adduct	–	+	II, III, F		Optimal reactivity at pH 8 and temperature range: 4–90 °C; CMCT is soluble up to 300 mg ml ^{–1} in water

Table 10.1 (Continued)

Probes	Molecular weight	Specificity, product	Detection		RNA structure, footprint	<i>In vivo</i> mapping	Considerations of buffers, pH, temperature, etc.
			Direct	Indirect			
CMCT	424	U(N3) N3-adduct	–	+	II, III, F		Optimal reactivity at pH 8 and temperature range: 4–90 °C; CMCT is soluble up to 300 mg ml ^{–1} in water
ENU	117	Phosphate (–O–CH ₂ –CH ₃)	s	s	III, F*	+	Optimal reactivity at pH 8 and temperature range: 4–90 °C; tris buffers should be avoided as ENU reacts with amine groups
NMIA	177	2'-OH ribose of unpaired nucleotide 2'-O-acylation	–	+	II, III, F*	+	Active under a wide range of conditions, rapidly hydrolyzed in water, has a half-life of 8.5 min in water at 37 °C
1M7	222	As for NMIA	–	+	II, III, F	+	Same comment as for NMIA, has a half-life of 14 s in water at 37 °C
BzCN	117	As for NMIA	–	+	II, III, F	?	Same comment as for NMIA, has a half-life of 0.2 s in water at 37 °C
Hydroxyl radicals and divalent metal ions							
Fe-EDTA (hydroxyl radicals)		Cleavage at ribose (C1', C4')	+	+	II, III, F	?	Reactivity relatively insensitive to buffer composition, variations in pH and temperature, and to the presence of Mg ²⁺ and other cations; avoid sodium phosphate and reduce glycerol to less than 0.5%
Pb(II)-acetate		Unpaired nucleotide and unstable helices phosphodiester cleavage	+	+	II, III, F	+	Formation of precipitates if used as chloride, use instead acetate salt, and avoid acidic pH; optimal pH: 7.5–8.0

(continued overleaf)

Table 10.1 (Continued)

Probes	Molecular weight	Specificity, product	Detection		RNA structure, footprint	<i>In vivo</i> mapping	Considerations of buffers, pH, temperature, etc.
			Direct	Indirect			
In-line		Unpaired nucleotide phosphodiester cleavage	+	+	II, III, F*	?	Works under a wide range of conditions (different salt and Mg ²⁺ concentrations, and different temperatures); optimal pH: 7.5–8.0

DMS, dimethylsulfate; DEPC, diethylpyrocarbonate; CMCT, 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfonate; NMIA, *N*-methylisatoic anhydride; 1M7, 1-methyl-7-nitroisatoic anhydride; BzCN, benzoyl cyanide; ENU, *N*-ethyl-*N*-nitrosourea.

Detection method: (direct) detection of cleavages on end-labeled RNA molecule or (indirect) detection by primer extension with reverse transcriptase using either a ³²P-labeled primer or a primer labeled at the 5'-end with a fluorophore. (+), the corresponding detection method can be used; (s), a chemical treatment is necessary to cleave the ribose-phosphate chain before detection; *in vivo* mapping: (+), probes which diffuse efficiently across cell membranes and walls; (–), probes which have been used only after permeabilization of the cell; (?), probes that might be adaptable to *in vivo* mapping. BzCN is too rapidly destroyed in water to be used *in vivo*. Caution should be taken with DEPC as it strongly modifies proteins with a potential effect on their activities. The molecular weight (MW) of the probe, its specificity, and generated products are indicated. Most of the probes provide information useful to build a secondary structure model (II), to decipher tertiary interactions (III), and to map the binding sites (footprints) of proteins, other RNAs, and metabolites (F). F*: owing to the long time of modification, the reagent may be inappropriate for mapping the sites of interactions. However, in-line probing has been a useful tool to visualize the conformational changes of RNAs upon metabolite binding (see text).

are able to identify secondary structure RNA elements such as hairpins. RNase V1 is the only probe that provides positive evidence of the existence of helical regions. Owing to their size, however, the RNases are sensitive to steric hindrance. RNase cleavages may potentially induce conformational rearrangements in the RNA that provide new targets for the enzyme, resulting in secondary cuts. Such secondary effects are especially observed when cleavage reactions have been too extensive, for example, owing to high RNase concentrations.

10.2.2

Base-Specific Chemical Probes

Base-specific reagents modify specific positions on bases. The combination of DMS and 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate (CMCT) allows probing of the four bases at one of their positions involved in Watson–Crick base pairing (Table 10.1). DMS methylates the N1 of adenines and, to a lesser extent, N3 of cytosines. CMCT modifies the N3 of uridines and, to a lesser extent, the N1 of guanines ([47] and references therein).

The N7 of purines can be probed by diethyl pyrocarbonate (DEPC) for adenines, and by DMS or the so-called nickel complex for guanines. The nickel complex [48] and DEPC [49] are very sensitive to the stacking of the base rings. Thus, the N7 of purines is unreactive within a helix unless the deep groove of the helix is widened.

10.2.3

Backbone-Specific Chemical Probes

A new generation of chemicals has been designed that modify 2'-OH groups of ribose units. These reagents include *N*-methylisatoic anhydride (NMIA) [26], 1-methyl-7-nitroisatoic anhydride (1M7) [37, 50], and benzoyl cyanide (BzCN) [36, 51]. All of them react in a very similar way, but their half-lives differ, decreasing from 8.5 min (NMIA) to 14 s (1M7) and 0.2 s (BzCN) (Table 10.1). All these reagents are highly sensitive to the local flexibility of the nucleotide backbone, which is required for the formation of a 2'-O-adduct between the 2'-OH group of ribose units and the hydroxyl-selective electrophile [51]. Hence, these reagents unambiguously identify unpaired nucleotides. Although purine residues are slightly more reactive than pyrimidines, it has been shown that these differences are subtle in contrast to the reactivity differences between paired and unpaired nucleotides [50]. These ribose modifications can only be detected as stops in reverse transcription reactions.

In-line probing exploits the fact that RNA phosphodiester linkages are spontaneously cleaved owing to a nucleophilic attack of the 2'-OH group, resulting in the formation of a transient 2', 3'-cyclic phosphate. It has been proposed that the maximum rate for this spontaneous cleavage is achieved when the 2'-oxygen, the phosphorus center, and the 5'-oxygen leaving group form a perfect 180° angle (see Chapter 13, Figure 13.1 therein). Thus, the ribose–phosphate linkages that participate in stable Watson–Crick base pairs exhibit lower rates of spontaneous cleavage than unpaired nucleotides [24, 52]. The in-line probing procedure involves a long incubation (>40 h) of the RNA in a basic buffer containing monovalent and divalent ions. This simple approach has been used to monitor the conformational states of a tRNA in solution and in crystals [25], and to visualize the conformational changes and structures of riboswitches, that is, *cis*-acting regulatory regions of mRNAs that respond to the intracellular concentration of specific metabolites or magnesium (e.g., [5, 9, 53]). The combination of in-line probing and sequence alignment led to the establishment of secondary structure models for several riboswitches, which correlated well with high-resolution X-ray and NMR structures [17, 54].

Experimental Protocols

Probing the conformation of RNAs with different enzymes and chemicals requires the use of defined buffer conditions (e.g., pH, ionic strength, magnesium concentration, and temperature). Indeed, the optimal conditions vary slightly with different probes, and the possibility exists that subtle conformational changes may occur under different incubation conditions (Table 10.1). The probe : RNA ratio

must also be adapted so that the experiments are conducted under limited conditions, that is, less than one cleavage or one modification occurred per molecule. For the initial experiments, different probe concentrations and timescales should be used, which is also necessary when the commercial source of the probes has been changed.

Equipment and Reagents

Equipment

Equipment for denaturing PAGE (Model S2, Gibco BRL) is used to size end-labeled RNA products and labeled cDNA fragments. If the primer extension reactions are sized on capillary gels, instruments from Applied Biosystems (ABI) or Beckman can be used. Eppendorf tubes, tips, and buffers should be sterilized before use.

Chemicals and Enzymes

- CMCT (Fluka, 29469).
- DMS (Acros Organics, 11682-0100).
- DEPC (Sigma-Aldrich, D578).
- BzCN (Sigma-Aldrich, 115959); should be stored in a desiccator at 4 °C.
- Calf intestinal phosphatase (Roche, 713023; 1 U μl^{-1}).
- T4 RNA ligase (New England Biolabs, M0204L; 20 U μl^{-1}).
- Avian myeloblastosis virus (AMV) reverse transcriptase (MP Biochemicals, EMAMV200; 20 U μl^{-1}).
- T4 polynucleotide kinase (Fermentas, EK0031; 10 U μl^{-1}).
- [γ - ^{32}P]ATP (3200 Ci mmol^{-1} ; GE Healthcare).
- [5'- ^{32}P]pCp (3000 Ci mmol^{-1} , Amersham).
- RNase T1 (Fermentas, EN0541; 1000 U μl^{-1}).
- RNase T2 (Life Technologies Invitrogen, 18031-013; 20 U μl^{-1}).
- RNase V1 (Ambion, 2275; 0.1 U μl^{-1}).

Safety Rules for Chemicals

Most of these chemical reagents are potential carcinogens; therefore, until the removal of the first ethanol supernatant, chemical modification procedures are carried out under a fume hood while wearing protective gloves. DMS and kethoxal solutions are discarded in 1 M sodium hydroxide waste, and CMCT in 10% acetic acid waste.

Buffers

The buffer conditions tailored to the study of RNAIII may be varied for other RNAs and ligand interactions.

- *Buffer N1*: 20 mM Tris-HCl pH 7.5, 10 mM MgCl_2 , and 150 mM KCl.
- *Buffer N2*: 50 mM sodium cacodylate pH 7.5, 10 mM MgCl_2 , and 100 mM KCl.
- *Buffer N3*: 50 mM sodium borate pH 8, 10 mM MgCl_2 , and 100 mM KCl.

- *Buffer N4*: 40 mM 3-(N-morpholino)propanesulfonic acid (MOPS) pH 8, 80 mM potassium acetate, and 10 mM MgCl₂.
- *Buffer in-line*: 50 mM Tris-HCl pH 8.3, 100 mM KCl, and 20 mM MgCl₂.
- *Buffer ΔT1*: 20 mM sodium citrate pH 4.5, 1 mM EDTA, 7 M urea, 0.025% xylene cyanol, and 0.025% bromophenol blue.
- *Ladder buffer*: Na₂CO₃/NaHCO₃, 0.1 M each, pH 9.
- *RNA loading buffer*: 0.025% xylene cyanol and 0.025% bromophenol blue in 8 M urea.
- *DNA loading buffer*: 0.025% xylene cyanol and 0.025% bromophenol blue in formamide.
- *RTB buffer*: 50 mM Tris-HCl pH 7.5, 20 mM MgCl₂, and 50 mM KCl.
- *Tris-borate-EDTA (TBE) buffer*: 90 mM Tris-borate pH 8.3 and 1 mM EDTA. All buffers are given as 1 x concentrations, but they are usually prepared as 5 x solutions.
- *RNA template hydrolysis buffer*: 50 mM Tris-HCl pH 7.5, 7.5 mM EDTA, 0.5% SDS, and 3.5 μl of 3 M KOH.

RNA Preparation

The RNA is usually transcribed *in vitro* from a DNA template using T7 RNA polymerase (see e.g., Chapter 1). It is recommended to produce amounts of RNA that are sufficient for the entire series of probing experiments (50–100 μg). RNA transcripts are then separated from the DNA template and excess nucleoside triphosphates (NTPs) by gel filtration [55], monoQ [56], or ion-pairing reversed-phase high-performance liquid chromatography (IP-RPLC) [57]. For large RNAs, a treatment with RNase-free DNase I should be performed to remove the DNA template. Denaturing PAGE is the best strategy to obtain homogeneous RNA [58], but the efficiency of the RNA elution from the gel decreases with the length of the RNA.

For 5'-end-labeling, the RNA is first dephosphorylated at its 5'-end and then labeled using [γ -³²P]ATP and T4 polynucleotide kinase [22, 59] (Chapter 9). The dephosphorylation step can be avoided if transcription is carried out in the presence of, for example, UpG (Sigma-Aldrich) (see also Chapter 1). 3'-End-labeling is performed with [5'-³²P]pCp and T4 RNA ligase [60] using 5 μg of RNA (Chapter 9). The labeled RNAs are purified by electrophoresis on 8% polyacrylamide (PAA) (0.5% bisacrylamide)/8 M urea slab gels, followed by elution from the gel slice in 100 μl of 500 mM ammonium acetate/1 mM EDTA plus 10 μl phenol. The RNA is then phenol-extracted and precipitated with 2.5 volumes of cold ethanol in the presence of 1 μg of glycogen. After two washing steps with 200 μl of cold 80% ethanol, the pellet is redissolved in sterile H₂O (to obtain about 50 000 cpm μl⁻¹). Given that the RNA is often in contact with denaturing reagents during its purification, it is worth making an effort to carry out a renaturation process before the probing experiments.

Enzymatic Cleavage and In-Line probing of End-Labeled RNA

This direct approach utilizes probes that cause RNA hydrolysis. The pattern of generated cleavage fragments is then resolved by denaturing PAGE (Table 10.1).

Enzymatic probing has been adapted for structure probing of the *S. aureus* RNAlII (514 nt) and for the determination of its footprint on its target mRNAs and vice versa. Exemplary experiments are illustrated in Figure 10.1. All reactions were conducted in a total volume of 10 μl . Appropriate dilutions of enzymes were made in buffer N1 (see the section titled “Equipment and Reagents”) just before use. For the study of RNA–ligand footprinting experiments, samples containing constant amounts of the RNA of interest and increasing concentrations of the ligand were incubated for 15 min at 20 °C in the appropriate buffer for optimal binding to allow the complexes to reach equilibrium before addition of the enzymatic or chemical probe. Incubation controls in the absence of the probes were performed with the RNA either free or bound to its ligand (e.g., another RNA or a protein) to detect spontaneous cleavages. In such controls, the RNase and the RNA ligand are replaced with sterile H₂O. Labeled RNA (50 000 cpm μl^{-1}) sufficient for the planned experiments was first denatured in sterile H₂O at 90 °C for 1 min and then cooled on ice for 1 min.

RNase Hydrolysis

- **RNase T1:** Labeled mRNA (1 μl , 50 000 cpm) is renatured in the presence of 5 μl of H₂O and 2 μl of *buffer N1* (5 x concentrated) at 20 or 37 °C for 20 min. 1 μl of total tRNA (2 $\mu\text{g} \mu\text{l}^{-1}$) is then added, and the reaction is started by addition of 1 μl of RNase T1 (0.2 U), followed by incubation for 5 min at 20 °C or 2 min at 37 °C.
- **RNase T2:** The same protocol as for RNase T1 is used, except that the reaction is performed with 1 μl of RNase T2 (0.05 U) for 5 min at 20 °C or 2 min at 37 °C.
- **RNase V1:** The same protocol as for RNase T1 is used, except that reaction is performed with 1 μl of RNase V1 (0.05 U) for 5 min at 20 °C or 2 min at 37 °C.

In initial experiments, at least three different concentrations of the enzymes should be tested: RNase T1 (0.1–0.2–0.5 U), RNase T2 (0.01–0.05–0.1 U), and RNase V1 (0.01–0.05–0.1 U).

Stopping the Reactions

Enzymatic assays are stopped by phenol extraction.

- 1) 40 μl of 0.3 M sodium acetate pH 6 and 50 μl of phenol saturated with chloroform are added to each sample. Samples are mixed for 1 min and then centrifuged (1 min at maximum speed in a desktop centrifuge).
- 2) The aqueous phase is carefully removed and transferred into a new sterile 1.5 ml Eppendorf tube, and 2.5 volumes of cold absolute ethanol (150 μl) is added to precipitate the RNA. After mixing, the samples are left in a dry ice–ethanol bath for 10 min, followed by centrifugation (13 000 rpm at 4 °C for 15 min in a desktop centrifuge).
- 3) The supernatant is discarded and the pellet is washed twice with 200 μl of cold 80% ethanol. After a short centrifugation (13 000 rpm for 5 min at 4 °C), the supernatant is removed, and the pellets are vacuum-dried (no more than 5 min), and dissolved in 6 μl of *RNA loading buffer*.

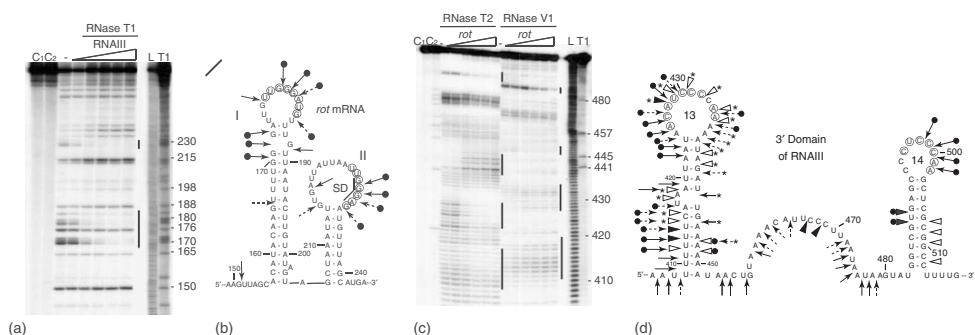


Figure 10.1 Enzymatic probing of *Staphylococcus aureus* *rot* mRNA (a) and the 3'-end domain of RNAIII (b). (a) RNase T1 hydrolysis performed on 5'-end-labeled *rot* mRNA either free (-) or in the presence of increasing concentrations of RNAIII (1, 10, 50, 100, and 250 nM). The concentration of RNase T1 was 0.2 U per assay. Lanes C₁, C₂: incubation controls of *rot* mRNA alone (C₁) and in the presence of RNAIII (50 nM final concentration) (C₂). Lanes L, T₁: alkaline ladder (L) and limited RNase T1 digestion performed under denaturing conditions (T₁). Changes induced by RNAIII binding to *rot* mRNA are shown by bars on the autoradiograph. (b) RNase T1 cuts are shown on the secondary structure model of the regulatory region of *rot* mRNA: medium cut, black arrow; weak cut, dashed arrow. Protections induced by RNAIII binding are indicated by black dots on arrow ends. SD indicates the Shine-Dalgarno sequence. For circled nucleotides, see below. (c) RNase T2 and RNase V1 hydrolysis performed on the 3'-end domain of RNAIII labeled at its 5'-end, either free (-) or in the presence of *rot* mRNA at a concentration of 1, 10, 50, 100, or 250 nM. RNase T2 (0.04 U) and RNase V1 (0.01 U). For further details, see legend to (a). Bars on the autoradiograph represent changes induced by RNAIII binding. (d) Secondary structure of the 3'-end domain of RNAIII: RNase T2, black and dashed arrows for medium and weak cleavages, respectively. RNase V1, black and empty arrowheads for strong and medium cleavages, respectively. The results are taken from Boisset *et al.* [45]. Nucleotides that are circled in the *rot* mRNA and RNAIII are complementary and form loop-loop interactions: the hairpin loop 14 of RNAIII interacts with the hairpin loop II of *rot* mRNA while the hairpin loop 13 of RNAIII interacts with the apical loop I of *rot* mRNA. Black dots on symbols mark protections induced by *rot* mRNA binding to RNAIII, while asterisks indicate cleavage enhancements.

In-Line Probing

In-line probing can be conducted under various conditions. The pH can range from 7.5 to 8.5, and the time of incubation should decrease if the temperature is increased. In-line probing can also be used to study magnesium-dependent effects on RNA folding [61].

- 1) Labeled mRNA (75 000–100 000 cpm) is incubated with 2 μ l of 5 x *buffer in-line* in a total volume of 10 μ l at 20 °C for 40 h.
- 2) The reaction is stopped by adding 10 μ l of *RNA loading buffer* and samples are frozen at –20 °C.
- 3) The in-line probing reaction and an aliquot of the same RNA not incubated in *buffer* (control) are loaded next to each other onto a denaturing PAA gel.

Fractionation of End-Labeled RNA Fragments

RNase T1 and alkaline ladders are required to identify the cleavage positions.

- RNase T1 ladder: Labeled mRNA (1 μ l, 50 000 cpm) is preincubated at 50 °C for 5 min in 5 μ l of *buffer* Δ T1 containing 1 μ g of total tRNA. The reaction is then carried out at 50 °C for 10 min in the presence of 1 μ l of RNase T1 (0.5 U).
- Alkaline ladder: Labeled mRNA (2 μ l, 100 000 cpm) is incubated at 90 °C for 3 min in the presence of total tRNA (2 μ g) in 5 μ l of *ladder buffer*. The duration of the hydrolysis reaction must be adapted according to the length of the RNA.

The end-labeled RNA fragments are sized by electrophoresis on 12% or 15% PAA (0.5% bisacrylamide)/8 M urea slab gels (0.5 mm $\times$ 30 cm $\times$ 40 cm) in 1 x TBE. Gels should be prerun (30 min at 75 W) and run warm (75 W) to avoid band compression. The migration conditions must be adapted to the length of the RNA, given that on a 15% gel, xylene cyanol migrates as a 39 nt long and bromophenol blue as a 9 nt long RNA fragment. A 15% denaturing PAGE is convenient to collect data on small fragments (1–50 nt length range). For a 250 nt RNA, a shorter and a longer gel run are required to interpret correctly the reactivity of nucleotides of the entire RNA molecule. At the end of the run, the 12% gel is fixed for 5 min in a 10% ethanol/6% acetic acid solution, transferred to Whatman 3 MM paper, and dried. The 15% gel is transferred without drying on a plastic support and wrapped with a plastic film. Exposure is carried out at –80 °C using an intensifying screen.

Chemical Modifications

Examples of chemical modifications detected for a *S. aureus* mRNA target of the regulatory RNAIII are shown in Figure 10.2. For enzymatic reactions, the same experimental conditions can be used as described above except that the reactions are carried out with 1 pmol of unlabeled RNA. Reactions are carried out in a total volume of 20 μ l, except for the modification of riboses by 1M7, which is carried out in 50 μ l. A control lacking the modification reagent is incubated in parallel under the same conditions to detect pauses of reverse transcriptase due to stable secondary structures and/or spontaneous cleavage. Unlabeled mRNA is first heated in sterile H₂O at 90 °C for 1 min, then cooled on ice for 1 min, and treated as described below.

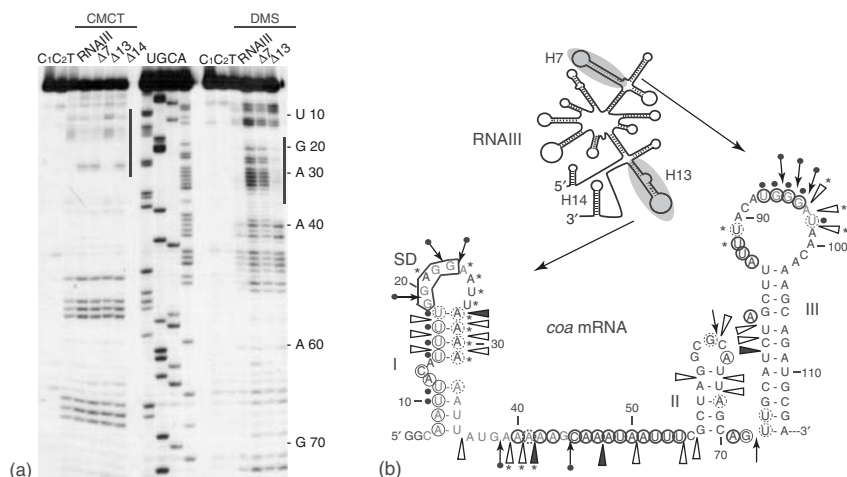


Figure 10.2 Chemical probing of *Staphylococcus aureus coa* mRNA free or bound to RNAIII. (a) Primer extension performed on *coa* mRNA modified by DMS and CMCT. The mRNA was either free (T) or bound to wild-type RNAIII (RNAIII) or various RNAIII derivatives carrying deletions of hairpin 7 ($\Delta 7$), hairpin 13 ($\Delta 13$), or hairpin 14 ($\Delta 14$) at a concentration of 50 nM. Lanes C₁, C₂: incubation controls of the *coa* mRNA either free (C₁) or in the presence of RNAIII (50 nM) (C₂). Lanes U, G, C, and A: DNA sequencing ladders. Note that, relative to sequencing ladders, reverse transcriptase stops in the other lanes occur one residue before the site of cleavage or modification. Further experimental details are given in the text. (b) A summary of the probing data shown in the context of the secondary structure of the first 114 nucleotides of *coa* mRNA: circled nucleotides are reactive toward DMS (N1A >> N3C) or CMCT (N3U >> N1G); circles with solid and dashed lines indicate moderate and low reactivity, respectively. Enzymatic cleavages (gel images not shown) are specified as follows: black arrows, RNase T1 cleavages; filled and empty arrowheads depict medium and weak RNase V1 cuts, respectively. Black dots, associated with symbols specified above, mark protections from chemical modification or enzymatic cleavage upon complexation with RNAIII; asterisks indicate enhanced reactivity owing to complex formation with RNAIII. A schematic view of the secondary structure

of RNAIII is given at the top, with hairpins 7 and 13 highlighted. The position of the primer used for reverse transcription is complementary to nucleotides A99–A116 of *coa* mRNA. The concomitant presence of double-strand-specific RNase V1 cleavages and of reactivity of nucleotides 10–70 at one of their Watson–Crick positions may indicate the formation of alternative structures in this region [62]. Nucleotides in gray of the *coa* mRNA are complementary to two domains of RNAIII: hairpin 13 of RNAIII binds to the 5′-untranslated region of the mRNA, including the Shine–Dalgarno (SD) sequence and AUG start codon (nucleotide 36–38), while hairpin 7 forms a loop–loop interaction with hairpin loop III of the *coa* mRNA. The data are adapted from Chevalier *et al.* [62]. In several experiments, additional protection at the N1 positions of A40–44 of *coa* mRNA induced by RNAIII binding was also observed [62]. It is worth noting that on the autoradiograph showing the DMS modifications, there are fewer labeled cDNA products in the lane corresponding to the free mRNA (T) compared to the other lanes. Although several explanations might be possible, it is likely that a fraction of the RNA was lost during the treatment before primer extension. The modification levels in lane T are expected to be very similar to those in lane $\Delta 13$ as the specific *coa* mRNA–RNAIII interaction is abolished in the case of the RNAIII $\Delta 13$ variant.

DMS Modification (N3C, N1A)

- 1) 1 pmol of mRNA (per assay) is renatured by incubation at 37 °C for 15 min in *buffer N2* in a total volume of 15 µl.
- 2) When probing mRNA–RNAIII interactions, increasing concentrations of RNAIII (1–250 nM in 1–3 µl) are added to the renatured mRNA, allowing complex equilibria to form by incubation at 37 °C for 15 min.
- 3) 1 µl of tRNA (2 µg µl⁻¹) and 1 µl of pure DMS or DMS freshly diluted in ethanol at 1 : 2 (1 + 1), 1 : 5, or 1 : 10 is added and the tubes are gently mixed. The reaction is performed in a total volume of 20 µl at 37 °C for 5 min. The optimal chemical modification is usually obtained with DMS diluted at 1 : 10 (Figure 10.2).

DMS Modification (N7G)

- 1) The same conditions are used as for modifications of N3C and N1A except that the reaction is carried out for 10 min at 37 °C.
- 2) After ethanol precipitation (see first step of the section “Stopping the Reactions” as part of “Enzymatic Cleavage and In-Line Probing of End-Labeled RNA”), the pellet is dissolved in 10 µl of 1 M Tris-HCl pH 8.3 and 10 µl of 8 mg ml⁻¹ sodium borohydrate (dissolved in H₂O extemporaneously). The reaction is carried out in the dark and on ice for 10 min, followed by ethanol precipitation of the modified RNA.
- 3) The RNA pellet is dissolved in 10 µl of aniline (100 µl of bidistilled aniline (Fluka), 60 µl acetic acid, and 930 µl H₂O) and incubated at 60 °C for 10 min in the dark. For further steps, see section “Stopping the Reactions” below.

CMCT Modification (N3U, N1G)

- 1) 1 pmol of mRNA (per assay) is first incubated at 37 °C for 10 min in *buffer N3* in a total volume of 15 µl.
- 2) When probing mRNA–RNAIII interactions, increasing concentrations of RNAIII (1–250 nM in 1–3 µl) are added to the renatured mRNA, allowing complex equilibria to form by incubation at 37 °C for 15 min.
- 3) 1 µl of tRNA (2 µg µl⁻¹) and 4 µl of CMCT (40 or 60 mg ml⁻¹ dissolved in water just before use) are added, and the samples are gently mixed. The modification is carried out in a total volume of 20 µl at 37 °C for 10 min or at 20 °C for 20 min. The optimal modification of mRNA is seen with CMCT at 40 mg ml⁻¹ (Figure 10.2).

DEPC Carboxylation (N7A)

- 1) 1 pmol of mRNA (per assay) is first renatured by incubation at 37 °C for 15 min in *buffer N2* in a total volume of 15 µl.
- 2) When probing mRNA–RNAIII interactions, increasing concentrations of RNAIII (1–250 nM in 1–3 µl) are added to the renatured mRNA, allowing complex equilibria to form by incubation at 37 °C for 15 min.

- 3) 1 μl of tRNA ($2\text{ }\mu\text{g }\mu\text{l}^{-1}$) and 4 μl of pure DEPC are added, and the tubes are mixed gently. The reaction is carried out at 37°C for 15 min.

Ribose Acetylation with 1M7

- 1) 2 pmol of RNA (per assay) are renatured in *buffer N4* at 20°C for 15 min in a total volume of 18 μl .
- 2) The reagent 1M7 is not commercially available but has been chemically synthesized by Mortimer and Weeks [23]. It undergoes rapid hydrolysis in buffer such that the reaction is complete after 70 s. Therefore, this reagent should be stored in a desiccator at 4°C . 1M7 must be diluted with anhydrous DMSO to obtain an 80 mM solution so that the reagent remains stable and can be stored at 20°C in a dessicator.
- 3) The sample is split into two tubes (9 μl each). Then, 1 μl of 80 mM 1M7 is added to one tube, and 1 μl of DMSO is added to the other tube (incubation control). The contents of the tubes are mixed gently, and the reaction is performed at 37°C for 70 s.
- 4) The optimal concentration of 1M7 can vary with RNA length. Thus, it is recommended to determine the concentration dependence ranging from 2 to 20 mM, with the typical optimum concentration being 8 mM. No quenching step is required.

Ribose Acetylation with Benzoyl Cyanide (BzCN)

- 1) 2 pmol of RNA (per assay) are renatured in *buffer N4* at 20°C for 15 min in a total volume of 20 μl .
- 2) BzCN solutions should be freshly prepared for each experiment. BzCN is dissolved in DMSO to obtain a $10\times$ (400 mM) stock solution.
- 3) To each RNA sample, 2 μl of 400 mM BzCN are added, and the tube contents are mixed gently. The reaction is performed at 37°C for 1 s or at 20°C for 3 s. Five microliters of DMSO are added to the control sample instead of BzCN. As for 1M7, it is also recommended to determine the BzCN concentration dependence in the range from 10 to 80 mM final concentration, with the typical optimum concentration being 40 mM. No quenching reaction is required [36].

Stopping the Reactions

All reactions are stopped by ethanol precipitation of the RNA. For RNA–protein footprinting experiments, the protein is removed by phenol extraction before ethanol precipitation.

- 1) First, 50 μl of 0.3 M sodium acetate pH 6 and 250 μl of cold absolute ethanol are added to each sample. The samples are then mixed, placed in a dry ice/ethanol bath for 15 min, and centrifuged (13 000 rpm $\approx$ 10 000 g) at 4°C for 15 min.

- 2) The supernatants are removed carefully (without touching the pellets), and 200 μl of cold 80% ethanol are added to each pellet. The samples are centrifuged (13 000 rpm at 4 °C for 5 min), and the supernatants are removed with the same caution. This step is repeated once.
- 3) The pellets are vacuum-dried (not more than 5 min) and dissolved in 4 μl of sterile H_2O .

Primer Extension Analysis

Primer extension with reverse transcriptase was originally developed to probe the structure of ribosomal RNAs [63]. Reverse transcriptase stops its incorporation of dNTPs at the residue preceding a cleavage or a modification at a Watson–Crick position. While carbethoxylation of A(N7) by DEPC is sufficient to stop reverse transcriptase, subsequent treatment is necessary to induce cleavage at G(N7) after DMS modification (Table 10.1).

The length of the primer varies, but 12- to 18-mers typically provide sufficient specificity, even if the primers are used on a mixture of RNAs. For long RNAs, primers are selected every 200 nt because of gel resolution limits. To obtain efficient hybridization and primer extension, pilot assays should be performed in advance to define the ideal RNA concentration, the choice of the primer sequence, and the hybridization conditions.

In most of the cases, ^{32}P -labeled DNA oligonucleotides are used for primer extension followed by separation of the labeled cDNA fragments by denaturing PAGE. The experimental conditions for this approach are given below. As an alternative strategy, 5'-fluorescently labeled DNA primers have also been used for primer extension, and the generated cDNA fragments have been fractionated using capillary electrophoresis and a classical DNA sequencing apparatus. For these experiments, at least three DNA primers labeled with a different fluorophore are required. Hence, the control, experimental, and sequencing reactions can be loaded on the same capillary gel. Detailed protocols have been published previously [35, 36].

Primer Hybridization and Extension

- 1) First, 1 μl of 5'- ^{32}P -end-labeled DNA primer (around 100 000 cpm) is added to 4 μl of modified mRNA (1–5 pmol). The samples are then heated at 90 °C for 1 min and quickly cooled on ice. Annealing of the primer can also be done by incubating the reaction at 65 °C for 5 min and then quickly placing it on ice for 1 min.
- 2) 1 μl of 5 x *RTB buffer* is added and the samples are incubated for 20 min at 37 °C.
- 3) The primer extension is carried out in a total volume of 15 μl ; 2 μl of 5 x *RTB buffer*, 2 μl of a dNTP mix (2.5 mM of each dNTP), 4 μl sterile H_2O , and 1 μl of AMV reverse transcriptase (2 U μl^{-1} freshly diluted in sterile H_2O) are added to the 6 μl hybridization mix from step 2. Primer extension is done at 37 °C for 30 min.

- 4) To improve gel resolution, RNA templates are hydrolyzed by alkaline treatment. Just after primer extension, 20 μ l of *RNA template hydrolysis buffer* is added (see section “Buffers” as part of “Equipment and Reagents”). The samples are incubated at 90 °C for 3 min and then at 37 °C for at least 1 h.
- 5) 6 μ l of 3 M acetic acid, 100 μ l 0.3 M sodium acetate pH 6, and 300 μ l of cold absolute ethanol are added to each sample. Samples are then mixed, placed in a dry ice/ethanol bath for 15 min, and centrifuged (13 000 rpm at 4 °C for 15 min).
- 6) The supernatants are carefully removed, and the pellets are washed twice with 200 μ l of cold 80% ethanol. After centrifugation (13 000 rpm at 4 °C for 5 min), the supernatants are removed, and the end-labeled DNA fragments are redissolved in 6 μ l of *DNA loading buffer*.

Gel Fractionation

The DNA fragments are denatured by incubating the samples at 90 °C for 3 min, followed by fractionation on 8% PAA (0.4% bisacrylamide)/8 M urea slab gels using 1 x TBE as electrophoresis buffer. As described above, the gels should be prerun (30 min at 75 W) and run warm (75 W). The migration conditions must be adapted to the size of the fragments being analyzed, given that on an 8% gel, xylene cyanol migrates as an 81 nt and bromophenol blue as a 19 nt RNA fragment.

The modification or cleavage positions are identified by running a sequencing reaction in parallel. The elongation step for sequencing reactions is performed as described above, except that one of the four nucleotides is present as a mixture of the dNTP (25 μ M) and ddNTP (2.5 μ M), while the other three nucleotides are provided as dNTPs only (100 μ M each). After migration, the gels are dried and subjected to phosphorimaging or autoradiography overnight at –80 °C with an intensifying screen.

Quantification of the Data

Chemical mapping provides information at the single nucleotide level; therefore, accurate gel quantification represents a key step in the analysis of the chemical reactivity data. A protocol has been recently published using the Semi-Automated Footprinting Analysis (SAFA) software [64]. This protocol permits accurate quantification of each band of an autoradiography scan. It includes several steps, namely, lane identification and band assignment, model fitting, and band-intensity normalization. This method was first developed to analyze hydroxyl radical footprints [64], but has also been used to quantify ribose modifications by 1M7 to probe RNA structure and protein footprints [65].

ShapeFinder (<http://bioinfo.unc.edu>) has been used to quantify the reactivity of each 2' hydroxyl of the ribose units toward NMIA, 1M7, or BzCN following high-resolution capillary electrophoresis [16, 66]. ShapeFinder quantifies the reactivity of each nucleotide on the basis of fluorescence intensity. To accurately predict the structure of long RNAs, the reactivity value of each ribose is considered to be a pseudo-free energy change value [67]. These experimental values are introduced

into a dynamic programming algorithm for the prediction of RNA secondary structure [68]. Using a combination of free energy minimization (reactivity of riboses) and nearest neighbor parameters, this strategy allows accurate prediction of the structure of long and short RNAs, with a predictive power similar to that of comparative sequence analysis [67].

10.3

In Vivo DMS Modification

10.3.1

Generalities

This reagent is capable of diffusing efficiently across the cell wall and membrane of prokaryotic and eukaryotic cells to modify unpaired adenines (N1), cytosines (N3), and guanines (N7). The protocol given here was used to probe the structure of *S. aureus* RNAIII *in vivo* [44]. *S. aureus* cells were grown to the mid-logarithmic phase and then treated with DMS. As for *in vitro* experiments, it is important to verify that the reaction occurred under limited conditions, such that less than one modification per molecule was statistically induced. Thus, a range of DMS quantities (100 μ l of DMS diluted 1 : 10, 1 : 5, 1 : 2, or pure) and time intervals for incubation (2–15 min) were initially tested. After treatment, the reaction was stopped just before total RNA extraction. Sites of DMS modification were detected by primer extension on total RNA extracts (5–20 μ g), using end-labeled primers specific for a chosen region of interest of the tested RNA.

10.3.2

In Vivo Probing

- *Cold stop buffer*: 100 mM Tris-HCl pH 8.0, 200 mM β -mercaptoethanol, and 5 mM EDTA; kept at 4 °C.
 - *Cold buffer*: 10 mM Tris-HCl pH 8, 100 mM NaCl, and 1 mM EDTA; kept at 4 °C.
 - *S. aureus lysis buffer*: 50 mM Tris-HCl pH 8.0, 50 μ g ml⁻¹ lysostaphine (Sigma-Aldrich), and 1% SDS.
- 1) Bacteria (20 ml of culture) are grown in LB medium in a 50 ml sterile tube to mid-logarithmic phase at 37 °C (until an OD₆₀₀ of 0.5 is reached).
 - 2) About 100 μ l of DMS (diluted 1 : 2 in ethanol) are added and the culture is incubated for an additional 5 min at 37 °C under gentle shaking.
 - 3) The reaction is stopped by adding 10 ml of *cold stop buffer*.
 - 4) The cells are pelleted (3000 g for 15 min at 4 °C), followed by resuspension in 1.5 ml of *cold buffer*. After transfer to a 1.5 ml Eppendorf tube, cells are centrifuged again at 13 000 g for 15 min at 4 °C.
 - 5) *S. aureus* cells are then resuspended in 200 μ l of *S. aureus lysis buffer* and incubated for 30 min at 37 °C. Cell debris are eliminated by centrifugation at

- 13 000 g for 15 min at 4 °C, and the supernatant is transferred to a new 1.5 ml Eppendorf tube.
- 6) 200 µl of phenol saturated with 0.1 M sodium acetate pH 5.5 and 10 mM EDTA are added. Cells are vortexed for 30 s at high speed. Samples are heated to 65 °C for 15 min, with mixing every 5 min in between.
 - 7) The mixture is cooled on ice and centrifuged for 10 min at 13 000 g at room temperature. The aqueous phase is carefully collected, and the phenol phase and interphase are re-extracted by vortexing the samples with 100 µl 0.1 M sodium acetate pH 5.5.
 - 8) After centrifugation, the aqueous phases are pooled and extracted once with phenol/chloroform (50/50%) previously saturated with sodium acetate (0.1 M, pH 5.5), and once with chloroform.
 - 9) The RNA is then precipitated twice with three volumes of cold ethanol in the presence of 0.3 M sodium acetate pH 5.5 (final concentration). The pellet is washed twice with 200 µl of 80% ethanol, vacuum-dried (no more than 5 min), and dissolved in a small volume of sterile H₂O. The RNA concentration is measured and 10 µg of material is used for primer extension.
 - 10) Primer hybridization and elongation by reverse transcriptase is performed as described above (see the section “Primer Extension Analysis”), except that elongation is conducted at 45 °C for 30 min with 5 U of reverse transcriptase. Sequencing reactions are prepared in parallel (see the section “Primer Extension Analysis”), which help to map the sites of modification.
 - 11) An incubation control is prepared from cells grown and treated under the same conditions as above but in the absence of DMS.
 - 12) In another control (stop control), step 2 above is omitted and DMS is added after the addition of the *cold stop buffer* (i.e., between steps 3 and 4 above) to exclude that any DMS modification occurred during steps 4–9.

10.4 Commentary

10.4.1 Critical Parameters

10.4.1.1 RNA Preparation

The RNA is usually synthesized by *in vitro* transcription using T7 RNA polymerase, therefore, the 3'- or 5'-end of the RNA might be heterogeneous (see Chapter 1). The incorporation of ribozymes at the 5'-and/or 3'-end of transcripts is one solution to generate homogenous ends ([69]; see Chapter 2). Furthermore, several abortive transcription products might accumulate. According to our experience, the method of choice to obtain homogeneous RNA molecules remains fractionation of transcripts by denaturing PAGE. For long RNA molecules, electroelution (Chapter 1) might help to increase the elution efficiency.

10.4.1.2 Homogeneous RNA Conformation

During purification, the RNA can be partially denatured. It is thus essential to design renaturation protocols to generate a “conformationally” homogeneous RNA population and to test whether this conformation is biologically relevant (enzymatic activity for ribozymes, or efficient ligand binding). The simultaneous presence of cleavages by single-strand- and double-strand-specific RNases may indicate the coexistence of alternative RNA conformations. If the conformers have different electrophoretic mobilities on a native PAA gel, chemical probing can be used to distinguish them. After chemical modification, the coexisting structures are separated by native PAGE, and the modification sites for each conformer are identified by primer extension [70].

10.4.1.3 Chemical and Enzymatic Probing

- The chemical reactions and RNase T1 hydrolysis can be conducted under a variety of experimental conditions. For instance, the influence of divalent ions (such as magnesium) on the folding of the RNA can be tested, and by varying the temperature (between 4 and 90 °C), one can follow thermal transitions of the RNA [71]. It is essential to adjust the extent of modification or enzymatic hydrolysis for each condition to less than one cut or modification per molecule (i.e., more than 80% of the RNA should not be modified or cleaved). For example, for DMS modification, the reaction is carried out at 4 °C for 20 min in the presence of 1 µl of DMS, but for 5 min with 1 µl of DMS diluted 1 : 16 at 50 °C.
- RNase cleavages can induce conformational rearrangements that are potentially able to provide new targets for secondary cleavages. These secondary cleavages occur when the reaction is too extensive, but they are usually of weak intensity and are not reproducibly found in all experiments. These cleavages can be distinguished from primary cuts by comparing the hydrolysis patterns obtained with 5'- versus 3'-end-labeled RNA.
- In contrast to alkaline treatment, RNase V1 hydrolysis generates RNA fragments with a 3'-OH group. Therefore, 5'-end-labeled fragments generated by alkaline treatment will migrate faster than fragments derived from cleavage by RNase V1 at the same position. Conversely, 3'-end-labeled fragments generated by RNase V1 will migrate faster than those generated by alkaline hydrolysis of the same phosphodiester linkage. This difference is essentially only detectable for the shortest RNA fragments.
- Appropriate incubation controls are required to identify cleavages or reverse transcriptase stops (e.g., due to stable secondary structures) that appear independent of the presence of the probe. Nucleotides for which strong bands are visible in the control lanes are not considered in the interpretation. For primer extension of structured RNAs, SuperScript III reverse transcriptase (Life Technologies Invitrogen, ref. no. 18080-093) may be used, as the enzyme is still active at 55 °C [36].
- Each experiment should be repeated several times, and only the reproducible cleavages and modifications should be considered in the interpretation.

- Footprinting experiments should be done in the presence of increasing concentrations of the ligand (Figure 10.1). Lead(II)-ions, hydroxyl radicals, and 1M7 or BzCN are appropriate probes to map ligand binding sites because of their small size and therefore low sensitivity to higher-order RNA structure that can impede accessibility in the case of large probes such as RNases. In-line probing is perhaps less appropriate for protein footprinting as the time of incubation is quite long. Chemicals such as DEPC react strongly with proteins; therefore, it is important to avoid the use of such compounds in footprinting assays. Results should be interpreted with care as decreased reactivity does not necessarily result from a direct shielding effect but could be due to an indirect steric hindrance effect (particularly observed with the bulky RNases) or a conformational change in the RNA.
- For the acylation of riboses, only NMIA and BzCN are commercially available (Sigma-Aldrich or Life Technologies Invitrogen). These reagents must be stored in a desiccator at 4 °C. Owing to its short reaction time, BzCN is particularly appropriate to study the RNA folding process. For NMIA, detailed protocols have already been described [22, 26]. One microliter of NMIA (30, 65, and 140 mM dissolved in DMSO) is added to 9 µl of *buffer N4* containing 1 pmol of renatured RNA. An incubation control is performed in the presence of DMSO. The samples are then mixed and incubated at 37 °C for 45 min.

10.4.1.4 *In Vivo* DMS Mapping

- Owing to the inability of many compounds to enter and freely diffuse within cells, only a few probes have been successfully used to map RNA structure intracellularly (Table 10.1). DMS and lead(II)-ions are to date the most commonly used probes. Other probes have been used *in vivo* after permeabilization of cells [32]. However, because of this additional treatment, particular caution has to be taken to ensure that the cells remain intact during the time of incubation. It is also essential to verify that the reaction was efficiently stopped before the RNA extraction procedure.
- Data from *in vivo* probing may be more complex to interpret than those obtained *in vitro*. One of the main reasons is that the studied RNA may be simultaneously engaged in several complexes (e.g., regulatory RNAs). However, *in vivo* mapping becomes powerful when it is used in a comparative manner. For example, conformational changes of mRNA induced by a *trans*-acting ligand may be identified. The mRNA structure may also be compared under different growth conditions (e.g., varying temperature) or when certain environmental cues are taking effect.

10.5

Troubleshooting

- Compression of bands due to stable secondary structures (helical regions rich in G–C base pairs). The end-labeled RNA samples can be heated at 90 °C for 3 min

before gel loading. The alkaline ladder and the RNase T1 ladder, however, should never be heated.

- Too many bands in the incubation controls of end-labeled RNA. This problem is likely due to RNase contamination. The RNA should be purified again by denaturing PAGE, and new sterile buffers should be prepared.
- Radiolysis of end-labeled RNAs. This can also be a source of noise, particularly for in-line probing. The end-labeled RNA must be kept in water with an activity below $200\,000\text{ cpm }\mu\text{l}^{-1}$. It is sometimes better to perform a short gel elution of the end-labeled RNA (several hours at 4°C) and to store the gel slice at -20°C for further elution.
- Aggregation of cDNA fragments in the gel pockets (only smaller fragments can be resolved). Degradation of the RNA template after primer extension prevents such gel migration artifacts.
- Samples not migrating correctly during electrophoresis: This complication is probably due to the presence of salt. Several washing steps with 80% ethanol should be included before gel loading.
- No full-length RNA after modification or enzymatic hydrolysis. The conditions should be adjusted by reducing the amount of probe to obtain more than 80% unmodified RNA molecules.
- Very weak signals for modifications or cleavages in spite of an intense full-length product being observed. This problem reflects insufficient modification or hydrolysis. It is always useful to determine the concentration dependence of the probes in order to find optimal conditions for probing. The preparation of the probe (enzyme dilution or preparation of working solutions of chemical probes) must be done just before the experiments. Enzymes must be diluted in the reaction buffer, and chemicals (DMS, DEPC, 1M7, and BzCN) must be stored under particular conditions.
- Absence of a signal after primer extension. It is likely that the modified RNA did not precipitate efficiently. As the RNA is not labeled, particular caution should be taken to prevent loss of the pellets after ethanol precipitation. The concentration of the RNA can also be increased (2–5 pmol per assay). Another explanation could be the quality of the reverse transcriptase, which is sensitive to both MgCl_2 concentration (3 mM is the optimal concentration) and freezing. Optimal conditions for primer hybridization should be established in a series of pilot experiments. The optimal temperature for annealing varies from RNA to RNA, depending on the G+C content, the propensity of the RNA to form stable secondary structures, length of primer, and stability of primer–target duplexes.
- No proper extension or weak bands in the sequencing lanes. The concentration of the ddNTP and the ratio of ddNTP/dNTP should be adjusted.
- Unsatisfactory gel resolution. Acrylamide, urea solutions, and particularly ammonium persulfate should be freshly prepared.
- Low yield of total RNA prepared for *in vivo* probing. Incomplete homogenization or lysis of cell samples or degradation of the RNA.

- Strong stops in the control lanes of the *in vivo* probing experiment. Degradation of RNA, or pauses of reverse transcriptase due to stable secondary structures (perform elongation at higher temperature, increase the concentration of enzyme and dNTPs, or change the primer sequence). Many RNA molecules carry post-transcriptional modifications that may interfere with reverse transcriptase elongation (primer should be changed in order to overlap and thus mask the modified base).

Acknowledgments

We thank P. Fechter and I. Caldelari for helpful discussions. We are grateful to E. Westhof for his constant support and helpful advice. This work was supported by the Centre National de Recherche (CNRS), the Agence Nationale de la Recherche (ANR-07-BLAN-0351-01 and ANR-09-BLAN-0024-01). EL received support from FEBS, CR from the Conseil Régional Alsace and the Délégation Générale de l'Armement, and CC from the Région Alsace.

References

1. Caron, M.P., Lafontaine, D.A., and Masse, E. (2010) Small RNA-mediated regulation at the level of transcript stability. *RNA Biol.*, **7**, 140–144.
2. Mattick, J.S., Taft, R.J., and Faulkner, G.J. (2010) A global view of genomic information—moving beyond the gene and the master regulator. *Trends Genet.*, **26**, 21–28.
3. Waters, L.S. and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell*, **136**, 615–628.
4. Dambach, M.D. and Winkler, W.C. (2009) Expanding roles for metabolite-sensing regulatory RNAs. *Curr. Opin. Microbiol.*, **12**, 161–169.
5. Roth, A. and Breaker, R.R. (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.*, **78**, 305–334.
6. Narberhaus, F. (2010) Translational control of bacterial heat shock and virulence genes by temperature-sensing mRNAs. *RNA Biol.*, **7**, 84–89.
7. Giuliodori, A.M., Di pietro, F., Marzi, S., Masquida, B., Wagner, R., Romby, P., Gualerzi, C.O., and Pon, C. (2010) The *cspA* mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Mol. Cell*, **37**, 21–33.
8. Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E., and Altuvia, S. (2009) A pH-responsive riboregulator. *Genes Dev.*, **23**, 2650–2662.
9. Ramesh, A. and Winkler, W.C. (2010) Magnesium-sensing riboswitches in bacteria. *RNA Biol.*, **7**, 77–83.
10. Green, N.J., Grundy, F.J., and Henkin, T.M. (2010) The T box mechanism: tRNA as a regulatory molecule. *FEBS Lett.*, **584**, 318–324.
11. Babitzke, P., Baker, C.S., and Romeo, T. (2009) Regulation of translation initiation by RNA binding proteins. *Annu. Rev. Microbiol.*, **63**, 27–44.
12. Vogel, J. (2009) A rough guide to the non-coding RNA world of Salmonella. *Mol. Microbiol.*, **71**, 1–11.
13. Figueroa-Bossi, N., Valentini, M., Malleret, L., and Bossi, L. (2009) Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target. *Genes Dev.*, **23**, 2004–2015.
14. Frank, J. and Gonzalez, R.L. Jr. (2010) Structure and dynamics of a processive Brownian motor: the translating ribosome. *Annu. Rev. Biochem.*, **79**, 381–412.

15. Reuter, J.S. and Mathews, D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinf.*, **11**, 129.
16. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M., and Giddings, M.C. (2009) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*, **14**, 1979–1990.
17. Serganov, A. (2010) Determination of riboswitch structures: light at the end of the tunnel?. *RNA Biol.*, **7**, 98–103.
18. Lescoute, A., Leontis, N.B., Massire, C., and Westhof, E. (2005) Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res.*, **33**, 2395–2409.
19. Leontis, N.B., Lescoute, A., and Westhof, E. (2006) The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.*, **16**, 279–287.
20. Latham, J.A. and Cech, T.R. (1996) Defining the inside and outside of a catalytic RNA molecule. *Science*, **245**, 276–282.
21. Romby, P., Moras, D., Bergdoll, M., Dumas, P., Vlassov, V.V., Westhof, E., Ebel, J.P., and Giege, R. (1985) Yeast tRNAAsp tertiary structure in solution and areas of interaction of the tRNA with aspartyl-tRNA synthetase. A comparative study of the yeast phenylalanine system by phosphate alkylation experiments with ethylnitrosourea. *J. Mol. Biol.*, **184**, 455–471.
22. Wakeman, C.A. and Winkler, W.C. (2009) Analysis of the RNA backbone: structural analysis of riboswitches by in-line probing and selective 2'-hydroxyl acylation and primer extension. *Methods Mol. Biol.*, **540**, 173–191.
23. Mortimer, S.A. and Weeks, K.W. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.*, **129**, 4144–4145.
24. Soukup, G.A. and Breaker, R.R. (1999) Relationship between internucleotide linkage geometry and the stability of RNA. *RNA*, **5**, 1308–1325.
25. Moras, D., Dock, A.C., Dumas, P., Westhof, E., Romby, P., Ebel, J.P., and Giege, R. (1986) Anticodon-anticodon interaction induces conformational changes in tRNA: yeast tRNAAsp, a model for tRNA-mRNA recognition. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 932–936.
26. Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
27. Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. (2005) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *J. Am. Chem. Soc.*, **127**, 4223–4231.
28. Fabbretti, A., Milon, P., Giuliodori, A.M., Gualerzi, C.O., and Pon, C.L. (2007) Real-time dynamics of ribosome-ligand interaction by time-resolved chemical probing methods. *Methods Enzymol.*, **430**, 45–58.
29. Ralston, C.Y., Sclavi, B., Sullivan, M., Deras, M.L., Woodson, S.A., Chance, M.R., and Brenowitz, M. (2000) Time-resolved synchrotron X-ray footprinting and its application to RNA folding. *Methods Enzymol.*, **317**, 353–368.
30. Solomatin, S.V., Greenfield, M., Chu, S., and Herschlag, D. (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature*, **463**, 681–684.
31. Cruz, J.A. and Westhof, E. (2009) The dynamic landscapes of RNA architecture. *Cell*, **136**, 604–609.
32. Balzer, M. and Wagner, R. (1998) A chemical modification method for the structural analysis of RNA and RNA-protein complexes within living cells. *Anal. Biochem.*, **256**, 240–242.
33. Lindell, M., Romby, P., and Wagner, E.G. (2002) Lead(II) as a probe for investigating RNA structure *in vivo*. *RNA*, **8**, 534–541.
34. Gherghe, C., Leonard, C.W., Gorelick, R.J., and Weeks, K.M. (2010) Secondary structure of the mature ex virio Moloney murine leukemia virus genomic RNA

- dimerization domain. *J. Virol.*, **84**, 898–906.
35. Mitra, S., Shcherbakova, I.V., Altman, R.B., Brenowitz, M., and Laederach, A. (2008) High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res.*, **36**, e63.
 36. Mortimer, S.A. and Weeks, K.M. (2009) Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat. Protoc.*, **4**, 1413–1421.
 37. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W.J., Swanstrom, R., Burch, C.L., and Weeks, K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.
 38. Skripkin, E., Paillart, J.C., Marquet, R., Ehresmann, B., and Ehresmann, C. (1994) Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 4945–4949.
 39. Schlegl, J., Gegout, V., Schlager, B., Hentze, M.W., Westhof, E., Ehresmann, C., Ehresmann, B., and Romby, P. (1997) Probing the structure of the regulatory region of human transferrin receptor messenger RNA and its interaction with iron regulatory protein-1. *RNA*, **3**, 1159–1172.
 40. Ryder, S.P., Ortoleva-Donnelly, L., Kosek, A.B., and Strobel, S.A. (2000) Chemical probing of RNA by nucleotide analog interference mapping. *Methods Enzymol.*, **317**, 92–109.
 41. Waldsich, C. (2008) Dissecting RNA folding by nucleotide analog interference mapping. *Nat. Protoc.*, **3**, 811–823.
 42. Joseph, S. and Noller, H.F. (2000) Directed hydroxyl radical probing using iron(II) tethered to RNA. *Methods Enzymol.*, **318**, 175–190.
 43. Culver, G.M. and Noller, H.F. (2000) Directed hydroxyl radical probing of RNA from iron(II) tethered to proteins in ribonucleoprotein complexes. *Methods Enzymol.*, **318**, 461–475.
 44. Benito, Y., Kolb, F.A., Romby, P., Lina, G., Etienne, J., and Vandenesch, F. (2000) Probing the structure of RNAlII, the *Staphylococcus aureus agr* regulatory RNA, and identification of the RNA domain involved in repression of protein A expression. *RNA*, **6**, 668–679.
 45. Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N., Possedko, M., Chevalier, C., Helfer, A.C., Benito, Y., Jacquier, A., Gaspin, C., Vandenesch, F., and Romby, P. (2007) *Staphylococcus aureus* RNAlII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev.*, **21**, 1353–1366.
 46. Wells, S.E., Hughel, J.M., Igel, A.H., and Ares, M.J. (2007) Use of dimethyl sulfate to probe RNA structure *in vivo*. *Methods Enzymol.*, **318**, 479–493.
 47. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.P., and Ehresmann, B. (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res.*, **15**, 9109–9128.
 48. Chen, X., Woodson, S.A., Burrows, C.J., and Rokita, S.E. (1993) A highly sensitive probe for guanine N7 in folded structures of RNA: application to tRNA(Phe) and *Tetrahymena* group I intron. *Biochemistry*, **32**, 7610–7616.
 49. Weeks, K.M. and Crothers, D.M. (1993) Major groove accessibility of RNA. *Science*, **261**, 1574–1577.
 50. Wilkinson, K.A., Vasa, S.M., Deigan, K.E., Mortimer, S.A., Giddings, M.C., and Weeks, K.M. (2009) Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *RNA*, **15**, 1314–1321.
 51. Mortimer, S.A. and Weeks, K.M. (2008) Time-resolved RNA SHAPE chemistry. *J. Am. Chem. Soc.*, **130**, 16178–16180.
 52. Dock-Bregeon, A.C. and Moras, D. (1987) Conformational changes and dynamics of tRNAs: evidence from hydrolysis patterns. *Cold Spring Harb. Symp. Quant. Biol.*, **52**, 113–121.
 53. Winkler, W.C. and Breaker, R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.*, **59**, 487–517.
 54. Buck, J., Furtig, B., Noeske, J., Wohnert, J., and Schwalbe, H. (2007) Time-resolved NMR methods resolving ligand-induced RNA folding at atomic

- resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 15699–15704.
55. Romaniuk, P.J., De Stevenson, I.L., and Wong, H.H. (1987) *Nucleic Acids Res.*, **15**, 2737–2755.
 56. Jahn, M.J., Jahn, D., Kumar, A.M., and Söll, D. (1991) Defining the binding site of *Xenopus* transcription factor IIIA on 5S RNA using truncated and chimeric 5S RNA molecules. *Nucleic Acids Res.*, **19**, 2786.
 57. Gelhaus, S.L., Lacourse, W.R., Hagan, N.A., Amarasinghe, G.K., and Fabris, D. (2003) Rapid purification of RNA secondary structures. *Nucleic Acids Res.*, **31**, e135.
 58. Milligan, J.F. and Uhlenbeck, O.C. (1989) Synthesis of small RNAs using T₇ RNA polymerase. *Methods Enzymol.*, **180**, 51–62.
 59. Silberklang, M., Gillum, A.M., and Rajbhandary, U.L. (1979) Use of *in vitro* ³²P labeling in the sequence analysis of nonradioactive tRNAs. *Methods Enzymol.*, **59**, 58–109.
 60. England, T.E. and Uhlenbeck, O.C. (1978) 3'-terminal labelling of RNA with T4 RNA ligase. *Nature*, **275**, 560–561.
 61. Dann, C.E., Wakeman, C.A., Sieling, C.L., Baker, S.C., Irnov, I., and Winkler, W.C. (2007) Structure and mechanism of a metal-sensing regulatory RNA. *Cell*, **130**, 878–892.
 62. Chevalier, C., Boisset, S., Romilly, C., Masquida, B., Fechter, P., Geissmann, T., Vandenesch, F., and Romby, P. (2010) *Staphylococcus aureus* RNAIII binds to two distant regions of *coa* mRNA to arrest translation and promote mRNA degradation. *PLoS Pathog.*, **6**, e1000809.
 63. Qu, H.L., Michot, B., and Bachelier, J.P. (1983) Improved methods for structure probing in large RNAs: a rapid 'heterologous' sequencing approach is coupled to the direct mapping of nuclease accessible sites. Application to the 5' terminal domain of eukaryotic 28S rRNA. *Nucleic Acids Res.*, **11**, 5903–5920.
 64. Laederach, A., Das, R., Vicens, Q., Pearlman, S.M., Brenowitz, M., Herschlag, D., and Altman, R.B. (2008) Semiautomated and rapid quantification of nucleic acid footprinting and structure mapping experiments. *Nat. Protoc.*, **3**, 1395–1401.
 65. Moreno, R., Marzi, S., Romby, P., and Rojo, F. (2009) The Crc global regulator binds to an unpaired A-rich motif at the *Pseudomonas putida alkS* mRNA coding sequence and inhibits translation initiation. *Nucleic Acids Res.*, **37**, 7678–7690.
 66. Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., and Weeks, K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.*, **6**, e96.
 67. Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 97–102.
 68. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 7287–7292.
 69. Walker, S.C., Avis, J.M., and Conn, G.L. (2003) General plasmids for producing RNA *in vitro* transcripts with homogeneous ends. *Nucleic Acids Res.*, **31**, e82.
 70. Schroder, A.R., Baumstark, T., and Riesner, D. (1998) Chemical mapping of co-existing RNA structures. *Nucleic Acids Res.*, **26**, 3449–3450.
 71. Jaeger, L., Westhof, E., and Michel, F. (1993) Monitoring of the cooperative unfolding of the *sunY* group I intron of bacteriophage T4. The active form of the *sunY* ribozyme is stabilized by multiple interactions with 3' terminal intron components. *J. Mol. Biol.*, **234**, 331–346.

11

Probing RNA Solution Structure by Photocrosslinking: Incorporation of Photoreactive Groups at RNA Termini and Determination of Crosslinked Sites by Primer Extension

Michael E. Harris

11.1

Introduction

An integral component of modern analysis of RNA structure and function is the incorporation of modified nucleotides and nucleotide analogs. Nucleotides bearing base and backbone modifications can be introduced randomly or at specific locations, and a variety of modifications are available, which facilitate mechanistic and structural studies of RNA [1–9]. The goal of this chapter is to outline common techniques to introduce nucleotide analogs as photocrosslinking agents into an RNA molecule *in vitro* and to provide examples of experiments in which these techniques have been proved useful for the analysis of ribonuclease P RNA solution structure. Techniques discussed include 5'-end modification by transcription priming [10–16], generation of nucleotide monophosphates, and monophosphorothioates for use in 5'-end modification [17], derivatization of a 5'-phosphorothioate modification [18], and posttranscriptional 3'-end modification.

11.1.1

Applications of RNA Modifications

The availability of a wide range of modified nucleotides provides a range of capabilities for the study of RNA structure–function relationships. Modified nucleotides that alter individual functional groups, or in some cases individual atoms, are available, which allow for a greater degree of experimental resolution than which is possible with nucleobase mutations alone [19]. Importantly, modifications with useful chemistry, such as photocrosslinking agents, can be introduced. Combined with the ability to randomly modify a population of RNAs and to position modifications in a site-specific manner, the investigator can design a wide range of experiments to probe the structure and function of essentially any RNA.

Site-specific incorporation of modified nucleotides is useful when the contribution or role of a single nucleotide or functional group is the focus of study. Site-specific placement of modified nucleotides is often an essential first step in a variety of further techniques. For example, site-specific fluorescent markers can

be introduced into RNA molecules for kinetic and folding studies [20–22] (see Chapter 28 of the first edition of this Handbook). Site-specific placement of phosphorothioate modifications is also necessary for many thiophilic metal ion rescue experiments [23, 24] (see Chapter 19 of the first edition of this Handbook and Chapter 14 of this edition). 5'-End modification with m⁷G(5')ppp(5')G or other cap analog is an essential step in the use of *in vitro* transcribed RNAs in translation or splicing experiments [25–28] (see Chapters 1 and 3). In addition, affinity tags can also be placed at specific positions in an RNA structure to facilitate the isolation of ribonucleoprotein (RNP) complexes (Chapters 1, 6, and 41).

This chapter describes random and site-specific modification methods to position photocrosslinking agents for studies of RNA structure. A number of modifications are available for RNA, which can form crosslinks to both RNA and proteins [29–33]. Commonly used photocrosslinking agents include 4-thiouridine and 6-thioguanosine, which can be internally incorporated into an RNA molecule [34–40] (see also Chapter 1). In addition, the 5'- and 3'-ends of RNA molecules are readily modified with aryl azide-containing photocrosslinking agents [11, 41–44]. Together with the use of circular permutation to relocate the 5'-end of complex RNA molecules to different positions in the RNA structure [45–50], these modifications have been used to provide distance constraints for the structures of complex RNAs, such as bacterial ribonuclease P RNA [14, 41, 44, 46, 51–53] and ribosomal RNA [30, 54–58].

In contrast to site-specific modification, which is useful when a substructure or functional group is already known to be of interest, random incorporation of modified nucleotides can be used to survey for important nucleotides and functional groups. The primary example of this approach is nucleotide analog interference mapping (NAIM), which has been used successfully in a number of systems to identify functionally important base and backbone functional groups [1,59–62] (Chapters 17 and 18). Similarly, photocrosslinking agents can also be positioned randomly throughout an RNA to survey for adjacent nucleotide positions within the folded structure. The triphosphates of 4-thiouridine and 6-thioguanosine, randomly incorporated during *in vitro* transcription, have been used in the study of RNA structure and RNA–protein interactions in such diverse systems as the HIV-I Rev protein, the ribosome, RNA polymerase, and the group II self-splicing intron [55, 57, 63–66].

11.1.2

Techniques for the Incorporation of Modified Nucleotides

The choice of technique for the incorporation of modifications is primarily dependent on the location of the desired modification, and thus, on the experiment in which the modified RNA will be used. Site-specific incorporation of modified nucleotides can be accomplished by a number of methods. Internal modifications are often introduced by chemical synthesis of all or part of the desired RNA. If the RNA in question is short enough, the simplest solution for internal site-specific incorporation is often chemical synthesis of the entire RNA. If not, chemical

synthesis can be combined with RNA ligation (Chapter 3). Ligation, however, has limitations that can often lead to low yields, and circular permutation combined with either 5'- or 3'-end modification can be a viable alternative, especially for highly structured RNAs.

RNAs with site-specific 5'-end modifications can be prepared in a straightforward manner by transcription priming [10–16], for which an example protocol is given below. During transcription with T7 RNA polymerase, nucleotides lacking a 5' triphosphate cannot be incorporated into an elongating RNA chain but can be used to initiate a transcript (Chapter 1). However, transcription priming can produce a mixed population of 5'-modified and unmodified RNAs. To maximize analog incorporation, it is essential to include an excess of modified nucleotide over the corresponding unmodified nucleotide triphosphate. While this has the effect of increasing the fraction of the population with the desired 5'-modification, it also tends to lower transcription efficiency. It is therefore generally necessary to empirically determine the appropriate analog concentration to balance these factors. For more details on modified nucleotides that can be incorporated by wild-type T7 RNA polymerase, or mutant enzyme versions that are more permissive to 2'-modifications in particular, see Chapter 1.

Site-specific 3'-terminal modifications are generally introduced posttranscriptionally. T4 RNA ligase can be used to add a modified nucleotide to the 3'-end of an RNA (Chapter 3). In addition, posttranscriptional chemical modification can be used to attach affinity selection or photocrosslinking agents to the 3'-terminus, an example protocol involving chemical modification is given below. In this method, the unique 2', 3'-*cis*-diol of the RNA 3'-terminus is oxidized to form a dialdehyde, which then reacts with an alkyldiamine under reducing conditions to yield a unique primary aliphatic amine (see below, also Chapter 6). This primary amine can be further derivatized with *N*-hydroxysuccinimidyl esters to introduce photocrosslinking agents or other useful modifications. The modifications that can be introduced to the 3'-end in this manner are only limited by the availability of reagents that will react with primary amines.

11.2

Description

11.2.1

5'-End Modification by Transcription Priming

The following protocol describes the procedure for the incorporation of a 5'-guanosine monophosphorothioate (GMPS) modification at the 5'-end of RNA generated by *in vitro* transcription. 5'-GMPS-modified RNA can be further derivatized, and an example of modification with an aryl azide photocrosslinking agent is also described. Other nucleotide monophosphates or nucleosides can be substituted in this protocol if a different 5'-modification is required. GMPS to GTP ratios of 10 : 1 to 40 : 1 result in a 70–90% yield of 5'-modified transcripts [12, 29, 63, 67]. The following protocol was designed to maximize the transcription

efficiency of a 76-nt bacterial tRNA and utilizes a ratio of 4.8 : 1. If modified nucleotides other than GMPS are used, the optimal ratio for efficient incorporation and maximal transcription efficiency will have to be determined empirically.

This protocol predominantly requires widely available reagents. Nucleotide triphosphates, as well as T7 RNA polymerase and its associated buffer are available from a number of sources, including Ambion. Yeast pyrophosphatase is available from Sigma-Aldrich. GMPS is also commercially available (Biolog, distributed by Axxora, LLC) but can be conveniently generated in large quantity by chemical phosphorylation of guanosine (see below).

Protocol: Transcription of RNAs with a 5'-Terminal GMPS Modification

- 10 x transcription buffer: 500 mM Tris pH 7.5, 150 mM MgCl₂, 20 mM spermidine, 50 mM dithiothreitol (DTT).

Mix the following components:

T7 transcription 100 µl		Final concentration
10 µl	10 x transcription buffer	1 x
3 µl	0.2 M DTT	6 mM
4 µl	100 mM ATP	1 mM
4 µl	100 mM CTP	1 mM
4 µl	100 mM UTP	1 mM
1.25 µl	100 mM GTP	0.5 mM
20 µl	30 mM GMPS	6 mM
x µl	Linearized DNA template	5 µg
1 µl	40 U µl ⁻¹ RNase inhibitor (Supersin; Ambion)	0.4 U µl ⁻¹
5 µl	5 U µl ⁻¹ yeast pyrophosphatase	0.25 U µl ⁻¹
5 µl	20 U µl ⁻¹ T7 RNA polymerase (Ambion)	2 U µl ⁻¹
To 100 µl	with RNase-free water	—

Incubate the reaction mixture overnight at 37 °C, recover products by ethanol precipitation, and gel-purify on an appropriate percentage denaturing polyacrylamide gel by standard procedures.

11.2.2

Chemical Phosphorylation of Nucleosides to Generate 5'-Monophosphate or 5'-Monophosphorothioate Derivatives

Generation of 5'-phosphorylated nucleosides can be achieved in a straightforward manner and in high yield by chemical phosphorylation of nucleosides with phosphoryl chloride and its derivatives [11, 16, 18, 68, 69]. The example protocol given below describes the generation of GMPS by the reaction of unphosphorylated guanosine with thiophosphoryl chloride. The reaction of phosphoryl chloride with nucleosides in triethylamine occurs almost exclusively with the 5'-hydroxyl group,

making protection of the 2'- and 3'-hydroxyl groups unnecessary. In addition, workup to generate nucleotide triphosphates from the resulting monophosphates has been described well [70]. As many useful analogs are only available as nucleosides, this general procedure can provide opportunities for probing RNA structure and function that might not otherwise be available.

This protocol begins with the generation of a saturated solution and slurry of the nucleoside in triethylamine. Phosphoryl chloride or one of its derivatives is then added to the slurry. A nucleophilic reaction results in the 5'-attachment of phosphoryl chloride to the nucleoside. As the nucleoside becomes phosphorylated, its solubility increases and the pH of the solution is lowered, resulting in more of the nucleoside becoming soluble until a clear solution is achieved. When this reaction is complete, the nucleotide 5'-phosphoryl chloride is hydrolyzed to the nucleotide monophosphate by the addition of an excess of water. The nucleotide can be purified from unreacted nucleoside by ion-exchange chromatography [11, 29, 46]. Alternatively, the barium salt can be incubated with sodium sulfate to generate the sodium salt that may be used to circumvent the chromatography step [18].

A wide variety of nucleoside analogs and phosphoryl chloride derivatives are commercially available and can be substituted for thiophosphoryl chloride in the protocol described below. Methyl-, phenyl-, and ethyl-dichlorophosphite, thiophosphoryl chloride, and 4-nitrophenyl phosphodichloridate are available from Sigma-Aldrich and may be used to yield useful analogs. For example, in addition to GMPS, we have used this procedure to generate other 5'-phosphorylated guanosine derivatives, including 6-thioguanosine monophosphate and guanosine 5'-*p*-nitrophenylphosphate [12, 71]. All other reagents used in the following procedure are commonly available from a variety of sources.

Protocol: Synthesis of GMPS from Guanosine

- 1) Mix 2 mmol of guanosine and 5 ml triethylamine in a small round-bottomed flask. Use a heating mantle to gently warm the flask to 50 °C and stir at that temperature for 10 min.
- 2) Next, cool the guanosine solution on wet ice for at least 10 min. With the resulting guanosine slurry on ice, add 0.6 ml (5.8 mmol) of thiophosphoryl chloride. Continue stirring on ice in a cold room at 4 °C overnight (or at least 7 h) until the guanosine slurry has become a clear solution.
- 3) The reaction mixture is next mixed with 500 ml of water to hydrolyze the resultant guanosine 5'-thiophosphoryl chloride to GMPS. Because the hydrolysis reaction is exothermic, the reaction mixture should be added to water in small aliquots and the solution should be stirred on ice between additions. Before chromatographic purification, adjust the pH of the solution to 7.5.

Protocol: Purification of GMPS Product

For purification of GMPS and other analogs, we have used a 2.5 × 18 cm column containing Supelco TSK-gel Toyopearl DEAE-650M resin prepared according to the manufacturer's instructions.

- 1) Load the sample in 50-ml aliquots at a flow rate of 1 ml min^{-1} and wash the column with three column volumes of water.
- 2) Elute the nucleoside and phosphorylated compound with a linear gradient of 0–100% B in 60 ml (A = water; B = 0.5 M ammonium bicarbonate, pH 7.0). A sample chromatogram with good separation between the nucleoside, which elutes first, and the thiophosphorylated nucleotide is shown in Figure 11.1a. Pool the peak fractions and recover products by rotary evaporation. The column can be regenerated by washing with several column volumes of 100% B followed by water.
- 3) Next, resuspend the residue in 200 ml of 10% ethanol and dry again. Repeat this step at least four times. Finally, resuspend the GMPS product in 5 ml of water.
- 4) The identity of the individual peaks can be confirmed by running the appropriate fractions on polyethyleneimine thin layer chromatography (TLC) plates developed in 1 M lithium chloride relative to standards. In addition, samples can be analyzed using mass spectrometry. An example of a mass chromatogram of final pooled GMPS fractions is shown in Figure 11.1b. A clear peak at 378 m.u. indicates that GMPS was the sole component of the second peak as shown in Figure 11.1a. Peaks at 379 and 380 m.u. are due to the natural abundance of carbon and oxygen isotopes. The final product concentration can be determined by UV absorbance ($\epsilon_{260} = 11.7 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$ at pH 7 [72]).

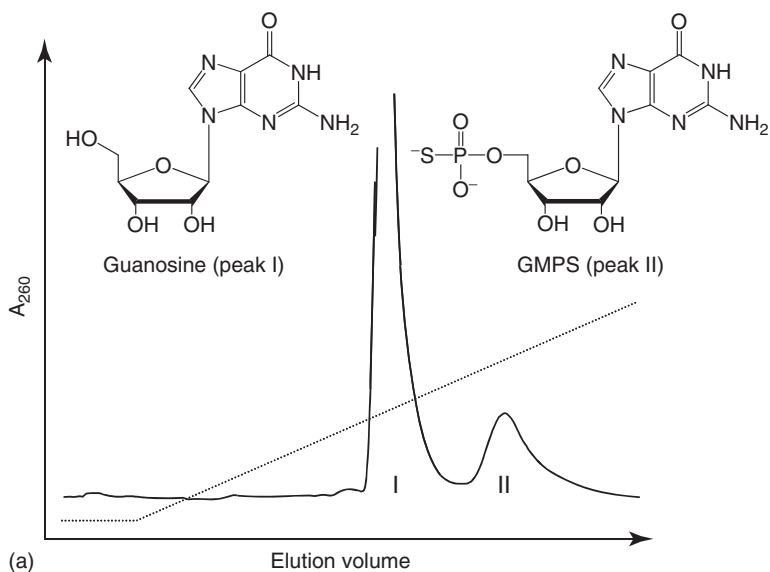
11.2.3

Attachment of an Aryl Azide Photocrosslinking Agent to a 5'-Terminal Phosphorothioate

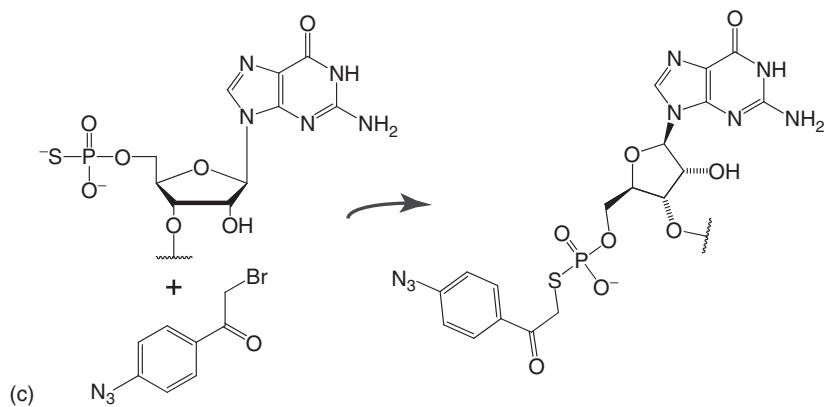
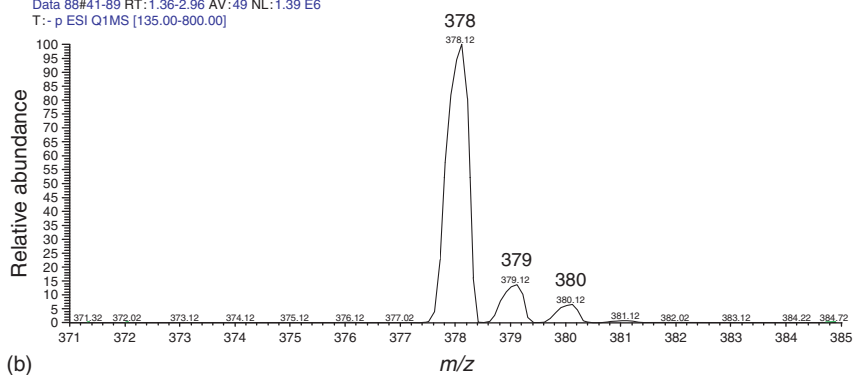
RNAs that are modified with a 5'-terminal phosphorothioate, as described above, can be further derivatized with reagents containing an acid bromide. The following example describes the attachment of an azidophenacyl photocrosslinking agent to a 5'-phosphorothioate-modified RNA [11] (Figure 11.1c). This crosslinking agent has been used to investigate the structures of various RNAs and has an effective crosslinking radius of approximately 10 Å.

Sodium bicarbonate, tris (hydroxymethyl) aminomethane (Tris), ethylenediaminetetraacetic acid (EDTA), sodium dodecyl sulfate (SDS), and methanol are commonly available from a number of suppliers. 4-Azidophenacyl bromide (APBr)

Figure 11.1 GMPS purification, analysis, and modification with the azidophenacyl photoagent. (a) Absorbance trace showing the purification of GMPS from guanosine by liquid chromatography. The dotted line indicates the fraction of buffer B (0.5 M ammonium bicarbonate, pH 7.0) in water. (b) Mass spectrographic analysis of GMPS from the indicated peak in (a). Data were acquired with a Thermoquest TSQ quadrupole mass spectrometer equipped with an electrospray ion source. (c) Azidophenacyl attachment to a 5' phosphorothioate. Reaction with an acid bromide generates a linkage through the sulfur atom of the phosphorothioate.



Data 88#41-89 RT: 1.36-2.96 AV: 49 NL: 1.39 E6
T: - p ESI Q1MS [135.00-800.00]



is available from Sigma-Aldrich (Cat. No. A6057) (Figure 11.1c). The methanol solution of APBr used for this reaction should be freshly prepared. In addition, reducing agents such as DTT reduce the azide moiety of this crosslinker to an amine and should be avoided. Finally, as this protocol involves the use of photosensitive reagents, care should be taken to avoid exposure to light.

Protocol: 5'-Addition of the Aryl Azide

- 1) Resuspend 10–40 µg of GMPS-primed RNA in 10 µl of water.
- 2) Add the following components:

24 µl	RNase-free water
20 µl	100 mM Sodium bicarbonate, pH 9.0
22 µl	0.05% SDS
40 µl	3 mg ml ⁻¹ APBr in methanol

- 3) Incubate the reaction mixture in the dark at room temperature for 1 h.
- 4) Add 100 µl of 10 mM Tris, 1 mM EDTA, pH 8.0.
- 5) Next, extract the reaction mixture with an equal volume of 50/50% phenol/chloroform to remove any unreacted APBr.
- 6) Finally, extract the reaction with an equal volume of chloroform and recover the RNA products by ethanol precipitation.

11.2.4

3'-Addition of an Aryl Azide Photocrosslinking Agent

Photocrosslinking agents and other modifications can be posttranscriptionally added to the 3'-end of RNA molecules by chemical modification [43, 44]. As described above, this procedure takes advantage of the unique *cis*-diol of the 3'-terminal ribose. The procedure described below is a three-step process. In the first step, the 3'-*cis*-diol is oxidized to a dialdehyde [73] (Figure 11.2). In the second step, a primary amine is introduced at the 3'-end. The example below uses ethylenediamine in this step of the reaction [74, 75]. If a longer carbon chain is desired, other compounds such as 1,6-diaminohexane can be used. In the third and final step, an *N*-hydroxysuccinimide-containing reagent is added to react with the modified RNA. The example given here utilizes *N*-hydroxysuccinimidyl-4-azidobenzoate for attachment of the aryl azide crosslinking agent. Other *N*-hydroxysuccinimide-containing reagents, such as a number of available fluorescent labels, should also be usable in this protocol.

As with the other protocols presented in this chapter, most of the reagents called for are commonly available. *N*-Hydroxysuccinimidyl-4-azidobenzoate is available from Sigma-Aldrich (Cat. No. A2907) (Figure 11.2). Similar to azidophenacyl bromide used in Section 11.2.3, care should be taken to avoid exposure of the

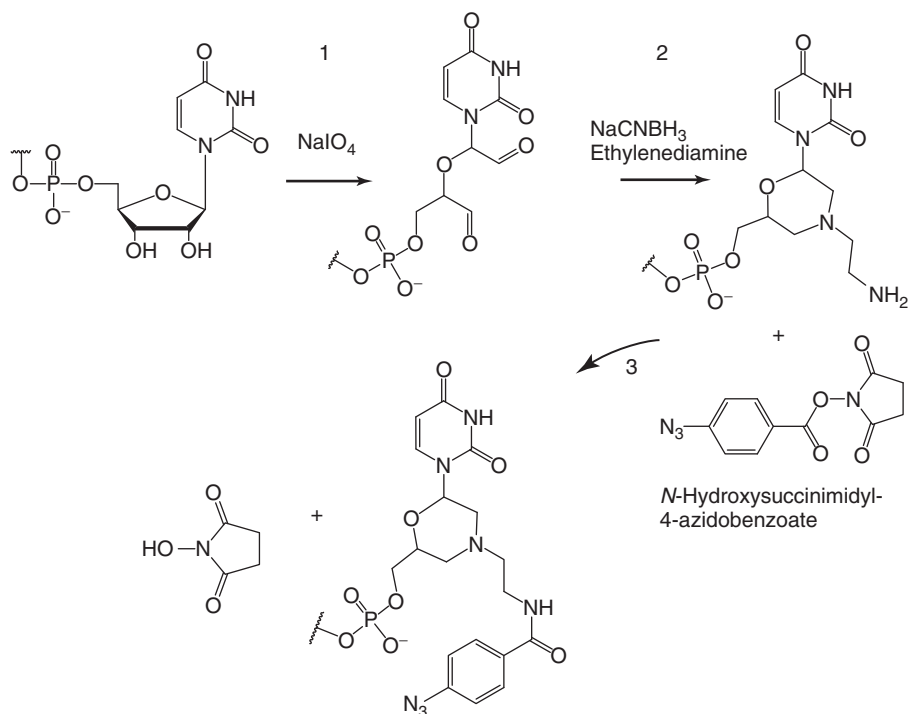


Figure 11.2 3'-End modification of RNA with a terminal primary amine followed by incorporation of the azidophenacyl photoagent. The chemical scheme for 3'-end modification by the method of Oh and Pace consists of three reactions steps. (1) Oxidation of the 2', 3'-*cis*-diol of the 3'-terminal ribose to a dialdehyde.

(2) Reduction of the dialdehyde in the presence of ethylenediamine to yield a primary amine with a two-carbon linker. (3) Reaction with the *N*-hydroxysuccinimide ester derivative to yield the final RNA product carrying the 3'-terminal azidophenacyl crosslinking functionality.

photoagent to light and to reducing agents such as DTT. In addition, appropriate safety precautions should be taken when handling ethylenediamine.

Protocol: 3'-Addition of the Aryl Azide

1) Begin by mixing the following components:

Final volume 100 μl		Final concentration
10 μl	1 M sodium acetate, pH 5.4	0.1 M
10 μl	30 mM NaIO_4	3 mM
5–10 μg	RNA	—
To 100 μl	with RNase-free water	—

- 2) Incubate the reaction mixture for 1 h at room temperature in the dark.
- 3) Recover the RNA by ethanol precipitation.
- 4) Redissolve the precipitate in 72 μl and add the following components:

Final volume 100 μl		Final concentration (mM)
10 μl	200 mM Imidazole, pH 8.0	20
10 μl	50 mM NaCNBH ₃	5
1.0 μl	100 mM EDTA	1
6.7 μl	15 mM Ethylenediamine	1

- 5) Incubate this solution at 37 °C for 1 h.
- 6) Add 10 μl of 50 mM NaBH₄ and continue incubating at 37 °C for an additional 10 min. Once again, recover reaction products by ethanol precipitation.
- 7) Redissolve the precipitate in 50 μl of 100 mM 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid(HEPES), pH 9.0, and add 50 μl of 20 mM *N*-hydroxy-succinimidyl-4-azidobenzoate. Allow this mixture to react at room temperature for 1 h in the dark. Finally, ethanol-precipitate to recover the 3'-modified RNA product.

11.3

Troubleshooting

The 5'-end modification protocol described in this section is dependent on T7 RNA polymerase to incorporate modified nucleotides. As such, a major concern is likely to be the efficiency of incorporation. As described above, the level of incorporation can be adjusted manually to suit the experiment by altering the ratio of modified nucleotide to its unmodified counterpart in the transcription reaction. However, increasing concentrations of modified nucleotide, especially if the modification does not lend itself to efficient incorporation, is likely to also reduce the overall yield. As mentioned earlier, mutant versions of T7 RNA polymerase are often useful in increasing the incorporation of modifications, especially those with 2' or minor groove constituents (Chapter 1).

11.4

Probing RNA Structure by Photoaffinity Crosslinking with 4-Thiouridine and 6-Thioguanosine

11.4.1

Introduction

Chemical crosslinking, including photoaffinity crosslinking, has been widely used to gain insight into structures associated with the biological function of large,

structurally complex RNAs and RNPs. Combined with continuing improvements in the ability to generate RNAs with site-specific modifications, crosslinking continues to be a key analytical method for investigating structure–function relationships. Together with kinetic and thermodynamic studies of structure variants, crosslinking can be used to reveal residues involved in catalysis and molecular recognition. If sufficient information is available from other biochemical and comparative analyses, it can be possible to use the information gained from crosslinking as constraints for molecular modeling. Although the resolution of structures obtained this way is necessarily low (generally of the order of ± 10 Å), they present an explicit context for designing new structure–function experiments and for interpreting structural information.

Important new insights into the validity and the limitations of crosslinking as an approach for exploring RNA structure come from the comparison of the recent three-dimensional structure of the ribosome and the extensive collection of biochemical structure probing data [76]. Overall, a large percentage of the crosslinking data were consistent with the structure from X-ray crystallography; however, the resolution of the structural information was less than the expected, given the chemical structure and size of the different crosslinking reagents used. Furthermore, no individual crosslinking reagent appeared to be superior with respect to data validity; however, the method of detection did have an important impact as most of the lower quality data were obtained by primer extension mapping. Most likely, this limited accuracy is due to misidentification of non-specific terminations as crosslink sites. The highest quality data was obtained by direct physical mapping of the crosslinked nucleotides. Nevertheless, primer extension mapping is still a convenient method because of its sensitivity and flexibility.

Crosslinking approaches are particularly valuable in those numerous instances when it is impossible to obtain material in adequate amounts or in sufficient purity for high-resolution structural analysis. Often, it is desirable to probe structure in a context, such as within cell extracts, where high-resolution studies are impractical. In principle, crosslinking reports on the structure or structures as they occur in solution, and in instances when conditions can be found to favor one conformation over another, it can be possible to use crosslinking to define the characteristic structural features of these different states. Another advantage is sensitivity, as relatively small amounts of crosslinked material are needed for mapping.

Here, we focus on simple procedures and considerations for generating and isolating crosslinked RNAs and for primer extension mapping of crosslinked nucleotides. In the examples given below, crosslinking using 6-thioguanosine (s^6G) is applied to identify active site components within the RNase P ribozyme–substrate complex [12] (Figure 11.3). The description is designed to be sufficiently general in order to be of maximum use as a guideline for an experimenter at least at the graduate level, who is considering the application of photocrosslinking of RNA in her/his research. However, a basic understanding of techniques for handling nucleic acids is assumed.

Although a wide variety of chemical and photocrosslinking reagents are available, s^6G and the widely used s^4U analog (Figure 11.3a) are excellent choices because of

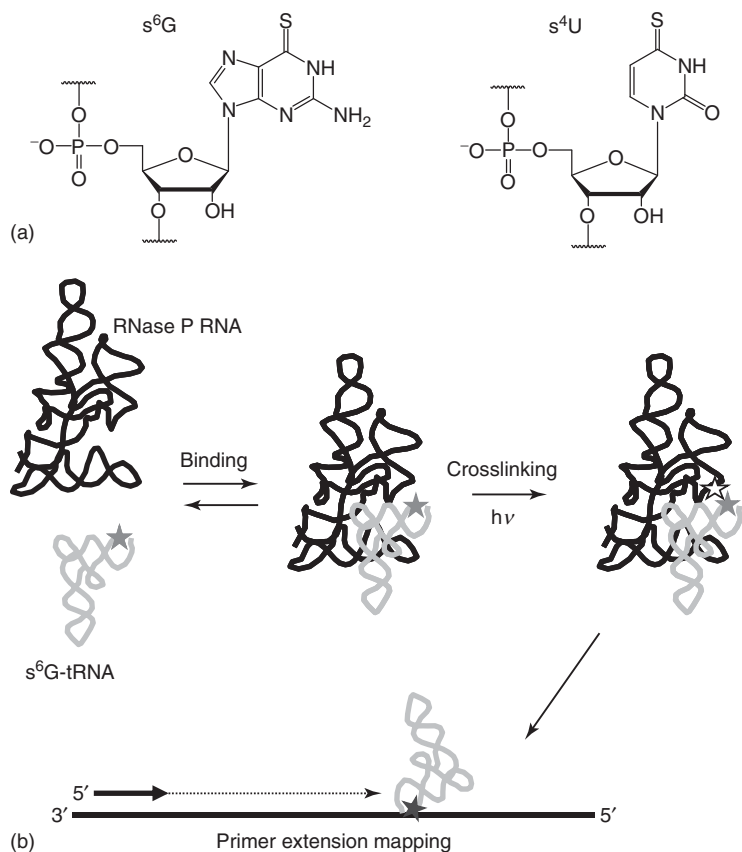


Figure 11.3 (a) Structures of 6-thioguanosine (s^6G) and 4-thiouridine (s^4U) 5'-monophosphates. (b) Overview of the photoaffinity crosslinking and primer extension mapping approach. The RNase P RNA is represented in this example as a black ribbon diagram. The photoagent-modified pre-tRNA substrate is shown in gray. The position of the photoagent is indicated by a dark-gray star. As described in the text, the two RNAs are allowed to bind (binding) and the photoagent is activated by exposure

to the appropriate wavelength of UV light (crosslinking). The crosslinked site in RNase P RNA is illustrated by a white star with black contour. Subsequently, the appropriate crosslinked RNA species are isolated by gel purification and the sites of crosslinking are determined by primer extension mapping. The radiolabeled primer, which is annealed to the RNase P RNA moiety of the crosslinked complex, is indicated by the thick arrow, and its extension by the thin arrow with dotted line.

their simple molecular structure, relative stability, and high reactivity. s^4U and s^6G introduce only minimal perturbations of the native structure, as they differ from their corresponding “parent” nucleoside by a single atomic substitution only. This substitution renders the reagent sensitive to UV light and exposure yields a reactive sulfur radical that can react efficiently with functional groups in its proximity. Crosslinking reactions involving these reagents can be very efficient, making it an easier task to isolate sufficient quantities of crosslinked species for mapping

of crosslinked nucleotides and assessment of retention of biological activity. In addition, these reagents are advantageous in that they have a relatively short range (around 3 Å), and thus, in principle provide spatial information of higher resolution than, for example, azido derivatives introducing a linker between the RNA and photoagent that is circa 10 Å.

Despite its conceptual simplicity, crosslinking reactions can be inefficient because of inherently unfavorable geometry or an adverse chemical environment at the site of photoagent incorporation. Although methods using radioactive labeling that have high sensitivity are used to map crosslink sites, the clearest and best results are obtained when nanogram quantities of the crosslinked species can be obtained. Similarly, it can sometimes be hard to generate high-quality primary data because several manipulations of RNA are required (e.g., photoagent modification, crosslinking, and gel purification). As described in more detail below, interpretation of primer extension–termination mapping of crosslinked sites (Figure 11.3b) can be difficult and great care must be taken to ensure that the data truly reflect the formation of crosslinks. Much of the ambiguity can be resolved through inclusion of appropriate controls as outlined below.

A critical step is testing whether the crosslinking data reflects the functional, native fold of the RNA and is therefore biologically meaningful. Optimally, this goal can be achieved by assaying directly whether the crosslinked RNA retains biological activity. This may become a problem when probing the functional core of an RNA, as the crosslink itself can alter chemical groups important for biological function. Alternatively, the proximity data from crosslinking can be considered in light of other structural constraints from, for example, phylogenetic comparative studies, chemical and enzymatic probing, and high-resolution structures of homologous molecules.

11.4.2

Description

11.4.2.1 General Considerations: Reaction Conditions and Concentrations of Interacting Species

Important issues are the significant condition dependence of crosslinking reactions and the variability in the physical behavior of different RNAs. Thus, crosslinking experiments usually require substantial effort toward optimization of different experimental parameters. Moreover, individual RNAs often adopt multiple conformations, and obviously, it is necessary that the crosslinking experiment be performed under conditions favoring the correct structural form or the structure of interest. For optimal folding of the RNA, it is important to have as detailed an understanding as possible about the influence of mono- and divalent ion identity and concentration, as well as pH on biological activity. It is also useful to examine the effect of these parameters on the crosslinking reaction as well, as gaining the highest efficiency possible is important for subsequent identification of crosslinked nucleotides and for analysis of the retention of biological activity of the purified crosslinked species.

For crosslinking of two RNAs or RNA and a specific protein, one should be aware that RNAs can self-associate or bind in non-productive ways. Similarly, even specific RNA binding proteins can interact weakly with RNA in a non-specific manner. Thus, it is important to consider the relative concentrations of the interacting species in the reaction to minimize the potential for the formation of non-specific complexes. Examining the effect of macromolecular concentration on the crosslinking reaction can provide insight into whether the information gained accurately reflects the formation of high-affinity or biologically active complexes.

11.4.2.2 Application Example – RNase P RNA and s⁶G-Modified Precursor tRNA

In the following example, the interaction between the RNase P ribozyme and one of its substrates was examined using intermolecular crosslinking with s⁶G-modified precursor tRNA (pre-tRNA; derived from *Bacillus subtilis* tRNA^{Asp}) (Figures 11.4 and 11.5). For crosslinking with pre-tRNAs carrying the related s⁴U analog, see [63]. RNase P is a widespread and essential RNP enzyme that generates the 5'-end of mature tRNAs *via* a site-specific phosphodiester bond hydrolysis reaction [77, 78]. In bacteria, RNase P enzymes are heterodimers composed of a small, but essential,

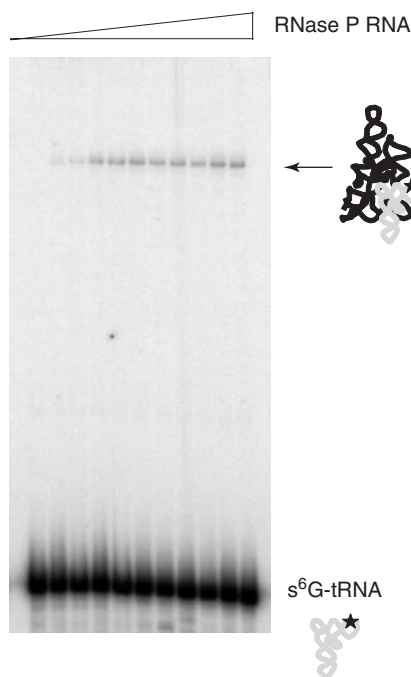


Figure 11.4 Analysis of the formation of crosslinked species by denaturing gel electrophoresis. In this example, radiolabeled and photoagent-modified pre-tRNA [s⁶G-tRNA(G-1)] was incubated with increasing concentrations of RNase P ribozyme

and the reactions exposed to UV light. The formation of a single crosslinked species resulting in anomalously slower migration that requires the presence of the ribozyme and is dependent on its concentration is indicated by the arrow.

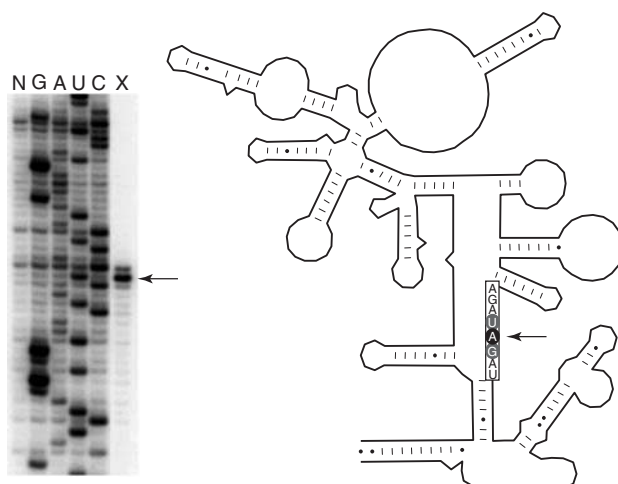


Figure 11.5 Primer extension mapping of crosslinked nucleotides. Products from reactions containing gel-purified crosslinked RNA (lane X) as well as RNA from a control reaction that was not irradiated (lane N) were resolved by gel electrophoresis. Lanes G, A, U, and C contain products from sequencing

reactions containing the appropriate dideoxynucleotide. An arrow indicates the position of the termination due to crosslink formation. A sketch of the secondary structure of *B. subtilis* RNase P RNA is shown on the right with the position of the crosslink indicated by an arrow.

protein subunit and a larger RNA component that is the catalytic subunit [79] of the enzyme. While most ribozymes catalyze self-cleavage or self-splicing reactions and have to be engineered to work *in trans*, catalysis of a multiple turnover reaction is intrinsic to the biological function of RNase P RNA.

To identify residues in the RNase P ribozyme that are proximal to the substrate cleavage site, we synthesized a model pre-tRNA with a 5'-terminal s⁶G modification in its 1-nt leader [s⁶G-tRNA(G-1)] immediately upstream of the reactive phosphodiester bond (Figure 11.3). Kinetic and thermodynamic studies demonstrated that the inclusion of the photoagent at the substrate cleavage site did not interfere with high-affinity binding and that the modified substrate was processed at a rate that was essentially identical to that of the unmodified substrate. For the construction of the s⁶G-tRNA(G-1) pre-tRNA, s⁶GMP was synthesized by chemical phosphorylation starting from 2 mmol 6-mercaptoguanosine (Sigma-Aldrich, Cat. No. 858412) and phosphoryl chloride as described in Section 11.2.2. The 5'-monophosphorylated analog was then used to initiate T7 transcripts (Section 11.2.1).

Protocol: Folding of Ribozyme and Substrate RNAs

To ensure proper folding of the two RNAs and efficient formation of the enzyme-substrate complex, the following procedure was used.

- 1) The RNAs are redissolved separately (e.g., 2 μ M RNase P RNA, 200 nM photoagent-containing pre-tRNA) in reaction buffer (here: 2 M ammonium

acetate; 50 mM Tris–HCl, pH 8.0) for refolding. The RNA-containing solutions are heated to 90 °C for 1 min in a programmable heating block (MJ Research) and then cooled to room temperature using a standard water bath over a period of approximately 20 min.

- 2) Divalent metal ions, in this example 25 mM CaCl_2 , are added and the RNAs are incubated at 37 °C for 15–30 min to ensure that as much of the RNA as possible has attained the native fold.
- 3) Equal volumes of substrate and enzyme RNA are mixed (for RNA concentrations, see below) and incubated for 2 min. Here, Ca^{2+} replaces the optimal metal ion for the reaction, Mg^{2+} , in order to slow the rate of catalysis and permit the assessment of the binding affinity of the substrate, and not of the mature tRNA product [80].

Preparative intermolecular crosslinking reactions generally contained 100 nM photoagent-containing pre-tRNA and 1 μM RNase P ribozyme to saturate the photoagent-modified substrate with ribozyme. These concentrations also resulted in the highest yield of crosslinked product. Importantly, it could be demonstrated that the formation of crosslinks was dependent on the presence of the ribozyme and occurred in a concentration-dependent manner over a broad range of concentrations (Figure 11.4). The concentration dependence clearly demonstrates that the crosslinks are intermolecular. In addition, the same crosslinked species were detected at both high and low concentrations of the ribozyme, indicating that a single type of RNA–RNA complex was probed here.

11.4.2.3 Generation and Isolation of Crosslinked RNAs

Once the conditions and concentrations of the reaction are set or optimized, crosslinking is easily initiated by irradiation with the appropriate wavelength of light. Subsequently, the reactions are analyzed for the formation of new crosslinked species. Identification and isolation is almost always accomplished by taking advantage of the altered mobility of the crosslinked RNAs relative to uncrosslinked RNA on denaturing polyacrylamide gels. The crosslinked RNAs are subsequently eluted from the gel and recovered by ethanol precipitation using standard methods.

Protocol: Analytical Crosslinking

- 1) 12 μl aliquots containing *B. subtilis* RNase P RNA and radioactively labeled s^6G -tRNA(G-1) from step 3 of the preceding protocol were transferred to a parafilm-covered aluminum block. A convenient source is the block from a standard dry bath incubator, precooled in ice for at least 1 h before the experiment. We found that crosslinking optimally occurred at 4 °C. Parafilm and samples were placed on the block just before irradiation to minimize dilution or contamination by condensation of water from the air on the cold surface.
- 2) Samples were irradiated for 5–15 min at 366 nm at a distance of 3 cm using a model UVGL-58 ultraviolet lamp from UVP (Upland, CA, USA). A standard (3–4 mm) thick glass plate was placed between the lamp and the sample to help filter out shorter wavelengths of UV light that can damage the RNA sample.

- 3) Aliquots were recovered from the block, diluted to 200 μl with 10 mM Tris-HCl, pH 8.0, 0.5 mM EDTA, 0.3 M sodium acetate, then extracted twice with 50/50% phenol/chloroform and once with chloroform, and precipitated by the addition of 3 vol. of ethanol.

Because both inter- and intramolecular crosslinking alter the linear topology of the targeted RNA, one can generally identify and isolate crosslinked species based on their slower mobility in denaturing acrylamide gels (Figure 11.4). Appropriate controls should be run in parallel in which the photoagent is omitted from the reaction in order to demonstrate that the formation of the more slowly migrating species depends on presence of the crosslinking reagent and not from adventitious crosslinking because of ambient UV light. Similarly, control samples that are not irradiated must also be compared as crosslinking can occur during sample workup, which may not necessarily reflect the functional structure. In addition, for intermolecular crosslinking, it is essential to demonstrate that formation of the crosslinked species requires the presence of the interacting partner RNA or protein and that its formation is concentration-dependent.

Once the specificity of the crosslinking reaction is established, the next step is to isolate sufficient quantities of the individual crosslinked species to map the crosslink sites. Keeping in mind that picomole amounts of material will be optimal for primer extension mapping, it is necessary to scale the crosslinking reaction up accordingly. We have had good success in simply “spiking” preparative reactions with a small quantity of radiolabeled RNA that is used as a marker for gel purification. For electrophoresis, the sample can be loaded in a continuous well across the top of the gel, applying electrophoresis conditions that were optimized in analytic reactions, to achieve the best degree of separation between crosslinked and uncrosslinked RNA. Standard methods are appropriate for the location of bands by autoradiography, excision of gel slices, and elution and recovery of the RNA. We have found that the addition of $0.01 \mu\text{g } \mu\text{l}^{-1}$ glycogen as a carrier greatly improves recovery from larger volumes of gel elution buffer and does not interfere in subsequent primer extension mapping experiments.

11.4.2.4 Primer Extension Mapping of crosslinked Nucleotides

The general principle behind primer extension mapping of crosslinked nucleotides is that reverse transcriptase will continue to synthesize a DNA strand up to, but not beyond, the site of crosslinking because of steric blockage of the template strand (Figure 11.3b). These specific terminations observed in reactions containing crosslinked RNA as a template and not observed in control, uncrosslinked RNA samples are interpreted as sites where the photoagent has formed a covalent bond (Figures 11.3 and 11.5). Terminations are interpreted as occurring 1 nt 5' to the site of crosslinking. A key advantage is that only relatively small amounts of template RNA are required (1–0.1 pmol); however, best results are obtained when at least picomolar amounts are available. In addition, the technique is relatively easy and rapid, and requires reagents and equipment that are widely available.

An inherent problem of the primer extension method is that reverse transcriptase will pause at specific sites on virtually any RNA template. Therefore, it is essential

to distinguish between a termination that is due to a crosslink and one that is due to RNA structure, non-specific radiation damage, or degradation. To control for these phenomena, it is important to perform a parallel analysis of RNA taken through the protocol but not subjected to irradiation, as well as a control RNA sample irradiated in the absence of ligand, or with a ligand population that has not been modified with the photoagent.

Corroborative information can also be obtained from RNase H mapping [58, 81] of crosslink sites. Crosslinking alters the mobility of RNAs in denaturing acrylamide gels, resulting in anomalously slower migration for fragments containing the crosslinked nucleotides. RNase H mapping involves the use of complementary DNA oligonucleotides to direct site-specific cleavage of the crosslinked RNA. By comparing the oligonucleotide-directed RNase H digestion products for crosslinked and uncrosslinked RNAs, information about the sites of crosslinking can be inferred. The altered electrophoretic mobility of RNase H cleavage fragments derived from crosslinked RNAs can in favorable cases be used directly to estimate the location of crosslinked sites [41].

Protocol: Primer Extension Analysis of Intermolecular Crosslinks between Photoagent-Modified Pre-tRNA and the RNase P Ribozyme

- 1) 0.2 pmol of 5'-³²P-end-labeled primer (here 5'-GTG GTC TAA CGT TCT GT complementary to the 3'-terminal region of *B. subtilis* RNase P RNA) are annealed to 0.05–0.2 pmol of gel-purified crosslinked RNA in a total volume of 5 μ l. The annealing solution is composed of 50 mM Tris–HCl, pH 8.3, 15 mM NaCl, and 10 mM DTT.
- 2) Individual samples are heated to 65 °C for 3 min and then set immediately on dry ice.
- 3) The annealed samples are thawed on ice and 1 μ l of 30 mM MgCl₂ is added, followed by the addition of 2 μ l dNTP mix (0.5 mM each dNTP). These reactions (8 μ l) are initiated by the addition of 2 μ l (2 U) of AMV reverse transcriptase (Roche Applied Science) and then incubated at 47 °C for 5 min.
- 4) Reactions are quenched by the addition of an equal volume of 0.5 M NaCl, 20 mM EDTA, and 0.5 mg glycogen, and the extension products are recovered by ethanol precipitation (addition of 50 μ l ethanol), followed by washing with 80% ethanol to remove excess salt from the radiolabeled products.
- 5) In parallel, prepare dideoxy sequencing standards for denaturing PAGE analysis of extension products. For this purpose, the primer extension assay described before is performed in the presence of one of the four ddNTP analogs. The added concentration of dideoxynucleotide can be varied from 5 to 100 μ M to adjust the resolution of sequencing ladders from several hundred to less than ten nucleotides away from the primer binding site, depending on the needs of the particular experiment.
- 6) After ethanol precipitation, the radiolabeled extension products are redissolved in a small volume (2–5 μ l) of formamide loading buffer (95% formamide, 150 mM Tris–HCl, pH 8.0, 15 mM EDTA, and trace amounts of bromophenol blue and xylene cyanol FF). Only a fraction of the reaction (e.g., 2 μ l) is loaded

in an individual lane such that the sample just covers the bottom of the well in order to generate a banding pattern as sharp as possible.

Figure 11.5 illustrates the result of a primer extension experiment to locate a crosslink of s⁶G-tRNA(G-1) pre-tRNA on *B. subtilis* RNase P RNA.

11.4.3

Troubleshooting

The key problems associated with these particular methods are inefficient formation of crosslinked species and ambiguous or unclear primer extension mapping results.

- 1) In the case of inefficient crosslinking yield, locating the photoagent to a nearby region of the molecule may overcome unfavorable geometric constraints. In addition, as mentioned above, it is usually important to optimize the folding of the RNAs of interest in order to ensure that the maximum fraction of the sample is in the correctly folded and biologically active form. If the particular site of photoagent attachment is sufficiently interesting, then trying additional, longer-range crosslinking agents such as phenylazides can be used to increase crosslinking efficiency. One of the key difficulties that we have encountered is in protecting the photoagent-modified RNA from ambient UV light. Although it can be somewhat awkward, working up the photoagent-modified RNA and performing the crosslinking reactions in a darkened laboratory environment can be important for obtaining clean results.
- 2) In our experience, by far the most challenging aspect is to generate clean primer extension mapping data. As with all procedures involving RNA, making sure that the sample is not exposed to heating in the presence of metal ions is essential. Degradation of the RNA sample is apparent as an intense background of terminations in the control RNA primer extensions. Beyond repetition of the experiment to confirm the reproducibility of the results, good experimental technique and careful handling of the RNA is implicit. Some pilot experiments addressing what conditions are most appropriate for annealing of the radiolabeled primer will also pay off in the long run. We have also observed large differences in the quality of data and the pattern of non-specific terminations with different primers, and comparing two or more sites of primer binding was necessary in some cases. In addition, the optimal primer concentration is often also idiosyncratic to individual oligonucleotide sequences. As a further option, oligonucleotide primers with locked nucleic acid modifications [82] may be used to minimize the off-rate of primer-RNA target duplexes for the sake of increased assay sensitivity.
- 3) Another optimization area is the annealing procedure where it is useful to compare slow cooling to rapid cooling.
- 4) We have found that longer incubation times for the reverse transcriptase reaction can increase the background, and thus, it is important to consider assessing the effects of varying the reaction time. In addition, we have found that increasing the reaction temperature can result in fewer non-specific

transcription terminations, but temperatures in excess of 50 °C result in enzyme denaturation or inhibition.

Despite the number of parameters that can be varied, generally, a few days spent optimizing these few aspects of the procedure using control, uncrosslinked RNA will be an excellent investment, as the payoff will be in obtaining clearer and therefore more convincing primary data.

References

1. Suydam, I.T. and Strobel, S.A. (2009) Strobel, Nucleotide analog interference mapping. *Methods Enzymol.*, **468**, 3–30.
2. Suydam, I.T. and Strobel, S.A. (2008) Fluorine substituted adenosines as probes of nucleobase protonation in functional RNAs. *J. Am. Chem. Soc.*, **130**, 13639–13648.
3. Yajima, R., Proctor, D.J., Kierzek, R., Kierzek, E. and Bevilacqua, P.C. (2007) A conformationally restricted guanosine analog reveals the catalytic relevance of three structures of an RNA enzyme. *Chem. Biol.*, **14**, 23–30.
4. Schwans, J.P., Cortez, C.N., Olvera, J.M., and Piccirilli, J.A. (2003) 2'-mercaptornucleotide interference reveals regions of close packing within folded RNA molecules. *J. Am. Chem. Soc.*, **125**, 10012–10018.
5. Lu, J., Li, N.S., Koo, S.C., and Piccirilli, J.A. (2009) Synthesis of Pyridine, Pyrimidine and Pyridinone C-Nucleoside Phosphoramidites for Probing Cytosine Function in RNA. *J. Org. Chem.* **74**, 8021–8030.
6. Dai, Q., Frederiksen, J.K., Anderson, V.E., Harris, M.E. and Piccirilli, J.A. (2008) Efficient synthesis of [2'-18O]uridine and its incorporation into oligonucleotides: a new tool for mechanistic study of nucleotidyl transfer reactions by isotope effect analysis. *J. Org. Chem.*, **73**, 309–311.
7. Das, S.R. and Piccirilli, J.A. (2005) General acid catalysis by the hepatitis delta virus ribozyme. *Nat. Chem. Biol.*, **1**, 45–52.
8. Liu, L., Cottrell, J.W., Scott, L.G. and Fedor, M.J. (2009) Direct measurement of the ionization state of an essential guanine in the hairpin ribozyme. *Nat. Chem. Biol.*, **5**, 351–357.
9. Da Costa, C.P., Fedor, M.J., and Scott, L.G. (2007) 8-Azaguanine reporter of purine ionization states in structured RNAs. *J. Am. Chem. Soc.*, **129**, 3426–3432.
10. Fang, X., Pan, T., and Sosnick, T.R. (1999) A thermodynamic framework and cooperativity in the tertiary folding of a Mg²⁺-dependent ribozyme. *Biochemistry*, **38**, 16840–16846.
11. Burgin, A.B. and Pace, N.R. (1990) Mapping the active site of ribonuclease P RNA using a substrate containing a photoaffinity agent. *EMBO J.*, **9**, 4111–4118.
12. Christian, E.L., McPheeters, D.S., and Harris, M.E. (1998) Identification of individual nucleotides in the bacterial ribonuclease P ribozyme adjacent to the pre-tRNA cleavage site by short-range photo-crosslinking. *Biochemistry*, **37**, 17618–17628.
13. Chen, J.L., Nolan, J.M., Harris, M.E. and Pace, N.R. (1998) Comparative photocross-linking analysis of the tertiary structures of *Escherichia coli* and *Bacillus subtilis* RNase P RNAs. *EMBO J.*, **17**, 1515–1525.
14. Harris, M.E., Nolan, J.M., Malhotra, A., Brown, J.W., Harvey, S.C., and Pace, N.R. (1994) Use of photoaffinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA. *EMBO J.*, **13**, 3953–3963.
15. Nolan, J.M., Burke, D.H., and Pace, N.R. (1993) Circularly permuted tRNAs as specific photoaffinity probes of ribonuclease P RNA structure. *Science*, **261**, 762–765.

16. Macosko, J.C., Pio, M.S., Tinoco, I.Jr., and Shin, Y.K. (1999) A novel 5 displacement spin-labeling technique for electron paramagnetic resonance spectroscopy of RNA. *RNA*, **5**, 1158–1166.
17. Zhang, B., Cui, Z., and Sun, L. (2001) Synthesis of 5'-deoxy-5'-thioguanosine-5'-monophosphorothioate and its incorporation into RNA 5'-termini. *Org. Lett.*, **3**, 275–278.
18. Wu, C.W., Eder, P.S., Gopalan, V., and Behrman, E.J. (2001) Kinetics of coupling reactions that generate monothiophosphate disulfides: implications for modification of RNAs. *Bioconjug. Chem.*, **12**, 842–844.
19. Wachowius, F. and Hobartner, C. (2010) Chemical RNA modifications for studies of RNA structure and dynamics. *Chembiochem*, **11**, 469–480.
20. Walter, N.G., Huang, C.Y., Manzo, A.J., and Sobhy, M.A. (2008) Do-it-yourself guide: how to use the modern single-molecule toolkit. *Nat. Methods*, **5**, 475–489.
21. Pagano, J.M., Clingman, C.C., and Ryder, S.P. (2011) Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. *RNA*, **17**, 14–20.
22. Zhao, R. and Rueda, D. (2009) RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods*, **49**, 112–117.
23. Frederiksen, J.K. and Piccirilli, J.A. (2009) Identification of catalytic metal ion ligands in ribozymes. *Methods*, **49**, 148–166.
24. Forconi, M., Lee, J., Lee, J.K., Piccirilli, J.A., and Herschlag, D. (2008) Functional identification of ligands for a catalytic metal ion in group I introns. *Biochemistry*, **47**, 6883–6894.
25. Contreras, R., Cheroutre, H., Degrove, W., and Fiers, W. (1982) Simple, efficient *in vitro* synthesis of capped RNA useful for direct expression of cloned eukaryotic genes. *Nucleic Acids Res.*, **10**, 6353–6362.
26. Konarska, M.M., Padgett, R.A., and Sharp, P.A. (1984) Recognition of cap structure in splicing *in vitro* of mRNA precursors. *Cell*, **38**, 731–736.
27. Yisraeli, J.K. and Melton, D.A. (1989) Synthesis of long, capped transcripts *in vitro* by SP6 and T7 RNA polymerases. *Methods Enzymol.*, **180**, 42–50.
28. Jemielity, J., Fowler, T., Zuberek, J., Stepinski, J., Lewdorowicz, M., Niedzwiecka, A., Stolarski, R., Darzynkiewicz, E., and Rhoads, R.E. (2003) Novel “anti-reverse” cap analogs with superior translational properties. *RNA*, **9**, 1108–1122.
29. Harris, M.E. and Christian, E.L. (2009) RNA crosslinking methods. *Methods Enzymol.*, **468**, 127–146.
30. Huggins, W., Ghosh, S.K., Nanda, K., and Wollenzien, P., (2005) Internucleotide movements during formation of 16 S rRNA-rRNA photocrosslinks and their connection to the 30 S subunit conformational dynamics. *J. Mol. Biol.*, **354**, 358–374.
31. Xu, Z. and Culver, G.M. (2009) Chemical probing of RNA and RNA/protein complexes. *Methods Enzymol.*, **468**, 147–165.
32. Buchmueller, K.L., Hill, B.T., Platz, M.S., and Weeks, K.M. (2003) RNA-tethered phenyl azide photocrosslinking via a short-lived indiscriminant electrophile. *J. Am. Chem. Soc.*, **125**, 10850–10861.
33. Wang, Z. and Rana, T.M. (1999) Probing RNA-protein interactions by psoralen photocrosslinking. *Methods Mol. Biol.*, **118**, 49–62.
34. Barritault, D., Buckingham, R.H., Favre, A., and Thomas, G. (1981) Photocrosslinking of thiolated aminoacyl-tRNA to ribosomal RNA and proteins. *Biochimie*, **63**, 587–593.
35. Kao, T., Miller, D.L., Abo, M., and Ofengand, J. (1983) Formation and properties of a covalent complex between elongation factor Tu and Phe-tRNA bearing a photoaffinity probe on its 3-(3-amino-3-carboxypropyl)uridine residue. *J. Mol. Biol.*, **166**, 383–405.
36. Hsu, L.M., Lin, F.L., Nurse, K., and Ofengand, J. (1984) Covalent crosslinking of *Escherichia coli* phenylalanyl-tRNA and valyl-tRNA to the ribosomal A site via photoaffinity probes attached to the 4-thiouridine residue. *J. Mol. Biol.*, **172**, 57–76.

37. Sontheimer, E.J. (1994) Site-specific RNA crosslinking with 4-thiouridine. *Mol. Biol. Rep.*, **20**, 35–44.
38. McGregor, A., Rao, M.V., Duckworth, G., Stockley, P.G., and Connolly, B.A. (1996) Preparation of oligoribonucleotides containing 4-thiouridine using Fpmp chemistry. Photo-crosslinking to RNA binding proteins using 350 nm irradiation. *Nucleic Acids Res.*, **24**, 3173–3180.
39. Yu, Y.T. (2000) Site-specific 4-thiouridine incorporation into RNA molecules. *Methods Enzymol.*, **318**, 71–88.
40. Yu, Y.T. (1999) Construction of 4-thiouridine site-specifically substituted RNAs for cross-linking studies. *Methods*, **18**, 13–21.
41. Harris, M.E., Kazantsev, A.V., Chen, J.L., and Pace, N.R. (1997) Analysis of the tertiary structure of the ribonuclease P ribozyme-substrate complex by site-specific photoaffinity crosslinking. *RNA*, **3**, 561–576.
42. Niranjanakumari, S., Day-Storms, J.J., Ahmed, M., Hsieh, J., Zahler, N.H., Venter, R.A., and Fierke, C.A. (2007) Probing the architecture of the *B. subtilis* RNase P holoenzyme active site by cross-linking and affinity cleavage. *RNA*, **13**, 521–535.
43. Oh, B.K., Frank, D.N., and Pace, N.R. (1998) Participation of the 3'-CCA of tRNA in the binding of catalytic Mg²⁺ ions by ribonuclease P. *Biochemistry*, **37**, 7277–7283.
44. Oh, B.K. and Pace, N.R. (1994) Interaction of the 3'-end of tRNA with ribonuclease P RNA. *Nucleic Acids Res.*, **22**, 4087–4094.
45. Harris, M.E. and Christian, E.L. (1999) Use of circular permutation and end modification to position photoaffinity probes for analysis of RNA structure. *Methods*, **18**, 51–59.
46. Harris, M.E. and Pace, N.R. (1995) Analysis of the tertiary structure of bacterial RNase P RNA. *Mol. Biol. Rep.*, **22**, 115–123.
47. Pan, T. and Zhong, K. (1994) Selection of circularly permuted ribozymes from *Bacillus subtilis* RNase P by substrate binding. *Biochemistry*, **33**, 14207–14212.
48. Pan, T. and Uhlenbeck, O.C. (1993) Circularly permuted DNA, RNA and proteins—a review. *Gene*, **125**, 111–114.
49. Gott, J.M., Pan, T., LeCuyer, K.A., and Uhlenbeck, O.C. (1993) Using circular permutation analysis to redefine the R17 coat protein binding site. *Biochemistry*, **32**, 13399–13404.
50. Pan, T., Gutell, R.R., and Uhlenbeck, O.C. (1991) Folding of circularly permuted transfer RNAs. *Science*, **254**, 1361–1364.
51. Sharkady, S.M. and Nolan, J.M. (2001) Bacterial ribonuclease P holoenzyme crosslinking analysis reveals protein interaction sites on the RNA subunit. *Nucleic Acids Res.*, **29**, 3848–3856.
52. Niranjanakumari, S., Stams, T., Crary, S.M., Christianson, D.W., and Fierke, C.A. (1998) Protein component of the ribozyme ribonuclease P alters substrate recognition by directly contacting precursor tRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 15212–15217.
53. Massire, C., Jaeger, L., and Westhof, E. (1998) Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.*, **279**, 773–793.
54. Seo, H.S. and Cooperman, B.S. (2002) Large-scale motions within ribosomal 50S subunits as demonstrated using photolabile oligonucleotides. *Bioorg. Chem.*, **30**, 163–187.
55. Juzumiene, D., Shapkina, T., Kirillov, S., and Wollenzien, P. (2001) Short-range RNA-RNA crosslinking methods to determine rRNA structure and interactions. *Methods*, **25**, 333–343.
56. Juzumiene, D.I. and Wollenzien, P. (2000) Organization of the 16S rRNA around its 5' terminus determined by photochemical crosslinking in the 30S ribosomal subunit. *RNA*, **6**, 26–40.
57. Sergiev, P.V., Lavrik, I.N., Wlasoff, V.A., Dokudovskaya, S.S., Dontsova, O.A., Bogdanov, A.A., and Brimacombe, R. (1997) The path of mRNA through the bacterial ribosome: a site-directed crosslinking study using new photoreactive derivatives of guanosine and uridine. *RNA*, **3**, 464–475.

58. Baranov, P.V., Dokudovskaya, S.S., Oretskaya, T.S., Dontsova, O.A., Bogdanov, A.A., and Brimacombe, R. (1997) A new technique for the characterization of long-range tertiary contacts in large RNA molecules: insertion of a photolabel at a selected position in 16S rRNA within the *Escherichia coli* ribosome. *Nucleic Acids Res.*, **25**, 2266–2273.
59. Cochrane, J.C., Batey, R.T., and Strobel, S.A. (2003) Quantitation of free energy profiles in RNA-ligand interactions by nucleotide analog interference mapping. *RNA*, **9**, 1282–1289.
60. Oyelere, A.K., Kardon, J.R., and Strobel, S.A. (2002) pK(a) perturbation in genomic Hepatitis Delta Virus ribozyme catalysis evidenced by nucleotide analogue interference mapping. *Biochemistry*, **41**, 3667–3675.
61. Ortoleva-Donnelly, L., Szewczak, A.A., Gutell, R.R., and Strobel, S.A. (1998) The chemical basis of adenosine conservation throughout the *Tetrahymena* ribozyme. *RNA*, **4**, 498–519.
62. Strobel, S.A. and Shetty, K. (1997) Defining the chemical groups essential for *Tetrahymena* group I intron function by nucleotide analog interference mapping. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 2903–2908.
63. Christian, E.L. and Harris, M.E. (1999) The track of the pre-tRNA 5' leader in the ribonuclease P ribozyme-substrate complex. *Biochemistry*, **38**, 12629–12638.
64. de Lencastre, A., Hamill, S., and Pyle, A.M. (2005) A single active-site region for a group II intron. *Nat. Struct. Mol. Biol.*, **12**, 626–627.
65. Podar, M. and Perlman, P.S. (1999) Photocrosslinking of 4-thio uracil-containing RNAs supports a side-by-side arrangement of domains 5 and 6 of a group II intron. *RNA*, **5**, 318–329.
66. Podar, M., Zhuo, J., Zhang, M., Franzen, J.S., Perlman, P.S., and Peebles, C.L. (1998) Domain 5 binds near a highly conserved dinucleotide in the joiner linking domains 2 and 3 of a group II intron. *RNA*, **4**, 151–166.
67. Christian, E.L., Zahler, N.H., Kaye, N.M., and Harris, M.E. (2002) Analysis of substrate recognition by the ribonuclease protein endonuclease RNase P. *Methods*, **28**, 307–322.
68. Kawamoto, S.A., Sudhahar, C.J., Hatfield, C.L., Sun, J., Behrman, E.J., and Gopalan, V. (2008) Studies on the mechanism of inhibition of bacterial ribonuclease P by aminoglycoside derivatives. *Nucleic Acids Res.*, **36**, 697–704.
69. Murray, A.W. and Atkinson, M.R. (1968) Adenosine 5'-phosphorothioate. A nucleotide analog that is a substrate, competitive inhibitor, or regulator of some enzymes that interact with adenosine 5'-phosphate. *Biochemistry*, **7**, 4023–4029.
70. Arabshahi, A. and Frey, P.A. (1994) A simplified procedure for synthesizing nucleoside 1-thiotriphosphates: dATP alpha S, dGTP alpha S, UTP alpha S, and dTTP alpha S. *Biochem. Biophys. Res. Commun.*, **204**, 150–155.
71. Cassano, A.G., Anderson, V.E., and Harris, M.E. (2004) Analysis of solvent nucleophile isotope effects: evidence for concerted mechanisms and nucleophilic activation by metal coordination in nonenzymatic and ribozyme-catalyzed phosphodiester hydrolysis. *Biochemistry*, **43**, 10547–10559.
72. Dawson, R.M.C., Elliott, D.C., Elliott, W.H., and Jones, K.M. (1986) *Data for Biochemical Research*, 3rd edn, Clarendon Press, Oxford, pp. 580.
73. Easterbrook-Smith, S.B., Wallace, J.C., and Keech, D.B. (1976) Pyruvate carboxylase: affinity labelling of the magnesium adenosine triphosphate binding site. *Eur. J. Biochem.*, **62**, 125–130.
74. Krieg, U.C., Walter, P., and Johnson, A.E. (1986) Photocrosslinking of the signal sequence of nascent preprolactin to the 54-kilodalton polypeptide of the signal recognition particle. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 8604–8608.
75. Rayford, R., Anthony, D.D.Jr., O'Neill, O.R.Jr., and Merrick, W.C. (1985) Reductive alkylation with oxidized nucleotides. Use in affinity labeling or affinity chromatography. *J. Biol. Chem.*, **260**, 15708–15713.

76. Whirl-Carrillo, M., Gabashvili, I.S., Bada, M., Banatao, D.R., and Altman, R.B. (2002) Mining biochemical information: lessons taught by the ribosome. *RNA*, **8**, 279–289.
77. Hartmann, R.K., Gössringer, M., Späth, B., Fischer, S., Marchfelder, A., (2009) The making of tRNAs and more-RNase P and tRNase Z. *Prog. Mol. Biol. Transl. Sci.*, **85**, 319–368.
78. Lai, L.B., Vioque, A., Kirsebom, L.A., and Gopalan, V. (2010) Unexpected diversity of RNase P, an ancient tRNA processing enzyme: challenges and prospects. *FEBS Lett.*, **584**, 287–296.
79. Ellis, J.C. and Brown, J.W. (2009) The RNase P family. *RNA. Biol.*, **6**, 362–369.
80. Smith, D., Burgin, A.B., Haas, E.S., and Pace, N.R. (1992) Influence of metal ions on the ribonuclease P reaction. Distinguishing substrate binding from catalysis. *J. Biol. Chem.*, **267**, 2429–2436.
81. Yu, Y.T. and Steitz, J.A. (1997) Site-specific crosslinking of mammalian U11 and u6atac to the 5' splice site of an AT-AC intron. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 6030–6035.
82. Fratzczak, A., Kierzek, R., and Kierzek, E. (2009) LNA-modified primers drastically improve hybridization to target RNA and reverse transcription. *Biochemistry*, **48**, 514–516.

12

Terbium(III) Footprinting as a Probe of RNA Structure and Metal Binding Sites

Dinari A. Harris, Gabrielle C. Todd, and Nils G. Walter

12.1

Introduction

Cations play a pivotal role in RNA structure and function. A functional RNA tertiary structure is stabilized by metal ions that neutralize and, in the case of multivalent ions, bridge the negatively charged phosphoribose backbone [1, 2]. This chapter describes the use of the trivalent lanthanide metal ion terbium(III), abbreviated Tb^{3+} in the following, as a versatile probe of RNA secondary and tertiary structure, as well as an indicator of high-affinity metal ion binding sites. Tb^{3+} was chosen over other lanthanide metals because aqueous Tb^{3+} complexes have a low $\text{pK}_a \sim 7.9$, which conveniently generates enough $\text{Tb}(\text{OH})(\text{aq})^{2+}$ around neutral pH to hydrolyze the RNA backbone. Cleavage occurs *via* deprotonation of the 2'-hydroxyl group and nucleophilic attack of the resulting oxyanion on the adjacent 3', 5'-phosphodiester to form 2', 3'-cyclic phosphate and 5'-hydroxyl termini (Figure 12.1) [3]. In addition, Tb^{3+} has a similar ionic radius (0.92 Å) as Mg^{2+} (0.72 Å) and a similar preference for oxygen and nitrogen ligands over softer ones, while further generating a specific fluorescence signature upon binding to RNA [4–6]. Thus, Tb^{3+} binds to similar sites on RNA as Mg^{2+} but with 2–4 orders of magnitude higher affinity, commensurate with its higher charge density. As a consequence, low (micromolar) concentrations of Tb^{3+} ions readily displace medium (millimolar) concentrations of Mg^{2+} ions from both specific and non-specific high-affinity binding sites and promote slow phosphodiester backbone cleavage, revealing the location of Mg^{2+} -binding sites in the RNA [6]. Higher (millimolar) concentrations of Tb^{3+} ions bind diffusely to RNA and result in backbone cleavage in a sequence-independent manner, preferentially cutting solvent-accessible, single-stranded, or non-Watson-Crick base-paired regions. Incubation with high concentrations of Tb^{3+} provides a footprint of the RNA's secondary and tertiary structure with nucleotide resolution [5–12].

Tb^{3+} can be a very straightforward and useful probe of metal binding and tertiary structure formation in RNA. However, there are several precautions that need to be considered in order to obtain a reliable and reproducible RNA footprinting pattern. As low (micromolar) concentrations of Tb^{3+} bind to high-affinity metal binding

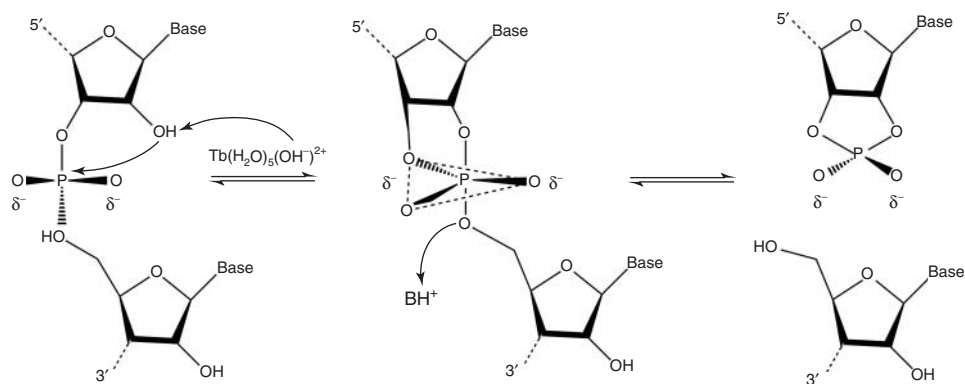


Figure 12.1 Mechanism of RNA backbone cleavage by $\text{Tb}(\text{OH})(\text{aq})^{2+}$. Cleavage occurs *via* deprotonation of the 2'-hydroxyl group and nucleophilic attack of the resulting oxyanion on the adjacent 3', 5'-phosphodiester to form 2', 3'-cyclic phosphate and 5'-hydroxyl termini.

sites within a folded RNA, while high (millimolar) concentrations produce a footprinting pattern of solvent-accessible regions, it is critical to perform Tb^{3+} -induced cleavage reactions over a wide range of Tb^{3+} concentrations. To ensure conformational homogeneity, prefolding the RNA under optimized buffer conditions and magnesium concentrations is necessary. This is especially important when trying to identify metal binding sites, as there will be relatively few cleavage events at low Tb^{3+} concentrations. All cleavage reactions should be performed near physiological pH (7.0–7.5) to allow for the accumulation of the cleavage-active $\text{Tb}(\text{OH})(\text{aq})^{2+}$ species [8]. Insoluble polynuclear hydroxo aggregates of Tb^{3+} can form at pH 7.5 and above [13, 14], which should be avoided. Another parameter that needs to be empirically optimized is the temperature and duration of the metal ion-induced cleavage reaction. Higher temperatures not only result in faster cleavage rates but also increase the amount of background degradation. Therefore, typical reaction temperatures range from 25 to 45 °C over a period of 0.5–2 h. All of these parameters need to be well established before carrying out Tb^{3+} footprinting experiments in earnest.

Tb^{3+} Footprinting of Short RNAs

Materials

Reagents and Buffers

- RNA of interest, end-labeled, and gel-purified
- Appropriate buffers to fold the RNA (usually Tris, MES, and/or HEPES of desired pH)
- 1 M MgCl_2 solution
- 100 mM TbCl_3 in 5 mM sodium cacodylate buffer, pH 5.5 (store in small aliquots at -20°C)

- 0.5 M Na₂EDTA, pH 8.0
- 3 M Na(OAc), pH 5.2
- Urea loading buffer: 8 M urea, 50 mM Na₂EDTA, pH 8.0, 0.01% bromophenol blue, 0.01% xylene cyanol.

Equipment

- Heating block at 90 °C
- Water bath
- Phosphor screens and phosphorimager with appropriate software (e.g., Typhoon 9410 Variable Mode Imager with ImageQuant software (GE Healthcare))
- Vacufuge concentrator (Eppendorf).

Protocol 1

- 1) Before performing Tb³⁺-mediated footprinting of an RNA molecule, the RNA should be end-labeled (typically with ³²P at either the 5' or 3' end), purified by denaturing gel electrophoresis, and stored in water (or an appropriate buffer) at –20 °C. For experimental details on ³²P-end-labeling and gel purification of RNA, see Chapters 3 and 9.
- 2) Prepare a single pool with 250 000–500 000 cpm (typically 0.5–2 pmol) of end-labeled RNA per reaction. The total pool volume should be sufficient for single or duplicate reactions at each of the desired Tb³⁺ concentrations. Heat-denature the RNA pool at 90 °C for 2 min under appropriate buffer conditions, but in the absence of divalent metal ions to avoid RNA hydrolysis.
- 3) Prefold the RNA by incubating the pool at an optimized temperature (typically 25–45 °C) for approximately 10 min to ensure structural homogeneity. Some RNAs fold best when a slow-cooling procedure is used, or when certain cations are already added at this stage.
- 4) To obtain the desired Mg²⁺ concentration, add an aliquot of MgCl₂ from an appropriately diluted stock solution and equilibrate at the optimized temperature for an additional 5–10 min. At this point, the total volume of the RNA pool should be 8 µl per reaction aliquot.
- 5) From the 100 mM TbCl₃ stock solution, make a serial set of TbCl₃ dilutions in water, ranging from micromolar to millimolar concentrations (5 × over the final reaction concentration). This wide range of TbCl₃ concentrations should be sufficient to probe for both high-affinity metal binding sites and secondary/tertiary structure formation. *Note: The 100 mM TbCl₃ stock solution is dissolved in a 5 mM sodium cacodylate buffer at pH 5.5 to prevent formation of terbium(III) hydroxide precipitates at higher pH. The TbCl₃ dilutions in water should be made immediately before use. A serial set of dilutions is recommended to ensure consistency in cleavage band intensity between gel lanes. Use a fresh aliquot of 100 mM TbCl₃ stock solution each time. Final TbCl₃ concentrations used in the cleavage reactions should be optimized together with other experimental conditions for the specific RNA and experimental goal.*

- 6) To initiate Tb^{3+} -mediated cleavage, mix an 8- μl aliquot from the refolded RNA pool with 2 μl of an appropriate dilution of TbCl_3 to achieve the desired final Tb^{3+} concentration (typically ranging from 5 μM to 5 mM, in addition to a 0 mM Tb^{3+} control). Continue to incubate at the optimized temperature for an optimized amount of time (typically 30 min to 2 h). *Note: The incubation times should be chosen to generate a partial digestion pattern of end-labeled RNA under single-hit conditions. Extended incubation times will increase secondary hits that may reflect structural distortions of the RNA due to the preceding cleavage event.*
- 7) Quench the cleavage reaction by adding EDTA, pH 8.0, to a final concentration of 50 mM (or at least a twofold excess over the total concentration of multivalent metal ions in the reaction aliquot).
- 8) Perform an ethanol precipitation of the RNA by adding Na(OAc) to a final concentration of 0.3 M and 2–2.5 vol of 100% ethanol and precipitate at -20°C overnight. Centrifuge 30 min at 12 000 g, 4°C . Decant supernatant, wash with 80% (v/v) ethanol, decant supernatant, and dry RNA in a Speedvac evaporator. Redissolve samples in 10–20 μl of urea loading buffer.
- 9) Partial alkaline hydrolysis and RNase T1 digestion reactions of refolded RNA from the same pool should be performed as calibration standards by incubating the end-labeled RNA in the appropriate buffers.
- 10) Heat samples at 90°C for 5 min and place on ice water to snap cool. Analyze the cleavage products on a high-resolution denaturing (8 M urea) polyacrylamide sequencing gel, using the partial alkaline hydrolysis and RNase T1 digestion reactions as size markers to identify the specific Tb^{3+} hydrolysis products at nucleotide resolution. *Note: In the example cited below, an 8 M urea, 20% polyacrylamide gel was poured between 35×45 glass plates with a 1-mm wedged spacer and was run at a constant power of 80 W for separating the cleavage products of a radiolabeled 39-mer RNA. Identical samples can be loaded at different times on the same gel to resolve different regions of longer RNA.*
- 11) Product bands are directly visualized by exposing a phosphor screen to the gel. *Note: The exposure can take several hours to overnight, depending on the level of radioactivity of the bands in the gel.*
- 12) Quantify the full-length RNA and cleavage product bands using a volume count method. (For a more qualitative evaluation, a line scan method can be used.) At every Tb^{3+} concentration, calculate a normalized extent of cleavage (χ) by substituting the peak intensities in the equation:

$$\chi = \frac{\left(\frac{\text{Band intensity at nucleotide } a_x}{\sum_{i=1}^n \text{Band intensity at nucleotide } a_i} \right) \gamma \text{ mM } [\text{Tb}^{3+}]}{\left(\frac{\text{Band intensity at nucleotide } a_x}{\sum_{i=1}^n \text{Band intensity at nucleotide } a_i} \right) 0 \text{ mM } [\text{Tb}^{3+}]}$$

where γ is the Tb^{3+} concentration in a particular cleavage reaction and x is the analyzed nucleotide position of the RNA. 0 mM $[\text{Tb}^{3+}]$ signifies a control reaction incubated in the same manner as the Tb^{3+} -containing ones except that water is added instead of Tb^{3+} solution. A χ value of ≥ 2 indicates significant cleavage over background degradation. *Note: By dividing the ratio of a single-band intensity over total RNA (that is, the sum of all individual bands including the intact RNA) in the presence of Tb^{3+} by the ratio of a single-band intensity over total RNA in the absence of Tb^{3+} , one normalizes for the effect of non-specific background degradation.*

Footprinting Long RNAs by Primer Extension

The protocol in the section “ Tb^{3+} Footprinting of Short RNAs” presents a simple and rapid method for determining the secondary structure and metal binding sites of small RNAs. However, when the length of the RNA of interest is greater than ~100 nt, it becomes difficult to resolve all of the footprinting data on a single sequencing gel, even when loading the same sample multiple times while the gel is running. This difficulty can be overcome by the use of primer extension analysis. Multiple primers complementary to different regions of the RNA of interest (~100 nt apart) can be used to obtain footprinting information for the entire length of a long RNA in a single sequencing gel [15, 16].

Although the primer extension method is very powerful, additional steps are necessary to correctly interpret footprinting data. First, dideoxynucleotide (ddNTP) sequencing reactions must be performed with each primer to determine the sequence of the RNA. When these sequencing ladders are resolved on a gel in parallel with primer extension reactions of footprinted RNA, they enable site-specific identification of the nucleotides cleaved by Tb^{3+} . A primer extension reaction on unmodified RNA must also be resolved in parallel, and any signal in this lane must be subtracted from experimental data lanes to account for low processivity or structurally induced stops of the reverse transcriptase during the primer extension assay [15, 16]. Finally, primer extension analysis has the disadvantage that information at the very ends of the RNA of interest is often lost. Structural information at the 5' end of the RNA can be masked by the heavy signal from the full-length product generated by the necessary single-hit footprinting conditions. Information at the 3' end of the RNA of interest cannot be obtained because the 3'-terminal 20–30 nt are masked by the primer. These problems can sometimes be overcome by adding spacer sequences to the 5' and 3' ends of the RNA [15], although additional experiments should be done to ensure that these regions do not alter the structure and function of the RNA of interest.

Materials

Reagents and Buffers

- RNA of interest, unlabeled, and gel-purified
- $0.5 \times \text{TE}$, pH 8.0 (5 mM Tris-HCl, 0.5 mM EDTA)

- SuperScript III Reverse Transcriptase and included 5× First-Strand Buffer (Life Technologies Invitrogen)
- DNA primer(s) (Life Technologies Invitrogen), 5'-end-labeled with ^{32}P
- 10 mM Deoxynucleotides (dNTPs) (GE Healthcare Life Sciences)
- ddNTP Set, 5 mM Solutions (ddATP, ddCTP, ddGTP, ddTTP) (GE Healthcare Life Sciences)
- 100 mM DTT.

Equipment

- Heating block at 90 °C
- Water bath
- Phosphor screens and phosphorimager with appropriate software (e.g., Typhoon 9410 Variable Mode Imager with ImageQuant software (GE Healthcare)).

Protocol 2

- 1) Design DNA oligomers that are complementary to the RNA of interest, 3' of the region to be probed. DNA oligomers should be 20–30 nt in length, with a melting temperature between 55 and 60 °C. *Note: Primers should be spaced ~100 nt apart along the RNA of interest to obtain reliable structural data for the entire molecule.*
- 2) Label the 5' end of primers with ^{32}P and gel-purify. Redissolve primers in water and store at –20 °C.
- 3) Refold unlabeled RNA of interest as described in Protocol 1, steps 2–4. Refold a pool of RNA (0.5–2 pmol RNA in 8 µl per reaction) accounting for the number of primers, number of Tb^{3+} concentrations to be tested, negative controls, and sequencing lanes (discussed below). The same pool of refolded RNA should be used for all reactions and controls.
- 4) Perform Tb^{3+} footprinting as described in Protocol 1, steps 5–7, but scale up each footprinting reaction by the number of primers needed to probe the entire RNA.
- 5) A negative control (0 mM Tb^{3+}), where water instead of a Tb^{3+} solution is added to the RNA, should be performed in parallel for each primer. This reaction should be scaled up such that there is enough RNA to generate a dideoxy sequencing ladder as well. *Note: A negative control is necessary as non-footprinted RNA will reveal natural stops in the primer extension reaction because of idiosyncrasies of the reverse transcriptase or stable secondary structures in the RNA.*
- 6) Precipitate footprinted RNAs as in Protocol 1, step 8, but redissolve pellets in 10 µl of 0.5 x TE buffer, pH 8.0, per primer extension reaction for probed and unprobed RNA. For dideoxy sequencing reactions, redissolve in only 9 µl of 0.5 x TE buffer, pH 8.0.
- 7) Anneal 3 µl of ^{32}P -end-labeled primer (150 000 cpm, or more for long RNAs) per 10 µl aliquot of footprinted RNA by heating to 65 °C for 5 min, incubating at 45 °C for 2 min, and then cooling on ice.

- 8) Prewarm RNA:primer complex and reverse transcriptase mix (4 μ l 5 $\times$ First-Strand buffer, 1 μ l 100 mM DTT, 1 μ l 10 mM each dNTP, and 1 μ l Superscript III reverse transcriptase) separately for 10 s at 45 °C. Transfer 7 μ l of reverse transcriptase mix to the RNA:primer complex and incubate tubes at 52 °C for 5 min to extend primers. Place reactions at 65 °C for 5 min to inactivate the reverse transcriptase and return reactions to ice.
- 9) Perform dideoxy sequencing reactions as just described in steps 7 and 8, except for the addition of 1 μ l of an individual ddNTP solution at a concentration empirically optimized to produce a banding pattern, before addition of the reverse transcriptase mix.
- 10) Precipitate reactions as described in Protocol 1, step 8, and analyze reactions by denaturing PAGE as described in Protocol 1, steps 10–12.

12.2

Application Example

Tb³⁺ has been successfully used on a number of RNAs as a probe for high-affinity metal binding site localization and tertiary structure determination. For example, taking advantage of its luminescent and RNA footprinting properties, Tb³⁺ has revealed subtle structural differences between the precursor and product forms of the hepatitis delta virus (HDV) ribozyme [5, 6]. The HDV ribozyme is a unique RNA motif found in the human HDV, a satellite of the hepatitis B virus that leads to frequent progression toward liver cirrhosis in millions of patients worldwide [17]. There is a strong interest, both for scientific and medical reasons, in understanding the structure–function relationship of this catalytic RNA. We found that the Tb³⁺-mediated footprinting pattern of the 3' product (3'P) complex of a *trans*-acting version of the HDV ribozyme (Figure 12.2a), obtained in the presence of millimolar Tb³⁺ concentrations, is consistent with a post-cleavage crystal structure. In particular, protection is observed in all five helical stems, P1 through P4 and P1.1, while the backbone of the L3 loop region and that of the J4/2 joining segment are strongly cut (Figure 12.2) [5]. Cuts in J4/2 are particularly relevant, as it encompasses the catalytic residue C75 and its neighboring G76, and the strong Tb³⁺ hits implicate it as a region of high negative charge density with high affinity for metal ions.

Strikingly, Tb³⁺ footprinting reveals the precursor (ncS3) structure as distinct; while P1, P2, P3, and P4 remain protected, both the 5' and 3' segments of the P1.1 stem (as well as U20, immediately upstream) are strongly hydrolyzed, suggesting that this helix in the catalytic core is formed to a lesser extent than in the product complex. In addition, scission in J4/2 extends to A77 and A78, implying that the ribose zipper motif involving these nucleotides may not be fully formed in the precursor complex (Figure 12.2b). These differences in the extent of backbone scission in the precursor *versus* the 3' product complexes show that a significant conformational change occurs upon HDV ribozyme catalysis and 5' product dissociation from the 3' product [5, 6, 11, 12].



Figure 12.2 Tb^{3+} footprinting of the *trans*-acting HDV ribozyme. (a) Synthetic HDV ribozyme construct D1. The ribozyme portion is composed of two separate RNA strands A and B. The 3' product (3'P) strand, shown in gray letters, interacts with the ribozyme by forming helix P1. The substrate variant S3 contains eight additional nucleotides (gray) 5' of the cleavage site (arrow). To generate non-cleavable substrate analogs, the 2'-OH of the underlined nucleotide immediately 5' of the cleavage site was modified to 2'-methoxy and the suffix "nc" added to their name (Figure 12.3). Dashed lines represent tertiary structure hydrogen bonds of C75 and the ribose zipper of A77 and A78 in J4/2. Note that although HDV ribozyme construct D1 lacks the U27 present in other HDV ribozyme variants, the canonical numbering of the HDV ribozyme is maintained here for better comparability. (b) Tb^{3+} - and Mg^{2+} -mediated footprint of the (5'- ^{32}P)-labeled HDV ribozyme strand A upon incubation with Tb^{3+} for 2 h in 40 mM Tris-HCl, pH 7.5, 11 mM MgCl_2 at 25 °C. From left to right as indicated: Strand A

fresh after radiolabeling; incubated in buffer without Tb^{3+} ; incubated with excess strand B in buffer without Tb^{3+} ; incubated with excess strand B and non-cleavable substrate analog ncS3 in buffer without Tb^{3+} ; RNase T1 digest; alkali (OH^-) ladder; footprint of strand A with increasing Tb^{3+} concentrations in the presence of excess strand B and ncS3; strand A incubated in buffer without Tb^{3+} ; incubated with excess strand B in buffer without Tb^{3+} ; footprint with increasing Tb^{3+} concentrations in the presence of excess strand B and 3' product (3'P). As the Tb^{3+} concentration increases, backbone scission becomes more intense. The 5' and 3' segments of P1.1 (boxed) footprint very differently in the precursor and product complexes. Last four lanes on the right: Mg^{2+} -induced cleavage at pH 9.5 and 37 °C; from left to right: precursor (ncS3) complex, control incubated at pH 7.5; precursor complex, footprinted at pH 9.5; product complex, control incubated at pH 7.5; product complex, footprinted at pH 9.5. (Reprinted with permission from Ref. [5].)

While previous evidence from fluorescence resonance energy transfer (FRET) [18], 2-aminopurine fluorescence quenching [19], and NMR spectroscopy [20, 21] had already hinted at structural differences between the precursor and 3' product forms of the *trans*-acting HDV ribozyme, Tb^{3+} -mediated footprinting complements these techniques by providing specifics of these rearrangements at nucleotide resolution. Particularly intriguing are the differences in the catalytic core structure around C75 and P1.1 that may control access to the cleavage transition state and may therefore explain differences in the catalytic rate constants for substrates with different 5' sequences (Figure 12.3) [5, 12]. In fact, the 5' substrate sequence subtly modulates the Tb^{3+} hydrolysis pattern, but all the substrates consistently show strong cuts in the P1.1 stem and the ribose zipper motif in J4/2 (Figure 12.3). This implies that in the precursor, these tertiary structure interactions are not fully formed, in contrast to the 3'P complex. Interestingly, these subtle differences in the catalytic core structure of the various precursor complexes translate into significant changes in FRET efficiency between fluorophores attached to the termini of P4 and P2 [5]. Taken together, these results indicate that the various precursor complexes differ in structure both locally (in the catalytic core) and globally (as measured by FRET), providing an explanation for the wide range of catalytic activities of substrates with varying 5' extensions [5, 12, 22].

Several other laboratories have also found Tb^{3+} to be a useful probe of high-affinity metal binding sites and RNA tertiary structure. Musier-Forsyth and

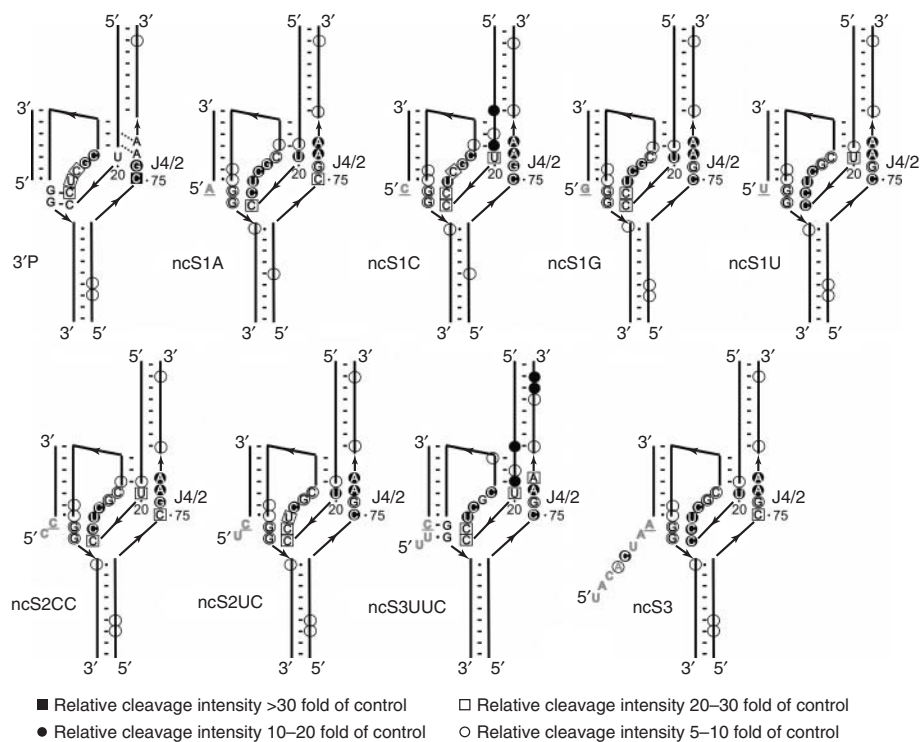


Figure 12.3 Sites of backbone scission mediated by 3 mM Tb^{3+} in 40 mM Tris-HCl, pH 7.5, 11 mM MgCl_2 at 25 °C and superimposed onto 2D representations of the precursor and product HDV ribozyme secondary structures. Only the catalytic core residues are explicitly shown. Relative scission intensities were calculated as described in Protocol 1 and are represented by the symbol code.

Scissions are located 3' of the indicated nucleotides. Only the product structure (3'P) is likely to fully form P1.1 and the ribose zipper of A77 and A78 in J4/2; consequently, dashes indicating the base pairs in P1.1 and dashed lines illustrating the ribose zipper interactions are solely shown for the structure of the 3'P complex. (Reprinted with permission from Ref. [5].)

coworkers were able to show that Tb^{3+} substitutes for several well-known metal binding sites in human tRNA^{Lys,3} and works as a sensitive probe of tertiary structure. At low Tb^{3+} concentrations, cleavage of tRNA^{Lys,3} is restricted to nucleotides that were previously identified by X-ray crystallography as specific metal binding pockets [7]. The use of higher Tb^{3+} concentrations resulted in an overall footprint of the L-shaped tRNA structure, showing increased cleavage in the loop regions (D and anticodon loop). HIV nucleocapsid protein could then be shown to disrupt the tRNA's metal binding pockets and, at higher concentrations, to destabilize the tRNA acceptor-T Ψ C stem domain, as inferred from increased Tb^{3+} cleavage in this region of the tRNA [9]. Other RNAs that have similarly been studied by Tb^{3+} -mediated footprinting include the hammerhead ribozyme [4], an aminoacyl-transferase ribozyme [23, 24], RNase P [25], and group II intron ribozymes [26].

12.3

Troubleshooting

Initial titration experiments will be necessary to obtain the optimal Tb^{3+} concentration(s) to be used for structure probing of any individual RNA (typical Tb^{3+} and RNA concentrations for determining tertiary structural features are $\sim 1\text{--}5\text{ mM}$ and $1\text{ }\mu\text{M}$, respectively). The trivalent Tb^{3+} has been shown to induce slight perturbations of RNA structure [7], but careful titration will reveal the optimal Tb^{3+} :RNA ratio needed for detecting unbiased secondary and tertiary structure features in a given RNA molecule.

To verify a high-affinity metal ion binding site, it is advisable to first decrease the Tb^{3+} concentration until a very limited cleavage pattern is observed (typically at $10\text{--}100\text{ }\mu\text{M}$ Tb^{3+}), and then to perform a competition experiment with increasing concentrations of Mg^{2+} . The intensity (or fraction of RNA cleaved at a particular nucleotide position) should decrease as the Mg^{2+} concentration increases. Quantifying the intensities of cleaved bands at each nucleotide position directly relates to the structure of RNA. It is critical to keep the extent of total cleavage lower than 20% of the uncleaved or full-length band to ensure that each RNA molecule is (at most) undergoing a single cleavage event. Finally, it is important to keep in mind that Tb^{3+} footprinting may not reveal all of the high-affinity metal ion binding sites. This complication may occur when there is steric hindrance preventing Tb^{3+} from binding close to the 2'-hydroxyl group on the ribose. For example, A-type RNA helices have an unfavorable geometry for Tb^{3+} binding, and therefore, strong metal ion binding sites that occur in RNA helical regions may be underestimated or remain undetected when using Tb^{3+} as probe [26].

12.4

Frontiers in Footprinting Data Analysis

A recent development in the field of RNA structure probing is the advent of a high-throughput method for the analysis of primer extension reactions by capillary electrophoresis (CE). This method, pioneered by Kevin Weeks and coworkers [27], bypasses the time-consuming steps of working with radioactivity and running sequencing gels and automates much of the data analysis. Primer extension reactions are carried out with 5'-fluorescently labeled DNA primers, and the cDNA fragments are resolved using a multichannel CE DNA sequencer [27]. The software package ShapeFinder [28] converts output electropherograms of peak intensity *versus* time into nucleotide reactivity *versus* location of that nucleotide. Data conversion subtracts background stops in unmodified RNA and aligns the data to sequencing reactions to determine the identity of reactive nucleotides. This method provides a rapid way to analyze many primer extension reactions simultaneously, and fewer overall primers are needed to sequence a long RNA because 300–600 nt can typically be resolved from a single primer [27, 29]. Any type of footprinting, including Tb^{3+} probing, that generates stops during primer

extension can be analyzed by this method, thus making it a very versatile and rapid way to study structure–function relationships in RNA.

References

- Pyle, A.M. (2002) Metal ions in the structure and function of RNA. *J. Biol. Inorg. Chem.*, **7**, 679–690.
- Draper, D.E., Grilley, D., and Soto, A.M. (2005) Ions and RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, **34**, 221–243.
- Ciesiolka, J., Marciniak, T., and Krzyzosiak, W. (1989) Probing the environment of lanthanide binding sites in yeast tRNA(Phe) by specific metal-ion-promoted cleavages. *Eur. J. Biochem.*, **182**, 445–450.
- Feig, A.L., Panek, M., Horrocks, W.D. Jr., and Uhlenbeck, O.C. (1999) Probing the binding of Tb(III) and Eu(III) to the hammerhead ribozyme using luminescence spectroscopy. *Chem. Biol.*, **6**, 801–810.
- Jeong, S., Sefcikova, J., Tinsley, R.A., Rueda, D., and Walter, N.G. (2003) Trans-acting hepatitis delta virus ribozyme: catalytic core and global structure are dependent on the 5' substrate sequence. *Biochemistry*, **42**, 7727–7740.
- Harris, D.A., Tinsley, R.A., and Walter, N.G. (2004) Terbium-mediated footprinting probes a catalytic conformational switch in the antigenomic hepatitis delta virus ribozyme. *J. Mol. Biol.*, **341**, 389–403.
- Hargittai, M.R. and Musier-Forsyth, K. (2000) Use of terbium as a probe of tRNA tertiary structure and folding. *RNA*, **6**, 1672–1680.
- Walter, N.G., Yang, N., and Burke, J.M. (2000) Probing non-selective cation binding in the hairpin ribozyme with Tb(III). *J. Mol. Biol.*, **298**, 539–555.
- Hargittai, M.R., Mangla, A.T., Gorelick, R.J., and Musier-Forsyth, K. (2001) HIV-1 nucleocapsid protein zinc finger structures induce tRNA(Lys,3) structural changes but are not critical for primer/template annealing. *J. Mol. Biol.*, **312**, 985–997.
- Harris, D.A. and Walter, N.G. (2003) Probing RNA structure and metal-binding sites using terbium(III) footprinting. *Curr. Protoc. Nucleic Acid Chem.*, **6.8**, 6.8.1–6.8.8.
- Sefcikova, J., Krasovska, M.V., Spackova, N., Sponer, J., and Walter, N.G. (2007) Impact of an extruded nucleotide on cleavage activity and dynamic catalytic core conformation of the hepatitis delta virus ribozyme. *Biopolymers*, **85**, 392–406.
- Sefcikova, J., Krasovska, M.V., Sponer, J., and Walter, N.G. (2007) The genomic HDV ribozyme utilizes a previously unnoticed U-turn motif to accomplish fast site-specific catalysis. *Nucleic Acids Res.*, **35**, 1933–1946.
- Baer, C.F. and Mesmer, R.E. (1976) *The Hydrolysis of Cations*, Wiley Interscience, New York.
- Matsumura, K. and Komiyama, M. (1997) Enormously fast RNA hydrolysis by lanthanide(III) ions under physiological conditions: eminent candidates for novel tools of biotechnology. *J. Biochem.*, **122**, 387–394.
- Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
- Tijerina, P., Mohr, S., and Russell, R. (2007) DMS footprinting of structured RNAs and RNA-protein complexes. *Nat. Protoc.*, **2**, 2608–2623.
- Hadziyannis, S.J. (1997) Review: hepatitis delta. *J. Gastroenterol. Hepatol.*, **12**, 289–298.
- Pereira, M.J., Harris, D.A., Rueda, D., and Walter, N.G. (2002) Reaction pathway of the trans-acting hepatitis delta virus ribozyme: a conformational change accompanies catalysis. *Biochemistry*, **41**, 730–740.

19. Harris, D.A., Rueda, D., and Walter, N.G. (2002) Local conformational changes in the catalytic core of the *trans*-acting hepatitis delta virus ribozyme accompany catalysis. *Biochemistry*, **41**, 12051–12061.
20. Luptak, A., Ferre-D'Amare, A.R., Zhou, K., Zilm, K.W., and Doudna, J.A. (2001) Direct pK(a) measurement of the active-site cytosine in a genomic hepatitis delta virus ribozyme. *J. Am. Chem. Soc.*, **123**, 8447–8452.
21. Tanaka, Y., Tagaya, M., Hori, T., Sakamoto, T., Kurihara, Y., Katahira, M., and Uesugi, S. (2002) Cleavage reaction of HDV ribozymes in the presence of Mg^{2+} is accompanied by a conformational change. *Genes Cells*, **7**, 567–579.
22. Shih, I. and Been, M.D. (2001) Energetic contribution of non-essential 5' sequence to catalysis in a hepatitis delta virus ribozyme. *EMBO J.*, **20**, 4884–4891.
23. Flynn-Charlebois, A., Lee, N., and Suga, H. (2001) A single metal ion plays structural and chemical roles in an aminoacyl-transferase ribozyme. *Biochemistry*, **40**, 13623–13632.
24. Vaidya, A. and Suga, H. (2001) Diverse roles of metal ions in acyl-transferase ribozymes. *Biochemistry*, **40**, 7200–7210.
25. Kaye, N.M., Zahler, N.H., Christian, E.L., and Harris, M.E. (2002) Conservation of helical structure contributes to functional metal ion interactions in the catalytic domain of ribonuclease P RNA. *J. Mol. Biol.*, **324**, 429–442.
26. Sigel, R.K., Vaidya, A., and Pyle, A.M. (2000) Metal ion binding sites in a group II intron core. *Nat. Struct. Biol.*, **7**, 1111–1116.
27. Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., and Weeks, K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.*, **6**, e96.
28. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M., and Giddings, M.C. (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*, **14**, 1979–1990.
29. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W. Jr., Swanstrom, R., Burch, C.L., and Weeks, K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.

13

Pb^{2+} -Induced Cleavage of RNA

Leif A. Kirseborn and Jerzy Ciesiolka

13.1

Introduction

Certain metal ions induce degradation of RNA in a non-oxidative manner, and in some RNA molecules this process is exceptionally efficient and specific. The best-known example, yeast tRNA^{Phe}, undergoes specific fragmentation in the D-loop in the presence of Pb^{2+} [1–3] and other ions, for example: Eu^{3+} [4, 5], Mn^{2+} [6], and Mg^{2+} [7, 8]. On the basis of X-ray analysis of yeast tRNA^{Phe} crystals, it was suggested that in order to promote cleavage, Pb^{2+} has to be positioned at an optimal distance from the 2'-OH that acts as the nucleophile [9, 10]. These findings gave rise to an experimental approach that uses Pb^{2+} and other ions to localize high-affinity metal ion binding sites as well as to probe the structure of RNA molecules.

Highly efficient and specific Pb^{2+} -induced cleavages are rather rarely observed. The majority of cleavages is weak and usually comprises several consecutive phosphodiester bonds. Most information on the specificity of Pb^{2+} -induced RNA fragmentation has been obtained from studies on ribosomal 16S RNA [11] and 5S RNAs [12–14]. Cleavages occur preferentially in bulges, loops, and other single-stranded RNA regions except those involved in stacking or other higher order interactions. Double-stranded RNA segments are essentially resistant to breakage. Cleavages are also observed in paired regions that are destabilized by the presence of non-canonical interactions, bulges, or other structural distortions. In general, it seems that the flexibility of the polynucleotide chain determines its sensitivity to Pb^{2+} -induced cleavage [11–15].

It has been suggested [16] that the mechanism proposed for the specific, Pb^{2+} -induced fragmentation of yeast tRNA^{Phe} [9, 10, 17] might account for all types of cleavages induced by metal ions. The simplified mechanism shown in Figure 13.1 is helpful for understanding the relation between RNA structure and sensitivity of a particular RNA region to cleavage.

Experimentally, the data suggests an inverse correlation between the pK_a values for different metal ion hydrates and cleavage rates: Pb^{2+} with a pK_a of 7.2 induces cleavage more efficiently than, for example, Mg^{2+} (pK_a 11.4). This would be in

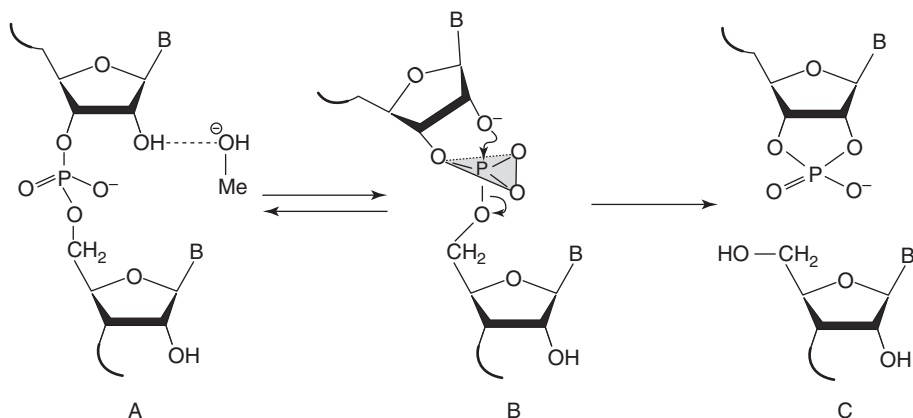


Figure 13.1 Mechanism of metal ion-induced cleavage of RNA. In the first step, the ionized metal ion hydrate acts as a Brønsted base and abstracts a proton from the 2'-OH group of the ribose. Subsequently, the activated anionic 2'-O⁻ attacks the phosphorus atom and a pentacoordinated intermediate is formed. The phosphodiester bond is cleaved, generating 2',3'-cyclic phosphate and 5'-hydroxyl groups as cleavage

products. However, one has also to consider the possibility that the metal ion acts as a Lewis acid by accepting electrons from the 2'-oxygen, thereby facilitating a nucleophilic attack on the phosphorus atom. Irrespective of mechanism, metal ion interaction with the 2'-OH results in a nucleophilic attack on the phosphorus atom and subsequent cleavage of the phosphodiester bond.

keeping with the suggestions that the metal ion either acts as a Lewis acid or Brønsted base. For metal ions with higher pK_a values, the reaction pH, time, or temperature has to be increased and/or, for example, ethanol has to be added to detect substantial cleavage.

The cleavage efficiency of a particular phosphodiester bond in an RNA molecule depends on (i) proper localization of the metal ion hydrate facilitating deprotonation of the 2'-OH group (Figure 13.1, transition A to B), (ii) sufficient conformational flexibility of the analyzed region allowing formation of the pentacoordinated intermediate/transition state and subsequent breaking of the phosphodiester chain (transition B to C). Optimal distance and correct orientation of the bound metal ion hydrate seems to be of primary importance when RNAs undergo efficient and highly specific fragmentation. The cleavage at these sites occurs at relatively low concentrations of Pb^{2+} , below 0.1 mM, conditions under which breakage of other phosphodiester bonds takes place only at significantly reduced rates. Cleavages with lower efficiencies are most likely induced by ions acting from the solution, from weak binding sites, and/or from sites at which the Pb^{2+} -ion is positioned suboptimally. Moreover, metal hydrates interact equally well with all accessible 2'-hydroxyl groups. Thus, differences in rigidity/flexibility of the phosphates, hindering or facilitating conformational transitions necessary for the reaction to occur, influence cleavage efficiency at individual phosphodiester bonds [16]. The contribution of rigidity/flexibility to the cleavage reaction is as yet difficult to assess.

However, the value of the potential rate enhancement derived from constraining a free RNA linkage to an optimal orientation for nucleophilic attack has recently been estimated not to be greater than 50- to 100-fold [18, 19].

13.2

Pb^{2+} -Induced Cleavage to Probe Metal Ion Binding Sites, RNA Structure, and RNA–Ligand Interactions

The Pb^{2+} cleavage approach has been used for the structural analysis of many RNAs and RNA complexes in various ways. For a compilation of some structural analyses of RNAs and RNA complexes using Pb^{2+} and other metal(II)-ions, we refer to Table 13.1 in the first edition of this handbook [20]. These studies can be classified into three groups according to the type of analysis and gain of information: metal ion-induced hydrolysis (i) to identify high-affinity metal ion binding sites, (ii) to probe RNA structure, and (iii) to map RNA–ligand interactions. In Figure 13.2, we show Pb^{2+} -induced cleavage of RNase P RNA in the presence of various divalent metal ions as a typical example. In this context, we also refer to the recent review by Forconi and Herschlag [21].

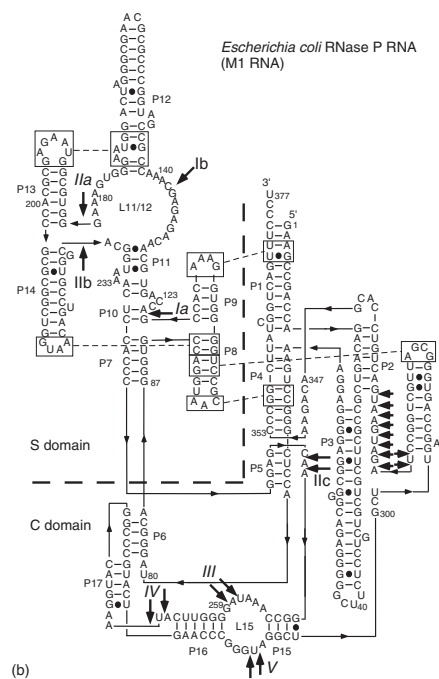
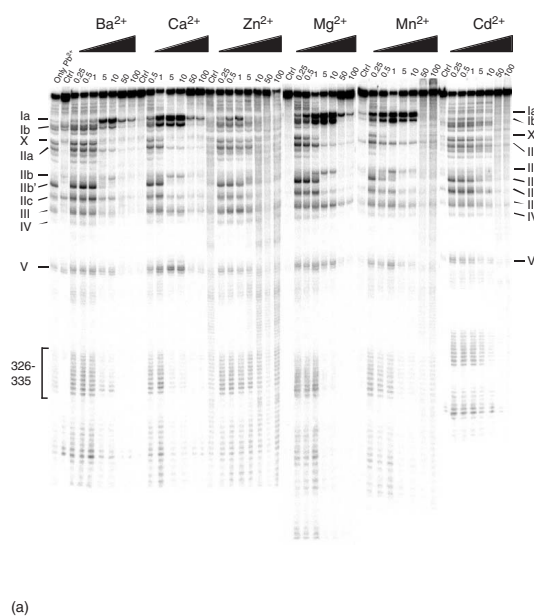
13.2.1

Probing High-Affinity Metal Ion Binding Sites

A strong, highly specific metal ion-induced cleavage suggests the presence of a tight metal ion binding site in the RNA. Cleavage occurring in a particular RNA region does not implicate, however, the direct involvement of that region in coordination of the metal ion. Also, tightly bound metal ions may not induce cleavage at all because of unfavorable distance constraints and/or high rigidity of the polynucleotide chain. For instance, in yeast tRNA^{Phe}, a Pb^{2+} ion induces cleavage in the D-loop but is bound in the TΨC-loop, while the ion positioned in the anticodon loop does not induce specific cleavage [9, 10, 17]. Furthermore, a metal ion binding pocket can usually accommodate different ions, thereby acting as a “general” metal ion binding site. To probe for a “general” metal ion binding site, the following two experimental approaches can be and have been used (Figure 13.2).

The first approach exploits the fact that cleavage by a metal ion (e.g., Pb^{2+}) acting from a “general” metal ion binding site is usually suppressed on addition of other metal ions (e.g., Mg^{2+} , Mn^{2+} , or Ca^{2+}). Quantitative analysis of the inhibition data can give the K_d value for binding of the natural ligand, for example Mg^{2+} , and information about the relative binding affinities of different metal ions for such binding sites in RNA [22, 27].

In the second approach, the RNA is probed in parallel with different metal ions that induce strong cleavages in the same region of the RNA chain. This was originally observed with Pb^{2+} , Mg^{2+} , Mn^{2+} , and Eu^{3+} ions binding in the D-TΨC region of yeast tRNA^{Phe}. Of note, cleavage induced by different metal ions does



not necessarily occur at exactly the same site. Rather, the cleavage sites usually differ by one or two nucleotides. This is rationalized considering the different coordination preferences and sizes of various metal ions, resulting in a slightly different arrangement of their hydrates in metal ion binding pockets. Note that the cleavage reactions should be performed under conditions that take the different pK_a values of metal ion hydrates into account (discussed earlier, also see below). Beside Pb^{2+} , other typical metal ions used in these experiments are Mg^{2+} , Mn^{2+} , Ca^{2+} , and Eu^{3+} .

In addition, Tb^{3+} , which has the same coordination geometry as Mg^{2+} , has been used to probe for metal ion binding sites in RNA. Cleavage of RNA with Tb^{3+} gives different cleavage patterns compared to cleavage with, for example, Pb^{2+} [28–30]. Thus, Tb^{3+} -induced cleavage can be used in combination with cleavage induced by other metal ions, and thereby more information concerning metal ion binding to RNA can be obtained.

13.2.2

Pb^{2+} -Induced Cleavage and RNA Structure

Although Pb^{2+} -induced cleavage is a sensitive tool in RNA structure analysis and permits identifying similarities and differences in related RNA molecules in the regions involved in metal ion binding, one should be aware that Pb^{2+} hydrolysis patterns do not always correspond precisely to RNA secondary structure models. Experimental results are most consistent with cleavage occurring preferentially in “flexible regions” of an RNA polynucleotide chain. Taking into account that our knowledge of RNA conformational dynamics is still insufficient, the term “*flexible regions of RNA*” should be used cautiously in the interpretation of experimental data.

Figure 13.2 Pb^{2+} -induced cleavage patterns of *E. coli* RNase P RNA. (a) Pb^{2+} -induced cleavage pattern in the presence of increasing concentrations of Ba^{2+} , Ca^{2+} , Zn^{2+} , Mg^{2+} , Mn^{2+} , or Cd^{2+} [22]. Cleavage was performed at 37 °C and at the metal ion concentration (in mM) specified above the lane. Lanes: Only Pb^{2+} , incubation in the presence of only 0.5 mM Pb^{2+} ; Ctrl, incubation in the absence of Pb^{2+} but in the presence of Ba^{2+} , Ca^{2+} , Zn^{2+} , Mg^{2+} , Mn^{2+} , or Cd^{2+} , 10 mM (final concentration). (b) Secondary structure of *E. coli* RNase P RNA [23]; Roman numerals refer to the sites of cleavage shown in (a). Roman numerals in *italic* refer to sites where Mg^{2+} -induced cleavage has also been observed [24]. The specificity (S)

and catalytic (C) domains are indicated and separated by the dashed line [25]. In (a) it is apparent that increasing the concentration of different divalent metal ions results in suppression of Pb^{2+} -induced cleavage at specific sites, although to different degrees. This suggests that these metal ions and Pb^{2+} bind to at least overlapping sites (for details, see text). Moreover, these data indicate that different divalent metal ions bind with different affinities to RNase P RNA and that the conformation of *E. coli* RNase P RNA is very similar in the presence of Ba^{2+} , Ca^{2+} , Mg^{2+} , and Mn^{2+} , while it is changed in the presence of others, for example, Cd^{2+} . For further details, see Brännvall *et al.* [22] and with respect to region 326–335, see Hardt *et al.* [26].

Recently, several well-defined RNA secondary structure motifs, such as bulges, hairpin loops, and single-stranded RNA, were characterized by Pb^{2+} -promoted cleavage [16]. These studies have revealed that the cleavage patterns of single nucleotide bulges depend on the structural context provided by adjacent base pairs. In general, a pyrimidine flanking the bulged nucleotide, particularly at its 5' side, facilitates cleavage, while a purine makes the bulge more resistant to cleavage. This effect seems to correlate with the ability of the bulge to form stacking interactions with its neighbors. Cleavage of two- and three-nucleotide bulges depends only slightly on their nucleotide composition. In case of terminal loops, cleavage usually increases with the loop size and strongly depends on its nucleotide composition. Particularly resistant are stable tetraloops, most likely due to their high conformational rigidity. Most single-stranded RNA regions are highly susceptible to Pb^{2+} -induced cleavage. However, tracts of G residues and, in most cases, also phosphodiester bonds at the junction of paired and unpaired RNA regions are more resistant. This can be attributed to extensive stacking interactions and increased conformational rigidity [16]. For some RNAs, however, efficient Pb^{2+} -induced cleavage at the junction between unpaired and double-stranded regions is seen, which can be attributed to increased “breathing” of the base pair at the junction.

Pb^{2+} cleavage is a very sensitive method to detect conformational changes in RNA molecules or to unravel alternative hairpin structures formed by transcripts composed of trinucleotide repeats ([31–33]; see also Figure 13.2 where it is shown that the Pb^{2+} -induced cleavage pattern changes with increasing concentration of other divalent metal ions indicating differences in folding). Several studies have further demonstrated that cleavage of the same structural motifs present in different RNAs results in essentially identical patterns. This raises the interesting possibility of using the Pb^{2+} -induced cleavage approach to identify certain RNA structural motifs in RNA molecules of unknown structure [16, 34–37].

Last, Pb^{2+} -induced cleavage in combination with genetics, that is, introduction of point mutations, has been used to provide support for the existence of long-range interactions in RNA. Note that point mutations may either result in increased/decreased cleavage at specific positions or in the appearance of cleavages at new positions [38, 39].

13.2.3

Pb^{2+} -Induced Cleavage to Study RNA–Ligand Interactions

Remarkable reductions of Pb^{2+} -induced cleavage intensities have been observed upon complexation of RNA with proteins, other RNAs, or low-molecular-weight ligands – amino acids, antibiotics, or other divalent metal ions (see for example [22, 40]).

In RNA–protein complexes, the shielding effect of a bound protein is most likely responsible for changes in cleavage intensities. However, it is not excluded that the RNA changes its conformation because of interaction with protein(s) such that the positioning of Pb^{2+} is altered or that the ion is displaced. In the case

of RNA–aminoglycoside complexes, structural studies have provided evidence for displacement of Pb^{2+} as a result of aminoglycoside binding [41]. Furthermore, there are several examples in the literature of enhanced Pb^{2+} cleavage upon complex formation. Here, moderately enhanced cleavages may suggest increased flexibility of the analyzed RNA regions. The appearance of a very strong cleavage may indicate formation of a new strong metal ion binding site or that a previously inactive metal ion has been repositioned such that efficient metal ion-induced cleavage becomes feasible. In both cases, the presence of a tightly bound ion needs to be verified by other methods, since it is conceivable that strand breakage may also occur with no metal ion bound. Needless to say, this mapping method is obviously restricted to RNA regions susceptible to Pb^{2+} -induced cleavage, mainly bulges, loops, and other single-stranded RNA stretches.

In some cases, changes in the Pb^{2+} cleavage pattern due to ligand binding may include unexpectedly large regions of the polynucleotide chain [40, 42]. This may be caused by a long-range loss of flexibility, that is, formation of a more rigid conformation of a large RNA fragment, which is unstructured in the absence of the ligand. The ligand may thus stabilize one of the near-isoenergetic RNA conformers.

13.2.4

Pb^{2+} -Induced Cleavage of RNA *In Vivo*

Most Pb^{2+} -cleavage studies of RNA have been performed *in vitro*. However, Lindell *et al.* ([43]; see below) developed a protocol for cleavage of RNA with Pb^{2+} *in vivo*. Comparison of *in vitro* with *in vivo* data revealed that essentially the same sites in a given RNA were susceptible to Pb^{2+} -induced cleavage both *in vivo* and *in vitro*, as observed for several non-coding RNAs, mRNAs, and *Escherichia coli* RNase P RNA [43, 44].

Protocols for Metal Ion-Induced Cleavage of RNA

From metal ion-induced cleavage one can conclude that a divalent metal ion(s) is positioned close in space to the site of hydrolysis, but this does not give any information about how the metal ion is coordinated. However, since the 2' hydroxyl in the immediate vicinity of the scissile phosphodiester is actively involved in the chemistry of cleavage (Figure 13.1), some structural constraints for the positioning of the divalent metal ion(s) can be derived. These aspects have to be kept in mind when interpreting the data. Here, we describe three protocols used to cleave RNA with divalent metal ions, *in vitro* using Pb^{2+} and Mg^{2+} and *in vivo* using Pb^{2+} [43, 44]. We use RNase P RNA, the catalytic subunit of the endoribonuclease P, as an example of an RNA that has been studied using these protocols (see for example [45], and unpublished data).

Protocol 1: Pb^{2+} -Induced Cleavage of RNA

- 1) The RNA is ^{32}P -labeled at the 3' end with $[5' \text{-}^{32}P]pCp$ or at the 5' end with $\gamma\text{-}[^{32}P]ATP$ using standard procedures.

- 2) The RNA is purified on a denaturing polyacrylamide gel containing 7 M urea in Tris/borate/EDTA (TBE) buffer (90 mM Tris-borate, pH 8.5, 2.5 mM EDTA) and eluted (see Chapters 1 and 3). The RNA is renatured by incubation for 5 min at 55 °C in water or a buffer of choice.
- 3) Pb^{2+} -induced cleavage of, for example, RNase P RNA: Typically, approximately 20 000–30 000 Cerenkov cpm of labeled RNA are mixed with around 2.5 pmol of unlabeled RNA and preincubated in 50 mM Tris-HCl pH 7.5, 100 mM NH_4Cl , and 10 mM $MgCl_2$ for 10 min at 37 °C. Note that when analyzing metal ion binding and/or the structure of an RNA, you have to adjust the conditions such that the RNA adopts a conformation relevant to what you would like to investigate.
- 4) Cleavage is initiated by the addition of a freshly prepared $Pb(OAc)_2$ solution to a final concentration of 0.5 mM. Depending on the nature of the experiment you can use other concentrations of $Pb(OAc)_2$, but usually not higher than 2 mM (see below). Chloride buffer salts can be used, but for higher concentrations of Pb^{2+} (>2 mM) acetate instead of chloride salts should be used to avoid precipitation of Pb^{2+} ions. The final volume of the reaction is 10 μ l.
- 5) The reaction is terminated after 10–15 min by the addition of 2 volumes of stop solution (9 M Urea, 25 mM EDTA, 0.1% bromophenol blue). The time of incubation in the presence of Pb^{2+} has to be optimized empirically and depends on the nature of the RNA, cleavage conditions, and also on the question you address, such as comparing (i) wild-type RNA *versus* mutant RNA, (ii) RNA with and without protein(s), (iii) RNA in the absence *versus* presence of small ligands such as aminoglycosides, or (iv) RNA with and without an interacting RNA molecule.
- 6) The cleavage products are separated on denaturing gels, where the percentage depends on the size of the RNA under study (usually 6–12% polyacrylamide gels).
- 7) The Pb^{2+} -induced cleavage sites are mapped using size markers and $[OH^-]$ ladders (Chapter 10). It is also possible to map cleavage sites by primer extension analysis (Protocol 4) using primers complementary to specific positions in the RNA. In the case of RNase P RNA, we use 15- to 20 nt long DNA primers.

We emphasize that an increase in the concentration of Mg^{2+} (or some other divalent metal ion such as Mn^{2+}) results in suppression of Pb^{2+} -induced cleavage as illustrated for cleavage of RNase P RNA (Figure 13.2a), suggesting that Mg^{2+} and Pb^{2+} bind, if not to the same, at least to overlapping sites [22]. In combination with the use of genetics (i.e., by using, in our example, RNase P RNA variants) or by studying cleavage of the RNase P RNA–substrate complex, it is also possible to use the Pb^{2+} -induced cleavage to probe for structural changes in RNase P RNA. Note that the formation of RNase P RNA-substrate complexes requires a higher concentration of Mg^{2+} (≥ 20 mM). Therefore, an increased concentration of Pb^{2+} is needed to detect cleavage. This might also apply when other RNA molecules are studied, and hence the conditions have to be adjusted accordingly.

Protocol 2: Cleavage of RNA in the Presence of Mg^{2+}

RNase P RNA is also cleaved by other divalent metal ions such as Mg^{2+} , first described by Kazakov and Altman [24]. However, Mg^{2+} -induced cleavage of RNase P RNA is less efficient compared to Pb^{2+} -induced cleavage (see above); in order to detect cleavage, the reaction has to be performed at a higher pH and in the presence of 10% ethanol. Hence, steps 3 and 4 of Protocol 1 are modified.

- 1) In our studies of RNase P RNA, we have used the following conditions: 50 mM CHES buffer, pH 9.5, 100 mM NH_4Cl , 10 MgCl_2 (higher concentrations of Mg^{2+} can be used), and 10% ethanol [24].
- 2) The reaction mixture is incubated at 37°C for 6 h.
- 3) The reaction is terminated and cleavage products are separated and characterized as described in Protocol 1.

Protocol 3: Pb^{2+} -Induced Cleavage of RNA *In Vivo*

Here, the protocol is adapted to study cleavage in growing bacteria, for example, *E. coli* [43, 44]. See also Chapter 10 by Huntzinger *et al.* [46] in the first edition of this handbook.

- 1) Typically, *E. coli* cells are grown in Luria-Broth (LB) medium overnight at 37°C (or another temperature of choice).
- 2) The culture is diluted 400-fold in LB and allowed to grow to an $\text{OD}_{600} \approx 0.5$ (mid-log phase).
- 3) Freshly prepared $\text{Pb}(\text{OAc})_2$ solutions in sterile H_2O (note that lead(II)-acetate precipitates at high concentrations in LB medium) of appropriate concentrations are prepared and diluted with prewarmed (37°C) 4 x concentrated LB medium (3 volumes of $\text{Pb}(\text{OAc})_2$ and 1 volume of 4 x LB). Hence, to give a final (theoretical) concentration of 100 mM in the *E. coli* cell suspension (step 4), that typically has been used, the freshly prepared $\text{Pb}(\text{OAc})_2$ solution should be 467 mM. For reproducibility, mixing has to be performed rapidly. Note that when LB and $\text{Pb}(\text{OAc})_2$ are mixed, there is always some precipitation and hence the final concentration of $\text{Pb}(\text{OAc})_2$ in solution is lower. Moreover, replacing LB with minimal media results in substantial precipitation and poor RNA cleavage.
- 4) Then 1 volume of the $\text{Pb}(\text{OAc})_2/\text{LB}$ solution ($[\text{Pb}(\text{OAc})_2] = 350 \text{ mM}$) is added to 2.5 volumes of cell culture ($\text{OD}_{600} \approx 0.5$, see above) and incubated for 7 min at 37°C under vigorous shaking (the total volume is 3.5 volumes = V_{tot}). The final concentration of $\text{Pb}(\text{OAc})_2 = 100 \text{ mM}$ (not considering the precipitate).
- 5) The reaction is stopped by adding excess of EDTA to a final concentration of 1.5 times the $\text{Pb}(\text{OAc})_2$ concentration, typically 150 mM final concentration of EDTA.
- 6) The cell suspension is put on ice.

- 7) Cells are harvested by centrifugation and the cell pellet is snap frozen in liquid nitrogen and stored at -70°C .
- 8) The cell pellet is resuspended in $0.5 \times V_{\text{tot}}$ preheated (65°C) lysis buffer (100 mM Tris-HCl pH 7.5, 40 mM EDTA, 200 mM NaCl, 0.5% [w/v] SDS) and incubated at 65°C for 3–5 min.
- 9) This is followed by the addition of preheated (65°C) phenol solution (volume: $0.5 \times V_{\text{tot}}$). To prepare the phenol solution, 1 volume phenol is saturated with 1 volume 10 mM Tris-HCl pH 8.0, 10 mM EDTA. The RNA is extracted at 65°C .
- 10) The phenol extraction is followed by chloroform/isoamylalcohol (24 : 1) extraction at room temperature and ethanol precipitation in the presence of 0.3 M sodium acetate, pH 6.0.
- 11) The RNA is dissolved in DNase buffer (40 mM Tris-HCl pH 7.9, 100 mM NaCl, 60 mM MgCl_2 , 1 mM CaCl_2) and 80 units of DNase I (RNase-free) are added. This mixture is incubated for 15 min at 37°C , followed by standard phenol extraction and ethanol precipitation in the presence of 0.3 M sodium acetate, pH 6.0.
- 12) The RNA is redissolved in RNase-free water, snap frozen in liquid nitrogen, and stored at -70°C .
- 13) The Pb^{2+} -induced cleavage sites are mapped by primer extension analysis using appropriate oligodeoxyribonucleotides as primers (Protocol 4).

Note that RNA samples prepared from untreated cells (no Pb^{2+} added) have to be analyzed in parallel to RNA prepared from Pb^{2+} -treated cells. Hence, $\text{Pb}(\text{OAc})_2$ is omitted in step 4 by replacing $\text{Pb}(\text{OAc})_2$ with RNase-free water, and the RNA is prepared according to the same procedure as outlined above. This is an essential control to be able to differentiate signals (stops) in the primer extension analysis that are related to Pb^{2+} -induced cleavage from those that are due to pausing or premature termination sites of reverse transcriptase. The following three comments are directly taken from Chapter 10 of the first edition of this handbook, with minor changes [46].

- 1) Pb^{2+} -induced hydrolysis *in vivo* can be performed under different growth conditions. However, lead(II)-acetate has some tendency to precipitate, in particular when chloride ions are present within the medium. It is thus essential to screen a wide range of lead(II)-acetate concentrations. One simple and reliable method for evaluating cleavage conditions is to fractionate total RNA derived from Pb^{2+} -treated cells on agarose gels [43]. With increasing lead(II) acetate concentrations (25–200 mM), the intensities of the bands corresponding to full-length 16S and 23S rRNA, the predominant cellular RNA species, should decrease substantially. Optical inspection of the changes in RNA fragmentation can be used to adjust the conditions such that Pb^{2+} -induced hydrolysis is significant but not too extensive.
- 2) Other protocols can be used to extract total RNA. For example, reagents combining phenol and guanidine thiocyanate have been shown to enable a

straightforward isolation of total RNA from samples of human, yeast, bacterial, and viral origin.

- 3) Data from *in vivo* probing may be more complex to interpret than *in vitro* probing. This will be the case if the studied RNA is simultaneously involved in several complexes (e.g., regulatory RNAs with multiple targets). However, *in vivo* mapping becomes powerful when it is used in a comparative manner, for instance, Pb^{2+} -induced cleavages can be used to monitor the conformational changes of RNA under different environmental conditions such as temperature variations.

Protocol 4: Primer Extension

- 1) An appropriate 5'- ^{32}P -end-labeled oligodeoxyribonucleotide is mixed with 10 μg of total cellular RNA from step 12 (Protocol 3) in RNase-free water and incubated for 1 min at 90°C .
- 2) The mixture is put on ice for 1 min followed by warming at 20°C for 5 min.
- 3) The actual primer extension is performed in a total volume of 15 μl in 50 mM Tris-HCl pH 8.5, 6 mM MgCl_2 , 40 mM KCl, and 5'-dNTPs (1.0 mM each) and 200 U of reverse transcriptase (for example, Superscript II, Life Technologies). The primer extension mixture is incubated at 45°C for 30 min.
- 4) The reaction is terminated by the addition of 20 μl stop buffer (50 mM Tris-HCl pH 7.5, 0.1% SDS) and 3.5 μl of 3 M KOH. This solution is incubated for 3 min at 90°C followed by incubation at 37°C for 3 h.
- 5) Add 6 μl of 3 M acetic acid to neutralize the KOH and ethanol-precipitate the cDNA in the presence of 0.3 M sodium acetate, pH 6.0. The products are separated on denaturing gels (step 6, Protocol 1).

13.3

Troubleshooting

13.3.1

No Pb^{2+} -Induced Cleavage Detected

- Old solution of Pb^{2+} . The action is to prepare a new solution.
- Your crystalline $\text{Pb}(\text{OAc})_2$ is old or has moistened. The action is to buy a new bottle of solid $\text{Pb}(\text{OAc})_2$.
- Cleavage conditions are not optimized with respect to time of incubation, concentration of other divalent metal ions (e.g., Mg^{2+}), or concentration of $\text{Pb}(\text{OAc})_2$. The action is to optimize the conditions: increase/decrease concentration of Mg^{2+} and/or Pb^{2+} , vary the time of incubation.
- Poor quality of RNA and/or the RNA solution contains metal ions that interfere with Pb^{2+} -induced cleavage or is contaminated with metal ion chelators, for example, EDTA. The action is to prepare a new batch of RNA.
- The pH is too low. The action is to increase the pH.

13.3.2

Complete Degradation of the RNA

- Too high concentration of Pb^{2+} . The action is to decrease $[Pb^{2+}]$ and/or time of incubation.
- Time of incubation is too long. The action is to decrease the time and/or to decrease $[Pb^{2+}]$.
- The pH of the reaction is too high. The action is to lower pH.
- Contamination of your solutions with RNase. The action is to change all solutions (by experience, the RNase-free water is usually the candidate that is most often contaminated).

13.3.3

In Vivo

This section is taken and adapted from Chapter 10 of the first edition of this handbook [46].

- Low yield of total RNA; incomplete homogenization or lysis of cells, degradation of the RNA.
- Strong stops in the control lanes; degradation of RNA or pauses of reverse transcriptase due to stable secondary structures. Suggested actions: perform the reverse transcription at a higher temperature, increase the enzyme and/or dNTP concentrations, or change the primer sequence.
- Many RNA molecules carry posttranscriptional modifications that may interfere with cDNA synthesis by reverse transcriptase. Suggested action: primer should be changed in order to cover the modified base.
- No more full-length RNA product after modification (the reaction was too extensive). Suggested actions: reduce either the quantity of the reagent and/or the time of incubation. Also, make sure that the reaction is stopped before the extraction of total RNA.
- Weak or smearing signal after primer extension. Suggested actions: increase the concentration of total RNA and/or check that the primer hybridization protocol is efficient.

Acknowledgments

We are grateful to Dr. Pascale Romby for parts of the text that relate to Pb^{2+} -induced cleavage *in vivo* and permission to use this in the present chapter. The ongoing research in the laboratories of LAK and JC are supported by the Swedish Research-Council (LAK) and Uppsala RNA Research Center (Linné support to LAK), and the Polish Ministry of Science and Higher Education (grant no. NN301272037 to JC).

References

1. Werner, C., Krebs, B., Keith, G., and Dirheimer, G. (1976) Specific cleavages of pure tRNAs by plumbous ions. *Biochim. Biophys. Acta*, **432**, 161–175.
2. Sampson, J.R., Sullivan, F.X., Behlen, L.S., Drenzo, A.B., and Uhlenbeck, O.C. (1987) Characterization of two RNA-catalyzed RNA cleavage reactions. *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 267–275.
3. Krzyzosiak, W.J., Marciniak, T., Wiewiórowski, M., Romby, P., Ebel, J., and Giegé, R. (1988) Characterization of the lead(II)-induced cleavages in tRNAs in solution and effect of the Y-base removal in yeast tRNAPhe. *Biochemistry*, **27**, 5771–5777.
4. Rordorf, B.F. and Kearns, D.R. (1976) Effects of europium (III) on the thermal denaturation and cleavage of transfer ribonucleic acids. *Biopolymers*, **15**, 1491–1504.
5. Ciesiołka, J., Marciniak, T., and Krzyzosiak, W.J. (1989) Probing the environment of lanthanide binding sites in yeast tRNA(Phe) by specific metal-ion-promoted cleavages. *Eur. J. Biochem.*, **182**, 445–450.
6. Wrzesinski, J., Michałowski, D., Ciesiołka, J., and Krzyzosiak, W.J. (1995) Specific RNA cleavages induced by manganese ions. *FEBS Lett.*, **374**, 62–68.
7. Wintermeyer, W. and Zachau, G. (1973) Mg^{2+} -catalyzed specific splitting of tRNA. *Biochim. Biophys. Acta*, **299**, 82–90.
8. Marciniak, T., Ciesiołka, J., Wrzesinski, J., and Krzyzosiak, W.J. (1989) The response of the double strand-specific nuclease V1 to Y-base removal in yeast tRNA(Phe). *Acta Biochim. Pol.*, **36**, 115–122.
9. Brown, R.S., Hingerty, B.E., Dewan, J.C., and Klug, A. (1983) Pb(II)-catalyzed cleavage of the sugar-phosphate backbone of yeast tRNAPhe—implications for lead toxicity and self-splicing RNA. *Nature*, **303**, 543–546.
10. Rubin, J.R. and Sundaralingam, M. (1983) Lead ion binding and RNA chain hydrolysis in phenylalanine tRNA. *J. Biomol. Struct. Dyn.*, **1**, 639–646.
11. Górnicki, P., Baudin, F., Romby, P., Wiewiórowski, M., Krzyzosiak, W.J., Ebel, J.P., Ehresmann, C., and Ehresmann, B. (1989) Use of lead(II) to probe the structure of large RNA's. Conformation of the 3' terminal domain of *E. coli* 16S rRNA and its involvement in building the tRNA binding sites. *J. Biomol. Struct. Dyn.*, **6**, 971–984.
12. Brunel, C., Romby, P., Westhof, E., Ehresmann, C., and Ehresmann, B. (1991) Three-dimensional model of *Escherichia coli* ribosomal 5S rRNA as deduced from structure probing in solution and computer modeling. *J. Mol. Biol.*, **221**, 293–308.
13. Ciesiołka, J., Lorenz, S., and Erdmann, V.A. (1992) Structural analysis of three prokaryotic 5S rRNA species and selected 5S rRNA-ribosomal-protein complexes by means of Pb(II)-induced hydrolysis. *Eur. J. Biochem.*, **204**, 575–581.
14. Ciesiołka, J., Lorenz, S., and Erdmann, V.A. (1992) Different conformational forms of *Escherichia coli* and rat liver 5S rRNA revealed by Pb(II)-induced hydrolysis. *Eur. J. Biochem.*, **204**, 583–589.
15. Jovine, L., Djordjevic, S., and Rhodes, D. (2000) The crystal structure of yeast phenylalanine tRNA at 2.0 Å resolution: cleavage by $Mg(2+)$ in 15-year old crystals. *J. Mol. Biol.*, **301**, 401–414.
16. Ciesiołka, J., Michałowski, D., Wrzesinski, J., Krajewski, J., and Krzyzosiak, W.J. (1998) Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs. *J. Mol. Biol.*, **275**, 211–220.
17. Brown, R.S., Dewan, J.C., and Klug, A. (1985) Crystallographic and biochemical investigation of the lead(II)-catalyzed hydrolysis of yeast phenylalanine tRNA. *Biochemistry*, **24**, 4785–4801.
18. Mikkola, S., Kaukinen, U., and Lönnberg, H. (2001) The effect of secondary structure on cleavage of the phosphodiester bonds of RNA. *Cell Biochem. Biophys.*, **34**, 95–119.

19. Emilsson, G.M., Nakamura, S., Roth, A., and Breaker, R.R. (2003) Ribozyme speed limits. *RNA*, **9**, 907–918.
20. Kirsebom, L.A. and Ciesiołka, J. (2005) Pb^{2+} -induced cleavage of RNA in *Handbook of RNA Biochemistry*, 1st edn, vol. 1 (eds R.K. Hartmann, A. Bindereif, A. Schön, and E. Westhof), Wiley-VCH Verlag GmbH & Co. KGa, Weinheim, pp. 214–228.
21. Forconi, M. and Herschlag, D. (2009) Metal ion-based RNA cleavage as a structural probe. *Methods Enzymol.*, **468**, 91–106.
22. Brännvall, M., Mikkelsen, N.E., and Kirsebom, L.A. (2001) Monitoring the structure of *Escherichia coli* RNase P RNA in the presence of various divalent metal ions. *Nucl. Acids Res.*, **29**, 1426–1432.
23. Massire, C., Jaeger, L., and Westhof, E. (1998) Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.*, **279**, 773–793.
24. Kazakov, S. and Altman, S. (1991) Site-specific cleavage by metal ion cofactors and inhibitors of M1 RNA, the catalytic subunit of RNase P from *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 9193–9197.
25. Pan, T. (1995) Higher order folding and domain analysis of the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry*, **34**, 902–909.
26. Hardt, W.-D., Schlegel, J., Erdmann, V.A., and Hartmann, R.K. (1993) Role of the D arm and the anticodon arm in tRNA recognition by eubacterial and eukaryotic RNase P enzymes. *Biochemistry*, **32**, 13046–13053.
27. Labuda, D., Nicoghossian, K., and Cedergren, R. (1985) Cooperativity in low-affinity Mg^{2+} binding to tRNA. *J. Biol. Chem.*, **260**, 1103–1107.
28. Flynn-Charlebois, A., Lee, N., and Suga, H. (2001) A single metal ion plays structural and chemical roles in an aminoacyl-transferase ribozyme. *Biochemistry*, **40**, 13623–13632.
29. Kaye, N.M., Zahler, N.H., Christian, E.L., and Harris, M.E. (2002) Conservation of helical structure contributes to functional metal ion interactions in the catalytic domain of ribonuclease P RNA. *J. Mol. Biol.*, **324**, 429–442.
30. Sigel, R.K., Vaidya, A., and Pyle, A.M. (2000) Metal ion binding sites in a group II intron core. *Nat. Struct. Biol.*, **7**, 1111–1116.
31. Napierała, M. and Krzyzosiak, W.J. (1997) CUG repeats present in myotonin kinase RNA form metastable “slippery” hairpins. *J. Biol. Chem.*, **272**, 31079–31085.
32. Jasinska, A., Michlewski, G., De Mezer, M., Sobczak, K., Kozłowski, P., Napierała, M., and Krzyzosiak, W.J. (2003) Structures of trinucleotide repeats in human transcripts and their functional implications. *Nucl. Acids Res.*, **31**, 5463–5468.
33. Sobczak, K., De Mezer, M., Michlewski, G., Krol, J., and Krzyzosiak, W.J. (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucl. Acids Res.*, **31**, 5469–5482.
34. Ciesiołka, J. and Krzyzosiak, W.J. (1996) Structural analysis of two plant 5S rRNA species and fragments thereof by lead-induced hydrolysis. *Biochem. Mol. Biol. Int.*, **39**, 319–328.
35. Matysiak, M., Wrzesinski, J., and Ciesiołka, J. (1999) Sequential folding of the genomic ribozyme of the hepatitis delta virus: structural analysis of RNA transcription intermediates. *J. Mol. Biol.*, **291**, 282–294.
36. Dutkiewicz, M. and Ciesiołka, J. (2005) Structural characterization of the highly conserved 98-base sequence at the 3' end of HCV RNA genome and the complementary sequence located at the 5' end of the replicative viral strand. *Nucl. Acids Res.*, **33**, 693–703.
37. Dutkiewicz, M., światkowska, A., Figlerowicz, M., and Ciesiołka, J. (2008) Structural domains of the 3'-terminal sequence of the hepatitis C virus replicative strand. *Biochemistry*, **47**, 12197–12207.
38. Mattsson, J.G., Svärd, S.G., and Kirsebom, L.A. (1994) Characterization of the *Borrelia burgdorferi* RNase P RNA gene reveals a novel tertiary interaction. *J. Mol. Biol.*, **241**, 1–6.

39. Svård, S.G., Mattsson, J.G., Johansson, K.E., and Kirsebom, L.A. (1994) Cloning and characterization of the RNase P RNA genes from two porcine mycoplasmas. *Mol. Microbiol.*, **11**, 849–859.
40. Ciesiołka, J. and Yarus, M. (1996) Small RNA-divalent domains. *RNA*, **2**, 785–793.
41. Mikkelsen, N.E., Johansson, K., Virtanen, A., and Kirsebom, L.A. (2001) Aminoglycoside binding displaces a divalent metal ion in a tRNA-neomycin B complex. *Nat. Struct. Biol.*, **8**, 510–514.
42. Wallace, S.T. and Schroeder, R. (1998) *In vitro* selection and characterization of streptomycin-binding RNAs: recognition discrimination between antibiotics. *RNA*, **4**, 112–123.
43. Lindell, M., Romby, P., and Wagner, E.G. (2002) Lead(II) as a probe for investigating RNA structure *in vivo*. *RNA*, **8**, 534–541.
44. Lindell, M., Brännvall, M., Wagner, E.G.H., and Kirsebom, L.A. (2005) Lead(II) cleavage analysis of RNase P RNA *in vivo*. *RNA*, **11**, 1348–1354.
45. Altman, S. and Kirsebom, L.A. (1999) in *The RNA World*, 2nd edn (eds R.F. Gesteland, T.R. Cech, and J.F. Atkins), Cold Spring Harbor Laboratory Press, New York, pp. 351–380.
46. Huntzinger, E., Possedko, M., Winter, F., Moine, H., Ehresmann, C., and Romby, P. (2005) Probing RNA Structures with Enzymes and Chemicals *In Vitro* and *In Vivo* in *Handbook of RNA Biochemistry*, 1st edn, vol. 1 (eds R.K. Hartmann, A. Bindereif, A. Schön, and E. Westhof), Wiley-VCH Verlag GmbH & Co. KGa, Weinheim, pp. 151–171.

14

Identification and Characterization of Metal Ion Coordination Interactions with RNA by Quantitative Analysis of Thiophilic Metal Ion Rescue of Site-Specific Phosphorothioate Modifications

Michael E. Harris

14.1

Introduction

The essential roles that RNA plays in biology depend on the association of divalent metal ions [1–4]. Most of these divalent metal ions associate with RNA *via* weak electrostatic interactions and are loosely bound, constituting a population that has been referred to as the *ion atmosphere* [1, 5–8]. These diffuse ion interactions are essential because they act to neutralize the enormous negative charge present on RNA. This charge neutralization allows RNA to fold into compact structures necessary for function. A smaller number of metal ions bind in a localized manner at specific sites within the folded RNA molecule [9–11]. These site-bound ions are involved in stabilizing RNA structure motifs and promoting the formation of long-range tertiary contacts. In addition, site-bound metal ions act as essential catalytic cofactors in the active sites of many RNA enzymes and protein enzymes that catalyze RNA phosphoryl transfer reactions [12–15].

Numerous site-bound ion interactions that include both monovalent and divalent ion binding sites have been revealed by high-resolution X-ray structures of representatives of most classes of RNAs [16–22]. Comparing these sites reveals that chemical interactions between ions and RNA include H-bonding between RNA functional groups and water molecules in the inner hydration sphere of the bound ion. In addition, for over 50% of site-specifically bound ions, their interactions include displacement of a coordinated water and direct inner sphere coordination to an electronegative RNA functional group primarily a purine N7 or non-bridging phosphoryl oxygen [19]. Computational simulations correlate regions of high negative electrostatic potential with these sites [22]. However, to understand how the binding of individual ions or classes of ions is linked to function presents significant experimental challenges due to the large number of ion interactions involved and the interdependence of their association free energies.

The use of RNA with a single phosphorothioate (PS) modification, in conjunction with functional measurements in the presence of metal ions with sufficiently high affinity for sulfur, provides a powerful means to correlate binding of individual ions with function [23–26]. Other biochemical methods described in this volume

including nucleotide analog interference mapping (NAIM) (Chapters 17 and 18) are highly complementary because they can be used to identify positions where nucleotide analog incorporation including phosphorothioates interfere with function. This information can be used to target incorporation of site-specific PS modifications by “DNA splint”-directed RNA ligation (Chapter 3) for more detailed kinetic and thermodynamic analyses including quantitative metal ion rescue analyses, as described in this chapter.

14.1.1

Thiophilic Metal Ion Rescue of RNA Phosphorothioate Modifications

NMR analysis of the stability constants of metal ions for nucleotide phosphates and phosphorothioates indicates that Mg^{2+} prefers oxygen over sulfur by a factor of >10 000-fold, while this difference is much smaller for other “thiophilic” ions such as Cd^{2+} [27]. Metal–ligand specificity is generally understood in terms of a framework developed by Pearson [28], which states that small non-polarizable (“hard”) Lewis acids interact more strongly with small non-polarizable Lewis bases. Conversely, highly polarizable (“soft”) Lewis acids interact more strongly with large highly polarizable Lewis bases. Mg^{2+} is an example of a hard Lewis acid, and oxygen is a hard Lewis base and so they interact strongly. Sulfur on the other hand is a softer Lewis base than oxygen and interacts preferentially with softer metal ions such as Mn^{2+} , Zn^{2+} , and Cd^{2+} .

These preferential interactions between different metal ions and their ligands permit putative contacts to be probed by incorporating a PS modification into the RNA at the site of interest. If the PS substitution prevents Mg^{2+} binding, then biological function will be proportionally decreased. Including in the reaction a softer more “thiophilic” metal ion such as Mn^{2+} or Cd^{2+} can result in binding of those ions at the site of PS modification, and therefore fulfilling the catalytic or structural role of the displaced Mg^{2+} ion. Rescue of the deleterious effects of PS modification by thiophilic metal ions is a good evidence for direct coordination. By quantifying the concentration dependence of the rescue, binding free energies of the rescuing ion can be obtained. By investigating how RNA variants with specific structural changes or how different reaction conditions (pH and varying concentrations of other divalent metal ions) affect the affinity of the rescuing ion, the proposed ion interactions at the PS site can be further explored.

PS modification and thiophilic metal ion rescue to reveal sites of functional divalent metal ion coordination interactions were originally developed for studies of protein enzymes by Cohn and also by Eckstein [29, 30]. Piccirilli and Cech first applied this approach to identify coordination interactions between active site metal ions and substrate phosphoryl oxygens in the group I intron ribozyme [23]. Herschlag and colleagues further developed this approach by quantitatively analyzing the concentration dependence of thiophilic metal ion rescue in order to distinguish different active site ions [31]. In the case of the hammerhead ribozyme, this method was used to provide evidence for interaction of the same active site metal ion with both the reactive phosphoryl and a ribozyme non-bridging

phosphoryl oxygen distant from the cleavage site [32]. Indeed, this interaction was not observed in initial structures of the minimal hammerhead ribozyme (HHRz) [33], yet it was ultimately observed in a high-resolution X-ray structure of the optimally active extended form of the hammerhead ribozyme [34, 35]. The most extensive application has been the combined effort of the Herschlag and Piccirilli groups to use quantitative thiophilic metal rescue of PS modification to reveal both ribozyme and substrate interactions in the group I intron ribozyme [15, 24, 25, 31, 36]. The transition state interactions identified in this manner were consistent with those inferred from an X-ray ground state structure of a group I intron ribozyme [14], illustrating the combined power of functional and structural investigations to reveal mechanistic detail. These correlations provided a strong validation of the crystal structure as an accurate picture of the catalytic form of the RNA structure.

Similar PS-sensitive sites have been identified in most RNAs, but quantitative analysis has not been widely applied. In addition, there have been only a few applications to date to protein enzymes where this method could have a significant impact. In this chapter, the general method is illustrated in its application to characterizing active site metal ion interactions in the bacterial ribonuclease P (RNase P) enzyme–substrate complex [37, 38]. RNase P is an essential ribonucleoprotein enzyme that catalyzes the formation of the tRNA 5'-end *via* endonucleolytic cleavage. The enzyme was historically proposed to act as a metalloenzyme [39, 40]. Hartmann and colleagues demonstrated interference by an Rp phosphorothioate and rescue by Cd^{2+} demonstrating active site metal coordination to the Rp non-bridging phosphoryl oxygen of precursor tRNA (ptRNA) [41, 42] (Figure 14.1). Cooperative concentration dependence was observed when total metal ion was titrated, indicating coordination of two metal ions to the Rp position. Our laboratory subsequently used the quantitative metal ion rescue approach to reproduce the observation of two active site metal ions coordinating to the Rp non-bridging oxygen. These results are an integral component together with a range of structural, biochemical, and biophysical data in the current model for catalysis by RNase P (Figure 14.1) (see [37, 43–45] and references therein). Furthermore, results obtained using the quantitative approach described here and combined with the use of ptRNA substrates containing an Rp phosphorothioate modification at the RNase P cleavage site revealed that binding of the RNase P protein subunit increases the affinity of these essential, active site metal ion cofactors [46].

The general scheme for the experiment is to follow the reaction of the unmodified RNA and the RNA containing the site-specific PS modification first in the presence of Mg^{2+} alone and then at the same concentration of Mg^{2+} with increasing concentrations of Mn^{2+} or Cd^{2+} . Here, the rescuing metal ion is given as “Me” and can represent whichever rescuing metal ion is optimally functional. A simple kinetic framework for the reaction is shown in Figure 14.2a. In this scheme, K_{Me} is the affinity of the rescuing metal ion and k_{Me} is the rate constant when a single rescuing ion or multiple rescuing ions with absolute cooperativity are bound to the enzyme-substrate (ES) complex. The ratio of the rate constants for the PS-modified *versus* unmodified substrate RNA is $k_{\text{rel}} = k_{\text{obs(PS)}}/k_{\text{obs(PO)}}$. The relative reaction rates are analyzed in order to control for any non-specific effects of the

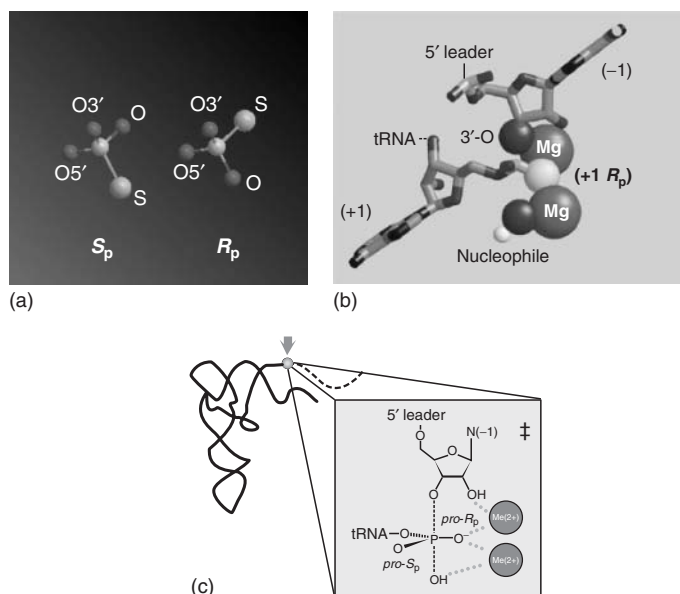


Figure 14.1 Phosphorothioate stereoisomers probe metal ion interactions by disrupting inner sphere coordination with non-bridging phosphate substituents. (a) Phosphorothioate stereoisomers. The $O3'$, $O5'$, and non-bridging O and S atoms are indicated. (b) Proposed active site interactions between metal ions and the reactive phosphorol group of ptRNA in the RNase P active site. The nucleotides $5'(-1)$ and $3'(+1)$ to the cleavage site are indicated. The sphere in light gray sandwiched between the two Mg^{2+} ions indicates the position of the sulfur group in the R_p phosphorothioate

at the reactive phosphodiester. The nucleophilic water molecule and $3'-O$ leaving group are indicated (dark sphere at the upper Mg^{2+} ion and dark sphere plus small white sphere at the lower Mg^{2+}). (c) The polynucleotide chain of tRNA is indicated by a black ribbon and the nucleotides upstream of the cleavage site that constitute the $5'$ -leader sequence removed by RNase P are indicated by a dashed line. A transition state model of the reactive phosphodiester is shown in the box, pinpointing the location of active site metal ion interactions.

rescuing metal ion on the function of the RNA under analysis. The observed k_{rel} will necessarily depend on the fraction of RNA with a metal ion bound to the site of PS modification.

With increasing concentrations of Me, the fraction of the RNA with Me bound to the PS modification site will increase and thus accelerate the reaction rate defined as $v = k_{Me} \times [ES^{Me}]$ (Figure 14.2a). Therefore, the affinity and binding cooperativity of the rescuing metal ion, Me, are obtained experimentally by plotting k_{rel} (measured at different Me concentrations) versus $[Me]$ and fitting the data to Eq. (14.3) or (14.4), or variations thereof as required by changes in reaction mechanism.

Thus, application of quantitative thiophilic metal ion rescue of PS modification faces the following experimental challenges: (i) purification of the R_p and S_p phosphorothioate stereoisomers, (ii) introduction of site-specific PS modification into the RNA of interest, (iii) accurate determination of the appropriate reaction

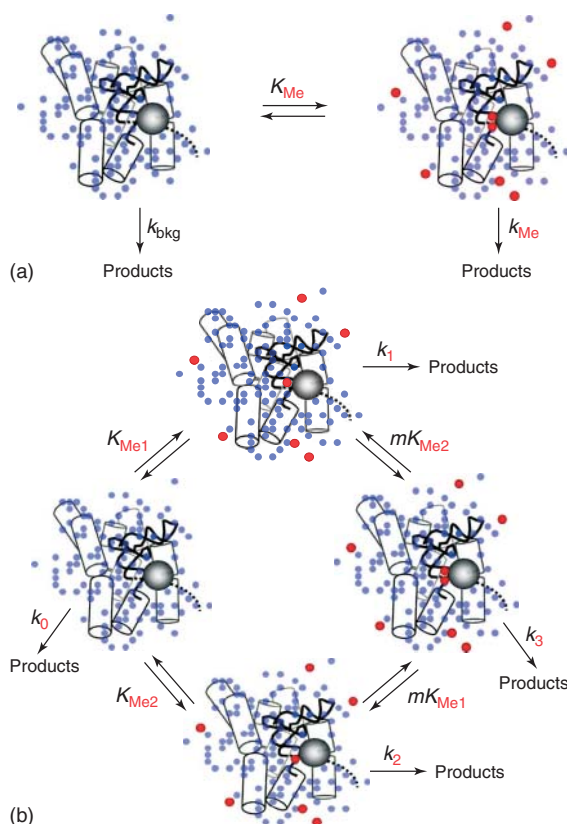


Figure 14.2 Minimal kinetic scheme for the rescue of a deleterious PS modification by site-specific binding of a thiophilic metal ion or ions. A cartoon of the structure of the RNase P holoenzyme is shown. The positions of helices of the RNase P RNA subunit (P RNA) are shown as cylinders. The ptRNA is shown as a dark ribbon as in Figure 14.1c. The RNase P protein subunit (P protein) is symbolized as a gray sphere of appropriate size relative to P RNA. (a) Simple binding model for interaction of a single ion or multiple ions with absolute cooperativity in which there are two states of the enzyme-substrate complex (ES) complex. The unbound state contains all of the functional metal ion interactions (blue spheres) except for the interaction disrupted by PS modification. The complex reacts with rate constant k_{bkg} . Binding of one or more thiophilic metal ions to the site of PS modification (illustrated by the two red spheres close to the

P protein on its left side), described by the equilibrium constant K_{Me} , is a prerequisite to attain the active state that reacts with the rate constant K_{Me} . Other red spheres indicate thiophilic metal ions binding somewhere else in the P RNA, which may affect the folding equilibrium of P RNA and thus its activity. (b) Scheme for binding of two ions independently or with cooperative or anti-cooperative effects on each other. K_{Me1} and K_{Me2} are the individual equilibrium constants for binding of ions Me1 and Me2, respectively. The rate constant for reaction of the ES complex in the absence of the thiophilic metal ion is k_0 , while k_1, k_2 represent the rate constants for the complexes with one and k_3 with two ions bound. In the case of cooperative or anti-cooperative effects, binding of Me1 and Me2 becomes interdependent indicated by the factor m (<1 or >1) that modifies K_{Me1} and K_{Me2} .

rate constants, and (iv) fitting the data to equilibrium binding models to determine apparent dissociation constants and assess cooperativity. The experimental strategies used to accomplish these goals are discussed in the following text in the context of catalytic metal ion binding at the cleavage site of the ptRNA substrate of RNase P.

14.2

Purification of Phosphorothioate Stereoisomers by RP-HPLC

Introduction of the sulfur atom at the non-bridging position results in a chiral center at the phosphorus with two stereoisomers, Rp and Sp (Figure 14.1). For targeting based on crystal structure, one needs to identify which phosphoryl oxygen is the proposed ligand. The two stereoisomers are readily separated by reversed phase (RP)-HPLC, and this allows the two positions to be interrogated individually. RNA oligonucleotides (<80 nt) containing site-specific PS modifications are available commercially from several sources. Shorter oligonucleotides resolve better in the HPLC procedure described below. This approach has been useful for lengths up to 15 nt. PS-containing RNAs from Thermo Scientific Dharmacon (<http://dharmacon.com>) up to the 1 μ mol scale were used directly after 2'-bis(2-acetoxyethoxy)methyl (ACE) deprotection by the manufacturer.

Because sensitive methods of radiochemical detection are used for kinetic studies, only relatively small amounts (hundreds of picomoles) of material are needed. Typically, a 0.2 μ mol scale after purification yields sufficient amounts of material if the PS-containing RNA is to be used as radiolabeled substrate and included in the reaction in trace amounts. Ligation into longer RNAs, see below, is also used depending on the application. Optimization of the ligation procedure can result in almost quantitative conversion of the oligonucleotide into the ligated product in some cases (see Chapter 3 for more details). As indicated earlier, the most convenient application is to set up the reaction and analysis such that the PS-modified RNA is used in trace amounts. This approach is tenable when the PS modification is located at the reactive phosphate within an RNA substrate, in this case, the ptRNA substrate of RNase P.

Protocol 1: RP-HPLC Purification of Phosphorothioate-Containing RNA Oligonucleotides

- 1) Deprotected RNA (0.2 μ mol) after lyophilization is resuspended in 200 μ l distilled and deionized water to create stocks of circa 1 mM (1 nmol/ μ l).
- 2) The Rp and Sp stereoisomers are separated on a 300 $\times$ 3.5 mm Bondclone C-18 column with 5 μ m particle size. A flow rate of 1 ml min⁻¹ is used with a mobile phase of 0.2 M ammonium acetate (pH 8.0) and 4% acetonitrile. Samples are injected in this mobile phase and eluted with a gradient from 4 to 80% acetonitrile in 25 min. This combination of C-18 Guard column, mobile phase, and inline 0.5 μ m micro filter runs with a backpressure of circa 1200 psi on the instrument in our laboratory. UV absorption at 260 nm is monitored to follow the elution of oligonucleotides.

- 3) Analytical injections of 100–200 pmol in 50 μl of water are used to determine the retention times of the stereoisomers. Some additional optimization of the HPLC method is generally necessary. Although a complete discussion of HPLC method optimization is beyond the scope of this chapter, making incremental changes relative to these starting conditions has been satisfactory for oligonucleotides up to 22 nt. Making the gradient shallower, which increases the retention times, generally helps to better resolve the two isomers if initial overlap of the peaks is observed. Most oligonucleotides in the range of 10–15 nt can be separated with “baseline” resolution after some optimization of the HPLC method.
- 4) The *Rp*-stereoisomeric oligonucleotide elutes first, and there is typically a bias toward synthesis of the *Sp* isomer, which represents 60–70% of the material synthesized by Dharmacon. For preparative chromatographic runs, circa 50 μmol of oligonucleotide are injected and fractions of 2–3 ml are collected. Care should be taken to collect the leading/lagging edges of the *Rp* and *Sp* peaks to minimize any potential for contamination between the purified materials. Fractions are placed on dry ice after collection, followed by lyophilization (step 5).
- 5) Use of ammonium acetate in the mobile phase allows the oligonucleotide to be recovered and desalted by lyophilization. As mentioned earlier, fractions are frozen on dry ice immediately as they are taken and dried down in a SpeedVac equipped with a carbon trap. The presence of acetonitrile necessitates frequent change of the pump oil (after every one to two weeks depending on usage). Samples should be redissolved in 1 ml distilled deionized water and subsequently dried to completion; this procedure should be repeated two or more times to remove residual salt and organic solvent. The final residue is then redissolved at concentrations convenient for subsequent labeling or ligation steps (typically 100 pmol μl^{-1}).

14.3

Techniques for Incorporation of Phosphorothioates into RNA

The choice of technique for incorporation of modifications is dependent primarily on the location of the desired modification. In favorable circumstances, it may be possible to make the RNA of interest entirely by solid phase synthesis and thus to use the purified oligonucleotide directly in functional assays. More commonly, it is necessary to incorporate the PS modification into a larger RNA, in the case of the current example into a tRNA with a length of approx. 100 nt. Most convenient is if the oligonucleotide can be joined to the end of the RNA in a single ligation reaction. Such is the case with tRNA where the PS modification at the reactive phosphoryl can be ligated onto an RNA generated by *in vitro* transcription that contains the remainder of the tRNA sequence. Although the subject of “DNA splint”-directed ligation is dealt with in detail in Chapter 3, the following protocol is included to

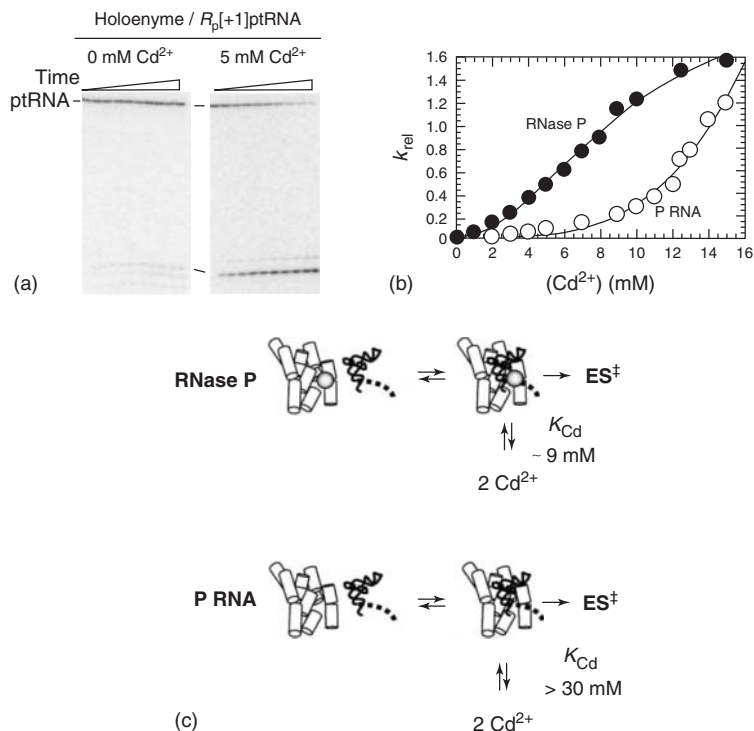


Figure 14.3 Determination of the apparent equilibrium binding constant for the binding of active site metal ions (here Cd^{2+}) to P RNA and the RNase P holoenzyme. (a) Determination of pseudo-first-order rate constants by quantification of reaction progress using radiolabeled ptRNA and denaturing gel electrophoresis. Reaction time courses for the $R_p[+1]$ ptRNA substrate in the absence (left) and presence (right) of 5 mM Cd^{2+} . (b) Analysis of k_{rel} as a function of

Cd^{2+} concentration for reactions catalyzed by P RNA alone (open circles) or by the RNase P holoenzyme (filled circles). The data are fit to the simple binding model (Protocol 5, Eq. (14.2)). (c) Interpretation of the Cd^{2+} rescue data shown in panel b supports a model in which two rescuing metal ions bind to the active site of the ES complex in the ground state. The apparent affinities of the rescuing ion interactions are indicated. $ES^\ddagger$ indicates the reaction transition state.

illustrate a method for efficient modification of ptRNA used in subsequent kinetic studies.

Protocol 2: Generation of ptRNA Containing a Cleavage Site PS Modification by “DNA Splint”-Directed Ligation

The *Escherichia coli* ptRNA^{Met608} carrying a single site-specific PS modification at the RNase P cleavage site ($R_p[+1]$ ptRNA, [46]) was generated by ligation of an 11-nt RNA oligonucleotide (covering positions -5 to $+6$ of ptRNA^{Met608}) to a T7 transcript (covering positions $+7$ to $+77$ of ptRNA^{Met608}) [47]. As shown in Figure 14.3, reaction of this substrate with RNase P gives rise to cleavage primarily at the correct phosphodiester bond 5' to nt $+1$ of the ptRNA substrate.

- 1) Equal amounts (10–50 pmol) of 5'-³²P-endlabeled oligonucleotide and the +7 to +77 RNA transcript plus a DNA oligonucleotide ("DNA splint") complementary to the 11-mer and the first 19 nt of the T7 transcript are combined in a total volume of 6–8 μ l in distilled deionized water.
- 2) Heat the combined RNAs to 60 °C for 2 min and then immediately freeze on dry ice for at least 2 min.
- 3) Thaw sample on wet ice and supplement with
 - 1 μ l 10 x T4 DNA ligase buffer (New England Biolabs),
 - 1 μ l 50% 8000 g mol⁻¹ polyethylene glycol (Fluka),
 - 0.5 μ l (200 U) SUPERase•In™ (Ambion, RNase inhibitor),
 - 0.7 μ l (280 U) T4 DNA ligase (New England Biolabs).
- 4) Incubate for 2 h at 30 °C.
- 5) Combine the reaction with an equal volume (10 μ l) of 95% formamide containing 10 mM EDTA and heat to 90 °C for 2 min. Purify the ligated RNA by 8% (19:1 acrylamide/*N,N'*-methylene bisacrylamide) denaturing (8 M urea) PAGE, visualize by autoradiography, excise the product from the gel, and recover it using standard approaches (see, e.g., Chapters 1 and 3).

14.4

Kinetic Analysis of Thiophilic Metal Ion Rescue

As described earlier (Section 14.2), the scheme requires the rate constants for the all-PO substrate ($k_{\text{obs}}(\text{PO})$) and the PS-containing substrate ($k_{\text{obs}}(\text{PS})$) to be determined in Mg^{2+} alone, and over a range of rescuing thiophilic metal ion concentrations. The 5'-endlabeled ptRNA from the ligation reaction can be used directly in subsequent kinetic experiments without quantification as long as the reaction conditions are pseudo-first-order. That is, the observed rate constant depends only on the concentration of the free enzyme because the substrate is present in trace amounts and the low amounts of formed ES complexes leave the free enzyme concentration essentially unchanged. An important control is to prepare an all-PO substrate by ligation in order to ensure that any differences are solely due to the presence of the PS modification.

Protocol 3: Determination of Pseudo-First-Order Single-Turnover Rate Constants for RNase P Cleavage of Radiolabeled Substrate ptRNA

- 1) The P RNA subunit of *E. coli* RNase P is synthesized by *in vitro* transcription using standard approaches and the P protein subunit is generated as described [48]. Briefly, the protein is overexpressed as a fusion to the maltose-binding protein *via* a peptide spacer including an intein. Intein cleavage generates the native amino terminus of the *E. coli* P protein. A detailed procedure is included in the above reference and stock concentrations of 100–200 μ M are used for the following analyses. The P protein storage buffer, which will contribute proportionally to the final reaction conditions, contains 20 mM Tris–HCl

(pH 8), 500 mM NaCl, 0.1 mM EDTA, 0.01% reduced Triton X-100, 5 mM dithiothreitol (DTT), and 50% glycerol.

- 2) Enzyme and substrate solutions are made separately at twice the final concentration.

The following are the final reaction conditions:

50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) (pH 5.75),
 100 mM NaCl,
 17.5 mM MgCl₂,
 0.005% Triton X-100,
 1 μM RNase P holoenzyme or 10–30 μM P RNA,
 1–4 nM 5′-³²P-endlabeled ptRNA.

The final reaction volumes and stock concentrations of the enzyme and substrate solutions should be adjusted depending on the aims of each experiment.

Preparation of substrate mix:

1–2 μl 0.1 μM 5′-³²P-endlabeled ptRNA,
 3 μl 4 x reaction buffer (200 mM MES, pH 5.75, 400 mM NaCl, 0.02% Triton X-100),
 5–6 μl deionized H₂O.

Preparation of enzyme mix:

1–5 μl P RNA (10–50 μM stock concentration),
 3 μl 4 x reaction buffer,
 2–6 μl deionized H₂O,

Volumes = 10 μl.

- 3) The enzyme and substrate solutions are renatured separately in the absence of divalent metal ions by incubation at 95 °C for 3 min, followed by shifting the temperature to 37 °C and continuing the incubation for a further 10–20 min.
- 4) MgCl₂ and the appropriate concentration of rescuing metal ion are then added to the desired concentration, and the mixture is further incubated for 20 min.
- 5) For holoenzyme reactions, P protein is then added (final concentration equal to P RNA), and the incubation is continued for another 10 min.
- 6) Enzyme and substrate solutions are mixed together to start the reaction, and aliquots are removed at defined intervals (ideally by homogeneously covering all phases of the substrate conversion curve) and quenched with EDTA at circa twice the divalent metal ion concentration in the reaction.
- 7) Products are resolved by denaturing PAGE (15%), and the conversion of substrate to product is quantified by phosphorimaging using a Molecular Dynamics system and the ImageQuant software (GE Healthcare).

14.5

Data Analysis by Fitting to Simple Equilibrium Models

In order to understand the limits to interpretation of such data, it is important to understand the underlying model and the assumptions that are made in order to simplify the quantitative analysis. Importantly, changes in reaction due to changes in structure and reaction conditions (as a consequence of varied metal ion composition and concentration) can complicate the interpretation of the data in terms of ion binding affinity and specificity. First, it is important to accurately determine the first-order rate constants, or pseudo-first-order rate constants if the reaction is bimolecular, as a function of metal ion concentration. Next, these data are used to compute k_{rel} that is subsequently evaluated as a function of rescuing metal ion concentration to determine the apparent metal ion binding affinity and cooperativity by fitting to simple equilibrium binding models. The approaches and equations for these two procedures are outlined in the following two protocols.

Protocol 4: Determination of Pseudo-First-Order Rates $k_{\text{obs(PS)}}$ and $k_{\text{obs(PO)}}$.

- 1) The procedure described earlier permits the fraction of substrate converted to product ($F_c = [P]/[S]_0$) to be determined as a function of time. The reaction is bimolecular and therefore formally it is second-order. However, as is commonly the case, the reaction is conveniently performed under pseudo-first-order conditions in which the enzyme is in excess and its free concentration is not significantly changed by binding the substrate, and does not change over the time course of the reaction. In this case, the data can be fit to a single exponential function to determine the rate constant.
- 2) F_c versus time data are plotted and fit to kinetic and equilibrium binding models. A suitable program that comes with a large range of useful functions and also permits input of user functions is *Origin*.
- 3) The most appropriate form of the function for fitting this kind of data is

$$F_c = A + Be^{-kt} \quad (14.1)$$

for substrate cleavage (decay) or $F_c = A + B(1 - e^{-kt})$ if product accumulation is plotted, where F_c is the fraction cleaved ($F_c = [P]/[S]_0$), A is the background (offset) at $t = 0$, B is the amplitude of the exponential, k is the observed pseudo-first-order rate constant, and t is time. Note that the function does not assume a fraction of reaction of zero at $t = 0$ nor a value of 1 when $t = \text{infinite}$. The variable A is included to account for non-zero F_c at time zero, which can occur when there is background in the signal used to derive F_c . The variable B is the amplitude of the exponential and will allow the function to fit reactions in which a portion of the substrate is unreactive. In this case, the rate constant will represent only the active substrate fraction. Evaluation of the magnitudes of A and B permit quantification of the level of background and any problems associated with incomplete reactions.

- 4) Individual kinetic experiments should be fit individually and constitute a single experimental trial for subsequent calculation of experimental error. Rate

constants should be determined from reactions with at least 70–80% substrate conversion, including one or more extended time points to better define the end point of the reaction and thus the overall fraction of substrate molecules that were reactive. Individual fitting errors are typically less than 30% with an adjusted R^2 of greater than 0.8.

- 5) Rate constants used to calculate k_{rel} should be the average of at least three independent determinations and at least two different preparations of the ligated substrate to control for any artifacts arising from the preparation of the modified RNA.

Protocol 5: Fitting $k_{\text{rel}} = k_{\text{obs}}(\text{PS})/k_{\text{obs}}(\text{PO})$ to Equilibrium Binding Models to Determine K_{Me} and Assess Cooperativity

- 1) The relative rate constant, k_{rel} , is calculated from the ratio of the observed pseudo-first-order rate constants for the substrate with and without the sulfur substitution, that is, $k_{\text{rel}} = k_{\text{obs}}(\text{PS})/k_{\text{obs}}(\text{PO})$. This operation controls for the effect of the rescuing ion concentration on the background activity of the unmodified RNA. For example, the rescuing ion species that binds to the site of PS modification may simultaneously bind to other sites where it inhibits activity, for example, by displacing functional Mg^{2+} ions.
- 2) The determined k_{rel} values are plotted as a function of increasing $[\text{Me}]$, in this case Cd^{2+} . In order to quantify binding, it is necessary to analyze a range of rescuing metal ion concentrations that are above and below the apparent binding equilibrium constant. Thus, there is some aspect of trial and error in determining the appropriate range of concentrations. Large errors in the fitting parameters described below indicate that additional data points are necessary. A convenient benchmark is to have concentrations that are at least 10-fold above and below the concentration of rescuing metal ion required to reach half of the maximum k_{rel} . Operationally, if there is a less than 2-fold change in the observed k_{rel} upon a 10-fold increase in metal ion concentration, the reaction can be considered to approximate saturation.
- 3) Plots of k_{rel} versus $[\text{Cd}^{2+}]$ are first fit to a simple equilibrium binding model:

$$k_{\text{rel}}(\text{observed}) = k_{\text{bkg}} \left(\frac{K_{\text{Me}}}{K_{\text{Me}} + [\text{Me}]} \right) + k_{\text{Me}} \left(\frac{[\text{Me}]}{[\text{Me}] + K_{\text{Me}}} \right) \quad (14.2)$$

where k_{rel} is the relative rate constant described earlier measured at each $[\text{Me}]$ concentration, k_{Me} is the cleavage rate constant at saturating $[\text{Me}]$, and K_{Me} is the apparent dissociation constant for binding of the rescuing metal ion(s). If the background rate constant is sufficiently low relative to k_{Me} (at least 100-fold lower), the first term can be omitted.

- 4) Theoretical models to distinguish between independent and cooperative and anti-cooperative interactions of metal ions in the observed binding data use a more complex reaction scheme (Figure 14.2b). The simplest model involves the binding of two or more rescuing metal ions with absolute cooperativity. In

this all-or-none binding scheme, the Hill equation is used.

$$k_{\text{rel}}(\text{observed}) = k_{\text{Me}} \left(\frac{[\text{Me}]^n}{(K_{\text{Me}})^n + [\text{Me}]^n} \right) \quad (14.3)$$

Here, the exponent is the Hill number that can be assigned a variable, n , and the data fit to examine the extent to which the data conform to this simple binding model. As it is well understood that in the Hill binding model fractional Hill coefficients have no physical meaning (the enzyme cannot bind one and a half metal ions, for example), the best approach is to fit the data with the Hill coefficient fixed at an increasing integer (1, 2, 3 ...) and analyze the residual parameters in order to assess the goodness of the fit.

- 5) However, metal binding is complex and cooperativity is not always absolute. In Figure 14.2b, the two binding sites can be partially occupied and binding of the individual ions can result in activity. Binding of two metal ions can be independent, cooperative, or anti-cooperative. The following model describes the predicted dependence of the observed cleavage rate constant on [Me]:

$$k_{\text{obs}} = \frac{k_0 + k_1 \frac{[\text{Me}]}{K_{\text{Me1}}} + k_2 \frac{[\text{Me}]}{K_{\text{Me2}}} + k_3 \frac{[\text{Me}]^2}{m K_{\text{Me1}} K_{\text{Me2}}}}{1 + \frac{[\text{Me}]}{K_{\text{Me1}}} + \frac{[\text{Me}]}{K_{\text{Me2}}} + \frac{[\text{Me}]^2}{m K_{\text{Me1}} K_{\text{Me2}}}} \quad (14.4)$$

Independent binding is indicated by m values equal to 1, while cooperative and anti-cooperative binding are indicated by m values <1 or >1 , respectively. Simulations are carried out over a range of m values (0.01–100), as well as with a range of values for K_{Me1} and K_{Me2} (0.001–100, respectively). Inspection and quantification of the residuals is used to evaluate the appropriate range of constants.

References

1. Draper, D.E. (2008) RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys. J.*, **95**, 5489–5495.
2. Sigel, R.K. and Pyle, A.M. (2007) Alternative roles for metal ions in enzyme catalysis and the implications for ribozyme chemistry. *Chem. Rev.*, **107**, 97–113.
3. Lonnberg, T. (2011) Understanding catalysis of phosphate-transfer reactions by the large ribozymes. *Chemistry*, **17**, 7140–7153.
4. Wedekind, J.E. (2011) Metal ion binding and function in natural and artificial small RNA enzymes from a structural perspective. *Met. Ions Life Sci.*, **9**, 299–345.
5. Draper, D.E. (2004) A guide to ions and RNA structure. *RNA*, **10**, 335–343.
6. Chu, V.B., Bai, Y., Lipfert, J., Herschlag, D., and Doniach, S. (2008) A repulsive field: advances in the electrostatics of the ion atmosphere. *Curr. Opin. Chem. Biol.*, **12**, 619–625.
7. Bai, Y., Greenfield, M., Travers, K.J., Chu, V.B., Lipfert, J., Doniach, S., and Herschlag, D. (2007) Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.*, **129**, 14981–14988.

8. Tan, Z.J. and Chen, S.J. (2011) Importance of diffuse metal ion binding to RNA. *Met. Ions Life Sci.*, **9**, 101–124.
9. Misra, V.K. and Draper, D.E. (1998) On the role of magnesium ions in RNA stability. *Biopolymers*, **48**, 113–135.
10. Holbrook, S.R., Sussman, J.L., Warrant, R.W., Church, G.M., and Kim, S.H. (1977) RNA-ligand interactions. (I) Magnesium binding sites in yeast tRNAPhe. *Nucleic Acids Res.*, **4**, 2811–2820.
11. Stahley, M.R., Adams, P.L., Wang, J., and Strobel, S.A. (2007) Structural metals in the group I intron: a ribozyme with a multiple metal ion core. *J. Mol. Biol.*, **372**, 89–102.
12. Golden, B.L. (2011) Two distinct catalytic strategies in the hepatitis delta virus ribozyme cleavage reaction. *Biochemistry*, **50**, 9424–9433.
13. Fedor, M.J. and Williamson, J.R. (2005) The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol.*, **6**, 399–412.
14. Stahley, M.R. and Strobel, S.A. (2006) RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr. Opin. Struct. Biol.*, **16**, 319–326.
15. Forconi, M., Lee, J., Lee, J.K., Piccirilli, J.A., and Herschlag, D. (2008) Functional identification of ligands for a catalytic metal ion in group I introns. *Biochemistry*, **47**, 6883–6894.
16. Quigley, G.J., Teeter, M.M., and Rich, A. (1978) Structural analysis of spermine and magnesium ion binding to yeast phenylalanine transfer RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 64–68.
17. Hanna, R. and Doudna, J.A. (2000) Metal ions in ribozyme folding and catalysis. *Curr. Opin. Chem. Biol.*, **4**, 166–170.
18. DeRose, V.J. (2003) Metal ion binding to catalytic RNA molecules. *Curr. Opin. Struct. Biol.*, **13**, 317–324.
19. Stefan, L.R., Zhang, R., Levitan, A.G., Hendrix, D.K., Brenner, S.E., and Holbrook, S.R. (2006) MeRNA: a database of metal ion binding sites in RNA structures. *Nucleic Acids Res.*, **34** (Database issue), D131–D134.
20. Kazantsev, A.V., Krivenko, A.A., and Pace, N.R. (2009) Mapping metal-binding sites in the catalytic domain of bacterial RNase P RNA. *RNA*, **15**, 266–276.
21. Philips, A., Milanowska, K., Lach, G., Boniecki, M., Rother, K., and Bujnicki, J.M. (2012) Metalion RNA: computational predictor of metal-binding sites in RNA structures. *Bioinformatics*, **28**, 198–205.
22. Misra, V.K. and Draper, D.E. (2000) Mg(2+) binding to tRNA revisited: the nonlinear Poisson-Boltzmann model. *J. Mol. Biol.*, **299**, 813–825.
23. Piccirilli, J.A., Vyle, J.S., Caruthers, M.H., and Cech, T.R. (1993) Metal ion catalysis in the *Tetrahymena* ribozyme reaction. *Nature*, **361**, 85–88.
24. Shan, S., Kravchuk, A.V., Piccirilli, J.A., and Herschlag, D. (2001) Defining the catalytic metal ion interactions in the *Tetrahymena* ribozyme reaction. *Biochemistry*, **40**, 5161–5171.
25. Hougland, J.L., Kravchuk, A.V., Herschlag, D., and Piccirilli, J.A. (2005) Functional identification of catalytic metal ion binding sites within RNA. *PLoS Biol.*, **3**, e277.
26. Frederiksen, J.K. and Piccirilli, J.A. (2009) Identification of catalytic metal ion ligands in ribozymes. *Methods*, **49**, 148–166.
27. Pecoraro, V.L., Hermes, J.D., and Cleland, W.W. (1984) Stability constants of Mg²⁺ and Cd²⁺ complexes of adenine nucleotides and thionucleotides and rate constants for formation and dissociation of MgATP and MgADP. *Biochemistry*, **23**, 5262–5271.
28. Pearson, R.G. (1988) Absolute Electronegativity and Hardness: Application to Inorganic Chemistry. *Inorg. Chem.*, **27**, 734–740.
29. Jaffe, E.K., Nick, J., and Cohn, M. (1982) Reactivity and metal-dependent stereospecificity of the phosphorothioate analogs of ADP and ATP and reactivity of Cr(III)ATP in the 3-phosphoglycerate kinase reaction. Structure of the metal nucleotide substrates. *J. Biol. Chem.*, **257**, 7650–7656.
30. Eckstein, F., Romaniuk, P.J., and Connolly, B.A. (1982) Stereochemistry of enzymic phosphoryl and nucleotidyl transfer. *Methods Enzymol.*, **87**, 197–212.

31. Shan, S., Yoshida, A., Sun, S., Piccirilli, J.A., and Herschlag, D. (1999) Three metal ions at the active site of the *Tetrahymena* group I ribozyme. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 12299–12304.
32. Peracchi, A., Beigelman, L., Scott, E.C., Uhlenbeck, O.C., and Herschlag, D. (1997) Involvement of a specific metal ion in the transition of the hammerhead ribozyme to its catalytic conformation. *J. Biol. Chem.*, **272**, 26822–26826.
33. Scott, W.G., Finch, J.T., and Klug, A. (1995) The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell*, **81**, 991–1002.
34. Martick, M. and Scott, W.G. (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, **126**, 309–320.
35. Shepotinovskaya, I.V. and Uhlenbeck, O.C. (2008) Catalytic diversity of extended hammerhead ribozymes. *Biochemistry*, **47**, 7034–7042.
36. Hougland, J., Piccirilli, J.A., Forconi, M., Lee, J., and Herschlag, D. (2006) How the group I intron works: a case study of RNA structure and function, in *The RNA World*, 3rd edn (eds R. Akins, R. Gesteland, and T. Cech), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
37. Hartmann, R.K., Gössringer, M., Späth, B., Fischer, S., and Marchfelder, A. (2009) The making of tRNAs and more-RNase P and tRNase Z. *Prog. Mol. Biol. Transl. Sci.*, **85**, 319–368.
38. Hsieh, J., Andrews, A.J., and Fierke, C.A. (2004) Roles of protein subunits in RNA-protein complexes: lessons from ribonuclease P. *Biopolymers*, **73**, 79–89.
39. Smith, D. and Pace, N.R. (1993) Multiple magnesium ions in the ribonuclease P reaction mechanism. *Biochemistry*, **32**, 5273–5281.
40. Guerrier-Takada, C., Haydock, K., Allen, L., and Altman, S. (1986) Metal ion requirements and other aspects of the reaction catalyzed by M1 RNA, the RNA subunit of ribonuclease P from *Escherichia coli*. *Biochemistry*, **25**, 1509–1515.
41. Warnecke, J.M., Held, R., Busch, S., and Hartmann, R.K. (1999) Role of metal ions in the hydrolysis reaction catalyzed by RNase P RNA from *Bacillus subtilis*. *J. Mol. Biol.*, **290**, 433–445.
42. Warnecke, J.M., Fürste, J.P., Hardt, W.D., Erdmann, V.A., and Hartmann, R.K. (1996) Ribonuclease P (RNase P) RNA is converted to a Cd(2+)-ribozyme by a single Rp-phosphorothioate modification in the precursor tRNA at the RNase P cleavage site. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 8924–8928.
43. Cuzic-Feltens, S., Weber, M.H., and Hartmann, R.K. (2009) Investigation of catalysis by bacterial RNase P via LNA and other modifications at the scissile phosphodiester. *Nucleic Acids Res.*, **37**, 7638–7653.
44. Cuzic, S., Heidemann, K.A., Wöhnert, J., and Hartmann, R.K. (2008) *Escherichia coli* RNase P RNA: substrate ribose modifications at G+1, but not nucleotide -1/+73 base pairing, affect the transition state for cleavage chemistry. *J. Mol. Biol.*, **379**, 1–8.
45. Cuzic, S. and Hartmann, R.K. (2007) A 2'-methyl or 2'-methylene group at G+1 in precursor tRNA interferes with Mg²⁺ binding at the enzyme-substrate interface in E-S complexes of *E. coli* RNase P. *Biol. Chem.*, **388**, 717–726.
46. Sun, L. and Harris, M.E. (2007) Evidence that binding of C5 protein to P RNA enhances ribozyme catalysis by influencing active site metal ion affinity. *RNA*, **13**, 1505–1515.
47. Zahler, N.H., Sun, L., Christian, E.L., and Harris, M.E. (2005) The pre-tRNA nucleotide base and 2'-hydroxyl at N(-1) contribute to fidelity in tRNA processing by RNase P. *J. Mol. Biol.*, **345**, 969–985. [Epub 2004 Dec 8].
48. Guo, X., Campbell, F.E., Sun, L., Christian, E.L., Anderson, V.E., and Harris, M.E. (2006) RNA-dependent folding and stabilization of C5 protein during assembly of the *E. coli* RNase P holoenzyme. *J. Mol. Biol.*, **360**, 190–203.

15

Probing RNA Structure and Ligand Binding Sites on RNA by Fenton Cleavage

Corina G. Heidrich and Christian Berens

15.1

Introduction

Once viewed as a mere messenger or scaffolding molecule involved in information transfer from DNA to protein, RNA is now recognized to be an extremely versatile molecule which is involved in many different processes in the cell [1]. The RNA molecules playing these different roles can be roughly divided into two groups. The first group, exemplified by mRNA, snoRNA, guide RNA, sRNA, miRNA, and siRNA, utilizes sequence-specific Watson–Crick base-pairing interactions to exert their activity. In contrast, the second group, comprising ribozymes, aptamers, and riboswitches, must fold into a specific three-dimensional structure to be functional. For these RNA molecules, knowledge of their structures and how they are formed is an absolute prerequisite to understand how they work mechanistically.

Because X-ray crystallography of flexible RNA molecules is still difficult, despite the many new structures of RNAs and RNA–protein complexes which have been solved and published [2–5], and because many interesting RNA molecules are still too large for NMR analysis, a large set of genetic, biochemical, and spectroscopic methods have been developed for the structural analysis of RNA [6–9]. One very versatile method to analyze RNA structures in solution is probing with hydroxyl radicals [10]. They are the smallest molecule species used for chemical probing and cleave nucleic acids with little or no sequence specificity [11, 12]. A significant secondary structure preference has not been observed in radical-induced cleavage of single- and double-stranded forms of both RNA and DNA [13].

Hydroxyl radicals are generated physically by radiolysis of water using synchrotron X-ray beams [14] or, most frequently and much more easily, chemically, by the reduction of peroxo-groups with iron(II) in the so-called Fenton reaction [15]. Like most transition metals, iron has more than one oxidation state besides its ground state and the valence electrons may be unpaired allowing one-electron redox reactions [16]. As such, iron(II) reacts with hydrogen peroxide (or other peroxo-molecules like peroxonitrous acid) to generate short-lived, highly reactive,

diffusible hydroxyl radicals. These cleave the bases of a nucleic acid, its phosphodiester backbone and also peptide bonds in spatial proximity of iron(II). The reach of these hydroxyl radicals, when generated by the FeBABE reagent (BABE: 1-(*p*-bromoacetamidobenzyl)-EDTA) and measured using protein substrates, was between 8.6 and 58 Å, with an average distance of 27.2 Å. The size and conformational freedom of the FeBABE moiety contribute about 6 Å to these distances. This demonstrates that, despite the hydroxyl radical's flexible reach, it can provide reliable local information [17]. Sodium ascorbate is often added to the reaction mixture in order to reduce the iron(III) generated in the reaction back to iron(II), thereby establishing a catalytic cycle and permitting low (micromolar) concentrations of iron(II) to be effective in substrate cleavage. Consequently, hydroxyl radical-based probing methods have been widely used for the structural analysis of RNA and also as a tool to study interactions of nucleic acids with proteins or other ligands. Table 15.1 summarizes the different methods used in iron(II)-mediated hydroxyl radical footprinting and the type of information that can be gained from these studies.

The classical interaction study is a footprinting experiment. The presence of the interaction partner protects the nucleic acid at the binding site from cleavage by the hydroxyl radicals. This approach was used to identify contact sites of 16S rRNA in 30S subunits with 50S subunits [47] or to determine the structural elements of an internal ribosomal entry site that interact with a 40S ribosomal subunit [48]. A detailed protocol for hydroxyl radical footprinting of protein–RNA complexes is given in [49]. Hydroxyl radical footprinting has also been and still is widely applied to study protein–DNA interactions as described in detail elsewhere [50].

The general applicability of this approach was greatly extended by tethering iron(II) to defined sites on proteins and RNA using the BABE reagent, which was originally synthesized by Meares and coworkers [34, 35]. Hydroxyl radical footprints with iron(II) tethered either to various ribosomal proteins or to rRNA gave important insights into the three-dimensional organization of the ribosome [36, 37], which were later confirmed by the crystal structure of the 70S ribosome (reviewed in [51]). Detailed protocols for interaction studies with hydroxyl radicals generated by iron(II) either tethered to proteins or RNA have been published [38, 39].

Hydroxyl radical cleavage is also used for RNA structure analysis. For example, Newcomb and Noller [40] tethered iron(II) to specific nucleotides in 23S rRNA to induce intramolecular self-cleavage and thus to probe their environment within 70S ribosomes. Likewise, Huq and coworkers [41] applied this approach to obtain structural information on the three-dimensional fold of the *trans*-activation-responsive region of HIV-1 RNA.

Besides tethered iron(II)/EDTA, free iron(II)/EDTA is a popular probe used to help determine the three-dimensional structure of an RNA by identifying solvent-accessible and solvent-excluded sugar moieties [25]. Protection of tRNA bound to the ribosomal P-site from hydroxyl radical cleavage gave important hints on the mechanism of tRNA–ribosome interaction [33]. In 2001, Pyle and coworkers presented hydroxyl radical footprints [26] demonstrating that group II introns also

Table 15.1 Applications using Fe(II)-mediated Fenton cleavage for macromolecule characterization.

Reagent used	Mode	Macromolecule targeted	Information mapped	References
Fe(II)	Free	DNA	Proximity to metal ion binding site	[18, 19]
		RNA	Macromolecule interior; proximity to metal ion binding site	[20]
			Proximity to metal ion binding site	[21]
		Protein	Proximity to metal ion binding site	[22–24]
Fe(II)–EDTA ^a	Free	DNA, Protein	Footprint	[11, 14]
		DNA	Macromolecule structure; solvent-exposed residues	[10, 12, 13]
		RNA	Macromolecule exterior; solvent-exposed residues	[25–29]
			Folding, stability	[28, 30–32]
			Footprint	[33]
Fe(II)–BABE ^b	Tethered to protein	Protein	Proximity to residue with the tether	[34, 35]
		RNA	Proximity to residue with the tether	[36–38]
	Tethered to RNA		Proximity to residue with the tether	[39, 40]
Fe(II)–CITC ^c	Tethered to RNA	RNA	Proximity to base pair with the tether	[41]
Fe(II)–tetracycline	Bound ligand	RNA	Proximity to ligand binding site	[42, 43]
		Protein	Proximity to ligand binding site	[44, 45]
Fe(II)–chromomycin A ₃	Bound ligand	DNA	Proximity to ligand binding site	[46]

^a Ethylenediaminetetraacetic acid.^b (S)-N-4-[2,3-Bis[bis(carboxymethyl)amino]propyl]phenyl bromoacetamide.^c (S)-4-[2,3-Bis[bis(carboxymethyl)amino]propyl]phenyl isothiocyanate.

have a tightly packed, solvent-inaccessible core like other large ribozymes [27, 28]. The cleavage patterns produced by hydroxyl radical footprinting can additionally be used to determine the relative stabilities of individual structural motifs by examining changes in the protection pattern as a function of added Mg^{2+} or urea [30, 31]. The high reactivity of the hydroxyl radical allows this method to be applied to address biological questions that require time-resolved analysis, like ligand-binding events or RNA folding. A fast Fenton footprinting method based on commonly used three-syringe quench-flow mixers was established and validated in a folding analysis of the *Tetrahymena* ribozyme [32]. Alternatively, hydroxyl radicals can be generated in a synchrotron to allow time-resolved monitoring of RNA folding [14] (a detailed protocol of the method is presented in [52]) or folding and assembly of ribonucleoprotein complexes, such as the 30S ribosomal subunit [53]. Physical generation of hydroxyl radicals is very powerful, because it is also applicable to study RNA structure and RNA–protein contacts *in vivo*. Irradiation of living *Escherichia coli* cells and subsequent RNA isolation and primer extension analysis gave results for the 16S rRNA that were consistent with data from Fe–EDTA treatment of 70S ribosomes in solution [54]. When combined with molecular dynamics simulations, hydroxyl radical footprinting data can be used to develop robust structural hypotheses for complex RNAs that have so far resisted their study by high-resolution methods [29].

Iron(II) is similar in size and coordination geometry [55] to magnesium(II) and has been used to replace the latter for mapping metal ion binding sites in proteins [22–24] (see also Chapter 19), in RGGG- or RGTR-containing DNA sequences [18, 19], and in eukaryotic rRNA from neurons vulnerable to Alzheimer disease [56]. Catalytic RNAs require divalent cations for achieving a stable tertiary structure as well as for catalysis [57–59]. The identification of metal ion binding sites is therefore essential for a thorough structure–function analysis of catalytic RNAs [60, 61]. Besides NMR studies (summarized in [62]), hydroxyl radical-induced cleavage based on limited replacement of magnesium(II) with iron(II) provides a powerful method for identifying RNA–metal ion interactions [20, 21].

Here, we present two different methods for hydroxyl radical probing of RNA. The first describes structural probing of large RNAs. Comparison of cleavage patterns obtained with iron(II) in the presence and in the absence of EDTA allows one to determine solvent-exposed and solvent-occluded regions of the RNA and to identify potential metal ion binding sites.

The second method describes an interaction study based on the ability of iron(II) to replace the magnesium(II) ion chelated to tetracycline. A subsequent hydroxyl radical digestion can then identify specific RNA residues in proximity of the (Fe^{2+} –tetracycline) chelate. This has already been done successfully with the tetracycline–proton-antiporter TetA [44], the tetracycline-binding regulatory protein TetR [45], and an *in vitro*-selected aptamer for tetracycline [42]. We used hydroxyl radical cleavage of 16S rRNA to identify tetracycline binding sites in the *E. coli* 70S ribosome [43].

Description of Methods

Fe²⁺-Mediated Cleavage of Native Group I Intron RNA

Reagents and Enzymes

- 5 x Group I intron cleavage buffer: 125 mM MOPS (3-morpholinopropane-1-sulfonic acid)-KOH, pH 7.0; 15 mM MgCl₂; 2 mM spermidine; 1 M NaCl
- FeCl₂ (Cat. No. 220299, Sigma-Aldrich)
- Sodium ascorbate (Cat. No. A4034, Sigma-Aldrich)
- Hydrogen peroxide (H₂O₂) (Cat. No. H1009, Sigma-Aldrich)
- Thiourea (Cat. No. T7875, Sigma-Aldrich)
- Glycogen (10 µg/µl; Cat. No. 37-1820, PEQLAB, Erlangen, Germany)
- Denaturing gel loading buffer: 7 M urea; 0.25% bromophenol blue; 0.25% xylene cyanol; in 1 x TBE, pH 8.3 (89 mM Tris base; 89 mM boric acid; 1 mM EDTA); optionally supplemented with 25% sucrose
- 5 x Ribosome cleavage buffer: 125 mM MOPS-KOH, pH 7.0; 15 mM MgCl₂; 0.5 mM spermidine
- Tetracycline hydrochloride (Cat. No. T3383, Sigma-Aldrich)
- Ribosome extraction buffer (REB): 0.3 M sodium acetate; 0.5% SDS; 5 mM EDTA; supplemented with 8 U RNase inhibitor just before use
- Protector RNase Inhibitor (40 U/µl; Cat. No. 03335402001, Roche Applied Science)
- Phenol (Cat. No. A980.1, ROTH, Karlsruhe, Germany), chloroform (Cat. No. 3313.1, ROTH), isoamylalcohol (Cat. No. T870.1, ROTH)
- One-step RT-PCR Kit (Cat. No. 210212, Qiagen)
- SuperScriptIII reverse transcriptase (200 U/µl; Cat. No. 1808080-044, Life Technologies Invitrogen)
- 5 x First-strand buffer: 250 mM Tris-HCl, pH 8.3 at room temperature; 375 mM KCl; 15 mM MgCl₂ (Life Technologies Invitrogen)
- Diethylpyrocarbonate, DEPC (Cat. No. D5758, Sigma-Aldrich)
- Millipore-filtered, DEPC-treated water

The method presented here was used to detect Mg²⁺-binding sites in the *Tetrahymena* LSU group I intron [20]. For a successful reaction, it is important to prepare the solutions of iron(II) chloride and hydrogen peroxide freshly. Sodium ascorbate can be prepared as a 10-fold stock solution and stored at −20 °C.

During the hydroxyl radical cleavage reaction, we recommend the use of a small desktop microcentrifuge, in order to mix each new reagent thoroughly with the components at the bottom of the Eppendorf tube. The appropriate reagent is deposited on the wall area of each Eppendorf tube, and the reagents in all tubes are then mixed simultaneously by brief centrifugation in the microcentrifuge. This way, six reactions can be processed in parallel within the 1 min time frame separating individual steps of the probing procedure. Then, during the subsequent 1 min incubation time, the reagent required next is deposited on the wall of the Eppendorf tube and the centrifugation step repeated.

Protocol 1: Hydroxyl Radical Cleavage

- 1) For experiments with the group I intron RNA, take 1 μ l RNA (5 pmol cold RNA, spiked with approximately 50 000 cpm of the RNA labeled with [^{32} P] at either its 5'- or 3'-end) and add 1 μ l of 5 x group I intron cleavage buffer. Renature the RNA for 2 min at 56 °C, followed by 3 min incubation at room temperature.
- 2) Add 1 μ l 1.25 mM FeCl₂ to the reaction tube, mix by centrifugation and incubate for 1 min before adding 1 μ l 12.5 mM sodium ascorbate.
- 3) After 1 min, add 1 μ l of 12.5 mM H₂O₂ and mix rapidly to initiate the reaction. The final concentrations are 250 μ M for Fe²⁺ and 2.5 mM for both sodium ascorbate and hydrogen peroxide.
- 4) Stop the cleavage reaction after 1 min by adding 1 μ l 1 M thiourea. The RNA is then precipitated with 1 μ l glycogen (10 μ g/ μ l) and 30 μ l 96% (v/v) ethanol.

Protocol 2: Mg²⁺-Competition of Fe²⁺-Mediated Cleavage

- 1) Proceed as described in step 1 of Protocol 1.
- 2) Mix equal volumes of a 2.5 mM FeCl₂ solution and an MgCl₂ solution (prepared from a 10 x stock solution), such that the desired, final Mg²⁺ concentration is reached in the step corresponding to step 3 of Protocol 1 (note that you have to take into account the magnesium already present in the reaction). Add 1 μ l of this Fe²⁺/Mg²⁺-mixture to the reaction tube containing the RNA, mix by centrifugation, and incubate for 1 min before adding 1 μ l 12.5 mM sodium ascorbate.
- 3) Proceed exactly as described in steps 3 and 4 of Protocol 1.

Protocol 3: Gel Electrophoresis

- 1) After precipitation, redissolve the RNA in denaturing gel loading buffer.
- 2) Separate the cleavage products on 6–20% denaturing polyacrylamide sequencing gels.

For obtaining sequencing ladders, carry out limited hydrolysis with RNase T1 and NaHCO₃ [63] (Chapters 9 and 10).

Visualization of Metal Ion Binding Sites in Group I Introns by Iron(II)-Mediated Fenton Reactions

Cleavage by Fe²⁺ is observed in distinct regions of the group I intron RNA and only with natively folded RNA (compare lanes 7 and 8 in Figure 15.1). It is competed by Mg²⁺ (compare lanes 8 and 10 in Figure 15.1) indicating that both ions interact with the same or overlapping binding sites. Lanes 8 and 12 allow comparison of the cleavage sites obtained using Fe²⁺ or Fe²⁺/EDTA. Most of the sites cleaved by Fe²⁺ are embedded in regions protected from cleavage by Fe²⁺/EDTA. They are, thus, located in the interior of the ribozyme where they would be expected to be if the metal ions they reflect bury phosphate oxygens in the interior of the RNA [64]. In a three-dimensional model of the bacteriophage T4-derived *td* intron and

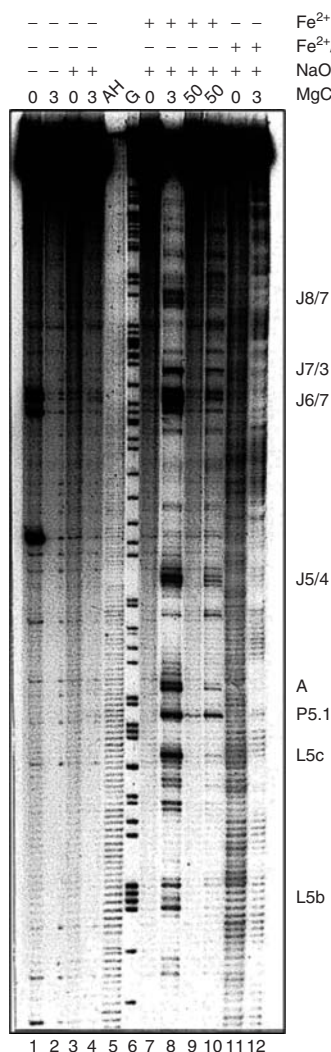


Figure 15.1 Mapping the Fe²⁺-cleavage sites in the *Tetrahymena* LSU group I intron. Autoradiogram of a 6% denaturing polyacrylamide (PAA) gel with 5'-end-labeled *Tetrahymena* L-21 RNA cleaved with 10 μ M Fe²⁺ (lanes 7, 9), 250 μ M Fe²⁺ (lane 8, 10), or with 250 μ M Fe²⁺/500 μ M EDTA (lanes 11, 12). Controls with untreated RNA (lanes 1, 2) and in which Fe²⁺ was omitted (lanes 3, 4), as well as competition of Fe²⁺-cleavage by 50 mM Mg²⁺ (lanes 9, 10) are also shown. The respective final concentrations of Mg²⁺, as well as

the presence (+) or absence (–) of Fe²⁺, sodium ascorbate, and hydrogen peroxide, are displayed above each lane. Renatured RNA was used in lanes 2, 4, 8, 10, and 12, denatured RNA in lanes 1, 3, 7, 9, and 11. Secondary structure elements cleaved by Fe²⁺ are marked on the right and the A-rich bulge in the *Tetrahymena* P4P6 domain as presented in Cate *et al.* [64] is indicated by a capital A. Sequencing markers are AH (alkaline hydrolysis) and G (RNase T1). Reprinted with permission from Ref. [20].

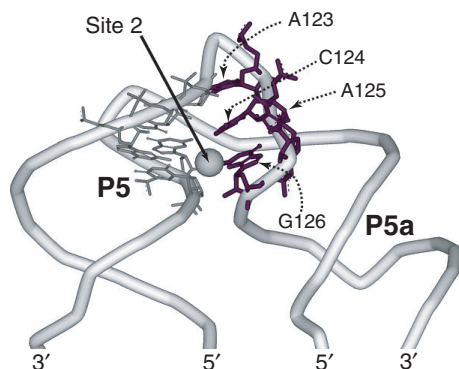


Figure 15.2 Correlation between a computationally predicted Mg^{2+} -binding site and Fe^{2+} -cleavage sites in the *Tetrahymena* P4P6 domain. The phosphodiester backbones of the two RNA strands in the hinge region of the P4P6 domain of the *Tetrahymena* LSU group I intron are shown as shaded ribbons with their polarity and the secondary

structure elements indicated. Site 2 (gray sphere) is a potential binding site for a diffusely bound Mg^{2+} ion [70] and candidate residues for its coordination are displayed as thin gray sticks. Residues that are cleaved by Fe^{2+} are shown as thick black sticks. Coordinates were taken from the RCSB entry 1HR2 [73] and Table 3 of Ref. [70].

in the crystal structure of the P4P6 domain of the *Tetrahymena* LSU intron [65], cleavage sites separated in secondary structure come together in three-dimensional space to form several distinct pockets (Figures 8 and 9 in [20]). There is also very good agreement between nucleotides cleaved by Fe^{2+} and nucleotides close to metal ions determined by phosphorothioate substitution [66–68], metal hydroxyl cleavage [69], or X-ray crystallography [65]. Figure 15.2 shows that cleavage sites in the hinge region of the P4P6 domain coincide nicely with a diffusely bound metal ion predicted from microenvironment analysis [70], but which was not observed in the crystal structure [65]. As another example of the potency of this method, the cleavage reaction of the hairpin ribozyme is facilitated by the presence of divalent metal ions. Earnshaw and Gait used Fenton chemistry to elucidate the locations of ion binding sites in this catalytic RNA molecule. They identified strong cleavage sites in loop B between nucleotides $\text{U}_{37}/\text{A}_{38}$, $\text{U}_{39}/\text{A}_{40}$, and $\text{U}_{42}/\text{A}_{43}$ [21]. Cleavage at $\text{U}_{37}/\text{A}_{38}$ and $\text{U}_{39}/\text{A}_{40}$ agrees nicely with two structure-stabilizing Ca^{2+} ions found in the major groove of stem B in the crystal structure of a hairpin ribozyme-inhibitor complex [71], and also with a terbium(III) cleavage site [72].

Fe^{2+} -Mediated Tetracycline-Directed Hydroxyl Radical Cleavage Reactions

This method describes the identification of tetracycline binding sites on rRNA in 70S ribosomes of *E. coli* [43]. Hydroxyl radical cleavage (Protocols 4 and 5) of the RNA in the vicinity of bound tetracyclines is detected by primer extension (Protocol 7). Fe^{2+} -mediated hydroxyl radical cleavage of the 70S ribosome is carried out similarly as described in Protocols 1 and 2.

Protocol 4: Hydroxyl Radical Cleavage

- 1) Add 4 μl of a ribosome solution (5 pmol ribosomes in a buffer containing 7 mM MgCl_2) to 2 μl 5 x ribosome cleavage buffer.
- 2) After addition of 1 μl of a freshly prepared 10 x antibiotic solution, incubate 30 min at 37 °C followed by 10 min incubation at room temperature.
- 3) Add 1 μl of 2.5 mM FeCl_2 to the reaction tube and mix by brief centrifugation in a microcentrifuge.
- 4) Incubate 1 min before adding 1 μl of 12.5 mM sodium ascorbate.
- 5) After 1 min, add 1 μl of 12.5 mM hydrogen peroxide to initiate the reaction and mix rapidly. The final concentrations are 250 μM for Fe^{2+} and 1.25 mM for both sodium ascorbate and hydrogen peroxide in the presence of 5.8 mM Mg^{2+} .
- 6) To stop the cleavage reaction, add thiourea after 1 min to a final concentration of 125 mM. Precipitate the RNA with 2 μl glycogen (10 $\mu\text{g}/\mu\text{l}$) and 300 μl of approximately 90% ethanol containing 0.3 M sodium acetate, pH 5.0.

Protocol 5: Mg^{2+} -Competition of Fe^{2+} -Cleavage

To assure that Mg^{2+} and Fe^{2+} share the same or overlapping binding sites, it is essential to carry out a Mg^{2+} -competition experiment. The range of $\text{Fe}^{2+}:\text{Mg}^{2+}$ ratios necessary for cleavage competition depends on the respective affinities of Fe^{2+} and Mg^{2+} to the specific binding site. In the case of tetracycline, Fe^{2+} binds to the antibiotic 100-fold more tightly than Mg^{2+} [45].

- 1) Proceed as described in steps 1 and 2 of Protocol 4.
- 2) Mix equal volumes of a 5 mM FeCl_2 solution and an MgCl_2 solution (prepared from a 20 x stock solution), such that the desired final Mg^{2+} concentration is reached in the step corresponding to step 5 of Protocol 4 (note that you have to take into account the magnesium already present in the reaction). Add 1 μl of this $\text{Fe}^{2+}/\text{Mg}^{2+}$ -mixture to the reaction tube containing the RNA, and mix by brief centrifugation in a microcentrifuge.
- 3) Proceed exactly as described in steps 4 to 6 of Protocol 4.

Protocol 6: Extraction of rRNA

The rRNA has to be extracted for the following primer extension analysis.

- 1) Resuspend the pellet obtained after ethanol precipitation (step 6, Protocol 4) in 200 μl REB at room temperature.
- 2) Store the resuspended pellets at 4 °C until phenolization to remove the ribosomal proteins.
- 3) Dissolve the precipitated SDS by shaking gently at room temperature for 10 min, then perform the phenol extraction, followed by a phenol/chloroform/isoamylalcohol (25:24:1) and a chloroform/isoamylalcohol (24:1) extraction.
- 4) After an ethanol precipitation, redissolve the RNA in 30 μl Millipore-filtered, DEPC-treated water.

Protocol 7: Primer Extension Reaction

- 1) Mix 2 μ l of the redissolved rRNA (0.5–0.7 μ g/ μ l) with 1.5 μ l 5 x Q-Solution (Qiagen, One-step RT-PCR Kit) and 1 μ l of 5'-[^{32} P]-labeled primer (labeled using standard methods; [74]; see also Chapter 9). Carry out the hybridization by heating the mixture for 1 min at $94 \pm 2^\circ\text{C}$ followed by continuous cooling to $50 \pm 2^\circ\text{C}$ (not longer than 20–30 min).
- 2) Start the extension reaction by addition of 10.5 μ l extension mix, containing 2.6 μ l Millipore-filtered, DEPC-treated H_2O , 2 μ l 2.5 mM dNTP mix (2.5 mM each dATP, dGTP, dCTP and dTTP at neutral pH), 3 μ l 5 x First-strand buffer, 1 μ l 0.1 M dithiothreitol, 1.5 μ l 5 x Q-Solution, 0.2 μ l RNase Inhibitor (40 U/ μ l), and 0.2 μ l SuperScriptIII reverse transcriptase (200 U/ μ l). The extension reaction is carried out in a final volume of 15 μ l.
- 3) Incubate the reaction for 60 min at $50 \pm 2^\circ\text{C}$. To stop the extension reaction, the RNA is degraded by adding 4 μ l 1 M NaOH and incubating for 60 min at 42°C .
- 4) Neutralize the reaction mixture by adding 4 μ l of 1 M HCl and precipitate the cDNA with 2 μ l EDTA (0.5 M, pH 8.0), 2 μ l glycogen (10 μ g/ μ l), and 100 μ l ethanol/0.3 M NaOAc, pH 5.0. Precipitate the samples at $-20 \pm 2^\circ\text{C}$ for at least 3 h or overnight.
- 5) Collect ethanol precipitates by centrifugation, wash the pelleted cDNA with 70% ethanol and centrifuge again.
- 6) Redissolve the cDNA pellet in 10 μ l denaturing gel loading buffer supplemented with 0.25% sucrose.
- 7) Separate the cleavage products on 6–10% denaturing polyacrylamide-sequencing gels.

Sequencing reactions are obtained using standard methods [74] with unmodified or total cellular RNA.

Mapping Tetracycline Binding to RNA by Drug-Directed Fenton Cleavage

Using Protocol 7 with Q-solution, a higher extension temperature and an improved reverse transcriptase, we not only obtained a clearer picture of the cleavage sites in the individual gels, but were also able to identify several new cleavage sites in addition to the three prominent Fe^{2+} -mediated cleavage sites in the 16S rRNA that had been described previously [43]. All cleavage sites are in good agreement with published data for tetracycline from genetics [75, 76], biochemistry [77–79], and crystallography [80, 81]. Figure 15.3 shows sections of denaturing polyacrylamide gels with cleavage sites mapped to A892 in helix (h) 27 (G885–C912) and to U965 in the internal loop of h31 (A964–A969) (helical numbering according to Mueller and Brimacombe [82]). On the basis of crystal structures of 30S subunits complexed with tetracycline [80, 81], the affected residues in h31 are within 10 Å distance of tetracycline bound to site-1, which is formed by h31 and h34. The signal detected in h27 maps to tetracycline site-5, which is formed by helices h11 and h27. For tetracycline, but not for tigecycline, we have now obtained signals at nucleotide A892 with DMS modification and Fenton cleavage. Till now, protection at this

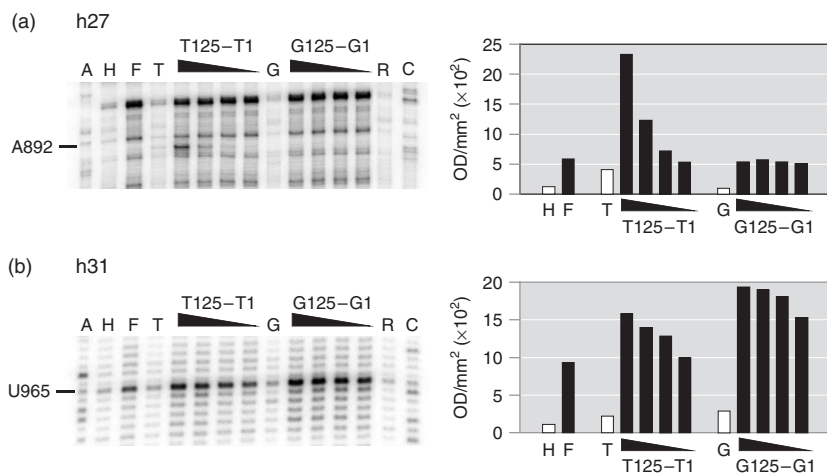


Figure 15.3 Fe^{2+} -mediated hydroxyl radical cleavage of the 16S rRNA. (a) Autoradiograph of a 6% denaturing PAA gel (left panel) showing cleavage sites in helix (h) 27 (tetracycline [Tc] binding site-5) and quantification of the cleavage signal intensity (right panel). (b) Autoradiograph of a 6% denaturing PAA gel showing cleavage sites in the internal loop of h31 (Tc binding site-1) and quantification of the cleavage signal intensity. **Autoradiographs** (left panels): Lanes A, C: A- or C-specific dideoxy sequencing lanes; R: unmodified RNA; H: control in which Fe^{2+} was omitted; F: RNA incubated with $\text{Fe}^{2+}/\text{H}_2\text{O}_2$ in the absence of antibiotic; T: control RNA incubated with 125 μM Tc and H_2O_2 ; T1–T125: $\text{Fe}^{2+}/\text{H}_2\text{O}_2$ cleavage in the presence of 1 μM , 5 μM , 25 μM , and 125 μM Tc; G: control RNA incubated with 125 μM tigecycline and H_2O_2 ; G1–G125, $\text{Fe}^{2+}/\text{H}_2\text{O}_2$ cleavage in the presence of

1 μM , 5 μM , 25 μM , and 125 μM tigecycline. *E. coli* 70S ribosomes were incubated with different amounts of tetracycline and tigecycline, and treated with $\text{Fe}^{2+}/\text{H}_2\text{O}_2$ as described in Protocol 4. Cleavage sites were detected by primer extension (Protocols 6 and 7) and analyzed by electrophoresis on a denaturing PAA gel. The nucleotide at which cleavage is affected by the respective antibiotic is indicated to the left of each gel. **Quantification plots** (right panels): the cleavage of 16S rRNA in the presence of rising amounts of tetracycline (T1–T125) or tigecycline (G1–G125) was quantified with a phosphorimager and compared with the control (lane F, cleaved rRNA in the absence of antibiotic). Quantification was adjusted for loading differences by standardization to regions unaffected by tetracycline or tigecycline.

position was interpreted to be due to the presence of a pseudoaxial OH group at position 6, which is lacking in tigecycline [43]. Based on this new data, we rather propose that tigecycline does not bind to site-5 at all. Because tigecycline has a lower minimum inhibitory concentration than tetracycline, we conclude that site-5 is not involved in the biological activity of tetracyclines.

Comparison of the cleavage patterns obtained at the two sites shown in Figure 15.3 highlights an important difference between a tetracycline molecule bound to site-1 and one bound to site-5 [80]. Cleavage at U965 (site-1) is quite pronounced even in the absence of the antibiotic. Cleavage intensity then increases with increasing amounts of tetracycline or tigecycline. Such a pattern suggests that a metal ion is already bound to this site prior to binding of the tetracycline–metal complex, which

is exactly what is seen in the crystal structure of the 30S subunit in the absence of tetracycline [80]. This metal ion's function is, most likely, to maintain the local RNA structure in this region. Tetracycline either binds to this metal ion or replaces it with its own complexed metal ion [45]. The additional contacts formed between functional groups of tetracycline and the RNA could stabilize the bound metal ion, leading to the increase in cleavage intensity. Alternatively, it cannot be excluded that the antibiotic binds at some distance to the prebound metal ion at site-1 and indirectly affects cleavage at U965 through a conformational effect. In contrast to U965, only weak background cleavage is observed at A892 (site-5) in the absence of tetracycline. Distinct cleavage is first detected in the presence of 5 μM tetracycline and increases in intensity with higher amounts of the drug. The crystal structure [80] shows no metal ion at this position and even does not identify one complexed to the bound tetracycline [80]. The Fenton cleavage observed here clearly indicates that this metal ion is brought to the binding site by the antibiotic. In conclusion, a careful analysis of the cleavage pattern allows one not only to detect sites that are close to the metal ion, and thus close to the drug, but also to reveal if these are preformed metal ion binding sites or if they instead represent ligand-mediated metal ion binding.

Fenton cleavage of an *in vitro*-selected tetracycline aptamer [83] with a tetracycline–iron(II) complex identified nucleotide A55 in the L3 loop as being close to the chelated ion in the drug–RNA complex [42]. This proximity was later confirmed in the crystal structure of the tetracycline–aptamer complex [2]. Importantly, the method is not just limited to the analysis of tetracycline binding to proteins or to RNA. Tight iron(II)-binding to the anticancer drug chromomycin A₃ resulted in cleavage of plasmid DNA *in vitro* in the presence of hydrogen peroxide, apparently only at a few sites. By mapping these cleavage sites, the DNA sequence specificity of the drug could be inferred [46].

15.2

Comments and Troubleshooting

- The native cleavage buffer and the denaturation/renaturation conditions should be adjusted to the requirements of the respective RNA molecule to be probed. For selection of buffer conditions, it is necessary to keep in mind that some buffer additives, for example, reducing reagents such as dithiothreitol and 2-mercaptoethanol, or EDTA, might scavenge radicals generated during the cleavage reaction.
- The crucial point for optimizing iron(II)-mediated cleavage is to adjust the iron(II) : magnesium(II) ratio to the requirements of the respective RNA molecule. On the one hand, magnesium(II) is needed for stabilizing the RNA. On the other hand, magnesium(II) competes with iron(II) for metal binding sites. One possible solution to circumvent RNA destabilization at lower magnesium concentrations is the substitution of magnesium by spermine and spermidine. These polyamines

did not disturb the cleavage reaction with the group I intron or small ribozyme RNAs (unpublished observations).

- Comparison of the reaction conditions for cleavage of the group I intron RNA and the 70S ribosomes indicated that the concentrations of the three chemical species (iron, sodium ascorbate, and hydrogen peroxide) might have to be varied to optimize the generation of hydroxyl radicals. We generally recommend using equimolar amounts of sodium ascorbate and hydrogen peroxide. For smaller RNAs (<500 nt) it is not necessary to precipitate the 5'- or 3'-labeled RNA for initial evaluation and optimization of the experimental protocol. Addition of an equal volume of denaturing loading buffer and immediate electrophoresis gives data of sufficient quality. However, a further ethanol precipitation step before gel loading is still recommended for final mapping of the cleavage sites and for quantification of the cleavage intensities.
- In contrast to the protocol described in the first edition of this Handbook, we have now altered the amounts of Fe^{2+} and spermidine when probing tetracycline binding sites on the ribosome. We doubled the Fe^{2+} concentration, and consequently also that of sodium ascorbate and hydrogen peroxide. We additionally reduced the spermidine concentration by a factor of 4. This improved the intensity of the Fenton cleavage signals, possibly due to quenching of singlet oxygen (which could interfere with free radical generation) by spermine and spermidine [84]. As mentioned above, we did not observe an effect of the spermine/spermidine concentration on cleavage intensity with RNA from group I introns or other small RNAs.
- An incubation time of 1 min before the addition of sodium ascorbate and hydrogen peroxide was the shortest period of time in which the six reaction tubes that fit into the desktop microcentrifuge (Qualitron) could be handled conveniently. Extending the incubation time for iron(II) and sodium ascorbate to 10 min did not result in different cleavage patterns. The extension of the reaction time after the addition of hydrogen peroxide might increase cleavage intensities, however it is recommended to change the FeCl_2 concentration if altered cleavage intensity is desired.
- Do not pre-mix the sodium ascorbate and iron(II) chloride solutions before adding them to the RNA, as they will form a chelate complex that effectively titrates the iron(II) from the reaction mixture.
- In order to distinguish between tetracycline-mediated cleavage and cleavage caused by metal ions either bound to distinct binding pockets or loosely associated with the negatively charged phosphodiester backbone of the RNA, it is necessary to perform control experiments with iron(II) and hydrogen peroxide in the absence of tetracycline. Tetracycline is known to bind RNA unspecifically. This may lead to additional cleavage sites at high concentrations. It is therefore advisable to titrate the reaction with tetracycline. In addition, one control should contain the highest amount of tetracycline used in the titration in the absence of iron(II) and hydrogen peroxide since binding of tetracycline to RNA could cause reverse transcription stops. For all controls, the compound omitted is substituted by Millipore-filtered, DEPC-treated water.

- *E. coli* mutant strains lacking the enzyme superoxide dismutase contain larger pools of intracellular free iron. If these cells are exposed to hydrogen peroxide, increased amounts of hydroxyl radicals are generated by Fenton chemistry leading to enhanced DNA damage [85]. Such strains might allow a transfer of the *in vitro*-probing approaches discussed here to *in vivo* analyses, as has been possible for kethoxal [86], dimethylsulfate [74], and lead [87]. Alternatively, synchrotron-generated hydroxyl radicals could be used for *in vivo* probing [54].

References

1. Atkins, J.F., Gesteland, R.F., and Cech, T.R. (eds) (2011) *RNA Worlds: From Life's Origins to Diversity in Gene Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
2. Xiao, H., Edwards, T.E., and Ferré-D'Amaré, A.R. (2008) Structural basis for specific, high-affinity tetracycline binding by an *in vitro* evolved aptamer and artificial riboswitch. *Chem. Biol.*, **15**, 1125–1137.
3. Chen, J.H., Yajima, R., Chadalavada, D.M., Chase, E., Bevilacqua, P.C., and Golden, B.L. (2010) Golden. A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry*, **49**, 6508–6518.
4. Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, **334**, 1524–1529.
5. Jenkins, J.L., Krucinska, J., McCarty, R.M., Bandarian, V., and Wedekind, J.E. (2011) Comparison of a *preQ1* riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J. Biol. Chem.*, **286**, 24626–24637.
6. Michel, F. and Westhof, E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.*, **216**, 585–610.
7. Piccirilli, J.A., Vyle, J.S., Caruthers, M.H., and Cech, T.R. (1993) Metal ion catalysis in the *Tetrahymena* ribozyme reaction. *Nature*, **361**, 85–88.
8. Ryder, S.P. and Strobel, S.A. (1999) Nucleotide analog interference mapping of the hairpin ribozyme: implications for secondary and tertiary structure formation. *J. Mol. Biol.*, **291**, 295–311.
9. Brunel, C. and Romby, P. (2000) Probing RNA structure and RNA-ligand complexes with chemical probes. *Methods Enzymol.*, **318**, 3–21.
10. Tullius, T.D. and Greenbaum, J.A. (2005) Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.*, **9**, 127–134.
11. Tullius, T.D., Dombroski, B.A., Churchill, M.E., and Kam, L. (1987) Hydroxyl radical footprinting: a high-resolution method for mapping protein-DNA contacts. *Methods Enzymol.*, **155**, 537–558.
12. Balasubramanian, B., Pogożelski, W.K., and Tullius, T.D. (1998) DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 9738–9743.
13. Celander, D.W. and Cech, T.R. (1990) Iron(II)-Ethylenediaminetetraacetic acid catalyzed cleavage of RNA and DNA oligonucleotides: similar reactivity toward single- and double-stranded forms. *Biochemistry*, **29**, 1355–1361.
14. Sclavi, B., Woodson, S., Sullivan, M., Chance, M.R., and Brenowitz, M. (1997) Time-resolved synchrotron X-ray “footprinting”, a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding. *J. Mol. Biol.*, **266**, 144–159.

15. Udenfried, S., Clark, C.T., Axelrod, J., and Brodie, R.B. (1954) Ascorbic acid in aromatic hydroxylation I. A model system for aromatic hydroxylation. *J. Biol. Chem.*, **208**, 731–739.
16. Cotton, F.A., Murillo, C.A., Bochmann, M., and Grimes, R.N. (1999) *Advanced Inorganic Chemistry*, 6th edn, John Wiley & Sons, Inc., New York.
17. Cheal, S.M., Ng, M., Barrios, B., Miao, Z., Kalani, A.K., and Meares, C.F. (2009) Mapping protein-protein interactions by localized oxidation: consequences of the reach of hydroxyl radical. *Biochemistry*, **48**, 4577–4586.
18. Rai, P., Cole, T.D., Wemmer, D.E., and Linn, S. (2001) Localization of Fe(2⁺) at an RTGR sequence within a DNA duplex explains preferential cleavage by Fe(2⁺) and H₂O₂. *J. Mol. Biol.*, **312**, 1089–1101.
19. Rai, P., Wemmer, D.E., and Linn, S. (2005) Preferential binding and structural distortion by Fe²⁺ at RGGG-containing DNA sequences correlates with enhanced oxidative cleavage at such sequences. *Nucleic Acids Res.*, **33**, 497–510.
20. Berens, C., Streicher, B., Schroeder, R., and Hillen, W. (1998) Visualizing metal-ion-binding sites in group I introns by iron(II)-mediated Fenton reactions. *Chem. Biol.*, **5**, 163–175.
21. Earnshaw, D.J. and Gait, M.J. (1999) The harpin ribozyme: the roles of ions in RNA cleavage. *Nucleic Acids Symp. Ser.*, **42**, 273–274.
22. Wei, C.H., Chou, W.Y., Huang, S.M., Lin, C.C., and Chang, G.G. (1994) Affinity cleavage at the putative metal-binding site of pigeon liver malic enzyme by the Fe²⁺-ascorbate system. *Biochemistry*, **33**, 7931–7936.
23. Ishida, H., Makino, A., and Mae, T. (1999) Fragmentation of the large subunit of ribulose-1,5-bisphosphate carboxylase by reactive oxygen species occurs near Gly-329. *J. Biol. Chem.*, **274**, 5222–5226.
24. Hlavaty, J.J., Benner, J.S., Hornstra, L.J., and Schildkraut, I. (2000) Identification of the metal-binding sites of restriction endonucleases by Fe²⁺-mediated oxidative cleavage. *Biochemistry*, **39**, 3097–3105.
25. Latham, J.A. and Cech, T.R. (1989) Defining the inside and outside of a catalytic RNA molecule. *Science*, **245**, 276–282.
26. Swisher, J., Duarte, C.M., Su, L.J., and Pyle, A.M. (2001) Visualizing the solvent-inaccessible core of a group II intron ribozyme. *EMBO J.*, **20**, 2051–2061.
27. Heuer, T.S., Chandry, P.S., Belfort, M., Celander, D.W., and Cech, T.R. (1991) Folding of group I introns from bacteriophage T4 involves internalization of the catalytic core. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 11105–11109.
28. Pan, T. (1995) Higher order folding and domain analysis of the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry*, **34**, 902–909.
29. Ding, F., Lavender, C.A., Weeks, K.M., and Dokholyan, N.V. (2012) Three-dimensional RNA structure refinement by hydroxyl radical probing. *Nat. Methods*, **9**, 603–608.
30. Fang, X., Pan, T., and Sosnick, T.R. (1999) A thermodynamic framework and cooperativity in the tertiary folding of a Mg²⁺-dependent ribozyme. *Biochemistry*, **38**, 16840–16846.
31. Ralston, C.Y., He, Q., Brenowitz, M., and Chance, M.R. (2000) Stability and cooperativity of individual tertiary contacts in RNA revealed through chemical denaturation. *Nat. Struct. Biol.*, **7**, 371–374.
32. Shcherbakova, I., Mitra, S., Beer, R.H., and Brenowitz, M. (2006) Fast Fenton footprinting: a laboratory-based method for the time-resolved analysis of DNA, RNA and proteins. *Nucleic Acids Res.*, **34**, e48.
33. Hüttenhofer, A. and Noller, H.F. (1992) Hydroxyl radical cleavage of tRNA in the ribosomal P site. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 7851–7855.
34. Rana, T.M. and Meares, C.F. (1990) Specific cleavage of a protein by an attached iron chelate. *J. Am. Chem. Soc.*, **112**, 2457–2458.
35. Greiner, D.P., Miyake, R., Moran, J.K., Jones, A.D., Negishi, T., Ishihama,

- A., and Meares, C.F. (1997) Synthesis of the protein cutting reagent iron (S)-1-(p-bromoacetamidobenzyl) ethylenediaminetetraacetate and conjugation to cysteine side chains. *Bioconjugate. Chem.*, **8**, 44–48.
36. Heilek, G.M. and Noller, H.F. (1996) Site-directed hydroxyl radical probing of the rRNA neighborhood of ribosomal protein S5. *Science*, **272**, 1659–1662.
37. Culver, G.M., Cate, J.H., Yusupova, G.Z., Yusupov, M.M., and Noller, H.F. (1999) Identification of an RNA-protein bridge spanning the ribosomal subunit interface. *Science*, **285**, 2133–2136.
38. Culver, G.M. and Noller, H.F. (2000) Directed hydroxyl radical probing of RNA from iron(II) tethered to proteins in ribonucleoprotein complexes. *Methods Enzymol.*, **318**, 461–475.
39. Joseph, S. and Noller, H.F. (2000) Directed hydroxyl radical probing using iron(II) tethered to RNA. *Methods Enzymol.*, **318**, 175–190.
40. Newcomb, L.F. and Noller, H.F. (1999) Directed hydroxyl radical probing of 16S ribosomal RNA in 70S ribosomes from internal positions of the RNA. *Biochemistry*, **38**, 945–951.
41. Huq, I., Tamilarasu, N., and Rana, T.M. (1999) Visualizing tertiary folding of RNA and RNA-protein interactions by a tethered iron chelate: analysis of HIV-1 Tat-TAR complex. *Nucleic Acids Res.*, **27**, 1084–1093.
42. Hanson, S., Bauer, G., Fink, B., and Suess, B. (2005) Molecular analysis of a synthetic tetracycline-binding riboswitch. *RNA*, **11**, 503–511.
43. Bauer, G., Berens, C., Projan, S.J., and Hillen, W. (2004) Comparison of tetracycline and tigecycline binding to ribosomes mapped by dimethylsulphate and drug-directed Fe²⁺ cleavage of 16S rRNA. *J. Antimicrob. Chemother.*, **53**, 592–599.
44. McMurry, L.M., Aldema-Ramos, M.L., and Levy, S.B. (2002) Fe(2⁺)-tetracycline-mediated cleavage of the Tn10 tetracycline efflux protein TetA reveals a substrate binding site near glutamine 225 in transmembrane helix 7. *J. Bacteriol.*, **184**, 5113–5120.
45. Ettner, N., Metzger, J.W., Lederer, T., Hulmes, J.D., Kisker, C., Hinrichs, W., Ellestad, G.A., and Hillen, W. (1995) Proximity mapping of the Tet repressor-tetracycline-Fe²⁺ complex by hydrogen peroxide mediated protein cleavage. *Biochemistry*, **34**, 22–31.
46. Hou, M.H., Lu, W.J., Lin, H.Y., and Yuann, J.M. (2008) Studies of sequence-specific DNA binding, DNA cleavage, and topoisomerase I inhibition by the dimeric chromomycin A3 complexed with Fe(II). *Biochemistry*, **47**, 5493–5502.
47. Merryman, C., Moazed, D., Daubresse, G., and Noller, H.F. (1999) Nucleotides in 23S rRNA protected by the association of 30S and 50S ribosomal subunits. *J. Mol. Biol.*, **285**, 107–113.
48. Nishiyama, T., Yamamoto, H., Shibuya, N., Hatakeyama, Y., Hachimori, A., Uchiumi, T., and Nakashima, N. (2003) Structural elements in the internal ribosome entry site of *Plautia stali* intestine virus responsible for binding with ribosomes. *Nucleic Acids Res.*, **31**, 2434–2442.
49. Ansel-McKinney, P. and Gehrke, L. (1997) Footprinting RNA in protein complexes with hydroxyl radicals, in *mRNA Formation and Function* (ed. J.D. Richter), Academic Press, San Diego.
50. Dixon, W.J., Hayes, J.J., Levin, J.R., Weidner, M.F., Dombroski, B.A., and Tullius, T.D. (1991) Hydroxyl radical footprinting. *Methods Enzymol.*, **208**, 380–413.
51. Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation. *Cell*, **108**, 557–572.
52. Ralston, C.Y., Sclavi, B., Sullivan, M., Deras, M.L., Woodson, S.A., Chance, M.R., and Brenowitz, M. (2000) Time-resolved synchrotron X-ray footprinting and its application to RNA folding. *Methods Enzymol.*, **317**, 353–368.
53. Adilakshmi, T., Bellur, D.L., and Woodson, S.A. (2008) Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly. *Nature*, **455**, 1268–1272.
54. Adilakshmi, T., Lease, R.A., and Woodson, S.A. (2006) Hydroxyl radical

- footprinting *in vivo*: mapping macromolecular structures with synchrotron radiation. *Nucleic Acids Res.*, **34**, e64.
55. Brown, I.B. (1988) What factors determine cation coordination numbers? *Acta Crystallogr. Sect. B*, **44**, 545–553.
 56. Honda, K., Smith, M.A., Zhu, X., Baus, D., Merrick, W.C., Tartakoff, A.M., Hattier, T., Harris, P.L., Siedlak, S.L., Fujioka, H. *et al.* (2005) Ribosomal RNA in Alzheimer disease is oxidized by bound redox-active iron. *J. Biol. Chem.*, **280**, 20978–20986.
 57. Woodson, S.A. (2005) Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.*, **9**, 104–109.
 58. Schnabl, J. and Sigel, R.K. (2010) Controlling ribozyme activity by metal ions. *Curr. Opin. Chem. Biol.*, **14**, 269–275.
 59. Auffinger, P., Grover, N., and Westhof, E. (2011) Metal ion binding to RNA. *Met. Ions Life Sci.*, **9**, 1–35.
 60. Frederiksen, J.K. and Piccirilli, J.A. (2009) Identification of catalytic metal ion ligands in ribozymes. *Methods*, **49**, 148–166.
 61. Philips, A., Milanowska, K., Lach, G., Boniecki, M., Rother, K., and Bujnicki, J.M. (2012) MetalionRNA: computational predictor of metal-binding sites in RNA structures. *Bioinformatics*, **28**, 198–205.
 62. Gonzalez, R.L. Jr. and Tinoco, I. Jr. (2001) Identification and characterization of metal ion binding sites in RNA. *Methods Enzymol.*, **338**, 421–443.
 63. Donis-Keller, H., Maxam, A.M., and Gilbert, W. (1977) Mapping adenines, guanines and pyrimidines in RNA. *Nucleic Acids Res.*, **4**, 2527–2538.
 64. Cate, J.H., Hanna, R.L., and Doudna, J.A. (1997) A magnesium ion core at the heart of a ribozyme domain. *Nat. Struct. Biol.*, **4**, 553–558.
 65. Cate, J.H. and Doudna, J.A. (1996) Metal-binding sites in the major groove of a large ribozyme domain. *Structure*, **4**, 1221–1229.
 66. Christian, E.L. and Yarus, M. (1992) Analysis of the role of phosphate oxygens in the group I intron from *Tetrahymena*. *J. Mol. Biol.*, **228**, 743–758.
 67. Christian, E.L. and Yarus, M. (1993) Metal coordination sites that contribute to structure and catalysis in the group I intron from *Tetrahymena*. *Biochemistry*, **32**, 4475–4480.
 68. Lindqvist, M., Sandström, K., Liepins, V., Strömberg, R., and Gräslund, A. (2001) Specific metal-ion binding sites in a model of the P4-P6 triple-helical domain of a group I intron. *RNA*, **7**, 1115–1125.
 69. Streicher, B., von Ahsen, U., and Schroeder, R. (1993) Lead cleavage sites in the core structure of group I intron-RNA. *Nucleic Acids Res.*, **21**, 311–317.
 70. Banatao, D.R., Altman, R.B., and Klein, T.E. (2003) Microenvironment analysis and identification of magnesium binding sites in RNA. *Nucleic Acids Res.*, **31**, 4450–4460.
 71. Rupert, P.B. and Ferré-D'Amaré, A.R. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, **410**, 780–786.
 72. Walter, N.G., Yang, N., and Burke, J.M. (2000) Probing non-selective cation binding in the hairpin ribozyme with Tb(III). *J. Mol. Biol.*, **298**, 539–555.
 73. Juneau, K., Podell, E., Harrington, D.J., and Cech, T.R. (2001) Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA-solvent interactions. *Structure*, **9**, 221–231.
 74. Liebeg, A. and Waldsich, C. (2009) Probing RNA structure within living cells. *Methods Enzymol.*, **468**, 219–238.
 75. Ross, J.I., Eady, E.A., Cove, J.H., and Cunliffe, W.J. (1998) 16S rRNA mutation associated with tetracycline resistance in a Gram-positive bacterium. *Antimicrob. Agents Chemother.*, **42**, 1702–1705.
 76. Trieber, C.A. and Taylor, D.E. (2002) Mutations in the 16S rRNA genes of *Helicobacter pylori* mediate resistance to tetracycline. *J. Bacteriol.*, **184**, 2131–2140.
 77. Moazed, D. and Noller, H.F. (1987) Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature*, **327**, 389–394.

78. Oehler, R., Polacek, N., Steiner, G., and Barta, A. (1997) Interaction of tetracycline with RNA: photoincorporation into ribosomal RNA of *Escherichia coli*. *Nucleic Acids Res.*, **25**, 1219–1224.
79. Noah, J.W., Dolan, M.A., Babin, P., and Wollenzien, P. (1999) Effects of tetracycline and spectinomycin on the tertiary structure of ribosomal RNA in the *Escherichia coli* 30 S ribosomal subunit. *J. Biol. Chem.*, **274**, 16576–16581.
80. Brodersen, D.E., Clemons, W.M., Carter, A.P., Morgan-Warren, R.J., Wimberly, B.T., and Ramakrishnan, V. (2000) The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell*, **103**, 1143–1154.
81. Pioletti, M., Schlünzen, F., Harms, J., Zarivach, R., Glühmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., Jacobi, C. *et al.* (2001) Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *EMBO J.*, **20**, 1829–1839.
82. Mueller, F. and Brimacombe, R. (1997) A new model for the three-dimensional folding of *Escherichia coli* 16S ribosomal RNA. I. Fitting the RNA to a 3D electron microscopic map at 20 Å. *J. Mol. Biol.*, **271**, 524–544.
83. Berens, C., Thain, A., and Schroeder, R. (2001) A tetracycline-binding RNA aptamer. *Bioorg. Med. Chem.*, **9**, 2549–2556.
84. Khan, A.U., Mei, Y.H., and Wilson, T. (1992) A proposed function for spermine and spermidine: protection of replicating DNA against damage by singlet oxygen. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 11426–11427.
85. McCormick, M.L., Buettner, G.R., and Britigan, B.E. (1998) Endogenous superoxide dismutase levels regulate iron-dependent hydroxyl radical formation in *Escherichia coli* exposed to hydrogen peroxide. *J. Bacteriol.*, **180**, 622–625.
86. Balzer, M. and Wagner, R. (1998) A chemical modification method for the structural analysis of RNA and RNA-protein complexes within living cells. *Anal. Biochem.*, **256**, 240–242.
87. Lindell, M., Romby, P., and Wagner, E.G. (2002) Lead(II) as a probe for investigating RNA structure *in vivo*. *RNA*, **8**, 534–541.

16

Measuring the Stoichiometry of Magnesium Ions Bound to RNA

Andrew J. Andrews and Carol A. Fierke

16.1

Introduction

Ribonucleic acids (RNAs) are large polyanions containing negative charges on the many backbone phosphodiester groups that interact with positively charged ions in the solution by charge-charge or Coulombic interactions [1]. This Coulombic field of negative charge attracts large numbers of positively charged counter ions. These counter ions typically consist of both monovalent ions, such as potassium or sodium, and divalent ions, such as magnesium. These ions can both loosely associate with the phosphodiester backbone of RNA mainly by electrostatic interactions, forming an “ionic cloud,” or form specific interactions with the RNA backbone and bases to bind in a unique position [2]. Specific binding sites can involve direct coordination of the cation by the RNA (inner-sphere interaction) or a hydrogen bond contact *via* a coordinated water molecule (outer-sphere interaction).

In this chapter we present a method to measure the total number of divalent Mg^{2+} ions, including ions that are either specifically or electrostatically bound, that interact with an RNA molecule under a given set of conditions. This method can be useful for determining how the composition of the electrostatic cloud is affected by solution conditions, and whether changes in the RNA structure or the addition of protein cofactors affect the number and composition of interacting ions [3]. For instance, alteration of a bulged nucleotide that results in the narrowing of the major groove or addition of a protein component may reduce the local or total ionic charge of the RNA and, therefore, the number or affinity of associated Mg^{2+} ions. Quantification of the number of ions that bind to or interact with polyanions, such as RNA or DNA, becomes more difficult as the size of the polyanion, and therefore the number of bound ions, increases. In some cases, ions bound to the RNA can be distinguished from free ions using spectroscopic analysis, for example, electron paramagnetic resonance spectroscopy has been used to measure the binding of manganese ions to RNA [4]. However, a more general approach is to physically separate the bound and free metal ions and then determine the concentration of ions in both fractions.

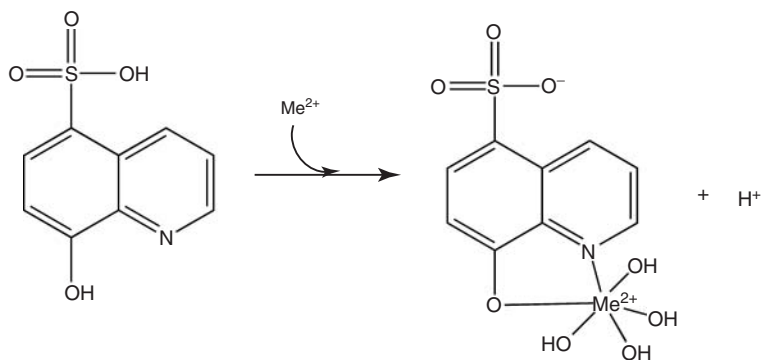


Figure 16.1 The chemical structure of 8-hydroxyquinoline-5-sulfonic acid (HQS) binding to a metal ion as seen in the nickel crystal structure [6]. The magnesium-HQS complex is assumed to have a similar structure.

The main methods for separating ions bound to large RNAs are gel filtration chromatography, equilibrium dialysis, and forced equilibrium dialysis [5]. Forced dialysis has given the most consistent results and is the easiest to perform; therefore this method will be the main focus of this chapter. Once the bound and free ions are separated, the total concentration of ions in each sample can be measured using a number of techniques (inductively coupled plasma emission mass spectrometry (ICP-MS), atomic absorption, or the fluorescent dye binding method) but most require expensive equipment and in-depth training, except for the fluorescent methods. A number of fluorophores have been used to detect metal ions including HQS (8-hydroxyquinoline-5-sulfonic acid). HQS at neutral pH is minimally fluorescent, but becomes highly fluorescent upon binding to magnesium and forming a magnesium-8-quinolinol complex (Figure 16.1). Here we describe in detail the use of the fluorophore HQS for determining magnesium concentrations following the separation of bound and free magnesium ions.

16.2

Separation of Free Mg^{2+} from RNA-bound Mg^{2+}

Equilibrium dialysis is still the best way to separate small molecules from ones associated with larger molecules under equilibrium conditions. While numerous methods of dialysis exist, from the traditional simple dialysis tubing to fully automated machines, the deciding factors are the concentration and volume of RNA required, and the amount of time needed for the dialysis experiment to reach equilibrium. Long equilibration times and large volumes often limit traditional equilibrium dialysis experiments. The ideal method would use a small amount of sample, have a high degree of precision, and be rapid. Forced dialysis is an equilibrium method with shorter equilibration times and smaller volumes than traditional dialysis methods. While new and more advanced methods for the

dialysis of small samples are available, the use of ultrafiltration devices provides a quick and cheap alternative.

16.3

Forced Dialysis Is the Preferred Method for Separating Bound and Free Mg^{2+}

The main advantages of forced dialysis compared to conventional dialysis are: (i) equilibration occurs in the absence of a dividing membrane, which significantly decreases the amount of time needed to reach equilibrium, (ii) small volumes (approximately 100 μl) of sample are required, and (iii) separation of the free and bound molecules is rapid. The forced dialysis method is accomplished by first allowing the sample to equilibrate at a given set of experimental conditions in the absence of a separating membrane. Separation is then carried out after equilibrium is reached. A simple method for carrying out this separation is to use a Microcon Ultra-0.5 ml centrifugal filter device (Millipore, Billerica, MA www.millipore.com) or other similar products. Microcon filter devices are manufactured with membranes that limit the size of the nucleic acid that can pass through the pores, with limitation sizes between approx. 10 and 300 single-stranded nucleotides (3–100 kDa). Liquid is forced through the membrane by centrifugation (up to 14 000 g) which separates unbound small molecules from those associated with the larger nucleic acid (Figure 16.2). Furthermore, equilibrium is maintained throughout the experiment as the concentration of unbound ligand remains constant. Two important notes are that Microcon devices, as manufactured, contain a small amount of glycerin in the filter and the filter has a $\sim 10 \mu\text{l}$ retention volume. Before beginning any dialysis experiment, it is necessary to confirm that your ligand of interest can pass through the membrane easily and that neither component will preferentially bind to the dialysis membrane.

The success of this experiment also depends significantly on the purity of the RNA and the reagents. RNA samples should be as clean as possible, lacking divalent metal ions or contaminating chelators (i.e., Ethylenediaminetetraacetic

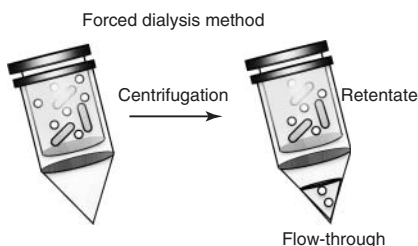


Figure 16.2 A cartoon illustrating the forced dialysis method. Spheres represent Mg^{2+} ions and rods symbolize RNA molecules. After enough time has elapsed to reach equilibrium, the device is centrifuged so that

a small amount of liquid (20% or less) is forced through the dialysis membrane. The retentate contains RNA-bound Mg^{2+} plus free Mg^{2+} , whereas the flow-through contains only free Mg^{2+} .

acid (EDTA)). All solutions should be prepared using metal-free tubes and pipet tips. Buffers should be as close to physiological or assay conditions as possible, with close to neutral pH, low monovalent salt concentrations and varying concentrations of Mg^{2+} ions. All small molecules with high affinity for Mg^{2+} , such as EDTA, should be removed. Excess metal ions can be removed from microfuge tubes by soaking in 100 mM EDTA overnight followed by extensive washing with metal-free water. To remove the glycerin found in the membranes, the device should be washed by centrifugation of at least 500 μ l of metal-free water followed by the same amount of buffer. The RNA is best prepared in a low salt buffer (such as Tris) and allowed to rebind in the experimental buffer using conditions best suited for the RNA of interest.

The minimum concentration of RNA needed for this experiment is dependent on both the binding affinity for Mg^{2+} and the stoichiometry for Mg^{2+} . There must be sufficient RNA to bind enough Mg^{2+} ions such that the total Mg^{2+} concentration ($[Mg^{2+}]_{tot} = [Mg^{2+}]_{bound} + [Mg^{2+}]_{free}$) is greater than the free Mg^{2+} concentration alone. The following Eq. (16.1) describing the fractional saturation of ligand binding demonstrates that for a stoichiometry (n) of 1, a $K_{1/2}^{Mg}$ of 1 mM and a measured $[Mg^{2+}]_{free}$ of 1 mM, the concentration of RNA would need to be 1 mM to see a twofold difference between the bound and free Mg^{2+} fractions:

$$\frac{[RNA : Mg^{2+}]}{[RNA]_{total}} = \frac{n [Mg^{2+}]_{free}}{([Mg^{2+}]_{free} + K_{1/2}^{Mg})} \quad (16.1)$$

For the just mentioned example, it follows:

$$[RNA : Mg^{2+}] = \frac{[RNA]_{total} n [Mg^{2+}]_{free}}{([Mg^{2+}]_{free} + K_{1/2}^{Mg})}$$

$$ergo, [RNA : Mg^{2+}] = 1 \text{ mM} \times 1 \times 1 \text{ mM} / (1 \text{ mM} + 1 \text{ mM}) = 0.5 \text{ mM}$$

However, most RNAs have a stoichiometry for bound Mg^{2+} that is much higher than 1, which lowers the needed concentration of RNA. For instance if $n = 100$, $K_{1/2}^{Mg} = 1 \text{ mM}$ and the measured $[Mg^{2+}]_{free} = 1 \text{ mM}$, then only 20 μ M RNA is required to achieve equal concentrations of free and bound Mg^{2+} :

$$[RNA : Mg^{2+}] = 0.02 \text{ mM} \times 100 \times 1 \text{ mM} / (1 \text{ mM} + 1 \text{ mM}) = 1 \text{ mM}.$$

Therefore, as the size of the RNA and the resulting Coulombic field increases, the concentration of RNA required to see a measurable difference in fluorescence decreases. The number of metal ions binding to RNA can be estimated to be between 0.3 and 0.7 M^{2+} /nucleotide [3, 7].

To initiate the experiment, the RNA is diluted into buffer containing Mg^{2+} . A recommended buffer volume for this experiment is 100 μ l, but this depends on the size of the filter device. This allows for the removal of sufficient volume for the analysis of free Mg^{2+} without changing the RNA concentration by a large amount. The half-time for equilibration can be estimated from the K_D ($\sim K_{1/2}^{Mg}$) and a

reasonable guess of the association rate constant (k_a), assuming a simple two-step binding reaction, as shown in Eq. (16.2).

$$t_{1/2} = \frac{\ln 2}{k_{\text{obs}}} = \frac{0.693}{k_a([\text{Mg}^{2+}] + K_D)} \quad (16.2)$$

Even assuming a value of K_D of 1 μM and a slow second-order association rate constant of $1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$, the calculated $t_{1/2}$ is 3 s suggesting that a 30 s equilibrium time should be sufficient. In practice, the best way to test whether the incubation time is adequate to achieve equilibrium is to demonstrate that doubling (or halving) the equilibration time does not alter the final result.

Once equilibrium is reached, the sample can be added to the pre-washed Microcon device. After adding the RNA to the device, centrifuge the microcon at 7000 g until a small volume has passed through ($\sim 50 \mu\text{l}$). The flow-through can then either be added back to the Microcon or replaced with an equal volume of the original Mg^{2+} buffer. Allow the reaction mixture to re-equilibrate (i.e., 15 min at 25°C) then spin again at 7000 g to collect a flow-through containing only the free Mg^{2+} . The flow-through, which should be a small percentage (20% or less) of the total sample volume, represents the $[\text{Mg}^{2+}]_{\text{free}}$ and is ready for analysis. The $[\text{Mg}^{2+}]_{\text{total}}$ sample can be obtained directly from the solution that is retained in the top of the Microcon.

16.4

Alternative Methods for Separating Free and Bound Mg^{2+} Ions

Size exclusion chromatography and equilibrium dialysis are alternative methods that can be used to separate free from bound Mg^{2+} ions. In all cases, care needs to be taken to make sure all solutions have no other divalent metals beside Mg^{2+} and no chelators of Mg^{2+} .

Gel filtration columns based on size exclusion matrices such as Sephadex G-50 and others can be used to separate small molecules from large RNAs. However, even small columns take on the order of 5 min to run by gravity flow which allows ample time for re-equilibration of the metal ions during the separation. Gel filtration spin columns are an alternative rapid separation method [8, 9]. Using gel filtration spin columns, separation can be accomplished on the order of seconds, which greatly reduces the likelihood of re-equilibration during separation. However, for affinities in the μM range, the dissociation rate constant for metal ions from RNA could be on the order of $1\text{--}10 \text{ s}^{-1}$ ($t_{1/2} = 0.1\text{--}1 \text{ s}$) indicating that re-equilibration can occur on the same time scale as the separation by the spin column method. Therefore, this technique is only applicable for tight, and thus slowly equilibrating, Mg^{2+} binding sites.

Equilibrium dialysis is probably the most well-known and common method of determining the stoichiometry and affinity of ligand binding sites. Equilibrium dialysis is often quite slow, taking many hours to complete since equilibration through the membrane can be slow [8]. Therefore, it is important to test both the

Mg^{2+} equilibration time and the RNA stability. As before, great care should be taken to make sure that all solutions, tubes, and equipment are free from metals or chelators that could interfere with the experiments. To exclude any RNase contamination, the equipment used should be autoclaved if possible and then soaked in 70% ethanol followed by a 100% ethanol wash. After choosing the correct dialysis membrane, the next important step is to determine the equilibration time. This can be estimated from measuring the time required for a Mg^{2+} solution to reach equal concentrations in both chambers of the equilibrium dialysis apparatus in the absence of RNA. After the dialysis experiment is complete, the RNA should be analyzed by gel electrophoresis to determine the extent of degradation that has occurred.

16.5

Determining the Concentration of Free Mg^{2+} in the Flow-Through

Mg^{2+} standards should be made in the same buffer as the RNA being tested. The standard solutions with varying Mg^{2+} concentrations will be used to generate a fluorescence standard curve from which the concentration of Mg^{2+} in the sample can be inferred. Three very important factors need to be addressed in the use of HQS to measure bound Mg^{2+} . (i) The RNA needs to be denatured to prevent high affinity sites from competing with HQS for Mg^{2+} binding. Therefore, guanidine-HCl is included in the assay buffer (5 M guanidine). (ii) The pH should be near neutral since, at high pH, HQS ionizes which increases the background fluorescence and decreases the sensitivity of the metal analysis. (iii) HQS should be in high enough concentration for quantitative binding of all Mg^{2+} present in the sample. Therefore, the HQS concentration should be much larger than both the Mg^{2+} dissociation constant (K_d) and the total Mg^{2+} concentration. A small volume of the sample to be tested (20 μl or less) is mixed with 150 μl of the HQS solution and the fluorescence is then measured in a 120- μl cuvette at 25 °C. The recommended assay conditions are 5 M guanidine-HCl, 0.1 M Tris pH 8 and 2 mM HQS. The excitation wavelength is set at 397 nm and the emission wavelength at 502 nm (Figure 16.3). At this point, the experimentally determined fluorescence can be compared to the standard curve of fluorescence as a function of $[\text{Mg}^{2+}]$ in the same buffer to calculate the concentration of Mg^{2+} in the experimental samples.

16.6

How to Determine the Concentration of Mg^{2+} Bound to the RNA and the Number of Binding Sites on the RNA

In the forced dialysis method, the concentration of $[\text{Mg}^{2+}]_{\text{bound}}$ can be calculated by subtracting $[\text{Mg}^{2+}]_{\text{free}}$, determined in the flow-through, from $[\text{Mg}^{2+}]_{\text{total}}$ determined in the retentate. Similarly, in a standard equilibrium dialysis experiment $[\text{Mg}^{2+}]_{\text{bound}}$ can be determined by measuring the difference between the Mg^{2+} concentrations on the two sides of the dialysis membrane. The $[\text{Mg}^{2+}]_{\text{free}}$ is measured on the

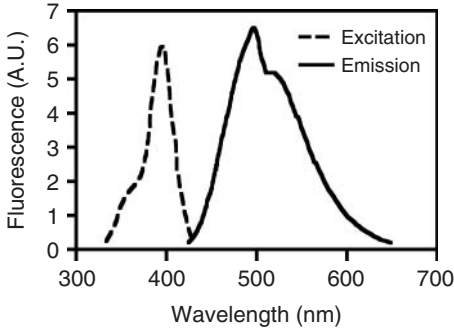


Figure 16.3 The excitation and emission spectra of Mg^{2+} -HQS. The intensity of the fluorescence emission spectrum of HQS (not shown) is ~ 70 -fold lower than the Mg^{2+} -HQS intensity. Emission spectra were collected on an Amico-Bowman Series 2

spectrometer (ThermoSpectronic, Rochester, N.Y.) by exciting the fluorophore at 397 nm. The slit width for both excitation and emission spectra was 8 nm. The sample was adjusted to 0.1 M Tris-HCL pH 8, 2 mM HQS, with or without 1 mM Mg^{2+} at 25 °C.

side without the RNA, while the concentration on the side with RNA equals $[Mg^{2+}]_{\text{bound}} + [Mg^{2+}]_{\text{free}}$. Therefore, the concentration of bound Mg^{2+} can be determined by subtracting the Mg^{2+} concentration on the “free” side from the concentration on the side with the RNA (Eq. (16.3)):

$$[Mg^{2+}]_{\text{bound}} = ([Mg^{2+}]_{\text{bound}} + [Mg^{2+}]_{\text{free}}) - [Mg^{2+}]_{\text{free}} \quad (16.3)$$

As stated before, important control experiments include demonstrating that the membrane does not bind Mg^{2+} ions to a significant extent and that the RNA of interest does not affect the determination of the Mg^{2+} concentration. If the RNA competes with HSQ for Mg^{2+} binding, you can either: (i) degrade the RNA using a non-specific RNase or alkaline hydrolysis to eliminate the high affinity Mg^{2+} binding sites, (ii) measure the Mg^{2+} standard curve in the presence of a comparable concentration of the competing RNA, or (iii) estimate the Mg^{2+} bound to RNA solely from the concentration of free and total Mg^{2+} by subtracting the total amount of free Mg^{2+} (equal to two times the amount of Mg^{2+} measured in the no RNA chamber to account for volume of both chambers of the dialysis apparatus) from the total amount of Mg^{2+} added (Eq. (16.4)):

$$\text{nmol } Mg^{2+}_{\text{bound}} = \text{nmol } Mg^{2+}_{\text{total}} - 2 \times \text{nmol } Mg^{2+}_{\text{free}} \quad (16.4)$$

Example: if one fills 100 μl of a 1 mM Mg^{2+} solution into one chamber of an equilibrium dialysis device (equals 100 nmol Mg^{2+}) and 100 μl of a 1 mM Mg^{2+} / 20 μM RNA solution into the other chamber, and these 2 nmol of RNA bind 50 nmol Mg^{2+} , then the free Mg^{2+} will equilibrate to 75 nmol Mg^{2+} in both chambers, and it follows:

$$\text{nmol } Mg^{2+}_{\text{bound}} = 200 \text{ nmol } Mg^{2+}_{\text{total}} - 2 \times 75 \text{ nmol } Mg^{2+}_{\text{free}} = 50 \text{ nmol}.$$

Once the concentrations of bound and free Mg^{2+} are determined at different Mg^{2+} concentrations, the $K_{1/2}^{Mg}$ and the stoichiometry, n , can be determined by fitting

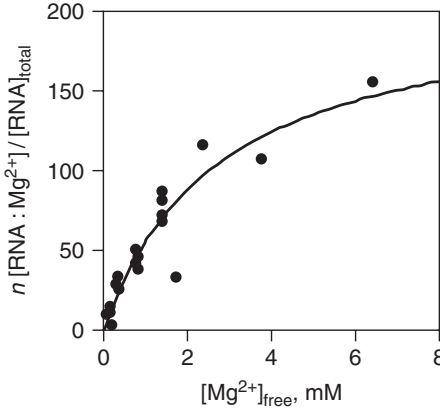


Figure 16.4 A plot showing the amount of Mg^{2+} bound to the RNase P holoenzyme at varying concentrations of free Mg^{2+} ions determined using the forced dialysis method [3]. The stoichiometry of magnesium bound to the holoenzyme is 160 Mg^{2+} per enzyme and the $K_{1/2}^{\text{Mg}}$ is 1.5 mM [3].

Eq. (16.5) to these data (Figure 16.4) [3, 8]. This analysis assumes that the Mg^{2+} ions bind to the RNA *via* multiple, independent binding sites [5], where ν equals Mg^{2+} bound divided by the total RNA added ($\nu = n[\text{RNA} : \text{Mg}^{2+}]/[\text{RNA}]_{\text{total}}$), such that ν varies from 0 to $n[\text{RNA}]_{\text{total}}$.

$$\nu = \frac{\left(\frac{n[\text{Mg}^{2+}]_{\text{free}}}{K_{1/2}} \right)}{\left(\frac{1 + [\text{Mg}^{2+}]_{\text{free}}}{K_{1/2}} \right)} \quad (16.5)$$

This assumption can be tested by making a Scatchard plot where $\nu/[\text{Mg}^{2+}]_{\text{free}}$ is plotted *versus* ν ; the slope of a linear fit equals $-1/K_{1/2}$ (Eq. (16.6)). If only one type of site is observed, the Scatchard plot will be linear. If multiple types of binding sites are observed, the Scatchard plots will not be linear [5]. In the case of most large RNAs, the ratio of specifically bound ions to loosely bound or interacting Mg^{2+} ions is so small that only one class of ions will be seen.

$$\frac{\nu}{[\text{Mg}^{2+}]_{\text{free}}} = \frac{n}{K_{1/2}} - \frac{\nu}{K_{1/2}} \quad (16.6)$$

So far, we have made the assumption that Mg^{2+} ions will bind to RNA in a non-cooperative fashion (Eq. (16.1)). However, if the Mg^{2+} ions bind cooperatively, the fraction of Mg^{2+} bound to RNA will be described by Eq. (16.7) (assuming that $[\text{Mg}^{2+}] \gg [\text{RNA}]$), so the binding can be measured and analyzed in a manner similar to what has been described for non-cooperative binding.

$$\frac{[\text{RNA} : \text{Mg}^{2+}]}{[\text{RNA}]_{\text{total}}} = \frac{[\text{Mg}^{2+}]_{\text{free}}^n}{([\text{Mg}^{2+}]_{\text{free}}^n + (K_{1/2})^n)} \quad (16.7)$$

16.7

Conclusion

This method is a straightforward way to quantitate the number of metal ions directly interacting with RNA by using common laboratory equipment. The important step in this method that is missing in other systems is the separation of free and bound magnesium ions. Other systems add HQS directly to the RNA and measure competition of the two ligands for the magnesium ions [10]. While these methods may work for one or two binding sites, the analysis of this type of competition experiment becomes extremely complicated as the number of binding sites increases. The measurement of the number of metal ions interacting with an RNA molecule under various conditions will advance our knowledge of RNA-metal ion interactions and will be useful for testing the validity of RNA modeling techniques.

16.8

Troubleshooting

- 1) If you are using Microcons, make sure that you have removed the glycerin from the filter, and you have not exceeded the maximum *g* force.
- 2) Confirm that RNA, but not Mg^{2+} , is retained by the membrane and that neither RNA nor Mg^{2+} sticks to the membrane.
- 3) Make sure that all solutions, tips, and tubes are free from contaminating metals, chelators, and RNases.
- 4) Determine that the RNA is folded and stable throughout the experiment.
- 5) Demonstrate that the reaction was incubated for sufficient time to reach equilibrium. Make sure that the Mg^{2+} in the experimental samples is determined under exactly the same conditions as the Mg^{2+} standard curve (buffer, temperature, wavelength range, etc.).

References

1. Manning, G.S. and Ray, J. (1998) Counterion condensation revisited. *J. Biomol. Struct. Dyn.*, **16**, 461–476.
2. Misra, V.K. and Draper, D.E. (1998) On the role of magnesium ions in RNA stability. *Biopolymers*, **48**, 113–135.
3. Kurz, J.C. and Fierke, C.A. (2002) The affinity of magnesium binding sites in the *Bacillus subtilis* RNase P x pre-tRNA complex is enhanced by the protein subunit. *Biochemistry*, **41**, 9545–9558.
4. DeRose, V.J. (2003) Metal ion binding to catalytic RNA molecules. *Curr. Opin. Struct. Biol.*, **13**, 317–324.
5. Cantor, C.R. and Schimmel, P.R. (1980) *Biophysical Chemistry*, W.H. Freeman and Company, New York.
6. Raj, S.B., Muthiah, P.T., Bocelli, G., and Righi, L. (2001) Metal-nucleobase interaction: dibromobis(cytosine)cadmium(II). *Acta Crystallogr.*, **E57**, m591–m594.
7. Feig, A.L. and Uhlenbeck, O.C. (1999) in *The RNA World*, 2nd edn (eds R.F. Gesteland, T.R. Cech, and J.F. Atkins), Cold Spring Harbor Laboratory Press, Cold Springs Harbor, pp. 287–319.

8. Beebe, J.A., Kurz, J.C., and Fierke, C.A. (1996) Magnesium ions are required by *Bacillus subtilis* ribonuclease P RNA for both binding and cleaving precursor tRNA^{ASP}. *Biochemistry*, **35**, 10493–10505.
9. Garcia, J.J., Gomez-Puyou, A., Maldonado, E., and Tuena De Gomez-Puyou, M. (1997) Acceleration of unisite catalysis of mitochondrial F1-adenosinetriphosphatase by ATP, ADP and pyrophosphate–hydrolysis and release of the previously bound [gamma-³²P]ATP. *Eur. J. Biochem.*, **249**, 622–629.
10. Serebrov, V., Clarke, R.J., Gross, H.J., and Kisselev, L. (2001) Mg²⁺-induced tRNA folding. *Biochemistry*, **40**, 6688–6698.

17

Nucleotide Analog Interference Mapping and Suppression (NAIM/NAIS): a Combinatorial Approach to Study RNA Structure, Folding, and Interaction with Proteins

Olga Fedorova, Marc Boudvillain, and Christina Waldsich

17.1

Introduction

17.1.1

NAIM: a Combinatorial Approach for RNA Structure–Function Analysis

Mutational analysis has traditionally been a powerful tool for elucidating the contribution of functional groups in RNA [1]. In these studies, the identity of an entire nucleobase is changed or, in a more precise adaptation, a single functional group is altered. The latter modifications are incorporated through chemical synthesis of RNA (if the RNA is no longer than ~50 nt) or a by a combination of chemical and enzymatic synthesis (for longer RNAs). The modified species are then analyzed for function in parallel with unmodified RNA (i.e., to evaluate binding or a reaction of interest). The relative importance of the functional group is then determined from the difference in activity between modified and unmodified molecules. This approach necessitates the synthesis of many different modified RNA molecules, in which important atoms are changed one at a time. The systematic screening of functionalities by this method requires an enormous synthetic effort. Furthermore, this method is simply inapplicable for screening the importance of functional groups throughout large RNA molecules such as group II introns, which are often ~1000 nt in length.

The problem of identifying specific atoms that are essential to the functioning of a large RNA is solved by nucleotide analog interference mapping (NAIM), which was pioneered by Gish and Eckstein [2, 3] and further developed by other researchers [4–6]. In a single experiment, this method screens a combinatorial library of modified RNA molecules for atoms that are important for function. The RNA of interest is prepared by *in vitro* transcription in the presence of a nucleoside analog thiotriphosphate that contains desired sugar or base modifications. Because it is susceptible to cleavage by iodine, the phosphorothioate linkage serves as a tag that reflects the location of a modified residue.

NAIM has a variety of applications including identification of atoms and functional groups critical for ribozyme folding and catalysis [4,6–9], determination

of free energy contributions of individual functional groups to ligand binding [10], elucidation of catalytic mechanisms by identifying nucleobases involved in ribozyme catalysis [11–13], analysis of RNA structural motifs by using their NAIM “signatures” [9, 14, 15], and location of metal binding sites in RNA [16, 17]. NAIM subsequently evolved into NAIS (nucleotide analog interference suppression), which permits the precise identification of tertiary interaction partners [6, 15, 18–20]. In combination, the two methods now represent the most powerful biochemical approach for RNA structure analysis.

17.1.1.1 Description of the Method

NAIM involves the following steps [21, 22]: (i) preparation, by *in vitro* transcription, of an RNA library containing statistically incorporated modified phosphorothioate nucleotide analogs (~1 incorporated analog per RNA molecule). (ii) Selection and separation: a process of interest (i.e., binding, folding, or catalytic reaction) is used to separate active from inactive RNA fractions. (iii) Visualization of interference and enhancement effects: iodine cleavage is used to detect the location of modified atoms that affect function (Figure 17.1a). In order to detect cleaved molecules, the RNA pool must be labeled at either the 3'- or the 5' terminus. In some cases, precursor molecules are labeled after transcription [22, 23], while in other studies, functional molecules are radiolabeled during the selection step [6].

For preparing the pool of randomly modified RNAs, the ratio of a thiotriphosphate analog to a corresponding unmodified nucleoside triphosphate (NTP) in the transcription mixture is adjusted so that the analog is incorporated with an efficiency of 1–2 modifications per molecule. Incorporation levels of up to 5% are considered satisfactory [21, 23]. For transcription efficiency and ease of analysis, only one modified analog is usually included at a time in a given transcription reaction [22, 23].

A typical NAIM experiment is exemplified as follows. To determine the functional importance of guanosine exocyclic amino groups in an RNA, it is transcribed in the presence of inosine α -thiotriphosphate. The resulting pool of RNA molecules has a statistical probability of including a single inosine substitution at each guanosine position in the RNA (Figure 17.1a). The RNA pool is then bound to another molecule or it undergoes a catalytic reaction of interest. If the inosine modification is located at a position where it interferes with binding or reaction, this RNA molecule will be depleted from the pool of functional molecules. The fraction of reacted molecules is then separated from unreacted molecules using electrophoresis or an affinity column, although separation methods vary widely depending on the application of interest. Then the reacted fraction (functional pool) as well as the precursor RNA pool and/or the unreacted fraction are treated with iodine and analyzed on a denaturing polyacrylamide gel (Figure 17.1a). If the modification at a certain position interferes with function, the corresponding band will be missing or underrepresented in the lane that represents the pool of active RNAs (Figure 17.1a). Conversely, if the modification at a certain position stimulates function, the corresponding band will be overrepresented in the active RNA pool (Figure 17.1a). Since the phosphorothioate linkage itself can interfere

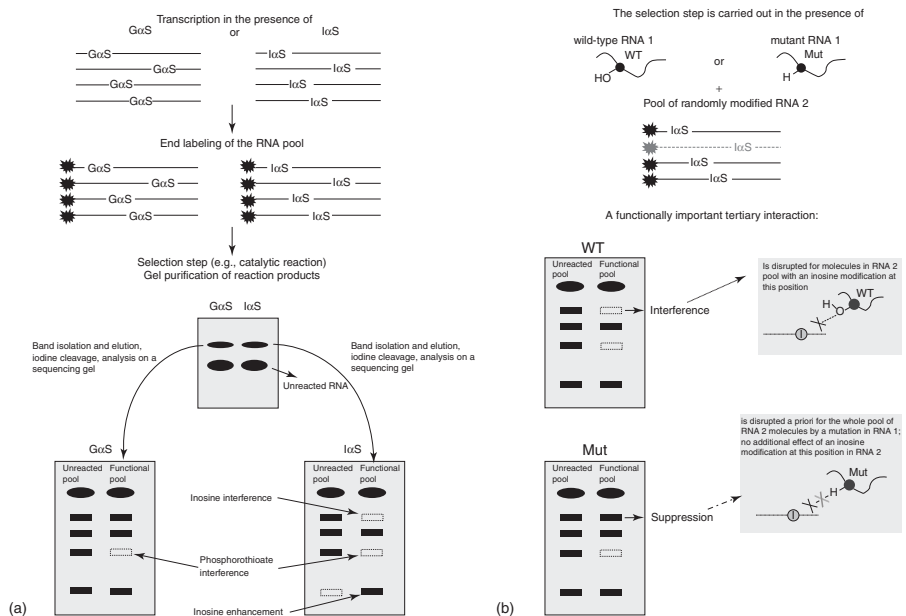


Figure 17.1 A schematic representation of NAIM (a) and NAIS (b) assays. Note that in (b) the term “mutant RNA 1” is used not to describe a classical base mutation but a single-atom substitution (2'-OH is changed to 2'-H). In NAIS literature, this is frequently done for the purpose of simplicity.

with or enhance activity, all NAIM assays contain a control experiment where only phosphorothioate is incorporated and its effects on function are analyzed in parallel.

17.1.2

NAIS: a Chemogenetic Tool for Identifying RNA Tertiary Contacts and Interaction Interfaces

17.1.2.1 General Concepts

Although NAIM is useful for identifying the locations of important atoms in an RNA molecule, it does not provide information on how these atoms interact and it does not reveal tertiary interaction partners. However, by comparing the NAIM profiles of a wild-type molecule with that of a mutant, one can discern patterns of energetic communication and thus RNA tertiary contacts. This variation of the NAIM experiment, which is called *nucleotide analog interference suppression*, is generally carried out after all interference (NAIM) data has been collected. The original NAIM patterns are used as a guide for the design of mutants with a critical nucleobase or a highly important atom altered, which are then produced by transcription or chemical synthesis. The resulting mutants are usually much less reactive than wild-type RNA, as important tertiary interactions are disrupted by introduced mutations.

In a two-part system for probing tertiary contacts along an interaction interface between RNA molecules (Figure 17.1b), one of the RNA molecules (RNA 1) is typically mutated. The other RNA molecule (RNA 2) is transcribed in the presence of nucleotide analogs. The latter RNA is then screened for changes in its interference pattern on interaction/reaction with the mutant *versus* the wild-type version of RNA 1. Assume, for example, that the 2'-OH group of a nucleotide X in RNA 1 interacts with the exocyclic amine of the guanosine residue at position Y in RNA 2 (Figure 17.1b), and that this interaction is critical for a selectable function. On reaction with wild-type RNA 1, RNA 2 will exhibit strong interference with an inosine and/or 2-methyl guanosine analog at position Y. However, when RNA 2 reacts with a mutant RNA 1 (under conditions altered to result in the same extent of reaction as with wild-type RNA 1) that contains a single 2'-deoxynucleotide at position X, the interaction between nucleotides X and Y is already disrupted and the penalty for this disruption has been paid in the form of lesser reactivity of mutant RNA 1. Therefore, when RNA 2 interacts with the specified mutant RNA 1, there will be no interference observed at position Y (Figure 17.1b). This effect is generally referred to as "*interference suppression*." Note that this is not a "rescue" because the disappearance of the interference effect is not attributable to a restored interaction, but rather to elimination of an interaction partner.

17.1.2.2 Applications: Elucidating Tertiary Contacts in Group I and Group II Ribozymes

NAIS as a method to identify long-range tertiary contacts in large ribozymes was first demonstrated during studies on group I intron ribozymes. There, it was used

to predict an extended minor groove triple helix between the P1 helix and the single-stranded J8/7 region, in which each triple appeared to be mediated by at least one 2'-hydroxyl group [19]. For group II intron ribozymes, NAIS identified three tertiary interactions, $\kappa-\kappa'$, $\lambda-\lambda'$, and $\mu-\mu'$, which are essential for active-site architecture in the group II intron core, as well as the interaction between the $\varepsilon-\varepsilon'$ duplex and an internal loop of domain 3 [6, 18, 20, 24]. Hydrogen-bonding contacts elucidated by NAIS have been extensively used as constraints for building models of group I and II ribozyme active sites [19, 25].

In Sections 17.2 and 17.3, we illustrate applications of NAIM and NAIS, respectively, by using three diverse systems as examples: (i) studies on catalysis by a group II intron ribozyme, (ii) RNA folding studies on a group II ribozyme construct, and (iii) studies of RNA-protein interactions that are involved in transcription termination. Although the provided protocols detail each step of the NAIM/NAIS procedure as it applies to these systems, they can be easily tailored to meet the specific needs of the reader. To avoid repetition, Section 17.2 begins with the basic protocols and methods that are used for all applications of NAIM. These are followed by specific examples, which differ primarily in the selection method used for NAIM. NAIM protocols are then followed by NAIS methodologies that were developed to study group II intron tertiary structure (Section 17.3).

17.2

Experimental Protocols for NAIM

17.2.1

Nucleoside Analog Thiotriphosphates

Many phosphorothioate nucleotide analogs are commercially available from Glen Research in a solution form (see below for details). Unmodified nucleotide phosphorothioates as well as 2'-deoxyribonucleotide phosphorothioates are also available from TriLink Biotechnologies as a dry powder. For European suppliers, see Chapter 18. Strobel and colleagues [21, 23] have expanded the existing collection by developing strategies and protocols for the efficient synthesis of other analogs (Figure 17.2). Most analogs can be synthesized from unprotected nucleosides (generally available from Sigma, R.I. Chemical; Orange, CA, USA) or ChemGenes (Wilmington, MA, USA) in two steps by adapting the procedure first described by Arabshahi and Frey [26]. The first step involves reaction of a nucleoside with thiophosphoryl chloride, resulting in the formation of 1,1-dichloro-1-thionucleoside phosphate, which is then converted to a nucleoside α -thiotriphosphate by reaction with tributylammonium pyrophosphate [23]. The synthetic procedures described above are not suitable for some cytidine analogs, which require the use of a salicyl phosphoramidite approach [27]. The synthesis of these analogs is described in detail by Oyeler and Strobel [28].

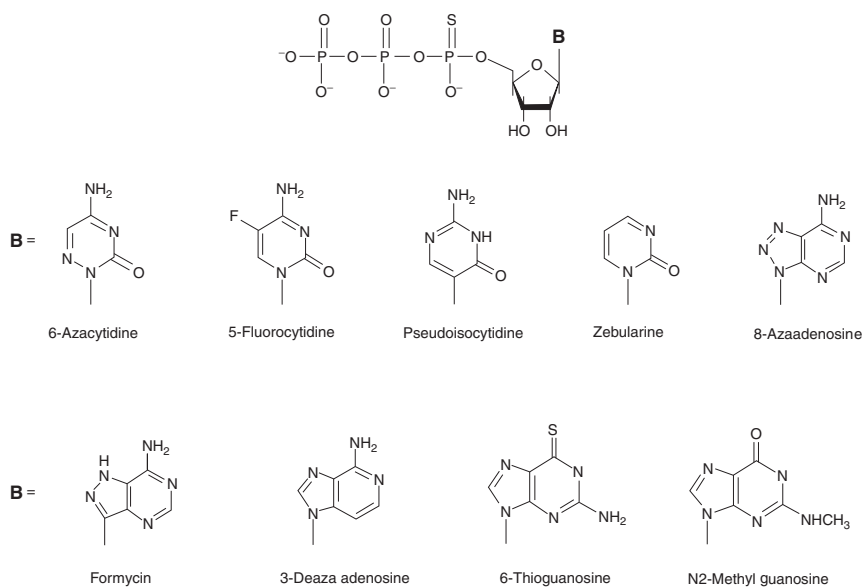


Figure 17.2 Nucleotide analog α -thiotriphosphates synthesized by the Strobel group. These analogs are not commercially available, but can be synthesized using published protocols [8, 28, 29]. Commercially available analogs can be found in the Glen Research catalog (www.glenresearch.com).

17.2.2

Preparation of Transcripts Containing Phosphorothioate Analogs

RNA molecules that contain randomly incorporated phosphorothioate analogs are prepared by *in vitro* transcription from a double-stranded DNA template (usually either a polymerase chain reaction (PCR) product or a linearized plasmid DNA). Transcription efficiency and fidelity tend to be low, however, unless the DNA template contains blunt or 5' overhanging ends. Thus, restriction enzymes must be chosen accordingly [30].

To optimize transcription efficiency, only one phosphorothioate NTP analog at a time is usually added to the transcription reaction. Since, in many cases, nucleotide analog α -thiotriphosphates are not incorporated into the transcript with the same efficiency as unmodified triphosphates, the ratio of NTP α S to NTP needs to be optimized in order to achieve the desired level of analog incorporation (generally 5%). For commercially available analogs sold by Glen Research, such an optimization has already been performed and they are sold as 10 x solutions at a concentration that is adjusted for 5% incorporation. For example, while conventional NTPs are added to the transcription at 1 mM final concentration, the modified analogs I α S and m⁶A α S are often added at a concentration of 0.4 mM. An optimal NTP/NTP α S ratio for many nucleotide analogs has been optimized and reported elsewhere (Table 17.1) [21, 23]. Wild-type T7 RNA polymerase is used for the incorporation of most analogs, except for those containing minor groove substitutions. The N2-methyl guanosine, 3-deaza adenosine, and sugar modifications (2'-deoxy, 2'-fluoro, 2'-O-methyl) are usually incorporated using mutant Y639F polymerase [31], which has an enhanced tolerance for minor groove modifications [22].

Table 17.1 Concentrations (in mM) of most common NTP α S analogs (and parental rNTP) required for ~5% N α S incorporations per transcript^a.

Incorporation	[NTP α S]	[rNTP]	RNAP
A α S, C α S, G α S, U α S (Sp-isomer only)	0.05	1	WT
dA α S, dC α S	1.5	1	Y639F
dG α S, dU α S	0.5	1	Y639F
I α S, N6MeA α S	0.4	1	WT
7-Deaza-A α S, 7-deaza-G α S	0.1	1	WT
Pur α S, 2AP α S	2	0.5	WT
DAP α S	0.025	1	WT
m ² G α S	1.5	0.5	Y639F

Unless otherwise specified, these concentrations of NTP α S analogs are for mixtures of Rp and Sp diastereoisomers.

^aFurther details on *in vitro* transcription with the Y639F mutant or wild-type T7 RNAP are described in the text.

Adapted from Ref. [23].

For most constructs employed in ribozyme studies, the optimal yield of RNA can be obtained using the following transcription protocol.

- 1) Prepare a reaction mixture containing 40 mM Tris-HCl, pH 7.5, 15 mM MgCl₂, 2 mM Spermidine, 5 mM dithiothreitol (DTT), 1 mM NTP (each), appropriate concentration of NTPαS depending on the analog (use Table 17.1 as a guide, determine experimentally or use 10 x NTPαS solution from Glen Research, which yields 5% incorporation of a respective NTPαS), 0.02 μg μl⁻¹ DNA template and 0.08 μg μl⁻¹ of wild-type T7 RNA polymerase or its mutant Y639F. In our hands, 2 ml transcription of exD123 or D56 was sufficient to carry out several NAIM/NAIS experiments. The same scale was used for folding studies on the D135 ribozyme.
- 2) Incubate the transcription mixture at 37 °C for 3–4 h.
- 3) Precipitate the transcript by the addition of 0.25 M NaCl, 0.02 M EDTA, and 3 volumes of ethanol followed by incubation at –80 °C for 1 h (one could pause here and leave it overnight) and centrifugation.
- 4) Redissolve the pellet in 10 mM 3-(N-morpholino)propanesulfonic acid (MOPS) (pH 6.0), 1 mM EDTA (M₁₀E₁ buffer) that is supplemented with an equal volume of denaturing loading buffer (8 M urea, 0.05% each of xylene cyanol and bromophenol blue, 17% sucrose, 83 mM Tris, pH 7.5, and 1.7 mM EDTA) and purify on a denaturing polyacrylamide gel. Visualize the band corresponding to the desired transcript by UV shadowing (for a detailed description, see Chapter 3), excise from the gel and elute in M₁₀E₁ buffer at 4 °C for 1–3 h.
- 5) Precipitate the RNA as above; redissolve the pellet in M₁₀E₁ buffer and store at –80 °C.

17.2.2.1 Tips and Troubleshooting

For the best yield of RNA, the transcription conditions always need to be optimized for each particular construct. One can vary, for example, the reaction time, the concentrations of DNA template, T7 RNA polymerase, MgCl₂ in the reaction buffer, and triphosphates. The addition of 0.1% Triton X-100 (final concentration), extra DTT, or RNase inhibitor can also help to increase the transcription efficiency. If it is necessary to optimize the incorporation level for a nucleotide analog, the transcription is performed using different NTP to NTPαS ratios. Raising the NTP concentration in the transcription mixture to 3–5 mM substantially increases the transcription yield, but this would also require a substantial increase of NTPαS concentration in order to maintain the desired incorporation level for the analog.

Note that it is practically impossible to separate a long (>200 nt) RNA transcript on a preparative 5% denaturing gel from any RNA molecules, which are 1 or 2 nt longer or shorter than the transcript of interest. Therefore, if a large transcript needs to be 5'-end-labeled for NAIM analysis, it is important that its 5' end is homogeneous; otherwise, multiple bands corresponding to the same phosphorothioate position will appear on the gel, thus making a NAIM pattern unreadable. In order to ensure homogeneity of the 5' end, the DNA template (if possible) should not contain more

than two guanosine residues after the T7 promoter (Chapter 1). We found that the starting sequence GGAG works well for the ai5 γ intron constructs.

The T7 RNA polymerase frequently adds untemplated nucleotides to the 3' end of an RNA transcript [32]. If such a transcript needs to be 3'-end-labeled for NAIM analysis, homogeneity of its 3' end can be achieved by either tethering a hammerhead or an hepatitis delta virus (HDV) ribozyme to the 3' end of the transcript (Chapter 2), or by using a PCR-generated DNA template containing a 2'-*O*-methyl nucleotide at the penultimate position (or two 2'-*O*-methyl nucleotides, one at the terminus and one at the adjacent penultimate position) [33–35] (Chapter 3). Alternatively, an RNA transcript can be selectively 3'-end-labeled *via* templated addition of a single α -³²P-labeled dNTP residue by Klenow polymerase (see below) [36].

17.2.3

Radioactive Labeling of the RNA Pool

In order to employ the NAIM assay, one component of the reaction must be radioactively labeled. This can be achieved by either labeling the entire RNA pool after transcription and purification, or by reaction of the unlabeled RNA pool with a radioactively labeled molecule during the selection step. For example, the selection step for a group II intron ribozyme is performed by reacting a labeled pool of one RNA with an unlabeled pool of a second RNA.

Radiolabeling of an RNA transcript can be performed either at the 5' end (kinase reaction) or at the 3' end using one of the following three methods: addition of [5'-³²P]pCp by T4 RNA ligase, templated addition of a single α -³²P-labeled dNTP residue by Klenow polymerase or addition of [α -³²P]Cordycepin by the Poly(A) polymerase [36, 37]. Note that in order to be suitable for NAIM, the first two methods of 3'-end labeling require that the 3' end of the transcript is homogeneous (see Chapter 2 for details).

When NAIM is used to study large RNAs, for example, group I and group II ribozymes, the respective transcripts are usually 5'-end-labeled to determine interference or enhancement effects in the 5' half of the molecule, and 3'-end-labeled to analyze its 3' half. This approach provides the necessary band resolution to screen the entire molecule.

The protocol for the 5'-end labeling of an RNA transcript is provided below.

- 10 x Annealing buffer:
 - 140 mM Tris-HCl, pH 7.5
 - 400 mM NaCl
 - 2 mM EDTA
 - 20 x MgCl₂/DTT mix:
 - 140 mM MgCl₂
 - 20 mM DTT
- 1) Before 5'-end labeling, dephosphorylate the transcribed RNA (~60 pmol) with calf intestinal alkaline phosphatase (CIAP, 40 U) in a final volume of

50 μl using the reaction buffer supplied by the manufacturer (Roche Applied Science). Incubate the reaction mixture at 37 °C for 60 min, then phenol-extract and ethanol-precipitate the RNA.

- 2) For 5'-end labeling, incubate 10 pmol of dephosphorylated RNA with 1 μl of [γ - ^{32}P] ATP (6000 Ci mmol^{-1} , 150 $\mu\text{Ci } \mu\text{l}^{-1}$) and 10 U of T4 polynucleotide kinase (PNK) in a final volume of 10 μl at 37 °C for 1 h.
- 3) Purify the labeled RNA on a denaturing polyacrylamide gel.

Templated addition of a single [α - ^{32}P]dCTP or [α - ^{32}P]dATP is the best method for 3'-end labeling of RNA transcripts with heterogeneous 3' ends because it permits one to selectively label only RNA transcripts of correct length. The protocol is provided below.

- 1) Design a 12–13 nt DNA splint, which is complementary to the 3' end of the RNA except for two 5'-terminal DNA residues. The penultimate of the two overhanging nucleotides (T or G) serves as the template residue for Klenow-catalyzed incorporation of [α - ^{32}P]dCTP or [α - ^{32}P]dATP at the RNA 3' end (step 4). The outermost 5'-terminal DNA residue increases the efficiency of this single nucleotide addition [36]. The splint needs to be desalted or gel-purified before use in the 3'-end labeling reaction.
- 2) In a total volume of 10 μl , combine 30–40 pmol of the RNA with 200 pmol of the DNA splint and 1 μl of 10 x annealing buffer (see above).
- 3) Denature by heating at 95 °C for 30 s and anneal by slowly cooling down to 42 °C.
- 4) Add 1 μl of 20 x MgCl_2 /DTT mix (see above), 5 μl of H_2O , 2 μl of [α - ^{32}P]dCTP or [α - ^{32}P]dATP (3000 Ci mmol^{-1} , 20 $\mu\text{Ci } \mu\text{l}^{-1}$) and 2 μl of Klenow Fragment (3'→5' exo-; 5000 U ml^{-1} , New England Biolabs (NEB) # M0212L). Incubate at 37 °C for 2 h.
- 5) Purify on a denaturing polyacrylamide gel (Section 17.2.2).

For RNA transcripts with homogeneous 3' termini, one could use pCp or cordycepin labeling because these methods are generally more efficient than dCTP or dATP labeling by Klenow polymerase.

For pCp labeling, the following protocol may be used.

- 1) In 20 μl of total reaction volume, combine 40 pmol of RNA, 2 μl of 10 x buffer for T4 RNA ligase provided by the manufacturer, 2 μl of dimethyl sulfoxide (DMSO), 10 μl of [$5'$ - ^{32}P]pCp (3000 Ci mmol^{-1} , 10 $\mu\text{Ci } \mu\text{l}^{-1}$), and 1 μl of T4 RNA ligase 1 (NEB, 20 000 U ml^{-1}).
- 2) Incubate at 4 °C for 12–16 h.
- 3) Purify on a denaturing polyacrylamide gel.

Below is the protocol for 3'-end labeling using cordycepin and Poly(A) polymerase:

- 10 x Buffer for Cordycepin labeling:
 - 100 mM MgCl_2
 - 500 mM Tris-HCl, pH 8.0

- 1 M NaCl
 - 10 mM DTT
 - 10 mM EDTA
- 1) In 10 μ l of total reaction volume, combine 1–5 pmol of RNA, 1 μ l of 25 mM MnCl_2 , 5 μ l of [α - ^{32}P]Cordycepin triphosphate (5000 Ci mmol^{-1} , 10 $\mu\text{Ci } \mu\text{l}^{-1}$), 1 μ l of 10 x reaction buffer (see above) and 1 μ l of yeast Poly(A) polymerase (Life Technologies Invitrogen, 2.5 U μl^{-1}).
 - 2) Incubate at 30 °C for 30 min.
 - 3) Purify on a denaturing polyacrylamide gel.

17.2.4

The Selection Step of NAIM: Three Applications to Studies of RNA Function

17.2.4.1 Group II Intron Ribozyme Activity: Selection through Transesterification

General Aspects Group II introns are complex ribozymes with a diverse catalytic repertoire [38–40]. Self-splicing of these ribozymes occurs in two steps, with the first step proceeding *via* competing parallel pathways: hydrolysis and branching [41, 42]. Despite the lack of primary sequence conservation, group II intron secondary structure is very conserved. It consists of six domains, each of which has a distinct catalytic function (Figure 17.3a) [38–40, 43, 44]. The largest, domain 1 (D1), folds independently [45, 46], serves as a scaffold for binding other intronic domains as well as the 5'-exonic substrate, and contributes key elements of the intron active site. Domain 5 (D5) is absolutely critical for any reaction catalyzed by the intron. Domain 3 (D3) is a catalytic effector that accelerates every reaction catalyzed by the intron. Domain 6 (D6) contains the bulged branch point adenosine, which is essential for the transesterification pathway of splicing. Domain 2 (D2) mediates conformational rearrangements along the splicing pathway and domain 4 (D4) often contains an open reading frame that encodes a protein cofactor.

One of the most remarkable features of group II intron ribozymes is their modularity, which allows reconstitution of an active ribozyme from two or more separately transcribed components. This modularity provides an opportunity to dissect the entire splicing pathway into separate reactions, which can be studied independently. One of the most informative constructs for studying group II intron catalysis is the *trans*-branching system, which divides the intron into two critical parts: exD123 (comprised of the 5' exon and intronic domains 1–3), and D56 (containing domains 5 and 6) [47] (Figure 17.3b). When these two RNA components are combined in the presence of Mg^{2+} , the 2'-OH of the D6 branch point (in the D56 construct) attacks the 5' splice site in exD123, resulting in the formation of a covalent linkage. The resulting 2'–3'–5' branched D56/D123 molecule is highly stable and can be isolated [47] (Figure 17.4). A pool of these branched molecules represents the population of species that are capable of undergoing the first step of group II intron self-splicing. Thus, *trans*-branching represents a useful selection assay for NAIM and NAIS studies on the first step of splicing. NAIM assays that



Figure 17.3 (a) Secondary structure of the ai5 γ group II intron. (b) Schematic representation of the *trans*-branching reaction used as a selection step in NAIM and NAIS studies of the ai5 γ group II ribozyme.

employ this *trans*-branching system have provided information on functionally important residues in both RNAs (exD123 and D56) [6, 15]. The obtained results were then used as a guide for subsequent NAIS studies to identify long-range tertiary contacts involving these residues [6, 15, 18, 20]. An example of a NAIM assay, which probes important functional groups in domain 3 of the ai5 γ intron [15], is shown in Figure 17.5. In this experiment, a predicted NAIM “signature” for the formation of non-canonical sheared *trans*-Hoogsteen-sugar edge base pairs

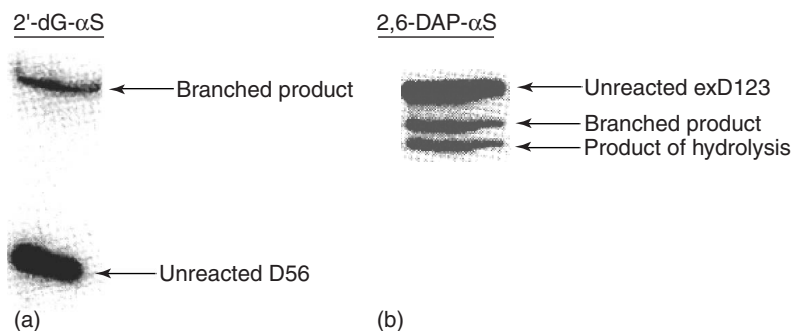


Figure 17.4 Separation of unreacted and branched products of a *trans*-branching reaction where D56 RNA is 5'-end-labeled (a) or exD123 RNA is 3'-end-labeled (b). A 2'-dGαS-modified pool of D56 RNA molecules was used in (a), and a 2,6-DAPαS-modified pool of exD123 RNA molecules was used in (b).

[48] helped determine the structure of the internal bulge, which is one of the most conserved regions within the intron.

Experimental Procedures The following protocol is used for identification of functional groups in D56 RNA that are important for branching.

- 1) Transcribe D56 in the presence of nucleotide analogs and 5'-end-label the transcript as described above.
- 2) Denature trace amounts of modified, labeled D56 (1–10 nM) and unlabeled, unmodified exD123 (1.5–3 μM) separately at 95 °C for 1 min in 40 mM MOPS pH 6.0. Cool to 42 °C by incubation for ~1 min on a benchtop and mix the two RNAs together with the simultaneous addition of salts to final concentrations of 100 mM MgCl₂, 2 mM Mn(OAc)₂, and 0.5 M (NH₄)₂SO₄. Incubate the mixture at 42 °C until the fraction of the branched product reaches ~20%. Reaction time may vary from 30 min to 2 h depending on the analog.
- 3) Separate unreacted D56 and the branched product on a 5% polyacrylamide denaturing gel (Figure 17.4a), recover both species from the gel, and subject to iodine cleavage as described below.

The experimental setup is essentially the same for mapping the 5' half of D123 in the exD123 construct. In this case, exD123 is transcribed with nucleotide analogs and D56 is unmodified. The reaction is carried out according to the above protocol and the branched product is purified on a 5% denaturing gel and subjected to sequencing by iodine. The cleavage pattern is compared to that of precursor exD123 instead of D56.

In order to map the 3' half of exD123, the procedure is modified as follows.

- 1) Transcribe exD123 in the presence of the nucleotide analogs of interest and label at the 3' end as described above (use templated addition of [α-³²P]dCTP by Klenow polymerase).

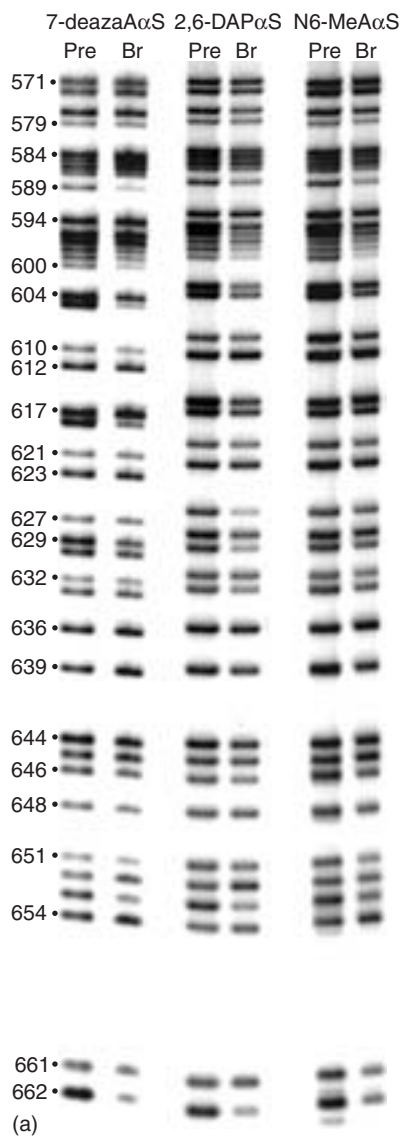
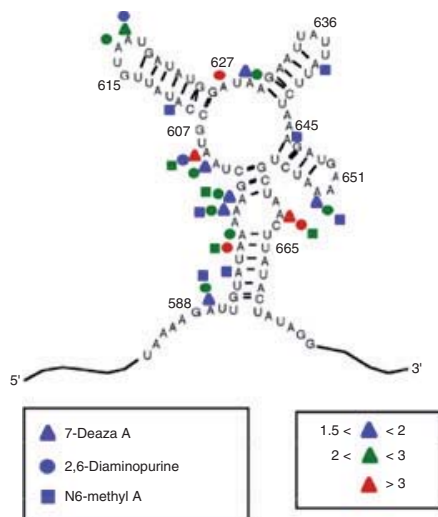
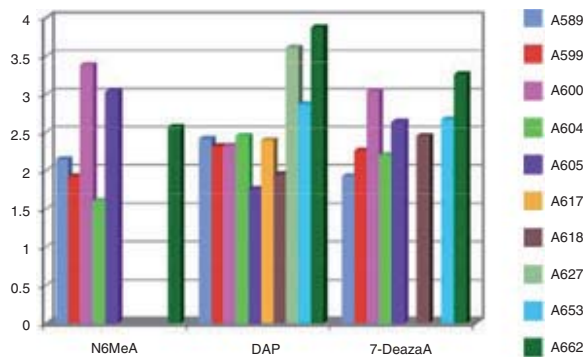


Figure 17.5 Interference effects in domain 3 of the ai5 γ group II intron. (a) A high-resolution sequencing gel summarizes the effects of 7-deazaA α S, 2,6-DAP α S, and N6-MeA α S substitutions in D3 after iodine cleavage of the unreacted (Pre) and branched (Br) fractions. (b) Summary of base analog interference effects in D3. (c) Representative average interference values

for different base analogs. (d) Predicted and observed NAIM effects for a sheared A.A pair formed by D3 internal loop residues A₅₉₉ and A₆₆₁ or A₆₆₂ and A₅₉₈ (brackets). The structure that fits the experimental data best is indicated by light gray shading. (Reprinted from Reference [15], Figures 3 and 8B therein.)



(b)

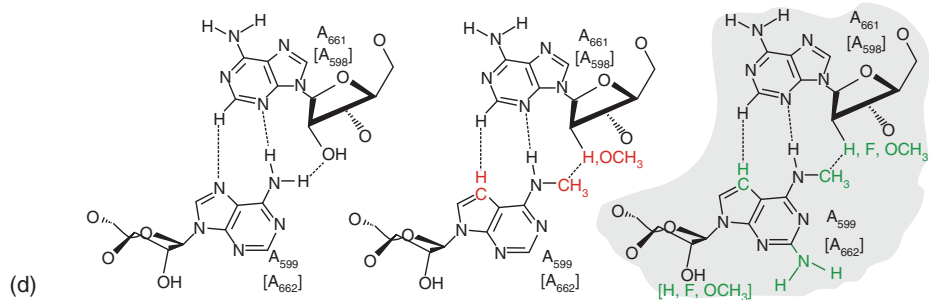


(c)

For a sheared A-A pair...

Expected interferences

Observed interferences



(d)

Figure 17.5 (continued).

- 2) React trace amounts of radiolabeled exD123 (1–10 nM) with 1.5–2 μ M unlabeled, unmodified D56 according to the same protocol as described above.
- 3) Separate the branched product from the unreacted exD123 on a 5% denaturing gel (Figure 17.4b); subject both RNAs to iodine cleavage.

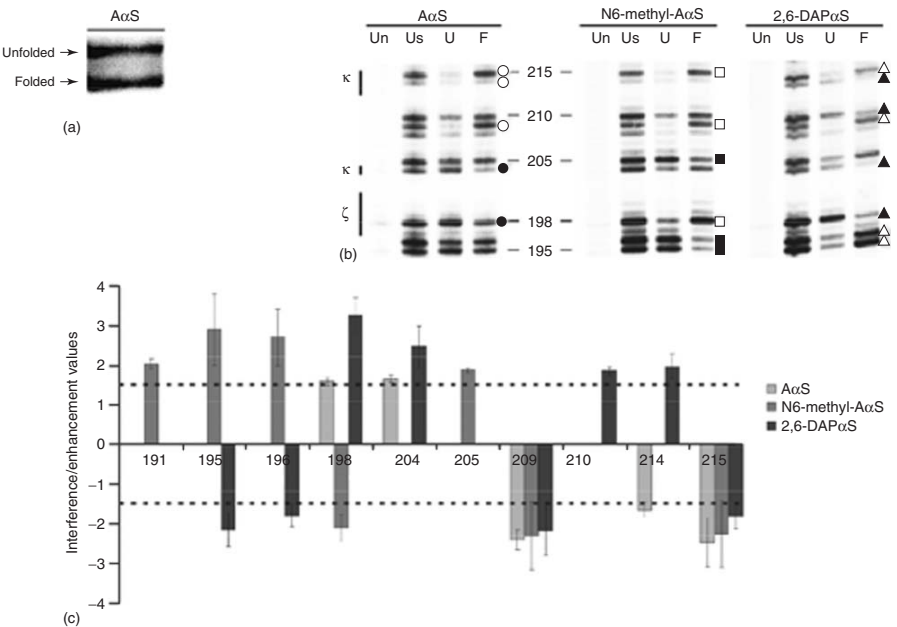
Since the 3' terminus of exD123 is labeled in this approach, one can observe the formation of both branched and hydrolysis products, both of which can be analyzed for interference effects (Figure 17.4b).

Tips and Troubleshooting To ensure the necessary sensitivity of the experiment, it is important that the branching reaction is carried out to an extent of 10–20%. We observe that RNAs transcribed in the presence of different nucleotide analogs may exhibit different reactivity; therefore, it is important to adjust reaction time for each case in order to achieve the desired extent of the reaction.

17.2.4.2 Group II Ribozyme Folding: Selection through Mg^{2+} -Induced Compaction of RNA

General Aspects A powerful model system for understanding the forces driving RNA folding *in vitro* has been a ribozyme derived from the yeast mitochondrial ai5 γ group II intron [49]. This ribozyme consists of all components known to contribute to the intron active site, that is, the scaffolding domain D1, the catalytic effector domain D3 and the active-site center D5 [50]. D135, as this ribozyme was named accordingly, is a highly reactive endonuclease capable of binding and cleaving a substrate analogous to the 5' exon *in trans* [50, 51]. As for other RNAs, an intricate network of tertiary contacts is essential for the assembly of individual intron domains into the splicing-competent structure [39, 52–54]. The ai5 γ ribozyme D135 achieves its complex architecture by following a direct path, in which formation of the scaffolding domain (D1) constitutes a compact intermediate on the way from the unfolded to the native conformation [49, 55]. Although D135 folding is devoid of kinetic traps [46, 55, 56], this RNA collapses slowly to the compact intermediate, while subsequent docking of other domains is rapid [46]. To shed light on the structural requirements and the driving forces of the Mg^{2+} -induced compaction of the D135 ribozyme, we set up a chemogenetic approach. Employing NAIM in conjunction with native gel electrophoresis (Figure 17.6), a small substructure within D1, the κ – ζ element, which displays numerous NAIM effects (Figure 17.6), was found to control D135 compaction and to initiate a cascade of folding events [7, 8]. This folding control element dictates both the Mg^{2+} -dependent stability [7] and the timescale [8] of D135 folding. During later stages of folding, this element serves as the platform for active-site assembly [6, 57]. Therefore, the group II intron appears to have evolved a sophisticated folding pathway, in which formation of an active-site docking region is rate limiting.

Optimizing the Selection Step Selection is the most fundamental step in the entire approach and can be modified to investigate a variety of aspects in RNA biochemistry (reviewed in [22]). Accordingly, this selection step defines the outcome of the NAIM



study and requires considerable optimization to ensure a successful outcome. For NAIM assays with folding as a selection step, it is important to establish the optimal concentration of metal ions, incubation time, temperature, and pH required to yield ~20–40% of folded RNA relative to the unfolded species – this ensures the necessary stringency of the selection process. For the D135 ribozyme, this was accomplished by folding it at a $[Mg^{2+}]$ below the K_{Mg} for the tertiary collapse (13 mM [46]), thereby allowing D135 to sample the collapsed intermediate state without proceeding to the native state (for which K_{Mg} is 20–40 mM [58]). To explore the structural requirements of the kinetic intermediate, the pool of modified D135 RNAs was allowed only a brief period of time to compact (approximately the $t_{1/2}$ for compaction [46], which is 15–45 s, depending on the analog).

Experimental Aspects The following experimental procedure is described below to provide an overview and guidelines on how to set up a NAIM assay to dissect RNA folding [7, 8].

- 1) *In vitro* transcribe D135 RNA in the presence of nucleotide analogs (Section 17.2.2), and after purification of the transcript 5'- or 3'-end-label the RNA (Section 17.2.3). For large RNA molecules (>250 nt), both end-labeling techniques need to be used to map the entire RNA.
- 2) Denature the end-labeled D135 RNA (1.5 pmol) in the presence of 500 mM KCl and 80 mM MOPS pH 7.0 by heating the sample at 95 °C for 1 min and cooling it for 2 min at room temperature.
- 3) To investigate the thermodynamic determinants for D135 compaction [7], add 4 mM $MgCl_2$ and 5 μ l 4 x glycerol-containing loading buffer to refold the RNA (final volume is 20 μ l). Incubate the sample at 42 °C for 10 min. Alternatively, add 100 mM $MgCl_2$ together with 5 μ l 4 x glycerol-containing loading buffer (final volume is 20 μ l) and incubate the sample at 42 °C for 15–45 s (depending on the analog) to obtain information on the kinetic intermediate.

Figure 17.6 A folding-based NAIM approach. (a) A representative preparative native gel. The unfolded and folded populations indicated here were excised and eluted from the gel. (b) Representative NAIM gels show interferences and enhancements within the folding control region encompassing the κ – ζ element. Lane 1 (Un, untreated) controls for background degradation of the labeled RNA; lane 2 (Us, unselected) shows the relative level of analog incorporation at a given position; this original RNA pool, not subjected to the selection step, was treated with iodine; lane 3 (U, unfolded) shows iodine cleavage products for the unfolded RNA fraction eluted from the native gel. Lane 4 (F, folded) shows iodine cleavage

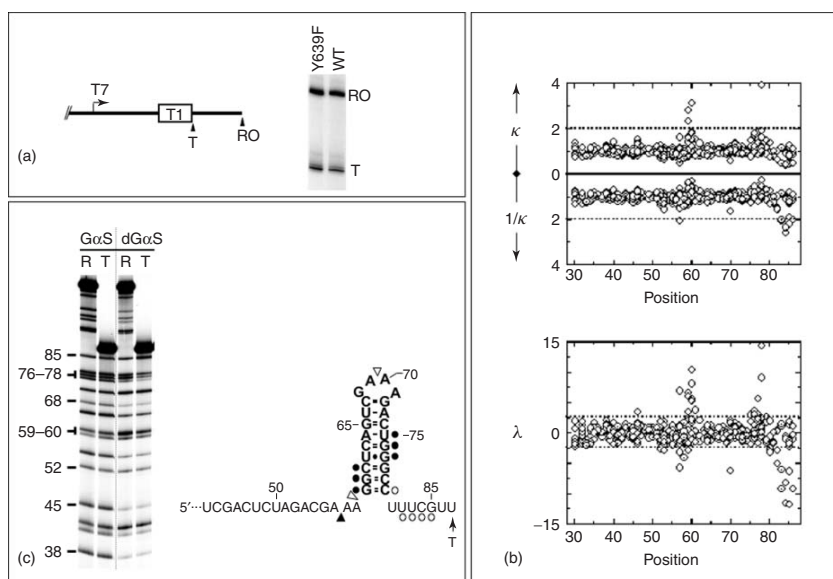
products for the corresponding folded RNA fraction purified from the native gel. Symbols are represented as follows: phosphorothioate adenosine analog (A α S, circles), N6-MeA α S (squares), 2,6-DAP α S (triangles). Closed symbols indicate interferences and open symbols represent enhancements. (c) Bar graph showing the average value and standard deviation for each raw interference or enhancement value (Section 17.2.6) found within the 5' region of the folding control element. The dotted lines indicate cutoff values; interferences in between were considered insignificant. Only interferences/enhancements exceeding the cutoff values are shown. (This figure was adapted from Ref. [7].)

- 4) To separate the unfolded from the compact population, load the sample immediately on a 4% native (non-denaturing) polyacrylamide gel containing 4 mM Mg^{2+} (0.4 mm spacer and 2 cm wide wells). Unlike 1 x TBE (Tris/Borate/EDTA), electrophoresis buffer for these gels (1 x TBE, pH 7) contains 34 mM Tris, 66 mM HEPES, 0.1 mM EDTA, and 4 mM MgCl_2 . Run the gel at 25 W for ~ 3 h in the cold room (4°C) until the xylene cyanol dye has reached the last third of the gel.
- 5) After the gel run, perform an autoradiography to detect the folded and unfolded populations (use an instant imager; alternatively, use a Phosphorimager or film exposure).
- 6) Cut out both bands (Figure 17.6a) with a sterile blade and recover each RNA species from the gel slice (Section 17.2.2).
- 7) Perform iodine cleavage as described in Section 17.2.5

Tips and Troubleshooting It may occur that the distinct RNA populations are not well resolved on the native gel. Therefore, the native gel assay has to be optimized for the RNA of interest transcribed without analog before conducting the NAIM experiment. However, it remains possible that in spite of the initial optimization the incorporation of analogs may lead to a decreased resolution of the different RNA species, such as the unfolded and folded populations. Some analogs can cause a massive smear between the bands, which may no longer be visible as defined bands. This problem can be resolved for most analogs by (i) adjusting the native gel setup (e.g., increasing Mg^{2+} concentration in the gel and running buffer, subtly changing the percentage of the native gel, decreasing the applied power, adjusting the loaded sample volume with respect to well size) and by (ii) reducing the nucleotide analog incorporation level. Nonetheless, some analogs may not be suitable for a folding-based NAIM approach. For example, the 2AP α S analog perturbed folding of the *Sc. ai5 γ* D135 ribozyme severely and could therefore not be used in our NAIM studies [7, 8].

17.2.4.3 RNA–Protein Interactions: a One-Pot Reaction for Studying Rho-Independent Transcription Termination

General Aspects During the processive phase of transcription in bacteria, the fast-moving transcription elongation complex (TEC) is held together by an intricate network of cooperative interactions between the DNA template, RNA polymerase, and the RNA product ([59] and references therein). To induce TEC dissociation, many transcription termination signals (termed *intrinsic* or *rho-independent terminators*) rely on the formation of a specific stem-loop structure within the nascent transcript that is directly upstream of a short U-rich 3' end. These RNA functional elements largely contribute to the disruption of TEC-stabilizing interactions, albeit by a mechanism that is not totally understood. To investigate the role of terminator components in the highly dynamic context of transcription elongation, NAIM experiments were applied [60]. At present, these experiments have been restricted to the study of RNA polymerases (RNAPs) that efficiently utilize NTP α S analogs,



which include certain mutants of bacteriophage RNAPs (prototypes are the Y639F and Y639F/H784A mutants of T7 RNAP [31]).

The implementation of a NAIM strategy for studying transcription termination has resulted in a simple “one-pot” reaction that combines the first two steps of NAIM (preparation of the RNA pool *and* transcript selection) [61]. In this approach, linear DNA templates containing a terminator sequence downstream from a phage promoter are transcribed with an appropriate RNAP variant (here we only consider the case of the Y639F mutant of T7 RNAP) and in the presence of a NTP α S analog. Owing to the incomplete efficiency of most termination signals, only a fraction of the TECs are released at the termination points, whereas other TECs continue transcription to the end of the template (Figure 17.7a). When both types of transcription products (i.e., the N α S-modified transcripts that are released at the terminator, and those that are released at the template end) are isolated, ^{32} P-labeled, and treated with iodine, one can identify and compare the RNA atoms and functional groups that are important for transcription termination and runoff transcription, respectively (Figure 17.7c). Because high-resolution crystal structures of the TEC now exist for various RNAPs ([59] and references therein), it is possible to link some of the N α S effects to specific known interactions between amino acid side chains of RNAP and the transcript [61].

Experimental Procedure

- 1) In a 0.5 ml tube, mix 1.1 pmol of DNA template (either linearized plasmid or purified PCR fragment) with 8 μ l of 5 x transcription buffer (30 mM MgCl₂, 50 mM NaCl, 100 mM HEPES, pH 7.5, 30 mM DTT, 0.05% Triton X-100, 10 mM spermidine), appropriate amounts of NTP α S analog and parental rNTP (Table 17.1), 20 nmol of each of the other rNTPs, and 3 pmol of Y639F RNAP. The final reaction volume should be 40 μ l, which usually yields sufficient transcript amounts for subsequent NAIM analysis.
- 2) Incubate transcription mixtures for 15–30 min at 37°C before addition of single-stranded M13 DNA (1 μ g) and KCl (250 mM) to prevent non-specific association of free RNAPs and transcripts [60].

Figure 17.7 NAIM analysis of transcription termination. (a) *In vitro* transcription of a linear DNA template (schematically depicted on the left) containing the sequence of the *rmB* T1 terminator from *Escherichia coli* (boxed) downstream from a T7 promoter (arrow). As shown by the gel (right), terminated (T) and runoff (RO) transcripts are formed in comparable amounts during transcriptions with the Y639F and wild-type (WT) T7 RNAPs. (b) Selection of NAIM interference effects on *rmB* T1 transcription termination [60] using either κ or $1/\kappa$ values (top) or λ discrimination factors (bottom).

Broken lines correspond to standard interference thresholds (see text). (c) A representative gel showing Rp-phosphorothioate (G α S) and 2'-deoxy (dG α S) interference effects at G positions during intrinsic termination of transcription at the major site (T) of the *rmB* T1 terminator; R, runoff transcripts. Positions of Rp-phosphorothioate (triangles) and 2'-deoxy (circles) modifications that favor (open symbols) or are detrimental (filled symbols) to transcript release at the T site are illustrated in the context of the terminator secondary structure.

- 3) Quickly load the mixtures onto Microcon columns (100 kDa cutoff for ~100 nt long transcripts) that have been saturated with bovine serum albumin (BSA) as described by the manufacturer (Millipore). After centrifugation of the columns for 2 min at 10 000 rpm in a desktop centrifuge, collect the filtrates, which should then be free of any unwanted high-molecular weight species [62]. Unwanted salts and nucleotides are also eliminated by gel filtration on a 1–2 ml Sephadex G-50 (Sigma-Aldrich) column.
- 4) To remove triphosphate moieties at the 5' ends of transcripts, mix the column eluate with 2 μ l of calf intestinal alkaline phosphatase (CIAP, 1 U μ l⁻¹, Roche Applied Science) and incubate for 30 min at 37 °C. The enzyme efficiently dephosphorylates transcripts under these conditions, and no additional buffer or salts are required.
- 5) Add sodium acetate to 0.3 M final concentration and extract twice with 1 volume of a phenol:chloroform:isoamyl alcohol mix (25:24:1; Amresco) and twice with 1 volume of chloroform buffered with 10 mM MOPS, pH 7. Then, add 3 volumes of ethanol and precipitate the RNA for 2 h at –20 °C.
- 6) Centrifuge for 30 min at 5000 g and 4 °C, discard the supernatant, briefly dry the pellet in a SpeedVac apparatus and redissolve in 14 μ l of M₁₀E₁.
- 7) Add 3 μ l of [γ ³²P]ATP (3000 Ci mmole⁻¹; Amersham Biosciences), 2 μ l of 10 x kinase buffer (100 mM MgCl₂, 50 mM DTT, 700 mM Tris-Cl, pH 7.5), and 1 μ l of T4 polynucleotide kinase (NEB). Incubate for 30 min at 37 °C, then phenol-extract the sample and precipitate it with ethanol as described above.
- 8) Redissolve the pellet in 10 μ l of denaturing buffer (95% formamide, 5 mM EDTA, 0.1% xylene cyanol) and incubate for 1 min at 95 °C before loading on a denaturing polyacrylamide gel (8–10% gels are adequate for ~100 nt long transcripts and should also contain 7 M urea and 30% formamide to ensure strongly denaturing conditions) that had been preheated to 60 °C. After gel migration (~2 h at 40 W for adequate separation of ~100 nt long transcripts on an 8% polyacrylamide, 20 x 40 cm gel), recover terminated and runoff transcripts from gel slices (Section 17.2.2) and then analyze through iodine sequencing as described below.

Tips and Troubleshooting Terminator stem-loop structures usually contain many G–C pairs. In some instances, complete unfolding of the structure does not occur within the polyacrylamide gel, resulting in anomalous electrophoretic migration (example in [60]) even under the harsh denaturing conditions described above. This may result in improper partitioning of the transcription products and in an inextricable mixing of N α S effects on termination and RNA structure. For similar reasons, N α S modifications that strengthen RNA base pairs (such as DAP α S) should be analyzed with great caution. In general, it is best to use terminators with no more than three to four consecutive G–C pairs in the hairpin stem such as in the *rrnB* T1 terminator of *Escherichia coli* [61]. The DNA template should also be designed to yield transcripts as short as possible; 70–100 nt should leave enough positions of neutral N α S incorporation (for calibration of NAIM effects; see Section 17.2.6) between the initiation and termination regions;

the terminator sequence should also be introduced sufficiently upstream from the template end to yield terminated and runoff transcripts with significantly different lengths. Appropriate DNA templates can be easily prepared by subcloning oligonucleotides into commercial vectors that bear a multicloning site surrounded by phage promoters such as the pSP73 plasmid (Promega).

In order to calibrate the assay, it is useful to compare NAIM signals obtained with both the Y639F and wild-type T7 RNAPs and analogs that are good substrates for the two enzymes (such as I α S or 7-deaza N α S analogs). With the *rrnB* T1 terminator, we did not observe significant differences (M. Boudvillain, unpublished observations).

Transcripts of various lengths are usually released in the termination window albeit with different efficiencies (Figure 17.7a). It is usually best to isolate and analyze those termination species separately as significant differences in NAIM patterns are likely to be observed [61].

The efficiency of transcript release by a given terminator may be affected by modifications of the experimental conditions [63]. This lack of control on the termination reaction precludes an easy adjustment of the detection threshold of NAIM effects. For this reason, the use of λ discrimination factors in place of κ interference values (Section 17.2.6) usually facilitates the identification of weak and moderate NAIM effects (Figure 17.7b; [61]). Although recent results suggest that, at least in some cases, the assay sensitivity may be controlled through variations of the RNA polymerase/DNA template ratio [64], we still prefer to rely on statistical discrimination (λ values) for detection of NAIM effects on transcription termination.

17.2.4.4 RNA–Protein Interactions: Elucidation of the Rho Helicase Activation Mechanism via Unwinding Activity

General Aspects In bacteria, transcription termination either relies on the formation of a specific stem-loop terminator motif (Section 17.2.4.3), or it involves the endogenous transcription termination factor (Rho protein). The Rho protein is a hexameric ring-shaped RNA–DNA helicase, which uses the energy of ATP hydrolysis to translocate toward the 3' end of the RNA transcript and dissociate transcription complexes consisting of the DNA template, RNA transcript, and RNA polymerase. Two recently solved crystal structures of Rho helicase [29, 65] do not fully explain the complex mechanism of RNA-dependent activation of this protein, which is believed to involve specific RNA-protein contacts formed inside the ring [66, 67]. Thus, additional biochemical experiments, such as NAIM, are necessary to understand the role of RNA–protein interactions in the function of Rho helicase. We have recently designed a series of multipiece DNA–RNA substrates to study helicase activity of the Rho protein [68] (Figure 17.8a,b). These tripartite substrates consist of a single-stranded RNA with a single Rho loading site (“anchoring” arm), and a second RNA that forms a hybrid duplex with a short complementary DNA oligo (“tracking arm”) (Figure 17.8a) [68]. These “arms” are separated by a 16-bp RNA stem formed by base-pairing of the two RNA oligonucleotides (Figure 17.8a).

These substrates were used to study the Rho helicase interaction with RNA secondary structure elements located on its translocation path by monitoring the unwinding of the RNA–DNA duplex formed downstream of such a structural obstacle [68]. We subsequently used these substrates in the NAIM assay to identify pro-Rp phosphate oxygen atoms and 2'-hydroxyl groups in the “anchoring” and “tracking” arms (Figure 17.8a) that are required for efficient unwinding of the RNA–DNA duplex by Rho helicase [69] (Figure 17.8c,d). The results of NAIM experiments indicate that 2'-deoxy interference effects are periodically repeated every ~7 nt along the RNA track. This suggests that 2'-OH-dependent Rho helicase activation events occur in a stepwise manner. Interestingly, Rho activation does not depend on the thermodynamic stability of the DNA–RNA helix, but it exhibits an RNA sequence bias that likely modulates Rho helicase activity *in vivo*.

Experimental Details The following protocol is used to identify functional groups critical for RNA-dependent Rho activation as determined by an unwinding assay.

- Hybridization buffer:
 - 150 mM KOAc
 - 20 mM HEPES, pH 7.5
 - 0.1 mM EDTA
 - Helicase buffer:
 - 20 mM HEPES, pH 7.5
 - 0.1 mM EDTA
 - 0.5 mM DTT
 - 150 mM sodium glutamate
- 1) Transcribe and purify the RNA of interest (anchoring or tracking arm) in the presence of NTP α S as described in Section 17.2.2. Dephosphorylate and 5'-end-label the resulting transcripts as described in Section 17.2.3.
 - 2) In a final volume of 20 μ l, mix 10 pmol of the labeled transcript with 11 pmol of the second unlabeled arm and 11 pmol of the DNA oligo in hybridization buffer. Denature the mixture for 2 min at 95 °C and slowly cool down to 20 °C.
 - 3) Add 5 μ l of a loading buffer for native polyacrylamide gel electrophoresis (PAGE) (15% Ficoll-400, 0.1 mM EDTA) and purify the substrate on a 7% native polyacrylamide gel. The substrate can be stored in helicase buffer (see above) at –20 °C.
 - 4) For the selection step (unwinding reaction), mix the assembled substrate (5 nM final concentration) with 4 M equiv. of Rho protein in 27 μ l of helicase buffer. Incubate at 30 °C for 3 min.
 - 5) Add 3 μ l of an ATP-MgCl₂ mix (1 mM final concentration of each component) and a DNA trap (400 nM final concentration; a DNA trap is an oligo complementary to the DNA part of the substrate) in helicase buffer. Incubate at 30 °C until ~20% of substrate is unwound.



on RNA–DNA unwinding of the substrate $R_{\text{anchor}}R_{\text{track}57B}D^{54\text{shift}}$, in which the length of the hybrid duplex is 54 bp, because the first 3 nt following the 18 nt single-stranded RNA linker remain unpaired (Figure 17.8a, bottom) [69]. (d) Summary of interference effects in the $R_{\text{anchor}}R_{\text{track}34}D^{34}$ substrate (Figure 17.8a, top) [69]. (Reprinted from Reference [69], Figures 1 and 2 therein.)

- 6) Add 4 volumes of quench buffer (27 mM EDTA, 0.7% (w/v) SDS, 4% (w/v) Ficoll-400). Separate unwinding products on a semi-native 7.5 % polyacrylamide gel containing 1 x TBE (89 mM Tris-base, 89 mM boric acid, 2 mM EDTA), and 0.5% (w/v) SDS. Isolate products from the gel as described above (Section 17.2.2)
- 7) Perform iodine cleavage and data analysis as described below (Sections 17.2.5 and 17.2.6, respectively). Since the protein may exhibit bias with respect to the sequence of the substrate, it is important to perform NAIM experiments using a series of substrates with different sequences, which would contain an A, C, G, or U at each nucleotide position of interest. These experiments would allow one to account for a potential sequence bias and generate sequence-weighted NAIM results. In order to evaluate sequence-weighted NAIM data for the RNA–DNA duplex, use the following formula:

$$\lambda_{\text{weighted}(p)} = 1/4i \times \sum_{0-i} [\lambda_A] + 1/4j \times \sum_{0-j} [\lambda_C] + 1/4m \times \sum_{0-m} [\lambda_G] + 1/4n \times \sum_{0-n} [\lambda_U]$$

$\lambda_A, \lambda_C, \lambda_G, \lambda_U$ are discrimination factors for the hybrid duplexes containing an A, C, G, or U residue at the position p ; i, j, m , and n are the number of molecules containing an A, C, G, or U residue at the position p .

Tips and Troubleshooting Interaction networks are frequently remodeled during helicase reactions. Interference effects may be caused by different factors [69], and caution as well as appropriate control experiments should be used to guard against overinterpretation of these effects. It is advisable to use NAIM only for RNA helicases, for which thermodynamic and kinetic frameworks have been well established. It is also judicious to assess the effect of the RNA track sequence and of the strength of the target RNA–DNA duplex on the NAIM patterns. This can be done by generating sequence-weighted NAIM signals from probing of a collection of RNA–DNA substrates wherein nucleotide identities have been varied systematically for all positions of interest (in the RNA–DNA helix region, for instance; [69]). If a set of different RNA–DNA sequences is analyzed, it is important to select a representation of nearest-neighbor combinations in the substrates as wide and unbiased as possible [70], while keeping sufficient sequence variability for optimal detection of NAIM signals.

17.2.5

Iodine Cleavage of RNA Pools

The preceding selection methods result in three different pools of RNA, all of which are examined by iodine cleavage in order to assay function: A, the selected pool of functional RNA molecules; B, the selected pool of non-functional RNA molecules; C, the unreacted starting pool of RNA molecules. By comparing Pool A to Pool B or Pool C (or both B and C), one can deduce the role of specific atoms on function.

17.2.5.1 Experimental Procedure

- 1) Redissolve RNAs in 10 μ l of a solution obtained by mixing 5 μ l M₁₀E₁ buffer (10 mM MOPS, pH 6.0, 1 mM EDTA) and 5 μ l of formamide, and denature at 95 °C for 1 min, followed by chilling on ice.
- 2) Add a freshly prepared solution of iodine (1 μ l of a 10 mM iodine (Sigma-Aldrich) solution dissolved in ethanol), and incubate the reaction mixture at 37 °C for 3 min.
- 3) Precipitate the RNA by adding 240 μ l of 0.3 M sodium acetate (pH 5.0), 1 μ g tRNA carrier and 750 μ l of ethanol.
- 4) Analyze samples on a denaturing polyacrylamide gel. Depending on the application and the length of the RNA transcript, the percentage of acrylamide varies from 4 to 20%. It is essential to load samples that have *not* been treated with iodine on the same gel to provide controls for non-specific RNA degradation. In many cases, photographs of NAIM gels in published articles do not show lanes of iodine-untreated samples in order to preserve space; however, it is always assumed that the researcher has performed this important control. In order to facilitate the comparison of iodine cleavage patterns in lanes corresponding to precursor RNA and selection product, it is advisable to determine the amount of radioactive material in each sample and to load equal amounts of radiolabeled RNA onto the gel. In this case, twofold or greater interference effects will be easily detectable after radioanalytic imaging, even before quantitation.

17.2.5.2 Tips and Troubleshooting

The stated iodine cleavage conditions may require optimization, depending on the RNA sample. The final concentration of iodine can vary from 0.1 to 1 mM, and often iodine cleavage reactions are carried out at room temperature. Incubation times can vary from 1 to 5 min. It is also possible to load the samples onto the gel directly without precipitation, thereby sparing additional sample handling. Frequently, however, this results in a salt front on the gel, which can adversely affect migration of the samples. Note that the salt front is sometimes caused by using old urea gel-loading buffers or old sucrose solutions used for the preparation of the urea gel-loading buffer.

17.2.6

Analysis and Interpretation of NAIM Results

17.2.6.1 Quantification of Interference Effects

Dried polyacrylamide gels are subjected to radioanalytic imaging (using, for example, a Phosphorimager (Molecular Dynamics)). For most applications, the Phosphorimager Imagequant program (Molecular Dynamics) is used to quantify the radioactive intensity of each band in precursor (pre) and product (pro) lanes for both unmodified (N α S) and modified (N' α S) nucleoside phosphorothioate-containing RNAs. After quantifying the band intensities, correcting for background

and for loading differences, NAIM effects are calculated. To determine whether an observed NAIM effect is due the phosphorothioate tag or the specific nucleotide modification, the NAIM effects for the modified nucleoside have to be compared to those for the corresponding phosphorothioate. This can be achieved in two different ways:

- 1) The interference at each position is calculated from intensities of respective bands using the following equation [9, 23]:

$$\text{Int} = \left(\frac{N\alpha S_{\text{pro}}/N'\alpha S_{\text{pro}}}{N\alpha S_{\text{pre}}/N'\alpha S_{\text{pre}}} \right)$$

This allows one to normalize the interference value for the effect of the phosphorothioate itself as well as for the incorporation differences between unmodified and modified nucleotides. After interferences have been calculated for each nucleotide position, their values are adjusted to correct for variable sample loading on the gel, as previously described [23] (Table 17.2): an average interference value is calculated for all positions that are within two standard deviations of the mean. Each interference value is then divided by this average. Adjusted interference values (usually referred to as κ values) that are higher than 2 or less than 0.5 are interpreted as interference or enhancement effects, respectively. Note that it is assumed in this case that effects of the non-bridging sulfur modification and of the secondary nucleoside modification are essentially independent of one another.

- 2) NAIM effects are calculated from the ratio of normalized counts in the precursor and product lanes for given positions. Raw interference values for each modification are presented next to that of the corresponding phosphorothioate [7]. This more conservative approach provides the following advantages. First, the effects from the non-bridging sulfur modification and the additional nucleoside modification might not be entirely independent of one another. Second, sometimes phosphorothioate enhancements may be observed, indicating that replacing oxygen with sulfur at certain positions is beneficial to the process of interest. Along the same line, modified nucleoside enhancements may also be detected. In these cases any correction for the parent phosphorothioate would bias the interpretation of the data, and it can therefore be preferable to provide the raw, uncorrected data for interference or enhancement at any given position [7, 8].

Utilization of λ Discrimination Factors When the stringency of the NAIM assay cannot be easily modulated, a statistical filter, such as the λ discrimination factor [61], may help to discriminate weak and moderate interference effects (Table 17.2, Figure 17.7b). The λ discrimination factors are derived from κ values [23] as follows: κ values deviating by more than two standard deviations from the mean of a homogeneous data set (inosine effects, for instance) are not included in a new calculation of the standard deviation (SD'). Then for every κ value of the data set,

Table 17.2 Analysis of NAIM data obtained upon selection of reactive molecules (pro) from a pool of AαS-containing transcripts (pre).

Imager counts ^a				
AαS _{pre}	AαS _{pro}	Int ^b	κ ^c	λ ^d
53385	49431	1.08	0.96	−0.47
82981	66920	1.24	1.11	1.25
69918	97108	0.68	0.61	► −7.10 ◄
62386	58855	1.06	0.95	−0.60
60916	59142	1.03	0.92	−0.99
51981	49982	1.04	0.93	−0.85
39672	20991	1.89	1.69	► 7.71 ◄
53039	44200	1.20	1.07	0.79
56126	41885	1.34	1.20	2.27
52541	53072	0.99	0.88	−1.55
16793	17313	0.97	0.87	−1.70
24755	23576	1.05	0.94	−0.72
28757	23571	1.22	1.09	1.02
18172	16225	1.12	1.00	0.00
35408	35057	1.01	0.90	−1.26
32444	29765	1.09	0.97	−0.35
39831	36210	1.10	0.98	−0.23
28660	26537	1.08	0.96	−0.47
27706	25654	1.08	0.96	−0.47
26325	21578	1.22	1.09	1.02
26404	21467	1.23	1.10	1.13
31698	26197	1.21	1.08	0.91
29349	24458	1.20	1.07	0.79
SD	—	0.21	0.18	—
Limits ^f	—	—	—	—
1 + 2*SD	—	1.42	1.36	—
1/(1 + 2*SD)	—	0.70	0.74	—
Mean ^g	—	1.12	1.00	—
SD ^{e,g}	—	—	0.09	—

^aVirtual imager counts were created to simulate a NAIM experiment. Each line corresponds to a single RNA position.

^bFor parental nucleotides such as AαS, Int values are determined with the formula:

$$\text{Int} = \text{N}\alpha\text{S}_{\text{pre}}/\text{N}\alpha\text{S}_{\text{pro}}.$$

^cThe κ values are obtained by dividing the corresponding Int values by the mean (1.12), thereby correcting for differences in loading of the gel lanes.

^dThe λ discrimination factors were determined with the formulas given in the text. Statistically significant interference effects (threshold = ±2.5) are enclosed by arrowheads. Such effects would not have been revealed by a standard analysis of the κ values (0.5 < κ < 2).

^eStandard deviations were determined with Microsoft Excel; here, values marked by gray boxes were included.

^fThe formula $L_{\text{low}} = 1 - 2 * \text{SD}$, which may also be used to calculate the lower limits, does not account for the inherently skewed distributions of Int and κ values.

^gValues marked by gray boxes were outside of the limits and thus excluded for mean and SD' calculations.

the discrimination factor is defined as

$$\lambda = (\kappa - 1)/SD' \text{ if } \kappa > 1$$

$$\lambda = (1 - 1/\kappa)/SD' \text{ if } \kappa < 1$$

Note that the sign of λ is completely arbitrary so that the above formulas can be formatted to suit an inverted reference system of “negative” and “positive” effects (as in [61]).

The use of λ factors presents several advantages. It normalizes NAIM signals for varying experimental quality and population extent among the different data sets and provides identical intensity scales for favorable ($\lambda < 0$) and detrimental ($\lambda > 0$) NAIM effects. Moreover, if one assumes that SD' adequately reflects the standard deviation of random (i.e., no interference) NAIM signals, then z tables, which are found in most statistics textbooks, may be used to select a confidence interval. For instance, an interference threshold of $|\lambda| \sim z = 2.5$ would correspond to a confidence interval for random NAIM signals of $> 98\%$ (the probability, p , that a value of $|\lambda| > 2.5$ is indicative of a random signal rather than interference would be < 0.0124). Of course, this would also depend on a number of assumptions such as signals being truly random, following a normal distribution, not being interdependent, and so on. In any case, the sample size should be sufficiently large (at least over 20 random signals for every data set) and the κ values determined with enough accuracy (in three to four independent experiments with an experimental error $< 20\%$) for the λ values to reflect statistically significant interference effects [61].

17.3

Experimental Protocols for NAIS

In this section, we describe experimental procedures specific for NAIS applied to the group II intron system. Some of the protocols used in this method, such as *in vitro* transcription, end labeling, and iodine treatment are shared with NAIM assays (Section 17.2).

17.3.1

Design and Construction of RNA Mutants

17.3.1.1 General Considerations

NAIS experiments are basically NAIM experiments on mutant and wild-type molecules that are performed in a side-by-side manner. It is commonly accepted that NAIS experiments are more meaningful when mutants contain single-atom changes. However, when studying molecules for which there is little structural information, it may be reasonable to initially perform NAIS using mutant constructs that contain full-base mutations. On the basis of the results in these coarse experiments, one can then refine the system by constructing single-atom mutants.

This approach was essential for identifying tertiary interactions in the group II intron active site (such as $\kappa-\kappa'$, $\lambda-\lambda'$, and $\mu-\mu'$) (Figure 17.9) [6, 18].

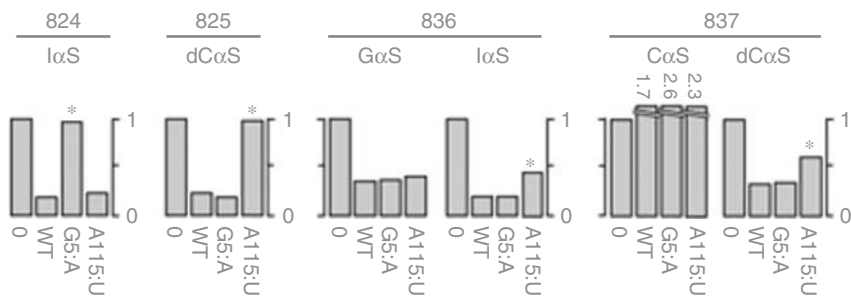
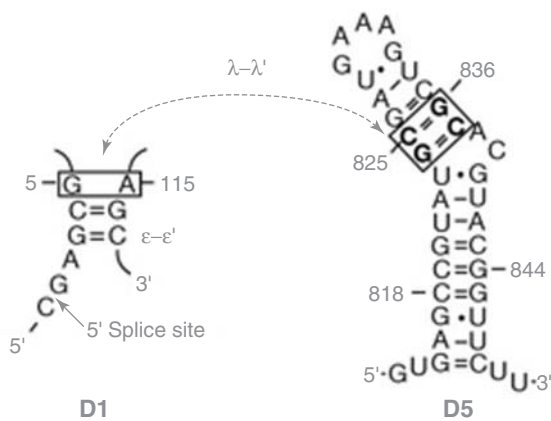
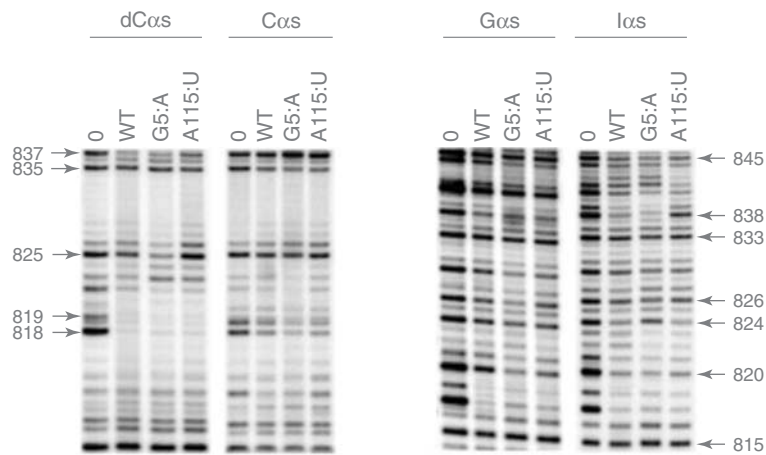
When NAIM was first performed on the ai5 γ group II intron, clusters of interference effects were observed in several regions of D1, including the very beginning of the intron (nucleotides 1–5) and an asymmetric bulge in the C1 stem (Figure 17.9) [6]. When full-base mutations were introduced in these regions at positions 5 (G5:A) and 115 (A115:U), respectively, the resulting mutants exhibited suppression of specific interferences in D5 (Figure 17.9), suggesting a complex tertiary interaction. This interaction ($\lambda-\lambda'$) was then studied at higher resolution by NAIS employing a series of D56 constructs containing single-atom changes [18]. A similar strategy was used when long-range tertiary contacts between D3 and domains 1 and 5 were identified by NAIS [20, 24]. Notably, when there is no appropriate analog available for studying certain atoms or functional groups by NAIM/NAIS (e.g., N1 of adenosine), NAIM/NAIS can be easily complemented by chemical modification interference (e.g., using dimethylsulfate (DMS) as a modifying reagent) [18].

17.3.1.2 Preparation of RNA Molecules Containing Single-Atom Substitutions

RNAs that contain single-atom substitutions can be either purchased commercially or synthesized on an automated DNA–RNA synthesizer using the standard solid-phase phosphoramidite method (Chapter 7). Although the variety of modified RNA oligonucleotides offered commercially is growing, one still has more options when synthesizing them in-house on an automated synthesizer (such as MerMade, Applied Biosystems, Pharmacia, etc.). A variety of modified phosphoramidites with single-atom or functional group modifications can be purchased commercially (from Glen Research, ChemGenes, etc.). When choosing a modified phosphoramidite for incorporation into an RNA oligonucleotide, it is important to ensure that base and/or 2'-OH protecting groups on the modified monomer are compatible with the synthetic cycle and can be easily and quantitatively removed by using standard RNA deprotection protocols. For example, if the desired modification is 2'-deoxy cytidine and the RNA oligo is to be deprotected by the Wincott procedure [71], it is advisable to use the phosphoramidite with an acetyl protecting group on the base, which is most compatible with the RNA deprotection protocol.

In our hands, the Wincott deprotection protocol is the most effective of all existing RNA deprotection procedures. Our protocol is similar but somewhat simpler than the procedure described in Chapter 7. Nevertheless, it allows one to synthesize the catalytic domain D5 of the group II intron, which is as active as the transcribed D5 RNA.

- 1) Our base-deprotection protocol is essentially the same as that described in Chapter 7, with the following exceptions.
 - a. We use 40% methylamine in water instead of 8 M ethanolic methylamine in our base-deprotection mixture. Our mixture consists of concentrated (28–30%) ammonium hydroxide : methylamine (40% in water) (1 : 1 by volume).



- b. We add 4 ml (not 2 ml) of base-deprotecting mixture to the polymer support, and incubate it at 65 °C for 10 min (not 20–40 min) with occasional stirring.
- 2) We generally synthesize trityl-off RNA, and our 2'-OH deprotection protocol also differs from the one described in Chapter 7. After base deprotection, we use the following procedure.
 - a. Separate the supernatant from the support, aliquot into six to eight Eppendorf tubes and dry in the SpeedVac.
 - b. Redissolve pellets in 250 μ l of triethylamine : triethylamine trihydrofluoride : *N*-methylpyrrolidinone mix (0.75 : 1 : 1.5) and incubate 1.5 h at 65 °C.
 - c. Add 0.5 ml of isopropoxytrimethylsilane to each tube and incubate with shaking at room temperature for 20 min to remove unreacted triethylamine trihydrofluoride [72]. This step allows further high-performance liquid chromatography (HPLC) purification of the oligo and substantially facilitates PAGE purification as well.
 - d. Precipitate the deprotected RNA by adding 1 ml of ether to each tube, followed by incubation at –20 °C overnight.
 - e. Purify the RNA on a denaturing polyacrylamide gel. In order to facilitate gel purification and prevent gel overloading with salts, one may consider desalting the oligo on a C₁₈ disposable cartridge (Sep-Pak, Waters, or OPC, Applied Biosystems).

For NAIS experiments on the group II intron system, D56 molecules containing single-atom substitutions were prepared by chemical synthesis. The D56 construct is 80 nt long, so it was synthesized in two pieces using the procedure described above. The pieces were joined by splint-directed ligation using T4 DNA ligase (see also Chapter 3), which has been particularly successful for preparing functionally active D56 molecules using the procedure described below.

Figure 17.9 Nucleotide analog interference suppressions within D5 and D1 of the ai5 γ group II intron. Autoradiographs correspond to iodine-cleaved transcripts modified with deoxy-C α S (dC α S), C α S, G α S, or I α S (inosine α S). Comparative data are shown for N α S-containing D56 RNAs that were unreacted (0) or branched to wild-type exD123 (WT), exD123_[G5:A] (G5:A), and exD123_[A115:U] (A115:U) mutants. Band intensities were quantitated and corrected for background phosphorothioate effects. The normalized intensities shown in the bar diagram at the bottom represent mean values, each having a maximum variance of 20%, based on two to four independent experiments. Asterisks indicate positions where NAIS effects were observed (loss or alleviation of

interference compared with the wild-type). For suppressions at positions 824 and 825, significant phosphorothioate-only effects were not observed, so corresponding bars are not shown. The band intensity observed for unreacted material was arbitrarily set to one. For branched products, the size of the bars is related to 1/(interference effect); that is, the smaller the bar, the stronger the interference. The values of band intensities over one are shown above the corresponding bars. In the middle of the figure, the intradomain I ϵ – ϵ' interaction (base pairs G3-C117 and C4-G116) as well as the tertiary interaction (λ – λ') with domain 5 are illustrated; nucleotides G5 and A115, which were mutated, are shown boxed on the left. (Reprinted from Reference [18], Figure 1 therein.)

- 1) Combine 5'-phosphorylated downstream RNA fragment, the upstream fragment, and a DNA splint oligonucleotide (60 nt) in equimolar ratio, denature in water at 95 °C for 1 min, and slowly cool down to 30 °C.
- 2) Supplement the mixture with 10 x buffer (500 mM Tris-HCl, 100 mM MgCl₂, 100 mM DTT, 10 mM ATP, 250 µg ml⁻¹ BSA, pH 7.5 at 25 °C; provided by the manufacturer, NEB), RNase Inhibitor (Protector RNase Inhibitor, Roche Applied Science), and T4 DNA ligase (NEB); incubate at 30 °C for 10–12 h.
- 3) Ethanol-precipitate and gel-purify the ligated RNA.

In this method, the RNA concentration, reaction volume, and amounts of RNase inhibitor and ligase must be optimized and conditions are highly dependent on the system. For the preparative ligation, we generally use 3'- and 5' fragments of the D56 construct at 20 µM each in a total reaction volume of 100 µl containing 8 µl of RNase Inhibitor (40 U µl⁻¹) and 8 µl of T4 DNA ligase (400 U µl⁻¹).

17.3.2

Functional Analysis of Mutants for NAIS Experiments

Before using mutant molecules in a NAIS experiment, it is essential to evaluate activity using the same functional assay as for the NAIS selection step. If the mutation does not affect activity, then it is unlikely to disrupt important interactions and therefore useless for NAIS analysis. By contrast, if the mutant retains only traces of activity, it becomes very difficult to harvest sufficient material for subsequent iodine treatment. In our hands, the optimal reduction in mutant activity is between 3- and 20-fold relative to the wild-type RNA. The *trans*-branching reaction was used both for selection and for analysis of D56 and exD123 mutant activity (Figure 17.3b) [6, 20, 24].

17.3.3

The Selection Step for NAIS

The selection step for NAIS studies was carried out by following essentially the same protocol as described for the NAIM experiments on this system (Section 17.2). However, the following considerations were incorporated into the experimental design.

- 1) In order to compare the interference pattern of the wild-type and mutant RNAs, the corresponding reactions must be carried out under exactly the same conditions (buffer, pH, ionic conditions).
- 2) The extent of the reaction for wild-type and mutant constructs must be the same (usually about 20% of product formation). This is generally achieved by varying the reaction time, which can be, for example, 20 min for the wild-type and 4 h for the mutant RNA. If increasing the reaction time does not result in the desired yield of the branched product, one can try raising the pH to 7 or 7.5 or the concentration of the monovalent ions to 1 M instead of 0.5 M [20, 24].

- 3) If the reaction conditions must be varied to make the mutant more reactive (e.g., the monovalent concentration is raised to 1 M instead of 0.5 M [24]), the wild-type RNA has also to be tested under these changed conditions to determine the reaction time at 20% product formation, and, most importantly, to make sure that the wild-type RNA still shows the interferences of interest.
- 4) Iodine treatment of the samples is performed according to the same protocol as described above. It is advised to have lanes of iodine-untreated reaction products for both wild-type and mutant RNAs on the gel next to the iodine-treated samples, to make sure that the band attributed to the interference suppression in the presence of the mutant is not simply a degradation product due to the prolonged incubation time.
- 5) In order to demonstrate that the tertiary interaction identified by NAIS is specific, it is helpful to carry out a reverse NAIS experiment. In this setup, the RNA that contains the single-atom modification and the pool of modified transcripts are switched. Ideally, a mutation or single-atom modification is introduced at the position where suppression was found in the previous experiment, and the other RNA is screened for interferences. If such an interference is observed, there is strong evidence that a specific interaction exists between the two RNAs. The validity of this approach has been successfully demonstrated for a group II intron system, where it was used to confirm tertiary contacts between D1 and D5 as well as D3 and D5 [18, 24]. If it is technically difficult to carry out the reverse NAIS experiment, the result of the “forward” NAIS can be confirmed by using other methods, for example, UV crosslinking [20].

17.3.4

Data Analysis and Presentation

NAIS results are analyzed and quantified essentially in the same manner as NAIM data (Section 17.2.6). The presentation of results usually includes a gel clearly showing analog interference with the wild-type RNA and suppression of this interference with the mutant RNA, and a bar graph with nucleotide analog effects for mutant molecules in comparison with the wild type [18, 19] (Figure 17.9). It is important to ensure that the mutation or a single-atom change results in a specific suppression of the analog interference at one or two positions with the rest of the interference effects remaining unchanged, and not in five or more suppression effects at various positions throughout the molecule. The latter indicates a problem with the selection step: that is, wild-type and mutant RNAs were either not treated under the same conditions or did not react to the same extent. If one mutation causes suppression of interference at more than one position, reverse NAIS (see above) may allow one to distinguish between a complex multicomponent tertiary interaction, such as λ – λ' in the group II intron [18], and an artifact or an indirect effect.

Exacerbation of interferences caused by a mutation is more difficult to interpret. The presence of a mutation sometimes makes a system more susceptible to

disturbance by modifications and can result in non-specific exacerbation of interferences with various analogs at different positions. Likewise, if a certain function is supported by a multicomponent interaction involving some redundant elements, a mutation at one of these elements can make the system more sensitive to modifications at other components of this interaction. In this case, the appearance of additional interferences compared to the wild-type RNA is selectively caused by a specific mutation. This type of specific exacerbation has also been observed during studies of the ai5 γ group II intron [73]. While wild-type RNA does not exhibit inosine interference at the θ' tetraloop receptor in domain 2 (Figure 17.3a), it is observed upon incorporation of a specific mutation in domain 3 (A627:G), suggesting a functional connection between this nucleotide and the θ – θ' tertiary contact.

Acknowledgments

O.F. thanks the Howard Hughes Medical Institute for the financial support of this work. Research on transcription termination (M.B.) has been supported in part by the Ligue Contre le Cancer (Comité du Loiret). C.W. acknowledges the financial support by the Austrian Science Foundation (FWF).

References

1. Abramovitz, D.L., Friedman, R.A., and Pyle, A.M. (1996) Catalytic role of 2'-hydroxyl groups within a group II intron active site. *Science*, **271**, 1410–1413.
2. Gish, G. and Eckstein, F. (1988) DNA and RNA sequence determination based on phosphorothioate chemistry. *Science*, **240**, 1520–1522.
3. Eckstein, F. and Gish, G. (1989) Phosphorothioates in molecular biology. *Trends Biochem. Sci.*, **14**, 97–100.
4. Strobel, S.A. and Shetty, K. (1997) Defining the chemical groups essential for *Tetrahymena* group I intron function by nucleotide analog interference mapping (NAIM). *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 2903–2908.
5. Hardt, W.-D., Erdmann, V.A., and Hartmann, R.K. (1996) Rp-deoxy-phosphorothioate modification interference experiments identify 2'-OH groups in RNase P that are crucial to tRNA binding. *RNA*, **2**, 1189–1198.
6. Boudvillain, M. and Pyle, A.M. (1998) Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.*, **17**, 7091–7104.
7. Waldsich, C. and Pyle, A.M. (2007) A folding control element for tertiary collapse of a group II intron ribozyme. *Nat. Struct. Mol. Biol.*, **14**, 37–44.
8. Waldsich, C. and Pyle, A.M. (2008) A kinetic intermediate that regulates proper folding of a group II intron RNA. *J. Mol. Biol.*, **375**, 572–580.
9. Ortoreva-Donnelly, L., Szewczak, A.A., Gutell, R.R., and Strobel, S.A. (1998) The chemical basis of adenosine conservation throughout the *Tetrahymena* ribozyme. *RNA*, **4**, 498–519.
10. Cochrane, J., Batey, R.T., and Strobel, S.A. (2003) Quantitation of free energy profiles in RNA-ligand interactions by nucleotide analog interference mapping. *RNA*, **9**, 1282–1289.
11. Ryder, S.P., Oyelere, A.K., Padilla, J.L., Klostermeier, D., Millar, D.P., and

- Strobel, S.A. (2001) Investigation of adenosine base ionization in the hairpin ribozyme by nucleotide analog interference mapping. *RNA*, **7**, 1454–1463.
12. Oyelere, A.K., Kardon, J.R., and Strobel, S.A. (2002) pK(a) perturbation in genomic Hepatitis Delta Virus ribozyme catalysis evidenced by nucleotide analogue interference mapping. *Biochemistry*, **41**, 3667–3675.
 13. Jones, F.D. and Strobel, S.A. (2003) Ionization of a critical adenosine residue in the *Neurospora* Varkud Satellite ribozyme active site. *Biochemistry*, **42**, 4265–4276.
 14. Strobel, S.A. (2002) Biochemical identification of A-minor motifs within RNA tertiary structure by interference analysis. *Biochem. Soc. Trans.*, **30**, 1126–1131.
 15. Fedorova, O. and Pyle, A.M. (2005) Linking the group II intron catalytic domains: tertiary contacts and structural features of Domain 3. *EMBO J.*, **24**, 3906–3916.
 16. Basu, S., Rambo, R.p., Strauss-Soukup, J., Cate, J.H., Ferré-D'Amaré, A.R., Strobel, S.A., and Doudna, J.A. (1998) A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat. Struct. Biol.*, **5**, 986–992.
 17. Basu, S. and Strobel, S.A. (1999) Thiophilic metal ion rescue of phosphorothioate interference within the *Tetrahymena* ribozyme P4-P6 domain. *RNA*, **5**, 1399–1407.
 18. Boudvillain, M., de Lencastre, A., and Pyle, A.M. (2000) A tertiary interaction that links active-site domains to the 5' splice site of a group II intron. *Nature*, **406**, 315–318.
 19. Szewczak, A., Ortoleva-Donnelly, L., Ryder, S.P., Moncoeur, E., and Strobel, S.A. (1999) A minor groove RNA triple helix within the catalytic core of a group I intron. *Nat. Struct. Biol.*, **5**, 1037–1042.
 20. Fedorova, O. and Pyle, A.M. (2008) A conserved element that stabilized the group II intron active site. *RNA*, **14**, 1048–1056.
 21. Cochrane, J.C. and Strobel, S.A. (2004) Probing RNA structure and function by nucleotide analog interference mapping. *Curr. Protoc. Nucleic Acid Chem.*, **6**, unit 6.9.
 22. Ryder, S.P., Ortoleva-Donnelly, L., Kosek, A.B., and Strobel, S.A. (2000) Chemical probing of RNA by nucleotide analog interference mapping. *Methods Enzymol.*, **317**, 92–109.
 23. Ryder, S.P. and Strobel, S.A. (1999) Nucleotide analog interference mapping. *Methods*, **18**, 38–50.
 24. Fedorova, O., Boudvillain, M., Kawaoka, J., and Pyle, A.M. (2005) in *Handbook of RNA Biochemistry* (eds R.K. Hartmann, A. Bindereif, A. Schoen, and E. Westhof), Wiley-VCH Verlag GmbH, Weinheim, pp. 259–293.
 25. Swisher, J., Duarte, C.M., Su, L.J., and Pyle, A.M. (2001) Visualizing the solvent-inaccessible core of a group II intron ribozyme. *EMBO J.*, **20**, 2051–2061.
 26. Arabshahi, A. and Frey, P.A. (1994) A simplified procedure for synthesizing nucleoside 1-thiotriphosphates. *Biochem. Biophys. Res. Commun.*, **204**, 150–155.
 27. Ludwig, J. and Eckstein, F. (1989) Rapid and efficient synthesis of nucleoside 5'-O-(1-thiotriphosphates), 5'-triphosphates and 2', 3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.*, **54**, 631–635.
 28. Oyelere, A. and Strobel, S.A. (2000) Biochemical detection of cytidine protonation within RNA. *J. Am. Chem. Soc.*, **122**, 10259–10267.
 29. Skordalakes, E. and Berger, T.M. (2003) Structure of the rho transcription terminator. Mechanism of mRNA recognition and helicase loading. *Cell*, **114**, 135–146.
 30. Schenborn, E. and Mierendorf, R.C. (1985) A novel transcription property of SP6 and T7 RNA polymerases: dependence on template structure. *Nucleic Acids Res.*, **13**, 6223–6236.
 31. Sousa, R. and Padilla, R. (1995) A mutant RNA polymerase as a DNA polymerase. *EMBO J.*, **14**, 4609–4621.
 32. Milligan, J., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic

- DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.
33. Kao, C., Zheng, M., and Rüdiger, S. (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3' terminus of RNAs transcribed by T7 RNA polymerase. *RNA*, **5**, 1268–1272.
 34. Kao, C., Rüdiger, S., and Zheng, M. (2001) A simple and efficient method to transcribe RNAs with reduced 3' heterogeneity. *Methods*, **23**, 201–205.
 35. Roitzsch, M. and Pyle, A.M. (2009) The linear form of a group II intron catalyzes efficient autocatalytic reverse splicing, establishing a potential for mobility. *RNA*, **15**, 473–482.
 36. Huang, Z. and Szostak, J.W. (1996) A simple method for 3'-labeling of RNA. *Nucleic Acids Res.*, **24**, 4360–4361.
 37. England, T.E. and Uhlenbeck, O.C. (1978) 3'-Terminal labelling of RNA with T4 RNA ligase. *Nature*, **275**, 560–561.
 38. Michel, F. and Ferat, J.-L. (1995) Structure and activities of group II introns. *Annu. Rev. Biochem.*, **64**, 435–461.
 39. Lehmann, K. and Schmidt, U. (2003) Group II introns: structure and catalytic versatility of large natural ribozymes. *Crit. Rev. Biochem. Mol. Biol.*, **38**, 249–303.
 40. Pyle, A.M. and Lambowitz, A.M. (2006) in *The RNA World* (eds R.F. Gesteland, T.R. Cech, and J.F. Atkins), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 469–506.
 41. Jarrell, K.A., Peebles, C.L., Dietrich, R.C., Romiti, S.L., and Perlman, P.S. (1988) Group II intron self-splicing: alternative reaction conditions yield novel products. *J. Biol. Chem.*, **263**, 3432–3439.
 42. Daniels, D., Michels, W.J., and Pyle, A.M. (1996) Two competing pathways for self-splicing by group II introns; a quantitative analysis of *in vitro* reaction rates and products. *J. Mol. Biol.*, **256**, 31–49.
 43. Michel, F., Umeson, K., and Ozeki, H. (1989) Comparative and functional anatomy of group II catalytic introns - a review. *Gene*, **82**, 5–30.
 44. Pyle, A.M. (1996) in *Nucleic Acids and Molecular Biology* (eds F. Eckstein and D.M.J. Lilley), Springer-Verlag, New York, pp. 75–107.
 45. Qin, P.Z.F. and Pyle, A.M. (1997) Stopped-flow fluorescence spectroscopy reveals that domain 1 of a group II intron is an independent folding unit. *Biochemistry*, **36**, 4718–4730.
 46. Su, L.J., Waldsich, C., and Pyle, A.M. (2005) An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic Acids Res.*, **33**, 6674–6687.
 47. Chin, K. and Pyle, A.M. (1995) Branch-point attack in group II introns is a highly reversible transesterification, providing a possible proof-reading mechanism for 5'-splice site selection. *RNA*, **1**, 391–406.
 48. Leontis, N.B., Stombaugh, J., and Westhof, E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
 49. Pyle, A.M., Fedorova, O., and Waldsich, C. (2007) Folding of group II introns: a model system for large, multidomain RNAs? *Trends Biochem. Sci.*, **32**, 138–145.
 50. Swisher, J. (2000) *Cellular, Molecular, and Biophysical Studies*, Columbia University, New York, p. 145.
 51. Fedorova, O., Su, L.J., and Pyle, A.M. (2002) Group II introns: highly specific endonucleases with modular structures and diverse catalytic functions. *Methods*, **28**, 323–335.
 52. Michels, W.J. and Pyle, A.M. (1995) Conversion of a group II intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides: elucidation of reaction mechanism and structure/function relationships. *Biochemistry*, **34**, 2965–2977.
 53. Qin, P.Z. and Pyle, A.M. (1998) The architectural organization and mechanistic function of group II intron structural elements. *Curr. Opin. Struct. Biol.*, **8**, 301–308.
 54. Michel, F., Costa, M., and Westhof, E. (2009) The ribozyme core of group II introns: a structure in want of partners. *Trends Biochem. Sci.*, **34**, 189–199.
 55. Steiner, M., Karunatilaka, K.S., Sigel, R.K., and Rueda, D. (2008)

- Single-molecule studies of group II intron ribozymes. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 13853–13858.
56. Fedorova, O., Waldsich, C., and Pyle, A.M. (2007) Group II intron folding under near-physiological conditions: collapsing to the near-native state. *J. Mol. Biol.*, **366**, 1099–1114.
 57. Costa, M. and Michel, F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.*, **14**, 1276–1285.
 58. Swisher, J., Su, L.J., Brenowitz, M., Anderson, V.E., and Pyle, A.M. (2002) Productive folding to the native state by a group II intron ribozyme. *J. Mol. Biol.*, **315**, 297–310.
 59. Murakami, K.S. and Darst, S.A. (2003) Bacterial RNA polymerases: the whole story. *Curr. Opin. Struct. Biol.*, **13**, 31–39.
 60. Kashlev, M. and Komissarova, N. (2002) Transcription termination: primary intermediates and secondary adducts. *J. Biol. Chem.*, **277**, 14501–14508.
 61. Schwartz, A., Rahmouni, A.R., and Boudvillain, M. (2003) The functional anatomy of an intrinsic transcription terminator. *EMBO J.*, **22**, 3385–3394.
 62. Gopal, V., Briebe, L.G., Guajardo, R., McAllister, W.T., and Sousa, R. (1999) Characterization of structural features important for T7 RNAP elongation complex stability reveals competing complex conformations and a role for the non-template strand in RNA displacement. *J. Mol. Biol.*, **290**, 411–431.
 63. Macdonald, L.E., Zhou, Y., and McAllister, W.T. (1993) Termination and slippage by bacteriophage T7 RNA polymerase. *J. Mol. Biol.*, **232**, 1030–1047.
 64. Epshtein, V., Toulmé, F., Rahmouni, A.R., Borukhov, S., and Nudler, E. (2003) Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J.*, **22**, 4719–4727.
 65. Skordalakes, E. and Berger, T.M. (2006) Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor. *Cell*, **127**, 553–564.
 66. Richardson, J. (1982) Activation of Rho protein ATPase requires simultaneous interaction at two kinds of nucleic acid-binding sites. *J. Biol. Chem.*, **257**, 5760–5766.
 67. Wei, R.R. and Richardson, J. (2001) Identification of an RNA binding site in the ATP binding domain of *Escherichia coli* Rho by H₂O₂-FeEDTA cleavage protection studies. *J. Biol. Chem.*, **276**, 28380–28387.
 68. Schwartz, A., Walmacq, C., Rahmouni, A.R., and Boudvillain, M. (2007) Noncanonical interactions in the management of RNA structural blocks by the transcription termination Rho helicase. *Biochemistry*, **46**, 9366–9379.
 69. Schwartz, A., Rabhi, M., Jacquinet, F., Margeat, E., Rahmouni, A.R., and Boudvillain, M. (2009) A stepwise 2'-hydroxyl activation mechanism for the bacterial transcription termination factor Rho helicase. *Nat. Struct. Mol. Biol.*, **16**, 1309–1316.
 70. Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, T., Ohmichi, M., Yoneyama, M., and Sasaki, M. (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, **34**, 11211–11216.
 71. Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., and Usman, N. (1995) Synthesis, deprotection, analysis and purification of RNA and ribozymes. *Nucleic Acids Res.*, **23**, 2677–2684.
 72. Song, Q. and Jones, R.A. (1999) Use of silyl ethers as fluoride scavengers in RNA synthesis. *Tetrahedron Lett.*, **40**, 4653–4654.
 73. Fedorova, O., Mitros, T., and Pyle, A.M. (2003) Domains 2 and 3 interact to form critical elements of the group II intron active site. *J. Mol. Biol.*, **330**, 197–209.

18

Nucleotide Analog Interference Mapping (NAIM): Application to the RNase P System

Simona Cuzic-Feltens and Roland K. Hartmann

18.1

Introduction

18.1.1

Nucleotide Analog Interference Mapping (NAIM) – the Approach

In classical mutational studies, only the base moiety of a nucleotide can be replaced with one of the three natural alternatives. Even a simple C to U transition affects more than one functional group, and exchange of the 4-amino for a keto group represents a rather radical chemical change that can have profound effects on RNA functionality. A more specific and versatile chemogenetic approach is nucleotide analog interference mapping (NAIM), which allows one to probe the functional consequences of changes as minor as single-atom substitutions in the base, sugar, or phosphate moiety. For example, in the case of a guanosine to inosine modification, the chemical alteration is restricted to deletion of the 2-amino group without additionally replacing the 6-keto with an amino group as in G to A mutations.

At the onset of NAIM studies, a pool of RNA molecules with limited numbers of randomly distributed nucleotide analogs is synthesized. The pool is then subjected to a selection procedure to separate active variants from those with impaired functions due to modification at a particular location. Subsequent comparative analysis of the distribution of modifications in the active RNA fraction and a reference fraction (e.g., the fraction of molecules with impaired function or the original unselected pool) reveals positions critical to function. The salient feature of the method is that all incorporated nucleotide analogs additionally carry a phosphorothioate modification (one non-bridging phosphate oxygen replaced with sulfur), which permits specific cleavage of the nucleic acid chain by iodine [1] exclusively at the sites of analog incorporation (for details, see Chapter 17). Iodine treatment thus results in A-, C-, G-, or U-specific sequence ladders on denaturing polyacrylamide (PAA) gels.

The partial modification of RNA is achieved by the presence of nucleoside α -thiotriphosphate analogs during *in vitro* transcription by T7 RNA polymerase (RNAP), resulting in the aforementioned pool of RNA molecules, each carrying a

low number of randomly distributed modifications. The elegance of the method lies in the capacity to simultaneously screen for the functional contribution of a particular chemical group at almost every, for example, A residue in an RNA chain. There are two major limitations: (i) residues that show a strong phosphorothioate interference effect *per se* will be insensitive to the effect of the additional modification, such as a 2'-deoxy substitution in case of 2'-deoxy NTP α S analogs; (ii) phage RNAPs do not accept all kinds of nucleotide analogs as substrates (Chapter 1) and not all analogs of interest are commercially available (Chapter 17). Chemical RNA synthesis expands the scope of possible modifications (e.g., introduction of Sp- in addition to Rp-phosphorothioates) [2], but such approaches are usually more tedious and require equipment for chemical RNA synthesis as well as special protocols in order to introduce a low level of randomly distributed modifications. Also, chemical RNA synthesis is practically limited to an RNA chain length of about 50 nt (nucleotides), thus excluding to directly screen larger RNAs, such as self-splicing introns or RNase P RNA.

The analogs available for NAIM studies can be divided into three categories [3], according to their main attribute: (i) primarily changing the chemical properties of the substituent, (ii) deleting a functional group, or (iii) introducing a bulky substituent. Depending on the type of modification introduced, NAIM experiments have the potential to reveal the following information:

- An Rp-phosphorothioate modification *per se* (AMP α S, GMP α S, CMP α S, UMP α S; abbreviated A α S, G α S, C α S, U α S, respectively, in the following) may identify crucial coordination sites for Mg²⁺ ions. Substitution of sulfur for a non-bridging phosphate oxygen essentially abolishes inner-sphere coordination to Mg²⁺, because Mg²⁺, a “hard” Lewis acid, prefers to bind oxygen, a “hard” Lewis base, relative to the much more polarizable and thus “softer” sulfur [4]. However, addition of more thiophilic metal ions (“softer” Lewis acids) such as Mn²⁺ or Cd²⁺ may restore, to varying extent, metal ion binding to the thiophosphate, leading to a (partial) rescue of the functional defect [5].
- c7-Deaza purine analogs (c7-AMP α S and c7-GMP α S) are used to reveal N7 positions involved in hydrogen binding or metal ion coordination. The latter aspect may be particularly relevant if RNA structure and function is probed in the presence of transition metal ions, such as Mn²⁺ or Zn²⁺, which form inner-sphere complexes with the N7 of purines [6, 7].
- Ribose 2'-deoxy modifications aid in probing for 2'-hydroxyls often involved in tertiary contacts as hydrogen bond donor and/or acceptor.
- Inosine α S (I α S), incorporated by T7 RNAP instead of G residues, is suited to probe the role of guanine exocyclic amino groups in hydrogen bonding. For *E. coli* RNase P RNA, relatively few inosine interference effects were detected in regular helices, suggesting that helix destabilization by this modification is of minor importance for the function of the RNase P ribozyme [8]. However, destabilization of secondary structure can become important if inosines are part of a short intermolecular hybrid helix required to bind the substrate to the ribozyme, particularly under conditions where ribozyme molecules compete for

a limited amount of substrate RNAs [9]. In such cases, combined analysis of I α S and *N*-2-methyl-G α S interference patterns was reported as a strategy to differentiate between helix destabilization and loss of important tertiary interactions as the cause of interference [3, 9]. The *N*-2-methyl group can still form a hydrogen bond with the O2 of cytosine in the Watson–Crick base pairs, but has lost its capacity to participate in bifurcated hydrogen bonding frequently observed in tertiary contacts that involve the 2-amino group of G residues [8–11].

- Analogs incorporated at A positions by T7 RNAP, such as purine, *N*-6-methyladenosine, 2-aminopurine, and 2,6-diaminopurine, which are all commercially available, probe the N6 position in terms of chemical properties and steric constraints and the tolerance for an additional 2-amino group on the minor groove edge of the base, respectively [12].

Analogs are available from Glen Research (Sterling, VA, USA; <http://www.glenresearch.com>; see their catalog under the keyword “NAIM”) through various European distributors (<http://www.glenresearch.com/Catalog/intro.html#distributors>) or from IBA (Göttingen, Germany) including several halogen-derivatized base analogs. For more information, see Chapter 17.

18.1.2

Critical Aspects of the Method

18.1.2.1 Analog Incorporation

T7 RNAP incorporates Sp-NTP α S analogs, yielding Rp-phosphorothioate-modified RNAs because of the inversion of configuration at the phosphorus atom during polymerization [13]. Sp-NTP α S analogs are incorporated with essentially the same efficiency as normal NTP substrates by T7 RNAP [14]. For phosphorothioate-tagged analogs carrying an additional modification at the base or sugar moiety, incorporation efficiency is usually lower. Also, incorporation efficiency differs if one uses the purified Sp diastereomer or a mixture of Rp and Sp diastereomers. For example, we observed more efficient incorporation of Sp-ITP α S (at G residues) than other researchers who used a mixture of both diastereomers [8, 15]. Many analogs (such as those with 2'-ribose modifications) are better accepted by the Y639F mutant T7 RNAP that shows a greater tolerance toward changes in the functional groups of the minor groove [16, 17]. A double mutation of T7 RNAP (Y639F/H784A) further increases the incorporation efficiency for modified NTPs with bulky 2'-substituents, such as 2'-OCH₃ and 2'-N₃ (Chapter 1). The reader is referred to Chapter 17 and the publication by Ryder and Strobel [16] for more details on individual analogs.

The extent of analog incorporation is adjusted to 2.5–10% (usually 5%) in NAIM studies, although incorporation efficiency may not exceed 1–2% in case of some analogs [16]. A modification extent of 5% permits good detection and quantitation of iodine hydrolysis bands, but avoids two problems associated with higher modification extents: (i) RNA inactivation due to phosphorothioate-tagged analog incorporation at multiple sites per molecule, as seen with fully (100%) AMP α S-, CMP α S-, or GMP α S-modified *E. coli* RNase P RNA [18], and (ii) an increased probability that each RNA molecule carries a modification at a site of strong interference

in addition to weakly interfering modifications; as a consequence, the sites of strong interference will, by themselves, fully determine the deficiency status of an RNA molecule irrespective of additional weaker interferences, thus masking the latter.

One should also be aware that analogs are not incorporated to the same extent at all transcript positions, the incorporation pattern being largely specific for the analog and the individual RNA under investigation. One observation is the lack or reduction of analog incorporation at homodinucleotide or homooligonucleotide stretches [16].

We usually follow a simple strategy to adjust the extent of analog incorporation, based on the roughly equal incorporation efficiency of NTP and Sp-NTPαS substrates. For example, to assess Sp-ITPαS incorporation efficiency, we transcribed RNase P RNA in the presence of 2.5, 5.0, and 10% ITPαS; in parallel, we performed transcriptions in the presence of 5% GTPαS and 95% GTP as the reference. Gel analysis of the samples after iodine hydrolysis revealed that transcripts obtained at 2.5% ITPαS and 97.5% GTP resulted in iodine hydrolysis band intensities comparable with those synthesized at 5% GTPαS.

18.1.2.2 Functional Assays

The second crucial step in NAIM studies, aside from analog incorporation, is the selection assay that partitions functional and defective RNA molecules. Examples of such assays that have been used in NAIM analyses of RNase P RNA are listed below (examples of selection assays developed for other RNA systems are described in Chapter 17).

- Separation of modified RNase P RNA pools into precursor tRNA-binding and non-binding fractions *via* adsorption of biotinylated precursor tRNA to streptavidin-agarose beads [12].
- Separation of modified RNase P RNA pools into RNase P protein-binding and non-binding fractions *via* adsorption of the His-tagged protein to Ni-NTA agarose beads [19].
- Partitioning of self-cleaving RNase P RNA ribozyme–substrate conjugates into reacted and unreacted fractions and separation by denaturing polyacrylamide gel electrophoresis (PAGE) [20–22].
- Separation of modified RNase P RNA pools into tRNA-binding and non-binding fractions by gel retardation [8, 18, 23, 24].

18.1.2.3 Factors Influencing the Outcome of NAIM Studies

Issues affecting the outcome of NAIM studies are described in the following.

- The functional aspect selected for. As an example, Boudvillain and Pyle [25] reported their *trans*-branching assay for group II intron self-splicing to be more sensitive to modification interference and thus perturbations of tertiary structure than *cis*-splicing assays. They argued that the *trans*-branching approach is so effective because the stabilization energy resulting from essential tertiary interactions has to counterbalance the entropic penalty inherent to the assembly of a two- or multicomponent system.

- Reaction conditions of the selection assay (nature and concentration of mono- and divalent cations, pH, and temperature). High salt conditions, for example, suppressed weaker interference effects in group II intron *cis*- and *trans*-splicing assays [26]. Variations in pH were shown to alleviate or exacerbate interference effects [20, 21], attributable to changes in rate-limiting steps of ribozyme-catalyzed reactions [24]. Several modifications that interfere with tRNA binding to RNase P RNA were suppressed at higher RNA concentrations [18]. Likewise, all factors that alter the dissociation constant of complexes in *trans*-binding assays (such as the temperature) will, to some extent, change the pattern and strength of interference effects.
- Inefficient analog incorporation at certain positions in the RNA chain (e.g., at homonucleotide stretches) and gel artefacts, such as band compressions at a string of G residues [8], can limit the information content of NAIM experiments. Furthermore, fluctuations in the strength of interference effects between individual experiments are considerable. The extent of such fluctuations is expected to rise with an increasing number of experimental steps involved in the NAIM procedure and difficulties to kinetically control a reaction catalyzed by an enzyme.
- In setups where an RNA binding protein is coupled to a matrix (for example, coupling of a His-tagged protein to a Ni-NTA affinity resin) [19], the proportion of correctly folded protein and its accessibility to specific RNA binding is hard to assess. Consistent data from other approaches may be required to increase confidence in the NAIM data.

18.1.3

Interpretation of Results

The Rp-phosphorothioate modification itself can cause interference effects that are due to disruption of H-bonding or inner-sphere coordination of Mg^{2+} normally involving the *pro*-Rp oxygen at this location. At most sites, the sulfur substitution has little or no effect on RNA structure; in rare cases, however, it may substantially change local structure [27] and, as a consequence, could induce global conformational changes. When NAIM is used to probe, for example, RNA–ligand interactions, a phosphorothioate modification could thus affect ligand binding even if the site of modification is at a considerable distance from the binding interface. The resulting interference effect will then be hard to interpret solely on the basis of NAIM data. Along the same lines, phosphorothioate interference effects are often found to be clustered in densely packed RNA core structures, which play a key role in RNA architecture [18, 23]. Such RNA regions are apparently very sensitive even to minor structural and chemical perturbations caused by the sulfur substitution, and thus, many interference effects detected therein are likely to reflect changes of global structure rather than direct contacts to an RNA or protein ligand.

Partial suppression of phosphorothioate interference in the presence of Mn^{2+} is often considered to indicate direct metal ion coordination to the respective phosphate oxygen [28]. The ability of Mn^{2+} , but the inability of Mg^{2+} , to coordinate

to the sulfur is made responsible for this effect. However, although Mg^{2+} and Mn^{2+} ions can occupy basically the same metal-ion-binding pockets, they may coordinate in a slightly different way, as originally observed in tRNA crystals [29]. Likewise, Mg^{2+} and Mn^{2+} interacted differently with the phosphate of A9 in two hammerhead ribozyme crystal structures [11, 30]. Nevertheless, Mn^{2+} rescue of phosphorothioate interferences in the central P4 helix of RNase P RNA [18, 20] provided the basis for detailed studies of metal ion coordination in this region using RNase P RNA constructs with single-site phosphorothioate modifications that were analyzed for kinetic defects as well as Mn^{2+} and Cd^{2+} rescue effects [31–33].

Generally, interpretation of NAIM results remains ambiguous without further information from other approaches. For example, if NAIM is performed to identify functional groups that are involved in binding a protein or another RNA, it is impossible to differentiate whether interference effects represent direct contact sites or indirect effects due to destabilization of the global RNA fold. However, additional information often enables the researcher to interpret interference data, as illustrated by two examples. For the *Tetrahymena* ribozyme, the crystal structure of the P4–P6 domain [10] revealed a hydrogen bonding interaction between the exocyclic amine of G212 in helix P4 and A184 in the A-rich bulge motif, which bridges the core helix P4 and the three-way junction of helices P5a, b, and c. NAIM identified *N*-2-methyl-G α S, but not I α S interference at G212 [9]. The authors concluded that elimination of the hydrogen bond due to inosine modification was not sufficient to significantly impair intron splicing, while introduction of the bulky *N*-2-methyl group in the P4 minor groove substantially destabilized this tightly packed region of the RNA. Another example comes from the RNase P RNA system. Comparative sequence analysis and results from biochemical studies in combination with computer-aided derivation of RNase P RNA architecture led to the proposal of the loop–helix tertiary interactions L18–P8 and L8–P4 [34, 35]. Furthermore, Easterwood and Harvey [36] proposed a model of tRNA 3'-CCA end binding to the P15 loop of *E. coli* RNase P RNA, which involves formation of two consecutive base triples. A variety of nucleotides in the corresponding structural elements of RNase P RNA showed Rp-phosphorothioate, 2'-deoxy, inosine, and/or c7-deaza interference effects in a tRNA binding assay [8, 18, 23, 24]. These interference data validated the proposed interactions and allowed us to refine their atomic details [8, 24]. Moreover, although largely consistent with the later determined X-ray structure of *Thermotoga maritima* RNase P RNA [37], the data have also been valuable for evaluating local details of the X-ray structure (data not shown).

NAIM approaches can indeed be instructive where it is unclear if a ribozyme crystal structure represents the catalytically competent conformer. A study combining point mutations and NAIM showed formation of a base triple in the core of the *Tetrahymena* ribozyme, which contributes to substrate helix docking and stabilization of active site conformation [38]. The key functional role of this base triple indicated that a crystallized form of the ribozyme [39] would require a substantial rearrangement to adopt an active conformation.

For *E. coli* RNase P RNA, results from a series of NAIM experiments led to the following conclusions.

- 1) Although largely different functional assays and assay conditions were used (self-cleavage of ribozyme–substrate conjugates at 3 M Na⁺ or NH₄⁺ and 1–25 mM Mg²⁺, gel-resolvable binding of mature tRNA to RNase P RNA in the presence of 0.1 M NH₄⁺ and 0.1 M Mg²⁺, RNase P RNA binding to immobilized precursor tRNA in the presence of 1 M NH₄⁺ and 25 mM Ca²⁺), a substantial number of identical interference effects have been observed ([24] and References therein; [22]). Some of these effects may identify residues that directly contribute to precursor and mature tRNA binding. Alternatively, modifications at these vulnerable sites may destabilize the tertiary fold, thus affecting substrate and product binding as well as catalysis more indirectly. Our estimation is that the vast majority of interference effects likely reflect perturbation of RNA tertiary structure rather than representing functional groups directly involved in catalysis or the binding of mature or precursor tRNA. The recently published X-ray structure of a bacterial RNase P holoenzyme in complex with mature tRNA and a soaked-in 5'-precursor segment [40] will be helpful in discerning direct and indirect effects.
- 2) Interference effects cluster in regions that (i) organize the conserved catalytic core of RNase P RNA (such as P4, J2/4, J3/4, J18/2, and J5/15), (ii) are tightly packed with the help of multiple metal ions (the P2–P3/P1–P4 four helix junction) [32], and (iii) organize the metal-ion-rich core of the specificity domain (J11/12, P11) [8, 12, 18, 23, 24, 41].

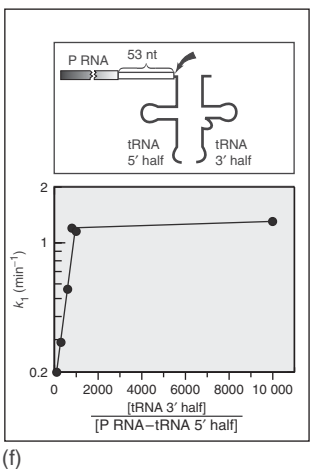
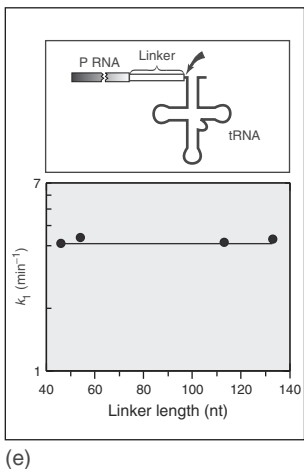
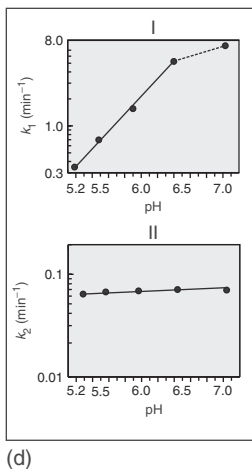
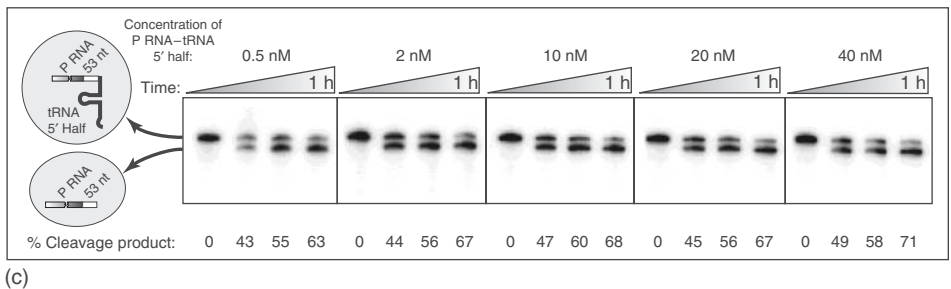
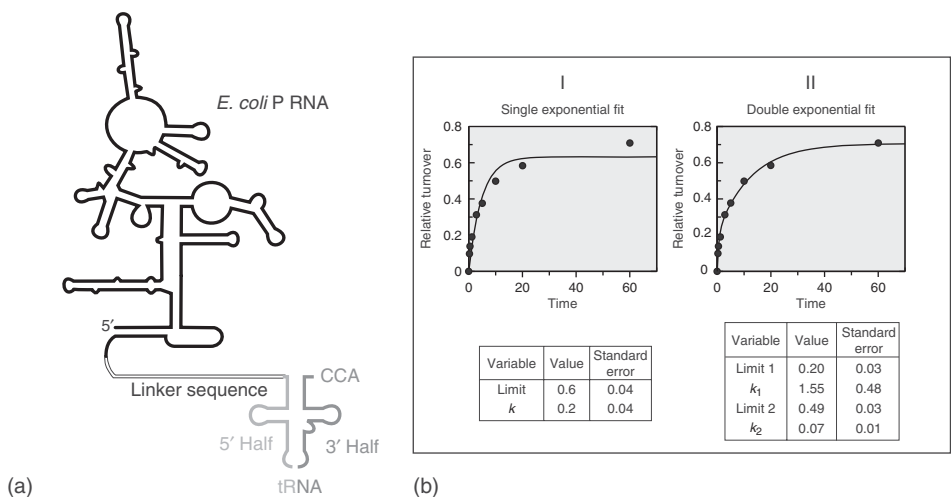
18.2

NAIM Analysis of *cis*-Cleaving RNase P RNA-tRNA Conjugates

18.2.1

Biochemical and kinetic characterization of a *cis*-Cleaving *E. coli* RNase P RNA-tRNA Conjugate

Cis-cleaving RNase P RNA (P RNA)-tRNA conjugates of the type shown in Figure 18.1a open up the perspective to identify functional groups that are crucial to cleavage chemistry, either applying NAIM or NAIS (*nucleotide analog interference suppression*, a technical extension of NAIM to identify the interaction partner of atoms or functional groups that give rise to interference effects; for details, see Chapter 17). In previous related approaches, the tRNA substrate was tethered to internal positions of *E. coli* P RNA [20–22]. However, these constructs required 3 M monovalent salt for efficient self-cleavage, and we were concerned that such conditions may suppress informative interference effects to an unwanted extent. We therefore pursued a somewhat different strategy based on a transcript consisting of *E. coli* P RNA, a linker region, and the 5' half of a bacterial tRNA^{Gly}, which is then annealed to the tRNA 3' half (added in excess over P RNA-tRNA 5' half) to reconstitute a self-cleaving P RNA-tRNA conjugate (Figure 18.1a). This bipartite system prevents self-cleavage at the tRNA 5' end already during RNA preparation and permits one to use tRNA 3' halves with single-site modifications for NAIS experiments. Before conducting NAIM experiments, it was essential to analyze



the kinetic behavior of the conjugate in order to determine the conditions under which exclusively *cis*-cleavage occurs. Our finding that the cleavage rate remained constant despite increasing P RNA-tRNA 5' half concentrations at a constant excess concentration of tRNA 3' half (Figure 18.1c) confirmed that this type of ribozyme only cleaves *in cis* and not *in trans* under dilute conditions. Exclusion of *trans*-cleavage is essential for the separation of active ribozymes from less active and inactive RNA variants. Moreover, we had to clarify what limits the reaction rate, for example, cleavage chemistry or a refolding step. The data points for the time course of the *cis*-cleavage reaction were best described by the sum of two first-order reactions (Figure 18.1b), suggesting that there are at least two populations of ribozymes reacting with different velocities: the correctly folded population (about 20%, see Limit 1 in Figure 18.1b, panel II) reacts fast, while the fraction of slower-reacting conjugates either has to change its conformation before *cis*-cleavage can occur or uses an alternative, albeit slower reaction pathway. The linear relationship for $\log k_1$ versus pH with a slope of about 1 in the range of pH 5.2–6.5 (Figure 18.1d, panel I) indicates that the chemical step is rate-limiting in the initial phase of the reaction [42]. In contrast, the cleavage rate constant k_2 describing the slow turnover (Figure 18.1d, panel II) was independent of pH, supporting the idea that k_2 reflects the rate of a slow refolding step. The rate constant k_1 was

Figure 18.1 Characterization of a model *E. coli* P RNA-tRNA conjugate. (a) Schematic representation of the construct consisting of the P RNA moiety (black), a linker sequence (thin double lines), and the tRNA 5' half (light gray); the complete tRNA substrate is reconstituted by annealing the tRNA 3' half (dark gray). (b) Self-cleavage of the conjugate shown in (a), equipped with a linker that is 53 nt long. Assays were performed as follows: the P RNA-tRNA^{Gly} complex was formed by annealing the tRNA 3' half to 5'-end-labeled P RNA-tRNA 5' half in the presence of 100 mM NH₄Cl and 5 mM CaCl₂ to avoid uncontrolled self-cleavage. The P RNA-tRNA 5' half concentration was 0.3 nM and that of the tRNA 3' half was 300 nM. The annealing mixture was heated to 95 °C for 2 min, then transferred to and cooled down in a heating block adjusted to 50 °C and preincubated for 30 min at 50 °C. After preincubation, the mixture was diluted to 0.1 nM P RNA-tRNA 5' half and 100 nM tRNA 3' half by adding NH₄Cl, urea, and CaCl₂ to a final concentration of 100, 100, and 5 mM, respectively. The reaction was started by adding MgCl₂ to a final concentration of 36 mM. Sample aliquots were withdrawn at different time points and analyzed by 8% denaturing PAGE. The time course was fit to either a single first-order (panel I) or two consecutive first-order reactions (panel II).

(c) No significant *trans*-cleavage occurred after annealing of P RNA-tRNA 5' half and tRNA 3' half under the conditions tested, since the cleavage rate was constant at different concentrations of P RNA-tRNA 5' half (0.5–40 nM in the presence of constantly 200 nM tRNA 3' half). (d) pH dependence of k_1 (measuring the rate of the fast initial phase of the reaction) and k_2 (measuring the rate of the second slower phase of the reaction; see panel II in (b)); the linear relationship of $\log k_1$ versus pH with a slope of about 1 in the range of pH 5.2–6.5 suggests that the chemical step is rate-limiting in the initial phase of the reaction. (e) Influence of the linker length on the cleavage rate constant k_1 . The *linker length* is defined as the number of nucleotides that separate the P RNA 3' end from the tRNA 5' end. No significant differences were observed in the cleavage rate constant k_1 among complexes with linker lengths between 45 and 133 nt. (f) To find the saturation limit of the tRNA 3' half, its concentration was varied up to a 10 000-fold excess over P RNA-tRNA 5' half (53 nt long linker) whose concentration was 0.1 nM; tRNA 3' half concentration exceeding that of P RNA-tRNA 5' half by more than a factor of 800 did not further increase the cleavage rate constant k_1 . For reaction conditions in (d–f), see legend of (b).

further shown to be independent of the linker length in the range of 45–130 nt (Figure 18.1e); also, tRNA 3' half concentrations exceeding that of P RNA-tRNA 5' half (0.1 nM) more than 800-fold did not further increase k_1 (Figure 18.1f).

18.2.2

Application Example

In the following section, we describe, as an application example, I α S modification experiments using the *E. coli* P RNA-tRNA conjugate with a 53 nt long linker connecting the P RNA and tRNA moieties. Usually, one compares I α S with G α S interference patterns to be able to ascribe interference effects to the thioate and/or additional base modification (Chapter 17). However, our initial intention was to determine to which extent I α S interference patterns obtained with the *cis*-cleaving conjugate overlap with those observed in a gel retardation assay selective for high-affinity tRNA binding to *E. coli* P RNA [8]. The protocol is tailored to NAIM analysis of the 3' half of P RNA. Usually, analysis from the 5' end is conducted in parallel in order to fully resolve interference effects for an RNA of the size of P RNA (about 400 nt). Corresponding analysis from the P RNA 5' end requires two major changes in the procedure outlined here and illustrated in Figure 18.2: (i) in step 1, the procedure is started exclusively with 5'-³²P-end-labeled pool RNA (instead of adding some 5'-end-labeled to predominantly unlabeled RNA) using about 15 times more radioactivity than specified in Protocol 5, and (ii) 3'-end-labeling (step 6) is omitted. In addition, the simplest strategy for this setup is to directly excise the cleaved and uncleaved fractions from the same gel lane (skipping step 4b), since 5'-end-labeled bands derived from iodine hydrolysis within the P RNA moiety will comigrate for the cleaved and uncleaved fraction in the final gel analysis (step 7).

The experimental procedure involved the following steps, illustrated in Figure 18.2.

- 1) Transcription of the P RNA-tRNA 5' half carrying a low degree (2.5%) of randomly distributed I α S modifications, 5'-³²P-end-labeling of an aliquot of the RNA pool for the detection of cleaved and uncleaved molecules on denaturing PAA gels, and transcription of tRNA 3' half (unmodified for NAIM). The concept for NAIS experiments is illustrated as well. For example, with the goal to identify functional groups in P RNA that interact with 2'-hydroxyls in the T arm, two tRNA 3' halves will be utilized: an all-ribose 3' half and a variant thereof with a single 2'-deoxyribose (indicated by a filled triangle). Interference patterns obtained with the two tRNA 3' halves will then be compared in order to identify interference suppression effects when using the tRNA 3' half carrying a single 2'-deoxy modification in the T arm. Experimental steps for NAIS will be identical to those of NAIM, except that the reaction conditions have to be adapted for the 2'-deoxy-modified 3' half to give the same extent of reaction as for the all-ribose tRNA 3' half.
- 2) Formation of the P RNA-tRNA^{Gly} complex by annealing the tRNA 3' half (8.8 μ M) to the P RNA-tRNA 5' half (91 nM, including trace amounts of 5'-end-labeled material) in the presence of 100 mM NH₄Cl and 5 mM CaCl₂ at

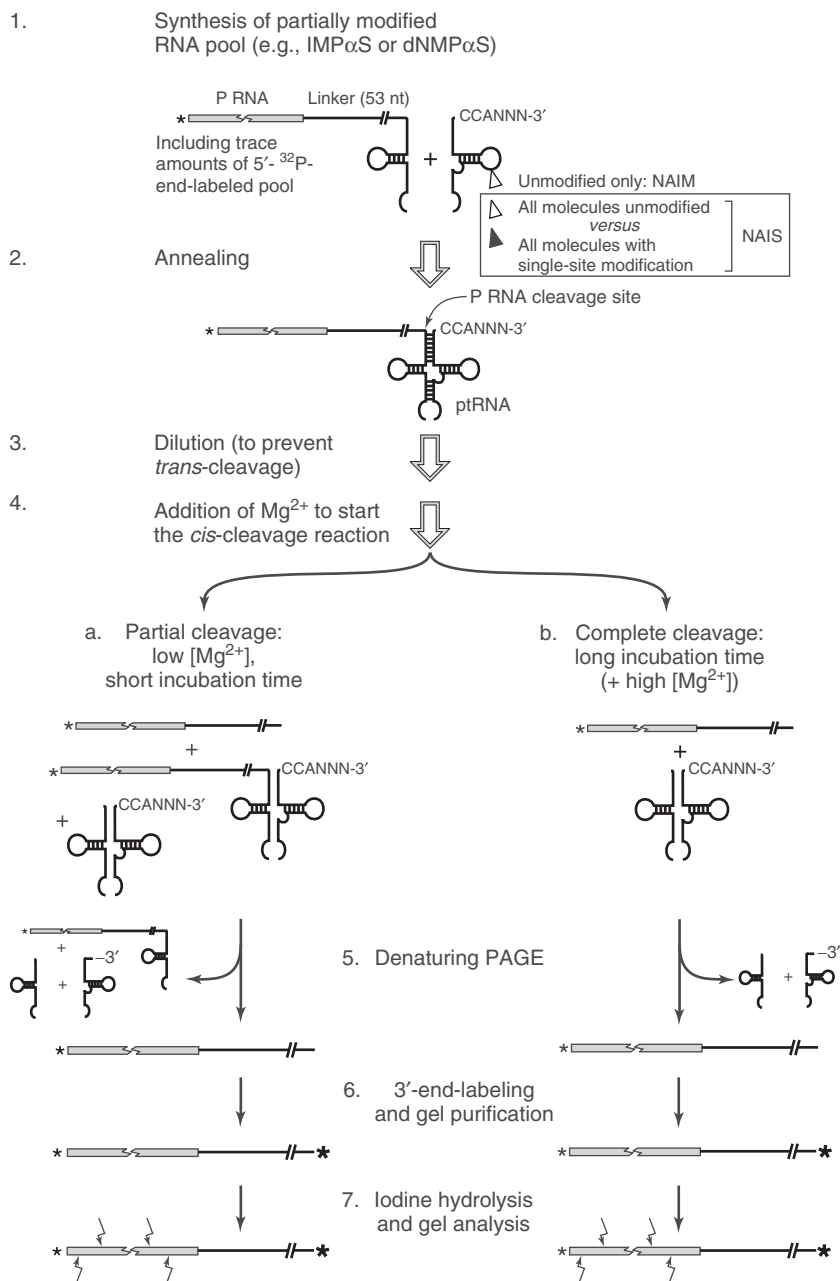


Figure 18.2 Flow scheme for NAIM (NAIS) analysis of a *cis*-cleaving *E. coli* RNase P RNA-tRNA conjugate. For details, see Section 18.2.2 and the protocols.

pH 5.9, conditions that prevent uncontrolled self-cleavage; the annealing mix is heated to 95 °C for 2 min, then cooled down to 50 °C in a heating block and preincubated for 30 min at 50 °C.

- 3) A 3.8-fold dilution of the annealed mix to 24 nM P RNA-tRNA 5' half and 2.3 μM tRNA 3' half by addition of NH₄Cl, urea, and CaCl₂ to final concentrations of 100, 100, and 5 mM, respectively.
- 4) Starting the *cis*-cleavage reaction by addition of MgCl₂ to a final concentration of 36 mM.
 - a. Stopping the reaction after 2 min at 50 °C, resulting in 20–30% product formation corresponding to the fast phase of the reaction (Figure 18.1b, panel II).
 - b. A parallel reaction was incubated for 2 h, resulting in essentially complete substrate conversion, serving as the reference RNA pool for NAIM analysis. The reason for taking this sample as the reference pool, and not the original untreated RNA pool, is that the same length species as in the short incubation is generated (P RNA plus linker). One potential drawback is that modification at some positions may entirely block *cis*-cleavage, resulting in the absence of an iodine hydrolysis band for the 2 min fraction as well as the reference pool, with the effect that strong interference effects would escape notice. A strategy to circumvent this problem is to incubate the reference pool for 2 h under *trans*-cleavage assay conditions by elevating the Mg²⁺ concentration. Yet another option is to first load the iodine-hydrolyzed starting pool (P RNA-tRNA 5' half), or possibly the uncleaved fraction from step 4a, onto the gel (step 7), let it run for some time, and then load the RNA fraction that was cleaved within 2 min to compensate for its reduced length due to the absence of the tRNA 5' half (Figure 18.2). A further elegant possibility is to cleave the starting pool with a DNase at a position that would bring the length of the precursor RNA as close to the length of the cleavage product as possible.
- 5) Denaturing PAGE and elution of 5'-end-labeled cleavage products (P RNA plus linker).
- 6) 3'-End labeling of eluted RNAs, such that the radioactivity of the 3'-³²P-label exceeds that of the 5'-label by a factor of around 100.
- 7) Iodine hydrolysis, denaturing PAGE analysis, phosphorimaging, and quantification of band intensities.

Protocols

Protocol 1: Transcription of P RNA-tRNA 5' Half Carrying Randomly Distributed IaS Modifications

- Sp-ITPaS was custom-synthesized and purified by IBA (Göttingen, Germany).

Transcripts can be initiated with the nucleoside guanosine that introduces a 5'-terminal hydroxyl group to permit direct 5'-end-labeling (Figure 18.2, step 1). Alternatively, one may perform a standard transcription initiating RNA chains

with 5'-guanosine triphosphate, followed by a dephosphorylation step with alkaline phosphatase before 5'-end-labeling, as described in Chapters 3 and 17. With the protocol outlined below, the average yield of P RNA-tRNA 5' half was in the range of 400–1000 μg (2.4–6 nmol) per 500 μl transcription mix.

- 1) Before starting to prepare the reaction mix incubate the guanosine stock solution at 75 °C in a thermoshaker until the solution becomes clear, then stop shaking, but leave the solution at 75 °C.
- 2) Prepare the reaction mix – except for guanosine and T7 RNAP – at room temperature, add the components in the order they are presented in the following table, and then prewarm to 37 °C before addition of guanosine. Add the preheated guanosine solution rapidly to the reaction mix and vortex to avoid guanosine precipitation; start the reaction by addition of enzyme.

Transcription reaction, 500 μl	Final concentration	
RNase-free water	83.5 μl	
HEPES pH 8.0, 1 M	40 μl	80 mM
DTT 100 mM	75 μl	15 mM
MgCl_2 3 M	5.5 μl	33 mM
Spermidine 100 mM	5 μl	1 mM
rNTP mix (25 mM each)	75 μl	3.75 mM (each)
Sp-ITP α S 4.33 mM	11 μl	0.095 mM
Template (linearized plasmid 3.2 kb) 0.5 $\mu\text{g } \mu\text{l}^{-1}$	40 μl	40 $\mu\text{g } \text{ml}^{-1}$
Pyrophosphatase 200 U ml^{-1}	5 μl	2 U ml^{-1}
Prewarm mixture to 37 °C, then add		
Guanosine (30 mM, kept at 75 °C)	150 μl	9 mM
T7 RNAP 200 U μl^{-1}	10 μl	4000 U ml^{-1}

- 3) Incubate for 4 h at 37 °C.
- 4) Extract RNA once with an equal volume of phenol (saturated with 10 mM Tris-HCl, 1 mM EDTA, pH 7.5–8.0) and twice with equal volumes of chloroform.
- 5) Precipitate by adding 0.2 volumes 2 M NH_4OAc (pH 7.0) and 2.7 volumes of ethanol. Mix and keep at –20 °C for at least 1 h; centrifuge at 4 °C and 16 000 g for 1 h in a desktop centrifuge.
- 6) Dissolve the pellet in 50 μl RNase-free water and 50 μl gel loading buffer (Protocol 2a).
- 7) Purify the RNA by denaturing PAGE as described in Protocol 2a.

Protocol 2a: Purification of Analog-Modified RNase P RNA-tRNA 5' Half Pools by Denaturing PAGE

- Gel loading buffer: 2.7 M urea; 1 x TBE; 67% (v/v) formamide; and 0.01% (w/v) each of bromophenol blue and xylene cyanol.

- RNA elution buffer 1: 200 mM Tris-HCl; 1 mM EDTA; 0.1% SDS; and pH 7.0 at room temperature (this buffer may also be prepared without SDS, e.g., when the eluted RNA is afterward subjected to 3'-end-labeling).
- RNA elution buffer 2: 1 M NH₄OAc and pH 7.0.

Prepare a 5% PAA (24 : 1 acrylamide: *N,N'*-methylene bisacrylamide) gel in 8 M urea and 1 x TBE, measuring 15 cm wide, 35 cm long, and 1 or 0.5 mm thick. The pocket size depends on the amount of RNA that has to be purified (about 10 cm pocket width for a 500 µl preparative transcription in case of 1 mm gel thickness).

- 1) Load the RNA sample from Protocol 1, step 6, onto the gel immediately after extensive rinsing of the pocket with a syringe to remove urea solution that has diffused from the gel matrix into the pocket; run the gel at 20–25 mA for about 3 h until the xylene cyanol has reached the bottom of the gel.
 - 2) Separate the glass plates and place the gel between two sheets of kitchen wrapping film.
 - 3) Visualize RNA band(s) by UV shadowing (for details, see Chapter 3). The exposure should be minimized to avoid UV-induced damage of the RNA. Mark the band of interest with a pen or marker and excise it with a sterile scalpel under normal light. Check correct excision by UV shadowing.
 - 4) Elute the RNA in the appropriate volume of RNA elution buffer 1 or 2 (3 ml for RNA purified from a 500 µl transcription assay) overnight at 4 °C with shaking.
 - 5) Collect the supernatant.
 - 6) For ethanol precipitation, add 0.2 volumes of 2 M NH₄OAc (pH 7.0) and 2.7 volumes of ethanol (omit NH₄OAc when using elution buffer 2). Mix and keep at –20 °C for at least 1 h and centrifuge at 4 °C and 16 000 g for 1 h in a desktop centrifuge.
 - 7) Wash the pellet with 100–200 µl ice-cold 70% ethanol, centrifuge at 4 °C and 16 000 g for 5 min; air-dry the pellet and redissolve in 200 µl RNase-free water for RNA derived from a 500 µl transcription assay (the RNA concentration should be 2–5 µg µl^{–1}).
 - 8) Measure the RNA concentration by UV spectroscopy (Chapter 3 and Appendix).
- The elution process (steps 4–8) may be repeated to recover higher amounts of RNA.

Protocol 2b: Purification of Aliquots of Analog-Modified Pool RNA after ³²P-End-Labeling

Follow Protocol 2a, with the following alterations:

- 1) Load the radiolabeled RNA (5–20 pmol) into a 0.5 (1 mm gel thickness) or 1.3 cm wide (0.5 mm gel thickness) gel pocket.
- 2) In step 3, visualize the radiolabeled RNA with a phosphorimager (instead of UV shadowing) after an image plate has been exposed to the gel for 1–20 min, depending on the amount of radioactivity loaded on the gel.
- 3) In step 4, elute the RNA in 500–1000 µl elution buffer 1 or 2.

- 4) Redissolve the RNA pellet after elution and ethanol precipitation in 10–20 μl RNase-free water.

Protocol 3: 5'-End-Labeling of Analog-Modified Pool RNA

- 1) Prepare the reaction mix by adding the components in the order as they are presented in the following table and then vortex, spin down, and incubate at 37 °C for 60–120 min.
 - 10 x T4 polynucleotide kinase (T4 PNK) buffer (forward reaction): 500 mM Tris-HCl, pH 7.6; 100 mM MgCl_2 ; 50 mM DTT; 1 mM spermidine; and 1 mM EDTA

Labeling reaction, 15 μl		Final concentration
10 x T4 PNK buffer (forward reaction)	1.5 μl	1 x
25 mM DTT	1.5 μl	2.5 mM
RNase-free water	6–7 μl	
Pool RNA purified according to Protocol 2a	1–2 μl (10–20 pmol)	0.66–1.33 μM
$[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (3000 Ci mmol^{-1} , 10 $\mu\text{Ci } \mu\text{l}^{-1}$, 3.3 μM)	3.0 μl	0.66 μM
T4 PNK, 10 U μl^{-1}	1 μl	0.66 U μl^{-1}

- 2) After incubation, add 35 μl of RNase-free water and then vortex; add 7 μl 2 M NH_4OAc and 94.5 μl ethanol for RNA precipitation; proceed as described in Protocol 2a, step 6. Wash the pellet with 100 μl ice-cold 70% ethanol, centrifuge at 4 °C and 16 000 g for 5 min and air-dry the pellet.
- 3) Redissolve the pellet in 10 μl gel loading buffer (Protocol 2a) and purify the radiolabeled RNA by 5% denaturing PAGE as described in Protocols 2a and 2b.
- 4) After gel elution and ethanol precipitation redissolve the RNA pellet in 10–20 μl RNase-free water and determine the overall yield of labeled RNA by measuring 1 μl using a scintillation counter.

Protocol 4: 3'-End-Labeling of Analog-Modified Pool RNA

- 10 x T4 RNA ligase buffer: 500 mM HEPES-NaOH, pH 8.0; 100 mM MgCl_2 ; and 100 mM DTT

3'-End-labeling reaction, 6 μl		Final concentration
Cleaved RNA after step 5 of Figure 18.2 (air-dried pellet)	5–10 pmol	0.83–1.66 μM
10 x T4 RNA ligase buffer	0.6 μl	1 x
1.5 mM ATP	0.33 μl	82.5 μM
$[\gamma\text{-}^{32}\text{P}]\text{pCp}$ (3000 Ci mmol^{-1} , 10 $\mu\text{Ci } \mu\text{l}^{-1}$)	4 μl	2.2 μM
T4 RNA ligase, 10 U μl^{-1}	1 μl	1.67 U μl^{-1}

- 1) Prepare the reaction mix by adding the components in the order given in the table; vortex, spin down, and incubate at 4 °C overnight.
- 2) Add 10 µl gel loading buffer and purify the radiolabeled RNA by 5% denaturing PAGE as described in Protocols 2a and 2b.
- 3) After gel elution and ethanol precipitation according to Protocols 2a and 2b, redissolve the RNA pellet in 10–20 µl RNase-free water and determine the overall yield of labeled RNA by measuring 1 µl using a scintillation counter.

Protocol 5: Selection for *cis*-Cleavage of P RNA-Substrate Conjugates

The analog-modified pool RNA (P RNA conjugated to tRNA 5' half, Figure 18.1a) is first annealed to the tRNA 3' half in order to reconstitute the full-length substrate, followed by dilution and concomitant addition of Mg²⁺ to start the *cis*-cleavage reaction (Figure 18.2).

- 4 x Annealing buffer: 400 mM MES-NaOH, pH 5.9 and 400 mM NH₄Cl.

- 1) Prepare the annealing master mix:

Annealing mix, 660 µl		Final concentration
4 x Annealing buffer	165 µl	1 x
30 mM CaCl ₂	110 µl	5 mM
24 µM IαS-modified P RNA-tRNA 5'-half	2.5 µl	91 nM
26 000 cpm µl ⁻¹ Radiolabeled IαS-modified P RNA-tRNA 5'-half	5 µl	197 cpm µl ⁻¹
166 µM tRNA 3'-half	35 µl	8.8 µM
RNase-free water	342.5 µl	

- 2) Distribute in 132 µl aliquots to five different tubes.
- 3) Heat to 95 °C for 2 min, then transfer to and cool down in a heating block adjusted to 50 °C and preincubate for 30 min at 50 °C.
- 4) Add to each aliquot of annealed mix the components listed in the following table; mix thoroughly and incubate for 2 min at 50 °C.

Self-cleavage reaction, 506 µl		Final concentration
Annealed mix	132 µl	0.26 x Annealing buffer
300 mM CaCl ₂	6.23 µl	5 mM
4 x Annealing buffer	93.5 µl	0.74 x Annealing buffer
5 M urea	10 µl	99 mM
RNase-free water	258.2 µl	

- 5) Start the *cis*-cleavage reaction by adding 6.1 µl 3 M MgCl₂ to each of the five mixes (final concentration 36 mM).
- 6) For four tubes, stop the reaction after 2 min (resulting in 20–30% product formation, representing the fraction of functional RNA) by placing them

- immediately on ice, followed by addition of 100 μl 2 M NH_4OAc , 2 μl of 20 $\mu\text{g } \mu\text{l}^{-1}$ glycogen as carrier, and 1.2 ml of ethanol.
- 7) Mix vigorously and store at -20°C overnight.
 - 8) Keep the fifth tube at 50°C for 2 h to allow the *cis*-cleavage reaction to proceed to quasi-completion (represents the endpoint of the reaction, where essentially all RNA molecules of the original RNA pool have been cleaved), and then prepare ethanol precipitation as in steps 6 and 7.
 - 9) Centrifuge all five samples for 30–60 min at 4°C and 16 000 g, wash the pellets each with 100 μl of 70% ethanol, briefly centrifuge and air-dry the pellets, redissolve (combine) the pellets of tubes 1–4 in 15 μl gel loading buffer, redissolve the pellet of the fifth tube separately in 15 μl gel loading buffer, and run the two samples on a 5% PAA/8 M urea gel (0.5 mm thick and pocket width 1.3 cm) as described in Protocol 2b.
 - 10) Expose an image plate to the gel for 10 min and visualize the bands using a phosphorimager. Excise the cleaved product band from the two lanes and elute each in 500–1000 μl elution buffer 2. Do not perform elution in buffer 1 because residual SDS may disturb the next step (3'-end-labeling). Ethanol-precipitate the eluted RNA and air-dry the pellets as in Protocol 2a.
 - 11) 3'-End-label the eluted RNA fractions according to Protocol 4.

Protocol 6: Iodine-Induced Hydrolysis of Analog-Modified RNA Fractions after Functional Selection

- 1) Prepare a fresh I_2 solution with reagents listed in the following table.

I_2 Solution, 50 μl		Final concentration
10 mg ml^{-1} I_2 solution in ethanol	5 μl	1 mg ml^{-1}
Ethanol	5 μl	10%
RNase-free water	40 μl	

- 2) Prepare the iodine hydrolysis reaction mix as specified in the following:

Iodine-induced hydrolysis reaction, 50 μl		Final concentration
3' (or 5')-End-labeled RNA	1–10 μl (50 000 cpm)	1000 cpm μl^{-1}
100 mM HEPES pH 7.5	5 μl	10 mM
1 mg ml^{-1} I_2 solution from step 1	5 μl	0.1 mg ml^{-1} (~ 0.4 mM)
RNase-free water	30–39 μl	

- 3) Incubate the reaction mix for 10–20 min at 37°C .
- 4) Add 150 μl RNase-free water, and ethanol-precipitate by addition of 40 μl 2 M NH_4OAc , 2 μl glycogen (20 mg ml^{-1}), and 540 μl ethanol.

- 5) Redissolve the pellet in 10 μ l gel loading buffer and apply to a 10% PAA gel containing 8 M urea (gel thickness, 0.5 mm; length, 35 cm; pocket width, 0.6–1.3 cm). Run the gel at 10 mA for 3–4 h until the distance of the xylene cyanol dye is 5 cm to the gel bottom. To also resolve longer iodine hydrolysis fragments, and thus a larger portion of the RNA molecule, the 10% denaturing PAA gel may be run for an extended period, or a lower gel percentage (e.g., 5%) may be used. However, according to our experience, longer runs or the use of 5% gels have the drawback of usually resulting in more diffuse bands.
- 6) Remove the glass plates. Place the gel between one layer of gel drying (Whatman) paper and one layer of kitchen wrapping film.
- 7) Dry the gel for 30 min at 70 °C in a gel dryer under vacuum. Switch off the heating and leave under vacuum for another 30 min.
- 8) Expose an image plate to the dried gel overnight.
- 9) Scan the image plate with a phosphorimager; encircle each band, either by a rectangle, an ellipsoid, or by following the individual contours of the band; quantify the image quants therein and evaluate interference data as outlined in Section 18.2.3.

18.2.3

Data Evaluation

Iodine hydrolysis bands were visualized and quantified using a phosphorimager, in our case a Bio-Imaging Analyzer BAS-1000 or FLA 3000-2R (FUJIFILM) and the analysis software PCBAS/AIDA (Raytest). An application example is illustrated in Figure 18.3a,b, where two experiments were run in parallel (lanes 1–4 and 5–8). The quantification boxes positioned with the program AIDA are drawn with thin black or white lines (white when black lines are masked by high band intensities) in Figure 18.3a (Figure 18.3b is a magnification of the region comprising nucleotides 291–350 with quantification boxes omitted for better visualization of bands). In the case of insufficient separation of individual bands (e.g., boxes 1 and 2), two or more bands are enclosed in a single box. The next step is to determine the normalization factor κ to compensate for differences in total radioactivity in lanes *a* versus *b*, where *a* represents the cleaved conjugate at the endpoint (here, after 2 h) and *b* represents the fraction of conjugates *cis*-cleaved in the initial fast phase of the reaction (within 2 min). *a'* and *b'* represent data from another individual experiment; data from a third experiment (*a''* and *b''*) were also included for the calculation of the average *R*-value, but are not shown in Figure 18.3 and Table 18.1. The normalization factor κ is then calculated from the ratio of the sum of all band intensities in lane *a* versus *b* ($\Sigma I_a / \Sigma I_b$). Interference and enhancement effects for individual bands (or two or more bands if quantified as one due to low gel resolution) are then determined by calculating the ratio $R = (\kappa \times I_b) / I_a$. Interference effects are associated with *R*-values below 1.0, whereas enhancement effects will result in *R*-values above 1.0. In a previous study [19], only *R*-values below 0.82 and above 1.2 were considered significant, but these cutoff values are arbitrary and depend on the quality of the data, such as the number of individual experiments and the

fluctuations between them. In the example of Table 18.1, the data are based on three individual experiments, which we consider to be the minimum for such studies. Figure 18.4a shows a graphical representation of the mean R -values (including errors). We also evaluated the data following the calculation procedure described in Chapter 17 (Figure 18.4b). Both evaluation procedures revealed G300 as a site of $I\alpha S$ interference, while several other weaker interferences are only identified according to the evaluation procedure presented in Figure 18.4a.

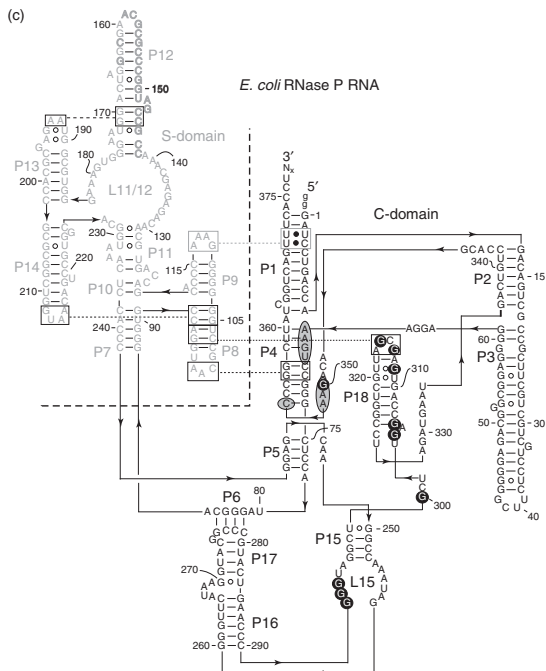
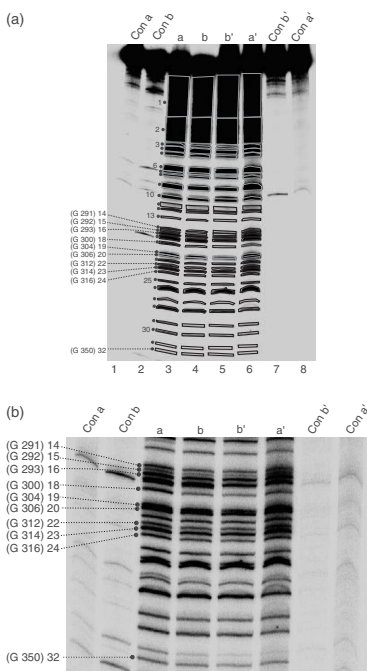
The knowledge gain derived from NAIM experiments can be largely extended when comparing NAIM results obtained for the same system but with different functional assays. For example, comparison of the interference results shown in Figure 18.4a with those observed in a gel retardation assay selective for high-affinity tRNA binding to *E. coli* P RNA [8] (Figure 18.5) reveals substantial overlap. In the region of nucleotides 291–350 (see secondary structure of *E. coli* P RNA in Figure 18.3c), $I\alpha S$ modifications at G291–293, G300, G304, G306, and G314 caused interference in both functional assays, although with different amplitudes. It should be noted that in the tRNA binding assay [8], interference at G300 was predominantly a phosphorothioate effect, already observable with the $G\alpha S$ modification alone, which, however, was not analyzed in the *cis*-cleavage assay (Figure 18.3). Modifications at G291–293 directly weaken the interaction with the 3'-CCA terminus of tRNA [8, 43], explaining why interferences at these positions are detected in both functional assays. Modifications at G300, G304, G306, and G314 apparently destabilize mature and precursor tRNA binding to the ribozyme, either directly or by inducing conformational changes of ribozyme structure. In contrast to the *cis*-cleavage assay applied here, the tRNA binding assay revealed additional $I\alpha S$ interferences at G329 and G356, but no effect at G312, G316, and G350 [8] (Figure 18.5). G312 and G316 are borderline cases due to the weakness of interference effects (Figure 18.4a). However, G350 represents a position at which $I\alpha S$ modification specifically interferes with the catalytic step. This is supported by a mutational study providing evidence that G350 contributes to the binding of catalytically important Mg^{2+} near the active site of *E. coli* P RNA [44]. In line with these biochemical findings, the recent crystal structure of a bacterial RNase P holoenzyme in complex with mature tRNA revealed that G350 is one of the eight P RNA nucleotides that form the immediate active site together with two putative catalytic metal ions [40].

18.3 Troubleshooting

18.3.1

RNA Transcription Reaction Did Not Work

- pH too low: check if all the reaction components were added in the required quantities, particularly the HEPES buffer; check the pH of the reaction mixture (it should be in the range of 7.5–8.0).
- For further troubleshooting, see Chapter 1.



18.3.2

RNA Degradation

RNase contamination: prepare all solutions freshly using RNase-free water (we use double-distilled H₂O).

18.3.3

Inefficient RNA Elution from Denaturing PAA Gels

- Use elution buffers described in Protocol 2a; do not use buffers containing HEPES instead of Tris.
- Lower the PAA concentration to an extent compatible with satisfactory gel resolution.

18.3.4

RNA Is Degraded after Elution

RNase contamination: prepare new elution buffer using RNase-free water; control pH (7.0).

18.3.5

Inefficient 3'- or 5'-End-Labeling

- Check the RNA you want to label by denaturing PAGE to analyze if it was already degraded before the labeling reaction.
- RNA is degraded during the labeling reaction: use fresh solutions for labeling reaction.
- Be sure that enzymes from previous steps (e.g., alkaline phosphatase) have been completely removed.

Figure 18.3 (a) NAIM experiment to identify λ S modifications that interfere with *cis*-cleavage of an *E. coli* P RNA-tRNA conjugate. 3'-End-labeled RNA samples, treated with iodine (lanes a, a', b, b') or not treated with iodine (lanes Con a, a', b, b') were loaded on a 10% PAA/8 M urea gel and separated by electrophoresis until xylene cyanol reached the bottom of the gel. Radioactive bands were visualized as described in Section 18.2.3. Lane a: pool of RNA molecules after 2 h of incubation, representing the endpoint of the reaction (Figure 18.2, step 4b); lane b: fraction of conjugates *cis*-cleaved in the initial fast phase of the reaction (within 2 min; Figure 18.2, step 4a). Lanes a' and b': same as lanes a and b, respectively, but representing a second individual

experiment. (b) Magnification of the informative section of the phosphorimage shown in (a). (c) Secondary structure presentation of *E. coli* RNase P RNA according to [35]. The C-domain (in black) and the S-domain (in gray) are separated by the thick dotted line (P, helical (paired) elements; L, loop regions). Tertiary interactions found in type A RNase P RNAs are indicated by boxes connected with thin dotted lines. Nucleotides for which λ S interference effects were inferred from experiments of the kind shown in (a, b) are highlighted as white letters on black circles. Gray-shaded ovals indicate nucleotides that belong to the immediate active site environment according to the recent crystal structure [40].

Table 18.1 Analysis of NAIM data obtained upon selection of fast-cleaving (within 2 min) *E. coli* P RNA-tRNA conjugates (fractions b and b') compared with the I α S-modified pool of RNA molecules after 2 h of incubation, representing the endpoint of the reaction (fractions a and a').

No	Position	Fraction a		Fraction b		Fraction b'		Fraction a'									
		I_a		I_b		$I_b \times \kappa$		$I_{a'}$		$I_{b'}$		$I_{b'} \times \kappa'$	R	R'	R''	$\varnothing R$	SD
1	—	267 325.7	268 508.1	285 370.41	205 632.4	247 931.9	229 510.56	1.07	1.16	1.02	1.07	0.03					
2	—	99 128.7	97 563	103 689.96	93 011.7	96 683.2	89 499.64	1.06	0.96	0.98	1.01	0.03					
3	—	5103.5	4019.6	4272.03	8152.5	3743.8	3465.64	0.84	0.43	1.08	0.78	0.19					
4	—	11 505.4	11 559.2	12 285.12	8920.8	9722.3	8999.93	1.07	1.01	1.04	1.04	0.02					
5	—	6894.3	6143.4	6529.21	6028.7	5262.1	4871.13	0.95	0.81	0.99	0.92	0.06					
6	—	16 869.3	15 114.1	16 063.27	11 072	13 779.9	12 756.05	0.96	1.15	1.11	1.07	0.06					
7	—	5823.9	3511.6	3732.13	3464.9	2963.2	2743.03	0.65	0.79	0.96	0.8	0.09					
8	—	6918.6	5624.5	5977.72	5825.2	5363.9	4965.36	0.86	0.85	0.96	0.89	0.03					
9	—	20 596.9	16 283.7	17 306.32	17 227.6	14 217.7	13 161.32	0.85	0.76	1.15	0.92	0.12					
10	271 + 270	9981.1	8461.5	8992.88	6623.5	8808	8153.57	0.90	1.23	1.21	1.11	0.11					
11	275 + 276	4298.9	4396.8	4672.92	4542.9	4306.1	3986.16	1.09	0.88	1.22	1.06	0.10					
12	280	3472.3	3499.8	3719.59	2768	3338	3089.99	1.07	1.12	0.93	1.04	0.06					
13	285	1660.9	1402.6	1490.68	1699.1	1788.3	1655.43	0.90	0.97	1.38	1.09	0.15					
14	291	2395.1	1871.8	1989.35	3471.2	2678.9	2479.84	0.83	0.71	0.86	0.80	0.04					
15	292	2225.7	1535	1631.40	3451.9	2099.5	1943.51	0.73	0.56	0.84	0.71	0.08					
16	293	5352.9	3555.6	3778.89	5085.5	4225.7	3911.73	0.71	0.77	0.80	0.76	0.03					
17	297 + 296	8472.7	7253.3	7708.81	5296.3	5313.3	4918.52	0.91	0.93	1.36	1.07	0.15					
18	300	3822.9	1686.3	1792.20	2301.7	1650.6	1527.96	0.47	0.66	0.63	0.59	0.06					
19	304	6238.6	4291.2	4560.69	4714.8	3548.7	3285.03	0.73	0.70	0.88	0.77	0.06					
20	306	7388.6	5131.3	5453.55	5342.3	4166.4	3856.84	0.74	0.72	0.81	0.76	0.03					
21	310	3425.2	2531	2689.95	3154.1	2732.1	2529.10	0.79	0.80	0.95	0.85	0.05					
22	312	2973	2091.6	2222.95	3304	2428.8	2248.34	0.75	0.68	0.97	0.80	0.09					
23	314	5925.1	3661	3890.91	4617.2	4307.2	3987.17	0.66	0.86	0.81	0.78	0.06					
24	316	2856.3	2112.4	2245.06	2818	1905.8	1764.20	0.79	0.63	1.02	0.81	0.12					
25	320	2667.7	2265.8	2408.09	2014.7	2228.7	2063.11	0.90	1.02	0.81	0.91	0.06					
26	323 + 324	10 050.2	9544.4	10 143.79	10 304.4	10 058.5	9311.15	1.01	0.90	0.93	0.95	0.03					
27	329	1942.9	1449.4	1540.42	1867.6	1696.1	1570.08	0.79	0.84	0.90	0.85	0.03					
28	332	6697.3	6058.8	6439.29	3723.5	5509.8	5100.42	0.96	1.37	0.80	1.04	0.17					
29	336	2334.2	2167.8	2303.94	2261.5	1788.6	1655.71	0.99	0.73	0.83	0.85	0.07					
30	340	2966	2978.4	3165.44	2224.3	2556.2	2366.27	1.07	1.06	0.71	0.95	0.12					
31	346	1264.6	804.2	854.70	702.1	632.2	585.23	0.68	0.83	1.14	0.88	0.14					
32	350	1944.3	1486.7	1580.06	1461.1	1236.6	1144.72	0.81	0.78	0.68	0.76	0.04					
Sum		540 522.8	508 563.9		443 085.5	478 672.1											
κ		1.06284146			κ'	0.92565558											

I , intensity; κ , normalization factor (see text).

Position: according to *E. coli* P RNA numbering system (Figure 18.3c).

$R = I_b \times \kappa / I_a$; $R' = I_{b'} \times \kappa' / I_{a'}$; $R'' = I_{b''} \times \kappa'' / I_{a''}$, primary data ($I_{a''}$; $I_{b''}$; $I_{b''} \times \kappa$) not included in the table; $\emptyset R$, the mean of R , R' , and R'' ; SD, standard deviation of the mean.

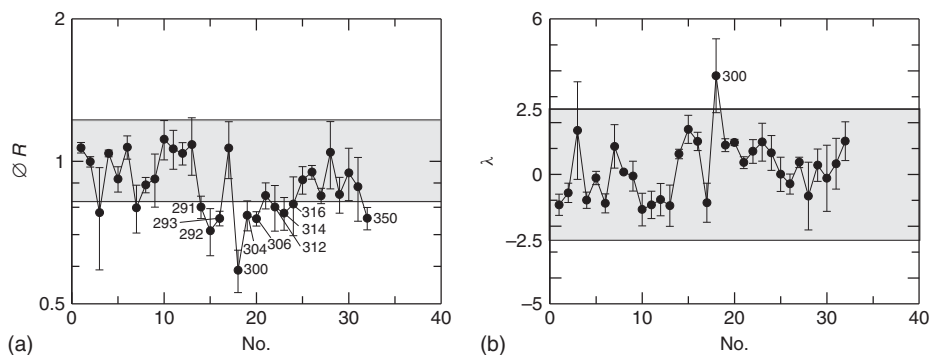


Figure 18.4 Quantification and statistical analysis of $I\alpha S$ modifications that interfere with *cis*-cleavage of an *E. coli* P RNA-tRNA conjugate, based on three independent experiments. (a) Mean R -values ($\bar{R}$; error bars: standard deviation of the mean) were plotted against the numbers (No.) of quantification units according to Table 18.1. Effects with R -values between 0.82 and 1.2 were arbitrarily considered insignificant (gray-shaded area). At positions where data

points correspond to individual G residues for which interference effects are suggested, the position according to the *E. coli* P RNA numbering system (see secondary structure in Figure 18.3c) is given next to the corresponding circle. (b) Data evaluation based on λ values as detailed in Chapter 17, starting from the primary intensity values (I_a and I_b ; $I_{a'}$ and $I_{b'}$; $I_{a''}$ and $I_{b''}$) according to Table 18.1.

- Be sure that the RNA sample is free of any residual chemicals that may inhibit the labeling reaction (e.g., SDS; use elution buffer 1 without SDS in Protocol 2a).
- Check enzyme activity: test old enzyme batch in parallel with a new one; also include a control substrate to rule out failure on the RNA level.

18.3.6

Iodine-Induced Hydrolysis Failed or Was Inefficient

- Increase the volume (decrease RNA concentration) of your iodine hydrolysis mix; with a total of 50 000 cpm of ^{32}P -labeled RNA, make sure not to exceed a radioactivity/volume ratio of 1000 cpm μl^{-1} in the iodine hydrolysis reaction; also consider shaking the iodine hydrolysis sample (we speculate that RNA solubility problems may be a cause for inefficient iodine hydrolysis).
- Always use freshly prepared iodine solutions.
- The incorporation of nucleotide analogs during *in vitro* transcription was insufficient.
 - Use freshly prepared stock solutions of analogs in RNase-free water, pH 7.0; use lithium salts if available; increase final iodine concentration to 1 mM (Protocol 6).
 - Modification extent is too low: increase ratio of analog/natural nucleotide; the degree of modification has to be optimized for each RNA and type of modification (Section 18.1.2.1).

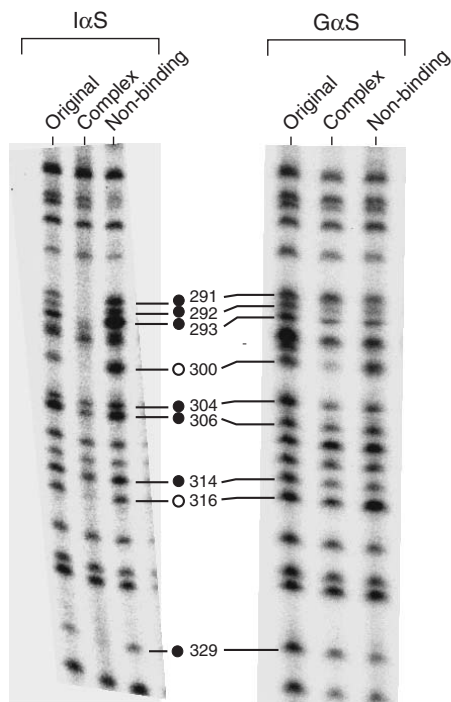


Figure 18.5 NAIM experiment to identify $I\alpha S$ and $G\alpha S$ modifications that interfere with tRNA binding to *E. coli* P RNA. (Source: Adapted from Ref. [8].) The comparison of $I\alpha S$ and $G\alpha S$ modifications permits to identify interference effects that are attributable to the phosphorothioate modification *per se* ($G\alpha S$). 3'-End-labeled RNA samples were subjected to iodine treatment and analyzed by denaturing PAGE: original, the original partially modified P RNA pool; complex, P RNA fractions recovered from the complex with a $tRNA^{Gly}$ derivative carrying an intron

in the anticodon arm [$tRNA^{Gly}_{(intron)}$]; non-binding, P RNA fractions that did not bind to $tRNA^{Gly}_{(intron)}$. The phosphorimage shows the region between nucleotides 270 and 332 of *E. coli* P RNA. Nucleotide positions indicated by circles are sites of tRNA-binding interference effects attributable to the inosine modification alone (black circles) or to both the *Rp*-phosphorothioate and the inosine modification (open circles); for details, see [8].

18.3.7

Unsatisfactory Gel Performance after Iodine Cleavage (Band Smearing, Curved Bands, Irregular Shape of Bands, Unequal Band Migration in Different Lanes, and Insufficient Band Separation)

- Always freshly prepare acrylamide-urea gel solutions and filter them before use; avoid heating up acrylamide-urea solutions for a prolonged time to dissolve the urea.
- Clean the glass plates thoroughly; make sure that the quality of glass plates is sufficient (plane surface, uniform thickness).

- After pulling the comb out, rinse the gel pockets extensively and immediately before loading the samples.
- Adjust electrophoretic conditions such that the temperature of the glass plates does not exceed about 50 °C, for example, let an 8% denaturing PAA gel (15 cm wide, 40 cm long, 0.5 mm thick) run at 10 mA.
- Unequal band migration in different lanes is frequently caused by different salt concentrations (leftovers from elution, ethanol precipitation, and iodine cleavage) in the compared samples. Washing the pellets with 70% ethanol after ethanol precipitation and using wider gel wells are appropriate actions to eliminate such problems.

References

1. Gish, G. and Eckstein, F. (1988) DNA and RNA sequence determination based on phosphorothioate chemistry. *Science*, **240**, 1520–1522.
2. Knöll, R., Bald, R., and Fürste, J.P. (1997) Complete identification of non-bridging phosphate oxygens involved in hammerhead cleavage. *RNA*, **3**, 132–140.
3. Strobel, S.A. (1999) A chemogenetic approach to RNA function/structure analysis. *Curr. Opin. Struct. Biol.*, **9**, 346–352.
4. Pearson, R.G. (1963) Hard and soft acids and bases. *J. Am. Chem. Soc.*, **85**, 3533–3539.
5. Warnecke, J.M., Fürste, J.P., Hardt, W.-D., Erdmann, V.A., and Hartmann, R.K. (1996) Ribonuclease P (RNase P) RNA is converted to a Cd(2+)-ribozyme by a single Rp-phosphorothioate modification in the precursor tRNA at the RNase P cleavage site. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 8924–8928.
6. Sigel, H. and Song, B. (1996) in *Metal ions in Biological Systems*, Vol. 32 (eds A. Sigel and H. Sigel), Marcel Dekker, Inc., New York, Chapter 5, pp. 135–198.
7. Rubin, J.R., Wang, J., and Sundaralingam, M. (1983) X-ray diffraction study of the zinc(II) binding sites in yeast phenylalanine transfer RNA. Preferential binding of zinc to guanines in purine-purine sequences. *Biochim. Biophys. Acta*, **756**, 111–118.
8. Heide, C., Pfeiffer, T., Nolan, J.M., and Hartmann, R.K. (1999) Guanosine 2'-NH₂ groups of *Escherichia coli* RNase P RNA involved in intramolecular tertiary contacts and direct interactions with tRNA. *RNA*, **5**, 102–116.
9. Ortoleva-Donnelly, L., Kronman, M., and Strobel, S.A. (1998) Identifying RNA minor groove tertiary contacts by nucleotide analogue interference mapping with N2-methylguanosine. *Biochemistry*, **37**, 12933–12942.
10. Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., and Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1685.
11. Pley, H.W., Flaherty, K.M., and McKay, D.B. (1994) Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature*, **372**, 111–113.
12. Siew, D., Zahler, N.H., Cassano, A.G., Strobel, S.A., and Harris, M.E. (1999) Identification of adenosine functional groups involved in substrate binding by the ribonucleosidic P ribozyme. *Biochemistry*, **38**, 1873–1883.
13. Griffiths, A.D., Potter, B.V., and Eperon, I.C. (1987) Stereospecificity of nucleases towards phosphorothioate-substituted RNA: stereochemistry of transcription by T7 RNA polymerase. *Nucleic Acids Res.*, **15**, 4145–4162.
14. Christian, E.L. and Yarus, M. (1992) Analysis of the role of phosphate oxygens in the group I intron from

- Tetrahymena*. *J. Mol. Biol.*, **228**, 743–758.
15. Strobel, S.A. and Shetty, K. (1997) Defining the chemical groups essential for *Tetrahymena* group I intron function by nucleotide analog interference mapping. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 2903–2908.
 16. Ryder, S.P. and Strobel, S.A. (1999) Nucleotide analog interference mapping. *Methods*, **18**, 38–50.
 17. Huang, Y., Eckstein, F., Padilla, R., and Sousa, R. (1997) Mechanism of ribose 2'-group discrimination by an RNA polymerase. *Biochemistry*, **36**, 8231–8242.
 18. Hardt, W.-D., Warnecke, J.M., Erdmann, V.A., and Hartmann, R.K. (1995) Rp-phosphorothioate modifications in RNase P RNA that interfere with tRNA binding. *EMBO J.*, **14**, 2935–2944.
 19. Rox, C., Feltens, R., Pfeiffer, T., and Hartmann, R.K. (2002) Potential contact sites between the protein and RNA subunit in the *Bacillus subtilis* RNase P holoenzyme. *J. Mol. Biol.*, **315**, 551–560.
 20. Harris, M.E. and Pace, N.R. (1995) Identification of phosphates involved in catalysis by the ribozyme RNase P RNA. *RNA*, **1**, 210–218.
 21. Kazantsev, A.V. and Pace, N.R. (1998) Identification by modification-interference of purine N-7 and ribose 2'-OH groups critical for catalysis by bacterial ribonuclease P. *RNA*, **4**, 937–947.
 22. Kaye, N.M., Christian, E.L., and Harris, M.E. (2002) NAIM and site-specific functional group modification analysis of RNase P RNA: magnesium dependent structure within the conserved P1-P4 multihelix junction contributes to catalysis. *Biochemistry*, **41**, 4533–4545.
 23. Hardt, W.-D., Erdmann, V.A., and Hartmann, R.K. (1996) Rp-deoxy-phosphorothioate modification interference experiments identify 2'-OH groups in RNase P RNA that are crucial to tRNA binding. *RNA*, **2**, 1189–1198.
 24. Heide, C., Feltens, R., and Hartmann, R.K. (2001) Purine N7 groups that are crucial to the interaction of *Escherichia coli* RNase P RNA with tRNA. *RNA*, **7**, 958–968.
 25. Boudvillain, M. and Pyle, A.M. (1998) Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.*, **17**, 7091–7104.
 26. Chanfreau, G. and Jacquier, A. (1994) Catalytic site components common to both splicing steps of a group II intron. *Science*, **266**, 1383–1387.
 27. Smith, J.S. and Nikonowicz, E.P. (2000) Phosphorothioate substitution can substantially alter RNA conformation. *Biochemistry*, **39**, 5642–5652.
 28. Christian, E.L. and Yarus, M. (1993) Metal coordination sites that contribute to structure and catalysis in the group I intron from *Tetrahymena*. *Biochemistry*, **32**, 4475–4480.
 29. Jack, A., Ladner, J.E., Rhodes, D., Brown, R.S., and Klug, A. (1977) A crystallographic study of metal-binding to yeast phenylalanine transfer RNA. *J. Mol. Biol.*, **111**, 315–328.
 30. Scott, W.G., Finch, J.T., and Klug, A. (1995) The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell*, **81**, 991–1002.
 31. Christian, E.L., Kaye, N.M., and Harris, M.E. (2000) Helix P4 is a divalent metal ion binding site in the conserved core of the ribonuclease P ribozyme. *RNA*, **6**, 511–519.
 32. Christian, E.L., Kaye, N.M., and Harris, M.E. (2002) Evidence for a polynuclear metal ion binding site in the catalytic domain of ribonuclease P RNA. *EMBO J.*, **21**, 2253–2262.
 33. Cray, S.M., Kurz, J.C., and Fierke, C.A. (2002) Specific phosphorothioate substitutions probe the active site of *Bacillus subtilis* ribonuclease P. *RNA*, **8**, 933–947.
 34. Brown, J.W., Nolan, J.M., Haas, E.S., Rubio, M.A., Major, F., and Pace, N.R. (1996) Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial populations reveals tertiary structural elements. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 3001–3006.

35. Massire, C., Jaeger, L., and Westhof, E. (1998) Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.*, **279**, 773–793.
36. Easterwood, T.R. and Harvey, S.C. (1997) Ribonuclease P RNA: models of the 15/16 bulge from *Escherichia coli* and the P15 stem loop of *Bacillus subtilis*. *RNA*, **3**, 577–585.
37. Torres-Larios, A., Swinger, K.K., Krasilnikov, A.S., Pan, T., and Mondragon, A. (2005) Crystal structure of the RNA component of bacterial ribonuclease P. *Nature*, **437**, 584–587.
38. Szewczak, A.A., Ortoleva-Donnelly, L., Zivarts, M.V., Oyeler, A.K., Kazantsev, A.V., and Strobel, S.A. (1999) An important base triple anchors the substrate helix recognition surface within the *Tetrahymena* ribozyme active site. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 11183–11188.
39. Golden, B.L., Gooding, A.R., Podell, E.R., and Cech, T.R. (1998) A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science*, **282**, 259–264.
40. Reiter, N.J., Osterman, A., Torres-Larios, A., Swinger, K.K., Pan, T., and Mondragon, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*, **468**, 784–789.
41. Krasilnikov, A.S., Yang, X., Pan, T., and Mondragon, A. (2003) Crystal structure of the specificity domain of ribonuclease P. *Nature*, **421**, 760–764.
42. Persson, T., Cuzic, S., and Hartmann, R.K. (2003) Catalysis by RNase P RNA: unique features and unprecedented active site plasticity. *J. Biol. Chem.*, **278**, 43394–43401.
43. Busch, S., Kirsebom, L.A., Notbohm, H., and Hartmann, R.K. (2000) Differential role of the intermolecular base-pairs G292-C(75) and G293-C(74) in the reaction catalyzed by *Escherichia coli* RNase P RNA. *J. Mol. Biol.*, **299**, 941–951.
44. Rasmussen, T.A. and Nolan, J.M. (2002) G350 of *Escherichia coli* RNase P RNA contributes to Mg^{2+} binding near the active site of the enzyme. *Gene*, **294**, 177–185.

19

Identification of Divalent Metal Ion Binding Sites in RNA/DNA-Metabolizing Enzymes by Fe(II)-Mediated Hydroxyl Radical Cleavage

Yan-Guo Ren, Niklas Henriksson, and Anders Virtanen

19.1

Introduction

The presence and requisite of divalent metal ions in the active sites of nucleic acid metabolizing enzymes, which participate in phospho(di)ester formation and breakage, has emerged as a common theme (reviewed in [1, 2]). One of the best-studied active sites involved in the cleavage of a phosphodiester bond is the 3' exonucleolytic site of *Escherichia coli* DNA polymerase (Pol) I, which has been characterized by a combination of genetics, biochemistry, and structural techniques (see [3–5] and references therein). Most importantly, crystallographic studies provided direct evidence that divalent metal ions are coordinated in this active site directly or *via* bridging water to oxygen atoms of four acidic amino acid residues as well as to a non-bridging oxygen at the scissile phosphodiester. The divalent metal ions in the 3' exonucleolytic site of DNA Pol I play a critical role during catalysis, and it has been proposed that the nucleophile (water or hydroxide ion) attacking the scissile phosphate during cleavage is oriented by one metal ion and two amino acid residues. After cleavage, a second divalent metal ion stabilizes the negative charge on the leaving group. A similar mechanism, where one divalent metal ion activates the initially attacking nucleophile while the other stabilizes the leaving group, is used by several other enzymes involved in breaking and forming phospho(di)ester bonds, such as endo- or exonucleases, kinases, phosphatases, and polymerases (reviewed in [1, 2, 6]). Taken together, it has become apparent that a very fruitful strategy to study the active site of any enzyme participating in phospho(di)ester formation and breakage is to identify and characterize its divalent metal ion binding sites.

Here, we describe protocols that we have used to characterize and map divalent metal ion binding sites in the active sites of the Klenow fragment of *E. coli* DNA Pol I and human poly(A)-specific ribonuclease (PARN) [7, 8]. PARN is a 3' exonuclease that efficiently degrades mRNA poly(A) tails [9–19] and belongs to the RNase D family of nucleases [8, 10, 20, 21], of which the 3' exonuclease domain of *E. coli* DNA Pol I is one of the best-studied examples. The method of Fe(II)-mediated hydroxyl radical cleavage [8] described here has been applied to map divalent metal

ion binding sites in a large variety of metalloenzymes (see, for example, [22–28] and references therein). We want to emphasize, however, that this method needs to be combined with several other approaches before a complete picture of a divalent metal ion binding site can be drawn. In our case, the Fe(II)-mediated hydroxyl radical cleavage assays of PARN were preceded by two important steps: (i) bioinformatic identification of amino acids potentially located in the active site of the enzyme and (ii) site-directed mutagenesis of amino acids expected to be important for catalysis. Several protocols and descriptions of these two steps are available elsewhere (see, for example, [29, 30]) and thus will not be provided here.

19.2

Probing Divalent Metal Ion Binding Sites

One of the most important prerequisites for the successful analysis of any enzyme is the availability of sufficiently large and pure preparations of the enzyme. It is therefore important to spend some time to figure out a simple and efficient protocol for the expression and purification of a recombinant form of the enzyme of interest. A large variety of expression systems are commercially available, and we have successfully used several of them (e.g., the pET system originally from Novagen, now Merck4Biosciences, or the pCAL system from Agilent Technologies). When choosing your expression system it is important to investigate if the recombinant form of the enzyme has the same key properties as the non-recombinant one, which is often not the case. For example, we observed when studying human poly(A) polymerases that the position, N- or C-terminus, of the tag used for affinity purification significantly affected the K_m parameter of the enzyme.

19.2.1

Fe(II)-Mediated Hydroxyl Radical Cleavage

The induction of hydroxyl radicals through the Fenton reaction [31] (Figure 19.1) in the vicinity of Fe(II) ion binding sites has become a powerful tool to identify divalent metal ion binding sites in protein and RNA enzymes. In the presence of reductants, such as ascorbic acid or dithiothreitol (DTT), Fe(II) generates hydroxyl radicals, which efficiently cleave the polypeptide or nucleic acid backbone in the vicinity of the Fe(II) binding site.

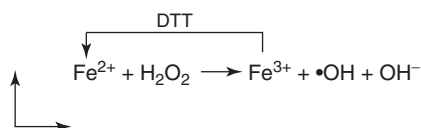


Figure 19.1 The Fenton reaction.

Before performing the Fe(II)-mediated hydroxyl radical cleavage assay, it is advisable to investigate if the enzyme under study is active in the presence of Fe(II) [32]. For PARN, we could readily detect enzymatic activity when we replaced the essential divalent metal ion Mg(II) with Fe(II) [8]. A positive result from this simple assay argues immediately that Fe(II) functions catalytically, which implies that some of the Fe(II) ion binding sites overlap with binding sites for Mg(II) ions. It is important to remember that the latter statement is one of the key assumptions of your analysis since you will argue that particular Fe(II) binding sites inferred from this cleavage assay correspond to binding sites for the natural metal cofactors in the active site of the enzyme.

Technically, the Fe(II)-mediated hydroxyl radical cleavage assay is easy to perform. The enzyme (0.5–10 μ M) is incubated in a buffer containing 50–100 mM HEPES, pH 7.0, 10 mM DTT, and 0.2–20 μ M Fe(NH₄)₂SO₄. The exact reaction volume, incubation time, and temperature as well as the amount of enzyme have to be determined empirically toward the goal to resolve visible and distinct cleavage products by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). As a rule, reaction volumes are usually about 10 μ l, and incubation times and temperatures vary between 2 and 30 min and 0 and 37 °C, respectively. Protocol 1 describes the conditions we used for PARN [8], while Protocol 2 describes our conditions for the Klenow Pol fragment [7]. Often, a small amount of H₂O₂ (~0.1–0.2% v/v) has to be added, as well as a small amount of NaCl. In the case of Fe(II)-mediated cleavage of PARN and Klenow Pol we could omit H₂O₂, while we included 5 mM NaCl in reactions containing the Klenow Pol fragment. The presence of substrate can also influence the cleavage pattern or the efficiency of cleavage. For PARN, the inclusion of the substrate poly(A₅₀) improved the cleavage reaction significantly, while the addition of DNA to Klenow Pol had a minor effect only.

The reaction is terminated by the addition of one reaction volume of 2 x SDS loading buffer and directly analyzed by SDS-PAGE. Subsequently, the cleavage products are visualized by silver staining if a non-radioactive polypeptide was reacted, or by autoradiography if the polypeptide was radioactively labeled. Figure 19.2 shows a typical result obtained when Klenow Pol is subjected to Fe(II)-mediated cleavage. Note the importance of the control lanes 2, 3, 5, and 6, which check for the appearance of non-specific cleavage products. In addition, lane 3 demonstrates the essential role of the reducing agent DTT, lane 6, including the chelator EDTA, documents the requirement for Fe(II) and/or possibly traces of other divalent metal ions brought in with the enzyme preparation, and lane 5 in comparison with lane 4 suggests that Fe(II) and Mg(II) occupy overlapping binding sites since addition of Mg(II) suppresses the appearance of the main cleavage product.

19.2.2

How to Map Divalent Metal Ion Binding Sites

The Fe(II)-induced cleavage site(s) will be in the vicinity, within a few Ångströms, of the binding site(s) for Fe(II). Thus, a major effort is dedicated to localizing the cleavage site. For this purpose, radioactive labeling of the N- or C-terminus of

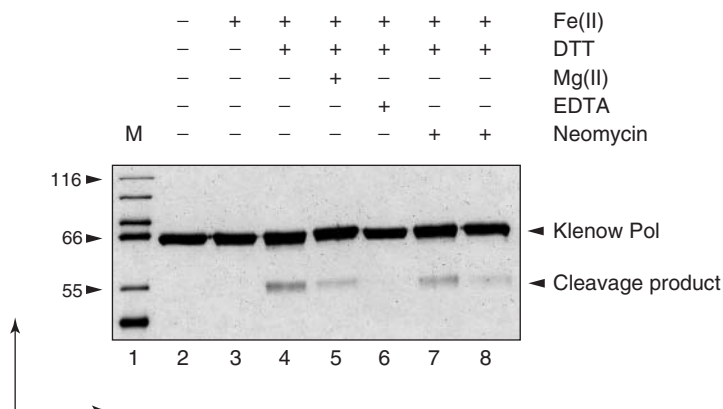


Figure 19.2 Fe(II)-mediated cleavage of recombinant Klenow Pol fragment; 2–4 μ g samples of Klenow Pol fragment were incubated as described in Protocol 2. The resulting cleavage products were analyzed by SDS-PAGE followed by silver staining. The presence (+) or absence (–) of the respective reagent is indicated at the top; Fe(II) (lanes 3–8), 20 μ M $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$; DTT (lanes 4–8), 10 mM DTT; Mg(II) (lane

5), 10 mM MgCl_2 ; EDTA (lane 6), 10 mM Na_2EDTA ; Neomycin (lanes 7 and 8), 2 and 10 mM neomycin B, respectively. The molecular size marker was fractionated in lane 1. The arrowheads on the right mark the positions of intact Klenow Pol and its specific cleavage fragment; arrowheads on the left depict selected size markers with their indicated molecular weight (in kilodaltons).

the enzyme and size fractionation of cleavage fragments by SDS-PAGE is often used [33–37]. A good assignment of the cleavage position then depends on an accurate correlation of molecular size and migrational distance. Another advantage of using radioactively labeled polypeptides is the possibility of quantifying the cleavage reaction. Protocol 3 describes our procedure for radioactive labeling of PARN at the N-terminus [8]. In this particular case, we have made use of a protein kinase recognition motif that was present in the N-terminal tag encoded in the commercially available expression plasmid (originally, pET33 from Novagen was used; a corresponding, now commercially available vector is pET-33b(+) from Merck4Biosciences).

However, owing to the tertiary structure of proteins, induced cleavages and Fe(II) binding sites are not always close to each other in the primary sequence. As a matter of fact, accurate mapping of the cleavage sites is not always required. This is very well exemplified by our studies of the active site of PARN. Here we could apply a different strategy instead, since bioinformatic characterization followed by site-directed mutagenesis had already revealed amino acids presumably located in the active site of the enzyme. Thus, we simply investigated if any of these introduced mutations affected the appearance of the Fe(II)-induced cleavage products in comparison with the wild-type (wt) enzyme. An altered cleavage pattern for the mutant polypeptide then indicated that the mutated amino acid is required for Fe(II)-mediated cleavage. The observed effects caused by different mutations ranged from minute decreases to complete disappearance of cleavage product. An

excellent way to quantify such effects is to determine an apparent K_d ($\text{app}K_d$) for the Fe(II) ions causing the cleavages [25]. For this purpose, an increasing amount of Fe(II) is added to the individual reactions, and the cleavage product at each concentration of Fe(II) is quantified. The saturation curve for the extent of specific cleavage product formation permits to derive an $\text{app}K_d$ of Fe(II) binding. It is worth mentioning that a difference in the calculated $\text{app}K_d$ for two or more cleavage sites in the same polypeptide implies that (i) the cleavages are induced by different Fe(II) ions and, thus, (ii) multiple binding sites for Fe(II) ions have been identified.

19.2.3

How to Use Aminoglycosides as Functional and Structural Probes

Aminoglycosides bind frequently to negatively charged binding pockets present in both protein enzymes and RNA (see [7, 38–41] and references therein). Often, these binding sites overlap with binding sites for divalent metal ions, and experimental evidence suggests that aminoglycosides displace functionally important divalent metal ions upon binding and thereby perturb the function of RNA and protein metalloenzymes (e.g., [7, 39–41]). Aminoglycosides have therefore turned out to be convenient probes in studies of divalent metal ion binding and function. For application of the strategy that we used in our studies of PARN, Klenow Pol, and poly(A) polymerase [7, 41], one may follow the experimental scheme outlined below (see [7, 41] for detailed descriptions):

- 1) Investigate if aminoglycosides inhibit the enzymatic activity by simply adding increasing amounts of aminoglycoside to the reaction. The chemical properties of the aminoglycoside will, of course, influence how efficiently it inhibits enzyme activity and one should therefore investigate a repertoire of commercially available aminoglycosides. As a rule of thumb: the higher its number of positively charged amino groups, the more efficiently the aminoglycoside will inhibit the enzyme. The interaction is highly electrostatic; the pH of the reaction therefore plays a decisive role and should usually be below 7.0. The inhibition constants are often in the micromolar range.
- 2) Once conditions for inhibition have been established, one should investigate if the aminoglycoside perturbs the Fe(II)-mediated cleavage reaction. For this purpose, include increasing amounts of aminoglycoside in the Fe(II)-mediated cleavage reaction, followed by SDS-PAGE. This is illustrated in Figure 19.2 (lanes 7 and 8) for neomycin B and the Klenow Pol fragment.
- 3) Finally, one should investigate if increasing amounts of a second divalent metal ion, such as Mg^{2+} , relieve the inhibition.

Provided certain aminoglycosides bind to the metalloenzyme of interest with reasonable affinity and specifically displace active site metal ion(s) as inferred from suppression of Fe(II)-mediated cleavage, one has established an elegant

experimental platform to investigate the structural and functional role of divalent metal ions in much detail.

Materials

Reagents and Stock Solutions

- 0.5 M HEPES-KOH, pH 7.0
- 20 or 100 μM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6 \text{H}_2\text{O}$, Sigma–Aldrich F3754], freshly prepared
- 50 mM DTT
- 0.3 μM poly(rA₅₀) (originally obtained from Sigma–Aldrich, which at present only offers bulk poly(rA) unfractionated in size, P9403)
- 2 x SDS-PAGE loading buffer supplemented with 0.2 M DTT: 0.25 mM Tris-HCl, pH 6.8 at room temperature, 20% glycerol, 2% SDS, 0.025% bromophenol blue, and 0.2 M DTT
- 50 mM NaCl
- 1 x spin column buffer A: 20 mM Tris-HCl, pH 7.5, 0.1 M NaCl, and 12 mM MgCl_2
- 1 x spin column buffer B: 25 mM HEPES-KOH, pH 7.0, and 0.1 M NaCl
- [γ -³²P]ATP (10 mCi ml⁻¹, 3000 Ci mmol⁻¹)
- 50 μM ATP

Consumables, Kits

- G-50 spin column (MicroSpin G-50 Columns, GE Healthcare)
- ProteoSilver™ Silver Stain Kit from Sigma–Aldrich, or PlusOne DNA Silver Staining Kit from GE Healthcare

Enzymes

- Pure recombinant human PARN ($\sim 10 \text{ pmol } \mu\text{l}^{-1}$). For preparation, see [8].
- Pure preparation (for details, see [7]) of Klenow Pol ($\sim 1\text{--}2 \text{ mg ml}^{-1}$) in 20 mM HEPES-KOH, pH 7.0, and 5 mM NaCl.
- Bovine heart protein kinase A catalytic subunit (Sigma–Aldrich P2645), supplied as lyophilized powder.

Protocol 1: Fe(II)-Mediated Cleavage of PARN

- 1) Label PARN with ³²P at the N-terminus using [γ -³²P]ATP and bovine heart protein kinase (see Protocol 3).
- 2) For each reaction, mix on ice 5 pmol of ³²P-labeled PARN, 1 μl 0.5 M HEPES-KOH, pH 7.0, 1 μl 20 μM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$, and adjust the volume to 8 μl with H₂O.
- 3) Start the reaction by adding 2 μl 50 mM DTT. Mix by gently flicking the tube with your finger and transfer to 37 °C. Incubate for 15–30 min.
- 4) Stop the reaction by adding 10 μl of 2 x SDS-PAGE loading buffer supplemented with 0.2 M DTT.
- 5) Boil the sample for 3 min.

- 6) Load on a 10% SDS-polyacrylamide gel with a 4% stacking gel.
- 7) Run the gel until the dye has reached the bottom.
- 8) Fix and dry the gel, and expose an X-ray film or Phosphorimage screen.

Protocol 2: Fe(II)-Mediated Cleavage of Klenow Pol

- 1) For each reaction, mix on ice 2–4 μg Klenow Pol, 2 μl 0.5 M HEPES-KOH, pH 7.0, 1 μl 50 mM NaCl, 2 μl 100 μM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$, and adjust the volume to 8 μl with H_2O . Replace $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ with H_2O for the control shown in lane 2 of Figure 19.2. Include additional reagents (see lanes 5–8 of Figure 19.2) at this step.
- 2) Start the reaction by adding 2 μl 50 mM DTT (2 μl H_2O for the control in lane 3 of Figure 19.2). Mix by gently flicking the tube with your finger and transfer to 37 °C. Incubate for 15–30 min.
- 3) Stop the reaction by adding 10 μl of 2 x SDS-PAGE loading buffer supplemented with 0.2 M DTT.
- 4) Boil the sample for 3 min.
- 5) Load on a 10% SDS-polyacrylamide gel with a 4% stacking gel.
- 6) Run the gel until the dye has reached the bottom.
- 7) Fix and silver stain the gel.

Protocol 3: Radioactive Labeling of Recombinant Polypeptide

The recombinant polypeptide should contain a protein kinase recognition motif, either at the N- or the C-terminus. A number of recombinant protein expression systems (Section 19.2.2) will provide such a motif in frame with the affinity tag and placed in the multiple cloning site. We have successfully labeled polypeptides expressed by the bacterial pET33 expression system (Novagen; now available as pET-33b(+)) from Merck4Biosciences).

- 1) Apply 30 μl of the purified recombinant polypeptide (~ 300 pmol) onto a G-50 spin column preequilibrated with 1 x spin column buffer A. Spin the column at 2000 g for 1 min.
- 2) Mix the eluate with 1 μl [$\gamma\text{-}^{32}\text{P}$]ATP (10 mCi ml^{-1} , 3000 Ci mmol^{-1}), 1 μl 50 μM ATP, and 3 μl (10 units μl^{-1}) of a freshly dissolved batch of bovine heart protein kinase A catalytic subunit.
- 3) Incubate for 30 min on ice.
- 4) Apply the labeling mixture onto a G-50 spin column equilibrated in 1 x spin column buffer B. Spin the column at 2000 g for 1 min. Collect the eluate and check the efficiency of labeling by SDS-PAGE analysis of a small aliquot of the eluate.

19.3

Notes and Troubleshooting

- 1) **No Fe(II)-mediated cleavage detected**
 - a. Use fresh Fe(II) and DTT solutions.

- b. Check if the pH of the reaction is altered. Usually, a higher pH (>7) facilitates Fe(II)-mediated cleavage, while a low pH (<6) abolishes cleavage.
 - c. Make sure that there is no metal ion-chelating compound such as EDTA present or check if contaminating divalent metal ions compete with Fe(II) in the reaction.
 - d. Try titrating H_2O_2 into the reaction.
- 2) **No distinct cleavage product(s) detected by SDS-PAGE after Fe(II)-mediated cleavage**
- a. The concentration of Fe(II) is too high. Optimize the concentration of Fe(II) in the reaction.
 - b. If H_2O_2 is used, optimize the concentration of H_2O_2 or omit it.
 - c. Shorten the incubation time and/or lower the temperature.
 - d. Try addition of enzyme substrate or product, which may contribute to the coordination of active site Fe(II) ions.
- 3) **Low efficiency of ^{32}P -incorporation by the kinase**
- a. Be sure to use a freshly dissolved batch of protein kinase (Protocol 3).
 - b. Check if the protein kinase recognition sequence tag is intact and has not been removed by protein degradation.

References

1. Steitz, T.A. and Steitz, J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 6498–6502.
2. Joyce, C.M. and Steitz, T.A. (1995) Polymerase structures and function: variations on a theme? *J. Bacteriol.*, **177**, 6321–6329.
3. Bernad, A., Blanco, L., Lazaro, J.M., Martin, G., and Salas, M. (1989) A conserved 3'-5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell*, **59**, 219–228.
4. Brautigam, C.A. and Steitz, T.A. (1998) Structural principles for the inhibition of the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I by phosphorothioates. *J. Mol. Biol.*, **277**, 363–377.
5. Joyce, C.M. and Steitz, T.A. (1994) Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.*, **63**, 777–822.
6. Steitz, T.A. (1998) A mechanism for all polymerases. *Nature*, **391**, 231–232.
7. Ren, Y.G., Martinez, J., Kirsebom, L.A., and Virtanen, A. (2002) Inhibition of Klenow DNA polymerase and poly(A)-specific ribonuclease by aminoglycosides. *RNA*, **8**, 1393–1400.
8. Ren, Y.G., Martinez, J., and Virtanen, A. (2002) Identification of the active site of poly(A)-specific ribonuclease by site-directed mutagenesis and $\text{Fe}(2^+)$ -mediated cleavage. *J. Biol. Chem.*, **277**, 5982–5987.
9. Körner, C.G. and Wahle, E. (1997) Poly(A) tail shortening by a mammalian poly(A)-specific 3'-exoribonuclease. *J. Biol. Chem.*, **272**, 10448–10456.
10. Körner, C.G., Wormington, M., Muckenthaler, M., Schneider, S., Dehlin, E., and Wahle, E. (1998) The deadenylating nuclease (DAN) is involved in poly(A) tail removal during the meiotic maturation of *Xenopus* oocytes. *EMBO J.*, **17**, 5427–5437.
11. Martinez, J., Ren, Y.G., Nilsson, P., Ehrenberg, M., and Virtanen, A. (2001) The mRNA cap structure stimulates rate of poly(A) removal and amplifies processivity of degradation. *J. Biol. Chem.*, **276**, 27923–27929.

12. Martinez, J., Ren, Y.G., Thuresson, A.C., Hellman, U., Astrom, J., and Virtanen, A. (2000) A 54-kDa fragment of the poly(A)-specific ribonuclease is an oligomeric, processive, and cap-interacting poly(A)-specific 3' exonuclease. *J. Biol. Chem.*, **275**, 24222–24230.
13. Åström, J., Åström, A., and Virtanen, A. (1991) *In vitro* deadenylation of mammalian mRNA by a HeLa cell 3' exonuclease. *EMBO J.*, **10**, 3067–3071.
14. Åström, J., Åström, A., and Virtanen, A. (1992) Properties of a HeLa cell 3' exonuclease specific for degrading poly(A) tails of mammalian mRNA. *J. Biol. Chem.*, **267**, 18154–18159.
15. Dehlin, E., Wormington, M., Körner, C.G., and Wahle, E. (2000) Cap-dependent deadenylation of mRNA. *EMBO J.*, **19**, 1079–1086.
16. Copeland, P.R. and Wormington, M. (2001) The mechanism and regulation of deadenylation: identification and characterization of *Xenopus* PARN. *RNA*, **7**, 875–886.
17. Gao, M., Fritz, D.T., Ford, L.P., and Wilusz, J. (2000) Interaction between a poly(A)-specific ribonuclease and the 5' cap influences mRNA deadenylation rates *in vitro*. *Mol. Cell*, **5**, 479–488.
18. Wu, M., Nilsson, P., Henriksson, N., Niedzwiecka, A., Lim, M.K., Cheng, Z., Kokkoris, K., Virtanen, A., and Song, H. (2009) Structural basis of m(7)GpppG binding to poly(A)-specific ribonuclease. *Structure*, **17**, 276–286.
19. Wu, M., Reuter, M., Lilie, H., Liu, Y., Wahle, E., and Song, H. (2005) Structural insight into poly(A) binding and catalytic mechanism of human PARN. *EMBO J.*, **24**, 4082–4093.
20. Mian, I.S. (1997) Comparative sequence analysis of ribonucleases HII, III, II PH and D. *Nucleic Acids Res.*, **25**, 3187–3195.
21. Ren, Y.G., Kirsebom, L.A., and Virtanen, A. (2004) Coordination of divalent metal ions in the active site of poly(A)-specific ribonuclease. *J. Biol. Chem.*, **279**, 48702–48706.
22. Farber, J.M. and Levine, R.L. (1986) Sequence of a peptide susceptible to mixed-function oxidation. Probable cation binding site in glutamine synthetase. *J. Biol. Chem.*, **261**, 4574–4578.
23. Ermacora, M.R., Delfino, J.M., Cuenoud, B., Schepartz, A., and Fox, R.O. (1992) Conformation-dependent cleavage of staphylococcal nuclease with a disulfide-linked iron chelate. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 6383–6387.
24. Grodsky, N.B., Soundar, S., and Colman, R.F. (2000) Evaluation by site-directed mutagenesis of aspartic acid residues in the metal site of pig heart NADP-dependent isocitrate dehydrogenase. *Biochemistry*, **39**, 2193–2200.
25. Mustaev, A., Kozlov, M., Markovtsov, V., Zaychikov, E., Denissova, L., and Goldfarb, A. (1997) Modular organization of the catalytic center of RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 6641–6645.
26. Lykke-Andersen, J., Garrett, R.A., and Kjems, J. (1997) Mapping metal ions at the catalytic centres of two intron-encoded endonucleases. *EMBO J.*, **16**, 3272–3281.
27. Godson, G.N., Schoenich, J., Sun, W., and Mustaev, A.A. (2000) Identification of the magnesium ion binding site in the catalytic center of *Escherichia coli* primase by iron cleavage. *Biochemistry*, **39**, 332–339.
28. Hlavaty, J.J., Benner, J.S., Hornstra, L.J., and Schildkraut, I. (2000) Identification of the metal-binding sites of restriction endonucleases by Fe²⁺-mediated oxidative cleavage. *Biochemistry*, **39**, 3097–3105.
29. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (eds) (2004) *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc., New York.
30. Sambrook, J. and Russel, D. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
31. Fenton, H.J.H. (1893) Oxidation of α -hydroxy acids with hydrogen peroxide and ferrous salts (Fenton's reagent) to α -keto acids or of 1,2-glycols to hydroxy aldehydes. *Proc. Chem. Soc.*, **9**, 113.
32. Brown, I.B. (1988) What factors determine cation coordination numbers? *Acta Crystallogr. Sect. B*, **44**, 545–553.

33. Jensen, T.H., Jensen, A., and Kjems, J. (1995) Tools for the production and purification of full-length, N- or C-terminal ³²P-labeled protein, applied to HIV-1 Gag and Rev. *Gene*, **162**, 235–237.
34. Jensen, T.H., Jensen, A., Szilvay, A.M., and Kjems, J. (1997) Probing the structure of HIV-1 Rev by protein footprinting of multiple monoclonal antibody-binding. *FEBS Lett.*, **414**, 50–54.
35. Jensen, T.H., Leffers, H., and Kjems, J. (1995) Intermolecular binding sites of human immunodeficiency virus type 1 Rev protein determined by protein footprinting. *J. Biol. Chem.*, **270**, 13777–13784.
36. Hori, R., Pyo, S., and Carey, M. (1995) Protease footprinting reveals a surface on transcription factor TFIIIB that serves as an interface for activators and coactivators. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 6047–6051.
37. Zhong, M., Lin, L., and Kallenbach, N.R. (1995) A method for probing the topography and interactions of proteins: footprinting of myoglobin. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 2111–2115.
38. Hermann, T. and Westhof, E. (1998) Aminoglycoside binding to the hammerhead ribozyme: a general model for the interaction of cationic antibiotics with RNA. *J. Mol. Biol.*, **276**, 903–912.
39. Mikkelsen, N.E., Brännvall, M., Virtanen, A., and Kirsebom, L. (1999) Inhibition of RNase P RNA cleavage by aminoglycosides. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 6155–6160.
40. Mikkelsen, N.E., Johansson, K., Virtanen, A., and Kirsebom, L.A. (2001) Aminoglycoside binding displaces a divalent metal ion in a tRNA-neomycin B complex. *Nat. Struct. Biol.*, **8**, 510–514.
41. Thuresson, A.C., Kirsebom, L.A., and Virtanen, A. (2007) Inhibition of poly(A) polymerase by aminoglycosides. *Biochimie*, **89**, 1221–1227.

20

RNA Structure and Folding Analyzed Using Small-Angle X-Ray Scattering

Nathan J. Baird, Jeremy West, and Tobin R. Sosnick

20.1

Introduction

Small-angle X-ray scattering (SAXS) is a solution technique that measures the size and the shape of an individual macromolecule or a complex of macromolecules. The method is well suited for the analysis of RNA structure and folding, as well as the association of RNA with proteins. For example, partially folded states that are not readily measured using other biophysical methods such as NMR or crystallography can be readily studied by SAXS.

In the typical SAXS experiment, X-rays are scattered by the sample and the scattering profile is measured at very low angles θ (Figure 20.1a). For globular objects, the profile of scattered intensity, approximated as a Gaussian $I(Q) = I_0 e^{-(Q^2 R_g^2)/3}$, is determined as a function of the momentum transfer $Q = 2\pi \sin \theta / \lambda$ where R_g is the radius of gyration, λ is the X-ray wavelength, and θ is the scattering angle. R_g is the root-mean-square of the distances of all regions to the center of mass of the particle appropriately weighed by its contrast (the excess electron density above that of the buffer solution, $\rho_{\text{RNA}} - \rho_{\text{solvent}}$). R_g can be obtained from the width of the Gaussian scattering profile at lower scattering angles (e.g., $R_g Q < 1.3$). Typically, R_g is obtained from the slope of the Guinier plot of $\ln I(Q)$ versus Q^2 .

The entire scattering profile can be used to obtain R_g and other shape information such as the pair-distance distribution function $P(r)$ and maximum distance within the particle, d_{max} [1]. $P(r)$ is the Fourier transform of the $I(Q)$ profile and it represents the probability distribution of vector distances between atom pairs within the macromolecule (Figure 20.1). It has a maximum at the most probable distance in the object (e.g., slightly larger than the radius for a sphere) and goes to zero at d_{max} (e.g., the diameter of a sphere). R_g can be calculated from the second moment of the $P(r)$ distribution, $R_g^2 = \frac{\int_0^\infty r^2 P(r) dr}{2 \int_0^\infty P(r) dr}$.

SAXS can be used to test the consistency of either high- or low-resolution structural models obtained from other experimental methods or from modeling.

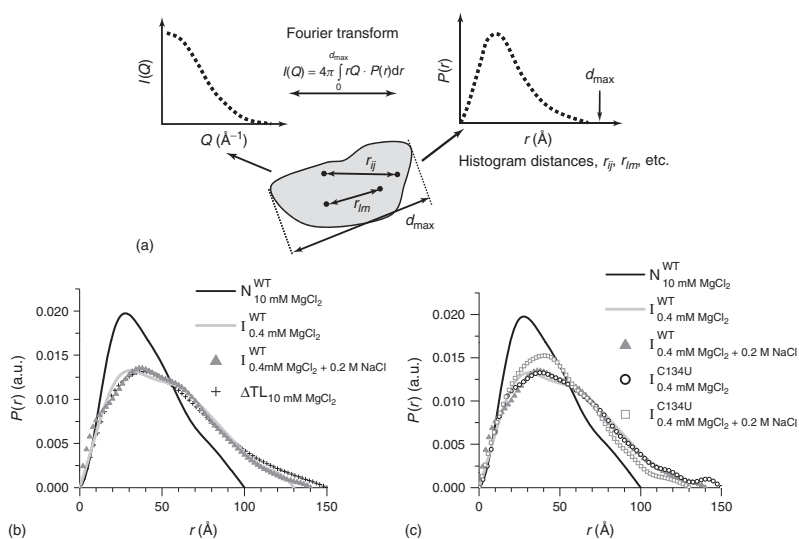


Figure 20.1 (a) Relationship between $I(Q)$, $P(r)$, and the macromolecule of interest sketched below the graphs. (b) Comparing the $P(r)$ function of the *B. subtilis* S-domain in its native (N) state (black), the intermediate (I, gray), the intermediate with 200 mM NaCl (gray triangles), and the Δ TL mutant in 10 mM Mg^{2+} (+). a.u., arbitrary units. (c) Comparing the $P(r)$ function as in (b), also including the folding intermediate of the C134U mutant S-domain either in the absence (open circles) or presence of 200 mM NaCl (open squares).

Both the experimental R_g value and the $P(r)$ function measured in a SAXS experiment can be compared with the predicted quantities [2]. As SAXS is a solution technique, this capability is particularly useful for determining whether a structure undergoes a conformational change upon crystallization or electron microscopy preparation. Multidomain RNAs or RNA–protein complexes can be modeled based on SAXS data using the crystal or NMR structures of the individual components. More recently, constructing a three-dimensional model from one-dimensional scattering data is common practice using the ATSAS suite of programs developed by Svergun and coworkers [3].

Another useful parameter derived from SAXS measurements is the absolute scattering intensity at zero angle, I_0 . For a monodispersed system, I_0 is related to the molecular weight (MW) of the scattering species, RNA concentration (C , in mg ml^{-1}), and the electron density (ρ) according to [4]

$$I_0 \propto (\rho_{\text{macromolecule}} - \rho_{\text{solvent}})^2 \times \text{MW} \times C \quad (20.1)^1$$

This proportionality helps determine the oligomerization state of an RNA or RNA–protein complex, when a suitable RNA mass standard is used, such as yeast tRNA^{Phe}. Because the crystal structure is known for this tRNA [5], its monodispersity can be confirmed by comparing the measured R_g and $P(r)$ values with those calculated from the crystal structure. Therefore, the I_0 value of the tRNA^{Phe} can be used as the MW and size standard for the study of the RNA or RNA–protein complex of interest [6].

An advantage of SAXS in studying RNA–protein complexes is that RNA scatters more strongly than protein due to its higher electron density or contrast. Generally, the relative scattering power of the RNA to the protein $(\rho_{\text{RNA}} - \rho_{\text{solvent}})^2 / (\rho_{\text{protein}} - \rho_{\text{solvent}})^2$ is approximately fivefold [7]. This property allows the size and the shape of the RNA to be analyzed in the presence of moderate amounts of protein, a feature particularly useful for identifying RNA conformational changes occurring upon protein binding.

The downside of the sensitivity of SAXS measurements to size and oligomeric state is that even mild amounts of aggregation can lead to spurious $P(r)$ and R_g values. Before synchrotron facilities were available, a typical SAXS experiment at a home source required relatively high RNA concentrations, for example, 10 mg ml^{-1} . A SAXS experiment at a synchrotron such as the Advanced Photon Source (APS) at the Argonne National Laboratory, however, only requires RNA samples in the range of $0.1\text{--}1 \text{ mg ml}^{-1}$ ($4\text{--}40 \text{ }\mu\text{M}$ yeast tRNA^{Phe}) and proportionally lower concentrations for larger RNAs. Data acquisition takes just a few seconds, which further reduces the chances of aggregation and sample degradation. In addition, lower concentrations can be critical for samples that are difficult to obtain, or permit measurements to be conducted over a variety of conditions.

¹) $\propto$ = proportional to.

20.2

Description of Method

20.2.1

General Requirements

Biological SAXS experiments require instruments designed specially for low-angle measurements and optimized for low background levels. As mentioned, synchrotron-based measurements have a significant number of advantages (Figure 20.2). In addition, a programmable titrator in conjunction with a flow-through sample cell enables an entire titration series (e.g., varying the ion or denaturant concentration) to be expediently carried out with a single sample in an hour or less.

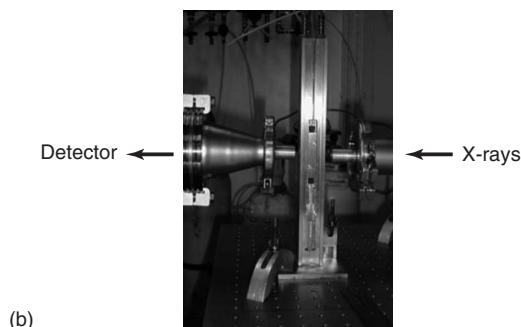
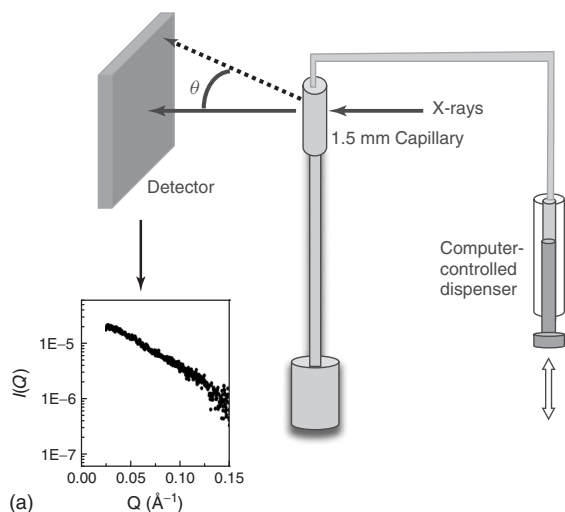


Figure 20.2 (a) Setup of a SAXS experiment at a synchrotron facility. (b) The sample holder used in our SAXS studies, sandwiched between the X-ray beam on the right and the CCD detector on the left.

The following steps are recommended for a SAXS experiment.

- 1) **Find a synchrotron-based beamline with personnel who have experience with biological samples.** As biological samples scatter weakly, their signal can be orders of magnitude weaker than background levels. Hence, minimizing background levels at the outset of an experiment is highly advantageous in terms of sample requirements and reproducibility. Frequent interactions are necessary between the novice user and the instrument scientists, and fortunately, many facilities now have established user programs that provide expert support.
- 2) **Sample requirement.** Different synchrotron beamlines have their own sample requirements and holders. In case of the APS, the required volume for a single experiment using current sample holders ranges from approximately 12 to 100 μl of $\sim 0.3 \text{ mg ml}^{-1}$ RNA with MW $\sim 50 \text{ kDa}$. Large RNAs scatter more strongly than small RNAs (intensity $\propto$ MW), so the operational concentration can be proportionally lower for large RNAs. For a titration series of 10–20 data points, at least 1 ml of sample is recommended.
- 3) **Time requirement.** The most time-consuming component is the instrumental setup that is done by the staff at the synchrotron facility. A single SAXS measurement takes about 10 min, most of which is spent on sample manipulation. At the BioCAT and BSSERC ID-12 beamlines at the APS, the actual data acquisition takes just a few seconds or less. A titration series can take approximately an hour, including the background (buffer) measurements.
- 4) **Data analysis.** To confirm that measurements are generating useful information, preliminary data analysis should be conducted during the course of the experiment. This analysis generally includes background subtractions and the Guinier analysis to obtain I_0 and R_g . These quantities enable the user to confirm that background levels are minimized, the signal is reproducible, and the sample is monodispersed – three critical elements of a high-quality experiment.

20.2.2

SAXS Application Example

RNase P is an essential enzyme required for the 5' maturation of all tRNAs and is conserved in all three kingdoms of life [8, 9]. The bacterial RNase P RNA subunit (P RNA) alone is catalytically active, but the protein subunit is required for full activity under physiological conditions. P RNA is composed of two independently folding domains, the specificity (S) domain and the catalytic (C) domain. We used SAXS to address the following questions focusing on the folding behavior of two constructs, the independently folding S-domain from *Bacillus subtilis* (for more details, see [10, 11]) and the full-length P RNA:

- 1) What is the global conformation of the stable S-domain folding intermediate?
- 2) What interactions contribute to the extended conformation of this intermediate?
- 3) To what extent are low-resolution SAXS reconstructions useful in RNA modeling studies?

20.2.3

General Information

RNA samples were obtained by standard transcription using T7 RNA polymerase and purification on denaturing polyacrylamide (PAA) gels. The RNA concentration was determined by UV absorbance using previously determined extinction coefficients.

The RNA was renatured as follows: (i) heat the RNA in the buffer without Mg^{2+} at 85 °C for 2 min, (ii) incubate at room temperature for 5 min, (iii) add MgCl_2 to the desired final concentration, (iv) incubate at 50 °C for 5 min, and (v) add NaCl to the desired concentration.

SAXS experiments were carried out using the SAXS instrument on the BioCAT ID-18 beamline of the APS [12]. Data were collected using a 165 mm MAR CCD (charge-coupled device) camera and exposure times were 1–6 s for each measurement. Sample to detector distance was 2.3–2.5 m; the energy of X-ray radiation was set to 12.4 keV. Computer-controlled Hamilton syringes injected the samples into a thermostated flow cell made of a 1.5 mm diameter cylindrical quartz capillary (Figure 20.2). The background scattering was collected from a buffer solution in the identical configuration. Samples were measured under constant flow conditions in order to reduce the possibility of radiation damage.

20.2.4

Question 1: The Global Conformation of the S-Domain Folding Intermediate

The crystal structure of the *B. subtilis* S-domain [13] contains four structural modules – a four-way junction at the base, a tetraloop–receptor interaction at the distal end, a non-canonical tertiary junction (J11/12), and a central core connecting these pieces (Figure 20.3). Our folding studies of the *B. subtilis* S-domain demonstrated that this RNA traverses a well-separated *unfolded-intermediate-native* (U-I-N) Mg^{2+} -induced folding pathway. We aimed at identifying the structural features involved in the final transition from the intermediate to the native state. Chemical and nuclease mapping experiments probe the local RNA structure at nucleotide resolution. Observed changes in the level of modification or extent of hydrolysis between the intermediate and native states were used to characterize the conformations of the four structural modules in the intermediate state relative to the native structure. The four-way junction and J11/12 modules are folded in the intermediate. The core and the tetraloop–receptor modules exhibit distinctly different hydrolysis and modification patterns between the intermediate and native states. These latter modules are not yet folded in the intermediate. The local structural mapping experiments are void of any information regarding the relative orientation of the four modules in the intermediate state. Global constraints are necessary to orient the various modules in space.

SAXS experiments were conducted to obtain global size and shape information. Folding from the intermediate state (0.4 mM MgCl_2) to the native state (10 mM MgCl_2) resulted in a global collapse as measured by the change in R_g (ΔR_g) of

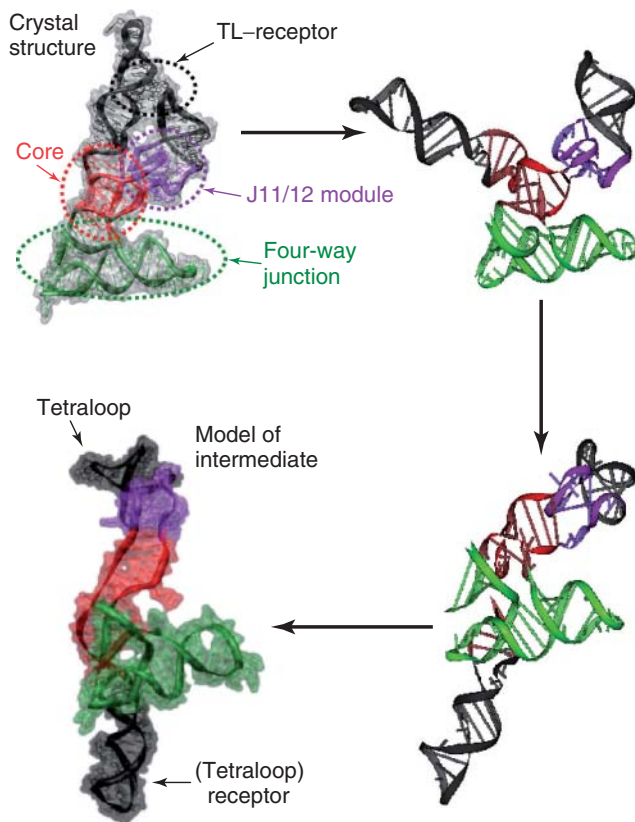


Figure 20.3 Process of modeling the *B. subtilis* S-domain folding intermediate beginning with the crystal structure (top left). The four modules of the native state are colored and indicated accordingly. Breaking of the tetraloop–receptor (TL–receptor) interaction (top right) alone is not sufficient to mimic the experimental SAXS data. Global

rearrangement of the helix containing the receptor (bottom right) is still insufficient to fit the experimental data. Movement of the helices containing both the tetraloop and the tetraloop receptor results in a model of the intermediate in good agreement with the experimental SAXS data (Figure 20.1b and Table 20.1).

$\sim 8\text{--}10\text{ \AA}$ (Table 20.1). In addition, the $P(r)$ plots exhibit distinct shapes for the intermediate ($d_{\text{max}} = 135\text{ \AA}$) and native states ($d_{\text{max}} = 110\text{ \AA}$) (Figure 20.1b). These global constraints complement the local constraints obtained from structural mapping. Working backward from the crystal structure, we used structural modeling software to disrupt the relative orientation of the four S-domain modules in agreement with the structural mapping data. Multiple models resulted from this process (Figure 20.3). Using available computational programs, we calculated theoretical scattering profiles, including R_g values and $P(r)$ plots, from the models we created. This process was iterated until good agreement was reached between the theoretical scattering parameters and those determined experimentally (Figure 20.1b). The

Table 20.1 Compaction of the *B. subtilis* S-domain RNA measured by SAXS.

Construct	Solution condition	R_g (Å)
Wild type	10 mM $MgCl_2$	32.4 ± 0.1
	0.4 mM $MgCl_2$	42.2 ± 0.3
	0.4 mM $MgCl_2$ + 0.2 M NaCl	41.8 ± 0.3
C134U	0.4 mM $MgCl_2$	43.4 ± 0.2
	0.4 mM $MgCl_2$ + 0.2 M NaCl	38.8 ± 0.1
$\Delta L12$	10 mM $MgCl_2$	44.6 ± 0.4

main feature of our final model is that it is quite extended with non-native core regions. In the intermediate, the tetraloop-containing helix is nearly 180° and >120 Å from the helix containing its receptor. Folding to the native state involves a dramatic rotation, over $\sim 120^\circ$, of these two helices involved in the tetraloop–receptor interaction and concomitant folding of the central core of the RNA.

20.2.5

Question 2: The Stable, Extended Conformation of the S-Domain Folding Intermediate

The orientation of the four modules of the *B. subtilis* S-domain intermediate results in an extended conformation in our model. The helices containing the tetraloop and its receptor are splayed $\sim 180^\circ$ apart (Figure 20.3). In the native structure, the apices of these helical arms interact to form the tetraloop–receptor interaction, which results in an angle of $\sim 60^\circ$ between them. The four-way junction is largely unchanged relative to the native conformation and is arranged approximately perpendicular to the extended axis. The core module is expanded in the intermediate and contains a 5 nt linker between the two extended helical arms (Figure 20.4). In the native structure, this linker folds into a U-turn and participates in an A-minor interaction with the helices in the four-way junction module.

Various models of the extended folding intermediate demonstrate that the orientation of the receptor helix may vary up to $\sim 30^\circ$ and still be in agreement with the experimental data [10]. Nonetheless, it remained uncertain what factors stabilized the extended nature of the intermediate. Using SAXS, we tested two hypotheses regarding the forces involved in maintaining the extended conformation of the intermediate [11]. The first hypothesis posits that electrostatic repulsion of the RNA backbone gives rise to the extended orientation of the helices. At low divalent metal ion concentrations where the intermediate is stable, this hypothesis asserts that the negative charges of the phosphate backbone are insufficiently shielded to allow for collapse of the RNA structure. The second hypothesis suggests that specific interactions within the core of the molecule are responsible for restricting the conformation of the intermediate, preventing compaction of the molecule.

To test the first hypothesis, we performed SAXS experiments of the S-domain intermediate at 0.4 mM Mg^{2+} and varying concentrations of NaCl. If electrostatic

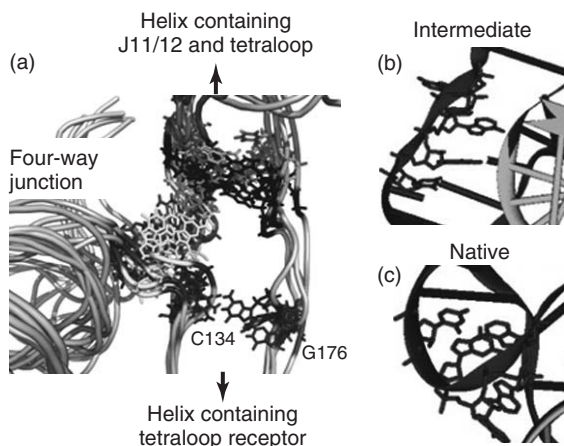


Figure 20.4 Models of the core in the intermediate and native states. (a) Langevin dynamics simulations demonstrate the formation of the non-native C134–G176 interaction in the intermediate. Note also the dramatic bend in the RNA backbone just

above C134. (b) Orientation of the 5 nt long linker (including G176) in the intermediate state. (c) Folding to the native state involves formation of a compact U-turn by this 5 nt long linker.

repulsion is limiting the collapse of the intermediate, the addition of positively charged Na^+ ions should result in a more compact intermediate state. SAXS experiments of the intermediate in up to 200 mM NaCl resulted in no global changes as inferred from R_g , d_{max} , and the $P(r)$ plot (Figure 20.1b, Table 20.1). To further investigate the role of electrostatic repulsion, we tested a construct with a mutated tetraloop (ΔTL), which is incapable of folding into the native structure. We probed this mutant construct in Mg^{2+} concentrations up to 10 mM Mg^{2+} , which is 5–10 times more Mg^{2+} than required to fold the wild-type S-domain to its native structure. The addition of 10 mM Mg^{2+} to this mutant resulted in $R_g = 44.6 \text{ \AA}$, in contrast to $R_g = 32.4 \text{ \AA}$ obtained for the native state. In fact, even at elevated Mg^{2+} concentrations, this construct remained extended, structurally similar to the wild-type intermediate ($R_g = 42.2 \text{ \AA}$; Figure 20.1b). From both results – the invariance of dimensions upon addition of either 200 mM NaCl to the wild-type S-domain or 10 mM MgCl_2 to the mutant RNA – we conclude that the contribution of electrostatic repulsion is insufficient to explain the stable, extended conformation of the S-domain intermediate.

Next we investigated the contribution of specific contacts within the core of the intermediate.

The geometry of the core in our model of the intermediate suggested that a non-native base pair (C134–G176) forms at the base of the helix containing the tetraloop receptor. The formation of this base pair was further supported by the Langevin dynamics simulations (Figure 20.4a). Closer examination of this non-native interaction revealed that the backbone undergoes a marked change in the direction of C134. This bend in the RNA backbone positions C134 to interact with

G176; in the native structure, C134 is not near G176 but makes a base–triple interaction with another base pair. We hypothesized that the C134–G176 base pair might be important in stabilizing the core of the intermediate. To test this hypothesis, we created a C134U mutant RNA and probed its folding using SAXS. In the absence of monovalent ions, the C134U mutant RNA yielded an R_g and $P(r)$ plot similar to the wild-type S-domain under Mg^{2+} concentrations that stabilize the intermediate state (Figure 20.1c, Table 20.1). Next, we equilibrated the C134U intermediate in 0.4 mM Mg^{2+} ($R_g = 43.4 \text{ \AA}$) and monitored the SAXS profile at various monovalent salt concentrations (Figure 20.1c, and data not shown). Addition of 200 mM NaCl resulted in the reduction of the C134U intermediate's R_g by more than 4 $\text{\AA}$ ($R_g = 38.8 \text{ \AA}$), whereas the wild-type RNA remained unchanged under identical conditions (Table 20.1). This substantial collapse of the C134U mutant S-domain supports the hypothesis that the C134–G176 interaction is important for stabilizing the intermediate conformation in the wild-type RNA. Disruption of this interaction results in sensitivity to the ionic conditions: in 200 mM NaCl, the mutant RNA structure is now able to collapse due to charge neutralization of the phosphate backbone.

20.2.6

Question 3: The Utility of Low-Resolution Real-Space Reconstructions in RNA Modeling

Doniach and colleagues [14] demonstrated the utility of Svergun's low-resolution SAXS reconstruction software for use in SAXS studies of RNA. They combined Svergun's DAMMIN/DAMMIF/DAMFILT software programs from the ATSAS suite of programs to generate bead models consistent with the experimental scattering data. For native conformations of tRNA^{Phe} and the P4–P6 domain of the group I intron, good agreement was found between the SAXS reconstructions and the available crystal structures. Recent studies of riboswitches have demonstrated the utility of SAXS reconstructions to characterize the conformational changes of these RNAs in the absence and presence of their cognate small molecule ligands [15–19]. In addition, Lilley and colleagues [20] successfully used real-space SAXS reconstructions in building a model of the VS ribozyme from its constituent parts.

We recently used SAXS and real-space reconstructions to investigate structural distinctions between two stable, native populations of the full-length P RNA. In the X-ray structures of the full-length P RNA, the two domains pack against each other in a side-by-side arrangement [21–23]. This arrangement is consistent with the current knowledge of how P RNA binds a pre-tRNA substrate and catalyzes the cleavage reaction [24, 25]. It is therefore commonly assumed that these domains have a similar arrangement in solution.

However, a previous SAXS result indicated that even at 10 mM $MgCl_2$, there is a large difference in the global conformation in the absence *versus* presence of monovalent ions [26]. We repeated the SAXS measurements and confirmed this puzzling observation (Figure 20.5). The radius of gyration, R_g , was obtained from

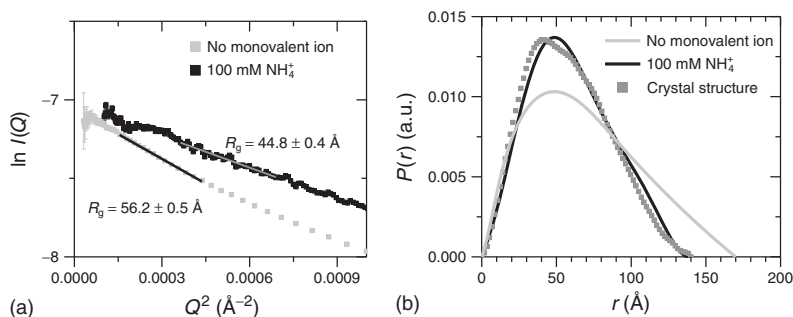


Figure 20.5 Compaction of the *B. subtilis* RNase P RNA upon the addition of 0.1 M monovalent ions. (a) The Guinier analysis to obtain R_g . The buffer contained 10 mM MgCl_2 , 20 mM Tris-HCl, pH 8.0, and either

0 or 100 mM NH_4Cl . Data were fit over the range marked by straight lines, with $Q_{\text{max}} \times R_g < 1.2$. (b) $P(r)$ analysis. $P(r)$ distributions are unit-normalized ($\int P(r) dr = 1$).

the Guinier analysis of the scattering data obtained at two ionic conditions. For the “ P_{0M} ” species populated in the absence of monovalent ions, $R_g = 56.2 \pm 0.5 \text{ \AA}$, which is substantially larger than $44.8 \pm 0.4 \text{ \AA}$, the R_g value for the “ $P_{0.1M}$ ” species formed in the presence of 0.1 M KCl or NH_4Cl . This difference in global dimensions is also apparent from the difference in the pair-distance distribution function, $P(r)$.

The addition of 0.1 M monovalent salt resulted both in small-scale changes localized to J5.1/7 and J11/5, the hinge regions between the two domains, as well as global shape changes, inferred from SAXS measurements. These changes indicate that the major structural difference between the P_{0M} and $P_{0.1M}$ native states resides in the relative orientation of the S- and C-domains. Crystallization of the full-length P RNA was carried out in the presence of 235 mM monovalent cations and 35 mM Mg^{2+} [21]. Hence, the crystal structure may represent the solution conformation observed in 0.1 M KCl or NH_4Cl ($P_{0.1M}$).

To identify the change in domain orientation, we used the crystal structure of *Bacillus stearothermophilus* P RNA (PDB 2A64) as a mimic of the $P_{0.1M}$ solution structure of *B. subtilis* P RNA. We then modeled the P_{0M} structure as a starting point, incorporating global and local information obtained from the SAXS and in-line probing data (not shown), respectively.

To visualize the ion-induced compaction, 3D reconstructions were performed with the program DAMMIF [27]. From 10 such reconstructions, a consensus reconstruction was generated using the DAMAVER/DAMFILT programs. The reconstruction is a space-filling bead model. This model was converted using the PDB2VOL program, a part of the SITUS suite of SAXS analysis programs [28], to create a final molecular envelope for the RNA (Figure 20.6). To examine its resolution and robustness in combination with our scattering data, the reconstruction process was repeated for each solvent condition. The envelopes resulting from the second reconstruction round were very similar to those obtained in the first round

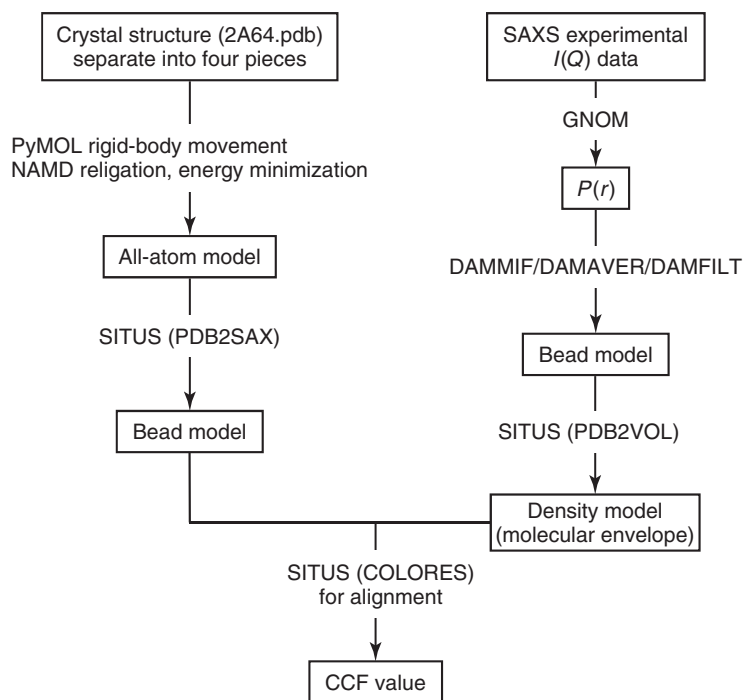


Figure 20.6 Flow diagram of modeling (left) and data analysis (right). Default parameters were used in all programs except for the Situs analysis section, for which 2–3 Å voxel spacing was used, and the DAMMIF reconstructions, for which a bead radius of 3.5 Å was chosen. The alignment was accomplished using the Situs program COLORES.

(Figure 20.7b). The difference between the envelopes at each condition (P_{0M} versus $P_{0.1M}$) provides a measure of the information content of the data.

The molecular envelope for the P RNA under “high ionic” conditions ($P_{0.1M}$) is clearly more compact than that for the “low ionic” conditions (P_{0M} , Figure 20.7b). The crystal structure of P RNA only fits well to the envelope for the high ionic conditions (Figure 20.7b), consistent with the agreement of their $P(r)$ distributions. From this agreement, we conclude that the crystal structure is indeed an adequate representation of the RNA under high ionic conditions, and no further adjustments are warranted given the resolution of the method.

To create a model of the P_{0M} species that agrees with its envelope, the C- and S-domains were repositioned using the PyMOL program (Schrödinger, LLC). The crystal structure was first split into four different PDB files to enable manual rigid body movement. The four pieces are the S-domain, the C-domain, the P8/P9 arm of the S-domain, and the bulge at the border of the C- and S-domains (residues 81–85). After adjustment, the pieces were re-ligated *in silico* and the local geometry improved by all-atom energy minimization using the NAMD program while allowing all atoms to move [29]. A crystal structure-derived model with the pieces

repositioned in an extended, nearly collinear arrangement of S- and C-domains was compatible with the experimental molecular envelope and the experimental $P(r)$ distribution of the P_{0M} structure (Figure 20.7).

To quantify the agreement between the reconstructed molecular envelopes and either the original crystal structure or our crystal structure-derived extended model of P_{0M} , the cross-correlation function (CCF) was calculated. The atomic coordinates of the model were converted to a bead model using the SITUS program PDB2SAX. The SITUS program COLORES performs a 6D rigid body search to overlay this bead model with the envelope created from the experimental bead model for the SAXS-based reconstruction (obtained using the program PDB2VOL; Figure 20.6). From the overlay of the two models, the COLORES program calculates an unnormalized CCF value.

The CCF value was 0.73 between the crystal structure and the $P_{0.1M}$ envelope and 0.72 between the crystal structure-derived extended model of P_{0M} and the P_{0M} envelope. The same CCF values were obtained when the reconstruction process was repeated (Figure 20.7b). In addition, the CCF value was reduced to 0.65 when the extended model of P_{0M} was correlated with the $P_{0.1M}$ envelope and to 0.64 when correlating the original crystal structure with the P_{0M} envelope.

To further test the robustness of the gross features of our extended model for the P_{0M} structure, the S-domain was repositioned in a slightly bent, nonlinear arrangement (Figure 20.7c). Two such models fit the molecular envelope slightly worse than the first model and have lower CCF values (0.71 and 0.65 *versus* 0.72). More importantly, their $P(r)$ functions exhibit substantial differences to the experimentally determined $P(r)$ distribution (Figure 20.7a). This disparity indicates that the structure(s) under low ionic conditions must have a highly extended geometry.

As a further validation of our quantitative approach that uses the CCF to gauge the quality of models *versus* SAXS data, we performed similar analyses of three other RNAs for which crystal structures and SAXS data are available (Figure 20.8). Crystal structures [13, 21] and SAXS data [6, 10] for the C- and S-domains from *B. subtilis*/*B. stearothermophilus* have been published previously. The S-domain of *Thermus thermophilus* [30] represents another architectural type different from the *B. subtilis* S-domain. The SAXS data for the *T. thermophilus* S-domain shown here is new. In all cases, the crystal structures largely match the respective envelopes obtained from the SAXS data. For the three RNAs, *B. subtilis* C- and S-domains and the *T. thermophilus* S-domain, the CCF values for two separate reconstructions were 0.72 & 0.75, 0.66 & 0.70, and 0.65 % 0.71, respectively. The similarity between the crystal structures and the SAXS-based reconstructions for these three RNAs supports our analysis procedure.

This and related previous studies of RNA tertiary structures [11, 15–17, 19, 20, 31, 32] demonstrate that the quality of molecular models can be scored by fitting into molecular envelopes obtained from SAXS measurements. It is clear that this mode of comparing and contrasting models with global size and shape data is a useful combination to more precisely refine molecular models.

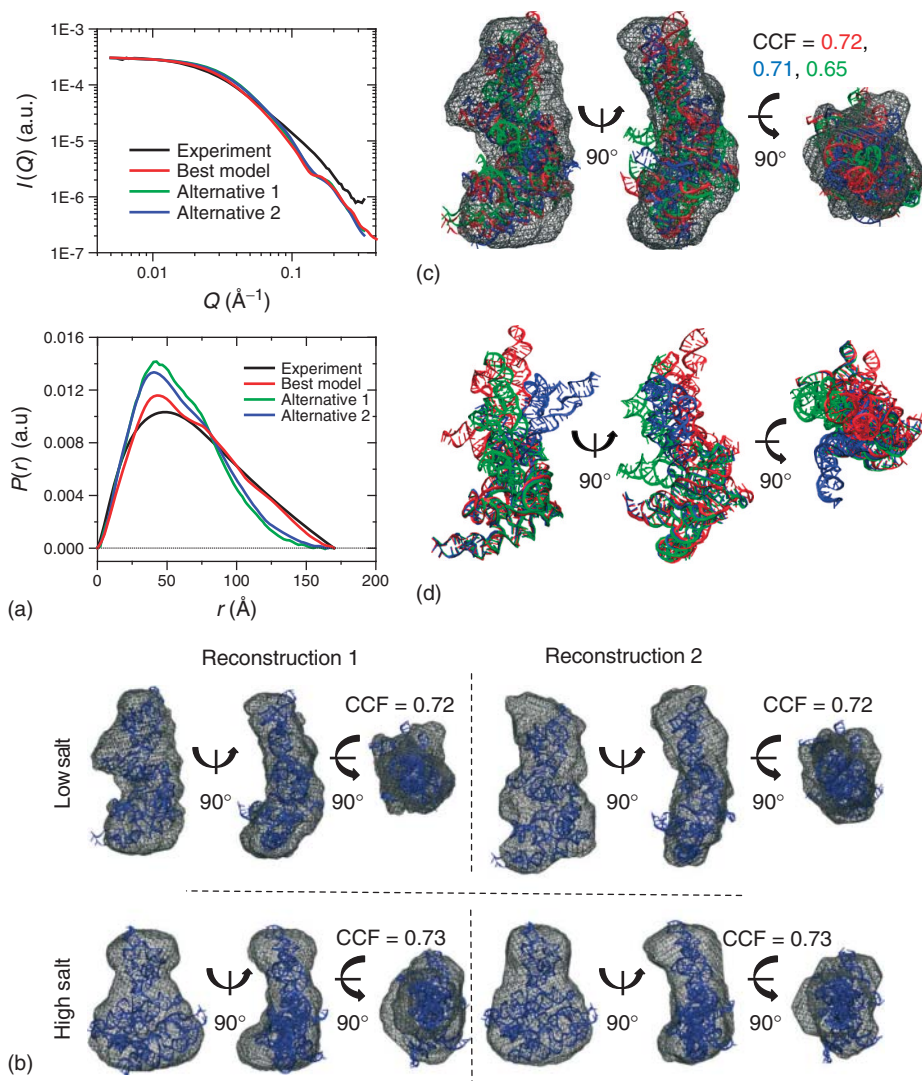


Figure 20.7 Comparing P RNA models and SAXS data under low (P_{0M}) and high ($P_{0.1M}$) monovalent ion concentrations. (a) Comparison between the experimental scattering curves and $P(r)$ functions for P_{0M} , and those from the best-fit model along with two alternative models built to test the robustness of the modeling process. $P(r)$ distributions are unit-normalized ($\int P(r) dr = 1$). (b) Superimposed on the molecular envelopes for the

high- and low-salt conditions are the crystal structure and our best-fit model derived from the crystal structure, respectively. (c) Three views (90° rotation) of the best-fit model (red) along with two alternative models, each independently fit to the P_{0M} molecular envelope. (d) Superposition of the two alternative models aligned to the lower half (C-domain) of the best-fit model to illustrate the differences among the three structures.

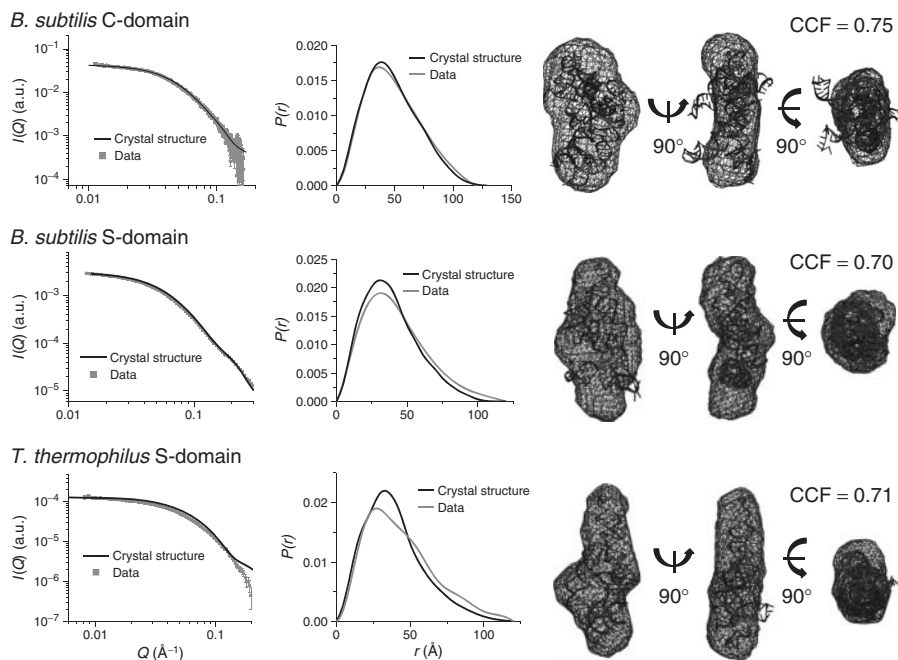


Figure 20.8 Comparison of molecular envelopes obtained from SAXS data and crystal structures for three RNAs. The size of these RNAs is 255 nucleotides for the *B. subtilis* C-domain, 154 nucleotides for the *B. subtilis* S-domain, and 161 nucleotides for the

T. thermophilus S-domain. For all three RNAs, SAXS data were obtained in 20 mM Tris-HCl, pH 8, 10 mM MgCl₂. $P(r)$ distributions are unit-normalized ($\int P(r) dr = 1$); a.u., arbitrary units.

20.3 Troubleshooting

20.3.1

Problem 1: Radiation Damage and Aggregation

Owing to the extremely high flux and tight focus at a synchrotron source, radiation damage is a serious concern. The intense X-ray beam generates hydroxyl radicals that react with the RNA and proteins in the sample to result in aggregation. This problem can often be identified by a change in the scattering profiles between successive exposures.

There are two simple remedies to deal with damage-induced aggregation issues. First, Tris and other organic buffers scavenge free radicals, so that their presence can significantly reduce radiation damage. Phosphate or other inorganic buffers are not recommended unless supplemented with a radical scavenger. Second, the measurement should be carried out under flow, so that a fresh solution of molecules is exposed to the X-ray beam at all times during measurement. Flow

speeds required depend on the intensity of the X-ray beam. Flow can be achieved using the sample handling device depicted in Figure 20.2.

20.3.2

Problem 2: High Scattering Background

Beyond optimizing beamline performance, RNA samples purified by denaturing gel electrophoresis sometimes contain minute amounts of PAA particles that strongly scatter X-rays. Because these particles may be much bigger than the RNA molecules, the scattering profile may be severely compromised by these unwanted particles. Passing the RNA samples through a 0.22 μm filter often alleviates this problem.

20.3.3

Problem 3: Scattering Results Cannot Be Fit to Simple Models

Before proposing complicated interpretations, it is often advisable to determine the level of sample heterogeneity. Even when a single RNA species is present, it may exist in two or more conformations. Such fractional folding or activity can be identified by other methods. This subject is particularly important for interpreting titration experiments where the concentration of one or more components is varied. In addition, the known propensity of RNA to misfold has reiterated the utility of connecting the output of a size exclusion column directly to the sample holder [19].

Another possible explanation for a complicated result is the loss of sample integrity during the experiment. Although conducting the measurements under flow conditions should preclude X-ray damage, samples after exposure to X-rays can be examined by other methods to ensure that the sample has remained intact.

Finally, the relatively high concentrations of RNA used in SAXS measurements can reduce the concentration of free divalent cations when measurements are performed in the absence of monovalent salts. Cations condense around the negatively charged RNA backbone. This ionic cloud reduces the number of free cations by approximately one Mg^{2+} ion per 2–4 nt. For example, a 10 mM RNA having 200 nt is neutralized by $\sim 0.5\text{--}1\text{ mM Mg}^{2+}$. This reduction in the free $[\text{Mg}^{2+}]$ can increase the apparent Mg^{2+} requirement for the folding of the RNA to its native structure [10]. Other methods such as absorbance, circular dichroism, or chemical probing can determine the folding behavior of the RNA under conditions identical to the planned SAXS experiments and identify the required divalent cation concentration needed for the SAXS measurement.

20.4

Conclusions – Outlook

The availability of high-flux synchrotron-based small-angle scattering beamlines along with reliable real-space reconstruction has renewed this relatively old biophysical method. Besides the RNase P work described here, SAXS has also been

applied recently to RNA folding studies (e.g., see [32] and references therein). The two advantages of the SAXS method are the determination of the shape and size change in real time under a broad range of solution conditions and the integration of SAXS data into structural modeling processes. We anticipate additional applications of SAXS to the understanding of other, even more complex biological systems in the near future.

Acknowledgments

All SAXS experiments were carried out at the Argonne National Laboratory at the BioCAT beamline. We thank Tao Pan for useful discussions, S. Zaheer and L. Guo for their assistance with the SAXS measurements, and X.-J. Yang and E. Haddadian for their assistance with the modeling and simulations of the native P RNA. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract No. W-31-109-ENG-38. BioCAT is a National Institutes of Health (NIH)-supported Research Center RR-08630. We thank NIH and the U.S. Department of Energy for their financial support.

The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Center for Research Resources or the NIH.

Abbreviations

C	Macromolecular mass concentration (mg ml^{-1})
I_0	The absolute scattering intensity at zero angle
MW	Molecular weight of the scattering species
nt	Nucleotide
$P(r)$	The pair-distribution function
ρ	The electron density
R_g	Radius of gyration
SAXS	Small-angle X-ray scattering

References

1. Moore, P.B. (1980) Small-angle scattering. Information content and error analysis. *J. Appl. Cryst.*, **13**, 168–175.
2. Svergun, D.I., Barberato, C., and Koch, M.H. (1995) CRY SOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, **28**, 768–773.
3. Svergun, D.I., Volkov, V.V., Kozin, M.B., and Stuhmann, H.B. (1996) New developments in direct shape determination from small-angle scattering. *Acta Crystallogr., Sect. A*, **52**, 419–426.
4. Glatter, O. and Kratky, O. (1982) *Small Angle X-Ray Scattering*, Academic Press, London.
5. Jovine, L., Djordjevic, S., and Rhodes, D. (2000) The crystal structure of yeast phenylalanine tRNA at 2.0 Å resolution: cleavage by Mg(2+) in 15-year old crystals *J. Mol. Biol.*, **301**, 401–414.
6. Fang, X., Littrell, K., Yang, X.J., Henderson, S.J., Siefert, S.,

- Thiyagarajan, P., Pan, T., and Sosnick, T.R. (2000) Mg^{2+} -dependent compaction and folding of yeast tRNA^{Phe} and the catalytic domain of the *B. subtilis* RNase P RNA determined by small-angle X-ray scattering. *Biochemistry*, **39**, 11107–11113.
7. Cantor, C. and Schimmel, P. (1980) *Biophysical Chemistry: Part II*, W.H. Freeman and Co, New York.
8. Frank, D.N. and Pace, N.R. (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.*, **67**, 153–180.
9. Altman, S. and Kirsebom, L. (1999) Ribonuclease P. In *The RNA World* (eds R.F. Gesteland, T.R. Cech, and J.F. Atkins), Cold Spring Harbor Laboratory Press, pp. 351–380.
10. Baird, N.J., Westhof, E., Qin, H., Pan, T., and Sosnick, T.R. (2005) Structure of a folding intermediate reveals the interplay between core and peripheral elements in RNA folding. *J. Mol. Biol.*, **352**, 712–722.
11. Baird, N.J., Gong, H., Zaheer, S.S., Freed, K.F., Pan, T., and Sosnick, T.R. (2010) Extended structures in RNA folding intermediates are due to nonnative interactions rather than electrostatic repulsion. *J. Mol. Biol.*, **397**, 1298–1306.
12. Seifert, S., Winans, R.E., Tiede, D.M., and Thiyagarajan, P. (2000) Design and performance of a SAXS instrument at the Advanced Photon Source. *J. Appl. Crystallogr.*, **33**, 782–784.
13. Krasilnikov, A.S., Yang, X., Pan, T., and Mondragon, A. (2003) Crystal structure of the specificity domain of ribonuclease P. *Nature*, **421**, 760–764.
14. Lipfert, J., Chu, V.B., Bai, Y., Herschlag, D., and Doniach, S. (2007a) Low-resolution models for nucleic acids from small-angle X-ray scattering with applications to electrostatic modeling. *J. Appl. Crystallogr.*, **40**, s229–s234.
15. Lipfert, J., Das, R., Chu, V.B., Kudaravalli, M., Boyd, N., Herschlag, D., and Doniach, S. (2007b) Structural transitions and thermodynamics of a glycine-dependent riboswitch from *Vibrio cholerae*. *J. Mol. Biol.*, **365**, 1393–1406.
16. Kulshina, N., Baird, N.J., and Ferré-D'Amaré, A.R. (2009) Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. *Nat. Struct. Mol. Biol.*, **16**, 1212–1217.
17. Ali, M., Lipfert, J., Seifert, S., Herschlag, D., and Doniach, S. (2010) The ligand-free state of the TPP riboswitch: a partially folded RNA structure. *J. Mol. Biol.*, **396**, 153–165.
18. Baird, N.J. and Ferré-D'Amaré, A.R. (2010) Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. *RNA*, **16**, 598–609.
19. Rambo, R.P. and Tainer, J.A. (2010) Improving small-angle X-ray scattering data for structural analyses of the RNA world. *RNA*, **16**, 638–646.
20. Lipfert, J., Ouellet, J., Norman, D.G., Doniach, S., and Lilley, D.M. (2008) The complete VS ribozyme in solution studied by small-angle X-ray scattering. *Structure*, **16**, 1357–1367.
21. Kazantsev, A.V., Krivenko, A.A., Harrington, D.J., Holbrook, S.R., Adams, P.D., and Pace, N.R. (2005) Crystal structure of a bacterial ribonuclease P RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 13392–13397.
22. Torres-Larios, A., Swinger, K.K., Krasilnikov, A.S., Pan, T., and Mondragon, A. (2005) Crystal structure of the RNA component of bacterial ribonuclease P. *Nature*, **437**, 584–587.
23. Kazantsev, A.V., Krivenko, A.A., and Pace, N.R. (2009) Mapping metal-binding sites in the catalytic domain of bacterial RNase P RNA. *RNA*, **15**, 266–276.
24. Kazantsev, A.V. and Pace, N.R. (2006) Bacterial RNase P: a new view of an ancient enzyme. *Nat. Rev. Microbiol.*, **4**, 729–740.
25. Torres-Larios, A., Swinger, K.K., Pan, T., and Mondragon, A. (2006) Structure of ribonuclease P – a universal ribozyme. *Curr. Opin. Struct. Biol.*, **16**, 327–335.
26. Fang, X.W., Yang, X.J., Littrell, K., Niranjanakumari, S., Thiyagarajan, P., Fierke, C.A., Sosnick, T.R., and Pan, T. (2001) The *Bacillus subtilis* RNase P holoenzyme contains two RNase P

- RNA and two RNase P protein subunits. *RNA*, **7**, 233–241.
27. Franke, D. and Svergun, D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **42**, 342–346.
 28. Wriggers, W. and Chacon, P. (2001) Using Situs for the registration of protein structures with low-resolution bead models from X-ray solution scattering. *J. Appl. Crystallogr.*, **34**, 773–776.
 29. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
 30. Krasilnikov, A.S., Xiao, Y., Pan, T., and Mondragon, A. (2004) Basis for structural diversity in homologous RNAs. *Science*, **306**, 104–107.
 31. Lipfert, J. and Doniach, S. (2007) Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.*, **36**, 307–327.
 32. Moghaddam, S., Caliskan, G., Chauhan, S., Hyeon, C., Briber, R.M., Thirumalai, D., and Woodson, S.A. (2009) Metal ion dependence of cooperative collapse transitions in RNA. *J. Mol. Biol.*, **393**, 753–764.

21

Temperature-Gradient Gel Electrophoresis of RNA

Detlev Riesner and Gerhard Steger

21.1

Introduction

As suggested by the name, temperature-gradient gel electrophoresis (TGGE) is a combination of a thermodynamic method and an analytical separation technique. The electrophoretic mobility of one or several RNA molecules is analyzed as a function of the temperature. A variety of parameters of an RNA affect this temperature-dependent mobility. These parameters are size, sequence, secondary and tertiary structure, structural stability, hydrodynamic flexibility, and electrical properties such as counterion condensation. As different electrophoretic mobilities result in well-separated bands in a gel, RNA molecules of different sizes and/or with differences in the other parameters mentioned above can be analyzed in one and the same experiment. Although this feature might appear trivial, one should keep in mind that in other well-established physical methods such as spectroscopy, in particular optical melting curves or most hydrodynamic methods (e.g., analytical ultracentrifugation, light scattering), different molecules or different conformers always result in a superimposed experimental readout that can hardly be deconvoluted into the properties of the individual molecules or conformers.

TGGE can be applied to a wide variety of nucleic acids, including single- and double-stranded DNA and RNA that vary in size from oligonucleotides to the resolution limit of polyacrylamide gels, that is, a few thousand nucleotides; the most relevant range is between 100 and 1000 nucleotides. Different staining protocols can be used for the detection of the nucleic acid, silver staining being the most common one, but also very specific detection methods such as hybridization of a specific oligonucleotide probe to its target sequence as part of a crude nucleic acid preparation may be applied. The present chapter focuses on RNA analysis, predominantly single-stranded RNA and a few more specialized examples of RNA–RNA complexes and RNA–protein complexes. For a detailed description of other examples and applications of TGGE to DNA and protein analysis, the reader is referred to several previous book articles [1–3].

21.2

Method

21.2.1

Principle

As shown in Figure 21.1, a nucleic acid sample is loaded onto a slab polyacrylamide gel at the side of the negative electrode, using a slot that spans nearly the entire width of the gel. A linear temperature gradient is established perpendicular to the direction of the electric field. Molecules on the left side of the gel migrate at low temperatures, molecules on the right at high temperatures, and those at positions in between at corresponding intermediate temperatures. Each individual molecule, however, migrates at constant temperature during the whole electrophoretic run. Figure 21.1 shows a schematic example of an RNA that undergoes three cooperative conformational transitions: one from the native to a slightly altered conformation, a second to a partially denatured state, and a third transition to the completely denatured state. The transitions occur at well-defined denaturation temperatures $T_{m1} < T_{m2} < T_{m3}$. Immediately above T_{m1} , the molecules begin to migrate with slightly retarded mobility; the molecules at $T > T_{m2}$ migrate much slower than those at $T < T_{m2}$. In the temperature range of the first transition, the molecules switch between two states reversibly and assume a mobility that is averaged

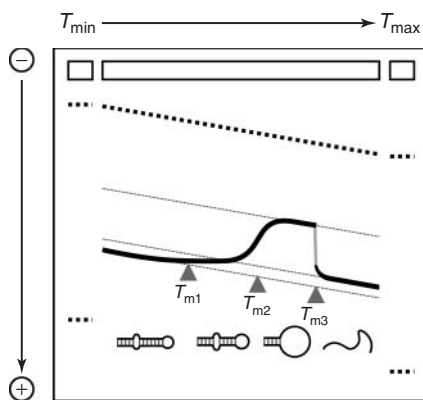


Figure 21.1 Schematic illustration of TGGE. A linear temperature gradient is applied perpendicular to the electric field. The sample is loaded into the broad central slot; small slots for marker loading are on the left and right side of the gel. In the given example, the mobility of an RNA with secondary structure is slightly decreased after a cooperative, reversible transition at T_{m1} , is drastically decreased after a further transition at T_{m2} , and is increased after the irreversible transition

at T_{m3} from a partially denatured molecule to the random-coil state without any base pairing. The secondary structure drawings below the transition curve symbolize the major structures before and after the different transitions. The stippled lines show the migration behavior of double-stranded nucleic acids that do not undergo any transition in the applied temperature range; their increase in mobility is due to a decrease in viscosity with increasing temperature.

according to the fraction of the states. The same holds for the second transition. The third transition is an example of an irreversible step: there is no thermodynamic equilibrium between the paired structure below T_{m3} and the completely denatured state above T_{m3} . The RNA denatures at T_{m3} , and as renaturation is impossible or extremely slow under the low salt condition of the gel electrophoresis, there is a jump in migration behavior associated with a discontinuity of the nucleic acid trace in between the two states.

Besides a gradient perpendicular to the electric field, one in parallel to the electric field can also be applied. In this case, the samples are loaded into narrow slots and run from low to high temperature. This mode of TGGE is applied mostly to the analysis of mutant RNAs. For example, mismatched heteroduplexes formed between wild-type and mutant sequences will give rise to retarded mobility owing to partial melting at lower temperatures than fully complementary homoduplexes [1, 2].

21.2.2

Instruments

The original instruments were homemade and the gradient was established by adjusting the temperature of the two edges of a thermostatic block with two independent liquid-based thermostats [4]. Such instruments worked quite well and were commercially available for some years. However, they are no longer on the market, so here the presently available system from Biometra (Göttingen, Germany) is described. To our knowledge, the only other supplier of a TGGE system, called *MicroTG*, is Taitec Corp. (Koshigaya City, Saitama, Japan; see http://e-taitec.com/products/products-detail-en.php?machine_name=Micro%20TG).

The Biometra TGGE system is a well-constructed instrument, which has been developed for routine use as well as for research. The microprocessor-driven gradient block on the basis of Peltier elements allows one to apply well-defined temperature gradients with good resolution and reproducibility. The Biometra TGGE system is available in two formats. The standard TGGE “mini” system operates small gels and is therefore ideally suited for fast initial experiments. The TGGE “maxi” system provides a large separation distance, allows high electrophoretic resolution and is well suited for systematic TGGE analyses of RNA.

21.2.3

Handling

Handling and protocols are intimately connected to the type of TGGE instrument. Thus, we refer to the detailed Manual (Version 09/10 TGGE MAXI System) from Biometra. The single steps are

- 1) casting of gels (assembly of the gel cuvette, preparation of gel and buffer solutions, gel pouring);

- 2) assembly of electrophoresis unit;
- 3) electrophoresis (sample loading, a pre-electrophoresis step after sample loading and before application of the temperature gradient, building up and equilibration of temperature gradient, electrophoresis);
- 4) RNA staining.

21.3

Optimization of Experimental Conditions

TGGE relies on the fact that the electrophoretic mobility of a biopolymer is altered because of a conformational change. The extent of this alteration in mobility critically depends on several parameters, such as pore size of the gel matrix, electric field, ionic strength, or presence of chaotropic substances. For a systematic study, it is essential to first optimize the basic parameters in order to observe mobility changes that are as pronounced as possible. Furthermore, a conformational transition may only be reversible under certain conditions. Therefore, the variables outlined below should be optimized, which has to be done primarily according to empirical rules.

21.3.1

Pore Size of the Gel Matrix

The pore size of the gel matrix can be varied by the concentration of polyacrylamide and the ratio of acrylamide to *N,N'*-methylenebisacrylamide (Bis). A compromise has to be found between a large change in electrophoretic mobility due to (partial) RNA denaturation and an acceptable migration velocity. For example, a concentration of 6% polyacrylamide might be too low because of only small visible changes in mobility, whereas 8% acrylamide may provide a pore size allowing good resolution of RNA conformers combined with still relatively fast electrophoretic mobility. The polyacrylamide concentration may be varied between 4% for larger RNA and up to 20% for short oligonucleotides, and the ratio of acrylamide/bisacrylamide between 20 : 1 and 40 : 1.

21.3.2

Electric Field

The electric field tends to stretch the molecules. These effects are very sensitive to the charge distribution, the conformation, and the flexibility of the molecule. The mobility of two conformers at a conformational transition may be differently affected by changes of the electric field, but in a hardly predictable manner. A strong electric field, although potentially increasing the mobility differences between two conformers, may also impair the overall TGGE profile (smeary RNA migration) and/or may decrease the resolution of other transitions due to molecule-stretching effects.

21.3.3

Ionic Strength and Urea

A variation of ionic strength always alters the T_m of a conformational transition and might result in complex changes of TGGE profiles. Therefore, this parameter cannot be varied independently from other features of TGGE. For nucleic acids, one may keep in mind as a general rule that changes in electrophoretic mobility are larger at low ionic strength. However, lowering ionic strength also shifts the conformational transitions to lower temperatures. The effect was found to be particularly large with circular RNA such as viroids (cf. Figure 21.2), where the change in mobility was reduced to 40% when the ionic strength was raised from 8.9 to 89 mM Tris-borate. With nucleic acids, high ionic strength always improves the reversibility of a transition, and reversible transitions may be evaluated more easily and more accurately than irreversible or discontinuous transitions. Some examples given below will illustrate this. The increase in transition temperature owing to an increase in ionic strength may be counteracted by addition of urea. For example, addition of 5–10 mM NaCl and 4–6 M urea to the standard electrophoresis buffer (8.9 mM Tris base, 8.9 mM boric acid) was found to be advantageous in terms of transition reversibility and band sharpness.

21.4

TGGE – General Interpretation Rules

In order to attribute secondary structures to transition curves obtained by TGGE, thermodynamic features as well as principles of gel-electrophoretic mobility have to be taken into account. While the transition temperatures may be calculated quite accurately, the interpretation of gel-electrophoretic mobilities has to rely more on qualitative criteria.

- 1) Branched structures migrate slower than extended structures. This effect is known from the denaturation of double-stranded nucleic acids, which leads to drastic retardation as long as the denaturation is incomplete [5]. The effect has also been described for dimeric transcripts of potato spindle tuber viroid (PSTVd) RNA [6].
- 2) Structures with large loops migrate extremely slowly. The low mobility of denatured circular viroids and the lower mobility of plasmids in the relaxed circle form, as compared to supercoils, are examples of this tendency.
- 3) Because of their higher molecular weight and their usually high degree of bifurcations, bimolecular complexes migrate much slower than the corresponding uncomplexed molecules [7–10].
- 4) Most RNA transitions appear as smooth curves; this is based on the reversible nature of the transitions. Irreversible transitions are recognized by a leap in gel mobility; irreversibility is primarily due to a low rate of renaturation under the low ionic strength conditions of gel electrophoresis.



Attempts have been made to derive quantitative relationships from the general rules [10, 11]. A set of optimal and suboptimal secondary structures was calculated for the RNA at several relevant temperatures T ; for each of these individual structures S_i , a retardation value $r(S_i) = \sum_j \sum_k r_j \cdot s_k$ summed up the retardation values r_j weighted by size s_k of all its structural elements of type $j \in \{\text{helix, internal loop, bulge loop, hairpin loop, junction, dangling end}\}$. The retardation of the structure ensemble in equilibrium was calculated by Boltzmann weighting $r(T) = \sum_i [r(S_i) \cdot p(S_i, T)] / \sum_i [p(S_i, T)]$ in dependence upon temperature T .

The relative mobility $\mu/\mu_0 = \exp(-\bar{p}(T)/L_r)$ of partly denatured, double-stranded nucleic acids has been calculated with fair success [12, 13]; μ_0 is the mobility of the completely double-stranded molecule with N base pair stacks; the retardation length L_r is a normalizing factor, which may be interpreted as proportional either to the length of the flexible unit of the nucleic acid or to the permeability of the gel; and $\bar{p}(T) = \sum_{i=1}^N p_i(T)$ is the sum of probabilities p_i of all base pair stacks in the denatured state, calculated according to [13–15]. This assumes that a change in mobility is independent of the position of the denatured bases; that is, the same effect on mobility is predicted for dangling ends as for internal loops.

21.5

Examples of TGGE Applications

We will illustrate the power of TGGE by four examples. All of them were produced with an apparatus commercially available from Qiagen GmbH (Hilden, Germany) up to about 1995 and which was very similar to the homemade instrument

Figure 21.2 TGGE analysis of a crude RNA extract from tomato plants infected by potato spindle tuber viroid (PSTVd). (a) Gel image. Conditions of electrophoresis: 0.2 x TAE (8 mM Tris-HCl, 20 mM NaOAc, 0.2 mM EDTA, pH 8.4), 5% (w/v) acrylamide, 0.125% (w/v) bisacrylamide (acrylamide/bisacrylamide 40 : 1), 0.1% (v/v) N,N,N',N' -tetra-methylethylenediamine (TEMED), and 0.04% (w/v) ammonium peroxydisulfate to initiate polymerization; sample loading was followed by pre-electrophoresis at 25 °C (300 V) for 10 min, application of temperature gradient (35–75 °C, equilibration for 10 min without current), and main electrophoresis for 90 min at 300 V; gel size: width 180 mm, length 190 mm, thickness 0.9 mm; slot size: width 130 mm, height 4 mm; RNAs were visualized

by silver staining; cPSTVd, lPSTVd: circular and linear forms of (+) strand PSTVd, respectively; 5S: 5S rRNA; (5S)₂: dimeric complex of 5S rRNA; 7S: 7S RNAs. Transitions described in the text are marked by numbered arrows. (b) Secondary structures of 5S rRNA, 7S RNA (GenBank AC Z29104), cPSTVd, and a sketch of the partially denatured two-hairpin structure of cPSTVd at the bottom. In the cPSTVd structure, the following regions are marked: T_L and T_R, terminal left and terminal right domain, respectively (cf. Figure 21.4); P, pathogenicity domain; C, conserved central domain; V, variable domain; gray lines marked I/I' and II/II' are the sequence stretches that form the stable hairpins I and II, respectively, at temperatures above the main transition (cf. Figure 21.3).

mentioned above [4]; the gel length in this apparatus was 190 mm. However, our examples should be reproducible on any TGGE apparatus after adapting the applied voltage to the length of the particular gel system in use.

Most of the following examples were obtained with PSTVd RNA (for reviews see [16–18]). PSTVd is a covalently closed RNA molecule of 359 nt which has no protein-coding capacity, but is able to infect certain plants. DNA-dependent RNA polymerase of the host plant replicates the PSTVd RNA genome by an asymmetric rolling circle mechanism; multimeric (+) strand intermediates are processed by plant RNases/ligases to monomeric mature circular molecules.

21.5.1

Example 1: Analysis of Different RNA Molecules in a Single TGGE

Optical melting curves of nucleic acids are based on structure-dependent extinction coefficients; consequently, any impurities in a sample as well as different conformations of a nucleic acid coexisting at the same temperature add up to the signal, leading to a non-interpretable transition curve. In TGGE, molecules are separated according to their hydrodynamic shape, as in any standard gel-electrophoretic method; that is, nucleic acid molecules of different length are separated according to their size, and molecules of identical length are separated on the basis of their differing conformations. Thus, sample impurities such as abortive transcripts derived from T7 transcriptions or residual DNA templates do not jeopardize the analysis of RNA denaturations by TGGE. For visualization, basically any gel-staining method can be used. While methods such as silver staining detect all kinds of nucleic acids, specific RNAs can be detected by oligonucleotide hybridization or double-stranded RNAs (dsRNAs) by dsRNA-specific antibodies.

An example of silver staining is shown in Figure 21.2a. A crude RNA extract, consisting of at least circular potato spindle tuber viroid (cPSTVd) and linear potato spindle tuber viroid (lPSTVd)(359 nt), 7S RNAs (about 305 nt; at least 14 variants with considerable sequence variation have been found in tomato [19]), 5S rRNA (120 nt), and tRNAs, was separated by TGGE and the RNA bands were silver-stained. At the lowest temperature of the gel, cPSTVd as well as lPSTVd adopt a rod-shaped secondary structure, which leads to relatively high gel mobility. At a defined temperature, about 57 °C under the applied electrophoresis conditions (see arrow 1 in Figure 21.2a), the native structure denatures completely and at least two stable hairpins are formed; the nucleotide regions forming these hairpins are indicated by bold characters in the cPSTVd secondary structure and are further illustrated by the sketch at the bottom of Figure 21.2b. The main transition (at 57 °C) is highly cooperative, leading to the sharp change in mobility at this temperature; mobility is decreasing because the branched hairpin-containing structure is much bulkier than the rod-like native structure. At higher temperature, the most stable hairpin II (HP II) starts to denature (see arrow 2 in Figure 21.2a); because fewer base pairs are involved, this transition takes place over a broader temperature

range.¹⁾ The completely denatured structure is an expanded circle (cPSTVd), which has strongly retarded mobility. lPSTVd molecules, which are either replication intermediates or derivatives of cPSTVd hydrolyzed during preparation of the RNA extract, migrate proportional to their length after full denaturation; completely denatured full-length lPSTVd migrates faster than fully denatured cPSTVd. The temperature of the main transition of lPSTVd depends on the point of linearization; because of the low concentration of individual lPSTVd molecules with defined 5'- and 3'-ends, these transitions are not visible on the gel.

Similarly, at low temperature bands corresponding to 7S RNA variants are not visible; only after full denaturation all 7S RNAs comigrate as a single distinct band owing to their very similar length. The 5S rRNA shows a single transition (see arrow 3). At the denaturation temperature of 5S rRNA, a dimeric 5S rRNA complex also dissociates. At high temperature, the band derived from dimeric 5S rRNA migrates slightly above that derived from monomeric 5S rRNA; the small migration difference corresponds to the migration distance of dimeric and monomeric 5S rRNA during the pre-electrophoresis step at low temperature before the temperature gradient was applied (for details, see legend to Figure 21.2a).

Note that all bands show an increase in mobility with increasing temperature. This general effect is not based on structural rearrangements, but is due to the decreasing buffer viscosity toward higher temperatures.

21.5.2

Example 2: Analysis of Structure Transitions in a Single RNA – Detection of Specific Structures by Oligonucleotide Hybridization

Structure formation of RNA is based on hydrogen bonds (base–base, base–ribose, ribose–ribose) and the thermodynamically favorable stacking interactions between neighboring bases and base pairs. While only highly specialized sequences favor the formation of a single structure, most sequences give rise to a structure distribution or ensemble of structures, where many different structures coexist in a thermodynamic equilibrium. Furthermore, structure distributions may arise from “sequential folding” based on the kinetics of folding: as local RNA structures form before RNA polymerase has synthesized the complete transcript, the RNA chain, immediately after synthesis, is folded into structures that may be thermodynamically suboptimal in the context of the full-length RNA molecule. These energetically unfavorable structures (which are metastable structures as they are overrepresented relative to thermodynamic equilibrium) can rearrange more or less rapidly to the equilibrium distribution provided that sufficient activation energy is available. Such structure distributions, either based on thermodynamics or kinetics, are often a problem in structure determination by chemical and enzymatic probing [20]. TGGE allows for separation of structure distributions, provided that the different structures neither interconvert rapidly nor comigrate.

1) The half-width $\Delta T_{1/2} = 4RT_m^2 / \Delta H^0$ of a cooperative transition is inversely proportional to the enthalpy ΔH^0 of the transition.

Figure 21.3 shows the analysis of sequential folding during synthesis of a (–) strand PSTVd transcript [20, 21]. Under thermodynamic equilibrium conditions, which can be established by complete denaturation and slow renaturation under low salt conditions to avoid formation of bimolecular complexes, the transcript forms a single dominant structure (Figure 21.3a). The native conformation is rod-like (Figure 21.3d, bottom) and exhibits a single main transition (Figure 21.3a), at which the rod-like structure denatures and a structure with the particularly stable hairpin HP II forms. The latter structure (Figure 21.3d, top) is bulky; when the temperature is further increased, additional helices and finally HP II melt, leading to an increase in mobility.

In vitro transcripts synthesized by T7 RNA polymerase may also be mixtures of several different structures (Figure 21.3b,c). The band of highest mobility, marked “S,” resembles the band obtained under thermodynamic equilibrium conditions as shown in Figure 21.3a. However, if analyzed directly after 30 s of transcription (Figure 21.3c), a thermodynamic equilibrium is not yet established and the rodlike structure makes up a minor fraction only within the population of different conformers. The other bands show lower mobility and therefore represent bulkier structures; they contain many thermodynamically unstable helices, and the slower migrating bands of the bulkier structures denature irreversibly at lower temperatures. When the T7 transcription reaction is extended from 30 s to 1 h, slow

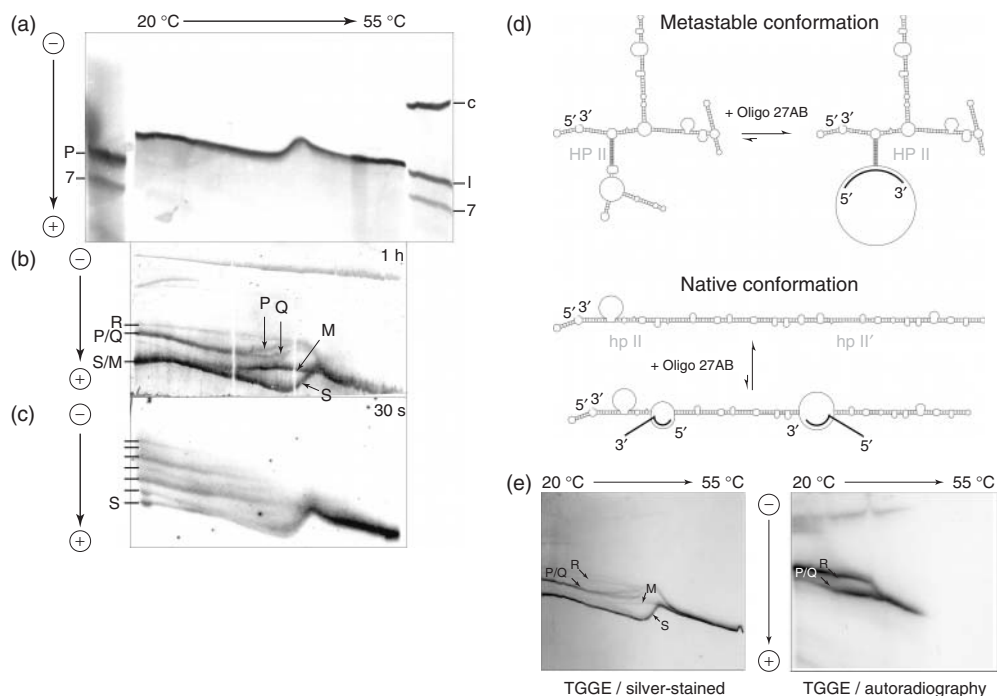


Figure 21.3 Detection of specific structures in a linear T7 transcript representing the first 337 nt of (–) strand PSTVd RNA (using a *Styl*-linearized DNA template) by silver staining or hybridization with a ^{32}P -labeled oligonucleotide. (a) TGGE of the transcript under equilibrium conditions. The transcript was denatured in 0.2 x TBE (17.8 mM Tris base, 17.8 mM borate, pH 8.4, 2.5 mM EDTA) at 70 °C for 15 min and slowly re-natured to room temperature with about 0.1 °C min^{–1}. Conditions of electrophoresis: 0.2 x TBE, 4% (w/v) acrylamide, 0.13% (w/v) bisacrylamide (acrylamide/bisacrylamide 30 : 1), 0.1% (v/v) TEMED, and 0.04% (w/v) ammonium peroxodisulfate to initiate the polymerization; after sample loading, 15 min pre-electrophoresis at 15 °C with 30 V cm^{–1}, 10 min equilibration for the 20–55 °C temperature gradient (without current), main electrophoresis at 30 V cm^{–1} for 120 min; gel size: width 180 mm, length 190 mm, thickness 0.2 mm; slot size: width 130 mm, height 4 mm; RNAs were visualized by silver staining. The marker lanes at the left and right contain a crude RNA extract (compare Figure 21.2) with 7 indicating 7 S RNA; P: PSTVd; L: linear PSTVd; c: circular PSTVd. Circular and linear PSTVd comigrate at low temperature because their shape is nearly identical; the completely denatured molecules migrate very differently because of the high retardation of the covalently closed circle. The synthetic PSTVd transcript has a length of 372 nt while the natural PSTVd has a length of 359 nt. (b,c) TGGE analysis of the transcript after different times of transcription/incubation. T7 transcription assays performed in the presence of 0.25 mM each NTP at 25 °C, corresponding to an elongation rate of about 130 nt s^{–1}, were stopped after 1 h (b) or 30 s (c). Conditions of electrophoresis were as in (a); detection of transcripts by electroblotting, hybridization with a digoxigenin-labeled probe, followed by immunodetection and colorimetry with 5-bromo-4-chloro-3-indolyl phosphate (BCIP) and 4-nitroblue tetrazolium salt (NBT) (b) or by autoradiography (c) after transcription in the presence of [α - ^{32}P]UTP; the total UTP concentration was identical to those of the other three NTPs. In (b),

the band at the top is the DNA template. For further details see [21]. (d) After synthesis by T7 RNA polymerase, the (–)-strand PSTVd transcript exists in different conformations due to sequential folding; one of these conformations, which is metastable, contains a long G–C-rich hairpin (HP11) that is thought to be critical for replication to (+) strands *in vivo*. In the native conformation, the two halves of the HP11 stem are involved in other base pairings at two distant locations. The oligonucleotide 27AB (oligo27AB; 5'-CUU ACU UGC UUC CUU UGC GCU GUC GCU-3'), complementary to the PSTVd (–) RNA sequence stretches 307–318 and 237–251, is designed to hybridize with its full length to the just mentioned nucleotide stretches that are juxtaposed in the loop enclosed by HP11 in the metastable conformation. In contrast, oligonucleotide 27AB binds to the native conformer at two different locations with only partial complementarity and thus with insufficiently weak affinity. (e) Hybridization of oligonucleotide was performed by incubating 200 ng of the *in vitro* transcript for 20 min in buffer (500 mM NaCl, 4 M urea, 1 mM sodium cacodylate, 0.1 mM EDTA, pH 7.0) with 10⁵ cpm oligonucleotide 27AB. After subsequent dialysis against 0.2 x TBE buffer, the TGGE analysis was performed. Conditions of electrophoresis: 0.2 x TBE, 5% (w/v) acrylamide, 0.17% (w/v) bisacrylamide (acrylamide/bisacrylamide 30 : 1); TEMED, ammonium peroxodisulfate, and electrophoresis conditions as in (a), except that the main electrophoresis was conducted for only 90 min. The gel was stained with silver (left image); additionally, an X-ray film (Kodak Xomat AR) was exposed to the gel (right image) for detection of hybrids between the transcript and ^{32}P -labeled oligonucleotide. On the silver-stained gel, several bands are detectable (S, M, Q/P, R), which represent different conformations of the same transcript; conformation S behaves as the transcript in the native conformation. In the autoradiograph of the same gel, only the RNA species Q/P and R are visible, as they expose a full-length target site for oligonucleotide 27AB (see (d), top). For further details see [20].

conformational rearrangements can equilibrate under the high salt conditions of the transcription buffer. As a result, structures of increased mobility and stability now prevail (Figure 21.3b *versus* c). Note that a single band in TGGE does not always represent a single structure; the bands marked “P/Q” and “S/M” comigrate at low temperature but separate at some temperature below the main transition.

The migration behavior of certain bands – that is, steepness of transitions, transition temperature, and mobility reduction or increase resulting from a transition – might be correlated with the results from structure predictions. For this purpose, the programs `mfold` [22] and `RNAfold` [23], which are based on thermodynamic equilibrium calculations (Chapter 26), or programs that consider folding kinetics or sequential folding pathways [24, 25], have been applied; in both cases, the ionic strength difference between gel electrophoresis and *in silico* calculations had been corrected for. A direct experimental confirmation of a structural model can then, for example, be obtained by “oligonucleotide mapping,” where an oligonucleotide is designed such that it hybridizes to its target sequence only in a specific conformational context. In the example shown in Figure 21.3d,e, an oligonucleotide (27AB) was designed to detect structures of (–) strand PSTVd transcripts that contain the hairpin structure HP11, which is thought to be critical for the synthesis of (+) strand replication intermediates *in vivo*. Oligonucleotide 27AB is able to simultaneously pair with both sequences protruding from the HP11 stem into the loop (Figure 21.3d, top). With structures containing HP11, the oligonucleotide pairs over its full length, which confers sufficient hybrid stability. In contrast, the oligonucleotide can only pair over half of its length with structures lacking HP11, such as the native rod-like structure (Figure 21.3d, bottom). Furthermore, the length of the oligonucleotide was carefully chosen to avoid shifting the mixture of existing structures toward more HP11-containing structures. In Figure 21.3e, a hybridization analysis of the structure distribution of the transcript (as in Figure 21.3b) using the radioactively labeled oligonucleotide 27AB is shown. Comparing the silver stain of the gel with its autoradiograph revealed that only structures in bands “P/Q” and “R” are marked by the oligonucleotide; these bands are visible in the autoradiograph up to a temperature at which the oligonucleotide dissociates. These bands, however, are not visible in the silver-stained gel because of the low concentration of the oligonucleotide–transcript complex. The assignment of bands was inferred from an overlay of autoradiograph and silver stain. Although the hybridized oligonucleotide had little effect on transcript migration in our case, one should be aware that the migration behavior of an oligonucleotide–RNA complex may differ from that of the uncomplexed RNA structure, considering that complex formation increases the molecular weight and may alter the hydrodynamic shape.

21.5.3

Example 3: Analysis of Mutants

Single nucleotide mutations might alter the structure and the structural stability of an RNA [26]. Visualization of such alterations by TGGE can be achieved either

by loading a mixture of mutant and wild-type RNA onto the gel or by loading them sequentially. The first method allows for detection of even subtle differences in hydrodynamic migration and/or thermodynamic stability, whereas the second makes it easier to identify the individual species.

An analysis based on sequential loading of three slightly different RNAs is shown in Figure 21.4. The terminal left domain (T_L) of PSTVd (Figure 21.2b) contains two repeats (marked in light and dark gray, respectively, in Figure 21.4), which are partially complementary to each other. This gives rise to two different structural arrangements of the T_L domain: a rod-like conformation, which is the prevalent conformation according to calculations, and a branched conformation. For an easy comparison of the different structures, the wild-type “native” structure was engineered by single base pair changes that drastically favor either the rod-like or the branched structure, respectively. In the “rod” mutant an A_{344} – U_{18} pair was changed to a G–C pair; as evident from inspection of the structures in Figure 21.4, this change stabilizes the rod-like conformation and destabilizes the branched conformation. In the “branched” mutant, an A_5 – U_{18} pair was changed to a G–C pair, which stabilizes the branched conformation and destabilizes the rod-like conformation. These specific mutant RNAs were analyzed experimentally by TGGE and silver staining (Figure 21.4a), and also by UV melting and NMR [27]. The stability of the “rod” mutant structure is increased in comparison to the “native” structure; note the increase in melting temperature and the slight decrease in half-width (corresponding to a slight increase in steepness of the transition), which is based on the increase in ΔH^0 . In contrast, the “branched” structure shows a much broader transition at relatively low temperature (filled arrowhead), which might originate from opening of one of the branch helices, overlapping with a second transition at higher temperature (open arrowhead in Figure 21.4a). TGGE, optical melting curves, and NMR experiments provided evidence for essentially identical conformations adopted by the “native” and “rod” mutant RNAs, thus verifying that the native conformation of PSTVd is indeed rod-like.

21.5.4

Example 4: Detection of Protein–RNA Complexes by TGGE

A relatively simple analysis of RNA–protein interaction is possible *via* gel retardation (gel shift) experiments. In contrast to DNA–protein interactions, RNA–protein interactions are usually determined by RNA structure rather than sequence. Thus, combining gel retardation with TGGE to simultaneously analyze the temperature dependence of RNA–protein complex formation and RNA structure may provide further insight into the nature of the RNA–protein interaction.

A corresponding analysis [10] of the interaction of the 5′ untranslated region (5′ UTR) of spinach chloroplast *psbA* mRNA, which encodes the D1 protein of photosystem II, and a stromal protein extract is shown in Figure 21.5. To establish complex formation, samples were incubated at 25 °C in the presence of heparin and tRNA to suppress non-specific RNA–protein interactions. In a gel shift assay

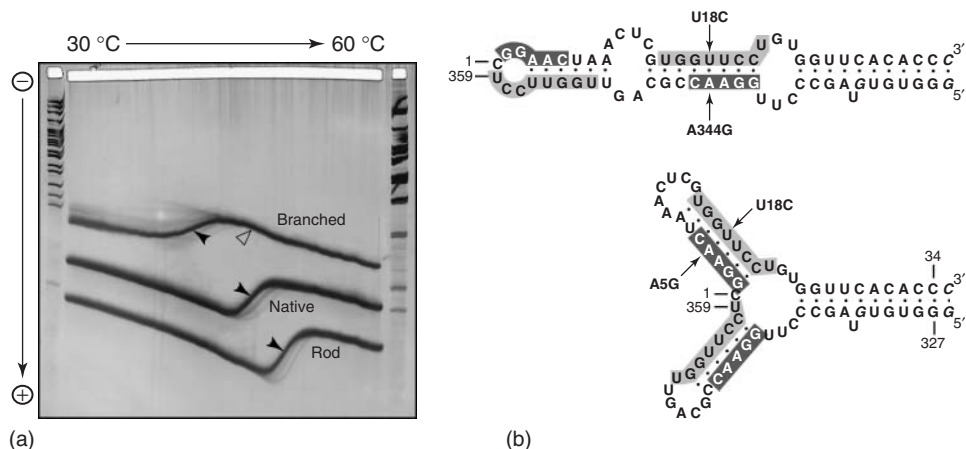


Figure 21.4 (a,b) Analysis of *in vitro* transcribed RNA variants derived from the T_L domain of PSTVd. The two major conformations of the “native” sequence are shown on the right; the numbering is that for circular (+) strand PSTVd; deviating from the native T_L domain sequence of PSTVd (Figure 21.2), the terminal helix of these short RNA transcripts was stabilized by addition of a terminal G–C base pair and a U₃₃₂G mutation (indicated by italic letters). In the “rod” mutant transcript, the rod-like conformation was stabilized by the two mutations U₁₈C and A₃₄₄G; in the “branched” mutant transcript, the branched conformation was stabilized by the two mutations A₅G and U₁₈C. Samples were loaded sequentially onto the gel in the order “rod”, “native”, and “branched” transcript. The very faint additional bands observed at the main transitions were attributed to (*n* + 1) transcripts that could not be removed completely during the purification procedure. The black arrowheads denote the main melting transitions; the open arrowhead denotes an additional transition in the “branched”

mutant. Conditions of electrophoresis: 0.2 x (v/v) TBE (17.8 mM Tris-HCl, 17.8 mM boric acid, 0.4 mM EDTA, pH 7.5), 5% (w/v) polyacrylamide, 0.17% (w/v) bisacrylamide (acrylamide/bisacrylamide 30 : 1); polymerization was started by adding TEMED to 0.1% (v/v) and ammonium peroxodisulfate to 0.04% (w/v). RNA samples (300 ng) were loaded into the broad sample slot (130 × 4 mm), while the small slots (5 × 4 mm) on both sides were used to load a DNA size marker (pBR322 digested with *Msp* I). Upon applying 500 V at a uniform temperature of 20 °C for 25 min, the RNA sample migrated a few millimeters into the gel matrix. This step was repeated for the loading of the second and third RNA samples, except that it was shortened to 10 min for the third sample. The DNA size marker was loaded immediately after the second RNA sample. Electrophoresis was paused for 15 min to apply and equilibrate the 30–60 °C gradient and then continued for 75 min at 500 V. The gel was stained with silver. For further details, see [26].

at 25 °C with 0.5 mg ml⁻¹ of protein extract, ~50% of the radioactively labeled RNA (0.5 pmol ml⁻¹) showed retarded mobility, and in the presence of 1 mg ml⁻¹ protein extract, the RNA was completely shifted into the complex (Figure 21.5a). Because the concentration of the RNA binding protein(s) in the extract was unknown, no binding constant could be estimated. The fraction of bound RNA, however, strongly depended on the temperature during incubation and electrophoresis: at 0.5 pmol RNA and 1 mg ml⁻¹ protein extract, no retardation of RNA mobility at all

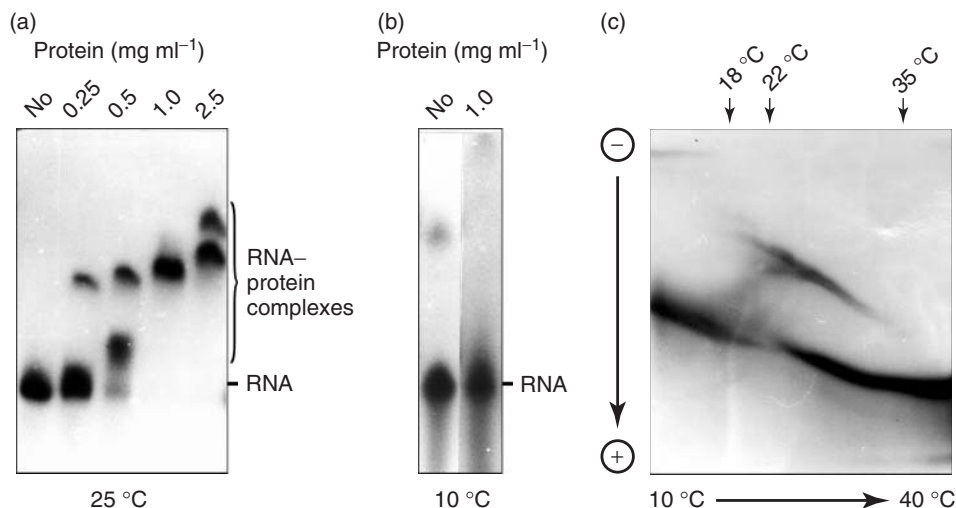


Figure 21.5 Analysis of protein(s) binding to a 95 nt transcript harboring the 5' untranslated region (5' UTR) of spinach chloroplast *psbA* mRNA. (a) Effect of protein amount on RNA–protein complex formation at 25 °C. Five fmol of radiolabeled RNA were incubated with varying amounts of a stromal protein extract for 15 min at 22 °C in a volume of 10 μ l containing 20 mM Tris-HCl, pH 8.5, 20 mM KCl, 10 mM MgCl₂, 5 mM dithiothreitol, 2 mg ml⁻¹ heparin and 0.5 mg ml⁻¹ tRNA. Free and complexed RNA were separated

on an 8% (w/v) polyacrylamide gel (acrylamide/bisacrylamide 20 : 1) in 2 x Laemmli buffer (50 mM Tris base, 0.38 M glycine) at 25 °C and 200 V for 75 min. (b) As in (a), but incubation and electrophoresis performed at 10 °C. (c) The radiolabeled *psbA* 5' UTR transcript (67 fmol) was combined with 260 μ g of protein (final volume 400 μ l) and directly incubated for 10 min in the gel slot (130 $\times$ 4 mm) in the presence of a temperature gradient from 10 to 40 °C. All other conditions were as in (a). For further details, see [10].

was observed at 10 °C, whereas complete complex formation was seen at 25 °C (cf. Figure 21.5a,b). To analyze the temperature dependence of complex formation in more detail, 67 fmol RNA was combined with 260 μ g protein in a total volume of 400 μ l and directly incubated for 10 min in the broad sample slot of a gel in the presence of a temperature gradient from 10 to 40 °C. After preincubation, the RNA was separated by electrophoresis and visualized by autoradiography (Figure 21.5c). The gel shows a distinct temperature range with retarded RNA mobility resulting from complex formation with protein(s). Retardation was observed between $\sim$ 18 and 35 °C, with a maximum at 22–25 °C. Additional experiments verified that it is neither the pH gradient produced by the temperature gradient, nor a protein conformational change (aggregation, precipitation) in the low temperature range, but RNA conformational changes near 22 and 33 °C that are the cause of the temperature dependence of complex formation. Thus, conformations of the RNA prevailing below 22 °C and above 33 °C are not adapted to protein binding [10]. In the temperature range above 20 °C, binding of protein to the RNA decreases with increasing temperature, suggesting that binding is driven by enthalpy.

21.5.5

Outlook

In the given examples, we focused on our own experiments with viroid RNA and an mRNA 5' UTR. Applications of TGGE by other groups, using similar experimental procedures as outlined here, can be found in [9, 28]. TGGE is not limited to the analysis of “exotic” RNAs such as viroids, but should be more generally viewed as a method that may replace or at least complement optical melting approaches (UV or Circular dichroism spectroscopy) in thermodynamic analyses of RNA structure, with the additional advantage to allow for analysis of RNA mixtures.

The Bevilacqua Lab [29, 30] used TGGE to select for thermodynamically more stable loop variants in combinatorial libraries of small RNA hairpins. They have mainly used a version of TGGE with the temperature gradient parallel to the direction of RNA migration. This allowed for an easy identification and isolation of molecules with structures more stable than the bulk of molecules, because at a certain temperature within the gradient, the already denatured molecules run slower than the still structured hairpins.

TGGE may also be used for the analysis of tertiary structures. Note, however, that Mg^{2+} ions, usually necessary for the stabilization of tertiary interactions, also catalyze the degradation of RNA at elevated temperatures, which exacerbates with increasing pH. The groups of Schroeder and Westhof [31] applied TGGE to monitor the tertiary structure transitions of the wild-type *td* intron of bacteriophage T4 and several mutants thereof. With two mutant RNAs loaded onto the same gel, similar to the experiment shown in Figure 21.4, small stability differences of the mutants could be detected, while simultaneously checking for deviating conformations and dimer formation.

Guo and Cech [32] searched for *Tetrahymena* ribozymes with enhanced activity at elevated temperature by *in vitro* evolution. Using TGGE, they selected for variants with the thermodynamically most stable (fast migrating) tertiary structures in the first step, and for activity in the second. In contrast to the selection procedure used by the Bevilacqua Lab [29, 30], Guo and Cech [32] used temperature gradients perpendicular to the migration direction and excised small rectangular regions close to the transition from the fully folded state (including the native tertiary interactions) to that with disrupted tertiary interactions. During eight rounds of selection, the transition temperature for tertiary structure unfolding increased from 45 to 52 °C in a buffer containing ~0.4 mM free Mg^{2+} ions. In six additional selection rounds at ~0.2 mM free Mg^{2+} ions, this transition was further shifted from 35 to 40 °C. Indeed, the final variants contained up to 11 mutations, which increased the maximum temperature compatible with catalytic activity by ~10 °C, mainly through strengthening of tertiary interactions.

As outlined in this chapter, TGGE is a versatile method for analyzing conformational transitions of single RNAs or multiple RNAs simultaneously, for studying the conformational heterogeneity of RNAs, to evaluate the structural effects of mutations, to characterize RNA–protein interactions, and for *in vitro* selection of stabilized RNA variants or even thermostable RNA–protein complexes.

We predict that TGGE will contribute to the understanding of many more RNA systems in the future.

References

1. Riesner, D., Henco, K., and Steger, G. (1991) in *Advances in Electrophoresis*, Vol. 4 (eds A. Chrambach, M.J. Dunn, and B.J. Radola), VCH Verlagsgesellschaft, Weinheim, 169–250.
2. Henco, K., Harders, H., Wiese, U., and Riesner, D. (1994) in *Methods in Molecular Biology*, Vol. 31 (ed. J.M. Walker), Humana Press, Clifton, NJ, 211–228.
3. Riesner, D. (1997) in *Antisense Technology: A Practical Approach*, Vol. 185 (eds C. Lichtenstein and W. Nellen), Oxford University Press, 1–24.
4. Rosenbaum, V. and Riesner, D. (1987) Temperature-gradient gel electrophoresis. Thermodynamic analysis of nucleic acids and proteins in purified form and in cellular extracts. *Biophys. Chem.*, **26**, 235–246.
5. Steger, G., Tien, P., Kaper, J., and Riesner, D. (1987) Double-stranded cucumovirus associated RNA 5: Which sequence variations may be detected by optical melting and temperature-gradient gel electrophoresis? *Nucleic Acids Res.*, **15**, 5085–5103.
6. Hecker, R., Zhi-min, W., Steger, G., and Riesner, D. (1988) Analysis of RNA structures by temperature-gradient gel electrophoresis: viroid replication and processing. *Gene*, **72**, 59–74.
7. Baumstark, T. and Riesner, D. (1995) Only one of four possible secondary structures of the central conserved region of potato spindle tuber viroid is a substrate for processing in a potato nuclear extract. *Nucleic Acids Res.*, **23**, 4246–4254.
8. Steger, G., Baumstark, T., Mörchen, M., Tabler, M., Tsagris, M., Sängner, H.L., and Riesner, D. (1992) Structural requirements for viroid processing by RNase T1. *J. Mol. Biol.*, **227**, 719–737.
9. Zhang, G., Zhang, J., George, A.T., Baumstark, T., and Simon, A.E. (2006) Conformational changes involved in initiation of minus-strand synthesis of a virus-associated RNA. *RNA*, **12**, 147–162.
10. Klaff, P., Mundt, S., and Steger, G. (1997) Complex formation of the spinach chloroplast *psbA* mRNA 5' untranslated region with proteins is dependent on the RNA structure. *RNA*, **3**, 1468–1479.
11. Hemmings-Mieszczak, M., Steger, G., and Hohn, T. (1998) Regulation of CaMV 35S RNA translation is mediated by a stable hairpin in the leader. *RNA*, **4**, 101–111.
12. Lerman, L.S., Fischer, S.G., Hurley, I., Silverstein, K., and Lumelsky, N. (1984) Sequence-determined DNA separations. *Ann. Rev. Biophys. Bioeng.*, **13**, 399–423.
13. Steger, G. (1994) Thermal denaturation of double-stranded nucleic acids: prediction of temperatures critical for gradient gel electrophoresis and polymerase chain reaction. *Nucleic Acids Res.*, **22**, 2760–2768.
14. Poland, D. (1974) Recursion relation generation of probability profiles for specific-sequence macromolecules with long-range correlations. *Biopolymers*, **13**, 1859–1871.
15. Fixman, M. and Freire, J.J. (1977) Theory of DNA melting curves. *Biopolymers*, **16**, 2693–2704.
16. Tabler, M. and Tsagris, M. (2004) Viroids: petite RNA pathogens with distinguished talents. *Trends Plant Sci.*, **9**, 339–348.
17. Hadidi, A., Flores, R., Randles, J.W., and Semancik, J.S. (eds) (2003) *Viroids*, CSIRO Publishing.
18. Schmitz, M. and Steger, G. (2007) Potato spindle tuber viroid (PSTVd). *Plant Viruses*, **1**, 106–115.
19. Riedel, L., Pütz, A., Hauser, M.T., Luckinger, R., Wassenegger, M., and Sängner, H.L. (1995) Characterization of the signal recognition particle (SRP)

- RNA population of tomato (*Lycopersicon esculentum*). *Plant Mol. Biol.*, **27**, 669–680.
20. Schröder, A.R.W. and Riesner, D. (2002) Detection and analysis of hairpin II, an essential metastable structural element in viroid replication intermediates. *Nucleic Acids Res.*, **30**, 3349–3359.
 21. Repsilber, D., Wiese, U., Rachen, M., Schröder, A.R., Riesner, D., and Steger, G. (1999) Formation of metastable RNA structures by sequential folding during transcription: time-resolved structural analysis of potato spindle tuber viroid (-)-stranded RNA by temperature-gradient gel electrophoresis. *RNA*, **5**, 574–584.
 22. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
 23. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.*, **36**, W70–W74.
 24. Schmitz, M. and Steger, G. (1996) Description of RNA folding by “simulated annealing”. *J. Mol. Biol.*, **255**, 254–266.
 25. Flamm, C., Fontana, W., Hofacker, I.L., and Schuster, P. (2000) Elementary step dynamics of RNA folding. *RNA*, **6**, 325–338.
 26. Dingley, A.J., Steger, G., Esters, B., Riesner, D., and Grzesiek, S. (2003) Structural characterization of the 69 nucleotide potato spindle tuber viroid left-terminal domain by NMR and thermodynamic analysis. *J. Mol. Biol.*, **334**, 751–767.
 27. Schuster, P., Fontana, W., Stadler, P.F., and Hofacker, I.L. (1994) From sequences to shapes and back: a case study in RNA secondary structures. *Proc. R. Soc. London, Ser. B Biol. Sci.*, **255**, 279–284.
 28. Owens, R.A. and Baumstark, T. (2007) Structural differences within the loop E motif imply alternative mechanisms of viroid processing. *RNA*, **13**, 824–834.
 29. Bevilacqua, J.M. and Bevilacqua, P.C. (1998) Thermodynamic analysis of an RNA combinatorial library contained in a short hairpin. *Biochemistry*, **37**, 15877–15884.
 30. Proctor, D.J., Schaak, J.E., Bevilacqua, J.M., Falzone, C.J., and Bevilacqua, P.C. (2002) Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions. *Biochemistry*, **41**, 12062–12075.
 31. Brion, P., Michel, F., Schroeder, R., and Westhof, E. (1999) Analysis of the cooperative thermal unfolding of the td intron of bacteriophage T₄. *Nucleic Acids Res.*, **27**, 2494–2502.
 32. Guo, F. and Cech, T.R. (2002) Evolution of *Tetrahymena* ribozyme mutants with increased structural stability. *Nature Struct. Biol.*, **9**, 855–861.

22

UV Melting Studies with RNA

Philippe Dumas, Eric Ennifar, Francois Disdier, and Philippe Walter

22.1

Introduction

Ultraviolet (UV) melting studies have been used for decades and have been the subject of a considerable number of reports. This is certainly the easiest method to obtain thermodynamic information on nucleic acids, in general, and on RNA, in particular. Alternative techniques, as those requiring the introduction of fluorescent probes [1], are not considered. Our goal is to recall some basic features in as simple a way as possible and to focus not only on well-known but also on much less well-known facts about this technique. Important practical problems are considered (buffers, evaporation, data processing, sloping baselines), and beyond those, we put emphasis on several pitfalls (overinterpretation of melting curves, lack of equilibrium, discerning unimolecular from bimolecular transitions).

It is significant that, in many cases, “old papers” will be referenced. As a matter of fact, given the impressive pace of emergence in the biological literature of new fields and new techniques, there is a strong propensity to ignore such “oldish” reports where the scientists collected their results on paper and made all calculations and figures by hand. As a result, in many instances, new users of this technique are not aware of many of these “well-known” old facts but sometimes rediscover them. Even on a rather well-delineated subject, as the one treated in this review, it is impossible to be exhaustive. Several other excellent reviews should be consulted [2–4]. For all general aspects around melting of nucleic acids, one should consult the thorough review by Turner [5]. Two papers are landmarks for the fundamental problem of the kinetics of helix formation from two strands [6, 7]. For general aspects on nucleic acids, see [8], and for a “reasonable” introduction to several physical problems, see [9].

22.2

A Simplified Account of the Physical Basis of UV Absorption

In this chapter, we are concerned with a spectroscopic method taking advantage of the interaction between UV light (in the wavelength range 180–300 nm) and the

delocalized π -electrons of the aromatic bases. As such, only a quantum mechanical approach would allow a full description of the underlying phenomena. However, there is no need for embarking on such considerations, and an outrageously simplified comparison will provide us with a useful intuitive view of these phenomena. The cloud of π -electrons, when perturbed by a periodic electromagnetic field, may be compared to a miniature ocean agitated by “electronic tides.” The interest of this comparison lies in its heuristic power.

First, the major components of these electronic tides are in the plane of the base. In the language of quantum theory, such movements correspond to the so-called $\pi - \pi^*$ transitions, the star being the mark of an excited π -orbital. These periodic collective movements give rise to oscillatory dipoles because of local accumulation and, concomitantly, local rarefaction of electrons [9, 10]. Second, for each base, there are particular frequencies of the excitation producing resonances that correspond to the maximum amplitude of the oscillatory dipoles. Finally, these oscillatory dipoles dissipate energy from the incoming electromagnetic field. In general, the electromagnetic energy is ultimately dissipated as heat, but with some aromatic systems (e.g., tryptophan in proteins, Y-base in tRNA, 2-amino purine), it may be first dissipated in part as fluorescence light at a longer wavelength [11]. Whatever the followed path, it is precisely this energy dissipation that manifests itself by absorption of the incoming UV light. It will not come as a surprise that larger amplitudes of the tide (in other words, larger amplitudes of the oscillating dipoles) produce more energy dissipation, and thus more absorption. Obviously, different aromatic systems, having different resonance characteristics, show different absorption spectra. Furthermore, any modification of the environment (pH, solutes, hydration, etc.), and any kind of interaction with other bases (base stacking, canonical base pairing, and non-canonical base pairing), is likely to influence the resonance characteristics of a given aromatic base (Section 22.4). As a consequence, depending on the environment, one may observe for each aromatic base a displacement of the absorption maxima and, more generally, a variation of absorption at a given wavelength. An increase and a decrease in absorption are referred to as *hyperchromism* and *hypochromism*, respectively. Such an effect is particularly interesting with nucleic acids because both the absorption of UV light by the bases around 260 nm and the hypochromic/hyperchromic effect are quite large.

22.3

Definitions and Nomenclature

For completeness, we recall basic definitions and the common nomenclature in use. The absorbance A , or optical density (OD), is defined as

$$A = \log_{10} \frac{I_0}{I} \quad (22.1)$$

where I_0 and I are the incident and transmitted light intensity, respectively. According to the Beer–Lambert law, the transmitted intensity through a length l

(in cm) is $I = I_0 10^{-\varepsilon l c}$, where ε is the so-called molar extinction coefficient (in $\text{M}^{-1} \text{cm}^{-1}$) of the absorbing molecules and c their concentration (in M). As a consequence, the absorbance is given by

$$A = \varepsilon l c \quad (22.2)$$

In agreement with the previous paragraph, ε is a function of not only the wavelength but also the temperature and all other parameters that can influence the amplitude of the oscillating dipoles. Beware of not being confused by ε values reported for concentrations in milligrams per milliliter, which is to be avoided for the sake of standardization (albeit this may be of practical interest, particularly for inhomogeneous samples). Most often, the optical path length l is 1 cm, and it is not mentioned explicitly. Such a sloppy convention should not be followed, and the path length should always be mentioned since various types of measuring cells exist. For strongly absorbing solutions, cells with 1-mm optical path can be used. Any cell must be made of quartz to avoid UV absorption.

22.4

Well-Known and Less Well-Known Characteristics of UV Absorption by Nucleic Acid Bases

The importance of UV absorption by an aromatic base around 260 nm is best appreciated by comparing it with the maximum possible absorption that would result if any photon passing through were absorbed. Elementary considerations [9] show that the observed ε values (about $1.4 \times 10^4 \text{ M}^{-1} \text{cm}^{-1}$ for purines and $1.0 \times 10^4 \text{ M}^{-1} \text{cm}^{-1}$ for pyrimidines) are about 10% of these maximum possible values, which is remarkably high. This means that the optical density (1-cm path, $\lambda = 260 \text{ nm}$) is 1 for an RNA or a DNA concentration as low as about $30\text{--}40 \mu\text{g ml}^{-1}$. This is only intended to yield an order of magnitude for practical purposes. It is important to mention that $\lambda = 260 \text{ nm}$ is the wavelength for which “normal” and AU-rich sequences have a maximum absorbance. However, for GC-rich sequences, the maximum is shifted close to 280 nm [2]. It may thus be useful to collect a full spectrum before performing a melting experiment, which may allow making a better choice for the wavelength than the usual 260 nm value. This sequence-dependent absorbance may also be of interest in particular circumstances to recognize which domain melts first within large RNA molecules [12].

Let us now consider in more detail the effect of temperature, which is the major experimental parameter in use. As it is well known, there is a strong hyperchromic effect (about 15–20%) upon base unstacking resulting from temperature increase (at $\lambda = 260 \text{ nm}$). This is easily explained, at least qualitatively, by a mutual counteracting influence of base stacking on the development of oscillating dipoles within each base. This is because more or less parallel dipoles borne by two bases in a stacking interaction, being forced to oscillate in phase, tend to accumulate partial charges of the same sign in close proximity. However, it is shown in the

following that significant effects arising from base solvation and base pairing, as well as simply from a change of wavelength, make the statement “temperature increase = absorbance increase” too simple.

First, significant hypochromic and hyperchromic effects also exist for mononucleoside monophosphates upon temperature variation (Fig. 1 in [13]). This work showed that an important decrease of absorbance (–12% for CMP at 265 nm) may result from a temperature increase. These far from minor effects were shown to arise from variations of base solvation with temperature, which correlates with the

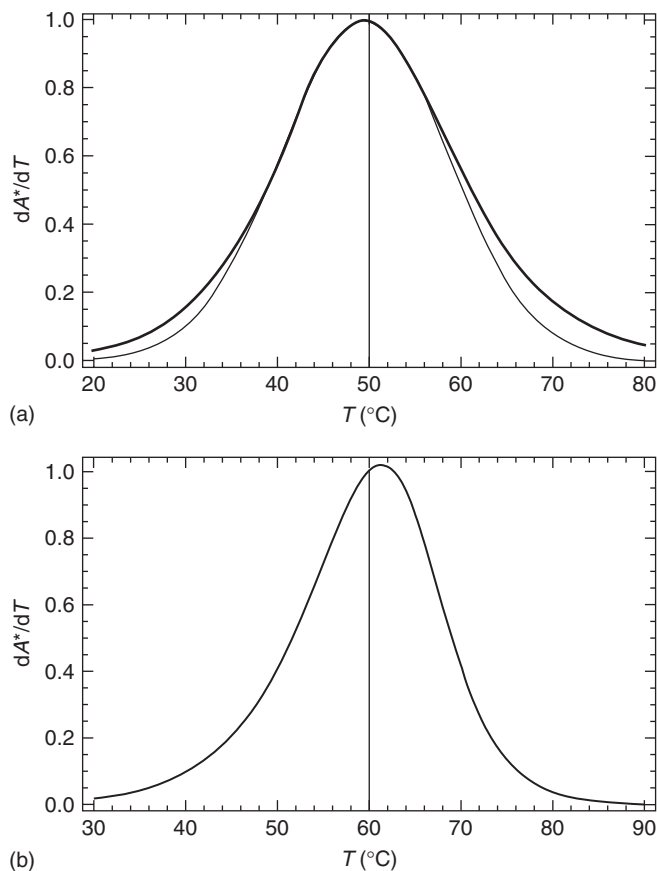


Figure 22.1 (a) Comparison of a theoretical DMC for a unimolecular process normalized in such a way that $dA^*/dT = 1$ for $T = T_{max}$ (Eq. 22.8, thick curve) with the best Gaussian approximation around $T = T_{max}$ (thin curve) to emphasize the very slight departure from exact symmetry. The curves were calculated with $T_m = 50$ °C (vertical line) and $h = -50$ ($\Delta H = -133$ kJ mol⁻¹). (b)

Theoretical DMC for a bimolecular process normalized in such a way that $dA^*/dT = 1$ for $T = T_m$ (Eq. 22.11). The curve was calculated with $T_m = 60$ °C (vertical line) and $h = -100$ ($\Delta H = -274.5$ kJ mol⁻¹). Note the clear asymmetry of the curve and the obvious difference between T_m and T_{max} ($T_{max} - T_m \approx 1.4$ °C).

strong dipolar moment attached to a water molecule. When as simple systems as dinucleoside monophosphates were further used to study the thermodynamics of base unstacking, relevant results could be obtained only after subtracting these solvation effects obtained separately for each nucleoside monophosphate [14]. Again, it clearly appeared with these dinucleoside monophosphates that a temperature increase is not always accompanied by an absorbance increase, even though base unstacking is taking place. A good example of the importance of keeping this in mind is furnished by a study on G-quartet formation, which was not performed at 260 nm because the hyperchromism was unusually low at that wavelength [15]. On the contrary, a much stronger hypochromism (50% absorbance *decrease* between 0 and 80 °C) was observed at 295 nm, which allowed obtaining good-quality melting curves.

Another effect is associated with partial disruption of Watson–Crick hydrogen bonds without significant base unstacking. Such an effect, known as *breathing*, can occur in the premelting temperature range when terminal base pairs oscillate between the paired and unpaired state, as inferable from hydrogen–deuterium exchange of imino protons [16, 17]. It was shown with thermal perturbation differential spectra on poly(G)-poly(C) and poly(G-C)-poly(C-G) (both polymers melt above 90 °C under the experimental conditions used) that a significant decrease of absorbance, and not an increase, occurred upon heating [18]. The hypochromism was as large as 3% for poly(G)-poly(C) and, importantly, the maximum effect occurred at 260 nm, the usual wavelength in most UV melting studies. Also noticeable is the sharp variation of this effect and its complete disappearance at 267 nm, followed by a hyperchromism beyond 267 nm.

These two effects, the former due to solvation variation with temperature and the latter due to Watson–Crick hydrogen bond breaking without base unstacking, imply that the exact function $\varepsilon(\lambda, T)$ for a given molecular species is rather complex and most often, if not always, unknown. It is thus precarious to draw conclusions from subtle melting curve variations (see Section 22.12 for details). For example, it may happen that a local minimum, instead of a monotonous increase, of absorbance is observed at moderate temperature: this is certainly insufficient evidence to conclude that this is because of a transient increase in base stacking.

Recently, temperature-dependent spectral variations were reconsidered systematically with DNA molecules and shown to contain robust signatures of their structures (duplex, triplex, quadruplex) [19]. We are not aware of an analogous systematic work for RNA, but particular examples for tRNAs can be found in [20].

22.5

The Basis of UV Melting Experiments for Thermodynamic Studies

Two simple hypotheses form the basis of UV melting experiments: (i) the two-state approximation is valid (Section 22.6) and (ii) thermodynamic equilibrium holds even though the temperature is variable (Section 22.7). Here, we suppose that these two hypotheses hold true and test the method in use with two important cases.

22.5.1

The Only Valid Definition of a Melting Temperature

Whatever the process being considered, there is no other definition of a melting temperature T_m than that derived from the equation below:

$$\Delta G = \Delta H - T_m \Delta S = 0 \quad (22.3a)$$

with the “Gibbs energies” (“*Gibbs energy*” is the term recommended by the International Union of Pure and Applied Chemistry (IUPAC) rather than the most commonly used “Gibbs free energy”) of the native and melted states being equal (i.e., $\Delta G = 0$) at T_m , and ΔH and ΔS being the variations in enthalpy and entropy, respectively, between the native and melted states. Therefore, *in all situations*,

$$T_m = \frac{\Delta H}{\Delta S} \quad (22.3b)$$

the ratio being calculated with ΔH and ΔS values corresponding to $\Delta G = 0$. One has to insist on that point because, quite often, it seems to be considered that Eq. 22.3b is only valid either for unimolecular transitions or for very sharp melting processes such as the one with long complementary DNA (or RNA) strands. Also, contrary to another common mistake, this is not equivalent to defining T_m as the inflection point of a differential melting curve (DMC). In order to emphasize the universality of Eq. 22.3b, it may be useful to recall that the temperature of any first-order phase transition obeys this relationship. For example, under a pressure of 1 atm, the heat necessary to melt crystalline ice at $T_m = 0^\circ\text{C}$ (or 273.15 K) is $\Delta H = 6.012 \text{ kJ mol}^{-1}$, which leads based on Eq. 22.3b to an increase of entropy ΔS upon melting of $6.012 \times 10^3 / 273.15 = 22.01 \text{ J mol}^{-1} \text{ K}^{-1}$. Note that Eq. 22.3b implies that at $T = T_m$, ΔH and ΔS must have the same sign (e.g., both ΔH and ΔS are positive for the transition of crystalline ice $\rightarrow$ liquid water at 0°C , as well as for any transition of the form hairpin or duplex $\rightarrow$ unfolded strands at $T = T_m$).

22.5.2

Reminders

In all theoretically sound formula, the temperature T must be expressed in Kelvin (K), which is related to the Celsius scale by $0^\circ\text{C} = 273.15 \text{ K}$. Also energies must be expressed in Joule (J), which is the SI unit. The universal molar “gas constant” R is equal to $8.31447 \text{ J K}^{-1} \text{ mol}^{-1}$. The unit “calorie” (cal), $1 \text{ cal} = 4.1868 \text{ J}$, is still widely used (against the IUPAC recommendations since it is not an SI unit). Because of such a wide use, it is of interest to recall that, by definition, one calorie is the heat necessary to increase the temperature of 1 g of pure water from 14.5 to 15.5°C under normal atmospheric pressure. We make use of it in this chapter only for a direct comparison with previous results. *Caveat*: mixing of different units is an endless source of mistakes! It is good to keep in mind the following orders of magnitude. The product RT , which is ubiquitous in thermodynamic expressions, varies from 2.3 to 3.1 kJ mol^{-1} between 0 and 100°C . The formation of one “average base pair” produces $\sim 35 \text{ kJ mol}^{-1}$ ($\sim 8.5 \text{ kcal mol}^{-1}$) of heat that

is released into the surrounding and that can be measured by a calorimeter. Such a process is termed *exothermic*. Conversely, the breaking of the same base pair absorbs the same amount of heat and is termed *endothermic*. At constant pressure, which may always be considered to be the case in usual laboratory experiments, such absorbed or released heat corresponds to the change of enthalpy ΔH . By convention, ΔH is counted negatively for an exothermic process and positively for an endothermic process. This must be respected strictly. Therefore, the ΔH accompanying the formation of an “average base pair” is about -35 kJ mol^{-1} .

22.5.3

Unimolecular Transitions

Let us consider, for example, the melting of a hairpin according to $H \rightarrow S$, where H and S represent the hairpin and single-stranded species, respectively. This unimolecular process is characterized by a dimensionless equilibrium constant for hairpin formation:

$$K = \frac{[H]}{[S]} = \frac{\theta}{1 - \theta} = e^{-\Delta G/RT} = e^{(\Delta S/R) - (\Delta H/RT)} \quad (22.4)$$

θ being the fraction of bases engaged in base pairs because of the formation of hairpins, and ΔH and ΔS stand for the molar enthalpy and molar entropy changes when going from the single-stranded to the hairpin state (ΔH and ΔS are thus negative). Solving Eq. 22.4 for θ yields¹⁾:

$$\theta(T) = \frac{K}{1 + K} = \frac{e^{\frac{\Delta H}{R} \left(\frac{1}{T_m} - \frac{1}{T} \right)}}{1 + e^{\frac{\Delta H}{R} \left(\frac{1}{T_m} - \frac{1}{T} \right)}} = \frac{1}{2} \left[1 + \tanh \frac{\Delta H}{2R} \left(\frac{1}{T_m} - \frac{1}{T} \right) \right] \quad (22.5)$$

At any stage of the melting process, the UV absorbance is given by

$$A = [\varepsilon_H \theta + \varepsilon_S (1 - \theta)] C_T = [\varepsilon_S + (\varepsilon_H - \varepsilon_S) \theta] C_T \quad (22.6)$$

with C_T being the total concentration in strand, and ε_H and ε_S standing for the molar UV absorbance of, respectively, hairpin and single-stranded species. Note that the optical path length l has been merged with ε_H and ε_S . In a first step, we make the simplifying assumption of ε_H and ε_S not varying with temperature (which is critically examined in the following paragraphs). This leads to the following expression for the DMC:

$$\frac{dA}{dT} = (\varepsilon_H - \varepsilon_S) C_T \frac{d\theta}{dT} = (\varepsilon_H - \varepsilon_S) C_T \frac{h T_m}{2T^2 \left[1 + \cosh h \left(1 - \frac{T_m}{T} \right) \right]} \quad (22.7)$$

with the reduced enthalpy $h = \Delta H/RT_m = \Delta S/R$. With $RT = 2.7 \text{ kJ mol}^{-1}$ at 50°C , this means that h lies commonly in the range of 40–120. To our knowledge, Eq. 22.7 has only been mentioned in the frame of high-resolution DNA melting

1) “Hyperbolic functions” appearing in that part are defined by $\tanh x = (e^x - e^{-x})/(e^x + e^{-x})$ and $\cosh x = (e^x + e^{-x})/2$.

studies [21] (see also the comments in that work about [22]). Normalizing this derivative to 1 at $T = T_m$ yields

$$\frac{dA^*}{dT} = \frac{2 \left(\frac{T_m}{T} \right)^2}{\left[1 + \cosh h \left(1 - \frac{T_m}{T} \right) \right]} = \frac{\left(\frac{T_m}{T} \right)^2}{\cosh^2 \frac{h}{2} \left(1 - \frac{T_m}{T} \right)} \quad (22.8)$$

where the star indicates the normalization. Because $T_m = T_{\max}$ is a very good approximation for such transitions (Section 22.5.4.4), Eq. 22.8 is directly applicable in practice. Although a DMC curve for a unimolecular process appears to have a Gaussian-like shape, it is not exactly symmetrical (Figure 22.1a).

22.5.4

Bimolecular Transitions

Such transitions are those involving, for example, the formation of a duplex or a loop–loop (or kissing) complex.

22.5.4.1 Entropic Considerations

The major difference to a unimolecular transition is that the overall change of entropy ΔS on melting does not only depend on the melting process itself but also on the total concentration of strands C_T : $\Delta S = \Delta S_{\text{melting}} + \Delta S_{\text{concentration}}$. In other words, once the two strands are separated, they become fully independent from one another. Equation (22.3b) thus implies $1/T_m = (\Delta S_{\text{melting}} + \Delta S_{\text{concentration}})/\Delta H$, where ΔH and $\Delta S_{\text{melting}}$ are concentration-independent, and the dependence of $\Delta S_{\text{concentration}}$ on C_T is given by the basic relationship $\Delta S_{\text{concentration}} = R \ln (C_T/C_T^0)$, C_T^0 being an arbitrary reference concentration. One thus obtains from Eq. 22.3b:

$$\frac{1}{T_m} = \frac{1}{T_m^0} + \frac{R}{\Delta H} \ln \frac{C_T}{C_T^0} \quad (22.9a)$$

T_m and T_m^0 corresponding to C_T and C_T^0 , respectively. Several important points should be stressed. First, one usually considers $C_T^0 = 1\text{M}$ as the reference state, which explains that the latter equation is somewhat sloppily written as $1/T_m = 1/T_m^0 + (R/\Delta H) \ln C_T$. This is sloppy in the sense that C_T , being the argument of a logarithm, should be read as the dimensionless quantity $(C_T/1\text{M})$. Second, Eq. 22.9a is valid for the melting of dimers made of identical self-complementary strands; for two complementary, but different strands, each at a total concentration $C_T/2$, there is another entropic factor related to the probability of encounter of two strands of different types in comparison to the probability of encounter of two identical strands at a total concentration C_T . Everything else being the same, the former probability is four times less than the latter, which is accounted for by the new equation:

$$\frac{1}{T_m} = \frac{1}{T_m^0} + \frac{R}{\Delta H} \ln \frac{C_T}{4C_T^0} \quad (22.9b)$$

Therefore, for a valid comparison of the melting temperatures of duplexes, one has to consider a fourfold excess of C_T for duplexes made of two different strands

Table 22.1 Comparison of results obtained with the different processing methods for the melting of a duplex made of two different complementary strands.

Method	ΔH (kcal mol ⁻¹)	T_m (°C)
Fit of DMC (no BLC)	-102.1 ± 0.8	65.2 ± 0.04
Fit of DMC (with BLC)	-102.4 ± 0.6	64.7 ± 0.025
Slope (with BLC)	-92.9	65.7
Half-width (no BLC)	-98.3	66.1
Half-width (with BLC)	-102.5	65.7
Concentration variation	-104.4	65.7
Fit of EHCC	-107.0 ± 1.4	73.7 ± 0.06
Integration of EHCC	-117.0	74.0

DMC, differential melting curve; BLC, baseline correction; EHCC, excess heat capacity curve from differential scanning calorimetry (DSC). Fit of DMC: processing of the DMC with Equation 22.11 modified for a correct normalization (see Appendix A). Slope: the slope method can only be used after baseline correction to obtain the f values. Integration of EHCC: result from [25].

The two last lines of the table correspond to ΔH_{cal} estimates and the other lines to ΔH_{VH} estimates. Note that the higher melting temperature obtained by DSC is due to a strand concentration much higher than that used for DMC. *Nota bene*, the results are given in the non-SI unit kcal mol⁻¹ for direct comparison with the original publication [25].

(heterodimers) *versus* duplexes made of identical strands with self-complementary sequences (homodimers).

A practically important consequence of Eqs. 22.9a,b is that varying the concentration C_T should induce a variation of T_m according to

$$\Delta \left(\frac{1}{T_m} \right) = \frac{R}{\Delta H} \Delta (\ln C_T) \quad (22.9c)$$

This allows estimating ΔH from experiments at different C_T . There are many examples of that in the literature. Such an estimate of ΔH , often referred to as ΔH_{VH} for *Van't Hoff* estimate, is said to be model-dependent since it relies on the two-state approximation (and also on this peculiar way of obtaining ΔH), contrary to its direct measurement as ΔH_{cal} by differential scanning calorimetry (DSC) or isothermal titration calorimetry (ITC) [23, 24] (note that ITC can only be used for a bimolecular transition involving heterodimers). A good example of excellent agreement between ΔH_{VH} and ΔH_{cal} is illustrated with the data from [25] reported in Table 22.1. Other examples of agreement are in [26–28]. One should be aware of the possibility of significant disagreements between ΔH_{VH} and ΔH_{cal} . This was the case, for example, with GTC repeats of sufficient length [29]. This is considered as an indication of a non-two-state character of the transition (Section 22.6).

As a conclusion of these considerations, it may be stated that if the “melting temperature” of a bimolecular process is not concentration-dependent, or substantially deviates from the expected temperature dependence, this means that either things are more complicated than for a simple bimolecular process, or that one is not at thermodynamic equilibrium (or both). This is illustrated in Section 22.7.

22.5.4.2 Basic and Less Basic Equations about Melting Curves Involving Bimolecular Transitions

The difference to a unimolecular transition is that the equilibrium constant K_c should be expressed as

$$K_c = \frac{[D]}{[S]^2} = \frac{\theta}{2(1-\theta)^2 C_T} \quad (22.10a)$$

for the dimerization of identical sequences and as

$$K_c = \frac{[D]}{[S_1][S_2]} = \frac{2\theta}{(1-\theta)^2 C_T} \quad (22.10b)$$

for the dimerization of different sequences. In these expressions, θ is as in Eq. 22.4, $[D]$ is the dimer concentration, and $C_T = 2[D] + [S]$ for Eq. 22.10a, or $C_T = 2[D] + [S_1] + [S_2]$ for Eq. 22.10b, is the total concentration of strands.

If we consider, in agreement with the previous entropic considerations, a fourfold excess in C_T for heterodimers *versus* homodimers, and assuming that the absorbance parameters do not vary with temperature, the following equation for the DMC normalized to 1 at $T = T_m$ is valid in the two situations:

$$\frac{dA^*}{dT} = \frac{3 \left[4 - e^{h\left(\frac{T_m}{T} - 1\right)} \left(\sqrt{1 + 8e^{h\left(1 - \frac{T_m}{T}\right)}} - 1 \right) \right]}{2\sqrt{1 + 8e^{h\left(1 - \frac{T_m}{T}\right)}}} \left(\frac{T_m}{T} \right)^2 \quad (22.11)$$

Note that the sign of $h = \Delta H/RT_m = \Delta S/R$ must be taken here as negative (which was without importance for Eq. 22.8). Importantly, the DMC is now significantly skewed (Figure 22.1b), contrary to the almost perfect symmetry in the previous case (Figure 22.1a). Such a difference is often overlooked, yet it allows recognizing at a glance a pure bimolecular transition and, more importantly, provides the basis to dismiss it when a DMC is not skewed (Section 22.8). Many erroneous conclusions exist in the literature because of ignoring this criterion. A theoretical curve analogous to that in Figure 22.1b was shown in [30], but with no mention of its equation. Particularly clean experimental examples of it are in [25] (examined in Section 22.5.4.5) and [28].

It is also worth mentioning that, at variance with unimolecular transitions, there is now a significant difference between T_m and T_{max} . Therefore, Eq. 22.11, unlike Eq. 22.8, is not directly usable in practice, because $T = T_m$ is not directly recognizable from mere inspection of the DMC. See Appendix A for obtaining a form normalized to 1 at $T = T_{max}$ and not at $T = T_m$. Also, for a given value of h , the derivative curve for bimolecular transitions extends farther on either side of $T = T_m$ than for monomolecular transitions (note that $h = -50$ in Figure 22.1a, and $h = -100$ in Figure 22.1b). One should thus be aware of introducing “sloping baselines” with caution (Section 22.11.2).

22.5.4.3 Higher Order Transitions

Transitions involving three, or more, molecules can also be studied. Expressions analogous to Eqs. 22.9a,b have been obtained [31]. They were shown to be of interest, for example, in the case of tetraplex formation from nonameric DNA fragments [32]. However, in many situations where a triplex or tetraplex species was formed, the hypothesis of equilibrium was far from being verified [33–37]. Therefore, these equations should be used with care. Also, be aware that a study on triplex formation may in fact involve dimer formation when the triplex results from complexation of a preformed hairpin with a single strand [38]. Interesting remarks and other references can be found in [3].

22.5.4.4 Influence of the Temperature Dependence of the Absorbance Parameters

The previous theoretical considerations assumed temperature-independent absorbance parameters associated with flat baselines. We now consider a more realistic situation by taking lower and upper baselines into account. Numerical tests showed that depending on the sign of the difference between the upper and lower slopes $\delta s = (dA/dt)_U - (dA/dt)_L$, the T_m obtained with the ideal equations without baseline correction is too high if $\delta s > 0$, or too small if $\delta s < 0$. For ΔH , the effect of temperature-dependent absorbance parameters is opposite and $|\Delta H|$, particularly with the “half-width method” (see next section), is too low if $\delta s > 0$ or too high if $\delta s < 0$. An example of this is given in Table 22.1 for duplex melting corresponding to $\delta s > 0$ with an increase of T_m by $\sim 0.5^\circ\text{C}$ (65.2 vs. 64.7°C) and a decrease of $|\Delta H|$ by $4.2 \text{ kcal mol}^{-1}$ (-98.3 vs. $-102.5 \text{ kcal mol}^{-1}$). Because $\Delta S = \Delta H/T_m$, the two opposite effects add up and $|\Delta S|$ behaves as $|\Delta H|$.

22.5.4.5 The Different Ways of Obtaining T_m , ΔH , and ΔS

We focus our attention on simple systems such as those corresponding to hairpin or duplex melting, and we assume that the hypothesis of thermodynamic equilibrium is valid. As seen above, ΔS is always obtained from Eq. 22.3b. Therefore, the question is how to obtain T_m and ΔH ? A first estimate of T_m can readily be obtained with the approximation $T_m \approx T_{\text{max}}$. For ΔH , a distinction must be made between unimolecular and bimolecular processes, since for the latter, there is the possibility to determine ΔH by varying the concentration of strands (Eq. 22.9c). How can the problem be solved by using a single melting curve?

The Slope Method Examination of Eqs. 22.8 and 22.11 reveals that, in both cases, the temperature dependence is governed by a global multiplicative term $(T_m/T)^2$ and different terms involving $h(T_m/T) = \Delta H/RT$ as the argument. Therefore, the higher $|\Delta H|$, the more important is the variation of the argument with the temperature and, in turn, the more abrupt is the temperature variation of the melting curve. In other words, increasing $|\Delta H|$ makes a transition more cooperative. This is a very general fact that is in no way limited to UV melting studies [39]. In agreement with this observation, one may derive ΔH from the slope df/dT of the melting curve at $T = T_m$, $f = 1 - \theta$ being the fraction of melted base pairs (θ as in Eq. 22.4). Note that this supposes that any upper and lower baseline

has been properly taken into account (Section 22.11.2 and [3]). A general formula is

$$\Delta H = -2(1 + N_s) RT_m^2 \left(\frac{df}{dT} \right)_{T=T_m} \quad (22.12)$$

where N_s is the number of strands ($N_s = 1$ for a hairpin and $N_s = 2$ for a duplex), and the sign is for native state formation [2, 31]. In theory, this is a rigorous expression. However, as previously noticed, this should be used with caution for $N_s > 2$.

The Half-Width Method Another method involving the measurement of the DMC half-width yields [30]

$$\Delta H \approx \frac{-\alpha}{\frac{1}{T_{\max}} - \frac{1}{T_{3/4}}} \quad (22.13)$$

where $T_{\max}$ (noted $T_{1/2}$ in [30]) has the same meaning as before, and $T_{3/4} > T_{\max}$ is the superior temperature for which the DMC has half its maximum value, that is, 1/2 when the DMC is normalized to 1 at $T = T_{\max}$ (see example in Figure 22.3). The values for α are $14.583 \text{ kJ mol}^{-1} \text{ K}^{-1}$ ($3.483 \text{ kcal mol}^{-1} \text{ K}^{-1}$) for $N_s = 1$ and $18.493 \text{ kJ mol}^{-1} \text{ K}^{-1}$ ($4.417 \text{ kcal mol}^{-1} \text{ K}^{-1}$) for $N_s = 2$. These values were carefully redetermined by numerical calculations with synthetic data and are slightly different from those in the original publication that, for theoretical reasons, used derivatives of melting curves *versus* $1/T$, and not *versus* T as usual in practice. Note that great accuracy in the small difference $\Delta T = T_{3/4} - T_{\max}$ is required because $\delta \Delta H / \Delta H = -\delta \Delta T / \Delta T$.

The Fitting, or Shape, Method Another solution is to fit the normalized DMC with the proper theoretical function: Eq. 22.8 for $N_s = 1$ and Eq. 22.11 modified for a correct normalization for $N_s = 2$ (Figure 22.2). Three remarks have to be made. First,

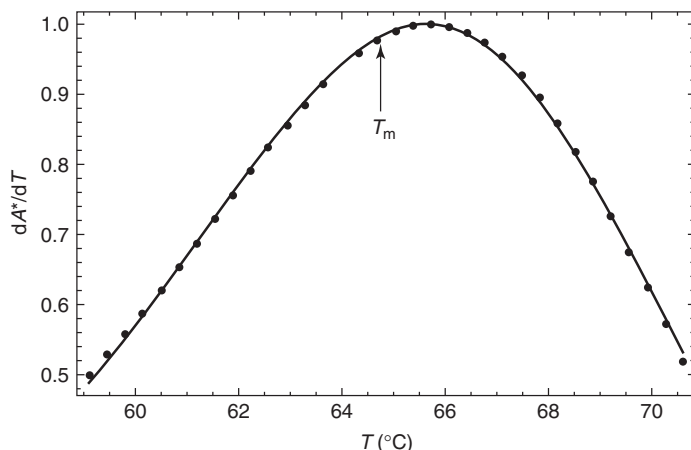


Figure 22.2 Fit of the experimental DMC data from [25] (dots) to Eq. 22.11 modified for a normalization to 1 at $T_{\max} > T_m$ (Appendix A). Such a fit can be performed on our web site [40]. The results of the fit are listed in Table 22.1.

one does not rely on the slope of the DMC at $T = T_m$, as with Eq. 22.12, but on its curvature. Second, at variance with Eq. 22.13, there is no need of using an important extension on both sides of $T = T_m$ to obtain a good estimate of this curvature, which minimizes the influence of error in the baseline correction. Finally, this fitting procedure also yields the best estimate of T_m , whereas the two other methods rely on $T_m \approx T_{\max}$, which is a good approximation only for unimolecular transitions.

Results Since all these ΔH values depend on the method in use, they are said to be “model-dependent”, unlike experimental values obtained by DSC or by ITC [23, 24]. As an example of such model-dependent treatments with data from *bona fide* duplex melting, we consider a 13-bp DNA duplex made of two different complementary strands [25]. First, the fact that this was DNA, and not RNA, suppressed all problems due to degradation, which is favorable for assessing and comparing the methods. Second, the two strands can only form unstable intramolecular hairpins that do not compete with duplex formation (compare with the HIV-1 DIS RNA as a good example of the contrary, Section 22.8). Third, data from DSC are also available, which allows a comparison to be made with UV melting data. In order to compare all processing methods, the experimental data were carefully retrieved from the figures in the publication [25]. The results are summarized in Table 22.1. Of course, they only differ marginally from those in the original work (when available). One first notices that the DSC and the UV melting curves compare remarkably well (Figure 22.3), which is a good mutual support for

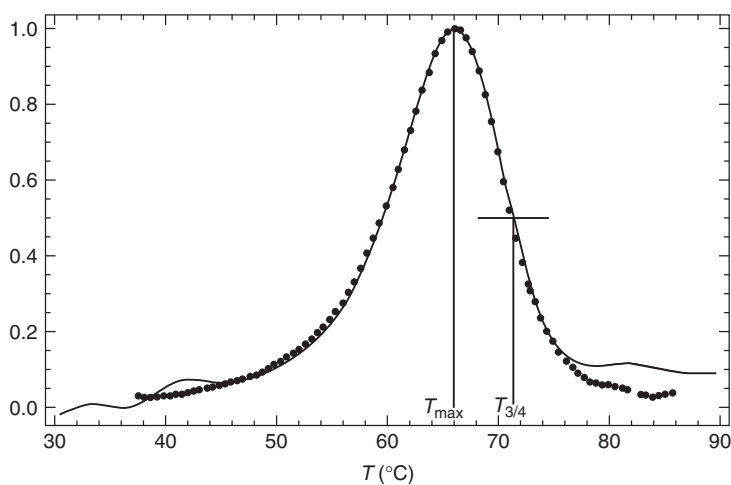


Figure 22.3 Comparison of an excess heat capacity curve from DSC (dots) with a DMC (continuous curve). (Data from Ref. [25].) The two curves were normalized for a maximum value equal to 1. The DSC curve was displaced by -9°C to take into account the difference in strand concentrations in the two experimental situations. To illustrate the

half-width method (Section 22.5.4.5), the temperatures $T_{\max}$ and $T_{3/4}$ necessary for Eq. 22.13 are highlighted. The small horizontal segment defining $T_{3/4}$ lies at 0.5. *Nota bene:* $T_{\max} \neq T_m$ and $T_{\max}$ and $T_{3/4}$ must be given in Kelvin: $T_{\max} = (66 + 273.15) \text{ K}$ and $T_{3/4} = (\sim 71 + 273.15) \text{ K}$.

the two methods. The data from DSC were processed by fitting the excess heat capacity curve, which yielded $\Delta H_{\text{DSC}} = (-107 \pm 1.4) \text{ kcal mol}^{-1}$, a value somewhat below that derived from integration of the same curve $\Delta H = -117 \text{ kcal mol}^{-1}$ [25]. The processing of the DMC (Figure 22.2) yielded $\Delta H_{\text{DMC}} = (-102.4 \pm 0.6) \text{ kcal mol}^{-1}$ in reasonable agreement with the analogous result from DSC. However, there is a slight difference $\Delta\Delta H = \Delta H_{\text{DSC}} - \Delta H_{\text{DMC}} = (-4.6 \pm 1.5) \text{ kcal mol}^{-1}$ between these two values obtained with the same fitting method but at different melting temperatures (Table 22.1). It is shown in the next section that this difference is related to a genuine variation of ΔH with the temperature related to a $\Delta C_p = \partial\Delta H/\partial T \neq 0$, as discussed in [23].

Now, considering the two sets of values ($\Delta H_{\text{DSC}} = -107.0 \text{ kcal mol}^{-1}$, $T_m = 73.7^\circ\text{C}$; $\Delta H_{\text{DMC}} = -102.4 \text{ kcal mol}^{-1}$, $T_m = 64.7^\circ\text{C}$; Table 22.1), corresponding to $465 \mu\text{M}$ in strands for the DSC, and $7.69 \mu\text{M}$ in strands for the DMC, one obtains $\Delta H = -104.4 \text{ kcal mol}^{-1}$ with the method exploiting the variation of T_m with strand concentration (Eq. 22.9c). The latter ΔH value, which can only be some average between the two extreme concentrations $465 \mu\text{M}$ and $7.69 \mu\text{M}$, indeed lies in between -107.0 and $-102.4 \text{ kcal mol}^{-1}$. There is thus a good general agreement among these methods, except for the “slope method” which is likely less reliable. The significantly lower $|\Delta H|$ value obtained with it ($92.9 \text{ kcal mol}^{-1}$ vs. 102.4 , 102.5 , and $104.4 \text{ kcal mol}^{-1}$) implies that the slope of the melting curve at $T = T_m$ was underestimated, which means that the baseline correction was not optimal even though the baselines seemed correct. On the contrary, depending on whether or not a baseline correction was made, the fitting method gave virtually the same ΔH estimates (and a shift in T_m by 0.5°C , which is small but significant in view of the estimated standard deviations, see Section 22.5.4.4). In conclusion, the procedure involving the fitting of the DMC seems better since it appears to be less dependent on baseline correction. One should be aware that not all experimental systems are amenable to such a simple and efficient treatment (see Section 22.8).

Consequence for the Determination of $\Delta C_p = \partial\Delta H/\partial T$ As just noticed, a $\Delta\Delta H = \Delta H_{\text{DSC}} - \Delta H_{\text{DMC}} = (-4.6 \pm 1.5) \text{ kcal mol}^{-1}$ and a $\Delta T_m = T_m^{\text{DSC}} - T_m^{\text{DMC}} = 9^\circ\text{C}$ were observed between the DSC and DMC estimates (Table 22.1, -107 vs. $-102.4 \text{ kcal mol}^{-1}$ and 64.7 vs. 73.7°C). Interestingly, this $\Delta\Delta H \neq 0$ did not result from two different techniques having been used, but is the consequence of $\Delta C_p = \partial\Delta H/\partial T \neq 0$ and of the measurement of ΔH at two different T_m values. At this point, a subtlety has to be stressed. Indeed, with this change of T_m resulting from a variation in strand concentration, it is legitimate to approximate the partial derivative $\partial\Delta H/\partial T = \Delta C_p$ with the full derivative $d\Delta H/dT_m \approx \Delta\Delta H/\Delta T_m$. Beware! This would not be valid if the change in T_m had been obtained by changing the ionic strength (see Appendix C for additional comments). With $\Delta T_m = 9^\circ\text{C}$ and $\Delta\Delta H = -4.6 \text{ kcal mol}^{-1}$, a $\Delta C_p \approx (-510 \pm 170) \text{ cal mol}^{-1} \text{ K}^{-1}$ is obtained, in reasonable agreement with ΔC_p values obtained within a systematic study [28]. Therefore, if a sufficient change in T_m and ΔH can be obtained through concentration variation (which is thus only applicable to a bimolecular transition), the ΔC_p can be calculated. This obviously requires to use the “fitting or shape method”,

since this is the only method yielding T_m and ΔH estimates from a single melting curve. Experimentally, a sufficiently large variation in total strand concentration C_T should be achieved. Toward low concentrations, the limitation will arise from a too low signal-to-noise ratio, and toward high concentration, it will arise from a lack of linearity of the spectrophotometer (it is safe to respect $OD < 1$). However, using cells with 1-mm optical path and/or lowering the molar extinction coefficient by changing the wavelength will allow increasing the concentration. Note that the result obtained with the two extreme concentrations should better be confirmed with a third experiment at an intermediate concentration.

22.6

The Two-State Approximation and Its Limitations

By “two-state approximation,” it is meant that a system (e.g., a duplex) is considered to be either fully in the paired form or fully melted. Intuitively, however, one easily admits that melting is a progressive process that cannot be exactly represented by such an approximation. For example, the ends of a helix will more easily melt than the central part of the same helix. Such a melting “from the ends” is named fraying. It is, therefore, strange that in many situations, this two-state approximation works rather well. In fact, such a success should not be viewed as proof for an unexpectedly large range of its validity. The reason is that, in usual situations, it is always possible to describe a gradual process by invoking an ideal two-state process. Indeed, as far as the evolution of absorbance is concerned, there is generally no way to differentiate a situation where all molecules have, on an average, 20% of their base pairs melted from a situation where 20% of the molecules are fully melted and 80% are not melted at all. However, there are situations where the differentiation is possible. For example, a duplex made of A-U base pairs in one-half and G-C base pairs in the other half would likely show two separate melting events and the two-state approximation would definitely not be valid. Such considerations are at the basis of many attempts to interpret melting curves of large RNA molecules assumed to melt domain by domain. The possible limitations and successes of such attempts are examined in Section 22.9.

22.7

Equilibrium and Non-equilibrium

It is extremely important that thermodynamic equilibrium is not compromised during temperature variation for a safe processing of results in Section 22.5. There are enough examples of departure from equilibrium to consider the problem seriously. Well-recognized cases of non-equilibrium are those observed in triplex [35] or tetraplex studies [33, 36, 41]. In such instances, one has to undertake a kinetic analysis as in [35] by considering the differential equations governing the evolution of the concentrations with time. This was first performed long time ago for the kinetics of association of poly(A) and poly(U) [42]. A serious error would

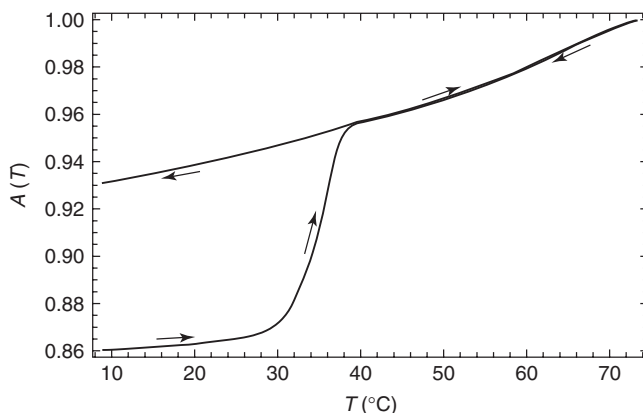


Figure 22.4 Illustration of a huge hysteresis arising from lack of equilibrium with a simple dodecamer duplex (7.6 μM in strands) at low salt conditions (10 mM NaCl). The heating rate was $0.3^\circ\text{C min}^{-1}$. $A(T)$ corresponds to the raw experimental values.

Note the very good superimposition of the ascending and descending branches above 40°C . This marks the absence of any evaporation at high temperature because of the setup shown in Figure 22.5 and described in Section 22.11.1.

be to assume that lack of equilibrium is restricted to long polymers, or to triplexes or quadruplexes. There are many examples of departure from equilibrium at low temperature even for simple duplexes [43–45].

Importantly, low salt conditions are very efficient at inducing departure from equilibrium because the rate of strand association, but not of strand dissociation, may become extremely slow. This is well illustrated by the huge hysteresis observed with a system as simple as the “Dickerson dodecamer” (dCGCGAATTCGCG) in 10 mM NaCl (Figure 22.4). To demonstrate hysteresis, we deliberately prepared the DNA to be close to 100% in its duplex form (slow cooling at high strand concentration in 1 M salt, followed by dialysis to remove the salt). Another evidence of lack of equilibrium is that the temperature at mid-transition varied very slightly with strand concentration, and in a way opposite to that expected from Eq. 22.9c. Rather, the temperature at mid-transition was significantly displaced toward lower/higher temperatures upon decreasing/increasing the heating rate (Section 22.12.3). Conclusion: this temperature at mid-transition is *not* the duplex melting temperature defined by Eq. 22.3b and ignoring that would yield totally erroneous results. The review by Mergny and Lacroix [3] should be consulted about these kinetic problems that would deserve a full chapter for themselves.

22.8

A Common Pitfall with Self-Complementary Sequences

It is a common mistake to think of a duplex made of two self-complementary strands as being intrinsically more stable than the related hairpin. This fallacious intuition results from the fact that there is more than twice the number of base

pairs in the duplex than in the hairpin form. However, Eq. 22.3b reminds us that the enthalpic term (the “strength of the glue”) has to compete with the entropic term (the counter influence of disorder). For example, a single strand “can attempt at closing on itself,” so to speak, at a very high frequency, which leads to a high effective hairpin closure rate (roughly 10^4 s^{-1}) [46]. In comparison, the concentration-dependent frequency of encounter of two separate strands is much lower and there are also many non-productive encounters, which leads to effective duplex formation at a much lower rate (roughly 1 s^{-1} at $1 \mu\text{M}$ in strands) [5]. As a consequence, the T_m of a hairpin may be higher than that of the related duplex. Of course, the exact outcome of such a comparison depends on several factors. The first factor is the loop size: increasing the loop length to beyond 4 nt makes hairpin closure less efficient [46, 47]. The second factor is the exact loop sequence: see [48] for A4, T4, G4, and C4 tetraloops and [49] for the particularly stable UNCG, GNRA, and CUUG tetraloops. The third factor is the ionic strength: a higher ionic strength favors more duplex formation than hairpin closure. The fourth factor is strand concentration: increasing strand concentration favors the duplex form. Quite often, the two T_m values are close to one another, and this may be misleading. A good example for this is the dimerization initiation site (DIS) of HIV-1 genomic RNA, which slightly varies in sequence depending on viral subtype. All sequences, however, allow duplex and hairpin formation, two hairpin loops being able to form a kissing complex (a natural consequence of the almost perfect self-complementarity of the 23-mers) [50, 51]. As explained in [52], the so-called Lai sequence (5'-CUUGCUGAAGCGCGCACGGCAAG-3') led to an inconsistency because the DMC showed a fully symmetric hairpin-like peak being concentration dependent ($T_m = 70^\circ\text{C}$ at $2 \mu\text{M}$ in strands). In contrast, the so-called Mal sequence (5'-CUUGCUGAGGUGCACACAGCAAG-3') also showed a fully symmetric hairpin-like peak being, as it should, independent of concentration ($T_m = 72^\circ\text{C}$). It was convincingly shown by using mutants that there is almost simultaneous melting of both the duplex and the hairpin forms for the Lai sequence (which explains the concentration dependence), whereas the Mal hairpin melts at a slightly higher temperature (which explains that it is not perturbed by duplex melting).

Therefore, one can conclude that self-complementary sequences often lead to mixed structures and thus misleading results because of the competition between hairpin and duplex formation, whereas only two complementary, but different, sequences that are unable to form stable hairpins give rise to pure duplex melting as seen in Figures 22.1b, 22.2, and 22.3.

22.9

Extracting Thermodynamic Information from Melting Curves of Large RNAs

Among the first attempts to extract information for large RNA molecules were those with tRNAs [20, 53], particularly tRNA^{Phe} [54]. Interestingly, UV melting analyses were performed soon after DSC had been used [55], and there was excellent agreement between results from the two approaches. Although the melting curves showed at most two distinct peaks, it was attempted to decompose them into five

distinct transitions. In this context, the authors of the DSC study stated “*it has to be admitted that the calorimetric studies alone cannot provide a unique and unambiguous assignment of the peaks to structural domains of the tRNA.*” Since the excess heat capacity curve showed somewhat more details than the DMC, this cautionary remark also applies to the latter. This clearly means that one should not play down the real danger of overinterpreting experimental data that do not contain enough information.

An important goal of such studies on large RNA is to distinguish tertiary from secondary interactions. If one or several divalent cations, typically Mg^{2+} , are specifically involved in locking the tertiary structure (such as that in the core of the *Tetrahymena* group I intron [56]), then this should show up by monitoring the dependence of T_m versus $\log[\text{cation}]$ for various cations (for example, [57]).

Thermal unfolding of the *td* intron of bacteriophage T4 [58] showed a highly cooperative 3D structure melting followed by the melting of 2D structure (hairpins) at higher temperatures, which prevented a detailed dissection of the 3D structural melting profile. However, the effect of several point mutations was studied and, in parallel to the UV melting profiles, temperature gradient gel electrophoresis (TGGE) experiments [59] were performed to detect possible multiple conformers. Furthermore, the effect of salts was examined in great detail, revealing a displacement of the melting toward higher temperatures and, concomitantly, a sharpening of the transition (increase of ΔH of melting). It was observed that 50 mM NH_4^+ supplemented with as little as 2 mM Mg^{2+} was as efficient as 0.9 M NH_4^+ alone. This was thought to result from the existence of specific magnesium binding site(s).

Several other examples (reviewed in [4]) document that, with thorough studies involving additional techniques, it is indeed possible to gain significant information. A good example is the study of an mRNA pseudoknot that showed two distinct transitions [12]. These two transitions could also be observed by DSC and were convincingly attributed to the two halves of the pseudoknot after a thorough experimental investigation (separate melting analyses of the two halves, enzymatic probing of the pseudoknot and the two halves, investigation of monovalent and divalent salt dependence).

The lesson from these examples is that gaining information for large and multidomain RNAs requires using UV melting supplemented with other techniques.

22.10

Parameters Influencing the Melting Temperature

The major parameter influencing the melting temperature of nucleic acids is the ionic strength. This is due to counterion screening of the negatively charged phosphates, which alleviates electrostatic repulsion during association of two strands. Manning [60, 61] developed the first comprehensive theory to explain the effect of cation condensation on elongated polyelectrolytes bearing negative charge with sufficient linear density. The theory explains that for monovalent salt concentrations for which the Debye–Hückel approximation is valid (i.e., <0.1 M), $dT_m/d(\log_{10} M)$ has a constant value irrespective of the sequence. For long helical molecules, the constant is around 18–20 °C per decade, and for shorter helices,

as well as for hairpins, it is less (see [62] for more quantitative considerations). Importantly, the screening effect of divalent cations, and particularly of magnesium cations, is two orders of magnitude stronger than that of monovalent salts. Finally, it should be kept in mind that, beyond some concentration (e.g., above 1.5 M NH_4^+ for the *td* intron mentioned in the previous section), further increases in salt concentration can destabilize RNA structures.

There also exists a large family of natural compounds called “*osmolytes*” (sugars, polyols, methylamines, urea, etc.) that may have significant (positive or negative) effects on the stability of RNA [63]. The small molecule trimethylamine oxide (TMAO) was studied in detail and shown to promote native-like tertiary folding (but not secondary structure formation) in the absence of magnesium ions. In line with this observation, a T_m increase of 14 °C was observed from 0 to 2 *m* TMAO (1 molal (*m*) = 1 mol kg⁻¹) for the aptamer domain of an adenine riboswitch [64].

Typical denaturants such as urea or guanidinium chloride obviously tend to decrease the melting temperature. This may be practically useful to make a transition observable that has an otherwise too high melting temperature in the absence of denaturant [12]. Beyond practical reasons, variable concentrations of denaturant (*D*) are commonly used to study the stability of proteins [65], an approach that is much less widely applied in the nucleic acid field. There do not seem to exist good reasons for this, since it has been shown that the method yields quite comparable results in both fields, particularly with respect to the correlation between $m = \partial \Delta G / \partial D$ and the variation of accessible surface area during denaturation [66].

Finally, it is of interest to recall that the melting temperature of DNA is sensitive to methylation or halogenation of bases at sites not involved in Watson–Crick pairing [67]. In the same line, melting temperature variations were also observed with RNA after uridine halogenation at the C5 position [68]. The most striking effect, however, was on the outcome of annealing of almost self-complementary sequences after their heat denaturation in water. Depending on the position of modification, an important and unpredictable change in the yield of the hairpin *versus* duplex form was obtained [68]. Such observations are of importance for attempts to favor the hairpin form in structural studies. In this context, it should be mentioned that, contrary to a common thought, the kinetics of cooling during an annealing procedure may have a negligible effect on the hairpin *versus* duplex yield, whereas the strand concentration may be the crucial parameter [52].

22.11

Practical Problems

22.11.1

Evaporation during Heating: an Important Improvement

A recurring and irritating problem is evaporation during the increase in temperature, which alters the concentration of the absorbing material in an unpredictable manner. To prevent this problem, the samples are usually covered with paraffin oil and the cuvettes are closed with polytetrafluoroethylene (PTFE,

brand name teflon) caps. Furthermore, the samples are often degassed before use to avoid bubble formation, a problem that tends to be exacerbated for cuvettes with path lengths smaller than 1 cm. However, even when respecting such good practices, one is still often faced with evaporation in an unpredictable manner, which makes it necessary to repeat experiments until a coherent data set is obtained. To solve this problem, we designed a new setup that was manufactured by Hellma (Mülheim – Germany) following our specifications.

The setup consists of new cells (Hellma: product number 176 729 QS, modification from catalog No. 176 700; see Appendix B) being closed by a hairpin tubing. Each end is attached to the cuvette by a PTFE screw (Figure 22.5). The sample volume is 390 μl (outer dimensions $35 \times 12.5 \times 12.5 \text{ mm}^3$). The position of the window had to be adapted to the UV beam dimensions in our UVIKON spectrophotometer. See Appendix B for more details.

This setup provides an additional free volume allowing water dilatation and humid air expansion. As a consequence, the evaporation problem is abolished (Figure 22.4) without having the burden to use oil. In addition, the setup allows linking the cuvettes in line, which facilitates their cleaning and sterilization (Figure 22.5c). For that, a peristaltic pump is used to flush the cuvettes with 0.1 M sodium hydroxide, followed by pure water. Eventually, the cuvettes are emptied and vacuum-dried thoroughly.

22.11.2

Sloping Baseline

The sloping baseline problem is a recurring problem in UV melting studies. It is thus common to subtract both lower and upper temperature baselines, $A_L(T)$ and $A_U(T)$, respectively, from the raw absorbance $A(T)$ to obtain the fraction f of melted strands as follows (Figure 22.6a):

$$f = \frac{A(T) - A_L(T)}{A_U(T) - A_L(T)} \quad (22.14)$$

Obviously, $A_L(T)$ or $A_U(T)$ is not well assignable if there is no clear linear part of $A(T)$ in the lower or upper temperature range, respectively (see Figure 22.6b for a problem with the upper temperature range). What is the order of magnitude of the slope of these functions, and what is their origin? At a wavelength such as 295 nm, which leads to hypochromism (Section 22.4), lower baselines may even have negative slopes.

A particularly pure example of a sloping baseline has been obtained in a study with the complementary strands dCA₅G and dCT₅G [70]. When each strand was used alone in 1 M NaCl, an almost perfect linear variation of $A(T)$ was obtained with a normalized slope $[1 - A(T_{\min})/A(T_{\max})]/(T_{\max} - T_{\min}) = 1.2 \times 10^{-3} \text{ } ^\circ\text{C}^{-1}$ ($\lambda = 260 \text{ nm}$). The linearity is the mark of the absence of any significant secondary structure, as expected with such sequences. It is an indication that a sloping lower baseline results from ensembles of low-energy structures, for example, involving unpaired residues in dangling ends or in more or less disordered loops. This is

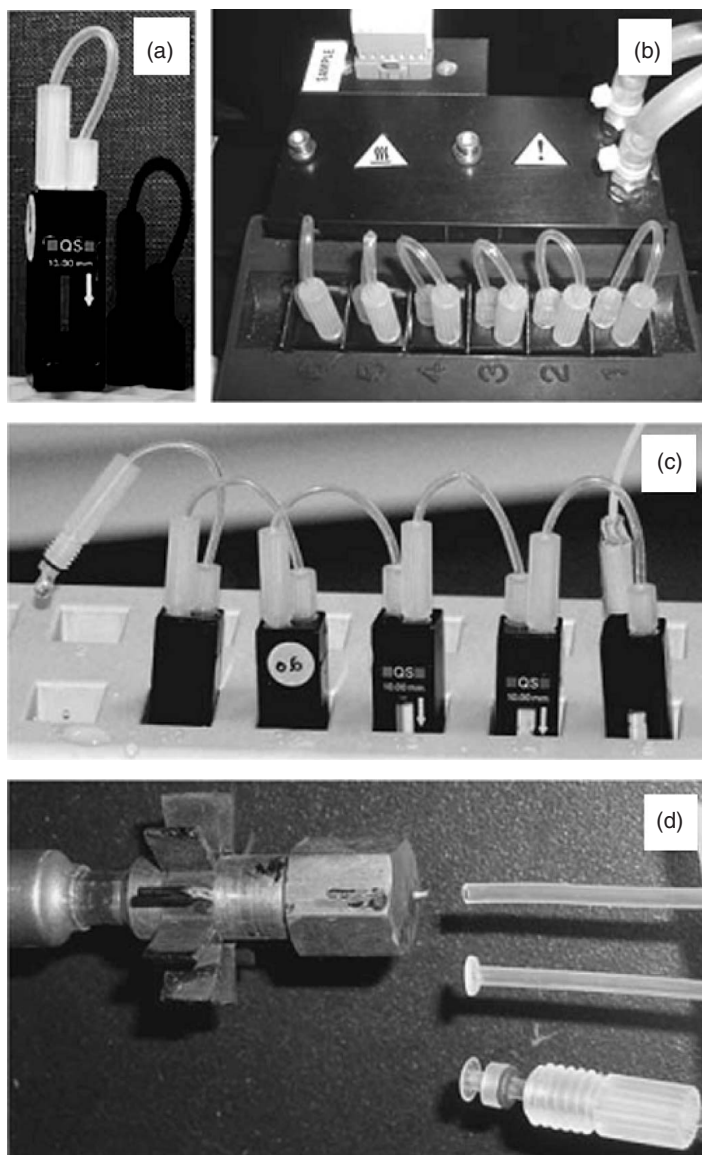


Figure 22.5 Photograph of the new setup preventing leakage and evaporation. (a) Spectrophotometric cell closed by the hair-pin PTFE tubing fit, (b) closed cells shown in the UVIKON-XL thermoregulated rack,

(c) in-line link of the cells for their washing between each experiment, and (d) heat-molding of the tubing for the second screw. The heating device is a standard tool from GE Healthcare.

strongly corroborated by a study on repeated sequences $(CTG)_n$ that can form in many ways imperfect hairpins, or imperfect duplexes, with single-stranded tails. With 25 repeats, lower baselines ($\lambda = 290$ nm) with comparable normalized slopes ($\sim 1.5 \times 10^{-3} \text{ } ^\circ\text{C}^{-1}$, see Figure 1A in Ref. [29]) were indeed observed. This is also illustrated by the melting of the duplex formed by 5'-GCGCGAGCGC-3' having a lower baseline with a large slope (Figure 22.6a). Clearly, such a sequence can form either a blunt-ended duplex with a central AG-GA mismatch or 4-bp duplexes involving the pairing of GCGC, each strand having a dangling tail (either 5'-GCGCGA-3' or 5'-GAGCGC-3').

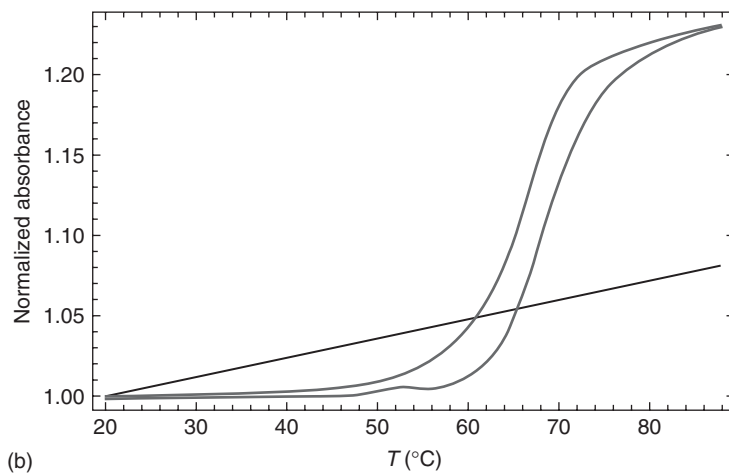
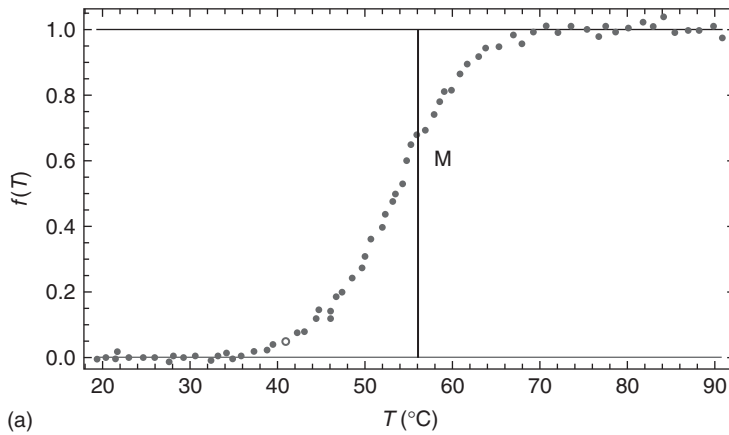
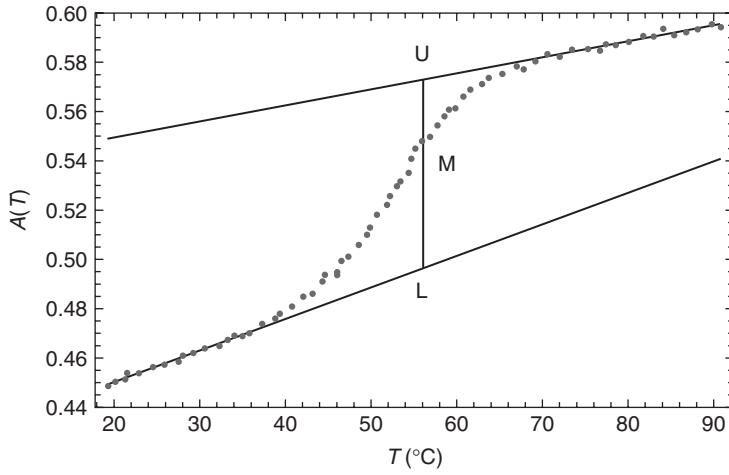
Strong evidence in favor of this interpretation can be obtained by preparing material for which no such low-energy structures are able to form and verifying that the lower baseline is flat. This can be neatly achieved by dissolving well-ordered crystals because such crystals have obviously grown from molecular species of a uniform kind. We have performed this with crystals of HIV-1 DIS RNA able to form either a kissing complex or a duplex [50, 51]. The two 3D structures showed all bases in stacking interactions, and most of them base paired in agreement with the expected secondary structures. Accordingly, the resulting melting curves with RNA from dissolved crystals had perfectly flat lower baselines at 260 nm (Figure 22.6b).

It is thus clear that a lower baseline with a large slope is a hallmark of the multiplicity of “native” forms with dangling ends and/or flexible internal loops (which should be predictable by sequence analysis). A sound theoretical treatment should consider all possible forms, and not only the ideal and most stable form. However, this is certainly not easy to perform and, to our knowledge, this is not often performed (except for studies explicitly addressing alternative folds, such as [71, 72]). It may at least be said that such ensembles of conformers should lead to a broadening of the melting curve and thus to underestimating the ΔH associated with the genuine native form.

Finally, in situations where two different strands S_1 and S_2 are mixed, it results from the previous considerations that a perfect stoichiometric ratio ($C_{T1} = C_{T2}$) implies a minimum value of the slope of the lower baseline. Indeed, any systematic excess of either strand will result in an additional contribution to the lower baseline.

Figure 22.6 (a) Lower and upper baseline correction. Top: raw absorbance data (at 280 nm) with well-defined lower and upper baselines for the sequence 5'-GCGCGAGCGC-3' (Source: Redrawn from Fig. 1 in Ref. [69]). Bottom: corrected data yielding $f(T)$ as defined by Eq. 22.14. The value of $f(T)$ at any point M on the corrected melting curve is derived from the original absorption data in the top figure by calculating $(M - L)/(U - L)$ at the same temperature. *Nota bene*: lower and upper baselines are not often so well defined; compare with the upper temperatures in (b). (b) Melting curves (after smoothing) obtained from dissolved crystals of subtype-F HIV-1

DIS in the duplex form (lower T_m) and the kissing-complex form (higher T_m) at 260 nm. The two melting curves, with absorbance normalized to 1 in the low-temperature range, have remarkably flat lower baselines. For comparison, a sloping lower baseline obtained from the single-stranded oligonucleotides dCA₅G or dCT₅G (Section 22.11.2) is shown. Note that accurate upper baselines would not be assignable in case of the HIV-1 DIS duplex and the kissing complex in particular. *Nota bene*: the two crystals were obtained in the presence of the antibiotic lividomycin, which further stabilized the complexes.



22.12

A Neat Experimental Solution to the Sloping Baseline

An insightful discussion of the sloping baseline problem is found in [30]. Gralla and Crothers used the T-jump technique to monitor the kinetics of return to equilibrium of duplexes with internal loops of different lengths. The resulting curves obtained by T-jumps of a few degrees showed a biphasic absorbance increase (at 266 nm) with an unresolved fast process being followed by a well-resolved slower process. In fact, the fast component “*which accounts for the sloping low- and high-temperature baselines of the integral melting transitions, depends strongly on the RNA sequence and the wavelength of observation,*” whereas the “*slower relaxation reflects only the dimer-to-coil transition.*” As a consequence, this provides a “*partitioning of the observed absorbance change into its various components*” and the amplitude of the slow component is just proportional to dA/dT without any influence of the sloping baseline. In other words, this slow component is proportional to a DMC fully devoid of baseline problems. The other great interest of this technique is that it also yields an experimental measurement of the relaxation time at the final temperature of each T-jump. This is strongly informative since a purely unimolecular strand-to-hairpin transition has a much shorter relaxation time (about 10–20 μ s) than a purely bimolecular strand-to-duplex transition (about 10–50 ms). The authors were thus able to avoid a misinterpretation of their results for one particular duplex in the series showing unexpectedly two resolvable long relaxation times, and not only one. They considered the possibility of triple helix formation being involved. No doubt, relying on melting curve data only would have made this situation undetectable and would have led to erroneous conclusions. One can only repeat the warning from the authors: “*We would like to urge caution in the thermodynamic interpretation of RNA melting curves that do not include partitioning of the observed absorbance change into its various components.*” Anyone who really wants to appreciate the subtleties and the possible pitfalls associated with UV melting studies is eagerly encouraged to read the two articles of this classical study [30, 73].

22.12.1

pH Variation and Buffers

The pH is an important parameter in itself and, more specifically, because of RNA degradation at high pH and temperatures, particularly in the presence of divalent cations (Section 22.12.2). Also, cytidine protonation can occur in RNA and DNA at moderately acidic pH. The pK_a of cytidine in isolation is 4.3, but this value can rise substantially when the protonated form is stabilized by a non-standard base pair, as with C.C⁺ in the DNA i-motif [74].

A buffer for a UV melting study obviously should not absorb UV light significantly. It is also commonly stated that a buffer has to be chosen for its capacity to maintain a constant pH. It may, of course, be of interest to maintain a constant pH in numerous situations, but this should not be considered as an indisputable rule for the following reasons. First, any buffer involves acid-base (de-)protonations that are exothermic/endergonic and generally lead to $\Delta H \neq 0$. As a consequence,

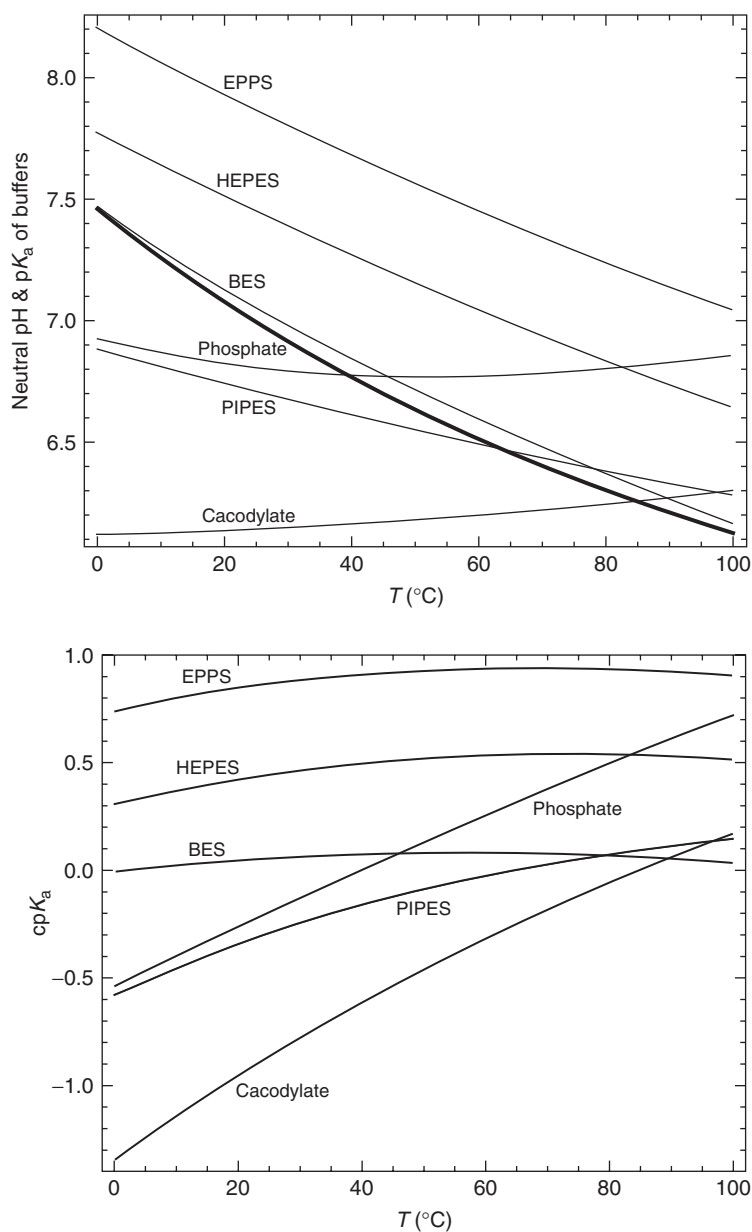


Figure 22.7 (a) Variation with temperature of the neutral pH (thick curve) and of the pK_a of different buffers (thin curves). (b) Variation with temperature of corrected pK_a (cpK_a) values for different buffers (Section 22.12.1 and Appendix D). See Ref. [76] for an indispensable compendium on biological buffers.

the pK_a of a buffer varies in general with the temperature according to the Van't Hoff equation (see [75] for a study on all usual buffers analyzed in the presence of 0.1 M KCl). Likewise, water dismutation $H_2O \rightleftharpoons H^+ + OH^-$ also has a $\Delta H \neq 0$ and the ionic product of water $K_w = [H^+][OH^-]$ is thus temperature-dependent. Since the neutral pH is defined by $[H^+] = [OH^-] = K_w^{1/2}$, one should be aware that *pH 7 is the neutral pH only at 25 °C*: for example, $[H^+] = [OH^-] = 10^{-7.45}$ M at 0 °C (neutral pH = 7.45) and $10^{-6.3}$ M at 80 °C (neutral pH = 6.3), as illustrated by the neutrality curve in Figure 22.7a. Thus, to adjust the pH of water to 7.0 at 0 °C, protons have to be added, resulting in non-neutral, slightly acidic conditions ($[H^+] > [OH^-]$); likewise, protons have to be removed to adjust the pH to 7.0 at 80 °C, yielding slightly basic conditions ($[H^+] < [OH^-]$). Overall, the neutral pH changes by as much as 1.5 units from 0 to 100 °C. This is taken into account by applying a “corrected pH” index (cpH) defined as

$$cpH(T) = pH - pH_{\text{neutral}}(T) = -\log_{10} \sqrt{\frac{[H^+]}{[OH^-]}} \quad (22.15)$$

A null cpH value is the mark of effective neutrality, whereas a negative/positive cpH value is the mark of effective acidic/basic conditions. It is thus legitimate to question the systematic attempt to maintain a constant pH when the temperature is strongly varied. This is particularly obvious if one seeks to remain as close as possible to *real* neutrality ($[H^+] = [OH^-]$), since this can only be achieved by using a buffer with a pK_a *versus* T curve as close as possible to the neutrality curve. A very good choice for this is N,N-bis(2-hydroxyethyl)-2-aminoethanesulfonic acid (BES) (Figure 22.7a), whereas phosphate buffer, which, in absolute terms, ensures a much more stable pH value close to 7, is very far from attaining that goal. In fact, compared to the neutrality curve, phosphate buffer produces an effective variation of one pH unit along the 10–90 °C temperature range. Since a buffer is best used at a pH equal to its pK_a , it is interesting to consider also a “corrected pK_a ” equal to $cpK_a(T) = pK_a(T) - pH_{\text{neutral}}(T)$, which is illustrated in Figure 22.7b. One may thus advocate in some situations the use of buffers with pK_a –temperature variations minimizing the variation of this cpK_a index, and not of the pK_a itself. Independent of these considerations, it should also be highlighted that phosphate buffer (but not BES) has a strong propensity to complex Mg^{2+} (and Ca^{2+}) ions, which definitely does not make it a very good choice when such divalent cations, particularly Mg^{2+} , are essential factors for the stability of the RNA being studied. To our knowledge, the influence of the buffer on the outcome of a UV melting study has not yet been analyzed in a systematic manner. The information contained in Figure 22.7, along with the additional explanations in Appendix D, should help the reader toward reaching a decision on this topic.

22.12.2

RNA Degradation

A major concern linked to melting studies of RNA is degradation, particularly in the presence of magnesium or other divalent cations. As a precaution, the measuring

buffer may be adjusted to 0.1–1 mM ethylenediaminetetraacetic acid (EDTA). For large RNAs (say, above 30 nt), the only way to test this is a classical polyacrylamide gel electrophoresis before and after the experiment. A simple ethidium bromide (EtBr) staining of the RNA within the gel followed by visualization of bands under UV light should suffice to assess whether or not degradation was significant. However, note that EtBr staining is more efficient on double- than single-stranded nucleic acids, not linearly correlated with the amount of nucleic acid, and generally inefficient with shorter oligoribonucleotides. Alternatively, RNA may be stained with SYBR[®] Gold (Life Technologies Invitrogen), Toluidine Blue O (Sigma-Aldrich), or silver (Silver Stain Plus Kit, Bio-Rad). An accurate picture of degradation would require ³²P-labeling of the RNA before the melting experiment. However, since making use of radioactive material in a spectrophotometer cannot be recommended, a separate heating experiment with the same kinetics of heating would be required. For shorter RNAs, high-pressure liquid chromatography (HPLC) with a Dionex Nucleopac PA-100 ion exchange column under denaturing conditions gives excellent results (see [77] for details). In order to minimize RNA degradation, the heating rate should be as fast as possible, particularly in the high-temperature range.

22.12.3

Heating Rate and Data Sampling

The rate of temperature change dT/dt is an important parameter. It is often referred to as the *heating rate* or *scanning rate* (particularly in DSC studies). *Stricto sensu*, it should be close to zero to maintain thermodynamic equilibrium and perfect temperature homogeneity in the cell. In practice, an experiment cannot last too long, and furthermore, it is often sufficient to ensure temperature homogeneity. In general, one considers heating rates in the order of or below $1\text{ }^{\circ}\text{C min}^{-1}$ (typically $0.1\text{--}0.5\text{ }^{\circ}\text{C min}^{-1}$). In order to assess whether the heating rate is slow enough to ensure thermodynamic equilibrium, one has to check first that the melting temperature is independent of it (which was not the case in the situation of Figure 22.4 with a clear hysteresis upon cooling). Obviously, such a hysteresis is also a clear mark of lack of equilibrium. Be aware that, in such a case, lowering the scanning rate might well be totally useless to suppress hysteresis and to return to equilibrium because this could require astronomically long experiments. Note that checking for the absence of hysteresis upon cooling may be difficult because of evaporation, which increases the absorbance and prevents one from obtaining superimposable heating and cooling curves. For this reason, the setup described in Section 22.11.1 is of great interest since evaporation can be completely avoided.

In order to obtain the best possible results from curve fitting, it is important to sufficiently sample the experimental absorbance curve, at least every $0.5\text{ }^{\circ}\text{C}$. In situations where the DMC is used for the fitting procedure, it is better to use a more frequent sampling ($\sim 0.25\text{ }^{\circ}\text{C}$) because the information lies in the part where the DMC has a maximum curvature.

22.12.4

Experimental Data Processing

It is often advisable to work with the derivative of the melting curve (DMC), as commonly performed. This makes it easier to assess whether or not the melting curve is symmetrical or skewed (Section 22.5) and also to perform a direct comparison with a DSC experiment (Figure 22.3). This also discloses much more efficiently the small accidents on a melting curve. However, experimental noise has a dramatic influence on a derivative obtained numerically. It is thus extremely important to perform a smoothing with great care. Various methods have been described [78]. We first explain briefly the principle of classical Fourier-based methods and, finally, how to minimize the effect of a possible distortion because of smoothing.

An absorbance curve $A(T)$, being bounded by the temperatures $T_{\min}$ and $T_{\max}$, may be viewed as one basic repeat of an imaginary periodic curve with the period $T_{\max} - T_{\min}$. This allows representing it by a discrete Fourier series:

$$A(\Theta) = \sum_{k=-\infty}^{k=+\infty} A_k e^{i2\pi k\Theta} \quad (22.16a)$$

with Θ being the new variable $(T - T_{\min})/(T_{\max} - T_{\min})$ varying from 0 to 1, and A_k being the k^{th} complex Fourier coefficient of $A(\Theta)$.²⁾ In practice, a function such as $A(\Theta)$ may be represented by a finite sum of sinusoids ($-k_{\max} \leq k \leq k_{\max}$). It is intuitively clear that the true signal information contained in each A_k lies more in the low k values (low frequencies), whereas the noise lies more in the high k values (high frequencies). A good way of cleaning up the signal from a part of the noise is thus to down-weight progressively the high-frequency coefficients by replacing Eq. 22.16a with

$$A(\Theta) = \sum_{k=-k_{\max}}^{k=k_{\max}} \omega_k A_k e^{i2\pi k\Theta} \quad (22.16b)$$

where the weights ω_k vary from 1 for $k = \pm 1$ to almost 0 for $k = \pm k_{\max}$. In fact, several common smoothing procedures are equivalent to determining a specific set of ω_k . Many textbooks and web sites illustrate the link between the ω_k used to multiply each Fourier coefficient and the nature of the smoothing (for example, [79]).

2) For those not familiar with the use of imaginary exponential functions, it suffices to know that each pair of terms $A_k e^{i2\pi k\Theta}$ and $A_{-k} e^{-i2\pi k\Theta}$ corresponds to a particular sinusoid oscillating **exactly** k times between $T_{\min}$ and $T_{\max}$, or equivalently between 0 and 1 for Θ . The amplitude and the phase of each sinusoid are contained in the pair of Fourier coefficients A_k

and A_{-k} . Because the absorbance function is real one has $A_{-k} = \overline{A_k}$, the bar indicating complex conjugation. The crystallographers will recognize this as the “Friedel law”. We did not address the so-called “Gibbs problem” arising from the discontinuity in the imaginary periodic function due to $A(T_{\min}) \neq A(T_{\max})$. This is beyond the scope of this summary.

We propose on our web site (<http://www-ibmc.u-strasbg.fr:8080/webMathematica/Meltingcurve/>) a tool for processing experimental curves using such methods [40].

An important practical remark should be made. Any smoothing procedure has to cope with the dilemma of performing enough smoothing without introducing undue distortion. In order to counteract the adverse effect of such a possible distortion of the smoothed experimental DMC, it is judicious to apply the same smoothing procedure to any theoretical DMC calculated during the fitting procedure (Section 22.5.4.5). The aim of this ultimate treatment is of course not to suppress noise in this noise-free theoretical function, but to distort it in the same way as was possibly distorted the experimental curve to fit. It is particularly efficient with this Fourier-based method since it reduces the variation of ΔH resulting from different extents of smoothing (the variation on T_m being negligible). For example, with the melting curve corresponding to the kissing-complex form in Figure 22.6b, the value of ΔH obtained by using this improved procedure changed marginally from (369 ± 6.3) to (375 ± 8.4) kJ mol⁻¹ when going from $k_{max} = 9$ to $k_{max} = 12$ in Eq. 22.16b, whereas ΔH changed significantly from (353.4 ± 5.4) to (371 ± 8) kJ mol⁻¹ without using it.

22.12.5

Softwares

Each spectrophotometer may be furnished with software tools, at least for visualization of the melting curves. The following articles presented software developments [80–83]. Softwares are also available either for download from web sites (such as <http://www.meltwin3.com/>, for Windows), or directly usable through the web sites as the one we propose for smoothing raw melting curves, obtaining DMCs and analyzing simple unimolecular or bimolecular transitions (<http://www-ibmc.u-strasbg.fr:8080/webMathematica/Meltingcurve/>).

Simulation and prediction softwares are available (non-exhaustive list):

<http://unafold.math.rpi.edu/rna/node3.html>

<http://www.ebi.ac.uk/compneur-srv/melting/>

<http://www.tbi.univie.ac.at/%7Eivo/RNA/man/RNAheat.html>

Acknowledgment

We warmly thank Roland K. Hartmann for a thorough reading of the manuscript and many valuable suggestions.

Appendix A: Difference between T_m and T_{max} and DMC Normalization

The difference between T_m and the temperature T_{max} corresponding to the maximum value of the DMC (i.e., to the inflection point of the melting curve defined by

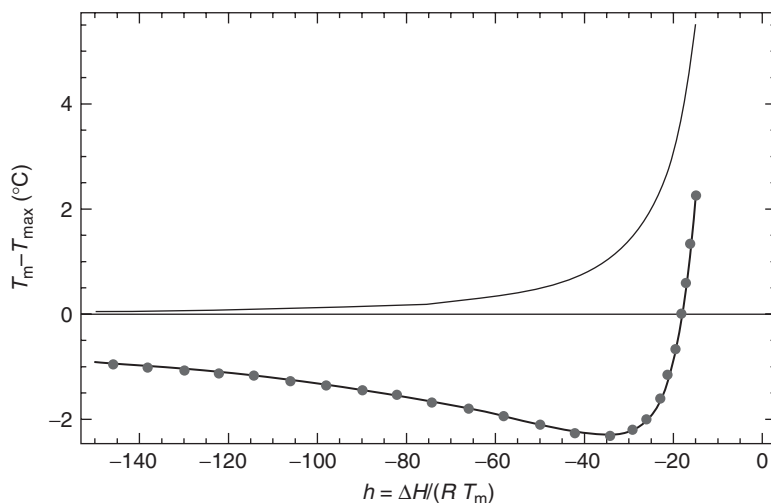


Figure 22.8 Variation of $\Delta T_m = T_m - T_{\max}$ versus $h = \Delta H/(R T_m)$ (with $T_m = 313$ K) for unimolecular transitions (Eq. 22.17, upper curve) and for bimolecular transitions (Eq. 22.18, lower curve). The dots correspond to exact values calculated numerically, which proves the quality of the approximation given by Eq. 22.18.

$dA^2/dT^2 = 0$) is obtained as

$$\Delta T_m = T_m - T_{\max} \approx \frac{4T_{\max}}{h^2} \approx \frac{4T_m}{h^2} = \frac{4R^2 T_m^3}{\Delta H^2} \quad (22.17)$$

for a unimolecular transition. This result is shown in Figure 22.8, where it is compared with its equivalent for bimolecular transitions (see below). For unimolecular transitions, $T_{\max}$ is thus always lower than the true melting temperature (*nota bene*, with ideal absorbance coefficients not varying with the temperature). However, the error is usually small since for a $T_m \sim 50^\circ\text{C}$ (323 K), ΔT_m attains 1°C only when $|\Delta H| \leq 25 \text{ kcal mol}^{-1}$ and is less than 0.25°C for $|\Delta H| \geq 50 \text{ kcal mol}^{-1}$. Furthermore, baseline variation has generally an opposite effect, which makes $T_m = T_{\max}$ a very good approximation for unimolecular transitions.

For bimolecular transitions, a more complex equation than Eq. 22.17 is required to calculate $\Delta T_m = T_m - T_{\max}$. The following three-term series expansion may eventually be obtained (with help from a software such as *Mathematica*, *Wolfram Research*):

$$\frac{\Delta T_m}{T_m} = \frac{T_m - T_{\max}}{T_m} \approx \frac{0.4980}{h} + \frac{7.8785}{h^2} - \frac{18.039}{h^3} \quad (22.18)$$

This is illustrated in Figures 22.1b and 22.2 showing that, contrary to a unimolecular transition, $\Delta T_m = T_m - T_{\max}$ is almost always negative (remember: $h = \Delta H/(RT_m)$ is negative). Furthermore, ΔT_m is now significant since $|\Delta T_m| > 1\text{K}$ in all common situations, but $|\Delta T_m|$ cannot exceed $0.00738 \times T_m$ (2.32 K for $T_m \approx 313$ K). The major purpose of Eq. 22.18 is to yield an explicit expression for $T_{\max}$ as a function of ΔH and T_m , which allows rewriting Eq. 22.11 (normalized to

1 at $T = T_m$) in a form involving only ΔH and T_m as free parameters, but correctly normalized to 1 at $T = T_{\max}$. Indeed, an experimental DMC can be normalized in view of its processing only by considering its maximum value at $T = T_{\max}$, a temperature that can be identified before any treatment, whereas $T = T_m$ is not exactly known before DMC processing. For obtaining this normalization, the function in Eq. 22.11 has to be divided by the value it takes at $T_{\max}$, $T_{\max}$ being obtained from Eq. 22.18 as a function of ΔH and T_m . These considerations were put into practice on our web site [40].

Appendix B: Experimental Setup against Evaporation

Anyone interested in using the new setup has to order new cuvettes from Hellma, with the window position and size (height and width) adapted to the beam position of the spectrophotometer in use. At the time of writing, the reference 176 729 QS for the new product was not present in the Hellma catalog. However, Hellma will manufacture on request new cuvettes adapted to any instrument because the position of the window may need to be adapted to the height and dimension of the UV beam. The additional cost for these modified cells is about 20% of the regular price. The tubing and screws are also furnished by Hellma but require some adaptation. Each screw visible in Figure 22.5 is provided with a 20-cm long tubing already attached to it. To obtain the final setup with the tubing making a loop between two screws (Figure 22.5a), the tubing has to be shortened to 8.5 cm and fixed to the other screw. For that, the free end of the tubing has to be heat molded (Figure 22.5d). Additional information is readily available from Philippe Walter (p.walter@ibmc-cnrs.unistra.fr).

Appendix C: The Subtleties with Partial Derivatives for ΔC_p Determination

The only valid thermodynamic definition of heat capacity is $C_p = \partial H / \partial T$, from which derives $\Delta C_p = \partial \Delta H / \partial T$. Estimating this term is of great interest because it is linked to the variation of ΔG with T by the rigorous expression $\Delta C_p = -T \partial^2 \Delta G / \partial T^2$. The partial derivatives are of fundamental importance since any parameter other than the temperature must be kept fixed for their calculation. Considering any other parameter X that can be varied independently of the temperature, one has $d\Delta H = (\partial \Delta H / \partial T) dT + (\partial \Delta H / \partial X) dX$, the two elementary variations dT and dX being mutually independent. The parameter X may be, for example, the salt concentration, or the strand concentration for a duplex. In both cases, a variation of X induces a variation of T_m . If we are interested in following the variation of ΔH with T_m , we may replace dT with dT_m in the previous $d\Delta H$ expression but, then, dT_m is no more an independent variable since it is now physically linked to dX . It then appears from this new expression that $\partial \Delta H / \partial T = d\Delta H / dT_m$ if, and only if, $\partial \Delta H / \partial X = 0$. When the parameter X is the strand concentration, this is indeed the case since the molar variation of enthalpy

during duplex formation does not depend on the strand concentration (everything else being the same!). As a consequence, $\Delta C_p = \partial \Delta H / \partial T = d\Delta H / dT_m$ is valid when a duplex melting temperature is varied by changing the strand concentration. However, this is not true in general if this variation is obtained by changing the salt concentration C_s because, in general, $\partial \Delta H / \partial C_s \neq 0$.

Appendix D: Buffer pK_a Variation with the Temperature

The Van't Hoff equation $\partial \ln K_a / \partial T = \Delta H / R T^2$ can be integrated by using a second-order Taylor expansion $\Delta H = \Delta H_0 + \Delta C_p (T - T_0) + \alpha/2 (T - T_0)^2$, with ΔH_0 being the enthalpy of acid dissociation at a reference temperature T_0 (25 °C), $\Delta C_p = \partial \Delta H / \partial T$ and $\alpha = \partial^2 \Delta H / \partial T^2$. This leads to

$$K_a(T) = K_a^0(T/T_0)^{\frac{\Delta C_p - \alpha T_0}{R}} e^{\frac{(1 - T_0/T)[\Delta H_0/T_0 - \Delta C_p + (\alpha/2)(T + T_0)]}{R}} \quad (22.19)$$

with $K_a^0 = K_a(T_0)$. This allows obtaining the curve $pK_a(T) = -\log_{10} K_a(T)$ for all buffers (Figure 22.7a) for which ΔH_0 , K_a^0 , ΔC_p , and α were determined [75]. Most often, the term α is null or negligible. The previous expression for $K_a(T)$ also holds for the variation with temperature of the ionic product of water $K_W(T) = [H^+][OH^-]$. This yields the equation for the neutral pH curve: $pH_{\text{neutral}}(T) = -\log_{10} \sqrt{K_W(T)}$ (Figure 22.7a). For water dismutation, the value of ΔH_0 at $T_0 = 25$ °C is 55.73 kJ mol⁻¹ [84] and, using the experimental values of $K_W(T)$ between 0 and 100 °C [85], it was determined $\Delta C_p = -190.2$ J mol⁻¹ K⁻¹ (assuming $\alpha = 0$). This yields the values necessary for evaluating the variation of pK_a and cpK_a for all buffers in [75] (Figure 22.7b).

References

1. Darby, R.A., Sollogoub, M., McKeen, C., Brown, L., Risitano, A., Brown, N., Barton, C., Brown, T., and Fox, K.R. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Res.*, **30**, e39.
2. Puglisi, J.D. and Tinocco, J.I. (1989) in *Methods Enzymology*, vol. 180 (eds J.E. Dahlberg and J.N. Abelson), Academic Press, Inc., pp. 304–325.
3. Mergny, J.L. and Lacroix, L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.
4. Draper, D.E. and Gluick, T.C. (1995) in *Methods in Enzymology* (eds M.L. Johnson and G.K. Ackers), Academic Press, San Diego, CA, pp. 281–305.
5. Turner, D.H. (2000) in *Nucleic Acids, Structure, Properties and Functions* (eds V.A. Bloomfield, D.M. Crothers, and I. Tinoco), University Science Books, Sausalito, CA, pp. 259–334.
6. Craig, M.E., Crothers, D.M., and Doty, P. (1971) Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.*, **62**, 383–401.
7. Porschke, D. and Eigen, M. (1971) Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic–oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.*, **62**, 361–381.

8. Bloomfield, V.A., Crothers, D.M., and Tinoco, J.I. (2000) *Nucleic Acids: Structures, Properties and Function*, University Science Books, Sausalito, CA, pp. 165–181.
9. Cantor, C.R. and Schimmel, P.R. (1980) *Biophysical Chemistry, Part III*, W.H. Freeman & Co, San Francisco, CA.
10. Tinoco, I. (1960) Hypochromism in polynucleotides. *J. Am. Chem. Soc.*, **82**, 4785–4790.
11. Lakowicz, J.R. (2006) 3rd edn, *Principles of Fluorescence Spectroscopy*, Springer.
12. Gluick, T.C., Wills, N.M., Gesteland, R.F., and Draper, D.E. (1997) Folding of an mRNA pseudoknot required for stop codon readthrough: effects of mono- and divalent ions on stability. *Biochemistry*, **36**, 16173–16186.
13. Frechet, D., Ehrlich, R., and Remy, P. (1979) Thermal perturbation differential spectra of ribonucleic acids. I. Hydration effects. *Nucleic Acids Res.*, **7**, 1965–1980.
14. Frechet, D., Ehrlich, R., Remy, P., and Gabarro-Arpa, J. (1979) Thermal perturbation differential spectra of ribonucleic acids. II. Nearest neighbour interactions. *Nucleic Acids Res.*, **7**, 1981–2001.
15. Mergny, J.L., Phan, A.T., and Lacroix, L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
16. Englander, S.W. and Englander, J.J. (1965) Hydrogen exchange studies of sRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **53**, 370–378.
17. Printz, M.P. and von Hippel, P.H. (1965) Hydrogen exchange studies of DNA structure. *Proc. Natl. Acad. Sci. U.S.A.*, **53**, 363–370.
18. Fréchet, D. (1980) PhD thesis. Université Louis Pasteur, Fusion fine des acides ribonucléiques Strasbourg.
19. Mergny, J.L., Li, J., Lacroix, L., Amrane, S., and Chaires, J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.
20. Cole, P.E., Yang, S.K., and Crothers, D.M. (1972) Conformational changes of transfer ribonucleic acid. Equilibrium phase diagrams. *Biochemistry*, **11**, 4358–4368.
21. Yen, W.S. and Blake, R.D. (1981) Analysis of high-resolution thermal dispersion profiles of DNA: Treatment as a collection of discrete subtransitions. *Biopolymers*, **20**, 1161–1181.
22. Azbel, Y. (1979) DNA sequencing and melting curve. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 101–105.
23. Mikulecky, P.J. and Feig, A.L. (2006) Heat capacity changes associated with nucleic acid folding. *Biopolymers*, **82**, 38–58.
24. Takach, J.C., Mikulecky, P.J., and Feig, A.L. (2004) Salt-dependent heat capacity changes for RNA duplex formation. *J. Am. Chem. Soc.*, **126**, 6530–6531.
25. Vesnaver, G. and Breslauer, K.J. (1991) The contribution of DNA single-stranded order to the thermodynamics of duplex formation. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 3569–3573.
26. Albergo, D.D., Marky, L.A., Breslauer, K.J., and Turner, D.H. (1981) Thermodynamics of (dG–dC)₃ double-helix formation in water and deuterium oxide. *Biochemistry*, **20**, 1409–1413.
27. Breslauer, K.J., Sturtevant, J.M., and Tinoco, I. Jr. (1975) Calorimetric and spectroscopic investigation of the helix-to-coil transition of a ribo-oligonucleotide: rA7U7. *J. Mol. Biol.*, **99**, 549–565.
28. Mikulecky, P.J. and Feig, A.L. (2006) Heat capacity changes associated with DNA duplex formation: salt- and sequence-dependent effects. *Biochemistry*, **45**, 604–616.
29. Amrane, S., Sacca, B., Mills, M., Chauhan, M., Klump, H.H., and Mergny, J.L. (2005) Length-dependent energetics of (CTG)_n and (CAG)_n trinucleotide repeats. *Nucleic Acids Res.*, **33**, 4065–4077.
30. Gralla, J. and Crothers, D.M. (1973) Free energy of imperfect nucleic acid helices. 3. Small internal loops resulting from mismatches. *J. Mol. Biol.*, **78**, 301–319.
31. Marky, L.A. and Breslauer, K.J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers*, **26**, 1601–1620.
32. Jin, R.Z., Breslauer, K.J., Jones, R.A., and Gaffney, B.L. (1990) Tetraplex

- formation of a guanine-containing non-america DNA fragment. *Science*, **250**, 543–546.
33. Mergny, J.L., De Cian, A., Ghelab, A., Sacca, B., and Lacroix, L. (2005) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.*, **33**, 81–94.
 34. Mergny, J.L., De Cian, A., Amrane, S., and Webba da Silva, M. (2006) Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes. *Nucleic Acids Res.*, **34**, 2386–2397.
 35. Rougée, M., Faucon, B., Mergny, J.L., Barcelo, F., Giovannangeli, C., Garestier, T., and Hélène, C. (1992) Kinetics and thermodynamics of triple-helix formation: effects of ionic strength and mismatches. *Biochemistry*, **31**, 9269–9278.
 36. Wyatt, J.R., Davis, P.W., and Freier, S.M. (1996) Kinetics of G-quartet-mediated tetramer formation. *Biochemistry*, **35**, 8002–8008.
 37. Arya, D.P. and Bruice, T.C. (1999) Triple-helix formation of DNA oligomers with methylthiouracil-linked nucleosides (DNmt): a kinetic and thermodynamic analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 4384–4389.
 38. Roberts, R.W. and Crothers, D.M. (1996) Prediction of the stability of DNA triplexes. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 4320–4325.
 39. Schellman, J.A. (1955) The stability of hydrogen-bonded peptide structures in aqueous solution. *C.R. Trav. Lab. Carlsberg Ser. Chim.*, **29**, 230–259.
 40. Dumas, P. and Disdier, F. (2008) A web site for melting curves analysis <http://www-ibmc.u-strasbg.fr:8080/webMathematica/Meltingcurve/>.
 41. Brown, N.M., Rachwal, P.A., Brown, T., and Fox, K.R. (2005) Exceptionally slow kinetics of the intramolecular quadruplex formed by the *Oxytricha* telomeric repeat. *Org. Biomol. Chem.*, **3**, 4153–4157.
 42. Blake, R.D. and Fresco, J.R. (1966) Polynucleotides. VII. Spectrophotometric study of the kinetics of formation of the two-stranded helical complex resulting from the interaction of polyriboadenylate and polyribouridylylate. *J. Mol. Biol.*, **19**, 145–160.
 43. Boulard, Y., Gabarro-Arpa, J., Cognet, J.A., Le Bret, M., Guy, A., Teoule, R., Guschlbauer, W., and Fazakerley, G.V. (1991) The solution structure of a DNA hairpin containing a loop of three thymidines determined by nuclear magnetic resonance and molecular mechanics. *Nucleic Acids Res.*, **19**, 5159–5167.
 44. Scheffler, I.E., Elson, E.L., and Baldwin, R.L. (1968) Helix formation by dAT oligomers. I. Hairpin and straight-chain helices. *J. Mol. Biol.*, **36**, 291–304.
 45. Xodo, L.E., Manzini, G., Quadrioglio, F., van der Marel, G.A., and van Boom, J.H. (1988) The duplex-hairpin conformational transition of d(CGCGGATCGCGCG) and d(CGCGGTACGCGCG): a thermodynamic and kinetic study. *J. Biomol. Struct. Dyn.*, **6**, 139–152.
 46. Bonnet, G., Krichavsky, O., and Libchaber, A. (1998) Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 8602–8606.
 47. Groebe, D.R. and Uhlenbeck, O.C. (1988) Characterization of RNA hairpin loop stability. *Nucleic Acids Res.*, **16**, 11725–11735.
 48. Senior, M.M., Jones, R.A., and Breslauer, K.J. (1988) Influence of loop residues on the relative stabilities of DNA hairpin structures. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 6242–6246.
 49. Sheehy, J.P., Davis, A.R., and Znosko, B.M. (2010) Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA*, **16**, 417–429.
 50. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C., and Dumas, P. (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol.*, **8**, 1064–1068.
 51. Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann, B., Ehresmann, C., and Dumas, P. (1999) The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two

- adenine bulges. *Structure Fold. Des.*, **7**, 1439–1449.
52. Bernacchi, S., Ennifar, E., Toth, K., Walter, P., Langowski, J., and Dumas, P. (2005) Mechanism of hairpin-duplex conversion for the HIV-1 dimerization initiation site. *J. Biol. Chem.*, **280**, 40112–40121.
 53. Cole, P.E. and Crothers, D.M. (1972) Conformational changes of transfer ribonucleic acid. Relaxation kinetics of the early melting transition of methionine transfer ribonucleic acid (*Escherichia coli*). *Biochemistry*, **11**, 4368–4374.
 54. Coutts, S.M., Riesner, D., Romer, R., Rabl, C.R., and Maass, G. (1975) Kinetics of conformational changes in tRNA Phe (yeast) as studied by the fluorescence of the Y-base and of formycin substituted for the 3'-terminal adenine. *Biophys. Chem.*, **3**, 275–289.
 55. Hinz, H.J., Filimonov, V.V., and Privalov, P.L. (1977) Calorimetric studies on melting of tRNA Phe (yeast). *Eur. J. Biochem.*, **72**, 79–86.
 56. Cate, J.H., Hanna, R.L., and Doudna, J.A. (1997) A magnesium ion core at the heart of a ribozyme domain. *Nat. Struct. Biol.*, **4**, 553–558.
 57. Heerschap, A., Walters, J.A., and Hilbers, C.W. (1985) Interactions of some naturally occurring cations with phenylalanine and initiator tRNA from yeast as reflected by their thermal stability. *Biophys. Chem.*, **22**, 205–217.
 58. Brion, P., Michel, F., Schroeder, R., and Westhof, E. (1999) Analysis of the cooperative thermal unfolding of the *td* intron of bacteriophage T4. *Nucleic Acids Res.*, **27**, 2494–2502.
 59. Rosenbaum, V. and Riesner, D. (1987) Thermodynamic analysis of nucleic acids and proteins in purified form and in cellular extracts. *Biophys. Chem.*, **26**, 235–246.
 60. Manning, G.S. (1972) On the application of polyelectrolyte “limiting laws” to the helix-coil transition of DNA. II. The effect of Mg^{++} counterions. *Biopolymers*, **11**, 951–955.
 61. Manning, G.S. (1972) On the application of polyelectrolyte “limiting laws” to the helix-coil transition of DNA. I. Excess univalent cations. *Biopolymers*, **11**, 937–949.
 62. Bloomfield, V.A., Crothers, D.M., and Tinoco, J.I. (2000) in *Nucleic Acids Structures, Properties and Functions* (eds V.A. Bloomfield, D.M. Crothers, and J.I. Tinoco), University Science Books, Sausalito, CA, pp. 475–534.
 63. Lambert, D. and Draper, D.E. (2007) Effects of osmolytes on RNA secondary and tertiary structure stabilities and RNA- Mg^{2+} interactions. *J. Mol. Biol.*, **370**, 993–1005.
 64. Lambert, D., Leippy, D., and Draper, D.E. (2010) The osmolyte TMAO stabilizes native RNA tertiary structures in the absence of Mg^{2+} : evidence for a large barrier to folding from phosphate dehydration. *J. Mol. Biol.*, **404**, 138–157.
 65. Myers, J.K., Pace, C.N., and Scholtz, J.M. (1995) Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.*, **4**, 2138–2148.
 66. Shelton, V.M., Sosnick, T.R., and Pan, T. (1999) Applicability of urea in the thermodynamic analysis of secondary and tertiary RNA folding. *Biochemistry*, **38**, 16831–16839.
 67. Gill, J.E., Mazrimas, J.A., and Bishop, C.C. Jr. (1974) Physical studies on synthetic DANs containing 5-methylcytosine. *Biophys. Biochim. Acta*, **335**, 330–348.
 68. Ennifar, E., Bernacchi, S., Wolff, P., and Dumas, P. (2007) Influence of C-5 halogenation of uridines on hairpin versus duplex RNA folding. *RNA*, **13**, 1445–1452.
 69. Morse, S.E. and Draper, D.E. (1995) Purine-purine mismatches in RNA helices: evidence for protonated G.A pairs and next-nearest neighbor effects. *Nucleic Acids Res.*, **23**, 302–306.
 70. Nelson, J.W., Martin, F.H., and Tinoco, J.I. (1981) DNA and RNA oligomer thermodynamics: the effect of mismatched bases on double-helix stability. *Biopolymers*, **20**, 2509–2531.
 71. Hobartner, C. and Micura, R. (2003) Bistable secondary structures of small RNAs and their structural probing by

- comparative imino proton NMR spectroscopy. *J. Mol. Biol.*, **325**, 421–431.
72. Micura, R. and Hobartner, C. (2003) On secondary structure rearrangements and equilibria of small RNAs. *Chembiochem*, **4**, 984–990.
 73. Gralla, J. and Crothers, D.M. (1973) Free energy of imperfect nucleic acid helices. II. Small hairpin loops. *J. Mol. Biol.*, **73**, 497–511.
 74. Gehring, K., Leroy, J.L., and Gueron, M. (1993) A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature*, **363**, 561–565.
 75. Fukada, H. and Takahashi, K. (1998) Enthalpy and heat capacity changes for the proton dissociation of various buffer components in 0.1 M potassium chloride. *Proteins*, **33**, 159–166.
 76. Good N.E., Winget, G.D., Winter, W., Connolly, T.N., Izawa, S., and Singh, R.M. (1966) Hydrogen ion buffers for biological research. *Biochemistry*, **5**, 467–477.
 77. Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann, B., Ehresmann, C., and Dumas, P. (1999) The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure*, **7**, 1439–1449.
 78. Gabarro, J. (1978) Numerical analysis of thermal denaturation of nucleic acids. *Anal. Biochem.*, **91**, 309–322.
 79. Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1986) *Numerical Recipes The Art of Scientific Computing*, Cambridge University Press, Cambridge, pp. 417–420.
 80. Kawakami, J., Tanaka, Y., and Kishimoto, K. (2009) Accurate curve fitting procedure for UV melting analysis of highly thermostable RNA hairpins. *Nucleic Acids Symp. Ser.*, **53**, 227–228, doi: 10.1093/nass/nrp114.
 81. Markham, N.R. and Zuker, M. (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.*, **33**, W577–W581.
 82. Tataurov, A.V., You, Y., and Owczarzy, R. (2008) Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids. *Biophys. Chem.*, **133**, 66–70.
 83. Draper, D.E. and Gluick, T.C. (1995) Melting studies of RNA unfolding and RNA-ligand interactions. *Methods Enzymol.*, **259**, 281–305.
 84. Bender, P. and Biermann, W.J. (1952) Heat of neutralization studies at high acid–base concentrations. I. Sodium hydroxide–hydrochloric acid. *J. Am. Chem. Soc.*, **74**, 322–325.
 85. Sweeton, F.H., Mesmer, R.E., and Baes, C.F. (1974) Acidity measurements at elevated temperatures. VII. Dissociation of water. *J. Solution Chem.*, **3**, 191–214.

23

RNA Crystallization

Jiro Kondo, Claude Sauter, and Benoît Masquida

23.1

Introduction

Recent crystal structures of large RNAs such as ribosomal subunits [1–4], RNase P [5, 6], group I [7–9] and group II [10] introns, or the class I ligase [11], as well as sequence analysis coupled to molecular modeling, have demonstrated that RNA structure is modular [12, 13]; in other words, RNA architectures can be decomposed into individual building blocks (modules), recurrently found in various RNA molecules, that can be assembled together to form the overall RNA fold [14].

Because ribosomal subunits are the most abundant native particles in a growing cell, there is no need to prepare ribosomal RNAs (rRNAs) by *in vitro* transcription or chemical synthesis. This is crucial, as it “reduces” molecular handling to biochemical purification of native particles. However, in most cases, overproduction techniques are needed, and the biochemist faces subsequent RNA folding or protein–ligand/RNA association problems, unless the RNA is rather short. This partly explains why small RNA structures can be solved fairly quickly while RNAs beyond 100 nt require time-consuming biochemical characterization before successful crystallization.

The RNAs of interest are usually extracted from a larger sequence context. RNA constructs can be engineered in order to favor a unique secondary structure or putative crystal packing contacts. This can be assessed by combining the results from *in silico* folding programs [15–17], native polyacrylamide gel electrophoresis (PAGE), and/or size exclusion chromatography. When the RNA binds a substrate, further biochemical characterization might be required. Then, the fragments are either cloned and *in vitro* transcribed or chemically synthesized. In this chapter we describe experimental procedures that are routinely used in the laboratory to obtain highly pure RNA molecules suitable for crystallization studies. Once the RNA has been synthesized, it has to be purified and concentrated. The correct length product is separated from contaminants by gel electrophoresis and/or chromatography. The RNA is then eluted, concentrated, and desalted. Folding assays are performed before crystal screening. Favorable crystallization conditions

identified in preceding screens are then further refined and upscaled. At this stage, chemical modifications allowing incorporation of heavy atoms can be undertaken in order to use heavy atom derivative-based methods to solve the phase problem [18]. RNA can be chemically modified by the halogens, iodine [19] or bromine [20], or the chalcogen selenium [21–23].

23.2

RNA Purification

RNA molecules can be purified by either PAGE or liquid chromatography (HPLC: High-performance liquid chromatography, or FPLC: fast protein liquid chromatography). These methods can even be coupled to improve the results. While PAGE is applicable to any RNA length, HPLC is appropriate for the purification of RNAs of up to 35 nt in length; independent of RNA length, HPLC is always useful to cleanse RNA preparations after gel purification. Routine techniques mentioned in this chapter are described in more detail in [24]. The reader is also referred to Chapter 5, describing the non-denaturing large-scale purification of RNA by weak anion-exchange chromatography.

23.2.1

HPLC Purification

When the RNA oligonucleotide is shorter than around 35 nt, it can be purified by FPLC or HPLC. The best results are obtained using salt gradients on anion-exchange columns bearing quaternary amines such as mono-Q matrices. HPLC presents the advantage that the column can be heated in an oven to temperatures up to 90 °C, thus promoting the unfolding of the RNA and increasing the retention time on the column for a better separation. The addition of chaotropic agents such as urea or formamide enhances the effect of heating the sample. However, formamide should be used with caution in the presence of RNA. Heat leads to formamide decomposition into carbon monoxide and ammonia, the latter can very quickly hydrolyze the RNA. Under denaturing conditions, the RNA mix is fractionated according to the size of the present species, similar to what can be achieved using gel electrophoresis. A typical protocol is described in Table 23.1. Other HPLC purification procedures for RNA have been described elsewhere [25–27].

Several pitfalls should nonetheless be avoided. If the RNA has been produced by *in vitro* transcription, proteins should be removed by phenol/chloroform extraction. Otherwise, the column bed may get coated with proteins and the column will lose its loading capacity over time. The sample should be assayed for precipitation by mixing with the highest salt buffer that is going to be used for separation to avoid clogging of the HPLC tubing. Since the sample is heated, divalent ion contamination should be avoided. As a precaution, the pK_a value of the buffer should be in the slightly acidic range (between pH 5.6 and 7.0; see Figure 23.2)

Table 23.1 Composition of HPLC buffers typical for RNA purification. A sodium perchlorate gradient is run over 70 min from 15 to 70% buffer B with a 1 ml min^{-1} flow rate (MES: 2-[N-morpholino]ethanesulfonic acid). In the laboratory, we run these gradients on an ÄKTA (GE Healthcare) system equipped with a Dionex Nucleopac PA-100 ($0.452 \times 25 \times \pi \text{ mm}^3$) column (now available from Thermo Scientific/Dionex as CarboPac PA100 columns). After a wash step at 90% buffer B, the column is reequilibrated in 15% of buffer A to prepare the next run.

	Buffer A	Buffer B
MES	20 mM pH 6.2	20 mM pH 6.2
Urea	4 M	4 M
NaClO ₄	1 mM	400 mM

so as to minimize spontaneous hydrolysis of phosphodiester bonds by acid/base catalysis enhanced by divalent cations. To achieve this, we recommend using peek-coated (polyether-ether-ketone) pumps and peek tubing.

23.2.2

Gel Electrophoresis

RNAs of any size (up to 500 nt) can be purified efficiently using PAGE under denaturing, semi-denaturing, or native conditions. Various urea concentrations can be tried at analytical scale before going to preparative scale in order to identify the most appropriate protocol. A sequencing electrophoresis apparatus with an aluminum back to homogenize glass plates' temperature allowing the use of $30 \times 40 \text{ cm}^2$ gel plates is recommended. The running temperature is usually set between 50 and 60°C . Of course, the gel thickness has to be significantly increased when going toward preparative scale (use at least 1.5 mm thick spacers) to well separate the RNAs of different lengths. The volume of gel to be prepared is thus around 250 ml.

The following equipment is needed to set up the experimental procedure:

- PAGE equipment: Electrophoresis apparatus, siliconized glass plates (approximately $30 \times 40 \text{ cm}^2$), comb, and spacers (at least 1.5 mm thick).
- Acrylamide 20%/urea 8 M stock solution; made from 500 ml acrylamide-bisacrylamide (38% (w/v) acrylamide, 2% (w/v) *N,N'*-methylene bisacrylamide) mixed with 480.5 g urea and millipore water to give a volume of 1 liter.
- 10 x TBE (Tris-borate-EDTA buffer [24])
- 8 M Urea solution.

Once a gel of the appropriate acrylamide percentage has been prepared [24], the RNA solution to be fractionated is mixed with 1 vol. of 8 M urea and then loaded onto the gel. The preparative gel electrophoresis usually requires a power value around 25 W. Progress of the migration is followed by the course of the bromophenol blue and xylene cyanol dyes loaded in a lane containing no RNA.

After electrophoresis, glass plates are removed, the gel is wrapped in plastic film, and the RNA bands are visualized by UV shadowing using a UV lamp and a silica plate as a screen (see Chapter 3 for details).

23.2.3

RNA Recovery

The RNA contained in the visualized bands has to be eluted from the gel and concentrated before subsequent experiments. To achieve this goal, the bands are first delineated with an indelible marker on the plastic wrap. Then, they are cut out of the gel using a sterile scalpel blade.

23.2.3.1 Elution of the RNA from the Gel

The oligoribonucleotides are recovered by passive elution at 4 °C in Millipore water. For this purpose, the excised gel slices are crushed in a mill (A11 basic analysis mill, IKA) and poured in a 50 ml polypropylene conical tube with water (approximately 30 ml for gel slices excised from a preparative gel). The RNA-containing tube is placed in a rock and roll stirrer at 4 °C overnight. Finally, the eluted RNA solution is filtered on a 0.22 µm sterile filtration unit (Nalgene) to get rid of the acrylamide particles.

23.2.3.2 Concentrating and Desalting

Whatever the technique employed to purify the RNA, it is necessary to desalt and concentrate it before use in crystallization trials. A very efficient way of achieving this is to use reverse-phase Sep-Pak columns that can be used on the bench (Waters Sep-Pak C18 Classic Cartridge). These are operated by gravity or with a syringe.

A classical protocol consists of the following steps:

- 1) Plug the column inlet to the luer of a 10 ml syringe and fix it to a bench stand.
- 2) Equilibrate the column using 10 ml of methanol.
- 3) Pass through 10 ml of Millipore water.
- 4) Load the sample.
- 5) Wash the sample with 10 ml of Millipore water.
- 6) Elute the sample with 5 ml of water/acetonitrile (1 : 4) in 1 ml fractions.

Three facts should be kept in mind when using Sep-Pak cartridges. The pH of the sample should not exceed 7 to guarantee efficient binding to the column bed. The loading step should not exceed 10 min to minimize loss of material due to driving by the mobile phase. If loading would take longer, the sample should be fractionated on more than one column. The column should never run dry to prevent sample loss. Hence, the syringe luer should be removed with caution in intermediate steps. The next solution should be added when there is still a small volume (100 µl) of the previous phase in the syringe.

The eluted RNA in water/acetonitrile is then evaporated to dryness in a SpeedVac. The pellet can be redissolved in a buffer of choice for further studies.

23.3

RNA Crystallization

The preparation of crystals for a new target molecule generally starts with a screening step to define solvent conditions favorable to crystallization. Various crystallants (salts, alcohols, polyethylene glycols (PEGs)) are added to the RNA solution in order to drive the macromolecule to an unstable state leading to either precipitation or crystallization. Best crystallants will drive the system to the nucleation zone of the phase diagram where RNA entities will first assemble to constitute so-called nuclei, and then grow to form crystals (Figure 23.1). The crystallization process is essentially multiparametric, and other critical parameters to be taken into consideration are pH and buffer, temperature, and additives such as cationic counterions (magnesium ions or polyamines). The outcome of screening must be examined carefully to identify all promising hits and possible polymorphs (different crystal forms and habits). Most of the time, an optimization step is required to improve the initial conditions and grow high-quality crystals suitable for structure determination at atomic resolution.

23.3.1

Renaturing the RNA

Before setting up crystallization experiments, the concentrated RNA has to be properly folded in the native state. This is performed by a heating step in a heating block for 1 min at 70–85 °C (depending on the melting temperature), in the presence of monovalent salts only. Then, the solution is left in the switched-off heating block to cool down slowly until room temperature is reached. In order to avoid self-cleavage of the RNA, the pH is usually chosen slightly acidic, and divalent cations are added only after the RNA solution has cooled down to at least 35 °C. A typical concentration of renatured RNA to start with for crystallization is about 5–10 mg ml⁻¹.

23.3.2

Search for Crystallization Conditions

Solvent conditions leading to the crystallization of a new RNA are hardly predictable *ab initio*. They are generally identified by screening a broad range of chemical mixtures. A survey of the RNA structures deposited in the PDB reveals general trends that can be used to narrow down the search: a majority of RNAs crystallize at pH 6–7 (slightly acidic to prevent RNA hydrolysis), in the presence of sodium cacodylate buffer, magnesium ions, and crystallants such as 2-methyl-2,4-pentandiol (MPD), PEG 400 or 4000, or ammonium sulfate (Figure 23.2). This observation triggered the design of first sparse matrices dedicated to the crystallization of nucleic acids [29–35]. Nowadays, many commercial screens are available in a 96-well format that facilitates the initial search for appropriate conditions. The number of individual conditions to be screened, designed as multiples of 96, will depend on

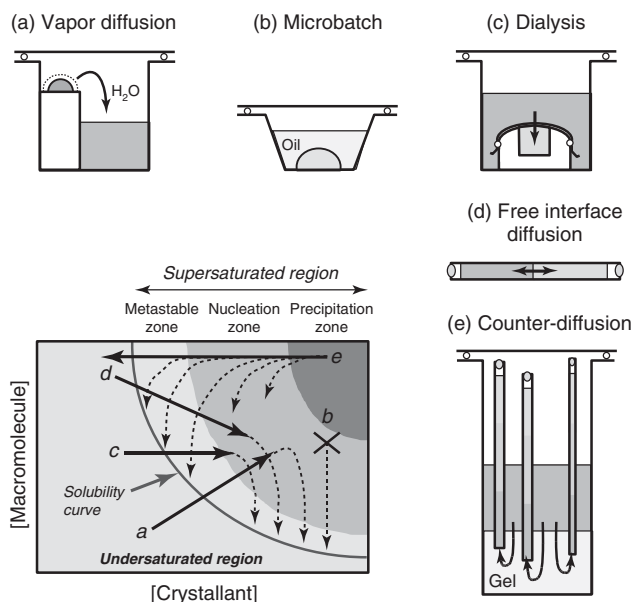


Figure 23.1 Crystallization, phase diagram, and crystallization methods. To define the conditions yielding high-quality crystals, the crystal grower can play with different parameters to drive a macromolecule, an RNA, for instance, from a solute to a crystalline state. This theoretical 2D phase diagram illustrates how this can be achieved. By modifying two parameters (macromolecule and crystallant concentrations), the system will reach the supersaturated region beyond the solubility curve where RNA molecules become unstable in solution. The system needs to cross an energy barrier (symbolized by the metastable zone) to produce stable molecular assemblies, the nuclei. This step only happens in the so-called nucleation zone. When stable nuclei are formed, they will capture more RNA entities from the

mother liquor and produce 3D crystals. Best conditions to grow crystals of high quality are close to the boundary between the metastable and the nucleation zones at low supersaturation. The trajectory in the diagram directly depends on the crystallization method. The most common crystallization setups are (a) vapor diffusion in hanging or sitting drops, (b) microbatch under oil, (c) dialysis in Cambridge buttons, (d) free interface diffusion in thin capillaries and microchannels, and (e) counter-diffusion in X-ray capillary tubes pinched in a gel layer. Typical equilibration trajectories are illustrated in the phase diagram by black arrows and the evolution of assays after nucleation and during crystal growth are indicated by dashed arrows. (For more details, see [28].)

sample quantity, crystallization method, and whether the assays are set up manually (0.5–1 μl /assay) or using pipetting robots (50–150 nL/assay). Indeed, robotic dispensers have been massively promoted by structural genomics initiatives over the past decade to perform high-throughput (HT) screening. They are becoming more and more common in crystallography laboratories because they significantly reduce sample consumption and thus allow more conditions to be screened with limited amounts of RNA. However, the experience shows that if the first 400–500 trials do not yield any promising result, it is probably the time to design a new RNA sequence with improved crystallization properties (see below).

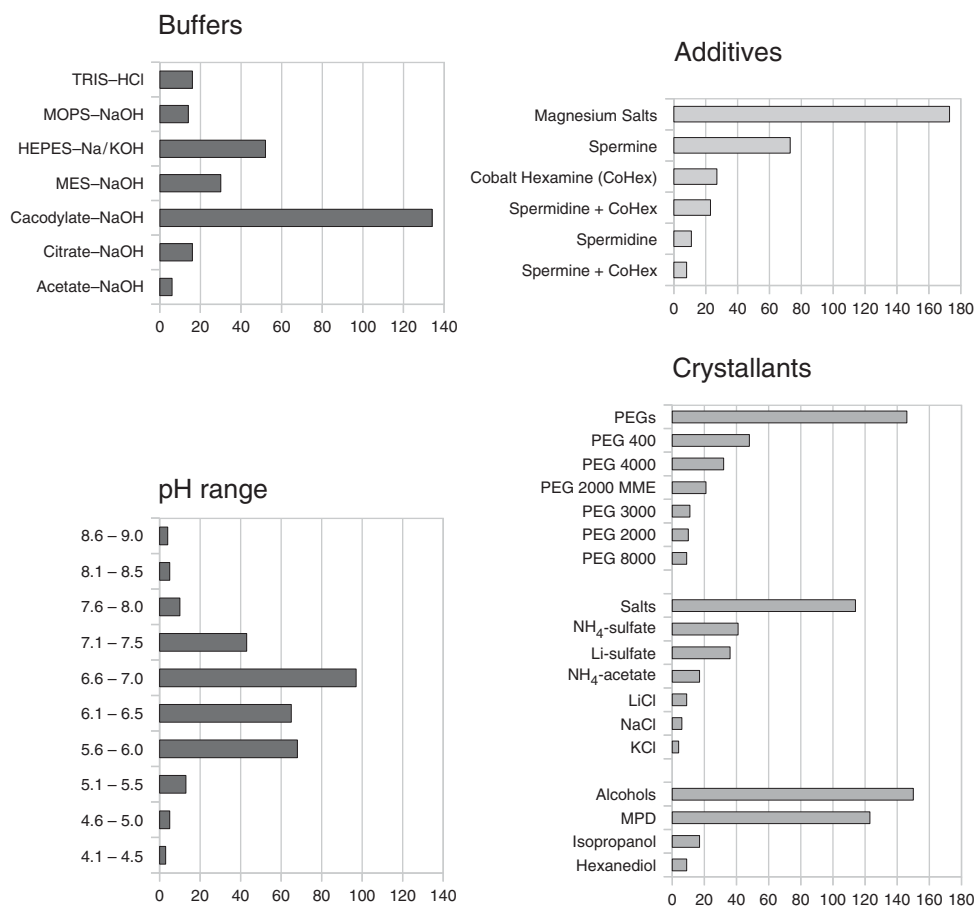


Figure 23.2 RNA crystallization conditions used for the structures deposited in the PDB. This figure indicates the most common buffers and pH values, crystallants (PEGs, salts, and alcohols), and specific additives – magnesium salts (MgCl_2 , $\text{Mg}(\text{OAc})_2$ and MgSO_4) and polyamines – used to crystallize RNA. These data were extracted from 356 structures (RNA only) available in the Protein Data Bank as of May 2010. They indicate a clear preference for slightly acidic pH values (which prevent RNA hydrolysis). A majority of structures was obtained from crystals grown in the presence

of a single crystallant. However, crystallants are often combined, such as PEGs and salts. The same observation is true for additives, for instance, spermine and cobalt hexamine (CoHex) are often used together. Noticeably, half of the structures were determined in the presence of magnesium ions that are known to stabilize RNA folds. MPD, 2-methyl-2,4-pentandiol; PEG, polyethylene glycol; and PEG 2000 MME = Polyethylene glycol monomethyl ether 2000. In the panel “Crystallants”, the bars for PEGs, salts, and alcohols are the sum of the individual bars listed below each of the three.

The initial search is often carried out using the method of vapor diffusion in sitting droplets made of a 1 : 1 mixture of screen and sample solutions. The droplets are then concentrated by the diffusion of the volatile species (vapor) toward the wells containing the concentrated test solution (Figure 23.1). This gentle equilibration process slowly drives the soluble RNA toward the supersaturated region of the phase diagram favorable to its crystallization. As mentioned above, crystallization robots have been developed to automate the preparation of the drops and to reduce their volume. Other methods that are commonly employed for RNA crystallization are microbatch or dialysis. Each of them leads to a different trajectory in the phase diagram. This property can be exploited during optimization to better explore the crystallization space [28].

Crystallization assays are stored at constant temperature to avoid any erratic change of growth regime. Indeed, temperature variations can severely affect crystal growth mechanisms and crystal quality [36]. A specificity of RNAs is that they often yield better results in crystallization at 30 or 37 °C. It is therefore recommended to run the same screen in parallel at 20 °C and at a higher temperature using a dedicated set of temperature-controlled rooms or incubators.

23.3.3

Evaluation of Crystallization Assays

First crystallization attempts rarely yield ready-to-use crystals. The results can range from large monocrystals to showers of microcrystals, precipitates, phase separations, or soluble drops (Figure 23.3). The optimization strategy will directly depend on the outcome of the screen. First, all drops must be inspected carefully

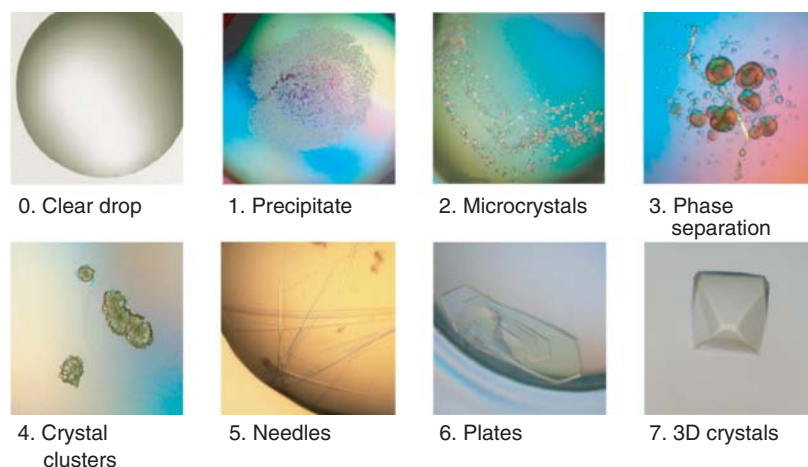


Figure 23.3 Visualization of crystallization droplets and associated numerical scoring terms. The scoring values are used to sort observations on excel-type spreadsheets in order to identify the successful conditions (Table 23.2). These can be used to design optimization screens.

using a binocular microscope hooked up to a digital camera to record observations. Second, a numerical scoring is assigned to each experiment to describe the content of the droplet. Dedicated databases or laboratory information management systems (LIMSS) have been designed along with automated imaging systems to facilitate both scoring and data mining. A low-cost version consists in collecting the observations in a spread sheet (Table 23.2). Screen conditions are sorted according to their score in order to highlight features (pH, crystallant) common to best hits.

Drop equilibration by vapor diffusion will last from a few days to more than a month depending on the crystallant and its concentration [37]. This is why droplets shall be inspected on a regular basis over a period of one to two months to follow up the appearance of crystals. In addition, RNAs or RNPs may start forming crystals long after equilibration because nucleation is a statistical process. Non-exhaustive guidelines to interpret the droplet content of crystallization screens and possible ways to optimize positive hits are provided below. See also [38] for more details.

- **Clear drop:** indicates that the RNA concentration is too low, the system is outside the nucleation zone (Figure 23.1). These experiments must be repeated with higher sample and/or salt concentrations. The temperature can also be lowered. A volume of concentrated salt solution can be added to the well to push the drop dehydration (concentration) by vapor diffusion.
- **Light precipitate:** indicates that the supersaturation is too high. Such a drop must be carefully inspected using polarized light with a magnification factor greater than 100, as it may contain a microcrystalline shower. Prepare new tests with a decreased RNA and/or crystallant concentration or dilute the droplet by vapor diffusion by means of adding water to the reservoir.
- **Strong precipitate:** may indicate that the sample has been partially denatured. The sample must be tested at a lower RNA concentration or less salt should be used. Note that a fresh test should be prepared in this case.
- **Phase separation:** indicates a need to increase the monovalent salt concentration and/or to decrease crystallant concentration (MPD, PEG) to make the RNA sample more soluble. The drop content can also be used for seeding.
- **Clusters of small crystals or needles:** try to slow down the nucleation and crystal growth by lowering RNA and crystallant concentrations or increase the monovalent salt concentration; test the same conditions at higher temperature, cover the reservoir solution with oil to slow down vapor diffusion, add agarose gel to get well-separated crystals in the gel matrix, and try seeding to trigger the nucleation of single crystals.
- **Large single crystals:** repeat the experiment to grow more samples and bring them to the next X-ray facility for a diffraction test!

23.3.4

The Optimization Process

The optimization of a first hit consists in modifying the crystallization parameters to define a zone in the phase diagram that yields the best diffracting crystals. A first

Table 23.2 Example of a screen evaluation spreadsheet. High-throughput (HT) crystallization screening kits are available in a 96-well format from various manufacturers (Hampton Research, Jena Biosciences, Qiagen). The results of the screening procedure are scored (Figure 23.3) and sorted to identify chemicals that enable successful crystallization of a given RNA. For example, the conditions can be sorted according to their score to highlight characteristic features such as a preference for low molecular weight PEG, pH 6–7, and a positive effect of spermine. The score is assigned to each assay after 1 (d1) and 30 days (d30). However, the time interval is at the user's discretion since crystals appear sometimes overnight.

Well	Buffer	Additive 1	Additive 2	Crystallant	Score d1	Score d30
C1	50 mM Na cacodylate pH 6.0	3 mM Spermine	15 mM MgCl ₂	10% (v/v) PEG 400	0	7
B12	50 mM Na HEPES pH 7.0	0.1 M KCl	10 mM CaCl ₂	10% (v/v) PEG 400	0	6
B10	50 mM MES pH 6.0	0.2 M KCl	10 mM MgSO ₄	10% (v/v) PEG 400	0	6
B11	50 mM MES pH 6.0	0.1 M KCl	10 mM MgCl ₂	10% (v/v) PEG 400	0	5
C2	50 mM Tris HCl pH 7.5	0.1 M KCl	15 mM MgCl ₂	10% (v/v) PEG MME 550	0	5
E10	50 mM Na HEPES pH 7.0	0.2 M KCl	25 mM MgSO ₄	20% (v/v) PEG 200	0	5
E12	50 mM Na cacodylate pH 6.5	0.1 M NaCl	200 mM MgCl ₂	20% (w/v) PEG 1000	0	4
D5	50 mM Na cacodylate pH 6.5	8 mM Mg(OAc) ₂		15% (v/v) PEG 400	0	4
C5	50 mM Na cacodylate pH 6.0	0.2 M KCl	10 mM CaCl ₂	10% (w/v) PEG 4000	0	4
C6	50 mM Na cacodylate pH 6.5	0.2 M KCl	10 mM MgCl ₂	10% (w/v) PEG 4000	0	4
C7	50 mM Na cacodylate pH 6.5	0.2 M NH ₄ OAc	10 mM CaCl ₂	10% (w/v) PEG 4000	0	3
F3	50 mM Na HEPES pH 7.0	5 mM MgCl ₂		25% (v/v) PEG MME 550	0	2
F7	50 mM Tris HCl pH 8.5	0.1 M KCl	10 mM MgCl ₂	30% (v/v) PEG 400	0	1
...	...	...	...	...	...	...

approach is a grid screening around the hit conditions: the RNA and crystallant concentrations are varied stepwise from -25 to $+25\%$ of their value in the screen, while keeping all other parameters constant. This will give a local description of the phase diagram and highlight regions at low supersaturation favorable to the growth of a limited number of large crystals (Figure 23.1). Temperature and pH may be adjusted in a second round of optimization. The effect of additives such as polyamines [39] or other cations (if not already present in the initial sample) should be probed, some of them being very useful for crystallographic phasing (cobalt hexamine, for instance).

Seeding is another powerful option. It consists in introducing a seed – a piece of crystal, microcrystals, or any type of material found in a drop containing a precipitate or separated phases – that may trigger nucleation at low supersaturation [40]. Seeding can rescue desperate cases when only clear drops or precipitates are observed and is recommended in conditions giving many small crystals. A set of drops are prepared under identical conditions, and they are sequentially seeded (after 0.5, 1, 2, 3 days, etc.) to determine the optimal supersaturation window during drop equilibration that leads to the growth of just a few large crystals.

Crystal growth and quality can also be improved by performing crystallization assays in agarose gel matrices (see [41] for a review). This reduces convection and the propensity for diffusive growth regime, thus making crystallization much less error-prone. Crystals grown within gel matrices have been shown to diffract X-rays more strongly and to present a less mosaic structure, which in turn facilitates the crystallographic analysis. In addition, gel-grown crystals may be less sensitive to temperature fluctuations, mechanical shock, or destabilization on soaking with a ligand, cryoprotectant, or heavy atom.

Such quasi-ideal convectionless conditions are also found in microfluidic systems that are powerful tools for screening and optimization using minimal amounts of sample [42]. In particular, the efficient miniaturization of crystallization methods such as free interface diffusion or counter-diffusion (Figure 23.1) in microchips provides excellent alternatives to vapor diffusion or batch procedures during the optimization step and the fine-tuning of crystal growth conditions. In addition, new generations of chips also allow the direct *in situ* crystal characterization by X-ray diffraction [43].

In the case of RNA crystallization, the target macromolecule itself is a parameter that can be optimized. As highlighted above, the sequence of the RNA can be designed to promote intermolecular contacts in regular crystal lattices. Small organic ligands can also be added to the RNA sample to modify its structural properties and improve its crystallizability. This is illustrated in the next section with the example of the aminoglycoside/ribosomal A-site complexes.

23.3.5

Designing RNA Constructs with Improved Crystallization Capabilities

If no crystals have been obtained during the first trials, it is often more efficient to vary the sequence instead of sampling a larger variety of crystallization conditions.

In the crystallization process, sequence and shape of the molecules will drive the nucleation and subsequent crystal growth through a network of packing interactions mediated among symmetry-related molecules constituting the asymmetric unit. With respect to RNA, these factors have even more drastic effect than for proteins, as the former are usually less globular in shape than the latter. Thus, various RNA constructs with different sequences and helix length, in other words with different shapes, should be tried when no crystals appear for a given construct [31, 44]. Chemical synthesis makes this process relatively straightforward for small RNAs (<30 nt). For longer RNAs, *in vitro* transcription of constructs flanked by ribozymes helps to control sequence homogeneity of the 5'- and 3'-ends (see Chapter 2 for details).

RNA structures can be seen as assemblies of building blocks of various shapes and complexities, which obey conservation rules at the level of sequence and structure. The 3D structure of an RNA or RNP is thus the consequence of building up intramolecular tertiary contacts between motifs that recognize each other specifically. At a larger scale, an RNA crystal can be seen as a supermolecular assembly of RNA molecules that form a regular 3D lattice through a recurrent pattern of intermolecular contacts. The structural motifs involved in tertiary interactions can thus in principle be used as tools to improve the capacity of RNAs to interact and to ultimately develop crystals.

Two interaction motifs have been proved useful to increase the crystallizability of RNA molecules. The first one is an RNA–RNA interaction motif, the GAAA tetraloop/tetraloop–receptor interaction motif historically unraveled by the crystal structure of the P4–P6 domain of the *Tetrahymena* group I ribozyme [45]. The second is an RNA–protein recognition motif consisting of a domain of the U1 spliceosomal RNA and the RNA-binding domain (RBD) of the U1A protein [46]. Conversion of an RNA to an RNP by inserting a U1A binding site has been a widely used strategy in RNA crystallization (see [47] for review). Since its proof of concept [48], the tetraloop/tetraloop–receptor motif has been less extensively utilized than the U1A system but was recently successfully applied to obtain well-diffracting crystals of a bacterial RNase P holoenzyme–tRNA complex [49]. Moreover, the A-minor motif [50, 51], a submotif of the GAAA tetraloop, has also shown its propensity to promote crystal packing through specific interactions with the shallow groove of symmetry-related molecules in the numerous minimal A-site crystal structures [52]. These observations support the idea that the A-minor motif may also be broadly applicable to improve the crystallizability of RNA molecules. Figure 23.4 gives an example of the structural diversity that can be achieved by bulging adenine tandems in different minimal A-site variants. Moreover, it has long been debated whether crystal structures can explain the structures that take place under physiological conditions. Evidence in favor of this idea is provided by the fact that biological macromolecules exist in highly crowded environments in living cells [53] where they adopt their native fold despite forming multiple interactions with other macromolecules. Crystal packing may in fact capture some of these potential contact sites and make them observable at atomic level.

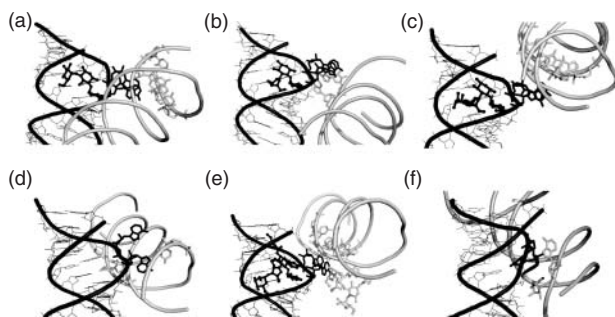


Figure 23.4 Various intermolecular crystal packing interactions in structures of A-site complexes. The black duplex is always at the same position to illustrate the different positions of the gray duplex in the individual crystal structures. The two consecutive bulged adenines are involved in

A-minor contacts, base pairings, and stacking interactions. Space group and PDB-ID of the individual structures are: (a) $P2_1$, 1LC4; (b) $P2_12_12$, 2ET5; (c) $P2_12_12_1$ form 1, 1J7T; (d) $P2_12_12_1$ form 2, 2O3X; (e) $P2_12_12_1$ form 3, 2PWT; and (f) $I4$, 3BNL.

23.3.6

Crystallizing Complexes with Organic Ligands: the Example of Aminoglycosides

Recent research has revealed that RNA function can rely on the binding of various organic ligands, as is the case for riboswitches [54]. Another class of ligands are antibiotics that interfere with the function of ribosomal RNA (rRNA). Different RNA fragments based on the A-site located in the penultimate helix of the small ribosomal RNA [52, 55–63] and derived from *E. coli*, human cytoplasmic and human mitochondrial rRNA, were investigated in the presence of their natural ligands, antibiotics of the aminoglycoside family. This work was part of an effort to better understand A-site architectures, their conformation switching, their different responsiveness to aminoglycosides, and the molecular basis of the non-syndromic hearing loss sometimes observed in recipients of aminoglycoside therapy. The RNA constructs were designed as self-complementary oligonucleotides so as to incorporate two A-sites in a head-to-head manner (Figure 23.5). This choice eliminates two drawbacks. First, as the internal loop is asymmetric, one would otherwise need to synthesize, purify, and mix 1 : 1 two different RNA strands in order to obtain a single site. Second, one could also use a single site capped by a stable hairpin of the GNRA family, for example. However, in such cases, the crystallized structure often reveals a full duplex with several non-Watson–Crick pairs, formed by complexation of two RNA molecules [64]. In order to take advantage of sequence variations for the crystallization of these complexes, different modifications have been performed, such as addition of a 5' UU overhang, insertion of a U–U pair, or moving the two A-sites closer to one another. An additional advantage of using a self-complementary oligonucleotide is to potentially solve the crystal structure of the A-site in two different conformations within the same crystal.

Routinely, the purified oligoribonucleotides are solubilized in a solution containing 2 mM RNA, 25 mM NaCl, 5 mM MgSO_4 , and 100 mM sodium cacodylate

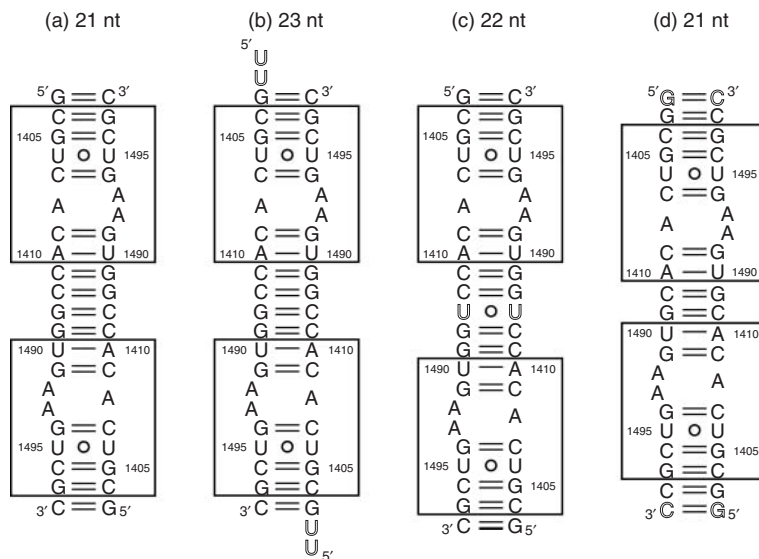


Figure 23.5 Secondary structures of four self-complementary RNA fragments designed for crystallization. The two *E. coli* 16S rRNA A-site motifs, flanked by G,C-rich stems, are boxed. Modifications deviating from the (a)

original sequence include (b) a 5'-UU overhang, (c) an additional central U-U pair, and (d) additional terminal G=C base pairs and a shortened central stem. The standard *E. coli* numbering is specified.

buffer pH 6.4. This solution is first heated to 85 °C for 2 min and then slowly cooled until a temperature of 37 °C is reached. One volume of a 4 mM aminoglycoside solution (equilibrated to 37 °C as well) is added to the RNA solution, followed by incubation for 2 h at 21 °C. As the RNA fragments contain two antibiotic binding sites, the aminoglycoside concentration is twice the RNA concentration. A general rule is that the organic ligand concentration should be 100 times higher than its dissociation constant (K_d) to ensure binding-site saturation, which is usually easy to achieve in the millimolar range.

In the example of the aminoglycoside/A-site complexes, different crystallization solutions were prepared to test various glycerol/MPD ratios. Temperatures of 37 or 20 °C were used to crystallize the fragments without affecting the diffraction power of the crystals. Using the vapor diffusion method in the hanging drop setup, the following conditions were successful: 1 μ l of solution containing the RNA-antibiotic complex is added to 1 μ l crystallization solution and equilibrated over a 40% MPD reservoir.

23.4

Conclusions

Protocols to purify and concentrate large amounts of RNA under controlled buffer and salt conditions for crystallization experiments have been described. The crucial

points to preserve the RNA sample are: the use of slightly acidic pH and the avoidance of divalent cations. Usually RNases, feared by most RNA scientists, are introduced into the solution by an upstream experiment such as plasmid and protein preparations (e.g., T7 RNA polymerase). It is recommended to assess the RNase activity of a solution before using it on the whole RNA sample. This is best done by mixing aliquots of solution and RNA, followed by incubation for a few hours and PAGE analysis to check for RNA integrity. Strategies to crystallize RNA oligonucleotides in the presence or absence of ligands have been presented.

A peculiarity of crystallization experiments is the absence of a negative control. To circumvent this, it is advisable to attempt to crystallize simultaneously several related RNA sequences in the same condition. If only one or a few of the tested RNA variants give rise to crystals, one can infer that those are most probably RNA crystals and not salt crystals. In cases where no crystal is observed, it is recommended to design a new set of RNAs bearing slight sequence changes and to incorporate RNA–RNA interaction motifs or a U1A protein binding site in order to enhance interactions between symmetry-related molecules that lead to regular crystal packing contacts.

References

1. Wimberly, B.T., Brodersen, D.E., Clemons, W.M. Jr., Morgan-Warren, R.J., Carter, A.P., Vonnrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
2. Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
3. Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H., and Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883–896.
4. Harms, J., Schlutzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001) High-resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, **107**, 679–688.
5. Torres-Larios, A., Swinger, K.K., Krasilnikov, A.S., Pan, T., and Mondragon, A. (2005) Crystal structure of the RNA component of bacterial ribonuclease P. *Nature*, **437**, 584–587.
6. Kazantsev, A.V., Krivenko, A.A., Harrington, D.J., Holbrook, S.R., Adams, P.D., and Pace, N.R. (2005) Crystal structure of a bacterial ribonuclease P RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 13392–13397.
7. Golden, B.L., Kim, H., and Chase, E. (2005) Crystal structure of a phage Twort group I ribozyme-product complex. *Nat. Struct. Mol. Biol.*, **12**, 82–89.
8. Guo, F., Gooding, A.R., and Cech, T.R. (2004) Structure of the *Tetrahymena* ribozyme: base triple sandwich and metal ion at the active site. *Mol. Cell*, **16**, 351–362.
9. Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J., and Strobel, S.A. (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature*, **430**, 45–50.
10. Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. (2008) Crystal structure of a self-spliced group II intron. *Science*, **320**, 77–82.
11. Shechner, D.M., Grant, R.A., Bagby, S.C., Koldobskaya, Y., Piccirilli, J.A., and Bartel, D.P. (2009) Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science*, **326**, 1271–1275.

12. Leontis, N.B., Stombaugh, J., and Westhof, E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
13. Leontis, N.B., Stombaugh, J., and Westhof, E. (2002) Motif prediction in ribosomal RNAs Lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie*, **84**, 961–973.
14. Masquida, B., Beckert, B., and Jossinet, F. (2010) Exploring RNA structure by integrative molecular modelling. *New Biotechnol.*, **27**, 170–183.
15. Zuker, M. and Jacobson, A.B. (1998) Using reliability information to annotate RNA secondary structures. *RNA*, **4**, 669–679.
16. Isambert, H. and Siggia, E.D. (2000) Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 6515–6520.
17. Ding, Y. and Lawrence, C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.*, **31**, 7280–7301.
18. Usman, N. and Cedergren, R. (1992) Exploiting the chemical synthesis of RNA. *Trends Biochem. Sci.*, **17**, 334–339.
19. Mueller, U., Muller, Y.A., Herbst-Irmer, R., Sprinzl, M., and Heinemann, U. (1999) Disorder and twin refinement of RNA heptamer double helices. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 1405–1413.
20. Anderson, A.C., O'Neil, R.H., Filman, D.J., and Frederick, C.A. (1999) Crystal structure of a brominated RNA helix with four mismatched base pairs: An investigation into RNA conformational variability. *Biochemistry*, **38**, 12577–12585.
21. Du, Q., Carrasco, N., Teplova, M., Wilds, C.J., Egli, M., and Huang, Z. (2002) Internal derivatization of oligonucleotides with selenium for X-ray crystallography using MAD. *J. Am. Chem. Soc.*, **124**, 24–25.
22. Pallan, P.S. and Egli, M. (2007) Selenium modification of nucleic acids: preparation of oligonucleotides with incorporated 2'-SeMe-uridine for crystallographic phasing of nucleic acid structures. *Nat. Protoc.*, **2**, 647–651.
23. Pallan, P.S. and Egli, M. (2007) Selenium modification of nucleic acids: preparation of phosphoroselenoate derivatives for crystallographic phasing of nucleic acid structures. *Nat. Protoc.*, **2**, 640–646.
24. Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. Usman, N., Egli, M., and Rich, A. (1992) Large scale chemical synthesis, purification and crystallization of RNA-DNA chimeras. *Nucleic Acids Res.*, **20**, 6695–6699.
26. Anderson, A.C., Scaringe, S.A., Earp, B.E., and Frederick, C.A. (1996) HPLC purification of RNA for crystallography and NMR. *RNA*, **2**, 110–117.
27. Golden, B.L. and Kundrot, C.E. (2003) RNA crystallization. *J. Struct. Biol.*, **142**, 98–107.
28. Sauter, C., Lorber, B., McPherson, A., and Giege, R. (2012) in *International Tables for Crystallography*, Crystallography of Biological Macromolecules, Vol. F (eds E. Arnold, D.A. Himmel, and M. Rossmann), John Wiley & Sons, Ltd, Chichester, 99–120.
29. Doudna, J.A., Grosshans, C., Gooding, A., and Kundrot, C.E. (1993) Crystallization of ribozymes and small RNA motifs by a sparse matrix approach. *Proc. Natl. Acad. Soc. U.S.A.*, **90**, 7829–7833.
30. Baeyens, K.J., Jancarik, J., and Holbrook, S.R. (1994) Use of low-molecular-weight polyethylene glycol in the crystallization of RNA oligomers. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 764–767.
31. Scott, W.G., Finch, J.T., Grenfell, R., Fogg, J., Smith, T., Gait, M.J., and Klug, A. (1995) Rapid crystallization of chemically synthesized hammerhead RNAs using a double screening procedure. *J. Mol. Biol.*, **250**, 327–332.
32. Berger, I., Kang, C.H., Sinha, N., Wolters, M., and Rich, A. (1996) A highly efficient 24-condition matrix for the crystallization of nucleic acid fragments. *Acta Crystallogr.*, **D52**, 465–468.

33. Cate, J.H. and Doudna, J.A. (1997) A sparse matrix approach to crystallizing ribozymes and RNA motifs. *Methods Mol. Biol.*, **74**, 379–386.
34. Golden, B.L., Podell, E.R., Gooding, A.R., and Cech, T.R. (1997) Crystals by design: a strategy for crystallization of a ribozyme derived from the *Tetrahymena* group I intron. *J. Mol. Biol.*, **270**, 711–723.
35. Nowakowski, J., Shim, P.J., Joyce, G.F., and Stout, C.D. (1999) Crystallization of the 10-23 DNA enzyme using a combinatorial screen of paired oligonucleotides. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 1885–1892.
36. Ng, J.D., Kuznetsov, Y.G., Malkin, A.J., Keith, G., Giege, R., and McPherson, A. (1997) Visualization of RNA crystal growth by atomic force microscopy. *Nucleic Acids Res.*, **25**, 2582–2588.
37. Mikol, V., Rodeau, J.L., and Giegé, R. (1990) Experimental determination of water equilibration rates in the hanging drop method of protein crystallization. *Anal. Biochem.*, **186**, 332–339.
38. Ducruix, A. and Giegé, R. (1999) *Crystallization of Nucleic Acids and Proteins: a Practical Approach*, IRL Press, Oxford.
39. Sauter, C., Ng, J.D., Lorber, B., Keith, G., Brion, P., Hosseini, W., Lehn, J.M., and Giegé, R. (1999) Additives for the crystallization of proteins and nucleic acids. *J. Cryst. Growth*, **196**, 365–376.
40. Bergfors, T. (2003) Seeds to crystals. *J. Struct. Biol.*, **142**, 66–76.
41. Lorber, B., Sauter, C., Theobald-Dietrich, A., Moreno, A., Schellenberger, P., Robert, M.C., Capelle, B., Sanglier, S., Potier, N., and Giege, R. (2009) Crystal growth of proteins, nucleic acids, and viruses in gels. *Prog. Biophys. Mol. Biol.*, **101**, 13–25.
42. Sauter, C., Dhoubib, K., and Lorber, B. (2007) From Macrofluidics to microfluidics for the crystallization of biological macromolecules. *Cryst. Growth Des.*, **7**, 2247–2250.
43. Dhoubib, K., Khan Malek, C., Pfleging, W., Gauthier-Manuel, B., Duffait, R., Thuillier, G., Ferrigno, R., Jacquamet, L., Ohana, J., Ferrer, J.L. *et al.* (2009) Microfluidic chips for the crystallization of biomacromolecules by counter-diffusion and on-chip crystal X-ray analysis. *Lab Chip*, **9**, 1412–1421.
44. Anderson, A.C., Earp, B.E., and Frederick, C.A. (1996) Sequence variation as a strategy for crystallizing RNA motifs. *J. Mol. Biol.*, **259**, 696–703.
45. Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., and Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1684.
46. Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
47. Ferré-D'Amaré, A.R. (2010) Use of the spliceosomal protein U1A to facilitate crystallization and structure determination of complex RNAs. *Methods*, **52**, 159–167.
48. Ferré-D'Amaré, A.R., Zhou, K.H., and Doudna, J.A. (1998) A general module for RNA crystallization. *J. Mol. Biol.*, **279**, 621–631.
49. Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B., and Steitz, T.A. (2001) RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 4899–4903.
50. Reiter, N.J., Osterman, A., Torres-Larios, A., Swinger, K.K., Pan, T., and Mondragon, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*, **468**, 784–789.
51. Doherty, E.A., Batey, R.T., Masquida, B., and Doudna, J.A. (2001) A universal mode of helix packing in RNA. *Nat. Struct. Biol.*, **8**, 339–343.
52. Kondo, J. and Westhof, E. (2008) The bacterial and mitochondrial ribosomal A-site molecular switches possess different conformational substates. *Nucleic Acids Res.*, **36**, 2654–2666.
53. Miyoshi, D. and Sugimoto, N. (2008) Molecular crowding effects on structure and stability of DNA. *Biochimie*, **90**, 1040–1051.
54. Tucker, B.J. and Breaker, R.R. (2005) Riboswitches as versatile gene control

- elements. *Curr. Opin. Struct. Biol.*, **15**, 342.
55. Vicens, Q. and Westhof, E. (2001) Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure (Camb)*, **9**, 647–658.
 56. Vicens, Q. and Westhof, E. (2002) Crystal structure of a complex between the aminoglycoside tobramycin and an oligonucleotide containing the ribosomal decoding A site. *Chem. Biol.*, **9**, 747–755.
 57. Vicens, Q. and Westhof, E. (2003) Crystal structure of geneticin bound to a bacterial 16S ribosomal RNA A site oligonucleotide. *J. Mol. Biol.*, **326**, 1175–1188.
 58. Francois, B., Russell, R.J., Murray, J.B., Aboul-ela, F., Masquida, B., Vicens, Q., and Westhof, E. (2005) Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. *Nucleic Acids Res.*, **33**, 5677–5690.
 59. Kondo, J., Francois, B., Russell, R.J., Murray, J.B., and Westhof, E. (2006) Crystal structure of the bacterial ribosomal decoding site complexed with amikacin containing the gamma-amino-alpha-hydroxybutyryl (haba) group. *Biochimie*, **88**, 1027–1031.
 60. Kondo, J., Francois, B., Urzhumtsev, A., and Westhof, E. (2006) Crystal structure of the *Homo sapiens* cytoplasmic ribosomal decoding site complexed with apramycin. *Angew. Chem. Int. Ed.*, **45**, 3310–3314.
 61. Kondo, J., Urzhumtsev, A., and Westhof, E. (2006) Two conformational states in the crystal structure of the *Homo sapiens* cytoplasmic ribosomal decoding A site. *Nucleic Acids Res.*, **34**, 676–685.
 62. Kondo, J., Hainrichson, M., Nudelman, I., Shallom-Shezifi, D., Barbieri, C.M., Pilch, D.S., Westhof, E., and Baasov, T. (2007) Differential selectivity of natural and synthetic aminoglycosides towards the eukaryotic and prokaryotic decoding A sites. *Chembiochem*, **8**, 1700–1709.
 63. Kondo, J., Pachamuthu, K., Francois, B., Szychowski, J., Hanessian, S., and Westhof, E. (2007) Crystal structure of the bacterial ribosomal decoding site complexed with a synthetic doubly functionalized paromomycin derivative: a new specific binding mode to an A-minor motif enhances *in vitro* antibacterial activity. *ChemMedChem*, **2**, 1631–1638.
 64. Kacer, V., Scaringe, S.A., Scarsdale, J.N., and Rife, J.P. (2003) Crystal structures of r(GGUCACAGCCC)2. *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 423–432.

24

Studying RNA Using Single Molecule Fluorescence Resonance Energy Transfer

*Felix Spenkuch, Olwen Domingo, Gerald Hinze, Thomas Basché,
and Mark Helm*

24.1

Introduction

24.1.1

The Advantages of Single Molecule Fluorescence Resonance Energy Transfer

RNA is intrinsically structured, and biological RNA structures are dynamic “by nature.” Pioneering work, for example, on the function of riboswitches in gene regulation [1] and the complicated dynamics of the ribosome during translation [2], has alerted many to the fact that structural dynamics are an important and ubiquitous feature of all RNAs. Among the numerous biochemical and biophysical methods commonly applied to structural dynamics, many suffer from serious restrictions with respect to size and quantity of the RNA under investigation. NMR (nuclear magnetic resonance, [3]), ESR (electron spin resonance, [4]), and DEER (double electron electron resonance [5]) shall serve as examples here. Methods based on fluorescence, such as fluorescence resonance energy transfer (FRET), are attractive because they combine small sample size with low RNA quantities, despite some drawbacks associated with, for example, bulk FRET experiments [6]. While typical sample amounts for NMR, ESR, and DEER lie in the micromole-range [7], fluorescence in bulk typically requires only about 1 pmol of material with conventional equipment.

The aforementioned approaches are so-called bulk or ensemble methods and do exhibit significant differences in comparison to single molecule experiments. However, fluorescence setups including confocal microscopy or prism-based wide-field microscopy, and particular detectors such as Avalanche Photodiodes (APDs), sensitive cameras, or spectrographs allow single molecule detection *via* fluorophores [8]. Such setups are able to visualize the dynamics of individual molecules and therewith subpopulations of an ensemble. These subpopulations get mixed in bulk experiments, resulting in an average over the whole ensemble, and accompanied by a total loss in sensitivity with respect to subpopulations of very low abundance.

Viewed from this perspective, it is obvious that a major breakthrough was achieved by the development of single molecule FRET (smFRET). This method

combines the advantages of FRET with the ability of a single molecule technique to resolve individual dynamics and subpopulations [9]. In confocal microscopy, two principal strategies toward smFRET experiments are commonly employed, with a time resolution in the 100 μ s range or better. The first strategy targets freely diffusing fluorescent molecules during their millisecond-short sojourn in the observation volume of a confocal microscope and is termed “burst mode” [10]. The second approach to smFRET uses immobilized molecules, which expands the observation time scale from milliseconds to minutes.

24.1.1.2

Chapter Scope

We focus here on smFRET approaches utilizing immobilized RNA molecules. At first, a brief overview on examples of RNA dynamics successfully addressed by smFRET is presented. After a short introduction into Förster theory, considerations for FRET construct design and dye choice are discussed, followed by the layout of the instrumental setup used by our group, including a summary of its advantages. On the experimental side, we (i) present methods for dye labeling of RNA constructs, (ii) discuss measurement chambers with respect to cleaning, assembly, and treatment, (iii) mention buffers that enhance performance, and (iv) describe the actual sample preparation. As smFRET experiments can be troublesome a section addressing possible problems closes the chapter.

Although the instrumental setup for smFRET experiments is quite demanding, the method and its applications has spread because of the quality and quantity of information obtained. Accordingly, several fundamentally different setups have been developed in the last years. This chapter puts the focus on a setup targeting surface-immobilized molecules by a confocal microscope coupled with two APDs and a spectrograph. Therefore, readers interested in smFRET of freely diffusing molecules by alternating laser excitation [11] in so-called burst measurements are referred to other overviews and method papers [10, 12]. Also, an introduction into wide-field microscopy on immobilized molecules using a prism-based total internal reflection fluorescence (TIRF) microscope and a CCD camera can be found elsewhere [13]. Undoubtedly, the most complicated form of FRET regarding construct design, experimental implementation, instrumental setup, and data analysis is multi-color FRET [14]; although this technique is not further detailed here, we have summarized in Table 24.2 some applicable dye combinations for three- to four-color FRET experiments.

24.1.1.3

Typical Topics of RNA Dynamics Addressed by Single Molecule FRET

The broad variety of RNA molecules that had been investigated by smFRET until 2008 included RNase P RNA, telomerase RNA, the Varkud satellite ribozyme, the *Tetrahymena* group I intron, the bI5 group I intron, the theophylline-activated hammerhead ribozyme, the hairpin ribozyme, the adenine riboswitch, and small

RNA motives such as the rRNA 3-way-junction, the kink-turn motif, and the Tat-tar motif (reviewed in Helm *et al.* [12]). Recently, the *Tetrahymena* ribozyme was found to possess multiple active states, which are interconvertible by unfolding and re-folding [15]. More recent studies also included riboswitches, for example, SAM riboswitches of type I [16] and II [17]. The field has clearly moved to the study of RNA-protein interactions, such as the DEAD-box helicase Mss-116 acting on RNA substrates [18] or telomerase [19]. Undoubtedly, the most challenging application of smFRET using immobilized substrates would at present be the investigation of the conformation and function of the ribosome during translation. Owing to its complexity, this system offers to apply a broad range of FRET-based approaches, and it is clearly impossible to discuss all of them here [20, 21]. Dye labeling of both ribosomal subunits made it possible to analyze intersubunit conformations by smFRET, demonstrating that during each elongation cycle, the ribosome unlocks upon peptide bond formation and then reverts to the locked state upon translocation onto the next codon; in addition, these analyses even revealed specific and cumulative effects of antibiotics [22]. Furthermore, smFRET on pretranslocation (PRE) complexes with dye labels positioned on tRNA and the ribosomal L1 stalk led to the observation that tRNA hybrid state formation and L1 stalk closure, events central to the unlocking mechanism, are not tightly coupled, but that the extent of conformational coupling rather depends on the nature of the tRNA substrates [23]. Fei *et al.* characterized the dynamics of ribosomal PRE complexes, carrying either wild-type or systematically mutagenized tRNAs, and found that the conformational flexibility of the tRNA molecule has a crucial role in directing the structural dynamics of the PRE complex during translocation [24]. Chen *et al.* investigated the dissociation of deacetylated tRNA from the exit site during early protein synthesis, revealing that the tRNA is often allosterically released when the next aminoacyl tRNA binds to the A site of the ribosome. This process is regulated by the length and sequence of the nascent peptide, leading to spontaneous dissociation of E-site tRNA in later cycles [25].

Our personal experiences concern small RNA molecules, which were investigated using internal labeling, and either immobilized molecules or the burst technique. Dye-labeled RNA constructs were assembled from multiple oligonucleotides by splint ligation (see Chapter 3), and immobilized for FRET analyses. This enabled Kobitski *et al.* to show that the Diels-Alderase ribozyme possesses three different states of folding that exhibit Mg^{2+} -dependent cooperative transitions while continuous fluctuations between an intermediate and the active state remain [26]. The work was expanded by Kraut *et al.* who, using the burst mode approach, revealed that three hydrogen bonds are critical for the catalytic activity of the ribozyme [27]. Using immobilized mitochondrial tRNA^{Lys}, Voigts-Hoffmann *et al.* discovered that the m¹A9 modification controls the conformational equilibrium of this tRNA [28] by favoring formation of its functional cloverleaf shape [29]. By carrying on these burst mode folding studies, Dammertz *et al.* could establish a thermodynamic model that quantitatively describes RNA tertiary structure stabilization by different types of cations [30]. Finally, the individual contributions of post-transcriptional modifications of tRNA^{Lys} to folding and stabilization were analyzed, showing an

unexpected destabilizing effect of two pseudouridines as opposed to an unprecedented cooperative effect of the m¹A9 and the m²G10 modifications in stabilizing the biologically functional cloverleaf shape [31].

24.2

Theory of Fluorescence Resonance Energy Transfer

Fluorescence Resonance Energy Transfer or FRET denotes a strong distance-dependent radiationless transfer of excitation energy from a donor dye to an acceptor dye. Measurement of the relative fluorescence of both dyes allows us to determine the spatial separation of the dyes under certain circumstances, because of the fact that the efficiency of the transfer is reciprocally proportional to the radius to the power of 6. This distance dependency is usually described by means of the so-called Förster radius R_0 , which is the most important property of a pair of FRET fluorophores and defines the distance at which the fluorescence energy of the donor fluorophore is transferred to the acceptor with 50% efficiency.

$$E_{\text{FRET}} = R_0^6 / (R_0^6 + R^6) = \frac{I_A}{I_A + I_D} = 1 - \frac{I_D}{I_A + I_D} \quad (24.1)$$

Here E_{FRET} is the efficiency of FRET-based energy transfer, R_0 is the Förster radius (see Figure 24.1), and R is the actual inter-dye distance. I_A and I_D denote the intensity of acceptor-emitted and donor-emitted light, respectively. In the experiment, E_{FRET} is a value calculated from the donor intensity and the acceptor intensity as specified in Eq. (24.1).

Two dyes form a usable FRET pair if a sufficient spectral overlap exists between the emission spectrum of the donor and the absorption spectrum of the acceptor. The Förster radius R_0 , which is characteristic for a given FRET pair, is measured in Angstrom. It depends on the quantum yield of the donor, Q_D , and the spectral overlap of donor emission and acceptor excitation spectra,

$$J(\lambda) = \int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^2 d\lambda = \frac{\int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda}{\int_0^\infty F_D(\lambda) d\lambda},$$

and, in addition, on the refractive index of the optical medium, n , and the orientation factor, κ^2 , which accounts for the relative orientation of the respective transition dipoles of both dyes [32],

$$R_0 = 0.211(\kappa^2 n^{-4} Q_D J(\lambda))^{1/6} \quad (24.2)$$

To detect intramolecular dynamics such as an interchange between two conformations, the dye pair should be attached to the RNA so that the difference in E_{FRET} between the two states (*i.e.*, the FRET change) is maximized. Thus, the fluorophores should be separated by 0.7–1.3 R_0 in at least one of the two conformations. In this range, the slope in Figure 24.1 is steepest and, consequently, the FRET efficiency is most sensitive to distance changes. Above 1.5 R_0 or below 0.5 R_0 , the transfer efficiency approaches 0 or 1, respectively, and sensitivity to distance changes is completely lost. As the structural models used as the basis for construct design are often static X-ray structures, starting construct designs with

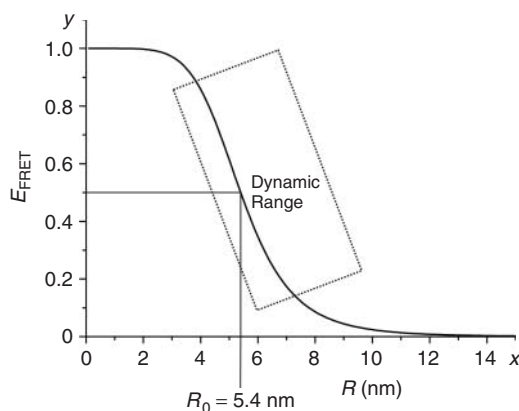


Figure 24.1 Distance dependency of the FRET efficiency E_{FRET} on the inter-dye distance R for a fictive dye pair with a Förster radius R_0 of 5.4 nm. A strong distance dependency can only be observed near the Förster radius as indicated by the dotted rectangle named “dynamic range.”

a dye distance of R_0 maximize the chance to observe alternative conformations. If structural information is available on two alternative conformations, attachment sites should be chosen such that the inter-dye distance passes from below to above R_0 , or vice versa, during conformational interconversion.

As the above-mentioned spectral properties that determine the Förster radius are affected by the medium and the direct chemical environment, potentially *via* the orientation factor κ^2 , an initial bulk spectroscopic characterization of the FRET construct is mandatory (see “Initial Spectroscopic Characterization by Bulk FRET Measurements”). Frequently, a value of $2/3$ is assumed for κ^2 , which corresponds to free rotation of both dyes. To facilitate free rotation, one should attach fluorophores by linkers and avoid 3'- or 5'-ends as attachment sites to prevent stacking of the dye to the bases (see Section 24.4.5). Free rotation can be proven by fluorescence anisotropy measurements [32]. For the above reasons, the Förster radii of fluorophore pairs given in Table 24.1 are approximate, and it may be necessary to determine more precise values for a particular system. Also, studies that satisfactorily clarify the influence of the refractive index are lacking, as the space between the interacting dyes may contain organic molecules or aqueous solution, both of which differ strongly in their refraction. Additional complications arise if the RNA does not assume a well-defined structure, but rather fluctuates among many different conformations.

24.3

Experimental Design

24.3.1

Considerations for Construct Design

Investigation of an RNA by smFRET starts with the search for potential dye attachment sites, which is ideally based on an X-ray or NMR structure, but requires

Table 24.1 The Förster radii of several FRET dye pairs used in investigations of nucleic acids.

Dye 1	Dye 2	R_0 (nm)	References
Alexa 350	Alexa 488	5.0	[5a]
Alexa 488	Cy3	3.8	[14e]
FAM	TMR	5.0	[44]
Alexa 488	Cy5	~5.1–5.4	[14e, 45]
Alexa 488	Cy7	~6.8	[14e]
Alexa 488	TMR	6.2	[14d]
Alexa 488	Alexa 647	5.6	[14d]
FAM	Cy5	3.9	[46]
Cy2	Cy3	~4.7	[14e]
Cy2	Cy5	~3.8	[14e]
Cy2	Cy7	~2.8	[14e]
Cy3	Cy5	5.4 or 4.7–6.5	[14d, 35, 47]
Cy3	Alexa 647	5.1	[48]
Alexa 546	Alexa 647	6.6	[49]
TMR	Alexa 647	6.2–6.6	[14b]
TMR	Atto647N	~6.7	[50]
TMR	Cy5	5.3	[51]
Cy3	Cy5.5	5.9	[14a]
Cy3	Cy7	3.8	[14d]
Cy3b	Atto647N	6.7	[52]
Fluorescein	Tetramethylrhodamine	5.0–5.5	[53]
TMR	Cy5	6.5–7.1	[35, 54]
Cy5	Cy5.5	7.3	[14a]
Cy5	Cy7	6.2	[14d]
Atto 610	Atto 680	~7.1	[55]

at least an educated estimate of the spatial distance between the prospective labeling sites. As outlined above, attachment sites for a given pair of FRET dyes are ideally separated by a distance corresponding to the Förster radius, R_0 , of the respective dye pair, such that the expected FRET efficiency is around 50%. Attachment sites with significantly different distances will yield a FRET signal that responds to changes of R with reduced sensitivity. The sites are preferably chosen inside helices to minimize potential interference of the dye with RNA structure and function, as well as to prevent an influence on folding pathways. Experimentally determined three-dimensional structures, or at least biochemically validated models of such structures are required to position dye pairs such that folding interference or suboptimal distances are avoided. Of note, intermolecular interactions with proteins, as, for example, in aminoacylation of tRNAs, can be severely disrupted depending on the label position, although the RNA structure itself may remain unaffected.

It is mandatory to compare the functional performance of the fluorophore-labeled construct relative to the unlabeled one. This allows one to detect whether the label

perturbs function, and if yes, whether this effect is of acceptable magnitude. In the case of tRNA, this can be addressed by analyzing the dye-labeled variant as substrate of tRNA modification enzymes [28]. Another approach uses aminoacylation and ribosomal translation assays. Here, the fidelity of the labeled tRNA compared to an unlabeled variant in ^{14}C -based aminoacylation assays is determined [33]. Once aminoacylated, the activity of such dye-labeled tRNAs can be tested in ensemble ribosome assays before their use in smFRET studies [34]. In addition, the dye-labeled RNA should be tested regarding its performance as FRET construct, which can be done by bulk FRET measurements (see “Initial Spectroscopic Characterization by Bulk FRET Measurements”). It has proved useful to design backup attachment sites for both dyes, in case the original construct turns out to be unsuitable. Construction of differently labeled variants is easily achieved by the splint ligation approach (see Chapter 3).

24.4

smFRET Experiments Using Immobilized Molecules

24.4.1

Instrumental Setup

In the home-built microscope setup depicted in Figure 24.2, the FRET signal, generated by excitation at 532 nm with a frequency-doubled ND-YAG laser, is detected twice. First, a split of the beam is reflected onto a spectrograph, which yields single-molecule spectra (Figure 24.2, top left inset) whose definition increases with the time during which photons are collected to compose a spectrum, that is, the binning time. The remainder of the split is registered by two APDs 1 and 2, one for the donor and one for the acceptor dye. The software interface is home-written in LabView. This setup is unique in combining the detection speed and sensitivity of APDs with the robust high-quality spectral information of a spectrograph. This allows for unambiguous spectral identification, while maintaining an excellent time resolution.

24.4.2

Means of Signal Correction and Data Analysis

The simplest way to correct a FRET signal is the introduction of a correction factor γ , which includes quantum yields of the fluorophores, spectral cross-talk, and differential detection efficiencies of the two channels. Once this factor is known, corrected E_{FRET} values can be obtained by using the following equation

$$E_{\text{FRET}} = \frac{I_A}{I_A + \gamma I_D} \quad (24.3)$$

Sabanyagam *et al.* [35] have used an approach that also corrects for leakage of the donor emission into the acceptor channel (“bleedthrough”) by using a factor β

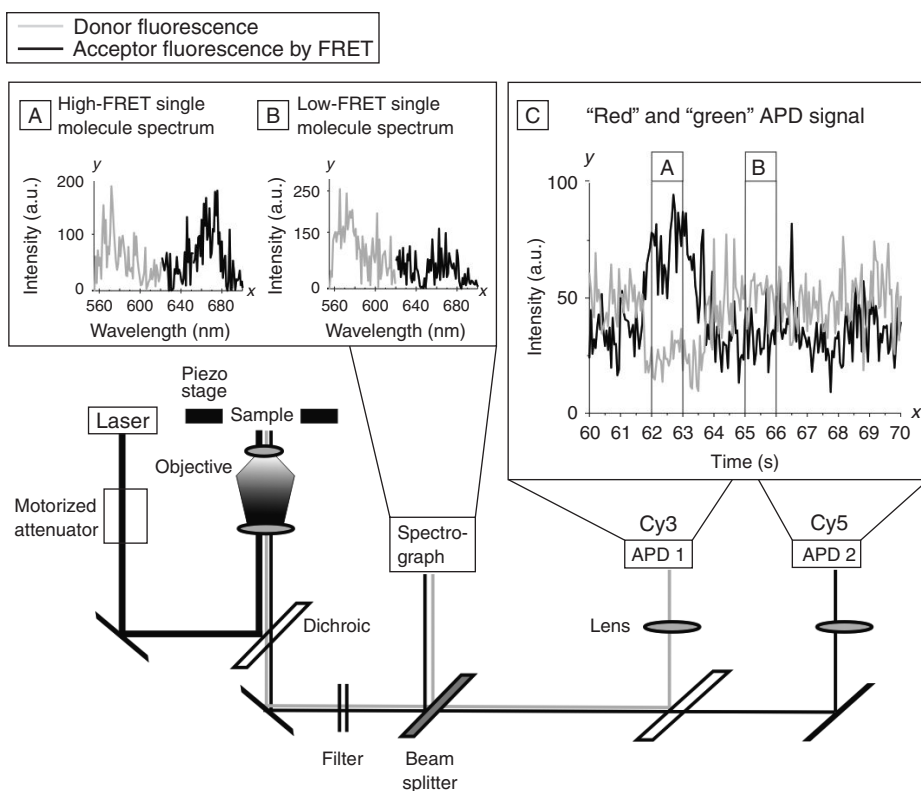


Figure 24.2 Scheme of our home-built microscope setup. The 532-nm laser line of a frequency-doubled ND-YAG laser is attenuated and channeled into the light-tight microscope box. Both excitation and emission light pathways are directed through a 1.49 NA (Numerical Aperture) objective. A dichroic mirror allows only emitted light to enter the detection pathway, where a beam-splitter cube promotes splitting of the signal between the spectrograph and two APDs. A second dichroic mirror reflects the donor emission light toward one detector (APD 1), and passes the acceptor emission light to a second one (APD 2). The inset C shows

instead of γ in the following equation:

$$E_{\text{FRET}} = \frac{I_A - \beta I_D}{I_A - \beta I_D + I_D} \quad (24.4)$$

Ha *et al.* [36] simply applied Eq. (24.1) as Cy3 and Cy5 resulted in a γ of 1 in their instrumental setup.

the data resulting from the APDs—one data point represents the photons collected in 0.05 s; dividing a dataset into time intervals is called “binning”; binning is used here to achieve a balance between an acceptable signal-to-noise ratio (favored by larger time intervals = bins) and sufficient time resolution (favored by smaller bins); insets A and B on the left show spectrograph data binned over the indicated time period depicted in C (1 s each); here, binning over 1 s was used as a compromise between acceptable signal-to-noise ratio and satisfactory spectral resolution.

In our instrumental setup, we use the robust signal of the spectrograph to correct the APD signals regarding their focus and wavelength dependency. This is done for every trace individually, thereby providing excellent correction under all circumstances. Still, the effect of the used objective filters and dichroic mirrors has to be taken into account. This is achieved by calibrating the spectrograph, using a certified third instrument. Bleedthrough of donor dye fluorescence into the acceptor channel is corrected by examining donor-only molecules, for example, by recording single molecule traces after the acceptor has been bleached, and extracting an appropriate correction factor from this analysis. This correction factor is then applied in Eq. (24.5).

From such corrected single-molecule time traces, rate coefficients of interconversion can be directly extracted either by dwell time analysis [37] for slow processes or by autocorrelation analysis for faster changes [38]. A newer approach relies on the Hidden Markov Modeling (HMM), which is implemented in the open source software packages HaMMY [39], vbFRET [40], and QuB [41]. The advantage of this approach when compared to visual inspections is the non-biased way to extract the number of FRET states. Detailed considerations on the analysis of smFRET trajectories in general and the use of these software packages can be found elsewhere [42]. A 3D-structure modeling based on experimental FRET data requires great care and has been discussed in detail by Brunger *et al.* [43].

24.4.3

The Choice of Dye Pairs for FRET

When choosing dye pairs for FRET, several considerations have to be taken into account. While it is desirable to use dyes with similar extinction coefficients and quantum yields, and of sufficient spectral separation, the cardinal parameter is the Förster radius of a specific FRET pair, because it governs the range of structural dynamics that can be investigated (see Figure 24.1). The Förster radii of selected FRET pairs are listed in Table 24.1. It is not unusual that reported Förster radii for the same dye pair vary by up to ~ 1 nm, depending on the instrumental setup and the experimental conditions.

Since the beginning of this millennium, the Cy3-Cy5 pair has been the dominating FRET pair in life sciences, due to its reasonable Förster radius and because both dyes are similar in terms of quantum yield and extinction coefficients. An even more important point is the performance of the acceptor dye, Cy5, together with antifading agents (see Section 24.4.7), which is the main reason for its use in our group.

If complex intramolecular or intra- as well as intermolecular interactions of different components are to be monitored by smFRET, one has to take a three- or even four-color FRET approach into consideration. Although the complexity of this approach is beyond the scope of this chapter, the dye combinations used so far are summarized in Table 24.2.

Table 24.2 Dye combinations used for three- or four-color FRET on nucleic acids.

Donor (Ex _{max} , Em _{max} (nm))	Acceptor 1 (Ex _{max} , Em _{max} (nm))	Acceptor 2 (Ex _{max} , Em _{max} (nm))	Acceptor 3 (Ex _{max} , Em _{max} (nm))	References
Cy3 (554, 573)	Cy5 (652, 672)	Cy5.5 (680, 710)	—	[14a, 23]
Alexa 488 (494, 519)	TMR (652, 672)	Alexa 647 (650, 668)	—	[14c]
Alexa 488 (494, 519)	Cy3B	Alexa 647 (650, 668)	—	[14b]
Cy3 (554, 573)	Cy5 (652, 672)	Cy7 (750, 780)	—	[14d]
Cy2 (489, 506)	Cy3 (554, 573)	Cy5 (652, 672)	Cy7 (750, 780)	[14e]
Atto488	Atto565	Atto647N	Alexa 750	[56]

24.4.4

Buffer Handling in Single Molecule Experiments

Ultrapure water should be used for all purposes. The pH adjustment of concentrated buffer stock solutions bears the risk of local heating of the solution as a result of the strongly exothermic nature of an acid-base reaction. Such local heating may result in the formation of aromatic fluorescent compounds, which will increase background fluorescence. Therefore, we recommend to prepare stock solutions of a defined pH by slowly mixing a buffering base with its conjugated acid of identical concentration, or vice versa. For example, stock solutions buffered with Tris can be prepared by mixing adequate volumes of a 1 M solution of Tris base with a 1 M solution of Tris-HCl (conjugated acid). If the resulting buffer shows fluorescence, the causative agents may be removed by the use of charcoal. To reduce light scattering and to avoid precipitates, all buffer solutions should be filtered through 0.22 μm syringe filters (Roth). It is equally recommendable to prepare fresh buffer solutions after a few days.

24.4.5

Strategies for Dye Labeling of RNA Constructs

This section will concentrate on the labeling of tRNAs, but the principal concepts can be applied to other RNAs as well. Connected to the growing interest in translational dynamics, a revival of research topics from the 1970s describing the site-specific labeling of modified nucleotides present in native tRNAs has been encountered [57, 58]. For example, it is possible to label the s⁴U8 position of tRNA^{Met} specifically by the use of maleimides [59]. The acp³U47 [3-(3-amino-3-carboxypropyl)U47] position of *Escherichia coli* tRNA^{Phe} can be specifically labeled by NHS-esters ([60], see also Chapter 8). A strategy applicable to various native tRNAs is the labeling of dihydrouridine (DHU) residues. This can be achieved with a stoichiometry of about one dye per tRNA, but may lead to an equal distribution of labels between the available DHU residues [61].

Site-specific introduction of fluorophores can be performed during solid-phase synthesis with dye-containing phosphoramidites (mostly for end-labeling) or *via* postsynthetic coupling to reactive moieties (modifiers) that are introduced during solid-phase synthesis (such as thiols and amino groups, see Chapter 8). For RNA-FRET experiments, labeling during synthesis is an option that, however, may bear several disadvantages. As certain fluorophores are unstable while facing the harsh conditions of RNA synthesis, the luminescence resulting from fluorophore incorporation may be lower than the one resulting from postsynthetic labeling. Furthermore, only a limited amount of fluorophores is available as phosphoramidites [62]. The most convenient approach is end-labeling, often applied during solid-phase synthesis, although one should be aware that this might lead to end-fraying [35] and to helix end-stacking [63] of the dye.

The modifiers mostly used for labeling in our group are alkyne-carrying nucleotides (commercially available or custom-made) that are introduced into in-house synthesized oligonucleotides. These moieties can be labeled postsynthetically with a dye-azide using click chemistry, as explained in the next section.

24.4.6

Postsynthetic Labeling of Alkyne-Containing RNA Oligonucleotides

The “click reaction” between terminal alkynes and azides [64] has become one of the standard procedures for the postsynthetic modification of RNA and DNA (Figure 24.3). This is mainly due to the bio-orthogonal coupling and stereospecific product formation of the 1,3-dipolar cycloaddition reaction in aqueous medium. Corresponding phosphoramidites for chemical RNA synthesis may contain the alkyne modification either directly on the base or on the sugar moiety. Azido-functionalized fluorescent markers are commercially available.

Cu(I) is used as catalyst in this reaction. As this oxidation state of copper is sensitive to oxygen, we prefer to use a copper (II) salt (CuSO_4) in the presence of a reducing agent (sodium ascorbate). The Cu(I) can be further protected from oxidation by the water-soluble, stabilizing THPTA [Tris-(hydroxypropyltriazolylmethyl)amine]. THPTA is not commercially available, but can be synthesized as described by Liu *et al.* [65]. NaH_2PO_4 (pH 8) is a suitable buffer for the deprotonation of the terminal alkyne during the click reaction between the crude RNA synthesis product and the fluorescent dye.

Gel filtration, HPLC, and precipitation are required for the purification of the labeled oligonucleotide. Unreacted dye molecules can be removed by means of size exclusion chromatography. This step also ensures the removal of all salts before HPLC separation of the dye-labeled full-length RNA strand from truncated strands that formed during oligoribonucleotide synthesis. Ethanol precipitation results in a pellet that contains the labeled RNA strand as a pure product.

An alternative approach uses RNA methyltransferases to site-specifically introduce an alkyne moiety. Here an alkyne-carrying S-Adenosylmethionine analogue is used as the cosubstrate. The herewith introduced alkyne moiety can then be labeled by click chemistry [67]. If the substrate contains only a single loop-exposed

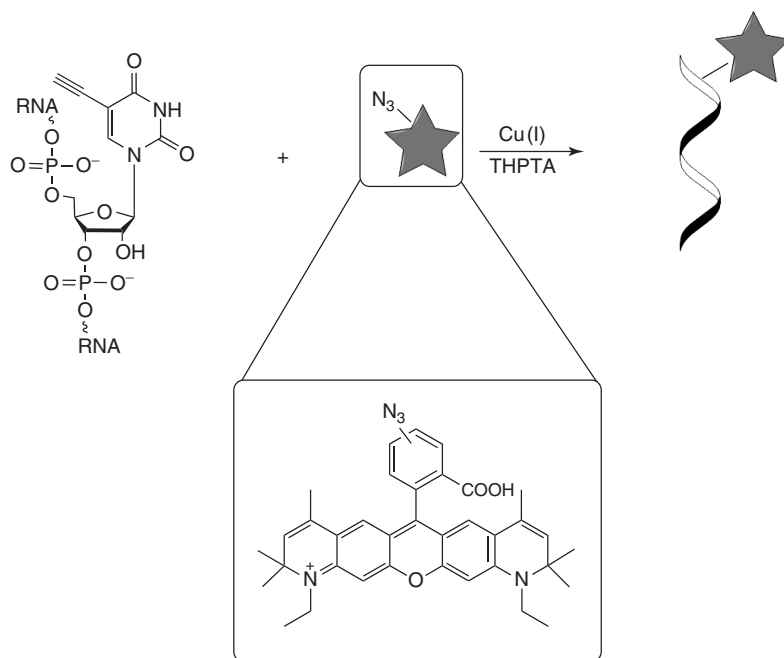


Figure 24.3 Reaction scheme of a click reaction between an RNA strand carrying a 5-ethynyluridine residue [66] and an ATTO 590 azide. Catalyzed by a THPTA-complexed Cu(I) generated *in situ*, this reaction yields a site-specifically labeled RNA strand.

uridine, or if dye attachment at multiple positions can be tolerated, the use of 7-Azido-4-(bromomethyl)coumarin as uridine-targeting azide may be an option for labeling *via* click chemistry [68].

Our laboratory also uses commercially synthesized oligonucleotides with site-specific dye labels. The internal modifiers used by commercial suppliers are usually ribothymidines or deoxythymidines carrying a C6-linker with a terminal amino group at the 5-position, or deoxyguanosines with the same amino linker at the 8-position. When positioned in helical regions, this type of label attachment has minimal effects on RNA stability, as the incorporation of deoxynucleotides into A-type RNA helices is well tolerated [62], and dyes are expected to project out of the helix into the major groove. In contrast, the labeling of cytosine (at the N4 position) has been reported to affect the melting properties of RNA duplexes [69].

24.4.7

Tuning Dye Endurance: Antifading Agents

The continuous observation of complex single-molecule dynamics for durations in the minute regime strongly relies on the use of antifading agents. These agents expand the time span until the fluorophore falls victim to photobleaching by up to several orders of magnitude.

Table 24.3 Antifading agents according to Cordes *et al.* [70].

Reductant, oxidant	Oxygen depletion system
Ascorbic acid, <i>N,N'</i> -methyl vilogen	Glucose-oxidase/catalase
Trolox & Trolox-quinone	Glucose-oxidase/catalase
<i>p</i> -phenylenediamine, <i>p</i> -chinoneimine (Supplier: Vectorlabs, Vectashield medium, glycerol-based) ^a	
Reductant: 1,4-diazabicyclo[2.2.2]octane = DABCO (Supplier: Ibidi, Ibidi mounting medium, glycerol-based) ^a	

^aThe Vectashield and Ibidi mounting media contain 70-90% glycerol. Compared with water, this increased viscosity leads to an at least threefold reduced oxygen concentration and thus a passive oxygen depletion effect.

The general principle of antifading reagents revolves around the two roles of oxygen with opposing effects: on the one hand, oxygen causes photobleaching by reacting with the fluorophore in its triplet state, and on the other hand, it functions as a triplet state quencher that reduces “dark states” and thus blinking of fluorophores. Consequently, antifading agents typically facilitate the depletion of oxygen in the solution, which is normally achieved by a reducing agent that consumes oxygen while being oxidized. This is frequently accompanied by addition of a substitute substance with triplet state quenching ability. The choice of the optimal fading agent depends on the nature of the dyes and the pH of the imaging buffer used. A comparison of the performance of antifading agents can be found elsewhere [70] (see Table 24.3 for the various possibilities). Owing to the predominant use of the FRET pair Cy3-Cy5 in our laboratory, an antifading system described in [71] has been adopted. This system relies on a combination of glucose, glucose oxidase, and catalase for oxygen depletion, and on Trolox for triplet state quenching (for details on Trolox, see [72]). The acidification caused by this system has also been analyzed recently [73]. A typical average of a fluorescence trace obtained by our setup is in the single-digit minute range. By combining this antifading system with the low-excitation intensities of prism-TIRF, several groups were able to extend the fluorescent lifetime even into the two-digit minute range [74].

Practical Protocols

The protocols presented here, mostly adopted from procedures described previously, were refined for our needs in order to achieve maximum performance.

Click Reaction between an Alkyne-containing RNA Oligonucleotide and an Azide-functionalized Fluorophore

Materials

- An appropriate dye azide dissolved in DMSO

- Your RNA (carrying a residue with an alkyne end group (that must not be confused with 3'- or 5'-end groups!))
- Ultrapure RNase-free H₂O, sodium phosphate NaH₂PO₄/Na₂HPO₄ buffer (pH 8), THPTA, sodium ascorbate, CuSO₄·5H₂O, acetonitrile (CH₃CN), Et₃NHOAc, absolute ethanol (−80 °C), NaOAc
- NAP 10 columns (GE Healthcare)
- Vortexer, thermomixer, centrifuge, HPLC and lyophilizer, Eppendorf SpeedVac, or similar
- Commercial or in-house synthesized RNA, carrying a residue with an alkyne group.

Click Reaction

- 1) Directly after cleavage from the support and deprotection, dissolve the crude chemically synthesized RNA in water. Adjust to a final concentration of ~100 μM RNA (typical for commercial RNAs).
- 2) Add 7.5 μl of the azido dye (2 mM in DMSO) to 100 μl of the 100 μM RNA.
- 3) Add 622.5 μl H₂O, 200 μl of 500 mM sodium phosphate (pH 8), 50 μl of 50 mM THPTA, 10 μl of 50 mM sodium ascorbate and, finally, 10 μl of 5 mM CuSO₄·5H₂O. The final concentration of RNA in the mixture is 10 μM, compared to 15 μM of the dye.
- 4) Vortex the mixture thoroughly and incubate at 25 °C overnight, with shaking. Be sure to protect the sample from light during this incubation time.

Purification

- 1) Load the total volume of the reaction mixture on a NAP 10 column pretreated according to the manufacturer's instructions. Elute with 1.5 ml of ultrapure RNase-free H₂O.
- 2) Concentrate the sample by means of freeze-drying and purify the crude mixture on a reversed-phase HPLC column. Use a flow rate of 3 ml/min, with CH₃CN as solvent A and 0.1 M Et₃NHOAc (pH 7) as solvent B. Use the following gradient of CH₃CN (Solvent A): linear gradient from 5 to 95% for 20 min; isocratic for 20 min; 5-min linear gradient to initial conditions; isocratic for 5 min. The labeled oligoribonucleotide elutes after the unreacted shorter strands and can be identified by its fluorescence.
- 3) Lyophilize the sample until dryness. Redissolve in 150 μl RNase-free H₂O.
- 4) Add 15 μl NaOAc (3 M; pH 5.3).
- 5) Add 500 μl ice-cold absolute EtOH (−80 °C) and gently invert the reaction tube.
- 6) Precipitate the RNA by storage at −20 °C for at least 1 h.
- 7) Centrifuge the sample (>13 000 g for 1 h at room temperature (RT)).
- 8) Remove the supernatant.
- 9) Gently rinse the pellet with 75% EtOH (50 μl, precooled at −20 °C).
- 10) Centrifuge the sample (>13 000 g for 30 min at RT)).
- 11) Dry the pellet *in vacuo*.

- 12) Redissolve the dry pellet in RNase-free H₂O to obtain the required concentration of labeled RNA.

Initial Spectroscopic Characterization by Bulk FRET Measurements

The following brief description of the acquisition and analysis of ensemble spectroscopic data allows a first assessment of the bulk FRET efficiency (E_{FRET}) of a construct [10]. If an E_{FRET} determined by such measurements significantly deviates from the expected value, this may either result from incomplete labeling, a large spatial separation of the dyes in the construct, or undesired interactions between dye and RNA. If such interaction impairs the rotation of the dye on the timescale of the excited state's lifetime, the resulting change in bulk fluorescence anisotropy may be revealed using polarization filters [32]. In any of the above cases, one should consider to redesign the RNA construct before performing single molecule experiments.

NOTE: Use only reagents of highest purity, such as ultrapure RNase-free water, and filter all buffers through a 0.22 μm syringe filter to avoid light scattering which results in high-intensity peaks at the excitation wavelength, or in peaks at multiples of the excitation wavelength.

Materials

- Ultrapure RNase-free H₂O
- Absolute ethanol
- ~2–10 pmol RNA sample (depending on the spectrometer's sensitivity)
- Appropriate folding buffer for the RNA under investigation
- Blackened fluorescence quartz cuvette (Hellma, Suprasil series), suitable for the fluorescence spectrometer used
- Thermoshaker (*e.g.*, Eppendorf)
- Emission-calibrated fluorescence spectrometer (*e.g.*, JASCO FP6500).

Bulk Fluorescence Measurement

- 1) Clean a quartz cuvette (30–100 μl) and rinse thoroughly with ultrapure RNase-free H₂O followed by rinsing with absolute ethanol, then blow dry with air [75].
- 2) Freshly prepare a 20–100 nM RNA solution in the appropriate folding buffer. Do not store dilute solutions for extended periods of time.
The folding buffer has to be selected based on biochemical knowledge of the target RNA.
- 3) Refold the RNA by heating to 60 °C in a thermoshaker for 5 min without agitation, and then slowly cool to RT over 10 min.
- 4) Adjust the spectrometer, such that a good signal-to-noise ratio is achieved, depending on the concentration of the sample (see example in Table 24.4).
- 5) Avoiding any air bubbles, fill the cuvette with measuring buffer (at least the nominal cuvette volume) and equilibrate for a few minutes in the spectrometer.

Table 24.4 Example spectrometer settings.^a

Emission spectrum	Data array type	Linear data array
Excitation wavelength: 540.0 nm	Data points	199
Measurement range: 550–748 nm	Data pitch	1 nm
Bandwidth (Ex.): 5 nm	Response	0.5 s
Bandwidth (Em.): 5 nm	Scanning speed	1000 nm/min
Sensitivity: medium	—	—

^aJasco FP6500.

- Record several blank spectra by scanning the emission intensity in the desired wavelength range.
- Equilibrate a freshly prepared sample (see step 3) in the spectrometer and record three to five spectra.
- Between measurements of different samples, thoroughly rinse the cuvette with ultrapure H₂O and record a spectrum to check for residual fluorescence.

Data Analysis

- Average at least 10 measurements for the blank and 3–5 measurements for each sample.
- Subtract the averaged blank from the averaged sample spectrum to reduce statistical noise.
- Correct the spectrum for the wavelength-dependent detection efficiency using a reference curve recorded according to the instructions of the manufacturer.
- Inspect the spectrum for donor and acceptor emission. *If the FRET efficiency of the construct is in the desired range, fluorescence of both fluorophores should be evident in the spectrum* (Figure 24.4).
- Calculate FRET efficiency values (E_{FRET}) from the following equation, where I_A and I_D are the fluorescence intensities emitted by the acceptor and donor at the wavelengths of their respective emission maxima (λ_{max}):

$$E_{\text{FRET}} = \frac{I_A}{I_A + \gamma I_D} \quad (24.3)$$

For initial characterization and comparison of different constructs, the correction factor γ can be ignored (i.e., set to 1). Accurate determination of FRET efficiencies requires correction for differential labeling efficiencies, detection efficiencies, quantum yields, and cross-talk of the fluorophores [35].

Antifading Agent Recipe [71]

Materials

- Ultrapure RNase-free H₂O
- Tris-HCl, Tris Base

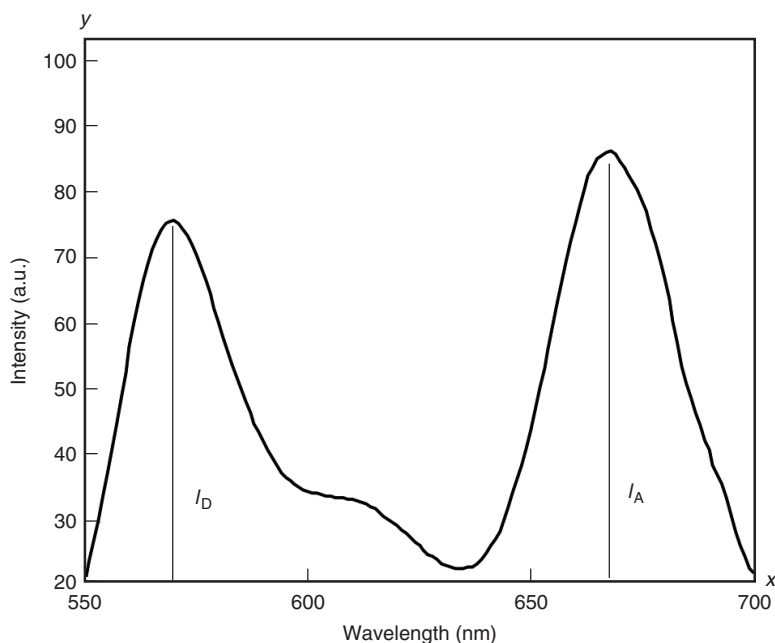


Figure 24.4 Bulk fluorescence spectrum of a tRNA labeled with the FRET pair Cy3-Cy5. The $\lambda_{\max}$ of donor intensity I_D and acceptor intensity I_A are marked by vertical lines; a.u., arbitrary units.

- NaCl
- Catalase from *bovine* liver (Merck CalBiochem, supplier especially important!)
- Glucose oxidase from *Aspergillus niger* (Sigma-Aldrich)
- Trolox or ($\pm$)-6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid (Sigma-Aldrich)
- β -D-Glucose (Sigma-Aldrich)
- 0.2 μ m syringe filters, 13 mm inner diameter, nylon membrane, PE shell (Roth).

Preparation of "Gloxy"-Oxygen-Scavenging Enzyme Stock

- 1) Weigh at least 5 mg glucose oxidase, dissolve in 30 μ l of T50 buffer (10 mM Tris-HCl, pH 8.0, 50 mM NaCl). The final glucose oxidase concentration of the 100 x stock will be 0.1 mg/ μ l. If you have weighed exactly 5 mg, the final volume of your 100 x stock will be 50 μ l.
- 2) Take another tube. Dissolve 1 mg of catalase in a final volume of 1 ml T50 buffer, and mix by cautiously tapping at the tube with your fingers (do NOT vortex!).
- 3) Supplement the 30 μ l from step 1 with 0.2 final volumes of catalase solution (10 μ l in the case of a final volume of 50 μ l).
- 4) Add T50 buffer to the final volume (50 μ l in the example above). Mix carefully by tapping (do NOT vortex!).

- 5) Centrifuge for 10 min at full speed in a small tabletop centrifuge to remove undissolved particles.
- 6) Carefully transfer the supernatant to a different tube.
- 7) This solution serves as a 100 x stock that is stable for about one week in a refrigerator ($\sim 4^{\circ}\text{C}$, do NOT freeze).

The final concentrations of the 100 x stock solution are 0.1 mg/ μl for glucose oxidase and 0.2 mg/ml for catalase. Final enzyme concentrations during FRET measurements are 1 $\mu\text{g}/\mu\text{l}$ glucose oxidase and 2.0 $\mu\text{g}/\text{ml}$ catalase.

Preparation of Imaging Buffer (Yield: $\sim 500\ \mu\text{l}$)

- 1) Put 1 spatula tip Trolox into a tube, add 200 μl of your desired 5 x imaging buffer, 250 μl 40% glucose solution and 550 μl ultrapure H_2O .
- 2) Vortex vigorously for at least 2 min.
- 3) Filter through two stacked small syringe filters. *If the solution is not homogeneous: Vortex and filter again.*
- 4) Protect imaging buffer from light; prepare newly every day (owing to oxygen and light sensitivity).

Directly before Start of Imaging

- 5) Add 1 μl of “gloxy” solution to imaging buffer, and mix gently by pipetting.
- 6) Exchange sample buffer against imaging buffer by pipetting a volume of imaging buffer several times larger than the measurement chamber volume into the cell. Enzyme-free buffer is rinsed out of the cell in this process.

For best performance, the enzyme system should always be prepared freshly. In our laboratory, one batch of this system can be used for up to 5 h without decreases in performance. Still, special considerations for measurement cell design have to be taken into account, as will be addressed in the next section.

Measurement Chamber Designs and Their Preparation [10, 29]

Materials

- Hellmanex cuvette cleaning solution (Hellma Analytics)
- Ethanol, ultrapure RNase-free H_2O , argon gas
- Parafilm (Pechiney Plastics Packaging, supplied by Roth)
- Heating block with removable insets (VWR)
- Scalpel (Braun, supplied by Roth)
- Fast drying Epoxy resin (Uhu)
- Electric drill and 0.8 mm diamond drill bit
- Glass-predetermined-breaking point producer (metal pen with diamond tip)
- Cover slides and 24 $\times$ 40-mm cover slips (Roth)
- Sponge, water-filled box (e.g., an empty pipette tip box).

Cleaning of Cover Slips and Slides

For setups using prism-TIRF or confocal detection of freely diffusing molecules, cover slips and slides can be cleaned by briefly burning off contaminants in a soot-free flame. However, if a plane surface is required, which is the case for a confocal microscope, one should avoid a hot flame, because the heat may lead to bending of the cover slips. Alternatives are the use either of an oven slowly heated to a temperature just below the softening temperature of the slips, or the Hellmanex cuvette cleaning solution that works nicely in our experience. Especially, if the surface is to be treated with PEG (Polyethylene glycol), a more complicated cleaning procedure may be used [76].

Fast Sample Cell Preparation

This type of sample cell (Figure 24.5a,b) is adjustable to up to five channels, which may contain different samples.

- 1) Incubate one cover slip and one cover slide for 40 min in 1% Hellmanex solution at 37 °C.
- 2) Heat the heating block to 85 °C.
- 3) Rinse cover slip and cover slide with ultrapure H₂O, blow dry in an argon stream.
- 4) Assemble sample cell immediately.
- 5) For assembly: Align two pieces of parafilm onto the coverslip perpendicular to the long axis, leaving a 2 mm wide channel in the middle. By adding more pieces of parafilm, up to five channels can be created.

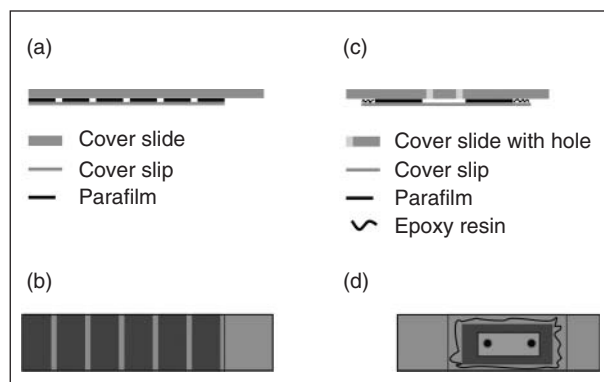


Figure 24.5 (a) Side view of the "fast" sample cell. It consists of a cover slide – parafilm – cover slip sandwich, in which several stripes of parafilm separate the individual channels. (b) Top view of an assembled fast sample cell. (c) Cross section through a "robust" sample cell. The holes inside the cover

slide make the chamber accessible that is again generated by a cover slide – parafilm – cover slip sandwich. Epoxy glue increases the stability. (d) Top view of a robust sample cell. Note that, for illustration reasons, the cover slide is shown in gray and the cover slip as transparent in (b) and (d).

- 6) Put the cover slip on top while taking care to maintain the parafilm arrangement.
- 7) Remove the inset out of the heating block and put the sample cell on the heating plate. Immediately, weight down the cell uniformly with the heated inset. Remove the cell from the heating block after 35 s.
- 8) The parafilm is now half melted and transparent, gluing slip and slide together. Let the cell cool to RT while weighted down appropriately. During cooling the parafilm turns opaque again.

The volume of the resulting cell is $\sim 5 \mu\text{l}$, so washing with $50 \mu\text{l}$ exchanges the solvent in the channel in a near-quantitative manner. The cell size is not of ultimate importance, as it will always exceed the recording area or volume during confocal microscopy measurements.

Robust Sample Cell Preparation

As the fast sample cell design turned out to be unsuitable for sample transport between different buildings, we developed the setup of a drilled cover slide and an otherwise totally sealed sample chamber (Figure 24.5c, d). Be aware that glass is an extremely hard surface with insulating properties, which makes drilling it a demanding task.

- 1) Cut a cover slide to a size slightly longer than the cover slip using the diamond pen.
- 2) Equip a drill with a 1 mm diameter diamond drill bit.
- 3) Mark the desired position of the holes in the cover slide using a waterproof pen and put scotch tape on top (the tape prevents slipping of the drill as glass cannot be punch-marked easily).
Preferably the holes are placed approximately in the middle of the width of the cover slide, spaced $\sim 1 \text{ cm}$ apart from each other.
- 4) Put the slide on a sponge under water (without cooling by water the glass will shatter).
- 5) Drill slowly.
- 6) Remove the scotch tape from the cover slide, clean by rinsing with ethanol and mechanically with a sponge.
- 7) Incubate slides and slips for 40 min in 1% Hellmanex solution at 37°C .
- 8) Heat the heating block to 85°C .
- 9) Rinse cover slip and cover slide with ultrapure H_2O , and blow dry in argon stream.
- 10) Assemble sample cell immediately: use a clean scalpel to cut a window ($\sim 1.6 \times 0.6 \text{ cm}$) into a piece of parafilm; the window should be significantly smaller than your cover slide, but large enough such that the parafilm isolates the holes drilled into the slide (see Figure 24.5c). *Be careful not to cut into the corners of the parafilm, because this will lead to leakage of the sample cell!*
- 11) Place the parafilm onto the slide such that the window contains the drilled holes.
- 12) Surround the parafilm sparsely with epoxy glue (this will enhance the long-term stability of the cell)

- 13) Place the cover slip on top, press gently.
- 14) Remove the inset from the heating block and put the sample cell into the tray of the heating block. Immediately weight down the cell uniformly with the heated inset. Remove the cell from the heating block after 35 s.
- 15) The parafilm is now half melted and transparent, gluing slip and slide together. Let the cell cool to RT while weighted down. Remove the weight after the glue has polymerized (duration depends on the glue). Upon cooling the parafilm turns opaque again. *Once assembled, the measurement cells can be stored several months if kept in a sealed container to keep off any dust.*

A sample chamber assembled in this manner should possess a volume of ~20–30 μl . Both presented sample chamber designs minimize exposure of the sample to air, which reduces oxygen consumption and therewith glucuronic acid generation. The robust sample cell design still allows one to adjust the buffer in the sample chamber in the course of an experiment. For both types of cells, their performance concerning quality of sample preparation and measured signal exceeded commercial alternatives significantly, in our experience.

Surface Passivation

To avoid artifacts caused by unspecific sample-surface interactions, surface passivation of the sample chamber is very important [10, 29]. Surface adsorption of RNA can be prevented by blocking hydrophobic surfaces with BSA (bovine serum albumine, 1 mg/ml in PBS (phosphate buffered saline), incubation for several minutes, followed by rinsing with ultrapure H_2O). In fact, biotinylated BSA is used anyway to immobilize the RNA molecules (see “Sample Preparation”). If unspecific surface interaction of proteins is to be prevented, the use of PEG-coated surfaces is recommended [76].

Sample Preparation

As described below, RNA molecules in surface-based smFRET experiments are usually immobilized using the biotin-streptavidin-biotin interaction. The surface is covered with biotinylated BSA (alternatively: biotinylated PEG), and streptavidin is added as a bridging molecule to finally couple the biotinylated substrate RNA.

Materials

- Heating Block (e.g., Thermoshaker from Eppendorf)
- Ultrapure H_2O
- 1 mg/ml BSA-biotin (Sigma-Aldrich) in 100 mM sodium phosphate buffer, pH 7.4 (Fluka)
- 200 $\mu\text{g}/\text{ml}$ streptavidin (New England Biolabs) in 100 mM sodium phosphate buffer, pH 7.4 (Fluka)
- 100 mM sodium phosphate buffer (PBS), pH 7.4 (Fluka)
- Imaging buffer without enzymes and, for the actual measurement, with enzymes
- 5 x folding buffer
- 2 fmol of biotinylated, fluorophor-labeled RNA

Sample Preparation Procedure

- 1) Prepare RNA solution (2 fmol in 40 μ l ultrapure H₂O).
- 2) Denature RNA by heating to 75 °C in a light-protected heating block for 4 min.
- 3) In the meantime fill 40 μ l of biotinylated BSA into your sample cell and incubate for 5 min.
- 4) Supplement the heated RNA with 10 μ l of 5 x concentrated folding buffer and mix by tapping or pipetting. Spin down briefly, and allow for light-protected cooling to RT for 15 min.
- 5) Wash sample cell with 100 μ l PBS.
- 6) Fill cell with $\sim$ 40 μ l streptavidin solution, incubate for 60 s at ambient temperature.
- 7) Wash with 100 μ l PBS.
- 8) Wash with 100 μ l enzyme-free imaging buffer.
- 9) Add RNA solution to sample cell and incubate for 5 min at ambient temperature.
- 10) Wash with 100 μ l enzyme-free imaging buffer.
- 11) Exchange against 100 μ l enzyme-containing imaging buffer just before the measurement.

24.5**Troubleshooting****24.5.1****RNase Contamination**

RNase contamination, which may, for example, result from protein impurities, can easily spoil an smFRET experiment. In experiments with surface-immobilized RNA, this can be detected by solely exciting the donor, because RNase-mediated release of the fluorophore will cause an explosion of background signal in the donor channel, which is accompanied by an almost total loss of the FRET signal. In measurements of freely diffusing fluorophore-labeled RNA molecules, RNase contamination would enhance the diffusion speed by several orders of magnitude and lead to a total loss of the FRET signal as well. A control for RNase contamination is easily done by incubating 1–10 pmol aliquots of the fluorescently labeled RNA with the different reagents in use for 40 min at 37 °C. The resulting mixture is then analyzed by denaturing PAGE. Scanning the gel with a gel scanner, such as GE Healthcare's Typhoon or Biorad's FX imager, will detect any degradation and identify the RNase-contaminated component.

24.5.2**Removal of Unbound Fluorophores**

After labeling, the RNA sample is typically purified by either HPLC or denaturing PAGE. Both methods should easily remove unbound fluorophores. If unbound fluorophores are not removed properly, one encounters fast diffusing species in

solution experiments and high-background fluorescence in surface-based measurements. In such cases, the use of NAP columns (see “Purification”) is recommended.

24.5.3

Drying of Samples

Normally the reduced air exposure should prevent drying. If that is not the case, one can seal the sample cell reversibly by parafilm or irreversibly by epoxy glue.

24.5.4

Donor-Only Populations

A prevalence of donor-only populations (*i.e.*, RNAs lacking any acceptor fluorescence) may have different reasons. One possibility is that excessive excitation intensity has been applied, which destroys the acceptor already during imaging. This may also be the case if the antifading system does not work properly. A third reason may be improper handling of the sample RNA during synthesis or isolation: UV light or the scanning by gel scanners (*e.g.*, GE Healthcare Typhoon or Biorad FX imager) almost instantaneously destroys significant amounts of fluorophore, making the samples essentially unsuitable for single molecule experiments. For detection of the labeled RNA during preparative PAGE, we recommend to excise the gel bands without use of the above-mentioned devices. As even the product band of a 400 pmol-scale ligation is easily visible to the naked eye (purified as a 50 μ l reaction + 25 μ l formamide loading buffer in a 160 μ l gel pocket), RNA band identification can be achieved by visual inspection if empty lanes are loaded with xylene cyanol and bromophenol blue to provide orientation concerning band sizes. Also, a small aliquot of the labeled RNA may be loaded as a reference into a well adjacent to the preparative well. After excision of the preparative band, the gel can be scanned by an imager, using the reference lane to make sure that the correct preparative band was excised. In any case, identity and integrity of the gel-purified product RNA should be finally evaluated by analytical PAGE.

24.5.5

Too Dense or Too Sparse Surface Coverage

Attachment of three fluorescent RNAs to the same streptavidin molecule represents the worst case of too dense surface coverage (note that only one of the four binding sites of streptavidin is needed for surface coupling). Here, incubation times and concentrations of streptavidin should be increased. Alternatively, if the streptavidin distribution is already dense, the sample concentration should be decreased. Several molecules at the same spot are identifiable by multiple distinct bleaching steps, but still they are unsuitable for experiments because their signals superimpose. In case of too sparse surface coverage, the addressable parameters are the same, but they have to be adjusted in the opposite sense.

References

1. Breaker, R.R. (2011) Prospects for riboswitch discovery and analysis. *Mol. Cell*, **43**, 867–879.
2. Frank, J. and Gonzalez, R.L. (2010) Structure and dynamics of a processive Brownian motor: the translating ribosome. *Annu. Rev. Biochem.*, **79**, 381–412.
3. Dominguez, C., Schubert, M., Duss, O., Ravindranathan, S., and Allain, F.H.T. (2011) Structure determination and dynamics of protein-RNA complexes by NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.*, **58**, 1–61.
4. Nguyen, P. and Qin, P.Z. (2011) *Wiley Interdisciplinary Reviews: RNA*, John Wiley & Sons, Ltd., **4**, 62–72.
5. (a) Grohmann, D., Klose, D., Klare, J.P., Kay, C.W.M., Steinhoff, H. Jr., and Werner, F. (2010) RNA-binding to archaeal RNA polymerase subunits F/E: a DEER and FRET study. *J. Am. Chem. Soc.*, **132**, 5954–5955; (b) DeRose, V.J., Kim, N.-K., Bowman, M.K., Green, B., Unger, A., Stoll, S., and Britt, R.D. (2010) Accurate Distance Constraints for RNA Structures Using Deer Spectroscopy *Biophys. J.*, **98**, 264a.
6. Clegg, R.M. (1995) Fluorescence resonance energy transfer. *Curr. Opin. Biotechnol.*, **6**, 103–110.
7. (a) Lee, M.K., Gal, M., Frydman, L., and Varani, G. (2010) Real-time multidimensional NMR follows RNA folding with second resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 9192–9197; (b) Tolbert, B.S., Miyazaki, Y., Barton, S., Kinde, B., Starck, P., Singh, R., Bax, A., Case, D.A., and Summers, M.F. (2010) Major groove width variations in RNA structures determined by NMR and impact of ^{13}C residual chemical shift anisotropy and ^1H - ^{13}C residual dipolar coupling on refinement. *J. Biomol. NMR*, **47**, 205–219; (c) Kim, N.K., Bowman, M.K., and DeRose, V.J. (2010) Precise mapping of RNA tertiary structure via nanometer distance measurements with double electron-electron resonance spectroscopy. *J. Am. Chem. Soc.*, **132**, 8882–8884.
8. Ha, T., Zhuang, X., Kim, H.D., Orr, J.W., Williamson, J.R., and Chu, S. (1999) Ligand-induced conformational changes observed in single RNA molecules. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 9077–9082.
9. Bokinsky, G., Rueda, D., Misra, V.K., Rhodes, M.M., Gordus, A., Babcock, H.P., Walter, N.G., and Zhuang, X. (2003) Single-molecule transition-state analysis of RNA folding. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 9302–9307.
10. Hengesbach, M., Kobitski, A., Voigts-Hoffmann, F., Frauer, C., Nienhaus, G.U., and Helm, M. (2008) RNA intramolecular dynamics by single-molecule FRET. *Current Protocols in Nucleic Acid Chemistry*, Chapter 11:Unit 11.12. John Wiley & Sons, Inc.
11. Kapanidis, A.N., Laurence, T.A., Lee, N.K., Margeat, E., Kong, X., and Weiss, S. (2005) Alternating-laser excitation of single molecules. *Acc. Chem. Res.*, **38**, 523–533.
12. Helm, M., Kobitski, A., and Nienhaus, G. (2009) Single-molecule Förster resonance energy transfer studies of RNA structure, dynamics and function. *Biophys. Rev.*, **1**, 161–176.
13. Zhao, R. and Rueda, D. (2009) RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods*, **49**, 112–117.
14. (a) Hohng, S., Joo, C., and Ha, T. (2004) Single-molecule three-color FRET. *Biophys. J.*, **87**, 1328–1337; (b) Lee, N.K., Kapanidis, A.N., Koh, H.R., Korlann, Y., Ho, S.O., Kim, Y., Gassman, N., Kim, S.K., and Weiss, S. (2007) Three-color alternating-laser excitation of single molecules: monitoring multiple interactions and distances. *Biophys. J.*, **92**, 303–312; (c) Lee, N.K., Koh, H.R., Han, K.Y., and Kim, S.K. (2007) Folding of 8-17 deoxyribozyme studied by three-color alternating-laser excitation of single molecules. *J. Am. Chem. Soc.*, **129**, 15526–15534; (d) Lee, S., Lee, J., and Hohng, S. (2010) Single-molecule three-color FRET with both negligible spectral overlap and long observation time. *PLoS ONE*, **5**, e12270; (e) Lee, J., Lee, S., Ragunathan, K., Joo, C., Ha, T., and Hohng, S. (2010) Single-molecule

- four-color FRET. *Angew. Chem. Int. Ed.*, **49**, 9922–9925.
15. Solomatin, S.V., Greenfeld, M., Chu, S., and Herschlag, D. (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature*, **463**, 681–684.
 16. Heppell, B., Blouin, S., Dussault, A.-M., Mulhbach, J., Ennifar, E., Penedo, J.C., and Lafontaine, D.A. (2011) Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nat. Chem. Biol.*, **7**, 384–392.
 17. Haller, A., Rieder, U., Aigner, M., Blanchard, S.C., and Micura, R. (2011) Conformational capture of the SAM-II riboswitch. *Nat. Chem. Biol.*, **7**, 393–400.
 18. Karunatilaka, K.S., Solem, A., Pyle, A.M., and Rueda, D. (2010) Single-molecule analysis of Mss116-mediated group II intron folding. *Nature*, **467**, 935–939.
 19. (a) Yeoman, J.A., Orte, A., Ashbridge, B., Klenerman, D., and Balasubramanian, S. (2010) RNA conformation in catalytically active human telomerase. *J. Am. Chem. Soc.*, **132**, 2852–2853; (b) Mihalusova, M., Wu, J.Y., and Zhuang, X. (2011) Functional importance of telomerase pseudoknot revealed by single-molecule analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 20339–20344.
 20. Petrov, A., Kornberg, G., O’Leary, S., Tsai, A., Uemura, S., and Puglisi, J.D. (2011) Dynamics of the translational machinery. *Curr. Opin. Struct. Biol.*, **21**, 137–145.
 21. Aitken, C.E., Petrov, A., and Puglisi, J.D. (2010) Single ribosome dynamics and the mechanism of translation. *Annu. Rev. Biophys.*, **39**, 491–513.
 22. Aitken, C.E. and Puglisi, J.D. (2010) Following the intersubunit conformation of the ribosome during translation in real time. *Nat. Struct. Mol. Biol.*, **17**, 793–800.
 23. Munro, J.B., Altman, R.B., Tung, C.-S., Cate, J.H.D., Sanbonmatsu, K.Y., and Blanchard, S.C. (2010) Spontaneous formation of the unlocked state of the ribosome is a multistep process. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 709–714.
 24. Brenner, M.D., Scanlan, M.S., Nahas, M.K., Ha, T., and Silverman, S.K. (2010) Multivector fluorescence analysis of the *xpt* guanine riboswitch aptamer domain and the conformational role of guanine. *Biochemistry*, **49**, 1596–1605.
 25. Chen, C., Stevens, B., Kaur, J., Smilansky, Z., Cooperman, B.S., and Goldman, Y.E. (2011) Allosteric vs. spontaneous exit-site (E-site) tRNA dissociation early in protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 16980–16985.
 26. Kobitski, A.Y., Nierth, A., Helm, M., Jäschke, A., and Nienhaus, G.U. (2007) Mg²⁺-dependent folding of a Diels-Alderase ribozyme probed by single-molecule FRET analysis. *Nucleic Acids Res.*, **35**, 2047–2059.
 27. Kraut, S., Bebenroth, D., Nierth, A., Kobitski, A.Y., Nienhaus, G.U., and Jäschke, A. (2011) Three critical hydrogen bonds determine the catalytic activity of the Diels-Alderase ribozyme. *Nucleic Acids Res.*, **108**, 20339–20344.
 28. Voigts-Hoffmann, F., Hengesbach, M., Kobitski, A.Y., van Aerschot, A., Herdewijn, P., Nienhaus, G.U., and Helm, M. (2007) A methyl group controls conformational equilibrium in human mitochondrial tRNA(Lys). *J. Am. Chem. Soc.*, **129**, 13382–13383.
 29. Kobitski, A.Y., Hengesbach, M., Helm, M., and Nienhaus, G.U. (2008) Sculpting an RNA conformational energy landscape by a methyl group modification - a single-molecule FRET study. *Angew. Chem. Int. Ed.*, **47**, 4326–4330.
 30. Dammertz, K., Hengesbach, M., Helm, M., Nienhaus, G.U., and Kobitski, A.Y. (2011) Single-molecule FRET studies of counterion effects on the free energy landscape of human mitochondrial lysine tRNA. *Biochemistry*, **50**, 3107–3115.
 31. Kobitski, A.Y., Hengesbach, M., Seidu-Larry, S., Dammertz, K., Chow, C.S., van Aerschot, A., Nienhaus, G.U., and Helm, M. (2011) Single-molecule FRET reveals a cooperative effect of two methyl group modifications in the folding of human mitochondrial tRNA(Lys). *Chem. Biol.*, **18**, 928–936.

32. Lakowicz, J.R. (2006) *Principles of Fluorescence Spectroscopy*, 3rd ed., Springer, New York.
33. Pan, D., Qin, H., and Cooperman, B.S. (2009) Synthesis and functional activity of tRNAs labeled with fluorescent hydrazides in the D-loop. *RNA*, **15**, 346–354.
34. Chen, C., Stevens, B., Kaur, J., Cabral, D., Liu, H., Wang, Y., Zhang, H., Rosenblum, G., Smilansky, Z., Goldman, Y.E., and Cooperman, B.S. (2011) Single-molecule fluorescence measurements of ribosomal translocation dynamics. *Mol. Cell*, **42**, 367–377.
35. Sabanayagam, C.R., Eid, J.S., and Meller, A. (2005) Using fluorescence resonance energy transfer to measure distances along individual DNA molecules: corrections due to nonideal transfer. *J. Chem. Phys.*, **122**, 061103.
36. Joo, C. and Ha, T. (2007) in *Single Molecule Techniques – A Laboratory Manual* (eds P. Selvin and T. Ha), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, chapter 2, p. 11–12.
37. Pond, S.J., Ridgeway, W.K., Robertson, R., Wang, J., and Millar, D.P. (2009) HIV-1 Rev protein assembles on viral RNA one molecule at a time. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 1404–1408.
38. Joo, C., McKinney, S.A., Lilley, D.M., and Ha, T. (2004) Exploring rare conformational species and ionic effects in DNA Holliday junctions using single-molecule spectroscopy. *J. Mol. Biol.*, **341**, 739–751.
39. McKinney, S.A., Joo, C., and Ha, T. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.*, **91**, 1941–1951.
40. Bronson, J.E., Fei, J., Hofman, J.M., Gonzalez, R.L. Jr., and Wiggins, C.H. (2009) Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.*, **97**, 3196–3205.
41. Qin, F. (2004) Restoration of single-channel currents using the segmental k-means method based on hidden Markov modeling. *Biophys. J.*, **86**, 1488–1501.
42. Blanco, M. and Walter, N.G. (2010) Analysis of complex single-molecule FRET time trajectories. In: *Single Molecule Tools, Part A: Fluorescence Based Approaches*, Methods in Enzymology, Elsevier Inc., Vol. **472**, p. 153–178.
43. Brunger, A.T., Strop, P., Vrljic, M., Chu, S., and Weninger, K.R. (2011) Three-dimensional molecular modeling with single molecule FRET. *J. Struct. Biol.*, **173**, 497–505.
44. Lorenz, M., Hillisch, A., and Diekmann, S. (2002) Fluorescence resonance energy transfer studies of U-shaped DNA molecules. *J. Biotechnol.*, **82**, 197–209.
45. Wozniak, A.K., Schröder, G.F., Grubmüller, H., Seidel, C.A., and Oesterhelt, F. (2008) Single-molecule FRET measures bends and kinks in DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 18337–18342.
46. Liu, J. and Lu, Y. (2002) FRET study of a trifluorophore-labeled DNAzyme. *J. Am. Chem. Soc.*, **124**, 15208–15216.
47. (a) Levitus, M. and Ranjit, S. (2011) Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments. *Q. Rev. Biophys.*, **44**, 123–151; (b) Stein, I.H., Schuller, V., Bohm, P., Tinnefeld, P., and Liedl, T. (2011) Single-molecule FRET ruler based on rigid DNA origami blocks. *ChemPhysChem*, **12**, 689–695.
48. Coban, O., Lamb, D.C., Zaychikov, E., Heumann, H., and Nienhaus, G.U. (2006) Conformational heterogeneity in RNA polymerase observed by single-pair FRET microscopy. *Biophys. J.*, **90**, 4605–4617.
49. Kuzmenkina, E.V., Heyes, C.D., and Nienhaus, G.U. (2005) Single-molecule Förster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15471–15476.
50. Di Fiori, N. and Meller, A. (2010) The effect of dye-dye interactions on the spatial resolution of single-molecule FRET measurements in nucleic acids. *Biophys. J.*, **98**, 2265–2272.
51. Grunwell, J.R., Glass, J.L., Lacoste, T.D., Deniz, A.A., Chemla, D.S., and Schultz, P.G. (2001) Monitoring the conformational fluctuations of DNA hairpins

- using single-pair fluorescence resonance energy transfer. *J. Am. Chem. Soc.*, **123**, 4295–4303.
52. Santoso, Y., Joyce, C.M., Potapova, O., Le Reste, L., Hohlbein, J., Torella, J.P., Grindley, N.D.F., and Kapanidis, A.N. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 715–720.
 53. (a) Clegg, R.M. (1992) Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol.*, **211**, 353–388; (b) Tuschl, T., Gohlke, C., Jovin, T.M., Westhof, E., and Eckstein, F. (1994) Three-dimensional model for the hammerhead ribozyme based on fluorescence measurements. *Science*, **266**, 785–789; (c) Chan, B., Weidemaier, K., Yip, W.T., Barbara, P.F., and Musier-Forsyth, K. (1999) Intra-tRNA distance measurements for nucleocapsid protein-independent tRNA unwinding during priming of HIV reverse transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 459–464; (d) Singh, R.N. and Dreher, T.W. (1998) Specific site selection in RNA resulting from a combination of nonspecific secondary structure and -CCR- boxes: initiation of minus strand synthesis by turnip yellow mosaic virus RNA-dependent RNA polymerase. *RNA*, **4**, 1083–1095; (e) Ha, T., Zhuang, X., Kim, H.D., Orr, J.W., Williamson, J.R., and Chu, S. (1999) Ligand-induced conformational changes observed in single RNA molecules. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 9077–9082.
 54. (a) Mujumdar, R.B., Ernst, L.A., Mujumdar, S.R., Lewis, C.J., and Waggoner, A.S. (1993) Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. *Bioconjugate Chem.*, **4**, 105–111; (b) Ha, T., Enderle, T., Ogletree, D.F., Chemla, D.S., Selvin, P.R., and Weiss, S. (1996) Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 6264–6268; (c) Lee, N.K., Kapanidis, A.N., Wang, Y., Michalet, X., Mukhopadhyay, J., Ebright, R.H., and Weiss, S. (2005) Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys. J.*, **88**, 2939–2953; (d) Ha, T., Ting, A.Y., Liang, J., Caldwell, W.B., Deniz, A.A., Chemla, D.S., Schultz, P.G., and Weiss, S. (1999) Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 893–898.
 55. Kupstat, A., Ritschel, T., and Kumke, M.U. (2011) Oxazine dye-conjugated DNA oligonucleotides: Förster resonance energy transfer in view of molecular dye-DNA interactions. *Bioconjugate Chem.*, **22**, 2546–2557.
 56. Stein, I.H., Steinhauer, C., and Tinnefeld, P. (2011) Single-molecule four-color FRET visualizes energy-transfer paths on DNA origami. *J. Am. Chem. Soc.*, **133**, 4193–4195.
 57. Sprinzl, M. and Faulhammer, H.G. (1978) Participation of X47-fluorescamine modified *E. coli* tRNAs in *in vitro* protein biosynthesis. *Nucleic Acids Res.*, **5**, 4837–4853.
 58. Behm-Ansmant, I., Helm, M., and Motorin, Y. (2011) Use of specific chemical reagents for detection of modified nucleotides in RNA. *J. Nucleic Acids*, **2011**, 408053.
 59. Watson, B.S., Hazlett, T.L., Eccleston, J.F., Davis, C., Jameson, D.M., and Johnson, A.E. (1995) Macromolecular arrangement in the aminoacyl-tRNA elongation factor Tu.GTP ternary complex. A fluorescence energy transfer study. *Biochemistry*, **34**, 7904–7912.
 60. Plumbridge, J.A., Bäumer, H.G., Ehrenberg, M., and Rigler, R. (1980) Characterisation of a new, fully active fluorescent derivative of *E. coli* tRNA Phe. *Nucleic Acids Res.*, **8**, 827–843.
 61. Kaur, J., Raj, M., and Cooperman, B.S. (2011) Fluorescent labeling of tRNA dihydrouridine residues: mechanism and distribution. *RNA*, **17**, 1393–1400.
 62. Qin, P.Z. and Pyle, A.M. (1999) Site-specific labeling of RNA with fluorophores and other structural probes. *Methods*, **18**, 60–70.

63. Dietrich, A., Buschmann, V., Muller, C., and Sauer, M. (2002) Fluorescence resonance energy transfer (FRET) and competing processes in donor-acceptor substituted DNA strands: a comparative study of ensemble and single-molecule data. *J. Biotechnol.*, **82**, 211–231.
64. (a) Kolb, H.C., Finn, M.G., and Sharpless, K.B. (2001) Click chemistry: diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed. Engl.*, **40**, 2004–2021; (b) Amblard, F., Cho, J.H., and Schinazi, R.F. (2009) Chem. Rev., Cu(I)-catalyzed Huisgen azide-alkyne 1,3-dipolar cycloaddition reaction in nucleoside, nucleotide, and oligonucleotide chemistry. *Chem. Rev.*, **109**, 4207–4220.
65. Liu, X.M., Thakur, A., and Wang, D. (2007) Efficient synthesis of linear multifunctional poly(ethylene glycol) by copper(I)-catalyzed Huisgen 1,3-dipolar cycloaddition. *Biomacromolecules*, **8**, 2653–2658.
66. Jao, C.Y. and Salic, A. (2008) Exploring RNA transcription and turnover *in vivo* by using click chemistry. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 15779–15784.
67. Motorin, Y., Burhenne, J., Teimer, R., Koynov, K., Willnow, S., Weinhold, E., and Helm, M. (2011) Expanding the chemical scope of RNA:methyltransferases to site-specific alkynylation of RNA for click labeling. *Nucleic Acids Res.*, **39**, 1943–1952.
68. Kellner, S., Seidu-Larry, S., Burhenne, J., Motorin, Y., and Helm, M. (2011) A multifunctional bioconjugate module for versatile photoaffinity labeling and click chemistry of RNA. *Nucleic Acids Res.*, **39**, 7348–7360.
69. Telser, J., Cruickshank, K.A., Morrison, L.E., and Netzel, T.L. (1989) Synthesis and characterization of DNA oligomers and duplexes containing covalently attached molecular labels: comparison of biotin, fluorescein, and pyrene labels by thermodynamic and optical spectroscopic measurements. *J. Am. Chem. Soc.*, **111**, 6966–6976.
70. Cordes, T., Maiser, A., Steinhauer, C., Schermelleh, L., and Tinnefeld, P. (2011) Mechanisms and advancement of antifading agents for fluorescence microscopy and single-molecule spectroscopy. *Phys. Chem. Chem. Phys.*, **13**, 6699–6709.
71. Rasnik, I., McKinney, S.A., and Ha, T. (2006) Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat. Methods*, **3**, 891–893.
72. Cordes, T., Vogelsang, J., and Tinnefeld, P. (2009) On the mechanism of Trolox as antiblinking and antibleaching reagent. *J. Am. Chem. Soc.*, **131**, 5018–5019.
73. Shi, X., Lim, J., and Ha, T. (2010) Acidification of the oxygen scavenging system in single-molecule fluorescence studies: in situ sensing with a ratiometric dual-emission probe. *Anal. Chem.*, **82**, 6132–6138.
74. (a) Rueda, D., Bokinsky, G., Rhodes, M.M., Rust, M.J., Zhuang, X., and Walter, N.G. (2004) Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 10066–10071; (b) McDowell, S.E., Jun, J.M., and Walter, N.G. (2010) Long-range tertiary interactions in single hammerhead ribozymes bias motional sampling toward catalytically active conformations. *RNA*, **16**, 2414–2426.
75. Hirsch, M., Kim, I.-H., Järve, A., Fischer, R., Trendelenburg, M.F., Massing, U., Rohr, K., and Helm, M. (2011) Tracing of siRNAs inside cells by FRET imaging. In: *RNA Interference Techniques* (ed. S.Q. Harper), Humana Press, Vol. 58, p. 141–157.
76. Lamichane, R., Solem, A., Black, W., and Rueda, D. (2010) Single-molecule FRET of protein-nucleic acid and protein-protein complexes: surface passivation and immobilization. *Methods*, **52**, 192–200.

25

Atomic Force Microscopy Imaging and Force Spectroscopy of RNA

Malte Bussiek, Antonie Schöne, and Wolfgang Nellen

25.1

Introduction

Atomic force microscopy (AFM) has proved to be an important and versatile method for investigations of biomolecular structures and interactions. As the basic principle, AFM measures very weak interaction forces between a sharp sensor (= stylus) tip and a sample surface. This makes it possible to image molecules adsorbed to the surface at high resolution without additional staining, and the method includes the option to image under physiological (or adjustable) conditions in solution. In a methodological extension, termed *force spectroscopy AFM*, the interaction forces between macromolecules, for example, RNAs and their binding proteins, can be measured. For this purpose, the nucleic acid may be attached to the sample surface, while the stylus is functionalized with the binding protein(s) or *vice versa*. Therefore, the diversity of RNA structures and functions throughout biology opens a wide field of applications for AFM, a technique potentially suitable to study the mechanisms of higher order folding of RNA. Different RNA structures have been examined by AFM imaging, among them hairpin ribozymes [1], viral RNA [2–4], and dsRNA [5]. Examples of imaged RNA–protein interactions include complexes of poly(A) binding protein and mRNA [6], a dsRNA antibody bound to dsRNA [7], iron response element (IRE) binding protein in complex with its IRE target site [8], and the RNA editing enzyme ADAR2 bound to natural editing sites [9].

Force spectroscopy AFM has been used to demonstrate that the overstretching behavior that was known for dsDNA also applies to dsRNA [10]. In a more recent work, AFM measurements could show different binding modes (specific and unspecific) in the interaction between a glycine-rich RNA binding protein from *Arabidopsis* and its RNA target sequence [11].

AFM may also be used for the detection and quantification of RNA expression. The motivation for research in this direction is to detect very small amounts of an RNA species. Gold surfaces modified with ssDNA-oligomers *via* thiol groups were used to quantify complementary microRNAs through the change in stiffness of the tethered oligomers after hybridization [12]. As another approach, cell

type-specific mRNA expression may be assayed directly on sectioned tissue by mapping interaction forces with a specific oligonucleotide bound to the tip [13].

Since stable AFM images of DNA and RNA were reported [14, 15], AFM has expanded considerably into research involving DNA, while applications to RNA have increased much slower. This is, at least in part, because interpretation of imaged RNA structures is much more difficult owing to their higher complexity. RNA structures may exist in different conformations with similar energies, and such conformational equilibria may vary with experimental conditions. Because of the structural diversity of RNA, successful imaging of a certain RNA is not necessarily a proof of principle for other RNA molecules. The question if and how AFM imaging can be applied to investigate RNA structures has to take two major limitations into account:

- The spatial resolution of AFM, although very high (10–20 nm), does not allow the observation of the molecular substructure of small folded RNA domains. However, it can detect the presence of a folded domain within a larger RNA molecule. This requires special selection of appropriate natural sample material or designed RNA constructs containing particular domains at defined sites so that the folding behavior or protein binding of the respective domain can be addressed.
- For AFM imaging, immobilization of RNA to a scanning surface is usually achieved *via* electrostatic interactions. However, interaction forces with the surface may interfere with intramolecular interactions, leading to altered RNA structures. In addition, RNA structures may be sensitive to the salt composition, particularly to the presence of divalent cations. For this reason, the selection of the surface chemistry and the conditions during surface adsorption are very important.

25.2

AFM Imaging of RNA Structures

25.2.1

General Preconditions: Mode of Operation, Data Analysis, and Resolution

The AFM setup (Figure 25.1) has three basic components [16]: (i) a sharp sensor attached to the end of a flexible cantilever and brought into contact with the sample surface, (ii) a piezo-electric scanner that moves the sample surface relative to the cantilever (or vice versa) in x , y , and z directions, and (iii) an electronic feedback loop that ensures a constant contact force between sensor and surface. This force, F , is proportional to the deflection, d , of the cantilever through Hooke's law, that is, $F = -kd$, where k is the spring constant of the cantilever. Movements of the scanner in x and y are performed to scan a surface in a line-by-line manner. The deflection is detected *via* a laser beam directed onto the back-side of the cantilever and reflected onto a position-sensitive photodetector. This detector is a photodiode split into sections that measure the light intensity separately (upper and lower

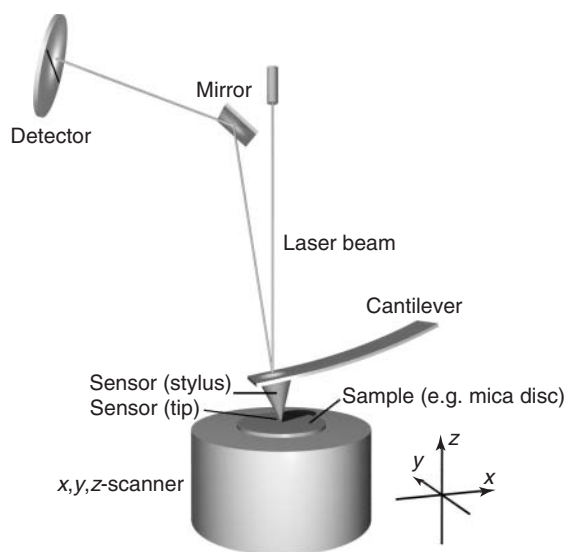


Figure 25.1 The principle of the atomic force microscope (see text for details). Note that the components are not shown to scale. The image was generated using Persistence of Vision™ Raytracer.

half sections of the detector shown in Figure 25.1, but typically four quadrants). A deflection of the laser beam from the center position due to bending of the cantilever causes a difference in voltage output measured at both halves. This difference, which is proportional to d (and F), is used as the input signal for the feedback loop that functions to keep a preset voltage difference constant by causing the scanner to retract or extend in z -direction, depending on the height changes along scan lines. As a result, the scanner follows the height profile along scan lines, which is used to build up the topographic image. However, biomolecules are mostly scanned in the amplitude modulation mode (also known as *tapping* or *intermittent contact mode*). Here, the cantilever oscillates vertically near its resonant frequency and the stylus is approached so that it barely touches the surface at the lower end of each oscillation cycle. Thereby, vertical and lateral forces acting on the molecules are reduced as compared with the continuous contact mode described above. The feedback loop uses the oscillation amplitude, which changes as a consequence of height differences along scan lines. The excitation of the cantilever takes place acoustically in most instruments by an additional piezo element situated near the cantilever holder or, in some instruments, by an oscillating magnetic field.

For data analysis, the AFM images have to be corrected first for line-by-line fluctuations, surface tilt and surface curvature. Such functions are usually included in the image-processing software provided by the manufacturers. This software allows one to perform some other basic operations, such as roughness calculations or measurements of height profiles along defined lines, but is often not sufficient for the needs of a specific experiment. Some expert laboratories use custom-made

(semi)automated image analysis software, for example, to analyze large sets of individual molecular structures. However, sophisticated software or software developers may not be available in many molecular biology laboratories. A versatile open-access image analysis software is ImageJ (NIH, USA; <http://rsbweb.nih.gov/ij/>), which allows one to perform a number of useful operations, although not in an automated manner, and that runs on all operating systems. A restriction is that it cannot open original AFM files, which must be first converted (e.g., to tiff format). For this reason, height values are not accessible in ImageJ. The conversion is done in the image-processing software of the AFM manufacturer. Free programs that are specifically designed for the analysis of scanning probe images (and are able to open original files generated by a large number of current AFM instruments) are ImageSXM (<http://www.liv.ac.uk/~sdb/ImageSXM/>) and Gwyddion (<http://gwyddion.net/>). ImageSXM is as versatile as ImageJ (both are conceptually equal) but only available for Macintosh operating systems.

The spatial resolution essentially depends on the sharpness of the sensor. The tip of the sensor is characterized by the end radius of curvature (R_c) of an assumed spherical shape (Figure 25.2). This radius is typically in the range of 5–15 nm for sensors mainly used for biological applications. Therefore, AFM is especially suitable for localizing substructures (a folded RNA domain or a bound protein) that are flanked by stretches of elongated RNA (dsRNA), but as mentioned above, the detailed molecular structure does not become visible. These substructures are rather recognized as peaks or globules (see below, Figure 25.3). The finite tip width induces an apparent broadening of molecules because any surface topography is a convolution of the shapes of tip and sample (outlined by the gray line above the obstacle in Figure 25.2). Because of this effect, dsRNA appears much broader (10–15 nm) than its real dimension. Substructures in very close neighborhood to each other may not be distinguishable for the same reason. The resolution may also be limited by the pixel size of the images. The maximum resolution is 512 pixels per scan line in most operating systems, but has been improved to 1024 pixels per

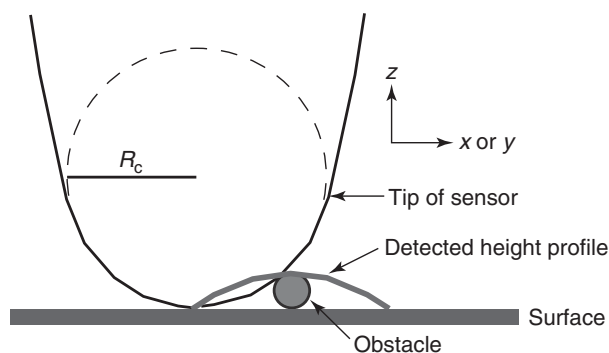


Figure 25.2 Schematic diagram showing the effect of tip–sample convolution in AFM imaging. R_c is the so-called radius of curvature which specifies the sharpness of the scanning tip. The detected height profile of an obstacle is broadened relative to its true width owing to a certain broadness of the scanning tip.

line in more recent systems. A trade-off is often necessary between scanning fields large enough to gauge a sufficiently large sample area within a reasonable time and an appropriate pixel resolution. For imaging RNA or DNA samples, scanning fields of 1–4 μm^2 are usually chosen, corresponding to pixel sizes between 1.9 and 7.8 nm (for 512 pixels per line). Thus, the tip width and pixel size are approximately on the same order of magnitude for these image sizes. Often, zooms of individual molecules are prepared from measurements, and analysis software provides the option to smooth images by pixel interpolation. For example, smoothing the zigzag contours that result from the image pixelization may be necessary for accurate length measurements (note that this does not increase the spatial resolution). The accuracy of measurement further depends on the type (size) of the x,y,z -scanner. Instruments designed for imaging larger scan fields, as for imaging cells, may not be optimal for imaging molecular structures.

A fragment of Turnip Yellow Mosaic Virus (TYMV) RNA that exhibits domains of secondary structure separated by stretches of C-rich ssRNA was examined by AFM to characterize the applicability of AFM for the analysis of complex natural RNAs [3]. Such structure was amenable to analysis by AFM imaging because the folded regions were distinguishable from the ssRNA stretches as globules. In addition, computer models of RNA folding [17] and other experimental data were available for comparison. In another study, RNA extracted from different icosahedral viral capsids was incubated for different time periods and then investigated by AFM [4]. Depending on the time of incubation in the absence of viral coat proteins, different compaction states were observed, indicating that secondary structural elements arranged in a sequential and linear manner along the RNA chain eventually interact to form a highly condensed structure. Here, AFM provided direct evidence that secondary structures can exist independent of tertiary structure and that folding occurs hierarchically.

25.2.2

Surface Preparation Conditions

Selecting between different sample preparation procedures is probably the most critical task for AFM imaging of RNA structures. Screening the literature indeed suggests a strong influence of preparation conditions, apparently stronger than in the case of dsDNA. Certain procedures are found inapplicable in one but appear successful in another study [3–5]. Nevertheless, surface preparations used for imaging RNA basically follow the same principles as for dsDNA. Thus, even though RNA secondary and tertiary structures require more specific precautions, the available AFM data on DNA are helpful in choosing appropriate conditions. Muscovite mica is mainly used as a surface for AFM of nucleic acids. This is an aluminosilicate mineral organized in thin layers that can be easily split. The mica sheet is mounted on a solid support (the type of which differs depending on the design of the AFM instrument). By pressing an adhesive tape against the top mica surface and then peeling off the tape, a fresh mica surface is “cleaved,” that is, some top layers are removed to obtain a clean, flat surface. The important



property of mica is its atomic flatness with a root-mean-square roughness of less than 0.1 nm. Layers of muscovite mica are highly negatively charged. Thus, nucleic acids, which are as well negatively charged at physiological pH, cannot be bound to mica without counterions. One widely used procedure for DNA is to supply the deposition buffer with divalent metal ions, usually Mg^{2+} , which mediate the interaction between the DNA and the mica. The interaction strength depends on pH, concentration, type of divalent cations, and the molar ratio of divalent to monovalent cations. The Mg^{2+} concentration is usually in the range of 1–30 mM and may have to be adjusted depending on the amount and type of monovalent ions which weaken the surface interactions. For imaging in air, the surface has to be rinsed with water before allowing it to dry because otherwise salts would remain on the surface and obscure the structures to be observed. While this simple method also worked for some RNAs [2, 3], it failed in another case where secure fixation of dsRNA was only successful by coating the mica with polylysine [5], as described below. Giro *et al.* [3] deposited TYMV RNA at varying Mg^{2+} concentrations and found only subtle structural changes below 10 mM, whereas higher order folding of the RNA seemed to be disrupted at 10 mM Mg^{2+} . The authors suggested that increased RNA–surface interactions are responsible for this effect.

Mica can be chemically modified in different ways to enhance the binding capacity. Glow discharging of mica for 1 min at 0.3 mbar was conducted before RNA adsorption in the presence of Mg^{2+} [7]. Rivetti *et al.* [18] also found enhanced binding of DNA strands to glow-discharged mica and to mica pretreated for several hours with water (H^+ exchange method) or with different salt solutions (KCl, LiCl, or NaCl). An efficient route to obtain an adhesive mica surface is its modification with basic peptides, poly-L-lysine (PL) [19] or poly-L-ornithine [20]. Coating with these agents generates a positively charged surface owing to protonated amino groups. A drop of an aqueous PL solution is placed on the mica, and the PL is allowed to bind for a certain time. Then, the surface is rinsed with a few milliliters of

Figure 25.3 Design and AFM imaging of a GUS-PSTVd-suppressor-tRNA^{Glu} RNA transcript. (a) Vector used for RNA transcription. The fragment to be transcribed by T7 or T3 RNA polymerase contains two β -glucuronidase (GUS) mRNA-derived segments that are complementary and form a continuous double helix, and three PSTVd segments that form imperfect double strands flanking the GUS hybrid. The single *Sma* I site within the PSTVd sequence was used to insert a sequence of interest (a tRNA^{Glu} domain in this case). The crossline at “11 o’clock” indicates the start of nucleotide numbering. (b) Schematic diagram of the structure of the folded RNA transcript synthesized from the plasmid

illustrated in panel (a). (c) Close-up views (136×136 nm) of the RNA transcript from the pT3T7 GUS-PSTVd-suppressor-tRNA^{Glu} vector imaged by AFM in air after adsorption to PL-coated mica. Observed structures were classified into the four groups denoted above the images. A focus of increased height is visible in the groups “globule” and “terminal globule.” The height scale ranges from 0 to 2.5 nm. (d) Histograms of the RNA backbone length measured separately for the four groups. The adsorption took place in 10 mM HEPES, pH 7.9, and 50 mM NaCl (left, without Mg^{2+}), or in 10 mM mM HEPES, pH 7.9, 10 mM NaCl, and 2 mM MgCl_2 (right, with Mg^{2+}).

clean water and dried under nitrogen before adsorbing the DNA or RNA. The minimal PL concentration that allows stable imaging of DNA is in the range of 1 (air) to 5 (liquid) $\mu\text{g ml}^{-1}$, as tested for a PL mixture of MW 500–2000 (Sigma-Aldrich) and an incubation time of 30 s before the drying step. The strength of DNA–surface interaction depends on the surface density of PL and the salt concentration during DNA deposition. This surface preparation is easy and the surface density of PL can be kept fairly constant between repeated preparations, so that reproducible binding conditions are obtained [19]. PL-coated mica was also used for imaging dsRNA [5]. (3-Aminopropyl)triethoxysilane (APTES) can be used to covalently modify mica, which also exposes amino groups on the surface [15]. APTES has to be deposited on mica from the vapor phase to obtain a smooth surface. For this purpose, freshly cleaved mica is placed in a dessicator together with small amounts of *N,N*-diisopropyl ethanolamine and APTES and left under argon for several hours. APTES-derivatized mica was shown to bind dsDNA and dsRNA [15]. The advantage of generating positively charged mica is that adsorption of nucleic acids can take place under a wide range of salt conditions. RNA–protein complexes may be deposited under conditions that were identified as optimal in previous *in vitro* experiments.

Depending on the properties of the surface, the adsorption mechanism of nucleic acids can be fundamentally different. The molecules may be subject to continuous conformational equilibration on the surface (before drying) or may be rather “frozen” in a certain conformation once bound, depending on the surface interaction strength. These scenarios were revealed by statistical analysis of surface conformations of dsDNA using the so-called worm-like (WLC) model of polymers in solution [18]: A useful parameter to describe the conformation of a polymer is the shortest distance R between its ends. WLC predicts the mean-squared distance $\langle R^2 \rangle$ for a population of filaments of given contour length L and bending rigidity B . The bending rigidity is described by the persistence length P ($P \sim B/(k_B T)$, with k_B representing the Boltzmann constant and T the absolute temperature). P is the length over which a filament is rather rigid (local deflections are governed by the competition between the thermal energy and the bending rigidity), whereas far above the length P , the filament behaves completely flexible. For such long filaments ($L \gg P$) in free solution, $\langle R^2 \rangle$ equals $2PL$ (see Ref. [18] for the complete formula necessary to describe filaments of intermediate length by WLC). The initial transfer to a surface may resemble a simple orthogonal compression of three-dimensional conformations, which would be verified by $\langle R^2 \rangle$ smaller than the theoretical value in free solution. However, it was shown [18] that dsDNA adsorbed to mica with Mg^{2+} relaxes on the surface and adopts ideal two-dimensionally equilibrated conformations; in that case, $\langle R^2 \rangle_{2D} = 4PL$. These conformations are more extended than in free solution because one dimension in which deflections can occur is removed. On the other hand, pretreatment of the mica to enhance the interaction strength, for example, by H^+ exchange or preincubation of the mica with salt solutions (see above), led to orthogonally compressed conformations as verified by reduced $\langle R^2 \rangle$ values. Corresponding observations

were made after adsorption of DNA and RNA to PL-mica [5, 19, 21]. Thus, knowledge of the adsorption scenario is a prerequisite to measure bending rigidity of dsRNA [5] and, generally, to optimize binding conditions for more complex RNAs.

25.2.3

Imaging in Liquid

The possibility to scan a sample surface in liquid could be very interesting with regard to investigations of RNA structures. This is not only because possible artifacts caused by the drying process are excluded but also because a better image quality can be achieved. Conformational states of hairpin ribozymes could be resolved owing to high-quality images of these small particles obtained in liquid [1]. In air, the surface and scanning tip are covered with a thin layer of water, leading to strong capillary forces, which have to be overcome each time the tip detaches from the surface while oscillating. Therefore, relatively stiff cantilevers with a spring constant in the range of 40 N m^{-1} are commonly used in air, whereas cantilevers for measurements in liquid have spring constants of less than 1 N m^{-1} . Probably because of the capillary forces, significantly reduced heights of dsDNA and dsRNA strands are usually observed in air (frequently 0.5 nm measured for dsDNA *versus* 2 nm expected). Measured DNA heights are typically two- to threefold larger in liquid. Measurements in liquid can thus give a better height contrast, and structural details may become more apparent. Furthermore, the drying procedure may distort RNA structures, firstly by mechanical forces and secondly by the sudden removal of ions during rinsing of the surface with water. Although no direct comparisons are available at present, it is conceivable that these factors influence fragile RNA structures. In the case of dsDNA, clear differences between wet and dried samples were observed in measurements of the differential flexibility along a DNA sequence [22]. The major limitations of imaging in liquid are that molecules are often not firmly attached, that the appropriate scanning parameters are more difficult to find, and that a sample often cannot be probed for a long period of time, because contaminants may slowly accumulate on the surface. Mg^{2+} is usually not sufficient for secure binding of DNA for imaging in liquid. Similar to DNA, RNA can be securely immobilized by adding Ni^{2+} to the deposition buffer or by pretreating the mica with Ni^{2+} [1, 23]. Modifying the mica with PL or APTES (see above) is another option. Finally, imaging in liquid may allow the observation of dynamic processes. Rhinovirus particles were adsorbed to mica in the presence of Ni^{2+} and the release of the RNA from the capsids on reducing the pH in the fluid cell could be followed *in situ* [23].

25.2.4

Experimental Example of Salt-Dependent RNA Folding Using a Designed RNA Construct

In order to observe protein binding to specific RNA sequences, a vector system was constructed into which an RNA motif of interest can be inserted (Figure 25.3a). *In vitro* transcription from this vector generates RNA that folds into an approximately

120-nm long rod-like structure consisting of RNA sequence elements derived from potato spindle tuber viroid (PSTVd), an internal extended dsRNA region derived from the beta-glucuronidase (GUS) gene and an inserted test sequence (Figure 25.3b). The test sequence protrudes from this backbone near one of its termini (Figure 25.3b). This system was used to observe binding of IRE binding protein to the IRE site [8] and binding of the RNA editing enzyme ADAR2 to natural editing sites [9]. Different tRNA sequences from the social amoeba *Dictyostelium discoideum* were recently inserted into the vector for studies of tRNA–protein interactions (unpublished results). The latter type of construct provides an example of a relatively simple RNA insertion domain that may display structural differences depending on the salt composition. Mg^{2+} is expected to promote folding of the tRNA domain, whereas the rigid secondary structures of the other parts of the transcript should not be affected by the absence or presence of Mg^{2+} .

To find out whether the expected Mg^{2+} influence is visible by AFM imaging, PL-coated mica was used as surface support to immobilize the RNA in the presence or absence of Mg^{2+} . After drying, imaging was done with a Multimode AFM (Veeco), with NCH Point probes (Nanosensors) operated in the amplitude modulation mode. The images revealed a certain heterogeneity of structures, which were classified into four groups, as shown in Figure 25.3c. The first group displayed a globule near one terminus and the second group a branch but no focus of significantly increased height. These two forms agree with the structure shown in Figure 25.3b, in which the protruding tRNA domain apparently exists in alternative conformations. The globules indicate a more compacted, and the branch an extended conformation of the tRNA domain. The third group showed a globule directly at one terminus, and the fourth group showed neither a globule nor a branch (termed “linear”). The lengths of all structures were determined by tracing the backbone contour using the software ImageJ, and length distributions were prepared separately for each group (Figure 25.3d). The frequencies in each group were normalized to the total number of molecules in a sample. The lengths for the groups “globule” and “branch” are in the expected range for the GUS-PSTVd backbone (about 120 nm). The number of molecules in the group “terminal globule” was small, and the lengths were in the range of the first two groups. In contrast, the “linear” form, which constituted a large fraction, showed a broad length distribution and was significantly shorter on average. The most frequent length of the linear form was 67 nm, which matches the length of the dsRNA region encompassing the GUS RNA of 240 bp (assuming a helical rise of 0.28 nm bp^{-1} for A-form RNA). Therefore, the linear form (and possibly also the “terminal globule” form) is most likely due to RNase degradation, primarily truncating the terminal regions that are characterized by more susceptible secondary structures such as bulges and loops. The activity of nucleases is always a problem in AFM analysis of RNA because RNase inhibitors would disturb the image. Comparison of the samples prepared with and without Mg^{2+} revealed that the branched form was predominant in the absence of Mg^{2+} , whereas structures with a globule were enriched in the presence of Mg^{2+} (Figure 25.3d). The most frequent backbone lengths did not differ between samples with and without Mg^{2+} . This indicates

that Mg^{2+} specifically promoted the formation of the globules near the terminus, suggesting a salt-dependent compaction of the tRNA domain.

Although these observations may reflect a salt-dependent equilibrium between folded and unfolded structures, it remains unknown whether, or to what extent, the tRNA domain was folded into a native structure; even different conformations that are not distinguishable on the images may have coexisted. In addition, effects originating from RNA–surface interactions and the drying procedure cannot be excluded. Further insight is expected by increasing the Mg^{2+} concentration and imaging in liquid. One should also take into account that investigation of protein binding to a certain RNA motif may be hampered if this motif folds into alternative three-dimensional structures that are not equally accessible for the binding protein(s). Finally, if the correct RNA structure is globular, the binding protein has to be larger than the globule in order to be distinguishable on the images.

The measured lengths of the terminal branches in the PSTVd-GUS-tRNA are in the range of 15–20 nm. Relating this size to dsRNA, the branches would correspond to about 50–70 bp. This may be considered as close to the lower limit of branches that are still clearly distinguishable as rod-shaped structures. The tRNA domain and the PSTVd(3) domain forming the terminal branch next to the tRNA are actually 89 and 159 nucleotides long, respectively. Therefore, although these domains are expected to contain bulges and loops in addition to double-stranded stretches, the measured dimensions of the terminal branches are in the expected range. The exact conformation cannot be addressed because of the resolution limit (the pixel size of 4 nm in the raw data of this experiment already corresponds to 14 bp of dsRNA).

25.3

Example Protocol: RNA Preparation for AFM Imaging in Air Using PL-Coated Mica

- To provide internal controls and to reduce the frequently observed preferential end binding of proteins to RNA, cloning of the (putative) protein interaction site into a vector such as pT3T7-GUS-PSTVd is recommended (Figure 25.3a). A single *Sma* I site within the PSTVd segment is used to insert the sequence of interest (usually 30–100 bp). Mfold analysis [17] of the expected transcript is carried out to increase confidence that the inserted sequence is exposed and does not disturb the structure of the backbone RNA portions.
- RNA is synthesized by *in vitro* transcription with T7 RNA polymerase after linearizing the pT3T7-GUS-PSTVd vector with *Sal* I in transcription buffer (40 mM Tris-HCl, pH 8.0, 30 mM MgCl_2 , 10 mM DTT, 10 mM NaCl, 10 mM spermidine; Fermentas). The RNA is extracted with phenol, precipitated with ethanol, and then further purified using Sephadex[®] G-50 columns. It is recommended to evaluate the RNA quality by native as well as by denaturing gel electrophoresis. In native agarose gels, the approximately 1000-nt long GUS-PSTVd backbone transcripts display a mobility similar to a 500-bp DNA fragment. The concentration is photometrically determined at 260 nm.

- Mica (mounted on an appropriate solid support) is freshly cleaved with an adhesive tape (Section 25.2.2). Modification with PL is done by incubating the mica with a 10 µg/ml aqueous PL solution (MW 500–2000; Sigma-Aldrich) for 30 s and subsequent rinsing with water and drying under a nitrogen stream. For binding the RNA to the surface, an RNA concentration of approximately 1 nM is appropriate (higher concentrations may be required for binding RNA to uncoated mica with Mg^{2+}). About 10–30 µl of the RNA solution in HEPES or Tris buffer at pH values in the range of 7.5–8.0 is applied to the surface for about 1 min or longer, and the surface is subsequently rinsed with clean water. After this step, the majority of the water is drawn off using a micropipette or tissue paper, and the surface is allowed to dry under a gentle flow of nitrogen.
- Reactions with protein (not included in the above experiment) are carried out under the appropriate buffer conditions in a reaction tube. Usually, it is not possible to apply the reaction mixture directly to the surface because complex formation requires protein concentrations that would result in overcrowding of the surface. Therefore, the mixture is diluted in the same buffer that is also used for binding the protein (or in the buffer supplemented with Mg^{2+} if the RNA is bound to unmodified mica) to an appropriate concentration (about 1 nM in 10–30 µl). Testing different protein concentrations is usually inevitable. The surface preparation for imaging in air is carried out as above. It may be advantageous to keep short the incubation time (1 min) with the surface, in order to minimize dissociation of the protein.
- For all aspects of sample preparation, MilliQ-purified or distilled water is used, including buffer preparations; all solutions are passed through a filter unit with a small pore size of 0.1 µm.
- Tapping mode scanning in air is done using silicon cantilevers with a nominal resonant frequency of 200–400 kHz, a spring constant of 10–50 N m⁻¹, and a specified end radius of curvature of <10 nm. High-quality images are taken at 0.5–1 Hz (5–10 min per image). The amplitude setpoint is typically adjusted to 5–15% below the amplitude measured without contact to the surface.

25.4

Troubleshooting

Problems during AFM imaging may not be specific for RNA and may as well be encountered in studies of DNA or DNA–protein interactions. RNA or DNA imaging requires a height contrast in the range of 1 nm, which poses high demands on the cleanliness of all solutions. However, troubleshooting may be more difficult when an RNA preparation is not homogeneous (e.g., owing to the presence of by-products from *in vitro* transcription or degradation products) or when the RNA adopts alternative structures. Insufficient image quality is usually due to the surface preparation or to a blunt or contaminated tip. In addition, tips differ in quality because of the manufacturing process. Typical observations and proposed solutions are listed below:

- *Structures on the surface appear broad*

A poor lateral resolution is due to a blunt or contaminated tip. Sharp images often require to further lower the amplitude setpoint after the surface has been approached. If this does not improve image sharpness, the tip has to be replaced. Tip contamination occurs primarily if the surface contains a high amount of impurities such as protein aggregates, which may be due to a too high sample concentration. Dilution of samples with clean solutions is the most effective way to reduce tip contaminations, which cause effects that may sometimes be mistaken as suboptimal instrument settings. Examples are “stripes” along single scan lines and low height contrast (even inverse images may occur). Such effects are likely due to stickiness of the tip. Washing or irradiation with UV light rarely removes contaminants. These may sometimes be stripped off by friction forces, which may be enhanced by scanning a few lines at a high contact force (using a strongly reduced amplitude setpoint) and high scanning speed, although this tends to wear out the tip.

- *“Shadow images”*

A duplication (or even multiplication) of structures indicates a double (or multiple) tip. The shadow image may not be obvious at the first glance. As multiple tips usually have different distances to the substrate, the parallel structures usually appear different in height and width. A weaker shadow image may be misinterpreted as single-stranded DNA or RNA. Depending on the design of the AFM instrument, the problem of double tips can sometimes be solved by tilting the measuring head in such a way that the shorter tip does not contact the surface.

- *Compacted or incompletely spread RNA on the surface*

A frequent artifact on air-dried surfaces is RNA or DNA appearing as blobs rather than as extended filaments. This kind of agglutination is sometimes difficult to distinguish from contaminants and potentially leads to the false conclusion that RNA does not bind to the surface or is fully degraded. Partial agglutination resulting in peaks along RNA or DNA strands can be misinterpreted as regions of higher order structure or protein binding. The compaction can be due to mechanical forces applied during rinsing of the surface and/or drying. Modifying the mica to enhance the surface interaction strength (as described above) can reduce the agglutination. However, an artificial compaction may as well be due to insufficient spreading of the RNA/DNA molecules before rinsing and drying. From experience, DNA that has been precipitated with ethanol or isopropanol should be dissolved for at least one day before AFM imaging.

- *Unexpected structures/aggregates*

High concentrations of molecules often result in non-specific aggregates that preclude a meaningful data interpretation. Such non-specific multimers may give the impression that the sample is inhomogeneous with many different sizes of protein. As only a “top view” of the molecules is possible, even specific dimers are frequently not observed as particles with twice the volume or twice the surface of a monomer. To reduce the sample concentration is a simple remedy for this problem. However, especially for RNA–protein interactions measured in air, concentrations in the reaction tube are not necessarily reflected in the

image because RNA, protein, and complexes may adhere with different efficiency to the surface. Usually, the lowest concentration showing complex formation in electrophoretic mobility shift assays is a good starting point to optimize imaging. As for biochemical assays, it may be helpful to briefly centrifuge the protein stock solution to spin down large aggregates. Small amounts of Nonidet-P40 (0.01% (v/v)) in the deposition buffer may avoid non-specific aggregation.

25.5

Force Spectroscopy AFM

In order to measure forces of molecular interactions, one interaction partner attached to the tip of the stylus is brought in contact with the surface that is modified with the other partner. The force–distance cycle is carried out by extension and subsequent retraction of the scanner in z direction (Figure 25.4). During approach, the cantilever is in its equilibrium state (phase 1) until it touches the surface. After contact, it is progressively bent upward (negative deflection) while the scanner continues to extend (phase 2). The direction of movement reverses

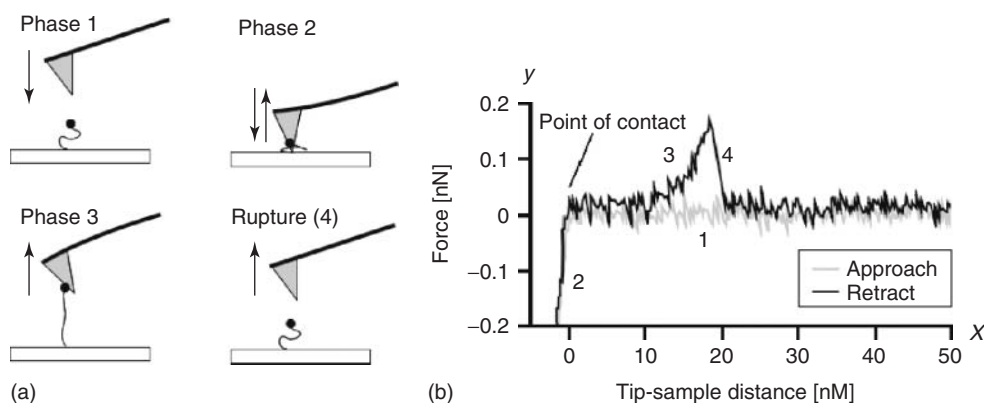


Figure 25.4 Force spectroscopy AFM. (a) Schematic illustration, where the arrows indicate the z -direction of the cantilever movement driven by the scanner. The tip is supposed to be functionalized such that it can form a specific interaction with the binding partner (black filled circle) at the end of a flexible linker (winding line) attached to the surface. Note that cantilever, tip, and size of the linker are not shown to scale. (b) Example of a force spectroscopy AFM measurement. The numbers in the force profile indicate the different phases

of a force extension cycle as depicted in (a) and described in the main text. For this measurement example, the sensor was functionalized with streptavidin to bind biotin attached to the 3'-end of 30-nt long ssDNA that was immobilized on a gold surface via a 5'-thiol group (see Section 25.4). The progression of the curve in phase 3 preceding the rupture of the interaction reflects the extension and overstretching of the ssDNA-linker conjugate (expected to be 18 nm long).

at a maximal deflection that the user specifies in the instrument settings. After withdrawal from the surface, the cantilever is bent downward (positive deflection) if an interaction between the functionalized tip and the partner molecule at the surface has formed (phase 3). The pulling force increases until the bond ruptures so that the cantilever snaps back to the equilibrium state.

To determine forces, the instrument must be calibrated to convert the voltage signal measured at the detector to deflection in nanometers and to force. Firstly, the deflection sensitivity (nanometers per volt) is obtained from the slope of the part of a force extension cycle where the stylus stays in contact with the surface and the cantilever is displaced linearly during extension and retraction of the scanner (phase 2 in Figure 25.4). The spring constant (N m^{-1}) can then be determined in different ways; a certain range of uncertainty must be taken into account. The thermal calibration method is included in recent operating systems. An alternative way is to measure the forces generated by pressing on a commercial cantilever with a known spring constant. In order to calculate the distance between tip and surface for each data point in a pulling measurement, the deflection d is subtracted from the scanner position $z((z - z_0) - d)$. The zero position z_0 (i.e., point of contact) is found at the intercept between the part of the curve showing the (constant) zero deflection out of contact with the surface (phase 1) and the part showing the (linearly) increasing deflection which is measured in contact with the surface (phase 2).

The smallest pulling forces that can be measured are in the range of a few piconewtons (pN). The force resolution is limited by the thermal noise of the cantilever, vibrations from the instrument, and electrical contributions. Force spectroscopy AFM has been further advanced by introducing different operation modes. Because of increasing processing power, it is possible to scan a surface by measuring force–distance curves pixel by pixel (force volume mode) and thus generate a force landscape. This allows one to establish maps of different properties, such as the spatial distribution of target molecules along a surface, material properties such as adhesion and elasticity and also the surface topography [24].

Intramolecular forces in RNA have been measured mainly with optical tweezers [25] or magnetic tweezers [5]. The detection of forces required to unfold RNA by AFM is difficult because these forces are close to the resolution limit of AFM. The force resolution is lower when compared with the first two techniques [10]. However, differences in the forces required to disrupt 30-bp long DNA/DNA and RNA/DNA hybrids with the same nucleotide sequence were detected [13]. A further important possibility is to probe RNA–protein interaction forces. Such measurements can provide new insight into the behavior of molecular bonds. The measured rupture forces depend not only on the interaction strength but also on the loading rate, that is, the speed at which force is applied to the interaction. This rate is the product of the retraction velocity and the spring constant of the cantilever if the bond remained at a fixed position (flexible linkers usually used to immobilize interaction partners add to the elasticity of the system). The most probable rupture force increases linearly with the logarithm of the loading rate because the contribution of thermal fluctuations to

the unbinding probability decreases. By varying the retraction velocity, the loading rate can be varied and the k_{off} rate constant can be determined from a plot of the mean rupture force against the logarithm of the loading rate. More than a single energy barrier for unbinding can be revealed when more than one linear regime is observed. In addition, varying the interaction time (the dwell time of the sensor at the surface) can reveal different interaction modes with different binding affinities [11, 24].

Protocol: Surface and Sensor Functionalization

The most critical requirement for force spectroscopy AFM is an appropriate chemistry for surface and sensor functionalization. Molecules have to be firmly fixed and have to be bound in such a way that the interacting domains remain accessible. The introduction of long linker molecules of polyethylene glycol (PEG) between the tip and one interaction partner will ensure that the latter can freely reorient while the sample surface exposing the complementary interaction partner is palpated [26]. In addition, unspecific interactions between the tip and surface have to be minimized. Designs must be adapted to the specific experiment and various alternative preparatory pathways are described in the literature. References [11, 26–29] offer a selection and are used in the following to give an insight into possible chemistries.

Functionalized Tips and Surfaces

- The covalent attachment of an interacting molecule to AFM sensors is often based on the modification of silicon nitride (Si_3N_4) tips with amino groups. In the protocol by Gruber and coworkers [26], amination is achieved by esterification of ethanolamine with the Si-OH groups present at the surface of the sensors. A total of 3.3 g ethanolamine are dissolved in 6 ml dimethyl sulfoxide (DMSO) by slowly heating to 60 °C in a clean glass beaker. After the solution has cooled down to room temperature, a molecular sieve monolayer (e.g., type 564, Roth, Karlsruhe, Germany) is added. When no more air bubbles are released from the sieve (which may take several hours), a clean cover glass is carefully immersed in the solution and placed on top of the beads. The cantilever is washed rigorously with chloroform, dried under a gentle nitrogen stream, and is placed on the immersed cover glass. The reaction is allowed to proceed for 14–16 h; subsequently, the cantilever is washed three times with DMSO, two times with 100% ethanol and is dried under a nitrogen stream. The amino groups at the tip can be used to attach a linker molecule or functionality of interest that carries *N*-hydroxysuccinimide (NHS) as the leaving group. In one approach, the bifunctional linker acetal-PEG-NHS was first coupled to the tip, followed by converting the acetal group to an aldehyde function for covalent attachment of a protein *via* lysine residues [27]. In the test experiment in Figure 25.4, the aminated tip was biotinylated by immersing the cantilever in a 1 mM biotin–NHS (Sigma-Aldrich) solution in DMSO for 2 h. After binding streptavidin, these cantilevers are appropriate to measure intramolecular forces

in nucleic acids attached to a surface and presenting a biotin moiety at their free ends [28]. Another approach to produce an aminated Si_3N_4 -tip is silanization. For coupling thiolated RNA [11], tips were activated by brief incubation with concentrated nitric acid and then with 2% APTES in toluene for 2 h. The linker NHS-PEG-maleimide is bound to the thiol-RNA through the maleimide function, and the aminated tip is subsequently functionalized with the RNA-PEG moiety using NHS as the leaving group [11]. Another possibility not based on amination is to coat AFM tips with a thin gold film and make use of the chemisorption of thiols to gold [29]. Gold is considered to be chemically inert, and resistance toward non-specific adsorption of molecules can be further increased by forming self-assembled monolayers (SAMs) of alkanethiols on the gold (thiol groups have a high affinity for gold). Mercaptohexanol (MCH) is often used for this purpose.

- Mica or glass surfaces can as well be activated by silanization to bind proteins or nucleic acids in subsequent reaction steps. An alternative substrate is gold, which may be advantageous for the reasons mentioned above. Thin gold films can be deposited by sputtering or evaporation. At least sputter coaters are quite easy to operate and are likely to be available at many research facilities. To produce a flat gold surface, the “template-stripped gold” (TSG) method has been introduced [30]. A gold film of 50–100 nm thickness is sputtered or evaporated onto freshly cleaved mica, followed by gluing the gold surface onto a glass surface using an epoxy adhesive. Immediately before usage, the mica is removed to bare the flat gold side that is in contact with the mica. This can take place chemically (e.g., by tetrahydrofurane), provided that such reagents do not dissolve the adhesive used to mount the gold surface onto the glass support after sputtering or evaporation (see above). The mica can also be removed by lifting it using a scalpel. The required complete mica removal is verified by conduction measurements (mica is a strong insulator, whereas gold is conductive). This is necessary even if the majority of mica layers were visibly detached because a few layers or patches of mica sometimes remain on the gold and are not easily distinguishable by eye. The TSG is cleaned by rinsing with 100% ethanol and dried with nitrogen. SAMs of MCH and thiolated biomolecules can be formed in one step. For the attachment of DNA or RNA, 100 μl of a buffered solution (5 mM sodium phosphate, pH 7.4) containing 9 μM MCH and 1 μM thiolated DNA or RNA are placed on the TSG, which is incubated with this solution for 4 h. The proportion of thiolated nucleic acid may have to be varied to yield an optimal probability of bond formation in the force spectroscopy measurements. The surface is rinsed with water and allowed to dry [31]. Alternatively, the thiolated nucleic acids are bound first and the surface is backfilled with MCH [28].
- The reactions described above have high demands on the purity of solutions and must, therefore, be prepared in clean containers. It should be pointed out that many protocols involve the cleaning of glassware with hazardous acids, bases or other mixtures. Usage of these substances strictly requires to gather information about special safety precautions.

25.6

Outlook

Observing systematic structural changes of RNA depending on different conditions is of interest with regard to investigations of structure–function relationships and folding mechanisms. For example, it has been suggested that, besides temperature and salt composition, varying the pH may provide valuable insight into the functioning of RNA molecules [3]. A requirement is sample material that is appropriate for the resolution power of AFM. The most significant difficulty is probably to discriminate between structural heterogeneity of RNA and surface-induced effects. By making use of the different available surface preparation procedures, a wide range of conditions can be tested, which may also allow the estimation of the extent of preparation artifacts. Force spectroscopy AFM can provide insight into the nature of molecular interactions, which is not possible with other techniques. Some modern instruments combine imaging and force spectroscopy options and include user-friendly operating software. Establishing surface and tip functionalization with RNA and protein is indeed labor-intensive but feasible because different chemistries are available and can be tested.

Acknowledgments

We thank Sara Müller and Nils Anspach for their experimental contributions and support.

References

1. Fay, M.J., Walter, N.G., and Burke, J.M. (2001) Imaging of single hairpin ribozymes in solution by atomic force microscopy. *RNA*, **7**, 887–895.
2. Hansma, H.G., Oroudjev, E., Baudrey, S., and Jaeger, L. (2003) TectoRNA and ‘kissing-loop’ RNA: atomic force microscopy of self-assembling RNA structures. *J. Microsc.*, **212**, 273–279.
3. Giro, A., Bergia, A., Zuccheri, G., Bink, H.H., Pleij, C.W., and Samori, B. (2004) Single molecule studies of RNA secondary structure: AFM of TYMV viral RNA. *Microsc. Res. Tech.*, **65**, 235–245.
4. Kuznetsov, Y.G., Daijogo, S., Zhou, J., Semler, B.L., and McPherson, A. (2005) Atomic force microscopy analysis of icosahedral virus RNA. *J. Mol. Biol.*, **347**, 41–52.
5. Abels, J.A., Moreno-Herrero, F., van der Heijden, T., Dekker, C., and Dekker, N.H. (2005) Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys. J.*, **88**, 2737–2744.
6. Smith, B.L., Gallie, D.R., Le, H., and Hansma, P.K. (1997) Visualization of poly(A)-binding protein complex formation with poly(A) RNA using atomic force microscopy. *J. Struct. Biol.*, **119**, 109–117.
7. Bonin, M., Oberstrass, J., Lukacs, N., Ewert, K., Oesterschulze, E., Kassing, R., and Nellen, W. (2000) Determination of preferential binding sites for anti-dsRNA antibodies on double-stranded RNA by scanning force microscopy. *RNA*, **6**, 563–570.
8. Bonin, M., Oberstrass, J., Vogt, U., Wassenegger, M., and Nellen, W. (2001) Binding of IRE-BP to its cognate RNA sequence: SFM studies on a universal

- RNA backbone for the analysis of RNA-protein interaction. *Biol. Chem.*, **382**, 1157–1162.
9. Klaue, Y., Kallman, A.M., Bonin, M., Nellen, W., and Ohman, M. (2003) Biochemical analysis and scanning force microscopy reveal productive and non-productive ADAR2 binding to RNA substrates. *RNA*, **9**, 839–846.
 10. Bonin, M., Zhu, R., Klaue, Y., Oberstrass, J., Oesterschulze, E., and Nellen, W. (2002) Analysis of RNA flexibility by scanning force spectroscopy. *Nucleic Acids Res.*, **30**, e81.
 11. Fuhrmann, A., Schoening, J.C., Anselmetti, D., Staiger, D., and Ros, R. (2009) Quantitative analysis of single-molecule RNA-protein interaction. *Biophys. J.*, **96**, 5030–5039.
 12. Husale, S., Persson, H.H., and Sahin, O. (2009) DNA nanomechanics allows direct digital detection of complementary DNA and microRNA targets. *Nature*, **462**, 1075–1078.
 13. Jung, Y.J., Park, Y.S., Yoon, K.J., Kong, Y.Y., Park, J.W., and Nam, H.G. (2009) Molecule-level imaging of Pax6 mRNA distribution in mouse embryonic neocortex by molecular interaction force microscopy. *Nucleic Acids Res.*, **37**, e10.
 14. Bustamante, C., Vesenska, J., Tang, C.L., Rees, W., Guthold, M., and Keller, R. (1992) Circular DNA molecules imaged in air by scanning force microscopy. *Biochemistry*, **31**, 22–26.
 15. Lyubchenko, Y.L., Gall, A.A., Shlyakhtenko, L.S., Harrington, R.E., Jacobs, B.L., Oden, P.I., and Lindsay, S.M. (1992) Atomic force microscopy imaging of double stranded DNA and RNA. *J. Biomol. Struct. Dyn.*, **10**, 589–606.
 16. Binning, G., Quate, C.F., and Gerber, C. (1986) Atomic force microscope. *Phys. Rev. Lett.*, **56**, 930–933.
 17. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
 18. Rivetti, C., Guthold, M., and Bustamante, C. (1996) Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.*, **264**, 919–932.
 19. Bussiek, M., Mucke, N., and Langowski, J. (2003) Polylysine-coated mica can be used to observe systematic changes in the supercoiled DNA conformation by scanning force microscopy in solution. *Nucleic Acids Res.*, **31**, e137.
 20. Podesta, A., Imperadori, L., Colnaghi, W., Finzi, L., Milani, P., and Dunlap, D. (2004) Atomic force microscopy study of DNA deposited on poly L-ornithine-coated mica. *J. Microsc.*, **215**, 236–240.
 21. Bussiek, M., Hoischen, C., Diekmann, S., and Bennink, M.L. (2009) Sequence-specific physical properties of African green monkey alpha-satellite DNA contribute to centromeric heterochromatin formation. *J. Struct. Biol.*, **167**, 36–46.
 22. Marilley, M., Sanchez-Sevilla, A., and Rocca-Serra, J. (2005) Fine mapping of inherent flexibility variation along DNA molecules: validation by atomic force microscopy (AFM) in buffer. *Mol. Genet. Genomics*, **274**, 658–670.
 23. Kienberger, F., Zhu, R., Moser, R., Blaas, D., and Hinterdorfer, P. (2004) Monitoring RNA release from human rhinovirus by dynamic force microscopy. *J. Virol.*, **78**, 3203–3209.
 24. Dufrene, Y.F. and Hinterdorfer, P. (2008) Recent progress in AFM molecular recognition studies. *Pflugers Arch.*, **456**, 237–245.
 25. Onoa, B., Dumont, S., Liphardt, J., Smith, S.B., Tinoco, I. Jr., and Bustamante, C. (2003) Identifying kinetic barriers to mechanical unfolding of the *T. thermophila* ribozyme. *Science*, **299**, 1892–1895.
 26. Ebner, A., Wildling, L., Kamruzzahan, A.S.M., Rankl, C., Wruss, J., Hahn, C.D., Hölz, M., Zhu, R., Kienberger, F., Blaas, D., Hinterdorfer, P., and Gruber, H.J. (2007) A new, simple method for linking of antibodies to atomic force microscopy tips. *Bioconjugate Chem.*, **18**, 1176–1184.
 27. Wildling, L., Unterbauer, B., Zhu, R., Rupprecht, A., Haselgrubler, T., Rankl,

- C., Ebner, A., Vater, D., Pollheimer, P., Pohl, E., Hinterdorfer, P., and Gruber, H. (2011) Linking of sensor molecules with amino groups to amino-functionalized AFM tips. *Bioconjugate Chem.*, **22**, 1239–1248.
28. Zang, W., Barbagallo, R., Madden, C., Roberts, C., Woolford, A., and Allen, S. (2005) Progressing single biomolecule force spectroscopy measurements for the screening of DNA binding agents. *Nanotechnology*, **16**, 2325–2333.
29. Krautbauer, R., Rief, M., and Gaub, H. (2003) Unzipping DNA oligomers. *Nanoletters*, **16**, 493–496.
30. Hegner, M., Wagner, P., Ungewickell, E., and Semenza, G. (1994) Immobilizing DNA on gold via thiol modification for atomic force microscopy imaging in buffer solutions. *FEBS Lett.*, **336**, 452.
31. Erts, D., Polyakov, B., Olin, H., and Tuite, E. (2003) Spatial and mechanical properties of dilute DNA monolayers on gold imaged. *J. Phys. Chem. B*, **107**, 3591–3597.

Part III

RNA Genomics & Bioinformatics, Global Approaches

26

Secondary Structure Prediction

Gerhard Steger

26.1

Introduction

The biological activity of RNA is quite diverse: messenger RNAs primarily code for proteins by sequence, but structural elements in the non-transcribed regions or even in the open reading frame may be of importance for translation; for most other RNAs (e. g., ribosomal RNAs, natural antisense RNAs, small interfering RNAs, and ribozymes), which are summarized by the terms structural or non-coding (nc) RNAs, the two- and/or three-dimensional structure is of utmost importance for their biological function. In this chapter, we concentrate on the prediction of such structures based on RNA sequence. Prediction of RNA secondary structure is much easier than the prediction of protein structure, because the stability of RNA secondary structure is dominated by nearest neighbor interaction (base pair stacking) of Watson–Crick and wobble base pairs, for which thermodynamic parameters are known. This thermodynamic background is introduced in Section 26.2. Given a complete set of thermodynamic parameters and the sequence, the problem to predict the structure of minimum free energy (mfe) or the structural equilibrium is an optimization problem that is solved by computational methods (Section 26.3). The background information should help the reader to make use of the “UNAFold Web Server” (with the underlying program packages *mfold* and *UNAFold*; see Section 26.4) and the “Vienna RNA Web Servers” (with programs such as *RNAfold* and *RNAcofold* from the “Vienna RNA package”; see Section 26.5), as well as to interpret the results more easily. The *RNAstructure* package [1] is an alternative to the *UNAFold* and *RNAfold* packages, respectively, but requires a local installation; a further description of *RNAstructure* is beyond the scope of this chapter. The next step after the prediction of a secondary structure – and its experimental verification – is computation and modeling of the tertiary structure (Chapters 30 and 31). This is in principle possible because the tertiary interactions contribute much less in energy than the basic secondary structure; in other words, nature builds an RNA tertiary structure on top of one of the best secondary structures.

26.2

Thermodynamics

Structure formation – from an unfolded, random coil structure C into the folded structure S – is a standard equilibrium reaction described best in an all-or-none manner by a temperature-dependent equilibrium constant K_T :

$$\begin{aligned} C &\rightleftharpoons S \\ K_T &= \frac{[S]}{[C]} \\ \Delta G_T^\circ &= -R T \ln K_T = \Delta H^\circ - T \Delta S^\circ \end{aligned} \quad (26.1)$$

At the denaturation temperature $T_m = \Delta H^\circ / \Delta S^\circ$ (melting temperature or midpoint of transition), the folded structure S has the same concentration as the unfolded structure ($K_{T_m} = 1$; $\Delta G_{T_m}^\circ = 0$). This is only true if the structure S denatures in an all-or-none transition. In most cases, however, temperature-dependent structural rearrangements and/or partial denaturations take place before the complete denaturation.

The number of possible secondary structures grows exponentially ($\sim 1.8^N$) [2] with the sequence length N . Accordingly, all possible structures S_i of a single sequence coexist in solution with concentrations dependent on their free energies $\Delta G_T^\circ(S_i)$; that is, each structure is presented as a fraction

$$f_{S_i} = \frac{\exp\left(-\frac{\Delta G_T^\circ(S_i)}{RT}\right)}{Q} \quad (26.2)$$

based on a partition function Q for the ensemble of all possible structures

$$Q = \sum_{\text{all structures } S_i} \exp\left(-\frac{\Delta G_T^\circ(S_i)}{RT}\right) \quad (26.3)$$

The structure S_0 of lowest free energy $\Delta G_T^\circ(S_0) = \Delta G_{T,\text{mfe}}^\circ$ is called the *optimal structure* or *structure of minimum free energy (mfe)*. But quite different structures with identical energies might exist for a single sequence. This is of special biological relevance for RNA switches [3]. Thus, one should not assume that an RNA folds into a single, static structure.

The relative concentration of the mfe structure may be very low when there are many structural alternatives (Eq. 26.2). Furthermore, the free energy of the ensemble is lower (more negative) than that of the mfe structure because of the entropy caused by the many different structures:

$$\begin{aligned} \Delta G_T^\circ(\text{ensemble}) &= -RT \ln Q \\ &= \Delta G_T^\circ(S_0) - RT \ln \left[1 + \sum_{i=1}^N \exp\left(-\frac{\Delta G_T^\circ(S_i) - \Delta G_T^\circ(S_0)}{RT}\right) \right] \\ &\leq \Delta G_T^\circ(S_0) \end{aligned}$$

The free energy ΔG_T° of a structure S is the sum of the free energy contributions of all structural elements of S :

$$\Delta G_T^\circ = \sum (\Delta H_{\text{stack}}^\circ - T \Delta S_{\text{stack}}^\circ) + \sum (\Delta H_{\text{loop}}^\circ - T \Delta S_{\text{loop}}^\circ) \quad (26.4)$$

Energetic contributions of base pairs are favorable ($\Delta G^\circ < 0$) because of stacking of the base pairs on top of each other, thus forming regular helices. Formation of loops is often, but not necessarily, unfavorable ($\Delta G^\circ > 0$); exact values depend on loop type, nucleotides adjacent to the loop-closing base pair(s), as well as the exact sequence of the loop and whether the loop nucleotides form a stable, structured motif. Loop types are classified according to the number of loop-closing base pair(s): a single base pair closes hairpin loops, while bulge loops (with no unpaired nucleotides in the other strand) and interior loops (with symmetric or asymmetric numbers of nucleotides in both strands) are closed by 2 bp, and multiloops (bifurcations, junctions) connect more than two helices (Figure 26.3). Note that for a bulge or interior loop of n nucleotides total length, there are (including the possibilities for the closing base pairs) up to $6 \times 6 \times 4^n$ different sequence combinations with different energetic contributions. A major set of these parameters has been measured by the group of Turner [4]. Parameters are known only within certain error limits; because these errors are smallest near $T = 37^\circ\text{C}$, mostly $\Delta G_{37^\circ\text{C}}^\circ$ values are reported.

A loop should not be thought of as a floppy structural element; in many cases, loop nucleotides have fixed orientations because of stacking and/or non-Watson–Crick (non-WC) interactions with other loop nucleotides. Famous examples are the loop E of eukaryotic 5S rRNA and the multiloop of tRNA. The loop E is an internal loop of four and five bases in its strands; all nucleotides are involved in non-WC interactions, including one triple-strand interaction. In tRNA, the stacking of multiloop-closing base pairs across the multiloop is a major energetic contribution to the stability of the cloverleaf and is critical for the formation of tRNA's tertiary structure.

Compensation of the negatively charged phosphate backbone of nucleic acids by positively charged counterions M^+ leads to stabilization of structural elements according to

$$C + n M^+ \rightleftharpoons S$$

$$K = \frac{[S]}{[C][M]^n}$$

From this, a logarithmic dependence between denaturation temperature T_m and salt concentration (ionic strength) follows:

$$\frac{dT_m}{d \ln[M]} = -n \frac{R T_m^2}{\Delta H^\circ}$$

All thermodynamic parameters for RNA structure formation were determined in 1 M NaCl. In terms of nucleic acid backbone neutralization, this is not far from the ionic conditions in cells except when specific interactions with divalent cations

play a role [5, 6]. If necessary, however, values for the ionic strength dependence of a structure or a structural element may be found in the literature, including functions accounting for the G:C-content of the RNA, or depending on various types of buffers (e.g., TRIS/borate) and cosolvents such as formamide or urea [7–14].

26.3

Formal Background

A structure of an RNA sequence R consists of base stacks and loops. It is defined as

$$R = r_1, r_2, \dots, r_N$$

with the indices $1 \leq i \leq N$ numbering the nucleotides $r_i \in \{A, U, G, C\}$ in the $5' \rightarrow 3'$ direction. Base pairs are denoted by $r_i : r_j$ or, in short, $i : j$ with $1 \leq i < j \leq N$. Allowed base pairs are A:U, U:A, G:C, C:G, G:U, and U:G. Formation of base pairs of a single secondary structure is restricted by

$$j - i > 3, \quad \text{which gives the minimum size of a hairpin loop (3 nt),}$$

and the order of two base pairs $i : j$ and $k : l$ has to be

$$i = k \text{ and } j = l \quad \text{or}$$

$$i < j < k < l \quad \text{or}$$

$$i < k < l < j$$

The second condition allows neighboring base pairs but disallows any triple strand formation; a base triple $j : k : l$ would force $i = k$ and $j \neq l$. The third condition allows the formation of several hairpin loops in a structure. The fourth condition explicitly disallows any tertiary interaction such as pseudoknots.

The search for an optimal secondary structure was initially solved by maximizing the number of base pairs for a given sequence [15, 16] using graph theory and dynamic programming [17]. The algorithm uses the possibility to dissect the problem into shorter and easier subproblems: if the maximum number of base pairs $M_{i,j}$ in the sequence region i, j ($1 \leq i < j < N$) and also the maximum number of all subsequences thereof is known, the following recursion does hold (Figure 26.1):

$$M_{i,j+1} = \max \left\{ M_{i,j}, \max_{i \leq k \leq (j-3)} [M_{i,k-1} + 1 + M_{k+1,j}] \rho(r_k, r_{j+1}) \right\} \quad (26.5)$$

with

$$r_k, r_{j+1} \in \{A, U, G, C\}$$

and

$$\rho(r_k, r_{j+1}) \begin{cases} 1 & \text{if } r_k \text{ and } r_{j+1} \text{ may form a base pair} \\ 0 & \text{otherwise} \end{cases} \quad (26.6)$$

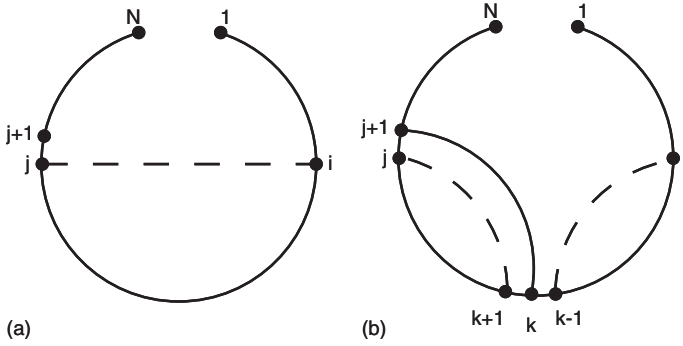


Figure 26.1 Principle of base pair maximization. The graphs symbolize a sequence of length N with given indices $1 \leq i \leq k < j < N$. The dashed line limits a sequence segment of which the maximum number of pairs is already known; the continuous line marks a base pair. (a) The maximum

number of base pairs for the sequence range i, j is $M_{i,j}$. For all subsequences within this range, the structures with maximum number of base pairs are also known. (b) The sequence range was increased to $i, (j+1)$ and a base pair $(j+1) : k$ with $1 \leq k \leq (j-3)$ should exist.

That is, the maximum number $M_{i,j}$ of base pairs for the subsequence from nucleotides i to j is already known (Figure 26.1a) and the subsequence is elongated to the subsequence i to $(j+1)$. The additional nucleotide r_{j+1} allows for two structural alternatives for the maximum number $M_{i,j+1}$ of base pairs for the subsequence i to $(j+1)$. Either the nucleotide r_{j+1} is not paired and the number of base pairs does not change ($M_{i,j+1} = M_{i,j}$; see Figure 26.1a) or the nucleotide r_{j+1} is paired to any nucleotide r_k in the subsequence i to $(j-3)$ (Figure 26.1b), provided that the nucleotide pair $r_k : r_{j+1}$ is a valid base pair ($\rho(r_k, r_{j+1}) = 1$ in Eq. 26.6). The base pair $r_k : r_{j+1}$ divides the subsequence from i to $(j+1)$ into two parts: the “excluded” fragment from i to $(k-1)$ and the “internal” fragment from $(k+1)$ to j ; for both fragments, the maximal numbers of base pairs are already known. The “new” base pair $r_k : r_{j+1}$ is restricted to the subsequence up to $k \leq (j-3)$ to allow for the formation of a hairpin-closing base pair with the minimum size (3 nt) of a hairpin loop $(j+1) - k - 1 \geq 3$. Next, from all possible structures with base pair $r_k : r_{j+1}$, the structure with the maximum number of base pairs is selected: $B = \max_{i \leq k \leq (j-3)} [M_{i,k-1} + 1 + M_{k+1,j}] \rho(r_k, r_{j+1})$. Finally, the maximum number of base pairs $M_{i,j+1}$ for the elongated subsequence i to $(j+1)$ is the maximum of both alternatives, $M_{i,j}$ or B : $M_{i,j+1} = \max\{M_{i,j}, B\}$.

The procedure starts with $i = 1$ and $j = 4$; no sequence range of this size may contain base pairs because of the minimal size of a hairpin loop. Next, the recurrent use of Eq. 26.5 allows the algorithm to determine all numbers $M_{i,j+1}$ by increasing j and i ($1 \leq i < j \leq N$). The maximum number of base pairs for the complete sequence is $M_{1,N}$. The corresponding structure is found by “backtracking”; that is, starting from the value $M_{1,N}$, the structure is reconstructed from the known values of the matrix M . The calculation cost for the algorithm is $O(N^3)$ because of the recursions of Eq. 26.5 with the three indices i, j , and k , which all are in

between 1 and N . The memory cost to store the matrix $M_{i,j}$ is $O(N^2/2)$ because $1 \leq i < j < N$.

With a matrix of size $N^2/2$, it is only possible to start a backtrack from $M_{1,N}$ toward the main diagonal. Structures with a lower, suboptimal number of base pairs are not found by the recursion (Eq. 26.5) because suboptimal solutions are overwritten by the optimal solution. The “trick” for not overwriting them is to use a virtually doubled sequence and a matrix V of size N^2 instead of the matrix M [18, 19].

- Each base pair $i : j$ divides a structure into two halves, called the *included fragment* from r_i to r_j and the *excluded fragment* from r_j to r_i . The optimal structure for the total sequence $r_1, \dots, r_N$ is obtained by joining of the two partial structures, which are optimal for the subsequences $r_i, \dots, r_j$ and $r_j, \dots, r_i$, respectively, from the two parts of matrix V : $V_{\max} = \max\{V_{i,j} + V_{j,i}\}$. This argument allows one also to compute structures for circular sequences.
- For storage of values for the partial structures with $r_i, \dots, r_j$ and $r_j, \dots, r_i$, a matrix V of size N^2 has to be used (Figure 26.2).
- To allow for multiloops, a further matrix W has to be used. In V at position i, j , structures are stored, which are optimal, and do contain the base pair $i : j$; in W , structures are stored, which are optimal for the sequence range $r_i, \dots, r_j$, but are not forced to have the base pair $i : j$.
- `mfold` and `RNAfold` use an algorithm similar to that sketched above. For energy minimization, however, the ρ values from Eq. 26.6 are replaced by energies according to the loop type (or base pair stack) under consideration. Additional cost in terms of memory and CPU time is needed because the energies depend on the type of loop nucleotides under consideration. This is especially critical for multiloops. `mfold` uses a simple linear dependence of energy on loop size during dynamic programming; only after the backtrack, corrections are applied

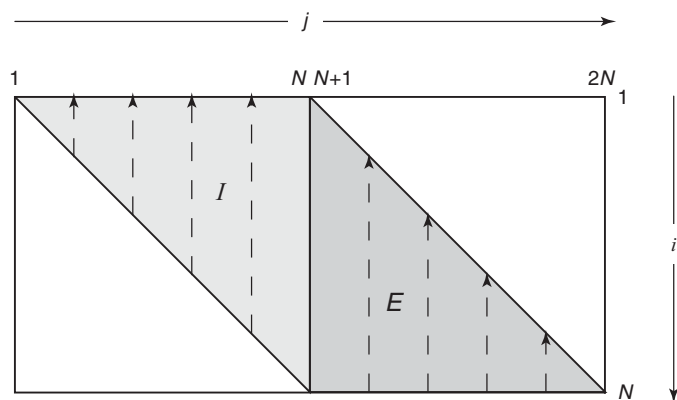


Figure 26.2 Memory allocation for matrices V and W in the algorithm for the prediction of suboptimal structures. The parts of the matrix filled with gray are used. Part I of the matrix contains values for the “included

fragment” from r_i to r_j . Part E of the matrix contains values for the “excluded fragment” from r_j to r_N to $r_{(N+i)} = r_i$. The matrix is filled during the recursion in direction of the arrows.

to account for the possibility of base pair stacking across the multiloop. *RNAfold* is able to use different models already in the dynamic programming step (see option “dangling end energies” in Section 26.4.1.4).

- *mfold* permits the user to find the mfe structure [20] and structures suboptimal in energy [19], which are optimal for any single, admissible base pair. Note that the calculable suboptimal structures do not cover all possible structures; for this another algorithm has to be used [21, 22].
- In *RNAfold* [23], two algorithms are used. The first algorithm computes only the mfe structure and produces the same structure as *mfold* when using identical energy parameters. The second algorithm computes the partition function (Eq. 26.3) as described by McCaskill [24]; this algorithm can be used to calculate the probability for the formation of each base pair instead of individual structures.

26.4

mfold and UNAFold

The UNAFold web server for detailed structural analysis of RNA, now located at the RNA Institute, SUNY Albany, NY, is the successor of the *mfold* web server, one of the oldest web services in computational molecular biology. The server is based on the UNAFold (“Unified Nucleic Acid Folding” [25]) and DINAMelt (“DI-Nucleic Acid hybridization and melting” [26]) program packages, which substitute and supersede the famous *mfold* package. More specialized variants of the main server – for folding many sequences at once (*Quikfold*), prediction of minimum folding energy only (*Zipfold*), prediction of two-state melting temperatures (T_m server), and hybridization server – will not be mentioned further. Authors who make use of the UNAFold service should cite the excellent description by the main developers [25]; the reference to DINAMelt is [26]. The standard reference to *mfold* is [27]; more detailed information is found in [19, 28] and references therein. The free energy parameters since *mfold* 3.2 are only valid at 37 °C; in case other temperatures are of importance, for example, for RNAs from plants or thermophilic organisms, one should use the *mfold* 2.3 server that uses a different and less complete set of parameters, which, however, allows for the calculation¹⁾ of free energies at other temperatures than 37 °C.

1) One needs ΔH° and ΔS° values to calculate ΔG_T° at any temperature T (Eqs. (26.1) and (26.4)). These ΔH° and ΔS° values are known by the older, less complete parameter set used in the *mfold* 2.3 server. For further structural elements, for example, internal loops with 2×2 nt in dependence on sequence, $\Delta G_{37^\circ\text{C}}^\circ$ values are additionally available in the parameter

set used in *mfold* 3.8 and *RNAfold*. The experimental standard deviations of the corresponding ΔH° and ΔS° values are quite high but largely compensate each other near 37 °C, which is not the case at significantly different temperatures. Thus, these additional parameters cannot be used at temperatures other than 37 °C.

The complete UNAFold package (version 3.8) is a free software and can be downloaded as source code that compiles easily on almost any flavor of Unix and Linux; precompiled binaries are also available (Table 26.1).

26.4.1

Input to the mfold Server

26.4.1.1 Sequence Name

A sequence name may be given to label any output from the server. If no sequence name is given but the sequence is in FASTA format, the FASTA header is used. Otherwise a date-based name is used. In any case, a few characters (“”, “<”, “>”, “’”, “\”) are deleted from the name.

26.4.1.2 Sequence

All characters except “A–Z” and “a–z” are removed from the text area box; that is, numbers and spaces are deleted but not headings such as those from a GenBank-formatted sequence. One may use sequence format conversion tools (Table 26.1) such as `readseq` or `sreformat` from the `SQUID` library to obtain either the raw sequence or a FASTA-formatted sequence. In the latter case, the header is used as the sequence name, if no sequence name is given in the corresponding box.

Case of characters has no special meaning but is retained in the output. T is converted to U. Certain characters have a special meaning, which does not coincide with the IUPAC ambiguous nucleic acid character convention:

- B, D, H, and V denote A, C, G, and U/T, respectively. These nucleotides may pair only if their 3′ neighbor is unpaired.
- W, X, Y, and Z denote again A, C, G, and U/T, respectively. These nucleotides may pair only at the end of a helix.
- N means any nucleotide but is not allowed to pair.

These conventions may help the user, for example, to incorporate knowledge from nuclease mappings.

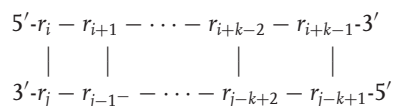
26.4.1.3 Constraints

Constraints forcing and/or prohibiting base pair(s) have to be given in a text area using a rigid format. Per line only one of the following constraints may be given:

- Forcing a base pair or a helix:

F i j k

will force the formation of a helix of k base pairs (or a single base pair with $k = 1$)



The first base pair of the forced helix is $r_i : r_j$ followed by $k - 1$ consecutive base pairs. For example, “F 3 11 2” would force the base pairs 3 : 11 and 4 : 10.

Table 26.1 Internet addresses of programs and manuscripts mentioned in the text.

Program/manuscript	URL	Reference/authors
UNAFold server	http://mfold.rna.albany.edu/	[25]
mFold server (at $T = 37^{\circ}\text{C}$)	http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form	—
mFold server (at $T \neq 37^{\circ}\text{C}$)	http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form2.3	[27]
mFold manual	http://mfold.rna.albany.edu/?q=mfold/documentation	—
UNAFold download	http://mfold.rna.albany.edu/?q=DINAMelt/software	[25]
mFold download	http://mfold.rna.albany.edu/?q=mfold/download-mfold	—
RNAfold server	http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi	[29]
RNA fold manuscript	http://www.tbi.univie.ac.at/papers/Abstracts/93-07-07.ps.gz	[23]
Vienna package	http://www.tbi.univie.ac.at/~ivo/RNA/	[30]
<u>Sequence conversion</u>		
readseq	ftp://ftp.bio.indiana.edu/molbio/readseq/	D. G. Gilbert
SQUID library	http://packages.ubuntu.com/maverick/readseq	
FASTA format	ftp://selab.janelia.org/pub/software/squid/	S. R. Eddy
<u>Graphics formats</u>	http://www.ncbi.nlm.nih.gov/BLAST/fasta.shtml	—
Png	http://www.libpng.org/	G. Roelofs
PostScript®	http://www.adobe.com/products/postscript/main.html	Adobe®
SVG®	http://www.w3.org/TR/SVG/	W3C®
<u>Structure drawing</u>		
R2R	http://breaker.research.yale.edu/R2R/	[31]
RNAMLview	http://mdbserver.rutgers.edu/services/	[32]
RnaViz	http://rnaviz.sourceforge.net/	[33]
XRNA	http://rna.ucsc.edu/rnacenter/xrna/xrna.html	H.F. Noller
Mac tct (loopdloop)	http://iubio.bio.indiana.edu/1UBio-Software+Data/molbio/loopdloop/	D. Gilbert

- Forcing a string of consecutive bases to pair (or a single base with $k = 1$):

`F i 0 k`

will force nucleotides $r_i, r_{i+1}, \dots, r_{i+k-1}$ to pair; the pairing partners are optimally selected by `mfold`. The third character of the input string is “zero”. For example, “`F 17 0 3`” would cause bases 17, 18, and 19 to pair.

- Prohibiting a base pair or a helix:

`P i j k`

will prevent the formation of a helix of k base pairs (or a single base pair with $k = 1$); the base pairs $r_{i+h} : r_{j-h}$, $0 \leq h < k$ are not allowed to form. For example, “`P 3 11 2`” would prevent the bases 3 : 11 and 4 : 10 to pair.

- Prohibiting a string of consecutive bases from pairing (or a single base with $k = 1$):

`P i 0 k`

will prevent nucleotides $r_i, r_{i+1}, \dots, r_{i+k-1}$ from pairing; the third character of the input string is “zero”. For example, “`P 3 0 2`” would prevent the bases 3 and 4 to pair.

- Prohibiting a string of consecutive bases from pairing with a string of consecutive bases:

`P i-j k-l`

will prevent nucleotides $r_i, r_{i+1}, \dots, r_j$ from pairing with nucleotides $r_k, r_{k+1}, \dots, r_l$ ($i \leq j, k \leq l$). If $i = k$ and $j = l$ then any base pair within the string $r_i, \dots, r_j$ is prohibited; that is, the segment $r_i, \dots, r_j$ is not allowed to pair with itself. For example, “`P 3-4 11-12`” would prevent pairing of bases 3 and 4 with bases 11 and 12.

Be careful with these constraints. For example, forcing a single base pair $r_i : r_j$ by the constraint “`F i j 1`” will fail when either no neighboring base pair is possible or this base pair itself cannot exist. Forcing too many base pairs, prohibiting too many nucleotides to be paired, or forcing an N nucleotide to pair will cause failure of the job.

26.4.1.4 Further Parameters

- RNA sequences may be linear or circular.

This is of importance because, for example, in a linear structure, the first and last nucleotides of the sequence may pair with each other without dangling ends; in a circular sequence, the first and last nucleotides of the linear input string may not base pair with each other but have to form a hairpin loop (or a different, energetically more favorable structure).

- The percent suboptimality number P .

This parameter controls the free energy range $\Delta G_{\text{mfe}}^\circ + \delta\delta G^\circ$ in which structures will be computed and their base pairs plotted in the energy dot plot (Section 26.4.2.1), that is, if you would like to see more suboptimal structures in the dot plot, increase the parameter, and vice versa. The default setting is $\delta\delta G^\circ = |\Delta G_{\text{mfe}}^\circ| P/100$ with $P = 5$, but the program limits $\delta\delta G^\circ$ values to $1 \text{ kcal/mol} \leq \delta\delta G^\circ \leq 12 \text{ kcal/mol}$ for any P and any $\Delta G_{\text{mfe}}^\circ$. Note that, according to Eq. 26.1, a suboptimal structure with $\Delta G^\circ = (\Delta G_{\text{mfe}}^\circ + 1 \text{ kcal/mol})$ at 37°C

will be by a factor of 5 lower in concentration than the mfe structure, and an energy difference of 10 kcal/mol will result in a factor of about 10^7 !

- The upper bound on the number of computed structures.
This number determines how many structural drawings are generated. It has no influence on the number of structures in the energy dot plot.
- The window parameter W .
This parameter controls how many structures will be automatically computed and how different they will be from one another. A smaller value of this parameter will result in more computed structures that may be quite similar to one another. A larger value will result in fewer structures that are very different from one another. More precisely, a distance between two base pairs $r_i : r_j$ and $r_k : r_l$ is defined as $W = \max\{|i - k|, |j - l|\}$. If $m - 1$ structures have already been predicted by mFold, the m th folding must have at least W base pairs that are at least a distance W from any of the base pairs in the first $m - 1$ structures.
- The maximum size of a bulge or interior loop.
This parameter is set per default to 30 nt. For bulge loops, it limits the maximum size. For interior loops with i and j single-stranded nucleotides on their two sides, the parameter limits the total size to $i + j \leq 30$. For most sequences at $T = 37^\circ\text{C}$, this is more than sufficient; larger values increase the computing time.
- The maximum asymmetry of a bulge or interior loop.
This parameter is set per default to 30 nt. For bulge loops, it limits the maximum size. For interior loops with i and j single-stranded nucleotides on their two sides, the parameter limits the size difference to $|i - j| \leq 30$. Again, for most sequences at $T = 37^\circ\text{C}$, this is more than sufficient and larger values increase the computing time.
Note, however, that even in natural sequences, long stretches of unpaired nucleotides do appear, and large loops may be present also at elevated temperatures before complete denaturation. In such cases, it is worthwhile to set the maximum-size and maximum-symmetry parameters to the sequence length.
- The maximum distance between paired bases.
If this parameter M is given, then any base pair $r_i : r_j$ with $i < j$ must satisfy $j - i \leq M$ for a linear sequence, and $\min\{j - i, N + i - j\}$ for a circular sequence of length N . Thus, a lower value of M favors more local pairings, whereas $M = N - 1$ allows for a pairing $r_1 : r_N$.
- Further parameters or options concern the types of output that are generated; most parameters may be given or modified also on the result page (Section 26.4.2). The structure drawing in Figure 26.3 was produced using default parameters.
 - Images are available in different image sizes and resolutions as png (portable network graphics; format with lossless compression) and jp(e)g (Joint Photographic Experts Group; format with not-lossless compression) files. Default (medium) values for size and resolution are 935×1210 pixels and 72×72 pixels per square inch, respectively. Most output is available also in PostScript and PDF (portable document format) format, which may be scaled to any size without loss in quality. Be aware, however, that the png and jpg hyperlinks of

the output page lead to clickable images, which allow the user to get additional information (Section 26.4.2.1).

- In structure drawings, nucleotides are shown as characters (“bases”) or not (“outline”); in the “automatic” mode, characters are drawn for sequences up to a length of 800 nt.

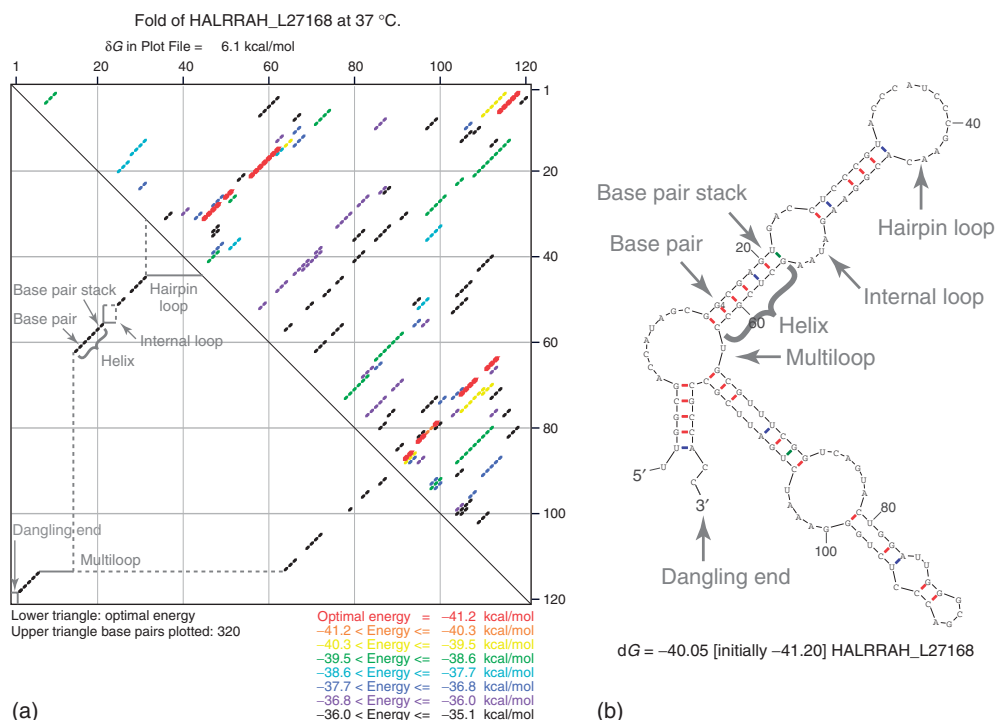


Figure 26.3 Major output produced by *mfold*. (a) Energy dot plot; the part below the diagonal shows the optimal (mfe) secondary structure; the part above the diagonal shows the mfe structure as red dots, and suboptimal structures—in order of decreasing stability—in dots of eight different colors (from red, orange, yellow, green, cyan, blue, purple to black); the free energy ranges for these are given below the plot. In this example, the best suboptimal structures (orange, yellow) are only variants of the mfe structure (red), whereas structures with more positive energy (green, cyan, blue, purple, and black) deviate significantly from the mfe structure. (b) Drawing of the optimal structure by *sir_graph*; base pairs are connected by lines (red for G:C, blue for

A:U, and green for G:U pairs). Note that the two neighboring pairs $^{80}\text{UG}^{81}$ and $^{98}\text{GU}^{97}$ are not marked by dots in the dot plot and are not connected by lines in the structure drawing (see also Figure 26.4a–d): *mfold* determines the energies of such “loops” from special 1×1 and 2×2 nt tables but not from the base pair stacking table; this allows *mfold* to use specific energy values for these “loops.” One should check the energy of these loops (see “thermodynamic details” and “png and jpg” hyperlinks in Section 26.4.2.5) and regard them as consisting of wobble pairs or non-canonical base pairs at least if $\Delta G_{\text{loop}} \leq 0$. Parameters: $T = 37^\circ\text{C}$, $P = 15$, $W = 1$; GenBank sequence accession: L27168.

- Nucleotides will be numbered automatically according to the length of the sequence or at user-selectable distances. A numbering frequency of 0 turns off numbering.
- The rotation angle, with positive values for counterclockwise rotation, modifies the orientation of structure drawings.
- Structure drawings may be annotated in different ways; for “p-num” and “ss-count” annotation see Figure 26.4. The “high-light” mode allows to enter region(s) of nucleotides, which are colored differently; regions may be given as a comma-separated list of positions (e.g., “10–20,30–40”).

26.4.1.5 Immediate versus Batch Jobs

Calculations for sequences of up to 800 nt are completed in several seconds, depending on the server load, and presented to the user; results will be deleted after about a day. For sequences up to 9000 nt, an e-mail address should be given and a “batch job” has to be started, which produces a page showing the URL for the results and the e-mail address to which a notification will be sent upon completion of the calculations, which needs at most a few hours. Results will be available for about two days after submission.

26.4.2

Output from the *mFold* Server

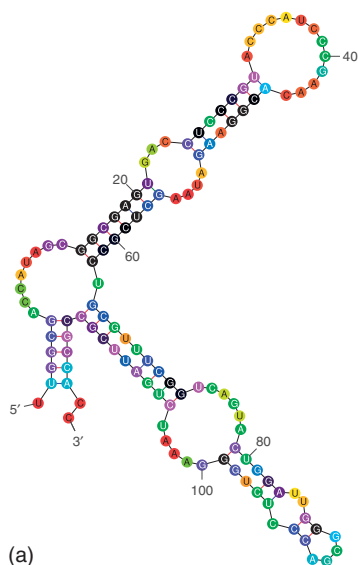
Computed results may be viewed by following links from the initial results page. On all result pages, two ΔG° values are given: the “initial” value is calculated as mentioned in Section 26.3; in a further reevaluation step, more complex energy rules (especially for multiloops) are applied, which are too time-consuming to be used in the first step. This reevaluation might alter the order of structures but not the structures themselves.

26.4.2.1 Energy Dot Plot

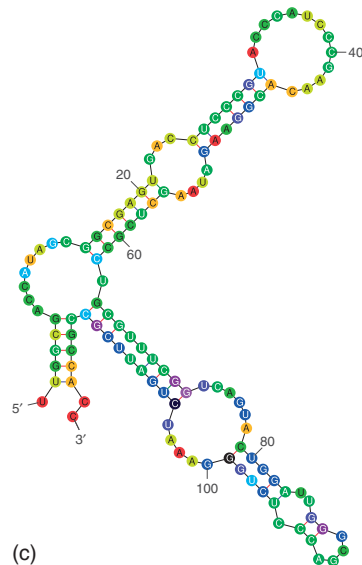
The mfe structure is presented in the part of the plot (Figure 26.3) below the diagonal (lower left triangle); a dot in row i and column j represents a base pair between the i th and j th nucleotide. The part of the plot above the diagonal (upper right triangle) contains the superposition of all computed structures within $P\%$ of the minimum energy, where P is the percent suboptimality parameter (Section 26.4.1.4). Red dots represent the mfe structure; dots of other colors represent superpositions of computed suboptimal structures. The colors are orange, yellow, green, cyan, blue, purple, and black, representing base pairs that are in structures with energies ΔG° in the ranges $\Delta G_{\text{mfe}}^\circ + (x - 1) \delta \delta G^\circ / 7 < \Delta G^\circ \leq \Delta G_{\text{mfe}}^\circ + x \delta \delta G^\circ / 7$, with $1 \leq x \leq 7$. Thus, the black dots represent base pairs that are least likely to form.

The “text” format of the energy dot plot gives the basic information on position of helices, the energy of the corresponding structure, and its energy range.

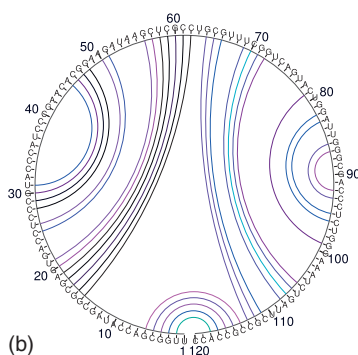
The png and jpg hyperlinks each lead to a new page with a png or jpg image of the energy dot plot and buttons that allow the user to get the position $i : j$ of a base pair by clicking on the corresponding dot (which also displays the optimal energy of the structure containing this base pair), to vary the number of colors (from 2 to



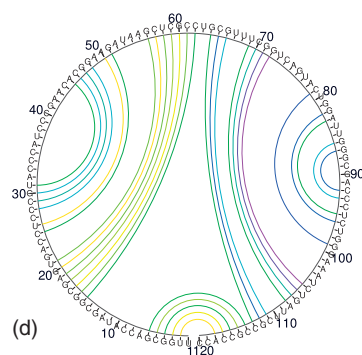
(a)



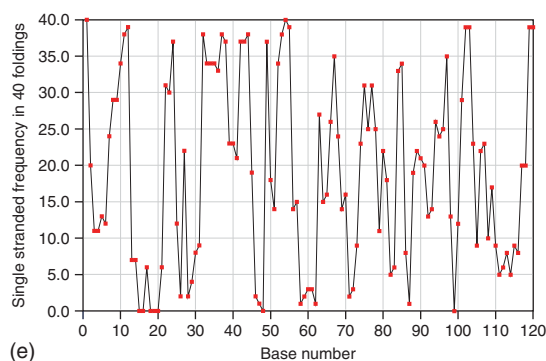
(c)



(b)



(d)



(e)

Color (Hex)	ss-count/%	p-num/%
ff0000	100.0	0.0
ff1f00	97.5	7.7
ffb000	85.0	15.4
ffff00	77.5	23.1
7fff00	70.0	30.8
1fff00	62.5	38.5
00ff7f	55.0	46.2
00bfff	47.5	53.8
0000ff	40.0	61.5
5f00ff	32.5	69.2
bf00ff	25.0	76.9
7f009f	17.5	84.6
1f003f	10.0	92.3
000000	0.0	100.0

(f)

8) used to illustrate suboptimal structures according to their energy, to define the energy window assigned to each color (energy increment in kcal mol⁻¹), and to limit the lower size of depicted helices (“filter”).

26.4.2.2 Extra Files

These are raw text files that make available the computed structures in “connect” (ct) format, the “p-num” values, also used for coloring structure representations, and “h-num” values [34, 35]. Note that p-num and h-num values depend on the chosen percent suboptimality parameter P .

p-num(i) is a measure on how likely the pairing of nucleotide i with any other nucleotide is

$$\delta(a) = \begin{cases} 1 & \text{if expression } a \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{p-num}(i) = \sum_{j < i} \delta \left(\Delta G_{j,i}^{\circ} \leq \Delta G_{\text{mfe}}^{\circ} + \delta \delta G^{\circ} \right) + \sum_{i < j} \delta \left(\Delta G_{i,j}^{\circ} \leq \Delta G_{\text{mfe}}^{\circ} + \delta \delta G^{\circ} \right) \quad (26.7)$$

That is, the p-num(i) value depends on the total number of dots in the i th row and i th column of the unfiltered upper triangle of the energy dot plot. Low values indicate “well-defined” nucleotides; in particular, values of 0 or 1 indicate that a nucleotide is single-stranded or paired with a defined partner, respectively, in all structures depicted in the energy dot plot. Higher values (closer to 1) indicate that the nucleotide is paired, but different alternatives exist for the pairing partner. The p-num values are used to annotate structure drawings by colors (Figure 26.4c,d): red denotes a nucleotide that is single-stranded in all computed structures, while black denotes a nucleotide paired to a defined partner in all structures (Figure 26.4f).

H-num(i,j) is a measure of the tendency of a nucleotide i to pair with a certain other nucleotide j :

$$\text{H-num}(i,j) = \frac{\text{p-num}(i) + \text{p-num}(j) - 1}{2}$$

Helices with lower h-num values are believed to be predicted more reliably.

26.4.2.3 Download All Foldings

All computed structures may be downloaded in a single zipped file (with extension “.zip”) or a compressed tar file (“.tar.Z”). Images are available in PostScript,

Figure 26.4 Structure output and annotation by mFold. (a,c) Secondary structure drawings by *sir_graph*; nucleotides are depicted as characters and colored dots, and base pairs are connected by lines. (b,d) Circle drawings by *sir_graph*; base pairs are connected by arcs. (a,b) Annotations by ss-count(i); for an explanation of ss-count values, see Section 26.4.2.4. (c,d) Annotations by p-num(i); for an explanation of p-num values, see Section 26.4.2.2. (e)

ss-count information useful for the comparison with structural probing data. (f) Coloring scheme for ss-count and p-num values. A value of 100% indicates that the corresponding nucleotide is predicted to be single-stranded in all structures (ss-count) or has a 100% probability to be base paired (p-num). Parameters: $T = 37^{\circ}\text{C}$, $P = 15$, $W = 1$; GenBank sequence accession: L27168.

pdf, png, or jpg format. Different formats useful for further processing by other programs (e.g., *RnaViz* [33], GCG, or *XRNA*; see Table 26.1) are available.

26.4.2.4 View ss-Count Information

The ss-count file contains statistics on single-strandedness of each nucleotide in all of the computed structures: $\text{ss-count}(i)$ is the number of structures in which the nucleotide i is single-stranded divided by the number of all computed structures. A corresponding plot is available, for which an “averaging window” parameter determines the nucleotide range over which the $\text{ss-count}(i)$ values are averaged (Figure 26.4e). For example, an averaging window of 9 means that the ss-count values are averaged for nt 1–9, then for nt 2–10, and so on. The plot is useful for comparison with structural probing data. The ss-count values are also used to annotate structure drawings by colors (Figure 26.4a,b): red denotes a nucleotide that is single-stranded in all computed structures, while black denotes a nucleotide paired in all structures (Figure 26.4f).

26.4.2.5 View Individual Structures

The link “Circular structure Plots” leads to a page that allows the user to produce circular graphs for structure representation (Figure 26.4b,d); this representation is quite useful for the comparison of different structures but fails for longer sequences.

Each of the following items – one for each individual structure – consists of links to different versions of structure representations:

- Clicking to “Structure i ” opens up an html file with the structure i in a simple, easily recognizable text format.
- The “thermodynamic details” assigned to each predicted structure list type, position, and free energy contribution of each structural element.
- The “PostScript” or “PDF” options give a static plot of structure i drawn in a manner as asked for on the input html page.
- The png and jpg hyperlinks each lead to a new page with a png or jpg image of structure i generated by *sir_graph* [27] and the “thermodynamic details” of the selected structure. Here, the structure representation may be magnified, differently annotated (for a description of the colors with $\text{p-num}(i)$ or $\text{ss-count}(i)$ annotation, see in the top right edge of the final html page; see example in Figure 26.4f), and labeled with dots and/or characters for the nucleotides (Figure 26.4a,c). Three versions for drawing loops are available (option “Algorithm”), and an option exists for different representations of the “exterior loop” (connecting 5'- and 3'-end).
- *ct* (connect format) [36], Vienna (Section 26.5.2.1), *RnaViz ct* [33], *Mac ct*, GCG connect, and *XRNA* are text formats useful for input into respective programs and further processing/refinement/labeling as wanted by the user.
- Computed structures are available in the RNAML (RNA markup language) format, which is the first standardized format for information exchange on RNA structure [37]. Up to now, only *RNAMLview* [32] and *VARNA* [38] are able to read this format, allowing the user to dynamically edit the two-dimensional structure drawing.

26.4.2.6 Dot Plot Folding Comparisons

This option plots several computed structures, selectable by the user, in a dot plot similar to the “energy dot plot option” for an easy recognition of those parts that are different between the depicted structures. Base pairs that occur in all selected structures are colored black, those that occur in two or more are colored gray, and base pairs unique to a structure are depicted in other colors. With the “Multicolor Overlap” option, the dots (squares) otherwise colored gray are depicted in different colors to allow for an easier discrimination between the different structures. For a comparison of three structures, base pairs occurring only in structures 1 and 2, only in 1 and 3, or only in 2 and 3 are plotted in three different colors in the presentation below the diagonal; base pairs common to all three structures are again colored black.

26.5

RNAfold

The RNAfold server (Table 26.1) is part of the Vienna package, which contains additional programs and web services related to RNA structure prediction and comparison as follows:

- RNAeval for calculation of the free energy of a given RNA secondary structure,
- RNAheat for calculation of specific heat of an RNA structure distribution (for comparison with measurements by differential scanning calorimetry),
- RNAinverse for inverse folding and design of sequences with a predefined structure,
- RNAdistance for the comparison of secondary structures,
- RNApdist for the comparison of base pair probability plots,
- RNAsubopt for complete suboptimal folding,
- RNAplot for RNA structure drawing,
- RNAcofold for prediction of the structural equilibrium of two RNA sequences A and B taking into account the five possible species (monomer structures A and B, dimeric structures $A \cdot A$ and $B \cdot B$, and the hybrid structure $A \cdot B$),
- RNAduplex for the prediction of optimal and suboptimal secondary structures of two RNA sequences A and B allowing only for intermolecular base pairs (hybrid $A \cdot B$).
- RNAalifold for the prediction of a consensus structure for a set of aligned RNA sequences, and
- RNAplfold and RNAfold to scan very long sequences for locally stable structures.

The server is hosted by the Theoretical Biochemistry group at the University of Vienna. The package is a free software and can be downloaded as C source code that compiles easily on almost any flavor of Unix and Linux. Authors who make use of the RNAfold service should cite the description by Hofacker [30]; for further details, see references therein.

26.5.1

Input to the `RNAfold` Server

In the following text, we concentrate on the web server; additional options, however, are available for a locally installed program, which will also be mentioned.

26.5.1.1 Sequence and Constraints

On the server, all non-alphabetic characters are removed from the text area box; that is, numbers and spaces are deleted but not headings such as those from a GenBank-formatted sequence. One may use sequence format conversion tools (see Table 26.1) such as `readseq`, which is also part of the Vienna package, or `sreformat` from the `SQUID` library to rewrite a sequence from any format into a FASTA-like format.

Lower case characters are converted to upper case; T is converted to U. Any symbols except AUCGTXKI will be interpreted as non-bonding bases.

- N means any nucleotide but is not allowed to pair.
- I can pair with A and U.
- K and X can only pair with each other.

With respect to the thermodynamic parameters, X is equivalent to G and K to C. A:I and I:U pairs use A:U parameters; an unpaired I is equivalent to N. A reasonable use of these nucleotide symbols is rare, but they may help the user to impose certain constraints.

The locally installed program reads sequences from stdin. The allowed sequence format is similar to the FASTA format; that is, the first line consists only of

```
>name
```

where “name” is a single string of up to 12 characters. The full sequence has to be given on the next line. Further sequences may follow, up to either an “@” in case of reading from stdin or end of file (EOF).

Constraints may be given by special characters (Section 26.5.2.1) in a third line, or in a separate input field in the web server:

- a dot “.” is an unconstrained nucleotide,
- a pair of matching brackets, “(” and “)”, forces the respective nucleotides to base-pair,
- a pipe symbol “|” forces this nucleotide to be paired,
- angle brackets “<” or “>” force pairing of the nucleotide with an upstream or downstream nucleotide, respectively,
- an “x” prohibits pairing of this nucleotide.

This constraint or structure string has to have the same length as the sequence string. In partition function folding, symbols except “.”, “)”, “(” and “x” are ignored.

For web server users, it is most convenient to align the constraint string below the sequence in a text editor and then to copy it into the corresponding `RNAfold`

server window (opened by clicking on “Show constraint folding” at the RNAfold input page). Local RNAfold users have to use the option “-C”. As an example, a file “mysequence.vie” with the content

```
>dummy
AGAGGGUUUCCUCU
.x.(.....)....
```

is used for calculation of the mfe structure and the partition function and the text output is written to the file “mysequence.log” with the command

```
RNAfold -C -p < mysequence.vie > mysequence.log
```

Because constrained folding works by assigning bonus energies to all structures complying with the constraints, sometimes single bracket pairs or a single angle bracket is not sufficient to get the desired output.

26.5.1.2 Further Parameters

In the following text, options that may be activated by appropriate buttons on the server or by command line options are outlined. A few additional options are only available *via* the command line; for these, check the manual page.

- Fold Algorithm.

RNAfold uses two algorithms as mentioned in Section 26.3. The first computes the optimal (mfe) structure only. The second, activated with option “-p” in the command line mode, computes the partition function and pair probabilities. The mfe structure and the pair probabilities are depicted in a probability dot plot (see below), which is similar to the energy dot plot of mfold. On the RNAfold web server input page, the default setting under “Fold algorithms and basic options” is “minimum free energy (MFE) and partition function”.

- Temperature “-T”.

The only possible temperature for thermodynamic parameters from Mathews *et al.* [4] is 37°C (as in mfold 3.5). When choosing other temperatures, parameters from Walter *et al.* [39], including ΔS values (Section 26.2), are used (as in mfold 2.3).

- No special tetraloops “-4” (option not available *via* web server).

Certain hairpin-loop sequences (especially UNCG and GNRA with N for any nucleotide and R for purine) are thermodynamically more stable than other sequences [40]. This option discards the appropriate tetraloop parameters, so that the influence of such sequences and parameters on a structure distribution is ignored.

The following options are accessible under “Show advanced options” on the web server input page.

- No dangling end energies “-d[0|1|2|3]”.

This option has influence only on the treatment of dangling ends (exterior loop at 5'/3' end) and of multiloops [41]: do unpaired nucleotides at free ends, or those that connect helices in multiloops, stack onto a neighboring base pair and is there coaxial stacking of helices in multiloops? For other unpaired nucleotides in bulge, interior, and hairpin loops, stacking is already considered by the thermodynamic parameters.

- Options “-d0” and “-d” (both are equivalent) ignore a possible influence of dangling ends or unpaired nucleotides connecting helices in multiloops.
- With “-d1”, which is the default for calculation of the mfe structure, only an unpaired nucleotide is allowed to stack on at most one base pair. For example, in a single structure, a nucleotide connecting two helices in a multiloop structure can either stack on the terminal base pair of helix 1 or on that of helix 2. Partition function folding will work as with “-d2”.
- With “-d2”, which is the default for partition function folding, the check from “-d1” is ignored. This treatment of dangling ends gives more favorable energies to helices adjacent to one another, which can be beneficial because such helices often do engage in stabilizing interactions through coaxial stacking. For example, an unpaired nucleotide connecting two helices in a multiloop can stack on the terminal base pairs of helix 1 and helix 2.
- With “-d3”, mfe folding will allow for coaxial stacking of adjacent helices in multiloops. Partition function folding will work as with “-d2.”

Since partition function and mfe folding treat dangling ends differently in the default mode, use “-d2” in addition to “-p” to ensure that both algorithms use the same energy model.

- No GU pairs at the end of helices “-noCloseGU” (accessible on the web input page under “Fold algorithms and basic options”).
No GU pairs “-noGU”.
Both are quite outdated options.
- Avoid isolated base pairs “-noLP” (default setting on the web server).
Avoid structures with lonely pairs (helices of length 1). For partition function folding, this disallows only pairs that are predicted to be isolated in all structures.
- Mountain plot (default setting on the web server).
A mountain representation of the mfe structure shows each base pair $i:j$ with $\Delta\gamma = 1$ and $\Delta\gamma = -1$ at positions $x = i$ and $x = j$, respectively, and $\Delta\gamma = 0$ for loops. This results in a mountainlike graph (Figure 26.5e), where helices are slopes, loops are plateaus, and hairpin loops are peaks. If the partition function is computed, the $\Delta\gamma$ values correspond to the pairing probability.

26.5.1.3 Immediate *versus* Batch Jobs

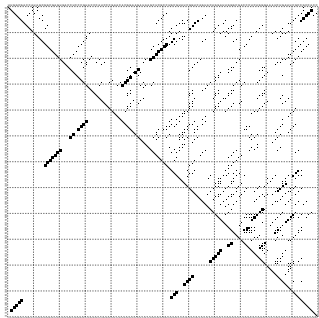
Calculations for sequences of a few hundred nucleotides are completed in several seconds, depending on the server load, and are directly presented to the user. For sequences up to 7500 nt (with partition function folding) and 10 000 nt (only mfe prediction), an e-mail address should be given for notification on completion of the calculations.

```

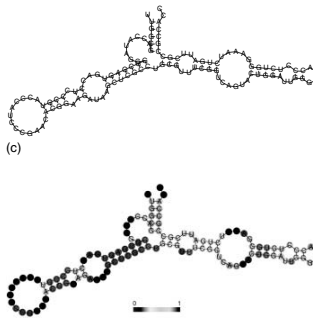
> HALRRAH_L27168
UUGGGGACCAUAGCGGAGUAGCCUCCGUAACCAUCCGAAACAGGAAGAUAAAGCUCGCGUUCGUGCAGUACUGGAUUGGGGACCCUCUGGGAAAUUGAUUCCGCGCACCC
·(((((((.....(((((((.....(((((((.....)))))).....)))))).....)))))).....))))))..... (-42.02)
·(((((((.....(((((((.....(((((((.....)))))).....)))))).....)))))).....))))))..... [~43.88]
·(((((((.....(((((((.....(((((((.....)))))).....)))))).....)))))).....))))))..... {~28.02 d=17.90}
frequency of mfe structure in ensemble 0.0489972; ensemble diversity 26.20

```

(a)

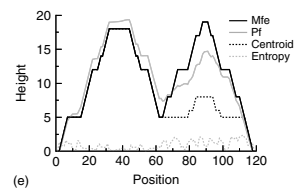


(b)



(c)

(d)



(e)

Figure 26.5 Major output produced by RNAfold. (a) Structure output in text format (local version); the first and second lines are identical to the input sequence; the third line is the optimal (mfe) structure in dot-bracket notation; the fourth line is a crude representation of the structure distribution; the fifth line is the centroid structure; (b) dot plot; the part below the diagonal presents the mfe structure, the part above the diagonal illustrates base-pairing probabilities; (c) mfe structure; (d) mfe structure annotated with base pair probabilities (colored in the output); and (e) mountain plot representation of mfe structure (black), partition function (gray), centroid structure (dotted black), and pairing entropy (dotted gray). Parameters: “-p -d2 -noLP”; GenBank sequence accession: L27168.

26.5.2

Output from the `RNAfold` Server**26.5.2.1 Text Output of Secondary Structure**

The text output (Figure 26.5a) is available only in parts from the web server.

For local `RNAfold` versions, the first output line presents the name of the input sequence. The second line is the sequence in uppercase letters with T converted to U if necessary. The third line contains the optimal (mfe) structure in dot-bracket notation with unpaired nucleotides indicated by dots and base pairs denoted by pairs of opening and closing bracket. At the end of the line is given the structure's free energy ΔG_T° (in kcal mol⁻¹). If the partition function was computed, further lines are given: the fourth line shows a representation of the partition folding with symbols denoting nucleotides that are essentially unpaired ("."), weakly paired (","), strongly paired without preference ("|"), weakly paired upstream ("{" or "}") or downstream ("[" or "]", or strongly paired upstream ("[" or "]", followed by the free energy of the structure distribution; the fifth line shows the centroid structure, which contains only base pairs predicted to pair with a probability above 0.5, followed by the free energy of the centroid structure and the distance d to all other structures ($d(S) = \sum_{(i,j) \in S} (1 - p_{ij}) + \sum_{(i,j) \notin S} p_{ij}$); the sixth line gives the relative concentration of the mfe structure in the structural distribution.

The results page of the `RNAfold` web server presents the mfe and centroid secondary structure as sequences in dot-bracket notation, their ΔG_T° values, the frequency of the mfe structure in the ensemble, and ensemble diversity. Then mfe and centroid secondary structures are presented, with coloring of base-pairing probabilities and download options. At the bottom, mountain plots for the mfe structure, centroid structure, and partition function are presented, followed by a positional entropy plot. The output page terminates with the corresponding `RNAfold` command line, which includes the name of the input sequence. From the web server, the mfe structure can be downloaded in dot-bracket notation (Vienna format) and in connect (ct) format for further processing. Furthermore, thermodynamic details of the mfe structure, such as type, position, and free energy contribution of each atomic structural element, are available *via* the `RNAeval` server.

26.5.2.2 Probability Dot Plot

If partition function folding ("-p") was selected, the major output of `RNAfold` is a dot plot (Figure 26.5b) that depicts the base pairs of the mfe structure below the diagonal. Above the diagonal, all probable base pairs from the partition function are depicted; the area of a dot in row i and column j is proportional to the probability of a base pair between the i th and j th nucleotide.

26.5.2.3 Graphical Output of Secondary Structure

The optimal (mfe) secondary structure is displayed (Figure 26.5c) in up to six different versions *via* an SVG viewer and is available in different graphic formats (eps, encapsulated postscript; pdf, portable document format; png, portable network graphics; and gif, graphic interface format). Note that all structures are given

by default in counterclockwise orientation; for mirroring and/or rotating the structures, the Vienna package contains a small Perl script (`rotate_ss.pl`). For annotation of a secondary structure by “reliability information” (Figure 26.5d), the Vienna package contains a small Perl script (`relplot.pl`) that colorizes the structure using either the positional entropy $S(i)$ by

$$S(i) = - \sum_{j \neq i} p_{i,j} \log_2(p_{i,j}) - \left(1 - \sum_{j \neq i} p_{i,j}\right) \log_2 \left(1 - \sum_{j \neq i} p_{i,j}\right)$$

or probability $p_{i,j}$ of forming the base pair $i:j$. Entropy and probability, respectively, are encoded as a color hue, ranging from red (low entropy, either well-defined loop nucleotides or base pairs) *via* yellow and green to blue and violet (high entropy, either ill-defined loop nucleotides or base pairs). This annotation is similar to the p-num annotation by `mfold` (Eq. 26.7). On the web server results page, the entropy and probability color codes can be switched directly by clicking respective buttons below the mfe and centroid structures (“Other display options”).

26.5.2.4 Mountain Plot

The Vienna server produces a mountain representation of the mfe structure (Figure 26.5c), of the partition folding, of the centroid structure, and of the pairing entropy. The pairing entropy is a measure of well-definedness:

$$S(i) = - \sum p_i \log_2(p_i)$$

with well-defined regions having low entropy. A PostScript image with the same information is available *via* a Perl script (`mountain.pl`) from the Vienna package (Figure 26.5e).

26.6

Troubleshooting

My major advice with any structure prediction program is to get some experience. For example, fetch two, relatively short and similar sequences from a data base such as GenBank, EMBL, or Rfam; either a consensus structure or an experimentally confirmed structure should be known for the sequence family from the literature or a database (e.g., [42]). The example used here is an archaeal 5S rRNA (AC: L27168); its predicted mfe structure is close to the 5S rRNA consensus structure. A second RNA could be another 5S rRNA (e.g., AC: AE010299, nts 1077681–1077815), for which the programs predict a deviating mfe structure. Next, reformat the sequences into a suitable format, and then play with the programs.

- Do not expect that a “good” structure has more base pairs than a “bad” structure; even for a random RNA sequence, about two thirds of its nucleotides are involved in base pairs.
- The quality of structure predictions decreases with increasing sequence length [43]. On the one hand, this may reflect uncertainties in the basic thermodynamic

parameters: errors, especially due to multiloops or other special loops, may accumulate with increasing sequence length. On the other hand, kinetics may increasingly influence the folding with increasing sequence length. The programs mentioned here are based on thermodynamics only; for programs that take into account the kinetics of folding, see [44–48].

- In case data from *in vitro* experiments such as chemical probing are available, the experimental conditions have to be taken into account. For example, most gel buffers have a very low ionic strength (e.g., 8.9 mM TRIS/borate has about 3 mM ionic strength), and temperatures other than 37°C may have been used.
- Be very careful with the usage of probing data in constraining any computations. For example, when multiple structures coexist in solution, mapping data would suggest more nucleotides to be single-stranded than actually are single-stranded in an individual structure. In my opinion, it is easier to compare probing data with pairing probabilities or ss-count information; if this does not fit, then add few constraints to favor individual suboptimal structures in order to reveal those structures that correlate best with the experimental data. An alternative is the use of pseudo-free energies for structure prediction, where energies are based on quantitative probing experiments [49].
- It may also be helpful if your sequence is one of a family of sequences, because all these homologous sequences should form a common structure; methods to search for homologous ncRNAs are described in Chapters 28, 29, and 33–38. To find such a common structure, one would like to simultaneously perform a sequence and structure alignment. However, this has an unaffordable computational cost of $O(N^{3m})$ for m sequences of length N [50]. Hence, several simplifying and more pragmatic approaches for consensus structure prediction have been developed, which can be classified according to their approach as follows [51]:

- Align the sequences first, then predict the structure common to the aligned sequences (e.g., KNetfold [52], EvoFold [53], ConStruct [54], RNAalifold [55]).

For the primary alignment step, pure sequence alignment programs or sequence + structure alignment programs can be used. In case of low sequence identity, however, the pure sequence alignment programs tend to fail; owing to the isostericity of all WC pairs (and other groups of non-WC pairs) [56, 57], the structure common to homologous RNAs can easily be conserved, while their sequences may differ to a large extent. Dynamic programming (for secondary structure prediction) [54, 55] or “maximum weighted matching” (for secondary structure prediction including pseudoknots or base triples) [54, 58] might be used in the structure prediction step based on a fixed alignment. Several RNA sequence + structure editors are available (e.g., RALEE [59], S2S [60], 4SALE [61], ConStruct [54]), which allow a user to refine the initial alignment.

- Predict structures for all single sequences, then align these structures (e.g., RNAforester [62], StrAl [63], RNASampler [64], R-Coffee [65]).
- Align and predict structures at the same time, using heuristics and/or restrict the alignment to two sequences to lower the computational cost (e.g.,

CARNAC [66], pmcomp/pmmulti [67], StemLoc [68], CMfinder [69], DynAlign [70], foldalign [71], LocARNA [72], LARA [73], MASTR [74], Murlet [75], PARTS [76], MAFFT/xinsi [77]).

Visualization of the additional information from RNA structural alignments is possible *via* MatrixPlot [78] or structure logo [79] (Table 26.1).

- A few programs are available that allow the user to compute tertiary interactions such as pseudoknots (e.g., PKNOTS [80], NUPACK [81], KNetfold [52], pknotsRG [82], ConStruct [54], DotKnot [83]); however, the user has to take into account the necessary computing resources and to judge the predicted output with a critical eye.
- The energy unit used within the programs mfold and RNAfold is kcal mol⁻¹; this unit should be replaced with the standard IUPAC unit kJ mol⁻¹ (1 kcal = 4.1868 kJ).

Acknowledgment

I thank R. K. Hartmann (Philipps-Universität Marburg), I. Hofacker (University of Vienna), and M. Zuker (Rensselaer Polytechnic Institute) for helpful discussions and critical reading of the manuscript.

References

1. Reuter, J.S. and Mathews, D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.
2. Waterman, M.S. (1995) *Introduction to Computational Biology. Maps, Sequences and Genomes*, Chapman & Hall, London.
3. Nagel, J.H. and Pleij, C.W. (2002) Self-induced structural switches in RNA. *Biochimie*, **84**, 913–923.
4. Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
5. Draper, D.E. (2008) RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys. J.*, **95**, 5489–5495.
6. Ramesh, A. and Winkler, W.C. (2010) Magnesium-sensing riboswitches in bacteria. *RNA Biol.*, **7**, 77–83.
7. McConaughy, B.L., Laird, C.D., and McCarthy, B.J. (1969) Nucleic acid reassociation in formamide. *Biochemistry*, **8**, 3289–3295.
8. Klump, H. (1977) Thermodynamic values of the helix-coil transition of DNA in the presence of quaternary ammonium salt. *Biochim. Biophys. Acta*, **475**, 605–610.
9. Steger, G., Müller, H., and Riesner, D. (1980) Helix-coil transitions in double-stranded viral RNA: Fine resolution melting and ionic strength dependence. *Biochim. Biophys. Acta*, **606**, 274–284.
10. Michov, B.M. (1986) Specifying the equilibrium constants in Tris-borate buffers. *Electrophoresis*, **7**, 150–151.
11. Riesner, D., Steger, G., and Saenger, W. (1990) Viroids and viroid-like RNA, In *Nucleic Acids, Subvolume d, Physical Data II, Theoretical Investigations, Landolt-Börnstein, Group VII Biophysics*, Vol. I, Springer-Verlag, Berlin, pp. 194–243.

12. Record, M.T. and Lohmann, T.M. (1978) A semiempirical extension of polyelectrolyte theory to the treatment of oligonucleotides: Application to oligonucleotide helix-coil transitions. *Biopolymers*, **17**, 159–166.
13. Sadhu, C. and Gedamu, L. (1987) *In vitro* synthesis of double stranded RNA and measurement of thermal stability: effect of base composition, formamide and ionic strength. *Biochem. Int.*, **14**, 1015–1022.
14. Shelton, V.M., Sosnick, T.R., and Pan, T. (1999) Applicability of urea in the thermodynamic analysis of secondary and tertiary RNA folding. *Biochemistry*, **38**, 16831–16839.
15. Nussinov, R., Pieczenik, G., Griggs, J.R., and Kleitman, D.J. (1978) Algorithms for loop matchings. *SIAM J. Appl. Math.*, **35**, 68–82.
16. Waterman, M.S. and Smith, T.F. (1978) RNA secondary structure: A complete mathematical analysis. *Math. Biosci.*, **42**, 257–266.
17. Bellman, R. and Kalaba, R. (1960) On k th best policies. *SIAM J. Appl. Math.*, **8**, 582–588.
18. Steger, G., Hofmann, H., Förtsch, J., Gross, H.J., Randles, J.W., Sängler, H.L., and Riesner, D. (1984) Conformational transitions in viroids and virusoids: Comparison of results from energy minimization algorithm and from experimental data. *J. Biomol. Struct. Dyn.*, **2**, 543–571.
19. Zuker, M. (1989) On finding all sub-optimal foldings of an RNA molecule. *Science*, **244**, 48–52.
20. Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.
21. Waterman, M. and Byers, T. (1985) A dynamic programming algorithm to find all solutions in a neighborhood of the optimum. *Math. Biosci.*, **77**, 179–188.
22. Wuchty, S., Fontana, W., Hofacker, I.L., and Schuster, P. (1999) Complete sub-optimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**, 145–165.
23. Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, S., Tacker, M., and Schuster, P. (1994) Fast folding and comparison of RNA structures. *Monatsh. Chem.*, **125**, 167–188.
24. McCaskill, J.S.M. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
25. Markham, N.R. and Zuker, M. (2008) UNAFold: software for nucleic acid folding and hybridization. In *Bioinformatics: Structure, Function and Applications*, Methods in Molecular Biology, Vol. 453 (ed. J.M. Keith), Humana Press, Totowa, NJ, pp. 3–31.
26. Markham, N.R. and Zuker, M. (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.*, **33**, W577–W581.
27. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
28. Zuker, M., Mathews, D.H., and Turner, D.H. (1999) Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In *RNA Biochemistry and Biotechnology*, NATO ASI Series (eds J. Barciszewski and B.F.C. Clark), Kluwer, pp. 11–43.
29. Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.
30. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.*, **36**, W70–W74.
31. Weinberg, Z. and Breaker, R.R. (2011) R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics*, **12**, 3.
32. Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H., and Westhof, E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.
33. De Rijk, P., Wuyts, J., and De Wachter, R. (2003) RnaViz 2: an improved representation of RNA secondary structure. *Bioinformatics*, **19**, 299–300.

34. Zuker, M. and Jacobson, A.B. (1995) "Well-determined" regions in RNA secondary structure prediction: analysis of small subunit ribosomal RNA. *Nucleic Acids Res.*, **23**, 2791–2798.
35. Zuker, M. and Jacobson, A.B. (1998) Using reliability information to annotate RNA secondary structures. *RNA*, **4**, 669–679.
36. Brucoleri, R.E. and Heinrich, G. (1988) An improved algorithm for nucleic acid secondary structure display. *Comput. Appl. Biosci.*, **4**, 167–173.
37. Waugh, A., Gendron, P., Altman, R., Brown, J.W., Case, D., Gautheret, D., Harvey, S.C., Leontis, N., Westbrook, J., Westhof, E., Zuker, M., and Major, F. (2002) RNAML: a standard syntax for exchanging RNA information. *RNA*, **8**, 707–717.
38. Darty, K., Denise, A., and Ponty, Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.
39. Walter, A.E., Turner, D.H., Kim, J., Lyttle, M.H., Müller, P., Mathews, D.H., and Zuker, M. (1994) Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 9218–9222.
40. Varani, G. (1995) Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 379–404.
41. Janssen, S., Schudoma, C., Steger, G., and Giegerich, R. (2011) Lost in folding space? Comparing four variants of the thermodynamic model for RNA secondary structure prediction. *BMC Bioinformatics*, **12**, 429.
42. Gardner, P.P., Daub, J., Tate, J., Moore, B.L., Osuch, I.H., Griffiths-Jones, S., Finn, R.D., Nawrocki, E.P., Kolbe, D.L., Eddy, S.R., and Bateman, A. (2010) Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Res.*, **39**, D141–D145.
43. Zuker, M., Jaeger, J.A., and Turner, D.H. (1991) A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. *Nucleic Acids Res.*, **19**, 2707–2714.
44. Gulyaev, A.P., van, Batenburg, F.H., and Pleij, C.W. (1995) The computer simulation of RNA folding pathways using a genetic algorithm. *J. Mol. Biol.*, **250**, 37–51.
45. Schmitz, M. and Steger, G. (1996) Description of RNA folding by "simulated annealing". *J. Mol. Biol.*, **255**, 254–266.
46. Flamm, C., Fontana, W., Hofacker, I.L., and Schuster, P. (2000) RNA folding at elementary step resolution. *RNA*, **6**, 325–338.
47. Xayaphoummine, A., Bucher, T., and Isambert, H. (2005) Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.*, **33**, W605–W610.
48. Geis, M., Flamm, C., Wolfinger, M.T., Tanzer, A., Hofacker, I.L., Middendorf, M., Mandl, C., Stadler, P.F., and Thurner, C. (2008) Folding kinetics of large RNAs. *J. Mol. Biol.*, **379**, 160–173.
49. Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. (2008) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 97–102.
50. Sankoff, D. (1985) Simultaneous solution of the RNA folding, alignment and protosequence problems. *SIAM J. Appl. Math.*, **45**, 810–825.
51. Gardner, P.P. and Giegerich, R. (2004) A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics*, **5**, 140.
52. Bindewald, E. and Shapiro, B.A. (2006) RNA secondary structure prediction from sequence alignments using a network of k-nearest neighbor classifiers. *RNA*, **12**, 342–352.
53. Pedersen, J.S., Bejerano, G., Siepel, A., Rosenbloom, K., Lindblad-Toh, K., Lander, E.S., Kent, J., Miller, W., and Haussler, D. (2006) Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput. Biol.*, **2**, e33.
54. Wilm, A., Linnenbrink, K., and Steger, G. (2008) ConStruct: Improved construction of RNA consensus structures. *BMC Bioinformatics*, **9**, 219.

55. Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R., and Stadler, P.F. (2008) RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics*, **9**, 474.
56. Leontis, N.B., Stombaugh, J., and Westhof, E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
57. Stombaugh, J., Zirbel, C.L., Westhof, E., and Leontis, N.B. (2009) Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.*, **37**, 2294–2312.
58. Tabaska, J.E., Cary, R.B., Gabow, H.N., and Stormo, G.D. (1998) An RNA folding method capable of identifying pseudoknots and base triples. *Bioinformatics*, **14**, 691–699.
59. Griffiths-Jones, S. (2005) RALEE - RNA Alignment editor in Emacs. *Bioinformatics*, **21**, 257–259.
60. Jossinet, F. and Westhof, E. (2005) Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics*, **21**, 3320–3321.
61. Seibel, P.N., Müller, T., Dandekar, T., Schultz, J., and Wolf, M. (2006) 4SALE - a tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinformatics*, **7**, 498.
62. Höchsmann, M., Voss, B., and Giegerich, R. (2004) Pure Multiple RNA Secondary Structure Alignments: A Progressive Profile Approach. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **1**, 53–62.
63. Dalli, D., Wilm, A., Mainz, I., and Steger, G. (2006) STRAL: progressive alignment of non-coding RNA using base pairing probability vectors in quadratic time. *Bioinformatics*, **22**, 1593–1599.
64. Xu, X., Ji, Y., and Stormo, G.D. (2007) RNA Sampler: a new sampling based algorithm for common RNA secondary structure prediction and structural alignment. *Bioinformatics*, **23**, 1883–1891.
65. Moretti, S., Wilm, A., Higgins, D.G., Xenarios, I., and Notredame, C. (2008) R-Coffee: a web server for accurately aligning noncoding RNA sequences. *Nucleic Acids Res.*, **36**, W10–W13.
66. Perriquet, O., Touzet, H., and Dauchet, M. (2003) Finding the common structure shared by two homologous RNAs. *Bioinformatics*, **19**, 108–116.
67. Hofacker, I.L., Bernhart, S.H., and Stadler, P.F. (2004) Alignment of RNA base pairing probability matrices. *Bioinformatics*, **20**, 2222–2227.
68. Holmes, I. (2005) Accelerated probabilistic inference of RNA structure evolution. *BMC Bioinformatics*, **6**, 73.
69. Yao, Z., Weinberg, Z., and Ruzzo, W.L. (2005) CMfinder - a covariance model based RNA motif finding algorithm. *Bioinformatics*, **22**, 445–452.
70. Harmanci, A.O., Sharma, G., and Mathews, D.H. (2007) Efficient pairwise RNA structure prediction using probabilistic alignment constraints in Dynalign. *BMC Bioinformatics*, **8**, 130.
71. Torarinsson, E., Havgaard, J.H., and Gorodkin, J. (2007) Multiple structural alignment and clustering of RNA sequences. *Bioinformatics*, **23**, 926–932.
72. Will, S., Reiche, K., Hofacker, I.L., Stadler, P.F., and Backofen, R. (2007) Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput. Biol.*, **3**, e65.
73. Bauer, M., Klau, G.W., and Reinert, K. (2007) Accurate multiple sequence-structure alignment of RNA sequences using combinatorial optimization. *BMC Bioinformatics*, **8**, 271.
74. Lindgreen, S., Gardner, P.P., and Krogh, A. (2007) MASTR: multiple alignment and structure prediction of non-coding RNAs using simulated annealing. *Bioinformatics*, **23**, 3304–3311.
75. Kiryu, H., Tabei, Y., Kin, T., and Asai, K. (2007) Murlet: a practical multiple alignment tool for structural RNA sequences. *Bioinformatics*, **23**, 1588–1598.
76. Harmanci, A.O., Sharma, G., and Mathews, D.H. (2008) PARTS: probabilistic alignment for RNA joint secondary structure prediction. *Nucleic Acids Res.*, **36**, 2406–2417.

77. Katoh, K. and Toh, H. (2008) Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics*, **9**, 212.
78. Gorodkin, J., Stærfeldt, H.H., Lund, O., and Brunak, S. (1999) MatrixPlot: visualizing sequence constraints. *Bioinformatics*, **15**, 769–770.
79. Gorodkin, J., Heyer, L.J., Brunak, S., and Stormo, G.D. (1997) Displaying the information contents of structural RNA alignments: the structure logos. *Comput. Appl. Biosci./Bioinf.*, **13**, 583–586.
80. Rivas, E. and Eddy, S.R. (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.*, **285**, 2053–2068.
81. Dirks, R.M. and Pierce, N.A. (2003) A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J. Comput. Chem.*, **24**, 1664–1677.
82. Reeder, J., Steffen, P., and Giegerich, R. (2007) pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. *Nucleic Acids Res.*, **35**, W320–W324.
83. Sperschneider, J. and Datta, A. (2010) DotKnot: pseudoknot prediction using the probability dot plot under a refined energy model. *Nucleic Acids Res.*, **38**, e103.

27

RNA Secondary Structure Analysis Using Abstract Shapes

Robert Giegerich and Björn Vofsi

27.1

Introduction to Abstract Shape Analysis

27.1.1

Looking Deeper into the RNA Folding Space

According to the rules of base pairing, an RNA molecule can fold into a very large number of structures. RNA secondary structure prediction is often based on a thermodynamic model. Established tools such as *Mfold* or *RNAfold* can predict the structure with minimum free energy (MFE). This is also the most likely individual structure for the given molecule, but tells us little about all its competing folding alternatives. Yet often, there is good reason not to rely on one predicted structure, albeit of minimal free energy. Inaccuracy of energy parameters, varied physiological conditions or temperature, unconsidered base modifications, interaction with RNA binding proteins or with another RNA molecule, and many other circumstances may cause the functional structure to be different from the MFE structure. But still – in these cases, the functional structure is hidden among the near-optimal foldings of our molecule, and we seek means to analyze this folding space more completely.

Abstract shape analysis is a systematic way to extract comprehensive information about the folding possibilities of an RNA molecule. Shape abstraction is a mathematical version of what humans do when they communicate about RNA structure: We speak of a hairpin structure for miRNA precursors, a cloverleaf structure for tRNAs, two adjacent hairpins for oxyS RNA, and, for lack of a compact name, “the arrangement of helices P3–P9 typical of a group I intron.” The number and arrangement of helices can characterize a particular structure quite comprehensively, while abstracting from details such as loop sizes or length of stems. (If, however, a precise size or distance of structural features or a sequence motif is characteristic for a class of RNA, then this characteristic is lost and shape abstraction is not useful.)

Postponing formal definitions to Section 27.1.3, we see that the idea of shape abstraction partitions the folding space into different classes of structures, each class characterized by an abstract shape. This idea can, in principle, be added to

any structure prediction program, but currently, only the programs *RNAshapes* [1, 2] and *RNAlishapes* [3] compute with abstract shapes.

27.1.2

Overview of Functions of Abstract Shape Analysis

Computing with abstract shapes allows us to perform the following analyses, which are detailed below as Protocols 1–4.

- 1) Computing a small set of representative structures: We can ask for near-optimal structures $s_1, s_2, \dots, s_n$ of different shapes $p_1, p_2, \dots, p_n$, such that each s_i is the optimal structure of shape class p_i . This gives a representative, but compact overview of the folding space.
- 2) Computing shape probabilities: We can accumulate Boltzmann probabilities for all structures per shape to find the probability that a molecule will fold into a structure of that shape. Or in other words, in a population of many copies of the same RNA in a cell, it tells us the proportion of them which attains that shape. This gives a weight to the representative structures – for example, if the shapes of s_1 and s_2 together consume (say) 95% probability, we may rule out that there are further relevant structures. This cannot be said when computing energies alone.
- 3) Computing a consensus structure from several aligned sequences: This technique combines the power of shape abstraction with that of comparative structure prediction, given that the sequences share enough similarity to align reasonably well.
- 4) Computing a consensus shape from several unalignable sequences: When sequences are highly divergent, but we still suspect a conserved structure, consensus shapes can point to such a conserved structure.

27.1.3

Definition of Shape Abstraction

Let x be an RNA sequence and $F(x)$ its folding space, that is, the set of all feasible secondary structures of x . For structure $s \in F(x)$, $E(s)$ denotes the free energy of s , and $P(s)$ is the probability for the molecule to attain structure s , which can be computed from its free energy and that of all competing structures according to Boltzmann statistics.

27.1.3.1 Shapes

Figure 27.1 gives a graphical explanation of abstract shapes.

Any $s \in F(x)$ can be described by a “dot-bracket” string of the same length as x , such as “(((...)))(...(((...)))(...)))”, where dots designate unpaired bases and matching parentheses designate paired bases. In abstract shapes, we represent each stack of base pairs by a single pair of square brackets. The abstract shape of the above structure would be “[[] [] []]”. This completely abstracts from

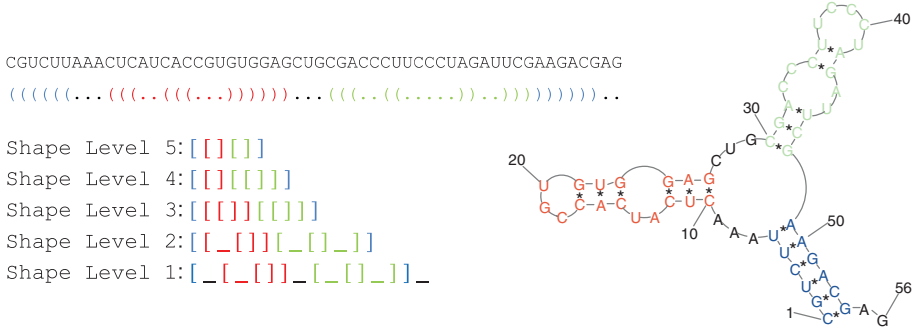


Figure 27.1 A secondary structure and its five shape representations, as implemented in *RNAshapes*. Graphics taken from Ref. [4]. Runs of base pairs are represented by a single pair of square brackets. Runs of unpaired bases are represented as a single underscore, or not at all, depending on the abstraction level.

the length of helices and the presence of unpaired bases. A special case is the completely unfolded structure “.....”, which is represented as shape “_”.

27.1.3.2 Shape Abstraction Function

Shape abstraction is a function π that maps a structure s to its shape $\pi(s)$. Below, we define several such shape abstraction functions.

27.1.3.3 Shape Representative Structures (shreps)

The shape representative structure $s_p^*(x)$ of shape p is the MFE structure of the given shape, that is, $s_p^* = \operatorname{argmin}_s \{E(s) | s \in F(x), \pi(s) = p\}$. Shape representative structures are often called “shreps.”

27.1.3.4 Levels of Abstraction

It depends on the application and the length of the sequences studied how much information detail we want to retain in the shapes. Therefore, we define shape abstraction functions for five levels of abstraction. Table 27.1 defines which information on unpaired bases is retained in the shape for each level, Figure 27.1 gives an example of the same structure seen at the five different levels of abstraction. In practice, the desired abstraction level is a program option to be selected by the user, where levels 5 and 3 are most frequently used.

If we care whether or not there are unpaired bases between two helices, we may use levels 1 or 2, which indicate the presence of (certain) unpaired bases by an underscore, yielding the (less abstract) shape “[[_]_][[_]][_]]” for the example structure introduced at the beginning of section 27.1.3.1. Referring to Figure 27.1 or Table 27.1, we see that it belongs to abstraction level 1, while the above shape “[[_]][_]]” belongs to abstraction level 5. (We never represent the hairpin loop by an underscore – it carries no information, as it *must* be there at the turn of a helix.) Representing the bulge in the leftmost helix has now led to the fact that the shape

Table 27.1 Runs of unpaired bases are represented by a single “_”, indicated only on levels 1 and 2.

Unpaired bases represented for	External loop	Multiloop	Bulge	Internal loop
Level 1	Yes	Yes	Yes	Yes
Level 2	No	No	Yes	Yes
Level 3	No	No	Implicit	Implicit
Level 4	No	No	No	Implicit
Level 5	No	No	No	No

Unpaired bases in the “external loop,” that is, not enclosed in any base pairs, are only indicated on level 1. On higher levels, internal loops (levels 3 and 4) and bulges (level 3 only) are indicated implicitly by splitting the enclosing helix. Note, for example, that a level 3 shape “ $[[[]]$ ” tells us that there is a stem interrupted by a single bulge or internal loop (but not which), while a level 5 shape “ $[[[]]]$ ” does not exist. Helix interruptions are ignored and the above level 3 shape becomes “ $[]$ ” on level 5.

tells us that this helix has two parts, interrupted by the bulge on the 3' side of the stem: "[]]".

27.1.3.5 Shape Probabilities

The shape probability of shape p is the accumulated probability of all structures of that shape: $P_p(x) = \sum \{P(s) | s \in F(x), \pi(s) = p\}$.

27.1.3.6 Consensus Shape

Given RNA sequences $x_1, \dots, x_k$, a shape p is a consensus shape if each x_i has a structure of shape p . Note that shape “_” is a consensus shape of any set of sequences, though not an interesting one. Interesting ones are consensus shapes of high probability $P_p(x_i)$ for $1 \leq i \leq k$, and, additionally, with similar shape representatives $s_p^*(x_i)$.

27.1.4

General Caveats when Working with Abstract Shapes

The nice thing about shape abstraction is that it introduces no heuristics (beyond the assumptions underlying thermodynamics-based folding anyway). Computed energies are exact and no part of the folding space is ignored. But there is an abstraction. This is not only the strength but also the main problem of the approach. Taking the shrep as a representative of all structures (of the same sequence) with that shape is normally a good idea, but there may still be considerable variation within the shape class. It generally holds that the longer the sequences are, the more structural variation within a shape you have to expect.

For example, for shape “[] [],” we may find structures such as the following:

	level 5	level 4
((((((((.....)))))))))((..))	[] []	[] []
((...)).((((((.....)).)).)))...	[] []	[] [[]]

where neither can be seen as a good representative of the other. Note that the two structures have different level-4 shapes, because of the internal loop in the second structure. Owing to the hierarchic definition of shape abstraction levels, two structures of the same level- i shape also have the same shape on abstraction level $i + 1$. This holds with the exception of going from level 3 to level 4. Conversely, if one wants to check how homogeneous the structures within a shape class at level i are, one can rerun the same analysis for shape level $i - 1$.

As we shall see in the applications below, shapes are also meaningful across sequences of different length and base composition. But still, sequence lengths should be in a similar range; otherwise, the observation of a common same shape has little meaning.

27.1.5

Applications of Abstract Shape Analysis

Before we describe, with some technical detail, four different ways to analyze RNA structure with abstract shapes, let us name a few scenarios in which these techniques have been applied.

In a large scale study, Berezikov *et al.* [5] investigated miRNA evolution in *Drosophilids*. They used hairpin shape probabilities of 90 and 60% to assign stringency levels to predicted miRNA precursors.

Bullerwell *et al.* [6] used consensus shapes and shape probabilities to characterize 5S rRNA-like transcripts encoded in the mitochondrial genomes of *Amoebozoa*. The consensus shape, termed *rabbit shape*, was found to be similar to the one known for conventional 5S rRNAs. Furthermore, the rabbit shape was, in all cases studied, the one achieving the highest probability.

Consensus structures for non-coding RNAs (ncRNAs) were predicted on a large scale with *RNALishapes* for *Synechocystis*, *Synechococcus*, and *Thermosynechococcus* [7]. The tool *RNAifier* [4] uses abstract shapes to speed up Rfam searches in efforts to classify RNA transcripts. It computes an index of top ranking shapes of the members of each Rfam family. It allows triggering the expensive search with Rfam models only when a family member has a high ranking shape in common with the query. This can save up to 90% computation time when a large number of ncRNA transcripts are to be checked against Rfam.

Some approaches use abstract shapes simply to communicate about structures. The tool *Locomotif* allows one to specify motif search programs by drawing structure graphics [8]. It informs the user about the overall shape of the motif under development by displaying a shape string.

Finally, let us take a look beyond abstract shape analysis as used at present. The idea of shape abstraction may be customized in manifold ways. Say, you are interested in RNAs that should exhibit a particular feature of interest in some low-energy (not necessarily MFE) structure. Say this feature is a compact hairpin of at least 10 base pairs, with a single bulged out C. It is not difficult to define a shape abstraction function that uses only two shapes: YES and NO. YES is the shape of all foldings that exhibit such a hairpin at any position in the overall structure, and

NO is the shape of all others. It only requires a bioinformatics student with some RNA background to replace our level 1–5 shape abstractions with your customized version and perform the analyses described in the sequel, specialized to your feature of interest.

27.2

Protocol 1: Computing Shape Representative Structures

We start from a single RNA sequence x of unknown structure. Folding it with any RNA folding program, using the standard thermodynamic model, we would obtain its MFE structure $\hat{s}(x)$. This, however, only tells us that the molecule *may* fold this way, but not whether there are competitive structural alternatives. Running *RNAshapes* with the mode “shape folding” (parameter $-a$) allows computing a small set of shape representative structures.

In our first example, we do so for two tRNA sequences and obtain the following result:

RNAshapes command line call: `RNAshapes -a -t 5 -c 20.0 -M 30 -# 1000`

```
>X54408.1/1-73
      GCCCCCAUCGUCUAGAGGCCUAGGACACCUCCUUCACGGAGAAAAC
-29.20 ((((((...((((((...))))))..((((.....)))))).....
-27.50 ((((((...((((((...((((((...))))))..((((.....)))))).....
-25.50 ((((((...((((((...))))))..((((.....))))))..((((.....))))))..

                                GCGGAUUCGAAUCCGCGUGGGGUA
                                (((((((.....)))))).. [][[]]
                                (((((((.....)))))).. [][[]]
                                (((((((.....)))))).. [][[]][[]]

>J05395.1/2325-2398
      GGUUUCGUGGUCUAGUCGGUUAUGGCAUCUGCUUAAACACGCAGAACGU
-23.10 (((((((((((((...(((...((((((...((((.....))))))..))))
-21.40 (((((((((((((...((((((...((((((...))))))..)))).....
-20.40 (((((((((((((...((((((...((((((...))))))..)))).....
-19.90 (((((((((((((...((((((...((((((...))))))..)))).....
-18.50 .((((((((((((((...((((((...((((((...))))))..))))..

                                CCCCAGUUCGAUCCUGGCGGAAAUCG
                                ..))..))..))..))..))..)).. [ ]
                                .((((((((((((((...((((((...((((((...))))))..)))).. [][[]][[]]
                                .....))..))..))..))..))..)).. [][[]]
                                .((((((((((((((...((((((...((((((...))))))..)))).. [][[]][[]][[]]
                                (((((((((((((((((...((((((...((((((...))))))..)))).. [ ] [ ]
```

In this call, we choose shape level 5 (parameter $-t$), an energy range of 20% (parameter $-c$) above the minimum, set a maximum loop length of 30 bases (parameter $-M$) as is commonly done in structure prediction and limit the potential output to a maximum of 1000 shapes (parameter $-#$). The latter is a general safeguard but not necessary with a short sequence similar to this one.

27.2.1

```
-e <value> Set energy range (kcalmol-1)
-c <value> Set energy range (%) [10]
-t <value> Specify shape type (1-5) [5]
-l        Allow lonely base pairs
```

Protocol 2: Probabilistic Shape Analysis

RNAshapes command line call: RNAshapes -q -t 5 -F 1.0E-6 -M 10 -# 1000

```
>X54408.1/1-73
GCCCCCAUCGUCUAGAGGCCUAGGACACCUCCCUUUCACGGAGAAAAAC
-29.20 ((((((...((((((...)))))).))..(((.....))))).)....
-27.50 ((((((...((((((...(((.....)))))).))....
-25.50 ((((((...((((((...))))).)....(((.....))))).)....
-23.30 ((((((...(((.....))))).)....(((.....)).(((.....))))).)....
-21.10 .....(((.....))))).(((.....(((.....))))).)....
-21.50 ((((((...(((.....))))).)....(((.....(((.....))))).)....
-21.46 ((((((...(((.....(((.....(((.....))))).)....

GCGGAUUCGAAUUCGCGUGGGGUA
(((.....))))).).... 0.9183351 [[[]]]
(((.....))))).).... 0.0795089 [[[]]]
(((.....))))).).... 0.0020713 [[[]]]
(((.....))))).).... 0.0000790 [[[]]]
(((.....))))).).... 0.0000027 [[[]]]
.(((.....))))).).... 0.0000018 [[[]]]
.))....(((.....))))).).... 0.0000011 []
```

[illegible]

In the output, we observe several interesting details:

First of all, this mode has no option `-c` to indicate an energy range. Probabilistic shape analysis accumulates probabilities across *all* shapes. Hence, more shapes are computed than we have seen shreps before.

Shapes ranked by probability mostly follow the rank order of their shreps (which are ordered by energy), but not always – for example, on the lower ranks of X54408.1/1–73, we see some “overtaking”: although the shrep energy of “[[] [[] []]]” on rank 6 is lower, there are more low-energy structures of shape “[[] [[] []]]” (rank 5) in the Boltzmann ensemble and their accumulated probability exceeds that of the other shape. In our example, this overtaking takes place only on the lower ranks, but for a different example, it might well occur between top ranking shapes.

Overall, the MFE shapes clearly dominate the Boltzmann ensemble in this example. Alternative shapes only account for 8% or 5% plus 4% probability. Taken together, this practically rules out that any other alternative structures play an important role.

Clearly, the probabilistic mode of abstract shape analysis gives more complete information than the previous protocol, but is much more expensive and therefore not used by default. Probabilities accumulate from many small contributions, and the dynamic programming algorithm must compute them for *all* the shapes in order to get them right. To speed up the program, shapes for (sub)structures with *very* small probability contributions can be ignored, without distorting the result. This is an optional heuristics, controlled by parameter F.

If your sequences are longer and you have many probabilistic shape analyses to compute, you may resort to a heuristics called *RapidShapes* [9], which normally gives good speedup when the sequence length exceeds 200 nt, without missing high-probability shapes. It is also available *via* the *RNASHapes* web site.

27.3.1

Useful Parameters

```
-t <value> Specify shape type (1-5) [5]
-F <value> Set probability cutoff filter (use with -p, -q, or -P)
-T <value> Set probability output filter (use with -p, -q, or -P)
-l Allow lonely base pairs
```

27.4

Protocol 3: Comparative Shape Analysis from Aligned Sequences

The so-called Plan A approach [10] to comparative structure prediction starts from a set of related RNA sequences. It works best when the sequences have 60–70% pairwise identity. This allows for a meaningful sequence alignment by standard methods. Then, the whole alignment is folded as a single metasequence to predict consensus structure. Of special interest are covarying positions, that is, pairs of alignment columns that vary in sequence but preserve the ability to form a base pair. Such occurrences are assigned a pseudo-energy contribution, which is added to the mean free energy of the structural element under consideration [11].

This route is also available with abstract shapes and has been implemented in the program *RNAlishapes*.

We want to exemplify the use of *RNAlishapes* and the interpretation of its results using sequences unrelated to the one presented in Sections 27.1 and 27.2. They are taken from Ref. [12] and aligned with ClustalW. The resulting alignment (*trp_attenuators.aln*) has the correct input format for *RNAlishapes*. Before executing the algorithm, we have to decide which shape type and which energy increment we want to use for the analysis. On the basis of our experience, we suggest to begin such an analysis with shape type 5 and an energy range of 3 kcal mol⁻¹. Thus, the command line call

```
RNAlishapes -t 5 -e 3 -s trp_attenuator.aln
```

will produce the following output:

```

      TGGTGGTGGCGCGCTTACCCGGCGGCCA-----CTGATTGCACGT-ACTC
-16.93 .....((((((((.....)))))))).
-15.6  .....(((((((((.(((.....)))))))).
-14.6  .....((((((((.....)))))))).
-14.17 .....(((.....)))))))).
      A-----AGGCTCG--CGCGTAGGCCGCGGAGCGGGCTTTTGTTCCT--
      .....((((((((.....)))))))).
      .....)))))))).).....
      .....((((((((.....)))))))).
      .....((((((((.....)))))))).
```

We observe that there are four shapes found within this energy range, and looking at the *shreps*, we see that the ones for shape 1, 3, and 4 are very similar. In other words, the *shreps* 3 and 4 each differ only by two base pairs from

shrep 1. As the inclusion of these base pairs in both cases increases the energy, we can reasonably neglect these two shapes from our further considerations. Note that the computation of shape 3 could have been prevented using the option `-u`, which prohibits substructures with positive energies, as is the case with the extra hairpin considered in shape “[] [] []” but not in shape “[] [] .”

On the basis of these arguments, we focus the interpretation of the results on shapes 1 and 2. Figure 27.2 summarizes the results and shows two conformations that are mutually exclusive, similar in score and also explain the attenuation mechanism. The stem-loop at the 3′ end of structure A (shape 1) is a terminator hairpin, leading to premature termination of transcription. It achieves a better score than structure B (shape 2), reflecting the fact that structure A corresponds to the native state of this conformational switch.

27.4.1

Useful Parameters for *RNAshapes*

```
-t <value>   Specify shape type (1-5) [5]
-e <value>   Set energy range (kcalmol-1)
-f <value>   Set fraction of aligned bases in two columns that must be
             able to actually pair
-c <value>   Set weight of covariance
-u           Ignore unstable (sub)structures
-i           Consensus sequence in IUPAC code
```

27.5

Protocol 4: Comparative Shape Analysis from Unaligned Sequences

Let us assume that we are given a set of sequences, where we hope for a conserved structure, but in contrast to the previous section, they are too diverged to be aligned meaningfully as sequences. Then, we can try a so-called Plan C approach¹⁾ [10], where we first fold sequences individually and then determine consensus shapes [13]. The rationale behind this approach is that if there is a consensus structure s for our sequences, then $\pi(s)$ is a consensus shape. Thus, a set of consensus shapes can be predicted efficiently, but deciding on the “true” one may still be difficult.

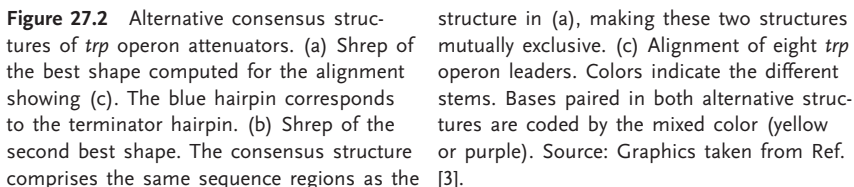
We use a set of 10 tRNA sequences that are quite difficult. Only half of them fold into a cloverleaf as their MFE shape. Our first run of *RNAshapes* in the consensus mode (parameter `-C`) with default parameters returns no result:

```
RNAshapes command line call: RNAshapes -C -t 5 -c 10.0 -F 0 -M 10 -# 1000
```

No consensus shapes found. Try to increase energy range with `-e` or `-c`.

As explained in the introduction, there *must* be consensus shapes. If none is found, this means that the energy interval chosen was too small with `-c 10.0`. For

1) Plan B according to the categories in Ref. [10] would be to simultaneously fold and align two sequences, which is considered the ideal, but most expensive approach.



any consensus shape, there is a sequence in our data set, where the shrep of the consensus shape has energy smaller than 90% of its MFE. Let us lower the cutoff to 20% of MFE and try again.

RNAshapes command line call: RNAshapes -C -t 5 -c 20.0 -F 0 -M 30 -# 1000

```

1) Shape: [[]] Score: -234.99 Ratio of MFE: 0.94
>X51423.1/1802-1874
    GGGCUUGUAGCUCAGCUGGUAGAGCGCCGCCUUUGCAAGGCGGAGGCCUGGG
-29.40 (((((((((((((((((((((((((((((((((((((((((((((((((((
>AF008220.1/6713-6785
    GCGGUUGUGGCGAAGUGGUUACGCACCAGAUUGUGGCUCUGGCAUUCUGUGG
-21.90 (((((((((((((((((((((((((((((((((((((((((((((((((((
>X54408.1/1-73
    GCCCCCAUCGUCUAGAGGCCUAGGACACCUCUUUACACGGAGAAAAACGCGGA
-27.50 (((((((((((((((((((((((((((((((((((((((((((((((((((
>J05395.1/2325-2398
    GGUUUCGUGGUCUAGUCGGUUAUGGCAUCUGCUUAAACACGAGAACGUCCCCA
-20.40 (((((((((((((((((((((((((((((((((((((((((((((((((((
>X12977.1/475-548
    CGGUGAUUAGCGCAGCCCGUAGCGCAUCUGGUUUUGGGACACAGGGUCAAAAG
-25.20 (((((((((((((((((((((((((((((((((((((((((((((((((((
>X03126.1/364-434
    AAGAGUUAUAGUUUAAAGGUAAAACAGAAAGCUUCAACCUUUAAUUCUUAAGU
-14.19 (((((((((((((((((((((((((((((((((((((((((((((((((((
>X52070.1/777-849
    UCCUCCGUAGCUCAAUUGGCAGAGCAGCCGGCUGUUAAACGGCAGGUUACUGG
-27.10 (((((((((((((((((((((((((((((((((((((((((((((((((((
>X13558.1/186-257
    UGGGGCGUAGCCAAGCGGUAAGGCAACGGGUUUUGGUCCCGCUAUUCGGAGGU
-24.80 (((((((((((((((((((((((((((((((((((((((((((((((((((
>X52392.1/3967-4040
    GCUAGCGUGGCAGAGCUCGGCAAAUGCAAAAGGCUUAAAGCCUUUAUCCAGAG
-21.80 (((((((((((((((((((((((((((((((((((((((((((((((((((
>K00287.1/1-73
    UCCUUGUUAAGCUCAGUUGGUAGAGCGUUCGGCUUUUAAACGAAAUUCAGGGG
-22.70 (((((((((((((((((((((((((((((((((((((((((((((((((((
    UCCGAAUCCAGCAAGUCCA
    .....)))).)))).) [ [] [] ] R = 3
    UUCGAUUCCTCAUCAAUCGCC
    .....)))).)))).) [ [] [] ] R = 3
    UUCGAAUUCGCGUGGGGUA
    .....)))).)))).) [ [] [] ] R = 2
    GUUCGAUCCUGGCGAAAUUCG
    .....)))).)))).) [ [] [] ] R = 3
    GUUCGAAUCCUUUAUACCGA
    .....)))).)))).) [ [] [] ] R = 2
    CGAGUCUAAAGUCUCUUG
    .....)))).)))).) [ [] [] ] R = 1
    UUCGAGUCCAGUCGGGGAG
    .....)))).)))).) [ [] [] ] R = 3
    UCGAAUCCUUCGUCCAG
    .....)))).)))).) [ [] [] ] R = 1
    GUUCAAUCCUCUCCCUAGCU
    .....)))).)))).) [ [] [] ] R = 1
    UUCGAGCCCCCUAUGAGGAG
    .....)))).)))).) [ [] [] ] R = 2

```

The suggested consensus shape is a Y-shape. Its score is the sum of shrep-energies, and the MFE ratio is the score divided by the sum of all MFE values. The value of R shown in the output indicates the rank of the consensus shape for each individual sequence. $R = 3$ for the first sequence, for example, means that this sequence has two other shapes with shreps of a better free energy than the consensus Y-shape.

Ranks lying between 1 and 3 seem quite good. However, looking at the concrete structures raises scepticism. They do not appear very similar, as sizes of the two inner helices vary a lot.

Is there any reason to look still deeper into the set of consensus shapes? In fact, if one of our sequences is sort of an outlier, it may fold rather poorly into a shape with which the other sequences tend to agree even better. Let us set the energy cutoff to 50%, leading to the following result:

```

1) Shape: [[]][[]] Score: -236.50 Ratio of MFE: 0.94
>X51423.1/1802-1874
    GGGCUUGUAGCUCAGCUGGUAGAGCGCCGCCUUUGCAAGGCGGAGGCCUUGGG
-32.60 ((((((((((((.....))))))...((((((((.....)))))))).((((
>AF008220.1/6713-6785
    GCGGUUGUGGCGAAGUGGUUAAACGCACCAAGUUGGUCUCUGGAUUCUGUGG
-21.80 ((((((((((.....(((.....))))).((((((((.....))))))....((((
>X54408.1/1-73
    GCCCCCAUCGUCUAGAGGCCUAGGACACCUCCCUUUCACGGAGAAAACGCGGA
-29.20 ((((((((((.....(((.....))))).((((((((.....))))))....((((
>J05395.1/2325-2398
    GGUUUCGUGGUCUAGUCGGUUAUGGCAUCUGCUUAAACGCAGAACGUCCCA
-21.40 ((((((((((.....(((.....))))).((((((((.....))))))....((((
>X12977.1/475-548
    CGGUGAUUAGCGCAGCCCGGUAGCGCAUCUGGUUUGGACCAGAGGGUCAAAG
-26.20 ((((((((((.....(((.....))))).((((((((.....))))))....((((
>X03126.1/364-434
    AAGAGUAUAGUUUAAAGGUAAAACAGAAAGCUUCAACCUUUAUUUCUAGUU
-8.20 ((((((((((.....(((.....))))).((((((((.....))))))....((((
>X52070.1/777-849
    UCCUCCGAGUCUCAAUUGGCAGAGCAGCCGGCUGUUAACCGGCAGGUUACUGG
-30.90 ((((((((((.....(((.....))))).((((((((.....))))))....((((
>X13558.1/186-257
    UGGGGCGUAGCCAAGCGGUAAGGCAACGGGUUUUGGUCCCGCUAUUCGAGGU
-24.30 ((((((((((.....(((.....))))).((((((((.....))))))....((((
>X52392.1/3967-4040
    GCUAGCGUGGCAGAGCUCGGCAAAUGCAAAAGGCUUAAGCCCUUUAUCCAGAG
-18.00 ((((((((((.....(((.....))))).((((((((.....))))))....((((
>K00287.1/1-73
    UCCUUGUUGAGCUCAGUUGGUAGAGCGUUCGGCUUUAACCGAAAUGUCAGGGG
-23.90 ((((((((((.....(((.....))))).((((((((.....))))))....((((

    UCCGAAUCCCGACAAGUCCA
    .....)))))))). [[]][[]] R = 1
    UUCGAUUCCAUCAAUCGCC
    ..)))).))))). [[]][[]] R = 4
    UUCGAAUCCGUGGGGUA
    .....)))))). [[]][[]] R = 1
    GUUCGAUCCUGGCGAAAUCG
    (.....)))))))). [[]][[]] R = 2

```

```

GUUCGAAUCCUUUAUCACCGA
(. . . . .)))). . . . .). [[[]]] R = 1
CGAGUCUAAGUGCUCUUG
. . . . .)))). . . . .). [[[]]] R = 3
UUCGAGUCCAGUCGGGGAG
. . . . .)))). . . . .). [[[]]] R = 1
UCGAAUCCUCCGUCCAG
. . . . .)))). . . . .). [[[]]] R = 2
GUUCAAAUCCUCUCCUAGCU
(. . . . .)))). . . . .). [[[]]] R = 2
UUCGAGCCCCUAUGAGGAG
(. . . . .)))). . . . .). [[[]]] R = 1

```

Now, we find the cloverleaf as a consensus, where the score has improved to $-236.50 \text{ kcal mol}^{-1}$ over the previous suggestion²⁾, which is now ranked second. At first glance, the helices in this consensus suggestion are in much better agreement with each other. A logical next step would be to align them with a structure alignment program, such as *RNAforester* [14, 15].

Currently, there is no master rule to decide on the most plausible consensus prediction, nor where to set the energy cutoff. Our example shows that by considerably relaxing the cutoff (from 0.2 to 0.5), we can find a higher scoring consensus. The safe way would of course be to set the cutoff to 100%, if your computer is powerful enough to compute *all* consensus shapes. In our given example, there would be only 13 shapes returned, the least-ranking one being the open shape “_”, with all sequences unfolded and with score 0.0, as expected. For longer sequences, however, the number of low-scoring consensi will be too large to compute.

27.5.1

Useful Parameters for *RNAshapes*

```

-e <value> Set energy range (kcalmol-1)
-c <value> Set energy range (%) [10]
-t <value> Specify shape type (1-5) [5]
-l        Allow lonely base pairs
-u        Ignore unstable structures (use with -a, -s, or -C)

```

27.6

RNAshapes Parameter Overview

Usage: *RNAshapes* [options] <input sequence> (or stdin)

Options:

```

-h          Display this information
-H <option> Display detailed information on <option>
-v          Show version

```

2) The ratio of MFE has improved from 0.938 to 0.944, but they both round to 0.94.

Sequence analysis modes:

- a Shape folding (standard mode)
- s Complete suboptimal folding
- p Shape probabilities
- q Shape probabilities (including shreps)
- P <value> Shape probabilities for mfe-best shapes
- i <value> Sampling with <value> iterations
- C Consensus shapes (RNACast)

Additional modes (use with any of the above):

- r Calculate structure probabilities
- w < value> Specify window size
- W <value> Specify window position increment (use with -w) [1]
- m <shape> Match shape (use with -a, -s, -p, -q, or -C)

Analysis control:

- e <value> Set energy range (kcalmol⁻¹)
- c <value> Set energy range (%) [10]
- t <value> Specify shape type (1-5) [5]
- y <value> Set minimal shape length
- F <value> Set probability cutoff filter (use with -p, -q, or -P)
- T <value> Set probability output filter (use with -p, -q, or -P)
- M <value> Set maximum loop length [30] (use -M n for unrestricted)
- l Allow lonely base pairs
- u Ignore unstable structures (use with -a, -s, or -C)

Input/Output:

- o <value> Specify output type (1-4,f) [2]
- O <string> Specify output format string
- S <value> Specify output width for structures
- # <value> Print only the first <value> results
- g <value> Generate structure graphs for first <value> structures
- L Highlight upper case characters in structure graphs
- N Do not include additional information in graph output file
- A Omit samples in output in sampling mode (-i)
- f <file> Read input from <file>
- B Show progress bar (use with -p, -q, or -P)
- z Enable colors (in interactive mode: disable colors)
- Z Enable colors for dot-bracket and shape strings
- D <string> Convert dot-bracket-string to shape (choose type with -t)
- U Start graphical user interface

27.7**RNAlshaped Parameter Overview**

```

RNAlshaped -t[t,0-5] -e[erange] -s[Infile] -o[Outfile] -p -S -u[no
    unstable] -f[fraction] -c[covariance weight] -i[iupac] -h
-t: shape_Type: 0=complete suboptimal folding, 1-5=different
    abstraction levels
-e: Energy range above MFE to analyze
-s: Alignment file in ClustalW format
-o: Redirect output to File
-p: Shape probabilistic mode (do not use option -u with this!)
-S: Boltzmann-weighted sampling of given no. of structures (do not use
    option -u with this!)
-u: ignore unstable (sub)structures in external loop and multiloops
-f: fraction of aligned bases in two columns that must be able to
    actually pair
-c: weight of covariance
-i: Consensus sequence in IUPAC code, default is most frequent base
-h: Print this message

```

References

1. Giegerich, R., Voss, B., and Rehmsmeier, M. (2004) Abstract shapes of RNA. *Nucleic Acids Res.*, **32**, 4843–4851.
2. Voss, B., Giegerich, R., and Rehmsmeier, M. (2006) Complete probabilistic analysis of RNA shapes. *BMC Biol.*, **4**, 5.
3. Voss, B. (2006) Structural analysis of aligned RNAs. *Nucleic Acids Res.*, **34**, 5471–5481.
4. Janssen, S., Reeder, J., and Giegerich, R. (2008) Shape based indexing for faster search of RNA family databases. *BMC Bioinformatics*, **9**, 131.
5. Berezikov, E. van, Tetering, G., Verheul, M., van Belt, J., van Laake, L., Vos, J., Verloop, R., van de Wetering, M., Guryev, V., Takada, S. *et al.* (2006) Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis *Genome Res.*, **16**, 1289–1298.
6. Bullerwell, C.E., Burger, G., Gott, J.M., Kourennaia, O., Schnare, M.N., and Gray, M.W. (2010) Abundant 5S rRNA-like transcripts encoded by the mitochondrial genome in amoebozoia. *Eukaryot. Cell*, **9**, 762–773.
7. Voß, B., Georg, J., Schön, V., Ude, S., and Hess, W.R. (2009) Biocomputational prediction of non-coding RNAs in model cyanobacteria *BMC Genomics*, **10**, 123.
8. Reeder, J., Reeder, J., and Giegerich, R. (2007) Locomotif: from graphical motif description to RNA motif search. *Bioinformatics*, **23**, i392–400.
9. Janssen, S. and Giegerich, R. (2010) Faster computation of exact RNA shape probabilities *Bioinformatics*, **26**, 632–639.
10. Gardner, P.P. and Giegerich, R. (2004) A comprehensive comparison of comparative RNA structure prediction approaches *BMC Bioinformatics*, **5**, 140.
11. Hofacker, I.L., Fekete, M., and Stadler, P.F. (2002) Secondary structure prediction for aligned RNA sequences *J. Mol. Biol.*, **319**, 1059–1066.
12. Seliverstov, A.V., Putzer, H., Gelfand, M.S., and Lyubetsky, V.A. (2005) Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria. *BMC Microbiol.*, **5**, 54.
13. Reeder, J. and Giegerich, R. (2005) Consensus shapes: an alternative to the Sankoff algorithm for RNA consensus structure prediction *Bioinformatics*, **21**, 3516–3523.
14. Höchsmann, M., Töller, T., Giegerich, R., and Kurtz, S. (2003) Local similarity in RNA secondary structures. *Proceedings of the IEEE Computational Society Bioinformatics Conference (CSB 2003)*, pp. 159–168.
15. Höchsmann, M., Voss, B., and Giegerich, R. (2004) Pure multiple RNA secondary structure alignments: a progressive profile approach. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, **1**, 53–62.

28

Screening Genome Sequences for known RNA Genes or Motifs

Daniel Gautheret

28.1

Introduction

In the mid 1980s, as genomic sequences in the Genbank database reached their first million nucleotides, a few intrepid biologists turned their attention toward screening DNA sequences for the presence of known RNA genes (mostly tRNA at this time) and RNA motifs such as ribozymes or specific stem–loop structures. They soon realized this task was harder than finding homologs of protein-coding genes, since non-coding RNAs were defined not only by their primary sequence but also by a combination of sequence and secondary structure.

Thirty years and hundreds of new RNA families later, this view still holds, albeit with wide variations between individual RNA molecules. For instance, some RNAs have sequence signatures that are stronger than those of most protein-coding genes. Figure 28.1a shows an alignment of the *Escherichia coli* tRNA^{Ala} and its homolog in the *Bacillus subtilis* genome. Although the two species are separated by more than two billions years of evolution, their tRNAs have retained more than 90% identity. Such homologs may be efficiently identified using a sequence alignment program such as Blast [1]. One may argue, however, that tRNAs are very special molecules: they sustain selective pressure at many sequence positions because of their multiple and universal functions. But experience has shown that a variety of RNA genes, including fast evolving ones such as microRNAs, are also well characterized at the sequence level and can be efficiently detected using sequence comparison programs [2]. Yet, for many other RNAs, primary sequence information is not sufficient for accurate detection. Certain RNA genes such as the one for RNase P RNA present little, if any, conserved sequence, and most RNA structural motifs such as GNRA- and E-loops, SECIS elements, U-turns, or Rho-independent transcription terminators are composed of a few key nucleotides associated with a larger secondary structure scaffold. Figure 28.1b shows consensus structures of the loop E [3] and U-turn [4] motifs, whose functions are conferred by a small set of conserved nucleotides embedded into a larger stem–loop structure. Such motifs cannot be identified using sequence comparison and require specialized programs that take into account both secondary structure and sequence constraints.

Score = 111 bits (122), Expect = 4e -25
 Identities = 70/76 (92%), Gaps = 0/76 (0%)
 Strand=Plus/Plus

```

Query 1      GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTTGCACGCAGGAGGTCTGCGGTTCGATC 60
          ||||| |||||||||||||||||||||||||||||||||||||||||||||
Sbjct 11550  GGGGCTTAGCTCAGCTGGGAGAGCGCCTGCTTTGCACGCAGGAGGTCTGCGGTTCGATC 11609

Query 61      CCGCATAGCTCCACCA 76
          ||||| |||||||
Sbjct 11610  CCGCTAGGCTCCACCA 11625
  
```

(a)

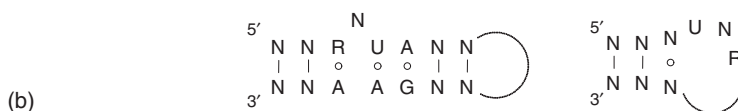


Figure 28.1 RNA motifs vary widely in their evolutionary constraints. (a) A strong primary sequence constraint is visible in this Blast alignment of tRNA^{Ala} from *E. coli* (top) and *B. subtilis* (bottom).

(b) Predominant secondary structure constraints are visible in the consensus loop E and U-turn motifs. Thin vertical lines indicate Watson–Crick base pairs, open circles indicate non-canonical base pairs.

Mining genomic sequences for RNA elements may thus entail using very different approaches depending on the type of element under scrutiny.

Freyhult *et al.* [5] published a detailed comparison of the performance of various RNA detection programs using simulated sequence sets, and Menzel *et al.* [2] proposed a more qualitative evaluation of three programs based on “real-life” sequence sets. Here, our purpose is not to evaluate programs but instead to help readers choose the most appropriate program for different situations and present practical execution examples for each of the programs. This article updates and expands our previous one in this series [6]. The proposed tasks are accessible to any beginner bioinformatician or to biologists with a basic knowledge of Unix commands. We provide links to official download sites for the various bioinformatics programs at the end of the chapter. We assume that all the required programs are properly installed by a system administrator on a computer running the Linux or another Unix operating system.

28.2

Choosing the Right Search Program

RNA search programs can be roughly divided in two classes: descriptor-based and homology-based programs. Descriptor-based programs allow biologists to describe, using a defined syntax, the sequence and structure of an RNA motif and search this motif in genome databases. The most widely used program in this family is RNAMOTIF [7]. Homology-based programs constitute a very diverse family. The most popular of them is Blast [1], which uses a single sequence as input and identifies segments of similarity in a sequence database. The HMMER program [8]

Table 28.1 Domains of application of different RNA search programs.

Program	Good for	Not good for	Pros/cons
Descriptor-based (RNAMOTIF)	Short motifs (terminators, hairpin loops, internal loops, etc.)	Complete RNA genes	Quick/requires learning descriptor language
BLAST	Long or short RNA genes, domains of genes	Short motifs, distant homologs	Quick and easy/boundaries not reliable
HMMER	Long or short RNA genes, domains of genes	Short motifs, distant homologs	Quick/a bit more complex than Blast
ERPIN	RNA motifs or genes <200 nt, domains of RNA genes, RNA with important pseudoknot	Long RNA genes, highly variable structure, irregular stems	Quick/somewhat tricky to use
INFERNAL	RNA motifs or genes <200 nt, distantly related RNAs	Long RNA genes, highly variable structure, pseudoknots	Most sensitive method/slow

uses a sequence alignment as input and, consequently, is able to exploit variation and conservation in the gene or motif. The last category of homology search programs, exemplified by ERPIN [9] and INFERNAL [10], use multiple sequences and a description of secondary structure to perform the search. Knowledge of the secondary structure allows these programs to identify parts of an RNA motif that are completely variable in sequence while retaining a constant structure. In Table 28.1, we present an overview of the qualities and domains of application of the different programs.

28.3

Overview of the RNA Search Procedure

A good RNA search job should address the question of secondary structure. Knowledge of an RNA secondary structure opens the way to the most sophisticated and sensitive search programs. It may also provide important clues about the function of an RNA. If a conserved structure is present in the RNA under study, then it must be identified. There are well-known programs for predicting an RNA secondary structure from its sequence. However, it is believed that on average about 30% of the base pairs predicted by these programs are incorrect. This false prediction rate is definitely too high if one wants to discover further RNAs based on this information. Much more reliable structure prediction can be achieved using a set of homologous RNAs and seeking a secondary structure that is compatible with all sequences in the set. A typical RNA search begins with a step of sequence collection to create this first training set. This can be done using

a purely sequence-based program such as Blast. Once this first set is completed, a common RNA structure is derived using a program that simultaneously folds and aligns the sequences in the set. Until recently, such algorithms were considered as too computationally expansive to be applicable to real-life situations and biologists were reduced to aligning RNAs using a sequence-based multiple aligner such as CLUSTALW [11], which often yields poorly aligned structures. However, recent progress allows reasonably accurate simultaneous alignment and folding. We chose the LOCARNA [12] program that presents the advantage of running on a web server.

Once a structured alignment is available, it needs to be carefully inspected. Are we sure all RNAs in the set are homologs? Are they complete? Are there obvious misalignments? Some manual curation may be helpful at this stage. Then, one should identify the most conserved sequence and structure elements. This consideration dictates which program can be used to seek further homologs. If no obvious conserved secondary structure stands out, we recommend pursuing with further rounds of BLASTN and the HMMER program. If short sequence/structure motifs appear, then one may use RNAMOTIF to seek further instances of these motifs. If a conserved structure including several stems and loops is visible, then one might resort to more sophisticated programs that will take these structures into account. ERPIN is useful when a core of regular stems is present. INFERNAL is preferred when structures are more variable. The latter is the most general and powerful of all programs used in this chapter, but it is also the slowest, although recent developments have brought its search speed closer to that of other programs.

As further RNAs of the family are discovered using any of the above tools, they can be used to enrich and improve the structural alignment so that further search iterations can be performed. The steps of a typical RNA search can be summarized as in Figure 28.2. All search programs accept sequence databases in the multi-Fasta format (with a pre-processing step for Blast). Recent versions have practically no limitation in the size or number of sequences in the input file.

28.4

Assessing Search Specificity

Most sequence and motif search programs compute a score and an expectation value (E-value) for each hit. An E-value tells us how many hits could be obtained by chance with the same score or higher in a random database of same size. An E-value above 1 or even 0.1 is likely to occur by chance. Any E-value below 10^{-2} should, in principle, convince peer scientists that a biologically related motif has been found. However, keep in mind that a low E-value does not demonstrate biological significance, as it is subject to artifacts caused, for instance, by low complexity regions in genomes. Conversely, hits with a poor E-value may turn out to be true homologs that have diverged a lot. Also, note that E-values depend on the database size. When a search program is run repeatedly on separate sequence files (e.g., with one file per chromosome), the E-value is valid only for each distinct file.

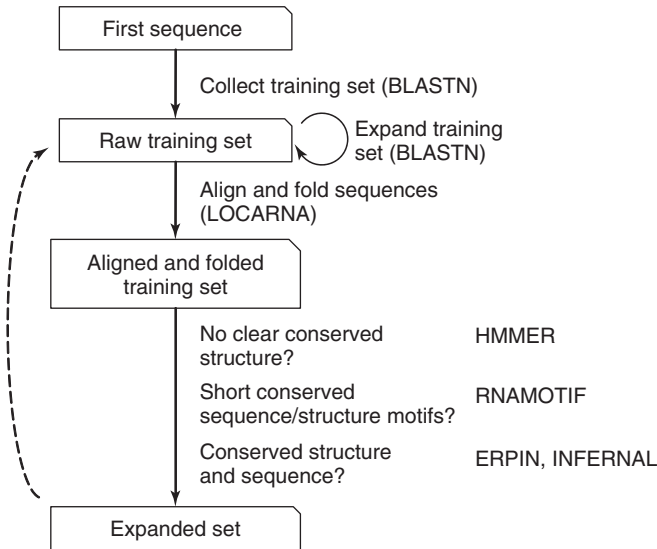


Figure 28.2 Outline of a typical RNA motif search protocol. The protocol starts with a BLASTN sequence collection followed by a structural alignment (align and fold) and the use of a specialized search software.

RNAMOTIF is the only program in our protocol that does not automatically compute E-values. In this case, or when E-values are not trusted for some reason, it is common to use an expected ratio of false positives (FP) as a measure of specificity (e.g., FP per Mb). To this end, we may run the search against a randomized sequence database. Random sequences with a uniform nucleotide composition (25% each) are not advised, since background compositions (single-nucleotide and dinucleotide frequencies) strongly affect the number of chance hits: for instance, an AU-rich motif is more likely to occur in an AU-rich sequence background. It is therefore important to reproduce in the negative control set the overall compositional biases of the target database. Several programs can do this. We will use *shuffle*, by S. Eddy (Section 28.8), that randomizes a sequence while preserving the single- and dinucleotide frequencies.

The randomization procedure can be performed on the target database, so that each genome in the database is shuffled using its own compositional characteristics. When the search database is not decided beforehand, an alternative option is to shuffle the training set itself. This is a very conservative approach, as a randomized training set, owing to its limited sequence space, is more likely to give rise to false positives than any other sequence.

If specificity has to be quantified, the negative set should, ideally, be large enough to find at least a few false positives. By default, the *shuffle* program produces a sequence of the same size as the input sequence. When using a short sequence

such as the training set, the shuffling procedure must be iterated to produce a random sequence of the desired size. For instance,

```
shuffle -d -n 100 trset.fasta > trset.rnd
```

Here, the sequence to be randomized is in file `trset.fasta` (multi-Fasta format) and the output is redirected to `trset.rnd`. The `-d` option is used to preserve dinucleotide composition in the random sequence, and `-n 100` specifies 100 shuffling iterations. The randomized sequences from each iteration round are concatenated and the random dataset will thus be 100 times larger than the original sequence.

28.5

A Test Case: Looking for Homologs of a Bacterial sRNA

Our test challenge here is to discover homologs of a bacterial small RNA (sRNA) gene named *RsaE*, a 113 bp RNA first identified in *Staphylococcus aureus* [13, 14]. We will try to stay as close as possible to a “real-life” situation, which means we assume that only one instance of the RNA molecule is known first (here, the *S. aureus* *RsaE*), and no structure information is available. This is the situation shown in the first box in Figure 28.2. The initial, experimentally confirmed, sequence is entered in the Fasta format in a file named `rsae.fa` (Section 28.7), as follows:

```
>Rsae
AUGAAAUUAACACUAACAACAUACCCUUUGUUUGAAGUGAAAAUUUCUCCCAUC
CCCUUUGUUUAGCGUCGUGUAUUCAGACACGACGUUUUUUUAU
```

Our objective is to find all other instances of this sequence in bacterial genomes. Our search database will contain all complete genomes retrieved from the NCBI web site. An initial set of genomes from 1080 species was downloaded in 2010 and reduced to 732 species by retaining a single instance of each species. The sequences are saved in a single multi-Fasta file of about 2.5 GB named `732bacs.fa` (Section 28.7). To simplify command lines, we assume that all data files are present in the working directory.

28.5.1

Building a First Training Set with BLASTN

BLAST is the program of choice for fast and easy identification of a first set of homologs. Biologists are familiar with the NCBI BLAST web interface. Here, we use the command line version of the program. An advantage of using the command line is that users can scan “private” sequences not yet included in the NCBI sequence databases. Furthermore, playing with parameters is faster with command lines. We assume a recent version of BLAST (such as 2.2.25) is installed on the reader’s workstation.

A local BLAST run requires a properly formatted database. Formatting our database file `732bacs.fa` is done using:

```
makeblastdb -in 732bacs.fa -dbtype 'nucl'
```

This command generates several index files required for BLAST runs. You should not worry about the names of these files, as you will always refer to the Fasta file name (here, `732bacs.fa`) for further operations. There are several programs in BLAST. Here, we use BLASTN, which is used to scan nucleic acid databases for nucleic acid sequences. With our data files, a simple BLASTN command would look like:

```
blastn -db 732bacs.fa -query rsae-sau.fa
```

The command takes about 0.7 s to execute and produces a lot of output. It is pretty difficult to understand how many hits were exactly recorded. A simpler table output can be obtained using the `-outfmt 7` parameter, such as in

```
blastn -db 732bacs.fa -query rsae.fa -outfmt 7
```

Now, we can count 12 hits from 12 different species. However, are they all statistically significant? The lowest scoring hit is from the *Geobacillus* strain WCH70. Its E-value of 3×10^{-11} is highly significant. We can control the E-value cutoff using the `-evalue` parameter, for instance, by appending `-evalue 0.01` to the BLAST command.

Can we find more distant homologs using BLASTN? This question is commonly addressed by reducing word size. In BLAST, both the database and the query sequences are broken down into k-length words (k-words). BLAST is able to quickly scan these word lists for exact or near-exact matches. Longer alignments with possible gaps and mismatches are generated only after one k-word match is found. This enables a very fast search at the risk of missing some hits that would contain no exact or near-exact k-word match. The default word size for BLASTN is 25. It can be changed using the `word_size` parameter. Shorter words will increase sensitivity, at the cost of longer runtimes. Let us reduce word size to 10:

```
blastn -db 732bacs.fa -query rsae.fa -outfmt 7 -evalue 0.01 -  
word_size 10
```

The command now takes 2.1 s to run, but finds 21 hits from different species, that is, 9 more than with the default word size. Note how significant these new alignments are. The lowest scoring hit is now found in *Geobacillus kaustophilus*, and has an E-value of 9×10^{-10} , which is still very significant. Reducing word size to its minimum value ($k = 5$) extends runtime to nearly 1 min but does not reveal further solutions.

This list of 21 BLAST hits is a good starting point for further analysis. We should now extract the corresponding sequences. The latest version of BLAST

provides useful output controls. It is now possible to retrieve automatically the matched sequences and other items using parameter `-outfmt 10` followed by the desired item names. For instance replacing `-outfmt 7` in the above command with `-outfmt "10 eval evalue sseqid sseq"` will output the E-value, database sequence Id, and matching sequence region in the database for each hit.

A visual inspection of these sequences reveals that several species share the exact RsaE sequence. After discarding the redundant hits, our raw training set is composed of 14 distinct sequences. Before going any further, we should pay attention to the species distribution of these sequences. We have found RsaE in *Staphylococcus*, *Macroccoccus*, *Bacillus*, *Geobacillus*, *Anoxybacillus*, and *Lysinibacillus*. These genera all belong to the “Bacilli” class of the Firmicute phylum. This is an important observation that supports the relevance of the hits. Hits distributed in multiple distant phyla would be a warning sign of possible false positives.

28.5.2

Alignment and Structure Prediction

As explained above, establishing the RNA secondary structure is an important step for further homology search. To this aim, we will use the LOCARNA program [12] through its Freiburg web server. LOCARNA belongs to a family of programs that simultaneously align and fold a set of RNA sequences. These programs are known to provide better RNA alignments than those relying only on sequence. The LOCARNA server is intuitive and works fine with default parameters. We can just paste the 14 Fasta-formatted sequences into the input window. To produce a Fasta file from the BLAST output, any text editor will do. As LOCARNA does not accept gap characters, we will use the editor to remove gaps too. The Fasta-formatted raw training set is available as Suppl. file `trset.fa` (Section 28.7). The LOCARNA server displays the structural alignment after a few minutes. The output is shown in Figure 28.3a with parentheses representing the secondary structure. This predicted structure is remarkably similar to the one established for *S. aureus* using chemical and enzymatic probing [13] (Figure 28.3b).

LOCARNA, like any other multiple alignment program, does not produce perfect alignments. For instance, the AAAC in the 5' region of *G. kaustophilus* RsaE would probably be aligned better with the AAAC segments found in all other sequences. To keep things simple, we will not try to improve this alignment manually, and we will proceed with the one in Figure 28.3a.

It appears from our alignment that RsaE homologs differ in their 3' regions. Either some RNAs are actually shorter, or this may be an effect of the BLAST program failing to extend the alignment to its actual ending. To settle this, it is necessary to retrieve the corresponding regions from the bacterial genomes and observe their sequence. This can be done using the coordinates from the BLAST output and any genome browser such as the Ensembl Bacterial Genome Browser (<http://bacteria.ensembl.org/>). We retrieved the missing 3' regions and realigned the sequences using LOCARNA. These 3' regions differed strongly from other RsaE sequences and aligned poorly both at the sequence and structure level (not

(a)



Figure 28.3 (a) An aligned and folded training set of RsaE sequences (Stockholm format) obtained from the initial BLAST screen and aligned using LOCARNA. Parentheses in the bottom line indicate base-paired positions, highlighted by shaded boxes. Bases in italics in “*Geobacillus_kaus*” were absent in the initial BLAST hit and were manually inserted before alignment. (b) The *S. aureus* (“*Staphylococcus_a*”) RsaE secondary structure derived from enzymatic and chemical probing [13].

shown). Therefore, we can assume that these 3' regions are either missing or too different from the original RsaE sequence to be kept in our alignment. We also noted that the *G. kaustophilus* sequence lacked about 15 nt in its 5' region. We retrieved the corresponding segment from the genome sequence and used it for a new LOCARNA alignment. The extended sequence aligned nicely with other RsaE sequences although parts of the long 5' stem appeared not to be conserved. The extra segment (in *italics*) was incorporated into Figure 28.3a.

The RsaE initial alignment presents a long stem on the 5' side. However, the sequence of this stem is highly conserved and there is little evidence of covariation that would indicate the base pairs are under evolutionary pressure. Therefore, we cannot be absolutely sure about this secondary structure. The 3' part, which is missing in half of the sequences, is a typical Rho-independent transcription terminator, with a stem-loop followed by a U-stretch. We may hypothesize that the RsaE sequences lacking this terminator use another termination mechanism.

28.5.3

Searching with HMMER

A sequence alignment is a powerful means of conducting homology searches, even in the absence of secondary structure information. The main reason is that an alignment permits distinguishing variable regions that are not crucial for RNA function from conserved, crucial regions, and thus enables weighting matches accordingly. A good way to exploit an alignment is to use the HMMER package [8]. HMMER converts the alignment into a profile Markov model that records both the statistics of base frequencies in each alignment column as well as that of consecutive bases in neighboring columns. Once a profile HMM model is built from the alignments, matches to this model in sequence databases can be sought using the dynamic programming algorithm in HMMER. First, the training alignment should be formatted in the Stockholm format, as shown in suppl. file `trset.stk`. Here is the command to build the profile HMM:

```
hmmbuild trset.hmm trset.stk
```

This creates a file named `trset.hmm`, which can be used for scanning genome sequences. The search command is then entered as follows:

```
hmmsearch trset.hmm 732bacs.fa
```

After a runtime of 42 s, we obtain 18 solutions. These include new *Bacillus* and *Geobacillus* species, as well as an entire new genus, *Oceanobacillus*, that was absent in the initial training set. This HMMER hit is shown in Figure 28.4. The figure legend explains the different output items. The default mode of `hmmsearch` uses several heuristics to speed up the search, at the cost of reducing search sensitivity. It is possible to switch off the heuristics using the `--max` option. This slows down the search by a factor of 10, but reveals no novel significant hit in our case.


```
>> Oceanobacillus_iheyensis_BA000028
#      score bias  c-Evalue  i-Evalue hmmfrom  hmm to    alifrom  ali to    envfrom  env to    acc
---
1 !    22.5   0.1   3.2e-07   1.4e-05    15     67 .. 1267728 1267782 .. 1267716 1267784 .. 0.80

Alignments for each domain:
== domain 1    score: 22.5 bits;  conditional E-value: 3.2e-07
               ...((((((-.....))))))..-))..)))))..... CS
               trset1    15 aaacaaaca.auaccccu.uuguuugaacgugaaaaauucucccaucccc.uuug 67
               a acaaaaca auaccccu uuguuuga c ug aaa uuucucccaucccc uuug
Oceanobacillus_iheyensis_BA000028 1267728 AUACAAACaaAUACCCCUuUUGUUUGAUG-AAAACUUUCUCCCAUCCCCcUUUG 1267782
               567888888889999996268*****.5678*****987515555 PP
```

Figure 28.4 Part of an HMMER output. The local hit is shown with five lines. The “CS” line is just a reminder of the secondary structure in the .stk file, which is not used by HMMER; the next three lines show the consensus profile, matching residues, and target sequence; and the last line shows the posterior probabilities of each aligned residue, which is an indication of alignment reliability. Values range between 0 and *, with 0 corresponding to a 0–5% probability, up to * for a 95–100% probability that this residue complies with the hidden Markov model created from the input alignment.

28.5.4

Searching with RNAMOTIF

The RNAMOTIF program permits the description of an RNA motif, including sequence and secondary structure constraints, and enables one to find all instances of this motif in a sequence database. Therefore, a major difference from the previous approaches is that we can now use secondary structure information. A typical RNAMOTIF descriptor is shown in Figure 28.5a. Let us take a quick look at this language. The first section (*parms*), says that G:U pairs are allowed in addition to Watson-Crick pairs. The next section (*descr*) is the actual descriptor, listing all secondary structure elements. There are three types of elements: *h5* stands for the 5' strand of a helix, *h3* for the 3' strand of a helix, and *ss* for a single strand. When no other indication is provided, RNAMOTIF assumes pairing between *h3* and *h5* elements in a purely nested fashion, just as in a parenthesis notation. Non-nested helices (i.e., pseudoknots) require a special tag that modifies this default behavior. Associated with each *h5*, *h3*, or *ss* element is a list of optional parameters that specify minimum and maximum length, conserved sequences, conserved base pairs, and so on. By default, RNAMOTIF reports all sequences matching the descriptor. Sometimes, however, additional filters that cannot be applied at the “*descr*” level are required. An optional score section is used for this purpose. In this section, tests can be performed using a C-like computer language, and matches are rejected if the tests fail. For an in-depth explanation of RNAMOTIF descriptors and scoring functions, refer to the detailed manual included in the program distribution.

Let us see how our knowledge of the RsaE RNA can be turned into an efficient descriptor. We will use only the most basic features of the RNAMOTIF language. To build our first descriptor (Figure 28.5a), we make the typical assumption that a regulatory RNA has variable sequences in helices and conserved sequences in single-stranded regions, as the latter are supposed to be involved in tertiary interactions. This is of course a simplistic view, but we would like to see how it performs. We set helix lengths as in the Figure 28.3a consensus (size = 6, 4, and 6 nt, respectively from 5' to 3'). To account for shorter helices, we allow one mismatch per helix (*mismatch=1*) and, since the default behavior of this parameter restricts mismatches to internal positions, we use *ends='mm'* to allow for mismatches at helix ends. We specify conserved sequences in several single-stranded segments. The most conserved single-stranded region is the apical loop where the conserved seven 3'-proximal bases are specified by *seq="ACCCCUU\$"*. In the fifth “*ss*” region, *seq="^A"* means that this strand should start with an A. We save this descriptor file as *descriptor.txt* and we are now ready to look for instances of this motif in our database. The basic RNAMOTIF command then looks like

```
rnamotif -descr descriptor.txt 732bacs.fa
```

The run takes about 1 min and produces a pretty large output. It is better to use a redirection to store hits in an output file. Also, owing to variations in helix

<pre> parms wc+=gu; descr ss(len=1, seq="U") h5(len=6, mispair=1, ends='mm') ss(len=2) h5(len=4, mispair=1, ends='mm') ss(minlen=2, maxlen=4, seq="^A") h5(len=6, mispair=1, ends='mm') ss(minlen=7, maxlen=14, seq="ACCCCUU\$") h3 ss(minlen=1, maxlen=3, seq="^A") h3 ss(minlen=1, maxlen=3, seq="^AA") h3 </pre>	(a)
<pre> descr ss(len=60, seq="^.....AAAC.\{2,10\}ACCCCUUUGUUUGA") </pre>	(b)
<pre> descr ss(len=20, seq="AAACS") ss(minlen=2, maxlen=10) ss(len=14, seq="ACCCCUUUGUUUGA", mismatch=1) </pre>	(c)

Figure 28.5 RNAMOTIF descriptors for RsaE. (a) This descriptor defines three stems with variable sequences and six single strands containing conserved sequences. (b) This descriptor defines a single 60 nt single strand with a long conserved 3' motif. (c) This descriptor splits the previous single strand in three parts. In the last part, a mismatch is allowed in the consensus motif.

length and position, RNAMOTIF tends to generate several overlapping solutions for each actual motif. The utility program `rmprune` is provided to filter out these overlapping solutions and to retain only the highest scoring one at each site. This program is executed in combination with RNAMOTIF using the “pipe” sign (`|`), as follows:

```
rnamotif -descr descriptor.txt 732bacs.fa |rmprune > outfile
```

From now on, we will assume that every RNAMOTIF run is performed using `|rmprune > outfile`. We can now count 16 solutions produced by the RNAMOTIF run. Sample outputs are shown in Figure 28.6. Most hits are from the genera *Bacillus*, *Macrococcus*, *Staphylococcus*, and *Anoxybacillus*. A few training set sequences such as the Geobacilli are missing, which indicates that our descriptor is too strict in some respects. Indeed, we can see in the alignment that shorter helices #2 and #3 in *G. kaustophilus* do not match the descriptor constraints.

```
>Anaerococcus_prevotii_DSM_20548_CP001708
Anaerococcus_prevotii_DSM_20548_CP001708    0.000 0 212356 74
t ttcttc aa ccaa at ccaaaa caactacccctt ctttgg agt ttg aa gagaaa
```

(a)

```
>Streptococcus_uberis_0140J_AM946015
Streptococcus_uberis_0140J_AM946015    0.000 0 430954 72
tgtactgatggtatcaaac cagactta acccttttgttga tgtaccttcagaaattggcatagcgcgcac

>Clostridium_phytofermentans_ISDg_CP000885
Clostridium_phytofermentans_ISDg_CP000885    0.000 0 2755278 69
gcacctggtttttacaaaac tacct accccttttttga atatattacaccctctaatctcaattgcgc
```

(b)

Figure 28.6 Extracts of RNAMOTIF outputs. Hits are formatted as follows. Line 1: title line of Fasta file entry. Line 2: Sequence name, score (here, our descriptors do not include a scoring section, hence all scores are 0), strand ("0" indicates plus strand and "1" minus strand), hit position, hit length

in nucleotides. Line 3: hit sequence, with stems and single-stranded elements separated by spaces. (a) A hit from the descriptor in Figure 28.5a. (b) The two hits were found by using the descriptor of Fig. 28.5c plus defining an additional "ss" element to include the 30 nt of downstream sequence.

Interestingly, only four sequences are from species outside our training set. Considering the large size of the database, this is a small number. These include one hit from *Anaerococcus prevotii* (Figure 28.6a). Albeit being distant in the phylogenetic tree, this species belongs to the Firmicutes, as do the other known RsaE sequences, and is therefore an interesting candidate to consider. However, we observed that the sequence just following the motif is not as CU-rich as in other RsaE sequences. The other new hits (not shown) are from phyla that are very distant from Firmicutes and therefore unlikely to contain an RsaE homolog. We can reasonably assume they are false positives.

To assess the significance of new RNAMOTIF hits, we advise using a shuffled database as a control. Here, we will randomize our bacterial genome database while preserving the dinucleotide frequencies, using the `shuffle` command:

```
shuffle -d 732bacs.fa > 732bacs.rnd
```

Running the same RNAMOTIF descriptor on the shuffled database `732bacs.rnd` produces six hits. This is a clear warning that the few hits obtained with real data can be false positives.

Examination of the RsaE alignment suggests that sequence alone may describe RsaE better than secondary structure. The stems are indeed highly conserved in sequence and show no sequence covariation to support their base-pairing potential. Let us see how efficient our RsaE search could be by using solely primary sequence constraints. All RsaE members harbor a constant AAAC motif followed by a highly conserved CU-rich sequence that is repeated in tandem. The descriptor in Figure 28.5b describes one copy of the tandem motif in a single line. The syntax for sequence constraints in RNAMOTIF is that of Unix regular expressions. The "^"

sign indicates the beginning of the element and the string `.\{2,10\}` describes a sequence of 2–10 unspecified bases, described by the dot character. This simple motif performs surprisingly well. After 75 s, RNAMOTIF identifies 27 hits, covering all the species in the training set plus several other species from the same genera and 1 hit in *Pectobacterium wasabiae*, an enterobacterium. Enterobacteria are very distant from Firmicutes, which makes this hit an unlikely RsaE candidate.

Running the descriptor of Figure 28.5b against the shuffled genome database produces a single hit, which shows that our simple sequence motif is more specific than the complex secondary structure motif of Figure 28.5a. Can we now relax this descriptor and find additional candidates? RNAMOTIF allows for mismatches in sequence motifs, using the `mismatch=n` parameter (n being the number of allowed mismatches). However, the mismatch parameter is compatible only with regular expressions of constant length. Therefore, we cannot allow for mismatches in a variable region of size 2–10 as above. We thus have to rewrite our descriptor using distinct `ss` elements, as shown in Figure 28.5c.

The descriptor in Figure 28.5c that tolerates one mismatch in the CU-rich region takes longer to run (2 min) than the exact descriptor and produces 71 solutions. In addition to new species from the same genera as above, these include a number of apparent false positives from Cyanobacteria or Enterobacteria, plus new Firmicutes candidates of the genera *Streptococcus* and *Clostridium* (Figure 28.6b). However, these results should be considered with caution, as the false-positive ratio has increased significantly. We now find 35 hits in the randomized database (not shown).

A possible control to evaluate the new candidates is to analyze the sequence downstream of the motif. We can see in the training set (Figure 28.3a) that the downstream sequence is a near-exact repeat of the central conserved region. Flanking sequences around a given motif are easy to extract using RNAMOTIF by adding new “`ss`” elements of desired length with no sequence constraint. We used this to obtain the 30 nt of downstream sequences. This sequence is not UC-rich in Streptococci, but is somewhat UC-enriched in one of the Clostridia, *C. phytofermentans* (Figure 28.6b). Therefore, this candidate may be considered further.

We can see that RNAMOTIF is a valuable tool to test ideas about what determines an RNA motif. We have developed highly specific descriptors in just minutes, using either primary or secondary structure information. One important limitation of RNAMOTIF, however, is its lack of an objective scoring function. Although the program accepts user-defined scoring functions, their quality will depend on the user’s appreciation of the RNA motif. The subsequent protocols address this issue.

28.5.5

Searching with ERPIN

ERPIN circumvents the descriptor design problem by extracting information directly from a structure-annotated alignment and building log-odds score profiles for each stem and single-stranded segment in the alignment.

First, we need to convert the alignment into the ERPIN format. Figure 28.7a shows how the first lines of our RsaE alignment should look like. The first three

[illegible]

(a)

[illegible]

(b)

Figure 28.7 ERPIN data formats. (a) typical ERPIN input file (.epn file) showing secondary structure (first three lines) and aligned sequences (next lines). The format is explained in the text. (b) A .epn file displayed using the `view` command. Each secondary structure element, S (single-stranded) or H (helical), is highlighted with element number (red vertically; e.g., $\frac{1}{18}$) and sequence conservation (asterisk).

lines provide secondary structure information. In this example, the secondary structure is the same as the parenthesized structure in Figure 28.3a, but each structural element (single-stranded or helical) is assigned a different number that is read vertically. When distant segments have the same number (e.g., 02 or 04 in this example), the corresponding element is understood to be helical; otherwise, it is single-stranded. The secondary structure lines can be created manually using a text editor, or automatically from a parenthesized secondary structure file using the `parent2epn.pl` script provided in the ERPIN distribution. The lines following the structure information contain the aligned sequences in the classic Fasta format.

The initial parenthesized secondary structure had a single long single-stranded segment composed of elements 17 and 18. Element 18 (the terminator) is not present in all sequences, whereas element 17 is present everywhere and is highly conserved. To enable searching for the conserved part while ignoring the other, it is preferable to split this single-stranded segment into two different elements. We thus created element 18 using a text editor. Our ERPIN-formatted alignment was saved as `trset.epn` (Section 28.7).

The `tview` command is useful to visualize constraints in the sequence alignment:

```
tview trset.epn
```

The result is shown in Figure 28.7b. ERPIN users must decide which part of the alignment is used for the search. A direct search for a multiple-element region could be prohibitive in terms of CPU and memory usage, especially when elements vary a lot in size. Here, the alignment suggests that the region from the 5' side of helix 02 to single strand 17 is the most conserved (asterisks on top). We will thus launch the search using this region (we assume all ERPIN commands below are complemented by `-silent > outfile` in order to send output to file `outfile`, devoid of trace text).

```
erpin trset.epn 732bacs.fa -2,17 -nomask
```

The minus sign before 2 is used to specify the 5' part of stem 02. The `-nomask` argument indicates that all elements in the region should be used for the search. We will see later how we can mask/unmask elements. The ERPIN command takes as much as four days to be executed, producing 21 solutions. A sample is shown in Figure 28.8a. The solutions cover all species from the training set, plus a few closely related species. The search is highly specific, with all E-values around 10^{-20} ; however, it takes a long time and does not find novel distant homologs.

Using the whole structure for search is thus not very practical here (four days' run!), and it can become worse for RNA motifs containing more elements, more gaps, or fewer conserved regions than the RsaE RNA. More gaps in single-stranded segments (helices cannot contain gaps) and more variation in the alignment means longer search times and reduced search specificity. One way to reduce runtime is to restrict the search to a smaller conserved region. For instance, we can use the region from single strand 9 to single strand 17:

```

>Anoxybacillus_flavithermus_WK1_CP000922
RC 1 2094941..2095011 108.53 1.23e-22
GAAGTG.AA.T.C.AC.-AATC.AA.ACAAACTTATACCCCTTTGT.TT.GACC.GT.G.A.AA.AATTC.TCCCATCCCCCTTTGT
>Bacillus_amyoliquefaciens_FZB42_CP000560
FW 1 1134164..1134229 115.95 5.80e-26
GAAATT.GA.T.C.AC.-AAAC.AA.ACA----T-TACCCCTTTGT.TT.GACC.GT.G.A.AA.AATTC.TCCCATCCCCCTTTGT
>Bacillus_anthraxis_str_Sterne_AE017225
FW 1 1170862..1170925 100.75 3.59e-20
GAACGA.AT.T.C.AC.-ATAC.AA.ACA----TATACCCCTTTGT.TT.GAAC.GT.G.A.-A.AATTC.TCCCATCCCCCTTT--
(a)

>Anoxybacillus_flavithermus_WK1_CP000922
RC 1 2094941..2094993 88.48 2.65e-16
ACAAACTTATACCCCTTTGT.TT.GACC.GT.G.A.AA.AATTC.TCCCATCCCCCTTTGT
>Bacillus_amyoliquefaciens_FZB42_CP000560
FW 1 1134182..1134229 90.94 2.15e-17
ACA----T-TACCCCTTTGT.TT.GACC.GT.G.A.AA.AATTC.TCCCATCCCCCTTTGT
>Bacillus_anthraxis_str_Sterne_AE017225
FW 1 1170880..1170925 82.78 3.41e-14
ACA----TATACCCCTTTGT.TT.GAAC.GT.G.A.-A.AATTC.TCCCATCCCCCTTT--
(b)

>Oceanobacillus_iheyensis_BA000028
FW 1 1267734..1267781 43.42 2.82e+00
ACA----AATACCCCTTTG.tttgatcatgaaactttc.....TCCCATCCCCCTTT-T
(c)

```

Figure 28.8 Extracts of ERPIN outputs. Hits are formatted as follows. Line 1: title line of Fasta file entry. Line 2: Forward (FW) or Reverse (RC) strand, a fixed tag, hit positions, Position Weight Matrix (PWM) score and E-value. Line 3: hit sequence. Structural elements are separated by dots. Consecutive dots indicate that corresponding elements were not included in the ERPIN command. Gaps inside elements are used to align the hit sequence onto the training alignment. (a–c) Results of different ERPIN runs (see text).


```
erpin trset.epn 732bacs.fa 9,17 -nomask
```

The search time is now down to a more reasonable 10 h and ERPIN finds the same 21 hits as before. However, because of the different search regions, the hit sequences and E-values are different (Figure 28.8b).

It is relatively easy to further decrease CPU time by implementing a search strategy. The main idea is to perform a stepwise search using masks. In the first step, one should mask most of the selected region and retain only a few key elements. ERPIN will disregard all masked elements and restrict the search to the unmasked part. The best parts to unmask during the first step are those conserved elements that occur rarely in the database and are in close proximity to each other so that they are quickly identified. Then, at each successive step, one unmask additional elements for ERPIN to look for. ERPIN will consider these latter elements only after elements have been identified in the first round. Any number of search steps can be specified on the same command line. For our RsaE motif, let us try the following two-step strategy:

```
erpin trset.epn 732bacs.fa 9,17 -umask 17 -nomask
```

In this command, we first unmask element 17, then unmask the whole region. The search time is reduced to 26 min and this does not change the number of solutions. To increase search sensitivity and find more distant homologs, we now need to play with scores and pseudocounts. Default cutoff scores are set for each element in order to capture the lowest scoring member of the training set. Pseudocounts are used to allow for additional variation in regions that have as yet displayed very little variation, such as strand 17 in Figure 28.7. Playing with scores and pseudocounts can become relatively complex and is beyond the scope of this chapter. Interested readers should refer to the ERPIN documentation. However, the Perl script `erpincommand.pl` provided with the ERPIN distribution can suggest a reasonable ERPIN command using cutoff and pseudocounts to achieve a balanced specificity/sensitivity ratio. The script is run as follows:

```
erpincommand.pl trset.epn
```

and it suggests the following command:

```
erpin trset.epn DATABASE -09,18 -add 09 17 -pcw 51 -cutoff 41.96
```

Running this command (replacing DATABASE with our actual database file name) produces 23 hits after a 176-min run. Some hits have poor E-values (1 or more). To filter hits, we can use the script `readerpin.pl`, as follows, assuming the ERPIN output was redirected to `outfile`:

```
readerpin.pl outfile -e 1
```

This prints out only those hits with an E-value below 1. There are 21 such hits, and the list is the same as before; hence, no new significant hit was found. However, one of the two high E-value hits ($E = 2.82$) is from *Oceanobacillus ihayensis*, a

member of Bacilli that was not in the training set and was also found by HMMER (Figure 28.8c). This confirms the strength of this candidate.

28.5.6

Searching with INFERNAL

INFERNAL uses covariance models to describe a structural RNA sequence alignment and subsequently locates instances of this model in a genome database. A covariance model is a probabilistic description of the sequence and structure constraints in the training set. It is more general than the model used by ERPIN with the caveat that it cannot take pseudoknot constraints into account. Early versions of INFERNAL required amounts of computing power that kept the program out of reach of most biology laboratories. However, version 1.0 implemented heuristic filters that increased search speed considerably and rendered INFERNAL searches accessible to anyone with basic command line skills.

First, the following command should be executed in order to produce the covariance model from the Stockholm structural alignment `trset.stk` we created above in the HMMER section.

```
cmbuild trset.cm trset.stk
```

The output covariance model is written in file `trset.cm`. In principle, a database scan is possible using this file. However, in order to benefit from E-value calculation and performance-enhancing filters, the model must be processed using the `cmcalibrate` command, as follows:

```
cmcalibrate trset.cm
```

This command is more CPU-intensive than the `cmbuild` command. It requires about 40 min in our example, but it can take much longer with complex multi-helix motifs. When it terminates, the `cm` file is changed to incorporate calibration information. We can now run the search over our database that comprises 732 genomes using the `cmsearch` command, as follows:

```
cmsearch trset.cm 732bacs.fa
```

As `cmsearch` is relatively slow, we highly recommend using output redirection and the `nohup` command that enables a program to keep running even if the terminal window is closed:

```
nohup cmsearch trset.cm 732bacs.fa > outfile &
```

Here the `&` sign is used to execute the task in the background. This run takes about 10 h and finds 26 solutions. In addition to the Firmicute species found by other programs, four identical hits are found in different *Listeria* species (Figure 28.9). *Listeria* are Firmicutes and E-values around 0.01 indicate that these are serious candidates. Some *Listeria* hits were found with HMMER but with E-values above 1.

- rsae.fasta: initial RsaE sequence from *S. aureus*
- trset.fa: training set of RsaE sequences in Fasta format
- trset.stk: training set of RsaE sequences and conserved secondary structure in Stockholm format
- trset.epn: training set of RsaE sequences and conserved secondary structure in ERPIN format
- 732bacs.fa: Fasta-formatted database of 732 bacterial genomes (2.5 GB uncompressed).

28.8

Program Versions and Download Sites

- BLAST 2.2.25 (or latest: 2.2.28) from <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/>
- LOCARNA web server: <http://rna.informatik.uni-freiburg.de:8080/LocARNA/Input.jsp>
- HMMER 3.0: <http://hmmerr.org/>
- ERPIN 5.4 : <http://rna.igmors.u-psud.fr/erpin/>
- RNAMOTIF: v3.0.7 <http://casegroup.rutgers.edu/>
- INFERNAL : v 1.0.2. <http://infernal.janelia.org/>
- SHUFFLE. Part of SQUID 1.9 library: <ftp://selab.janelia.org/pub/software/squid/>

Acknowledgments

I thank Magali Naville and Antonin Marchais for their help in running the different software and Alain Denise for his assistance with INFERNAL runs and useful comments on the manuscript.

References

1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
2. Menzel, P., Gorodkin, J., and Stadler, P.F. (2009) The tedious task of finding homologous noncoding RNA genes. *RNA*, **15**, 2075–2082.
3. Leontis, N.B. and Westhof, E. (1998) A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J. Mol. Biol.*, **283**, 571–583.
4. Gutell, R.R., Cannone, J.J., Konings, D., and Gautheret, D. (2000) Predicting U-turns in ribosomal RNA with comparative sequence analysis. *J. Mol. Biol.*, **300**, 791–803.
5. Freyhult, E.K., Bollback, J.P., and Gardner, P.P. (2007) Exploring genomic dark matter: a critical assessment of the performance of homology search methods on noncoding RNA. *Genome Res.*, **17**, 117–125.
6. Legendre M. and Gautheret D. (2008) Seeking RNA motifs in genomic

- sequences, in *Handbook of RNA Biochemistry* (eds R.K. Hartmann, A. Bindereif, A. Schön, and E. Westhof), Wiley-VCH Verlag GmbH, doi: 10.1002/9783527619504.ch35.
7. Macke, T.J., Ecker, D.J., Gutell, R.R., Gautheret, D., Case, D.A., and Sampath, R. (2001) RNAMotif - an RNA secondary structure definition and search algorithm. *Nucleic Acids Res.*, **29**, 4724–4735.
 8. Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L. *et al.* (2003) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–D141.
 9. Gautheret, D. and Lambert, A. (2001) Direct RNA definition and identification from multiple sequence alignments using secondary structure profiles. *J. Mol. Biol.*, **313**, 1003–1011.
 10. Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009) Infernal 1.0: Inference of RNA alignments. *Bioinformatics*, **25**, 1335–1337.
 11. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.*, **31**, 3497–3500.
 12. Smith, C., Heyne, S., Richter, A.S., Will, S., and Backofen, R. (2010) Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA. *Nucleic Acids Res.*, **38**, W373–W377.
 13. Geissmann, T., Chevalier, C., Cros, M.J., Boisset, S., Fechter, P., Noirot, C., Schrenzel, J., François, P., Vandenesch, F., Gaspin, C., and Romby, P. (2009) A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation. *Nucleic Acids Res.*, **37**, 7239–7257.
 14. Bohn, C., Rigoulay, C., Chabelskaya, S., Sharma, C.M., Marchais, A., Skorski, P., Borezée-Durant, E., Barbet, R., Jacquet, E., Jacq, A. *et al.* (2010) Experimental discovery of small RNAs in *Staphylococcus aureus* reveals a riboregulator of central metabolism. *Nucleic Acids Res.*, **38**, 6620–6636.

29

Homology Search for Small Structured Non-coding RNAs

Manja Marz, Stefanie Wehner, and Peter F. Stadler

29.1

Introduction

Homology search is a basic computational technique to identify a biopolymer sequence. In the simplest case, an experimentally determined *query* sequence is compared against the content of a database such as GenBank. A close match of the *query* to a database entry identifies the *query* as a homolog of a known molecule. In the case of small non-coding RNAs (ncRNAs), the search should be conducted against the Rfam database. For some ncRNA classes, such as microRNAs, snoRNAs, and tRNAs, also specific databases are available, which provide search tools *via* the web. But what if a novel ncRNA sequence cannot be readily annotated by comparison to previously published ones? In many cases, such as RNase P RNA, only secondary structure is well conserved, while the underlying sequence has evolved very rapidly. Standard sequence search tools such as BLAST will fail in such a setting. In this chapter, we outline a general protocol to characterize ncRNAs bioinformatically. This entails, in particular, the systematic search for homologous sequences in genomic and possibly other sequence databases and the construction of a secondary structure model.

29.2

Materials

29.2.1

Sequence Data

Initial seed sequences for the family of interest can be retrieved from GenBank¹⁾ or one of the dedicated ncRNA databases, in particular Rfam²⁾ and Noncode³⁾, or in

1) <http://www.ncbi.nlm.nih.gov/Genbank/>

2) <http://rfam.sanger.ac.uk/>

3) <http://www.noncode.org/>

the case of bacterial ncRNAs, from the small noncoding RNAmapping.⁴⁾ Of course, a newly discovered RNA can also serve as seed even if it is not yet contained in a public database.

Genomic sequences, ESTs, or shotgun trace sequences can typically be searched both online *via* web interfaces or locally after download. There is no central collection that holds all data in a convenient format. Vertebrate genomes are best accessible through ENSEMBL,⁵⁾ prokaryotic and organellar genomes are easily searched *via* NCBI.⁶⁾ For many eukaryotic genome projects, dedicated websites are in use, for example, WormBase, FlyBase, and PlasmoDB. The UCSC Genome Browser⁷⁾ unfortunately does not offer sensitive BLAST-based searches; the mapping tool `blat` can retrieve only very similar sequences. We prefer the UCSC browser to view genomic context information and to produce publication-quality figures.

29.2.2

Web Services

In addition to BLAST, one needs tools for predicting and evaluating RNA secondary structure, tools for computing and viewing alignments, and a text editor. Many, but not all of these tasks, can be performed conveniently *via* web interfaces.

In our example of a coherent homology search project outlined below, we use the Vienna RNA Web Suite⁸⁾ [1], that is, the web version of the Vienna RNA Package [2], to compute and manipulate RNA secondary structures. This toolkit provides an instructive tutorial-like feature that can be used to learn how to use the command line-based software locally. Of central importance are the included software tools `RNAfold` and `RNAalifold` [3] to compute the secondary structure of individual sequences and the consensus structure of an alignment, respectively, and `RNAz` [4, 5] to determine whether the alignment contains a conserved secondary structure.

Alignment programs are also available *via* the web. For example, EBI offers access to `ClustalW`⁹⁾ [6]. Sequence logos are a convenient way to summarize conservation information in an alignment, which can be generated using `Weblogo`¹⁰⁾ [7]. While not used in the example below, structure-based information may be useful when highly divergent candidate sequences are to be analyzed. `LocARNA`¹¹⁾ [8], provided as part of the Freiburg RNA Tools¹²⁾, and the Vienna RNA Web Suite¹³⁾ are convenient options for this task.

4) <http://srnamap.mbc.nctu.edu.tw/>

5) <http://www.ensembl.org/>

6) <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

7) <http://genome.ucsc.edu/>

8) <http://rna.tbi.univie.ac.at/>

9) <http://www.ebi.ac.uk/Tools/clustalw2/>

10) <http://weblogo.berkeley.edu/logo.cgi>

11) <http://www.bioinf.uni-freiburg.de/Software/LocARNA/>

12) <http://rna.informatik.uni-freiburg.de:8080/>

13) <http://rna.tbi.univie.ac.at/cgi-bin/LocARNA.cgi>

29.2.3

Web Service-Independent Software

The software behind these web tools can also be downloaded for local use, a strategy that we highly recommend in particular when homology searches for many different queries are to be performed. Simple programs based on scripting languages can then automatize a large part of the protocol below and literally save thousands of mouse clicks.

Some crucial tasks cannot be performed with web tools. In particular, an alignment editor is required to handle the candidate sequences and to facilitate their initial analysis. There are many tools available.

We prefer to use the generic powerful text editor system `emacs`, because, with the RALEE mode [9], `emacs`¹⁴⁾ is turned into a specialized editor for RNA sequence-structure alignments that also provides direct access to a local installation of the Vienna RNA Package. Emacs is part of most standard unix installations and also freely available for all other commonly used operating systems. Alternatively, the platform-independent `SARSE`¹⁵⁾ [10] can be used.

For more sensitive searches beyond the realm of `BLAST`, several tools can be employed, which can only be run on a local computer system. Instead of `BLAST`, semi-global dynamic programming can be performed with `GotohScan`¹⁶⁾ [11], and fragmented sequence patterns can be searched with `fragrep`¹⁷⁾ [12].

Secondary structure features can also be utilized to search for distant homologs. We favor Sean Eddy's `rnabob` [13], owing to its search speed and simple query language. The most sophisticated automated way for RNA homology search is the use of so-called covariance models (CMs), as implemented in `infern` [14]. In a nutshell, CMs are statistical models that represent the likely substitutions both at the sequence and secondary structure level. In particular, they account for the correlated substitution patterns in helices [15]. CMs are trained from a multiple sequence alignment that is annotated with a consensus structure. Their only disadvantage is the high computational cost, hence they are used when other less costly approaches are unsuccessful.

29.3**Protocol: mascRNAs**

In the following text, we describe our standard protocol for RNA detection step by step. We use mascRNAs as an example because of their short length of only 60 nt, which makes it possible to show intermediate results as a whole in view of the limited amount of available space.

14) <http://www.gnu.org/software/emacs/>

15) <http://sarse.ku.dk/>

16) <http://www.bioinf.uni-leipzig.de/Software/GotohScan/>

17) <http://www.bioinf.uni-leipzig.de/Software/fragrep/>

29.3.1

The Seed

Every annotation project based on homology by definition starts out with at least one known representative sequence. Here, we use the two recently discovered mascRNAs from human and mouse (accession numbers FJ209302.1 and FJ209303.1) as an example. In general, it helps to have some background on the ncRNAs that are searched for. In particular, what is the phylogenetic range in which we expect to find homologs? Do we expect single copies per genome or multiple copies as in the case of tRNAs; or do we have to deal with repetitive elements such as ALUs? Are there any special properties that we can use to validate our computational predictions?

Hence, we start with a summary of the information that can be gleaned from literature: The cytoplasmic mascRNA [16] is released from the nascent MALAT-1 transcript by means of RNase P cleavage downstream of a genomically encoded poly-A tract. The MALAT-1 transcript is a long ncRNA found in the nucleus of several mammals [17, 18]. The mature mascRNA is only 61 nt long, has a tRNA-like secondary structure, and carries a CCA 3'-end that is attached by the tRNA processing machinery [16]. Only the two experimentally validated sequences from human and mouse mentioned above can be found in GenBank and form our “seed” set to start with. In addition, Figure 29.1D of Ref. [16] reports homologs in chimp, rhesus, dog, lizard, and stickleback. As they have been found by BLAST

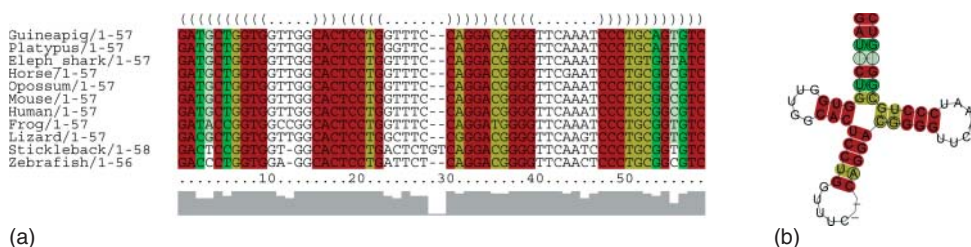


Figure 29.1 (a) Structure-annotated multiple alignment of the two mascRNA seed sequences (human, mouse) and a selection of the BLAST (nucleotide blast) results. It was obtained by first aligning the sequences with ClustalW and using this alignment as input for RNAalifold. The full alignment contains more than 50 sequences. Note that the fourth base pair in the terminal stem is a mismatch in the two teleost species (zebrafish, stickleback). This is a general feature in this clade. Parentheses in the first line of

the alignment indicate the base-pairing nucleotides. For instance, C5 pairs with G55 and T9 with A17. Colors code conservation and compensatory substitutions in the stems: red color – strictly conserved base identities; pale colors indicate that some sequences cannot form the pair; ochre, green, and blue colors indicate an increasing number of different types of base pairs. (b) Conventional drawing of the secondary structure, with base pairs annotated in the same coloring scheme. Both panels are produced by RNAalifold.

according to the publication, we can just leave them out of the seed set and check if we recover them in the first step of our protocol.

29.3.2

Low-Hanging Fruits: Initial **BLAST** Search

As **BLAST** [19] is well known to experimental researchers, it is applied to all easy homology search problems. It can be installed locally and conveniently run from a little script across all genomes of interest, in our case the vertebrate ones. Alternatively, homology searches are easily performed using the web interface of ENSEMBL or NCBI. Using ENSEMBL, simply click on the “BLAST/BLAT” page, paste the two seed sequences (in FASTA format) into the “Query Sequence” window, and mark the genomes to be searched by mouse click. Choose “BLASTN” as search tool and set the search sensitivity to “Distant homologies”. Then *configure*: we switch off repeat masking to see the true number of hits in the genome, reduce the *E*-value cutoff to 0.01 to reduce the noise and hit the “run” button (jobs are queued now). After a couple of minutes, the results are ready for inspection. In case of *maskRNA*, there is only a single hit in the human genome. Pressing the *view* button, the genomic location and four links at the lower left of the page are displayed. Clicking “[A]” shows the pairwise alignments between query and genomic sequence, and “[G]” gives the genomic sequence with some context, a convenient way to extract the sequence by copy and paste. The “[C]” link leads to the genome browser graphics, showing the genomic context.

We collect the sequences of the putative *maskRNA* homologs in a simple text file, giving each of them a FASTA style header indicating its source. At this point, we realize several peculiarities:

- 1) There are no **BLAST** hits for invertebrate genomes.
- 2) While most mammals show **BLAST** hits with *E*-values between 10^{-5} and 10^{-10} , nothing is found in several other mammals. Conspicuously, no hit is found in rat (*Rattus norvegicus*). As we see in the following, this is because of gaps in the current genome assembly.
- 3) There is one clear hit in the frog (*Xenopus tropicalis*) and 41(!) hits in the lizard *Anolis carolinensis*, but nothing in the birds *Gallus gallus* and *Taeniopygia guttata*.
- 4) Two hits in the zebrafish *Danio rerio* have a rather mediocre *E*-value of about 0.0001, the candidate in the pufferfish *Tetraodon nigroviridis* is found with the mouse query only, and nothing at all is found in the other pufferfish *Takifugu rubripes*.

Aligning these sequences (by pasting the FASTA files into the **ClustalW** web-server) shows quite convincingly that the top hits from the lizard, the stickleback, and the zebrafish sequences fit together with our seed sequences. In order to obtain a more complete set of *maskRNAs* of all organisms, we repeat the search with a phylogenetically diverse set of queries. In the *maskRNA* example, this consists of

some of the teleost fishes, the frog, and mammalian sequences. We obtain mostly the same sequences again. Still no bird homologs are detected.

We also screen GenBank for additional homologs by clicking on “BLAST” and then “nucleotide blast”. We search not only the “nr/nt” database but also genome survey sequences (GSS), high-throughput genomic sequences (HTGS), whole-genome shotgun reads (WGS), and ESTs, which are related databases that serve as additional sources for the discovery of novel ncRNAs. Setting the default program selection “megablast” to “blastn” increases the computation time but enhances sensitivity to also detect more divergent homologs. This adds a few more sequences, in particular, from some teleost fishes for which no complete genomes are available, as well as a rat sequence that is nearly identical to the mouse seed sequence. A cross check with the rat genome in ENSEMBL shows that the sequence is indeed missing in the current genome assembly.

For more complex search problems, the BLAST searches might need to be iterated several times until no further sequences are found. The end result of this first step is a multiple alignment of clearly recognizable homologous sequences similar to the one illustrated in Figure 29.1.

29.3.3

Initial Secondary Structure Model

The next step is to analyze the fruits collected in the first round. First, we check whether there is a conserved secondary structure common to the mascRNAs. This can be done using RNAalifold; for the purpose of illustration, we again used a web service (Figure 29.1). Other tools could be RNashapes¹⁸⁾, a comparative prediction tool of consensus structures, RNAforester¹⁹⁾ [20], an alignment tool for RNA secondary structures, or Locarna for a sequence and secondary structure alignment.

We use RNAz to check whether the secondary structure prediction is significant. Pasting the mascRNA alignment of Figure 29.1 into the web server (just using the default setting), the HTML output tells us that the prediction is “RNA” with a “class probability” of 0.902, that is, the program has 90% confidence in its classification. We conclude that evolution of mascRNA has been under pressure to select for stable secondary structures, a feature that should be explicitly used in the subsequent search steps.

Next, we evaluate our sequences against the predicted consensus secondary structure. A few minor manual adjustments of the alignment can possibly improve the consistency with the consensus structure, or the secondary structure prediction itself. This can be checked with RNAz, where a classification score and structure conservation index (SCI) should improve the edited alignment. Next, we identify sequences that do not fit well into the alignment. In our mascRNA example, this pertains, for example, to some of the more deviant candidates from the

18) <http://bibiserv.techfak.uni-bielefeld.de/rnashapes/>

19) <http://bibiserv.techfak.uni-bielefeld.de/rnaforester/>

lizard genome, and an EST sequence from the flounder *Paralichthys*, which has a conspicuous deletion in the “anticodon” stem.

Alternatively, other secondary structure prediction programs, such as *Locarna* and *RNAforester*, can be used to compare the previous results. FASTA sequences of mascRNA are taken as input for *Locarna*, resulting in slightly better sequence/structure alignments. Furthermore, there is *RNAforester* that compares RNA secondary structures and, therefore, needs besides the sequences their known or predicted structures as input. In the case of mascRNA, *RNAforester* predicts formation of three longer hairpins instead of the cloverleaf structure calculated by *RNAalifold* (Figure 29.1).

At this point, it is rewarding to spend some time on investigating the genomic locations of the seed and the candidate sequences. Among multiple hits in the same genome, the most likely genuine candidate is the one located in a syntenically conserved region relative to the seed sequences. In our case, this is simple: we can use a genome browser to locate the homologs of MALAT-1 and compare their locations with those of our BLAST hits. In the case of the lizard, for instance, we can identify the sequence on scaffold 443 as our best guess, as it represents the longest sequence stretch on which MALAT-1 is annotated. The UCSC genome browser annotates (using “BLAT”) five additional loci as MALAT-1 homologs; however, their conservation is restricted to the length of the mascRNA. We, therefore, use the sequence on scaffold 443 as the representative of the lizard mascRNA in the further analysis. MALAT-1 homolog(s) are not annotated in the frog or stickleback genomes. Here, we can use the presence of the adjacent SCYL1 gene to conclude that the mascRNA is located at the syntenically conserved genomic location. In fact, upstream of the mascRNA, there is a 10 kb region of the frog scaffold 398 that is full of unspliced ESTs, suggesting that there is a homolog of the MALAT-1 transcript also in the frog [21].

Before we proceed, we consider the phylogenetic distribution of the sequences that we have retained in our alignment. What is missing? We lack candidates in some mammalian genomes, mostly those with poor coverage and in the early stages of assembly. In some cases, a simple BLAST search in the NCBI Trace archive can fill the gap. We also lack candidates in birds and for all species outside the vertebrates. Even if mascRNA is a vertebrate innovation, we should have found avian homologs. Have they been lost in evolution, or are they just too divergent to be picked up by BLAST?

29.3.4

Drilling Deep – Structure-Based Searches

To obtain more candidates for mascRNA in other eukaryotes, we have defined conserved sequences and secondary structures. To this end, we computed the consensus secondary structure from the mascRNA alignment using *RNAalifold*. The inspection of the conservation patterns of both sequence and structure (e.g., using the *emacs* RALEE mode) shows a rather well conserved cloverleaf structure

with many nicely conserved sequence blocks. These can be converted into an `rnabob` descriptor:

```
# structural arrangement
s1 h1 s2 h1' h2 s3 h2' s4
# sequence patterns
s1 0 NNNNNNNN
h1 0:0 GTG:CAC
s2 0 NNN**
h2 1:1 TCCTGG:CCAGGA
s3 0 NNN*****
s4 0 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

The pattern consists of a single line that designates the structural architecture, s_i denotes unpaired sequences, and h_j and h_j' refer to the two sides of helical regions. As an example, we model two hairpins, h_1 s_2 h_1' thus refer to the first hairpin: the stem has to be 5'-GTG-loop-CAC-3' and the loop should consist of three to five nucleotides. The second hairpin is also very conserved and, therefore, specified as 5'-TCCTGG-loop-CCAGGA-3' with a single substitution allowed on each side in the helix h_2 . Length variations can be specified with asterisks (*), where each asterisk will allow either 0 or 1 N. The `rnabob` search runs over the genomes in minutes. All candidates are added to our structure-annotated alignment and are inspected. Using this approach, we obtain a homolog in the squirrel *Spermophilus tridecemlineatus* that has a divergent 3' end:

Spermophilus ATTGTTGGTGGTTGGCACTCCTGGCTTCAGGACGGGGTGATACACTTAGCCGAGAT
 |||||||
Homo GATGCTGGTGGTTGGCACTCCTGGTTTCAGGACGGGGT--TCAAATCCCTGCGCCGCTC

We still failed, however, to find mascRNA candidates in the avian genomes.

We next used `infern` 1.0 [14] in an attempt to find more distant homologs in the avian genomes and in the basal deuterostomes. To this end, we start from the structure-annotated alignment of the sequences collected so far and construct a CM with `cmbuild`. This program produces a profile that needs to be calibrated by running `cmcalibrate`. This step, which takes several hours, adjusts internal parameters in the CM model in order to generate realistic estimates of *E*-values in the subsequent search. With `cmsearch`, we compare this model against the genomes of interest. In practice, this amounts to the following three simple commands:

```
cmbuild mascRNA.cm mascRNA.stk
cmcalibrate mascRNA.cm
cmsearch mascRNA.cm genome.fa
```

As we expect that the hit will be unspliced and the query is already short, we insist on global alignments. While `cmsearch` returns several candidates, their *E*-values

are close to 1 and a closer inspection shows that they do not at all fit well to our alignment. We, therefore, discard them.

Many of the well-known ncRNAs appear in multiple copies in the genome, and several of them have rather divergent paralogs. For instance, vault RNAs appear in four expressed copies in the human genome, the last of which was identified only recently after having been misannotated as a microRNA [22]. Similarly, several paralog groups of Y RNAs have evolved in vertebrates [23, 24]. In some cases, these related ncRNAs appear in similar genomic locations. It makes, therefore, sense to check also the genomes that have already been harvested by BLAST with more sensitive methods, such as *cmsearch*.²⁰⁾

In our example, running *cmsearch* against the human chromosome 11 yields a single convincing hit with our mascRNA CM:

```
((((((((<<<_____>>><<<<<<.____>>>>>><<<<_____>>>>))))))):
1  GAcaCcGGuGGUUGGCaCUCCuGg..uUucCaGGACgGGGUUCAaUCCCuGCgGugUCU 58
   C:C:GGUGGU GGCAC: C::G    C::G :C:GGGUUC A UCCC:GC:G:G
65213012 GGCUCUGGUGGU-GGCACGUCCAGcaCGGCUGGGCCGGGUUCGAGUCCCCGAGUGUUG ..070
```

Although the *cmsearch* *E*-value of about 10^{-3} is still not stellar, we investigated the candidate further. To this end, we first searched for homologs of this new candidate (termed *menRNA*) as described earlier for mascRNAs. With BLAST, one easily identifies orthologs throughout the eutherians and uncovers a diverged hit in the genome of the wallaby *Macropus eugenii*, a marsupial. As in human, the location of this hit is close to the mascRNA in the wallaby genome. A representative subset of these results is shown in Figure 29.2.

As for the mascRNA, the structure of *menRNA* is clearly a well conserved tRNA-like cloverleaf. Mapping the candidate sequence back to the genome shows that it is located adjacent to the 3'-end of the *MEN β* transcript (also known as *NEAT1*), an abundant transcript localizing to the paraspeckles [25, 26]. We have,

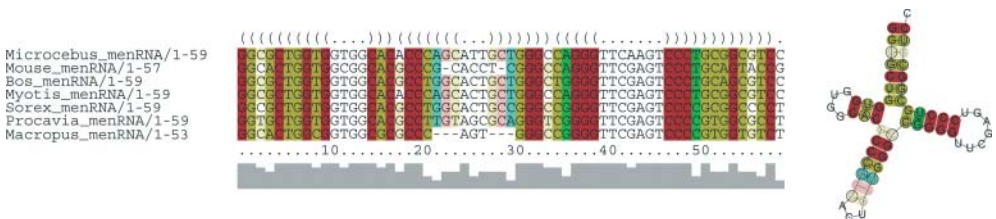


Figure 29.2 Structure-annotated alignment and secondary structure of *menRNAs*, which are homologs of mascRNAs.

20) The current *infern* version has problems with very large stretches of N's in the subject sequence. We thus removed the N's from the sequences, ran *cmsearch*, and recovered the correct coordinates by mapping the hits onto the changed sequences with *Blat*.

therefore, designated the mascRNA homolog menRNA. The human menRNA is described in some detail in the supplemental material of Ref. [27]: Similar to mascRNA, it is cleaved off the primary NEAT1 transcript by RNase P and its 3'-end is processed by RNase Z with subsequent addition of a CCA tail. Interestingly, MEN/NEAT1, MALAT-1/NEAT2, and Xist were identified as the only polyadenylated long RNA transcripts displaying nuclear enrichment [18]. For both MEN/NEAT1 and MALAT-1/NEAT2, the poly-A tail is genomically encoded and exposed by cleaving off the tRNA-like mascRNA and menRNA, respectively [27]. An overview of the locus is given in Figure 29.3.

The combined alignment of the menRNAs and mascRNAs was again fed into `cmbuild` and `cmcalibrate` to obtain a CM that would recognize both homologous groups, and thus might be general enough to detect the missing bird homologs, if they exist. We searched against the avian genomes, the lamprey genome, the urochordates (*Ciona intestinalis*, *Ciona savignyi*, and *Oikopleura dioica*), as well as the cephalochordate *Branchiostoma floridae*. The CM produced only hits with *E*-values close to 1. The top-scoring `cmsearch` results were compared against the genome annotation and found to be either tRNAs or repetitive elements. As the sensitivity of the descriptor is already so high that it starts to pick up *bona fide* tRNAs, we suggest that (i) the mascRNA family is a gnathostome innovation (although our failure to find a lamprey homolog may be an artifact of the genome assembly) and (ii) that Aves have lost their mascRNAs or, alternatively, still have descendants of mascRNA that diverged beyond recognition. Such extreme deviations in sequence and structure are unusual but have been observed, for example, for 7SK RNAs [28].

The remaining task is to check the genomic locations of our candidate sequences. As both the human and the wallaby masc- and menRNAs are encoded in the same orientation and not more than 65 kb apart from each other, we expect a similar pattern also for the other mammalian genomes. As summarized in Figure 29.4, this is indeed the case for all genomes where we have found homologs of both candidates.

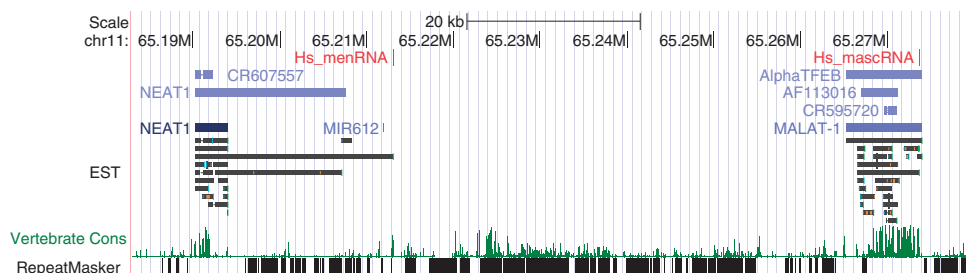


Figure 29.3 Genome browser image of the NEAT1 and MALAT-1 locus, showing the position of the tRNA-like elements at the ends of the mature long ncRNAs. EST data are indicated in the middle, and sequence conservation and repetitive elements are depicted at the bottom.

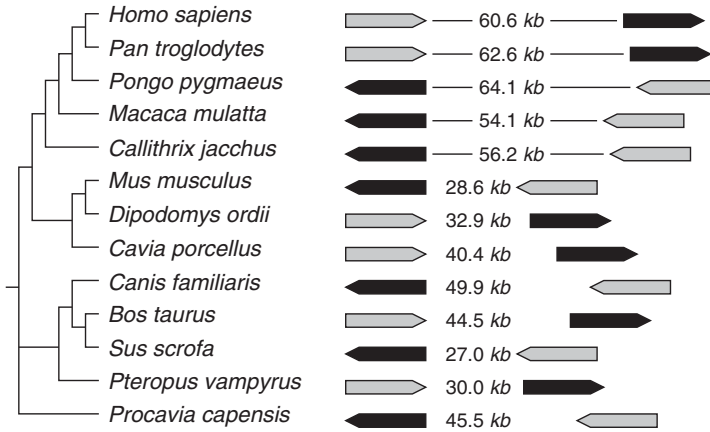


Figure 29.4 MenRNA (gray) and mascRNA (black) are located on the same chromosome in all eutherians at a distance of 27 to 64.1 kb. The fact that mascRNA is always located downstream of menRNA suggests that the two RNAs have position-specific roles in a common functional context.

29.4

Concluding Remarks

The analysis of the mascRNAs used here as an example took about three weeks. The final results, namely, a manually curated alignment of the RNA family, a well-supported consensus structure, and the knowledge about its phylogenetic distribution are valuable information.

The Rfam database, a so-called reference database, collects exactly this type of information, which is based on homology and is utilized, for example, by the ENSEMBL annotation pipeline. In particular, it provides structure-annotated alignments for many, but by no means all, of the better known RNA families. These often serve as starting points for homology search projects similar to the one outlined in this chapter. Alignments of RNA families can be submitted directly to the curators of the Rfam database.

The journal *RNA Biology* publishes careful homology search studies that characterize new ncRNA families in its *RNA Family* section, provided that the alignment is also submitted to Rfam [29].

The protocol we have described here is limited to methods and software that can be applied fairly routinely. While sufficient in many cases, homology-based annotation can be extremely difficult for particular RNA families (such as telomerase RNA), requiring custom-made tools to recognize and evaluate candidates with even higher sensitivity. We refer to Ref. [30, 31] for a more detailed discussion of the limitations and for hints on how to proceed in such hard cases.

Acknowledgments

We thank our “guinea pigs” Benedikt Beckmann, Maren Thomas, the students of the pharmaceutical “Compulsory subject choice Practical Course”-Marburg SS10 for testing this protocol, and the GRK 1384 for funding.

References

1. Gruber, A.R., Lorenz, R., Bernhart, S.H., Richard, N., and Hofacker, I.L. (2008) The Vienna RNA Websuite. *Nucleic Acids Res.*, **36**, 70–74.
2. Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, S.L., Tacker, M., and Schuster P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, **125**, 167–188.
3. Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R., and Stadler, P.F. (2008) RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics*, **9**, 474.
4. Washietl, S., Hofacker, I.L., and Stadler, P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc. Natl. Acad. Sci. USA*, **102**, 2454–2459.
5. Gruber, A.R., Findeiß, S., Washietl, S., Hofacker, I.L., and Stadler, P.F. (2010) RNAz 2.0: improved noncoding RNA detection. *Pac. Symp. Biocomput.*, **15**, 69–79.
6. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
7. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004) WebLogo: A sequence logo generator. *Genome Res.*, **14**, 1188–1190.
8. Will, S., Missal, K., Hofacker, I.L., Stadler, P.F., and Backofen, R. (2007) Inferring non-coding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comp. Biol.*, **3**, e65.
9. Griffiths-Jones, S. (2005) RALEE - RNA alignment editor in Emacs. *Bioinformatics*, **21**, 257–259.
10. Andersen, E.S., Lind-Thomsen, A., Knudsen, B., Kristensen, S.E., Havgaard, J.H., Torarinsson, E., Larsen, N., Zwieb, C., Sestoft, P., Kjems, J., and Gorodkin, J. (2007) Semiautomated improvement of RNA alignments. *RNA*, **13**, 1850–1859.
11. Hertel, J., de Jong, D., Marz, M., Rose, D., Tafer, H., Tanzer, A., Schierwater, B., and Stadler, P.F. (2009) Non-coding RNA annotation of the genome of *Trichoplax adhaerens*. *Nucleic Acids Res.*, **37**, 1602–1615.
12. Mosig, A., Chen, J.L., and Stadler, P.F. (2007) Homology search with fragmented nucleic acid sequence patterns. In *Algorithms in Bioinformatics (WABI 2007)*, volume 4645 of Lecture Notes in Computer Science, (eds R. Giancarlo and S. Hannenhalli), Springer Verlag, Berlin, Heidelberg, pp. 335–345.
13. Eddy, S.R. *RNABOB: A Program to Search for RNA Secondary Structure Motifs in Sequence Databases*, (1996) <http://selab.janelia.org/software.html>.
14. Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics*, **25**, 1335–1337.
15. Eddy, S.R., and Durbin, R. (1994) RNA sequence analysis using covariance models. *Nucleic Acids Res.*, **22**, 2079–2088.
16. Wilusz, J.E., Freier, S.M., and Spector, D.L. (2008) 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell*, **135**, 919–932.
17. Ji, P., Diederichs, S., Wang, W., Böing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B., Buerger, H., Bulk, E. et al. (2003) MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage

- non-small cell lung cancer. *Oncogene*, **22**, 8031–8041.
18. Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007) A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*, **8**, 39.
 19. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
 20. Höchsmann, M., Töller, T., Giegerich, R., and Kurtz, S. (2003) Local similarity in RNA secondary structures. *Proc IEEE Comput Soc Bioinform Conf*, **2**, 159–168.
 21. Stadler P.F. (2010) Evolution of the long non-coding RNAs MALAT1 and MEN β/ϵ . In *Advances in Bioinformatics and Computational Biology, 5th Brazilian Symposium on Bioinformatics*, volume 6268 of Lecture Notes in Computer Science, (eds Carlos Eduardo Ferreira, Satoru Miyano, and Peter F. Stadler), Springer Verlag, Heidelberg, pp. 1–12.
 22. Stadler, P.F., Chen, J.J., Hackermüller, J., Hoffmann, S., Horn, F., Khaitovich, P., Kretschmar, A.K., Mosig A., Prohaska, S.J., Qi, X., Schutt, K., and Ullmann, K. (2009) Evolution of vault RNAs. *Mol. Biol. Evol.*, **26**, 1975–1991.
 23. Mosig, A., Guofeng, M., Stadler, B.M., and Stadler, P.F. (2007) Evolution of the vertebrate Y RNA cluster. *Theory Biosci.*, **126**, 9–14.
 24. Perreault, J., Perreault, J.P., and Boire, G. (2007) Ro-associated Y RNAs in metazoans: evolution and diversification. *Mol Biol Evol*, **24**, 1678–1689.
 25. Bond, C.S. and Fox, A.H. (2009) Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.*, **186**, 637–644.
 26. Sasaki, Y.T., Ideue, T., Sano, M., Mituyama, T., and Hirose, T. (2009) MEN ϵ/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci USA*, **106**, 2525–2530.
 27. Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S., and Spector, D.L. (2009) MEN ϵ/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.*, **19**, 347–359.
 28. Marz, M., Donath, A., Verstaete, N., Nguyen, V.T., Stadler, P.F., and Bensaude, O. (2009) Evolution of 7SK RNA and its protein partners in metazoa. *Mol. Biol. Evol.*, **26**, 2821–2830.
 29. Gardner, P.G., and Bateman, A.G. (2009) A home for RNA families at *RNA Biology*. *RNA Biology*, **6**, 2–4.
 30. Menzel, P., Gorodkin, J., and Stadler, P.F. (2009) The tedious task of finding homologous non-coding RNA genes. *RNA*, **15**, 2075–2082.
 31. Mosig, A., Zhu, L., and Stadler, P.F. (2009) Customized strategies for discovering distant ncRNA homologs. *Brief. Funct. Genomics Proteomics*, **8**, 451–460.

30

Predict RNA 2D and 3D Structure over the Internet Using MC-Tools

Stephen Leong Koan, Jonathan Roy, Marc Parisien, and François Major

30.1

Introduction

Determining RNA tertiary (3D) structures at high resolution is the key to studying RNA function, but is not always possible using physical methods such as X-ray crystallography and nuclear magnetic resonance. Over approximately the last twenty years, our research group at the Université de Montréal has developed, and made available over the Internet, computational tools and services for the rapid and reliable modeling of RNA 2D and 3D structures from sequence information. Here, we introduce a series of protocols to model structures of an RNA for which we know either a single sequence or multiple sequences. We also outline how to introduce experimental data in the modeling steps so as to restrain the conformational space. The protocols use web services integrated under a single and simple dashboard architecture, which simplifies the managing of modeling projects remotely over the Internet.

The tools that are necessary for applying the modeling protocols are MC-Fold, MC-Sym, and MC-Cons [1]. MC-Fold predicts 2D structures of single RNA sequences and duplexes formed from two separate RNA single strands. MC-Sym uses the results of MC-Fold to predict 3D structures from MC-Fold's output. MC-Cons uses the results of MC-Fold to select among the set of predicted 2D structures one for each sequence that maximizes structural similarities with all others. Compared to classical 2D structure prediction methods that account for the energetic contributions of canonical base pairs (i.e., A:U and G:C Watson-Crick, and G:U wobble pairs), our tools make use of nucleotide cyclic motifs (NCMs) that incorporate canonical and non-canonical base pairs. NCMs are small RNA motifs of eight or less nucleotides made of either a single strand closed by a base pair (i.e., hairpin loops with their flanking base pair), or two strands closed by base pairs (e.g., tandems of adjacent base pairs, or bulge and interior loops with their flanking base pairs). Since non-canonical base pairs are present in almost all RNA structures, prediction of 2D structures that include them provides more information and links better to 3D structure prediction using MC-Sym [2, 3].

We provide 12 step-by-step protocols allowing you to obtain RNA 2D and 3D structures from a single, or a set of related sequences, irrespective of structural constraints being available or not. Specifically, if you want to:

- Connect and register a new user to the MC-Tools web site, apply Protocol 1
- Predict 2D structures of single RNA sequences, apply Protocol 2a
- Predict 2D structures of single RNA sequences using structural data, apply Protocol 2b
- Add non-canonical base pairs to a 2D structure obtained from third-party software, apply Protocol 2c
- Predict 2D structures of RNA duplexes (interactions between two separate RNA strands), apply Protocol 2d
- Predict 2D structures of several RNA sequences, apply Protocol 2e
- Produce a consensus structural assignment of multiple-sequence data, apply Protocol 3
- Predict 3D structures from a 2D structure, apply Protocol 4a
- Energy-minimize 3D structures, apply Protocol 4b
- Partition 3D structures, apply Protocol 4c
- Analyze 3D structures, apply Protocol 4d
- Fit experimental structural data to 3D structures, apply Protocol 4e.

30.2

Materials

30.2.1

Equipment

- 1) Any workstation, personal computer, laptop, or other electronic device connected to the Internet (e.g., Mac running Mac OS X, PC running Linux or Windows, tablets, smartphones).
- 2) A web browser. The web pages described here were tested using the following browsers and computers: Safari 5.1.7 on an iMac running Mac OS X 10.7.4 and 10.8.2, and Google Chrome 19.0.1084.56 and Firefox 13.0.1 on a PC running Windows 7 Enterprise x64 (other operating systems and/or web browsers should render the web pages properly, but as they have not been tested, we cannot guarantee their compatibility; please report any problem to mctools@iric.ca).

30.2.2

Data

- 1) The RNA sequences in electronic format.
- 2) If available, base-pairing and reactivity constraints, such as those obtained from chemical and enzymatic probing methods, in electronic format.

30.3

MC-Tools

30.3.1

MC-Fold

MC-Fold predicts a number of 2D structures associated with an RNA sequence, ranked by their free energy in kcal/mol (Protocol 2a). The best or minimum free energy (MFE) structure often represents a ground state structure, which does not always correspond to the *in vivo* or active structure. The latter can often be found among the remaining structures predicted as suboptimal. As RNA structures may be induced or modulated by cofactors, such as proteins [4], selecting an active structure merely on computational grounds and without knowledge of these cofactors is often not an easy task.

The number of 2D structures grows exponentially with the length of the sequence [5]. In MC-Fold, you decide how many predicted structures you are willing to work with. When possible, obtaining nucleotide reactivity constraints is the best way to identify active structures among a set of predicted 2D structures (Protocol 2b). Such constraints can be derived from chemical and enzymatic probing data. In general, reactive nucleotides are associated with unpaired states, and non-reactive residues with paired states. In Table 30.1, we provide a list of experimental methods that can be applied to produce such structural constraints.

We also provide a protocol to add non-canonical base pairs to classical 2D structures (Protocol 2c), and to predict the structure of RNA duplexes, that is, the interactions between two separate RNA strands (Protocol 2d).

Another version of MC-Fold, MC-Fold2, has been developed and made available as an option. MC-Fold2 delivers better performances and can be invoked with longer sequences compared to MC-Fold. However, while of writing this chapter, MC-Fold2 is still in a beta-testing phase and you should use it at your own risk. The protocols that apply to MC-Fold apply to MC-Fold2 as well.

Table 30.1 Experimental methods for deriving base-pairing constraints.

Method	References
Detect G=U base pairs in A-form RNA duplexes.	[6]
Measure the flexibility of nucleotides by selective 2'-hydroxyl acylation and primer extension (SHAPE). Low flexibility of a nucleotide may indicate pairing.	[7–11]
Analyze susceptibility to enzymatic cleavage. Enzymes attacking unpaired nucleotides include nuclease S1 and RNases A, U2, T1, and T2. An enzyme that attacks double-stranded regions is RNase V1. For corresponding protocols, see also Chapter 10.	
Analyze susceptibility to chemical cleavage. This is similar to enzymatic cleavage, but the probes are smaller, and thus less affected by the RNA's global structure. Commonly used chemical probes are Dimethyl sulfide (DMS), 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho- <i>p</i> -toluenesulfonate (CMCT), and Kethoxal. For corresponding protocols, see also Chapter 10.	[12–15]

30.3.2

MC-Cons

Collecting sequences of the same family (i.e., evolutionarily related sequences) is another way to get information helping with the identification of active structures. Since more than one structure can be shared between sequences of the same family, MC-Cons does not find a single consensus structure that fits all sequences. Instead, it assigns to each sequence one of its suboptimal predictions that globally optimize the sum of pair-wise similarities with all others (Protocol 3).

30.3.3

MC-Sym

The MC-Tools web site contains a button to submit to MC-Sym any 2D structure for 3D modeling. MC-Sym predicts a set of 3D structures for such a 2D structure (Protocol 4a). A set of 3D structures is called a *Decoy*. MC-Tools contains many actions that can be applied to *Decoys*, including creating new decoys from existing ones, energy minimization (Protocol 4b), clustering (Protocol 4c), various 3D structure analyses (Protocol 4d), as well as fitting global geometry, distance restraints, footprinting data, and excluded volumes (Protocol 4e).

Procedure**Protocol 1: Connect and Register a New User to the MC-Tools web Site**

- 1) **CONNECT** to the MC-Tools web site: www.major.irc.ca/Web/MC-Tools

The MC-Tools front page, also called *Dashboard*, is made of three sections. The first one is for Login or Registration. When connecting to MC-Tools, you will automatically have the status of an anonymous user (“Hi anon”). You can use the MC-Tools by staying anonymous, but your work will be lost as soon as you disconnect from the MC-Tools web site. Alternatively, you can register as a new user:

- 2) **CLICK** the Register button
- 3) **FILL** all fields with the appropriate information
- 4) **CLICK** the Register button at the bottom of the form

MC-Tools will then return to the Dashboard and say “Hi <your first name >”. The Register button at the top-right corner will be changed to a new button called Change personal info. The advantage of registering as a user is that your work will be saved and reloaded each time you Login to the MC-Tools web site.

The second section of the Dashboard is the Toolbar, which gives access to the tools: MC-Fold, MC-Fold2(beta), MC-Sym, MC-Cons, and MC-Fold BATCH (upper part below the MC-Tools title). New tools may appear over time in the toolbar. To inspect the protocol of changes for each update, click on News and Troubleshooting at the bottom of the web page. When first

connecting to MC-Tools, the Dashboard has no items in it, and displays the message “No project has been found for this account”. The “Projects” section of the Dashboard is displayed only if you submitted a modeling job to anyone of the tools and is for showing and accessing the results of these jobs. When logging in to MC-tools the second time after having been active at this web site, the “Projects” section appears, containing either a single project termed “Main” or additional projects that you created and named yourself. For each project (which can be opened and closed by clicking on the project name), the Dashboard will display the date and time of its creation and last activity (“Created on”, “Last Activity on”). A list of buttons allows for changing the name of the project, clone it (make a copy of), download it, and delete it: Change Name, Clone, Download, and Delete. The jobs related to the project are listed below the project actions bar.

For each job, the Dashboard displays the associated Tool, Job Name, Status (e.g., Queuing, Running, Done), Start and End date and time, and a list of icons that allows one to perform Actions on the results of the jobs, that is, Inspect, Download, Clone parameters, Trace, and Delete. The time displayed in the Dashboard is the time zone of our servers in Montreal, QC, Canada (UTC-05:00).

Now that you are registered and connected to the MC-Tools web site, you can start ***Modeling RNA 2D and 3D structure over the Internet using MC-Tools!***

Protocol 2a: Predict 2D Structures of a Single RNA Sequence

To predict 2D structures of a single RNA sequence (Figure 30.1a):

- 1) **CLICK** the MC-Fold tab
- 2) **SELECT** the name of the project or create a new project in the Project menu
- 3) **TYPE** the name of your sequence in the Name field
- 4) **TYPE** or **PASTE** the RNA sequence into the Sequence field

In the example, we used:

Name	Rat 28S rRNA loop E
Sequence (5' to 3')	GGGUGCUCAGUACGAGAGGAACCGCACCC (29 nt)

Your sequence may contain spaces, dashes “-”, and underscores “_”, which are automatically removed before submission to the MC-Fold Server. The nucleotides can be in upper or lower case characters. However, only the following nucleotide symbols are accepted:

- Adenosine: “a” or “A”
- Cytosine: “c” or “C”
- Guanosine: “g” or “G”
- Uridine: “u” or “U”
- Thymidine: “t” or “T” (which are systematically converted to Uridine).

Dashboard MC-Fold MC-Fold2(beta) MC-Sym MC-Cons MC-Fold Batch

▼ General

Project: Create New Project Protocol 2a

Name: Rat 28S rRNA loop E

Sequence (5'→3'): GGGUGCUCAGUACGAGAGGAACCGCACCC

Annealed Sequence (5'→3') (optional):

▶ Algorithm

▶ Structural Constraints

▶ Single-Stranded Chemical/Enzymatic Reactivity Data

Submit Reset

(a)

Dashboard MC-Fold MC-Fold2(beta) MC-Sym MC-Cons MC-Fold Batch

Projects:

Protocol 2a

Created on: 2012.10.18 11:20
Last Activity on: 2012.10.18 11:20

Project actions: Change Name Clone Download Delete

Showing 1 to 1 of 1 entries
Show 10 entries
Search:
First Previous 1 Next Last

Tool	Job Name	Status	Start	End	Actions
MC-Fold	Rat_28S_rRNA_loop_E	Done! (50 Sol. MFE -28.38kcal/mol)	2012.10.18 11:20	2012.10.18 11:21	

Showing 1 to 1 of 1 entries
Show 10 entries
Search:
First Previous 1 Next Last

(b)

Figure 30.1 Predict 2D structures of a single RNA sequence. (a) The MC-Fold job submission form and (b) the Dashboard displaying the projects and jobs associated with the user account.

- 5) **CLICK** the `SUBMIT` button to send your job to the MC-Fold server

After having clicked the `SUBMIT` button, you will be redirected to the Dashboard. For each job submitted, the Dashboard displays a new entry, and follows its progress. The Status of MC-Fold jobs goes from `Queuing` to `Running` to `Finalizing` to `Done`. The Actions available during the status `Queuing`, `Running`, and `Finalizing` are: `Kill` and `Trace`. The `Kill` action removes the job from the server, but leaves the job on the Dashboard, so that you can do the following Actions: `Download`, `Clone params`, `Trace`, and `Delete`. `Download` allows you to download on your computer the files associated with the job. `Clone params` redirects you to the MC-Fold tab with the fields pre-filled with the same parameters as the ones you used previously for the aborted job. This is useful, for instance, when you need to submit similar jobs, for example, when you want to modify the sequence or change the Algorithm parameters. `Trace` brings you to the `Trace` page. Finally, `Delete` removes the job from the Dashboard. You `Kill` and then `Delete` a job when you submitted it by error, for instance.

You `Trace` a job to follow its execution. By clicking the `Trace` action, you are redirected to a page that refreshes every 60 s. This page gives you the job Status, Elapsed time, command line (that can be used with the Linux version of MC-Fold). Two windows give information about the running job: Standard output and Standard Error.

Once the job is completed, the Status `Done` appears, allowing you to perform the following Actions: `Inspect`, `Download`, `Clone params`, `Trace`, and `Delete` (Figure 30.1b). `Inspect` allows you to access the results of your MC-Fold job in a table, where each entry is a predicted 2D structure.

- 6) **CLICK** the `Inspect` button to access the 2D predicted structures

The predicted structures are sorted by energy (Figure 30.2a). Structure Rank 1 has the lowest energy, then structure Rank 2 the second lowest, and so on. The dot bracket representation for each structure is given, then the free energy in kcal/mol, the free energy contribution of coaxial stacking in kcal/mol, and three actions: `Display`, `Submit`, and `Edit`.

- 7) **CLICK** the `Display` button to access one of four representations of the 2D structure
- 8) **CLICK** the `PS`, `JPG`, or `PDF` button to display the 2D structure in planar representation

Additionally, an output in Connectivity Table (CT) format is provided, which is an ASCII representation (Figure 30.2b). It is made of three columns. The first column is the nucleotide number. The second is the nucleotide type (A, C, G, or U). The last column is the number of its base-pairing partner. When the base-pairing partner is 0, it means that the nucleotide is not base-paired (i.e., equivalent to the “.” in the dot-bracket notation).

The `PS` corresponds to the postscript format, `JPG` to the Joint Photographic Experts Group format, and `PDF` to the Portable Document Format. These three formats are images that can be visualized directly (Figure 30.2c). We produce

MC-Fold results

The following structures were obtained using the following parameters:

- Job Name: Rat_28S_rRNA_loop_E
- Sequence Name: Rat 28S rRNA loop E
- Sequence: GGGUGCUCAGUACGAGAGGAACCGCACCC

Showing: 1 to 10 of 50 entries

Show entries

Search:

First Previous 1 2 3 4 5 Next Last

Rank	Structure	Energy (kcal/mol)	Coaxial Stacking	Actions
	GGGUGCUCAGUACGAGAGGAACCGCACCC			
1	((((((((.-(((.-.))))))))))))	-28.38	0.00	
2	((((((((.-(((.-.))))-))))))	-27.48	0.00	
3	(((((((-.-(((.-.))))-.))))))	-27.25	0.00	
4	(((((((((-.-.))((.-.))))))	-27.17	-0.54	
5	((((((((((-.-.))))))))))	-27.15	0.00	
6	((((((((((-.-.))))))))))	-26.99	0.00	
7	(((((((-.-(((.-.))))))))))	-26.94	0.00	
8	(((((((-.-(((.-.))))))))))	-26.88	0.00	
9	((((((((.-.))))((.-.))))))	-26.72	-0.54	
10	(((((((-.-.))))((.-.))))))	-26.66	-0.54	

Rank Structure Energy (kcal/mol) Coaxial Stacking Actions

Showing: 1 to 10 of 50 entries

Show entries

Search:

First Previous 1 2 3 4 5 Next Last

[Download Results](#) [Download Trace](#)

(a)

1 G 29

2 G 28

3 G 27

4 U 26

5 G 25

6 C 24

7 U 23

8 C 22

9 A 21

10 G 0

11 U 20

12 A 19

13 C 18

14 G 17

15 A 0

16 G 0

17 A 14

18 G 13

19 G 12

20 A 11

21 A 9

22 C 8

23 C 7

24 G 6

25 C 5

26 A 4

27 C 3

28 C 2

29 C 1

30 C 0

Rat 28S rRNA loop...

15

A

G

G

A

C

A

A

G

U

G

10

G

A

20

A

C

C

C

U

C

C

G

5

G

C

25

U

A

G

C

G

C

G

C

dG = -28.38

>Rat_28S_rRNA_loop_E

GGGUGCUCAGUACGAGAGGAACCGCACCC

((((((((.-(((.-.)))))))))))) -28.38 (+0.00)

(((((((-.-(((.-.))))-)))))) -27.48 (+0.00)

(((((((-.-(((.-.))))-.)))))) -27.25 (+0.00)

(((((((-.-(((.-.))))((.-.)))))) -27.17 (-0.54)

((((((((((-.-.)))))))))) -27.15 (+0.00)

(((((((-.-(((.-.)))))))))) -26.99 (+0.00)

(((((((-.-(((.-.)))))))))) -26.94 (+0.00)

...

[40 lines]

...

(((((((-.-(((.-.)))))))))) -24.78 (+0.00)

(((((((-.-(((.-.)))))))))) -24.72 (+0.00)

(((((((-.-(((.-.)))))))))) -24.70 (+0.00)

(((((((-.-(((.-.)))))))))) BP >= 50%

(((((((-.-(((.-.)))))))))) BP >= 55%

(((((((-.-(((.-.)))))))))) BP >= 60%

(((((((-.-(((.-.)))))))))) BP >= 65%

(((((((-.-(((.-.)))))))))) BP >= 70%

(((((((-.-(((.-.)))))))))) BP >= 75%

(((((((-.-(((.-.)))))))))) BP >= 80%

(((((((-.-(((.-.)))))))))) BP >= 85%

(((((((-.-(((.-.)))))))))) BP >= 90%

(((((((-.-(((.-.)))))))))) BP >= 95%

(((((((-.-(((.-.)))))))))) BP >= 100%

(b)

(c)

(d)

the 2D structure images using the `plot_rna` program of the CONTRAfold package [16].

Two buttons at the bottom of the MC-Fold results page allow for downloading the results in ASCII format (Figure 30.2d).

- 9) **CLICK** the `Download Results` button at the bottom of the page

The results contain the 50 structures predicted by MC-Fold, as well as a pattern of dot-bracket symbols at different percentages of the suboptimal space, from 100% to 50% of the free energy of the MFE structure. The parentheses at a given percentage mean that the respective nucleotides are paired in this percentage of the predicted structures. For instance, the pattern

```
.(((.....))). BP >= 100%
```

indicates that the nucleotide regions 2–5 and 25–28 are predicted to be paired in all structures. Here, nt 2–5 are always paired with nt 25–28, but this may not always be the case. Thus, an open parenthesis in a pattern allows for base pairing with different partners. A dot in a pattern of a given percentage means that the nucleotide is unpaired in that or a greater percentage of the predicted structures. The `Download Trace` option returns the complete execution trace, which includes the results at the end.

After returning to the Dashboard, the `Download` action allows you to download the files associated with the job. The download file is an archive in `.tgz` (tar) format, which can be read on most computers. As indicated above, `Clone params` redirects you to the MC-Fold tab with the fields pre-filled. The `Trace` action allows you to return to the Trace page. Finally, the `Delete` action removes the job from the Dashboard.

Algorithm Options

You can open the `Algorithm` section of the MC-Fold page to access algorithmic options. If you believe your structure may contain a pseudoknot, then activate this option, which is inactivated by default.

- 10) **CLICK** the `Consider H-type Pseudoknots` check box

If activated, the `pseudoknot` option increases the execution time as more calculations are needed to check and add stems that form pseudoknots.

- 11) **CHANGE:** Return the best 50 structures

This will modify the number of output structures. You can change the default 50 structures to 1, 10, 20, 100, or 1000. As a rule of thumb, output more

Figure 30.2 Inspect MC-Fold results. (a) The structures predicted by MC-Fold are displayed in the dot-bracket representation and sorted by energy. For each structure, different actions are available: `Display`, `Submit to MC-Sym`, and `Edit`. (b) CT-formatted representation of a secondary structure. The first column is the nucleotide number, the second

is the nucleotide type (A, C, G, or U), and the third is the number of its base-pairing partner (when this number is 0, it means that the nucleotide is not base-paired). (c) Planar representation of a secondary structure. (d) MC-Fold result file obtained by clicking on the `Download Results` button.

structures for longer sequences. Changing the number of output structures has no additional computational cost.

- 12) **CHANGE:** Explore the best 30% sub-optimal structures

This will modify the fraction of the conformational space explored by MC-Fold. You can change the default 30% to 5%, 10%, 15%, or 50%. Reducing or increasing the fraction of the conformational space explored reduces or increases computational cost, respectively.

Caution

Execution times grow exponentially with the length of the sequence. The G/C content of the sequences is a critical aspect. For a G/C content of 50%, execution times do not vary much for sequences of 50, 100, 150, or 200 nt, with execution times of around 3, 20, 40, or 80 min, respectively. When the G/C content approaches 75%, then the execution times with exploration percentages over 5% grow more rapidly. For example, a 200-nt sequence with 75% G/C and an exploration percentage of 50% takes about 80 min. A 150-nt sequence with 80% G/C and exploration percentages over 5% takes over 12 h.

Alternatively, for long sequences, use MC-Fold2 instead of MC-Fold. A 200-nt sequence with the default parameters in MC-Fold2 takes approximately 20 min. MC-Fold2 runtimes are proportionally subject to G/C content, as well as to two *Algorithm* parameters: the number of solutions and the Waterman-Byers criterion.

Protocol 2b: Predict 2D Structures of Single RNA Sequences Using Structural Data

An active or ligand-bound *in vivo* structure might not always correspond to the MFE structure. Identifying such a structure in a set of suboptimal structures is often impossible without experimental data. Most of the experiments listed in Table 30.1 produce data that can be easily interpreted as and converted into base-pairing constraints, which can then be used to help MC-Fold to select predictions that satisfy these constraints. This is particularly useful for modeling long RNA sequences.

A particularly difficult RNA sequence for all RNA 2D prediction tools is the 5S rRNA of *Escherichia coli*. Here, we show how DMS probing data can help MC-Fold to predict 2D structures resembling those that were obtained by X-ray crystallography (Protein DataBank (PDB) code 2AW4).

To fold a single RNA sequence using experimental data (Figure 30.3):

- 1) **CLICK** the MC-Fold tab
- 2) **TYPE** (under “Create New Project”) or **SELECT** the name of your project in the Project menu (e.g., “Protocol 2b”)
- 3) **TYPE** the name of your sequence in the Name field
- 4) **TYPE** or **PASTE** the RNA sequence into the Sequence field
- 5) **OPEN** and **FILL** the Single-Stranded Chemical/Enzymatic Reactivity Data option

Dashboard
MC-Fold
MC-Fold2(beta)
MC-Sym
MC-Cons
MC-Fold Batch

General

Project: Create New Project Protocol 2b

Name:
E. coli 5S rRNA

Sequence (5'→3'):
UGCCUGGCGGCCUUAGCGCGGUGGUCCACCUGACCCCAUGCCGAACUCAGAAGUGAAACGCCGAGCGCCGAUGGUAGUGU
GGGGUCUCCCCAUGCGAGAGUAGGGAACUGCCAGGCAU

Annealed Sequence (5'→3') (optional):

Algorithm

Structural Constraints

Single-Stranded Chemical/Enzymatic Reactivity Data

High Reactivity Mask:
.....x.....x.....

Medium Reactivity Mask:
.....x.....xx.....x.....

Low Reactivity Mask:

Submit Reset

Figure 30.3 2D structure prediction using structural data. To help MC-Fold in selecting predictions that satisfy constraints obtained from experimental data, one can provide these constraints as parameters through the Single-Stranded Chemical/Enzymatic Reactivity Data section.

In the example, we used:

Name	E. coli 5S rRNA
Sequence (5' to 3')	UGCCUGGCGGCCUUAGCGCGGUGGUCCACCUGACCCCAUGCCGAACUCAGAAGUGAAACGCC (120 nt) GUAGCGCCGAUGGUAGUGUGGGGUCUCCCCAUGCGAGAGUAGGGAACUGCCAGGCAU
High reactivity mask	x.....x..... (14 dots; 1 x; 29 dots; 1 x; 75 dots)
Medium reactivity mask	x.....xx.....x..... ..x..... (38 dots; 1 x; 12 dots; 2 x; 5 dots; 1 x; 6 dots; 1 x; 54 dots)

Here, the masks were derived from DMS constraints. A dot indicates that no constraint is applied to the nucleotide (no DMS signal). The “x” indicates DMS reactivity, that is, the nucleotide is assumed to be unpaired. Three different masks (high, medium, and low reactivity) can be used for different strengths of DMS signals.

- 6) **CLICK** the **SUBMIT** button to send your job to the MC-Fold server

The folding of this 120-nt sequence takes somewhat more than 25 min. Although DMS data were used, no predicted structure matches perfectly that obtained by X-ray crystallography of *E. coli* 5S rRNA at 3.5 Å resolution. Nevertheless, the predicted structures generated by MC-Fold are among the best one can get computationally for this RNA.

Exercises

We propose the identification of the yeast tRNA-ASP crystal structure using high and medium reactivity data generated by the selective 2'-hydroxyl acylation and primer extension (SHAPE) experimental method [8]. The corresponding Name, Sequence, and High and Medium reactivity mask fields are:

Name	Yeast tRNA-ASP
Sequence	UCCGUGAUAGUUUAAUGGUCAGAAUGGGCGCUUGUCGCGGCCAGAUCCGGGUCAAUU
(5' to 3')	CCCCGUCGCGGAGCCA
	(75 nt)
High	xx.....xxxx.....x.....
reactivity	
mask	(17 dots; 2 x; 13 dots; 4 x; 17 dots; 1 x; 21 dots)
Medium	xx..x.....xx.....xx...
reactivity	
mask	(15 dots; 2 x; 2 dots; 1 x; 16 dots; 2 x; 16 dots; 2 x; 19 dots)
X-ray	((((((((.....))))))(((.....))))))...((((.....))..
structure	)))))))))...))
(PDB file	(7 open; 2 dots; 5 open; 6 dots; 5 close; 6 open; 7 dots; 6 close; 3 dots; 6 open; 3 dots;
2TRA)	1 close; 2 dots; 12 close; 4 dots)

Question 1: Do you find the X-ray structure among the structures predicted by MC-Fold?

Compare the MFE structure and the one corresponding to the X-ray crystal structure, if among your solutions:

Question 2: What is the free energy of the MFE structure?

Question 3: What is the free energy of the X-ray structure?

Question 4: What is the Matthews correlation coefficient of the MFE structure?

Question 5: Do you find the structure corresponding to the X-ray crystal structure if you omit any reactivity data?

Expected Results

Computing the results should take less than 30 min.

Answer 1: Yes, the structure corresponding to the X-ray crystal structure ranks eighth. You can easily find this out by copy and paste of the X-ray structure dot-bracket notation into the `Search` field at the top-right corner of the result table (empty the search field to get back to the original results table).

Answer 2: The MFE structure (rank 1) has a free energy of -59.23 kcal/mol, including -1.45 for coaxial stacking.

Answer 3: The eighth structure has a free energy of -57.92 kcal/mol, including -0.89 for coaxial stacking.

Answer 4: To answer question 4, visualizing the MFE and X-ray structures helps. For both structures, **CLICK** the `Inspect` button of `MC-FOLD Yeast_tRNA-ASP`, and then the `Display` button of the structures ranked first and eighth. For each, **CLICK** either `PS`, `JPG`, or `PDF`. Note that you can print the images if your computer is linked to a printer.

Then, looking at both structures, count how many of the 24 bp in the X-ray structure are identical to or different from those in the MFE structure. Identical ones are True Positives (*TP*), those base-paired in the MFE but not in the X-ray structure are False Positives (*FP*), and those base-paired in the X-ray but not in the MFE structure are False Negatives (*FN*). The answer is 18 *TP*, 6 *FP*, and 6 *FN*. This gives a Matthews correlation coefficient of $\text{SQRT}(TP/(TP + FP) \times TP/(TP + FN)) = \text{SQRT}(18/24 \times 18/24) = 3/4 = 0.75$ (SQRT = square root).

Answer 5: To answer this question, you must use Protocol 2a to fold the tRNA-ASP. Then use the `Search` field at the top-right corner of the result table to see if the X-ray structure dot-bracket notation is found among the structures you generated with Protocol 2a. Folding the tRNA-ASP without any structural data can be very time-consuming, in the range of hours. Unfortunately, even if you generate 1000 structures for this tRNA, you will not find the X-ray structure. There are too many possible structures with better energies. You would have to generate many more than 1000 to have it included. This indicates that for some RNAs, of even just 70 nt, folding without structural data or inspirational hints is simply futile!

Protocol 2c: Add Non-canonical Base Pairs to a 2D Structure Obtained Using Third-Party Software

Sometimes, one has obtained 2D structures using one of the numerous RNA 2D structure prediction programs [15, 17–22] (see also Chapters 26, 27, and 29). However, these programs do not predict non-canonical base pairs (except for G:U wobble pairs), making it difficult to predict authentic 3D structures. Instead of refolding the sequence from scratch using MC-Fold, it is possible and legitimate to add non-canonical base pairs to the 2D structure predicted by the third-party software. For this purpose, we superimpose the non-canonical base pairs onto the “imported” 2D structural constraints that are enforced by a structural mask. MC-Fold can thus focus its computation on the addition of the non-canonical base pairs.

To add non-canonical base pairs to a 2D structure obtained with third-party software (Figure 30.4):

- 1) **CLICK** the `MC-Fold` tab
- 2) **TYPE** or **SELECT** the name of your project in the `Project` menu (e.g., “Protocol 2c”)
- 3) **TYPE** the name of your sequence into the `Name` field
- 4) **TYPE** or **PASTE** the RNA sequence into the `Sequence` field
- 5) **OPEN** and **FILL** the `Structural Constraints` option

In the example, we used:

Name	Rat 28S rRNA loop E (Mfold)
Sequence (5' to 3')	GGGUGCUCAGUACGAGAGGAACCGCACCC (29 nt)
Structural mask	(((((.....)))))) (6 open; 17 dots; 6 close)

Here, the mask is the MFE structure returned by the `mfold` software for the rat 28S rRNA loop E. The large number of dots in the dot-bracket notation indicates a large loop.

- 6) **CLICK** the `Convert unpaired to unconstrained` check box, which allows MC-Fold to add base pairs in the loop region
- 7) **CLICK** the `SUBMIT` button to send your job to the MC-Fold server

Once the job status is `Done`, you have access to the results, which you can compare with those gained in Protocol 2a. You will see that the dot-brackets are almost the same as the MC-Fold predictions obtained without the third-party software mask. All predicted structures contain the six base pairs predicted by the third-party software and enforced by the mask.

In MC-Fold, the `Structural Mask` is applied at the end of computation as a filter of the solutions. This filter ensures that the opened and closed parentheses are satisfied regardless of the base pairs suggested by the mask. The mask can even be unbalanced. In MC-Fold2, the mask is enforced and considered during the computation. Also, the notion of a `Balanced Mask` has been introduced, where the base pairs suggested by the mask are enforced during computation. The parentheses in this mask must be balanced.

Alternative Procedure

If you do not get a structure by applying MC-Fold (Protocols 2a–c), then select MC-Fold2 instead of MC-Fold in STEP 1. When using Protocol 2c, paste the third-party 2D structure in the `Balanced Mask` of MC-Fold2 in STEP 5. At this moment, applying the `Balanced Mask` with MC-Fold2 is the only way to theoretically and practically fulfil the intent of Protocol 2c.

Dashboard
MC-Fold
MC-Fold2(beta)
MC-Sym
MC-Cons
MC-Fold Batch

General

Project:
Create New Project
Protocol 2c

Name:
Rat 28S rRNA loop E (mfold)

Sequence (5'→3'):
GGGUGCUCAGUACGAGAGGAACCGCACCC

Annealed Sequence (5'→3') (optional):

Algorithm

Structural Constraints

Structural Mask:
(((((((.....))))))

Convert Unpaired to Unconstrained: ☒

Single-Stranded Chemical/Enzymatic Reactivity Data

Submit
Reset

Figure 30.4 Adding non-canonical base pairs to a classical 2D structure obtained with a third-party software. Simply paste the structure obtained with a third-party software (here mfold) as a mask in the Structural Constraints section.

Protocol 2d: Predict 2D Structures of RNA Duplexes

An RNA duplex is formed from two single-stranded RNAs. In MC-Fold, one is called the *Sequence*, and the other is the *Annealed sequence*.

To fold an RNA duplex (Figure 30.5):

- 1) **CLICK** the *MC-Fold* tab
- 2) **TYPE** the name of your duplex into the *Name* field
- 3) **TYPE** or **PASTE** the first strand of the RNA duplex into the *Sequence* field (5' to 3')
- 4) **TYPE** or **PASTE** the second strand of the RNA duplex into the *Annealed sequence* field (5' to 3')

In the example, we used:

Name	R2 retrotransposon internal loop
Sequence (5' to 3')	GUGAAGCCCGU (11 nt)
Annealed sequence (5' to 3')	CGGAGGACACU (11 nt)

- 5) **CLICK** the *SUBMIT* button to send your job to the MC-Fold server
- Once the job status is *Done*, you have access to the results. At present, duplex dot-brackets cannot be converted to planar representation images, and thus cannot be visualized on the MC-Tools web site. Also, RNA duplexes cannot be submitted to the server for 3D modeling.
- The result of our example shows that, although the internal loop of this retrotransposon contains a 4 × 4 interior loop, the nucleotides in the loop form non-canonical base pairs.

Exercise

We propose, as an exercise, to evaluate the energy of hybridization between the seed of a microRNA (miRNA), miR-124, and a targeted messenger RNA (mRNA). Consider the mRNA target sequence 5'-GUGGCCUU and the miR-124 seed (nt 2–8) sequence 5'-AAGGCAC. Use MC-Fold to evaluate the different hybridization possibilities.

The suggested *Name*, *Sequence*, and *Annealed Sequence* are:

Name	mRNA miR-124 duplex
Sequence (5' to 3')	GUGGCCUU (8 nt)
Annealed sequence (5' to 3')	AAGGCAC (7 nt)

The screenshot shows the MC-Fold2(beta) web interface. At the top is a navigation bar with tabs: Dashboard, MC-Fold, MC-Fold2(beta), MC-Sym, MC-Cons, and MC-Fold Batch. The MC-Fold2(beta) tab is selected. Below the navigation bar is a 'General' section with a dropdown arrow. It contains a 'Project' field with a dropdown menu showing 'Create New Project' and 'Protocol 2d'. Below this is a 'Name:' label followed by a text input field containing 'R2 retrotransposon internal loop'. Then is a 'Sequence (5'→3):' label followed by a text input field containing 'GUGAAGCCCCU'. Below that is an 'Annealed Sequence (5'→3') (optional):' label followed by a text input field containing 'CGGAGGACACU'. At the bottom of the 'General' section are three expandable sections: 'Algorithm', 'Structural Constraints', and 'Single-Stranded Chemical/Enzymatic Reactivity Data'. At the very bottom are 'Submit' and 'Reset' buttons.

Dashboard MC-Fold MC-Fold2(beta) MC-Sym MC-Cons MC-Fold Batch

General

Project: Create New Project Protocol 2d

Name:
R2 retrotransposon internal loop

Sequence (5'→3):
GUGAAGCCCCU

Annealed Sequence (5'→3') (optional):
CGGAGGACACU

Algorithm

Structural Constraints

Single-Stranded Chemical/Enzymatic Reactivity Data

Submit Reset

Figure 30.5 2D structure prediction of RNA duplexes. Enter the two single-stranded sequences supposed to form a duplex in the Sequence and Annealed Sequence windows, respectively (here the R2 Retrotransposon internal loop).

Computing the results should take less than 1 min.

Question: According to your results, what would be the hybridization structure and why?

Answer: The MFE structure “(. ((((((-----)))))))” bulges the second U in the mRNA, with an energy of -17.08 kcal/mol. This structure includes an A:G mismatch between nucleotide 3 of the mRNA and nucleotide 6 of the miRNA sequence. It is favored over the Rank 2 structure by -1.22 kcal/mol. Although the Rank 2 structure “(((. (((-----))))))” contains only canonical Watson-Crick base pairs (G4 of the mRNA sequence is bulged in the Rank 2 structure), non-canonical A:G base pairs are often observed in RNA structures, owing to the augmented pairing and stacking potentials of G residues; because of these properties, G residues are less often observed in single-nucleotide bulges. Therefore, the MFE structure would be our best guess for the hybridization structure of miR-124, if it was happening at this precise site.

Protocol 2e: Predict 2D Structures of Several RNA Sequences

It is possible to run multiple MC-Fold jobs without submitting them one by one. To do so, one can use the MC-Fold Batch tool.

- 1) **CREATE** a text file using Microsoft WordPad or notepad containing the sequences in a FASTA format and set, after each sequence and one per line, the individual parameters using the following format:

```
[parameter_name] <tab> [value]
```

- 2) **CLICK** the MC-Fold Batch tab
- 3) **SELECT** or create a new project in the Project menu (e.g., “Protocol 2e”)
- 4) **CHOOSE** the version of MC-Fold
- 5) **UPLOAD** the input file
- 6) (optional) **CHANGE** the Algorithm settings
- 7) **CLICK** the **SUBMIT** button to send the job to the MC-Fold server

All the jobs will then appear in the Dashboard as if they were submitted individually.

In the example, we used:

MC-Fold version 1.6.0

Algorithm Return the best: 1000 structures
options

File content:

>ASN

GACUCCAUGGCCAAGUUGGUUAAAGCGUGCGACUGUUAACGCAAGAUCGUGAGUUAACCCUCACUGGGGUGGCCA

singlehigh XX.XX.....XXX.....X.....

>GLY

GCGCAAGUGGUUAGUGGUAAAAUCCAACGUUGCCAUCGUUGGGCCCCGGUUCGAUUCGGGCUUGCGCACCA

singlehigh X.....X.X.....XXX.....

```

> ILE
GGUCUCUUGGCCAGUUGGUUAAAGGCACCGUGC UAAUAACGCGGGAUCAGCGGUUCGAUCCCGUAGAGACCACCA
singlehigh      .....x.....xx..xx.....xxx.....x.....

> LYS
UCCUUGUUGAGCUCAGUUGGUAGAGCGUUCGGCUUUUAACCGAAAUGUCAGGGGUUCGAGCCCCUUGAGGAGGCCA
singlehigh      .....xx..x.....xxx.....x.....

> MET
GCUUCAGUAGCUCAGUAGGAAGAGCGUCAGUCUCAUAAUCUGAAGGUUCGAGAGUUCGAACCUCCUGGAGCACCA
singlehigh      .....x.....xxx.....x.....

> THR
GCUUCUAUUGGCCAAGUUGGUAAAGGCGCCACACUAGUAAUGUGGAGAUCAUCGGUUCAAAUCCGAUUGGAAGCACCA
singlehigh      .....xx..x.....xxx.....x.....

> TRP
GAAGCGGUGGCUCAAGGUAGAGCUUUCGACUCCAAUUCGAAGGGUUCGAGGUUCAAUUCCUGUCCGUUUCACCA
singlehigh      .....x.....x..x.....xxx.....x.....

> ALA
GGGCGUGUGGCGUAGUCGGUAGCGCGCUCUCCUAGCAUGGGAGAGGUCUCCGGUUCGAUUCGGACUCGUCCACCA
singlehigh      .....x.....x..x.....xxx.....x.....

> ARG
UUCCUCUGUGGCCAAUGGUCACGGCGUCUGGCUACGAACCAGAAGAUUCCAGGUUCAAGUCCUGCGGGGAAGCCA
singlehigh      .....x.....x..x.....xxx.....x.....

> ASP
UCCGUAUAGUUUAAUGGUCAGAAUGGGCGCUUGUCGCGUGCCAGAUCCGGGUUCAAUUCCCCGUCGCGAGCCA
singlehigh      .....x..x.....xxx.....

> GLU
UCCGAUAUAGUGUAACGGCUAUCACAUACGCUUUCACCGUGGAGACCGGGUUCGACUCCCCGUUUCGGAGCCA
singlehigh      .....x.....xxx.....

> HIS
GGCCAUCUUGAUUAGUGGUUAGUACACAACAUUGUGGCUUGUUGAAACCCUGGUUCGAUUCUAGGAGGUGGCACCA
singlehigh      .....x..xx.....xxx.....

> PHE
GCGGAUUUAGCUCAGUUGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUGCAUCCACAGAAUUCGCACCA
singlehigh      .....xx.....xxx.....

> VAL
GGUUUCGUGGUCUAGUCGGUUAUGGCAUCUGCUUAAACACGCAGAACGUCCCCAGUUCGAUCCGGGCGAAAUCACCA
singlehigh      .....x.....x..xx.....xxx.....x.....

```

Parameters

Different parameters are accepted and may vary depending on the version of MC-Fold that is used; get the documentation by clicking on the `Help` button at the bottom of the MC-Fold Batch page.

Protocol 3: Produce a Consensus Structural Assignment of Multiple-Sequence Data

When we have a set of related RNA sequences (e.g., different tRNAs from the same organism, or group I introns from different organisms), then it is reasonable to consider the structural information encoded in these multiple sequences. MC-Cons computes a structural assignment by determining for each sequence

the structure among all that maximizes the overall sum of pairwise structural similarities.

As an example, we will produce here a consensus structural assignment of the results obtained from Protocol 2e above:

- 1) **CLICK** the `MC-Cons` tab

This action opens a new MC-Cons page, where you can upload MC-Fold results (Figure 30.6a)

- 2) **SELECT** or create a new project using the `Project` menu (e.g., “Protocol 3”)
- 3) **ENTER** a name for the MC-Cons job in the `Name` box
- 4) **SELECT** MC-Fold’s results by either:
- 5) **CLICKING** the `Choose Files` button and select the previously downloaded files

This operation must be done only after your MC-Fold’s result files are saved on your computer (see Protocol 2a, Action 9).

and/or:

- 6) **SELECTING** the MC-Fold jobs whose results you want to use as the input for MC-Cons by clicking the corresponding project checkbox or individual job checkboxes in the table

These two ways of selecting MC-Fold results in MC-Cons are equivalent, although the second approach is much easier as it does not require downloading the results on your computer. By selecting a project *via* clicking its box, you select all MC-Fold’s jobs related to this project, and you can then:

- 7) **ADJUST** the algorithmic options (for this example, as tRNA sequences are relatively long, we will consider the top 1000 structures)
- 8) **CLICK** the `SUBMIT` button to send your job to the MC-Cons server

You will be directed to the Dashboard, where you will see the new MC-Cons job. As for the other MC-Tools, the Dashboard informs about the status of the job. The time required by MC-Cons to compute the structural assignment for these 14 tRNAs is a few hours.

When the status of the job is `Done`, three actions are possible: `Inspect`, `Download`, and `Delete`.

- 9) **CLICK** the `Inspect` button to access the structures predicted by MC-Cons

For each sequence, the rank and the energy of the structure among the original structures predicted by MC-Fold is given (column `Index`, and `Energy`), and three actions for 2D structures are provided: `Display`, `Submit` (to MC-Sym), and `Edit` (the MC-Sym parameters), as well as `Go back to MC-Fold` (a link to the original MC-Fold results) (Figure 30.6b). As you can see, for each tRNA sequence, MC-Cons was able to select among the numerous 2D structures one 2D structure that optimizes its overall alignment with the 2D structures of all sequences. Interestingly, each selected 2D structure has the typical four-stem cloverleaf structure of tRNAs (Table 30.2). MC-Cons selected the Rank 20 structure for tRNA-ASP, whereas the tRNA-ASP X-ray crystal structure is found at Rank 25.

Dashboard MC-Fold MC-Fold2(beta) MC-Sym MC-Cons MC-Fold Batch

Input

Project:

Name:

Select MARNA-Formatted MC-Cons file(s) and/or MARNA-Formatted MC-Fold result file(s):

- Make sure that your input file contains only printable ASCII characters
- Do not use Microsoft Word to edit the script, use WordPad instead

No file chosen

and/or select MC-Fold results from project(s):

☒ Protocol 2e ☐ Protocol 2d ☐ Protocol 2c ☐ Protocol 2b ☐ Protocol 2a

Protocol 1e

Created on: 2012.11.03 16:20
Last Activity on: 2012.11.03 20:21

Showing 1 to 10 of 14 entries
Show entries

Search:

Select	Job Name	Start	End	Actions
☒	ILE (77 nts, 1000 sol.)	2012.11.03 16:20	2012.11.03 16:39	P E B I S
☒	GLY (73 nts, 1000 sol.)	2012.11.03 16:20	2012.11.03 16:43	P E B I S
☒	TRP (75 nts, 1000 sol.)	2012.11.03 16:20	2012.11.03 16:53	P E B I S
☒	ARG (76 nts, 1000 sol.)	2012.11.03 16:20	2012.11.03 16:59	P E B I S
☒	ASN (77 nts, 1000 sol.)	2012.11.03 16:20	2012.11.03 17:07	P E B I S
☒	ASP (75 nts, 1000 sol.)	2012.11.03 16:20	2012.11.03 17:20	P E B I S
☒	PHE (76 nts, 1000 sol.)	2012.11.03 16:20	2012.11.03 17:24	P E B I S
☒	GLU (75 nts, 1000 sol.)	2012.11.03 16:20	2012.11.03 17:27	P E B I S
☒	THR (76 nts, 1000 sol.)	2012.11.03 16:20	2012.11.03 17:43	P E B I S
☒	MET (76 nts, 1000 sol.)	2012.11.03 16:20	2012.11.03 17:55	P E B I S

Showing 1 to 10 of 14 entries
Show entries

Search:

First Previous 1 2 Next Last

Protocol 2d

Protocol 2c

Protocol 2b

Protocol 2a

Algorithm (Options)

Use at most gaps

Tolerate at most unbalanced positions

Consider only the top structures

(a)

Figure 30.6 2D structure prediction using multiple-sequence data. (a) The job submission page for MC-Cons is shown. (b) After the job is done, the MC-Cons Inspect action opens the results page displaying the consensus structure for each sequence.

Table 30.2 MC-Cons selection of yeast tRNA 2D structures.

Amino acid Energy Rank	Sequence Reactivity mask MC-Cons 2D structure prediction
ASN -55.68 112	GACUCCAUGGGCCAAGUUGGUUAAGGCGUGCGACUGUUAUACGAAGCAGUGAGUUAACCCUCACUGGGGUGCCCA xx..xx......xxx......x..... (((((((.....))))((((.....)))).....((((.....)))))).....
GLY -62.65 6	GCGCAAGUGGUUAGUGGUAAAUAUCCAAACGUUGCCAUUCGUUGGGCCCGGCUUCGAUCCGGGCUUGCGCACCA x.....x..x......xxx..... (((((((.....))))((((.....)))).....((((.....)))))).....
ILE -58.96 33	GGUCUCUUGGGCCAGUUGGUUAAGGCACCGUGCUAAUAACGCGGGGAUCAGCGGUUCGAUCCCGCUAGAGACCACCA x.....xx..xx......xxx......x..... (((((((.....))))((((.....)))).....((((.....)))))).....
LYS -59.93 15	UCCUUGUAGCUCAGUUGGUAGAGCGUUCGGCUUUUAACCGAAAUUGCAGGGGUUCGAGCCCCCUAUGAGGAGCCA xx..x......xxx......x..... (((((((.....))))((((.....)))).....((((.....)))))).....
MET -58.65 130	GCUUCAGUAGCUCAGUAGGAAGAGCGUCAGUCUAUAUCUGAAGGUCGAGUUCGAAACCUCCUCGAGGAGCACCA x.....x......xxx......x..... (((((((.....))))((((.....)))).....((((.....)))))).....
THR -57.11 23	GCUUCUAGGGCCAAGUUGGUUAGGCGCCACACUAGUAAUUGUGGAGAUCAUCGCGUUAACCAUCCGAUUGGAAGCACCA xx..x......xxx......x..... (((((((.....))))((((.....)))).....((((.....)))))).....
TRP -59.85 13	GAAGCGGUGGCUCAAUGGUUAGACUUCGACUCCAAUACGAAGGGUUCAGGUUCAAUCCGCUUCACCA x.....x..x......xxx......x..... (((((((.....))))((((.....)))).....((((.....)))))).....

(continued overleaf)

Table 30.2 (Continued)

Amino acid Energy Rank	Sequence Reactivity mask MC-Cons 2D structure prediction
ALA -61.01 130	GGGCGUGUGGCGUAGCGGUAGCGCGCUCUUAAGCAUGGGAGAGGUCUCCGGUUCGGAUUCGGACUCGUCACCA x.....x.....x.....xxxx.....x..... (((((((.....))))((((.....))))((((.....))))((((.....))))((((.....))))
ARG -59.44 267	UUCUCUGUGGCCAAUGGUCACGGCGUCUGGCUACGAACCAAGAUUCCAGGUUAGUCCUGGCGGGAACCA x.....x.....x.....xxx.....x..... (((((((.....))))((((.....))))((((.....))))((((.....))))((((.....))))
ASP -58.15 20 X-ray structure (Rank 25)	UCCGUGAUAGUUUAUGGUCAGAAUGGCGCUUGUCGCGUGCCAGAUCCGGGUUCAAUUCCCGUCGCGGAGCCA x.....x.....xxx..... (((((((.....))))((((.....))))((((.....))))((((.....))))((((.....)))) (((((((.....))))((((.....))))((((.....))))((((.....))))((((.....))))
GLU -57.92 269	UCCGAUAUAGUGUAACGGCUAUACAUCACGCUUUCACCGUGGAGACCGGGGUUCGACUCCCGUAUCGGAGCCA x.....xxx..... (((((((.....))))((((.....))))((((.....))))((((.....))))((((.....))))
HIS -54.59 386	GGCCAUCUUAUAGUAGUAGUACACAACAUAUGUGCGUGUUGAAACCCUGGUUCGAUUCUAGGAGGUGGCACCA x.....xx.....xxx..... (((((((.....))))((((.....))))((((.....))))((((.....))))((((.....))))
PHE -61.58 339	GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUUGGAGGUCCUGUUCGUAUCCACAGAAUUCGACCA xx.....xxx..... (((((((.....))))((((.....))))((((.....))))((((.....))))((((.....))))
VAL -56.67 317	GGUUUCGUGGUCUAGUCGUUAUGGCAUCUGCUUAACACGCAGAACGCCCGCCAGUUCGAGUCCUGGGCGAAUACCA x.....x.....xx.....xxx..... (((((((.....))))((((.....))))((((.....))))((((.....))))((((.....))))

To generate a 3D structure from a 2D structure:

First, find any MC-Fold results in your Dashboard that generated 2D structures, except for RNA duplexes (see Protocol 2d). As an example, retrieve from the above Protocols the results for “Rat_28S_rRNA_Loop_E”, and then:

- 1) **CLICK** the `Inspect` button to access the 2D predicted structures
- 2) **CLICK** under `Actions` the `Submit to MC-Sym` button of any 2D structure to submit it to the MC-Sym server

Here, if we click on the `Rank 1` structure, the button becomes yellow to indicate that this structure has been submitted to the MC-Sym server. If you go back to the Dashboard, you will see the MC-Sym job, with the `Status` field indicating how many 3D structures (PDB files), termed *Solutions*, have been generated so far. 100 *Solutions* are provided when the job is `Done`, which takes only a few seconds. The `Job Name` is “Rat_28S_rRNA_Loop_E_1_d”. The “_1” at the end indicates that this job was for the `Rank 1` 2D structure of the “Rat_28S_rRNA_Loop_E” predicted by MC-Fold and the “_d” indicates that this job was submitted using the `Submit to MC-Sym` button.

The `Actions` available for the results of a MC-Sym job are: `Inspect`, `Get`, `Download`, `Trace`, and `Delete`. The `Get` action downloads the MC-Sym input script to your computer, and the `Download` option does the same for an archive containing all MC-Sym output files, including a PDB file [35] for each 3D structure. The `Trace` action goes to the MC-Sym output page.

- 3) **CLICK** the `Inspect` button to access the predicted 3D structures

The `Inspect` action opens a MC-Sym results page, which offers one action per 3D structure, that is, the `Get` action for downloading the corresponding PDB file to your computer. In our case, 10 of the 100 3D structures are listed on the `Results` page by default. You can change the number of structures per page to up to 100, and navigate through the pages using the menu located at the top and the bottom of the result table (`First` – `Previous` 1 – `Next` – `Last`)

- 4) **CLICK** the `Get` button to download a predicted 3D structure

Once on your computer, the PDB file can be viewed using a 3D structure visualization program, such as PyMOL [36] (Figure 30.7a). The PyMOL “show cartoon” mode is a convenient option for visualizing the backbone of RNAs (Figure 30.7b).

Modify MC-Sym's parameters

After clicking the `Inspect` action of your MC-Fold job displayed on the Dashboard, you may alternatively use the `Edit` button (instead of `Submit to MC-Sym`) to modify the default parameters, before submitting a job to the MC-Sym server. This opens a MC-Sym parameter page, which allows you to change some of the program's parameters. Here, we will focus on the `Explore` subsection, which guides the user to the most frequently changed parameters.

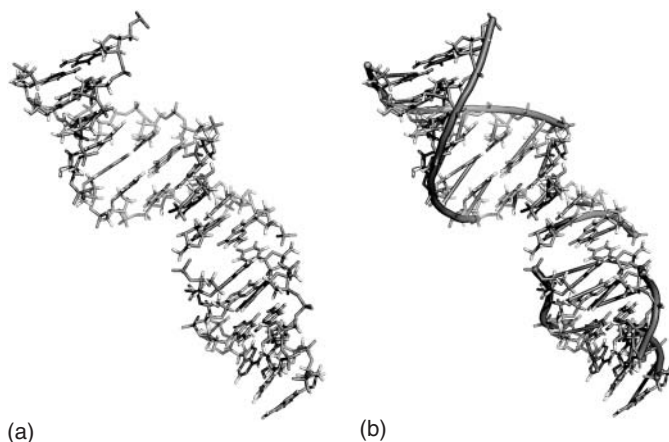


Figure 30.7 3D structure prediction from 2D structures. (a) A 3D structure generated by MC-Sym visualized using PyMOL. (b) Adding the backbone using the “show cartoon” option in PyMOL.

Exploration Method

MC-Sym has two exploration methods: probabilistic (by default) and exhaustive. The probabilistic method makes use of random choices during the exploration of the conformational space, which does not guarantee that all possibilities will be tested, but usually yields satisfactory results. Change the default setting to exhaustive when you have rather small RNAs (<50 nt) and have difficulties to obtain 3D structures with the probabilistic method.

Model Limit

This is the maximum number of 3D structures MC-Sym will generate. If this limit is reached, then MC-Sym stops. The default setting is 100. Adjust this parameter according to the size of the RNA you are modeling. Usually, the larger the RNA is, the more 3D structures one should generate.

Time Limit

This is the maximum time MC-Sym will run. If this limit is reached, then MC-Sym stops, whether 3D structures have been generated or not. The default setting is 30 min. Adjust the time as a function of your RNA. If you do not get any 3D structure within the defined time limit, then increase it. For short RNAs (<50 nt), 30 min is usually long enough, but not always.

Model Diversity

This is the minimum root-mean-square deviation between any two generated structures (default setting 3 Å). The smaller the value is, the more structures will be generated. Adjust this parameter as a function of your RNA and the

number of structures you want. Small values are appropriate for small RNAs, because you will get more slightly different structures. Large values are good for large and multi-branched RNAs, because you may want to explore larger motions of the branches rather than small differences between globally very similar structures.

After adjustment of the four `Explore` parameters, click the `Submit` button at the bottom of the page.

Alternative Procedure

If you need to modify more parameters or want to use your own building strategy, you will need to manually edit an MC-Sym input script. By doing so, you will become an Advanced MC-Sym user. An edited MC-Sym script can be uploaded to the MC-Sym server under the MC-Sym page that is accessible from the Dashboard. As the MC-Sym input scripting language is not described here, the reader is referred to the *MC-Sym* wiki page at www.major.irc.ca/wiki/index.php/MC-Sym as a good starting point.

Protocol 4b: Energy Minimization of 3D Structures

As you can see in Figure 30.7b, the backbone trace is broken at several places. This is due to MC-Sym's discrete backbone building procedure. You can solve this problem by energy minimization. Go back to the MC-Sym's results page to minimize the energy of the first 3D structure:

- 1) (optional) **CLICK** the `Uncheck All` button
- 2) (optional) **SELECT** the check box of `Rat_28S_rRNA_Loop_E_1_d-0001` (the one at the top); you can also select several 3D structures simultaneously
- 3) **OPEN** the `Minimization` subsection
- 4) **CLICK** the `Refine` button

This will submit the 3D structure(s) to the `Refine` server and bring you back to the Dashboard, where you can follow the progress of the job. Minimizing a single Rat 28S rRNA loop E structure is a question of seconds. The minimized 3D structure(s) is (are) now listed in the Dashboard, indicated by `Refine` in the `Tool` field.

- 5) **CLICK** the `Inspect` button of the `Refine` job to access the refined 3D structure(s)
- 6) **CLICK** the `Get` button to download the individual minimized 3D structure to your computer

Use your 3D structure visualization software (here PyMOL) to look at it, and using the “show cartoon” action, see how the backbone is now connected throughout.

Decoys

3D structure collections generated by MC-Sym are called *Decoys*. In the Dashboard, a *Decoy* is linked to at least the following actions: `Inspect`, `Download`, `Trace`, and `Delete`. The `Inspect` action opens a *Decoy* page (Figure 30.8),

Back
This task's base UUID is b9209fb3-7db7-4cc0-8e7b-20690dff52d8

Results

The structures are stored as PDB formatted files. [Click here](#) for more information about the PDB format.

For the post-processes, a selection of structures may apply. You can Select and UnSelect All, Invert a selection, and Select the structures one by one using the check boxes.

Showing 1 to 10 of 100 entries
Show 10 entries Search:

First Previous 1 2 3 4 5 Next Last

Select	Structure	Action
☒	Rat_28S_rRNA_loop_E_1_d-0001	
☒	Rat_28S_rRNA_loop_E_1_d-0002	
☒	Rat_28S_rRNA_loop_E_1_d-0003	
☒	Rat_28S_rRNA_loop_E_1_d-0004	
☒	Rat_28S_rRNA_loop_E_1_d-0005	
☒	Rat_28S_rRNA_loop_E_1_d-0006	
☒	Rat_28S_rRNA_loop_E_1_d-0007	
☒	Rat_28S_rRNA_loop_E_1_d-0008	
☒	Rat_28S_rRNA_loop_E_1_d-0009	
☒	Rat_28S_rRNA_loop_E_1_d-0010	

Select Structure Action

Showing 1 to 10 of 100 entries
Show 10 entries Search:

First Previous 1 2 3 4 5 Next Last

Uncheck All Invert Selection

Download Clone Download CSV

- File Upload
- Minimization
- Analysis
- K-Means Clustering
- Global Geometry
- Distance Restraints
- Hydroxyl Radical Footprinting
- Molecular Envelope
- Excluded Volume
- Renumbering

Figure 30.8 The MC-Sym Inspect page showing a set of MC-Sym 3D structures (*Decoy*). Many decoy actions and tools are available (for details, see text).

where many other tools are available for transforming (e.g., Minimization, Renumbering) and analyzing *Decoy* structures (e.g., K-Means Clustering, Analysis, Root-Mean-Square Deviation (RMSD), Global Geometry, Distance Restraints, Hydroxyl Radical Footprinting, Molecular Envelope, and Excluded Volume). RMSD appears only if a File Upload has been done.

The energy minimization of Protocol 4b is a transforming operation, as it changes the atomic coordinates and the corresponding PDB files of the

structures in the *Decoy*. Renumbering also changes the files of the structures, as it modifies the `Residue Id` (see the PDB format) of the nucleotides.

For the analysis of *Decoys*, one must be familiar with the structure selection functions. Above, when you unchecked all and then clicked the check box of the structure at the top of the list, you made a selection. Selection of structures is an important step if one wants to apply the operations to a subset of structures. You did apply the minimization to a single structure, but you could have selected more than one, and then the minimization would have been applied to the selected structures. A button allows you to `Uncheck All`, which is useful when you want to select only a few structures; clicking `Invert Selection` is useful when you want to apply an action to all but those that are already selected.

Next to the selection buttons, you find buttons to `Download` the *Decoy* (this is the same function as MC-Sym's `Download` action), `Clone`, and `Download CSV` (Comma Separated Values). The `Clone` button creates a new *Decoy* in the Dashboard with the structures that are selected (or all, if all are selected). Cloned *Decoys* have the same name as their master *Decoy* with the letter "C" at the end. This is an important operation because most analysis operations apply to all structures of a *Decoy*. Therefore, to apply an operation to a subset of structures, select the structures, clone the *Decoy*, and then apply the operation to the cloned *Decoy*. Each time an analysis operation is applied, it adds a new column in the *Decoy*. The `Download CSV` action produces a downloadable spreadsheet that includes the values of each column in a format that can be read by Microsoft Excel and R.

Protocol 4c: Partitioning of 3D Structures

The number of 3D structures in a *Decoy* can be high (up to 9999), which makes it impossible to visualize them all. By clustering the structures by their geometrical similarity (`K-Means Clustering`), you identify groups or partitions of structures that "look" alike, and then you can visualize only one representative structure per partition, instead of inspecting all of them. `K-Means Clustering` is thus an important analysis operation, and it differs from the others by generating its results in a subsection.

To partition the structures in *K* groups:

- 1) **CLICK** the `Inspect` button to access the predicted 3D structures of any *Decoy*
- 2) **OPEN** the `K-Means Clustering` subsection
- 3) **CHOOSE** a number of clusters (5 by default)
- 4) **CLICK** the `Submit` button

You will be redirected to the Dashboard, where you will see the `K-Means Clustering` job progression. Let us call the *Decoy* to which we applied the clustering

the Master *Decoy*. When the Status of the clustering job is Done, then it auto-deletes itself and the results are saved in the Master *Decoy*. Now inspect the Master *Decoy* to see the new column called *cluster*, which you can use to sort the *Decoy*. Open the K-Means Clustering Results subsection to see a report of the results. This is where you will get the information about which structures are the centroids of the clusters, that is, those you can visualize to get an idea of how the structures of the same cluster look like.

Protocol 4d: Analyze 3D Structures

There are two types of analysis operations: those that require an input file, and those that do not. The latter are similar to partitioning a *Decoy* using K-Mean Clustering. To analyze a Decoy of 3D structures:

- 1) **CLICK** the Inspect button to access the predicted 3D structures of any *Decoy*
- 2) **OPEN** the Analysis and Global Geometry subsections
- 3) Under Global Geometry, **INSERT** the Target value (estimate) for Radius of Gyration and/or Maximum Width
- 4) **CLICK** one of the buttons in the Analysis subsection (Entropy, Score, P-Score, Amber'99, Volume or Analyze) or in the Global Geometry subsection (Radius of Gyration or Maximum Width)

You will be redirected to the Dashboard, where you will see the progression of the job you just submitted. When the job's Status is Done, then it auto-deletes itself and the results are saved in the Master *Decoy*, where a new column is added for each type of analysis you run. Any column can become the sort criterion of the *Decoy*.

Protocol 4e: Fit Structural Data to 3D Structures

Analysis operations that require input files fit the structural data described in these files to the 3D structures of the *Decoy*. To fit structural data to a *Decoy* of 3D structures:

- 1) **CLICK** the Inspect button to access the predicted 3D structures of any *Decoy*
- 2) **OPEN** the Distance Restraints, Hydroxyl Radical Footprinting, Molecular Envelope, or Excluded Volume subsection
- 3) **CLICK** the Choose File button to locate the structural data file on your computer
- 4) **CLICK** the Submit button

You will be redirected to the Dashboard, where you will see the progression of the job you just submitted. When the job's Status is Done, then it auto-deletes itself and the results are saved in the Master *Decoy*, where a new column is added for each type of structural fit you run. Any column can become the sort criterion of the *Decoy*.

30.4

Troubleshooting

My job is stuck on the Finalizing status	This most likely means that our server encountered an error associated with your job. You can: (i) download the last state of the job by clicking on Download; (ii) the job is too important and you cannot afford to restart it again (e.g., it computed for weeks), then send us an e-mail (mctools@iric.ca) and we'll do our best to help. If you do so, then we will need your username, the server used, the start date, as well as the name you gave to the job; or (iii) Click on Delete and restart the job again.
I deleted a job from the Dashboard by error	Deleted jobs can be retrieved within 10 business days. Send us an e-mail with your username, the server, and the name you gave to the job; we will restore the job on your Dashboard.
The time that is displayed on the Dashboard does not match the time on my computer	The time displayed in the Dashboard is the time zone of our servers in Montreal, QC, Canada (UTC-05:00).
The table on my page seems to be loading forever	It has been noticed that Internet Explorer has some difficulties with the DataTables plug-in for jQuery. We strongly recommend using a browser other than Microsoft Internet Explorer, such as Mozilla Firefox, Google Chrome, or Safari.
MC-Sym did not generate any structure and returns an error message similar to "Failed to expand pattern 'MCSYM-DB/4/UACG/*_pdb.gz': No such file or directory"	Unfortunately, it means the script generated by the MC-Symizer is invalid, and you have to edit it manually to make it work. Please refer to: http://www.major.iric.ca/wiki/index.php/MC-Sym and http://www.major.iric.ca/MC-Sym/faq.html for more information about the MC-Sym scripting language.

Acknowledgments

We thank all members of the RNA Engineering Laboratory for useful comments and discussions. This work was supported by Grants from the Canadian Institutes of Health Research (Grant number MOP-93679) and the National Institutes of Health (Grant number 1R01GM088813-01A1). FM is a member of the Centre Robert-Cedergren.

References

1. Parisien, M. and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.
2. Major, F. (2003) Building three-dimensional ribonucleic acid structures. *IEEE Comput. Sci. Eng.*, **5**, 44–53.

3. Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E., and Cedergren, R. (1991) The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science*, **253**, 1255–1260.
4. Williamson, J.R. (2000) Induced fit in RNA-protein recognition. *Nat. Struct. Biol.*, **7**, 834–837.
5. Waterman, M.S. (1978) *Secondary Structure of Single-Stranded Nucleic Acids*, Academic Press, New York, pp. 167–212.
6. Burgstaller, P., Hermann, T., Huber, C., Westhof, E., and Famulok, M. (1997) Isoalloxazine derivatives promote photocleavage of natural RNAs at G.U base pairs embedded within helices. *Nucleic Acids Res.*, **25**, 4018–4027.
7. Chamberlin, S.I. and Weeks, K.M. (2000) Mapping local nucleotide flexibility by selective acylation of 2'-amine substituted RNA. *J. Am. Chem. Soc.*, **122**, 216–224.
8. Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
9. Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
10. Mortimer, S.A. and Weeks, K.M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.*, **129**, 4144–4145.
11. Mortimer, S.A., Johnson, J.S., and Weeks, K.M. (2009) Quantitative analysis of RNA solvent accessibility by N-silylation of guanosine. *Biochemistry*, **48**, 2109–2114.
12. Moine, H., Ehresmann, B., Ehresmann, C., and Romby, P. (1998) in *RNA Structure and Function* (eds R.W. Simons and M. Grunbert-Manago), Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 77–115.
13. Knapp, G. (1989) Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol.*, **180**, 192–212.
14. Krol, A. and Carbon, P. (1989) A guide for probing native small nuclear-RNA and ribonucleoprotein structures. *Methods Enzymol.*, **180**, 212–227.
15. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 7287–7292.
16. Do, C.B., Woods, D.A., and Batzoglou, S. (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, **22**, e90–e98.
17. Ding, Y., Chan, C.Y., and Lawrence, C.E. (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.*, **32**, W135–W141.
18. Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.
19. Zuker, M.M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
20. Dawson, W.K., Fujiwara, K., and Kawai, G. (2007) Prediction of RNA pseudoknots using heuristic modeling with mapping and sequential folding. *PLoS ONE*, **2**, e905.
21. Knudsen, B. and Hein, J. (2003) Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.*, **31**, 3423–3428.
22. Ruan, J., Stormo, G.D., and Zhang, W. (2004) ILM: a web server for predicting RNA secondary structures with pseudoknots. *Nucleic Acids Res.*, **32**, W146–W149.
23. Lerman, Y.V., Kennedy, S.D., Shankar, N., Parisien, M., Major, F., and Turner, D.H. (2011) NMR structure of a 4×4 nucleotide RNA internal loop from an R2 retrotransposon: identification of a three purine-purine sheared pair motif

- and comparison to MC-SYM predictions. *RNA*, **17**, 1664–1677.
24. Major, F. and Thibault, P. (2007) in *Bioinformatics: From Genomes to Therapies* (ed. T. Lengauer), Wiley-VCH Verlag GmbH, Weinheim, pp. 491–539.
 25. Gendron, P., Lemieux, S., and Major, F. (2001) Quantitative analysis of nucleic acid three-dimensional structures. *J. Mol. Biol.*, **308**, 919–936.
 26. Cedergren, R. and Major, F. (1998) Modeling the tertiary structure of RNA, in *RNA Structure and Function* (eds R.W. Simons and M. Grunbert-Manago), Cold Spring Harbor Press, Cold Spring Harbor, NY, pp 37–75.
 27. Wang, Z., Parisien, M., Scheets, K., and Miller, W.A. (2011) The cap-binding translation initiation factor, eIF4E, binds a pseudoknot in a viral cap-independent translation element. *Structure*, **19**, 868–880.
 28. Reymond, C., Lévesque, D., Bisaillon, M., and Perreault, J.P. (2010) Developing three-dimensional models of putative-folding intermediates of the HDV ribozyme. *Structure*, **18**, 1608–1616.
 29. Zhang, Q., Kim, N.K., Peterson, R.D., Wang, Z., and Feigon, J. (2010) Structurally conserved five nucleotide bulge determines the overall topology of the core domain of human telomerase RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 18761–18768.
 30. Yang, S., Parisien, M., Major, F., and Roux, B. (2010) RNA structure determination using SAXS data. *J. Phys. Chem.*, **114**, 10039–10048.
 31. Harris, S. and Schroeder, S.J. (2010) Nuclear magnetic resonance structure of the prohead RNA E-loop hairpin. *Biochemistry*, **49**, 5989–5997.
 32. Ali, M., Lipfert, J., Seifert, S., Herschlag, D., and Doniach, S. (2010) The ligand-free state of the TPP riboswitch: a partially folded RNA structure. *J. Mol. Biol.*, **396**, 153–165.
 33. McGraw, A.P., Mokdad, A., Major, F., Bevilacqua, P.C., and Babitzke, P. (2009) Molecular basis of TRAP-5'SL RNA interaction in the *Bacillus subtilis trp* operon transcription attenuation mechanism. *RNA*, **15**, 55–66.
 34. Parisien, M. and Major, F. (2012) Determining RNA three-dimensional structures using low-resolution data. *J. Struct. Biol.*, **179**, 252–260.
 35. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
 36. DeLano, W.L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA, USA, <http://www.pymol.org>.

31

S2S-Assemble2: a Semi-Automatic Bioinformatics Framework to Study and Model RNA 3D Architectures

Fabrice Jossinet and Eric Westhof

31.1

Introduction

RNA molecules play key roles at all stages of cell life through key sequential motifs and structural modules involved in intermolecular interactions (other RNAs, ligands, and proteins) and embedded in the intricate three-dimensional folds adopted by RNA molecules. This has stimulated many bioinformatics initiatives dedicated to the automated search for non-coding RNA (ncRNA) genes in sequenced genomes. Databases such as NONCODE [1], RNAdb [2], or “The noncoding RNA database” (<http://goo.gl/Z0zxa>) have been developed to store and diffuse these ncRNA candidates. Others, such as the Rfam database [3], organize them into families and provide them as multiple alignments and covariance models. Despite the fact that the number of RNA tertiary structures has increased dramatically in the last years (strengthening our knowledge of RNA structure and folding), a large majority of RNA molecules discovered so far still await their structures to be solved. As an alternative, a three-dimensional architectural model for an RNA molecule can be produced using a molecular modeling strategy. Such theoretical approaches have proved to be valuable in the past. One method is based on a constraint satisfaction algorithm that searches the conformational space, so that all possible models compatible with a given set of constraints are calculated. In a second approach, the researcher strives to identify, in an iterative process, the structural constraints from a mixture of theoretical and experimental data, in order to be able to construct the model.

Among the different sorts of information pieces, the availability of a solved tertiary structure for at least one RNA molecule within a family of homologous sequences is the richest. As a molecular 3D architecture evolves much more slowly than the sequences, structural data can be inferred for all the members of the RNA family with the same function. RNA 3D modeling by homology requires the alignment of the sequence to be modeled against the solved tertiary structure (a.k.a. the “reference structure”) (Figure 31.1). The construction of such a structural alignment is a difficult and iterative task that is up to now performed most frequently by hand. This has prompted us to develop the S2S tool as a way to optimize such

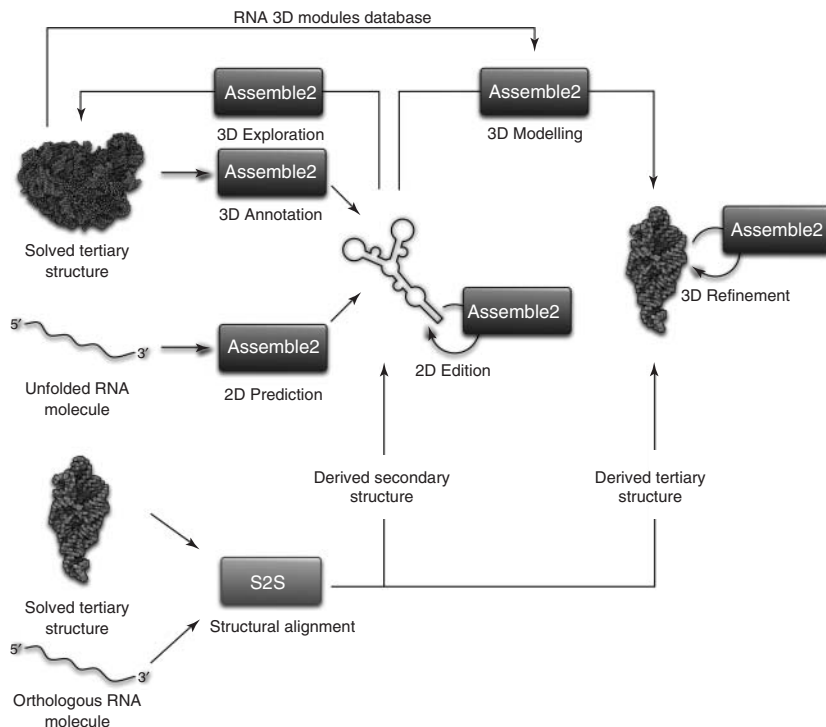


Figure 31.1 Workflows that can be realized with S2S and Assemble2.

processes [4]. For any orthologous sequence aligned to the reference structure, S2S can derive a secondary and a tertiary structure. The modeling process can then be pursued in a dedicated 3D modeler named Assemble [5], for which a new version has recently been released (Assemble2). With its ability to retrieve and apply local folds from solved tertiary structures, Assemble2 is also a valuable tool to realize an *ab initio* 3D modeling process that can be initiated from a pre-computed secondary structure or an unfolded RNA sequence (Figure 31.1). Combined together, these tools realize a bioinformatics platform that is a good compromise between usability and feasibility for the construction and understanding of RNA architectures.

The next three sections are dedicated to a general description of the features available with the S2S and Assemble2 tools. Then, the reader will find how to install these tools, along with the description of several scenarios corresponding to classical tasks for any modern RNA biologist (see Figure 31.1).

31.2

S2S: an Interactive RNA Alignment Viewer and Editor

Toward understanding the diverse functional roles of an RNA molecule, it is necessary to decipher its underlying architecture. As, at least energetically, the secondary

structure is the main component of an RNA architecture, and as it produces the skeleton by which the tertiary structure will be constrained, its definition constitutes a first and essential step. Such secondary structures can be retrieved from dedicated web sites such as the “Comparative RNA web site” [6] or RNA STRAND [7]. They can also be computed from one or several unfolded RNA sequences. Among several strategies, comparative sequence analysis is the most valuable one. Finding covariations within an RNA family allows one to define the locations of regular helices and consequently the locations of the single-stranded regions (hair-pin and internal loops, bulges, and multiple junctions). Unfortunately, automated approaches to produce RNA structural alignments and to deduce a consensus structure thereof have several limitations. First, these approaches are limited to the secondary structure level. Second, in most cases, the results cannot be used *per se* but have to be re-evaluated and curated by the expert in a dedicated environment. Several tools have been released recently to cope with this iterative and time-consuming task (4SALE [8], Construct [9], S2S [4], SARSE [10]).

S2S has been designed to align a set of RNA sequences against a solved tertiary structure or a “simple” secondary structure (without any tertiary information). To do so, it provides the following unique functionalities (Figure 31.2):



Figure 31.2 Global view of S2S. The right panel (the “2D panel”) displays the reference structure computed from a tertiary one with the RNAVIEW algorithm. The secondary interactions are displayed in blue color and the tertiary ones in green. The current selection is displayed in red. The upper middle panel (the “alignment panel”) displays the structural alignment.

The sequence of the reference structure is labeled as S1. Six orthologous sequences are displayed (labeled from S2 to S7). The lower middle panel (the “sequences panel”) lists all the RNA sequences present in the alignment. The left panel (“Explorer panel”) displays a tree of all the RNA objects making up the reference structure.

- The interconnection between an alignment panel and a 2D panel displaying the reference structure.
- The display of long-range base–base interactions: the alignment can be split into several views, that can be manipulated independently. Between these views, the secondary and tertiary interactions are displayed if their residues are present simultaneously in the display.
- The display of a structural mask: for each view, the graphical engine of S2S highlights the compatibility between the orthologous sequence(s) and the local interaction schemes. More precisely, S2S calculates, from a set of isostericity rules [11], the ability of the orthologous sequence to conserve the folding of the reference structure. Technically, two different colors can be chosen by the user to highlight the secondary and tertiary interactions (Figures 31.2 and 31.3). For each color, the darkest coloring indicates that the combination of the orthologous residues chosen by the user for the type of interaction observed in the reference structure will be isosteric with the reference base pair. The lightest coloring indicates that the combination is geometrically possible but not necessarily isosteric with the reference base pair, whereas the absence of color indicates that such combination has never been observed in solved structures so far. As the base of an RNA residue presents three edges for base pairing, the background area of each residue can be split into two or three equal segments (Figure 31.4).

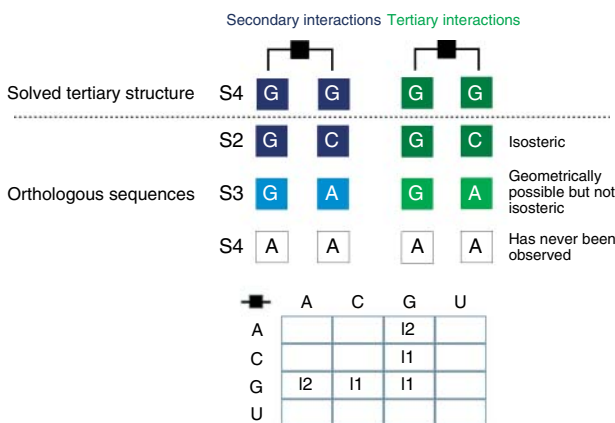


Figure 31.3 Example of the structural mask used by S2S for the *cis* Hoogsteen/Hoogsteen (cHH) geometric family. This family contains two isosteric subfamilies (I1 and I2). Two different colors can be chosen by the user to differentiate secondary and tertiary interactions. For each color, the darkest coloring indicates that the combination of the orthologous residues chosen by the user for the type of interaction observed in the reference structure will be isosteric with

the reference base pair (above the stippled line). The lightest coloring indicates that the combination is geometrically possible but not necessarily isosteric with the reference base pair, whereas the absence of color indicates that such combination has never been observed in solved structures so far. Note that the isostericity rules for the shown base–base interaction are the same no matter if it represents a secondary or tertiary interaction.

If a residue in the solved tertiary structure participates in a triple interaction, the background color is split and we have 9 possible combinations:



Figure 31.4 The structural mask used by S2S to display the structural conservations for a residue participating in a triple interaction. The secondary interaction is displayed in blue color, and the tertiary one in green.

31.3

Assemble2: an Interactive RNA 3D Modeler

In its first version, published in 2010, Assemble was made with two interconnected panels:

- a 3D panel displaying the RNA 3D architecture and
- a 2D panel displaying the RNA's "extended" secondary structure and including the tertiary structural constraints.

For any selection done in one panel, its counterpart is highlighted in the second panel. With its version 2 (Assemble2), introduced in this chapter for the first time, Assemble has been reduced to its 2D panel and is now interconnected to the program UCSF Chimera for 3D visualization and manipulation (Figure 31.5). Consequently, Assemble2 can be seen as a kind of "plugin," extending Chimera with functionalities adapted to the peculiarities of RNA:

- the ability to easily highlight any part of the structure, going from a single residue to a helix or single-stranded region (termed *single-strand* in the following),
- the ability to generate regular helices and single-strands in Chimera, along with the ability to assemble them into a compact RNA 3D model, and
- the refinement of an RNA 3D model: Assemble2 provides a geometric refinement function to optimize structural parameters such as nucleotide stereochemistry, Watson-Crick and Wobble base pairs, sugar pucker, and atom distances. The structural constraints used during this refinement step are deduced from the set of base–base interactions linked to the secondary structure displayed in the 2D panel.

The construction paradigm used by Assemble2 is based on the modular architecture that has been observed for all RNA structures solved so far. It is now well known that several local 3D foldings such as tetraloops (e.g., GNRA, UUCG), internal loops (loop E, sarcin/ricin motif), kink-turns [12], three-way junctions [13], or A-minor motifs [14] are recurrently observed in RNA architectures. These RNA motifs can be seen as 3D patterns reused by RNA architectures having different biological functions [15, 16]. As these RNA modules are recurring and occur across all phylogenetic domains, it can be speculated that any structured ncRNA contains one or several of these motifs along with the regular A-form double helices. It

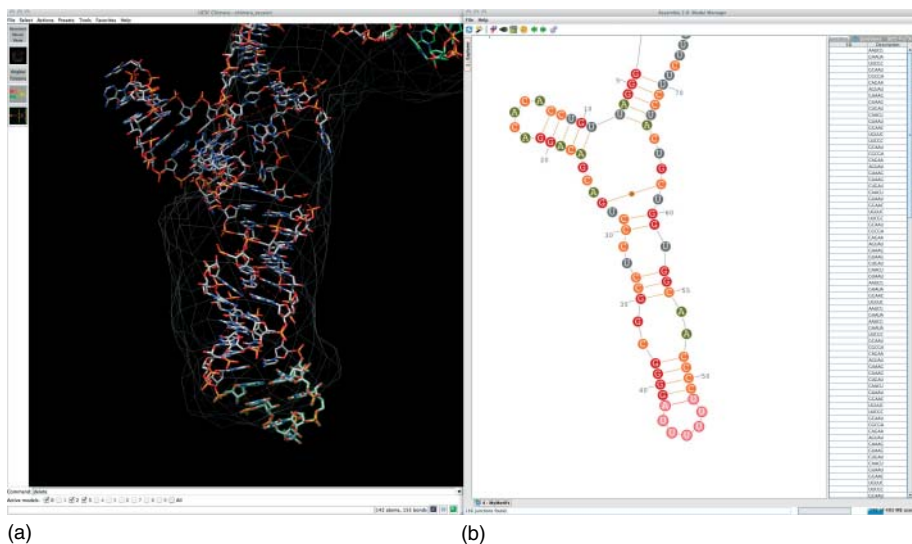


Figure 31.5 Global view of Assemble2.

Assemble2 extends the rendering and analysis abilities of the UCSF Chimera program (a) by means of an interactive 2D panel (b) allowing one to visualize and manipulate the RNA secondary structure to be modeled.

In this screenshot, an apical loop has been selected (nucleotides marked with pink background in the 2D panel), invoking a list (to

the right of the 2D panel) of 3D folds of apical loops of the same length and extracted from solved tertiary structures. With a single click, one of these folds (residues highlighted with contours in the 3D panel) was selected, exported to Chimera and stacked automatically above its enclosing helix. In the 2D panel, A, C, G, and U residues are highlighted by different background colors.

should be noted that, up to now, Assemble2 is limited to the modeling of a single molecular architecture.

31.4

The Semi-Automatic Architecture of S2S and Assemble2

During the last years, the bioinformatics community has accumulated a large amount of algorithms working on RNA data. These algorithms, which have contributed to a faster and deeper understanding of biological RNA systems, mainly address the structural aspects of RNA molecules: 2D and 3D prediction, 2D and 3D visualization, 3D annotation, 2D and 3D comparison, and so forth. Unfortunately, the installation and usage of these bioinformatic algorithms is in general a limiting step for biologists. Frequently, one needs to recompile them to make them functional. A large majority are also provided as “simple” command-line tools, which is an important limitation for many scientists not trained in this exercise. In recent years, the bioinformatics community has released some solutions to allow biologists to use such tools hosted on remote computers [17]. These “Web services” are now provided by an increasing number of laboratories (for an exhaustive list, the reader can consult the following web site: <http://www.biocatalogue.org/>).

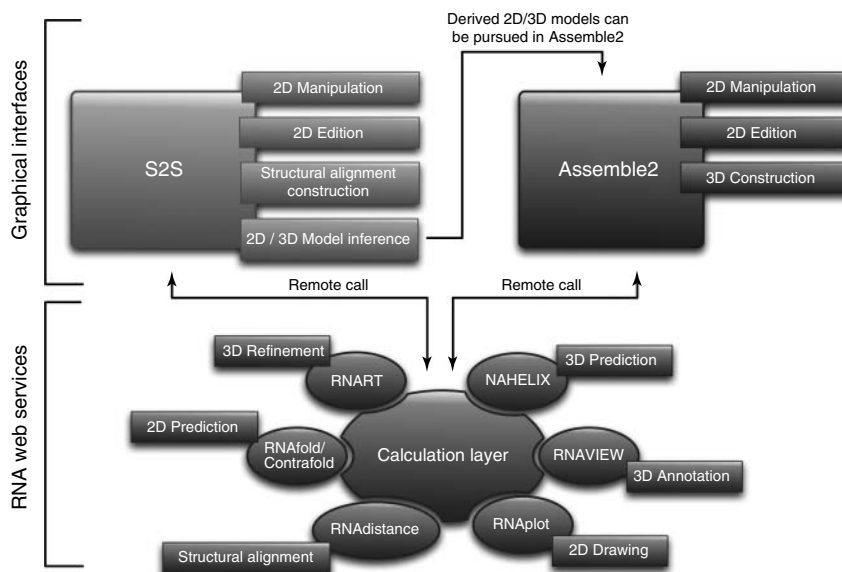


Figure 31.6 The semi-automatic architecture of S2S and Assemble2.

As modern bioinformatics tools, Assemble2 and S2S delegate several recurrent tasks to RNA web services hosted on a dedicated server in our laboratory (Figure 31.6):

- **RNA 2D prediction:** by using Contrafold [18] or RNAfold from the Vienna RNA package [19],
- **RNA 2D plotting:** by using RNApIot from the Vienna RNA package [19],
- **RNA 3D annotation:** by using RNAVIEW [20],
- **RNA alignment:** by using RNAdistance from the Vienna RNA package [19], and
- **RNA 3D motifs query and usage:** by using in-house algorithms (RNART and NAHELIX).

These RNA web services are set up on a 2×2.26 GHz Quad-Core Intel Xeon “Nehalem” processors machine with 16 GB memory and running under Ubuntu 12.04 (<http://sbgarn1.in2p3.fr/api/>).

31.5

Installation of S2S and Assemble2

To install S2S or Assemble2, you first need to download them at <http://bioinformatics.org/s2s/> and <http://bioinformatics.org/assemble/>, respectively. To use them efficiently, you will need an Internet connection (to use the RNA web services) and a three-button mouse. You will also need a Java environment, installed and running on your computer. S2S and Assemble2 can be used with Java 1.5 and 1.6. You need to use the official Java environment from Oracle (<http://java.sun.com>). You can type “java version” in a terminal to test your running version.

Once downloaded, unzip the archive by double-clicking on it. Then open the file “launch_S2S_for_XXX” or “launch_Assemble2_for_XXX” (“XXX” corresponding to your operating system) with a text editor. Set the location of your Java environment and/or modify the amount of memory allocated to the tool (500 MB are allocated by default). Thanks to the remote architecture of S2S and Assemble2, 256 MB of memory is enough to load structures of eukaryotic and bacterial ribosomal RNAs. Concerning Assemble2, the UCSF Chimera program has to be installed on the computer as well. After its first launch, Assemble2 will ask you to set its location.

S2S and Assemble2 are open-source projects. The source code is freely available at <http://goo.gl/0gFbM> and <http://goo.gl/Yd1td>, respectively.

Protocols

Scenario #1: How to Predict, Visualize, and Edit an RNA Secondary Structure

Prerequisites

- the Assemble2 tool,
- an RNA molecule stored in a FASTA file or a pre-computed RNA secondary structure stored in a CT, BPSEQ, or FASTA + bracket notation file, and
- an Internet connection.

Scientific Context

As the biological function of an RNA molecule is directly linked to its structure, the experimental approaches applied to an RNA candidate have to be designed in the light of its folding landscape. The latter can be explored theoretically using prediction algorithms, for which the offer has greatly increased during recent years. Unfortunately, this offer is mainly focused on the prediction of the secondary structure of an RNA sequence. New tools predicting the tertiary structure have been made available recently, but their applications are restricted to short RNA molecules and, consequently, to RNA molecules with reduced complexity (for recent reviews, see [21–23]).

Owing to the inherent limitations of these prediction tools, a biologist frequently needs to re-evaluate and re-edit the results in a dedicated environment, according to her/his own assumptions and/or experimental data. In this section, we will illustrate how Assemble2 can be used as such an environment.

Practical Workflow

- 1) Step #1: launch Assemble2
Launch Assemble2 by double-clicking on the file “launch_Assemble2_for_XXX” or by typing in a terminal “./launch_Assemble2_for_XXX”
- 2) Step #2: load an RNA secondary structure
Use “File -> Load...” to open your file. Choose:
 - “RNA molecule”: to load an unfolded RNA sequence stored in a FASTA file. Assemble2 will ask you to select a 2D prediction algorithm.

Table 31.1 File formats supported by S2S and Assemble2.

File format	Example
FASTA	<pre>>my_rna TAACAATCTGCTGAAAGGTACCGTCGGAGGGAGCTTTGTTGCCAGCGCCA GAAACGCCGGTTTAACCAGCGCCGAAGTGAGCGCAGTGATTAAAGCCATG CAGTGGCAAATGGATTTCGCAAACCTGAAAAAAGGCGATGAATTTGCGGT</pre>
FASTA + bracket notation	<pre>>Secondary structure((.....)) >my_rna TAACAATCTGCTGAAAGGTACCGTCGGAGGGAGCTTTGTTGCCAGCGCCA GAAACGCCGGTTTAACCAGCGCCGAAGTGAGCGCAGTGATTAAAGCCATG CAGTGGCAAATGGATTTCGCAAACCTGAAAAAAGGCGATGAATTTGCGGT</pre>
CT	<pre>50 ENERGY = -8.5 my_rna 1 G 0 2 0 1 2 A 1 3 0 2 3 G 2 4 0 3 4 A 3 5 0 4 47 U 46 48 18 47 48 G 47 49 17 48 49 G 48 50 16 49 50 U 49 0 15 50</pre>
BPSEQ	<pre>1 U 0 2 A 0 3 A 0 4 C 0 ... 146 G 121 147 C 120 148 G 119 149 G 118 150 U 0</pre>
PDB	A complete description can be found at http://goo.gl/DVoSA

At present, Contrafold and RNAfold from the Vienna package are supported. New prediction algorithms will be added in the near future;

- “RNA secondary structure”: to load a pre-computed RNA secondary structure stored in a BPSEQ, CT, or FASTA + bracket notation file (see Table 31.1 for details about these file formats).

3) Step #3: translate and zoom the RNA secondary structure

You can act on the RNA secondary structure at two different levels:

- **globally:** the structure can be translated (right mouse button pressed and dragged) and magnified (using the mouse wheel);
 - **locally:** by selecting a subpart of the structure, which is described in Step #4.
- 4) Step #4: edit the layout of the RNA secondary structure
- By clicking on a residue in the 2D panel, you can select subparts of the secondary structure. The accumulation of mouse clicks on the same residue allows to iterate over the following selection process:
- **one mouse click:** selection of a single residue;
 - **two successive mouse clicks:** selection of the secondary interaction (if any) containing this residue;
 - **three successive mouse clicks:** selection of the structural domain (single-strand or helix) containing this residue;
 - **four successive mouse clicks:** if the residue selected is inside a single-strand, this selects the corresponding junction.

By double-clicking outside of the structure, the current selection is cleared, allowing you to restart the selection process.

If a helix is selected, it can be translated and rotated independently of the rest of the secondary structure by keeping the Alt key pressed (Alt + right mouse button pressed and dragged to translate, or Alt + left mouse button pressed and dragged to rotate).¹⁾

- 5) Step #5: alter helices and base–base interactions
- Assemble2 provides a mode to change the definition of the secondary structure. Once the editing mode is activated, a user can:
- remove a selected helix,
 - change the isostericity family of a selected interaction,
 - create a new interaction between two selected residues, or
 - create a new helix with four selected residues.

The Assemble2 website provides several screencasts explaining these different steps in more detail.

Scenario #2: How to Easily Study and Manipulate a Solved RNA Tertiary Structure

Prerequisites

- the Assemble2 tool,
- a solved tertiary structure stored in a PDB file, and
- an Internet connection.

Scientific Context

The recent progress made in the field of structural biology has greatly improved the number and the complexity of solved tertiary structures. Since 2010, several

1) In Ubuntu, the “Alt key + left mouse button” is used to move a window, and thus does not work for Assemble. This behavior, however, might be modified by menus, depending on the window manager.

high-resolution structures of eukaryotic ribosomal RNAs have become available, making them the largest and most complex RNA molecules solved so far.

Consequently, the exploration and manipulation of such 3D architectures need to be done nowadays with the help of new abstraction levels. Just as topographic maps are used to facilitate the exploration and the understanding of the three-dimensional surface of the earth, RNA secondary projections of tertiary architectures serve a similar purpose. This supposes the ability to automatically identify all the building blocks (helices, single-strands, junctions, base–base interactions) contributing to the 3D architecture. This has been made possible recently with the availability of a new class of RNA algorithms converting a complete atomic RNA structure into an “extended” secondary structure. By “extended,” we mean a secondary structure supplemented with all the secondary and tertiary interactions stabilizing the 3D packaging of this secondary structure. As will be illustrated here, by computing such an “extended” secondary structure and interconnecting it to its original all-atom 3D structure, S2S and Assemble are unique tools to improve the exploration of complex tertiary structures.

Practical Workflow

- 1) Step #1: launch Assemble2
Launch Assemble2 by double-clicking on the file “launch_Assemble2_for_XXX” or by typing in a terminal “./launch_Assemble2_for_XXX”
- 2) Step #2: load an RNA tertiary structure
Use “File -> Load. . .” to open your file. Choose “RNA Tertiary Structure” to load a solved tertiary structure stored in a PDB file (see Table 31.1). Assemble2 will use the RNAVIEW algorithm [20] to annotate the structure and to produce an “extended” secondary structure. New algorithms will be added in the near future.
The computed “extended” secondary structure is displayed in the 2D panel and the tertiary structure is displayed *via* the UCSF Chimera program. In the “extended” secondary structure, all the secondary and tertiary interactions identified are described with the symbols of the Leontis–Westhof classification [24].
- 3) Step #3: translate and zoom the RNA secondary structure
You can act on the secondary structure at two different levels:
 - **globally**: the structure can be translated (right mouse button pressed and dragged) and magnified (using the mouse wheel);
 - **locally**: by selecting a subpart of the structure, which is described in Step #4.
- 4) Step #4: explore the RNA 3D structure from its secondary structure
By clicking on a residue in the 2D panel, Assemble2 highlights the 3D counterpart in the Chimera program. By accumulating mouse clicks on the same residue, the following selection process can be traversed:
 - **one mouse click**: selection of a single residue; you can also shift-click on a second residue to select both residues and all residues in between; hold down the ctrl key to select multiple residues;

- **two successive mouse clicks:** selection of the secondary interaction (if any) containing this residue;
- **three successive mouse clicks:** selection of the structural domain (single-strand or helix) containing this residue;
- **four successive mouse clicks:** if the residue selected is inside a single-strand, this selects the corresponding junction.

By double-clicking outside of the structure, the current selection is cleared, allowing you to restart the selection process.

Scenario #3: How to Derive a New RNA Tertiary Structure from a Solved One

Prerequisites

- the S2S tool,
- an orthologous RNA molecule to be modeled and stored in a FASTA, FASTA + bracket notation, CT, or BPSEQ file,
- a solved tertiary structure stored in a PDB file, and
- an Internet connection.

Scientific Context

Frequently, RNA biologists are confronted with the availability of a tertiary structure that has been solved for an RNA molecule orthologous to their own system. As with proteins, the sequence of an RNA is less conserved than its structure, leading to sequence rearrangements due, for example, to indels and point mutations. Owing to the plasticity of RNA architectures, orthologous sequences within a given RNA family can also contain additional domains with sizes varying from about 10 to more than 100 nt (as observed for ribosomal RNAs, [25] or more recently, for RNase P RNA [26]). Such additional domains are often independent folding units that can be added to or removed from the common core, with minimal alterations to the fold [27].

However, the threading of an RNA sequence into a known structure, even if it is an orthologous one, can be a tricky task without adapted tools. We detail in this section how S2S, with its unique graphical features, can be used to “structurally” align an RNA sequence more easily to an orthologous one extracted from a solved tertiary structure, and how this structural alignment can be used to automatically infer a tertiary structure for the orthologous RNA of interest.

The fact that RNA architectures often “accumulate” new modules makes RNA alignments easily unmanageable without adapted tools [28]. We will see how S2S, with its ability to display a single alignment in several independent views, allows one to circumvent such difficulties.

Practical Workflow

1) Step #1: launch S2S

Launch S2S by double-clicking on the file “launch_S2S_for_XXX” or by typing in a terminal “./launch_ S2S _for_XXX”

- 2) Step #2: load an RNA tertiary structure
Use “File -> Load...” to open your file. Choose “RNA Tertiary Structure” to load a solved tertiary structure stored in a PDB file (see Table 31.1). S2S will use the RNAVIEW [20] algorithm to annotate the RNA structure and to produce an “extended” secondary structure.
The “extended” secondary structure computed is displayed in the 2D panel and its primary sequence in the alignment panel (Figure 31.2).
- 3) Step #3: load an orthologous sequence
An orthologous sequence has to be imported into the working session using “File -> Load...”. To avoid the necessity to produce the full alignment from scratch, S2S can pre-compute a first structural alignment by estimating the distance between the reference secondary structure and the one of the new sequence. If this new sequence has been imported from a file without any structural information (such as a FASTA file), a secondary structure will be computed by S2S (and S2S will ask you to choose a prediction algorithm). If the sequence has been loaded from a BPSEQ or CT file, S2S will use the secondary structure described in the file to estimate the structural distance.
- 4) Step #4: “structurally” align your orthologous sequence
To align your orthologous sequence to the reference structure, you can change any residue position with keyboard shortcuts:
 - press the “i” key to switch to the “gap insertion” mode; click on the bracket notation or on the “alignment meter” above it to insert a column of gaps; the gaps allow you to move your residues;
 - select the view you want to use; to do so, click on the black circle in the upper-left corner of this view.
 - select a residue inside this view by clicking on it; you can also shift-click on a second residue to select both residues and all residues in between. Press the “o” or “p” keys to move all the selected residues to the left or to the right;
 - press the “d” key to delete a column containing only gaps; click on the bracket notation or on the “alignment meter” above it to mark such columns for deletion.
 - press the “s” key to come back to the “standard” mode (this is a secure mode, and the default one when S2S is starting, precluding accidental insertions or deletions).

In order to help during the alignment of residues participating in long-range interactions, S2S allows the presentation of several independent views of the same alignment (using the following buttons  ). As illustrated in Figure 31.2 (sequence alignment panel in the middle), the same sequence stretch (approximately nt 1640–1710) is displayed on top of each other in antiparallel orientation to visualize local secondary and tertiary interactions. Likewise, the sequence stretch of nt 120–180 is positioned at the bottom in antiparallel orientation to the nt 1710–1640 region to illustrate long-range tertiary interactions.

- 5) **Step #5: derive new 2D and 3D models for your orthologous sequence**
Once the structural alignment is done, choose your target sequence in the “Sequences” panel, click on it with the right mouse button and choose “Set as reference molecule.” S2S will ask you to choose a name for the derived 2D and 3D structures. Once done, the target sequence will become the new reference one (labeled as S1), and its derived secondary structure will be displayed in the 2D panel.

It should be noted that, if additional domains are defined during the construction of the structural alignment, they will be treated as mere single-strands in the resulting secondary structure, without any 3D information. Owing to the existence of such additional domains, or to structural inconsistencies (Figure 31.7), these derived 2D and 3D models generally need to be improved with Assemble2. To do so, save your working session as an S2S project with your target sequence as the reference one, and reopen it with Assemble2.

Scenario #4: How to Construct an RNA 3D Model from Its Secondary Structure

Prerequisites

- the Assemble2 tool,
- a pre-computed RNA secondary structure, and
- an Internet connection.

Scientific Context

As, at least energetically, the secondary structure is the main component of an RNA architecture and because it produces the skeleton constraining the tertiary structure, its definition constitutes a first and essential step in the construction of

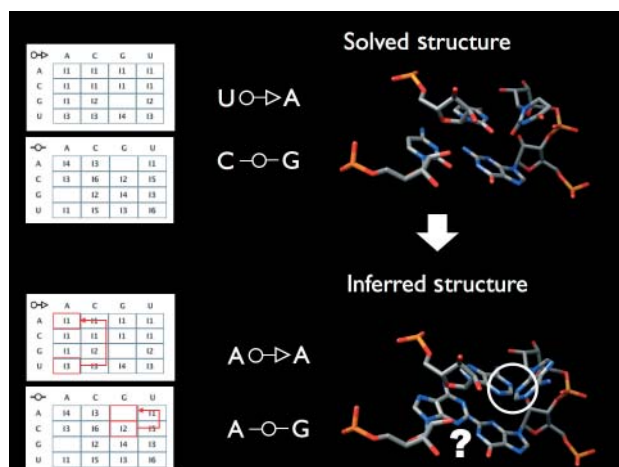


Figure 31.7 Examples of steric clashes in a 3D structure inferred from S2S. Such structural inconsistencies can be easily fixed with Assemble2.

an RNA 3D model [29, 30]. This secondary structure can then be broken down into “building blocks” such as the helical regions and the non-helical linking elements (a.k.a. single-stranded regions or “single-strands”: apical/internal loops, 5'/3' dangling ends, and multiple junctions). These “building blocks” can then be added step by step to the 3D model. The main difficulty of the approach is to identify the final folding of these building blocks and their relative organization in the 3D space. In the near future, we plan to exploit the semi-automatic architecture of Assemble2 for connecting it to novel algorithms that are able to find such structural modules automatically [31].

Thanks to its interconnection with a dedicated 3D modules database and to its ability to generate regular modules such as helices, we will now detail how Assemble can be used to easily generate 3D modules based on their secondary structure definition and how they can be easily assembled in a 3D space toward a final 3D model.

Practical Workflow

1) Step #1: launch Assemble2

Launch Assemble2 by double-clicking on the file “launch_Assemble2_for_XXX” or by typing in a terminal “./launch_Assemble2_for_XXX”

2) Step #2: load an RNA secondary structure

Use “File -> Load...” to open your file. Choose:

- **“RNA molecule”**: to load an unfolded RNA sequence stored in a FASTA file; Assemble2 will ask you to select a 2D prediction algorithm; at present, Contrafold and RNAfold from the Vienna package are supported;
- **“RNA secondary structure”**: to load a pre-computed RNA secondary structure stored in a BPSEQ, CT, or FASTA + bracket notation file (see Table 31.1 for details about these file formats);
- **“S2S/Assemble2 projects”**: to load a previous Assemble2 working session or 2D/3D structures derived from S2S (see scenario #3).

Whatever option used, the UCSF Chimera program should start and display the RNA 3D structure (if any).

3) Step #3: add new 3D building blocks within the Chimera environment

In a first step, you need to select a subpart of the RNA secondary structure from the 2D panel. By accumulating mouse clicks on the same residue, you will iterate over the following selection process:

- **one mouse click**: selection of a single residue; you can then export it to Chimera using the button ;
- **two successive mouse clicks**: selection of the secondary interaction (if any) containing this residue; you can then export it to Chimera using the button ;
- **three successive mouse clicks**: selection of the structural domain (single-strand or helix) containing this residue; you can then export

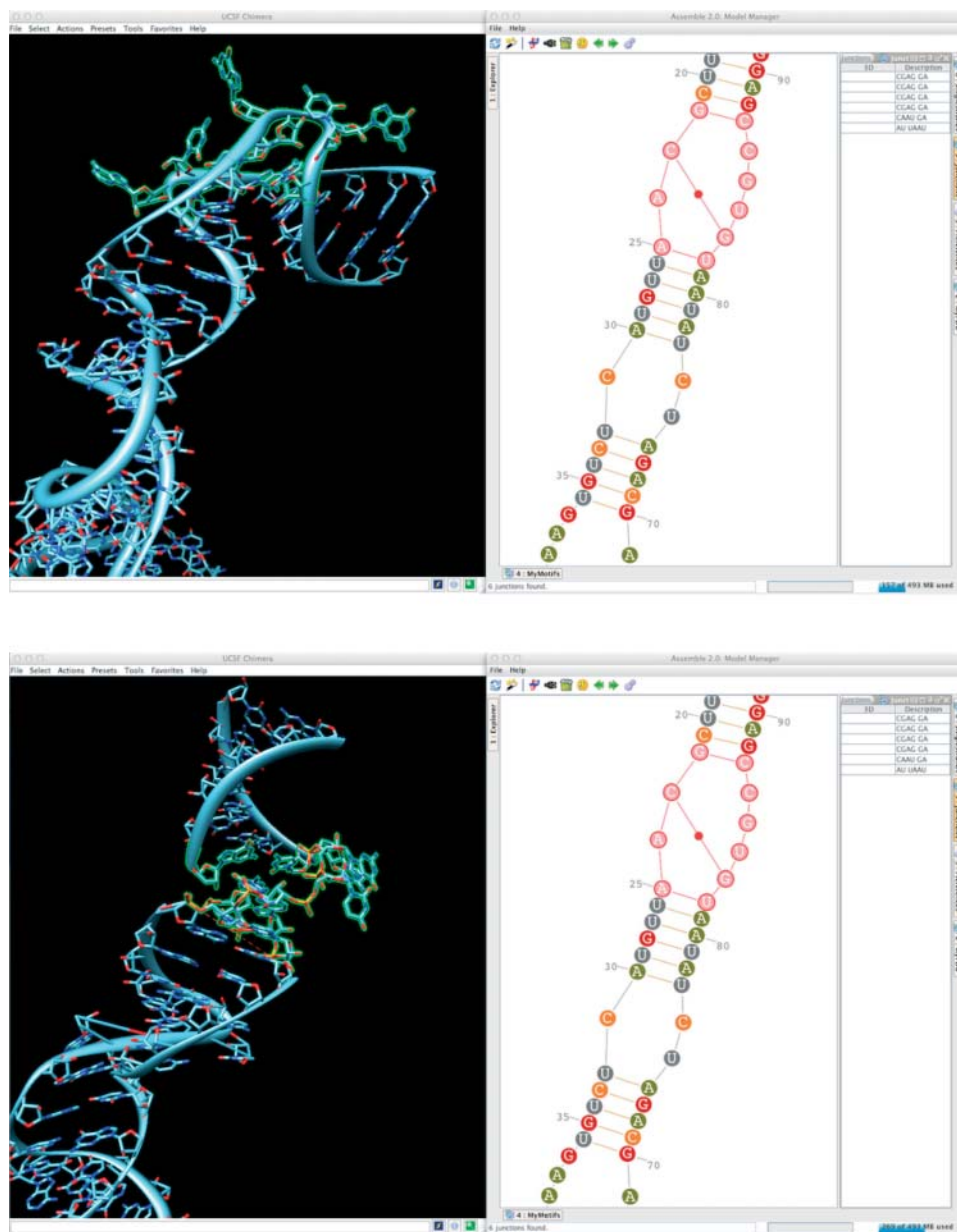


Figure 31.8 Assemble2 allows one to switch easily between alternative 3D folds extracted from solved tertiary structures. In this example, an internal loop was selected in the 2D panel and the fold of two isosequential loops from the list at the right image margin was implemented, resulting in the different 3D structure models shown on the left.

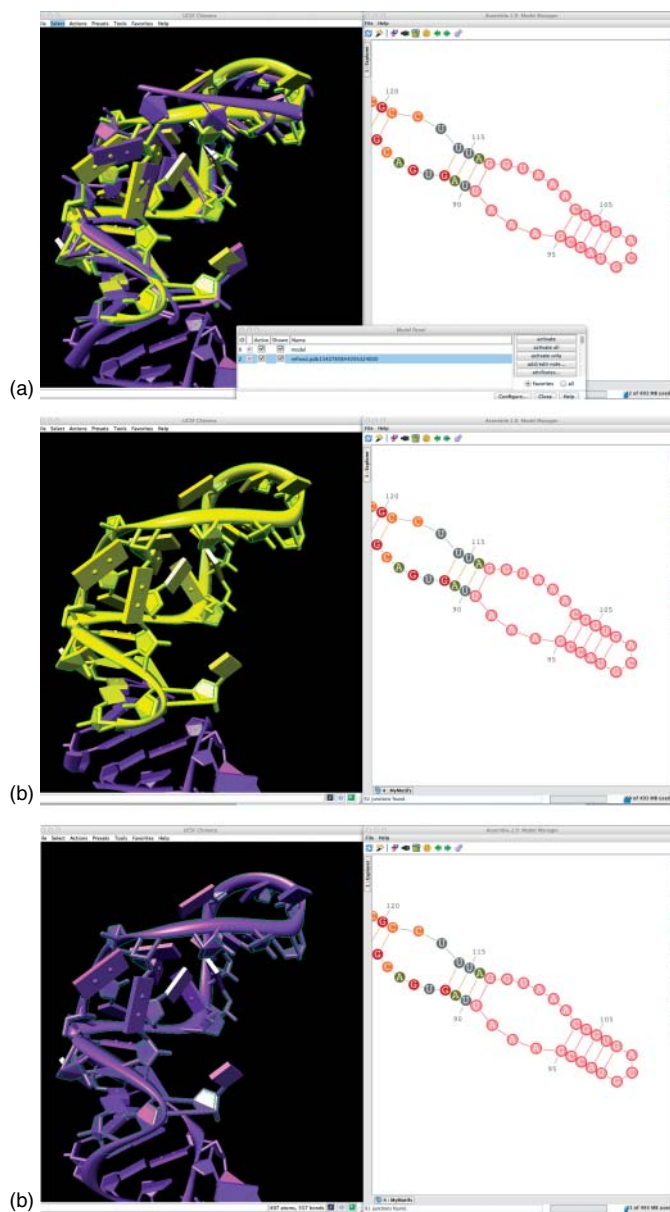


Figure 31.9 Refinement of an RNA 3D model. Assemble2 can refine a sub-part of an RNA 3D model after selection of its residues in the 2D panel. At the end of the process, the refined residues are exported into a new layer of Chimera (displayed in yellow color on

the left side of panel a). This allows the user to easily compare the folding before and after the refinement process. The refined residues can be easily included in the 3D model by removing the old ones (b) and merging the layers (c). The three 2D panels remain identical during this operation.

it to Chimera, with a regular helical fold, using the button ; concerning single-strands, the selection done will also open a “Single-Strands” panel listing all the single-strands with the same length and extracted from solved tertiary structures; click with the right mouse button on an entry and choose “Apply Hit”;

- **four successive mouse clicks:** if the residue selected is inside a single-strand, this selects the corresponding junction. The selection done will also open a “Junctions” panel listing all the junctions of the same type (apical and internal loops, three-way to n-way junctions) and extracted from solved tertiary structures; click with the right mouse button on an entry and choose “Apply Hit.”

By double-clicking outside of the structure, the current selection is cleared, allowing you to restart the selection process. If some residues were already present in Chimera at the beginning of Step #3, the new ones will replace them, allowing you to easily switch between alternative 3D folds (Figure 31.8).

4) Step #4: assemble the 3D building blocks

By default, a 3D building block created is automatically stacked on the previous residue in the RNA sequence; but you can also easily stack a selection of residues using the icon . It should be noted that the phosphate backbone of the building blocks can be connected directly within Chimera.

5) Step #5: refine the RNA 3D model (Figure 31.9)

Once finished, several structural inconsistencies can exist within the 3D model. Consequently, Assemble2 provides an embedded refinement tool named RNART (for RNA Refinement Tool), setting structural constraints such as nucleotide stereochemistry, Watson-Crick and Wobble base pair geometry, sugar pucker and atom distances. RNART is a Java rewriting of the NUCLIN-NUCLSQ auxiliary programs used by MANIP [32].

Assemble2 allows you to easily refine any subpart of your RNA 3D model (Figure 31.9). As a first step, one has to select the residues to be refined in the 2D panel. At the end of the process, the refined residues are exported to a new layer of Chimera, allowing you to easily compare the folding before and after the refinement process. To replace the original residues with the refined ones, you have to:

- select the original residues from the 2D panel,
- remove these residues using the “recycle bin” icon,
- from Chimera, merge the two layers (layers #0 and #1) using the “copy/combine” option in the Model Panel. The following options should be used:
 - new model’s name: *name*,
 - new model’s ID: *number*
 - select the option “close source models”
 - synchronize the 2D model with the merged structure in Chimera (using the  icon); this step is necessary to let Assemble2 know that the 3D model in Chimera is now the default 3D state.

References

1. He, S., Liu, C., Skogerbø, G., Zhao, H., Wang, J., Liu, T., Bai, B., Zhao, Y., and Chen, R. (2008) NONCODE v2.0: decoding the non-coding. *Nucleic Acids Res.*, **36** (Database issue), D170–D172.
2. Pang, K.C., Stephen, S., Dinger, M.E., Engström, P.G., Lenhard, B., and Mattick, J.S. (2007) RNAdb 2.0—an expanded database of mammalian non-coding RNAs. *Nucleic Acids Res.*, **35** (Database issue), D178–D182.
3. Gardner, P.P., Daub, J., Tate, J., Moore, B.L., Osuch, I.H., Griffiths-Jones, S., Khanna, A., and Marshall, M. (2011) Rfam: Wikipedia, clans and the “decimal” release. *Nucleic Acids Res.*, **39** (Database issue), D141–D145.
4. Jossinet, F. and Westhof, E. (2005) Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics*, **21**, 3320–3321.
5. Jossinet, F., Ludwig, T.E., and Westhof, E. (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics*, **26**, 2057–2059.
6. Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Müller, K.V. *et al.* (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinf.*, **3**, 2.
7. Andronescu, M., Bereg, V., Hoos, H.H., and Condon, A. (2008) RNA STRAND: the RNA secondary structure and statistical analysis database. *BMC Bioinf.*, **9**, 340.
8. Seibel, P.N., Müller, T., Dandekar, T., and Wolf, M. (2008) Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. *BMC Res. Notes*, **1**, 91.
9. Wilm, A., Linnenbrink, K., and Steger, G. (2008) ConStruct: improved construction of RNA consensus structures. *BMC Bioinf.*, **9**, 219.
10. Andersen, E.S., Lind-Thomsen, A., Knudsen, B., Kristensen, S.E., Havgaard, J.H., Torarinsson, E., Larsen, N., Zweib, C., Sestoft, P., Kjems, J., and Gorodkin, J. (2007) Semiautomated improvement of RNA alignments. *RNA*, **13**, 1850–1859.
11. Leontis, N.B., Stombaugh, J., and Westhof, E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
12. Klein, D.J., Schmeing, T.M., Moore, P.B., and Steitz, T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214–4221.
13. Lescoute, A. and Westhof, E. (2006) Topology of three-way junctions in folded RNAs. *RNA*, **12**, 83–93.
14. Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B., and Steitz, T.A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 4899–4903.
15. Leontis, N.B., Lescoute, A., and Westhof, E. (2006) The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.*, **16**, 279–287.
16. Hendrix, D.K., Brenner, S.E., and Holbrook, S.R. (2005) RNA structural motifs: building blocks of a modular biomolecule. *Q. Rev. Biophys.*, **38**, 221–243.
17. Curcin, V., Ghanem, M., and Guo, Y. (2005) Web services in the life sciences. *Drug Discovery Today*, **10**, 865–871.
18. Do, C.B., Woods, D.A., and Batzoglou, S. (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, **22**, e90–e98.
19. Hofacker, I.L. (2009) RNA secondary structure analysis using the Vienna RNA package. *Curr. Protoc. Bioinf.*, **12**, 12.2.
20. Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H., and Westhof, E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.

21. Laing, C. and Schlick, T. (2011) Computational approaches to RNA structure prediction, analysis, and design. *Curr. Opin. Struct. Biol.*, **21**, 306–318.
22. Cruz, J.A., Blanchet, M.-F., Boniecki, M., Bujnicki, J.M., Chen, S.-J., Cao, S., Das, R., Ding, F., Dokholyan, N.V., Flores, S.C. *et al.* (2012) RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*, **18**, 610–625.
23. Parisien, M., Cruz, J.A., Westhof, E., and Major, F. (2009) New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA*, **15**, 1875–1885.
24. Leontis, N.B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
25. Yokoyama, T. and Suzuki, T. (2008) Ribosomal RNAs are tolerant toward genetic insertions: evolutionary origin of the expansion segments. *Nucleic Acids Res.*, **36**, 3539–3551.
26. Kachouri, R., Stribinskis, V., Zhu, Y., Ramos, K.S., Westhof, E., and Li, Y. (2005) A surprisingly large RNase P RNA in *Candida glabrata*. *RNA*, **11**, 1064–1072.
27. Westhof, E. and Massire, C. (2004) Structural biology. Evolution of RNA architecture. *Science*, **306**, 62–63.
28. Brown, J.W., Birmingham, A., Griffiths, P.E., Jossinet, F., Kachouri-Lafond, R., Knight, R., Lang, B.F., Leontis, N., Steger, G., Stombaugh, J., and Westhof, E. (2009) The RNA structure alignment ontology. *RNA*, **15**, 1623–1631.
29. Brion, P. and Westhof, E. (1997) Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 113–137.
30. Tinoco, I. and Bustamante, C. (1999) How RNA folds. *J. Mol. Biol.*, **293**, 271–281.
31. Cruz, J.A. and Westhof, E. (2011) Sequence-based identification of 3D structural modules in RNA with RM Detect. *Nat. Methods*, **8**, 513–521.
32. Massire, C. and Westhof, E. (1998) MANIP: an interactive tool for modelling RNA. *J. Mol. Graphics Modell.*, **16**, 197–205, 255–257.

32

Molecular Dynamics Simulations of RNA Systems

Pascal Auffinger

32.1

Introduction

As a result of important methodological advances, the number of molecular dynamics (MD) simulations related to RNA systems has become really significant while still being behind those of DNA and proteic systems. Brief historical accounts of the progress in MD simulations in general and more specifically in MD simulations of RNA systems have been published [1–3].

The simulation of the dynamics of RNA systems presents specific challenges owing to the complex three-dimensional folds adopted by their highly charged polyanionic backbone, their very specific interactions with metals [4] and the large variety of naturally occurring modified nucleotides [5]. It is, thus, timely to present the current state of the art of MD simulations of RNA systems in a comprehensive and practically oriented manner. Several reviews on nucleic acid MD simulations [6–18] and other biomolecular systems [19–21] already address important methodological issues (note that the present work is an update of that published in 2005 in the first edition of this Handbook [22]).

Here, we focus on how to set up an MD simulation of an RNA system using explicit representation of solvent and Ewald summation methods for the treatment of the long-range electrostatic interactions. Specific features related to MD simulations of RNA systems using implicit solvation models can be found in [7, 23], while details on coarse-grained MD simulations can be found in [24, 25]. The simulation protocols that are described in the following text refer mainly to our own experience with the Assisted Model Building with Energy Refinement (AMBER) program and force fields with which we are most familiar and which are used in many laboratories working on nucleic acids. They should be easily transposable to other MD packages. Relevant web sites of structural databases, programs, and other tools cited in this chapter are noted in Table 32.1.

Table 32.1 List of web sites associated with structural databases, programs, and other tools cited in this chapter.

Structural databases and web tools to check experimental structures			References
PDB	http://www.rcsb.org/pdb/	Protein Data Bank	[26]
PDB_REDO	http://www.cmbi.ru.nl/pdb_redo/	A databank with updated and optimized macromolecular X-ray diffraction structures	[27]
RNAmods	http://rna-mdb.cas.albany.edu/RNAmods/	The RNA modification database	[28]
MolProbity	http://molprobity.biochem.duke.edu/	A site that provides numerous tools to evaluate X-ray and NMR structures	[29]
RNABC	http://kinemage.biochem.duke.edu/software/rnabc.php	RNA backbone correction tool	[30]
SwS	http://tatooine.u-strasbg.fr/%7Esws/SwS.html	“Solvation web service for nucleic acids”	[31]
<i>Some visualization tools</i>			
Chimera	http://www.cgl.ucsf.edu/chimera/	An extensible molecular modeling system	[32]
Jmol	http://jmol.sourceforge.net/index.en.html	An open-source Java viewer for chemical structures in 3D	N.A.
PyMOL	http://www.pymol.org/	The PyMOL visualization program	N.A.
VMD	http://www.ks.uiuc.edu/Research/vmd/	Visual Molecular Dynamics	[33]
<i>Molecular dynamics simulation packages, force fields, and parameterization tools</i>			
AMBER	http://ambermd.org/	Assisted Model Building with Energy Refinement	[34]
CHARMM	http://www.charmm.org/	A versatile and widely used molecular simulation program with broad application to many-particle systems	[35]
GROMACS	http://www.gromacs.org/	An efficient molecular dynamics simulation package	[36]
GROMOS	http://www.gromos.net/	GRONingen MOlecular Simulation computer program package	[37]
NAMD	http://www.ks.uiuc.edu/Research/namd/	A scalable molecular dynamics package	[38]
GLYCAM	http://glycam.ccruc.uga.edu/ccrc/pages/3dspt.html	A web site with AMBER force field parameters devoted to carbohydrates	[39]
R.E.D.	http://q4md-forcefieldtools.org/RED/	A method designed to calculate electrostatic charges for molecular fragments	[40]
<i>Other tools</i>			
3DNA	http://3dna.rutgers.edu/	A software suite for the analysis, rebuilding, and visualization of three-dimensional nucleic acid structures	[41]
Babel	http://openbabel.org/	A versatile format conversion program	N.A.
Curves+	http://gbio-pbil.ibcp.fr/Curves_plus/Curves+.html	A tool to analyze nucleic acid structures	[42]

N.A., not available.

32.2

MD Methods

Common force field-based simulation methods make use of an empirical potential energy function and parameter set to describe the physical properties of the system to be modeled. In MD simulations, the “possible” time evolution of the system, starting from an initial set of coordinates and velocities, is calculated by integrating Newton’s equations of motion over a given time frame [2, 43]. Descriptions of the integration algorithms associated with such programs can be found in [44–47]. These are common to all current simulation packages and will not be further discussed here. Various methods aim at expanding presently accessible time scale limits using multiple time-step integrators that are currently under development [36, 48, 49].

32.3

Simulation Setups

32.3.1

Selecting an Appropriate Starting Structure

Starting structures can be derived from four sources: (i) model-built systems, (ii) low- to high-resolution X-ray and neutron diffraction data, (iii) cryo-EM (cryo-electron microscopy) structures, and (iv) NMR data. Also a combination of pertinent experimentally derived and/or model-built molecular fragments may be used. Experimental structures can be retrieved from the PDB (Protein Data Bank) [26]. By the end of 2011, about 8% (≈ 6000 in total; ≈ 3500 with resolution ≤ 3.0 Å) of the close to 75 000 “PDB” structures were nucleic acid structures.

32.3.1.1 Model-Built Structures

These structures can be used to study specific issues related to (i) the conformation and solvation of simple RNA building blocks such as Watson–Crick base pairs and base pair steps [50, 51], (ii) the differences between RNA and DNA structures of similar sequence [50–52], or (iii) the effects associated with the introduction of natural [53, 54] and non-natural modified nucleotides [55, 56]. In order to generate larger structures, model-built fragments can also be assembled using motifs derived from X-ray and/or NMR structures [57–59]. Yet, in most cases, these structures are not reliable enough in their atomic details [60] and should consequently be used as starting structures only in exceptional instances.

32.3.1.2 X-Ray and Neutron Diffraction Structures

When derived from moderate- to high-resolution data (0.5–2.5 Å), X-ray structures certainly provide the most accurate source of starting structures, as, besides precise coordinates for the solute, they also supply important information related to its hydration and ionic environment [61, 62]. The main source of X-ray structures is

the PDB [26, 63]. In structures with resolution >3.0 Å, local regions are often relatively inaccurate and should be considered with caution. Neutron diffraction structures provide complementary information on solvation of biomolecular systems [64, 65]. Yet, only four DNA structures have been deposited in the PDB, and no neutron diffraction structure of an RNA system has yet been published.

32.3.1.3 Cryo-Electron Microscopy (Cryo-EM) Structures

Cryo-EM structures are sometimes also considered as starting points for MD simulations, as their resolution nowadays approaches that of low-resolution X-ray structures [66, 67]. These EM data are sometimes the only structural source for large systems and can, in combination with X-ray data, be used to provide “dynamical images” of processes such as the movement of transfer and messenger RNA through the ribosome [68]. Such “dynamical images” may serve as a framework for the “steering” of MD simulations [69, 70].

32.3.1.4 NMR Structures

NMR structures are similarly deposited in the PDB. Owing to the lower amount of detail contained in these structures compared to those derived from X-ray data, they are generally selected only when no X-ray data for a given system are available. Furthermore, NMR provides data that is rarely related to the hydration and ionic environment of the solute. However, as in most other fields, NMR made significant progress in recent years and an astonishing high-resolution structure of a 14-mer RNA hairpin with a cUUCGg tetraloop and a root mean square deviation of 0.3 Å has been published, giving details on the orientation of 2'-OH groups [71] that are in excellent agreement with an earlier modeling study [72]. A recent publication claims that the best currently available NMR structures are now comparable to crystal structures at 1.0 Å resolution [73].

32.3.2

Checking the Starting Structure

The imprecision associated with most model-built and NMR structures is well characterized. While X-ray structures are generally very precise, they may nonetheless enclose local errors that, if not noticed and corrected, could seriously impair the quality of the calculated MD trajectories.

32.3.2.1 Conformational Checks

For proteins, the precise orientation of several residues (essentially HIS, ASN, and GLU) cannot be correctly assessed by a single examination of the electron density maps obtained at resolutions above ≈ 1.0 Å. Some groups have addressed this issue [74] and proposed various automated solutions based on an examination of “contact maps” [29, 75] for detecting incorrect or suspicious orientations or labeling of amino acid residues such as HIS, ASP/ASN, and GLU/GLN in existing experimental structures and, subsequently, suggesting alternative conformations. In the case of HIS residues, the correct orientation can generally be deduced

from simple stereochemical considerations involving the number and type of hydrogen bonds that link these residues to neighboring ones. These methods can also be used to detect incorrect orientations of natural or modified nucleic acid residues such as pseudouridines (Ψ) that have been identified in a large number of RNA structures [76] (see also the RNA modification database [28]). Checking and correcting the orientation and labeling of such residues is therefore essential if one wishes to perform accurate simulations of protein/RNA complexes. In cases where the orientation of such residues remains ambiguous, it is good to recall that the PDB proposes the special “ASX” and “GLX” residue names for the ASP/ASN and GLU/GLN pairs. Similarly, “UNK” and “A” stand for undetermined residues and atoms, respectively. The “A” code applies generally to the terminal atoms of ASP and GLU side chains and to the ring atoms of HIS residues. The UNK code should also be used more often for solvent atoms when some ambiguity related to their nature is suspected. Thus, when working with nucleic acid structures, it is strongly advised to refer to publications and programs that have implemented controls for structural integrity [77–81]. The RNABC program aims at reducing steric clashes in RNA backbones by proposing alternative conformations [30]. It is also important to note that in RNA/protein systems, checking the starting structure for any flaws in the proteic part is as important as checking the nucleic acid part itself and/or the interface. Recently, stereochemical errors were uncovered in PDB structures and their implications for MD simulations discussed [82]. Tools for detecting such and other problems are integrated in the VMD package (Visual Molecular Dynamics [33]). At last, the files in CIF format from the PDB (partially reflected in the PDB header) contain many indications related to the correctness of a structure and should always be checked during the first stages of preparing an MD run. It has also been stated that biomolecules are dynamic and exhibit structural heterogeneity in the form of anisotropic motion and discrete conformational sub-states, often of functional importance. The authors of this study suggest that the accuracy of crystallographic structures has been widely overestimated [83].

Finally, it is worth mentioning here the PDB-REDO suite of programs and database that pursues the objective to automatically refine PDB structures, sometimes more than 30 years old, to improve their fit with existing experimental data and to obey the appropriate stereochemical constraints [27, 84, 85].

32.3.2.2 Rare Non-covalent Interactions

Non-covalent interactions are not just restricted to hydrogen bonds [86, 87] but include many rare and “exotic” contacts, most of which have remained unknown to biophysicists [88–92]. Halogen bonds belong to this category. These interactions that are of the $C-X \cdots O/N/S$ ($X = Cl, Br, I$) type were first described in the 1950s by the chemistry Nobel Prize winner Odd Hassel [93] but were found only recently to have an impact on biomolecular systems [94]. These interactions can compete in strength with regular hydrogen bonds [95]. Their main characteristic is that the $X \cdots O/N$ distance is shorter than the sum of the van der Waals radii of the involved atoms and these interactions might be classified as cases where automated checking procedures notify refinement errors. In specific instances, caution must

be exerted because some rare non-covalent interactions might also be marked as faulty contacts by structure checking programs. A non-exhaustive list of rare non-covalent interactions comprises $O/N-H \cdots \pi$ [96, 97], $C-H \cdots O/N/S/\pi$ [98, 99], cation/ π [91, 100, 101], carbohydrate/aromatic interactions [102, 103], or aromatic/aromatic interactions [104]. Besides, stacking interactions in general, and more specifically in nucleic acids, are not yet well understood [105, 106]. When very odd interactions not yet taken into account by current force fields are suspected to be present in a given system, such as halogen bonds, it is advised to either design appropriate force field patches or refrain from trying to simulate such systems.

32.3.2.3 Protonation Issues

It is now recognized that some nucleic acid residues such as adenines or cytosines can be found in their protonated state and that protonation of such bases can contribute significantly to RNA catalytic mechanisms [107–111]. Yet, X-ray structures rarely provide direct evidence for such protonation states. These have to be derived from chemical intuition by inspecting the structural environment of the involved residues. If the most probable protonation state cannot be identified, it is advisable to perform simulations of both the protonated and unprotonated states [112–115]. Evidently, protonation states of charged amino acids and of drugs such as aminoglycosides [116, 117] have to be clearly defined before starting a simulation [118]. A recent publication discusses the effects induced by the use of different protonation states on MD trajectories [119]. Note that numerous pK_a prediction methods have been developed for ligands and amino acids [120].

Different experimental pH conditions can also alter the structure of a given molecular fragment. For example, the NMR structure of a tRNA^{Lys} anticodon stem-loop obtained at acidic pH is different from those extracted from tRNA crystal structures, essentially because of the formation of a wobble A^+-C pair [121] instead of a more standard (in this context) bifurcated A-C pair [122]. Consequently, great care must be taken in choosing a starting structure that genuinely reflects the experimental conditions that one wants to reproduce.

32.3.2.4 Solvent

The interpretation of solvent density maps derived from X-ray experiments are often difficult. A recent PDB survey has revealed that it is not unusual for water molecules to be assigned to sites with very large electron densities (beyond those expected for a water molecule), which may be indicative of the presence of a metal ion at those locations [77]. Recently, it has also been emphasized that some density spots resulting from the presence of anions (SO_4^{2-} , Cl^- , etc.) in the vicinity of nucleic acids might have been wrongly assigned to divalent cations (Mg^{2+} or Mn^{2+}) [123]. Incorrectly assigned metal ions can represent an important source of errors in MD simulations. Progress in the unambiguous assignment of SO_4^{2-} anions in crystallographic structures has been made through the use of SeO_4^{2-} substitutes [124]. Statistical data related to the structure of the first solvation shell around the most common nucleic acid base pairs should contribute to a better

understanding of crucial features of water and ion binding to nucleic acids (see the SwS site or Solvation web Service for nucleic acids [31]).

32.3.3

Adding Hydrogen Atoms

Most of the structures deposited in the PDB, with the exception of ultrahigh resolution X-ray and neutron diffraction structures [125–127], are devoid of hydrogen atoms. Hence, before starting a simulation, it is essential to add them using either the tools provided by the main MD packages or external programs such as Mol-Probity [29, 75] or visualization programs such as PyMOL or VMD. As mentioned before, one has to be aware of protonation issues that are, up to now, not generally addressed for nucleic acids by such programs.

32.3.4

Choosing the Environment (Crystal, Liquid) and Ion Types

Two types of environment may be chosen: the crystal or the liquid phase. Simulations of the crystal phase are rare for nucleic acids [128–136] even though such simulations probably provide the most accurate comparison between experimental and theoretical data [137]. Simulations of RNA systems in the aqueous phase is the most common approach. The ions used to neutralize the nucleic acid can be of various types (Na^+ , K^+ , NH_4^+ , Mg^{2+} , etc.) [18] and different ionic conditions can be selected, ranging from a number of ions sufficient to neutralize the charge carried by the nucleotidic backbone to a number of ions that would appropriately represent a solution containing 0.1–1.0 M or more of added salt. In this last instance, co-ions (Cl^- , SO_4^{2-} , etc.) must be placed in the simulation box as well. Note that anions can, under specific conditions, break into the first solvation shell of nucleic acids and should, consequently, be included in the modeling scheme [123].

32.3.5

Setting the Box Size and Placing the Ions and Water

32.3.5.1 Box Size

For simulations of the crystal phase, the box size is determined by the size of the crystal cell [138]. In solution, the box size is usually allowed to extend by approximately 8–12 Å around the solute, thus ensuring three to four hydration layers in each direction.

32.3.5.2 Monovalent Ions

Monovalent ions are then placed in the simulation box by replacing the water molecules with the lowest (for cations) and highest (for anions) electrostatic potential [50, 139]. If one wants to avoid biases because of the initial placement of ions, it is also essential to set limits ensuring that no ions are closer than a certain

distance (usually 5–6 Å) to any nucleic acid atom or to any other ion. In this manner, no ions are initially found in nucleic acid grooves, in direct proximity of the phosphate groups, or in positions where they can form ion pairs.

32.3.5.3 Divalent Ions

The limited length of the simulations does not allow reproduction of dehydration processes of Mg^{2+} ions and formation of direct $\text{Mg}^{2+} \cdots \text{O}$ contacts with the RNA that take microseconds to occur in bulk solvent [140, 141]. Thus, divalent ions cannot be placed in the simulation box as described above. Such systems can only be realistically simulated if the Mg^{2+} ions occupy their crystallographic positions already at the beginning of the MD run [142–146]. It has to be noted that, while the Mg^{2+} ions occupy fixed positions, the surrounding water molecules are allowed to rapidly find their equilibrium positions in the first hydration shell of the ion during the equilibration phase. They form a hexahydrated cation when the cation is far from the solute or a $\text{Mg}(\text{H}_2\text{O})_n$ cation ($n < 6$) when the cation is already in close contact with one or more solute atoms. A more realistic and complete setting takes into consideration the divalent (Mg^{2+}) and monovalent (K^+) cations, as well as anions (Cl^-) as in a study of the 5S loop E motif [142].

32.3.5.4 Minimal Salt Conditions

From a historical point of view, most MD simulations of nucleic acids were performed in “minimal salt” conditions implying the placement of a number of monovalent cations just sufficient for neutralizing the negatively charged nucleic acid backbone. Yet, under “*in vitro*” and “*in vivo*” conditions, cations and anions are present in excess, and there is no further justification for using minimal salt conditions if one is careful in choosing appropriate force field parameters for the ionic particles [147, 148]. Furthermore, using a finite ionic concentration (from 0.1 to 1.0 M or even above) allows a much better sampling of the configurational space accessible to these ions.

32.3.5.5 Water Molecules

Water is an essential part of biomolecular structure, function, and dynamics [149]. As such, it is also found at RNA/drug interfaces [118] and supposed to play an important role in recognition phenomena [150, 151]. The accurate placement of water molecules at biomolecular interfaces should thus matter for those involved in drug design studies [152].

32.3.5.6 Building Initial Solute and Solvent Configurations

Most simulation packages provide programs for helping to place all the relevant molecules in the simulation box. “Packmol” is an independent package that provides a large set of options to achieve this task [153]. Besides, the “Simulaid” package can also be of help in the early stages of setting up an MD simulation and for analyzing the trajectories [154].

32.3.6

Choosing the Program and Force Field**32.3.6.1 Programs**

Two widely used MD simulation packages are AMBER (current version AMBER 11; [34, 155]) and CHARMM (Chemistry at HARvard Molecular Mechanics; [35]). Each of those is associated with its own force field. Other non-commercial MD packages such as GROMOS [37], GROMACS [36], or NAMD [38] are also available. Package choice is a very subjective matter, usually governed by traditional preferences of the respective research group.

32.3.6.2 Force Fields

AMBER [155] and CHARMM [134, 135] are the most commonly used force fields for simulating the dynamics of nucleic acids and proteins [156]. They coexist with the less frequently used OPLS [157] and GROMOS [158] force fields. Yet, it is essential to be aware of their limitations. Although they are all able to reproduce quite well most of the important structural and dynamic features of biological systems, they do differ in specific aspects. Early comparisons of the AMBER and CHARMM force fields performed on DNA duplexes [159, 160] have led to the conclusion that, in some instances, a “force field-dependent polymorphism” [6] is observed. This polymorphism contributes to an artifactual drift of the duplexes toward an A- or B-form that is independent of the starting conformation. Some of these problems have been corrected in recent versions of the AMBER and CHARMM force fields [161, 162], while new issues have been raised; see, for example, the unnatural “ladder-like” geometries for hairpin stems appearing in some RNA simulations and the attempts to understand and correct them [163–165].

Thus, as force fields are constantly evolving, it is recommended to check for the latest available updates and patches while keeping in mind that most of these modifications will have to be thoroughly tested before being accepted by the majority of researchers in the community. In this respect, see the pyramidalization issue at the glycosidic nitrogens in purine bases [166] or the modifications related to the torsional barrier around the χ -axis [165, 167, 168].

It can be concluded that, up to now, none of the widely used force fields are able to fully reproduce all of the nucleic acid structural features experimentally observed in solution [10, 162]. For instance, the dynamics of Z-DNA in solution are still difficult to simulate and a recent modification of the parameters for the χ torsion angle is recommended for RNA but not B-DNA [165]. Many recent studies have evaluated the respective impact of available force fields on MD runs for proteins [138, 169] and nucleic acids [163, 164, 170–174].

Polarizable force fields [175–178], such as those recently implemented in AMBER, CHARMM, or the AMOEBA force field [179], are expected to significantly improve the informative and predictive power of MD simulations (although at the expense of a requirement for expanded computer resources). These force fields are still at an initial stage of development and will have first to be thoroughly parametrized and tested on small systems.

32.3.6.3 Parameterization of Modified Nucleotides, Ligands, and Ions

Most empirical force fields used in common MD packages are equipped with parameters for modeling all canonical nucleotides and amino acids but not modified nucleotides and ligands. For simulation of such systems, new parameter sets must be developed. Although some general sets of procedures for automated parameterization based on fitting to both experimental and quantum chemical data have been proposed, force field parameterization remains a matter for experts. A classical review on force field parameterization has been written by Dinur and Hagler [180]. Fortunately, some tools and methods for generating reasonable to high-quality force field parameters are available [181, 182]. Among them, a new set of scripts (R.E.D.) has been made available to facilitate the derivation of partial charge sets from electrostatic potential calculations [40]. In addition to those for nucleic acids and amino acids, some force fields have been developed for specific classes of molecules, such as carbohydrates (OPLS-AA [183]), sulfates and sulfamates [39], polyphosphates [184], or small organic molecules [185]. Ion parameters compatible with the chosen force field can easily be retrieved from the literature [186] and polarizable force fields are being developed for monovalent [187] and divalent cations [188, 189]. Finally, it is worth mentioning a study that developed a set of parameters for the 107 modified nucleotides known to occur in nature [190] and for thio-substituted nucleotides [191].

32.3.6.4 Clustering Artifacts and Ion Parameters

As just mentioned, parameters for the ions that are compatible with the chosen force field can be easily retrieved from the literature. Most of the recent parameters take care of a recurrent issue in the parameterization of monovalent ions affecting numerous force fields [148, 186]. It was reported that when Na^+ or K^+ parameters were used, ionic clusters start to form and grow rapidly in the vicinity of biomolecules even at physiological concentrations ($\approx 0.25\text{M}$ of KCl) [118, 192, 193], thus defying current knowledge related to dilute ionic solutions. When more recent parameters for monovalent cations are used (keeping everything else unchanged) [148, 186], no such ionic clusters are observed even on the nanosecond scale. Users should be especially careful about the quality of parameters associated with these “ubiquitous” and essential ionic species.

32.3.6.5 Water Models

Among the available water models, the TIP3P (Transferable Interatomic Potential 3-Point [194]) model is generally associated with the AMBER force field. The computationally more “expensive” four- (TIP4P [194]) and five- (TIP5P [195]) point models have been developed and tested on various systems. SPC (Single Point Charge [196]) and SPC/E (SPC/Extended [197]) models are also very popular. For DNA, it has been shown that the use of TIP3P and SPC/E water models resulted in comparable hydration patterns although the water densities associated with the TIP3P model that applies a higher diffusion rate appear to be more blurred

than those calculated with the SPC/E model that utilizes diffusion rates in better agreement with experimental data [12]. Besides, the MD simulation of a protein in a crystal environment has led to the conclusion that results obtained with the SPC/E model are in much better agreement with neutron-scattering data than those collected with the TIP3P model [198]. Thus, a particular water model can significantly alter the results of MD simulations, and more studies are needed to precisely evaluate the influence of such models. One must also consider that the results obtained by one or the other model are strongly affected by the type of treatment of the electrostatic interactions [199]. More recent simulations of a protein crystal established that the choice of a water model, although not being neutral, matters less than the choice of an appropriate force field [138].

32.3.7

Treatment of Electrostatic Interactions

The ensemble of parameters in the above referenced force fields is usually considered to be the most important factor in determining the quality of an MD trajectory. Yet, if the empirical equation that models the intermolecular forces at play in biomolecular systems is not accurate enough or if one of its mathematical terms is not correctly evaluated, then even the most precisely developed set of parameters cannot realistically account for the dynamics of these systems. Indeed, numerous studies have revealed that severe artifacts emerge when the long-range electrostatic contributions are neglected, that is, when only a part of the Coulomb term $q_i q_j / 4\pi\epsilon_0 r_{ij}$ is estimated [200–202]. It has been shown that an accurate treatment of the long-range electrostatic interactions is of paramount importance for generating realistic trajectories of nucleic acid systems. The Particle Mesh Ewald (PME) summation method [203, 204] has proven to lead to very stable nucleic acid trajectories [202, 205] and to be efficient in correcting layer structure artifacts in simulations of liquid water [199]. Further improvements of the Ewald summation method are still under development [138].

The program default parameters of MD programs are generally adequate for using the PME method. These parameters are such that a cubic interpolation scheme and a 10^{-5} Å tolerance for the direct space sum cutoff is chosen. To speed up the fast Fourier transform in the calculation of the reciprocal sum, the size of the PME charge grid is chosen to be an integer power of 2, 3, and 5 and to be slightly larger than the size of the periodic box. This leads to grid spacing of ≈ 1 Å or less.

32.3.8

Other Simulation Parameters

32.3.8.1 Thermodynamic Ensemble

MD simulations can be conducted in a microcanonical (N,V,E) ensemble with constant number of particles, volume and energy. Yet, the isothermal–isobaric or

(N,P,T) ensemble for constant number of particles, pressure and temperature is more commonly chosen. Working at constant pressure is particularly important during the equilibration step because during the building stage of the system, the water molecule positions are not optimized at the solute–solvent interface. This can lead to the generation of holes in the simulation box when constant volume options are selected.

Attempts to perform simulations in the “grand canonical ensemble” are rare given their complexity [206], as in this thermodynamic ensemble, the number of particles is allowed to change. There is little likelihood that such methods, despite the great interest they attract, will be used for biomolecular simulations in the near future.

32.3.8.2 Temperature and Pressure

In order to maintain a constant temperature and pressure, the Berendsen temperature coupling scheme and an isotropic molecule-based constant pressure scaling with a time constant of 0.2–0.5 ps for both parameters are often used [207], although other coupling schemes and time constants are found in the literature.

The target temperature and pressure values are usually set to 1 atm and 298 K or 25 °C also called “room temperature.” At this point, it is interesting to note that the choice of this target temperature is dictated by early studies that used experimental data obtained by chemists to parameterize force fields. This is close to the temperature at which many *in vitro* experiments are performed in a laboratory. Nonetheless, simulations using different target temperatures (5, 25, and 37 °C, for example) have been used to reveal some differences in the hydration of RNA systems [208].

32.3.8.3 Shake, Time Steps, and Update of the Non-bonded Pair List

A tolerance of 0.0005 Å is generally used for the SHAKE algorithm [209] that allows one to “safely” use a 2-ps time step (instead of 1 ps) by artificially freezing the most rapid vibration motions (C–H, O–H, N–H, etc. elongations) observed in biomolecular systems. A high-tolerance value has been found to reduce the so-called “flying ice cube” phenomenon described below [210, 211]. Another time-saving device is associated with the use of a pair list for calculating non-bonded interactions between all atoms in the system. This pair list is not calculated at each time step, which would be computationally expensive, but only once every *n*th step (usually every 10 steps).

32.3.8.4 The “Flying Ice Cube Problem”

“Since the non-bonded pair list is not updated every step (in order to save CPU resources) and constant pressure algorithms as well as uniform scaling of velocities by the Berendsen coupling scheme are used, some small energy drain during the simulation can occur and the center of mass velocity can slowly grow. Therefore, periodically in the simulation this center of mass velocity has to be removed otherwise the “flying ice cube” syndrome may appear [210, 211]. It is also advised to update the non-bonded pair list more frequently [210, 211].”

The preceding paragraph describes the “flying ice cube” issue, as it was reported before 2005. It is satisfying to note that the current literature no longer mentions this syndrome suggesting that it can be considered as no longer relevant.

32.3.9

Equilibration

The main aim of equilibration procedures is to alleviate tensions created in the system during the early building stages. These tensions, if neglected, would lead to unrealistic trajectories. In the following, a typical equilibration protocol used by us will be described.

First, 500 steps of steepest descent minimization are applied to the entire system with periodic boundary conditions in the (N,P,T) ensemble. This is followed by 25 ps of simulation in the same thermodynamic ensemble, where only the water molecules and the hydrogen atoms of the solute are allowed to move (the heavy atoms of the solute and the ions are frozen by the BELL option in AMBER). During the next 25 ps, the constraints placed on the monovalent ions are released so that they can start to equilibrate around the RNA system, while the BELL option is still used to freeze the heavy atoms of the solute. Then, several rounds of 50-ps MD simulations are performed during which positional constraints of 10, 5, 2, 1, 0.5, 0.1, 0.01, and 0.001 kcal mol⁻¹ Å² are applied to the heavy atoms of the solute, yielding a partially constrained 400-ps MD trajectory. The equilibration phase is generally extended by another 0.5–1.0 ns in order to allow a better sampling of the conformational space adopted by the solute and the ions [50]. Then the production phase can take place.

The length of the equilibration phase can be extended or shortened at will. However, it has to be noted that the phase during which constraints are applied ensures that the solvent is appropriately equilibrated in the vicinity of the solute and throughout the simulation box. The part of the simulation during which no constraints are applied then allows the solute to relax under the influence of a relatively well equilibrated solvent environment. Other equilibration protocols start at low temperature (50 K); then, the temperature is raised to the target temperature at discrete time intervals [145]. An apparent advantage of the procedure described in the preceding paragraph is that the sampling of the conformational space by the solvent is conducted at the target temperature and, hence, probably more efficient.

32.3.10

Sampling

32.3.10.1 How Long Should a Simulation Be?

Ideally, MD simulations should be long enough to sample all the conformational transitions occurring in a particular biomolecular system. In an earlier version of this review published in 2005, we wrote: *“Given present and predictable computational means, this goal will certainly not be achieved in the next decade. It is, thus, impossible to address in the near future biochemical processes with very long relaxation times by using*

classical MD simulations with explicit solvent representations. Up to now, the longest MD simulations of RNA systems have barely overcome the 10 ns time scale.” Yet, owing to the huge progress made in computational science and the use of GPUs (Graphical Processor Units) or high-potential computational facilities, almost unforeseeable progress has been made. At present, 1–100 μ s and even millisecond-long MD simulations of small proteins with explicit solvent are possible for studying folding processes [212–215] or ligand-binding mechanisms in an *ab initio* manner [216]. Simulations of large molecular systems also made stunning progress and 100-ns MD trajectories of multimillion atom systems such as ribosomes [66, 217] and complete virus particles are no longer exceptional [218]. By extrapolation, it has been suggested that 1-s simulations will be at reach by 2015 [2]. Thus, at first glance and assuming perfection of the software tools, it could be stated that a simulation should be just as long as one can afford. Yet, in practice, we have still to cope with significant force field issues or in other words, “*With modern computing capabilities, however, another round of modifications and improvements to molecular modeling force-fields is clearly required to maintain an accurate description of the simulated systems*” [219].

Although force fields have been significantly improved since the beginning of the MD journey, error propagations manifest themselves on long-time scales by altering the structure of a system that should remain stable under equilibrium conditions. Folding of proteins also fails sometimes, even with the best sampling technologies currently available [169, 219]. For a microsecond DNA simulation, global deformations of a duplex were optimistically interpreted as suggesting “... *that the intrinsic flexibility of DNA has been carefully designed by evolution to exploit its biological function* ...” [220].

Nevertheless, many processes associated with limited conformational perturbations lie within the reach of present methods. Among them, those that involve the binding of the smallest ligands to nucleic acids (monovalent ions and water molecules [50, 51, 142]) or the orientation of the 2'-OH hydroxyl groups [72] can trustfully be evaluated on the current nanosecond time scale. Sampling techniques have been recently evaluated [221].

32.3.10.2 When to Stop a Simulation

Another important and often overlooked issue is: *when to stop an MD simulation or when does it start to be a waste of time to continue sampling?* Indisputably, a very large number of approximations are included in MD simulations. These approximations are, among others, related to (i) imprecision in the force field parameters, (ii) incomplete evaluation of the intermolecular forces, (iii) neglect of polarization and charge transfer effects, and (iv) an incomplete knowledge of the starting conditions. This last point by itself can easily lead to the generation of partially or totally unrealistic simulations. For example, an instability of the tertiary core structure of the yeast tRNA^{Asp} molecule characterized by a reordering of several base triples was observed during a 500-ps MD simulation [222]. This had first been attributed to the absence of Mg²⁺ ions in the model. Later on, it was proposed that an adenine involved in the above-mentioned base triples is protonated [112].

Probably, both effects led to the calculated transitions and further sampling would, in this case, not have contributed to a better understanding of the structure and the dynamics of this molecule. Hence, there is a certain risk for long MD simulations (depending on the applied force field and MD package) to oversample regions of the “configurational space” that are not overlapping with the “true” configurational space adopted by the investigated system. A further example is provided by the formation of salt crystals in undersaturated conditions resulting from the use of inappropriate ionic parameters [147].

32.3.10.3 Multiple Molecular Dynamics (MMD) Simulations

An alternative to long MD simulations is to generate an ensemble of several “shorter” trajectories using the Multiple Molecular Dynamics (MMD) simulation technique. This method introduces slight perturbations (such as different initial velocities) in the initial conditions and, thus, exploits the chaotic nature of MD simulations in order to generate several uncorrelated trajectories [223–231]. It has been stated that several trajectories generated from similar but slightly different starting conditions may provide, on a statistical basis, more information than a single long trajectory. A parallel view developed by other authors consists of generating and comparing several trajectories obtained by starting from different initial configurations [229]. In short, MMD methods allow a statistical evaluation of MD simulations that cannot be achieved from the analysis of a single trajectory [202].

The effects of modifying various starting parameters (velocity; solvation; location of protons; conformation of HIS, ASN, GLN residues; protonation and titration of HIS residues; treatment of alternative conformations) to enhance conformational sampling have been recently evaluated [119, 232, 233].

32.3.10.4 Simulations of Large Systems

It has always been very attractive to simulate large systems, and simulations of multimillion atom systems are now possible [217]. Yet, such system present very specific challenges associated with their size. As it is already difficult to spot structural incongruences in small RNA or DNA systems, it is even much more so in very large systems for which it seems almost impossible to follow the paths taken by each of their molecular fragments and to assess their correctness. The only viable option here seems to improve force fields by analyzing MD runs conducted on much smaller systems and to transfer the gained knowledge to large-scale simulations.

32.4

Analysis

32.4.1

Evaluating the Quality of the Trajectories

Before trying to extract information from MD simulations, it is imperative to evaluate the quality of the generated trajectories by checking their internal consistency and accordance with all available experimental data [20, 21, 234].

32.4.1.1 Consistency Checks

Internal consistency checks are related to the detection of unphysical behaviors such as those resulting from the use of truncation methods for the evaluation of electrostatic interactions [199–201], such as the “flying ice cube” syndrome [210, 211] or the artifactual aggregation of monovalent cations [147]. When such issues come to the foreground, it is clear that the simulation can and should no longer be used to derive biologically relevant information. Rather, the focus of the study must shift toward the correction of the detected artifacts. By doing so, the importance of the neglected contributions is generally brought to light. For example, the stabilizing role of “hydration forces” could be assessed by comparing trajectories from simulations that include or discard long-range electrostatic interactions [6, 202, 225]. Consequently, “*in silico*” experiments in which important components of natural systems are switched “on” or “off” should reveal the effects linked to the presence or absence of the respective atomic interactions.

32.4.1.2 Comparison with Experimental Data

Confrontation with available experimental data (that are most frequently of a structural type) is in all cases mandatory. For instance, if important tertiary interactions present in the starting structure break during the simulation over short time scales, it is probable that the simulation protocols are not adequate and the reasons underlying such behavior must be uncovered [6, 201]. Early simulations performed under “*in vacuo*” conditions or with truncation of the electrostatic interactions and where disruptions of important tertiary interactions were observed clearly illustrate this point [201, 235]. Moreover, if the structure of the investigated system is significantly diverging, in part or entirely, from the initial structure during the time course of the simulation, it must be considered that this observation may result from force field inaccuracies. Only comparisons with experimental data on the same or related systems can help to resolve such issues.

The CASP (Critical Assessment of techniques to protein Structure Prediction) experiment [236] represents a good attempt to compare theoretical data with experiments. An assessment of CASP results performed in 2008 (CASP8) led to the conclusion that there is still ample opportunity for methods development in structure and function predictions [2]. Similar attempts to assess the performance of RNA folding methods [237] are currently under development. Yet, the fact that we are currently unable to solve this problem speaks in itself about our current knowledge of the forces that structure biomolecular systems and emphasizes the necessity to pursue force field developments.

32.4.1.3 Visualization

Visual checks are also mandatory and can be performed using the programs that are delivered with each main MD package or using programs such as VMD [33] or Chimera [32] that can read most of the available MD outputs. In rare cases, the Open BABEL utility is needed in order to convert data written in different formats. PyMOL is also a very efficient and complete visualization tool for experimental structures.

32.4.1.4 Validation through Statistical Survey of Structural Databases

It is necessary to develop tools enabling the starting structure of an MD run to be checked as well as the resulting trajectories. We developed the “Solvation web Service for nucleic acids” also called SwS [31], that is based on a statistical survey of crystallographic structures deposited in the PDB. SwS is designed to provide an exhaustive overview of the solvation of nucleic acid structural elements through the generation of three-dimensional solvent density maps. However, it is important to note that such databases contain some inaccurate information concerning, for example, misassigned solvent densities. Hopefully, structural databases, if well maintained, “heal” themselves through a process of constant reinterpretation of available experimental data [27, 85].

32.4.2

Convergence Issues

Next, the convergence of the simulation with respect to the investigated properties has to be assessed. For this, it is necessary for the simulation length to exceed the average relaxation time of the investigated property, which is not necessarily the longest relaxation time of the system that can largely exceed the currently accessible time scales [234, 238]. For example, it is possible to estimate on a statistical basis the residence times of water molecules (between 10 and 500 ps) around a regular RNA helix from nanosecond MD simulations, although the statistics will obviously be less good for the water molecules with the longest residence times [50, 51]. For less regular RNA systems, water molecules may be trapped in specific pockets for much longer times and only lower bounds for the residence times can be obtained [239]. A similar problem is associated with the estimation of the binding properties of monovalent and divalent ions. While it is feasible to estimate the exchange time of monovalent ions located in RNA grooves from nanosecond MD simulations [143], it is impossible on such time scales to estimate the same values for divalent ions [142].

32.4.3

Conformational Parameters

Most of the analysis programs delivered with major MD packages (such as “cpptraj” for AMBER) can monitor nearly all of the main conformational parameters, such as interatomic distances, angles, dihedral angles, and also parameters that are specific to nucleic acids, for example, sugar puckers or other user-defined variables. Programs such as CURVES [42] or 3DNA [41], which allow one to estimate parameters specific for nucleic acids, such as bending, shift, slide, rise, tilt, roll, twist, and so on, can also be very useful in the analysis process. The majority of these parameters are now attached to the experimental structures deposited at the PDB.

32.4.4

Data Analysis32.4.4.1 **Clustering**

The huge quantity of data generated by current MD runs is also a challenging issue. In this respect, conformational clustering methods can be of interest. These techniques allow classifying frequently occupied conformations with respect to user-defined criteria [219]. Yet, it has been found that geometric clustering strongly depends on the type of algorithms that are used [240]. A new algorithm (nMDS) based on PCA (Principal Component Analysis) and a non-metric multidimensional scaling has been presented [241].

32.4.4.2 **Analysis Packages**

Besides the analysis packages associated with the main MD or visualization programs [242, 243], independent packages such as “Simulaid” [154] and “MD-analysis” [244] can provide interesting options to fulfill this task.

32.4.4.3 **Solvent Analysis**

Solvent analysis is an important aspect of the evaluation of MD trajectories [14]. It is related to the characterization of water- and ion-binding sites and to the estimation of their residence times. Several reviews have already addressed these issues for nucleic acid systems [237, 239, 245–247], and we will not discuss them in further detail. The SwS web service [31] provides a statistical overview of the solvation of nucleic acid base pairs that is useful for better understanding crystal structures and analyzing MD results.

32.5

Perspectives

The current simulation protocols are now able to deliver stable trajectories of RNA systems ranging from small duplexes to large RNA/protein complexes over nanosecond to microsecond and even millisecond time scales. Important conformational transitions and solvation processes can thus be studied in detail. The current trend is to head toward longer time scales and multimillion atom systems. This would allow addressing, among others, folding, unfolding, and important molecular recognition issues. Nevertheless, one should be aware that by extending the currently accessible time scales, new artifacts that remain up to now hidden will emerge and necessitate the development of refined force fields and simulation methods that will evidently incorporate polarizability and charge transfer effects. The detection and correction of artifacts that will manifest themselves on the longer time scales must be regarded as an unconditional modeler duty. In this respect, new experimental data related to the structure and dynamics of RNA systems associated with a better knowledge of non-covalent interactions will be of paramount importance for the future development of MD techniques.

Yet, it is strongly advised to carefully check the starting experimental structure as, in most cases, MD simulations will expand initial errors rather than correcting them. As stated by colleagues, “. . . *advanced programs and simulation methods like MD and folding are far from generally applicable automated procedures; their application relies on user expertise as well as biological intuition.*” [2].

Acknowledgments

P.A. wishes to thank Prof. Eric Westhof for ongoing support.

References

1. Levitt, M. (2001) The birth of computational structural biology. *Nat. Struct. Biol.*, **8**, 392–393.
2. Schlick, T., Collepardo-Guevara, R., Halvorsen, L.A., Jung, S., and Xiao, X. (2011) Biomolecular modeling and simulation: a field coming of age. *Q. Rev. Biophys.*, **44**, 191–228.
3. Hashem, Y., Westhof, E., and Auffinger, P. (2008) in *Computational Structural Biology*, vol. 13 (eds T. Schwede and M.C. Peitsch), World Scientific Publishing Co. Pte. Ltd, London, pp. 363–399.
4. Auffinger, P., Grover, N., and Westhof, E. (2011) Metal ion binding to RNA. *Met. Ions Life Sci.*, **9**, 1–35.
5. Grosjean, H. and Benne, R. (1998) *Modification and Editing of RNA*, ASM Press, Washington, DC, 2005.
6. Auffinger, P. and Westhof, E. (1998) Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.*, **8**, 227–236.
7. Zacharias, M. (2000) Simulation of the structure and dynamics of nonhelical RNA motifs. *Curr. Opin. Struct. Biol.*, **10**, 311–317.
8. Norberg, J. and Nilsson, L. (2002) Molecular dynamics applied to nucleic acids. *Acc. Chem. Res.*, **35**, 465–472.
9. Giudice, E. and Lavery, R. (2002) Simulations of nucleic acids and their complexes. *Acc. Chem. Res.*, **35**, 350–357.
10. Cheatham, T.E. III and Young, M.A. (2000) Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers*, **56**, 232–256.
11. Cheatham, T.E. III and Kollman, P.A. (2000) Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.*, **51**, 435–471.
12. Cheatham, T.E. III and Kollman, P.A. (1998) in *Structure, Motion, Interaction and Expression of Biological Macromolecules, Proceedings of the Tenth Conversation, State University of New York, Albany, 1997* (eds R.H. Sarma and M.H. Sarma), Adenine Press, New York, pp. 99–116.
13. Hashem, Y. and Auffinger, P. (2009) A short guide to molecular dynamics simulations of RNA systems. *Methods*, **47**, 187–197.
14. Auffinger, P. and Hashem, Y. (2007) Nucleic acid solvation: from outside to insight. *Curr. Opin. Struct. Biol.*, **17**, 325–333.
15. MacKerell, A.D. Jr. and Nilsson, L. (2008) Molecular dynamics simulations of nucleic acid-protein complexes. *Curr. Opin. Struct. Biol.*, **18**, 194–199.
16. Ditzler, M.A., Otyepka, M., Spomer, J., and Walter, N.G. (2010) Molecular dynamics and quantum mechanics of RNA: conformational and chemical change we can believe in. *Acc. Chem. Res.*, **43**, 40–47.
17. McDowell, S.E., Spackova, N., Spomer, J., and Walter, N.G. (2007) Molecular dynamics simulations of RNA: an in silico single molecule approach. *Biopolymers*, **85**, 169–184.

18. Auffinger, P. (2012) Ions in molecular dynamics simulations of RNA systems, in *RNA 3D Structure Analysis and Prediction* (eds E. Westhof and N.B. Leontis), Springer Verlag, pp. 299–318.
19. Adcock, S.A. and McCammon, J.A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.*, **106**, 1589–1615.
20. van Gunsteren, W.F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D.P., Glättli, A., Hünenberger, P.H. *et al.* (2006) Biomolecular modeling: goals, problems, perspectives. *Angew. Chem. Int. Ed.*, **45**, 4064–4092.
21. van Gunsteren, W.F., Dolenc, J., and Mark, A.E. (2008) Molecular simulation as an aid to experimentalists. *Curr. Opin. Struct. Biol.*, **18**, 149–153.
22. Auffinger, P. and Vaiana, A.C. (2005) in *Handbook of RNA Biochemistry*, (eds R.K. Hartmann, A. Bindereif, A. Schön, and E. Westhof), Wiley-VCH Verlag, Weinheim, pp. 560–576.
23. Dong, F., Wagoner, J.A., and Baker, N.A. (2008) Assessing the performance of implicit solvation models at a nucleic acid surface. *Phys. Chem. Chem. Phys.*, **10**, 4889–4902.
24. Poursina, M., Bhalerao, K.D., Flores, S.C., Anderson, K.S., and Laederach, A. (2011) Strategies for articulated multibody-based adaptive coarse grain simulation of RNA. *Methods Enzymol.*, **487**, 73–98.
25. Xia, Z., Gardner, D.P., Gutell, R.R., and Ren, P. (2010) Coarse-grained model for simulation of RNA three-dimensional structures. *J. Phys. Chem. B*, **114**, 13497–13506.
26. Rose, P.W., Beran, B., Bi, C., Bluhm, W.F., Dimitropoulos, D., Goodsell, D.S., Prlic, A., Quesada, M., Quinn, G.B., Westbrook, J.D. *et al.* (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.*, **39**, D392–D401.
27. Joosten, R.P., Joosten, K., Cohen, S.X., Vriend, G., and Perrakis, A. (2011) Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank., *Bioinformatics*, **27**, 3392–3398.
28. Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A., Fabris, D., and Agris, P.F. (2011) The RNA modification database, RNAMDB: 2011 update., *Nucleic Acids Res.*, **39**, D195–D201.
29. Chen, V.B., Arendall, W.B. III, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography., *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
30. Wang, X., Kapral, G., Murray, L., Richardson, D., Richardson, J., and Snoeyink, J. (2008) RNABC: forward kinematics to reduce all-atom steric clashes in RNA backbone. *J. Math. Biol.*, **56**, 253–278.
31. Auffinger, P. and Hashem, Y. (2007) SwS: a solvation web service for nucleic acids. *Bioinformatics*, **23**, 1035–1037, <http://www-ibmc.u-strasbg.fr/arn/sws.html>.
32. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004) UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
33. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD - visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
34. Case, D.A., Darden, T.A., Cheatham, T.E. III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M. *et al.* (2012) *AMBER 12*. University of California, San Francisco.
35. Brooks, B.R., Brooks, C.L. III, Mackerell, A.D. Jr., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S. *et al.* (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545–1614.
36. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: algorithms for highly efficient,

- load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435–447.
37. Christen, M., Hünenberger, P.H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D.P., Heinz, T.N., Kastenholz, M.A., Krautler, V., Oostenbrink, C. *et al.* (2005) The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.*, **26**, 1719–1751.
 38. Phillips, J.C., Braun, R., Wang, W., HGulbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
 39. Huige, C.J.M. and Altona, C. (1995) Force field parameters for sulfates and sulfamates based on *ab initio* calculations: Extensions of AMBER and CHARMM fields. *J. Comput. Chem.*, **16**, 56–79.
 40. Dupradeau, F.Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., Lelong, D., Rosanski, W., and Cieplak, P. (2010) The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.*, **12**, 7821–7839.
 41. Lu, X.J. and Olson, W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.
 42. Blanchet, C., Pasi, M., Zakrzewska, K., and Lavery, R. (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures *Nucleic Acids Res.*, **39**, W68–W73.
 43. van Gunsteren, W.F., and Berendsen, H.J.C. (1990) Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed.*, **29**, 992–1023.
 44. Allen, M.P. and Tildesley, D.J. (1987) *Computer Simulation of Liquids*, Clarendon Press, Oxford.
 45. Rapaport, D.C. (1995) *The Art of Molecular Dynamics Simulation*, Cambridge University Press, Cambridge.
 46. McCammon, J.A. and Harvey, S.C. (1987) *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, New York.
 47. Leach, A.R. (2001) *Molecular Modeling: Principles and Applications*, Pearson Education Limited, Essex.
 48. Schlick, T. (2001) Time-trimming tricks for dynamics simulations: splitting force updates to reduce computational work. *Structure*, **9**, R45–R53.
 49. Morrone, J.A., Zhou, R., and Berne, B.J. (2010) Molecular dynamics with multiple time scales: how to avoid pitfalls. *J. Chem. Theory Comput.*, **6**, 1798–1804.
 50. Auffinger, P. and Westhof, E. (2000) Water and ion binding around RNA and DNA (C,G) oligomers. *J. Mol. Biol.*, **300**, 1113–1131.
 51. Auffinger, P. and Westhof, E. (2001) Water and ion binding around r(UpA)₁₂ and d(ApT)₁₂ oligomers - Comparison with RNA and DNA (CpG)₁₂ duplexes. *J. Mol. Biol.*, **305**, 1057–1072.
 52. Cheatham, T.E. III and Kollman, P.A. (1997) Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc.*, **119**, 4805–4825.
 53. Auffinger, P. and Westhof, E. (2001) Hydrophobic groups stabilize the hydration shell of 2'-O-methylated RNA duplexes. *Angew. Chem. Int. Ed.*, **40**, 4648–4650.
 54. Kulinska, K., Kulinski, T., Lyubartsev, A., Laaksonen, A., and Adamiak, R.W. (2000) Spatial distribution functions as a tool in the analysis of ribonucleic acids hydration - molecular dynamics studies. *Comput. Chem.*, **24**, 451–457.
 55. Nagan, M.C., Beuning, P., Musier-Forsyth, K., and Cramer, C.J. (2000) Importance of discriminator base stacking interactions: molecular dynamics analysis of A73 microhelix(Ala) variants. *Nucleic Acids Res.*, **28**, 2527–2534.
 56. Nagan, M.C., Kerimo, S.S., Musier-Forsyth, K., and Cramer, C.J. (1999) Wild-type RNA microhelix^{Ala} and 3:70 variants: molecular dynamics

- analysis of local helical structure and tightly bound water. *J. Am. Chem. Soc.*, **121**, 7310–7317.
57. Schneider, C. and Sühnel, J. (2000) A molecular dynamics simulation study of coaxial stacking in RNA. *J. Biomol. Struct. Dyn.*, **18**, 345–352.
 58. Westhof, E., Masquida, B., and Jossinet, F. (2011) Predicting and modeling RNA architecture. *Cold Spring Harb. Perspect. Biol.*, **3**, a003632.
 59. Jossinet, F., Ludwig, T.E., and Westhof, E. (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics*, **26**, 2057–2059.
 60. Shih, A.Y., Sligar, S.G., and Schulten, K. (2008) Molecular models need to be tested: the case of the solar flares discoidal HDL Model. *Biophys. J.*, **94**, L87–L89.
 61. Sundaralingam, M. and Pan, B. (2002) Hydrogen and hydration of DNA and RNA oligonucleotides. *Biophys. Chem.*, **95**, 273–282.
 62. Egli, M. (2002) DNA-Cation Interactions. *Chem. Biol.*, **9**, 277–286.
 63. Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., and Jain, S. (2002) The protein data bank. *Acta Crystallogr.*, **D58**, 899–907.
 64. Auffinger, P. and Vaiana, A.C. (2005) in *Hydrogen- and Hydration-Sensitive Structural Biology*, (eds N. Niimura, H. Mizuno, J. Hellivell, and E. Westhof), Kubapro Co, Ltd, Tokyo, pp. 227–234.
 65. Blakeley, M.P., Langan, P., Niimura, N., and Podjarny, A. (2008) Neutron crystallography: opportunities, challenges, and limitations. *Curr. Opin. Struct. Biol.*, **18**, 1–8.
 66. Sanbonmatsu, K.Y., Joseph, S., and Tung, C.S. (2005) Simulating movement of tRNA into the ribosome during decoding. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15854–15859.
 67. Li, W. and Frank, J. (2007) Transfer RNA in the hybrid P/E state: correlating molecular dynamics simulations with cryo-EM data. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 16540–16545.
 68. Dunkle, J.A. and Cate, J.H. (2010) Ribosome structure and dynamics during translocation and termination. *Annu. Rev. Biophys.*, **39**, 227–244.
 69. Sanbonmatsu, K.Y. (2006) in *Computational Studies of DNA and RNA*, vol. II (eds J. Sponer and F. Lankas), Springer-Verlag, p. 327–342.
 70. Munro, J.B., Sanbonmatsu, K.Y., Spahn, C.M., and Blanchard, S.C. (2009) Navigating the ribosome's metastable energy landscape. *Trends Biochem. Sci.*, **34**, 390–400.
 71. Nozinovic, S., Furtig, B., Jonker, H.R., Richter, C., and Schwalbe, H. (2010) High-resolution NMR structure of an RNA model system: the 14-mer cUUCG tetraloop hairpin RNA. *Nucleic Acids Res.*, **38**, 683–694.
 72. Auffinger, P. and Westhof, E. (1997) Rules governing the orientation of the 2'-hydroxyl group in RNA. *J. Mol. Biol.*, **274**, 54–63.
 73. Wylie, B.J., Sperling, L.J., Nieuwkoop, A.J., Franks, W.T., Oldfield, E., and Rienstra, C.M. (2011) Ultrahigh resolution protein structures using NMR chemical shift tensors. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 16974–16979.
 74. McDonald, I.K. and Thornton, J.M. (1995) The application of hydrogen bonding analysis in X-ray crystallography to help orientate asparagine, glutamine and histidine side chains. *Protein Eng.*, **8**, 217–224.
 75. Lovell, S.C., Davis, I.W., Arendall, W.B. III, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003) Structure validation by $\text{C}\alpha$ geometry: Φ , Ψ and $\text{C}\beta$ deviation. *Proteins*, **50**, 437–450.
 76. Auffinger, P. and Westhof, E. (2005) in *Modification and Editing of RNA* (eds H. Grosjean and R. Benne), American Society for Microbiology, Washington, DC, 1998, pp. 103–112.
 77. Das, U., Chen, S., Fuxreiter, M., Vaguine, A.A., Richelle, J., Berman, H.M., and Wodak, S.J. (2001) Checking nucleic acid crystal structures. *Acta Crystallogr.*, **D57**, 813–828.
 78. Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B. III,

- Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.
79. Kleywegt, G.J. (2009) On vital aid: the why, what and how of validation. *Acta Crystallogr.*, **D65**, 134–139.
 80. Williams, L.D. (2005) Between Objectivity and Whim: Nucleic Acid Structural Biology in *Top. Curr. Chem.*, **253**, 77–88.
 81. Davis, A.M., St-Gallay, S.A., and Kleywegt, G.J. (2008) Limitations and lessons in the use of X-ray structural information in drug design. *Drug Discov. Today*, **13**, 831–841.
 82. Schreiner, E., Trabuco, L.G., Freddolino, P.L., and Schulten, K. (2011) Stereochemical errors and their implications for molecular dynamics simulations. *BMC Bioinformatics*, **12**, 190.
 83. DePristo, M.A., de Bakker, P.I., and Blundell, T.L. (2004) Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure*, **12**, 831–838.
 84. Joosten, R.P., Womack, T., Vriend, G., and Bricogne, G. (2009) Re-refinement from deposited X-ray data can deliver improved models for most PDB entries. *Acta Crystallogr. D Biol. Crystallogr.*, **65**, 176–185.
 85. Joosten, R.P., Salzemann, J., Bloch, V., Stockinger, H., Berglund, A.C., Blanchet, C., Bongcam-Rudloff, E., Combet, C., Da Costa, A.L., Deleage, G. *et al.* (2009) PDB_REDO: automated re-refinement of X-ray structure models in the PDB. *Appl. Crystallogr.*, **42**, 376–384.
 86. Jeffrey, G.A. (1997) *An Introduction to Hydrogen Bonding*, Oxford University Press, New York.
 87. Steiner, T. (1996) C–H–O Hydrogen Bonding in Crystals *Crystallogr. Rev.*, **6**, 1–57.
 88. Ma, J.C. and Dougherty, D.A. (1997) The cation- π Interaction. *Chem. Rev.*, **97**, 1303–1324.
 89. Riley, K.E. and Hobza, P. (2011) Non-covalent interactions in biochemistry. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **1**, 3–17.
 90. Brandl, M., Weiss, M.S., Jabs, A., Suhnel, J., and Hilgenfeld, R. (2001) C-H... π interactions in proteins. *J. Mol. Biol.*, **307**, 357–377.
 91. Meyer, E.A., Castellano, R.K., and Diederich, F. (2003) Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem. Int. Ed.*, **42**, 1210–1250.
 92. Bissantz, C., Kuhn, B., and Stahl, M. (2010) A medicinal chemist's guide to molecular interactions. *J. Med. Chem.*, **53**, 5061–5084.
 93. Hassel, O. (1972) *Nobel Lectures, Chemistry 1963–1970*, Elsevier Publishing Company, Amsterdam.
 94. Auffinger, P., Hays, F.A., Westhof, E., and Ho, P.S. (2004) Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 16789–16794.
 95. Voth, A.R., Hays, F.A., and Ho, P.S. (2007) Directing macromolecular conformation through halogen bonds. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 6188–6193.
 96. Levitt, M. and Perutz, M.F. (1988) Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.*, **201**, 751–754.
 97. Steiner, T. and Koellner, G. (2001) Hydrogen bonds with pi-acceptors in proteins: frequencies and role in stabilizing local 3D structures. *J. Mol. Biol.*, **305**, 535–557.
 98. Nishio, M., Hirota, M., and Umezawa, Y. (1998) *The CH/ π Interaction. Evidence, Nature, and Consequences*, Wiley-VCH Verlag GmbH, New York.
 99. Desiraju, G. and Steiner, T. (1999) *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford University Press, New York.
 100. Zacharias, N. and Dougherty, D.A. (2002) Cation- π interactions in ligand recognition and catalysis. *Trends Pharmacol. Sci.*, **23**, 281–287.
 101. Dougherty, D.A. (2007) Cation- π interactions involving aromatic amino acids. *J. Nutr.*, **137**, 1504S–1508S.
 102. del Carmen Fernandez-Alonso, M., Canada, F.J., Jimenez-Barbero, J., and Cuevas, G. (2005) Molecular

- recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. *J. Am. Chem. Soc.*, **127**, 7379–7386.
103. Vandenbussche, S., Diaz, D., Fernandez-Alonso, M.C., Pan, W., Vincent, S.P., Cuevas, G., Canada, F.J., Jimenez-Barbero, J., and Bartik, K. (2008) Aromatic-carbohydrate interactions: an NMR and computational study of model system. *Chemistry*, **14**, 7570–7578.
 104. Bhattacharyya, R., Samanta, U., and Chakrabarti, P. (2002) Aromatic-aromatic interactions in and around α -helices. *Protein Eng.*, **15**, 91–100.
 105. Hunter, C.A., Lawson, K.R., Perkins, J., and Urch, C.J. (2001) Aromatic interactions. *J. Chem. Soc., Perkin Trans. 2*, 651–669.
 106. Sponer, J., Riley, K.E., and Hobza, P. (2008) Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys. Chem. Chem. Phys.*, **10**, 2595–2610.
 107. Acharya, S., Barman, J., Cheruku, P., Chatterjee, S., Acharya, P., Isaksson, J., and Chattopadhyaya, J. (2004) Significant pKa perturbation of nucleobases is an intrinsic property of the sequence context in DNA and RNA. *J. Am. Chem. Soc.*, **126**, 8674–8681.
 108. Strobel, S.A. and Cochrane, J.C. (2007) RNA catalysis: ribozymes, ribosomes, and riboswitches. *Curr. Opin. Chem. Biol.*, **11**, 636–643.
 109. Bevilacqua, P.C., Brown, T.S., Nakano, S., and Yajima, R. (2004) Catalytic roles for proton transfer and protonation in ribozymes. *Biopolymers*, **73**, 90–109.
 110. Wilcox, J.L., Ahluwalia, A.K., and Bevilacqua, P.C. (2011) Charged nucleobases and their potential for RNA catalysis. *Acc. Chem. Res.*, **44**, 1270–1279.
 111. Roitzsch, M., Anorbe, M.G., Miguel, P.J., Müller, B., and Lippert, B. (2005) The role of intramolecular hydrogen bonding on nucleobase acidification following metal coordination: possible implications of an “indirect” role of metals in acid-base catalysis of nucleic acids. *J. Biol. Inorg. Chem.*, **10**, 800–812.
 112. Csaszar, K., Spackova, N., Stefl, R., Sponer, J., and Leontis, N.B. (2001) Molecular dynamics of the frame-shifting pseudoknot from beet western yellows virus: the role of non-Watson-Crick base-pairing, ordered hydration, cation binding and base mutations on stability and unfolding. *J. Mol. Biol.*, **313**, 1073–1091.
 113. Banas, P., Walter, N.G., Sponer, J., and Otyepka, M. (2010) Protonation states of the key active site residues and structural dynamics of the *glmS* riboswitch as revealed by molecular dynamics. *J. Phys. Chem. B*, **114**, 8701–8712.
 114. Mlynsky, V., Banas, P., Hollas, D., Reblova, K., Walter, N.G., Sponer, J., and Otyepka, M. (2010) Extensive molecular dynamics simulations show that canonical G8 and protonated A38H⁺ forms are most consistent with crystal structures of hairpin ribozyme. *J. Phys. Chem. B*, **114**, 6642–6652.
 115. Ditzler, M.A., Sponer, J., and Walter, N.G. (2009) Molecular dynamics suggest multifunctionality of an adenine imino group in acid-base catalysis of the hairpin ribozyme. *RNA*, **15**, 560–575.
 116. Kaul, M., Barbieri, C.M., Kerrigan, J.E., and Pilch, D.S. (2003) Coupling of drug protonation to the specific binding of aminoglycosides to the A site of 16 S rRNA: elucidation of the number of drug amino groups involved and their identities. *J. Mol. Biol.*, **326**, 1373–1387.
 117. Pilch, D.S., Kaul, M., Barbieri, C.M., and Kerrigan, J.E. (2003) Thermodynamics of aminoglycoside-rRNA recognition. *Biopolymers*, **70**, 58–79.
 118. Vaiana, A.C., Westhof, E., and Auffinger, P. (2006) A molecular dynamics simulation study of an aminoglycoside/A-site RNA complex: conformational and hydration patterns. *Biochimie*, **88**, 1061–1073.
 119. Genheden, S. and Ryde, U. (2011) Comparison of different initialisation protocols to generate independent

- molecular dynamics simulations. *J. Comput. Chem.*, **32**, 187–195.
120. Stanton, C.L. and Houk, K.N. (2008) Benchmarking pK_a prediction methods for residues in proteins. *J. Chem. Theory Comput.*, **4**, 951–966.
 121. Durant, P.C. and Davis, D.R. (1999) Stabilization of the anticodon stem-loop of tRNA^{Lys,3} by an A⁺-C base pair and by pseudouridine. *J. Mol. Biol.*, **1999**, 115–131.
 122. Auffinger, P. and Westhof, E. (2001) An extended structural signature for the tRNA anticodon loop. *RNA*, **7**, 334–341.
 123. Auffinger, P., Bielecki, L., and Westhof, E. (2004) Anion binding to nucleic acids. *Structure*, **12**, 379–388.
 124. Kieft, J.S., Chase, E., Costantino, D.A., and Golden, B.L. (2010) Identification and characterization of anion binding sites in RNA. *RNA*, **16**, 1118–1123.
 125. Petrova, T. and Podjarny, A. (2004) Protein crystallography at subatomic resolution. *Rep. Prog. Phys.*, **67**, 1565–1605.
 126. Niimura, N., Arai, S., Kurihara, K., Chatake, T., Tanaka, I., and Bau, R. (2006) Recent results on hydrogen and hydration in biology studied by neutron macromolecular crystallography. *Cell. Mol. Life Sci.*, **63**, 285–300.
 127. Niimura, N. and Bau, R. (2008) Neutron protein crystallography: beyond the folding structure of biological macromolecules. *Acta Crystallogr.*, **A64**, 12–22.
 128. Hutter, J., Carloni, P., and Parrinello, M. (1996) Non-empirical calculations of a hydrated RNA duplex. *J. Am. Chem. Soc.*, **118**, 8710–8712.
 129. Bevan, D.R., Li, L., Pedersen, L.G., and Darden, T.A. (2000) Molecular dynamics simulations of the d(CCAACGTTGG)₂ decamer: Influence of the crystal environment. *Biophys. J.*, **78**, 668–682.
 130. Lee, H., Darden, T.A., and Pedersen, L.G. (1995) Molecular dynamics simulation studies of a high resolution Z-DNA crystal. *J. Chem. Phys.*, **102**, 3830–3834.
 131. York, D.M., Yang, W., Lee, H., Darden, T., and Pedersen, L.G. (1995) Toward the accurate modeling of DNA: the importance of long-range electrostatics. *J. Am. Chem. Soc.*, **117**, 5001–5002.
 132. Korolev, N., Lyubartsev, A.P., Laaksonen, A., and Nordenskiöld, L. (2002) On the competition between water, sodium ions and spermine in binding to DNA: a molecular dynamics computer simulation study. *Biophys. J.*, **82**, 2860–2875.
 133. Lee, H., Darden, T., and Pedersen, L. (1995) Accurate crystal molecular dynamics simulations using particle-mesh-Ewald: RNA dineucleotides – ApU and GpC. *Chem. Phys. Lett.*, **243**, 229–235.
 134. MacKerell, A.D. and Banavali, N. (2000) All-atom empirical force field for nucleic acids: RNA in solution. *J. Comput. Chem.*, **21**, 105–120.
 135. Foloppe, N. and MacKerell, A.D. Jr. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.*, **21**, 86–104.
 136. Martick, M., Lee, T.S., York, D.M., and Scott, W.G. (2008) Solvent structure and hammerhead ribozyme catalysis. *Chem. Biol.*, **15**, 332–342.
 137. Cerutti, D.S., Freddolino, P.L., Duke, R.E. Jr., and Case, D.A. (2010) Simulations of a protein crystal with a high resolution X-ray structure: evaluation of force fields and water models. *J. Phys. Chem. B*, **114**, 12811–12824.
 138. Cerutti, D.S. and Case, D.A. (2010) Multi-level Ewald: a hybrid multigrid / Fast Fourier Transform approach to the electrostatic particle-mesh problem. *J. Chem. Theory Comput.*, **6**, 443–458.
 139. Nina, M. and Simonson, T. (2002) Molecular dynamics of the tRNA(Ala) acceptor stem: comparison between continuum reaction field and particle-mesh Ewald electrostatic treatments. *J. Phys. Chem. B*, **106**, 3696–3705.
 140. Ohtaki, H. (2001) Ionic solvation in aqueous and nonaqueous solutions. *Monatsh. Chem.*, **132**, 1237–1268.
 141. Ohtaki, H. and Radnai, T. (1993) Structure and dynamics of hydrated ions. *Chem. Rev.*, **93**, 1157–1204.

142. Auffinger, P., Bielecki, L., and Westhof, E. (2004) Symmetric K^+ and Mg^{2+} ion-binding sites in the 5S rRNA loop E inferred from molecular dynamics simulations. *J. Mol. Biol.*, **335**, 555–571.
143. Auffinger, P., Bielecki, L., and Westhof, E. (2003) The Mg^{2+} binding sites of the 5S rRNA loop E motif as investigated by molecular dynamics simulations. *Chem. Biol.*, **10**, 551–561.
144. Reblova, K., Spackova, N., Stefl, R., Csaszar, K., Koca, J., Leontis, N.B., and Sponer, J. (2003) Non-Watson-Crick basepairing and hydration in RNA motifs: molecular dynamics of 5S rRNA loop E. *Biophys. J.*, **84**, 3564–3582.
145. Hermann, T., Auffinger, P., and Westhof, E. (1998) Molecular dynamics investigations of hammerhead ribozyme RNA. *Eur. J. Biophys.*, **27**, 153–165.
146. Reblova, K., Spackova, N., Sponer, J.E., Koca, J., and Sponer, J. (2003) Molecular dynamics simulations of RNA kissing-loop motifs reveal structural dynamics and formation of cation-binding pockets. *Nucleic Acids Res.*, **31**, 6942–6952.
147. Auffinger, P., Cheatham, T.E. III., and Vaiana, A.C. (2007) Spontaneous formation of KCl aggregates in biomolecular simulations: a force field issue? *J. Chem. Theory Comput.*, **3**, 1851–1859.
148. Joung, I.S. and Cheatham, T.E. III (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, **112**, 9020–9041.
149. Frauenfelder, H., Chen, G., Berendzen, J., Fenimore, P.W., Jansson, H., McMahon, B.H., Stroe, I.R., Swenson, J., and Young, R.D. (2009) A unified model of protein dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 5129–5134.
150. Ladbury, J.E. (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.*, **3**, 973–980.
151. Pal, S.K. and Zewail, A.H. (2004) Dynamics of water in biological recognition. *Chem. Rev.*, **104**, 2099–2123.
152. Luccarelli, J., Michel, J., Tirado-Rives, J., and Jorgensen, W.L. (2010) Effects of water placement on predictions of binding affinities for p38 α MAP kinase inhibitors. *J. Chem. Theory Comput.*, **6**, 3850–3856.
153. Martinez, L., Andrade, R., Birgin, E.G., and Martinez, J.M. (2009) PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.*, **30**, 2157–2164.
154. Mezei, M. (2010) Simulaid: a simulation facilitator and analysis program. *J. Comput. Chem.*, **31**, 2658–2668.
155. Case, D.A., Cheatham, T.E. III., Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
156. MacKerell, A.D. Jr. (2004) Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.*, **25**, 1584–1604.
157. Kaminski, G.A., Friesner, R.A., Tirado-Rives, J., and Jorgensen, W.L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, **105**, 6474–6487.
158. Soares, T.A., Hunenberger, P.H., Kastenzholz, M.A., Krautler, V., Lenz, T., Lins, R.D., Oostenbrink, C., and van Gunsteren, W.F. (2005) An improved nucleic acid parameter set for the GROMOS force field. *J. Comput. Chem.*, **26**, 725–737.
159. Feig, M. and Pettitt, B.M. (1997) Structural equilibrium of DNA represented with different force fields. *J. Chem. Phys. B*, **101**, 7361–7363.
160. Feig, M. and Pettitt, B.M. (1998) Structural equilibrium of DNA represented with different force fields. *Biophys. J.*, **75**, 134–149.

161. Cheatham, T.E. III, Cieplak, P., and Kollman, P.A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, **16**, 845–862.
162. Reddy, S.Y., Leclerc, F., and Karplus, M. (2003) DNA polymorphism: a comparison of force fields for nucleic acids. *Biophys. J.*, **84**, 1421–1449.
163. Banas, P., Hollas, D., Zgarbova, M., Jurecka, P., Orozco, M., Cheatham, T.E. III, Spomer, J., and Otyepka, M. (2010) Performance of molecular mechanics force fields for RNA simulations. Stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.*, **6**, 3836–3849.
164. Sklenovsky, P., Florova, P., Banas, P., Reblova, K., Lankas, F., Otyepka, M., and Spomer, J. (2011) Understanding the RNA flexibility using explicit solvent simulations: the ribosomal and group I intron reverse kink-turn motifs. *J. Chem. Theory Comput.*, **7**, 2963–2980.
165. Zgarbova, M., Otyepka, M., Spomer, J., Mladek, A., Banas, P., Cheatham, T.E. III, and Jurecka, P. (2011) Refinement of the Cornell et al. nucleic acid force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.*, **7**, 2886–2902.
166. Sychrovsky, V., Foldynova-Trantirkova, S., Spackova, N., Robeyns, K., Van Meervelt, L., Blankenfeldt, W., Vokacova, Z., Spomer, J., and Trantirek, L. (2009) Revisiting the planarity of nucleic acid bases: pyramidalization at glycosidic nitrogen in purine bases is modulated by orientation of glycosidic torsion. *Nucleic Acids Res.*, **37**, 7321–7331.
167. Ode, H., Matsuo, Y., Neya, S., and Hoshino, T. (2008) Force field parameters for rotation around χ torsion axis in nucleic acids. *J. Comput. Chem.*, **29**, 2531–2542.
168. Yildirim, I., Stern, H.A., Kennedy, S.D., Tubbs, J.D., and Turner, D.H. (2010) Reparameterization of RNA chi torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *J. Chem. Theory Comput.*, **6**, 1520–1531.
169. Freddolino, P.L., Park, S., Roux, B., and Schulten, K. (2009) Force field bias in protein folding simulations. *Biophys. J.*, **96**, 3772–3780.
170. Ricci, C.G., de Andrade, A.S., Mottin, M., and Netz, P.A. (2010) Molecular dynamics of DNA: comparison of force fields and terminal nucleotide definitions. *J. Phys. Chem. B*, **114**, 9882–9893.
171. Denning, E.J., Priyakumar, U.D., Nilsson, L., and Mackerell, A.D. III (2011) Impact of 2'-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.*, **32**, 1929–1943.
172. Jha, S., Coveney, P.V., and Laughton, C.A. (2005) Force field validation for nucleic acid simulations: Comparing energies and dynamics of a DNA dodecamer. *J. Comput. Chem.*, **26**, 1617–1627.
173. Priyakumar, U.D. and MacKerell, A.D. (2006) Base flipping in a GCGC containing DNA dodecamer: a comparative study of the performance of the nucleic acid force fields, CHARMM, AMBER, and BMS. *J. Chem. Theory Comput.*, **2**, 187–200.
174. Besseova, I., Otyepka, M., Reblova, K., and Spomer, J. (2009) Dependence of A-RNA simulations on the choice of the force field and salt strength. *Phys. Chem. Chem. Phys.*, **11**, 10701–10711.
175. Halgren, T.A. and Damm, W. (2001) Polarizable force fields. *Curr. Opin. Struct. Biol.*, **11**, 236–242.
176. Babin, V., Baucom, J., Darden, T.A., and Sagui, C. (2006) Molecular dynamics simulations of DNA with polarizable force fields: convergence of an ideal B-DNA structure to the crystallographic structure. *J. Phys. Chem. B*, **110**, 11571–11581.
177. Jiao, D., Golubkov, P.A., Darden, T.A., and Ren, P. (2008) Calculation of protein-ligand binding free energy using a polarizable potential. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 6290–6295.
178. Baker, C.M., Anisimov, V.M., and MacKerell, A.D. Jr. (2011) Development

- of CHARMM polarizable force field for nucleic acid bases based on the classical drude oscillator model. *J. Phys. Chem. B*, **115**, 580–596.
179. Ponder, J.W., Wu, C., Ren, P., Pande, V.S., Chodera, J.D., Schnieders, M.J., Haque, I., Mobley, D.L., Lambrecht, D.S., DiStasio, R.A. Jr. *et al.* (2010) Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B*, **114**, 2549–2564.
 180. Dinur, U. and Hagler, A.T. (1991) in *Reviews in Computational Chemistry*, Vol. 2 (eds K.B. Lipkowitz and D.B. Boyd), VCH Publishers, New York, pp. 99–164.
 181. Wang, J. and Kollman, P. (2001) Automatic parameterization of force field by systematic search and genetic algorithms. *J. Comput. Chem.*, **22**, 1219–1228.
 182. Njo, S.L., van Gunsteren, W.F., and Müller-Plathe, F. (1995) Determination of force field parameters for molecular simulation by molecular simulation: an application of the weak-coupling method. *J. Chem. Phys.*, **102**, 6199–6207.
 183. Kony, D., Damm, W., Stoll, S., and van Gunsteren, W.F. (2002) An improved OPLS-AA force field for carbohydrates. *J. Comput. Chem.*, **23**, 1416–1429.
 184. Meagher, K.L., Redman, L.T., and Carlson, H.A. (2003) Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.*, **24**, 1016–1025.
 185. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004) Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.*, **25**, 1157–1174.
 186. Joung, I.S. and Cheatham, T.E. III (2009) Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *J. Phys. Chem. B*, **113**, 13279–13290.
 187. Lamoureux, G. and Roux, B. (2006) Absolute hydration free energy scale for alkali and halide ions established from simulations with a polarizable force field. *J. Phys. Chem. B*, **110**, 3308–3322.
 188. Jiao, D., King, C., Grossfield, A., Darden, T.A., and Ren, P. (2006) Simulation of Ca^{2+} and Mg^{2+} solvation using polarizable atomic multipole potential. *J. Phys. Chem. B*, **110**, 18553–18559.
 189. Piquemal, J.P., Perera, L., Cisneros, G.A., Ren, P., Pedersen, L.G., and Darden, T.A. (2006) Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure. *J. Chem. Phys.*, **125**, 054511.
 190. Aduri, R., Psciuk, B.T., Saro, P., Taniga, H., Schlegel, H.B., and SantaLucia, J. (2007) AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J. Chem. Theory Comput.*, **3**, 1464–1475.
 191. Mayaan, E., Moser, A., MacKerell, A.D. Jr., and York, D.M. (2007) CHARMM force field parameters for simulation of reactive intermediates in native and thio-substituted ribozymes. *J. Comput. Chem.*, **28**, 495–507.
 192. Cheatham, T.E. III (2004) Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.*, **14**, 360–367.
 193. Chen, A.A., Marucho, M., Baker, N.A., and Pappu, R.V. (2009) Simulations of RNA interactions with monovalent ions. *Methods Enzymol.*, **469**, 411–432.
 194. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–936.
 195. Mahoney, M.W. and Jorgensen, W.L. (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, **112**, 8910–8922.
 196. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., and Hermans, J. (1981) in *Intermolecular Forces Interaction Models for Water in Relation to Protein Hydration* (ed. B. Pullman), Reidel, Dordrecht, pp. 331–342.
 197. Berendsen, H.J.C., Grigera, J.R., and Straatsma, T.P. (1987) The missing

- term in effective pair potentials. *J. Phys. Chem.*, **97**, 6269–6271.
198. Tarek, M. and Tobias, D.J. (2000) The dynamics of protein hydration water: a quantitative comparison of molecular dynamics simulations and neutron-scattering experiments. *Biophys. J.*, **79**, 3244–3257.
 199. van der Spoel, D. and van Maaren, P.J. (2006) The origin of layer structure artifacts in simulation of liquid water. *J. Chem. Theory Comput.*, **2**, 1–11.
 200. Auffinger, P. and Beveridge, D.L. (1995) A simple test for evaluating the truncation effects in simulations of systems involving charged groups. *Chem. Phys. Lett.*, **234**, 413–415.
 201. Ravishanker, G., Auffinger, P., Langley, D.R., Jayaram, B., Young, M.A., and Beveridge, D.L. (1997) in *Reviews in Computational Chemistry*, Vol. 11 (eds K.B. Lipkowitz and D.B. Boyd), VCH Publishers, New York, pp. 317–372.
 202. Louise-May, S., Auffinger, P., and Westhof, E. (1996) Calculations of nucleic acid conformations. *Curr. Opin. Struct. Biol.*, **6**, 289–298.
 203. Sagui, C. and Darden, T.A. (1999) Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 155–179.
 204. Darden, T., Perera, L., Li, P., and Pedersen, L. (1999) New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure*, **7**, R55–R60.
 205. Cheatham, T.E. III, Miller, J.L., Fox, T., Darden, T.A., and Kollman, P.A. (1995) Molecular dynamics simulations on solvated biomolecular systems: the Particle Mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.*, **117**, 4193–4194.
 206. Eslami, H. and Müller-Plathe, F. (2007) Molecular dynamics simulation in the grand canonical ensemble. *J. Comput. Chem.*, **28**, 1763–1773.
 207. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., and DiNola, A. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
 208. Auffinger, P. and Westhof, E. (2002) Melting of the solvent structure around a RNA duplex: a molecular dynamics simulation study. *Biophys. Chem.*, **95**, 203–210.
 209. Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, **23**, 327–336.
 210. Harvey, S.C., Tan, R.K.Z., and Cheatham, T.E. III (1998) The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.*, **19**, 726–740.
 211. Chiu, S.W., Clark, M., Subramaniam, S., and Jakobsson, E. (2000) Collective motion artifacts arising in long-duration molecular dynamics simulations. *J. Comput. Chem.*, **21**, 121–131.
 212. Piana, S., Lindorff-Larsen, K., and Shaw, D.E. (2011) How robust are protein folding simulations with respect to force field parameterization?. *Biophys. J.*, **100**, L47–L49.
 213. Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., Eastwood, M.P., Bank, J.A., Jumper, J.M., Salmon, J.K., Shan, Y., and Wriggers, W. (2010) Atomic-level characterization of the structural dynamics of proteins. *Science*, **330**, 341–346.
 214. Freddolino, P.L., Liu, F., Gruebele, M., and Schulten, K. (2008) Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys. J.*, **94**, L75–L77.
 215. Lindorff-Larsen, K., Piana, S., Dror, R.O., and Shaw, D.E. (2011) How fast-folding proteins fold. *Science*, **334**, 517–520.
 216. Shan, Y., Kim, E.T., Eastwood, M.P., Dror, R.O., Seeliger, M.A., and Shaw, D.E. (2011) How does a drug molecule find its target binding site?. *J. Am. Chem. Soc.*, **133**, 9181–9183.
 217. Sanbonmatsu, K.Y. and Tung, C.S. (2006) High performance computing in

- biology: multimillion atom simulations of nanoscale systems. *J. Struct. Biol.*, **157**, 470–480.
218. Freddolino, P.L., Arkhipov, A.S., Larson, S.B., McPherson, A., and Schulten, K. (2006) Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, **14**, 437–449.
 219. Freddolino, P.L., Harrison, C.B., Liu, Y., and Schulten, K. (2010) Challenges in protein folding simulations: timescale, representation, and analysis. *Nat. Phys.*, **6**, 751–758.
 220. Perez, A., Luque, F.J., and Orozco, M. (2007) Dynamics of B-DNA on the microsecond time scale. *J. Am. Chem. Soc.*, **129**, 14739–14745.
 221. Zuckerman, D.M. (2011) Equilibrium sampling in biomolecular simulations. *Annu. Rev. Biophys.*, **40**, 41–62.
 222. Auffinger, P., Louise-May, S., and Westhof, E. (1999) Molecular dynamics simulations of solvated yeast tRNA(Asp). *Biophys. J.*, **76**, 50–64.
 223. Elofson, A. and Nilsson, L. (1993) How consistent are molecular dynamics simulations? Comparing structure and dynamics in reduced and oxidized *Escherichia coli* thioredoxin. *J. Mol. Biol.*, **233**, 766–780.
 224. Auffinger, P., Louise-May, S., and Westhof, E. (1995) Multiple molecular dynamics simulations of the anticodon loop of tRNA^{Asp} in aqueous solution with counterions. *J. Am. Chem. Soc.*, **117**, 6720–6726.
 225. Auffinger, P., Louise-May, S., and Westhof, E. (1996) Molecular dynamics simulations of the anticodon hairpin of tRNA^{Asp}: structuring effects of C-H...O hydrogen bonds and of long-range hydration forces. *J. Am. Chem. Soc.*, **118**, 1181–1189.
 226. Auffinger, P. and Westhof, E. (1996) H-bond stability in the tRNA^{Asp} anticodon hairpin: 3 ns of multiple molecular dynamics simulations. *Biophys. J.*, **71**, 940–954.
 227. Auffinger, P. and Westhof, E. (1997) Rules governing the orientation of the 2'-hydroxyl group in RNA. *J. Mol. Biol.*, **269**, 326–341.
 228. Braxenthaler, M., Unger, R., Auerbach, D., Given, J.A., and Moul, J. (1997) Chaos in protein dynamics. *Proteins*, **29**, 417–425.
 229. Caves, L.S.D., Evanseck, J.D., and Karplus, M. (1998) Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Sci.*, **7**, 649–666.
 230. Zhou, H.B. and Wang, L. (1996) Chaos in biomolecular dynamics. *J. Phys. Chem.*, **100**, 8101–8105.
 231. Villani, V. (2003) Complexity of polypeptide dynamics: chaos, Brownian motion and elasticity in aqueous solution. *J. Mol. Struct. THEOCHEM*, **621**, 127–139.
 232. Zagrovic, B. and van Gunsteren, W.F. (2007) Computational analysis of the mechanism and thermodynamics of inhibition of phosphodiesterase 5A by synthetic ligands. *J. Chem. Theory Comput.*, **3**, 301–311.
 233. Genheden, S., Diehl, C., Akke, M., and Ryde, U. (2010) Starting-condition dependence of order parameters derived from molecular dynamics simulations. *J. Chem. Theory Comput.*, **6**, 2176–2190.
 234. van Gunsteren, W.F. and Mark, A.E. (1998) Validation of molecular dynamics simulation. *J. Chem. Phys.*, **108**, 6109–6116.
 235. Swaminathan, S., Ravishanker, G., and Beveridge, D.L. (1991) Molecular dynamics of B-DNA including water and counterions: a 140 ps trajectory for d(CGCGAATTCGCG) based on the GROMOS force field. *J. Am. Chem. Soc.*, **111**, 5027–5040.
 236. Kryshchuk, A., Moul, J., Bartual, S.G., Bazan, J.F., Berman, H., Casteel, D.E., Christodoulou, E., Everett, J.K., Hausmann, J., Heidebrecht, T. *et al.* (2011) Target highlights in CASP9: Experimental target structures for the critical assessment of techniques for protein structure prediction. *Proteins*, **79**, (suppl.10), 6–20.
 237. Cruz, J.A., Blanchet, M.F., Boniecki, M., Bujnicki, J.M., Chen, S.J., Cao, S., Das, R., Ding, F., Dokholyan, N.V., Flores, S.C. *et al.* (2012), RNA-Puzzles: a CASP-like evaluation of RNA

- three-dimensional structure prediction. *RNA*, **18**, 610–625.
238. Smith, L.J., Daura, X., and van Gunsteren, W.F. (2002) Stimulation and inhibition of fibril formation by a peptide in the presence of different concentrations of SDS. *Protein Struct. Funct. Genet.*, **48**, 487–496.
 239. Auffinger, P. and Westhof, E. (2000) RNA solvation: a molecular dynamics simulation perspective. *Biopolymers*, **56**, 266–274.
 240. Keller, B., Daura, X., and van Gunsteren, W.F. (2010) Comparing geometric and kinetic cluster algorithms for molecular simulation data. *J. Chem. Phys.*, **132**, 074110.
 241. Rajan, A., Freddolino, P.L., and Schulten, K. (2010) Going beyond clustering in MD trajectory analysis: an application to villin headpiece folding. *PLoS ONE*, **5**, e9890.
 242. Eichenberger, A.P., Allison, J.R., Dolenc, J., Geerke, D.P., Horta, B.A.C., Meier, K., Oostenbrink, C., Schmid, N., Steiner, D., Wang, D., and van Gunsteren, W.F. (2011) GRO-MOS++ software for the analysis of biomolecular simulation trajectories. *J. Chem. Theory Comput.*, **7**, 3379–3390.
 243. Knapp, B., Lederer, N., Omasits, U., and Schreiner, W. (2010) vmdICE: a plug-in for rapid evaluation of molecular dynamics simulations using VMD. *J. Comput. Chem.*, **31**, 2868–2873.
 244. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., and Beckstein, O. (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.*, **32**, 2319–2327.
 245. Auffinger, P. and Westhof, E. (1999) in *Water Management in the Design and Distribution of Quality Food* (eds Y.H. Ross, R.B. Leslie, and P.J. Lillford), Technonic Publishing Co. Inc., Basel, pp. 165–198.
 246. Feig, M. and Pettitt, B.M. (1998) A molecular simulation picture of DNA hydration around A- and B-DNA. *Biopolymers*, **48**, 199–209.
 247. Makarov, V.A., Pettitt, B.M., and Feig, M. (2002) Solvation and hydration of proteins and nucleic acids: a theoretical view of simulation and experiment. *Acc. Chem. Res.*, **35**, 376–384.

33

Identification and Characterization of Small Non-coding RNAs in Bacteria

Dimitri Podkaminski, Marie Bouvier, and Jörg Vogel

33.1

Introduction

The systematic discovery and characterization of non-coding RNA molecules that exert regulatory functions in the cell have become major fields in biology. In the traditional view, the RNA population of a cell fell into three major categories with distinct roles: mRNAs, tRNAs, and rRNAs. Small non-coding RNAs (sRNAs) outside these major classes have been known in bacteria for more than 40 years, but despite a growing appreciation of both their specific biological roles, and of the catalytic and regulatory potential of RNA in general, they were regarded as a minor addition until recently. About a decade ago, however, systematic searches in the genomes of prokaryotic and eukaryotic model organisms showed that these molecules are far more prevalent than assumed, and along with a staggering list of newly discovered functions, have firmly established sRNAs as a class of regulators of gene expression that is comparable to transcription factors.

This chapter is focused on bacterial sRNAs [1, 2]. Except function, there are presently no easily discernible criteria such as size (they are typically between 50 and 250 nt in length), genomic organization, common structures or structural motifs (although most seem structured) which would permit to sort sRNAs into meaningful categories. Most sRNAs are encoded in the intergenic regions (IGRs) of bacterial chromosomes and at a different genomic locus relative to target mRNAs or proteins. If current trends hold, most of the sRNAs regulate their *trans*-encoded mRNAs through short and imperfect antisense complementarity. At least in Gram-negative γ -proteobacteria such as *Escherichia coli* or *Salmonella*, this type of antisense regulation often requires the RNA chaperone Hfq, which both stabilizes the sRNAs and facilitates their annealing to target mRNAs [3–7]. Nonetheless, the methodologies of finding sRNAs that regulate mRNAs *in trans* or *cis*, or act to modulate the activity of proteins, overlap considerably, and most of the discovery approaches that are discussed in the first part of this chapter identify sRNAs independent of their mechanism or function.

The second part of the chapter aims to provide the reader with experimental approaches for the functional characterization of sRNAs, with a focus on those

sRNAs that regulate *trans*-encoded target mRNAs by antisense mechanisms. This area of the sRNA field has seen a dramatic growth since the original report of MicF sRNA as a translational repressor of *ompF* mRNA in *E. coli* [8, 9]. Several mechanisms of target recognition and regulation different from the MicF-mediated sequestration of the ribosome binding site (RBS) have recently been described, for example, translational repression or induction of mRNA decay in the upstream 5' untranslated region (5' UTR) [10–16] or the mRNA coding sequence [17–19]. Furthermore, since the original examples, RNAIII [20] and DsrA [21, 22], were reported, numerous additional examples and mechanisms of target activation have been discovered [23, 24]. Most importantly, however, many new target mRNAs have been identified over the past few years, and it has become clear that most of the Hfq-associated sRNAs seem to regulate multiple rather than a single mRNA [25, 26]. We have also learned that multiple target regulation is governed by short, hyperconserved subregions or “domains” of sRNAs [11, 27–29], and this has made target predictions more reliable [29] and led to new strategies for target validation.

We describe here both large-scale and small-scale approaches for the identification and characterization of sRNAs, and explain the underlying strategies and benefits or pitfalls. In the second part, we provide selected protocols for the study of sRNA–target interactions.

33.2

Expression-Based Discovery of sRNAs

The following sections deal with the most commonly used global strategies for the identification of sRNAs (microarray analysis, high-throughput sequencing, and RNA–protein coimmunoprecipitation). Further below we extend this by approaches with more predictive power such as biocomputational search strategies and genomic SELEX (Systematic Evolution of Ligands by EXponential enrichment). Global mining of expressed sRNAs was fostered by the so-called experimental RNomics techniques, wherein size-selected extracted RNA was reverse-transcribed, cloned, and analyzed by Sanger sequencing [30, 31]. However, RNA sequencing (RNA-seq)-based approaches have made the labor-intensive cDNA cloning and costly Sanger sequencing obsolete, so the traditional type of RNomics is omitted from this chapter.

33.2.1

Microarray

Microarrays are a valuable and affordable tool for transcriptome analysis, but often come with the caveat that probes are limited to the protein-coding, tRNA and rRNA genes. More recently, high-density (tiling) arrays, which can carry tens of thousands of DNA oligonucleotides and systematically cover the sense and antisense strand of a genome, including the IGRs (where most sRNAs are expressed from), have

become common for sRNA searches. Tiling arrays were first used in *E. coli* to verify the expression of sRNAs predicted by comparative genome analysis [32] and came into full play in a study predicting new *E. coli* sRNAs also in 5' and 3' UTRs [33]. Subsequently, tiling arrays led to the discovery of sRNAs in *Caulobacter crescentus* [34], *Prochlorococcus* species [35], *Listeria monocytogenes* [36], and *Mycobacterium leprae* [37].

Tiling arrays can combine sRNA discovery and gene expression profiling; once they are produced, their usefulness in a laboratory clearly transcends the individual sRNA project. As of this writing, quantification of gene expression (both sRNAs and other transcripts) for profiling under various conditions still seems more straightforward with arrays than with RNA deep sequencing (see next section), especially since there are many commercial computer programs for the data analysis. However, there are clear disadvantages in using arrays: first, the production costs of tiling arrays remain high, making them a rather expensive tool. Second, efficient cDNA construction or direct labeling of structured RNAs is an intrinsic technical problem, and poor labeling compromises the detection of sRNAs above the noise range. Note, however, that an alternative detection method with antibodies against RNA:DNA hybrids might circumvent the problems associated with labeling of small structured RNAs [34, 38, 39]. Third, for reasons of costs, most tiling arrays are limited to one probe per 30 nt or more, and thus do not permit to map the ends of transcripts at nucleotide resolution. Given these disadvantages and the fast progress in new sequencing technologies, microarrays and tiling arrays are expected to become less important in the future.

33.2.2

High-Throughput Sequencing and RNA-Seq

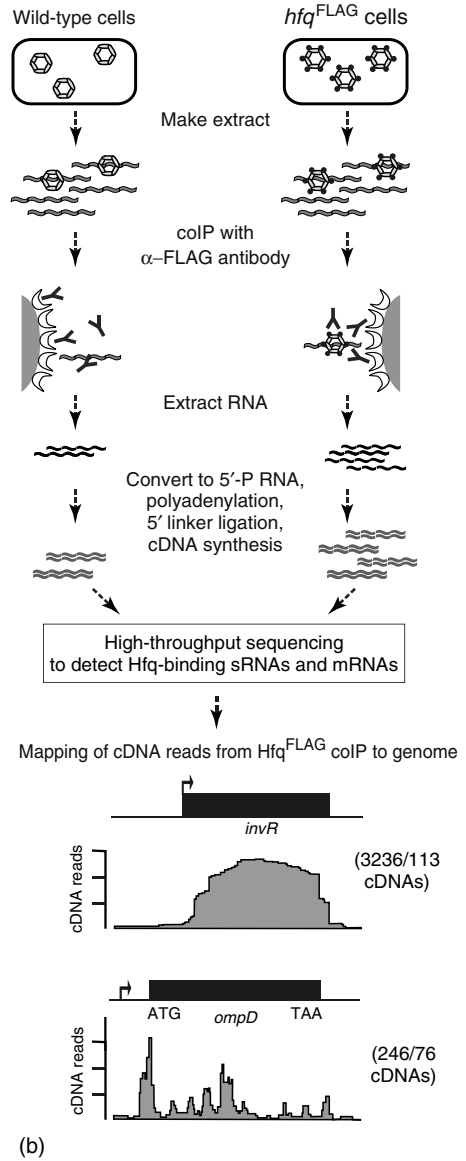
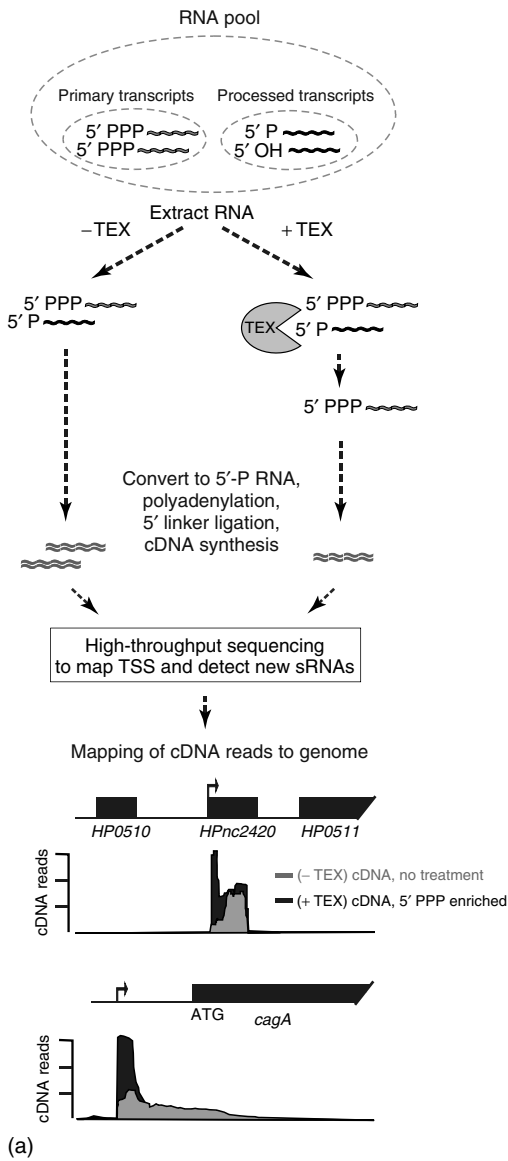
RNomics approaches aimed to identify sRNAs through cloning and sequencing of size-fractionated, reverse-transcribed RNA [30, 31, 40]. Nowadays, high-throughput technologies based on DNA sequencing-by-synthesis strategies have obviated the time-consuming cDNA cloning and costly Sanger sequencing. Currently, three sequencing platforms are used: 454 (Roche), SOLiD (ABI), and Solexa (Illumina). In the choice of a suitable platform, the two main criteria are the desired read length and the required number of reads. As of this writing, the 454 technology can sequence up to 400 nt cDNA and achieve ~1 million reads per run, at comparatively high costs. The SOLiD and Solexa machines only produce reads in the 100 nt range, but deliver tens of millions of reads per sequencing run, at much lower costs. Even though 454 sequencing was used for the pioneering sRNA discovery screens in bacteria [41], the cheaper short-read technologies are preferred nowadays because the majority of reads representing longer RNA species are shorter anyway owing to RNA fragmentation or premature termination during polymerization steps.

To discover sRNAs in a rather unbiased manner by high-throughput RNA-seq, total RNA is extracted and converted to cDNA. In bacterial cells, rRNA and tRNA represent the majority of transcripts, and since they usually are of lesser interest,

they might be depleted before reverse transcription (RT); a twofold reduction in rRNA and tRNA fractions is expected to yield a 10-fold higher coverage of mRNA species [42]. Several techniques have been used to deplete rRNAs and tRNAs: (i) capture of rRNA using antisense oligonucleotides to conserved positions along the 16S and 23S bacterial rRNAs, attached to magnetic beads [43, 44]; (ii) size selection using gel electrophoresis [45, 46]; and (iii) coimmunoprecipitation (coIP) of RNAs associated with a specific RNA binding protein such as Hfq ([41, 47]; see also the following section). For the magnetic capture strategy, the MICROBExpress kit (Ambion), which is capable of removing rRNA from numerous, but not all, bacteria, can be used, although with varying efficiencies (see MICROBExpress product information by Ambion); in fact, MICROBExpress was optimized for *E. coli*, and has been shown to remove 90–99% of total rRNA in three other species (*Bacillus subtilis*, *Staphylococcus aureus*, and *Mycobacterium bovis*). Size selection generally runs the risk of losing interesting long non-coding RNAs. Finally, the coIP approach requires prior knowledge on the particular RNA binding protein, and RNAs that do not associate with this protein will not be recovered by the procedure.

Our laboratory has developed a so-called differential RNA sequencing (dRNA-seq) approach for the global mapping of transcriptional start sites, a procedure that does not generally require prior rRNA removal [48]. The underlying principle is the depletion of processed transcripts, and therefore relative enrichment of primary species (Figure 33.1a). In a bacterial cell, primary transcripts possess a 5' triphosphate, whereas processed transcripts have either a 5' monophosphate or 5' hydroxyl group. This feature allowed us to use the 5' monophosphate-dependent

Figure 33.1 Detection of sRNAs using dRNA-seq or Hfq coimmunoprecipitation (coIP). (a) sRNAs can be identified by high-throughput transcriptomic analysis of total RNA via dRNA-seq. The cellular RNA pool consists of primary transcripts (5' triphosphate ends) and processed transcripts (5' monophosphate or 5' hydroxyl ends). Primary transcripts can be enriched by TEX treatment (the main steps of the approach are schematized; for details, see text). Schematic genome browser profiles illustrate cDNA enrichment patterns for TEX-treated (black) versus untreated samples (gray). sRNAs typically appear as blocks in IGRs, as shown for the HPnc2420 sRNA of *H. pylori*. *cagA* mRNA appears in the characteristic “sawtooth” shape with an elevated sharp 5' flank, since TEX treatment shifts the read distribution toward the nuclease-protected primary 5' end. (b) Strategy to identify sRNA- and mRNA-binding partners of the widespread RNA chaperone Hfq (sketched as hexameric rings). RNAs are coimmunoprecipitated with Hfq in extracts of chromosomal *hfq*^{FLAG} strains using an anti-FLAG antibody (an untagged isogenic strain serves as a control). Comparable to the dRNA-seq approach, the extracted RNAs are converted to 5' monophosphate RNAs by TAP treatment, polyadenylated and ligated to a 5' adaptor before reverse transcription into cDNA. After high-throughput sequencing, the distribution of cDNA reads is mapped to the genome and visualized by a genome browser. Shown are schematic read distributions for the Hfq-dependent InvR sRNA and its target *ompD* from *Salmonella*. Numbers in parentheses indicate the number of cDNA reads derived from coIPs with the *hfq*^{FLAG} strain relative to those obtained with the isogenic untagged control strain (wild type). (Figure adapted from Ref. [41].)



terminator exonuclease (TEX, EPICENTRE Biotechnologies) to degrade processed RNAs, including the abundant rRNAs and tRNAs, before cDNA synthesis.

The workflow of a typical dRNA-seq experiment (Figure 33.1a) starts with total RNA extraction. One half of the RNA sample is treated with TEX to remove processed RNAs while the other half is mock-treated. Then, both samples are treated in parallel as follows: the 5' triphosphate ends of RNA primary transcripts are converted to 5' monophosphates with tobacco acid pyrophosphatase (TAP), followed by poly(A) tailing of RNAs at the 3' end, ligation of a 5' adaptor RNA oligonucleotide, and RT. Finally, the cDNA is amplified and sequenced in a strand-specific manner on one of the three above-described platforms (note that the term “RNA-seq” is misleading since it is the cDNA that is sequenced). After biocomputational removal (clipping) of the linker and poly(A) sequences, the inserts are mapped to the respective bacterial chromosome (and plasmids, if present). We are currently using the Affymetrix Integrated Genome Browser (IGB) to visualize the cDNA sequences and compare the TEX-treated *versus* mock-treated libraries. Primary transcripts of sRNAs are usually enriched in TEX-treated libraries and often pop up as short read blocks in IGRs, whereas mRNAs typically show a sawtooth-like pattern with a sharp 5' end that coincides with the transcriptional start site (Figure 33.1a).

The dRNA-seq approach has been used to map the primary transcriptome and identify sRNAs in diverse species. In *Helicobacter pylori*, a human pathogen previously thought to lack riboregulators, more than 60 sRNAs were discovered [48]. The analysis of the archaeon *Methanosarcina mazei* Gö1 predicted 208 sRNA candidates [49]. In *Chlamydia trachomatis*, an obligate intracellular human pathogen, more than 40 sRNA were found [50]. More recently, dRNA-seq was applied to the cyanobacterium *Synechocystis* sp. PCC6803 [51].

Despite the success of the dRNA-seq approach, some aspects need improvement. Often, 5S RNA or tRNAs are not fully digested by TEX, thus still contributing a substantial fraction of reads. Also, stable secondary structures involving the 5' end might compromise the capability of TEX to initiate RNA degradation at the 5' monophosphate. Furthermore, RNAs carrying a 5' hydroxyl group will not enter the cDNA libraries, as the 5' adaptor ligation requires a 5' monophosphate. In addition, there is a bias toward the 5' ends of transcripts, which is partly attributable to the current read length. Of interest, others have combined TEX treatment with a size selection of transcripts in the 50–350 nt range to identify sRNAs in the plant symbiotic bacterium *Sinorhizobium meliloti* [52].

33.2.3

Hfq Coimmunoprecipitation

Many sRNAs require a protein for their activity and/or intracellular stability [1, 3–7, 53], while other sRNAs even target proteins to modulate their activity [54–56]. Both scenarios imply RNA–protein interactions, which can be exploited to isolate sRNAs *via* the interacting protein. The most commonly used protein so far has been Hfq, a bacterial Sm-like protein that is present in many bacterial species [3–7, 57].

In two pioneering studies, a polyclonal antiserum against Hfq was used to precipitate the protein from *E. coli* lysates, and the associated RNAs were analyzed by direct labeling and electrophoresis in a denaturing gel [32] or by hybridization to tiling arrays [32, 38]. The latter proved to be very successful, since they not only led to the identification of new *E. coli* sRNAs but also produced the first catalog of mRNA target candidates in *E. coli*. A study in the Gram-positive pathogen *L. monocytogenes* also used immunoprecipitation with a polyclonal serum raised against the respective Hfq protein. The sequences of three sufficiently abundant Hfq-associated sRNAs were determined by direct enzymatic sequencing [58].

Recently, our laboratory developed a generic approach, which combines coIP of a FLAG epitope-tagged Hfq (chromosomal tagging at the C-terminus) with deep sequencing of bound RNAs [41] (Figure 33.1b). This method overcomes the limited availability of species-specific Hfq antibodies and tiling arrays, and can be applied to any genetically tractable bacterium, provided that the epitope tag does not interfere with cellular Hfq function. Unlike tiling arrays, which only give a rough estimate of transcript termini, deep sequencing permits to infer the precise 5' and 3' termini of the Hfq-associated sRNAs. Similar to the tiling array approach in *E. coli* [38], the method discovered new sRNAs in *Salmonella*, many of which were impossible to predict *via* sequence conservation since they were absent from *E. coli* and apparently acquired by horizontal gene transfer [41]. Moreover, deep sequencing also identified hundreds of Hfq-associated mRNAs, several of which were subsequently confirmed as sRNA targets, for example, *ompX* mRNA, which was discovered to be repressed by the CyaR sRNA [59].

In an extension of the above approach, we expressed heterologous Hfq variants of *Neisseria meningitidis*, *Aquifex aeolicus*, and *Methanocaldococcus jannaschii* in an *hfq* deletion strain of *Salmonella* [47]. There was a striking enrichment of *Salmonella* sRNAs even for the remotely related Hfq proteins of *Aquifex* and *Methanocaldococcus*, indicating a conserved preference of Hfq for this class of transcripts. In addition, the strategy discovered new sRNAs, several of which barely accumulated in wild-type *Salmonella* but seem to be stabilized or expressed in the presence of a heterologous Hfq protein. Furthermore, the “foreign” Hfq protein of *M. jannaschii* suppressed a crucial processing event in the ArcZ sRNA maturation pathway, increasing the levels of its precursor while depleting the levels of the mature sRNA. More generally, these observations suggest that the converse strategy might also be successful, that is, the well-studied Hfq proteins of *E. coli* or *Salmonella* might be exploited as a bait to fish sRNAs in organisms where an endogenous Hfq protein cannot be identified.

The importance of the coIP/RNA-seq approach will surely grow as more unusual Hfq proteins are identified, for example, a presumably dimeric Hfq protein in *Borrelia burgdorferi* [60]. Moreover, the number of bacterial RNA binding proteins found to be targeted by sRNAs is growing, examples being the regulatory proteins CsrA [54] and Crc [61]. Taking into account the steadily increasing discovery of new sRNA–protein, sRNA–mRNA, and antagonistic sRNA–sRNA interactions, we expect that future research activities will focus on the development of refined

global and affordable analysis tools for the identification and characterization of such regulatory RNA interactions.

33.3

Expression-Independent Searches

The above methods are largely limited to the discovery of sRNAs that are significantly expressed under the applied experimental conditions. This section discusses global approaches for sRNA candidate prediction, which then usually require experimental validation of their expression and predicted size/termini.

33.3.1

Biocomputational Approaches

About a decade ago, newly developed *in silico* methods for the prediction of sRNAs in the genome sequences of related bacteria (e.g., *E. coli*, *Salmonella*, *Yersinia*, and *Klebsiella*) led to a breakthrough in sRNA research similar to what is happening with the revolutionary tool of RNA-seq today. In 2001, three pioneering studies quadrupled the number of known *E. coli* sRNAs [32, 62, 63]. Common to all these approaches was the search for conservation of sequence stretches and putative transcriptional elements (promoters, terminators), or putative RNA secondary structure, in the previously neglected IGRs of bacterial genomes where some sRNA genes had been stumbled on in the 1990s (Figure 33.2). Most importantly, these studies fostered an understanding that a substantial number of sRNAs is conserved in a broader set of species, arguing for physiological importance. The general concept of predicting orphan transcription signals, either with or without concomitant scoring of evolutionary conservation, was picked up by a number of studies. For example, a study in 2002 identified new sRNAs almost solely through the prediction of orphan terminators in the IGRs of *E. coli* [64].

The subsequent development of new software tools was also motivated by an endeavor to find sRNAs in bacterial pathogens unrelated to the narrow group of Gram-negative enterobacteria. For example, the software sRNAPredict2 combined the prediction of sequence conservation and Rho-independent terminators in IGRs to predict new sRNAs in *Vibrio cholerae* [65] and *Pseudomonas aeruginosa* [66]; it was also used to build a kingdom-wide catalog of bacterial sRNAs [67]. The Intergenic Sequence Inspector (ISI) is an sRNA screening algorithm, which automatically downloads genome files from NCBI and creates a library of IGRs. The retrieved sequences are compared with related species using BLAST, and are scored on the basis of variable input parameters, such as length, G/C content, putative promoters, terminators, and conserved secondary structure. The program was instrumental in predicting sRNAs in horizontally acquired virulence regions of *S. aureus* [68].

Predictions based on conservation of potential RNA secondary structure elements rather than primary sequence, as pioneered by the QRNA algorithm [63], were also

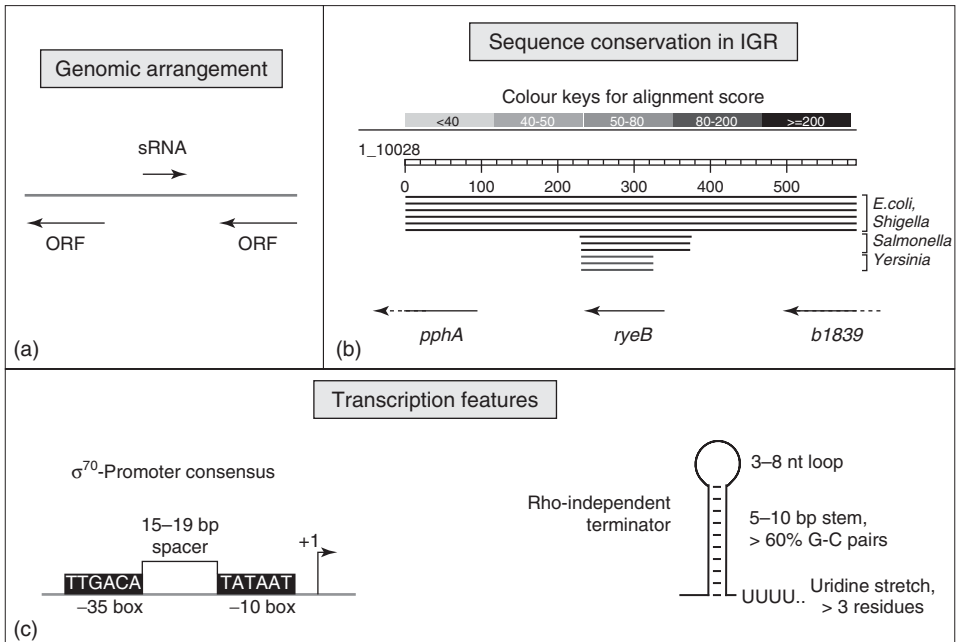


Figure 33.2 Individual criteria of the predictive scheme by Argaman *et al.* [62]. (a) sRNA candidates in intergenic regions that were orientated opposite to both neighboring genes were considered as particularly promising, that is, they could not correspond to mRNA leaders or trailers. (b) BLASTN search result for the intergenic region (including 100 bp of upstream and downstream coding sequence) that harbors the small

RNA gene *ryeB*. Conservation of the entire region from *E. coli* MG1655 is limited to the genomes of pathogenic *E. coli* and *Shigella* strains. However, the *ryeB* sequence itself displays high homology values in related enterobacteria such as *Salmonella* and *Yersinia* species. (c) Transcription features shared by many *E. coli* sRNAs, which served as input to search for promoter and terminator sequences in intergenic regions.

developed further. For example, a study [69] first identified conserved sequence stretches in related cyanobacterial genomes before evaluating RNA structure with the ALIFOLDZ algorithm [70]. In the Gram-positive pathogen *L. monocytogenes*, an *in silico* RNA folding (mfold) step was included in the prediction pipeline [71].

As of 2011, there has been a wealth of biocomputational studies pertaining to bacterial sRNAs, and their strengths and weaknesses were recently reviewed [72]. The growing availability of bacterial genome sequences will surely improve predictions of conserved sRNAs *via* comparative genomics. Biocomputation will become increasingly important for sRNA mining in the accumulating expression data sets generated by RNA-seq of single organisms or communities (meta-transcriptome data), including those of non-culturable organisms. Obviously, these predictions generate numbers of candidates that challenge the current validation schemes tailored to small-scale analyses.

As opposed to global and large-scale predictions, biocomputational searches remain particularly useful to identify a missing sRNA within a given regulon. It appears that the promoters of many sRNAs often carry nearly ideal recognition motifs for the cognate transcription factor; for example, the promoters of the *micA* and *rybB* sRNA genes include almost perfect boxes for binding of σ^E , their transcriptional activator, and are amongst the most highly expressed ones in the σ^E regulon [73]. Accordingly, a dedicated search for conserved orphan promoters with downstream terminators in IGRs is likely to succeed in identifying the sRNAs associated with a given regulon, and indeed discovered the σ^{54} -controlled Qrr sRNAs in *Vibrio* species [74] or the Fur-regulated RyhB analogs in *P. aeruginosa* [75] and *N. meningitidis* [76].

33.3.2

Genomic SELEX

Another approach that can discover sRNAs in an expression-independent manner is “Genomic SELEX”, as recently shown for Hfq-dependent sRNAs [77, 78]. A library of 50–500 nt long RNA fragments was generated from genomic DNA of *E. coli* by random priming, followed by *in vitro* transcription. The genomic RNA pool was then incubated with an RNA binding protein of interest, here Hfq, and high-affinity binders were selected following stringent washes, RT, and PCR amplification. Reiterated cycles aim at gradually enriching RNA species with high-affinity binding motifs, so-called aptamers, and thereby reveal which cellular transcripts might be targeted by the protein *in vivo*. The sensitivity of this approach is evident from the fact that it uncovered an unexpected number of low-abundance *cis*-antisense transcripts to protein-coding genes in *E. coli* [78], which had been missed in the global analyses based on Hfq coIP [38, 41, 47]. As a caveat, however, the same study did not identify many known abundant Hfq-associated sRNAs; only very few (GcvB, OmrA/B, and Tpk70) of these sRNAs were recovered. Perhaps genomic SELEX has a bias toward RNA species that can be well transcribed *in vitro* or that bind with very high affinity to Hfq. Although the enrichment does not necessarily reflect the pool of Hfq-bound RNAs *in vivo*, the approach contributes to an understanding of what types of RNAs the protein might potentially interact with in the cell, especially with respect to low-abundance RNA molecules.

33.4

Deciphering the Biological Role of an sRNA

Here, we outline experiments that will help elucidate the physiological function of a selected sRNA candidate. Unless already known from the discovery phase, the expression profile of an sRNA should be determined first, not only to identify relevant conditions for experimental target searches but also to get a hint in which regulatory network(s) an sRNA might operate. Next, both deletion and

constitutive overexpression strains for the sRNA should be constructed, and their protein profiles compared to the wild-type strain by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Although this strategy is prone to identifying pleiotropic and secondary regulatory effects, it can be a quick means to find direct targets among abundant proteins [18, 79, 80]. A more global and sensitive approach with growing importance for the identification of direct targets is sRNA pulse expression combined with genome-wide mRNA profiling on microarrays [81, 82]. All the above approaches will benefit from the availability of sRNA deletion or overexpression strain libraries for a given organism, which we recommend to construct early if a laboratory plans to specialize in sRNA research. Finally, for sRNAs that are likely to regulate proteins or act as regulators of other RNA molecules in an organism without Hfq, we refer to experiments that have been instrumental in identifying interaction partners. Obviously, biocomputational methods are helpful to predict a first set of putative targets, which can then be compared with the sRNA deletion and pulse-expression approaches.

33.4.1

sRNA Expression Profile

Variations in the expression level of an sRNA of interest can be instructive with respect to biological function, for example, if an sRNA is abundantly present during the response to certain adverse growth or environmental conditions, such as SgrS under glucose-phosphate stress [83] or OxyS under oxidative stress [84]. Northern blots are a good tool because they do not only measure the presence/abundance of the full-length sRNA but also provide information on changes in sRNA processing pathways under specific physiological conditions. In contrast, RNA processing is commonly overlooked in reverse transcription polymerase chain reaction (RT-PCR)-based approaches. Northern blot detection of longer and shorter species of an sRNA will most often reflect processing, although sRNA genes with alternative promoters are also known, for example, *E. coli istR* [13]. Knowledge of the expression profile of an sRNA is also important for subsequent research on its target(s) because regulations will be seen best under conditions when the sRNA is highly expressed. Protocol 1 describes a Northern blot procedure that has worked well in our hands. A variety of probes can be used, from *in vitro*-synthesized long antisense RNAs to custom-synthesized DNA, RNA, and LNA oligonucleotides [18]. To confirm that the obtained hybridization signals are specific, a deletion strain of the sRNA gene in question should be included early in the analysis of RNA expression patterns.

33.4.2

sRNA Deletion

Functional analysis of an sRNA benefits from the availability of a corresponding deletion (knockout) strain, to be complemented by an sRNA expression plasmid. Since there is no conceptual difference to the deletion of a protein-coding gene, we

recommend that readers use the available genetic tools of their favorite organism. As usual, the deletion strain should be otherwise isogenic with the parental strain, which in *E. coli* or *Salmonella* is achieved by phage transduction of the sRNA knockout into a fresh wild-type background. Mutations in bacterial strains can also be generated in a variety of other ways, for example, by chemically induced mutagenesis or transposon insertion, and sometimes mutant libraries will be available through large-scale projects (Section 33.4.5).

For sRNA deletion strategies, several potential pitfalls should be kept in mind. If an sRNA gene is essential, a deletion strain cannot be constructed directly by transformation with linear DNA. On the basis of available data, this is a rare case; exceptions are the RNA subunit of RNase P, 4.5S (SRP) RNA and tmRNA (deletion of the latter was lethal in some bacteria). If one suspects an essential role (i.e., if transformation with linear DNA failed repeatedly), a chromosomal deletion can be introduced in the presence of a plasmid carrying the intact sRNA gene to supply the essential function *in trans*. A second problem can arise from a tight interdigitation of genetic information: an sRNA gene may overlap DNA sequences that are required for proper expression or control of neighboring genes. It is worth noting that intergenic location, as defined by most researchers, only considers the position relative to neighboring open reading frames (ORFs), not to the genes *per se*. Therefore, the promoter of a neighboring gene, or its regulatory sites, may lie within the DNA segment of an sRNA deletion. If so, obtained phenotypic effects can be due to either the absence of the sRNA, a defect in expression of a neighboring gene, or both. In addition, the expression of the neighboring genes might be affected by the transcription of the inserted resistance cassette; such artificial effects can be avoided by subsequent removal of the marker gene using, for example, the FLP recombinase system [85] or by constructing a control strain that carries the sRNA gene plus the resistance cassette.

Once an sRNA deletion strain has been constructed, a first quick analysis of protein profile changes on 1D or 2D protein gels, combined with mass spectrometric identification of protein bands, might already reveal some targets. Examples of successful target identification by one-dimensional SDS-PAGE using sRNA deletion strains include the ABC transporters regulated by GcvB (analysis of periplasmic fractions) [11, 86], the *Vibrio* proteins OmpA and TCP, which were upregulated in a VrrA sRNA-deficient strain [87], or the LamB porin, which highly accumulated upon deletion of the *micA* sRNA gene in a *Salmonella* strain with chronic σ^E stress [88].

However, one may also fail to observe a robust accumulation of the derepressed target proteins in chromosomal sRNA deletion strains. For example, the envelope stress-induced RybB sRNA acts to repress many *omp* mRNAs, and these targets have up to twofold elevated levels in an *rybB* deletion strain. Yet, target protein levels are largely unaltered in this strain, likely because the overproduced fraction is degraded by another surveillance pathway that acts in parallel at the protein level [82]. Even more important, most of the current standard techniques cover only a small proportion of a bacterium's proteome; therefore, unlike microarray-based analysis

of mRNA, the profiling of abundant proteins does not provide a whole-genome view on potential targets.

33.4.3

sRNA Overexpression

Despite the risk of inducing non-physiological global changes, overexpression of an sRNA is usually the best way to find targets. A positive example is the *OmpF*–*MicF* sRNA system: here, a 300 bp DNA fragment proximal to the *ompC* promoter caused depletion of the *OmpF* protein when present on a multicopy plasmid. The DNA fragment turned out to encode *MicF*, the first *trans*-acting sRNA discovered [8, 89].

Typically, the cloning of an sRNA locus into a plasmid with more than one copy per cell should result in overexpression, unless there is feedback regulation. However, it is more straightforward to simply clone the sRNA gene sequence in a plasmid downstream of a strong heterologous promoter. The choice of plasmid may be important since very high copy numbers of an sRNA gene may cause global deregulations that may even culminate in toxicity. For example, constitutive overexpression of *ArcZ* sRNA impacts the expression of >16% of all *Salmonella* genes, likely due to a combination of off-targeting effects of the sRNA itself and intracellular sequestration of *Hfq* [17]. Toxicity can be species-dependent: strong overexpression of *GcvB* sRNA is toxic in *Salmonella*, yet tolerated in *E. coli* [11, 90]. In addition, the host range and the compatibility with potential other replicons must be taken into account when choosing an expression plasmid. Another issue is the choice of promoter to drive expression of the cloned sRNA gene. A strong constitutive promoter we use routinely in our laboratory is the $P_{LlacO-1}$ promoter, which can be tightly controlled with IPTG, particularly in a *lacI*⁺ background [90, 91]. The P_{BAD} (aka *araC*) promoter is also often used but needs special consideration since the inducer *l*-arabinose is a metabolizable sugar [92].

The sRNA overexpression plasmid can be used for both complementation of an already existing sRNA deletion strain and target discovery. Figure 33.3a shows an SDS-PAGE analysis of *Salmonella* strains with wild-type or overexpression levels of *RybB* sRNA, illustrating the readily visible depletion of the major porins in the strain that constitutively overexpresses *RybB* (from a plasmid-borne $P_{LlacO-1}$ promoter).

Secondary structures are often crucial for sRNA function [93–95]. In addition, an increasing number of sRNAs has been shown to use their 5′ end for target selection. Thus, it is advisable to map the transcription start site (+1 site) of an sRNA and to clone the sRNA such that transcripts from the heterologous promoter carry the native 5′ terminus to avoid 5′ extensions or truncations that may compromise the sRNA’s regulatory activity [82, 90]. Transcriptional start sites should be mapped using specialized 5′-RACE (rapid amplification of cDNA 5′ ends) procedures [62, 96, 97] (see also Protocol 2). The +1 sites of expression plasmid promoters are usually known or can be mapped by 5′-RACE as well. However, as an aside to the aficionados of the P_{BAD} system, we have looked into the +1 site of the *E. coli* *araC* promoter [82] and identified a starting nucleotide different from the original publication [98].

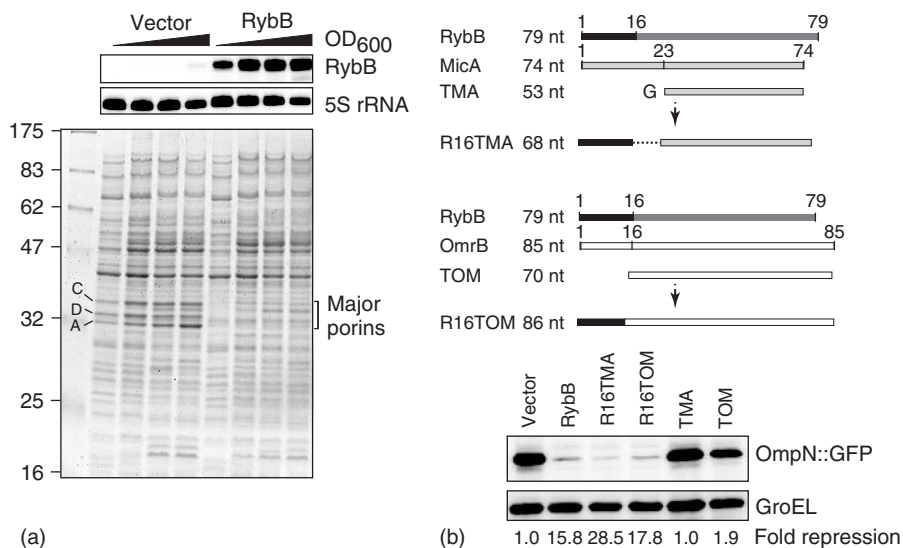


Figure 33.3 Multiple target regulation by RybB sRNA. (a) SDS-PAGE analysis of the global protein profile upon constitutive RybB overexpression in *Salmonella typhimurium*. Protein samples were collected from early logarithmic phase to late stationary phase from a wild-type strain harboring the control vector pJV300 (vector) and a strain that constitutively expresses RybB from a plasmid-borne P_{LacO-1} promoter (RybB). RybB overexpression was confirmed by Northern blotting using a RybB-specific oligonucleotide probe (upper panel). RybB overexpression leads to depletion of the major porins OmpC, OmpD, and OmpA (indicated with capital letters). (Figure adapted from Ref. [29].) (b) Top panel: Schematic drawing of sRNA chimeras R16TMA and R16TOM constructed by fusing the first 16 nt of RybB to truncated versions of *Salmonella* MicA (TMA) and OmpB sRNAs (TOM). An additional guanosine residue was added to the 5' end of the TMA in order to

guarantee efficient transcription. An alignment of RybB sequences from closely related species had highlighted a hyperconserved module composed of the first 16 nt [82]. Using the presented strategy (construction of chimeric sRNAs), these 16 nt were then experimentally confirmed to be the functional module of RybB that regulates the *omp* targets in *Salmonella* [29]. Bottom panel: Western blot analysis of OmpN::GFP regulation in a *Salmonella* $\Delta rybB$ strain transformed with the empty control vector pJV300 (vector; lane 1), the plasmid expressing full-length RybB (lane 2), the chimeras R16TMA and R16TOM (lanes 3 and 4, respectively), or the parental control RNAs TMA and TOM (lanes 5 and 6, respectively) [19]. Fold repression of OmpN::GFP synthesis by the sRNA variants relative to the control (vector only) was calculated based on OmpN::GFP signal intensity and normalized to the corresponding GroEL signal. (Figure adapted from Ref. [19].)

Moreover, our recently developed dRNA-seq method can potentially map all transcriptional start sites in a genome, including those of sRNA genes [48]. Since this or other deep sequencing-based methods will become more affordable, and likely the method of choice in initial sRNA discovery phases, we expect that transcriptional start site information will become readily available for most sRNAs of a given bacterium under study.

33.4.4

sRNA Pulse Expression Combined with Transcriptome Analysis

Transcriptional profiling is now routine, and for many bacterial genomes low-density microarrays are commercially available or can easily be fabricated in a custom-made format. However, wild-type and sRNA deletion strains often show little differences in mRNA expression, if any. Conversely, constitutive overexpression of an sRNA is prone to potential off-targeting and global effects on gene expression, especially if the primary targets include a global transcriptional regulator such as RpoS [22, 99–102].

A pulse expression of the sRNA (for 10–15 min) can counteract these limitations. The basic strategy is to induce the sRNA long enough to cause changes in target mRNA levels, while still keeping the expression phase as short as possible to avoid changes at the protein level, which might cause secondary effects. Changes in mRNA levels are analyzed on whole genome microarrays. For example, pulse expression revealed a role of the RyhB sRNA in global iron homeostasis in *E. coli* [81]. Likewise, the *Salmonella* RybB and *E. coli* OmrAB sRNAs negatively regulated mRNAs of outer membrane proteins [82, 103], which have meanwhile been confirmed as direct targets.

Inducible promoters that can be chosen for pulse-expression approaches are mostly derived from controlled bacterial operons, and include P_{Lac} , the hybrid P_{tac} and P_{BAD} [98, 104, 105] and the bacteriophage λP_L and T7 promoters [91, 106]. These promoters have often been modified to enhance transcription rates and often carry additional regulatory elements (operators) for tight control. The P_{lac} and P_{tac} promoters, which are derepressed by IPTG, are strong yet prone to leakiness when only the standard chromosomally encoded *lacI* (repressor) gene is present. However, a chromosomally or plasmid-encoded *lacI^q* allele greatly reduces the leakiness. Organisms with a codon usage different from that in *E. coli* might benefit from a synthetic *lacI* gene designed for efficient translation.

The P_{BAD} promoter, derived from the *ara* operon, is controlled by the AraC protein and has the beneficial property of being tightly controllable by arabinose, such that strong induction (– glucose, + arabinose) but also tight repression (+ glucose, – arabinose) can be achieved. To avoid significant fluctuations in gene expression in individual cells, *E. coli* strains carrying mutations in transport functions are useful [105]. In general, for P_{BAD} -containing sRNA constructs, host strains must be chosen in which addition of arabinose is not toxic. For example, *E. coli* strains such as MC4100 that are *ara* negative and encode only *araD* are sensitive to arabinose [107]. In addition, proper induction of P_{BAD} requires the host factor CRP (cAMP receptor protein) [108].

A main uncertainty in target discovery by transcriptome profiling is whether a target is expressed under the chosen experimental conditions. Thus, it might be necessary to study several growth conditions to capture all possible targets. Furthermore, in order to be able to see effects, the sRNA must alter the cellular level of its mRNA target(s) rather than affecting protein synthesis only. This is

often fulfilled for repressed targets, but less pronounced in cases of translational activation.

33.4.5

sRNA Libraries

If a given mRNA shows evidence of being an sRNA target (e.g., if regulated at the posttranscriptional level and/or affected by Hfq), a reverse target screening approach interrogating libraries of sRNA deletion strains or overexpression plasmids can be taken to identify the cognate regulator. A library of *Salmonella* strains deleted for 35 conserved sRNA genes was screened for a repressor of OmpX [59], a porin that had been known to accumulate in a Δhfq strain [109] and to be highly enriched (at the mRNA level) by coimmunoprecipitation with Hfq [38, 41]. Here, the CRP-dependent CyaR sRNA was identified as a direct, negative regulator of OmpX [59]. In a search for sRNAs associated with acid resistance, screening of a library comprising 79 sRNA deletion mutants of *E. coli* revealed GcvB as a factor that helps the bacterium to survive under low pH conditions; however, note that the molecular basis – a responsible mRNA target of GcvB – is still unknown [110]. The contributions of *Salmonella* sRNAs to successful host infection were evaluated using a global insertion library that also included 44 sRNA disruption mutants [111]. Strains devoid of the IstR or SroA sRNAs showed compromised replication rates in mice, whereas deficiency in OxyS sRNA caused a surprising and so far unexplained increase in fitness as compared to wild-type bacteria [111]. Another deletion strain library comprising 125 *E. coli* sRNA and small peptide-encoding genes is available from the Storz laboratory [112]. Each mutant carries a unique 20-mer barcode sequence that is introduced together with the resistance marker. This strategy enables simultaneous large-scale screening of the mutant strain collection in growth competition experiments, for example, by selecting for sensitivity to certain stresses. Subsequently, strains that became enriched (= stress-resistant) or depleted (= stress-sensitive) under the respective conditions are identified *via* DNA quantification with standardized microarrays recognizing the introduced barcodes. Screening this library, almost a dozen mutants were discovered to be sensitive or more resistant to envelope stress [112].

A large-scale phenotypic profiling study [113] was performed in *E. coli* using the Keio single-gene deletion library [114], the essential gene hypomorphs [115] and the above sRNA/peptide gene knockout library [112]. A quantitative growth score, which reflects the impact of each gene deletion on *E. coli*'s fitness under diverse stresses, was calculated based on colony size on agar plates. The collected data on synthetic lethal phenotypes, that is, effects seen only when two gene deletions are combined, will surely give valuable hints as to which pathways the *E. coli* sRNAs might have roles in. The database is available at <http://ecoliwiki.net/tools/chemgen/>.

sRNA overexpression libraries have been less common, perhaps because of the risk to induce non-native global and secondary effects, and were initially used to determine the specificity of target regulation in *E. coli* [90]. More recently, a set of 26 Hfq-dependent sRNAs of *E. coli* were cloned under the control of an inducible

P_{BAD} promoter in a multicopy vector, and individually tested for their effects on an *rpoS-lacZ* fusion [101]. The underlying cloning strategy combined 5' end mapping and precise plasmid insertion of the sRNA gene to transcribe it with its native 5' end. This overexpression plasmid library approach identified ArcZ sRNA as a posttranscriptional activator of *rpoS* expression [101], confirming previous hints obtained with the same sRNA gene expressed from its native promoter [32].

Starting with an interesting target or physiological pathway, sRNA deletion or overexpression libraries can be helpful in identifying sRNAs that have roles in the biological process of interest. The conditions for screening should be chosen with care; some sRNAs may regulate different targets at different growth stages, as shown for ArcZ sRNA in *Salmonella* [17].

33.4.6

Finding sRNA-Associated Proteins

Most RNAs are associated with proteins in living cells. These interactions may be fortuitous owing to non-specific electrostatic interactions, or functionally relevant; either the RNA requires a certain protein for its own activity and/or stability or it regulates the activity of the targeted protein [1, 5–7, 53].

One approach to find RNA-associated proteins is to use “bait oligonucleotides” complementary to the sRNA of interest; this approach toward identifying pre-formed RNA–protein complexes from a native source was successfully applied to the isolation of cellular complexes of 6S RNA and RNA polymerase [116]. Conversely, RNA binding proteins can be fished out of a cellular lysate with an *in vitro*-synthesized, immobilized RNA species. The RNA–protein complex then needs to form *de novo*, which requires the RNA to fold properly under the *in vitro* conditions and to withstand nucleolytic activities present in the cell lysate. An example of this approach is the identification of the enzyme tryptophanase as a target of the Rcd RNA encoded by the ColE1 plasmid [56]. Tryptophanase was enriched from cell lysates on a CNBr-sepharose column to which Rcd RNA had been crosslinked.

Aptamer-tagging strategies have increasingly been employed as well. Aptamers are RNA molecules with distinct folds that bind specific ligands (small molecules, proteins) with high affinity, such as streptavidin [117], streptomycin [118, 119], tobramycin [120], or the coat protein of bacteriophage MS2 [121–123], to name just a few. Again, both *in vivo* and *in vitro* approaches can be pursued. As an example for the *in vivo* approach, we tethered aptamer sequences to several *Salmonella* Hfq-dependent sRNA genes of interest to express these fusions in *Salmonella* either from a plasmid or the chromosome. We fused aptamers (MS2, boxB, eIF4) to different (terminal and internal) positions of the sRNA using a plasmid system that can be employed for other bacteria as well [124]. An initial assessment of these fusions with respect to *in vivo* stability and target regulation of the modified sRNAs revealed a tandem repeat of the 20 nt MS2 aptamer, which is recognized by the MS2 coat protein [122, 123], as the best candidate for further analysis. Adapting protocols from the eukaryotic field, we developed an MS2-based copurification

approach employing an MBP-MS2 protein (maltose-binding protein fused to the MS2 coat protein) immobilized *via* the MBP moiety on an amylose column. Proteins associated with *in vivo* expressed MS2-tagged sRNAs were isolated from cell lysates applied to the column. For example, *Salmonella* Hfq was successfully copurified with MS2-tagged GcvB, InvR, or RybB sRNAs [124]. Note, however, that copurification of the Hfq protein does not by itself prove that functional sRNA-Hfq complexes have been isolated.

The Schroeder laboratory *in vitro* transcribed several *E. coli* sRNAs as fusions with a streptomycin aptamer tag for affinity isolation of proteins binding to the respective sRNA moiety [125]. Mass spectrometry analysis of lysate-derived proteins retained by the aptamer-tagged sRNA transcripts on a streptomycin column identified several interaction partners, including Hfq, ribosomal protein S1, and the RNA polymerase β -subunit. Potential disadvantages of this approach include the failure of the *in vitro*-transcribed RNA to properly fold on the column, and the risk to isolate aberrant RNA–protein complexes that form under the applied *in vitro* conditions but not within bacteria.

Both the *in vivo* and *in vitro* aptamer approaches need empirical testing. Since most sRNAs are in a size range of 50–250 nt, the addition of an aptamer sequence – typically between 40 and 80 nt – can impair folding and mask the protein binding site. As a pro of the *in vivo* approach, the stability and regulatory function of the tagged sRNA can easily be tested *via* its target [124]. We recommend not to overexpress the sRNA-aptamer fusion transcript, since this might give rise to toxic protein sequestration effects or artificial interactions. Future applications might combine protein capture by aptamer-tagged sRNAs with crosslinking strategies, which should permit identification of more transient or comparatively weak RNA–protein interactions as well.

33.4.7

Biocomputational Approaches to Find Targets

Whenever target recognition by an sRNA involves formation of an extended intermolecular RNA duplex (>10 bp), target RNAs may be identified by BLASTN searches *via* sequence complementarity. This was successful in the case of RyhB and Spot42 sRNAs [126, 127], where additional information supported the selection of promising candidates from a long candidate list. Another example is the 70 nt long *E. coli* IstR-1 sRNA, which includes a region of >20 nt with almost perfect complementarity to its *tisAB* target mRNA [13]. Nevertheless, a survey of the known sRNA–mRNA interactions shows that this simple approach has a great probability to fail, since base-paired regions can vary considerably in length, are often non-contiguous as in the case of OxyS-*fhfA* [128], may contain bulges, internal loops, or non-canonical base pairs (mainly G-U). Several biocomputational algorithms to predict regulatory sRNA–mRNA interactions have been developed, and the most prominent ones are discussed below. Algorithms to predict RNA interactions with proteins in a genome-wide manner are not available.

The *RNAhybrid* algorithm considers context-dependent stacking interactions of base pairs to predict energetically favorable structures [129]. *TargetRNA2* [130, 131] is similar and calculates optimal hybridization scores based on RNA secondary structure prediction by *mfold* [132]. The output is given in the form of a ranking list of putative targets. *TargetRNA2* has become very popular, perhaps owing to its user-friendly interface, and has identified sRNA targets in several species including *E. coli* [130], *V. cholerae* [133], *N. meningitides* [76], *S. typhimurium* [11] and, more recently, in *H. pylori* [48]. The *TargetRNA2* web platform was recently upgraded with the tool *RNAtarget*, which enables searches for putative sRNA regulators of a given mRNA [131].

The above methods do not consider intramolecular base pairing, and thus do not discriminate predicted binding sites in sRNAs or mRNAs that might be buried in stable intramolecular secondary structures and therefore remain inaccessible *in vivo*. A strategy to solve this problem is to concatenate sRNA and target mRNA sequences *via* a short linker, followed by RNA secondary structure analysis using a suitable algorithm. The web server *sRNAtarget*, which uses this so-called concatenation approach, also considers other features for the prediction of mRNA targets, for example, A/U stretches as potential Hfq binding sites, or the distance between the sRNA seed region and the translation initiation region [134, 135]. Nevertheless, the concatenation method fails to predict complex structures such as double kissing hairpins as present in the OxyS-*flhA* interaction. *RNAup* [136] and *IntaRNA* [137] endeavored to overcome this problem by calculating the accessibility – or “single-strandness” – of regions in sRNA and target mRNA. Other programs are available as well, for example, an algorithm that evaluates base pairing, bulges, internal loops and stacking effects (see [71] for an sRNA screen in *L. monocytogenes*).

In general, the available biocomputational tools still tend to give many false-positive results. Therefore, it is advisable to combine several prediction algorithms and complement the *in silico* predictions with an experimental component (see above). More importantly, many sRNAs are known to be conserved in a wider range of species, and an alignment of the homologs usually reveals highly conserved subregions or “domains” in these sRNAs. These conserved seed regions of 10–20 nt in length usually recognize mRNAs by short seed pairings, for example, the G/U-rich consensus region R1 in GcvB sRNA that base-pairs with many mRNAs of genes involved in amino acid uptake and metabolism [11, 138] or the 5′ proximal R16 domain of RybB sRNA that recognizes *omp* mRNAs by seed pairings of 7 bp [28, 29]. If merely these short regions representing only 10–20% of the entire sRNA are used as query sequence, target predictions are often much more reliable and successful [29].

33.5

Experimental Target Validation

Predicted candidate target mRNAs require an independent experimental validation to confirm that the regulation occurs at the posttranscriptional level through

direct base pairing. This section outlines *in vivo* and *in vitro* approaches for target validation.

33.5.1

Reporter Gene Fusions and sRNA Chimera

Plasmid-borne or chromosomal fusions of predicted target mRNAs to a reporter gene (e.g., *gfp*, *lacZ*, or *luc*) are commonly used to confirm and quantify the regulatory effect of an sRNA. Given that most sRNAs act in the 5' regions of targets to regulate protein synthesis (regardless of whether they impact translation or mRNA stability), reporter genes are often constructed such that the entire 5' UTR and AUG start codon of the target are fused to the early start codon-less ORF of the reporter gene. Since an increasing number of sRNAs are found to act downstream of the start codon [17–19, 27, 29, 139], we strongly advise that a considerable part of the target's coding sequence (CDS) – at least the first 15 amino acids – is included in the reporter gene fusion construct. This will also increase the chance that all Hfq-binding or ribonuclease recognition sites necessary for target regulation are contained. However, it also needs to be considered that the 5' CDS of some genes includes sorting sequences for membrane targeting or export, and the full inclusion of such sequences might result in lethality (*lacZ* fusions) or inactivity of the reporter protein (*gfp* fusions).

Our laboratory has developed a modular plasmid system for target validation that utilizes translational fusions to *gfp* [90]. In this system, the sRNA and target mRNA of interest are each expressed from constitutive heterologous promoters to uncouple them from their native transcriptional control in order to primarily unveil posttranscriptional effects. Several plasmid variants are tailored to different purposes, such as the direct cloning of targets as 5'-RACE products, or the integration of internal gene fragments from operons to mimic sRNA targeting within polycistronic transcripts (Figure 33.5). This plasmid series has been successfully used for the validation of many targets in *E. coli* and *Salmonella* (Figure 33.3b and Figure 33.6c) [18, 19, 29, 59, 90, 140] as well as in more distant organisms such as *Neisseria* [141]. The advantages of this system are the ease with which compensatory base-pair exchanges can be introduced into sRNA and target [59], and the fast and statistically highly robust data acquisition from tens of thousands of single cells by flow cytometry [29]. Detailed descriptions of the available plasmids and cloning procedures are found in Protocol 3.

A conceptually similar approach uses a modified 5'-RACE protocol, which enables the direct insertion of putative targets into the chromosome, under the control of a P_{BAD} promoter fused to *lacZ* [142]. An advantage of this system is that β -galactosidase assays for the *lacZ* reporter are not only a standard technique in many laboratories but also provide precise quantitative readouts.

Once regulation has been established, the modular plasmid system provides quick tools for more detailed mechanistic studies. For the identification of the binding sites, an alignment of the sRNA sequences from closely related species can pinpoint hyperconserved sRNA “domains” involved in RNA duplex formation

with targets [11, 28, 29]. To test if such a “domain” does indeed determine target recognition, it can be either deleted from the sRNA or its sequence fused to an unrelated sRNA to create an sRNA chimera (Figure 33.3b) [18, 19, 29]. The latter requires that the sequence of the scaffold sRNA must not by itself regulate the target of interest. Furthermore, the intracellular stability of the newly generated artificial sRNA should be assessed on a Northern blot (Figure 33.3b). As an example, a fusion of the first 13 nt of MicC sRNA (~100 nt in length) to an unrelated scaffold RNA was fully sufficient to regulate the *ompD* target mRNA, which then aided the prediction of an 11 bp MicC-*ompD* duplex [18].

Another valuable approach is random mutagenesis of the target fusion or the sRNA. For example, in the case of MicA sRNA and *lamB* mRNA [88], a translational fusion of *lamB* to *lacZ* under control of the P_{BAD} promoter was integrated into the chromosome. The *micA* gene was amplified under error-prone PCR conditions, integrated into the chromosome as well, and bacteria were screened for loss of regulation of the *lamB::lacZ* fusion. Recently, this approach was also successfully used to define multiple-target interactions of RybB [28].

The gold standard in proving an sRNA–mRNA interaction is restoration of a regulatory interaction disrupted owing to mutations in one RNA through introduction of compensatory mutations into the other RNA. For sRNA–mRNA pairs with short conserved seed pairings, single point mutations often suffice to abrogate the regulation [29, 143]. Figure 33.6c illustrates the example of RybB-mediated repression of an *ompC::gfp* fusion [29]. A single C to G mutation was introduced at the second position of the highly conserved R16 sequence of RybB, causing loss of regulation, which was then fully restored with a compensatory G to C change in the conserved target site of *ompC* mRNA.

Longer target interactions might require more extensive changes; a good example is MicC sRNA, wherein a point mutation fully abrogates the 11 bp interaction with *ompD* mRNA, but not that with *ompC* mRNA recognized by a longer pairing [18, 144]. There are also cases where sRNAs seem to be able to form alternative interactions within the same or nearby target sites, as in ABC transporter mRNA regulation by GcvB [11], or in RybB-mediated repression of *ompD* mRNA which carries a similar alternative seed sequence [28, 29]; in these cases, compensatory base pair changes are difficult to design. In other cases, mutations in sRNA or target might result in unfavorable changes in RNA structure, which jeopardizes the success of compensatory mutations, as it has been argued for RyhB sRNA pairing with *iscRSUA* mRNA [145].

33.5.2

***In vitro* RNA–RNA Footprinting**

Despite many years of research into antisense regulation, we are still unable to predict the potential interaction between a given sRNA and its target with high confidence based on structural considerations. Nonetheless, knowing the structure of the interaction partners might give hints as to accessible regions or the presence

of RNA elements favorable for antisense pairing, for example, C-rich apical loops [59, 95]. For short RNAs, <120 nt in length, widely used biocomputational algorithms such as *mfold* [132], *RNAStar* [146], or *RNAfold* [147] can often predict secondary structure elements with reasonable confidence. This is much harder for longer RNAs, owing to the exponential increase of energetically favorable folding schemes with molecule length. Regarding target mRNA structure, one should bear in mind that in an *in vivo* situation, the sRNA faces a population of nascent target mRNAs of different lengths and folding states.

In vitro probing has successfully predicted secondary structures of sRNAs or cognate targets, alone and in complex, which were proved to be biologically relevant because they could subsequently be validated through mutational analysis *in vivo* [11, 18]. Following *in vitro* T7 transcription and gel purification (Chapter 1), the denatured RNA is first refolded and then subjected to partial cleavage using ribonucleases with specificity for single-stranded (RNase A, T1, or T2) or double-stranded (RNase III or V1) regions. Similarly, chemical backbone cleavages by lead(II) ions or hydroxyl radicals, or chemical modification of RNA bases by dimethylsulfate (DMS), kethoxal, or 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-*p*-toluenesulfonate (CMCT), have been popular methods to obtain structural information [148–150]. For more information on these enzymatic and chemical probes, including specific protocols, the reader is referred to Chapters 10, 13 and 15. Depending on the length of the analyzed RNA molecules, two experimental strategies are common. Short RNAs are usually labeled at the 5' or 3' end, and are directly analyzed by denaturing polyacrylamide gel electrophoresis (PAGE) after the probing reaction. Longer RNAs as well as RNA molecules that were subjected to modification rather than cleavage reactions are usually analyzed by RT, and the abortion of cDNA synthesis then provides a means to deduce cleaved or modified positions. Primer walking, that is, the use of several different RT primers, is necessary to cover longer RNA molecules over their entire length since standard PAGE is limited to a resolution of 200 nt or less. The information derived from probing experiments can then be used to evaluate and refine the results of biocomputational predictions of sRNA–mRNA pairings.

The interactions of *in vitro* synthesized RNAs can also be assessed by other means. The kinetics of binding of antisense-target RNA pairs may be determined by gel-shift assays using denaturing or non-denaturing gels, depending on the expected length and the G/C content of the base-paired regions [151, 152]. Such gel-shift assays can be carried out in the presence or absence of RNA chaperones such as Hfq. Also, cross-validation of probing data can be achieved when labeled sRNA is probed in the presence of unlabeled mRNA target in one setup, and labeled mRNA in the presence of unlabeled sRNA in a second setup. If the interaction is long enough, an sRNA will protect the binding region in its target from cleavage – in other words, leave a “footprint” – which can then be used in combination with the reciprocal footprint of the mRNA on the sRNA to determine a minimal binding region [11, 16, 153]. Note that we have often seen footprints in sRNAs or targets even under conditions when the two partners seemed to interact

only weakly, as judged by gel-shift assays (e.g., MicC-*ompD* [18]). In other words, *in vitro* RNA structure probing of sRNA-target complexes is indeed a sensitive method to localize interaction sites.

33.5.3

***In vitro* Characterization of sRNA Function**

The control of translational initiation at the level of 30S ribosome binding is the paramount mechanism of mRNA regulation by base-pairing sRNAs. Whether or not an sRNA impairs or promotes the formation of 30S–mRNA complexes can be studied in a so-called 30S toeprinting assay [154], first applied by Altuvia and coworkers [155] to demonstrate inhibition of 30S binding to *fhlA* mRNA by OxyS sRNA. In brief, an *in vitro* synthesized fragment of the 5′ proximal mRNA region, including parts of the CDS, is incubated with purified 30S ribosomal subunits and initiator tRNA. Next, RT initiating from a radiolabeled primer that binds in the early CDS is performed. If the 30S subunit has stably associated with the RBS, RT will terminate ~15 nt downstream of the start codon because of steric hindrance; this produces a characteristic “toeprint” signal, observable after electrophoresis of the radiolabeled cDNA on a denaturing gel along with a sequencing ladder. If the prior annealing of an sRNA inhibits 30S binding, RT will read through the sRNA–mRNA complex to produce a full-length cDNA, or generate an sRNA-specific toeprint on the target mRNA if the sRNA–mRNA duplex is sufficiently stable (Figure 33.7a). A protocol that has worked well in our hands is provided in Protocol 6.

The toeprint assay has been successfully used to characterize many different sRNAs that negatively regulate their targets [11, 14–16, 127, 144, 149, 153, 155, 156], as well as to define a “five-codon window” downstream of the start codon in which sRNAs are likely to achieve translational repression in the early CDS, without sequestration of the Shine–Dalgarno (SD) sequence or start codon [19]. In principle, the toeprint assay should also be useful to study the activation of translation by sRNAs pairing to the 5′ UTR. Successful toeprint experiments should be preceded by gel-shift assays (as described above) to determine optimal binding conditions for the sRNA–mRNA pair and to clarify if Hfq is involved in the annealing process. We include several control reactions to verify that the RT stop indeed reflects the formation of a ternary ribosome/tRNA^{fMet}/mRNA complex (Figure 33.7b; lanes 1–3). Furthermore, a control sRNA – either unrelated to the sRNA under study or the same sRNA with its mRNA interaction site disrupted by mutation – should be included in the experiment in order to rule out that changes in the toeprint signal are due to non-specific RNA competition effects.

sRNA-mediated regulation of protein synthesis can also be studied in a cell-free translation system [157, 158], although in the more classical approaches using crude bacterial extracts the added mRNAs and sRNAs may be rapidly degraded. Such limitations are overcome with more recently developed cell-free translation

systems composed of purified RNA polymerase, ribosomes, tRNA, and ribonucleotides [159]. The corresponding commercial system is commonly known as the PURESYS[®]TEM[®], or PURE[®]Express[®]. Details of the study of translational control by sRNAs with the PURESYS[®]TEM[®] are given below (Protocol 5). This *in vitro* system has also enabled us and others to study the contribution of Hfq in sRNA-mediated translational repression, since RNAs transcribed *in vitro* in the PURESYS[®]TEM[®] are stable for 30 min with or without Hfq [11, 160]. Such experiments are difficult *in vivo* because Hfq-dependent sRNAs are usually unstable in strains lacking the protein.

Hfq-dependent mRNA repression by sRNAs has been studied with the PURESYS[®]TEM[®] for GcvB-*gltI* [11], SgrS-*ptsG* [160], and MicC-*ompC* [18]; in the latter case, a 4.7-fold repression of *ompC* mRNA was observed. Also, the PURESYS[®]TEM[®] provided the first direct evidence that GlmZ sRNA alone can promote translational activation, regardless of enhanced activation in the presence of Hfq: GlmZ sRNA was shown to increase *glmS* mRNA translation 1.8-fold, while a 14.8-fold activation was observed when both GlmZ and Hfq were added [140].

As with toeprints, optimal conditions for sRNA–mRNA complex formation should be established first using gel-shift assays. Furthermore, it is important to adjust the mRNA concentration and time window such that the linear range of the protein synthesis reaction is measured. Finally, one needs to bear in mind that not all sRNA-mediated regulations can be recapitulated with the PURESYS[®]TEM[®]. For example, MicC binds its *ompD* target in the CDS to repress it at the level of mRNA stability without impacting translation [18]. As expected, MicC was unable to repress translation of an *ompD::gfp* fusion mRNA, but downregulated a fusion of its other target, *ompC*, through masking of the *ompC* RBS.

Given the increasing importance of mRNA decay [161], it will be important to develop *in vitro* systems in which induced mRNA degradation by sRNAs can be detected, perhaps by complementation of the PURESYS[®]TEM[®] with purified RNase E and other degradosome components. A disadvantage of this commercial *in vitro* translation system is that details of its precise composition/ingredients are not disclosed, which may hamper the development of new assays.

33.6

Conclusions

In the first part of this chapter, an overview of strategies for the identification and characterization of sRNAs was provided. Since the first edition of this handbook, the number of characterized sRNAs and the methods for their investigation have grown considerably. Several aspects are expected to be important for future studies. First, high-throughput sequencing is likely to further complement or fully replace microarray-based approaches in the near future, taking into account that the availability of sequencing facilities is growing and the costs are decreasing; we also expect the scope of applications for these technologies to expand, for example, in efforts to screen for specific RNA–RNA or RNA–protein interactions in cells.

So far, our efforts have been focused on target identification, reflecting our conviction that most sRNAs of so far unknown function will turn out to act as regulators, either acting on RNAs (as antisense RNAs) or on proteins (by sequestration or by modulating their activities). At the same time, more “structured RNAs” likely await their discovery [162, 163], which may serve more general functions similar to tmRNA, SRP RNA, or 6S RNA. Whether or not these sRNAs will require different approaches for discovery or characterization is difficult to predict, but the identification of molecular interaction partners will be useful in any case. Regulatory RNAs in bacteria cover a broad variety of targets, functions, and mechanisms. Thus, additional mechanisms of riboregulation are expected to be discovered in bacteria, and will in turn drive the development of new methodology. In addition, more sRNAs with dual functions as mRNA and antisense regulator, such as SgrS or RNAIII [164, 165], are likely to be stumbled on, and the analysis of such molecules – specifically, an understanding of how they can act as regulators of translation when also being decoded by the ribosome – may require new tools for their investigation. Similarly, riboswitch-derived regulatory RNAs with *trans*-targets as recently postulated for *L. monocytogenes* [166] call for new methods to investigate how their association with tiny ligands impacts regulation *in vivo*. Finally, CRISPR/Cas systems, which constitute a rapidly growing class of molecular machines using small RNA species as guides for their specific activities, are expected to prompt new approaches for the study of how RNA interacts with DNA [53, 167].

Many questions in the sRNA field remain to be answered. For example, recent studies of clonal bacterial populations increasingly demonstrated a high degree of phenotypic variability between single cells (reviewed in [168–170]). Interestingly, most heterogenic bacterial populations seem to originate from responses to specific environmental stresses. One may expect sRNAs to be important in cell-to-cell variances since they act faster than transcription factors, thanks to their short time of RNA synthesis and the lack of requirement to be translated into protein. There have been first mathematical attempts to simulate the differences between sRNAs and proteins as regulators, but advanced methodology to address these issues is still lacking [171–173]. Another intriguing phenomenon is bacterial aging, as demonstrated in *C. crescentus* [174], *E. coli* [175], and *B. subtilis* [176]. Could sRNAs possibly help a bacterial cell to face stresses that are associated with aging, such as the accumulation of aggregated proteins [177]? The development of single-cell sequencing technologies will allow us to address these questions in a global manner. Eventually, there is an increasing demand for methods to investigate the intracellular localization and perhaps dynamic behavior of sRNAs in single live cells. Except for tmRNA [178], the intracellular sites where sRNAs exert their function in the bacterial cell are unknown, and the localization of important associated proteins such as Hfq or RNase E has remained controversial [179–184]. Thus, sRNA localization studies, at best using single-molecule detection methods as done for proteins [185], will be a major challenge of future research in the sRNA field.

Protocols

General Comments

Working with RNA does not require special handling. In our laboratory, we try to follow good laboratory practice by working with gloves and trying to keep materials clean using 70% ethanol. However, all reagents and materials should be RNase-free. We generally use Millipore water (TKA GenPure UV-TOC/UF, #08.2227; TKA, Niedelbert, Germany; conductivity: $0.055 \mu\text{S cm}^{-1}$), which is further autoclaved before use. Eppendorf tubes and tips are autoclaved as well. From our point of view, DEPC treatment of water is unnecessary.

Protocol 1: Northern Blot Protocol

We routinely use the RNA extraction part of the protocol provided below for preparation of larger quantities of RNA (100 μg to 1 mg) from *E. coli* and other enterobacteria. If smaller amounts of RNA are needed for small-scale analysis, the more expensive TRIzol (Life Technologies Invitrogen) reagent is recommended (see Protocol 2).

Required Reagents

- STOP mix: 95% ethanol, 5% phenol (v/v)
- RNA lysis buffer: 100 mM Tris-HCl, pH 7.5, 40 mM EDTA, 200 mM sodium chloride, 0.5% SDS
- Water- or buffer-saturated phenol
- Chloroform
- 100 and 75% ethanol solutions
- RNase-free DNase I (Roche Diagnostics; 20 U μl , #04716728001)
- RNA loading dye: 95% (v/v) formamide, 0.1% (w/v) bromophenol blue, 2 mM EDTA
- DNA size ladder (we used the Fermentas* pUC Mix Marker, 8 (not offered anymore), with 67, 110/1, 147, 190, 242, 331, 404 bp plus several larger fragments)
- Hybond N+ blotting membrane (GE Healthcare, #RPN203B)
- Hybridization buffer: 0.5 M phosphate buffer, pH 7.2, 7% SDS, 10 mM EDTA
- Hybridization buffer for RNA probes: 50% deionized formamide, 7% SDS, 250 mM sodium chloride, 120 mM sodium phosphate buffer, pH 7.2
- Spin columns, for example, illustra MicroSpinTM G25 (GE Healthcare, #27-5325-01)
- 20 x SSC: 3 M sodium chloride, 300 mM trisodium citrate, adjusted to pH 7.0 with HCl
- Membrane wash buffers: 2 x SSC, 0.1% SDS; 1 x SSC/0.1% SDS; 0.1 x SSC/0.1% SDS
- Phosphorimager (Typhoon FLA 7000, GE Healthcare)
- AIDA software (Raytest, Germany) or ImageJ (<http://rsbweb.nih.gov/ij/>)

Step 1: RNA Extraction

- 1) Dilute an overnight bacterial culture in 50–100 volumes of growth medium.
- 2) Grow to the OD₆₀₀ of choice.
- 3) For RNA extraction, withdraw the required aliquot of culture volume and stop growth by adding 0.2 volumes of pre-chilled STOP mix (95% ethanol/5% phenol (v/v)).
- 4) Mix thoroughly.
- 5) Pellet the bacterial culture by centrifugation (~3500 g, 10 min, 4 °C) and discard the supernatant (at this point, samples can be stored at –20 °C after snap freezing in liquid nitrogen).
- 6) Resuspend the pellet in 0.5 volumes of RNA lysis buffer preheated to 65 °C, by gently pipetting up and down. For example, if you have started with a 10 ml culture, 5 ml of RNA lysis buffer is added. For samples of OD₆₀₀ > 1.5, use 1 volume of RNA lysis buffer.
- 7) Immediately place at 65 °C for 5 min.
- 8) Add an equal volume of water- or buffer-saturated phenol preheated to 65 °C and vortex thoroughly.
- 9) Centrifuge (~10 000 g, 30 min, 4 °C) and transfer the aqueous phase to a fresh RNase-free tube.
- 10) Add an equal volume of chloroform and vortex thoroughly.
- 11) Centrifuge (~10 000 g, 30 min, 4 °C) and transfer the aqueous phase to a fresh tube.
- 12) Add 0.1 volumes of 3 M sodium acetate (pH 6.5) and 2.5 volumes of ice-cold EtOH.
- 13) Precipitate at –20 °C for 1–3 h.
- 14) Pellet the RNA by centrifugation (at least 12 000 g, 30 min, 4 °C), discard the EtOH supernatant, wash the pellet with ice-cold 75% EtOH and centrifuge as above but for 10 min. Remove the supernatant and air-dry the RNA pellet at room temperature. Redissolve the pellet in, for example, 500 µl of H₂O when RNA has been isolated from 50 ml of an *E. coli* culture grown to an OD₆₀₀ of 1.0.

Step 2: DNase I Treatment

- 1) Treat RNA samples with RNase-free DNase I for a maximum of 1 h according to the manufacturer's protocol (see also Protocol 2, Step 4 "Removal of DNA").
- 2) Repeat phenol extraction as above, but use phenol adjusted to room temperature.

Step 3: Gel Separation and Northern Blot

- 1) Quantify the RNA by measuring absorbance at 260 and 280 nm in a UV spectrophotometer.

- 2) Analyze a small aliquot of each RNA sample on a 1% RNase-free agarose gel with 1 x Tris/Borate/EDTA (TBE) as electrophoresis buffer. This permits a rough assessment of RNA integrity. The three ribosomal RNA bands (5S, 16S and 23S rRNA) should be clearly visible and without substantial smear. Comet-like tails from the 16S down to the 5S rRNA bands are indicative of degradation and, at this stage, one should consider repeating the RNA preparation.
- 3) Prepare a 5–10% denaturing polyacrylamide (PAA) gel (7–8 M urea, 1 x TBE), depending on the size of your RNA of interest.
- 4) Pre-run your gel at a sufficiently high voltage, such that the gel temperature is perceived as hot (50–60 °C). This ensures a continuously denatured state of your RNAs during electrophoresis.
- 5) Load onto the gel 10 µg of RNA per sample after addition of an equal volume of RNA loading dye. Also load a lane with an appropriate DNA or RNA ladder that has been end-labeled with [γ -³²P]ATP. In our hands, the Fermentas pUC Mix Marker, 8, fulfilled this purpose, including diverse size markers ranging from 19 to 1118 bp. RNA size markers are expensive and may not be worth the more precise size prediction.
- 6) Cut out six sheets of 3MM paper of the size of your gel region of interest; pre-wet them in 1 x TBE.
- 7) Cut a blotting membrane of the same size (GE Healthcare Hybond N+ recommended); pre-wet in 1 x TBE.
- 8) After the gel run, detach one of the glass plates and apply one of the pre-wetted 3MM papers to the desired region of the gel. Start from one end and avoid trapping of bubbles between the paper and the gel; stack two more 3MM paper sheets on the first. Peel the gel off the second glass plate and turn around such that the naked gel surface is facing up.
- 9) Carefully place the membrane on the naked gel surface. Once in contact with the gel, avoid any lateral movement! Using a Pasteur pipette, gently roll out any bubbles that may have been trapped between the gel and the membrane.
- 10) Complete the sandwich setup by placing the last three sheets of 3MM paper on top of the membrane, rolling out any bubbles introduced during this process.
- 11) Transfer the RNA to the membrane by either electroblotting (preferred) or capillary-blotting techniques. Consult the manual of your blotting apparatus for an appropriate protocol. If a radiolabeled size marker is used, its successful transfer to the membrane can be evaluated with a hand-held β -radiation detector.
- 12) After the transfer, disassemble the sandwich and crosslink the RNA to the membrane either by baking at 80 °C for 1 h or preferably by using a UV crosslinker (for energy settings and crosslinking times, see manual of your UV crosslinker). After crosslinking, rinse the membrane with H₂O.
- 13) Place the membrane in a hybridization tube and wet it with hybridization buffer (20 ml for a membrane of 20 × 15 cm). Pre-hybridize at the chosen hybridization temperature for at least 30 min while preparing your radioactive

- probe. Note that the ends of your rolled membrane should not overlap in the tube.
- 14) Denature the probe by heating to 95 °C for 5 min. For single-stranded DNA oligonucleotide probes, 5–50 pmol of a 5'-end-labeled oligonucleotide should be used. Although DNA may suffice to detect most sRNAs, single-stranded antisense RNA probes are recommended if higher sensitivity is required. Such riboprobes can be prepared by *in vitro* transcription with T7 RNA polymerase in the presence of labeled rUTP from PCR templates (created such that the sRNA region is transcribed in antisense direction from a T7 phage promoter introduced by the downstream PCR primer). Approximately 10⁶ cpm of riboprobe per ml hybridization buffer should be used. To decrease background signals, labeled DNA oligonucleotides as well as riboprobes should be freed from unincorporated nucleotides by column purification (e.g., illustra MicroSpinTM G25).
 - 15) Add denatured probe to the tube and hybridize for at least 6 h at the temperature of choice. We generally hybridize DNA oligonucleotide probes at 42 °C, and riboprobes at 60–70 °C using the hybridization buffer for RNA probes.
 - 16) Discard the liquid containing the radioactive probe appropriately or save for another membrane.
 - 17) Rinse the membrane briefly in a small volume of 2 x SSC, 0.1% SDS.
 - 18) Wash the membrane at the hybridization temperature in three subsequent 15 min washing steps, using at least 100 ml for each step (first wash: 2 x SSC/0.1% SDS, second wash: 1 x SSC/0.1% SDS, and third wash: 0.1 x SSC/0.1% SDS).
 - 19) Remove any liquid draining off the membrane. Seal the membrane in plastic and expose a Phosphorimager screen to it overnight.
 - 20) Analyze the screen using a Phosphorimager (Typhoon FLA 7000, GE Healthcare). We routinely use the commercial AIDA software (Raytest, Germany) to analyze our screens; however, the free software ImageJ (<http://rsbweb.nih.gov/ij/>) is a good alternative.

Protocol 2: 5'-RACE Protocol for Mapping of Transcription Start Sites and Processed 5' Ends

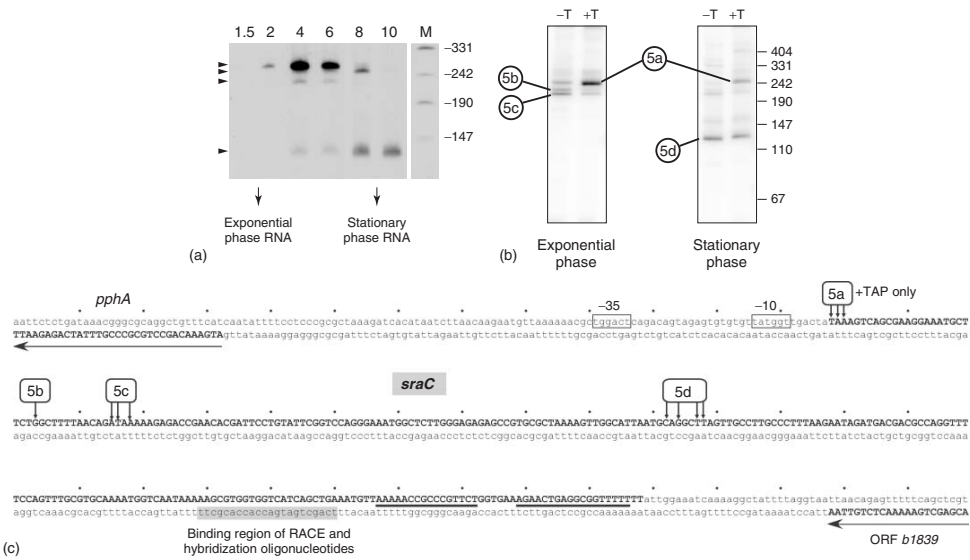
Although primer extension analysis is commonly employed for 5' end mapping of RNA, it is less well appreciated that this method fails to distinguish between primary 5' ends and those generated by processing. To identify the start of transcription and to distinguish it from processed 5' ends of a small RNA, a 5'-RACE protocol that includes treatment with TAP is more suitable. Additional advantages of 5'-RACE are the small amount of RNA required and that it obviates the use of radiolabeled oligonucleotides. Primary transcripts in bacteria carry a 5' triphosphate, which is hydrolyzed by TAP specifically between the α - and β -phosphate groups. Through the resulting 5' monophosphate, these RNAs are subsequently ligated to the 3' hydroxyl group of an RNA oligonucleotide (5' adaptor), which is followed by RT with

a gene-specific deoxyoligonucleotide, and subsequent PCR amplification using a 5' adaptor primer and a nested gene-specific primer. TAP treatment is expected to yield a specific or at least strongly enhanced signal for primary transcripts in the amplification step, as compared to untreated RNA samples. However, 5' ends that result from processing and if retaining a 5' monophosphate will also be amplified and analyzed in parallel (Figure 33.4).

The protocol below is the one commonly used in our laboratory. Given are conditions and primers for 5' end analysis of the small *SraC* RNA of *E. coli*, which may serve as a control when establishing the method for other sRNAs. The *sraC* locus yields four different RNAs (Figure 33.4a) that vary in size and abundance depending on growth conditions, as detected by Northern hybridization [62]. Our RACE analysis revealed that only one of these four species represents a primary transcript (band 5a in Figure 33.4). This protocol derives from [96], but has been optimized for sRNAs. For example, we use a thermostable reverse transcriptase enzyme to perform cDNA synthesis at an elevated temperature, which should significantly increase the yield of full-length cDNAs with authentic 5' ends, particularly in the case of highly structured RNAs. If 5' ends of several sRNAs need to be determined in parallel, it is possible to prime cDNA synthesis with several gene-specific oligonucleotides in the same reaction tube (see Step 7

Figure 33.4 Identifying primary and processed 5' ends of an sRNA. (a) Northern blot analysis of *SraC* sRNA expression in *E. coli* cells at different growth stages. (Figure adapted from Ref. [62].) Total RNA was prepared 1.5, 2, 4, 6, 8, and 10 h after dilution of an overnight culture into fresh LB medium. Four distinct RNA species are observed, indicated by arrowheads on the left. Approximate growth points at which RNA was sampled for 5'-RACE experiments (exponential and stationary phase RNA) are indicated by arrows below the blot picture. M, size marker (in nucleotides). (b) RACE mapping of *SraC* 5' ends at the two growth points indicated. Total *E. coli* RNA was linked to a 5' adaptor RNA without or after treatment with TAP (lanes -T and +T, respectively). TAP converts the 5' triphosphate of an RNA to a monophosphate. A comparison can therefore identify RNAs derived from primary transcripts versus 5'-processed RNAs (carrying 5' monophosphates). Subsequently, cDNA was made and the 5'-end sequence of the respective sRNA was amplified using gene- and adaptor-specific primers. PCR products were then separated on a native 5% polyacrylamide gel, and bands of interest

were excised, cloned, and sequenced. Only one out of four products, that is, 5a, was strongly enhanced after TAP treatment, thus identifying 5a as the primary transcript of the *sraC* gene. TAP treatment had little effect on the yield of RT-PCR products 5b–5d, indicating that they represent processing products of the primary *sraC* transcript. Note that bands on Northern blots (panel A) and amplified cDNAs somewhat differ in size because 5'-RACE products lack the 3'-terminal part of *SraC* but include the 5' adaptor RNA oligonucleotide portion. The size marker indicated at the right margin was the same as used in panel A, lane M. (c) Sequence of the *sraC* locus of *E. coli* (*sraC* gene on the plus strand) in the intergenic region between *pphA* and ORF *b1839* (both on the minus strand). Coding sequences of the adjoining genes and the *sraC* gene are capitalized and in bold face. Horizontal lines between plus and minus strand sequences delineate the transcription terminator of *sraC*. Labeled vertical arrows denote the 5' ends deduced from sequencing of the four 5'-RACE products (5a–5d). Putative -10 and -35 sequence elements of the *sraC* promoter are boxed.



below). However, we do not recommend the use of random oligonucleotides (e.g., hexamers), since this usually results in a lower yield of specific 5'-RACE products.

Required Reagents

- STOP mix: 95% ethanol, 5% phenol (v/v)
- TRIzol reagent (Life Technologies Invitrogen, #15596018)
- Isopropyl alcohol
- RNase-free DNase I (Roche Diagnostics; 20 U μl^{-1} , #04716728001)
- 10 x DNase I digestion buffer: 100 mM sodium acetate (pH 5.2), 5 mM magnesium sulfate
- RNase inhibitor SUPERase-In (Ambion; 20 U μl^{-1} , #AM2696)
- P:C:I (water-saturated phenol/chloroform/isoamylalcohol mixture, 25 : 24 : 1, v/v)
- Ethanol containing 0.3 M sodium acetate (pH 5.7)
- TAP (EPICENTRE Biotechnologies; 10 U μl^{-1} , #T19050); the enzyme is provided with 10 x TAP digestion buffer: 500 mM sodium acetate (pH 6.0), 10 mM EDTA, 1% β -mercaptoethanol, 0.1% Triton X-100
- Dimethyl sulfoxide (DMSO)
- T4 RNA ligase (New England Biolabs (NEB); 10 U μl^{-1} , #M0204S); the enzyme is provided with 10 x RNA ligation buffer: 500 mM Tris-HCl (pH 7.9), 100 mM magnesium chloride, 10 mM dithiothreitol (DTT) and 1 mM ATP
- SuperScript III Reverse Transcriptase (Life Technologies Invitrogen; 200 U μl^{-1} , #18080044); supplied with 5 x cDNA synthesis buffer (250 mM Tris-acetate, pH 8.4, 375 mM potassium acetate, 15 mM magnesium acetate, stabilizer) and 0.1 M DTT
- 10 mM dNTP mix (10 mM each dATP, dGTP, dCTP and dTTP at neutral pH)
- RNase H (NEB; 5 U μl^{-1} , #M0297S)
- *Taq* DNA polymerase (NEB; 5 U μl^{-1} , #M0273) provided with 10 x Standard *Taq* Reaction Buffer (100 mM Tris-HCl, 500 mM KCl, 15 mM MgCl_2 , pH 8.3 at 25°C)
- Small-size DNA marker (e.g., pUC19 DNA/MspI marker from Thermo Scientific Fermentas)
- TOPO TA cloning kit: PCR2.1-TOPO vector cloning kit with competent TOP10F' cells (Life Technologies Invitrogen, #K450001)
- 10 x PCR buffer (100 mM Tris-HCl, 500 mM potassium chloride, pH 8.3), 25 mM magnesium chloride solution
- 5' RNA adapter oligonucleotide A3 (5'-AUAUGCGCGAAUCCUGUAGAACGAACACUAGAAGAAA-3') – any other RNA oligonucleotide may be used as long as the 5' adapter-specific DNA oligonucleotide for subsequent PCR amplification is designed accordingly; note, however, that for efficient ligation with T4 RNA ligase, the 3' end of the RNA oligonucleotide should be purine-rich
- Gene-specific DNA oligonucleotide jB-101 for RT (5'-CAGCTGATGACCACCA-3')
- Gene-specific DNA oligonucleotide jB-101-B for PCR amplification (5'-TGATGACCACCACGCTT-3')

- 5' Adapter-specific DNA oligonucleotide B6 for PCR amplification (5'-GCGCGAATTCCTGTAGA-3')
- DNA oligonucleotide Seq-UNI-61 for colony PCR (5'-ACGACGTTGTAAAACGA CGG-3')
- DNA oligonucleotide Seq-REV for colony PCR (5'-TTCACACAGGAAACAGCTA TGAC-3')
- 6 x agarose gel loading buffer: 10 mM Tris-HCl, pH 7.6, 0.03% (w/v) each bromophenol blue and xylene cyanol, 60% (v/v) glycerol, 60 mM EDTA
- DNA oligonucleotide Seq-REV-50 for sequencing (5'-CACAGGAAACAGCTAT GAC-3')
- Phase Lock Gel (PLG) tubes, PLG Heavy, 0.5 ml and 2 ml (5 Prime Eppendorf, #2302830) – these tubes are recommended to facilitate phase separation, but ordinary Eppendorf tubes may alternatively be used as well; keep in mind that PLG tubes must not be vortexed – shake samples vigorously by hand instead
- Gel purification kit (e.g., QIAEX II; Qiagen, #20021)
- 1 x TAE buffer: 40 mM Tris, 40 mM acetic acid, 1 mM EDTA (pH 8.0)
- Low melting agarose (e.g., Biozym Sieve GP Agarose, #850080)

Step 1: Growth of Bacterial Cultures

- 1) Dilute a fresh overnight culture of *E. coli* K12 cells 1/100 into 25 ml Luria-Broth (LB) medium in a 100 ml flask and keep shaking at 37 °C and 200 rpm until cells have reached the desired OD₆₀₀ value.

Step 2: Harvest of Bacteria

- 1) Harvesting bacterial cultures as described below usually yields 20–30 µg of total RNA.
 - a. *For RNA preparation from exponential phase cells* (OD₆₀₀ = 0.3). Transfer 15 ml of culture into a 50 ml plastic tube. Add 3 ml (20% v/v) of STOP mix. Mix thoroughly and immediately freeze in liquid nitrogen. Thaw frozen suspension on ice with occasional mixing. Once bacteria are completely thawed, pellet them by centrifugation for 10 min at 3500 g and 4 °C. After discarding the bulk of the supernatant, remove any residual traces of supernatant.
 - b. *For RNA preparation from stationary phase cells* (OD₆₀₀ = 2). Remove 2 ml of culture into a 2 ml Eppendorf tube. Immediately collect cells by centrifugation in a table-top microcentrifuge for 1 min at maximum speed. After discarding the bulk of the supernatant, remove any residual traces of supernatant.

Step 3: RNA Isolation

- 1) Immediately resuspend the pellet in 1 ml of TRIzol reagent by pipetting up and down. Incubate at room temperature for 5 min. Transfer to a 2 ml PLG tube; it is important to collect the gel at the bottom of the PLG tube by centrifugation

in a table-top microcentrifuge for 30 s at maximum speed immediately before addition of 200 μ l chloroform, followed by vigorous shaking for 30 s and incubation for 2 min at room temperature. Separate organic and aqueous phase by centrifugation in a microcentrifuge at 10 000 g or higher for 15 min at room temperature (it is important not to centrifuge PLG tubes at low temperature since this will interfere with proper placement of the phase lock gel between the two phases). Transfer 500 μ l of the aqueous (upper) phase to a fresh Eppendorf tube. Add 400 μ l of isopropyl alcohol, mix, and incubate the sample at room temperature for 10 min. Centrifuge ($\sim$ 12 000 g, 10 min, 4 °C) and remove the supernatant. Wash the RNA pellet once by adding 500 μ l of ice-cold 75% ethanol. Pellet RNA by centrifugation ($\sim$ 7500 g, 5 min, 4 °C). Remove supernatant, and air-dry the pellet at room temperature. Redissolve the RNA pellet in 133 μ l H₂O. Incubate for 5 min at 65 °C with occasional vortexing. Centrifuge briefly to collect the RNA solution at the bottom of the tube.

Step 4: Removal of DNA

- 1) Add to the tube:
 - a. 15 μ l of 10 x DNase I digestion buffer
 - b. 0.5 μ l RNase inhibitor
 - c. 2 μ l DNase I
- 2) Mix and incubate for 20 min at 37 °C. Transfer to a 0.5 ml PLG tube and add 150 μ l of P:C:I. Shake vigorously for 30 s and separate phases by centrifugation in a table-top centrifuge at maximum speed for 15 min at room temperature. Transfer aqueous phase (150 μ l) to a fresh Eppendorf tube and add 450 μ l of ethanol containing 0.3 M sodium acetate (pH 5.7). Precipitate for 30 min on ice and pellet RNA by centrifugation (at least 12 000 g, 30 min, 4 °C). Wash the pellet once with 75% ethanol (see Step 3) and finally redissolve the RNA in 50 μ l H₂O. Determine the RNA concentration by UV spectroscopy.

Step 5: Treatment with TAP

- 1) Bring 12 μ g of dissolved RNA to a volume of 87.5 μ l by adding the required volume of H₂O. Add 10 μ l of 10 x TAP digestion buffer and 0.5 μ l RNase inhibitor. Mix and split into two tubes (49 μ l each). Add to one tube 1 μ l of TAP enzyme (treated sample) and incubate both samples at 37 °C for 30 min. At the end of the incubation, add 500 pmol of RNA adapter oligonucleotide A3 and 100 μ l H₂O to each tube. Remove enzyme and buffer by P:C:I extraction followed by ethanol precipitation as in Step 4.

Step 6: 5' RNA Adapter Ligation

- 1) Dissolve RNA pellets in 12 μ l H₂O. Heat to 90 °C for 5 min. Immediately place on ice for 5 min. Spin briefly and add to each tube:
 - a. 2 μ l of 10 x RNA ligation buffer

- b. 2 μ l ATP
 - c. 2 μ l DMSO
 - d. 2 μ l T4 RNA ligase premixed 10 : 1 with RNase inhibitor
- 2) Incubate overnight at 17 °C. At the end of the incubation, add 2 pmol of the gene-specific DNA oligonucleotide for RT (here oligonucleotide jb-101) and H₂O to bring the total volume in each tube to 150 μ l. Remove enzyme and buffer by P:C:I extraction followed by ethanol precipitation as in Step 4.

Step 7: Reverse Transcription

- 1) Dissolve the pellet from Step 6, containing the ligated RNA and the gene-specific DNA oligonucleotide, in 20 μ l H₂O. Transfer 10 μ l to a 0.2 ml PCR tube and place in the PCR block. The following steps can be programmed in a single PCR program, but it is important to add the required reagents and enzymes as specified below.
- 2) Denature the sample for 5 min at 65 °C. Place on ice. Add 9 μ l of buffer mix (18 μ l premixed for two reactions: 8 μ l of 5 x cDNA synthesis buffer, 2 μ l of 10 mM dNTP mix, 5 μ l H₂O, 2 μ l 0.1 M DTT, 1 μ l RNase inhibitor). Mix and incubate for 10 min at room temperature. Add 1 μ l SuperScript III Reverse Transcriptase. *Note that SuperScript III transcriptase is active at a temperature of 50 °C. Usage of this higher reaction temperature increases specificity compared to SuperScript II, for example.*
- 3) Incubate for 15 min at 42 °C, then 15 min at 50 °C, 15 min at 55 °C, and 15 min at 60 °C. Terminate RT by heating for 5 min at 85 °C. Cool down to 37 °C and add 1 μ l RNase H. Incubate at 37 °C for 20 min. Then keep on ice or at –20 °C until PCR amplification.

Step 8: PCR Amplification of cDNA

- 1) Set up the following PCR reactions (50 μ l final volume per reaction). As template, use cDNAs from untreated and TAP-treated samples, and a 50% dilution of the ligated RNA that was taken for RT (the remaining 10 μ l from Step 7) as negative control.

5 μ l	10 x PCR buffer without Mg ²⁺
3 μ l	25 mM MgCl ₂
1 μ l	10 mM dNTP mix
2.5 μ l	Oligo jb-101-B (10 pmol μ l ⁻¹)
2.5 μ l	Oligo B6 (10 pmol μ l ⁻¹)
2 μ l	Template (cDNA or RNA)
0.25 μ l	<i>Taq</i> DNA polymerase
34 μ l	H ₂ O

- 2) Run PCR with the following cycles:
 - a. 95 °C/10 min

- b. 35 x (95 °C/40 s, 58 °C/40 s, 72 °C/40 s)
 - c. 72 °C/7 min
- 3) Add 10 µl of 6 x agarose gel loading buffer to each reaction and run 20 µl on a 3% low melting agarose gel in 1 x TAE buffer, along with an appropriate DNA size marker.

Step 9: Cloning of Amplified cDNAs and Determination of 5' Ends

- 1) Excise bands of interest from the agarose gel and purify with a gel extraction kit. Consult the kit manual for lower cutoff size of DNA fragments. This is because sRNAs will frequently give rise to RACE products of 100 bp or less, which most column-based kits fail to purify efficiently. We have routinely used the QIAEX II Gel Extraction Kit (Qiagen), which is based on DNA binding to glass milk. As a slight alteration of Qiagen's protocol, we use 7 µl of glass milk per gel piece (instead of the recommended volume of ≥ 10 µl) and elute the purified DNA in only 15 µl Tris buffer (10 mM, pH 8). It is important to ensure that the eluate is free of glass milk since the latter may interfere with subsequent ligation into a cloning vector.
- 2) For a TOPO TA cloning kit standard reaction, 4 µl of the eluted DNA is used. Approximately 100 clones are usually obtained when using the pCR2.1-TOPO vector and chemically competent *E. coli* TOP10F' cells, and more than 95% of the clones will contain an insert. Keep in mind that the cloned fragments are present in very high copy numbers (several hundred per cell). Hence, we do not advise use of blue/white screening with X-gal to select positive clones, since this implies transcription of the cloned sRNA fragment, which in some cases may be lethal to the bacteria at this copy number. Instead, use colony PCR to identify inserts; the generated PCR products can then be sequenced, thereby obviating isolation of the plasmid. A 25 µl colony PCR reaction is usually sufficient for both gel visualization and subsequent sequencing. We recommend that six colonies from each ligation be analyzed.
- 3) Set up the following PCR reaction mixture (150 µl final volume):

15 µl	10 x PCR buffer without Mg ²⁺
9 µl	25 mM MgCl ₂
3 µl	10 mM dNTP mix
3 µl	Oligonucleotide Seq-UNI-61 (10 pmol µl ⁻¹ ; for ligation into plasmid pCR2.1)
3 µl	Oligonucleotide Seq-REV (10 pmol µl ⁻¹ ; for ligation into plasmid pCR2.1)
1 µl	<i>Taq</i> DNA polymerase
116 µl	H ₂ O

- 4) Split into six PCR tubes (6 × 25 µl) and place them in a PCR block set to 4 °C. Slightly touch a colony with a pipette tip (10–200 µl tip) and then swirl the tip at the bottom of the PCR tube. Avoid transferring too much of the colony

since this will inhibit amplification or result in high background. Once all PCR tubes are provided with “template,” start the PCR with the following cycles:

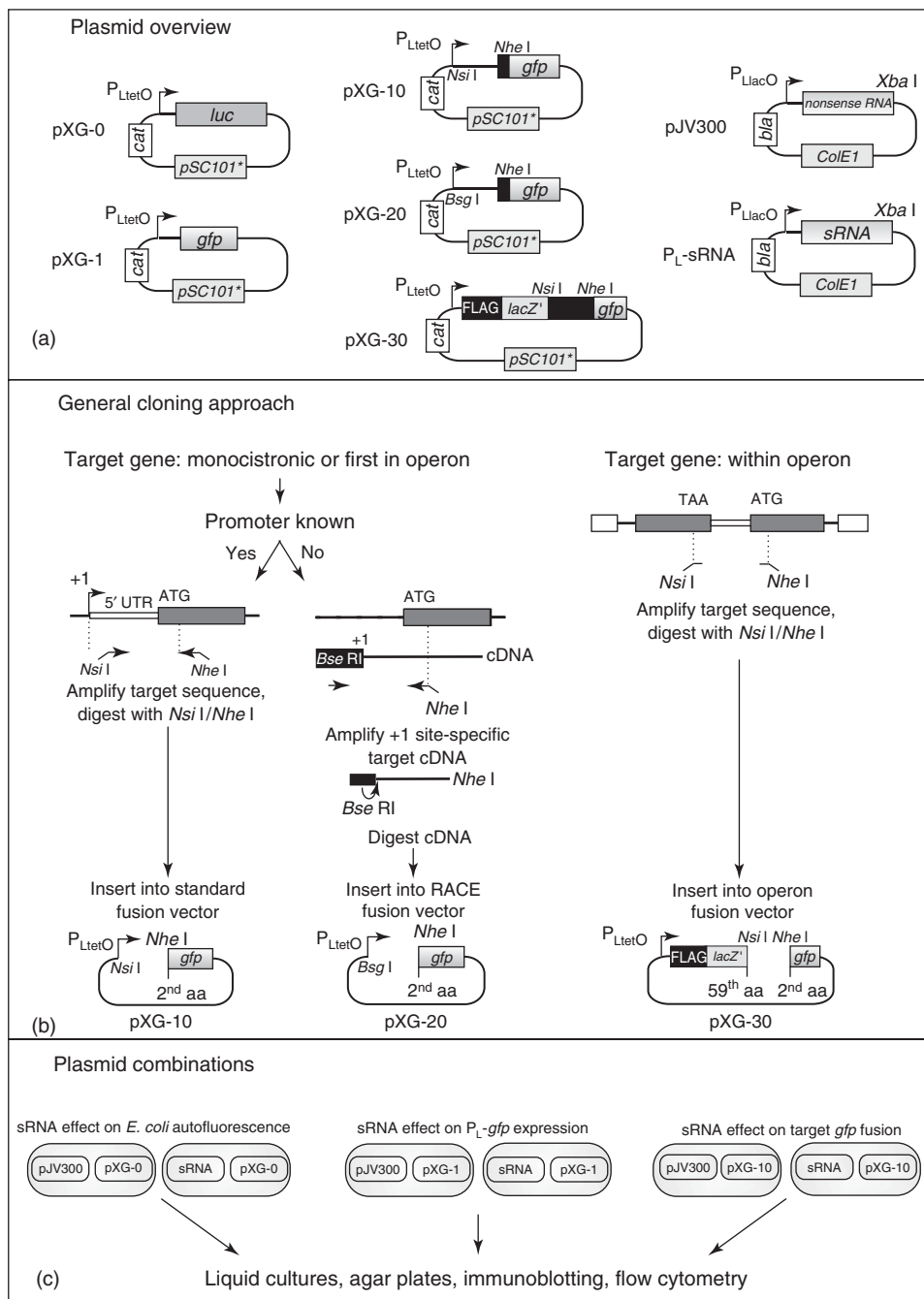
- a. 95 °C/10 min
 - b. 30 × (95 °C/40 s, 65 °C/40 s, 72 °C/40 s)
 - c. 72 °C/7 min
- 5) Add 5 µl of 6 x agarose gel loading buffer to each reaction, and analyze for inserts by running 4 µl on a 2% agarose gel. The size of colony PCR products from religated vector, that is, without insert, is approximately 220 bp. Hence, positive clones should give rise to a size of 220 bp plus the size of the DNA fragment that was cloned. Purify and sequence PCR products according to the standard protocol of your sequencing service. For sequencing, use oligo Seq-REV-50, which has a lower annealing temperature than oligonucleotide Seq-REV. It is highly recommended that the sequence of at least four inserts per excised PCR band be determined since both primary and processed 5' ends may vary by a few nucleotides (Figure 33.4c). For primary transcripts (band intensity increased after TAP treatment), the most upstream 5' nucleotide is regarded as the transcription initiation site.

Protocol 3: Target Validation by Application of a Green Fluorescent Protein (GFP)-Based Reporter System

Validation of posttranscriptional regulation in general and elucidation of the underlying sRNA–mRNA interactions are crucial steps in an sRNA project. Our laboratory has developed a modular plasmid system that utilizes the green fluorescent protein (GFP) as a reporter to study posttranscriptional control. The method is based on constitutive coexpression of the sRNA and the target fused to GFP. Putative target sequences are translationally fused to *gfp+* and expressed from a low-copy vector (pSC101* replicon, 3–4 copies per cell; carries a chloramphenicol resistance gene [*catI*] under the control of a constitutive $P_{\text{LtetO-1}}$ promoter. The sRNA is usually expressed from a high-copy vector (ColE1 replicon, ~70 copies per cell; carries an ampicillin resistance gene [*bla*]) under the control of a constitutive $P_{\text{LlacO-1}}$ promoter (Figure 33.5a). *Note that the overexpression of an sRNA might be toxic. In such cases, one should lower the copy number of the sRNA plasmid or use inducible expression of the sRNA. As for the latter, we use the E. coli Top10 F' strain with an episomal lacI^q allele for cloning, and induce sRNA expression from a $P_{\text{LlacO-1}}$ promoter with IPTG. Alternatively, the L-arabinose-inducible P_{BAD} promoter may be utilized.*

Required Reagents

- PlacoB: 5'-CGCACTGACCGAATTCATTAA-3' (binds to the *luc* gene in sense orientation)
- PlacoD: 5'-GTGCTCAGTATCTTGTATCCG-3' (binds antisense to the promoter starting with the –1 site)



- DNA polymerase with proofreading activity (e.g., Phusion Polymerase, Thermo Scientific Finnzymes, #F-530S)
- *Dpn* I (NEB; 20 U μl^{-1} , #R0176S)
- Gel Extraction Kit (e.g., QIAquick, Qiagen)
- PCR Purification Kit (e.g., QIAquick, Qiagen, #28104)
- *Xba* I (NEB; 20 U μl^{-1} , #R0145S)
- SAP = shrimp alkaline phosphatase (Thermo Scientific Fermentas; 1 U μl^{-1} , #EF0511)
- T4 DNA ligase (NEB; 400 U μl^{-1} , #M0202S)
- *E. coli* TOP10F' chemical competent cells (Life Technologies Invitrogen, #C303003)
- Primer pZE-A: 5'-GTGCCACCTGACGTCTAAGA-3' (binds upstream of the promoter $P_{\text{LacO-1}}$)
- Primer pZE-XbaI: 5'-TCGTTTTATTTGATGCCTCTAGA-3' (binds in the region of the *Xba* I site (underlined) in pZE12-*luc*)
- *Nhe* I-HF™ (NEB; 20 U μl^{-1} , #R3131S)
- *Nsi* I (NEB 10 U μl^{-1} , #R0127S)
- Primer pZE-CAT: 5'-TGGGATATATCAACGGTGGT-3' (binds in the *cat* gene)
- Primer JVO-155: 5'-CCGTATGTAGCATCACCTTC-3' (binds to *gfp* in antisense orientation)
- *Bsg* I (NEB; 5 U μl^{-1} , #R0559S)
- RNA linker A4 : 5'-GACGAGCACGAGGACACUGACAUGGAGGAGGGAGUAGAAA-3' (*Bse* RI site underlined)
- Primer JVO-367 (5'-ACTGACATGGAGGAGGGA-3')

Figure 33.5 Modular plasmid system for the validation of sRNA targets. (a) Plasmid overview. pXG-0 and pXG-1: control plasmids (see panel c and Protocol 3.3). Putative sRNA target sequences are cloned as translational fusions to *gfp* into one of the low-copy vector (pSC101* replicon) variants (pXG-10, 20, or 30) that confer chloramphenicol resistance (*cat*). The sRNA is cloned into a high-copy vector (ColE1 replicon) conferring ampicillin resistance (*bla*), such that transcription will precisely start at the native +1 site of the sRNA; pJV300, expressing a ~50 nt long RNA derived from the *rmB* terminator, controls for sRNA-independent, vector-related effects; sRNA and translational *gfp* fusion are expressed under control of the constitutive $P_{\text{LacO-1}}$ or $P_{\text{LtetO-1}}$ promoter, respectively. (b) General cloning approach. Depending on the chromosomal context of a putative target gene and whether the native promoter is known, different cloning strategies are pursued; for details, see

Protocol 3.2. (c) Plasmid combinations. On co-transformation into a cell, the effect of a given sRNA on a target fusion can be determined by monitoring GFP fluorescence of liquid cultures (standard laboratory flasks or microtiter plates) or colonies grown on agar plates, by Western blot analysis or by flow cytometry. Fusions that exhibit higher fluorescence levels in the presence of an sRNA as compared to the mock control would be considered activated, whereas reduced fluorescence would indicate target repression. Several controls are required to ensure proper interpretation of the results. As a control plasmid for sRNA expression, the vector pJV300 should be used (panel a); pXG-0 and pXG-1 probe the sRNA effect on cell auto-fluorescence and constitutive *gfp* expression, respectively; finally, pXG-10, -20 or -30 reveal if the sRNA is able to regulate the target-*gfp* fusion construct. (Figure adapted from Ref. [90].)

Protocol 3.1: Construction of the sRNA Expression Plasmid

The sRNA gene of interest is cloned under the control of the $P_{\text{LlacO-1}}$ promoter. Knowledge of the native transcriptional start site and of the approximate location of the transcription terminator is required. The transcriptional start is determined as described in Protocol 2. The prediction of the transcription terminator can be aided by secondary structure prediction algorithms such as `mfold` (<http://mfold.rna.albany.edu/?q=mfold>) [132].

Step 1: Preparation of Vector DNA

The sRNA overexpression vectors are based on plasmid pZE12-*luc* [91].

- 1) Prepare plasmid DNA of pZE12-*luc*.
- 2) Amplify by PCR using primers PlacoB and PlacoD.
Note: Use a polymerase with proofreading activity that generates blunt ends (e.g., Phusion polymerase, Thermo Scientific Finnzymes). One should get a product of ~3.9 kb.
- 3) Load the PCR reaction on a 0.8% agarose gel.
- 4) Excise and extract the PCR product from the gel using a commercial kit (e.g., QIAquick).
- 5) Perform a *Dpn* I digest by adding 1 μl enzyme ($20 \text{ U } \mu\text{l}^{-1}$) to the purified product and incubate for 3 h at 37°C . *Dpn* I recognizes sites in methylated DNA and will destroy the template plasmid.
- 6) Purify the reaction using a PCR purification kit (e.g., QIAquick).
- 7) Digest 2–3 μg of DNA using 20 U *Xba* I. Incubate for 3 h at 37°C .
- 8) Add 1 U of SAP directly to the reaction and incubate for 1 h at 37°C . SAP creates 5' hydroxyl termini, which will avoid recircularization of the vector.
- 9) Load the digested PCR-amplified vector DNA onto a 1% agarose gel.
Note: Two DNA fragments of ~2.2 and ~1.7 kb should be generated by Xba I digestion. The ~2.2 kb fragment carries the $P_{\text{LlacO-1}}$ promoter (from position -1), a bla resistance gene, a ColE1 replicon and a strong rrnB terminator followed by the sticky end created by Xba I digestion.
- 10) Excise the ~2.2 kb band from the gel and perform a gel extraction using a commercial kit.

Step 2: Preparation of sRNA Insert

- 1) Design primers to amplify the sRNA gene of interest. The sense primer should start with the +1 site of the sRNA and be 5' phosphorylated for blunt-end ligation. The antisense primer should pair ~40 nt downstream of the transcription terminator and carry an *Xba* I restriction site (5'-TCTAGA). In addition, the primer should be designed with a 5'-TTTTT extension (5'-TTTTTCTAGA. . .) to facilitate digestion of the PCR product.
- 2) Amplify by standard PCR using genomic DNA as the template.
- 3) Analyze 5 μl of the PCR reaction on a 2–3% agarose gel.
- 4) Purify the PCR reaction using a commercial kit (e.g., QIAquick).

- 5) Digest 1–2 µg of the DNA with 10 units of *Xba* I for 3 h at 37 °C.
- 6) Load the digested DNA onto a 2–3% agarose gel, excise the DNA band from the gel, and extract it with a commercial kit (e.g., QIAquick).

Step 3: Ligation and Transformation

- 1) Prepare a ligation reaction containing ~20 ng of digested and dephosphorylated pZE12-*luc* (from Step 1 “Preparation of vector DNA”), ~30 ng digested insert and 2 U T4 DNA ligase in a total volume of 20 µl; incubate for 1 h at room temperature. Run a control ligation (insert volume replaced with H₂O) to check for religation of the vector.
- 2) Transform *E. coli* TOP10F' chemical competent cells with 10 µl of the ligation reaction. Typically, ~20–200 colonies will appear on a selective plate, while the number of religation events is negligible.
- 3) Control the success of the cloning by PCR on the transformants using the primers pZE-A/pZE-Xba I. The PCR product size for successful insert integration depends on the insert size; for unwanted religation of pZE12, a 166 bp PCR band will appear (a ~1.8 kB fragment is generated on the original circular pZE12-*luc*). For colony PCR, see Protocol 2, Step 9.
- 4) Sequence your insert using primer pZE-A or pZE-XbaI to confirm the absence of mutations.

Protocol 3.2: Cloning of a Translational GFP Fusion

Before cloning of a translational target-GFP fusion, one should clarify the following questions:

- Is the transcriptional start site of the gene of interest known?
- Is the gene of interest located in an operon?
- Where is the anticipated location of the putative sRNA binding site?

If the transcriptional start site is known and the gene of interest is either monocistronic or the first gene of an operon, one should use the pXG-10 vector. If the transcriptional start site is unknown, pXG-20 is the vector of choice, because it is tailored to the cloning of 5'-RACE products. In case the putative target gene lies inside an operon, one should work with the pXG-30 vector (for a schematic representation of the different cloning strategies, see Figure 33.5b). The putative sRNA binding site should be based at least on strong evidence from *in silico* predictions. Here, programs such as *RNAhybrid* may be helpful (<http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/>) [129]. We routinely clone the 5' UTR plus at least the first 10–15 amino acids of the target ORF.

Protocol 3.2.1: Cloning of a Standard Translational GFP Fusion Using pXG-10

Step 1: Preparation of Vector pXG-10

- 1) Prepare plasmid DNA of pXG-10.
Note: Since pXG-10 is a low-copy plasmid a midi-prep should be performed in order to obtain a sufficient amount of plasmid DNA for subsequent cloning steps.

- 2) Digest 2–4 µg pXG-10 with 20 units *Nhe* I-HF for 2 h at 37 °C.
Note: Nhe I-HF has a reduced star activity compared to standard *Nhe* I.
- 3) Purify the digested plasmid using a commercial PCR purification kit (e.g., QIAquick).
- 4) Digest the DNA with 10 units *Nsi* I for 3 h at 37 °C.
- 5) Following the restriction digest, add 20 units SAP and continue the incubation for 1–2 h at 37 °C.
- 6) Following digestion, run the reaction on a 1% agarose gel, excise the ~4120 bp DNA band corresponding to the vector fragment and extract the DNA using a commercial kit.

Step 2: Preparation of *gfp* Fusion Inserts

- 1) Design primer pairs to amplify the DNA fragment to be fused to *gfp*. The sense primer should start at the gene of interest's transcriptional start site (+1) and carry an *Nsi* I restriction site (5'-ATGCAT) at its 5' end. The antisense primer should bind in the coding region of the gene of interest and carry an *Nhe* I restriction site (5'-GCTAGC).

Notes:

- The *Nhe* I restriction site of the antisense primer must be in frame with the target gene.
 - The amplified sequence should not contain an internal *Nsi* I or *Nhe* I site.
 - The primers should be extended at their 5' end by a stretch of five T residues for efficient restriction enzyme cleavage of the PCR product.
 - The insert ought to cover the complete 5' UTR and about 30–45 bp of the coding region (encoding amino acids 10–15). The translational fusion site should be adapted such that the putative sRNA binding site is entirely represented and authentically folded in the fusion construct.
- 2) Amplify the insert by standard PCR, check an aliquot on a 2–3% agarose gel, and purify the PCR reaction using a commercial kit. We usually obtain a total amount of 1.5–3 µg DNA.
 - 3) Digest 1–2 µg of the prepared DNA with *Nhe* I-HF and *Nsi* I as for preparation of pXG-10.
 - 4) Following digestion, run the reaction on a 2–3% agarose gel, excise the DNA band from the gel, and purify the DNA using a commercial kit.

Step 3: Ligation and Transformation

- 1) Ligate ~12 ng of digested, dephosphorylated pXG-10 to ~30 ng digested insert using 2 U T4 DNA ligase in a total volume of 20 µl for 1 h at room temperature. Run a control ligation (insert volume replaced with H₂O) to check for religation of the vector.
- 2) Transform *E. coli* TOP10F' chemically competent cells with 10 µl of ligation reaction. Typically, ~20–200 colonies should appear on a selective plate, while the number of religation events should be negligible. Visualization of

the colony fluorescence under UV light can be used to screen for positive colonies; pXG-10 itself gives rise to a low fluorescence level (pXG-10 contains a *lacZ186:gfp* fusion).

- 3) Control the cloning success by PCR on colonies using the primers pZE-CAT and JVO-155. PCR product size for a successful insertion depends on your insert size; for unwanted religation of pXG-10 in the absence of insert, a ~410 bp PCR fragment is obtained; on the original circular pXG-10, a ~1 kbp fragment is generated. For colony PCR, see Protocol 2, Step 9.
- 4) Sequence the insert using primer pZE-CAT or JVO-155 to confirm the absence of mutations.

Protocol 3.2.2: Cloning of a Standard Translational GFP Fusion Using pXG-20

The vector pXG-20, suitable for the cloning of RACE (Protocol 2) products, was developed to construct fusions without knowing the transcriptional start site of the target gene (we highly recommend that Protocol 2 be read before proceeding to the cDNA cloning protocol below). A direct cloning of RACE products into the pXG-20 vector is possible when using the special A4 RNA linker (in Protocol 2, Step 6) that contains a *Bse* RI restriction site. *Bse* RI has the peculiarity of cleaving the sense strand at any phosphodiester linkage located 10 nt downstream of its recognition sequence. The A4 RNA linker was designed such that the cleavage site matches the RNA linker ligation site, and produces a 5'-AA overhang. The pXG-20 vector carries a *Bsg* I and an *Nhe* I restriction site. *Bsg* I cleaves the sense strand at any phosphodiester linkage located 14 nt upstream of its recognition site. The *Bsg* I site was designed to produce a 3'-TT overhang that represents the -1 and -2 position of the promoter and is compatible with the extremity generated by *Bse* RI. For direct cloning, the cDNA should be amplified using a sense primer that is specific to the RNA linker and an antisense primer that binds in the coding sequence of the target of interest and contains an *Nhe* I restriction site.

Step 1: Preparation of Vector pXG-20

- 1) Proceed as for the preparation of pXG-10 described in Protocol 3.2.1, Step 1, (1)–(3).
- 2) Digest the product with 20 U *Bsg* I for 2 h at 37 °C.
- 3) Add 20 U SAP and incubate for another 1–2 h at 37 °C.
- 4) Apply the reaction to a 1% agarose gel, excise the ~4110 bp DNA band corresponding to the vector fragment (concomitantly, a ~988 bp fragment is generated) and extract the DNA using a commercial kit.

Step 2: Preparation of Insert cDNA

- 1) Prepare cDNA using the 5'-RACE Protocol 2 with RNA linker A4.
- 2) Amplify the fusion insert by standard PCR with 1 µl of the prepared cDNA using the sense primer JVO-367 identical in sequence to an internal portion of RNA linker A4 that contains the *Bse* RI restriction site, and an antisense primer specific to the target gene that carries a 5'-proximal *Nhe* I restriction

site. Perform the PCR reaction on the TAP-treated and untreated samples. Furthermore, genomic DNA should be used as a control template.

- 3) Analyze the reactions on a 3% agarose gel. The product that is enriched in the TAP-treated sample likely represents the primary transcript. Excise this band and extract the DNA using a commercial kit. Digest the purified PCR product with 10 U each *Nhe* I-HF and *Bse* RI for 2 h at 37°C in NEB Buffer 4 + BSA (bovine serum albumin).

Note: If not enough DNA is extracted for digestion, a second round of PCR can be performed using the previous PCR product as a template.

- 4) Following digestion, run the reaction on a 3% agarose gel, excise the band and extract the DNA using a commercial kit.

Step 3: Ligation and Transformation

- 1) Set up a ligation reaction as described in Protocol 3.2.1, Step 3 (1)–(2), for pXG-10.
- 2) Control the cloning success by PCR on colonies using the primers pZE-CAT and JVO-155. PCR product size for a successful insertion depends on your insert size; for unwanted religation of pXG-20 in the absence of insert, a ~430 bp PCR fragment is obtained; a ~1.4 kbp fragment is generated on the original circular pXG-20. For colony PCR, see Protocol 2, Step 9. As pXG-10, pXG-20 gives rise to a low fluorescence phenotype; thus, a visualization of the colony fluorescence under UV light can be used to screen for positive colonies.
- 3) Sequence your insert using primer pZE-CAT or JVO-0155 to confirm the absence of mutations.

Protocol 3.2.3: Cloning of a Standard Translational GFP Fusion Using pXG-30

In some cases, sRNAs are able to target untranslated regions of polycistronic mRNAs to specifically regulate the expression of one cistron. The operon fusion vector pXG-30 enables the cloning of putative target sequences by mimicking a dicistronic operon. The application of a dual reporter system enables to monitor the expression of both genes simultaneously. The first cistron is a 3 × FLAG-tagged artificial open reading frame derived from the coding region of the *E. coli lacZ* gene. The second cistron is *gfp*. A cloned intraoperon target site is then positioned between the upstream *FlacZ'* cistron and the downstream *gfp* cistron. The expression of *FlacZ'* and *gfp* cistrons can be studied simultaneously by immunoblotting using anti-FLAG and anti-GFP antibodies.

Step 1: Preparation of Vector pXG-30

- 1) Proceed as for the preparation of pXG-10 described in Protocol 3.2.1, Step 1 (1)–(5).
- 2) Run the reaction on a 1% agarose gel, excise the ~4400 bp DNA band corresponding to the vector fragment (concomitantly, a ~320 bp band is generated) and extract the DNA using a commercial kit.

Step 2: Preparation of gfp Fusion Inserts

- 1) Design primers to amplify the putative intercistronic target region. The sense primer should match the upstream gene (the last 10–50 amino acid residues should be included) and should carry an *Nsi* I restriction site, which needs to be in frame with the upstream gene. The antisense primer binds to the N-terminal part of the downstream gene of the locus and should carry an *Nhe* I restriction site. The *Nhe* I site of the antisense primer must be in frame with the downstream gene (see also Figure 33.5b).
- 2) Proceed as described in Protocol 3.2.1, Step 2, for pXG-10 insert preparation.

Step 3: Ligation and Transformation

- 1) Set up a ligation reaction as described in Protocol 3.2.1, Step 3 (1)–(2), for pXG-10.
- 2) Control the cloning success by PCR on colonies using the primers pZE-CAT and JVO-155. PCR product size for a successful insertion depends on your insert size; for unwanted religation of pXG-30 in the absence of insert, a ~716 bp PCR fragment is obtained; on the original circular pXG-30, a ~1 kbp fragment is generated. For colony PCR, see Protocol 2, Step 9. Unlike pXG-10 and pXG-20, the pXG-30 vector gives rise to a high fluorescence phenotype; thus, colony fluorescence visualized under UV light cannot serve as a measure for successful ligation.
- 3) Sequence your insert using primer pZE-CAT or JVO-155 to confirm the absence of mutations.

Protocol 3.3: Read Out Approaches

The plasmid carrying the target-*gfp* fusion and the sRNA overexpression plasmid are sequentially introduced into the desired bacterial background by transformation. Construction of control strains harboring (i) pXG-0 (expressing the *luc* gene and serving as control for non-specific sRNA effects on cell autofluorescence), (ii) pXG-1 (expressing full-length *gfp*, used to control for non-specific sRNA effects on GFP expression), and (iii) pJV300 (expressing a ~50 nt long RNA derived from the *rrnB* terminator and used as a control plasmid for sRNA expression) is of primary importance (Figure 33.5c). In particular, the strain combining pXG-1 and the sRNA overexpression plasmid permits one to rule out that the sRNA is regulating the *gfp* mRNA independent of the target sequence under study.

In order to rapidly assess target regulation, we use plate fluorescence imaging. To this end, the strain harboring the target-*gfp* fusion combined with pJV300 and the strain harboring the target-*gfp* fusion combined with the sRNA overexpression plasmid are restreaked next to each other. Pictures are obtained with a fluorescence plate reader equipped with a GPF filter using exposure times between 1 and 3 s (depending on the system). An additional picture should be taken in the visible light mode, in order to compare colony morphology of the different strains. However, besides being quick and illustrative, plate imaging can only provide a qualitative result, since growth on plates does not reflect a defined growth stage.

In case the fluorescence level is too low to be observed with a plate reader, Western blotting is a useful method to increase sensitivity and to obtain a semi-quantitative result. Here, monoclonal antibodies against GFP can be used to test for sRNA-mediated regulation (for Western blotting, see Protocol 5, Step 6). A fast and very reliable way to monitor GFP fluorescence is flow cytometry, which has become routine in our laboratory and is detailed in Protocol 4 below.

Protocol 4: Flow Cytometry

Flow cytometry is a reliable and very sensitive method to measure GFP fluorescence. However, the fluorescence signal from different GFP translation fusions can vary considerably. Therefore, it is essential to include a negative control (a non-fluorescent cell clone) and a positive control, which is usually a clone with the most strongly expressed GFP fusion in the absence of sRNA expression (the brightest clone on plate when exposed to UV light). The ratio between the negative and positive controls must be the same when independent measurements are done under the same conditions. Three independent experiments should always be performed for statistical relevance. Once the raw data have been saved, cell populations can be displayed in several different formats using commercially available software tools (FCS Express, FlowJo). If such analysis tools are not available, we can recommend the use of the free software Cyflogic (www.cyflogic.com/) or WinMDI (<http://facs.scripps.edu/software.html>). For a comprehensive overview on flow cytometry and potential technical problems, we refer the reader to “Current Protocols in Cytometry,” ISSN 1934-9297 (<http://cda.currentprotocols.com/WileyCDA/CPTitle/isbn-0471161314.html>).

Required Reagents

- 10 x Phosphate-buffered saline (PBS) solution: 1.37 M sodium chloride; 27 mM potassium chloride; 43 mM disodium hydrogen phosphate; 14.7 mM potassium dihydrogen phosphate
- 4% paraformaldehyde (PFA) solution in 1 x PBS at 4 °C. For 100 ml, dissolve 4 g of PFA in 90 ml dH₂O in a flask. Add 8 µl of 10 N sodium hydroxide. Place the flask on a hotplate and warm the solution up to 65 °C (do not boil!). Add 10 ml of 10 x PBS. The pH of the solution should be ~7.4. After cooling to room temperature, filter the solution through a 0.2 µm filter to remove any debris. Keep at 4 °C; aliquots can also be kept at –20 °C but avoid refreezing. After thawing, if necessary, warm the solution to clear it. (*PFA is a carcinogen and should be handled with precaution: work under a fume hood wearing a lab coat, a facemask, and gloves*)
- Flow Cytometer equipped with a blue laser and an FITC filter (e.g., FACSCalibur, BD Biosciences)
- Flow Cytometer polystyrene or polypropylene tubes of correct size (according to the instruction guide of your flow cytometer)
- BD FACS Flow Sheath Fluid (BD Biosciences, #342003)

- BD FACS Clean Solution (BD Biosciences, #340345)
- BD FACS Rinse Solution (BD Biosciences, #340346)

Step 1: Sample Collection and Fixation of Cells

The optical density for cell harvest depends on the individual experimental design. It should be noted that the measured fluorescence is affected by the oxygen level and will therefore be lower in the stationary phase independent of the protein level. Nevertheless, we recommend conditions close to the physiological ones.

- 1) Collect the cells by centrifugation (7000 rpm for 2 min at 4 °C) when the culture has reached the appropriate optical density.
- 2) Carefully resuspend the cells in a fresh solution of 4% PFA at 4 °C. Use a volume that results in an optimal concentration of 5×10^6 cells/ml.
- 3) Leave the cell suspension on ice for 10 min.
- 4) Collect the cells by centrifugation (7000 rpm for 2 min at 4 °C).
- 5) Resuspend the cells in 1 volume of 1 x PBS at 4 °C.
- 6) After subsequent removal of fixative by washing with 1 volume of 1 x PBS, the samples may be stored for weeks at 4 °C in the dark.

Step 2: Setting up the Flow Cytometer

- 1) Control the level of the Flow Sheath Fluid and of the Clean Solution.
- 2) Turn on the computer, start the software for data acquisition and the flow cytometer.
- 3) Check the laser power.
- 4) Clean the sample system before and after usage. Refer to the instruction guide of your flow cytometer.
- 5) The Flow Sheath Fluid keeps the sample at the center of the flow. Since the LOW, MEDIUM, and HIGH flow parameters do not change the flow rate, but the density of the sample, it is recommended to work on LOW or MEDIUM to avoid measurement of doublets and aggregates.
- 6) Depending on the apparatus, the FSC (forward scatter, measures cell size) detector can detect narrow- or wide-angle FSC signals. Adjust to the narrow angle when working with bacteria.
- 7) (Optional) A control of the correct alignment of the flow cytometer optical instrument can be performed using commercially available flow check beads (10–0.5 µm YG beads, Polysciences). Refer to the instruction guide of your flow cytometer.

Step 3: Sample Measurement

Create templates for sample measurements and save it as a protocol. The protocol should be saved since it will be the standard for further measurements. We propose the following acquisition templates:

- a. A two-parameter dot plot displaying the FSC *versus* SSC (side scatter, proportional to cell complexity and granularity) allows setting up the correct voltage. Particle

counts are shown by dot density. Adjust the voltage of the FSC and SSC photomultiplier until the majority of cells materialize in the lower left quadrant. An example of a two-parameter dot plot showing a *Salmonella typhimurium* overnight culture is given in Figure 33.6a.

- b. A one-parameter histogram measuring the cell count on the y-axis and the fluorescence level (FL1-FITC) on the x-axis (for most cytometers the FL1 channel corresponds to the fluorescence emitted by FITC, GFP, or Cy2; however, this should be checked in the cytometer's manual). It gives the fluorescence level of individual cells as detected by the FL1 photomultiplier. A homogeneous sample should give a distinct peak, since the fluorescence emitted by each cell should be comparable (Figure 33.6b,c).
- c. A statistic table for the histogram plot containing at least the number of events and the geometric mean (average value of all measured fluorescence values derived from individual cells).
 1. Add a minimum of 0.5 ml cell suspension from Step 1, (6), to a flow cytometer tube. During measurement, it is important not to exceed 3000 events/s. If the samples are too concentrated, dilute with 1 x PBS.
 2. Gently vortex the sample.
 3. Start the Sheath Fluid Flow and collect data.
- d. Measure a minimum of 30 000 events. An electronic threshold can be used to limit the number of events that the flow cytometer processes. In the FSC mode, it eliminates events such as debris or cell doublets.
- e. Save the raw data.

Troubleshooting

- 1) If the fluorescence level is too low, increasing the voltage of the photomultiplier or the amplifier gain will increase the signal; however, it will increase the background noise as well.
- 2) If events, for example, attributable to cell debris, need to be eliminated, thresholds should be set. This is an electronic cutoff that will limit the number of counted events. However, thresholds should be set carefully, since events of interest might be lost as well.
- 3) If the analyzed population is heterogeneous, the events of interest can be gated by any combination of criteria you choose.

Protocol 5: *In vitro* Translation with the PURESYSTEM

Minimal *in vitro* systems are often useful to characterize regulatory mechanisms of sRNAs. The PURESYSTEM, a commercially available cell-free translation assay based on reconstituted 70S ribosomes, was helpful in monitoring both repression [11, 160] and activation [140] processes mediated by sRNAs. Here, we describe how to obtain *in vitro* transcribed RNA for *in vitro* translation, how to set up the reaction, and how to read out the information.

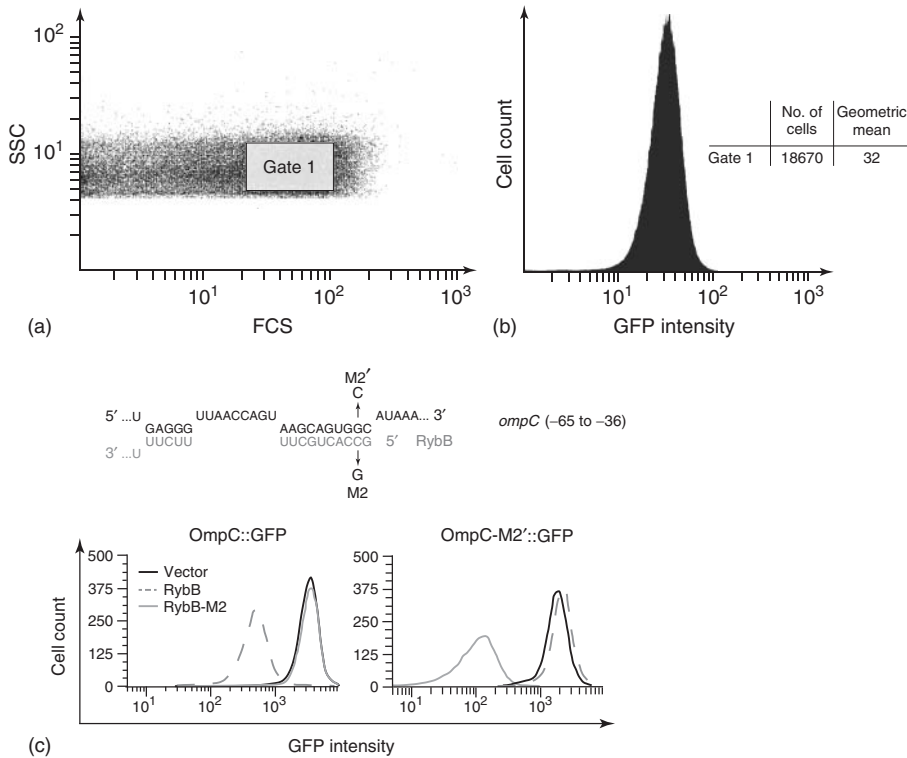


Figure 33.6 Analyzing target regulation by flow cytometry. (a) A scatter plot, showing the sideward scatter (SSC) on the y-axis and the forward scatter (FSC) on the x-axis, illustrates the size distribution of cells in a bacterial population. Here, the cell size distribution of a representative *Salmonella typhimurium* overnight culture harboring an OmpF::GFP fusion is shown. For fluorescence signal analysis, a region in the right corner of the population was chosen, which permits scoring of bacterial cells that are homogeneous in size (Gate 1). (b) Subsequently, the fluorescence intensity of the gated bacterial subpopulation was analyzed in a histogram plot. Cellular fluorescence is given in arbitrary units (GFP intensity). Different samples can be compared by the geometric mean of the population's fluorescence or by overlaying different histograms in one plot. (c) Top panel: Antisense complementarity of RybB hyperconserved residues (gray) to *ompC* mRNA (black) predicted by RNAhybrid

[129]. Numbers in parentheses denote mRNA nucleotide positions relative to the AUG start codon (A is +1). To confirm the physical interaction between the two RNAs, a single nucleotide mutation was introduced into the core interaction site (M2; C2G mutation in RybB yielding variant RybB-M2). A second compensatory mutation was introduced into the *ompC* target region (M2'; G42C mutation yielding variant *ompC*-M2') to analyze if restoration of base-pairing may rescue the loss of *ompC* repression by RybB-M2 RNA. Bottom panel: The mutations were introduced into the *rybB* and *ompC* genes on the plasmids of the modular *gfp*-reporter system (Figure 33.5, Protocol 3). Data was processed and plotted as described above. As expected for the base-pairing interaction proposed at the top of panel c, the M2 point mutation abrogates regulation of the *ompC* wild-type target. However, the inhibitory capacity of RybB-M2 is restored when acting on the *ompC*-M2' mutant target.

Required Reagents

- A sense primer containing the T7 promoter fused to the transcription start of the gene of interest (GOI): 5' gtttttttTAATACGACTCACTATAG::GOI. *Note:* T7 transcription starts with the G in bold
- An antisense primer corresponding to the 3' extremity of the mRNA
- DNA polymerase (0.5–1.0 U per reaction) and 5–10 x PCR buffer (e.g., *Taq* DNA polymerase with 10 x ThermoPol® buffer: 200 mM Tris-HCl, 100 mM ammonium sulfate, 100 mM potassium chloride, 20 mM magnesium sulfate, 1% Triton X-100, pH 8.8 at 25°C; NEB #M0267L)
- Gel extraction kit (e.g., QIAquick, Qiagen, #28704)
- T7 transcription kit (MEGAscript® T7 kit, Ambion, #AM1334)
- TURBO DNase I (2 U µl; MEGAscript® T7 kit, Ambion, #AM1334)
- Illustra MicroSpin™ G-50 columns (GE Healthcare, #27-5330-01)
- Phase lock gel (PGL) heavy 2.0 ml tubes (5 PRIME, #2302830)
- P:C:I (water-saturated phenol/chloroform/isoamylalcohol mixture, 25 : 24 : 1, v/v)
- Denaturing gel loading buffer for PAGE (e.g., Ambion, #AM8546G, contains 95% formamide, 18 mM EDTA and 0.025% each of SDS, xylene cyanol, and bromophenol blue)
- RNA size marker (e.g., RiboRuler™ low or high range RNA ladder, Thermo Scientific Fermentas)
- Denaturing polyacrylamide gel, 4–5%. *For details, see Sambrook and Russell, Molecular Cloning, A Laboratory Manual (Third Edition)*
- 10 x TBE buffer (0.9 M Tris base, 0.9 M boric acid, 20 mM EDTA)
- RNA staining solution (Stains-All, Sigma-Aldrich #E9379, or SYBR® Gold from Life Technologies Invitrogen, #S-11494)
- (Optional) RNA elution buffer (0.1 M sodium acetate pH 5.7, 0.1% SDS, 10 mM EDTA pH 8.0)
- Reconstituted cell-free expression system: PURESYSTEM CLASSIC II 96 from Cosmo Bio Co., #BCO-PURE2048C, or PURExpress™ from NEB, #E6800 S; both kits are provided with Solution A and Solution B, which contain all necessary components for the *in vitro* translation procedure
- 2 x protein sample loading buffer (2 x SLB: 40% glycerol, 6% 2-mercaptoethanol, 6% SDS, 10 mM DTT, 0.1% bromophenol blue, and 60 mM Tris-HCl pH 8.0)
- SDS-PAGE, 10–12%; *for details, see Sambrook and Russell, Molecular Cloning, A Laboratory Manual (Third Edition)*
- 10 x SDS running buffer: 0.25 Tris base, 1.918 M glycine, 1% SDS
- PolyScreen® PVDF Hybridization Transfer Membrane (PerkinElmer, #NEF1002001PK)
- 1 x Transfer buffer: 25 mM Tris Base, 192 mM glycine, 20% methanol
- 1 x TBS-T: Tris-buffered saline pH 7.4 containing Tween-20: 50 mM Tris-HCl pH 7.4, 150 mM sodium chloride, 0.1% Tween-20
- BSA (Roth, #8076.2)
- Anti-GFP antibody (Roche Diagnostics, #11814460001)

- HRP-conjugated secondary antibody, sheep anti-mouse IgG (GE Healthcare, #NXA931)
- ECL Plus™ Western Blotting Detection Reagents (GE Healthcare, #RPN2132)
- ECL film or Fluorimager

Step 1: Preparation of the DNA Template

The DNA template may encode the full-length mRNA of interest if a specific antibody against the protein is available. If there is no antibody available, we recommend the use of a GFP fusion in combination with commercially available anti-GFP antibodies. The *gfp* reporter fusion has the additional advantage of keeping the size of the full-length transcript below 1000 nt.

- 1) Amplify the entire or partial protein gene of interest from genomic DNA or from a previously constructed *gfp* fusion by standard PCR using a high-fidelity thermostable DNA polymerase in a 50 µl reaction.
- 2) Load the PCR reaction on a 1% agarose gel.
- 3) Gel-extract the correct band using a commercially available kit (e.g., QIAquick, Qiagen). Elute in 25–35 µl H₂O.

Note: If the PCR product is very pure, the preparative agarose gel can be omitted and a simple PCR purification step using a commercially available kit is sufficient.

- 4) Quantify the purified DNA template by UV spectroscopy.

Step 2: T7 Transcription

- 1) Thaw the frozen reagents.
- 2) Combine the following components:
 - a. H₂O to a final volume of 20 µl
 - b. 8 µl of an equimolar ATP, CTP, GTP, and UTP mix (each 18.75 mM)
 - c. 2 µl 10 x T7 reaction buffer (Ambion)
 - d. 1 µg linear template DNA
 - e. 2 µl enzyme mix (Ambion)

Note: The transcription reaction should be assembled at room temperature.

- 3) Incubate the reaction at 37 °C for 4–6 h or overnight at room temperature.
- 4) Add 2 U of TURBO DNase I and incubate for 15 min at 37 °C. At the end of the incubation, add 15 µl of ammonium stop solution (5 M ammonium acetate, 100 mM EDTA) and 65 µl H₂O.
- 5) Purify on an illustra MicroSpin™ G-50 column.
- 6) Transfer to a 2 ml PLG tube, add 100 µl of P:C:I. Shake vigorously for 30 s and separate the phases by centrifugation in a table-top centrifuge at maximum speed for 15 min at room temperature.
- 7) Transfer the aqueous phase (100 µl) to a fresh Eppendorf tube and add 300 µl of ethanol containing 0.3 M sodium acetate (pH 5.7). Precipitate for 2 h at –20 °C and pellet the RNA by centrifugation (at least 12 000 g, 30 min, 4 °C). Wash the pellet once with 500 µl of 75% ethanol and finally dissolve in 30 µl of H₂O. Determine the RNA concentration by UV spectroscopy.

Step 3: RNA Quality Control by Denaturing PAGE

To assess the quality of the *in vitro* synthesized RNA, run an aliquot on a denaturing polyacrylamide gel (4–5% depending on RNA size).

Note: Transcripts larger than about 1.5 kb should be run on denaturing agarose gels. For details, see Sambrook and Russell, Molecular Cloning, A Laboratory Manual (Third Edition).

- 1) Add 1 volume of denaturing gel loading buffer to 1 µg RNA. Add H₂O to a final volume of 10–15 µl. In parallel, prepare 5 µg of an RNA size marker.
- 2) Before gel loading, denature the RNA for 5 min at 95 °C, then cool on ice for 1 min.
- 3) The gel should be run at ~25 mA in 1 x TBE. Stop electrophoresis when the bromophenol blue has reached the bottom of the gel.
- 4) Remove the glass plates and stain the RNA using an RNA staining solution. If the *in vitro* synthesized RNA is pure, go to Step 5. If other bands appear, gel-purify the product of interest as described in Step 4.

Step 4 (Optional): Gel Purification of the *in vitro* Transcript

Note: Load at least 1 µg RNA. Consider distributing the RNA on several lanes to increase RNA elution recovery.

- 1) Put the gel slice containing the RNA of interest into a fresh 2 ml Eppendorf tube. Add 750 µl of RNA elution buffer. Incubate overnight in a cooling shaker (600 rpm, 4 °C).
- 2) Pellet polyacrylamide gel pieces by centrifugation in a table-top centrifuge at maximum speed for 5 min at 4 °C.
- 3) Transfer the supernatant to a 2 ml PLG tube, add 1 volume of P:C:I. Shake vigorously for 30 s and separate phases by centrifugation in a table-top centrifuge at maximum speed for 15 min at room temperature.
- 4) Transfer the aqueous phase to a fresh Eppendorf tube and add 3 volumes of ethanol containing 0.3 M sodium acetate (pH 5.7). Precipitate for 2 h at –20 °C and pellet the RNA by centrifugation (at least 12 000 g, 30 min, 4 °C). Wash the pellet once with 500 µl of 75% ethanol and finally dissolve in 30 µl of H₂O. Determine the RNA concentration by UV spectroscopy.

Step 5: *In vitro* Translation Reaction

The following protocol describes the experimental setup successfully applied to the sRNA GcvB and one of its target mRNAs fused to *gfp* (*gltI-gfp*) [11]. Note that *in vitro* translation efficiency greatly differs from one mRNA to the other. For instance, *in vitro* translation of the *ompR::gfp* fusion required 250 nM of *in vitro* transcribed mRNA for proper readout [11]. Therefore, it is recommended that several mRNA concentrations (ranging from 20 to 300 nM) be tested and the translation allowed to proceed for 30–45 min before stopping the reaction. The time course of protein synthesis should be followed to determine the incubation period necessary to produce detectable amounts of protein without reaching the saturation phase of the translation reaction. Finally, in some cases the Hfq protein

is essential for sRNA-mediated regulation, and it should be noted that addition of Hfq usually decreases translation efficiency.

The following protocol is optimized for the PURESYSYSTEM CLASSIC II 96 from Cosmo Bio Co., #BCO-PURE2048C kit; however, it can be adapted to the PURExpress™ from NEB (#E6800 S) as well.

- 1) The PURESYSYSTEM kit provides Solution A and B that contain all the necessary components for the *in vitro* reaction. Thaw the necessary amount of Solution A and Solution B aliquots on ice. Avoid multiple freeze–thaw cycles.
- 2) Denature 40 nM of full-length *in vitro* transcribed mRNA in the presence of a 10-fold excess of sRNA (400 nM) for 1 min at 95 °C in H₂O and then chill on ice for 5 min.
- 3) Incubate the sRNA–mRNA mix in the presence of 40 nM of purified Hfq-hexamer for 10 min at 37 °C in H₂O.
- 4) Prepare a PURESYSYSTEM mix (upscale according to the number of reactions you intend to perform). Per reaction, combine:
 - a. H₂O to final volume of 10 µl
 - b. 5 µl of Solution A
 - c. 2 µl of Solution B
 - d. sRNA–mRNA–Hfq mix
- 5) Mix gently, centrifuge briefly, and keep on ice.
- 6) If a master mix was prepared in (4), distribute 10 µl aliquots to fresh 1.5 ml Eppendorf tubes.
- 7) Incubate the tubes at 37 °C. Stop the reactions sequentially after 15, 30, and 45 min with 4 volumes of ice-cold acetone.
- 8) Chill on ice for 15 min. Collect the proteins by centrifugation using a table-top centrifuge at 10 000 g for 10 min at 4 °C. Redissolve the pellets in 2 x protein sample loading buffer.

Step 6: Protein Quantification by Western Blotting

- 1) Denature the sample for 10 min at 95 °C. Chill on ice. Centrifuge briefly.
- 2) Load half of each sample onto a 10–12% SDS-PAA gel. Run in 1 x SDS running buffer at 15–30 mA depending on the gel size (we usually apply 15 mA to 9 × 9 cm gels and 30 mA to 9 × 19 cm gels).
- 3) Transfer on PVDF membrane. *For details, see Sambrook and Russell, Molecular Cloning, A Laboratory Manual (Third Edition).*
- 4) Block the membrane with 1 x TBS-T containing 10% skim milk for 1 h at room temperature.
- 5) Incubate with an anti-GFP antibody diluted 1:1000 in 1 x TBS-T containing 3% BSA (or using another specific antibody at the appropriate dilution) for 1 h at room temperature.
- 6) Wash three times for 10 min with 1 x TBS-T.
- 7) Incubate for 1 h at room temperature with the HRP-conjugated secondary antibody diluted 1 : 5000 in 1 x TBS-T containing 3% BSA.
- 8) Wash three times for 10 min with 1 x TBS-T.

- 9) Develop the membrane in ECL detection solution. Initial film exposure times of 1–5 min are recommended. A Fluorimager can replace the film.

Protocol 6: 30S Ribosome Toeprinting Assay

Here, we provide the protocol for a 30S toeprint assay that has worked well in our hands. It is based on the protocol published by Udekwi *et al.* [156]. A schematic overview of the method is shown in Figure 33.7a. 30S ribosome toeprinting allows assessing the effect of an sRNA on translation initiation *in vitro*. To this end, the target mRNA is incubated with the 30S ribosome subunit and initiator tRNA^{fMet}. Reverse transcription on a ribosome-bound mRNA will stop at position +15/ +16 because of steric hindrance, which leads to the characteristic toeprint signal (Figure 33.7b). In case 30S binding is inhibited by an sRNA, the toeprint signal will decrease, since reverse transcriptase has the capability to read-through the sRNA–mRNA complex.

Required Reagents

- A sense primer containing the T7 promoter fused to the transcription start of the gene of interest (GOI): 5' gtttttttTAATACGACTCACTATAG::GOI. *Note: T7 transcription starts with the G in bold*
- An antisense primer corresponding to the 3' end extremity of the mRNA.
- DNA polymerase (0.5–1.0 U per reaction) and 5–10 x PCR buffer (e.g., *Taq* DNA polymerase with 10 x ThermoPol® buffer: 200 mM Tris-HCl, 100 mM ammonium sulfate, 100 mM potassium chloride, 20 mM magnesium sulfate, 1% Triton X-100, pH 8.8 at 25°C; NEB #M0267L)
- T7 transcription kit (MEGAscript® T7 kit, Ambion, #AM1334)
- A DNA primer complementary to the 3' end of the mRNA fragment
- T4 polynucleotide kinase (Thermo Scientific Fermentas, #EK0031)
- [γ -³²P]ATP (Hartmann Analytic, #SRP-501)
- IllustraMicroSpin™ G-25 columns (GE Healthcare, #27-5325-01)
- A DNA Sequencing Kit (we used the SequiTherm EXCEL™ kit from EPI-CENTRE Biotechnologies, or the CycleReader™ kit from Thermo Scientific Fermentas; both have recently been withdrawn from the market; replacements are the “Thermo Sequenase Cycle Sequencing Kit” of Affimetrix/USB, or identical/similar kits as well as individual sequencing components offered by Isogen Life Science)
- PLG heavy 2.0 ml tubes (5 PRIME, #2302830)
- 5 x sample buffer without magnesium (5 x SBMg0): 50 mM Tris-acetate pH 7.6, 5 mM DTT, 500 mM potassium acetate
- 1 x sample buffer containing 10 mM magnesium (SBMg10): 10 mM Tris-acetate pH 7.6, 1 mM DTT, 100 mM potassium acetate, 10 mM magnesium acetate
- 1 x sample buffer containing 60 mM magnesium (SBMg60): 10 mM Tris-acetate pH 7.6, 1 mM DTT, 100 mM potassium acetate, 60 mM magnesium acetate
- A dNTP mix containing 5 mM of each dATP, dTTP, dGTP, and dCTP

- 30S ribosomal subunits from *Escherichia coli*
- tRNA^{fMet}, N-formylmethionine-specific transfer ribonucleic acid from *E. coli* MRE 600 (Sigma-Aldrich, #R8019)
- Reverse transcriptase (SuperScript II, Life Technologies Invitrogen, #18064-014)
- STOP buffer: 50 mM Tris-HCl pH 7.5, 0.1% SDS, 10 mM EDTA pH 8.0

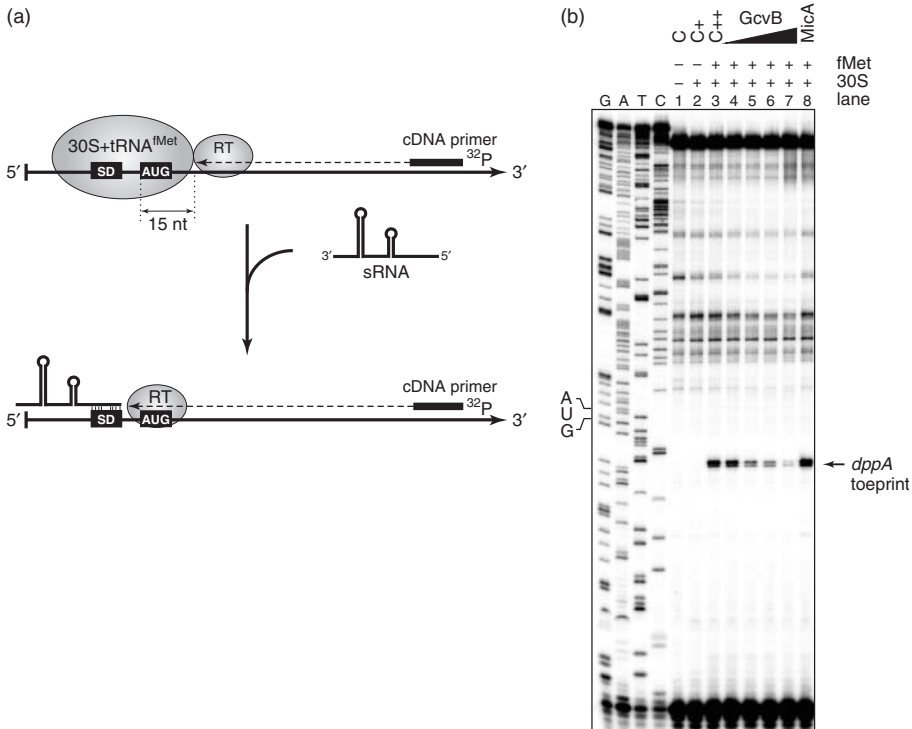


Figure 33.7 30S ribosome toeprinting assay. (a) An *in vitro* synthesized fragment of the mRNA including parts of the CDS is preincubated with purified 30S ribosomal subunits and initiator tRNA allowing formation of the ternary complex. A reverse transcription initiating from a radiolabeled primer that binds in the early CDS is performed. If the 30S subunit has stably associated with the Shine-Dalgarno sequence (SD) and the AUG start codon, reverse transcriptase (RT) will terminate around 15 nt downstream of the start codon due to steric hindrance; this leads to the characteristic “toeprint” signal. Binding of an sRNA at the ribosome binding site inhibits the formation of the ternary complex, thus allowing RT to read through, producing a full-length cDNA, or generate an

sRNA-specific toeprint on the target mRNA if the sRNA–mRNA duplex is stable enough to stop the elongation reaction. (b) 30S ribosome toeprinting of *dppA* mRNA. Incubation of *dppA* mRNA (20 nM) with 30S subunits (200 nM) and tRNA^{fMet} (1 μ M) led to the formation of a 30S/tRNA^{fMet}/mRNA complex and the appearance of the characteristic toeprint signal due to inhibition of primer extension (lane 3) as described above. 30S binding is progressively inhibited upon addition of increasing amounts of GcvB sRNA (lanes 4–7: 20, 60, 100, and 200 nM), as inferred from a decrease of the toeprint signal. The unrelated MicA sRNA, unable to bind to the *dppA* mRNA, did not affect the toeprint of the 30S subunit (lane 8). (Figure adapted from Ref. [11].)

- P:C:I (water-saturated phenol/chloroform/isoamylalcohol mixture, 25 : 24 : 1, v/v)
- 3 M potassium hydroxide solution
- 3 M acetic acid solution
- Glycoblue (Ambion, #AM9515)
- Absolute ethanol
- 3 M sodium acetate solution at pH 6.5
- Denaturing gel loading buffer for PAGE (e.g., Ambion #AM8546G, contains 95% formamide, 18 mM EDTA and 0.025% of SDS, xylene cyanol and bromophenol blue)
- Denaturing polyacrylamide gel, 4–5%. *For details, see Sambrook and Russell, Molecular Cloning, A Laboratory Manual (Third Edition)*
- 10 x TBE buffer: 0.9 M Tris base, 0.9 M boric acid, 20 mM EDTA
- Sequencing gel electrophoresis apparatus
- Gel dryer (Biorad, Model 583) and vacuum pump
- Phosphorimager (Typhoon FLA 7000, GE Healthcare)
- AIDA software (Raytest, Germany) or ImageJ (<http://rsbweb.nih.gov/ij/>)

Step 1: Synthesis of the mRNA Fragment

For the preparation of the DNA template and the T7 transcription reaction refer to Steps 1 and 2 of Protocol 5.

Step 2: DNA Primer 5'-End-Labeling

Use 2 pmol of the DNA primer which is antisense to the 3' extremity of the mRNA with 10 U of T4 polynucleotide kinase and 20 μ Ci of [γ - ^{32} P]ATP (use fresh radioactivity). Incubate for 1 h at 37 °C. After inactivation of the reaction by incubation for 3 min at 95 °C, purify the labeled DNA primer on a G-25 column following the manufacturer's instructions.

Step 3: Sequencing Ladder

A sequencing ladder should be generated on the DNA that served as the template for T7 transcription using the same 5'-end-labeled primer as for the toeprinting reactions. The ladder is necessary to map the 30S ribosome toeprint, which should be located at position +15 to +16 downstream of the start codon.

For the DNA sequencing step follow the manufacturer's protocol.

Step 4: Toeprinting Reaction

A 30S ribosome toeprint experiment should always include three controls: the negative control C in the absence of both the 30S subunit and tRNA^{fMet} (Figure 33.7b, lane 1), the negative control C+ in the absence of tRNA^{fMet} and the presence of the 30S subunit (Figure 33.7b, lane 2), and, finally, the positive control C++ in the presence of both the 30S subunit and tRNA^{fMet} (Figure 33.7b, lane 3). Only the toeprint appearing in the C++ lane at position +15 to +16 nt corresponds to ribosome binding.

To begin with, the mRNA (0.2 pmol), the sRNA (2 pmol), the 30S ribosomal subunit (2 pmol) and the tRNA^{fMet} (10 pmol) can be combined in this 1 : 10 : 10 : 50 ratio. Later, the ratio may be adjusted in order to improve toeprint quality.

1) Prepare the following annealing mixture:

	For one reaction	For five reactions
mRNA (1 μ M)	0.2 μ l	1 μ l
5 x SBMg0	0.8 μ l	4 μ l
H ₂ O	2.5 μ l	12.5 μ l
Primer (1 μ M)	0.5 μ l	2.5 μ l
→ Heat 1 min at 95 °C		
→ Chill on ice for 5 min		
dNTPs (5 mM each)	1 μ l	5 μ l
SBMg60	1 μ l	5 μ l
Final volume	6 μ l	30 μ l

2) For the toeprint reaction follow the pipetting scheme below:

Sample number	1	2	3	4
Sample name	C	C+	C + +	sRNA
Annealing mixture	6 μ l	6 μ l	6 μ l	6 μ l
sRNA (2 μ M) ^a	—	—	—	1 μ l
H ₂ O	1 μ l	1 μ l	1 μ l	—
→ Incubate 5 min at 37 °C				
30S subunit (1 μ M) ^b	—	2 μ l	2 μ l	2 μ l
SBMg10	2 μ l	—	—	—
→ Incubate 5 min at 37 °C				
tRNA ^{fMet} (10 μ M) ^b	—	—	1 μ l	1 μ l
SBMg10	1 μ l	1 μ l	—	—
→ Incubate 15 min at 37 °C				
Superscript II (100 U μ l ⁻¹) ^b	1 μ l	1 μ l	1 μ l	1 μ l
→ Incubate 20 min at 37 °C				
Add 100 μ l of STOP buffer				

^aBefore use, the sRNA must be denatured for 1 min at 95 °C and chilled for 5 min on ice.

^bDilute with SBMg10 buffer.

3) Transfer the whole reaction to a 2 ml PLG tube and add 110 μ l of P:C:I mix. Shake vigorously for 30 s and separate phases by centrifugation in a table-top centrifuge at maximum speed for 15 min at room temperature. Transfer aqueous phase (110 μ l) to a fresh Eppendorf tube and add 5 μ l of 3 M KOH. Heat 5 min at 95 °C. Add 10 μ l of 3 M acetic acid, 1 μ l of glycoblue, and 300 μ l of ethanol containing 0.3 M sodium acetate (pH 6.5). Precipitate at least for 2 h at -20 °C and pellet the RNA by centrifugation ($\geq 12\,000$ g, 30 min, 4 °C). Wash

the pellet once with 70% ethanol and dissolve the pellet in 11 μ l denaturing gel loading buffer.

- 4) Load 5 μ l of each reaction on a heated 6% sequencing gel, along with 3 μ l of each individual sequencing reaction and run the gel at 40 W for about 90 min. Before loading, the sequencing gel should be pre-run for at least 45 min, and the samples must be denatured for 2 min at 95 °C.
- 5) Dry the gel for 30 min at 80 °C in an appropriate gel dryer.
- 6) Analyze the gel with a Phosphorimager (for technical details, see Protocol 1, Step 3, (20)).

Acknowledgments

The authors gratefully acknowledge funding from the German Priority Program Sensory and regulatory RNAs in prokaryotes (SPP1258) and from EMBO (long-term fellowship to M.B.). We thank Kai Papenfort and Masatoshi Miyakoshi for helpful comments on the manuscript.

References

1. Waters, L.S. and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell*, **136**, 615–628.
2. Papenfort, K. and Vogel, J. (2010) Regulatory RNA in bacterial pathogens. *Cell Host Microbe*, **8**, 116–127.
3. Aiba, H. (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr. Opin. Microbiol.*, **10**, 134–139.
4. Brennan, R.G. and Link, T.M. (2007) Hfq structure, function and ligand binding. *Curr. Opin. Microbiol.*, **10**, 125–133.
5. Chao, Y. and Vogel, J. (2010) The role of Hfq in bacterial pathogens. *Curr. Opin. Microbiol.*, **13**, 24–33.
6. Valentin-Hansen, P., Eriksen, M., and Udesen, C. (2004) The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol. Microbiol.*, **51**, 1525–1533.
7. Vogel, J. and Luisi, B.F. (2011) Hfq and its constellation of RNA. *Nat. Rev. Microbiol.*, **9**, 578–589.
8. Mizuno, T., Chou, M.Y., and Inouye, M. (1984) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 1966–1970.
9. Schmidt, M., Zheng, P., and Delihias, N. (1995) Secondary structures of *Escherichia coli* antisense micF RNA, the 5'-end of the target *ompF* mRNA, and the RNA/RNA duplex. *Biochemistry*, **34**, 3621–3631.
10. Majdalani, N., Vanderpool, C.K., and Gottesman, S. (2005) Bacterial small RNA regulators. *Crit. Rev. Biochem. Mol. Biol.*, **40**, 93–113.
11. Sharma, C.M., Darfeuille, F., Plantinga, T.H., and Vogel, J. (2007) A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev.*, **21**, 2804–2817.
12. Prevost, K., Salvail, H., Desnoyers, G., Jacques, J.F., Phaneuf, E., and Massé, E. (2007) The small RNA RyhB activates the translation of *shiA* mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. *Mol. Microbiol.*, **64**, 1260–1273.
13. Vogel, J., Argaman, L., Wagner, E.G., and Altuvia, S. (2004) The small RNA IstR inhibits synthesis of an

- SOS-induced toxic peptide. *Curr. Biol.*, **14**, 2271–2276.
14. Vecerek, B., Moll, I., and Blasi, U. (2007) Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. *EMBO J.*, **26**, 965–975.
 15. Holmqvist, E., Reimegård, J., Sterk, M., Grantcharova, N., Römling, U., and Wagner, E.G. (2010) Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J.*, **29**, 1840–1850.
 16. Darfeuille, F., Unoson, C., Vogel, J., and Wagner, E.G.H. (2007) An antisense RNA inhibits translation by competing with standby ribosomes. *Mol. Cell.*, **26**, 381–392.
 17. Papenfort, K., Said, N., Welsink, T., Lucchini, S., Hinton, J.C., and Vogel, J. (2009) Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol. Microbiol.*, **74**, 139–158.
 18. Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J.C., and Vogel, J. (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat. Struct. Mol. Biol.*, **16**, 840–846.
 19. Bouvier, M., Sharma, C.M., Mika, F., Nierhaus, K.H., and Vogel, J. (2008) Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol. Cell.*, **32**, 827–837.
 20. Morfeldt, E., Taylor, D., von Gabain, A., and Arvidson, S. (1995) Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAlII. *EMBO J.*, **14**, 4569–4577.
 21. Sledjeski, D.D., Gupta, A., and Gottesman, S. (1996) The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in *Escherichia coli*. *EMBO J.*, **15**, 3993–4000.
 22. Majdalani, N., Cunnig, C., Sledjeski, D., Elliott, T., and Gottesman, S. (1998) DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 12462–12467.
 23. Fröhlich, K.S. and Vogel, J. (2009) Activation of gene expression by small RNA. *Curr. Opin. Microbiol.*, **12**, 674–682.
 24. Podkaminski, D. and Vogel, J. (2010) Small RNAs promote mRNA stability to activate the synthesis of virulence factors. *Mol. Microbiol.*, **78**, 1327–1331.
 25. Papenfort, K. and Vogel, J. (2009) Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Res. Microbiol.*, **160**, 278–287.
 26. Beisel, C.L. and Storz, G. (2010) Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol. Rev.*, **34**, 866–882.
 27. Guillier, M. and Gottesman, S. (2008) The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. *Nucleic Acids Res.*, **36**, 6781–6794.
 28. Balbontin, R., Fiorini, F., Figueroa-Bossi, N., Casadesús, J., and Bossi, L. (2010) Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in *Salmonella enterica*. *Mol. Microbiol.*, **78**, 380–394.
 29. Papenfort, K., Bouvier, M., Mika, F., Sharma, C.M., and Vogel, J. (2010) Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 20435–20440.
 30. Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jäger, J.G., Hüttenhofer, A., and Wagner, E.G. (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, **31**, 6435–6443.
 31. Willkomm, D.K., Minnerup, J., Hüttenhofer, A., and Hartmann, R.K. (2005) Experimental RNomics in *Aquifex aeolicus*: identification of small non-coding RNAs and the putative 6S RNA homolog. *Nucleic Acids Res.*, **33**, 1949–1960.
 32. Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S.

- S. (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.*, **15**, 1637–1651.
33. Tjaden, B., Saxena, R.M., Stolyar, S., Haynor, D.R., Kolker, E., and Rosenow, C. (2002) Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res.*, **30**, 3732–3738.
34. Landt, S.G., Abeliuk, E., McGrath, P.T., Lesley, J.A., McAdams, H.H., and Shapiro, L. (2008) Small non-coding RNAs in *Caulobacter crescentus*. *Mol. Microbiol.*, **68**, 600–614.
35. Steglich, C., Futschik, M.E., Lindell, D., Voss, B., Chisholm, S.W., and Hess, W.R. (2008) The challenge of regulation in a minimal photoautotroph: non-coding RNAs in *Prochlorococcus*. *PLoS Genet.*, **4**, e1000173.
36. Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K. *et al.* (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature*, **459**, 950–956.
37. Akama, T., Suzuki, K., Tanigawa, A., Kawashima, H., Wu, N., Nakata, Y., Osana, Y., Sakakibara, Y., and Ishii, N. (2009) Whole-genome tiling array analysis of *Mycobacterium leprae* RNA reveals high expression of pseudogenes and noncoding regions. *J. Bacteriol.*, **191**, 3321–3327.
38. Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G., and Gottesman, S. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.*, **50**, 1111–1124.
39. Hu, Z., Zhang, A., Storz, G., Gottesman, S., and Leppla, S.H. (2006) An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res.*, **34**, e52.
40. Kawano, M., Reynolds, A.A., Miranda-Rios, J., and Storz, G. (2005) Detection of 5'- and 3'-UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.*, **33**, 1040–1050.
41. Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., Hinton, J.C., and Vogel, J. (2008) Deep sequencing analysis of small non-coding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.*, **4**, e1000163.
42. Sorek, R. and Cossart, P. (2010) Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat. Rev. Genet.*, **11**, 9–16.
43. Passalacqua, K.D., Varadarajan, A., Ondov, B.D., Okou, D.T., Zwick, M.E., and Bergman, N.H. (2009) Structure and complexity of a bacterial transcriptome. *J. Bacteriol.*, **191**, 3203–3211.
44. Yoder-Himes, D.R., Chain, P.S., Zhu, Y., Wurtzel, O., Rubin, E.M., Tiedje, J.M., and Sorek, R. (2009) Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 3976–3981.
45. Liu, J.M., Livny, J., Lawrence, M.S., Kimball, M.D., Waldor, M.K., and Camilli, A. (2009) Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.*, **37**, e46.
46. Mraheil, M.A., Billion, A., Mohamed, W., Mukherjee, K., Kuenne, C., Pischmarov, J., Krawitz, C., Retey, J., Hartsch, T., Chakraborty, T., and Hain, T. (2011) The intracellular sRNA transcriptome of *Listeria monocytogenes* during growth in macrophages. *Nucleic Acids Res.* **39**, 4235–4248.
47. Sittka, A., Sharma, C.M., Rolle, K., and Vogel, J. (2009) Deep sequencing of *Salmonella* RNA associated with heterologous Hfq proteins *in vivo* reveals small RNAs as a major target class and identifies RNA processing phenotypes. *RNA Biol.*, **6**, 266–275.
48. Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., Stadler, P.F., and Vogel, J. (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, **464**, 250–255.
49. Jager, D., Sharma, C.M., Thomsen, J., Ehlers, C., Vogel, J., and Schmitz,

- R.A. (2009) Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 21878–21882.
50. Albrecht, M., Sharma, C.M., Reinhardt, R., Vogel, J., and Rudel, T. (2009) Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. *Nucleic Acids Res.*, **38**, 868–877.
 51. Mitschke, J., Georg, J., Scholz, I., Sharma, C.M., Dienst, D., Bantscheff, J., Voss, B., Steglich, C., Wilde, A., Vogel, J., and Hess, W.R. (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 2124–2129.
 52. Schluter, J.P., Reinkensmeier, J., Daschkey, S., Evguenieva-Hackenberg, E., Janssen, S., Jänicke, S., Becker, J.D., Giegerich, R., and Becker, A. (2010) A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. *BMC Genomics*, **11**, 245.
 53. Karginov, F.V. and Hannon, G.J. (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell.*, **37**, 7–19.
 54. Liu, M.Y., Gui, G., Wei, B., Preston, J.F., Oakford, L., Yüksel, U., Giedroc, D.P., and Romeo, T. (1997) The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J. Biol. Chem.*, **272**, 17502–17510.
 55. Heeb, S., Blumer, C., and Haas, D. (2002) Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct formation in *Pseudomonas fluorescens* CHA0. *J. Bacteriol.*, **184**, 1046–1056.
 56. Chant, E.L. and Summers, D.K. (2007) Indole signalling contributes to the stable maintenance of *Escherichia coli* multicopy plasmids. *Mol. Microbiol.*, **63**, 35–43.
 57. Jousset, A., Metzinger, L., and Felden, B. (2009) On the facultative requirement of the bacterial RNA chaperone, Hfq. *Trends Microbiol.*, **17**, 399–405.
 58. Christiansen, J.K., Nielsen, J.S., Ebersbach, T., Valentin-Hansen, P., Sogaard-Andersen, L., and Kallipolitis, B.H. (2006) Identification of small Hfq-binding RNAs in *Listeria monocytogenes*. *RNA*, **12**, 1383–1396.
 59. Papenfort, K., Pfeiffer, V., Lucchini, S., Sonawane, A., Hinton, J.C., and Vogel, J. (2008) Systematic deletion of *Salmonella* small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Mol. Microbiol.*, **68**, 890–906.
 60. Lybecker, M.C., Abel, C.A., Feig, A.L., and Samuels, D.S. (2010) Identification and function of the RNA chaperone Hfq in the Lyme disease spirochete *Borrelia burgdorferi*. *Mol. Microbiol.*, **78**, 622–635.
 61. Sonnleitner, E., Abdou, L., and Haas, D. (2009) Small RNA as global regulator of carbon catabolite repression in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 21866–21871.
 62. Argaman, L., Hershsberg, R., Vogel, J., Bejerano, G., Wagner, E.G., Margalit, H., and Altuvia, S. (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.*, **11**, 941–950.
 63. Rivas, E., Klein, R.J., Jones, T.A., and Eddy, S.R. (2001) Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr. Biol.*, **11**, 1369–1373.
 64. Chen, S., Lesnik, E.A., Hall, T.A., Sampath, R., Griffey, R.H., Ecker, D.J., and Blyn, L.B. (2002) A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *Biosystems*, **65**, 157–177.
 65. Livny, J., Fogel, M.A., Davis, B.M., and Waldor, M.K. (2005) sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes. *Nucleic Acids Res.*, **33**, 4096–4105.
 66. Livny, J., Brenic, A., Lory, S., and Waldor, M.K. (2006) Identification of 17 *Pseudomonas aeruginosa* sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. *Nucleic Acids Res.*, **34**, 3484–3493.

67. Livny, J., Teonadi, H., Livny, M., and Waldor, M.K. (2008) High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. *PLoS ONE*, **3**, e3197.
68. Pichon, C. and Felden, B. (2003) Intergenic sequence inspector: searching and identifying bacterial RNAs. *Bioinformatics*, **19**, 1707–1709.
69. Axmann, I.M., Kensche, P., Vogel, J., Kohl, S., Herzel, H., and Hess, W.R. (2005) Identification of cyanobacterial non-coding RNAs by comparative genome analysis. *Genome Biol.*, **6**, R73.
70. Washietl, S. and Hofacker, I.L. (2004) Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. *J. Mol. Biol.*, **342**, 19–30.
71. Mandin, P., Repoila, F., Vergassola, M., Geissmann, T., and Cossart, P. (2007) Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. *Nucleic Acids Res.*, **35**, 962–974.
72. Backofen, R. and Hess, W.R. (2010) Computational prediction of sRNAs and their targets in bacteria. *RNA Biol.*, **7**, 33–42.
73. Mutalik, V.K., Nonaka, G., Ades, S.E., Rhodius, V.A., and Gross, C.A. (2009) Promoter strength properties of the complete sigma E regulon of *Escherichia coli* and *Salmonella enterica*. *J. Bacteriol.*, **191**, 7279–7287.
74. Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., Wingreen, N.S., and Bassler, B.L. (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell*, **118**, 69–82.
75. Wilderman, P.J., Sowa, N.A., FitzGerald, D.J., FitzGerald, P.C., Gottesman, S., Ochsner, U.A., and Vasil, M.L. (2004) Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 9792–9797.
76. Mellin, J.R., Goswami, S., Grogan, S., Tjaden, B., and Genco, C.A. (2007) A novel fur- and iron-regulated small RNA, NrrF, is required for indirect fur-mediated regulation of the *sdhA* and *sdhC* genes in *Neisseria meningitidis*. *J. Bacteriol.*, **189**, 3686–3694.
77. Lorenz, C., von Pelchrzim, F., and Schroeder, R. (2006) Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat. Protoc.*, **1**, 2204–2212.
78. Lorenz, C., Gesell, T., Zimmermann, B., Schoeberl, U., Bilusic, I., Rajkowsch, L., Waldsich, C., von Haeseler, A., and Schroeder, R. (2010) Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Res.*, **38**, 3794–3808.
79. Opdyke, J.A., Kang, J.G., and Storz, G. (2004) GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J. Bacteriol.*, **186**, 6698–6705.
80. Urban, J.H., Papenfort, K., Thomsen, J., Schmitz, R.A., and Vogel, J. (2007) A conserved small RNA promotes dis-coordinate expression of the *glmUS* operon mRNA to activate GlmS synthesis. *J. Mol. Biol.*, **373**, 521–528.
81. Massé, E., Vanderpool, C.K., and Gottesman, S. (2005) Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J. Bacteriol.*, **187**, 6962–6971.
82. Papenfort, K., Pfeiffer, V., Mika, F., Lucchini, S., Hinton, J.C., and Vogel, J. (2006) SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global *omp* mRNA decay. *Mol. Microbiol.*, **62**, 1674–1688.
83. Vanderpool, C.K. and Gottesman, S. (2004) Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol. Microbiol.*, **54**, 1076–1089.
84. Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L., and Storz, G. (1997) A small, stable RNA induced by oxidative stress: role as a pleiotropic

- regulator and antimitator. *Cell*, **90**, 43–53.
85. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 6640–6645.
 86. Urbanowski, M.L., Stauffer, L.T., and Stauffer, G.V. (2000) The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol. Microbiol.*, **37**, 856–868.
 87. Song, T., Mika, F., Lindmark, B., Liu, Z., Schild, S., Bishop, A., Zhu, J., Camilli, A., Johansson, J., Vogel, J., and Wai, S.N. (2008) A new *Vibrio cholerae* sRNA modulates colonization and affects release of outer membrane vesicles. *Mol. Microbiol.*, **70**, 100–111.
 88. Bossi, L. and Figueroa-Bossi, N. (2007) A small RNA downregulates LamB maltoporin in *Salmonella*. *Mol. Microbiol.*, **65**, 799–810.
 89. Andersen, J. and Delihias, N. (1990) micF RNA binds to the 5' end of *ompF* mRNA and to a protein from *Escherichia coli*. *Biochemistry*, **29**, 9249–9256.
 90. Urban, J.H. and Vogel, J. (2007) Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res.*, **35**, 1018–1037.
 91. Lutz, R. and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.*, **25**, 1203–1210.
 92. Schleif, R. (2010) AraC protein, regulation of the L-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.*, **34**, 779–796.
 93. Wagner, E.G., Altuvia, S., and Romby, P. (2002) Antisense RNAs in bacteria and their genetic elements. *Adv. Genet.*, **46**, 361–398.
 94. Wagner, E. and Vogel, J. (2003) in *Noncoding RNAs* (eds J. Barciszewski and V. Erdmann), Landes Bioscience, Austin, TX, pp. 243–259.
 95. Geissmann, T., Marzi, S., and Romby, P. (2009) The role of mRNA structure in translational control in bacteria. *RNA Biol.*, **6**, 153–160.
 96. Bensing, B.A., Meyer, B.J., and Dunny, G.M. (1996) Sensitive detection of bacterial transcription initiation sites and differentiation from RNA processing sites in the pheromone-induced plasmid transfer system of *Enterococcus faecalis*. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 7794–7799.
 97. Clepet, C., Le Clainche, I., and Caboche, M. (2004) Improved full-length cDNA production based on RNA tagging by T4 DNA ligase. *Nucleic Acids Res.*, **32**, e6.
 98. Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.*, **177**, 4121–4130.
 99. Lease, R.A., Smith, D., McDonough, K., and Belfort, M. (2004) The small noncoding DsrA RNA is an acid resistance regulator in *Escherichia coli*. *J. Bacteriol.*, **186**, 6179–6185.
 100. Majdalani, N., Chen, S., Murrow, J., St John, K., and Gottesman, S. (2001) Regulation of RpoS by a novel small RNA: the characterization of RprA. *Mol. Microbiol.*, **39**, 1382–1394.
 101. Mandin, P. and Gottesman, S. (2010) Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *EMBO J.*, **29**, 3094–3107.
 102. Zhang, A., Altuvia, S., Tiwari, A., Argaman, L., Hengge-Aronis, R., and Storz, G. (1998) The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. *EMBO J.*, **17**, 6061–6068.
 103. Guillier, M. and Gottesman, S. (2006) Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. *Mol. Microbiol.*, **59**, 231–247.
 104. Amann, E., Brosius, J., and Ptashne, M. (1983) Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*. *Gene*, **25**, 167–178.

105. Morgan-Kiss, R.M., Wadler, C., and Cronan, J.E. Jr. (2002) Long-term and homogeneous regulation of the *Escherichia coli* araBAD promoter by use of a lactose transporter of relaxed specificity. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 7373–7377.
106. Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.*, **189**, 113–130.
107. Englesberg, E., Anderson, R.L., Weinberg, R., Lee, N., Hoffee, P., Huttenhauer, G., and Boyer, H. (1962) L-Arabinose-sensitive, L-ribulose 5-phosphate 4-epimerase-deficient mutants of *Escherichia coli*. *J. Bacteriol.*, **84**, 137–146.
108. Dunn, T.M. and Schleif, R. (1984) Deletion analysis of the *Escherichia coli* ara P_C and P_{BAD} promoters. *J. Mol. Biol.*, **180**, 201–204.
109. Sittka, A., Pfeiffer, V., Tedin, K., and Vogel, J. (2007) The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol. Microbiol.*, **63**, 193–217.
110. Jin, Y., Watt, R.M., Danchin, A., and Huang, J.D. (2009) Small noncoding RNA GcvB is a novel regulator of acid resistance in *Escherichia coli*. *BMC Genomics*, **10**, 165.
111. Santiago, C.A., Reynolds, M.M., Porwollik, S., Choi, S.H., Long, F., Andrews-Polymenis, H.L., and McClelland, M. (2009) Analysis of pools of targeted *Salmonella* deletion mutants identifies novel genes affecting fitness during competitive infection in mice. *PLoS Pathog.*, **5**, e1000477.
112. Hobbs, E.C., Astarita, J.L., and Storz, G. (2010) Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in *Escherichia coli*: analysis of a bar-coded mutant collection. *J. Bacteriol.*, **192**, 59–67.
113. Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., Wong, A. *et al.* (2011) Phenotypic landscape of a bacterial cell. *Cell*, **144**, 143–156.
114. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.0008.
115. Butland, G., Peregrín-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N. *et al.* (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*, **433**, 531–537.
116. Wassarman, K.M. and Storz, G. (2000) 6S RNA regulates *E. coli* RNA polymerase activity. *Cell*, **101**, 613–623.
117. Srisawat, C. and Engelke, D.R. (2001) Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA*, **7**, 632–641.
118. Locker, N., Easton, L.E., and Lukavsky, P.J. (2006) Affinity purification of eukaryotic 48S initiation complexes. *RNA*, **12**, 683–690.
119. Windbichler, N. and Schroeder, R. (2006) Isolation of specific RNA-binding proteins using the streptomycin-binding RNA aptamer. *Nat. Protoc.*, **1**, 637–640.
120. Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16719–16724.
121. Bardwell, V.J. and Wickens, M. (1990) Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res.*, **18**, 6587–6594.
122. Zhou, Z., Sim, J., Griffith, J., and Reed, R. (2002) Purification and electron microscopic visualization of functional human spliceosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12203–12207.
123. Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Lührmann, R. (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature*, **452**, 846–850.

124. Said, N., Rieder, R., Hurwitz, R., Deckert, J., Urlaub, H., and Vogel, J. (2009) *In vivo* expression and purification of aptamer-tagged small RNA regulators. *Nucleic Acids Res.*, **37**, e133.
125. Windbichler, N., von Pelchrzim, F., Mayer, O., Csaszar, E., and Schroeder, R. (2008) Isolation of small RNA-binding proteins from *E. coli*: evidence for frequent interaction of RNAs with RNA polymerase. *RNA Biol.*, **5**, 30–40.
126. Massé, E. and Gottesman, S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 4620–4625.
127. Möller, T., Franch, T., Udesen, C., Gerdes, K., and Valentin-Hansen, P. (2002) Spot 42 RNA mediates dis-coordinate expression of the *E. coli* galactose operon. *Genes Dev.*, **16**, 1696–1706.
128. Argaman, L. and Altuvia, S. (2000) *fhlA* repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. *J. Mol. Biol.*, **300**, 1101–1112.
129. Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.
130. Tjaden, B., Goodwin, S.S., Opdyke, J.A., Guillier, M., Fu, D.X., Gottesman, S., and Storz, G. (2006) Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res.*, **34**, 2791–2802.
131. Tjaden, B. (2008) TargetRNA: a tool for predicting targets of small RNA action in bacteria. *Nucleic Acids Res.*, **36**, W109–W113.
132. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
133. Davis, B.M. and Waldor, M.K. (2007) RNase E-dependent processing stabilizes MicX, a *Vibrio cholerae* sRNA. *Mol. Microbiol.*, **65**, 373–385.
134. Cao, Y., Zhao, Y., Cha, L., Ying, X., Wang, L., Shao, N., and Li, W. (2009) sRNATarget: a web server for prediction of bacterial sRNA targets. *Bioinformatics*, **3**, 364–366.
135. Zhao, Y., Li, H., Hou, Y., Cha, L., Cao, Y., Wang, L., Ying, X., and Li, W. (2008) Construction of two mathematical models for prediction of bacterial sRNA targets. *Biochem. Biophys. Res. Commun.*, **372**, 346–350.
136. Muckstein, U., Tafer, H., Hackermüller, J., Bernhart, S.H., Stadler, P.F., and Hofacker, I.L. (2006) Thermodynamics of RNA-RNA binding. *Bioinformatics*, **22**, 1177–1182.
137. Busch, A., Richter, A.S., and Backofen, R. (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, **24**, 2849–2856.
138. Sharma, C.M., Papenfort, K., Pernitzsch, S.R., Mollenkopf, H.J., Hinton, J.C., and Vogel, J. (2011) Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol. Microbiol.*, **81**, 1144–1165.
139. Heidrich, N., Moll, I., and Brantl, S. (2007) *In vitro* analysis of the interaction between the small RNA SR1 and its primary target *ahrC* mRNA. *Nucleic Acids Res.*, **35**, 4331–4346.
140. Urban, J.H. and Vogel, J. (2008) Two seemingly homologous noncoding RNAs act hierarchically to activate *glmS* mRNA translation. *PLoS Biol.*, **6**, e64.
141. Fantappie, L., Oriente, F., Muzzi, A., Serruto, D., Scarlato, V., and Delany, I. (2011) A novel Hfq-dependent sRNA that is under FNR-control and is synthesised in oxygen limitation in *Neisseria meningitidis*. *Mol. Microbiol.*, **80**, 507–523.
142. Mandin, P. and Gottesman, S. (2009) A genetic approach for finding small RNAs regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-component system. *Mol. Microbiol.*, **72**, 551–565.
143. Kawamoto, H., Koide, Y., Morita, T., and Aiba, H. (2006) Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration

- of duplex formation by Hfq. *Mol. Microbiol.*, **61**, 1013–1022.
144. Chen, S., Zhang, A., Blyn, L.B., and Storz, G. (2004) MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. *J. Bacteriol.*, **186**, 6689–6697.
 145. Desnoyers, G., Morissette, A., Prévost, K., and Massé, E. (2009) Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. *EMBO J.*, **28**, 1551–1561.
 146. van Batenburg, F.H., Gultyaev, A.P., and Pleij, C.W. (1995) An APL-programmed genetic algorithm for the prediction of RNA secondary structure. *J. Theor. Biol.*, **174**, 269–280.
 147. Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.
 148. Lindell, M., Romby, P., and Wagner, E.G. (2002) Lead(II) as a probe for investigating RNA structure *in vivo*. *RNA*, **8**, 534–541.
 149. Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., Lina, G., Etienne, J., Ehresmann, B., Ehresmann, C. *et al.* (2005) *Staphylococcus aureus* RNAIII and the endoribonuclease III coordinately regulate *spa* gene expression. *EMBO J.*, **24**, 824–835.
 150. Durand, S. and Storz, G. (2010) Reprogramming of anaerobic metabolism by the FnrS small RNA. *Mol. Microbiol.*, **75**, 1215–1231.
 151. Persson, C., Wagner, E.G., and Nordstrom, K. (1988) Control of replication of plasmid R1: kinetics of *in vitro* interaction between the antisense RNA, CopA, and its target, CopT. *EMBO J.*, **7**, 3279–3288.
 152. Persson, C., Wagner, E.G., and Nordstrom, K. (1990) Control of replication of plasmid R1: formation of an initial transient complex is rate-limiting for antisense RNA-target RNA pairing. *EMBO J.*, **9**, 3777–3785.
 153. Chevalier, C., Boisset, S., Romilly, C., Masquida, B., Fechter, P., Geissmann, T., Vandenesch, F., and Romby, P. (2010) *Staphylococcus aureus* RNAIII binds to two distant regions of *coa* mRNA to arrest translation and promote mRNA degradation. *PLoS Pathog.*, **6**, e1000809.
 154. Hartz, D., McPheeters, D.S., Traut, R., and Gold, L. (1988) Extension inhibition analysis of translation initiation complexes. *Methods Enzymol.*, **164**, 419–425.
 155. Altuvia, S., Zhang, A., Argaman, L., Tiwari, A., and Storz, G. (1998) The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *EMBO J.*, **17**, 6069–6075.
 156. Udekwu, K.I., Darfeuille, F., Vogel, J., Reimegård, J., Holmqvist, E., and Wagner, E.G. (2005) Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes Dev.*, **19**, 2355–2366.
 157. Nirenberg, M.W. and Matthaei, J.H. (1961) The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.*, **47**, 1588–1602.
 158. Katzen, F., Chang, G., and Kudlicki, W. (2005) The past, present and future of cell-free protein synthesis. *Trends Biotechnol.*, **23**, 150–156.
 159. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nat. Biotechnol.*, **19**, 751–755.
 160. Maki, K., Uno, K., Morita, T., and Aiba, H. (2008) RNA, but not protein partners, is directly responsible for translational silencing by a bacterial Hfq-binding small RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 10332–10337.
 161. Caron, M.P., Lafontaine, D.A., and Massé, E. (2010) Small RNA-mediated regulation at the level of transcript stability. *RNA Biol.*, **7**, 140–144.
 162. Weinberg, Z., Perreault, J., Meyer, M.M., and Breaker, R.R. (2009) Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature*, **462**, 656–659.
 163. Weinberg, Z., Wang, J.X., Bogue, J., Yang, J., Corbino, K., Moy, R.H., and Breaker, R.R. (2010) Comparative

- genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.*, **11**, R31.
164. Wadler, C.S. and Vanderpool, C.K. (2007) A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 20454–20459.
 165. Novick, R.P. and Geisinger, E. (2008) Quorum sensing in *Staphylococci*. *Annu. Rev. Genet.*, **42**, 541–564.
 166. Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., and Johansson, J. (2009) A *trans*-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell*, **139**, 770–779.
 167. Deveau, H., Garneau, J.E., and Moineau, S. (2010) CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.*, **64**, 475–493.
 168. Dubnau, D. and Losick, R. (2006) Bistability in bacteria. *Mol. Microbiol.*, **61**, 564–572.
 169. Avery, S.V. (2006) Microbial cell individuality and the underlying sources of heterogeneity. *Nat. Rev. Microbiol.*, **4**, 577–587.
 170. Veening, J.W., Smits, W.K., and Kuipers, O.P. (2008) Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.*, **62**, 193–210.
 171. Levine, E., Zhang, Z., Kuhlman, T., and Hwa, T. (2007) Quantitative characteristics of gene regulation by small RNA. *PLoS Biol.*, **5**, e229.
 172. Shimoni, Y., Friedlander, G., Hetzroni, G., Niv, G., Altuvia, S., Biham, O., and Margalit, H. (2007) Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol. Syst. Biol.*, **3**, 138.
 173. Hao, Y., Xu, L., and Shi, H. (2011) Theoretical analysis of catalytic-sRNA-mediated gene silencing. *J. Mol. Biol.*, **406**, 195–204.
 174. Ackermann, M., Stearns, S.C., and Jenal, U. (2003) Senescence in a bacterium with asymmetric division. *Science*, **300**, 1920.
 175. Stewart, E.J., Madden, R., Paul, G., and Taddei, F. (2005) Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.*, **3**, e45.
 176. Veening, J.W., Stewart, E.J., Berngruber, T.W., Taddei, F., Kuipers, O.P., and Hamoen, L.W. (2008) Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 4393–4398.
 177. Lindner, A.B., Madden, R., Demarez, A., Stewart, E.J., and Taddei, F. (2008) Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 3076–3081.
 178. Russell, J.H. and Keiler, K.C. (2009) Subcellular localization of a bacterial regulatory RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 16405–16409.
 179. Montero Llopis, P., Jackson, A.F., Sliusarenko, O., Survtsev, I., Heinritz, J., Emonet, T., and Jacobs-Wagner, C. (2010) Spatial organization of the flow of genetic information in bacteria. *Nature*, **466**, 77–81.
 180. Liou, G.G., Jane, W.N., Cohen, S.N., Lin, N.S., and Lin-Chao, S. (2001) RNA degradosomes exist *in vivo* in *Escherichia coli* as multicomponent complexes associated with the cytoplasmic membrane via the N-terminal region of ribonuclease E. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 63–68.
 181. Taghbalout, A. and Rothfield, L. (2007) RNaseE and the other constituents of the RNA degradosome are components of the bacterial cytoskeleton. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 1667–1672.
 182. Khemici, V., Poljak, L., Luisi, B.F., and Carpousis, A.J. (2008) The RNase E of *Escherichia coli* is a membrane-binding protein. *Mol. Microbiol.*, **70**, 799–813.
 183. Azam, T.A., Hiraga, S., and Ishihama, A. (2000) Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells*, **5**, 613–626.

184. Diestra, E., Cayrol, B., Arluison, V., and Risco, C. (2009) Cellular electron microscopy imaging reveals the localization of the Hfq protein close to the bacterial membrane. *PLoS ONE*, **4**, e8301.
185. Elf, J., Li, G.W., and Xie, X.S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, **316**, 1191–1194.

34

The Identification of Bacterial Non-coding RNAs through Complementary Approaches

Björn Voß and Wolfgang R. Hess

34.1

Introduction

Bacterial genomes encode a plethora of non-coding RNAs (ncRNAs) that are highly divergent in sequence, length, and secondary structure. This divergence is further extended owing to the fact that some “ncRNAs” with a known regulatory function have additional functions, such as coding for peptides. Moreover, bacterial ncRNAs, which are often termed small RNAs (sRNAs), are not necessarily always small in size (<100 nt). For practical reasons, we will use here the term *ncRNA* throughout for all types of RNA in bacteria that are not mRNA, tRNA, or rRNA. Biochemically characterized examples from various bacteria suggest that many of them fulfill important regulatory purposes, frequently as part of an adaptive stress response. Thus, it will be impossible to understand a large variety of bacterial regulatory systems without the correct integration of regulatory RNAs into the existing models of such regulatory networks.

Genes for putative ncRNAs can be computationally predicted by various methods and with variable efficacy but frequently relying on comparative genome analysis. Although their biocomputational prediction provides a fairly good starting point for the analysis of ncRNAs in a bacterial genome, none of the existing sRNA gene finders is able to identify all the experimentally validated ncRNA genes. Therefore, the combination of biocomputational and experimental methods is the best approach to identify the full set of ncRNAs. Comprehensive approaches for the experimental discovery and characterization of bacterial ncRNAs are mainly tiling microarrays and deep sequencing of the transcriptome (Section 34.3).

34.2

Computational Prediction

Owing to divergence in sequence and structure, there are no common identifiers for bacterial ncRNAs. Therefore, parameters used in conventional genome annotation and gene modeling are inappropriate, and a universal method for the detection of all classes of bacterial ncRNAs is lacking. Nevertheless, there are several algorithms

that provide reliable ncRNA predictions in a more focused context as well as program packages that can be applied to large-scale biocomputational analyses.

Here, we focus on comparative methods, as they turn out to be the most widely and most successfully used approaches. The principle idea behind all these methods is that a functional entity of the genome is conserved on the sequence level. However, ncRNAs have the drawback that their function is in most cases primarily determined by their structure that can be achieved with various sequences. This enables structure-neutral mutations to accumulate, which blurs the signal of sequence conservation. In addition, the degree of conservation cannot be directly inferred from the sequence but rather requires measures for structural conservation.

With the development of methods such as *QRNA* [1], *RNAz* [2], and *evofold* [3], such measures became available, permitting researchers to identify ncRNAs in comparative studies. For a recent overview on the computational prediction of bacterial ncRNAs and their targets, including the theoretical background, see Backofen and Hess [4].

The major bottleneck in genome-wide comparative ncRNA annotation is to locate related entities within the analyzed genomes, which can then be subjected to the above-described methods. Obviously, structure-based methods would be preferable, but they are computationally expensive and, therefore, not applicable to large-scale studies. The most common strategy is to use standard or specialized tools for the assessment of sequence conservation, such as *BLAST* [5] or *NcDnAlign* [6] as an example of a specialized tool. As already mentioned above, signals on the sequence level are weaker than on the structure level, and thus, they are harder to get. Particularly important here is the choice of genomes from related species that are included in the comparative analysis. If the evolutionary distance is too short, nearly every part of the genome will be conserved, thus requiring a huge (computational) effort to filter out the ones also showing structural conservation. On the other hand, sequence conservation might be hardly detectable or undetectable at all when only distantly related genomes are compared, leaving no or only few candidates for structural analysis.

34.2.1

Workflow

In general, the workflow is as follows:

- 1) Select the set of genomes for comparison and extract genomic regions to be analyzed; for example, all intergenic regions with a length ≥ 50 nt.
- 2) Compute conserved regions using *BLAST* or a similar method, which will result in pairwise homology information.
- 3) In order to provide comprehensive information for the entire set of genomes, the pairwise homologies have to be combined into multispecies homologies, which is a crucial step of the analysis. This clustering can be performed using, for example, *BLASTclust* or other methods, such as single linkage.
- 4) For each cluster, a multiple sequence alignment is computed, for example, by *ClustalW*, *T-COFFEE*, or another multiple sequence alignment program.

- 5) Each cluster alignment is subjected to a measure of structural conservation, for example, using *RNAz*, *QRNA*, or *evofold*. The scoring is also performed for the antisense sequences of the alignment, as the ncRNA may be encoded on the sense or antisense strand of the corresponding genomic region. On the basis of the derived scores, it is possible to decide whether or not a conserved sequence is actually a promising ncRNA candidate.

An example of this workflow with data from [7] is given in Figure 34.1, and the results are discussed in the next section. Steps (2)–(4) are related to the general problem of generating multiple sequence alignments of whole genome data sets. Several approaches to this have been developed, for example, *NcDNAlign* [6]. Similar to the above-described workflow, *NcDNAlign* makes use of *BLAST* and applies a maximum clique approach to assemble pairwise *BLAST* homologs into clusters. In addition, it makes use of heuristic filters to improve alignment quality. *NcDNAlign* requires the user to define one genome as the reference and computes homologies of this to all other genomes in the study, while the approach presented in [7] makes an all-against-all comparison.

34.2.2

Results and Interpretation

In principle, the result of a comparative analysis will be a list of numerous ncRNA candidates that need to be verified by other computational methods, for example, search for homology to known ncRNAs present in databases such as *Rfam*, or by experimental methods such as Northern Blots. In publication [7], initially 443 sequence clusters were found. Six of these seemed to be related to repetitive elements, as they comprised 16, 42, 78, 83, 87, and 153 individual sequences, respectively, and were thus removed from the collection. Filtering for high-scoring (*RNAz* $P \geq 0.5$ or $Z \leq -2.0$) elements reduced the number to a final set of 109 clusters. The candidate list likely includes riboswitches, which can be identified by their location close to the 5' end of protein-coding sequences. Nevertheless, not all candidates directly upstream of protein-coding sequences need to be riboswitches. Another type of element, which may be represented quite frequently, is the Rho-independent terminator that often consists of a GC-rich hairpin followed by an oligo-U stretch. Such elements should not be ignored, because they may terminate the transcription of ncRNAs, thus facilitating the localization of ncRNA genes. As a rule of thumb, terminator elements directly downstream of protein-coding genes can be neglected, while free-standing ones are of interest.

For the analysis presented in publication [7], experimental verification and existing positive controls (e.g., matches to *Rfam*) directly validated 9 candidates among the top 52 candidates. Including the results of a later transcriptome analysis, the expression of a vast majority of predicted RNA elements was confirmed [8]. It is important to note that ncRNAs in bacteria are often involved in stress responses, thus they are frequently expressed only under special conditions. For this reason, experimental verification should include as many stress conditions as possible.

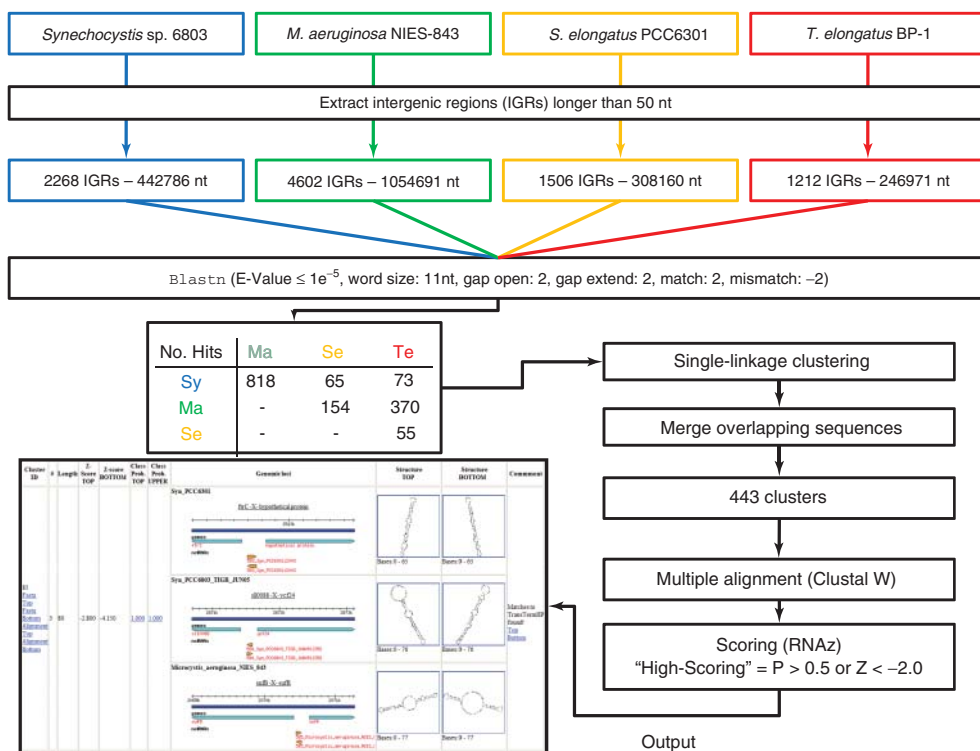


Figure 34.1 Pipeline for comparative ncRNA prediction (according to [7], modified). On the basis of the existing annotation, intergenic regions of at least 50 nt were searched for four cyanobacterial genomes, and their sequences extracted. The exact numbers and total lengths of IGRs for the individual genomes are given in the corresponding boxes. Next, these were locally aligned using blastn with the parameters shown in the box. For the blastn search, we used the smallest possible “word size” to obtain the highest sensitivity; values of 2 for the gap open and gap extend costs, and a mismatch score of -2 indicate that equal penalties were applied to all gap columns and mismatches. The match score of 2

allows one to compensate a single gap or mismatch with a single match. Sequences producing a significant pairwise blast alignment (e-value $\leq 10^{-5}$) were gathered into initial clusters. Whenever two such clusters contained at least one pair of overlapping sequences, such clusters were merged (“single-linkage clustering”). Note that such a cluster may now include sequences that initially did not appear as significant pairs. After merging overlapping sequences from the same IGR within each cluster (to eliminate sequence redundancies), the resulting clusters and their complementary clusters were aligned with ClustalW and scored using RNAz [2].

34.2.3

Alternative Approaches

The fact that transcription of ncRNAs is often terminated by Rho-independent terminators may be used as a key feature for their prediction. The web service

SIPHT (accessible under <http://newbio.cs.wisc.edu/sRNA/> [9]), for example, reduces the number of sequences of a genome for which homologs are found in other genomes by selecting only sequences in front of Rho-independent terminators predicted by TransTermHP [10]. This allows the scanning of all available prokaryotic genomes for conservation, although excluding transcriptional units for which different transcription termination mechanisms are utilized. Likewise, ncRNAs released from longer primary transcripts by processing events may escape detection. Nevertheless, the ease of use and its good performance make this approach a reasonable starting point when searching for ncRNAs in newly sequenced genomes. The results in publication [7] were also analyzed using TransTermHP to rule out that predicted ncRNAs are simply transcription terminator hairpins or to provide additional evidence for cluster expression. For 276 out of the initial 443 clusters and 68 out of the 109 high-scoring clusters, a terminator hairpin was predicted, including all validated ncRNA clusters.

34.2.4

Troubleshooting

34.2.4.1 Choice of Genomes

Of major importance for the success of a comparative genome-wide ncRNA prediction is the choice of genomes. Commonly, the relationship is inferred from a phylogenetic comparison of suitable marker sequences, such as the set of protein sequences of core genes shared by all genomes under study, or simply by comparing ribosomal RNA sequences. The maximum likelihood tree in Figure 34.2a shows the reconstructed phylogeny for a set of *Psychrobacter*, *Acinetobacter*, *Pseudomonas*, and *Shewanella* species (all γ -proteobacteria) on the basis of 16S ribosomal RNA sequences and using *Prochlorococcus marinus* MED4 as outgroup. But how does this relate to the number of intergenic elements conserved between the compared organisms? As a measure for this relation, we use the number of pairwise `blastn` hits (e-value cutoff $1e^{-5}$) between intergenic regions normalized to the mean of the cumulated lengths of intergenic regions of the two analyzed genomes, which we call hit frequency from now on. The normalization is needed due to the dependence of `blastn` hit numbers on the lengths of analyzed sequences. The numbers of pairwise `blastn` hits range from 2, in the case of *Pseudomonas aeruginosa* and *Psychrobacter cryohalolentis* (evolutionary distance: 0.170), to 32.224 in the case of *Psychrobacter arcticus* and *P. cryohalolentis* (evolutionary distance: 0.08). These pairs of organisms are both unsuitable: the first two are too distantly and the latter two are too closely related for a meaningful prediction. For the remaining pairs, we have analyzed the relationship of the hit frequency to (i) the pairwise differences in 16S rRNA, reflecting a very basic measure of evolutionary relationship, and (ii) to the more sophisticated evolutionary distance according to Jukes–Cantor. The same type of analysis was also performed for a set of cyanobacterial genomes, whose phylogenetic tree is given in Figure 34.2b. The results are presented in Figure 34.3 and show that for both data sets a weak correlation (coefficient of determination (R^2) 0.5848, 0.5867, 0.6561, and 0.6551, respectively) between 16S rRNA and the evolution of

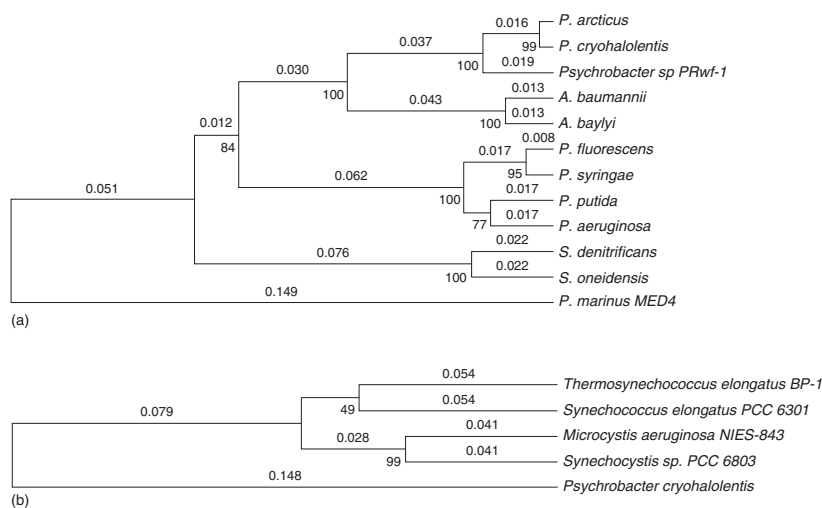


Figure 34.2 16S rRNA-based phylogenies. 16S rRNA sequences were obtained from the Ribosomal Database Project (RDP) and aligned using ClustalW. On the basis of these alignments, phylogenetic trees were estimated using the maximum likelihood method with the Jukes–Cantor substitution model and considering uniform mutation rates, as available in MEGA5 [11]. Statistical values are based on bootstrap analysis with 500 samples. (a) Phylogeny of a set of *Psychrobacter*, *Acinetobacter*, *Pseudomonas*, and *Shewanella* species using the cyanobacterium *Prochlorococcus marinus* MED4 as outgroup. (b) Phylogeny of a set of cyanobacterial species using the γ -proteobacterium *Psychrobacter cryohalensis* as outgroup.

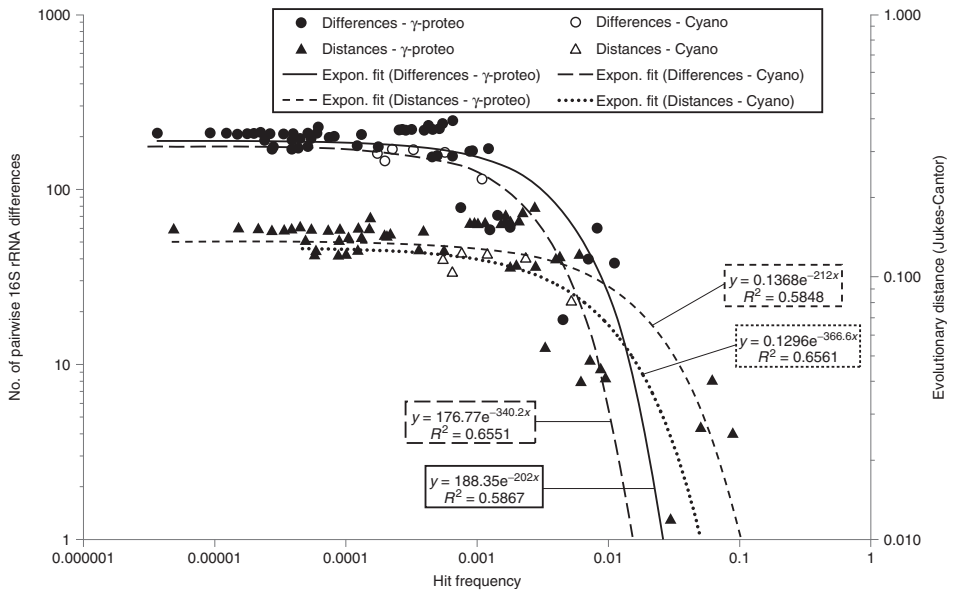


Figure 34.3 Relationship of 16S rRNA evolution and hit frequency. Relation of the number of pairwise 16S rRNA differences (left y-axis) and pairwise evolutionary distances (right y-axis, Jukes–Cantor, uniform rates; considers back mutations) to the hit frequency (ICR length-normalized number of pairwise blastn hits). Filled symbols correspond to the γ -proteobacterial dataset and open symbols to the cyanobacterial dataset.

intergenic regions can be identified. Interestingly, the strength of the correlation differs between the two data sets, as the fitted exponential functions differ in the coefficient of the exponent, that is, the fitted curves have different slopes. Altogether, the results show that the evolutionary distance assessed by comparative 16S rRNA analyses provides a rough estimate for choosing a “good” starting set of genomes. Comparative ncRNA predictions for several combinations of genomes from this starting set should be performed in order to find a (near-)optimal genome set.

34.2.4.2 Short mRNAs and Dual-Function RNAs

There are a few characterized cases of dual-function RNAs, such as RNAlIII in *Staphylococcus aureus* [12], SgrS in enteric bacteria [13], or SR1 in *Bacillus subtilis* [14], in which an RNA acts as a regulatory RNA and, in addition, encodes a (usually short) protein or peptide. Moreover, some of the predicted loci might give rise to short mRNAs that had remained unidentified during genome annotation. For these reasons, it is advisable to apply an additional processing step in which the coding potential of the predicted ncRNAs is analyzed. A suitable program is *RNAcode* [15] that differentiates between non-coding and possibly coding regions based on nucleotide substitution patterns in an alignment of at least three homologous sequences.

34.3

Experimental Approaches for High-Throughput RNomics in Bacteria

Numerous, frequently hundreds, of ncRNA candidates are obtained from a computational prediction. The prediction likely contains not only false positives but also many false negatives that were not identified because of their presence in only one organism (Section 34.2.4.1) or for other reasons. Eventually, any ncRNA candidate has to be verified experimentally, at least to prove its cellular expression. Experiments can be focused on a small number of candidates, which appear of special interest, for example, using Northern hybridizations or methods for the rapid amplification of cDNA ends. However, if the focus of a study is a global prediction of ncRNAs, this should be correlated with a global experimental transcriptome analysis, namely, microarrays (Section 34.3.1) or deep sequencing (Chapters 33 and 38). Such analyses will not only identify classical ncRNAs – those which act in *trans* on their mRNA (or protein) targets originating from loci usually far away in the genome – but also have the potential to further unveil *cis*-antisense RNAs and short gene-internal transcripts in sense orientation, representing two additional and much less characterized classes of ncRNAs. Experimental transcriptome analyses may be used in combination with or without computational predictions, depending on the method and scope of the chosen analysis.

34.3.1

Microarray Analysis

An advantage of microarrays arises from the fact that this technology is based on more than two decades of expertise with respect to microarray handling and data

analysis, including various technologies for their design and probe synthesis, that is, the synthesis or immobilization of oligonucleotide probes on the microarray surface and the selection of suitable sequences.

Most microarrays that are commercially available for model microorganisms cannot be used for the detection of ncRNAs because of their restriction to protein-coding sequences. However, there are several ways to design custom-tailored microarrays (Sections 34.3.1.1 and 34.3.1.2).

Another consideration is the high risk of measuring experimental artifacts that arise from unintended second strand synthesis when the samples for microarray analyses are generated by the classical cDNA synthesis labeling procedure [16]. This problem is potentially higher for short transcripts such as ncRNAs, which sometimes are detected by only a single oligonucleotide probe, whereas probe sets can be defined for longer mRNAs minimizing the impact of single probes. Moreover, the risk of unintended second strand synthesis renders the differentiation between “true” and “false” antisense transcripts impossible. For these reasons, we prefer direct labeling of the RNA to be hybridized, avoiding cDNA synthesis.

34.3.1.1 Considerations for the Design of Tiling Microarrays

There are several options for determining the length and distribution of probes, genome coverage, and the technical platform, for example, Agilent (see below) or Combimatrix (Figure 34.4). In any case, probe design should aim at generating a suitable set of oligonucleotide probes that are overlapping, are non-redundant, and have similar physicochemical properties. To fulfill the latter criterion, it is critical that all oligonucleotide/target duplexes have very similar stabilities (e.g., $T_m = 78 \pm 2^\circ\text{C}$) rather than having a uniform length. The redundancy of probe sequences within the genome and possibly existing plasmids should be checked and sequences with multiple perfect or 1-mismatch hits should be discarded. This approach can easily lead to several hundred thousand oligonucleotide probes.

As an example, for a *Synechocystis* PCC6803 tiling microarray [16], we chose sequences of 50 nt overlapped by 25–30 nt, with trimming of sequences to get a T_m as close to 72°C as possible (within a window of $70\text{--}74^\circ\text{C}$) and a minimum length of 25 nt. This procedure resulted in 102 739 probes, overlapping on an average by 28 nt. These probes were synthesized on a $2 \times 104\text{K}$ -Agilent custom microarray together with control probes detecting mouse actin sequences, with the limitation that this probe set covered only about 40% of the genome.

If the focus of the tiling microarray is solely on the detection of ncRNAs, then arrays carrying probes exclusively for non-annotated intergenic regions could be prepared. An example of such a microarray is shown in Figure 34.4. Although the array contained not a single probe for an annotated gene, a large number of transcript signals were detected. These included all *trans*-acting ncRNAs, and also the bulk of 5'- and 3'-flanking sequences of protein-coding transcripts. The two microarrays shown in Figure 34.4 report substantial and global changes in the expression levels of these “intergenic” regions upon induction of cold stress.

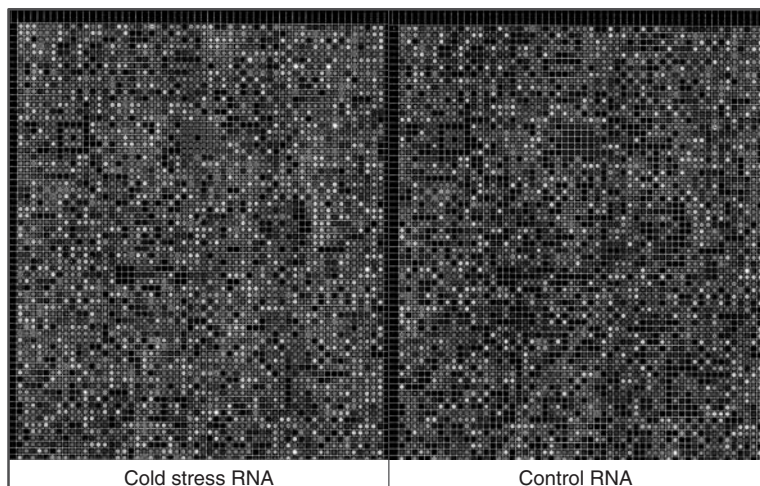


Figure 34.4 Scan of a hybridized “Inter-genic spacer only” microarray. Overlapping oligonucleotide probes covering all intergenic spacers in the genome of the cyanobacterium *Synechococcus* WH 7803 on both strands were designed and synthesized on

a Combimatrix custom microarray in 12 K format (tiling factor = 11). This array was hybridized with cDNA generated from total RNA of cells grown under cold stress and standard growth conditions [17].

34.3.1.2 Considerations for the Design of Expression Microarrays

If there is evidence for the presence of ncRNAs, for example, from a biocomputational prediction (Section 34.2) or inferred from an RNomics-type transcriptome analysis (see Protocols below), one might, as an alternative to whole genome tiling microarrays, consider to design more specialized microarrays that target specific regions. Such an expression microarray, not necessarily composed of overlapping probes, may include probe sets for all annotated genes and, additionally, for each genomic region whose expression has been predicted or verified by deep sequencing. The advantage is that this type of array is much more compact than a whole genome tiling microarray but can deliver almost as much information.

As an example, we chose for such an expression microarray targeting the cyanobacterium *Synechocystis* PCC6803 [18] on average three to five probes per transcript. These were directly designed using the Agilent eArray system (<https://earray.chem.agilent.com/earray/>). The design criteria were selected according to the “best distribution method,” with a T_m of 80 °C and a probe length between 45 and 60 nt, resulting in 20 293 probes. These probes were manufactured in a 44K Agilent custom microarray format with an internal duplication of all probes, hence providing an internal obligatory technical replicate.

34.3.1.3 Direct Labeling of RNA for Microarray Hybridization

For microarray hybridizations, total RNA was labeled directly (without cDNA synthesis) in 5 µg aliquots with the “ULS™ Fluorescent Labeling Kit for Agilent arrays (with Cy3 and Cy5)” (Kreatech Diagnostics, Cat. No. EA-021). In this protocol,

RNA is labeled non-enzymatically by binding of a platinum complex containing the fluorescent dye for coordinative binding to the N7 of guanines. Fragmentation and hybridization was performed following the manufacturer's instructions for Agilent one-color microarrays with 3–5.5 μg of labeled RNA. For the expression microarray, we directly labeled 2 μg RNA using the Cy3 labeling kit mentioned above. Hybridization was done with 1.5 μg RNA per array according to the Agilent protocol for $4 \times 44\text{K}$ single-color microarrays.

Protocols

Protocol 1: Labeling Reaction

- 1) Take 2 μg of DNA-free bacterial total RNA. Please note that the final concentration of the RNA in the labeling mixture must not be below $50 \text{ ng } \mu\text{l}^{-1}$.
- 2) Add 2 μl of Cy3-ULS per 2 μg RNA. Please note, when different amounts are required, the ratio of RNA in micrograms to ULS reagent in microliters should always be 1 : 1.
- 3) Add 2 μl of 10 x labeling solution (1/10th of the final volume) provided with the ULS™ Fluorescent Labeling Kit.
- 4) Adjust with RNase-free water to a final volume of 20 μl and mix.
- 5) Incubate for 15 min at 85°C (labeling reaction).
- 6) Place samples on ice, centrifuge briefly to collect the contents of the tube before proceeding with purification.

Protocol 2: Purification to Remove Unincorporated Label

- 1) Resuspend column material in KREApure columns by vortexing.
- 2) Loosen cap 1/4 turn, snap off the bottom closure, and place the column in a 2 ml collection tube.
- 3) Pre-spin the column for 1 min at 20 800 g (corresponds to 14 000 rpm on an Eppendorf 5417C centrifuge).
- 4) Discard flow-through and wash the column with 300 μl RNase-free water by centrifugation at 14 000 rpm for 1 min.
- 5) Discard collection tube and flow-through and place column in a new RNase-free 1.5 ml microcentrifuge tube.
- 6) Load the ULS-labeled RNA directly onto the center of the KREApure column bed; avoid pipetting on the sides of the column.
- 7) Spin column for 1 min at 20 800 g.
- 8) Collect flow-through containing the purified labeled RNA.

Protocol 3: Determination of Labeling Efficiency (Percentage of Nucleotides Labeled with the Dye)

- For Cy3-ULS-labeled RNA: Measure absorbance at 260 nm (A_{260}) and 550 nm (A_{550})
- For Cy5-ULS-labeled RNA: Measure A_{260} and A_{650}

$$\text{RNA in ng } \mu\text{l}^{-1} = \frac{A_{260} \times \text{dilution factor} \times 38}{\text{cuvette length (in cm)}}$$

$$\text{Dye in pmol } \mu\text{l}^{-1} = \frac{A_{550 \text{ or } 650 \text{ nm}} \times \text{dilution factor}}{\text{cuvette length (cm)} \times \epsilon_{\text{dye}} \times 10^{-6}}$$

$$\epsilon_{\text{dye}} \text{ at } 550 \text{ nm for the Cy3 reagent} = 150\,000 \text{ M}^{-1} \times \text{cm}^{-1}$$

$$\epsilon_{\text{dye}} \text{ at } 650 \text{ nm for the Cy5 reagent} = 250\,000 \text{ M}^{-1} \times \text{cm}^{-1}$$

$$\text{Labeling percentage} = \frac{0.34 \times \text{pmol dye}}{\text{ng}_{\text{RNA}}} \times 100\%$$

The factor of 0.34 derives from the fact that a ribonucleotide monophosphate has an average molecular mass of 340, thus $0.34 \text{ ng pmol}^{-1}$.

Protocol 4: Fragmentation

To decrease the fragment size to 60–200 nt, the “RNA Fragmentation Reagents” from Life Technologies Ambion (Cat. No. AMB8740) were used.

- 1) Transfer the mixture to a microcentrifuge tube and add 1/10 vol of 10 x fragmentation buffer (Ambion).
- 2) Incubate at 70 °C for exactly 15 min.
- 3) Spin the vial briefly and add 1 μl of stop solution (Ambion), mix by pipetting (the labeled RNA can form aggregates that dissolve by pipetting), place on ice, and use for microarray hybridization as soon as possible.

Protocol 5: Hybridization of a One-Color Agilent 4 x 44K Microarray

Using the components of Agilent’s Hi-RPM Gene Expression Hybridization Kit (Cat. no. 5190-0404).

- 1) Load clean gasket slide into chamber base with label facing up. Slowly dispense 100 μl of sample onto the gasket well in a “drag and dispense” manner, take care not to wet the gasket as this may cause leakage.
- 2) Carefully place an array active side down onto the gasket slide so that the numeric barcode is facing up, ensure that the slides are properly aligned. Place the chamber cover over the slides and secure with the clamp assembly. Vertically rotate the assembled chamber to wet the gasket and check the mobility of bubbles (if required tap the assembly on a hard surface to move stationary bubbles).
- 3) Load the slide chamber into the rotisserie of a hybridization oven set to 65 °C (Agilent model G2545A Hybridization Oven, manufactured by Sheldon Manufacturing, Inc., Cornelius, OR 97113, USA). Set the rotational speed to 10 rpm and hybridize at 65 °C for 17 h. Place Wash Buffer 2 (see below) into a 37 °C water bath to preheat overnight.

Protocol 6: Washing of the Microarray Slide

Using Agilent’s Gene Expression Wash Buffer Kit (Cat. no. 5188-5327).

- 1) Clean three washing dishes, a slide rack and a magnetic stirring bar in three changes of RNase-free water. Use Wash Buffer 1 to fill half dish 1. Place stirring bar and slide rack in dish 2 and fill to two-third with Wash Buffer 1.
- 2) Remove the hybridization chamber from the oven and inspect the slide to confirm that the bubble in each array slot was able to rotate freely and that no leakage occurred. Release the slide sandwich from the hybridization chamber, submerge both slides into dish 1, and use plastic forceps to separate the array from the gasket slide. Transfer the array slide to the rack in dish 2 ensuring that it is covered fully by Wash Buffer 1 and stir at 600 U min^{-1} for 1 min.
- 3) Pour prewarmed Wash buffer 2 (37°C) into dish 3 until it is filled to two-thirds. Transfer the slide rack (with array slide) and the stirring bar to dish 3 and stir at 600 U min^{-1} for 1 min.
- 4) Slowly remove the array slide from Wash buffer 2 (should take 5–10 s) to minimize droplets on the slide. Gently touch an absorbent paper with one edge of the slide to remove any excess liquid. The slide can be scanned immediately or stored in a plastic slide box in the dark.

34.4

Troubleshooting

- **Inefficient labeling reaction:** The RNA samples need to be free of divalent cations (e.g., Mg^{2+}) and other salts and chemicals such as guanidinium compounds that can disturb the labeling reaction. Possible other reasons for inefficient labeling include an incorrect ratio of labeling reagent to RNA or that the RNA concentration in the labeling reaction was $< 50 \text{ ng } \mu\text{l}^{-1}$. To correct, ensure a ratio of 1 : 1 between RNA in micrograms to ULS reagent in microliters; make sure that the RNA concentration in the labeling reaction is above $50 \text{ ng } \mu\text{l}^{-1}$; before labeling, subject the RNA to an additional ethanol precipitation and 80% ethanol wash step.
- **Background on slide:** The most frequent reasons for background on microarray slides include an excess amount of sample, insufficient blocking and partial drying during hybridization due to insufficient amounts of liquid in the hybridization vessel. To overcome these problems, add more or alternative blockers to prehybridization or hybridization buffers, for example, KREATECH's *KREAblock* solution should be added to one-fourth of the final volume (e.g., $25 \mu\text{l}$ of *KREAblock* in a total volume of $100 \mu\text{l}$) of the hybridization mixture; ensure that the correct amount of labeled RNA is used, that sufficient hybridization solution is added to the hybridization chamber, and that the vessel is sealed tightly.

Acknowledgments

We gratefully acknowledge funding from the German Federal Ministry of Education and Research for support (BMBF grant 0313921 to the “Freiburg Initiative in Systems Biology”), the German Research Foundation (DFG) through the Priority program SPP1258 “Sensory and regulatory RNAs in Prokaryotes” (grant HE 2544/4-2), and grant VO 1450/2-1.

References

1. Rivas, E. (2005) Evolutionary models for insertions and deletions in a probabilistic modeling framework. *BMC Bioinformatics*, **6**, 63.
2. Washietl, S., Hofacker, I.L., and Stadler, P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 2454–2459.
3. Babak, T., Blencowe, B.J., and Hughes, T.R. (2007) Considerations in the identification of functional RNA structural elements in genomic alignments. *BMC Bioinformatics*, **8**, 33.
4. Backofen, R. and Hess, W.R. (2010) Computational prediction of sRNAs and their targets in bacteria. *RNA Biol.*, **7**, 33–42.
5. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
6. Rose, D., Hertel, J., Reiche, K., Stadler, P.F., and Hackermüller, J. (2008) NcDNAAlign: plausible multiple alignments of non-protein-coding genomic sequences. *Genomics*, **92**, 65–74.
7. Voss, B., Georg, J., Schön, V., Ude, S., and Hess, W.R. (2009) Biocomputational prediction of non-coding RNAs in model cyanobacteria. *BMC Genomics*, **10**, 123.
8. Mitschke, J., Georg, J., Scholz, I., Sharma, C.M., Dienst, D., Bantscheff, J., Voss, B., Steglich, C., Wilde, A., Vogel, J., and Hess, W.R. (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 2124–2129.
9. Livny, J., Teonadi, H., Livny, M., and Waldor, M.K. (2008) High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. *PLoS ONE*, **3**, e3197.
10. Kingsford, C.L., Ayanbule, K., and Salzberg, S.L. (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.*, **8**, R22.
11. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, **28**, 2731–2739.
12. Novick, R.P., Ross, H.F., Projan, S.J., Kornblum, J., Kreiswirth, B., and Moghazeh, S. (1993) Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J.*, **12**, 3967–3975.
13. Wadler, C.S. and Vanderpool, C.K. (2007) A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 20454–20459.
14. Gimpel, M., Heidrich, N., Mader, U., Krügel, H., and Brantl, S. (2010) A dual-function sRNA from *B. subtilis*: SR1 acts as a peptide encoding mRNA on the *gapA* operon. *Mol. Microbiol.*, **76**, 990–1009.
15. Washietl, S., Findeiss, S., Müller, S.A., Kalkhof, S., von Bergen, M., Hofacker, I.L., Stadler, P.F., and Goldman, N. (2011) RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA*, **17**, 578–594.
16. Perocchi, F., Xu, Z., Clauder-Münster, S., and Steinmetz, L.M. (2007) Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res.*, **35**, e128.
17. Gierga, G., Voss, B., and Hess, W.R. (2012) Non-coding RNAs in marine *Synechococcus* and their regulation under environmentally relevant stress conditions. *ISME J.*, **6**, 1544–1557.
18. Georg, J., Voss, B., Scholz, I., Mitschke, J., Wilde, A., and Hess, W.R. (2009) Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Mol. Syst. Biol.*, **5**, 305.

35

Experimental RNomics, a Global Approach to Identify Non-coding RNAs in Model Organisms, and RNPomics to Analyze the Non-coding RNP Transcriptome

Mathieu Rederstorff and Alexander Hüttenhofer

35.1

Introduction

Non-coding RNAs (ncRNAs) do not encode proteins but function directly at the level of RNA in the cell. Over the past ~10 years, the importance of this surprisingly diverse class of molecules has been widely recognized [1–4]. They play key roles in a variety of fundamental processes in all three domains of life, that is, eukarya, bacteria, and archaea. ncRNAs have functions in DNA replication and chromosome maintenance, regulation of transcription, RNA processing (not only RNA cleavage and ligation but also RNA modification and editing), translation and stability of mRNAs, and even regulation of stability and translocation of proteins [4–15]. Many of them have been discovered fortuitously, suggesting that they merely represent the tip of the iceberg. The vast majority of ncRNAs are relatively small, much shorter than most mRNAs, hence their generic denomination as small non-messenger or non-coding RNAs.

Recent reports by the ENCODE project, focusing in high resolution on the analysis of about 1% of the human genome, have shown that up to 90% of the genome is being transcribed [16], with only a minor portion of RNA transcripts (1.5%) coding for proteins. Hence, it was suggested that the remaining 88.5% might serve as a source of regulatory ncRNAs. These findings implied the presence of a, so far, “hidden layer” of regulatory elements within the human and other eukaryal genomes, represented by ncRNA species [17], with more than 450 000 ncRNAs genes predicted to be encoded by the human genome [18]. However, it has also been argued that many of the ncRNA transcripts from the human or other higher eukaryal genomes merely represent spurious non-functional transcription products [19, 20]. Therefore, it is of utmost importance to elucidate which of these RNA transcripts represent functional ncRNAs.

Systematic searches for ncRNAs, dubbed experimental RNomics, have been initiated in various model organisms by screening cDNA libraries constructed from sized RNA fractions [21–23]. First, we describe procedures for the construction and screening of a general purpose cDNA library from sized RNA fractions. Next, to illustrate a more specialized RNomics strategy, we use as a paradigm one of

the two large families of small nucleolar RNAs (snoRNAs), both associated with a small set of specific proteins. The two snoRNA subtypes, termed *C/D* and *H/ACA* snoRNAs, guide two prevalent types of nucleotide modification, 2'-O-methylation and pseudouridylation, respectively, by forming a specific duplex around each target RNA modification site. Cellular RNAs targeted by guide snoRNAs include rRNAs in Eukarya and Archaea, small nuclear RNAs and archaeal tRNAs, and likely a range of still unidentified RNAs [4, 6, 8].

In eukarya, most if not all known ncRNAs are associated with RNA binding proteins, thus forming ribonucleoprotein particles or RNPs [24]. Prominent examples of these RNPs are represented by the classes of miRNPs or the abovementioned snoRNPs [6, 25, 26]. Therefore, identification of functional, and thus biologically relevant, ncRNAs can be achieved by isolating such RNAs as part of complexes with proteins, termed *ncRNPs*.

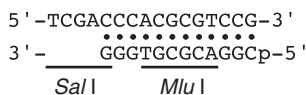
As an alternative to the library construction protocol detailed in the first part of this chapter, we describe in the second part procedures for the generation of a cDNA library derived from RNPs that were size-selected on glycerol gradients, thereby increasing the proportion of functional species in the library [27, 28].

All libraries described here can be subsequently analyzed by high-throughput sequencing techniques, provided the proper adapters are used, as outlined in the second part of the chapter. However, for some applications, libraries can also be cloned and sequenced by traditional Sanger sequencing, as described in the first part of the chapter.

Materials for Experimental RNomics

Oligonucleotide Primers

- Primer-adapter: 5'-TCG CGA GCG GCC GCG GGG GGG GGG GGG GG-3' (*Not* I site underlined)
- *Sal* I adapter: provided with the SuperScript™ Plasmid System; see Section "Reagents, Kits, Vectors, and Bacterial Cells"



- M13 fsp: 5'-GCT ATT ACG CCA GCT GGC GAA AGG GGG ATG TG-3'
- M13 rsp: 5'-CCC CAG GCT TTA CAC TTT ATG CTT CCG GCT CG-3'
- Primer 1: 5'-ATA AAG CGG CCG CGG ATC CAA-3' (RNA residue underlined)
- Primer 2: 5'-TTG GAT CCG CGG CCG CTT TAT TNN NNT CAG-3' (*Bam* HI site underlined)
- Primer 3: 5'-AAT AAA GCG GCC GCG GAT CCA ANN NNN RTG ATG A-3' (*Bam* HI site underlined; N = A, C, G or T; R = A or G)

Enzymes

- A-Plus™ poly(A) Polymerase (Epicentre Biotechnologies)
- T4 polynucleotide kinase (New England Biolabs)
- T4 RNA ligase (New England Biolabs)
- *Taq* DNA Polymerase I (Life Technologies Invitrogen)
- Other enzymes used but not listed here were supplied with the “SuperScript™ Plasmid System for cDNA Synthesis and Plasmid Cloning” kit (Life Technologies Invitrogen)

Buffers

For all protocols described in this chapter, use DEPC-treated RNase-free water to dissolve RNA pellets after ethanol precipitation. To prepare DEPC-treated RNase-free water, mix double-distilled water with 0.01% (v/v) DEPC in RNase-free bottles. Let it stand overnight and then autoclave. RNase-free water can be stored at room temperature (20 °C) for several months. DEPC-treated RNase-free water should be used to prepare RNase-free solutions, for example, 70% ethanol to wash RNA pellets after precipitation.

- Electrophoresis buffer: 1 x TBE (90 mM Tris-HCl, pH 8.3, 90 mM boric acid, 2.5 mM EDTA)
- 1 x tailing buffer: 50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 10 mM MgCl₂, 2 mM MnCl₂, 0.4 mM EDTA, 1 mM DTT, 2 mM CTP
- 5 x first strand buffer: 250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM MgCl₂
- 5 x second strand buffer: 100 mM Tris-HCl, pH 6.9, 450 mM KCl, 23 mM MgCl₂, 0.75 mM β-NAD⁺, 50 mM (NH₄)₂SO₄
- 5 x T4 DNA ligase buffer: 250 mM Tris-HCl, pH 7.6, 50 mM MgCl₂, 5 mM DTT, 5 mM ATP, 25% (w/v) PEG 8000
- 10 x *Not* I buffer: 500 mM Tris-HCl, pH 8.0, 100 mM MgCl₂, 1 M NaCl
- Hybridization buffer 1: 0.5 M sodium phosphate, pH 7.2, 7% SDS, 1 mM EDTA
- Hybridization buffer 2: 1 M sodium phosphate buffer, pH 6.2, 7% SDS
- Prehybridization buffer = Hybridization buffer 2
- Wash buffer 1: 2 x SSC (20 mM sodium phosphate, pH 7.4, 0.3 M NaCl, 2 mM EDTA), 0.1% SDS
- Wash buffer 2: 0.1 x SSC (1 mM sodium phosphate, pH 7.4, 15 mM NaCl, 0.1 mM EDTA), 0.5% SDS
- 1 x RNA ligation buffer: 50 mM Tris-HCl, pH 7.8, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP
- 1 x kinase buffer: 70 mM Tris-HCl, pH 7.6, 10 mM MgCl₂, 5 mM DTT, 1 mM ATP
- 1 x *Taq* I buffer: 10 mM Tris-HCl, pH 9.0, 50 mM KCl, 0.1% Triton X-100
- NET-150 buffer: 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.05% Nonidet P-40.

Reagents, Kits, Vectors, and Bacterial Cells

- TRIzol® Reagent (Life Technologies Invitrogen)

- Phenol/chloroform: water-saturated phenol/chloroform (1 + 1)
- cDNA synthesis and cloning kit: SuperScript® Plasmid System for cDNA Synthesis and Plasmid Cloning with Gateway™ Technology (Life Technologies Invitrogen; pSPORT 1 vector included)
- *Escherichia coli* strain DH10B™ competent cells (Life Technologies Invitrogen)
- BigDye terminator cycle sequencing reaction kit (Life Technologies Applied Biosystems)
- Protein A-Sepharose (PAS) (Sigma–Aldrich)
- Glycogen, RNA grade (Thermo Scientific): 20 mg ml⁻¹
- [γ -^{32/33}P]ATP
- Positively charged nylon membranes “Zeta-Probe” (BIO-RAD) or Amersham Hybond-N+ (GE Healthcare Life Sciences)
- pKS vector as part of the “pBluescript® II KS(+) Phagemid Kit” (Agilent Technologies)

Protocols for Library Construction and Analysis

The methods described below outline the construction and screening of a cDNA library from sized RNA fractions (Protocol 1), and the construction of a more specialized cDNA library (Protocol 2), either starting from an RNA pool enriched in the ncRNA subtype of interest and/or using specialized oligonucleotide primers for preferential cDNA synthesis of this ncRNA subtype.

As for the construction of the two different cDNA libraries (Protocols 1 and 2, respectively), two different methods for reverse transcription of ncRNAs into cDNAs are applied (Figure 35.1). In the first method (Protocol 1), 3'-tailing of ncRNAs by CTP and poly(A) polymerase is followed by reverse transcription of ncRNAs by an oligo-d(G) primer (see below). Second strand synthesis is initiated at internal sites *via* RNase-H-generated short RNA primers that are extended by DNA polymerase I (not shown in Figure 35.1). As an alternative to the C-tailing approach, we also used a “linker-ligation” method for cloning small RNA molecules (Protocol 2, see Figure 35.1, right path): both ends of ncRNAs are joined to a short 5'-phosphorylated DNA or RNA linker by T4 RNA ligase. Subsequently, ligated RNAs are transcribed into cDNA by RT-PCR using primers derived from the linker sequence. A potential disadvantage of this method is, however, the possibility of oligomerization of linker oligonucleotides and the limitation of linker attachment to RNAs with 5'-monophosphate ends, excluding primary transcripts with 5'-triphosphate termini or 5'-cap structures. The advantage of this compared with the C-tailing method is that full-length cDNA clones can be obtained whereas in the C-tailing method, the second strand synthesis by the RNase H/DNA polymerase I approach results in truncated 5' termini.

Protocol 1: cDNA Library Construction and Screening

This section describes the isolation of total RNA from the organism of interest, the purification and size selection, and the reverse transcription of ncRNAs into cDNA (Figure 35.1). Since ncRNAs do not contain a poly(A) tail (as do mRNAs), ncRNAs

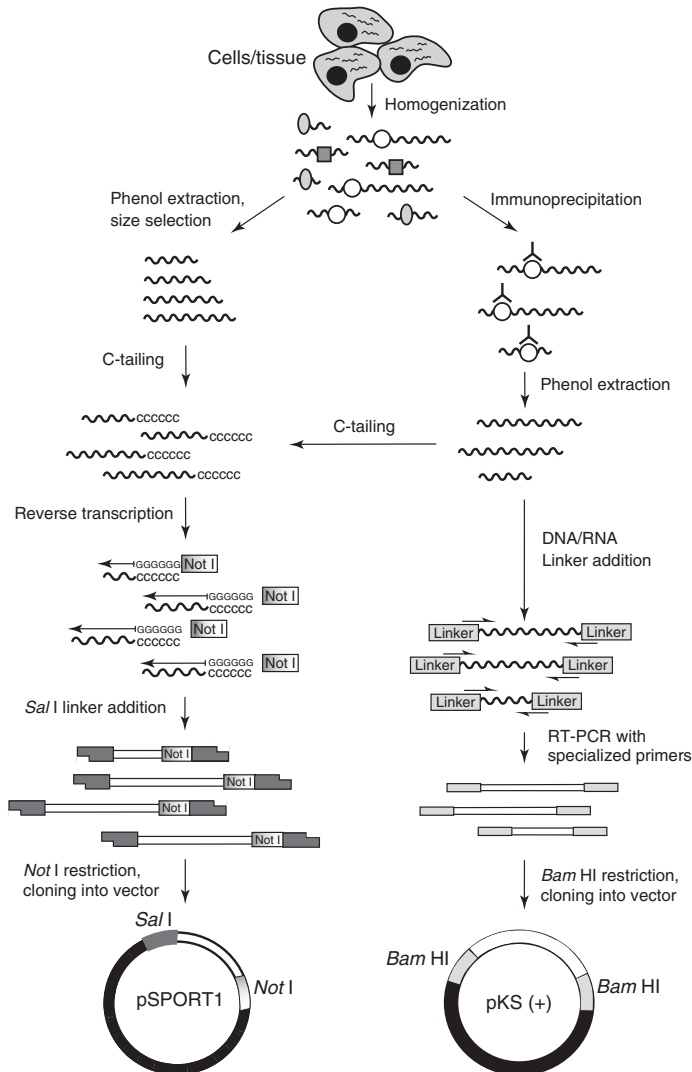


Figure 35.1 Construction of an unbiased, general purpose cDNA library (upper left) of ncRNAs or a more specialized library encoding RNAs of a certain ncRNA subclass, as for example, snoRNAs (upper right). Here, ncRNA subclass enrichment was achieved by immunoprecipitation of a cell lysate with an antibody directed against an RNA binding protein specific to this subclass (indicated by open circles at the top; gray ovals and squares represent other RNA binding proteins). Two different ways to generate

a cDNA library are illustrated in the bottom part. For the general purpose cDNA library, the C-tailing approach is the method of choice (left path; for second strand synthesis not illustrated here, see text); for the specialized cDNA library, a “linker-ligation” protocol (right path) represents a preferable alternative in cases where primers specific to the RNA subclass of interest can be used in RT-PCR reactions (for details, see Figure 35.2).

are, in the first step, tailed with CTP and poly(A) polymerase (which also uses CTP as a substrate). After the addition of C-tails, ncRNAs are reverse-transcribed employing an oligo-d(G) primer and the SuperScript™ II Reverse Transcriptase. This procedure also tags small poly(A)-tailed messenger RNAs that cofractionate with the ncRNA pool, but their cDNAs will be identified by the feature of an A-tail followed by a C-tail.

Protocol 1.1: Construction of the cDNA Library Encoding ncRNAs

- 1) Freeze 1–2 mg of tissue in liquid nitrogen, and prepare total RNA with the TRIzol reagent (Life Technologies Invitrogen) according to the protocol provided by the manufacturer.
- 2) Fractionate about 200 µg of total RNA on a denaturing 8% polyacrylamide gel (7 M urea, 1 x TBE buffer); run the gel (20 cm wide, 30 cm long, 2 mm thick) at 400 V for 1.5 h.
- 3) Excise RNA from the gel corresponding to the desired size range (for example, 50–500 nt) and passively elute the RNA by vigorous shaking in 0.3 M sodium acetate (pH 5.2), 0.1% SDS, 0.2% phenol at 4 °C overnight.
- 4) To the eluate, add an equal volume of phenol/chloroform to extract the RNA by brief vortexing.
- 5) After centrifugation (16 000 g, 10 min at 4 °C), recover the aqueous phase, add 40 µl 3 M sodium acetate (pH 5.3) and 1 ml of cold 100% ethanol, and chill at –20 °C for 30 min. Centrifuge at 4 °C and 16 000 g for 30 min; wash the pellet with 1 ml 70% ethanol, briefly centrifuge and redissolve the air-dried pellet in 10 µl RNase-free water. Store at –20 °C.
- 6) Tail 5 µg of RNA with CTP using poly(A) polymerase. The C-tailing reaction is carried out in a volume of 50 µl containing 1 x tailing buffer and 1.5 U of poly(A) polymerase. Incubate the reaction mix at 37 °C for 1.5 h. After incubation, add 20 µl 3 M sodium acetate (pH 5.2) and 130 µl RNase-free water. Extract the RNA with phenol/chloroform, precipitate with ethanol, and dissolve in 5 µl of RNase-free H₂O.
- 7) First and second cDNA strand synthesis: All enzymes and reagents, the *Sal* I adapter, and the pSPORT1 vector used in steps 5–7 are supplied with the Life Technologies Invitrogen SuperScript™ Plasmid System (see Section “Reagents, Kits, Vectors, and Bacterial Cells”). However, it is important not to use the *Not* I primer-adapter included in this kit since it contains an oligo-d(T) stretch suitable for reverse transcription of mRNAs only, but to use the primer-adapter oligonucleotide (custom-made, Section “Oligonucleotide Primers”) instead. Before first strand cDNA synthesis, combine 5 µl C-tailed RNA (obtained in step 4) with 2 µl primer-adapter oligonucleotide (0.5 µg µl^{–1}). Denature the mixture at 90 °C for 1 min, immediately transfer to 37 °C, and incubate for 10 min. Subsequently, add 4 µl of 5 x first strand buffer, 2 µl of 0.1 M DTT, 1 µl of 10 mM dNTP mix, and 1 µl of H₂O, and incubate at 37 °C for 2 min. Start first strand synthesis by addition of 5 µl of SuperScript™ II (AMV) Reverse Transcriptase (200 U µl^{–1}) and let the reaction proceed for 1.5 h at 45 °C. For

second strand DNA synthesis, add to the first strand cDNA mix 30 μl of 5 x second strand buffer, 1 μl of *E. coli* DNA ligase (10 U μl^{-1}), 4 μl of *E. coli* DNA polymerase I (10 U μl^{-1}), 1 μl of *E. coli* RNase H (2 U μl^{-1}), 3 μl of 10 mM dNTPs, and 91 μl H₂O to give a final volume of 150 μl . Gently mix and incubate at 16 °C for 2 h. Add 2 μl of T4 DNA polymerase (5 U μl^{-1}) to the reaction mix, continue incubation at 16 °C for another 5 min, and place the mixture on ice for 5 min. Stop the reaction by adding 10 μl of 0.5 M EDTA (pH 8.0), followed by one phenol/chloroform extraction and ethanol precipitation. Subsequently, dissolve the pellet in 25 μl of RNase-free H₂O.

- 8) Addition of *Sal* I adapters: Before cloning into the pSPORT1 vector, add *Sal* I adapters to the 5' and 3' ends of cDNAs as follows: combine the 25 μl of cDNA from step 7 with 10 μl of 5 x T4 DNA ligase buffer, 10 μl of *Sal* I adapter, and 5 μl of T4 DNA ligase (1 U μl^{-1}), resulting in a total volume of 50 μl . Incubate at 16 °C for 16 h. Extract the DNA with phenol/chloroform, precipitate with ethanol and dissolve in 41 μl of RNase-free H₂O. Subsequently, digest the DNA with *Not* I in a total volume of 50 μl containing 5 μl of 10 x *Not* I buffer and 4 μl *Not* I (10 U μl^{-1}) for 2 h at 37 °C. The restriction fragments are separated on a native 6% polyacrylamide/1 x TBE gel. The cDNA is visualized by staining with ethidium bromide, subsequently excised from the gel, and passively eluted in 0.3 M sodium acetate buffer (pH 7.2) at 4 °C overnight. The DNA is extracted with phenol/chloroform, precipitated with ethanol, and dissolved in 15 μl of RNase-free H₂O.
- 9) Cloning of cDNAs into the pSPORT 1 vector employing the Life Technologies Invitrogen SuperScript™ Plasmid System: Of the 15 μl of *Not* I-digested cDNA prepared in step 8, use 1.5 μl (about 100 ng) for the ligation reaction containing 1 x T4 DNA ligase buffer, 50 ng pSPORT1 vector, and 1 U T4 DNA ligase in a total volume of 20 μl ; incubate at room temperature for 3 h. Extract the ligation mixes with phenol/chloroform, precipitate the DNA with ethanol, and redissolve it in 10 μl H₂O; 3 μl are used for transformation of *E. coli* strain DH10B™ competent cells.
- 10) Amplify cDNA inserts from individual plasmid clones by PCR using primers M13 fsp and M13 rsp (see Section “Oligonucleotide Primers”) and spot PCR products on filters, if available with robots in high-density arrays [29].

Protocol 1.2: Analysis of cDNA Clones: Sequencing and Sequence Analysis

- 1) Sequence a subset of cDNA clones using the M13 rsp reverse primer and, for example, the BigDye terminator cycle sequencing reaction kit on an ABI Prism 3700 (Life Technologies Applied Biosystems) sequenator (see also Section 35.3, Note 3).
- 2) On the basis of sequence data obtained in step 1, design oligonucleotide probes to identify clones of abundant known ncRNAs, such as fragments of ribosomal RNA, by filter hybridization screening (see Protocol 1.3). Such clones are excluded from the second round of sequencing.

- 3) Then sequence the remaining cDNA clones and compare them to each other using, for example, the Lasergene Seqman II program package to identify clones with identical or overlapping cDNA sequences.
- 4) Search the GenBank database (NCBI) for all novel sequences using the BLASTN program. All sequences that have not been previously annotated in this database can be treated as potential candidates for novel ncRNAs.

Protocol 1.3: Exclusion of cDNA Clones for the Most Abundant, Known RNA Species

- 1) End-label DNA oligonucleotides (20–30 nt in length) derived from abundant sequences identified in Protocol 1.2, step 1, with [γ - ^{33}P]ATP and T4 polynucleotide kinase using standard molecular biology techniques [30].
- 2) Hybridize the labeled oligonucleotides to cDNA arrays spotted on filters (see above). Before hybridization, pretreat the membrane by rinsing twice with prehybridization buffer. Dot-blot hybridization is performed in hybridization buffer 1 at 53 °C for 12 h.
- 3) Wash the filters twice for 15 min in wash buffer 1 at room temperature, followed by another washing step in wash buffer 2 for 1 min at 53 °C.
- 4) Expose a phosphorimaging screen, and rank intensities of hybridization signals by computer-aided analysis as described [31].

Protocol 1.4: Verification of the Expression of Cloned ncRNAs by Northern Blot Analysis

- 1) Fractionate total RNA prepared from the organism or tissue of interest using the TRIzol method described above on an 8% denaturing polyacrylamide gel (7 M urea, 1 x TBE buffer).
- 2) Transfer onto nylon membranes, for example, using a Bio-Rad semi-dry blotting apparatus (Trans-blot SD, Bio-Rad, Germany), and immobilize the RNAs (e.g., for 1.5 min in a Stratagene UV crosslinker at 1200 J cm⁻²).
- 3) Incubate nylon membranes for 1 h in prehybridization buffer.
- 4) End-label 22–24 nt long oligonucleotides complementary to potential novel RNA species with [γ - ^{32}P]ATP and T4 polynucleotide kinase according to standard procedures and hybridize the nylon membranes for 12 h at 58 °C in hybridization buffer 2.
- 5) Wash blots twice at room temperature for 15 min in wash buffer 1, and subsequently at 58 °C for 1 min in wash buffer 2.
- 6) Expose Kodak MS-1 film to membranes for 1 h to 5 days or use a phosphor-imager.

Protocol 2: Construction of a Specialized cDNA Library

A specialized cDNA library is generated by using RNA isolated from a defined subcellular fraction [32], or RNA enriched in the ncRNA type of interest because of prefractionation of a total cellular extract (Figure 35.1, Protocol 2.1). Alternatively,

or in addition, enrichment can be achieved at the level of cDNA synthesis (see Protocol 2.2).

Protocol 2.1: Using an RNA Sample Enriched through Immunoprecipitation

The approach is illustrated for a cDNA library specific to C/D snoRNAs. Enrichment was achieved by using antibodies against fibrillarin, a protein that specifically binds all members of this large ncRNA family. The monoclonal antibody 72B9 (provided by M. Pollard) has been successfully employed for this purpose.

- 1) Rapidly cut freshly prepared or frozen tissues (stored at -80°C) into small pieces using a sterile cutter, and resuspend in 8 ml of cold NET-150 buffer.
- 2) Homogenize the cellular sample, for example, by using a Homogenizer One Shot Model (Constant Systems Ltd.) at 2.0 kbar, and clarify the extract by centrifugation (12 000 g, 10 min, at 4°C).
- 3) Immobilization of IgG antibodies on sepharose beads: In siliconized RNase-free Eppendorf tubes, incubate 10–50 μl of antibody solution (here a supernatant of a monoclonal culture producing antibody 72B9) with 2.5 mg of swollen Protein A-Sepharose (PAS) in 0.5 ml NET-150 buffer for 2 h at room temperature with gentle agitation. Pellet the immunoglobuline-coated PAS (PAS-Ig) matrix by brief centrifugation (at 500 g for 20 s) and wash the beads twice with 1 ml NET-150 buffer before resuspension in 0.5 ml NET-150.
- 4) Add 0.5 ml of cell extract supernatant from step 2 to 0.5 ml of resuspended PAS-Ig material. After 60 min at 4°C with gentle agitation, pellet, and wash the PAS-Ig material eight times as in step 3.
- 5) To the washed pellet, add 300 μl NET-150, 1 μl glycogen (20 mg ml^{-1}), 30 μl SDS (10%), and 300 μl phenol/chloroform. Incubate for 15 min at 37°C with occasional vortexing.
- 6) Precipitate with ethanol (for details, see Protocol 1.1, step 5)
- 7) Then proceed as described in Protocol 1.1 (starting from step 6) or in Protocol 2.2.

Protocol 2.2: Using an RT-PCR Procedure with Specialized Primers

Extracted RNAs can be oligonucleotide-tagged at both ends by T4 RNA ligase and amplified by RT-PCR [32]. Note that the reaction catalyzed by T4 RNA ligase requires RNA substrates with 5' monophosphate and 3' OH termini. In a variation of this approach, the RT-PCR primers are designed in such a way as to preferentially amplify RNA molecules of the selected ncRNA subtype. This can be performed whenever the ncRNA type of interest exhibits a pair of sequence signatures at a short, fixed distance from the 5' and 3' termini. This is illustrated below for the C/D snoRNAs, which contain the box C and D motifs only a few nucleotides away from their 5' and 3' termini, respectively [33]. Oligonucleotide-tagged RNA molecules are then subjected to RT-PCR using a pair of primers matching the tag as well as the adjacent terminal nucleotides of a typical C/D snoRNA specimen (Figure 35.2). The 3'-terminal sequence of both primers, termed 2 and 3, matches the conserved

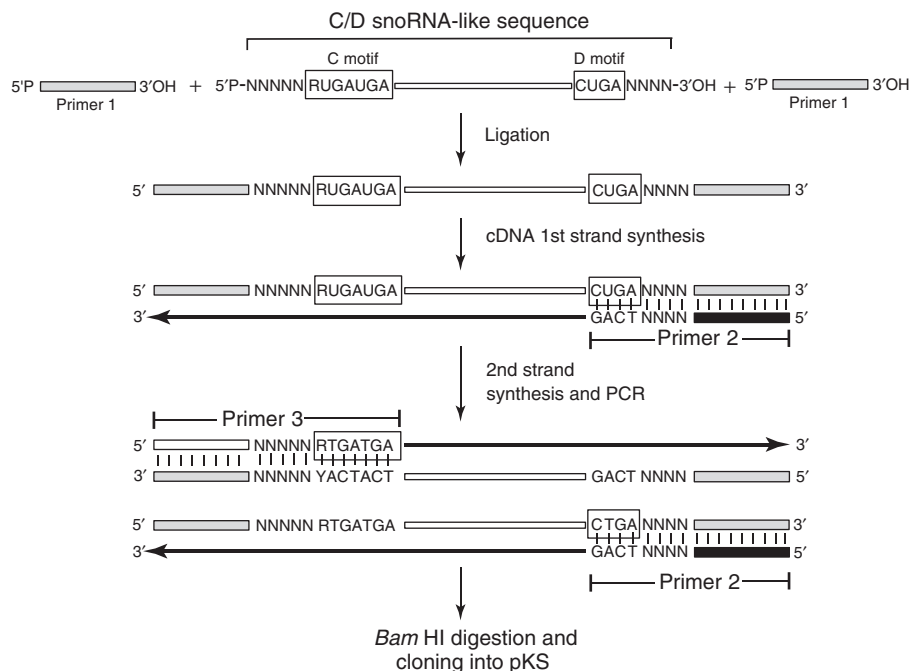


Figure 35.2 Preferential amplification of a specific RNA subset by an RT-PCR procedure with specialized primers. C/D snoRNAs immunoprecipitated by the anti-fibrillarin antibody are tagged with the same oligonucleotide at both ends. They are then amplified by RT-PCR using primers 2 and 3, which are designed to amplify only RNA molecules

with the prevalent terminal sequences of C/D snoRNAs, that is, RUGAUGA (box C) and CUGA (box D) located 5 nt from their 5' and 4 nt from their 3' ends, respectively. Accordingly, the two primers contain 5 and 4 degenerated positions, respectively, to cover all sequence variants in the region upstream of box C and downstream of box D.

box D and C motifs, respectively, thereby ensuring the preferential amplification of C/D snoRNAs.

- 1) Primer 1 (preferentially RNA, or DNA with a 3'-terminal ribonucleotide; see Chapter 3) must be phosphorylated at its 5' end. For the kinase reaction, incubate 60 pmol of primer 1 with 10 U of T4 polynucleotide kinase in 15 μ l 1 x kinase buffer for 60 min at 37 °C. Dilute with 200 μ l of RNase-free H₂O, extract with an equal volume of phenol/chloroform, and precipitate with ethanol using 1 μ l of glycogen (20 mg ml⁻¹) as a carrier. The pellet is washed with 70% ethanol, air-dried, and redissolved in 10 μ l RNase-free H₂O.
- 2) Ligate the extracted RNAs (from Protocol 2.1, step 6) to phosphorylated primer 1 (see also Section 35.3, Note 4). The reaction is carried out in 20 μ l of 1 x RNA ligation buffer, with 20 pmol of phosphorylated primer 1 (3.3 μ l from step 1) and 20 U of T4 RNA ligase. Incubate overnight at 4 °C. Then dilute with 200 μ l of RNase-free H₂O and proceed with a phenol/chloroform extraction

- and ethanol precipitation as in step 1 above. Redissolve in 10 μ l RNase-free H_2O .
- 3) The ligation product is then reverse-transcribed using primer 2, in a final reaction volume of 20 μ l. For this purpose, add 6 μ l of RNase-free H_2O to 5 μ l of the oligonucleotide-tagged RNA from step 2 and 1 μ l of primer 2 (0.5 μ g μ l⁻¹). Heat the mixture at 70 °C for 5 min, and quickly chill on ice. Immediately add 4 μ l of 5 x first strand buffer, 2 μ l of 0.1 M DTT, and 1 μ l of 10 mM dNTP mix (10 mM each). Mix and incubate for 2 min at 42 °C. Add 200 U of SuperScript™ II Reverse Transcriptase and incubate for 2 h at 42 °C.
 - 4) Use the resulting cDNA product as template for PCR by *Taq* DNA polymerase I with primers 2 and 3. The PCR reaction is carried out in a final volume of 100 μ l in 1 x *Taq* I buffer, after addition of 6 μ l $MgCl_2$ (25 mM), 2 μ l dNTP mix (10 mM each), 2 μ l of the reverse transcription reaction mix from step 3, 1 μ l each of primer 2 and 3 solutions (each 0.5 μ g μ l⁻¹), and 2.5 U of *Taq* DNA polymerase I. Perform 30 cycles, each consisting of 1 min at 90 °C, 1 min at 55 °C, and 1 min at 72 °C. Extract with an equal volume of phenol/chloroform and precipitate with ethanol.
 - 5) Digest the PCR product overnight at 37 °C with *Bam* HI (the PCR amplification primers are designed to contain a *Bam* HI site) and clone into the pKS vector by standard recombinant DNA techniques [30]. Individual clones can be manually sequenced (T7 sequenase version 2.0, USB Affymetrix). About 70–80% of them correspond to C/D snoRNA-like sequences.

35.2

Computational Analysis of ncRNA Sequences

The cloning strategy described in Protocol 1.1 will result in 5'-truncated cDNAs that are significantly shorter than the corresponding ncRNAs detected by Northern blot analysis (see above). For organisms with a completely sequenced genome, the missing 5'-terminal sequence of the ncRNA can be inferred from the genomic sequence, taking into account the size determined by the Northern analysis. All ncRNA sequences are then searched for the presence of known sequence or structure motifs typical of a particular ncRNA subtype, especially the two widespread C/D and H/ACA snoRNA families (see below). Motif searches can be carried out with the DNAMAN sequence analysis program. Other relevant software tools are SnoReport (computational identification of snoRNAs with unknown targets [34]), PLEXY (efficient target prediction for box C/D snoRNAs [35]), as well as general motif search programs, such as fragrep (homology search with fragmented nucleic acid sequence patterns [36]), rnabob (fast pattern searching for RNA secondary structures), and RNAmotif (sequence analysis code, for finding structural motifs in nucleic acid sequences [37]). Folding of ncRNA structures can be predicted using the mfold version 3.0 by Zuker [38] or rnafold (Vienna RNA web servers [37]).

35.3

Notes

- 1) After isolation of the RNA of choice by the TRIzol method, we find it useful to additionally perform one or two phenol/chloroform extractions to remove residual amounts of proteins still attached to RNAs.
- 2) When running total RNA on denaturing polyacrylamide gels for size separation, try to minimize the running distance of RNA into the gel because otherwise you will have too much gel material from which the RNA has to be passively eluted: the more gel material you end up with, the larger is the elution volume and hence the more difficult it will be to quantitatively precipitate the size-selected RNA.
- 3) When employing the C-tailing method, sequence analysis of clones will be severely impeded when a primer downstream from the C-tail is used. According to our experience, sequencing through the C-tail to determine the ncRNA gene of interest gives unsatisfactory results. Hence, for sequence analysis, a primer upstream from the ncRNA gene of interest is recommended.
- 4) As the ligation reaction is not restricted in terms of primer 5' and 3' end reactivity (Protocol 2.2, step 2), this step can lead to the production of various forms of primer–RNA concatemers, which may result in the cloning of inserts spanning two (or more) C/D snoRNA sequences.

Materials for RNPomics

Oligonucleotide Primers

- 5'-primer-adaptor: 5'-GTC AGC AAT CCC TAA C GAG-3' (ribonucleotides underlined)
- Anchor primer: 5'-AGG AGC CAT CGT ATG TCG GGG GGG GH-3' (H = A, C, or T)
- 5'-PCR primer*: 5'-(high-throughput specific sequence, ~30 nt)-GTC AGC AAT CCC TAA CGA G-3'
- 3'-PCR primer*: 5'-(high-throughput specific sequence, ~30 nt)-AGG AGC CAT CGT ATG TCG-3'

(Asterisk indicates that an additional specific sequence for high-throughput sequencing (454[®], Illumina[®], Solid[®]) has to be introduced 5' to both PCR primer sequences. As these sequences are often changed by high-throughput sequencing companies in the course of improving sequencing technologies, we prefer not to specify any sequence in this protocol.)

Enzymes

- A-Plus[™] poly(A) Polymerase (Epicentre Biotechnologies)
- RiboLock[™] RNase Inhibitor (Thermo Scientific Fermentas)

- TAP: Tobacco Acid Pyrophosphatase (Epicentre Biotechnologies)
- T4 RNA Ligase (New England Biolabs)
- SuperScript™ II Reverse Transcriptase (Life Technologies Invitrogen)
- *Taq* DNA Polymerase I (Life Technologies Invitrogen)

Buffers

- Buffer A: 10 mM Hepes, pH 7.9, 1.5 mM MgCl_2 , 10 mM KCl, 0.5 mM DTT, 0.5 mM PMSF (phenylmethylsulfonyl fluoride)
- Buffer B: 300 mM Hepes, pH 7.9, 30 mM MgCl_2 , 1.4 M KCl
- Buffer D: 20 mM Hepes, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 20% glycerol
- High-salt buffer: 20 mM Hepes, pH 7.9, 1.2 M KCl, 1.5 mM MgCl_2 , 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 25% glycerol
- Low-salt buffer: 20 mM Hepes, pH 7.9, 20 mM KCl, 1.5 mM MgCl_2 , 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 25% glycerol
- Gradient dilution buffer: 20 mM Hepes, pH 7.9, 100 mM KCl, 1 mM MgCl_2
- Glycerol solutions: 20 mM Hepes, pH 7.9, 100 mM KCl, 1 mM MgCl_2 , 10–30% glycerol
- 1 x Tailing buffer: 50 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 mM MgCl_2 , 10 mM CTP
- 1 x TAP buffer: 50 mM sodium acetate, pH 6.0, 1 mM EDTA, 1% β -mercaptoethanol, 0.01% Triton® X-100
- 1 x RNA ligation buffer: 50 mM Tris-HCl, pH 7.8, 10 mM MgCl_2 , 10 mM DTT, 1 mM ATP, $0.1 \mu\text{g} \mu\text{l}^{-1}$ BSA
- 5 x First strand buffer: 250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM MgCl_2
- 1 x *Taq* buffer: 10 mM Tris-HCl, pH 9.0, 50 mM KCl, 0.1% Triton® X-100

Reagents and Tools

- Teflon-glass Dounce homogenizer (for example, Potter S Accessories, Thomas Scientific).
- Phenol/chloroform/isoamyl alcohol: 25/24/1 solution, abbreviated PCI
- Dulbecco's Phosphate Buffered Saline (PBS) without CaCl_2 and MgCl_2 (Sigma-Aldrich)
- Glycoblue™: 15 mg ml^{-1} glycogen (Life Technologies Ambion)
- cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets (Roche Applied Science)
- PMSF: 200 mM solution in formamide

Protocol for Library Construction

The method below describes the construction of a specialized cDNA library from size-fractionated RNPs from HeLa cells on 10–30% glycerol gradients (adapted with permission of Ref. [28]). The method can also be employed to prepare cDNA libraries from any desired tissue or organism. For the construction of the

cDNA library, 3'-tailing of RNP-derived ncRNAs by CTP and poly(A) polymerase is followed by 5'-primer-adaptor ligation by T4 RNA ligase and reverse transcription of ncRNAs using an oligo-d(G) anchor primer. RNP selection significantly increases the probability to include biologically relevant ncRNA species in the library. However, a potential limitation of cDNA expression libraries is that ncRNAs of very low abundance or ncRNAs expressed in other tissues during a different developmental stage or upon particular conditions of stress might be missed. Their identification would require the generation of additional libraries for each condition in order to be detected. In addition, some ncRNAs might be involved in RNP complexes larger than 30S or might dissociate from their protein partners during library generation and thus would be lost as well. Therefore, taking these pitfalls into account, it is clear that the RNP approach, as observed for other ncRNA cloning methods, will likely not reveal all ncRNAs forming RNPs in a tissue or an entire organism. It is also important to note that this approach rather focuses on the small ncRNA/ncRNP transcriptome. Longer ncRNA species (e.g., in antisense orientation to protein coding genes), which might function even in the absence of protein binding partners, could represent yet another large class of so far unidentified ncRNA species [39].

Protocol 3: RNP-Derived cDNA Library Construction

This section describes the preparation of RNP extracts from the cells (or tissue) of interest, the sedimentation of the RNPs on a 10–30% glycerol gradient, the isolation of the RNA from the gradient fractions, and the reverse transcription of ncRNAs into cDNA. Since they do not contain a poly(A) tail, ncRNAs are tailed with C residues by poly(A) polymerase in the first step of the library generation process. After the addition of C-tails, a primer-adaptor is ligated to the 5'-end of ncRNAs and the latter are reverse-transcribed employing an oligo-d(G) primer and SuperScript™ II Reverse Transcriptase.

Protocol 3.1: Preparation of the RNP Extracts and Glycerol Gradient Centrifugation

- 1) Harvest HeLa cells from cell culture media and wash them twice in five volumes of ice-cold Dulbecco's PBS medium. All following steps must be carried out at 4 °C.
- 2) Preparation of the cytoplasmic RNP extract [40]: Suspend the cells in five packed cell pellet volumes of buffer A and incubate them on ice for 10 min. Resuspend the cells in two packed cell pellet volumes of buffer A and lyse the cells by 10 strokes of a Teflon-glass Dounce homogenizer or until 80% of the cells are lysed, which can be microscopically verified employing Trypan blue. Centrifuge the homogenate for 15 min at 700 g to pellet nuclei (save the pellet for step 3). Mix the supernatant with 0.11 volumes of buffer B and centrifuge it for 60 min at 100 000 g. Dialyze the high-speed supernatant two times for 2 h against 50 volumes of buffer D, and centrifuge the dialysate for 30 min at 9000 g. The supernatant corresponds to the S100 or cytoplasmic fraction.

- 3) Preparation of the RNP nuclear extract [40]: Suspend the pelleted nuclei obtained from the low-speed centrifugation (step 2) in 0.5 ml of low-salt buffer. Add 0.5 ml of high-salt buffer, and gently stir the resulting suspension with a magnetic stirring bar for 30 min. Centrifuge the homogenate for 30 min at 9000 g. Dialyze the supernatant against 50 volumes of buffer D two times for 2 h. Centrifuge the dialysate at 9000 g for 30 min. The supernatant is designated as the nuclear extract.
- 4) Sedimentation of the RNP extracts: Mix the cytoplasmic, nuclear, or a mixture of both extracts with an equal volume of gradient dilution buffer and layer the mixture onto a 10–30% glycerol gradient (Beckman Coulter polyallomer tubes, 14 × 89 mm, 13.2 ml). Spin the gradient at 140 000 g (34 000 rpm) for 18 h at 4 °C in a Beckman SW41 rotor and collect fractions.

Protocol 3.2: RNA Preparation and Construction of the cDNA Library [41]

- 1) Extract RNA from the fractions with PCI, and precipitate the RNA with ethanol.
- 2) C-tailing of ncRNAs: Tail 150 ng to 1 µg of RNA with CTP using poly(A) polymerase. The C-tailing reaction is carried out in a volume of 20 µl containing 1 x tailing buffer, 4 U of A-Plus™ poly(A) polymerase, and 20 U of RiboLock™ RNase Inhibitor. Incubate the reaction mix at 37 °C for 1.5 h. After incubation, add 20 µl 3 M sodium acetate (pH 5.2) and 160 µl RNase-free water. Extract the RNA with PCI, precipitate with ethanol (for details, see Protocol 1.1, steps 4 and 5, and Section 35.5, Note 3), and dissolve in 7 µl of RNase-free water.
- 3) Decapping of ncRNAs: Treat 7 µl of C-tailed RNA (obtained in step 2) with tobacco acid pyrophosphatase (TAP) so that all ncRNAs feature a 5'-monophosphorylated terminus. The decapping reaction is carried out in a volume of 10 µl containing 1 x TAP buffer, 10 U of TAP, and 40 U of RiboLock™ RNase Inhibitor. Incubate the reaction mix at 37 °C for 1 h. After incubation, add 20 µl 3 M sodium acetate (pH 5.2) and 170 µl RNase-free water. Extract the RNA with PCI, precipitate with ethanol, and dissolve in 10 µl of RNase-free water.
- 4) 5'-primer-adapter ligation: Combine 10 µl TAP-treated and C-tailed RNA (obtained in step 3) with 2 µl 5'-primer-adapter (100 pmol µl⁻¹). Preincubate the mixture at 65 °C for 5 min. The ligation reaction is carried out in a volume of 20 µl containing 1 x RNA ligation buffer, 40 U of RiboLock™ RNase Inhibitor, and 10–20 U of T4 RNA Ligase overnight at 4 °C. After incubation, add 180 µl RNase-free water. Extract the RNA with PCI, precipitate with ethanol, and dissolve in 10 µl of RNase-free water.
- 5) Reverse transcription: Combine 10 µl of 5'-primer-adapter-ligated RNA (obtained in step 4) with 1 µl anchor primer (100 pmol µl⁻¹) and 1 µl dNTP mix (10 mM each). Denature the mixture at 65 °C for 5 min and immediately chill on ice. Add 4 µl 5 x first strand buffer, 2 µl DTT (100 mM), 40 U of RiboLock™ RNase Inhibitor and preincubate at 42 °C for 2 min. Add 200 U of SuperScript™ II Reverse Transcriptase (final volume 20 µl) and incubate the reaction for 1 h at 42 °C.

- 6) cDNA amplification: Amplify 4 μ l of the cDNA from step 5 in a reaction volume of 50 μ l containing 1 x *Taq* buffer, 1 pmol μ l⁻¹ of both PCR primers, 0.25 mM MgCl₂, 0.1 mM each dNTP, and 1.0–2.5 U of *Taq* DNA Polymerase I. Heat samples in a thermal cycler at 94 °C for 2 min to denature the template. Perform 20–25 cycles of PCR amplification (1 min at 94 °C, 1 min at 54 °C, and 1 min at 72 °C). Then incubate for an additional 5 min at 72 °C and keep the reaction mixture at 4 °C. The samples can be stored at –20 °C until use, or directly analyzed by deep sequencing. Material suitable for direct use in the emulsion PCR step can be generated as described previously [42]. Quality control (agarose gel electrophoresis, cloning, and Sanger sequencing to check for the presence of the deep sequencing adaptors) can be realized but is usually performed by the company/platform in charge of the deep sequencing.

35.4

Computational Analysis of ncRNA Sequences

The 5'- and 3'-adaptors have to be removed from the raw data obtained before sequences can be mapped to the corresponding genomes. Sequences are subsequently compared to existing annotations, using data from different gene banks (e.g., Ensembl, tRNADB, UCSC), and can be further analyzed for conservation of secondary structure features. Numerous software tools are available to automatically process deep sequencing data (mapping, annotation, alignment, e.g., *segemehl* [43]). Expression of novel ncRNA candidates can next be verified by Northern blotting and precise subcellular localization by fluorescence *in situ* hybridization (FISH).

35.5

Notes

- 1) Aliquots from both S100 and nuclear extracts can be snap frozen in liquid nitrogen and stored at –80 °C for several months.
- 2) After glycerol gradient ultracentrifugation, bottom and top fractions (1 ml each) should be discarded. These fractions contain rRNA and various RNA degradation products, respectively. The remaining fractions can be pooled if no particular RNP size (10–30S) is requested. Otherwise, specific subfractions (e.g., 12S) may be selected for further processing.
- 3) The pH of the first PCI extraction can be acidic (below 7, ideally 4.5 or 5) in order to denature DNA that may possibly be present as a contaminant. At pH 7, DNA and RNA will appear in the water phase. At pH 5, DNA is partly denatured and will be directed to the organic phase, while RNA will remain in the water phase. For the precipitation step, glycogen or Glycoblue™ can be added as a carrier since the amount of material retrieved after the gradient may be rather low. Throughout the protocol, RNA pellets should be gently washed after ethanol precipitation to avoid washing out of small RNAs (e.g., miRNAs).

- 4) For the best efficiency of overnight ligation, samples should be incubated in melting ice (water + ice) rather than in crushed ice only. To avoid too fast melting of the ice, the incubation should additionally take place in a cold room. After overnight ligation at 4 °C, adding another 10 U of T4 RNA ligase followed by incubation for 30 min at 37 °C increases the yield of the reaction.
- 5) The PCR step can also be carried out with a homemade *Taq* polymerase if available. However, high-fidelity enzymes are recommended to avoid too many errors.

Acknowledgments

This work was supported by the grant HU 467/5-1 of the Deutsche Forschungsgemeinschaft (DFG), grant FWF 171370 of the Österreichischer Fonds für Wissenschaft und Forschung (FWF) to A. H., by the Medical University of Innsbruck, and the Austrian Genome Research (GEN-AU, <http://www.gen-au.at/>) funding program (Grants D-110420-011-013 and D-11420-011-015 to A. H.).

References

1. Couzin, J. (2002) Breakthrough of the year. Small RNAs make big splash. *Science*, **298**, 2296–2297.
2. Dennis, C. (2002) Small RNAs: the genome's guiding hand? *Nature*, **420**, 732.
3. Dennis, C. (2002) The brave new world of RNA. *Nature*, **418**, 122–124.
4. Hüttenhofer, A., Brosius, J., and Bachellerie, J.P. (2002) RNomics: identification and function of small, non-messenger RNAs. *Curr. Opin. Chem. Biol.*, **6**, 835–843.
5. Ambros, V. (2001) microRNAs: tiny regulators with great potential. *Cell*, **107**, 823–826.
6. Bachellerie, J.P., Cavaillé, J., and Hüttenhofer, A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
7. Gottesman, S. (2002) Stealth regulation: biological circuits with small RNA switches. *Genes Dev.*, **16**, 2829–2842.
8. Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
9. Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
10. Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, **294**, 862–864.
11. Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002) Prediction of plant microRNA targets. *Cell*, **110**, 513–520.
12. Storz, G. (2002) An expanding universe of noncoding RNAs. *Science*, **296**, 1260–1263.
13. Tuschl, T. (2002) Expanding small RNA interference. *Nat. Biotechnol.*, **20**, 446–448.
14. Tuschl, T. (2003) Functional genomics: RNA sets the standard. *Nature*, **421**, 220–221.
15. Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, **297**, 1833–1837.
16. Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder,

- M., Dermitzakis, E.T., Thurman, R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
17. Mattick, J.S. and Makunin, I.V. (2005) Small regulatory RNAs in mammals. *Hum. Mol. Genet.*, **14** (Spec No. 1), R121–R132.
 18. Willingham, A.T. and Gingeras, T.R. (2006) TUF love for “junk” DNA. *Cell*, **125**, 1215–1220.
 19. Brosius, J. (2005) Waste not, want not—transcript excess in multicellular eukaryotes. *Trends Genet.*, **21**, 287–288.
 20. Hüttenhofer, A., Schattner, P., and Polacek, N. (2005) Non-coding RNAs: hope or hype? *Trends Genet.*, **21**, 289–297.
 21. Hüttenhofer, A., Kiefmann, M., Meier-Ewert, S., O’Brien, J., Lehrach, H., Bachellerie, J.P., and Brosius, J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, **20**, 2943–2953.
 22. Marker, C., Zemann, A., Terhörst, T., Kiefmann, M., Kastenmayer, J.P., Green, P., Bachellerie, J.P., Brosius, J., and Hüttenhofer, A. (2002) Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr. Biol.*, **12**, 2002–2013.
 23. Tang, T.H., Bachellerie, J.P., Rozhdestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J., and Hüttenhofer, A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 7536–7541.
 24. Eddy, S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.
 25. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
 26. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
 27. Rederstorff, M., Bernhart, S.H., Tanzer, A., Zywicki, M., Perfler, K., Lukasser, M., Hofacker, I.L., and Hüttenhofer, A. (2010) RNPomics: defining the ncRNA transcriptome by cDNA library generation from ribonucleo-protein particles. *Nucleic Acids Res.*, **38**, e113.
 28. Rederstorff, M. and Hüttenhofer, A. (2010) cDNA library generation from ribonucleoprotein particles. *Nat. Protoc.*, **6**, 166–174.
 29. Schmitt, A.O., Herwig, R., Meier-Ewert, S., and Lehrach, H. (1999) in *PCR Applications: Protocols for Functional Genomics* (eds. M.A. Innis, D.H. Gelfand, and J.J. Sninsky), Academic Press, San Diego, CA, pp. 457–472.
 30. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, 2nd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
 31. Maier, E., Meier-Ewert, S., Ahmadi, A.R., Curtis, J., and Lehrach, H. (1994) Application of robotic technology to automated sequence fingerprint analysis by oligonucleotide hybridisation. *J. Biotechnol.*, **35**, 191–203.
 32. Kiss-Laszlo, Z., Henry, Y., Bachellerie, J.P., Caizergues-Ferrer, M., and Kiss, T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*, **85**, 1077–1088.
 33. Cavaillé, J., Vitali, P., Basyuk, E., Hüttenhofer, A., and Bachellerie, J.P. (2001) A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *J. Biol. Chem.*, **276**, 26374–26383.
 34. Hertel, J., Hofacker, I.L., and Stadler, P.F. (2008) SnoReport: computational identification of snoRNAs with unknown targets. *Bioinformatics*, **24**, 158–164.
 35. Kehr, S., Bartschat, S., Stadler, P.F., and Tafer, H. (2010) PLEXY: efficient target prediction for box C/D snoRNAs. *Bioinformatics*, **27**, 279–280.
 36. Mosig, C., Chen, J.L., and Stadler, P.F. (2007) in *WABI 2007* (eds. R. Giancarlo and S. Hannehalli), Springer-Verlag, Berlin, pp. 335–345.
 37. Hofacker, I.L. (2004) RNA secondary structure analysis using the Vienna

- RNA package. *Curr. Protoc. Bioinformatics*, Chapter 3: Unit 12.2.
38. Zuker, M. (1994) Prediction of RNA secondary structure by energy minimization. *Methods Mol. Biol.*, **25**, 267–294.
 39. Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.*, **10**, 155–159.
 40. Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.
 41. Hüttenhofer, A. and Vogel, J. (2006) Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res.*, **34**, 635–646.
 42. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bembien, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.
 43. Hoffmann, S., Otto, C., Kurtz, S., Sharma, C.M., Khaitovich, P., Vogel, J., Stadler, P.F., and Hackermüller, J. (2009) Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput. Biol.*, **5**, e1000502.

36

Computational Methods for Gene Expression Profiling Using Next-Generation Sequencing (RNA-Seq)

John C. Castle

36.1

Introduction

A fundamental goal of molecular biology is to understand the transcriptome present in a sample, including the identification and occurrence (expression) of RNA transcripts. Transcript identification has been done by sequencing RNA-derived libraries, both in gene-directed ways, such as found in Genbank mRNAs [1], and in non-directed ways, such as found in dbEST [2]. Gene expression studies have been done using low-throughput gene-directed protocols, such as quantitative polymerase chain reaction (qPCR) and Northern blots, and using high-throughput, high-content platforms, such as oligonucleotide microarrays. Intensities resulting from microarray images reflect the amount of nucleotides that bind to each oligonucleotide probe. Insofar as the RNA derived from a specific transcript binds selectively to the expected probe, the results provide semiquantitative expression levels.

Next-generation sequencing (NGS), which is also called *deep sequencing*, is a platform that simultaneously decreases the cost and time of sequencing while dramatically increasing the number of sequence reads generated. In the past 27 years, 8 301 471 human expressed sequence tag (EST) sequences have been deposited in dbEST [2]. For comparison, a single NGS instrument can now generate over 800 000 000 sequence reads in two days. Individual sequence reads can be mapped to transcripts based on the nucleotide sequence itself, and counting the number of sequence reads mapped to a transcript enables an accurate, digital determination of the transcript expression level. This platform is frequently called “RNA-Seq”.

As a graphical example of RNA-Seq, Figure 36.1 shows sequence read alignments from two human samples, a pool of brain tissues, and a pool of multiple cell lines universal human reference (UHR), across a locus on chromosome 9. One can identify regions with no transcriptional activity (Figure 36.1, #1); regions with sample and strand-specific transcription, such as the brain-transcribed region under gene SLC24A2 (Figure 36.1, #2); expression of ncRNA (Figure 36.1, #3); transcript-specific expression of overlapping 3' untranslated regions (UTRs) on opposite strands (Figure 36.1, #4); and novel transcription such as in this locus implicated in myocardial infarction (Figure 36.1, #5) ([3]; data from [4]).

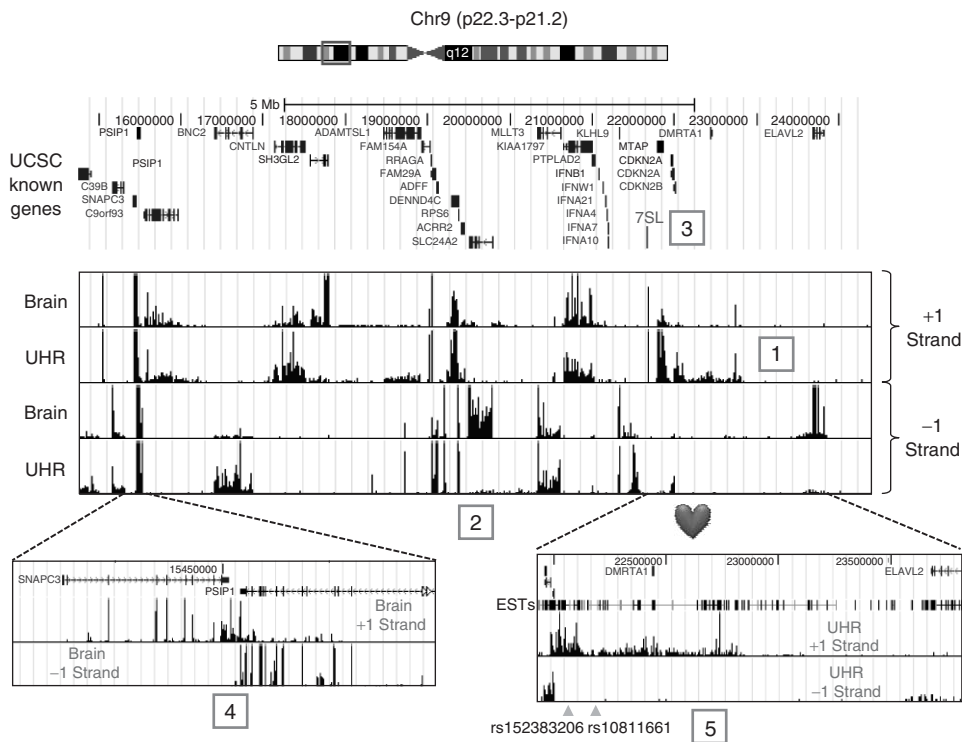


Figure 36.1 Read alignments from a pool of brain tissues and a pool of multiple cell lines (UHR) across a locus on chromosome 9. There are regions with no transcriptional activity (#1); regions with sample and strand-specific transcription (#2); expression

of ncRNAs (#3); transcript-specific expression of overlapping 3' UTRs on opposite strands (#4); and novel transcription such as in this locus implicated in myocardial infarction (#5). (Figure adapted from C. Raymond, personal communication.)

Several reports have demonstrated the attractiveness and application of RNA-Seq [5–8]. However, the joy of experimenters running their first NGS experiment can quickly turn to horror when they receive their first hard drive containing billions of sequence reads. This chapter describes computational methods to convert sequence reads into meaningful gene expression values. I focus on reads derived from Illumina sequencing instruments and explicitly assume that a reference genome exists. Many of the computational tools described are shown in Table 36.1.

36.2

Procedure Overview

Several computational steps are required to determine gene expression profiles. Figure 36.2 shows steps (left) and the subsequent output (right). First, the computational biologist must understand the purpose of the experiment and the laboratory

Table 36.1 Computational tools for RNA-Seq.

Name	Website	Notes
BOWTIE	http://bowtie-bio.sourceforge.net/index.shtml	Alignment code
TOPHAT	http://tophat.cbcb.umd.edu/	Alignment code
CUFFLINKS	http://cufflinks.cbcb.umd.edu/	Alignment code
BWA	http://bio-bwa.sourceforge.net/bwa.shtml	Alignment code
QPALMA	http://www.fml.tuebingen.mpg.de/raetsch/suppl/qpalma	Alignment code
GSNAP	http://research-pub.gene.com/gmap/	Alignment code
SOAP	http://soap.genomics.org.cn/	Alignment code
SAM format	http://samtools.sourceforge.net/SAM1.pdf	Alignment format
BED format	http://genome.ucsc.edu/FAQ/FAQformat.html	Alignment format
UCSC code	http://genomewiki.ucsc.edu/index.php/Kent_source_utilities	Utilities
BEDTools	http://code.google.com/p/bedtools/	Utilities
HTSeq	http://www-huber.embl.de/users/anders/HTSeq	Utilities
UCSC Genome Browser	http://genome.ucsc.edu/	Visualization
Integrative Genomics Viewer	http://www.broadinstitute.org/igv/	Visualization
Seqanswers	http://seqanswers.com/	Useful blog discussing NGS

protocols used. Second, reads are mapped to the genome and transcriptome. Third, mapped reads are associated with genes. Fourth, expression values and uncertainties are determined for each gene. Fifth, expression values and uncertainties are normalized for each transcript and each tissue. Finally, the results are formatted for viewing and downstream use.

36.2.1

Understand the Experiment and the Molecular Biology Protocol

The computational biologist may be handed a hard drive containing sequence reads and asked to analyze them. Before starting analysis, it is crucial for the computational biologist to understand the experiment goal and aims, what the reads represent, and the protocols that were used to generate the reads. Ideally, the computational biologist will be part of the experimental design team and will help identify the optimal protocol for the project.

36.2.1.1 Library Generation

There are many common variations in the RNA-Seq library generation protocols. These include using random primed total RNA, polyA+ purified RNA, and ribosomal depleted RNA. In each case, the protocol will impact the percentage of reads that align to rRNA, polyA-, and polyA+ transcripts. As many ncRNA and histone transcripts are not polyadenylated, monitoring them requires a polyA-neutral

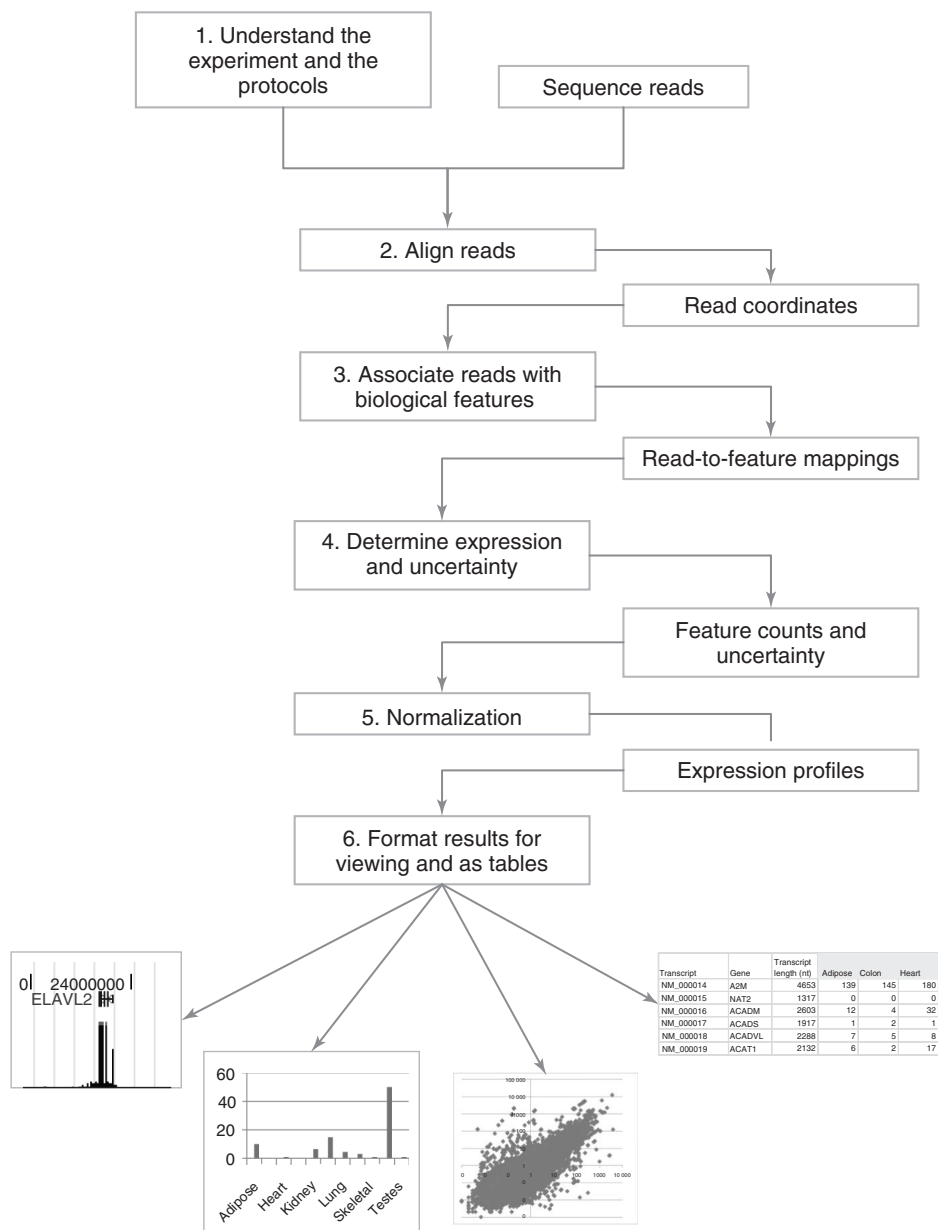


Figure 36.2 Flow chart to generate expression profiles from RNA-Seq data. Shown are the steps (left) and the subsequent output (right).

protocol, for example, one that amplifies polyA+ and polyA– transcripts (Figure 36.1, #3). Conversely, as mRNA transcripts are typically polyadenylated, use of a polyA+ enrichment step results in a greater percentage of reads that map to mRNAs. Some protocols preserve the strand of the original RNA transcripts, while others lose this information, typically in order to simplify the laboratory protocol. When the strand is preserved, protocols may generate either reads with the same strand or the reverse complement of the original RNA transcript. Use of a strand-specific protocol improves data quality and allows differentiation of reads that map to overlapping transcripts, as can occur in 3' UTRs (Figure 36.1, #4).

Transcripts with lengths less than 50 nt are not sequenced in common RNA-Seq protocols, either because of the use of amplification steps that deplete short transcripts or as the result of an explicit size selection before sequencing. Small transcripts, such as miRNAs, can be detected and profiled, but currently only with specialized protocols.

36.2.1.2 Sequencing

When sequencing, there are also several experimental parameters that can be modified. Foremost is the read length. While early experiments sometimes generated reads shorter than 32 nt in length, most experiments now generate reads of lengths that are at least 35 nt, up to 100 nt on the Illumina instrument, and over 400 nt on the Roche/454 instrument. Longer reads are more likely to align unambiguously to the genome and, given that the median protein-coding exon length is near 110 nt, are more likely to overlap exon–exon junctions.

Sequencing can be done to generate “single-end”, “paired-end”, and “mate-pair” reads. During library preparation, the resultant nucleotide sequences can be size selected to have lengths of 200–500 nt. Single-end reads are a single read from one end of the sequence. Paired-end reads are a pair of sequence reads derived from both ends of the sequence. “Mate-pairs” read are generated using a different laboratory protocol but, similar to paired-end reads, result in a pair of sequence reads with an expected separation (usually over 1000 nt). The use of paired-end and mate-pair reads provides information enabling connecting distal transcript fragments.

The instruments can generate over 50 million sequence reads per sample, such as in one flow cell lane of an Illumina HiSeq instrument. Depending on the experiment, significantly fewer reads may suffice, and thus multiple samples can be combined – multiplexed – in one flow cell lane to increase throughput and decrease costs. This can be accomplished through the use of “molecular barcodes”, in which case sample-specific nucleotides are ligated to the end of the sequence before sequencing. For a given sequence, the first nucleotides sequenced thus identify the sample from which the sequence was derived.

Before analysis, it is crucial that the computational biologist understands which protocols were used. For example, it is obviously critical to know if there was barcoding and, if used, the size (nucleotide) of the barcode, both in order to associate reads with samples and remove the barcode before alignment. The earlier the computational biologist engages the project team, the more likely that useful sequence data will be generated.

36.2.2

Align Reads

After preprocessing, the sequence reads for any protocol-specific steps, such as dealing with barcodes, the reads must be aligned to the genome. While potentially intimidating, any computational biologist who is able to use “blast” on the command line will be able to use NGS alignment programs. Conceptually, read alignment is similar to a blast similarity search, where each NGS sequence read is compared to a database containing a formatted reference genome. Commonly used alignment programs include BOWTIE [9], TOPHAT [10], BWA [11], QPALMA [12], SOAP [13], and GSNAP [14]. Each program has its own strengths, such as speed; memory requirements; ability to handle single nucleotide polymorphisms (SNPs), edits, mutations, insertions and deletions, and sequencing errors; ability to map junction reads; and ability to handle paired-end reads. Also, each program has its own set of parameters that the user can optimize for project-specific applications.

Sequences that fall entirely within one exon will align continuously to the genome. Sequences crossing exon–exon splice junctions, however, will not align continuously to the genome and thus require special treatment. The most common way to align these reads is to align sequence reads to a second sequence database, one that is not derived from the genome but composed of splice junctions from established transcripts. Given alignments of transcripts to the genome, the read-to-transcript alignment coordinates can be translated to read-to-genome coordinates. Following this path, the database of splice junctions can be expanded from established transcripts, such as RefSeq transcripts, to include putative splice junctions, such as those from ESTs, cross-species predictions, and even removing exons from established transcripts to generate new junctions.

Use of transcripts as an intermediate to align junction-spanning reads is effective but depends on a priori knowledge of the transcriptome. However, it will fail to help align junction-spanning reads in novel, unpredicted genes, and for species whose transcriptome is little known. Fortunately, alignment of junction-spanning reads to the genome without a priori transcriptome knowledge is an active field of research. Algorithms such as TOPHAT [15] and QPALMA [12] take different approaches to align junction reads, such as in their use of exon-aligning reads, of large gaps in alignments (e.g., introns between splice sites), and of preassembly of reads before genome alignment.

From the alignment steps, the output is typically a file listing the reads and the genomic coordinates to which they align. When a read aligns to multiple locations, the programs typically return all alignments up to a maximum value. Programs output results in various formats, with the SAM [16] and BED (<http://genome.ucsc.edu/FAQ/FAQformat.html>) formats being common. The alignments can also be used to show how many reads overlap each nucleotide position in the genome (Figure 36.1). One common file format is the “wiggle track” format, which can be readily loaded and viewed in the UCSC Genome Browser (<http://genome.ucsc.edu/>).

36.2.3

Associate Reads with Transcripts

Given the genomic coordinates of reads, the next step is to associate the reads to biological features of interest. For conciseness, I will describe the process using transcripts as the biological feature; however, features can just as easily be individual exons, junctions, alternative splice forms, ncRNAs, or genome blocks. In all cases, the features should be defined in terms of genomic coordinates in the same format as the sequence read alignments.

Conceptually, the goal of this step is simple: examine the coordinates of the sequence reads and determine if they fall within the coordinates of any transcript. While this coordinate intersection can be done in coding languages such as Perl, faster algorithms fortunately exist, such as BEDTools [17] and overlapSelect (http://genomewiki.ucsc.edu/index.php/Kent_source_utilities). OverlapSelect, for example, can intersect 25 000 transcripts with 50 million sequence reads in minutes. There are many options associated with these programs, including the required level of overlap to call a “hit” and how to account for the strand of the reads and the transcripts. Again, it is crucial that the computational biologist knows how the strand of the sequence reads is associated with the original RNAs.

The computational biologist must decide how to treat reads that align to multiple locations, and should make a deliberate and explicit decision. Paralogous genes with highly homologous sequence, low complexity and repetitive elements, and pseudogenes can all confound mappings between reads and features. The simplest solution is to use only reads that map uniquely to the genome. However, even that classification requires defining what a “unique” alignment is. Given that sequencing is not error-free and that SNPs and mutations exist, a read that aligns perfectly to only one location may, when allowing for variation, align to multiple locations. Indeed, this situation, similar to “cross-hybridization” on oligonucleotide microarrays, is observed, in particular, when sequencing errors occur in highly expressed genes that have highly homologous genes. An alternative strategy is to assign multiply mapping reads proportionally to associated transcripts based on the observed expression levels from uniquely mapping reads. For simplicity, I suggest using only uniquely aligning reads, where unique is defined to mean that there is no other alignment for the read even allowing for two mismatches (termed two “edit distance” in information theory, meaning two operations required to transform one sequence string into the other).

ncRNA such as 7SK present another challenge. There are hundreds of 7SK-associated transcripts with identical or almost identical sequences located in multiple locations throughout the genome. A read derived from a 7SK transcript may therefore align to multiple genomic locations. In order to nevertheless determine an expression level for 7SK, the analyst can first define gene clusters, such as the cluster of “all 7SK-associated genes” and their genomic coordinates, and then identify the reads that align to any – or all – of the cluster-associated coordinates [5]. This procedure will not determine the expression of each genomic locus, but rather the expression of a specific transcript as defined by the transcript sequence.

The output from this step in the process is an association between sequence reads and biological feature (transcripts, exons, junctions, splice forms, etc.).

36.2.4

Determine Expression and Uncertainty

Given the associations between reads and transcripts, the next step is to count the number of unique reads associated with transcript. This straightforward step results in the read count for each transcript, representing the expression level.

The uncertainty, or error, of the count can be modeled using Poisson statistics and is, simply, the square root of the counts [5]. Note that this uncertainty reflects the variability in the counting but does not reflect biological or biotechnology uncertainties.

The output from this step is a count and associated error estimate for each transcript, for each sample.

36.2.5

Normalization

Every sequencing run generates a different number of reads. The length of each transcript varies and longer transcripts, simply by being longer, will generate more sequence reads. Thus, to make comparisons between multiple samples or between transcripts, the read counts and uncertainties must be normalized. The most common normalization currently is to normalize the transcript and sample-specific read counts by the transcript length and the number of sample-specific mappable reads to obtain normalized expression in reads per kilobase per million (RPKM) units [6]. The same scaling factor can be used to normalize the uncertainties.

This simple method is effective but has distinct characteristics. If a major component, such as globin, is expressed at significantly higher levels, the RPKM values of other transcripts will decrease, reflecting the changes in transcript molarity. In some cases, transcript-specific up- and downregulation may be better reflected through use of alternative normalizations, such as adaptation of methods previously developed for oligonucleotide microarray data, such as use of robust multiarray averaging (RMA) and quantile normalization [18].

The output from this step should be a table containing normalized transcript expression and normalized uncertainty for each sample.

36.2.6

Output and Viewing

At this point, there should be transcript expression values and uncertainties for each transcript in each sample. This output must be formatted to enable easy interrogation of the results. Again, it is crucial that the computational biologist analyzing the sequence reads understands the purpose of the experiment as there

are multiple ways to format the results. The goal is to process the results in a way that leads to clear interpretations.

Viewing results in a genome browser is possible using, for example, the wiggle data format and the UCSC Genome Browser (Figure 36.1 and Figure 36.2, lower left) and the Integrative Genomics Viewer (<http://www.broadinstitute.org/igv/>). This genomic view of the data is particularly useful for exploring transcription within a specific genomic locus. Alternatively, the expression patterns of individual genes can be plotted across multiple tissues (Figure 36.2, bottom, second from left) and transcript expression patterns can be compared between sample pairs (Figure 36.2, bottom, second from right). Finally, data can be formatted as a table and interrogated in Microsoft Excel (Figure 36.2, lower right). Over the past decade, many computational tools and algorithms have been created for analysis of high-throughput microarray gene expression data; it is advantageous to leverage these tools by saving results in a common format, such as the BAM format.

36.2.7

Troubleshooting

Inclusion of quality control steps is crucial. Basic measurements for each data set include number of reads, overall base composition, and base composition and quality score at each position. After alignment, useful metrics include the percentages of reads that can be aligned, uniquely aligned, aligned to exons, and aligned to junctions. In particular, this metrics should be compiled and compared across samples for quality control. There are tools available to perform many QC steps, such as HTSeq (<http://www-huber.embl.de/users/anders/HTSeq>).

Changes in laboratory protocols can significantly impact the measured expression levels; so, again, it is crucial that the computational biologist knows the protocol used. For example, changing from a polyA+ enrichment protocol to a non-enriching protocol will dramatically change the expression levels of all non-polyadenylated transcripts. Whether and how such datasets can be integrated critically depends on the purpose of the experiment.

When reads do not align or do not map to genes, verify that the correct genome is being used. This includes not only verification that the genome from the correct species was used, but also that all analysis was done using the same genome assembly, including both read alignments and transcript coordinates. The UCSC Genome Browser is an outstanding tool; loading results into the browser enables fast and easy validation.

Typically, some transcripts are expressed at non-zero levels in one sample and zero in another sample. This can lead to one of the most oddly gratifying problems – that RNA-Seq can truly identify transcripts with zero expression – leading to errors when generating plots with logarithmic scales and when calculating fold changes. When needed, one solution is to add a minimum background value to each expression value. However, statistics for determining differential expression, such as *p*-values, should be made robust to zeros and use the measured values.

36.2.8

The Future Is Bright!

This chapter discussed the basics of RNA expression profiling using NGS. This is an exciting and evolving field, with rapid advances in biotechnology, analysis, and applications. The sequencing technology continues to improve at an extremely rapid pace, with increasing throughput, decreasing costs and time, and generating more, longer, and higher quality reads. New protocols and algorithms that use RNA-derived sequence reads to not only interrogate expression of known transcripts and exons but also profile alternative splice forms, alternative 5' and 3' exons, and alternative isoforms [10] are being developed. Furthermore, as the sequencing is relatively unbiased in terms of preselection, the reads enable transcriptome discovery – novel genes and isoforms – in both normal and mutated (cancerous) tissues. That the reads contain sequence information enables, for example, examination of allele-specific expression, RNA editing, and mutation/SNP discovery and monitoring [19].

36.3

Protocols: Useful Algorithms, Formats, and Tools

Table 36.1 lists useful alignment algorithms, formats of sequence read alignments, utilities, visualization tools, and a useful message board/blog. The list is obviously incomplete; the reader can find more information on the Seqanswers blog.

References

1. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2009) GenBank. *Nucleic Acids Res.*, **37**, D26–D31.
2. Boguski, M.S., Lowe, T.M., and Tolstoshev, C.M. (1993) dbEST – database for “expressed sequence tags”. *Nat. Genet.*, **4**, 332–333.
3. Helgadóttir, A., Thorleifsson, G., Manolescu, A., Gretarsdóttir, S., Blondal, T., Jonasdóttir, A., Sigurdsson, A., Baker, A., Pálsson, A., Masson, G. *et al.* (2007) A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*, **316**, 1491–1493.
4. Armour, C.D., Castle, J.C., Chen, R., Babak, T., Loerch, P., Jackson, S., Shah, J.K., Dey, J., Rohl, C.A., Johnson, J.M., and Raymond, C.K. (2009) Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. *Nat. Methods*, **6**, 647–649.
5. Castle, J.C., Armour, C.D., Lower, M., Haynor, D., Biery, M., Bouzek, H., Chen, R., Jackson, S., Johnson, J.M., Rohl, C.A., and Raymond, C.K. (2010) Digital genome-wide ncRNA expression, including snoRNAs, across 11 human tissues using polyA-neutral amplification. *PLoS ONE*, **5**, e11779.
6. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.
7. Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–1415.
8. Wang, E.T., Sandberg, R., Luo, S., Khrebukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008) Alternative isoform

- regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
9. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
 10. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
 11. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
 12. De Bona, F., Ossowski, S., Schneeberger, K., and Rättsch, G. (2008) Optimal spliced alignments of short sequence reads. *Bioinformatics*, **24**, i174–i180.
 13. Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics*, **24**, 713–714.
 14. Wu, T.D. and Nacu, S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.
 15. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.
 16. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
 17. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
 18. Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, **11**, 94.
 19. Pepke, S., Wold, B., and Mortazavi, A. (2009) Computation for ChIP-seq and RNA-seq studies. *Nat. Methods*, **6**, S22–S32.

37

Characterization and Prediction of miRNA Targets

Jean Hausser and Mihaela Zavolan

37.1

Introduction

MicroRNAs (miRNAs) are short, non-coding RNAs that repress gene expression at the posttranscriptional level. The first miRNA to be discovered, *lin-4*, plays an important role in the development of *Caenorhabditis elegans* by repressing the *lin-14* heterochronic gene [1, 2] at the transition between the first and second larval stage. To date, hundreds of miRNAs have been discovered in a broad range of species ranging from plants to metazoans. A substantial fraction of them are conserved over long evolutionary distances [3]. Even some DNA viruses encode miRNAs [4] that act to regulate viral life cycle in the host cell as well as the expression of host genes [5–8].

miRNA biogenesis progresses through multiple steps and involves a collection of enzymes and transport proteins. miRNAs are processed by the Drosha enzyme [9] from hairpin structures that occur in longer coding or non-coding transcripts. The pre-miRNA hairpins are exported out of the nucleus, further sliced into double-stranded RNAs by the Dicer enzyme, and loaded into the Argonaute protein of the RNA-induced silencing complex (RISC) [10–12]. They confer target recognition specificity to the Argonaute protein and guide the RISC to miRNA recognition elements located mostly in the 3' untranslated regions (UTRs) of mRNAs. The binding of RISC to an mRNA results in the rapid repression of translation, decapping, deadenylation, and ultimate degradation of the target mRNA [13, 14].

Many fundamental biological processes, such as metabolism [15, 16], embryogenesis [17], cell cycle [18], cancer [19, 20], epigenetic modification [21, 22], and immunity [23, 24] are now known to be regulated by miRNAs. Given that the functions of many miRNAs that have been isolated in sequencing studies remains to be characterized, one can speculate that many more biological processes will be found to be under miRNA control. Therefore, characterization of miRNA expression and identification of miRNA targets is an important problem. A variety of platforms such as microarrays [25, 26], Sanger sequencing of small RNA libraries [3, 27], or next-generation sequencing [28, 29] can be deployed to identify miRNAs and

characterize their expression. These methods are not discussed in this chapter. Rather, this chapter focuses on the question of predicting the genes that are subject to miRNA-dependent repression.

Many studies addressed the question of how miRNAs find their targets. While in plants, miRNAs bind to nearly perfectly complementary targets [30], in metazoans, it appears that most of the targeting specificity comes from the seven to eight nucleotides at the 5' end of the miRNA, also known as the miRNA "seed" [31–33]. Because miRNAs are conserved in evolution and were originally discovered because of their fundamental role in development, one can make the assumption that functional, physiologically relevant miRNA-binding sites are under evolutionary selective pressure. Indeed, many miRNA target prediction methods make this assumption, aside from requiring extensive pairing of the miRNA seed and/or unusually low free energy of binding between the miRNA and the mRNA [31, 34–36]. The average number of predicted targets per miRNA is in the range of hundreds [37, 38]. This number is in striking contrast with the number of targets that have so far been validated for any individual miRNA. Why the validation rate of miRNA targets appears to be rather low is unclear. One possibility that has been explored in some detail is that not only the seed but also the sequence context in which the miRNA target site resides is important for target site functionality [39, 40]. In this chapter, we describe methods that one can use to uncover properties of functional miRNA sites from high-throughput data and illustrate how these properties can be combined to yield more accurate methods for predicting miRNA target sites *in silico*.

The success of such an approach depends on several factors. First, since the definition of relevant properties relies on experimentally validated sites, it is important that such data be available. For miRNA target prediction in metazoans, this is no longer an issue. Second, the data sets should be of good quality, that is, they should have good reproducibility and exhibit efficient miRNA perturbation, with little side effects from perturbation of the expression of other miRNAs or transcription factors. This is not always the case, not only because of the limited accuracy of high-throughput measurements but also because a miRNA typically acts on many targets and triggers secondary effects that obscure the direct targets. Finally, although the data may allow one to determine whether or not a given property generally characterizes miRNA target sites, it is the analyst who needs to define the set of possibly relevant properties, and thus the success of the approach crucially depends on the creativity of the analyst and his or her understanding of the biological system.

37.2

Description

Our goal is to predict *in silico* functional miRNA-binding sites, which are stretches of mRNA that are bound by the given miRNA *in vivo* to induce posttranscriptional gene silencing. We use for this purpose the so-called binary classification approach,

which is inspired from the machine-learning field [41], and has three main ingredients.

The first ingredient is a *training set* of “positive” and “negative” examples of miRNA-binding sites. In our case, the positive sites are miRNA binding sites that we know – or strongly believe – to be functional. The negative sites, on the other hand, are sites that we know – or strongly believe – to be non-functional. Both sets are equally useful: by analyzing the set of positives we can find out what properties characterize functional miRNA binding sites, while from the set of negatives we can learn what is *specific* to functional sites, that is, what properties discriminate between functional and non-functional sites.

The second ingredient is a set of *features*, or in other words, a list of properties that we think might characterize functional miRNA binding sites. Practically, given a miRNA binding site, we need to be able to compute a set of well-defined quantities associated with the site. These properties may or may not be anchored in the molecular mechanism of miRNA target recognition. For example, the energy of interaction between the miRNA and the target site is a quantity that can in principle “be measured” within a cell, whereas the probability that the site has been under evolutionary selection pressure is not. Which of these properties are indeed relevant in discriminating functional from non-functional will be determined in the analysis.

The last ingredient is a mathematical framework to optimally combine the features into a model that is as accurate as possible in discriminating positive, functional, miRNA binding sites from negative, non-functional yet apparently similar, sites. The complexity of the framework can range from a simple classifier based on a unique feature and an arbitrary cutoff to a sophisticated non-linear classifier allowing for high-order interactions between features. The machine-learning field has developed diverse tools and machines to optimally classify sets of positives and negatives on the basis of a set of features [42]. However, none of these tools is guaranteed to perform better than the others in general: which approach works best depends on the problem. Therefore, designing a good predictor of miRNA targets involves a certain amount of experimentation and trial and error using different classification frameworks.

We now examine each of these three ingredients in more details.

37.2.1

Building a Set of “Positives” and “Negatives”; Obtaining Examples of Functional and Non-functional miRNA Binding Sites

In order to build a reliable model for what a functional miRNA target site looks like, we need to start from a fairly large collection of sites, which is not trivial to obtain. The first known miRNA binding sites were discovered unexpectedly by genetic screens and then validated by immunoprecipitating the protein and northern blotting the cognate mRNA [1, 2, 43]. While genetic screens were fundamental to the discovery of miRNAs and to the characterization of their biological function, they are not well suited for uncovering the targets of a particular miRNA that may act on hundreds of targets to slightly change their level of expression. An

approach that was widely used to uncover miRNA targets consists of expressing a reporter mRNA that contains the coding domain of a fluorescent protein coupled to the 3' UTR of the putative target while the miRNA is expressed in an inducible, tissue-specific manner [34, 44]. Comparing the fluorescence intensity of cells in which the miRNA is expressed with that of cells in which the miRNA is not expressed enables one to answer the question of whether the miRNA indeed acts on the tested construct. However, this approach, which relies on initial predictions of miRNA–mRNA interactions is low throughput.

37.2.1.1 Comparative genomics

Comparative genomics was one of the first computational approaches used to predict sites of miRNA–mRNA interaction. The underlying assumption of this approach is that miRNA binding sites that are critical to the cell are under evolutionary selective pressure and should exhibit stronger sequence conservation across related species relative to genomic regions that are not under selective constraints. Initial studies also determined that the pattern of sequence conservation suggests a model in which the 5' end of the miRNA has extensive base pairing with the target. For this reason, the seven to eight nucleotides at the 5' end of the miRNA have got a special name, that is, the “miRNA seed” [31]. The comparative genomic approach then consists of identifying miRNA seed matches across the 3' UTRs of transcripts and then evaluating the extent of conservation or the degree of selective pressure on these sites. Different miRNA target prediction programs used this idea, either on its own [37, 45, 46] or in combination with other criteria [31, 34, 35, 40, 47, 48]. The comparative genomics approach was instrumental in establishing the principles of miRNA target recognition in metazoans [31] and in assigning biological functions to miRNAs [15, 34, 45].

There are, however, a few conceptual and methodological complications in using comparative genomics to identify functional miRNA binding sites. At the conceptual level, there is first the issue that species- and clade-specific miRNAs have been described [49]. For these, the evolutionary signal may be too weak or, in the extreme, inexistent. In addition, even the deeply conserved miRNAs may acquire novel targets during the evolutionary process, and these recently emerged targets, that may form the substrate of evolutionary innovations, will not be detected by the approach we outlined above. Also from a conceptual point of view, the fact that a sequence conservation pattern is consistent with miRNA-dependent regulation need not exclude the possibility that the site is under selection pressure for an entirely different reason (see [50, 51] for examples of such elements).

On the methodological side, there is the practical issue of defining conservation: some initial studies (e.g., [31] and [52]) in mammals required that the sites be conserved in a specific set of species. The question that arises is in how many and what precise species should one require sequence conservation? More recent comparative-genomics-based methods have set a more rigorous quantitative basis for answering this question [45, 46], although issues such as changes in the pattern of hybridization of a miRNA to conserved sites in different species have not been addressed adequately yet. Last but not least, the comparative genomics approach

does not provide an answer to the question “in what tissue or cell type is a particular target site used or functionally relevant.” Examples of RNA proteins whose binding to an mRNA prevent or compete miRNA-mediated regulation only in particular tissues have been described [53, 54]. Along the same lines, how strongly an mRNA is repressed through the action of a specific miRNA depends on how strongly the miRNA is expressed in the tissue of interest [55].

The factors outlined above set a bound on the usefulness of the comparative genomics approach for miRNA target prediction. To circumvent such limitations, experimental methods have been designed to measure the impact of miRNAs at different levels in high throughput.

37.2.1.2 miRNA perturbation and omics

The most common high-throughput experimental approach to the identification of miRNA targets rests on the observation that miRNAs repress gene expression partly by triggering the degradation of the target mRNAs [56]. The experiment then consists in perturbing the expression of a miRNA and measuring the changes in mRNA levels that are induced by the perturbation with microarrays [57]. The level of the miRNA can be perturbed in different ways: transfection in cell cultures [18, 57] or *in vivo* infection with a viral vector [15] to induce miRNA overexpression, or genetic knockout [58] as well as knockdown with antisense oligos [16] to reduce the miRNA expression. In the case of a miRNA overexpression experiment, likely targets of the miRNA would be found among the mRNAs that harbor a seed match to the overexpressed miRNA and are downregulated following the overexpression of the miRNA.

The major advantages of this approach is that the microarray technology is very accessible, both in terms of availability and cost. miRNA perturbation experiments combined with microarray analysis have greatly contributed to understanding the biological function of miRNAs, and many data sets generated by such experiments are publicly available. However, one should keep in mind that these experiments also have limitations in accuracy (see Section 37.3.2 for details). One limitation stems from the fact that perturbing the expression of a miRNA likely induces perturbations in an entire network that controls transcription and stability of mRNAs. The complexity and our incomplete understanding of that network hinder the interpretation of data obtained from such experiments. In addition, by focusing on changes in mRNA expression levels, the microarray approach may miss targets that are only repressed at the translation regulatory layer, with mRNA expression levels remaining unaffected. The proteomics approach described in the next paragraph provides a solution to this later problem.

Whether miRNAs repress gene expression by increasing the decay rate or only through repressing the translation of the target mRNA, the level of the cognate protein should reflect the action of miRNAs. Based on this idea, another approach to identifying miRNA targets consists of measuring the changes in protein levels following the perturbation of a miRNA by shotgun quantitative proteomics [59, 60]. mRNAs that carry a seed match to the overexpressed miRNA and that encode

proteins that are downregulated following the overexpression of that miRNA are likely to be functional targets of the transfected miRNA.

In theory, high-throughput quantitative proteomics should provide the optimal readout for miRNA-dependent regulation because it directly measures the final outcome of the effect of miRNAs. However, high-throughput proteomics experiments are demanding in terms of equipment and expertise [61], which explains why only a handful of high-throughput proteomics data sets pertaining to the impact of miRNAs has been generated [59, 60]. Furthermore, the coverage of the proteome in quantitative shotgun proteomics experiments is still limited; only 3000–4000 proteins can be measured, compared to 15 000–20 000 transcripts that are measured in microarray experiments. Therefore, this method does not currently enable an exhaustive identification of miRNA targets. Finally, the proteomics approach shares the problem that the perturbation of a miRNA has many secondary effects at the level of transcription, mRNA stability, mRNA translation, or protein decay, which can make the interpretation of the results difficult (see Section 37.3.2).

37.2.1.3 Immunoprecipitation of RISC components

An alternative to measuring the impact of a miRNA perturbation at the mRNA or protein level is to monitor the upstream process of target finding by the miRNA-containing silencing complex. This can be achieved by immunoprecipitating the crucial components of RISC such as the Argonaute proteins [62–66] or GW protein family members such as AIN-1 and AIN-2 [67, 68]. It has been demonstrated that this procedure enables the isolation of not only the protein but also the miRNA and miRNA targets. Strong enrichment of a given mRNA in the immunoprecipitate compared to the lysate indicates a good affinity to the RISC. To isolate more specifically targets of a particular miRNA, these experiments are often combined with miRNA overexpression. mRNAs that are enriched in the immunoprecipitate obtained after overexpression relative to the lysate and that carry seed matches to the overexpressed miRNA are likely targets of the miRNA. Although a systematic comparison with other methods has not been performed, immunoprecipitation of RISC components should be less sensitive to secondary effects of the miRNA perturbation and should enable isolation of targets that would subsequently be repressed either through mRNA decay or at the level of translation. Recent improvements to this approach made it possible to isolate binding sites at very high resolution (basically of individual binding sites) [69, 70], which may in the longer term enable one to infer the miRNA that was bound to the site based on the site sequence and the relative abundance of the miRNAs in the sample. Thus, the need for perturbation experiments for precisely mapping sites of miRNA–target interaction may be alleviated in the future.

Taken together, the existing body of work is consistent with a model in which the binding of RISC to target mRNAs induces mRNA degradation. So far, only a minority of mRNA that are translationally repressed without their mRNAs being destabilized in response to miRNAs have been identified [64, 66]. This suggests that transcriptomics measurements after miRNA perturbation, particularly in

combination with immunoprecipitation of RISC components, remain a reasonable approach for miRNA target identification.

37.2.1.4 Measuring translation repression directly with polysome profiles

Finally, the latest refinement to the experimental identification of miRNA targets came from profiling the translation of different mRNAs in high throughput with sucrose gradients. In the study that introduced this approach to the miRNA field [71], the technique was combined with a miRNA perturbation experiment. Most of the observed changes at the protein level appeared to be explained by the changes in the cognate mRNA level, again supporting the idea that miRNA perturbation experiments combined with microarray analysis identifies most miRNA targets.

Sucrose gradients were only recently applied to study the regulatory impact of miRNAs, and the one published data set that was generated in such a study is insufficient for identifying the general properties of miRNA targets. As more data sets are generated, one will be able to give more definitive answers to questions such as what properties determine whether the binding of a miRNA to a site induces mRNA degradation in addition to translational repression.

37.2.1.5 Which data set should one use for inferring properties that characterize functional miRNA binding sites?

Each of the various computational and experimental approaches to miRNA target identification that we reviewed so far has its strengths and weaknesses in terms of measurement error and biases. Therefore, in order to obtain as general and accurate a picture as possible of the properties of functional miRNA target sites, one would like to compare results across data sets generated by different approaches. The number of published data sets that were obtained with any particular method varies from just one to dozens, and the effect of the miRNAs generally varies from experiment to experiment. There are several criteria that one can use to further zoom into the experiments that give us the most reliable information about miRNA effects, as detailed below.

First of all, as most approaches described so far start with a perturbation of the miRNA expression, one would first like to show that the perturbation was sufficiently strong to have measurable consequences. For example, in a miRNA knockdown experiment, we expect that upregulated genes are enriched in the miRNA seed motif. In the absence of such evidence, it would be difficult to distinguish between a situation in which the experimental conditions were not optimal and one in which the miRNA has only a very small number of targets that cannot give a statistical signal. Similarly, when replicate experiments are performed, we expect that the measurements of mRNA levels are reproducible across replicates. If this is the case, we can focus on targets that show a consistent change across replicates to define the “positives” and targets that are consistently left unchanged by the miRNA perturbation to define the “negatives.” Furthermore, we would like the miRNA perturbation to have as few side effects as possible. This is difficult to predict, but it may be possible to determine *a posteriori* that such effects occurred (see Section 37.2.3).

Finally, we would like the data sets of regulated mRNAs to be sufficiently large. Even when the technology supports the profiling of a large number of genes, some miRNAs may have only few targets and experiments with these miRNAs will not be helpful in uncovering general properties of target sites. Additionally, for any type of experimental measurement (microarray, proteomics, immunoprecipitation, polysome profiles), it is generally desirable to incorporate data sets from different laboratories, to make sure the inferences made from the analysis are not sample or laboratory specific.

37.2.2

Properties of Functional miRNA Binding Sites

Various studies aimed to uncover properties that could explain why, despite exhibiting similar complementarity to the miRNA seed, some miRNA-complementary sites are functional, that is, trigger a change in the mRNA or protein level of the target, while others are not. We will now review the main properties that have been proposed to discriminate functional miRNA binding sites from non-functional ones.

37.2.2.1 The “seed” binding criterion

In plants, miRNAs recognize their mRNA targets through near-perfect base pairing interactions that involve the entire miRNA sequence [30]. A similar model was initially proposed in animals [1, 2], but comparative genomics analyses showed that only the mRNA regions that hybridize to the 5′ end of the miRNA are strongly conserved across species [31]. This led to the view that this region of the miRNA is the most important for contacting the mRNA target. The seven to eight nucleotide region at the 5′ end of the miRNA was consequently termed miRNA “seed” (Figure 37.1).

Various experimental and computational studies lent further support to this view [33, 37, 55, 57]. In addition, it became apparent that the more extensive the pairing between the miRNA and its target, the stronger is the regulatory effect [40, 72], leading to the notion of a “seed hierarchy.” Another twist to the “seed” paradigm of miRNA target recognition is that opposite of the first nucleotide of the miRNA, functional mRNA target sites tend to have an adenosine, whether the first nucleotide of the miRNA is a uridine or not [37, 72]. A similar observation was

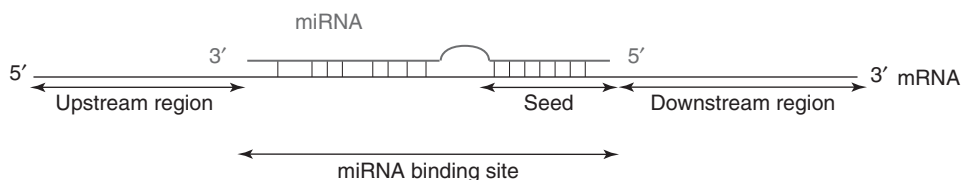


Figure 37.1 “Anatomy” of a miRNA binding site. The miRNA is represented bound to a target mRNA.

made for the nucleotide opposite to the ninth nucleotide of the miRNA: here an A or a U is beneficial for the miRNA-dependent regulation [72].

There are notorious exceptions to the seed binding paradigm, such as the let-7 interaction with lin-41, which was one of the first miRNA–target interactions to be described [43]. Also, introducing G:U base pairs in the region involving the miRNA seed does not appear to always abolish miRNA regulation [39]. In spite of these exceptions, “miRNA seed matching” still appears to account for most of the miRNA targeting specificity, and it is therefore widely used as the starting point for most miRNA target prediction algorithms [35, 37, 38, 40, 45–47, 52, 68, 72–76]. Most functional miRNA binding sites appear to be located in 3′ UTRs, although several studies reported miRNA-dependent repression through sites located in the coding region of mRNAs [22, 77, 78].

37.2.2.2 Evolutionary conservation

A second important criterion that was used to distinguish functional from non-functional miRNA-complementary sites was the conservation of the miRNA binding region [30, 31, 34] or of the surrounding region [72]. One should note, however, that in contrast to all other features that we will be discussing, evolutionary conservation is not a mechanistic feature. If the molecular mechanism of miRNA targeting were understood in its details, one would predict targets on the basis of this mechanism and would not rely on sequence conservation; after all, a cell does not determine whether to respond or not on the basis of a site being conserved, and functional binding sites emerge continuously in evolution. Nonetheless, by analyzing conserved sites, we can learn what properties characterize functional binding sites. Moreover, to the extent to which the information about sequence conservation further improves the accuracy of target predictions that are generated solely on the basis of mechanistic features, we know that we are still to capture properties that are relevant to target recognition by miRNAs.

37.2.2.3 Stability of the miRNA–mRNA duplex

On the basis of the first identified binding sites in *C. elegans* (e.g., the let-7 binding in the lin-41 transcript [43]) and the paradigm of miRNA targeting in plants [30], it was proposed that functional sites are characterized by stable pairing between the miRNA and mRNA. Several approaches to the estimation of miRNA–target interaction stability have been tried, ranging from sequence-alignment-like approaches using *ad hoc* scoring schemes (either for the pairing of the entire miRNA [35] or only of the miRNA 3′ end region [40]), to free energy minimization of single RNA molecules [34] or RNA duplexes [79].

37.2.2.4 Structural accessibility

RNA molecules can form complex internal secondary structures, and if a miRNA binding site is located in a region of the mRNA that forms a highly stable secondary structure such as a stem loop, it is expected that RISC binding may be strongly hindered [80]. In the case of siRNAs, the structural accessibility of binding sites

affects the efficiency of siRNA-mediated gene silencing [81], and the same principle has been proposed to apply to miRNAs [68, 75].

However, these studies do not entirely agree on what region within and around the miRNA target site needs to be accessible to RISC binding. Robins *et al.* [82], for instance, conjectured that three consecutive nucleotides of the miRNA seed need to be paired with accessible nucleotides of the mRNA in order to nucleate the interaction. Tafer *et al.* [81] proposed that the accessibility of the region of the binding site that corresponds to the 8–16 nucleotides at the 5' of an siRNA is critical for RISC binding, while Kertesz *et al.* [75] suggested that the region that needs to be accessible starts three nucleotides upstream of the miRNA binding site and ends 15 nucleotides downstream of the miRNA binding site. Finally, Hammell *et al.* [68] argued that the 25 nucleotides upstream of binding site need to be structurally accessible.

37.2.2.5 Sequence composition

Based on the observation that human genes can be split in two classes, namely, those with AU-rich 3' UTRs and those with GC-rich 3' UTRs, Robins and Press [83] hypothesized that the AU content of the 3' UTR influences the efficacy of miRNA target sites. Grimson *et al.* [40] then showed that the AU composition 30 nucleotides upstream and downstream of the seed match correlated better with miRNA binding site functionality than the overall AU content of the 3' UTR, with the nucleotides close to the seed match being more crucial than the nucleotides more distant from the seed match. After correcting for the effects of sequence conservation, overall 3' UTR AU content, mRNA expression level, and seed match type, Nielsen *et al.* [72] found that an increased AU content of the 50 nucleotides downstream of the seed match correlated with the magnitude of miRNA regulation. Finally, Wang and El Naqa [76] proposed that the dinucleotide composition of the surrounding of the miRNA binding site differs in functional and non-functional miRNA binding sites.

AU-rich elements (AREs) in mRNAs are known to control mRNA stability [84], and at least two RNA binding proteins (HuR and FXR1) that recognize such elements have been reported to interfere with miRNA action [53, 85]. It has also been proposed that a higher AU content implies weaker local secondary structure, making miRNA binding sites more accessible to RISC [40, 72]. However, at the moment, there is no mechanistic explanation for how the AU content in the vicinity of miRNA target sites relates to the efficacy of the site.

37.2.2.6 Spatial effects

Various studies have shown that there are spatial aspects to miRNA targeting. Functional miRNA binding sites are not uniformly distributed within 3' UTRs but tend to be located close to the extremities of 3' UTRs [40, 45, 86], although avoiding the 15 nucleotides directly downstream of the stop codon [40]. In addition, the presence of binding sites for coexpressed miRNAs in close proximity of each other leads to stronger mRNA repression relative to cases where the binding sites are spread farther apart [40].

The properties that we mentioned so far are relatively well known and have already been used in miRNA target prediction. One would typically start with a large number of properties that could conceivably distinguish between functional and non-functional miRNA-complementary site and discover from the data which of these properties are indeed relevant.

37.2.3

Combining Properties and Examples into a Predictive Model

At this point, we have outlined the procedure for constructing sets of miRNA-complementary sites that we believe to be functional (“positives”) or non-functional (“negatives”) and defined a set of properties (“features”) that we think may characterize functional miRNA binding sites and may distinguish them from the non-functional sites. We can now combine these two ingredients to find out what features best discriminate functional miRNA binding sites from non-functional ones in general, and then proceed to training a miRNA target prediction model.

37.2.3.1 Inferring properties that consistently predict miRNA targeting across data sets

To illustrate how different data sets can be combined in order to understand better what properties characterize functional miRNA binding sites, we obtained examples of functional and non-functional miRNA binding sites from 46 experimental and computational data sets that were generated while exploring the impact of miRNA regulation at the level of mRNA stability, protein translation rate, and evolutionary sequence conservation. The experiments were performed on a few different human cell lines in different laboratories, using different measurement technologies. The power that the features defined in the previous section have in predicting the functionality of miRNA binding sites in different experiments was quantified by computing the *t*-value comparing positive and negative sites for each feature and each experiment. The results of the analysis are shown as a heat map in Figure 37.2, where columns represent experiments and rows represents features. For each experiment (column), the color on the heat map indicates whether the value of a given feature (row) correlated positively (red, *t*-value > 0) or negatively (blue, *t*-value < 0) with functionality. The intensity of the color indicates the strength of that correlation, that is, the predictive power of the feature in the considered data set. For instance, the third column starting from the right of Figure 37.2 summarizes the comparisons of features of miR-124-complementary sites with a high and a low inferred probability of being under evolutionary selection [45] in human. The third cell from the top of that column, labeled “target site Eopen” is dark blue, meaning that the energy required to open the secondary structure of the putative binding site is significantly smaller for sites with high probability of being under evolutionary selection. This in turn suggests that evolutionary selection favored miR-124-complementary sites that are more accessible to RISC binding at the level of mRNA secondary structure. The fifth cell from the top of this column,

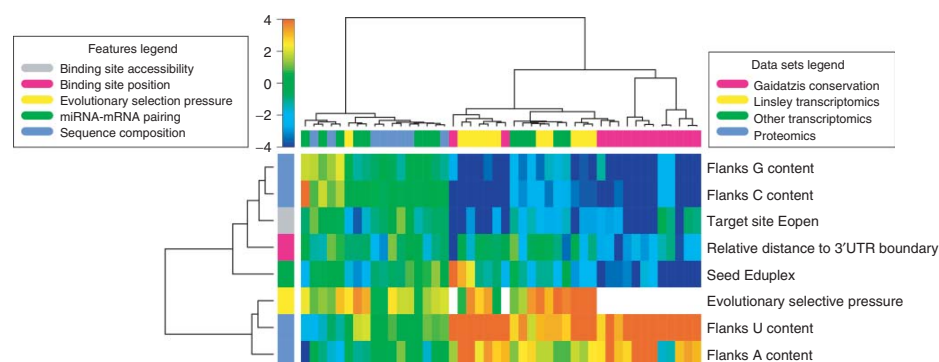


Figure 37.2 Predictive power of eight features of putative miRNA binding sites (rows) in predicting functional sites across 46 data sets (columns). A more complete version of this figure with 38 features and 72 data sets can be found in Hausser *et al.* [66]. The data sets cover transcriptomics measurements ("Linsley transcriptomics" [18] and "other transcriptomics" [40, 59, 60]) and proteomics measurements ("proteomics" [59, 60]) following miRNA overexpression transfection as well as target prediction based on comparative genomics ("Gaidatzis conservation" [45]). The heat map shows the t-values comparing the distributions of feature values in functional and non-functional miRNA binding sites. The red color indicates positive predictors of miRNA functionality, while the blue color, negative predictors of miRNA functionality. The dendrograms of features and data sets were produced through hierarchical clustering using Ward linkage on the euclidean space of t-values. This analysis suggests that miRNA binding sites that trigger mRNA degradation tend to be structurally accessible, feature extensive pairing between the seed region of the miRNA and the mRNA, and occur in A/U-rich 3' UTRs regions, preferentially close to the stop codon or the poly-A tail. The same properties characterize miRNA-complementary sites that are under evolutionary selective pressure. On the other hand, in terms of these properties, miRNA-complementary sites leading to a decrease in translation rate do not seem to be distinguishable from the sites that do not.

labeled “seed Eduplex” is also dark blue, indicating that the free energy of binding between the miRNA seed and the putative binding site is significantly lower – that is, the interaction is more stable – for sites with high probability of being under evolutionary selection. In contrast, the second cell from the bottom of the column, labeled “flanks U content” is dark red. This indicates that the frequency of U nucleotides is significantly higher in the regions flanking miR-124-complementary sites that have high probability of being under evolutionary selection.

Determining which data sets are characterized by similar features yields additional insights. Hierarchical clustering of the columns shows that data sets obtained from miRNA perturbation and microarray measurements of mRNA stability are most similar to data sets obtained from comparative genomics analysis (right side of Figure 37.2). In these data sets, functional miRNA binding sites are consistently characterized by relatively high A and U contents, both locally (in the 50 nucleotides surrounding the miRNA-complementary site, as shown in the figure) as well as globally (in the entire 3' UTR or mRNA that harbors the site – see *Hausser et al.* [66]). In addition, functional sites appear to be structurally more accessible and to allow more extensive pairing to the miRNA seed. Other features – among which those related to location of miRNA-complementary sites in the transcript – appear to be weaker predictors of functional miRNA binding sites, or to be specific to certain data sets. The strong consensus in the properties that characterize miRNA-complementary sites under evolutionary pressure and those leading to mRNA degradation suggests that sites that can induce mRNA degradation play an important role and are maintained in evolution.

The second cluster of data sets is more heterogeneous. It contains a group of experiments in which none of the studied features were predictive of the functionality of miRNA-complementary sites and experiments in which the correlation between features sequence related to sequence composition and functionality was in opposite direction to the one expected (left side of Figure 37.2). These experiments mostly involved miRNA overexpression and measurements of protein translation rate and a subset of the corresponding microarray-based measurements of mRNA stability. In all these experiments, the miRNA seed-complementary motif was enriched in genes that were downregulated following the miRNA overexpression, ruling out the possibility that the miRNA overexpression was too weak to trigger a measurable effect on gene regulation. At present, it is unclear why miRNA binding sites that are functional in these experiments do not share the properties that typically characterize functional miRNA binding sites. In the case of the experiments in which the direction of the correlation between site functionality and sequence-composition-related features was opposite to the one expected, the observed inconsistency is probably due to a side effect of the transfection, namely, the inhibition of the RNAi machinery by the overexpressed miRNA [87]. This analysis indicates that for the sake of training a miRNA target prediction algorithm, one cannot consider all data sets as a whole. Instead, it may be more appropriate to exclude the data sets of the later cluster, or to consider them separately from the rest.

To summarize, by comparing features of “positive” and “negative” examples of miRNA-complementary sites, we identified properties that characterize miRNA-complementary sites that are functional in a large number of data sets and are therefore unlikely to be the consequence of biases introduced of a specific technology or experimental method. In addition, this analysis enabled us to identify outlier data sets that we may want to exclude from training the model and/or we may want to analyze in depth to uncover miRNA-induced effects that have not been described so far. Finally, this analysis provided us with insights into the mechanisms that shape the evolution of 3' UTR sequences and miRNA-complementary sites.

37.2.3.2 Training a miRNA target prediction model

There are several ways to combine features that were found to distinguish “positive” from “negative” miRNA-complementary sites into a model for miRNA target prediction.

One approach that was previously used in the literature was to simply select a set of relevant features and cutoffs on the values of these features, on the basis of prior knowledge (biochemical measurements, analyses of known sites). Some of the first algorithms for miRNA target prediction started from 7- or 8 mers matches to miRNA seeds, then screened these candidates for sequence conservation (the number and identity of the species used in the prediction were tuned depending on the study), and then evaluated the conserved sites in terms of the binding free energy. Those candidates with a free energy of binding smaller than a cutoff value were retained as predicted miRNA binding sites. This type of approach had a strong empirical aspect and relied on a small set of known binding sites that were available at the time. Nonetheless, the approach was successfully used to uncover more examples of miRNA binding sites, which in turn helped in further understanding the molecular function of miRNAs [15, 16, 88].

As high-throughput measurement technologies started to be applied to the miRNA targeting problem and more data sets of miRNA binding sites were generated [57, 59, 60, 62, 67, 71], it became apparent that no simple combination of criteria sufficiently well characterized functional miRNA binding sites [40]. Rather, to improve the accuracy of miRNA target prediction, linear models that combined a number of carefully chosen criteria to predict the effect of miRNAs on the degradation rate of target mRNAs [40, 66] have been developed. Contrary to the filter-based approach that employs “hard” cutoffs, a linear model allows to smoothly combine properties and thereby account for trade-offs between different properties of miRNA binding sites. For instance, even when the A/U content of the sequence around a miRNA-complementary site is low, the site may still be functional if the miRNA can form a very stable duplex with the mRNA. Deciding what features to use in the prediction of miRNA-induced effects is critical, because it is typical that not all of the properties that have been studied are informative on one hand, while other features, for instance, the stability of the secondary structure around the target site and the nucleotide composition in this region, are correlated.

This problem, which in the machine-learning field is known as the *feature selection* problem, is beyond the scope of this chapter [89].

The availability of larger sets of “positive” and “negative” miRNA-complementary sites also enables the development of more elaborate classification frameworks such as support vector machines that can account for interactions between features. For instance, the position of the miRNA-complementary site may be predictive of its functionality but only in long 3′ UTRs [45]. Aside from support vector machines, the machine-learning field has developed a variety of classification frameworks (e.g., decision trees, neural networks, etc. [42]), which have already been applied to miRNA target prediction [76]. Nevertheless, such elaborate methods did not yet provide a strong improvement in miRNA target prediction accuracy compared to simpler linear models, whose accuracy is currently comparable to that of biological duplicates with low reproducibility [66].

To summarize, we have so far described the concepts behind a large number of miRNA target prediction methods. Some of these methods, that can be accessed online through web browsers, are listed in Table 37.1.

37.3 Troubleshooting

What we have not addressed up to this point are the practical aspects of using miRNA target predictions in the framework of a specific biological question. We have also not elaborated on some of the steps involved in designing miRNA target prediction algorithms that are not straightforward and are dependent on aspects of gene regulation, which are at present only partially understood.

In this section, we address these questions. We start by examining issues that arise when using miRNA target predictions for answering specific biological questions in an experimental setting. Then, we discuss how some of the uncertainties regarding gene regulation are reflected in computational miRNA target predictions and to what extent one can deal with these uncertainties.

37.3.1 Using miRNA target predictions in an experimental setting

A typical setting in which miRNA target predictions are useful is when the miRNAs that are involved in a specific process have been identified by miRNA expression profiling or genetic screens, and the question becomes what targets respond to these miRNAs. Target predictions are then necessary to guide target discovery. This approach has been used in numerous studies aiming to understand the role of miRNAs in development [1, 2, 43], insulin secretion [15], cholesterol biosynthesis [16], or pathologies [7, 20], to mention only a few cases.

Table 37.1 A non-exhaustive list of online miRNA target prediction methods, listed in alphabetical order. The columns indicate the name of the method, the URL of the web server, the organisms for which predictions are available (cel: *C. elegans*, dme: *Drosophila melanogaster*, dre: *Danio rerio*, hsa: *Homo sapiens*, mmu: *Mus musculus*, rno: *Rattus norvegicus*, ... : others), and the reference article.

Method and URL	Organisms	Reference
DIANA-microT http://diana.cslab.ece.ntua.gr/microT/	hsa, mmu	[74]
ELMMo http://www.mirz.unibas.ch/ElMMo3/	cel, dme, dre, hsa, mmu, rno	[45]
EMBL http://www.russell.embl.de/miRNAs/	dme	[47]
MicroCosm http://www.ebi.ac.uk/enright-srv/microcosm/	cel, dme, dre, hsa, mmu, rno, ...	[35]
microrna.org http://www.microrna.org/	hsa, mmu, rno	[52]
miRDB http://mirdb.org/	hsa, mmu, rno, ...	[76]
mirWIP http://146.189.76.171/query.php	cel	[68]
PicTar http://pictar.mdc-berlin.de/	cel, dme, hsa, mmu	[90]
Pita http://genie.weizmann.ac.il/pubs/mir07/rna22	cel, dme, hsa, mmu	[75]
TargetRank http://genes.mit.edu/targetrank/	hsa, mmu	[72]
TargetScan context http://www.targetscan.org/	hsa, mmu, rno, ...	[40]
TargetScan PCT http://www.targetscan.org/	cel, dme, hsa, mmu, rno, ...	[46]

37.3.1.1 How accurate are miRNA target predictions?

If miRNA target prediction is essential for the identification of miRNA targets, the immediate question is what target prediction program should an experimental biologist use. The literature does not provide a clear answer to this question for a number of reasons. First is that it is still unclear what experimental data is suitable to assess the quality of miRNA target predictions. Are measurements of mRNA stability sufficient, or does one really need measurements of the protein levels in order to identify miRNA targets? Second, given that miRNAs appear to act at multiple levels, it is unclear how we should treat the predictions. Should one require that the response to miRNA perturbations is predicted quantitatively or would it be sufficient to predict whether the target responds or not? Another complication in comparing target prediction methods is that there is no standard

set of potential targets that is always considered. That is, some authors use Refseq transcripts, others use transcripts from specialized databases (such as WormBase [91] or FlyBase [92]); even the content of a given database (such as Refseq) changes in time, and thus one method may have a prediction that another is missing simply because the transcript was not even considered. Nonetheless, we attempted to perform such a comparison on a subset of the available, commonly used, miRNA target prediction programs and a set of genes and transcripts that were used in all of these programs. Our results [66] indicated that there are a number of miRNA target prediction programs that perform comparably well and explain around 20% of the variance in the changes in gene expression induced in a miRNA perturbation experiment. This means that (i) there is no clear “best miRNA target prediction method” and (ii) the predictive power of the best miRNA target prediction methods available today is comparable to that of duplicated experiments of low reproducibility [66]. One category of targets that is generally not predicted by the methods that are currently available is that of the so-called non-canonical miRNA binding sites, that is, sites that cannot extensively pair with the miRNA seed. While such sites were in fact among the first to be identified [43], at present, attempts to predict them come at the cost of a dramatic loss of specificity [37, 45] and, on average, their impact on the mRNA stability appears to be limited *in vivo* [70].

However, one thing that we need to keep in mind is that high-throughput experiments of miRNA perturbations were frequently done on (cancer) cell lines. It is likely that in such systems the miRNA effects are more easily interpretable compared to *in vivo* situations, and moreover, that miRNA transfections usually lead to large changes in miRNA expression that may not be typical of the *in vivo* situations. Additional pitfalls arising from the use of high-throughput data sets for assessing the accuracy of miRNA target predictions are discussed in the second part of Section 37.3.2. To circumvent the current limitations of high-throughput methods, one can focus instead on high-quality, experimentally validated “positives” and “negatives” [93], although the set of such targets is significantly smaller and perhaps not even representative for the entire set of miRNA targets.

37.3.1.2 Which miRNA target prediction method should I use?

Because no method can currently predict miRNA targets very accurately, some authors attempted to obtain high-confidence predictions by intersecting the results of several prediction methods. While this idea may sound reasonable in theory, it may not necessarily result in more accurate predictions in practice, as can be illustrated by a simple example.

Let us imagine that we are given the list of genes that are predicted to be targeted by a certain miRNA by two different methods. Let us further assume that method A is a very accurate prediction method, while method B simply consists of tossing a coin for each gene in the human genome, calling the gene a predicted target when the coin toss returns heads. Intersecting the two lists of “predictions” we obtain a “random” subset of the targets predicted by method A. The fraction of real targets within this subset (the positive predictive value) will remain the same as for method A. The fraction of real targets that will indeed be predicted as targets

(the sensitivity) in the intersection will be half of that in the list of predictions from method A alone. Thus, the accuracy of this method that combines prediction lists is lower than the accuracy a single method (method A). In fact, the situation may be even worse when the current best method is not among those whose prediction lists will be intersected. Thus, in the context of an experiment-driven project, when the aim is to find what genes regulated by a given miRNA explain a certain phenotype, one should rather start by considering the assumptions made by the different prediction methods that are available. For instance, if the miRNA of interest is conserved in evolution, we may expect that its targets are also conserved and we could consider miRNA target prediction methods that rely on sequence conservation. If the miRNA is itself poorly conserved, one should rather consider methods that aim to predict miRNA-complementary sites that induce mRNA destabilization, even though it is presently unclear whether these methods have higher or lower accuracy relative to comparative-genomics-based methods.

37.3.1.3 How many targets does a miRNA have?

One of the main aims of miRNA target prediction is to generate a list of genes sorted in descending order of confidence of the gene being a miRNA target. Depending on the method and on the miRNA, one typically finds that the number of predicted targets ranges from a dozen to thousands of genes. Therefore, a natural question arising from such lists is how many targets does a miRNA have?

This is again a difficult question to answer in principle, not in the least because, as we discussed above, it is not entirely clear how to define a miRNA target. From the point of view of computational predictions, some miRNA target prediction methods propose criteria that are typically based on statistical considerations (signal to noise ratio, a posteriori probability, etc.) to decide where to cut off the list of predicted targets to be considered for experimental validation. With this approach, it has been inferred that a miRNA targets on an average hundreds of genes, the number varying between a couple and thousands of genes [31, 46, 57, 59, 60] for individual miRNAs.

37.3.1.4 Why does a particular high-confidence predicted target not change in response to miRNA overexpression?

It is not uncommon to find out that a substantial fraction of high-confidence predicted targets do not respond in a particular validation experiment. Of course, a trivial possibility is that the prediction is erroneous. In Section 37.3.2, we discuss several scenarios in which the target cannot be validated experimentally, even though it is indeed a target.

37.3.1.5 Transcript x is a target of miRNA y according to method z, yet it does not have an “miRNA y seed match” in the 3' UTR.

A frequently overlooked cause for discrepancies between target prediction methods or between the predictions and validation experiments is that the sequences associated with specific transcript identifiers change between database releases. Thus, predicted miRNA targets that have been downloaded from the web sites

associated with specific methods may differ in sequence from the transcripts that one can download at a later date from databases such as NCBI, Ensembl, or DDBJ.

37.3.1.6 The list of targets predicted by method x has a different type of identifiers (Entrez Gene ID/RefSeq ID/Ensembl transcript/ . . .) than the list predicted by method y or the list that one obtains in a large-scale validation experiment (e.g., microarray measurement)

Different miRNA target prediction methods may start from different databases of transcripts (putative targets). The websites of most databases of biological sequence data are typically cross-referenced and can be searched with different types of identifiers. When dealing with unknown sequence identifiers, the NCBI's search tool at <http://www.ncbi.nlm.nih.gov/> and EBI's "EB-Eye" tool at <http://www.ebi.ac.uk/> provide a good starting point.

37.3.2

The Complexity of Gene Regulation and its Impact on Designing Accurate miRNA Target Prediction Methods

In order to establish that there is a direct interaction between a miRNA and an mRNA target, one possibility is perturbing the miRNA binding site by deletion or mutation. This approach was largely used to understand the fundamentals of miRNA target recognition [33], and is well suited for confirming experimentally a small set of putative targets. However, this approach is not practical for the identification and validation of all target sites of a miRNA. For this purpose, the alternative approach that consists in perturbing the miRNA by overexpression, knockdown, or mutation was widely used to find what genes are targeted by miRNAs [16, 59, 60]. The downside of such experiments is that the perturbation of the miRNA will percolate through the regulatory networks of the cell, which respond at different time scales, which in the end will complicate the interpretation of the perturbation experiment. This is one reason why for instance, not all mRNAs that are downregulated following miRNA overexpression harbor a miRNA seed match [57]. "Secondary" effects superimpose with the "direct" miRNA-induced gene silencing effects, complicating the selection of "positive" and "negative" binding sites that one would use to train the predictive model. We review briefly the scenarios in which an incorrect selection of "positive" or "negative" binding sites is made, particularly those due to such secondary effects.

The most obvious causes for an incorrect identification of positives and negatives are the technical measurement error (noise) and the intrinsic biological variability. Such errors can be reduced by improving the technology and performing a larger number of replicated measurements. Another obvious factor is the concentration of the reactants (mRNA, miRNAs). The target mRNA needs to be expressed at a level that allows its detection and an accurate estimation of the change in response to the miRNA. If these conditions do not hold for the cell line used in the miRNA perturbation experiment, the mRNA will not be identified as a target. The concentration of the miRNA matters as well: knocking out a miRNA that

is only present in trace amount or overexpressing a miRNA whose expression is already very high in the cell line in which the experiment is performed is unlikely to produce a measurable change in the expression of the targets.

The “secondary effects” come from the interaction of the perturbed miRNA with the gene regulation network of the host cell. Because regulatory networks have not been sufficiently characterized quantitatively, one cannot simply predict the secondary effects of a miRNA overexpression. Nonetheless, given that miRNAs and transcription factors coregulate targets, with the miRNA often regulating the transcription factor or the transcription factor regulating the miRNA [94], such effects are expected to be important. Consider the example of a hypothetical experiment in which one overexpresses a miRNA that silences a transcription factor, which activates the transcription of mRNA A. We assume that A harbors a non-functional miRNA binding site (Figure 37.3). By inferring “positive” miRNA binding sites from mRNAs whose levels go down following the miRNA overexpression, one would treat mRNA A as a functional target of the overexpressed miRNA (a false “positive”). On the other hand, we can consider the hypothetical case of a miRNA that silences a transcription factor that represses an mRNA B. Let us further assume that mRNA B harbors a functional binding site for the overexpressed miRNA (Figure 37.3). Overexpressing the miRNA will lead to a downregulation of the transcriptional repressor, which could then result in an upregulation of mRNA B despite the direct regulation by the miRNA. In this case, from the measured changes in expression, we would treat mRNA B as non-functional miRNA target (a false “negative”). Other known examples of such secondary effects include the targeting of components of the miRNA pathway itself such as Dicer [60, 77] or TNRC6 [66], or of enzymes involved in chromatin structure [21, 22] by the perturbed miRNA. In addition, the overexpression of a miRNA has been shown to impact the posttranscriptional regulatory network of a cell through competition with the endogenously expressed miRNAs [95].

Other posttranscriptional regulatory mechanisms (that may be triggered in the miRNA perturbation experiment) have been reported to interfere with miRNA regulation. For instance, under stress conditions, RNA binding proteins can relieve miRNA-mediated translational repression [53], and miRNAs may even switch from acting as repressors to acting as activators [85]. Along the same lines, the RNA binding protein Dnd1 can relieve miRNA-dependent inhibition by blocking the

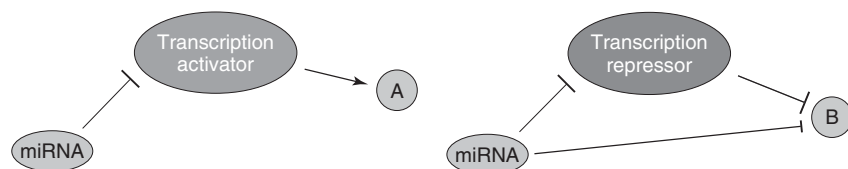


Figure 37.3 Hypothetical networks illustrating the coregulation of a gene (A, B) by a miRNA and a transcription factor, with the miRNA regulating the transcription factor. Gene A harbors a non-functional miRNA binding site, while gene B carries a functional miRNA binding site.

access of the miRNA to its site [54]. Thus, there are a variety of factors that can change the functionality of miRNA target sites in a context-dependent manner.

Strongly expressing a transcript with multiple sites complementary to a given miRNA was shown to derepress the targets of that miRNA [96, 97], which suggests that miRNA targeting is transcriptome dependent; in the context of a certain transcriptome, miRNA binding may be functional, but in the presence of another strongly expressed mRNA that recruits most copies of the miRNA, the same target may be derepressed and therefore appear non-functional [98, 99]. As the transcriptome largely depends on the tissue and the experimental conditions, taking into account the cell-type-specific transcriptome may lead to more accurate identification of positive and negative sites and more accurate miRNA target predictions.

Finally, although it is generally accepted that miRNAs silence their target genes both through repressing the translation of the transcript and by promoting deadenylation leading to the degradation of the target mRNA [13, 100], it can not be excluded that a set of target mRNAs are only repressed translationally without undergoing degradation [66]. For these sites, the transcriptomics approach is bound to generate false negatives.

These examples illustrate that obtaining sets of positive and negative miRNA binding sites from miRNA perturbation and omics experiments is not trivial and is necessarily associated with a certain amount of “noise.” The advantage of the large amount of data produced by omics experiments is that this noise is expected to cancel out if enough sites from different experiments with different miRNAs are used to train miRNA target prediction algorithms. The fact that miRNA target sites obtained from heterogeneous sources share similar properties argues for this scenario [66].

The discussion above also suggests that there are miRNA targets that have been predicted computationally but have not been validated experimentally because in the experimental context used to validate the target, the real effect of the miRNA on the putative target is masked by one or more factors mentioned above. Reconciling miRNA target predictions with validation experiments will require taking the precise context of the validation experiments into account.

References

1. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, **75**, 843–854.
2. Wightman, B., Ha, I., and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*, **75**, 855–862.
3. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.
4. Pfeffer, S., Zavolan, M., Grässer, F.A., Chien, M., Russo, J.J., Ju, J., John, B., Enright, A.J., Marks, D., Sander, C., and Tuschl, T. (2004) Identification of virus-encoded microRNAs. *Science*, **304**, 734–736.

5. Sullivan, C.S., Grundhoff, A.T., Tevethia, S., Pipas, J.M., and Ganem, D. (2005) SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature*, **435**, 682–686.
6. Nair, V. and Zavolan, M. (2006) Virus-encoded microRNAs: novel regulators of gene expression. *Trends Microbiol.*, **14**, 169–175.
7. Samols, M.A., Skalsky, R.L., Maldonado, A.M., Riva, A., Lopez, M.C., Baker, H.V., and Renne, R. (2007) Identification of cellular genes targeted by KSHV-encoded microRNAs. *PLoS Pathogens*, **3**, e65.
8. Gottwein, E., Mukherjee, N., Sachse, C., Frenzel, C., Majoros, W.H., Chi, J.-T.A., Braich, R., Manoharan, M., Soutschek, J., Ohler, U., and Cullen, B.R. (2007) A viral microRNA functions as an orthologue of cellular miR-155. *Nature*, **450**, 1096–1099.
9. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., and Kim, V.N. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415–419.
10. Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell*, **106**, 23–24.
11. Hutvágner, G. and Zamore, P.D. (2002) A microRNA in a multiple-turnover RNAi enzyme complex. *Science*, **297**, 2056–2060.
12. Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.*, **16**, 720–728.
13. Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.*, **9**, 102–114.
14. Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008) Getting to the root of miRNA-mediated gene silencing. *Cell*, **132**, 9–14.
15. Poy, M.N., Eliasson, L., Krützfeldt, J., Kuwajima, S., Ma, X., Macdonald, P.E., Pfeffer, S., Tuschl, T., Rajewsky, N., Rorsman, P., and Stoffel, M. (2004) A pancreatic islet-specific microRNA regulates insulin secretion. *Nature*, **432**, 226–230.
16. Krützfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M., and Stoffel, M. (2005) Silencing of microRNAs *in vivo* with ‘antagomirs’. *Nature*, **438**, 685–689.
17. Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science*, **312**, 75–79.
18. Linsley, P.S., Schelter, J., Burchard, J., Kibukawa, M., Martin, M.M., Bartz, S.R., Johnson, J.M., Cummins, J.M., Raymond, C.K., Dai, H. *et al.* (2007) Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol. Cell. Biol.*, **27**, 2240–2252.
19. Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K. *et al.* (2002) Frequent deletions and down-regulation of microRNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 15524–15529.
20. He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J., and Hammond, S.M. (2005) A microRNA polycistron as a potential human oncogene. *Nature*, **435**, 828–833.
21. Sinkkonen, L., Hugaschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C.G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008) MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.*, **15**, 259–267.

22. Duursma, A.M., Kedde, M., Schrier, M., le Sage, C., and Agami, R. (2008) miR-148 targets human DNMT3b protein coding region. *RNA*, **14**, 872–877.
23. Thai, T.-H., Calado, D.P., Casola, S., Ansel, K.M., Xiao, C., Xue, Y., Murphy, A., Frendewey, D., Valenzuela, D., Kutok, J.L. *et al.* (2007) Regulation of the germinal center response by microRNA-155. *Science*, **316**, 604–608.
24. O'Connell, R.M., Taganov, K.D., Boldin, M.P., Cheng, G., and Baltimore, D. (2007) MicroRNA-155 is induced during the macrophage inflammatory response. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 1604–1609.
25. Calin, G.A., Liu, C.-G., Sevignani, C., Ferracin, M., Felli, N., Dumitru, C.D., Shimizu, M., Cimmino, A., Zupo, S., Dono, M. *et al.* (2004) MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 11755–11760.
26. Miska, E.A., Alvarez-Saavedra, E., Townsend, M., Yoshii, A., Sestan, N., Rakic, P., Constantine-Paton, M., and Horvitz, H.R. (2004) Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biol.*, **5**, R68.
27. Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., Lin, C. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
28. Hafner, M., Landgraf, P., Ludwig, J., Rice, A., Ojo, T., Lin, C., Holoch, D., Lim, C., and Tuschl, T. (2008) Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods*, **44**, 3–12.
29. Olson, A.J., Brennecke, J., Aravin, A.A., Hannon, G.J., and Sachidanandam, R. (2008) Analysis of large-scale sequencing of small RNAs. Pacific Symposium on Biocomputing, pp. 126–136.
30. Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002) Prediction of plant microRNA targets. *Cell*, **110**, 513–520.
31. Lewis, B.P., Shih, I.-H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.
32. Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003) The microRNAs of *Caenorhabditis elegans*. *Genes Dev.*, **17**, 991–1008.
33. Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005) Principles of microRNA-target recognition. *PLoS Biol.*, **3**, e85.
34. Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003) Identification of *Drosophila* microRNA targets. *PLoS Biol.*, **1**, E60.
35. Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003) MicroRNA targets in *Drosophila*. *Genome Biol.*, **5**, R1.
36. Rajewsky, N. and Socci, N.D. (2004) Computational identification of microRNA targets. *Dev. Biol.*, **267**, 529–535.
37. Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
38. Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A.M., Lim, B., and Rigoutsos, I. (2006) A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell*, **126**, 1203–1217.
39. Didiano, D. and Hobert, O. (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat. Struct. Mol. Biol.*, **13**, 849–851.
40. Grimson, A., Farh, K.K.H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cells*, **27**, 91–105.
41. Hastie, T., Tibshirani, R., and Friedman, J. (2001) *The Elements of Statistical Learning*. Springer.
42. Duda, R.O., Hart, P.E., and Stork, D.G. (2000) *Pattern Classification*, Wiley-Interscience.

43. Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, **403**, 901–906.
44. Schagat, T. and Vidugiriene, J. (2007) microRNA biosensors: application for the psiCHECK-2 vector. *Cell Notes*, **19**, 4–7.
45. Gaidatzis, D., Nimwegen, E., Hausser, J., and Zavolan, M. (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics*, **8**, 69.
46. Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
47. Stark, A., Brennecke, J., Bushati, N., Russel, R.B., and Cohen, S.M. (2005) Animal microRNAs confer robustness to gene expression and have a significant impact on 3' UTR evolution. *Cell*, **123**, 1133–1146.
48. Krek, A., Grün, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., and Rajewsky, N. (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.
49. Berezikov, E., Thüemmler, F., van Laake, L.W., Kondova, I., Bontrop, R., Cuppen, E., and Plasterk, R.H.A. (2006) Diversity of microRNAs in human and chimpanzee brain. *Nat. Genet.*, **38**, 1375–1377.
50. Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.
51. Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, **446**, 926–929.
52. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. (2004) Human MicroRNA targets. *PLoS Biol.*, **2**, e363.
53. Bhattacharyya, S.N., Habermacher, R., Martine, U., Closs, E.I., and Filipowicz, W. (2006) Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell*, **125**, 1111–1124.
54. Kedde, M. and Agami, R. (2008) Interplay between microRNAs and RNA-binding proteins determines developmental processes. *Cell Cycle*, **7**, 899–903.
55. Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev*, **18**, 504–511.
56. Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. (2005) Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell*, **122**, 553–563.
57. Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769–773.
58. Zhao, Y., Ransom, J.F., Li, A., Vedantham, V., Drehle, M., Muth, A.N., Tsuchihashi, T., McManus, M.T., Schwartz, R.J., and Srivastava, D. (2007) Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell*, **129**, 303–317.
59. Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.
60. Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.
61. Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.

62. Beitzinger, M., Peters, L., Zhu, J.Y., Kremmer, E., and Meister, G. (2007) Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol.*, **4**, 76–84.
63. Easow, G., Teleman, A.A., and Cohen, S.M. (2007) Isolation of microRNA targets by miRNP immunopurification. *RNA*, **13**, 1198–1204.
64. Karginov, F.V., Conaco, C., Xuan, Z., Schmidt, B.H., Parker, J.S., Mandel, G., and Hannon, G.J. (2007) A biochemical approach to identifying microRNA targets. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 19291–19296.
65. Hendrickson, D.G., Hogan, D.J., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2008) Systematic identification of mRNAs recruited to argonaute 2 by specific microRNAs and corresponding changes in transcript abundance. *PloS One*, **3**, e2126.
66. Hausser, J., Landthaler, M., Jaskiewicz, L., Gaidatzis, D., and Zavolan, M. (2009) Relative contribution of sequence and structure features to the mRNA binding of Argonaute/Eif2C-miRNA complexes and the degradation of miRNA targets. *Genome Res.*, **19**, 2009–2020.
67. Zhang, L., Ding, L., Cheung, T.H., Dong, M.-Q., Chen, J., Sewell, A.K., Liu, X., Yates, J.R., and Han, M. (2007) Systematic identification of *C. elegans* miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Mol. Cell*, **28**, 598–613.
68. Hammell, M., Long, D., Zhang, L., Lee, A., Carmack, C.S., Han, M., Ding, Y., and Ambros, V. (2008) mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein enriched transcripts. *Nat. Methods*, **5**, 813–819.
69. Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.
70. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
71. Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009) Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biol.*, **7**, e1000238.
72. Nielsen, C.B., Shomron, N., Sandberg, R., Hornstein, E., Kitzman, J., and Burge, C.B. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, **13**, 1894–1910.
73. Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., and Hatzigeorgiou, A. (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.*, **18**, 1165–1178.
74. Maragkakis, M., Alexiou, P., Papadopoulos, G.L., Reczko, M., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., Simossis, V.A. et al. (2009) Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, **10**, 295.
75. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
76. Wang, X. and El Naqa, I.M. (2008) Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics*, **24**, 325–332.
77. Forman, J.J., Legesse-Miller, A., and Collier, H.A. (2008) A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 14879–14884.
78. Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, **455**, 1124–1128.

79. Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.
80. Ameres, S.L., Martinez, J., and Schroeder, R. (2007) Molecular basis for target RNA recognition and cleavage by human RISC. *Cell*, **130**, 101–112.
81. Tafer, H., Ameres, S.L., Obernosterer, G., Gebeshuber, C.A., Schroeder, R., Martinez, J., and Hofacker, I.L. (2008) The impact of target site accessibility on the design of effective siRNAs. *Nat. Biotechnol.*, **26**, 578–583.
82. Robins, H., Li, Y., and Padgett, R.W. (2005) Incorporating structure to predict microRNA targets. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 4006–4009.
83. Robins, H. and Press, W.H. (2005) Human microRNAs target a functionally distinct population of genes with AT-rich 3' UTRs. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15557–15562.
84. Barreau, C., Paillard, L., and Osborne, H.B. (2005) AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res.*, **33**, 7138–7150.
85. Vasudevan, S., Tong, Y., and Steitz, J.A. (2007) Switching from repression to activation: microRNAs can up-regulate translation. *Science*, **318**, 1931–1934.
86. Majoros, W.H. and Ohler, U. (2007) Spatial preferences of microRNA targets in 3' untranslated regions. *BMC Genomics*, **8**, 152.
87. Hausser, J., Berninger, P., Rodak, C., Jantscher, Y., Wirth, S., and Zavolan, M. (2009) MirZ: an integrated microRNA expression atlas and target prediction resource. *Nucleic Acids Res.*, **37**, W266–W272.
88. Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003) *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell*, **113**, 25–36.
89. Guyon, I. and Elisseeff, A. (2003) An introduction to variable and feature selection. *J. Mach. Learn. Res.*, **3**, 1157–1182.
90. Lall, S., Grün, D., Krek, A., Chen, K., Wang, Y.-L., Dewey, C.N., Sood, P., Colombo, T., Bray, N., Macmenamin, P. *et al.* (2006) A genome-wide map of conserved microRNA targets in *C. elegans*. *Curr. Biol.*, **16**, 460–471.
91. Harris, T.W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W.J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R. *et al.* (2010) WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res.*, **38**, D463–D467.
92. Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal, R., and Zhang, H. (2009) FlyBase: enhancing *Drosophila* Gene Ontology annotations. *Nucleic Acids Res.*, **37**, D555–D559.
93. Papadopoulos, G.L., Reczko, M., Simossis, V.A., Sethupathy, P., and Hatzigeorgiou, A.G. (2009) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res.*, **37**, D155–D158.
94. Shalgi, R., Lieber, D., Oren, M., and Pilpel, Y. (2007) Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comput. Biol.*, **3**, e131.
95. Khan, A.A., Betel, D., Miller, M.L., Sander, C., Leslie, C.S., and Marks, D.S. (2009) Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat. Biotechnol.*, **27**, 549–555.
96. Ebert, M.S., Neilson, J.R., and Sharp, P.A. (2007) MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat. Methods*, **4**, 721–726.
97. Gentner, B., Schira, G., Giustacchini, A., Amendola, M., Brown, B.D., Ponzoni, M., and Naldini, L. (2009) Stable knockdown of microRNA *in vivo* by lentiviral vectors. *Nat. Methods*, **6**, 63–66.

98. Seitz, H. (2009) Redefining microRNA targets. *Curr. Biol.*, **19**, 870–873.
99. Arvey, A., Larsson, E., Sander, C., Leslie, C.S., and Marks, D.S. (2010) Target mRNA abundance dilutes microRNA and siRNA activity. *Mol. Syst. Biol.*, **6**, 363.
100. Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A. *et al.* (2009) Mammalia miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Mol. Cells*, **35**, 868–880.

38

Barcoded cDNA Libraries for miRNA Profiling by Next-Generation Sequencing

Markus Hafner, Neil Renwick, John Pena, Aleksandra Mihailovic, and Thomas Tuschl

38.1

Introduction

MicroRNAs (miRNAs) are 20–23 nt long RNAs that guide sequence-specific posttranscriptional gene regulation in animals and plants [1–4]. Some viruses also express miRNAs [5]. These regulatory RNAs are expressed in tissue- and cell-type-specific patterns. Dysregulation or mutation of miRNA genes causes or contributes to a wide range of human diseases [5–13]. Therefore, methods are needed to record the abundance of miRNAs in biological or clinical samples with high sequence specificity.

Currently, standard methods for miRNA profiling from diseased tissues rely on microarray analysis [14–18] or quantitative reverse transcription polymerase chain reaction (qRT-PCR) [19]. These approaches are limited to the study of previously identified miRNAs deposited at www.mirbase.org [20]. Microarray assays are also hampered by cross-hybridization, preventing the identification of individual members of miRNA sequence families or mutant variants of miRNAs, whereas qRT-PCR methods are limited to a preselected subset of miRNAs. Here, we describe a method to derive miRNA expression profiles from sequencing of small RNA cDNA libraries. Sequence analysis also allows for the discovery of new or mutated miRNAs and novel families of small RNAs [21–31]. Using deep sequencing, essentially every miRNA can be detected, albeit most of them in low frequency, by sequencing a limited set of tissues [28, 32]. Typically, a small subset of miRNAs (30–50 in mammalian tissues) represent 90% of all sequence reads annotated as miRNAs [33–35]. Given that miRNAs regulate hundreds of transcripts, it is those abundantly expressed miRNAs that have been shown to control gene expression in a quantifiable manner [36–38].

The 150 million sequence reads currently obtained for the smallest scale Illumina HiSeq sequencing run provide more sequencing depth than required to monitor the relative abundance of miRNAs that are critical for gene regulation. At the same time, small RNA cDNA library preparation and sequencing is time-consuming and more expensive on a per sample basis than other standard assays. These

disadvantages are eliminated by the introduction of a barcoding approach at the first step of small RNA cDNA library preparation followed by pooling of multiple samples to increase sample throughput and reduce sequencing costs per sample [35, 39–43].

38.2

Overview of the Method

The experimental process (Figure 38.1) consists of small RNA isolation, ligation of barcoded 3'-adapters to up to 20 individual samples, pooling of samples, ligation of a 5'-adapter, cDNA library preparation, and sequencing. Bioinformatic analysis comprises separation of sequence reads according to the barcode sequence,

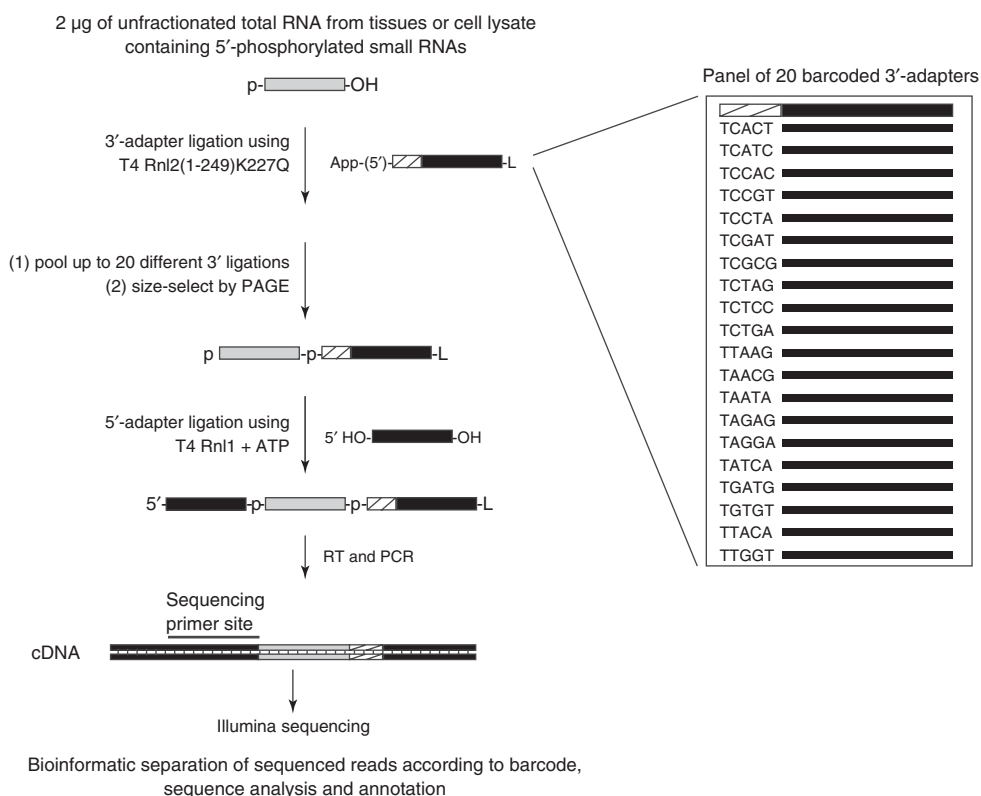


Figure 38.1 Schematic representation of the barcoded small RNA cDNA library preparation. The barcode represents a unique sequence of 5 nt at the 5' end of the 3'-adapter oligodeoxynucleotides. Illumina sequencing at its smallest scale yields

approximately 150 million sequence reads of which more than 70% contain recognizable barcode sequences [35], resulting in more than 5 million sequence reads per sample. L, 3' aminoethyl blocking group that prevents adapter circularization.

followed by mapping and annotation of the extracted sequence reads as described previously [44].

Most classes of small regulatory RNAs are characterized by the presence of a 5' monophosphate and a 3' hydroxyl group, whereas RNA turnover and hydrolysis products carry a 5' hydroxyl group and a 2',3' cyclic phosphate or 2' or 3' monophosphate. The protocol we developed is designed to capture only small RNAs with 5' monophosphate and 3' hydroxyl termini. Owing to the use of RNA ligases for joining adapters, precautions have to be taken to prevent circularization of small RNAs as well as adapters carrying 5' phosphate/3' hydroxyl termini during those steps [45]. (i) We use 3'-adapter deoxyoligonucleotides blocked at their 3' hydroxyl terminus to prevent adapter circularization. The 5'-adapter has 5' and 3' hydroxyl termini and the absence of a 5' monophosphate prevents its circularization. (ii) To minimize adenylation and subsequent circularization of the input RNA during 3'-adapter ligation, we use preadenylated 3'-adapters together with a truncated and mutated form of T4 RNA ligase 2, Rnl2(1–249)K227Q, in the absence of ATP.

Some classes of small RNAs other than miRNAs may carry 3'-terminal 2'-O-methyl modifications, for example, piRNAs. This 3' end modification does not prevent 3'-adapter ligation and cDNA library preparation [46]. Samples containing 2'-O-methylated RNA may be pretreated with periodate and base to oxidize and eliminate any unmodified ends with 2',3' hydroxyls, yielding 3' monophosphate termini that cannot be joined to the 3'-adapter [47, 48]. These libraries are then specifically depleted for miRNAs.

Some reports indicate that the overall abundance of miRNAs might be altered under certain conditions [49–51]. To capture global variations in miRNA content, we propose the use of internal standards, such as the addition of calibrator oligonucleotides. The ratio of calibrator sequence reads to total miRNA sequence reads provides a measure for absolute miRNA abundance in the cellular RNA input.

Materials

All reagents should be free of RNases. RNA samples should be stored at -20°C or below and kept on ice while the reactions are being set up to minimize hydrolysis. Importantly, use siliconized tubes for all manipulations of small RNAs subsequent to pooling of the 3'-adapter ligation reactions. Oligonucleotides in the nanomolar concentration range tend to adsorb to surfaces of non-siliconized tubes and tips.

Oligonucleotide Length Markers

We add trace amounts of ^{32}P -end-labeled markers of 19 and 24 nt length to the samples of total RNA to visualize adapter ligation and to guide the recovery of the expected length fraction of RNA ligation products.

- 19 nt size marker: 5'-CGUACGCGGGUUUAAACGA
- 24 nt size marker: 5'-CGUACGCGGAUAGUUUAAACUGU

The RNA size markers contain a *Pme* I restriction endonuclease recognition site (indicated in italics). Before sequencing, the PCR-amplified cDNA libraries are digested with *Pme* I to exclude the length markers from being sequenced.

Oligoribonucleotide Calibrator Sequences

The calibrator oligoribonucleotides have no match to the mammalian genome [14]. We recommend the addition of 0.5 fmol each of the 10 following calibrator oligoribonucleotides to 2 µg of total RNA. The preparation of a calibrator cocktail requires the use of a carrier oligonucleotide to prevent surface adsorption in the nanomolar concentration range. We recommend the use of an 11 nt oligodeoxynucleotide such as 5'-TCGAAGTATTC.

- Cal01: r(pGUCCACUCCGUAGAUCUGUUC)
- Cal02: r(pGAUGUAACGAGUUGGAAUGCAA)
- Cal03: r(pUAGCAUAUCGAGCCUGAGAACA)
- Cal04: r(pCAUCGGUCGAACUUAUGUGAAA)
- Cal05: r(pGAAGCACAUCGCACAUCAUUAU)
- Cal06: r(pUCUUAACCCGGACCAGAAACUA)
- Cal07: r(pAGGUUCCGGAUAAGUAAGAGCC)
- Cal08: r(pUACUCCUUAAGCGAAUCUCGC)
- Cal09: r(pAAAGUAGCAUCCGAAAUACGGA)
- Cal10: r(pUGAUACGGAUGUUAUACGCAGC).

where r indicates ribonucleotide residues and p a 5' monophosphate.

Preadenylated 3'-Adapters for Illumina Sequencing

We use a set of preadenylated 3'-adapter oligodeoxynucleotides bearing a pentamer barcode at the 5' end.

- 3'Ad01: rAppTCACTTCGTATGCCGTCTTCTGCTTG-L
- 3'Ad02: rAppTCATCTCGTATGCCGTCTTCTGCTTG-L
- 3'Ad03: rAppTCCACTCGTATGCCGTCTTCTGCTTG-L
- 3'Ad04: rAppTCCGTTCTGTATGCCGTCTTCTGCTTG-L
- 3'Ad05: rAppTCCTATCGTATGCCGTCTTCTGCTTG-L
- 3'Ad06: rAppTCGATTCGTATGCCGTCTTCTGCTTG-L
- 3'Ad07: rAppTCGCGTCGTATGCCGTCTTCTGCTTG-L
- 3'Ad08: rAppTCTAGTCGTATGCCGTCTTCTGCTTG-L
- 3'Ad09: rAppTCTCCTCGTATGCCGTCTTCTGCTTG-L
- 3'Ad10: rAppTCTGATCGTATGCCGTCTTCTGCTTG-L
- 3'Ad11: rAppTTAAGTCGTATGCCGTCTTCTGCTTG-L
- 3'Ad12: rAppTAACGTCGTATGCCGTCTTCTGCTTG-L
- 3'Ad13: rAppTAATATCGTATGCCGTCTTCTGCTTG-L
- 3'Ad14: rAppTAGAGTCGTATGCCGTCTTCTGCTTG-L
- 3'Ad15: rAppTAGGATCGTATGCCGTCTTCTGCTTG-L
- 3'Ad16: rAppTATCATCGTATGCCGTCTTCTGCTTG-L
- 3'Ad17: rAppTGATGTCGTATGCCGTCTTCTGCTTG-L

- 3'Ad18: rApp**IGTGTT**CGTATGCCGTCTTCTGCTTG-L
- 3'Ad19: rApp**TTACAT**CGTATGCCGTCTTCTGCTTG-L
- 3'Ad20: rApp**TTGGTT**CGTATGCCGTCTTCTGCTTG-L

where L is the 3' aminoethyl blocking group; rApp is the 5'-terminal adenosine residue connected via a 5',5'-diphosphate bridge, which activates the adapter for ligation (for more details, see Chapter 3). Residues marked in bold letters constitute the pentameric barcode, followed by the common RT primer-binding site. The preparation and purification of the preadenylated 3'-adapters has been described previously [29].

Oligoribonucleotide 5'-Adapter Compatible with Illumina Sequencing

- 5'-adapter: 5'-r(GUUCAGAGUUCUACAGUCCGACGAUC) (r indicates RNA residues)

Primers for Amplification of the Barcoded cDNA Library

- Forward primer: 5'-AATGATACGGCGACCAACGACAGGTTCTAGAGTTCTACAGTCCGA
- Reverse primer: 5'-CAAGCAGAAGACGGCATACGA

Enzymes

- T4 polynucleotide kinase (T4 PNK; New England Biolabs (NEB))
- T4 RNA ligase 2 truncated K227Q = T4 Rnl2(1–249)K227Q (NEB, M0351L). The plasmid for expression of recombinant, His-tagged T4 Rnl2(1–249)K227Q can also be obtained from www.addgene.org
- T4 Rnl1 (Thermo Scientific Fermentas or NEB).
- SuperScript III reverse transcriptase (Life Technologies Invitrogen)
- *Taq* DNA polymerase (any supplier)
- *Pme* I (NEB)

Buffers and Solutions

- 10 x Ligation buffer without ATP: 0.5 M Tris-HCl, pH 7.6, 0.1 M MgCl₂, 0.1 M 2-mercaptoethanol, 1 mg ml⁻¹ acetylated BSA (Sigma-Aldrich, B-8894)
- Denaturing acrylamide gel loading solution: 98.8% formamide, 1% (v/v) 0.5 M Na₂EDTA, pH 8.0, 0.2% bromophenol blue
- 5 x TBE buffer: 445 mM Tris base, 445 mM boric acid, 10 mM Na₂EDTA-NaOH, pH 8.0
- 10 x Ligation buffer with ATP: 0.5 M Tris-HCl, pH 7.6, 0.1 M MgCl₂, 0.1 M 2-mercaptoethanol, 1 mg ml⁻¹ acetylated BSA (Sigma-Aldrich, B-8894), 2 mM ATP
- 5 x First-strand buffer (supplied with SuperScript III reverse transcriptase): 250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM MgCl₂, 100 mM DTT
- 10 x dNTP mix: 2 mM dATP, 2 mM dCTP, 2 mM dGTP, 2 mM dTTP

- 10 x PCR buffer: 100 mM TrisHCl, pH 8.0, 500 mM KCl, 1% Triton-X100, 20 mM MgCl₂, 10 mM 2-mercaptoethanol
- 5 x Agarose gel loading solution: 0.2% (w/v) bromophenol blue, 0.2% (w/v) xylene cyanol FF, 50 mM Na₂EDTA, pH 8.0, 20% Ficoll type 400

Other Materials

- Siliconized 1.5 ml reaction tubes (BioPlas, 4165SL)
- QiaQuick gel extraction kit (Qiagen, 28704)
- TRIzol[®] reagent (Life Technologies Invitrogen)
- 25 bp DNA Ladder (Life Technologies Invitrogen).

Procedure

The following step-by-step procedure describes the generation of a small RNA cDNA library for Illumina sequencing. We use 2 µg of total RNA as input per biological or clinical sample and pool up to 20 samples directly after the 3'-adaptor ligation step that introduced the pentameric barcode sequence.

Isolation of Total RNA

Total RNA is isolated from either freshly collected cultured cells or tissues, flash-frozen samples stored below -70 °C, or formalin-fixed tissues [52]. As a rule of thumb, 1 mg of tissue or 10⁵ cells yield about 1 µg of total RNA. We recommend isolation of total RNA from tissue or cultured cells using either the guanidinium isothiocyanate (GITC)/phenol method [53] or the commercial TRIzol reagent. Avoid aqueous LiCl precipitation, as small RNAs, including tRNAs, will not precipitate. Other protocols designed to enrich for small RNAs should also be avoided when calibrators are used to determine the ratio of miRNA to total RNA.

Preparation of Calibrator Oligoribonucleotide Cocktail

- 1) Prepare 7 ml of carrier solution containing 500 nM carrier oligodeoxyribonucleotide in water. The carrier is necessary to prevent surface adsorption of calibrator oligoribonucleotides.
- 2) Prepare 50 µl of a calibrator cocktail containing 1 µM of each calibrator oligoribonucleotide (corresponding to a total calibrator concentration of 10 µM).
- 3) Dilute the calibrator cocktail 1 : 10 in carrier solution (step 1) resulting in a total calibrator concentration of 1 µM.
- 4) Further dilute the calibrator cocktail dilution (from step 3) 1 : 100 in carrier solution resulting in a total calibrator concentration of 10 nM.

Preparation of Radioactive Length Markers

- 5) The length markers are radiolabeled individually in a 10 µl reaction containing 1 µM RNA, 10 U T4 PNK, and 50 µCi [γ -³²P]ATP (6000 Ci/mmol); incubate at 37 °C for 15 min.

- 6) Quench the reaction by addition of 10 μ l of denaturing acrylamide gel loading solution.
- 7) Incubate the sample at 90 °C for 30 s.
- 8) Load samples on a 15% denaturing PAA gel (15 cm wide, 17 cm long, 0.5 mm thick; 30 ml gel solution required). Run the gel approximately 45 min at 30 W using 0.5 x TBE buffer until the bromophenol blue dye is close to the bottom of the gel.
- 9) Dismantle the gel, leaving it mounted on one glass plate. Wrap the gel in plastic film (e.g., Saran wrap) and expose an X-ray film for 1 min to the gel; develop the film.
- 10) Align the gel on glass plate to the X-ray film. Cut out the radioactive bands corresponding to the length markers and transfer the gel slices into 1.5 ml siliconized tubes (one for the 19 nt and one for the 24 nt marker). Also collect some weakly radioactive gel pieces from the gel running front and store in a radioactive storage container. These pieces will be implanted later to facilitate alignment of gels with phosphorimager printouts.
- 11) Add 300 μ l of 0.3 M NaCl and elute the RNA from the gel by incubating the tube overnight at 4 °C under constant agitation.
- 12) Collect the supernatant and add 900 μ l (3 vol.) of absolute ethanol. Keep the sample on ice for 1 h or overnight at -20 °C.
- 13) Centrifuge in a tabletop centrifuge at 4 °C and maximum speed (approximately 14 000 g) for 15 min.
- 14) Discard the supernatant. Collect residual ethanol by centrifugation at 14 000 g for 1 min. Air-dry the RNA pellets for 2 min.
- 15) Dissolve each pellet in 10 μ l of ddH₂O.
- 16) Combine 19 and 24 nt length marker solutions.
- 17) Dilute the combined length markers 1 : 50 in Carrier solution (step 1).

3'-Adapter Ligation

- 18) Provide 2 μ g of total RNA in 8 μ l of ddH₂O in a siliconized reaction tube.
- 19) Prepare a Mastermix for ligating the 3'-adapter, multiplying the volume for one reaction (9 μ l) by the number of samples being processed (up to 20 samples plus one control for the combined 19 and 24 nt length markers); increase the volume of the Mastermix by another 10% to compensate for dispensing errors. Each 3'-adapter ligation reaction requires:
 - a. 2 μ l of 10 x RNA ligase buffer without ATP
 - b. 6 μ l of 50% aqueous DMSO
 - c. 0.5 μ l of a 10 nM calibrator cocktail solution
 - d. 0.5 μ l of 5'-³²P-labeled length marker oligoribonucleotide mix.
- 20) Add 9 μ l of the Mastermix to each RNA sample prepared in step 18 (for the length marker control reaction add Mastermix to 8 μ l ddH₂O)
- 21) Add 2 μ l of 50 μ M adenylated 3'-adapter with a unique barcode to each sample.

- 22) Incubate the tube for 30 s at 90 °C to denature the RNA and immediately place on ice for 20 s.
- 23) Add 1 µl T4 Rnl2(1–249)K227Q (1 µg µl⁻¹), mix gently, and incubate overnight on ice.
- 24) Add 2 µl of 3 M NaCl to each sample to adjust to approximately 0.3 M NaCl.
- 25) Add three times the total volume of the combined 3'-adapter ligation reactions of absolute ethanol (i.e., 1.3 ml ethanol for 20 samples) to a new 2 ml reaction tube. Then transfer the individual adapter ligation reactions into the same ethanol-containing tube, which stops the ligation reaction and combines the barcoded libraries.
- 26) Precipitate the ligation products by incubation on ice for 1 h and collect the pellet by centrifugation in a tabletop centrifuge at 4 °C and maximum speed (approximately 14 000 g) for 15 min.
- 27) Discard the supernatant. Collect residual ethanol by centrifugation at 14 000 g for 1 min. Air-dry the RNA pellet for 10 min.
- 28) Dissolve the RNA pellet in 20 µl ddH₂O.
- 29) Add 20 µl of denaturing acrylamide gel loading solution and load the samples in two adjacent wells in the center of a 20-well 15% denaturing PAA gel (15 cm wide, 17 cm long, 0.5 mm thick; 30 ml gel solution). When processing more than one barcoded library (i.e., more than 20 individual small RNA samples at once), leave a space of two wells between samples to avoid cross-contamination. Load the control length marker ligation reaction separated by one blank lane flanking the samples in the center of the gel.
- 30) Run the gel at 30 W for approximately 45 min in 0.5 x TBE buffer until the bromophenol blue dye reaches the lower third of the gel. Do not run the gel much further in order to contain the ligation products within a gel area as small as possible for efficient elution.
- 31) Dismantle the gel, leaving it mounted on one glass plate.
- 32) To facilitate the alignment of the gel to the phosphorimager paper printout, implant three tiny radioactive gel pieces (step 10) asymmetrically at three of the four corners of the gel. Radioactive gel pieces were collected from the gel that was used to purify the size markers after ³²P-labeling.
- 33) Wrap the gel in plastic film (e.g., Saran wrap) to avoid contamination and expose a phosphorimaging screen for 45 min.
- 34) Print out a 100%-scaled image of the phosphorimaged gel and align the gel on top of the printout according to the positions of the three radioactive gel pieces. For gel excision in sample lanes, use the positions of the 19 nt/3' adapter and 24 nt/3' adapter ligation products as margins. Transfer the gel piece(s) into a siliconized 1.5 ml tube. Also excise the ligation products for the length marker control.
- 35) Add 350 µl of 0.3 M NaCl to each tube and elute the ligated RNAs from the gel slices by incubating overnight at 4 °C under constant agitation.
- 36) Collect the supernatant and add 3 vol. absolute ethanol to precipitate the RNA as described in steps 26 and 27.

5'-Adapter Ligation

- 37) Dissolve the RNA pellets in 9 μ l ddH₂O. This also pertains to the control length marker ligation product.
- 38) Prepare a Mastermix according to the rules described in step 19, containing the following components per reaction:
 - a. 1 μ l of 100 μ M 5'-adapter
 - b. 2 μ l of 10 x RNA ligase buffer with ATP
 - c. 6 μ l of 50% aqueous DMSO
- 39) Add 9 μ l of the Mastermix to each RNA sample from step 37. Incubate the tubes for 30 s at 90 °C to denature the RNA and immediately place on ice for 20 s.
- 40) Add 2 μ l of T4 RNA ligase 1 (Rnl1), mix gently, and incubate for 1 h at 37 °C.
- 41) Add 20 μ l of denaturing acrylamide gel loading solution and load the samples in two adjacent wells of a 20-well 15% denaturing PAA gel (15 cm wide, 17 cm long, 0.5 mm thick; 30 ml gel solution). When processing more than one barcoded library, leave a space of two wells between samples to avoid cross-contamination. Load the control length marker ligation reaction separated by one blank lane flanking the samples in the center of the gel. Run the gel at 30 W using 0.5 x TBE buffer until the bromophenol blue dye is close to the bottom of the gel. Image the gel as described in steps 31–34 and excise the new ligation products.
- 42) Add 350 μ l of 0.3 M NaCl and 1 μ l of 100 μ M reverse PCR primer as carrier and elute the ligated RNAs from the gel by incubating the tube overnight at 4 °C under constant agitation. The carrier facilitates the recovery of the ligation product.
- 43) Add 3 vol. absolute ethanol to precipitate and collect the RNA as described in steps 26 and 27.

Reverse Transcription (RT)

- 44) Dissolve pellets in 5.6 μ l ddH₂O. This also pertains to the control length marker ligation product.
- 45) Prepare an RT Mastermix (according to the rules described in step 19) containing the following components per reaction:
 - a. 1.5 μ l of 0.1 M DTT
 - b. 3 μ l of 5 x first-strand buffer
 - c. 4.2 μ l of 10 x dNTPs.
- 46) Denature the RNA from step 44 by incubating the tube for 30 s at 90 °C, followed by transfer of the tube to a 50 °C incubator.
- 47) Add 8.7 μ l of the RT Mastermix to each sample and incubate for 3 min at 50 °C. Add 0.75 μ l of Superscript III reverse transcriptase and incubate for 30 min at 42 °C.
- 48) To hydrolyze the RNA template, add 40 μ l of 150 mM KOH/20 mM Tris base and incubate for 10 min at 90 °C.

- 49) Neutralize the solution by addition of 40 μl 150 mM HCl and adjust the pH to a value between 7.5 and 9.0. Monitor the pH change by spotting 1 μl of cDNA solution on a pH paper. The pH of the solution should be slightly alkaline to not inhibit the subsequent PCR.

PCR Amplification

- 50) Prepare a PCR mix containing the following components:

- a. 10 μl of the cDNA solution
- b. 0.5 μl of 100 μM forward primer
- c. 0.5 μl of 100 μM reverse primer
- d. 10 μl of 10 x dNTP mix
- e. 10 μl of 10 x PCR buffer
- f. 68 μl of ddH₂O
- g. 1 μl *Taq* DNA polymerase (5 U μl^{-1}).

Also prepare a non-template control PCR (H₂O instead of cDNA) to check for DNA contamination.

- 51) Program the following cycle conditions: 45 s at 94 °C, 85 s at 50 °C, and 60 s at 72 °C. Remove 12 μl aliquots every other cycle following cycle number 10 by temporarily putting the PCR cyclor on hold at the end of the 72 °C step. This is done to determine the necessary number of cycles for amplifying the cDNA library. In our experience, it is usually not necessary to amplify for more than 15 cycles.
- 52) Analyze the PCR products on a 2.5% agarose gel. These products might appear as a double band with a higher band running at the expected length of about 90–95 nt and a lower band corresponding to 3'-adapter-to-5'-adapter ligation side products running at about 70 nt. The ligation of adapters occurs because of incomplete separation of the excess of 3'-adapter from ligation products before 5'-adapter ligation.
- 53) Define the optimal cycle number for cDNA amplification (usually not more than 15 cycles), which has to be within the exponential amplification phase of the PCR, that is, approximately five cycles away from reaching the saturation level of PCR amplification. It is important to limit the PCR to the exponential phase, otherwise extreme sequence-specific distortions will be introduced into the miRNA profiles.
- 54) Perform a 300 μl PCR according to steps 50/51 with the previously determined optimal cycle number by distributing the volume over three PCR tubes. After the reaction, remove 5 μl and combine with 5 μl agarose loading dye solution to verify product formation on a 2.5% agarose gel. If the PCR product is visible, proceed to the next step, otherwise add a few cycles of PCR or repeat the pilot experiment (steps 50/51) and the large-scale PCR (this step).
- 55) Transfer the PCR solution to a 1.5 ml tube, add 30 μl of 3 M NaCl and 330 μl of phenol/chloroform, and vortex for 20 s.

- 56) Separate the phases by centrifugation at 14 000 g in a tabletop centrifuge for 2 min.
- 57) Take off the upper, aqueous phase and transfer to a new tube. Make sure not to withdraw the interphase where denatured proteins accumulate.
- 58) Extract with 330 μ l of chloroform to remove residual phenol from the aqueous phase, vortex for 20 s.
- 59) Separate phases by centrifugation at 14 000 g in a tabletop centrifuge for 2 min.
- 60) Take off the upper, aqueous phase and transfer to a new tube.
- 61) Add 1 ml of absolute ethanol and incubate on ice for 1 h or overnight at -20°C .
- 62) Collect the pellet by centrifugation in a tabletop centrifuge at 4°C at maximum speed (approximately 14 000 g) for 15 min.
- 63) Discard the supernatant. Collect residual ethanol by centrifugation at 14 000 g for 1 min. Remove all of the supernatant but do not dry the pellet as this will cause the DNA to denature. Immediately proceed to the next step.

***Pme* I Digestion**

This step cleaves PCR products originating from the radiolabeled length markers used during small RNA cDNA library preparation. Be careful not to denature the double-stranded PCR product before or during the *Pme* I digestion. Denaturation and subsequent reannealing of a complex sequence pool will result in imperfect rehybridization and formation of DNA duplexes with internal bulges that might compromise *Pme* I digestion. As control, the PCR product obtained from the ligation of adapters to the marker oligonucleotides alone (marker control sample) must be digested completely.

- 64) Prepare a *Pme* I digestion cocktail containing the following components per reaction:
 - a. 2 μ l of 10 x *Pme* I buffer (NEB)
 - b. 0.2 μ l of 100 x BSA (10 mg ml $^{-1}$, NEB)
 - c. 17.3 μ l of ddH $_2$ O
 - d. 0.5 μ l (5 U) of *Pme* I (NEB)
- 65) Dissolve the DNA pellet in the 20 μ l *Pme* I digestion mixture and incubate for at least 2 h at 37°C . Do not vortex vigorously to avoid enzyme denaturation.
- 66) Next, to separate miRNA-containing adapter ligation products from shorter 5'-adapter-to-3'-adapter ligation side products, prepare a 2.5% agarose gel containing 0.4 μ g ml $^{-1}$ ethidium bromide. Load the entire restriction digest into two wells of the agarose gel, as well as the 25 bp DNA ladder in a separate well. Run the gel in 0.5 x TBE buffer for approximately 1.5 h at 180 V until the 25 bp ladder is sufficiently resolved.
- 67) Visualize the DNA in the gel using a 360 nm UV transilluminator and excise the upper band of approximately 90–95 bp in size.

- 68) Transfer the gel slice to a preweighed 1.5 ml reaction tube and reweigh it. Elute the DNA from the gel using the QiaQuick gel extraction kit or a comparable kit according to the manufacturer's instructions. The DNA will be recovered in 30 µl of 1 x TE buffer (Qiagen).
- 69) The DNA is then ready for Illumina sequencing.

38.3

Troubleshooting

- **No PCR product:** (i) Ligation did not work. Take care to cool the reaction mixture after heat shock and before addition of ligase. Especially, Rnl2(1–249)K227Q is inactive at temperatures above 37 °C and may be irreversibly inactivated at higher temperatures. (ii) After precipitation, take care that residual ethanol is evaporated before addition of reaction mix, as residual ethanol will inhibit ligases and reverse transcriptases. (iii) Revisit reverse transcription.
- **Only adapter–adapter products visible after PCR:** (i) Make sure adapter concentrations in the ligation reactions are correct. (ii) Make sure to only excise ligation products migrating between those obtained for the ³²P-labeled 19 and 24 nt length markers. Do not cut below the ligation product obtained for the 19 nt length marker.

Acknowledgments

M.H. was supported by the Deutscher Akademischer Austauschdienst and is currently funded by a fellowship of the Charles H. Revson Foundation. N.R. is supported through a K08 award (NS072235) from the National Institute of Neurological Disorders and Stroke. T.T. is an HHMI investigator, and work in his laboratory was supported by NIH grant MH08442 and the Starr Foundation.

References

1. Bhattacharyya, S.N. and Filipowicz, W. (2007) Argonautes and company: sailing against the wind. *Cell*, **128**, 1027–1028.
2. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
3. Stefani, G. and Slack, F.J. (2008) Small non-coding RNAs in animal development. *Nat. Rev. Mol. Cell. Biol.*, **9**, 219–230.
4. Voinnet, O. (2009) Origin, biogenesis, and activity of plant microRNAs. *Cell*, **136**, 669–687.
5. Skalsky, R.L. and Cullen, B.R. (2010) Viruses, microRNAs, and host interactions. *Annu. Rev. Microbiol.*, **64**, 123–141.
6. Hébert, S.S. and De Strooper, B. (2007) Molecular biology. miRNAs in neurodegeneration. *Science*, **317**, 1179–1180.
7. Ryan, B.M., Robles, A.I., and Harris, C.C. (2010) Genetic variation in microRNA networks: the implications for cancer research. *Nat. Rev. Cancer*, **10**, 389–402.

8. Croce, C.M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.*, **10**, 704–714.
9. Calin, G.A. and Croce, C.M. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, **6**, 857–866.
10. Latronico, M.V. and Condorelli, G. (2009) MicroRNAs and cardiac pathology. *Nat. Rev. Cardiol.*, **6**, 419–429.
11. Inui, M., Martello, G., and Piccolo, S. (2010) MicroRNA control of signal transduction. *Nat. Rev. Mol. Cell. Biol.*, **11**, 252–263.
12. Zhang, C. (2009) Novel functions for small RNA molecules. *Curr. Opin. Mol. Ther.*, **11**, 641–651.
13. Poy, M.N., Spranger, M., and Stoffel, M. (2007) microRNAs and the regulation of glucose and lipid metabolism. *Diabetes Obes. Metab.*, **9**, 67–73.
14. Bissels, U., Wild, S., Tomiuk, S., Holste, A., Hafner, M., Tuschl, T., and Bosio, A. (2009) Absolute quantification of microRNAs by using a universal reference. *RNA*, **15**, 2375–2384.
15. Bissels, U., Wild, S., Tomiuk, S., Hafner, M., Scheel, H., Mihailovic, A., Choi, Y., Tuschl, T., and Bosio, A. (2011) Combined characterization of microRNA and mRNA profiles delineates early differentiation pathways of CD133(+) and CD34(+) hematopoietic stem and progenitor cells. *Stem Cells*, **29**, 847–857.
16. Barad, O., Meiri, E., Avniel, A., Aharonov, R., Barzilai, A., Bentwich, I., Einav, U., Gilad, S., Hurban, P., Karov, Y. *et al.* (2004) MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res.*, **14**, 2486–2494.
17. Nelson, P.T., Baldwin, D.A., Scearce, L.M., Oberholtzer, J.C., Tobias, J.W., and Mourelatos, Z. (2004) Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat. Meth.*, **1**, 155–161.
18. Baskerville, S. and Bartel, D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
19. Fiedler, S.D., Carletti, M.Z., and Christenson, L.K. (2010) Quantitative RT-PCR methods for mature microRNA expression analysis. *Methods Mol. Biol.*, **630**, 49–64.
20. Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36** (Suppl. 1), D154–D158.
21. Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D.S., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003) The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell*, **5**, 337–350.
22. Chen, P.Y., Manninga, H., Slanchev, K., Chien, M., Russo, J.J., Ju, J., Sheridan, R., John, B., Marks, D.S., Gaidatzis, D. *et al.* (2005) The developmental miRNA profiles of zebrafish as determined by small RNA cloning. *Genes Dev.*, **19**, 1288–1293.
23. Aravin, A.A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T. *et al.* (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, **442**, 203–207.
24. Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006) Characterization of the piRNA complex from rat testes. *Science*, **313**, 363–367.
25. Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007) Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev.*, **21**, 1603–1608.
26. Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell*, **12**, 503–514.
27. Ruby, J.G., Jan, C.H., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, **127**, 1193–1207.

28. Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E. *et al.* (2010) Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.*, **24**, 992–1009.
29. Hafner, M., Landgraf, P., Ludwig, J., Rice, A., Ojo, T., Lin, C., Holoch, D., Lim, C., and Tuschl, T. (2008) Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods*, **44**, 3–12.
30. Bar, M., Wyman, S.K., Fritz, B.R., Qi, J., Garg, K.S., Parkin, R.K., Kroh, E.M., Bendoraitis, A., Mitchell, P.S., Nelson, A.M. *et al.* (2008) MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. *Stem Cells*, **26**, 2496–2505.
31. Friedländer, M.R., Adamidi, C., Han, T., Lebedeva, S., Isenbarger, T.A., Hirst, M., Marra, M.A., Nusbaum, C., Lee, W.L., Jenkin, J.C. *et al.* (2009) High-resolution profiling and discovery of planarian small RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 11546–11551.
32. Friedländer, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., and Rajewsky, N. (2008) Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotech.*, **26**, 407–415.
33. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M. Jr., Jungkamp, A.-C., Munschauer, M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
34. Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
35. Farazi, T.A., Horlings, H.M., Ten Hoeve, J.J., Mihailovic, A., Halfwerk, H., Morozov, P., Brown, M., Hafner, M., Reyat, F., van Kouwenhove, M. *et al.* (2011) MicroRNA sequence and expression analysis in breast tumors by deep sequencing. *Cancer Research*, **71**, 4443–4453.
36. Selbach, M., Schwanhaussner, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.
37. Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.
38. Krutzfeldt, J., Kuwajima, S., Braich, R., Rajeev, K.G., Pena, J., Tuschl, T., Manoharan, M., and Stoffel, M. (2007) Specificity, duplex degradation and subcellular localization of antagomirs. *Nucleic Acids Res.*, **35**, 2885–2892.
39. Xu, Q., Schlabach, M.R., Hannon, G.J., and Elledge, S.J. (2009) Design of 240,000 orthogonal 25mer DNA barcode probes. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 2289–2294.
40. Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M., and Fire, A.Z. (2007) A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res.*, **35**, e130.
41. Schulte, J.H., Marschall, T., Martin, M., Rosenstiel, P., Mestdagh, P., Schlierf, S., Thor, T., Vandesompele, J., Eggert, A., Schreiber, S. *et al.* (2010) Deep sequencing reveals differential expression of microRNAs in favorable versus unfavorable neuroblastoma. *Nucleic Acids Res.*, **38**, 5919–5928.
42. Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J., and Knight, R. (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Meth.*, **5**, 235–237.
43. Vigneault, F., Sismour, A.M., and Church, G.M. (2008) Efficient microRNA capture and bar-coding via enzymatic oligonucleotide adenylation. *Nat. Meth.*, **5**, 777–779.
44. Berninger, P., Gaidatzis, D., van Nimwegen, E., and Zavolan, M. (2008) Computational analysis of small RNA cloning data. *Methods*, **44**, 13–21.
45. Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser,

- F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M. *et al.* (2005) Identification of microRNAs of the herpesvirus family. *Nat. Meth.*, **2**, 269–276.
46. Ebhardt, H.A., Thi, E.P., Wang, M., and Unrau, P.J. (2005) Extensive 3' modification of plant small RNAs is modulated by helper component-proteinase expression. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 13398–13403.
 47. Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V.A., and Zamore, P.D. (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, **313**, 320–324.
 48. Alefelder, S., Patel, B.K., and Eckstein, F. (1998) Incorporation of terminal phosphorothioates into oligonucleotides. *Nucleic Acids Res.*, **26**, 4983–4988.
 49. Krol, J., Buskamp, V., Markiewicz, I., Stadler, M.B., Ribi, S., Richter, J., Duebel, J., Bicker, S., Fehling, H.J., Schübeler, D. *et al.* (2010) Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. *Cell*, **141**, 618–631.
 50. Kumar, M.S., Lu, J., Mercer, K.L., Golub, T.R., and Jacks, T. (2007) Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat. Genet.*, **39**, 673–677.
 51. Hwang, H., Wentzel, E.A., and Mendell, J.T. (2009) Cell-cell contact globally activates microRNA biogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 7016–7021.
 52. Masuda, N., Ohnishi, T., Kawamoto, S., Monden, M., and Okubo, K. (1999) Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Research*, **27**, 4436–4443.
 53. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156–159.

39

Transcriptome-Wide Identification of Protein Binding Sites on RNA by PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation)

Jessica I. Hoell*, Markus Hafner*, Markus Landthaler, Manuel Ascano, Thalia A. Farazi, Greg Wardle, Jeff Nusbaum, Pavol Cekan, Mohsen Khorshid, Lukas Burger, Mihaela Zavolan, and Thomas Tuschl

39.1

Introduction

The human genome encodes several hundred RNA binding proteins (RBPs) expressed in a tissue- and cell-type-specific manner [1, 2], each of which can specifically bind thousands of mRNAs (or other RNAs) at defined RNA recognition elements (RREs) [3]. For the dissection and understanding of these complex networks, the development of powerful tools to determine the bound transcripts and the RREs is critical in order to enhance our understanding of posttranscriptional mRNA regulation.

A combination of genetic, biochemical, and computational approaches has been available to identify mRNA–RBP or mRNA–ribonucleoprotein (RNP) interactions. Microarray profiling of mRNA associated with immunopurified RBPs (RNA binding protein immunoprecipitation - chip (RIP-chip)) [4] has been shown to enrich targeted mRNAs and RREs. For example, binding sites for the Pumilio RBP or miRNPs were identified within mature mRNAs that are hundreds to thousands of nucleotides long [5, 6].

An alternative approach involves *in vivo* UV 254 nm crosslinking and immunoprecipitation (CLIP) [7] combined with high-throughput sequencing of the bound RNA segments (HITS-CLIP) [8, 9]. The covalent bond between RBP and target RNA facilitates isolation of crosslinked RNAs under denaturing conditions, yet the crosslinked positions are not revealed after sequencing of cDNAs corresponding to the UV-crosslinked RNA fragment. Moreover, to separate crosslinked sites from background uncrosslinked noise, additional control experiments, such as crosslinking in cells deficient for the studied RBP, have to be performed in order to define the background.

We recently developed a new method for the identification of binding sites of RBPs, referred to as photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) (Figure 39.1a) [10]. It relies on photoactivatable

* These authors contributed equally to this chapter.

nucleoside analogs, such as 4-Thiouridine (4SU) and 6-Thioguanosine (6SG) that are taken up by cultured cells from the supplemented media [11, 12] to be incorporated into nascent RNAs. At concentrations up to 1 mM, neither of the tested photoreactive nucleosides shows any detectable toxic effects. Irradiation of the cells by UV light of 365 nm leads to crosslinking of photoreactive nucleoside-labeled cellular RNAs to interacting RBPs. Using similar irradiance levels, 4SU incorporation enhances RNA crosslinking efficiencies between 100- and 1000-fold while 6SG incorporation increases crosslinking efficiencies by ~10-fold compared to UV 254 nm crosslinking. Furthermore, the precise crosslinking position can be identified by mapping characteristic T to C mutations (G to A when using 6SG, although to a lesser extent) in the sequenced cDNA libraries that are generated from the recovered RNA. We presume that crosslinking of the photoreactive nucleoside to aromatic amino acid side chains induces a structural change that directs the incorporation of a non-cognate deoxynucleoside during reverse transcription of crosslinked RNAs. The presence of the mutations in the sequenced cDNA reads, together with the observation that multiple positions within a cluster of sequence reads can be altered, facilitates the separation from clusters of unaltered background sequences typically derived from abundant cellular RNAs. For details on the bioinformatic analyses, refer to our recent publication [10].

PAR-CLIP is readily applicable to any protein or RNP directly contacting RNA (including RBPs interacting transiently with their RNA targets, such as helicases, polymerases, or nucleases), which may be expressed naturally or ectopically in cell culture models.

It offers new opportunities for understanding both gene regulation and consequences of genetic variation in transcript regions aside from the coding sequence.

Materials

Special Equipment

- UV Stratalinker 2400 equipped with 365 nm light bulbs for crosslinking (Agilent Technologies Stratagene, or equivalent)

Oligonucleotides

- Preadenylated 3'-adapter (DNA except for the 5' riboadenylate (rApp) residue): 5' rAppTCGTATGCCGTCTTCTGCTTGT (for more details on preadenylated ligation substrates, see Chapter 3)
- 5'-Adapter (RNA): 5' GUUCAGAGUUCUACAGUCCGACGAUC
- Polymerase chain reaction (PCR) reverse primer (DNA): 5' CAAGCAGAAGACG-GCATACGA
- PCR forward primer (DNA): 5' AATGATACGGCGACCACCGACAGGTTTCAGAGTTCTACAGTCCGA
- 19 nt size marker (RNA): 5' CGUACGCGGGUUAACGA
- 24 nt size marker (RNA): 5' CGUACGCGGAUAGUUUAACUGU

- 33 nt size marker (RNA): 5' CAUCUUGGUCGUACGCGGAUAGUUUAA ACUGU
- 35 nt size marker (RNA): 5' CUCAUCUUGGUCGUACGCGGAUAGUUUAA ACUGU

Enzymes

- Calf intestinal alkaline phosphatase (CIP, New England Biolabs (NEB), M0290L)
- Proteinase K (Lyophilizate; Roche, 03115801001)
- RNase T1 (Thermo Scientific Fermentas, EN0541); concentration $1000 \text{ U } \mu\text{l}^{-1}$
- SuperScript III Reverse Transcriptase (Life Technologies Invitrogen, 18080-044); includes 5 x first-strand buffer
- T4 PNK (T4 polynucleotide kinase, NEB, M0201)
- T4 RNA Ligase 1 (NEB, M0204L)
- T4 RNA Ligase 2 truncated K227Q = T4 Rnl2(1-249)K227Q (NEB, M0351L)
- *Taq* DNA polymerase ($5 \text{ U } \mu\text{l}^{-1}$) (NEB, M0267L)

Consumables, Kits, and Chemicals

- Agarose, electrophoresis grade (SeaKem LE Agarose, Lonza, 50004)
- Agarose, low melting (NuSieve GTG Agarose, Lonza, 50080)
- Antibody (e.g., for FLAG-tagged RBPs: mouse monoclonal anti-FLAG M2 (Sigma–Aldrich, F1804))
- Ammonium persulfate (APS)
- Adenosine-5'-triphosphate (ATP)
- 2-Mercaptoethanol (14.3 M; Sigma–Aldrich, M6250)
- Bromophenol blue
- Bovine serum albumin (BSA), acetylated (Ambion, AM2614)
- $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$
- Cell scraper (Corning Life Sciences)
- Chloroform
- Citric acid monohydrate
- Complete EDTA-free protease inhibitor cocktail tablets (Roche, 11836170001)
- Dimethyl sulfoxide (DMSO)
- DNA ladder (25 bp) (Life Technologies Invitrogen, 10597-011)
- dNTPs: dATP, dCTP, dGTP, dTTP (0.1 M each; Thermo Scientific Fermentas, R0182)
- Dithiothreitol (DTT)
- D-Tube Dialyzer Midi, molecular weight cutoff (MWCO) 3.5 kDa (EMD Biosciences, 71506-3)
- Dynabeads Protein G (Life Technologies Invitrogen, 100-03D)
- EDTA disodium salt dihydrate (Sigma–Aldrich, E5134)
- EGTA, $\text{C}_{14}\text{H}_{20}\text{N}_2\text{O}_{10}\text{Na}_4$ (Sigma–Aldrich, E8145-10G)
- Ethanol (100%)
- Ethidium bromide

- Ficoll type 400
- Formamide
- [γ - ^{32} P]ATP, 10 mCi ml $^{-1}$, 6000 Ci (222 TBq) mmol $^{-1}$ (Perkin Elmer, NEG-002Z500UC)
- Glycerol
- Glycoblue or glycogen
- HCl, concentrated (Fisher Scientific, A144S)
- HEPES
- Isoamyl alcohol (IAA)
- Isopropyl alcohol
- KCl
- KOH
- MgCl $_2$ ·6H $_2$ O
- MOPS SDS running buffer (20 x; Life Technologies Invitrogen, NP0001)
- Na $_2$ HPO $_4$ ·7H $_2$ O (Sigma–Aldrich, S9390-100G)
- NaCl
- NaF
- NaOH
- NP40 substitute (100%; Sigma–Aldrich, 74385)
- NuPAGE Novex 4–12% BT Midi 1.0 gel (Life Technologies Invitrogen, WG1401BOX)
- Phosphate-buffered saline (PBS) (10 x, commercially available)
- Acidic phenol (saturated with 0.1 M citrate buffer, pH 4.3 $\pm$ 0.2, Sigma–Aldrich, P4682)
- Photoreactive nucleoside (Sigma–Aldrich; 4-Thiouridine, T4509/6-Thioguanosine, 858412)
- Protein ladder (e.g., BIO-RAD, 161-0374; 10–250 kDa)
- QIAquick gel purification kit (Qiagen)
- SDS (Fisher Scientific, BP166-500)
- Supor membrane syringe filter, 5 μ m (Pall Acrodisc)
- Tris-borate-EDTA (TBE buffer solution)
- Tetramethylethylenediamine (TEMED)
- Tris base (Fisher Scientific, BP152-1)
- Tris-HCl (Promega, PR-H5121)
- Triton X-100
- UreaGel – SequaGel – System (National Diagnostics, EC-833).

Buffers and Solutions

Note: All pH measurements of buffers are performed at room temperature. Buffers and all perishable reagents should be refrigerated for storage. We use water purified by a Millipore water purification system. Buffers are listed in the order of usage.

- 1 M 4-thiouridine or 6-thioguanosine in DMSO
- 1 x NP40 lysis buffer: 50 mM HEPES–KOH, pH 7.5, 150 mM KCl, 2 mM EDTA–NaOH, pH 8.0, 1 mM NaF, 0.5% (v/v) NP40 substitute, 0.5 mM

DTT, complete EDTA-free protease inhibitor cocktail (one tablet per 50 ml buffer)

- 1 x Citrate-phosphate buffer pH 5.0: 4.7 g l⁻¹ citric acid monohydrate, 9.2 g l⁻¹ Na₂HPO₄·7H₂O
- 1 x IP-wash buffer: 50 mM HEPES–KOH, pH 7.5, 300 mM KCl, 0.05% (v/v) NP40 substitute, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (one tablet per 50 ml buffer)
- 1 x High-salt wash buffer: 50 mM HEPES–KOH, pH 7.5, 500 mM KCl, 0.05% (v/v) NP40 substitute, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (one tablet per 50 ml buffer)
- 10 x Dephosphorylation buffer: 50 mM Tris-HCl, pH 7.9, 100 mM NaCl, 10 mM MgCl₂, 1 mM DTT
- 1 x Phosphatase wash buffer: 50 mM Tris-HCl, pH 7.5, 20 mM EGTA–NaOH, pH 7.5, 0.5% (v/v) NP40 substitute
- 1 x T4 PNK buffer without DTT: 50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 10 mM MgCl₂
- 1 x T4 PNK buffer with DTT: 50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 10 mM MgCl₂, 5 mM DTT
- 1 x SDS-polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer: 50 mM Tris-HCl, pH 6.8, 2 mM EDTA–NaOH, pH 8.0, 10% (v/v) glycerol, 2% (v/v) SDS, 100 mM DTT, 0.1% (w/v) bromophenol blue
- 1 x Proteinase K in storage buffer: 20 mg ml⁻¹ proteinase K, 50 mM Tris-HCl, pH 8.0, 30 mM CaCl₂, 50% glycerol
- 2 x Proteinase K buffer: 100 mM Tris-HCl, pH 7.5, 12.5 mM EDTA–NaOH, pH 8.0, 150 mM NaCl, 2% (v/v) SDS
- Acidic phenol/chloroform/IAA (25 : 24 : 1)
- 10 x RNA ligase buffer without ATP: 0.5 M Tris-HCl, pH 7.6, 0.1 M MgCl₂, 0.1 M 2-mercaptoethanol, 1 mg ml⁻¹ acetylated BSA
- 2 x Formamide loading solution: 50 mM EDTA–NaOH, pH 8.0, 0.05% (w/v) bromophenol blue, in formamide
- 5 x TBE: 445 mM Tris base, 445 mM boric acid, 10 mM Na₂EDTA–NaOH, pH 8.0
- 10 x RNA ligase buffer with ATP: 0.5 M Tris-HCl, pH 7.6, 0.1 M MgCl₂, 0.1 M 2-Mercaptoethanol, 1 mg ml⁻¹ acetylated BSA, 2 mM ATP
- SuperScript III Reverse Transcriptase 5 x first-strand buffer (supplied by Life Technologies Invitrogen together with the enzyme): 250 mM Tris-HCl (pH 8.3 at room temperature), 375 mM KCl, 15 mM MgCl₂
- 10 x dNTP solution: 2 mM dATP, 2 mM dCTP, 2 mM dGTP, 2 mM dTTP
- 1 x Alkaline hydrolysis solution: 150 mM KOH, 20 mM Tris base
- 1 x Neutralization solution: 150 mM HCl
- 10 x PCR buffer: 100 mM Tris-HCl, pH 8.0, 500 mM KCl, 10 mM 2-Mercaptoethanol, 1% (v/v) Triton X-100, 20 mM MgCl₂
- 5 x DNA loading solution: 50 mM EDTA–NaOH, pH 8.0, 0.2% (w/v) bromophenol blue, 20% (w/v) Ficoll type 400

Protocol for PAR-CLIP cDNA Library Construction

In the first step, the RBP of interest is crosslinked to its bound RNA targets, which incorporated the photoactivatable ribonucleoside into nascent transcripts in live cells. The cells are collected, the pellet is lysed, and the cell lysate is used for the immunoprecipitation (IP). When using the anti-FLAG antibody for which this protocol was originally developed, we conjugate it to protein G magnetic beads since this represents the optimal Ig-binding protein for the anti-FLAG antibody. However, the choice of protein A *versus* protein G should be considered depending on the individual antibody.

Two RNase T1 digestions ensure that only the RNA segment bound, crosslinked and protected by the RBP is recovered, which will later on allow the exact definition of the RNA binding sites. The short RNAs, bound by the RBP of interest, are radiolabeled and made visible by autoradiography after SDS-PAGE fractionation. The band corresponding to the protein of interest is excised from the SDS-polyacrylamide (SDS-PAA) gel and the mRNA–RBP complex is eluted from the gel piece. After proteolysis, the crosslinked RNA is recovered and used as input for the cDNA library preparation.

PAR-CLIP Preparations

Expansion of Cell Lines to Be Used as Starting Material for PAR-CLIP

Expand cells in appropriate growth medium containing selection antibiotics as suitable to maintain your stable cell line. For one experiment, we usually prepare lysates from 3 to 5 ml wet pellet of crosslinked cells, which corresponds to twenty to fifty 15 cm cell culture plates (for HEK293). However, if material is limiting, we have performed successful PAR-CLIP experiments with less than 0.5 ml wet cell pellet. An amount of 2×10^8 HEK293 cells (10×15 cm plates) will yield ~ 1 ml of wet cell pellet.

Cells are grown to $\sim 80\%$ confluency. Fourteen hours before crosslinking, add 4SU to a final concentration of $100 \mu\text{M}$ directly to the cell culture medium. 6SG (final concentration of $100 \mu\text{M}$) can also be used as the photoactivatable ribonucleoside. Induce expression of protein, if necessary.

Testing of Antibodies Used for Immunoprecipitation

This protocol was originally developed for the use of anti-FLAG antibodies; use of a different antibody will likely require optimization of immunoprecipitation and wash conditions before starting a large-scale experiment. Ensure that the optimal salt concentration for antibody binding is maintained throughout the protocol; washes of the immunoprecipitate with high salt may disrupt antibody/antigen interactions.

If in doubt, you can use NP40 lysis buffer instead of the IP-wash and the high-salt wash buffers; however, removal of non-specifically interacting RNAs might be less efficient. When using, for example, an anti-AGO2 antibody from Millipore (04-642), we substitute the three wash steps with IP-wash buffer with NP40 lysis buffer (150

instead of 300 mM KCl), and also the three wash steps with high-salt wash buffer again with NP40 lysis buffer (150 instead of 500 mM KCl). Except for these two buffers, all the other washing steps are performed as described.

Radiolabeling of RNA Size Markers

5'-³²P-radiolabeled RNA size markers are needed for the cDNA library preparation. It is advisable to prepare them before starting the PAR-CLIP experiment. Perform a standard 5'-end-labeling procedure with T4 PNK and [γ -³²P]ATP according to the manufacturer's guidelines and gel-purify the markers (e.g., phosphorylate 1 μ M RNA size marker in a reaction volume of 10 μ l using 1 μ l of conventional [γ -³²P]ATP). Keep radioactive gel pieces from the running front of this gel as markers to implant into gels for alignment of phosphorimager printouts after image plate exposure to gels.

Isolation of RNA Fragments Bound by the RNA Binding Protein of Interest

UV-Crosslinking of 4SU-Labeled Cells

For Adherent Cells

- 1) Decant the growth medium.
- 2) Wash cells once with 5 ml ice-cold PBS per plate and remove PBS completely by decanting and inverting the cell culture dish.
- 3) Place plates on a tray filled with ice to keep cells cold and irradiate uncovered with 0.15 J cm⁻² total energy of 365 nm UV light in a Stratalinker 2400 or similar device. (*Note:* the display of the Stratalinker says " μ J $\times$ 100"; however, it indicates " μ J $\times$ cm⁻² $\times$ 100". To irradiate with 0.15 J cm⁻², set the device to "1500"; the bulbs have a nominal power of 75 W and the period of irradiation is automatically determined by the sensor of the Stratalinker, making the irradiation time-dependent on the age of the bulbs; the distance of the cells to the radiation source is dictated by the position of the sensor and the fact that they have to be on ice; in our hands, this translates into a distance of about 15 cm.
- 4) Add 3 ml PBS per plate and detach cells with a cell scraper. Transfer to pre-chilled 50 ml centrifugation tubes on ice. After the cells from the last plate have been collected, centrifuge at 500 g for 5 min at 4 °C, and discard the supernatant. Expect to obtain about 2–5 ml of wet cell pellet from twenty to fifty 15 cm plates.
- 5) Unless you want to continue directly with cell lysis, snap freeze the cell pellet in liquid nitrogen, and store at –80 °C. Cell pellets can be stored for at least 12 months.

For Cells Grown in Suspension Culture

- 1) Collect cells by centrifugation at 500 g for 5 min at 4 °C.
- 2) Wash cells by resuspending in 20 ml ice-cold PBS and spin again at 500 g for 5 min at 4 °C.

- 3) Resuspend cells in 20 ml ice-cold PBS and transfer into one 15 cm cell culture plate.
- 4) Place plate on a tray with ice and irradiate uncovered with 0.2 J cm^{-2} of 365 nm UV light in a Stratalinker 2400 or similar device. (Note: the display of the Stratalinker says " $\mu\text{J} \times 100$ ". However, it indicates " $\mu\text{J} \times \text{cm}^{-2} \times 100$ ". To irradiate with 0.2 J cm^{-2} , set the device to "2000". For more information, see step 3 for adherent cells.)
- 5) Transfer cells into a 50 ml centrifugation tube and collect by centrifugation at 500 g for 5 min at 4°C , and discard the supernatant.
- 6) Unless you want to continue directly with cell lysis, snap freeze the cell pellet in liquid nitrogen, and store at -80°C . Cell pellets can be stored for at least 12 months.

Preparation of Cell Lysate for Immunoprecipitation

- 1) Thaw the crosslinked cell pellet (3–5 ml, see Section "Expansion of Cell Lines to Be Used as Starting Material for PAR-CLIP") on ice. Prepare the magnetic beads (see Section "Preparation of the Magnetic Beads for Immunoprecipitation") while the pellet thaws. Then resuspend the cell pellet in three cell pellet volumes of NP40 lysis buffer and incubate on ice for 10 min.
- 2) Clear the cell lysate by centrifugation at 13 000 g for 15 min at 4°C .
- 3) Clear the lysate further by filtering it through a $5 \mu\text{m}$ membrane syringe filter. Attach the syringe filter to a 20 ml syringe, remove the plunger, and transfer the supernatant into the syringe. Be careful to hold the syringe above the 50 ml conical tube since the lysate will start to drip through the filter. Then insert the plunger and gently apply pressure until all of the lysate has been filtered. Depending on the initial viscosity of the lysate, it might be necessary to replace a clogged filter with a fresh one.
- 4) Add RNase T1 ($1000 \text{ U } \mu\text{l}^{-1}$) to a final concentration of $1 \text{ U } \mu\text{l}^{-1}$ and incubate in a water bath for 15 min at 22°C . Mix by inversion from time to time. Both incubation temperature and duration are critical at this step for a controlled partial RNase T1 digestion. Cool reaction for 5 min on ice before proceeding.
- 5) Remove a 10 μl aliquot for immunoblotting as a control for the protein levels used as input and freeze it at -20°C .

Preparation of the Magnetic Beads for Immunoprecipitation

- 1) Transfer 10 μl of Protein G magnetic particles (Dynabeads Protein G) per milliliter cell lysate (typically $\sim 100\text{--}150 \mu\text{l}$ of beads) to a 1.5 ml microtube. Put the magnetic rack on ice. Wash the beads twice with 1 ml of citrate-phosphate buffer. Be careful not to let the magnetic beads dry out.
- 2) Resuspend the beads in twice the volume of citrate-phosphate buffer relative to the original volume of bead suspension (i.e., 200–300 μl).
- 3) Add antibody to a final concentration of 0.25 mg ml^{-1} and incubate on a rotating wheel for 40 min at room temperature.
- 4) Collect the beads and wash twice in 1 ml of citrate-phosphate buffer to remove unbound antibody.

- 5) Resuspend beads in twice the volume of citrate-phosphate buffer relative to the original volume of bead suspension.

Immunoprecipitation and Second RNase T1 Treatment

- 1) Add 20 μl of freshly prepared antibody-conjugated magnetic beads (see Section “Preparation of the Magnetic Beads for Immunoprecipitation”) per milliliter of partially RNase T1-treated cell lysate from the Section “Preparation of Cell Lysate for Immunoprecipitation” and incubate in a 15 ml centrifugation tube on a rotating wheel for 1 h at 4 °C.
- 2) Collect magnetic beads on a magnetic particle collector for 15 ml centrifugation tubes (Life Technologies Invitrogen) and remove the supernatant. Keep an aliquot of about 50 μl for immunoblotting.
- 3) Add 1 ml of IP-wash buffer and transfer to 1.5 ml polypropylene tubes.
- 4) Wash beads twice in 1 ml IP-wash buffer.
- 5) Resuspend beads in one original bead volume of IP-wash buffer.
- 6) Add RNase T1 (1000 U μl^{-1}) to a final concentration of 100 U μl^{-1} and incubate the bead suspension in a water bath for 15 min at 22 °C. Both RNase T1 incubation temperature and time are crucial to avoid overdigestion of RNA, which could result in RNA segments too short to be mapped unambiguously to transcript or genomic sequences. Cool on ice for 5 min.
- 7) Wash beads three times with 1 ml high-salt wash buffer.
- 8) Resuspend beads in 1 volume of 1 x dephosphorylation buffer.

Dephosphorylation and Radiolabeling of RNA Segments Crosslinked to Immunoprecipitated Proteins

- 1) Add CIP to a final concentration of 0.5 U μl^{-1} , and incubate the suspension for 10 min at 37 °C while mixing at 800 rpm in a thermomixer.
- 2) Wash beads twice in 1 ml of 1 x phosphatase wash buffer.
- 3) Wash beads twice in 1 x T4 PNK buffer without DTT.
- 4) Resuspend the beads in one original bead volume of 1 x T4 PNK buffer containing DTT.
- 5) Add [γ - ^{32}P]ATP to a final concentration of 0.1 $\mu\text{Ci } \mu\text{l}^{-1}$ and T4 PNK to 1 U μl^{-1} to the bead suspension. Incubate the suspension for 30 min at 37 °C and 800 rpm in a thermomixer with thorough mixing by hand every 5–10 min.
- 6) Add 100 μM non-radioactive ATP and incubate for another 5 min at 37 °C. This ensures that all RNAs are fully 5'-phosphorylated, which is required for the 5'-adapter ligation. (see Section “5'-Adapter Ligation for cDNA Library Preparation”)
- 7) Wash the magnetic beads five times with 800 μl of 1 x T4 PNK buffer without DTT; dispose of the radioactive buffer according to the local guidelines.
- 8) Resuspend the beads in 65 μl of 1 x SDS-PAGE loading buffer and incubate for 5 min in a heat block at 90 °C to denature and release the immunoprecipitated RBP with the crosslinked radiolabeled RNAs from the beads. Vortex.

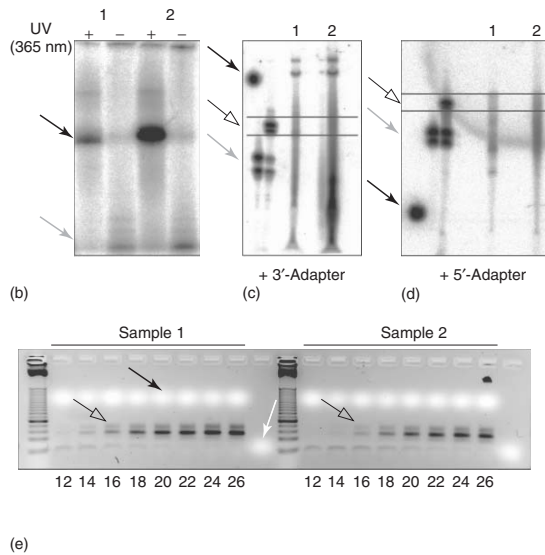
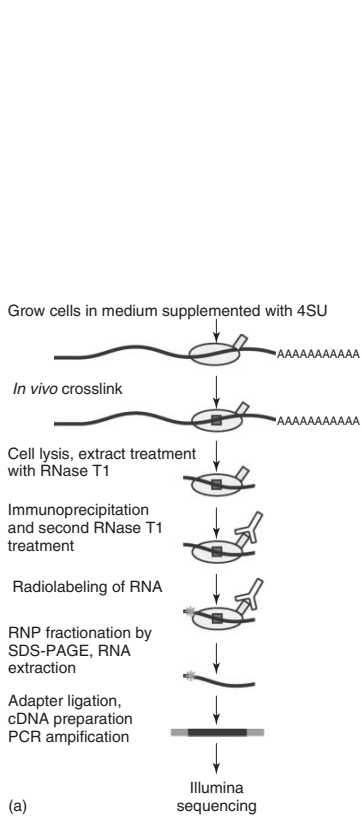
- 9) Remove the magnetic beads on the separator and transfer the supernatant to a clean 1.5 ml microtube. (*Pause point*: you can freeze the supernatant and continue with the protocol at another time)

SDS-PAGE and Electroelution of Crosslinked RNA–Protein Complexes from Gel Slices

- 1) Load $2 \times 30 \mu\text{l}$ of the supernatant into two adjacent wells on a Novex Bis-Tris 4–12% (Life Technologies Invitrogen) precast SDS-PAA gel and leave at least one lane empty between different samples/different proteins of interest (Figure 39.1b). On both sides of the gel, load a protein ladder. Keep the remaining $5 \mu\text{l}$ of the bead eluate for immunoblotting.
- 2) Run the gel in $1 \times$ MOPS SDS running buffer for 45–60 min at 200 V until the bromophenol blue has reached the bottom of the gel.
- 3) Disassemble the gel chamber and gently dismantle the gel, leaving it mounted on one plate. Cut the protruding bottom of the gel so that the gel lies flat on the phosphorimager screen.
- 4) To facilitate the alignment of the gel to the phosphorimager paper printout later on, place three tiny radioactive gel pieces (which you collected earlier when radiolabeling the size markers; Section “Radiolabeling of RNA Size Markers”) asymmetrically into three of the four corners of the gel.
- 5) After placing the radioactive gel pieces, wrap the gel in plastic wrap and expose a blanked phosphorimager screen to the gel for 15 min. Visualize it

Figure 39.1 Selected PAR-CLIP experimental steps. (a) Illustration of PAR-CLIP. 4SU-labeled transcripts are crosslinked to RBPs, and partially RNase-digested RNA–protein complexes are immunopurified and size-fractionated. RNA molecules are then recovered and converted into a cDNA library analyzed by deep sequencing. (b) SDS-PAGE of crosslinked RNPs after immunoprecipitation and 5'-end-labeling of RNA components. The black arrow points to the radioactive bands corresponding to the expected size of the RBP (IGF2BP1, running at 65 kDa) and the gray arrow points to the radioactive running front. (c) Denaturing (8 M urea) PAGE after 3'-adapter ligation. The black filled arrow indicates one of the inserted little radioactive gel pieces to facilitate alignment of gel to printout, the open arrow depicts the size markers ligated to the 3'-adapter, and the black lines mark the area that was excised from the gel and further processed; the gray arrow indicates the unligated size markers. (d) Denaturing (8 M urea) PAGE after 5'-adapter ligation. The black filled arrow indicates one of the

inserted little radioactive gel pieces to facilitate alignment of gel to printout, the open arrow depicts the size markers after additional ligation to the 5'-adapter, and the black lines mark the area that was excised from the gel and further processed; the gray arrow indicates the size markers ligated to the 3'-adapter, but not to the 5'-adapter. Lanes 1 and 2 in panels b-d indicate two independent samples. (e) Agarose gel after small-scale trial PCR. The black filled arrow points to the position of the xylene cyanol loading dye, and the white one to bromophenol blue running close to the gel front. Bands of about 75 and 100 bp are detectable, representing an insert-less 5'-adapter–3'-adapter PCR side product and expected insert-containing PCR products, respectively. The open arrows indicate the lanes of samples after 16 PCR cycles, conditions that were chosen for the large-scale experiment. A 25 bp ladder is loaded to the left of each set of experiments; the fourth band from the bottom corresponds to 100 bp. A negative control (water instead of cDNA) was performed but is not shown.



on a phosphorimager. Have a second blanked screen ready and expose it during the scanning process, should the first exposure indicate that a longer exposure is necessary.

- 6) Print the scanned image file in its original size (100%). Align the transparently wrapped gel on top of the printout guided by the implanted gel pieces for precise positioning. Cut out the bands that correspond to the expected size of the RBP (Figure 39.1b).
- 7) Add 800 μl H_2O to each D-Tube Dialyzer Midi Tube used for electroelution and allow to stand at room temperature for 5 min. Remove the water. Take care not to pierce the membrane.
- 8) Transfer each excised band to a dialyzer tube and add 800 μl 1 x MOPS SDS running buffer.
- 9) Place the tubes on the electroelution rack in a standard and sufficiently sized agarose gel chamber, such that the membrane is exposed to the flow of the current (for details, see manufacturer's instructions). Make sure that the membrane of the dialyzer tube is aligned correctly to allow flow of current. Fill the chamber with 1 x MOPS SDS running buffer until the tubes are covered by buffer.
- 10) Electroelute the crosslinked RNA–RBP complex at 100 V for 1.5 h. Reverse the current for 2 min to release any protein attached to the dialysis membrane.
- 11) Transfer the solution to two siliconized tubes so that each contains around 350 μl (you will not be able to fully recover the original 800 μl). (*Pause point:* freeze the solution at -20°C and continue the next day).

Proteinase K Digestion

- 1) Add 1 vol. of 2 x Proteinase K buffer, followed by the addition of Proteinase K (Roche) to a final concentration of 1.2 mg ml^{-1} . Incubate for 30 min at 55°C . If the volume per tube exceeds 800 μl at this stage, split the sample once more into two tubes per original tube.
- 2) Add 1 vol. of acidic phenol/chloroform/IAA, vortex and spin at 20 000 g for 10 min at 4°C . Recover the upper aqueous phase without disturbing the interphase and pipette into two siliconized tubes. Repeat the phenol/chloroform/IAA extraction until there is no precipitate visible in the interphase (usually once is sufficient but two or more times might be needed).
- 3) Add an equal volume of chloroform, vortex and spin at 20 000 g for 10 min at 4°C . Again, recover the aqueous phase without disturbing the interphase.
- 4) Add 1/10 vol. 3 M NaCl, 1 μl glycogen (10 mg ml^{-1} stock), and 3 vol. 100% ethanol.
- 5) Precipitate the RNA for 1 h on ice and spin at 20 000 g for 15 min at 4°C . (*Pause point:* precipitate the RNA overnight at -20°C)
- 6) Take off the supernatant, air-dry the pellets, and dissolve in a total of 10 μl H_2O . Monitor the radioactivity of the supernatant and the pellet to assess the efficiency of the ethanol precipitation.

cDNA Library Preparation from Recovered RNA Fragments

Illumina sequencing adapters are then ligated to the recovered RNA. Subsequently, the RNA pieces with adapters on both sides are reverse-transcribed and amplified by PCR to create sufficient starting material for the next-generation sequencing. To minimize the distortion of the cDNA library composition by excessive PCR and to recognize possible failure during reverse transcription leading to false positive PCR results, we monitor the accumulation of the PCR product during a pilot PCR. To determine the minimal cycle number, a small-scale trial PCR is performed before the final large-scale PCR. For purification, load the large-scale reaction mixture into at least two wells of an agarose gel to avoid overloading of individual lanes. After electrophoresis, the PCR product is excised and extracted from the gel, reanalyzed by analytical agarose gel electrophoresis, and submitted to sequencing. A second gel extraction should be performed if any 5'-adapter–3'-adapter products should still be detectable after the first gel extraction.

3'-Adapter Ligation for cDNA Library Preparation

- 1) Prepare the following reaction mixture for ligating the 3'-adapter, multiplying the volumes by the number of ligation reactions (plus one for the size markers, see below) to be performed plus another extra volume to account for pipetting errors:
 - a. 2 μ l 10 x RNA ligase buffer (without ATP)
 - b. 6 μ l 50% DMSO
 - c. 1 μ l 100 μ M preadenylated 3'-adapter
- 2) Add 9 μ l of the reaction mixture to each sample (so that you have 19 μ l per tube).
- 3) Prepare ~40 fmol from a 1 : 100 dilution of 5'-³²P-labeled RNA size markers (19 and 24 nt size marker at equimolar concentrations; see above). This controls for successful ligation and indicates the length of the bands that will later on be cut out from the gel.
- 4) Denature the RNA to disrupt secondary structures by incubating for 30 s at 90 °C. Place the tubes on ice immediately for 30 s.
- 5) Add 1 μ l T4 Rnl2(1-249)K227Q (1 μ g μ l⁻¹) to the ligation reactions, mix gently and incubate overnight on ice in the cold room or in an insulated ice bucket covered with a lid.
- 6) The next morning, cast a 15% denaturing 8M urea PAA gel (we use the UreaGel system from National Diagnostics) and wait until the polymerization process is complete. Our gels measure 15 cm $\times$ 17 cm $\times$ 0.8 mm and contain about 25 ml gel volume, and we use a 20-well comb.
- 7) Pre-run the gel for 30 min at 30 W using 0.5 x TBE buffer. After the pre-run, flush the wells with 0.5 x TBE.
- 8) Add 20 μ l of 2 x formamide loading solution to the samples to stop the ligation reactions.
- 9) Denature the RNA for 30 s at 90 °C.

- 10) Load each sample into one well (or two) of the gel. Load the size markers symmetrically on both sides of the gel to allow for approximation of the length of the ligated samples between them. Use the center of the gel to guarantee even running of the gel. Make sure to space different samples appropriately, typically at a two-well distance, to avoid cross-contamination. Ensure that the overall loading of the gel is asymmetrical.
- 11) Run the gel for 45 min at 30 W in 0.5 x TBE buffer until the bromophenol blue dye has almost reached the bottom of the gel.
- 12) Dismantle the gel, leaving it mounted on one glass plate. To facilitate the alignment of the gel to the phosphorimager paper printout, again implant three tiny radioactive gel pieces asymmetrically at three of the four corners of the gel. Cover the gel in plastic wrap.
- 13) Expose a phosphorimager screen to the gel for at least 1 h. If the radioactivity of the recovered RNA is weak, you may expose the screen overnight, placing the exposure cassette in a -20°C freezer.
- 14) Align the gel on top of a printout scaled to 100% according to the position of the three radioactive gel pieces. In sample lanes, use the position of the 19 nt/3'-adapter ligation product and a position somewhat above the 24 nt/3'-adapter ligation product as margins for gel excision (Figure 39.1c). (*Note:* We do not recommend excision of RNA that is running below the 19 nt marker line. For our bioinformatic analyses, all sequences shorter than 20 nt are discarded because of the increased probability of mapping to multiple locations, which would lead to uncertain locus assignment. In case you would like to cut out a larger size range, two longer size markers (33 and 35 nt), which we also used successfully, have been included in the Section "Oligonucleotides". Also, cut out the ligated 19 and 24 nt size markers, which will serve once more as a ligation control in the next step (Figure 39.1d).
- 15) Place the excised gel pieces in siliconized tubes and add 350 μl 0.4 M NaCl (ensure that the gel pieces are covered by the NaCl solution). Elute the ligation products overnight at 4°C with shaking at 800 rpm in a thermomixer.
- 16) Transfer the supernatant into a new siliconized tube and add 1 ml 100% ethanol. Precipitate the RNA for 1 h on ice and spin at 20 000 g for 15 min at 4°C .
- 17) Take off the supernatant, air-dry the pellets, and dissolve in a total of 9 μl H_2O . Dissolve the ligated markers in 12 μl H_2O .

5'-Adapter Ligation for cDNA Library Preparation

- 1) Prepare the following reaction mixture for the ligation of the 5'-adapter, multiplying the volumes by the number of ligation reactions to be performed (again, also include the positive control) plus one extra volume to account for pipetting errors:
 - a. 1 μl 100 μM 5'-adapter
 - b. 2 μl 10 x RNA ligase buffer with ATP
 - c. 6 μl 50% DMSO

- 2) Combine 9 μl of this mixture with the 9 μl of the gel-eluted 3'-adapter ligation sample (step 17 above). Do not forget to also process the ligated markers. Ligate 9 μl out of the 12 μl (step 17 above) and keep 3 μl as an unligated control for the next gel.
- 3) Denature the RNA by incubation for 30 s at 90 °C. Place the tube immediately on ice for 30 s.
- 4) Add 2 μl T4 RNA ligase 1 (10 U μl^{-1}), mix gently, and incubate for 1 h at 37 °C.
- 5) In the meantime, cast a 12% denaturing 8 M urea PAA gel and wait until the polymerization process is complete. We again use 0.8 mm spacers and a 20-well comb.
- 6) Pre-run the gel for 30 min at 30 W in 0.5 x TBE buffer. After the pre-run, gently flush the wells with 0.5 x TBE.
- 7) Add 20 μl of 2 x formamide loading solution, incubate the samples at 90 °C for 30 s, and load them onto the gel. Make sure to space different samples appropriately, typically at a two-well distance, to avoid cross-contamination.
- 8) Load 50% of the ligated markers on the left side and 50% on the right side of the sample lane(s). Load the remaining 3 μl unligated marker on one side but remember which (Figure 39.1d). Make sure that the loading of the gel is asymmetrical.
- 9) Run the gel for 45 min at 30 W using 0.5 x TBE buffer until the bromophenol blue dye has almost reached the bottom of the gel. Disassemble and image the gel as described earlier for the 3' ligation (start with an exposure roughly twice as long as for the 3' ligation) and excise the new ligation product (again include the ligated markers). You can also recover unligated material by excising the gel region below the position of the 3'-adapter/19 nt marker/5'-adapter ligation product since this represents 3'-ligated, 5'-unligated RNA fragments. Freeze these gel pieces as a backup, which permits you to perform another 5'-adapter ligation with RNAs eluted from these gel pieces.
- 10) Elute the ligation products from the gel slices overnight at 4 °C with shaking at 800 rpm in a thermomixer in 350 μl 0.4 M NaCl. Add 1 μl 100 μM PCR reverse primer as a carrier to facilitate the recovery of the ligation products.
- 11) Transfer the supernatant into a new siliconized tube and add 1 ml 100% ethanol. Precipitate the RNA for 1 h on ice and spin at 20 000 g for 15 min at 4 °C.
- 12) Remove the supernatant, air-dry the pellets, and dissolve in a total of 5.6 μl H₂O.

Reverse Transcription

- 1) Prepare the following reaction mix (multiplied by the number of samples plus one to account for pipetting errors):
 - a. 1.5 μl 0.1 M DTT
 - b. 3 μl 5 x first-strand buffer
 - c. 4.2 μl 10 x dNTPs

- 2) Denature the RNA from step 12 above by incubating the tube for 30 s at 90 °C and transfer the tube to a 50 °C thermomixer.
Add 8.7 µl of reaction mix (step 1) to each RNA sample and incubate for 3 min at 50 °C. Add 0.75 µl SuperScript III Reverse Transcriptase and incubate for 1 h at 42 °C.
- 3) Prepare 150 mM KOH/20 mM Tris base and 150 mM HCl and check on pH paper that a 1 + 1 mix results in a pH between 7.0 and 9.5. If not, change the ratios until the pH is within that range.
- 4) To hydrolyze the RNA, add 40 µl of 150 mM KOH/20 mM Tris base and incubate for 10 min at 90 °C.
- 5) Neutralize the solution by adding ~ 40 µl 150 mM HCl (the exact volume depends on the ratio determined in step 3 above) and check the pH of the mixture by spotting 1 µl on a pH paper. The pH should be between 7.0 and 9.5 so that the subsequent PCR is not inhibited. If necessary, readjust the pH by adding more base or acid.

PCR Amplification of cDNA Library and Sample Preparation for Sequencing

- 1) Prepare the following mix multiplied by the number of samples plus one for the negative control:
 - a. 40 µl 10 x PCR buffer
 - b. 40 µl 10 x dNTPs
 - c. 2 µl 100 µM PCR forward primer
 - d. 2 µl 100 µM PCR reverse primer
 - e. 272 µl H₂O
- 2) Use 89 µl of reaction mix for the pilot PCR reaction to determine the minimal cycle number; the remainder is needed for the large-scale PCR (freeze the reaction mix if you do not plan to run the large-scale PCR on the same day). To the 89 µl of reaction mix, add 10 µl cDNA and 1 µl *Taq* polymerase (5 U µl⁻¹). Remember to include a negative control (H₂O instead of cDNA).
- 3) Use the following cycle conditions: 45 s at 94 °C, 85 s at 50 °C, 60 s at 72 °C.
- 4) To determine the necessary number of cycles for amplifying the cDNA library, remove 12 µl aliquots after each cycle starting with cycle number 12 and up to cycle number 26. To remove aliquots from the PCR tube, temporarily pause the PCR cycler at the end of the 72 °C step. You may use a multichannel pipette to remove the aliquots.
- 5) Analyze 6 µl of each sample on a 2.5% agarose gel containing 0.4 µg ml⁻¹ of ethidium bromide. Load a 25 bp ladder on each side and load all cycle aliquots from one PCR sample next to each other in an ascending order. The PCR products might appear as a double band with the higher band running at the expected length of about 95–110 bp and a lower band corresponding to the 3'-adapter-to-5'-adapter ligation side products running at about 75 bp. Figure 39.1e illustrates a typical small-scale PCR. (The open arrows indicate the chosen number of cycles for the large-scale experiment.)

- 6) Define the minimal cycle number for the cDNA amplification, which should be within the exponential amplification phase of the PCR, about five cycles away from reaching the saturation level of PCR amplification. For a typical PAR-CLIP experiment, the minimal number of cycles is between 16 and 20. (*Pause point:* you can pause at any time before or after the large-scale PCR.)
- 7) Perform a 300 μl large-scale PCR (100 μl per well) with the determined minimal number of cycles and combine all three PCR reactions. Again, include a negative control (H_2O instead of cDNA).
- 8) Analyze 6 μl of the products next to the corresponding products from the pilot PCR (the 6 μl aliquots that remained after steps 4 and 5) on a 2.5% agarose gel containing 0.4 $\mu\text{g ml}^{-1}$ of ethidium bromide to check for consistency.
- 9) To the remaining 264 μl , add 26.4 μl 3 M NaCl and 1 ml 100% EtOH. Precipitate for 1 h on ice and spin at 20 000 g for 30 min at 4 °C. Take off the supernatant, air-dry the pellet, and dissolve in 40 μl 1 x DNA loading solution (5 x DNA loading solution diluted in 0.5 x TBE).
- 10) Divide the sample into two wells of a 2.5% low melting agarose gel containing 0.4 $\mu\text{g ml}^{-1}$ ethidiumbromide. Run the gel at 120 V for 2–3 h.
- 11) Do not overload the gel as this will compromise its separating capacity.
- 12) Visualize the DNA on a 365 nm transilluminator and excise the band corresponding to 95–110 bp with a clean scalpel.
- 13) Purify the DNA using the QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. Include the isopropanol step as described for short fragments. Elute in 30 μl elution buffer.
- 14) Analyze 5 μl of the eluate on a 2.5% agarose gel containing 0.4 $\mu\text{g ml}^{-1}$ ethidium bromide to ensure the removal of any unwanted amplified 5'-adapter–3'-adapter PCR products.
- 15) Submit 10 μl of the purified cDNA to Illumina sequencing.
- 16) Run 1 μl of the sample, for example, on an Agilent Bioanalyzer using the high-sensitivity kit to accurately determine its concentration. The sample is then diluted to a 10 nM stock of which 2 μl are employed in the denaturing step in a total volume of 20 μl . Four microliters of this reaction are brought up to 1000 μl and 1/20 of this reaction is loaded on a single lane of a flow cell.

Determination of Incorporation Levels of 4SU into Total RNA

To optimize crosslinking of protein to RNA, it is useful to determine the fraction of substitution of uridine by 4SU. This is especially necessary when changing cell growth conditions or cell type. Total RNA is isolated and enzymatically degraded to monomeric ribonucleosides, which are separated and quantified by high-performance liquid chromatography (HPLC) analysis [13].

Materials

Some of the required materials are given in the first “Materials” section of this chapter and are not listed here again.

Equipment

- Speed Vac
- HPLC with a Supelco Discovery C18 (bonded phase silica 5 μ M particle, 250 $\times$ 4.6 mm) reverse phase column (Bellefonte PA, USA)

Reference Oligoribonucleotides

- 5' CGUACGCGGAUACUUCGA(4SU)U (e.g., from Thermo Scientific)
- 5' CGUACGCGGAUACUUCGAUU

Enzymes

- Bacterial alkaline phosphatase (Worthington Biochemical, LS006344)
- Snake venom phosphodiesterase I (Worthington Biochemical, LS003926).

Consumables, Kits, and Chemicals

- Acetonitrile
- rA, rG, rC, rU, and 4SU (Sigma–Aldrich, T4509)
- Acetic acid – triethylamine solution (1 + 1) (TEAA, Sigma–Aldrich, 09748)
- TRIzol reagent (Life Technologies Invitrogen, 15596-026)
- Dry ice.

Buffers and Solutions

- 0.5 M Tris-HCl (pH 7.5 at room temperature)
- 3 M NaOAc (pH 5.2)
- HPLC solution A: 3% Acetonitrile, 0.1 M TEAA
- HPLC solution B: 90% Acetonitrile.

Protocol***Preparation of Cells***

- 1) Grow HEK293 cells on a 10 cm plate in regular medium supplemented with 100 μ M 4SU for 16 h before harvest. As a control, also grow cells without adding 4SU.
- 2) Decant the growth medium; wash cells once with 1 x PBS.

Extraction of Total RNA

- 1) Add 1 ml TRIzol reagent directly onto the plate and isolate total RNA according to the manufacturer's instructions.
- 2) Include 1 mM DTT (prepare a 1 M stock that can be frozen for storage) in isopropanol and ethanol wash steps to prevent oxidization of the thiocarbonyl group, yielding disulfides or uridine.
- 3) Dissolve the RNA pellet in 60 μ l H₂O containing 1 mM DTT.
- 4) Determine the concentration of the obtained RNA. Expect to obtain about 50–100 μ g total RNA per 10 cm plate.

Dephosphorylation and Enzymatic Hydrolysis of Total RNA to Ribonucleosides

For HPLC analysis, 0.2 OD₂₆₀ (8.0 µg RNA) of total RNA is digested and dephosphorylated to single nucleosides.

- 1) Set up the following reaction:

Reagent or solution	Final concentration	Volume (µl)
RNA	0.2 OD ₂₆₀	x
10 mM DTT solution	0.1 mM	1.3
1.0 M MgCl ₂ solution	13.8 mM	1.8
0.5 M Tris-HCl buffer (pH 7.7)	34.6 mM	9.0
Bacterial alkaline phosphatase	1.6 U	x
Snake Venom Phosphodiesterase I	0.2 U	x
H ₂ O	—	To 130 µl

- 2) Digest for 16 h at 37 °C. As an additional control, also digest and analyze synthetic RNAs with and without 4SU, such as the reference oligonucleotides listed above under “Materials”.

Sample Preparation for HPLC

- 1) Add 2.3 µl 100 mM DTT, 4 µl of 3 M NaOAc (pH 5.2), and 100 µl of cold 100% EtOH to each hydrolysis sample; incubate on dry ice for 10 min; and centrifuge the sample at 12 500 g for 5 min at 25 °C.
- 2) Collect the supernatant, add 3 µl of 100 mM DTT and 300 µl of cold 100% EtOH, cool on dry ice for 10 min, centrifuge the sample at 12 500 g for 5 min at 25 °C, and collect the supernatant.
- 3) Evaporate the supernatant to complete dryness in a SpeedVac. If the sample is not completely dried, the retention times during HPLC analysis are affected.
- 4) Dissolve the sample in 50 µl H₂O, which is the volume of one HPLC injection. *Note:* One injection should correspond to 0.2 OD₂₆₀ or 8.0 µg of RNA.

Analysis of Ribonucleosides by HPLC

- 1) Separate ribonucleosides on a Supelco Discovery C18 reverse phase column (bonded phase silica 5 µm particles, 250 × 4.6 mm).
- 2) Use an isocratic gradient of 0% B for 15 min, 0–10% B for 20 min, 10–100% B for 30 min, and a 5 min 100% B wash applied between runs to clean the HPLC column.
- 3) Calculate the absorption ratios from the known sequence first, which are then used to estimate the incorporation rate for 4SU (in our experiments ~2.5% of U is substituted with 4SU) (Figure 39.2).

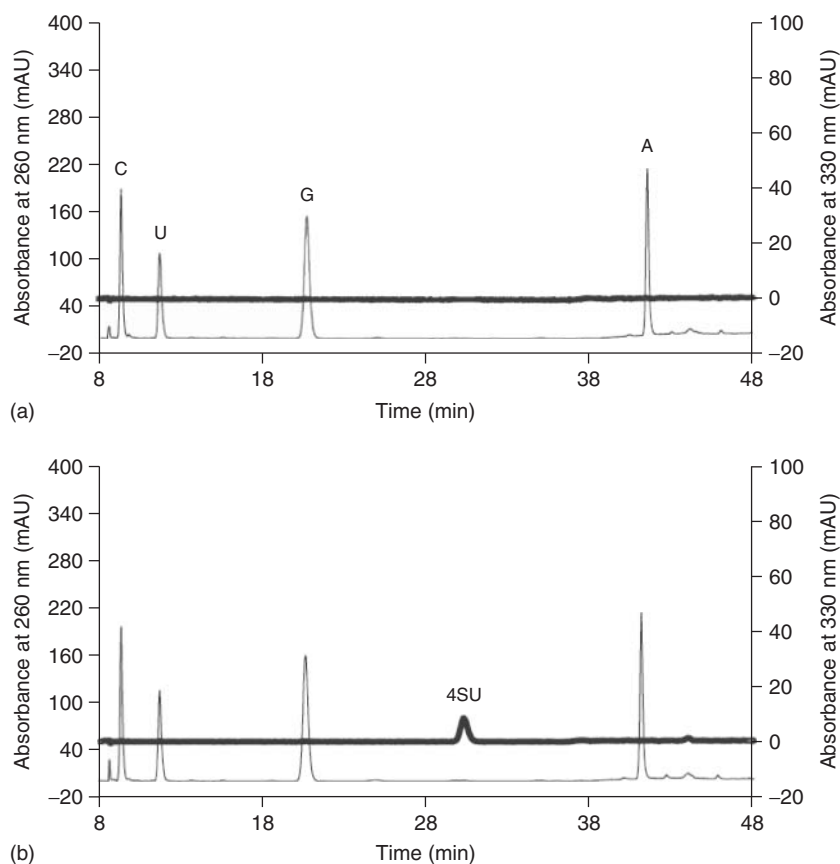


Figure 39.2 Reversed phase HPLC (RP-HPLC) analysis of extracted total RNA (0.2 OD₂₆₀, 8.0 µg RNA) to estimate 4SU incorporation in HEK293 cells. (a) As a control, total RNA extracted from HEK293 cells grown without 4SU addition was digested and analyzed in parallel. Before analyzing hydrolysates of total RNA from HEK293 cells, the positions of peaks corresponding to C, U, G, and A were identified by RP-HPLC of mononucleosides

derived from the unmodified reference oligoribonucleotide. (b) Enzymatic digestion of total RNA from HEK293 cells grown in medium supplemented with 0.1 mM 4SU for 16 h at 37 °C. Analysis of the enzymatic digest at 330 nm showed the presence of a new peak corresponding to 4SU with an estimated incorporation efficiency of ~2.5% (~2.5% of all U residues substituted with 4SU). Thin trace, 260 nm; thick trace, 330 nm.

- 4) Confirm U and 4SU retention times by coinjection of corresponding nucleoside standards.
- 5) Calculate the substitution ratio of 4SU by dividing the peak areas for 4SU at 330 nm and rU at 260 nm by the corresponding extinction coefficients (see below) and determining the ratio of the two quotients: Percentage U substituted with 4SU = $(\text{Area}_{4\text{SU}, 330 \text{ nm}} / \epsilon_{4\text{SU}, 330 \text{ nm}}) \times (\epsilon_{\text{U}, 260 \text{ nm}} / \text{Area}_{\text{U}, 260 \text{ nm}}) \times 100$.

Nucleoside	Extinction coefficient at 260 nm (pH 7.0) (M ⁻¹ cm ⁻¹)	Extinction coefficient at 330 nm (M ⁻¹ cm ⁻¹)
rA	15400	0
rC	7200	0
rG	11500	0
rU	9900	0
4SU	4250	17000

39.2

Troubleshooting

- 1) Designate a set of pipettes for working with RNases to avoid contamination at later RNA isolation and cDNA library preparation steps. Use aerosol barrier tips and take general precautions to avoid any RNase contaminations.
- 2) To confirm that the correct band was excised from the SDS-PAA gel (Figure 39.1b), run another small-scale SDS-PAA gel with 1 or 2 μ l of the remaining 5 μ l of your sample (see Section “SDS-PAGE and Electroelution of Crosslinked RNA–Protein Complexes from Gel Slices”, step 1). After transferring the SDS gel to a nitrocellulose membrane, first perform an autoradiography (exposure for 1–2 h) before conducting a standard immunoblot with protein-specific antibodies. Overlaying the resulting images allows one to verify that the protein of interest has been excised before proceeding with the next protocol steps.
- 3) Use siliconized tubes until you have obtained your cDNA library; at low concentrations, nucleic acids have a tendency to stick to the tube walls.
- 4) Keep supernatants from any ethanol precipitations. In case no pellet is detectable after the precipitation, add 1 μ l glycogen (1 mg ml⁻¹ stock solution) to the supernatant and precipitate again.

Acknowledgments

J.I.H. is supported by the Deutsche Forschungsgemeinschaft (DFG). M.H. was supported by the Deutscher Akademischer Austauschdienst (DAAD) and is currently funded by a fellowship of the Charles Revson, Jr. Foundation. T.T. is an HHMI investigator, and work in his laboratory was supported by NIH grant MH08442 and the Starr Foundation.

References

1. McKee, A.E., Minet, E., Stern, C., Riahi, S., Stiles, C.D., and Silver, P.A. (2005) A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. *BMC Dev. Biol.*, 5, 14.
2. Lunde, B.M., Moore, C., and Varani, G. (2007) RNA-binding proteins: modular design for efficient

- function. *Nat. Rev. Mol. Cell. Biol.*, **8**, 479–490.
3. Keene, J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.*, **8**, 533–543.
4. Tenenbaum, S.A., Carson, C.C., Lager, P.J., and Keene, J.D. (2000) Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 14085–14090.
5. Gerber, A.P., Luschnig, S., Krasnow, M.A., Brown, P.O., and Herschlag, D. (2006) Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. *Proc. Nat. Acad. Sci. U.S.A.*, **103**, 4487–4492.
6. Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P.Y., Soll, S.J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M., and Tuschl, T. (2008) Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA*, **14**, 2580–2596.
7. Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science*, **302**, 1212–1215.
8. Licatalosi, D.D. and Darnell, R.B. (2010) RNA processing and its regulation: global insights into biological networks. *Nat. Rev. Genet.*, **11**, 75–87.
9. Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.
10. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M. Jr., Jungkamp, A.-C., Munschauer, M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
11. Favre, A., Moreno, G., Blondel, M.O., Kliber, J., Vincens, F., and Salet, C. (1986) 4-thiouridine photosensitized RNA–protein crosslinking in mammalian cells. *Biochem. Biophys. Res. Commun.*, **141**, 847–854.
12. Bezerra, R. and Favre, A. (1990) *In vivo* incorporation of the intrinsic photolabel 4-thiouridine into *Escherichia coli* RNAs. *Biochem. Biophys. Res. Commun.*, **166**, 29–37.
13. Andrus, A. and Kuimelis, R.G. (2001) Base composition analysis of nucleosides using HPLC. *Curr. Protoc. Nucleic Acid Chem.*, Chapter 10, Unit 10.6.

40

Global Analysis of Protein–RNA Interactions with Single-Nucleotide Resolution Using iCLIP

Julian König, Nicholas J. McGlincy, and Jernej Ule

40.1

Introduction

The majority of transcripts are subject to diverse regulatory activities that determine their cellular fate [1]. Throughout their lifetime, transcripts are associated with a plethora of RNA binding proteins (RBPs) forming diverse types of ribonucleoprotein particles (RNPs). The unique composition and spatial arrangement of these RBPs on a particular RNA dictate most aspects of the RNAs function [2]. Therefore, an essential step toward understanding transcript regulation at the molecular level is to gain positional information on the binding sites of RBPs within the RNA [3].

Protein–RNA interactions can be studied using biochemical methods such as SELEX, electrophoretic mobility shift, and RNA protection assays, or genetic methods such as the yeast three-hybrid system [4–6]. However, these approaches do not address RNA binding in its native cellular context. RNA immunoprecipitation combined with microarray analysis (RIP-CHIP) was developed to recover protein–RNA complexes from their cellular environment [7]. However, this method is prone to identifying indirect interactions and is limited to studying stable RNPs since protein–RNA complexes can reassociate after cell lysis [8]. Furthermore, RIP-CHIP does not allow direct identification of the protein's binding site within the RNA. To increase the resolution and specificity, a more sophisticated strategy referred to as *UV crosslinking and immunoprecipitation* (CLIP) was developed [9, 10]. CLIP combines UV crosslinking of proteins and RNA molecules with rigorous purification schemes including denaturing polyacrylamide gel electrophoresis. In combination with high-throughput sequencing (HITS) technologies, CLIP has proven as a powerful tool to study protein–RNA interactions on a genome-wide scale (referred to as *HITS-CLIP* or *CLIP-seq*) [11, 12]. So far, CLIP has mostly been used to study splicing regulators in mammalian systems and offered a deeper understanding of the positional rules that determine the outcome of alternative splicing [12–14]. In addition, CLIP has been successfully applied to study protein–micro-RNA interactions and subcellular RNA localization in organisms as diverse as *Caenorhabditis elegans* and the fungus *Ustilago maydis* [15–17]. Recent modifications of the protocol

include the use of photoreactive ribonucleoside analogs (PAR-CLIP) [18, 19] and affinity purification under denaturing conditions (CRAC) [20].

Despite the high specificity of the obtained data, CLIP experiments are difficult to perform and commonly generate cDNA libraries of limited complexity. This is partly due to the restricted amount of copurified RNA and the multitude of enzymatic and purification steps required for library preparation. In addition, primer extension assays indicated that many cDNAs prematurely truncate at the crosslinked nucleotide [21]. Such truncated cDNAs are lost during the standard CLIP library preparation protocol. We recently developed individual-nucleotide resolution crosslinking and immunoprecipitation (iCLIP) to overcome this limitation [22]. To capture the truncated cDNAs, we replaced one of the inefficient intermolecular RNA ligation steps with a more efficient intramolecular cDNA circularization. Importantly, sequencing the truncated cDNAs provides direct insight into the position of the crosslink site, allowing us to map protein–RNA interactions with single-nucleotide resolution. We have successfully applied iCLIP to study the impact of binding position on the regulation of alternative splicing by hnRNP C and TIA1/TIAL1 [22, 23]. In the following sections, we describe the iCLIP protocol and discuss the key features of this technology.

40.2

Procedure

40.2.1

Overview

To preserve *in vivo* protein–RNA interactions for subsequent biochemical purification steps, living cells or tissue are irradiated with UV-C light (Figure 40.1). This covalently crosslinks proteins to RNA molecules at positions of close contact and therefore the crosslink sites will represent the positions of direct protein–RNA interactions. Following cell lysis, the RNA is partially fragmented using low concentrations of RNase I. Next, the protein–RNA complex of interest is immunoprecipitated with a specific antibody immobilized on magnetic beads. After stringent washing, an RNA adapter is ligated to the RNAs 3′ ends while their 5′ ends are radioactively labeled. The complex is then subject to denaturing gel electrophoresis and transfer to a nitrocellulose membrane. This removes free RNAs that are not covalently linked to the protein. The radioactively labeled RNA allows the visualization of the purified protein–RNA complex using autoradiography, before the RNA is recovered from the membrane and the protein removed by proteinase K digestion. Since the covalent bond formed by UV crosslinking is irreversible, a short peptide remains covalently attached to the RNA at the crosslink site after proteinase K digestion. Consequently, most reverse transcription events will truncate at this position, thereby inheriting the information about the crosslinked nucleotide. To amplify these truncated cDNAs, reverse transcription is performed with oligonucleotide

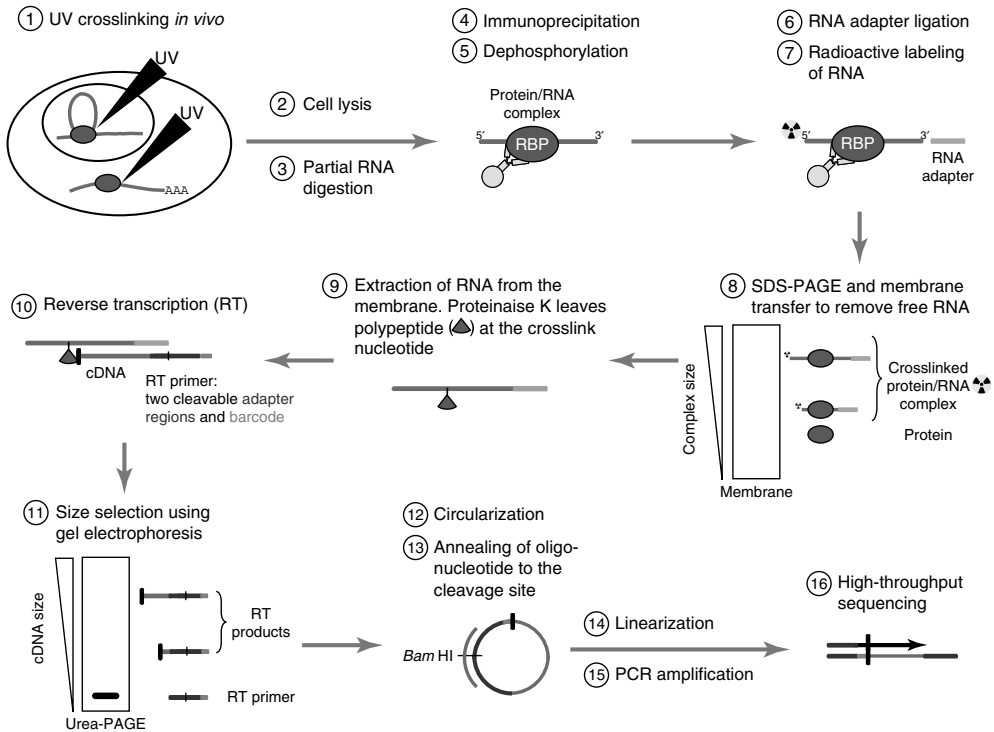


Figure 40.1 Schematic representation of the iCLIP protocol. Protein–RNA complexes are covalently crosslinked *in vivo* using UV irradiation (step 1). The protein of interest is purified together with the bound RNA (steps 2–5). To allow for sequence-specific priming of reverse transcription an RNA adapter is ligated to the 3' end of the RNA, whereas the 5' end is radioactively labeled (steps 6 and 7). Crosslinked protein/RNA complexes are purified from free RNA using SDS-PAGE and membrane transfer (step 8). The RNA is recovered from the membrane by digesting the protein with proteinase K leaving a polypeptide remaining at the crosslink

nucleotide (step 9). Reverse transcription (RT) truncates at the remaining polypeptide and introduces two cleavable adapter regions and barcode sequences (step 10). Size selection removes free RT primer before the RT products are circularized. The following relinearization generates suitable templates for PCR amplification (steps 11–15). Finally, high-throughput sequencing generates reads in which the barcode sequences are immediately followed by the last nucleotide of the cDNA (step 16). Since it is located 1 nt upstream of the crosslinked nucleotide, the binding site can be deduced with high resolution.

primers that contain two inversely oriented adapter regions separated by a *Bam* HI restriction site. To mark individual experiments and to control for PCR artifacts, the primers contain a barcode region at the 5' end consisting of 3 nt specific for each experiment and four random nucleotides individualizing single cDNA molecules. Next, cDNAs are size-purified using denaturing gel electrophoresis and circularized by single-stranded DNA ligase. The circularized cDNA is relinearized by *Bam* HI digestion between the two adapter regions. Since the circularized cDNA is single

stranded, this is achieved by annealing an oligonucleotide complementary to the restriction site. The linearized cDNAs are PCR amplified using primers complementary to the adapter regions and analyzed using polyacrylamide gel electrophoresis. Finally, samples are subjected to HITS on the Illumina Genome Analyzer.

40.2.2

Antibody and Library Preparation Quality Control

Before HITS, the success of an iCLIP experiment can be monitored at two steps: (i) the autoradiograph, which allows the study of the size distribution of the purified protein–RNA complexes and (ii) size distribution analysis of the final PCR products. It is of vital importance to use these two steps to control for the correct size of the protein–RNA complex, antibody specificity, and quality of the library preparation.

In the first step, the size distribution of the signal after partial RNase digestion (low RNase; used for library preparation) is compared to that after complete digestion of the crosslinked RNA (high RNase). In the low-RNase sample, the radioactive signal should be broad, extending from the size of the protein into higher-molecular-weight areas. In the high-RNase sample, a sharp radioactive band should appear directly above the expected size of the protein. A lack of this RNase-dependent size change might indicate that conditions for partial digestion are inappropriate, that cell extracts contain high endogenous RNase activity, or that the RBP itself has been labeled by the polynucleotide kinase reaction. The latter can be tested by omitting UV irradiation, since a radioactive signal originating from crosslinked RNA should be lost in this control.

To assess specificity of the immunoprecipitation, the low- and high-RNase samples should be compared to the respective no-antibody controls, which should give no signal in the autoradiograph. In addition, immunoprecipitations of the high-RNase samples should be carefully examined to detect signals from non-specific RBPs that in some cases co-immunoprecipitate with the protein of interest. Ideally, these contaminating signals should be avoided or minimized by optimizing the immunoprecipitation conditions or changing the antibody. Finally, specificity of the antibody itself should be controlled with samples from knockout or knockdown cells (or tissue). In the latter samples, a decrease in radioactive signal should correlate with the knockdown efficiency.

To control for the quality of library preparation, it is important to maintain one or more negative controls throughout the complete protocol. Optimal input material for these controls are the no-antibody samples or an immunoprecipitation from knockout cells. Control samples should show no product after PCR amplification and should result in very few unique sequences when submitted to HITS. Knockdown cells are not recommended as a sequencing control, since these still contain the protein, albeit in smaller quantities.

After PCR amplification, the length distribution of the products should reflect the size range of cDNAs that were purified from the polyacrylamide gel after reverse transcription. Note that the PCR primers introduce an additional 76 nt; therefore,

40.2.3 Oligonucleotide Design

To allow for sequence-specific priming of the reverse transcription reaction, the RNA oligonucleotide oRNA is ligated to the 3' end of the purified RNA. The sequence of oRNA is complementary to the 3' end of the PCR primer oP3, which

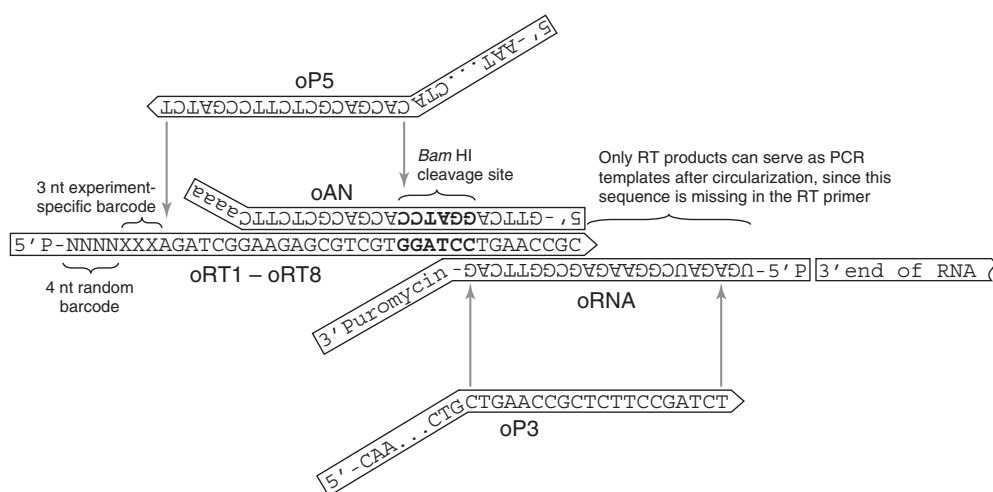


Figure 40.2 RNA and DNA oligonucleotide design. The RNA adapter (oRNA) is phosphorylated at the 5' end (5'P) to allow ligation to the crosslinked RNA. The 3' end is protected with puromycin to prevent concatenation. The reverse transcription primer (oRT1–oRT8) is complementary to the 3' half of oRNA to allow sequence-specific priming. The 5' end is phosphorylated to enable circularization and contains the barcode sequences. NNNN indicates the 4 nt random barcode, while XXX is the experiment-specific 3 nt barcode, which is

unique for each oRT oligonucleotide. oAN is complementary to the *Bam* HI cleavage site in oRT and contains four adenosines at its 3' end (aaaa) to prevent the oligo from acting as primer during the subsequent PCR amplification. The oligonucleotides oP3 and oP5 are used for PCR amplification. Complementary sections are delimited by gray arrows. It is important to note that oRT can only serve as a template for PCR amplification when acquiring sequence from oRNA during reverse transcription. This minimizes primer dimer formation during PCR.

is used to amplify the library for HITS. The 3' end of oRNA is blocked with a puromycin group to prevent oligomerization during the ligation reaction.

Reverse transcription is primed with either of the oligonucleotides oRT1–oRT8 (abbreviated as oRTN), which are complementary to the last 9 nt at the 3' end of oRNA. Since annealing of primer oP3 during the subsequent PCR reaction requires the full sequence introduced by oRNA, this incomplete overlap ensures that oRTNs cannot be amplified unless oRNA served as a template during reverse transcription. This is vital since excess of circularized oRTN oligonucleotides would otherwise generate primer dimers during the PCR amplification. The oRTNs also contain an adapter region complementary to the 3' end of the second PCR primer oP5. The regions of oRTN corresponding to oP3 and oP5 are separated by a *Bam* HI restriction site that allows later linearization of the circularized cDNAs. At the 5' end, oRTNs contain two distinct barcode regions. The 3-nt experiment-specific barcode is unique to each of the oRTNs and marks individual experiments or replicates. The 4-nt random barcode is unique to each oRTN molecule and allows distinguishing products of different reverse transcription events from mere PCR duplicates. To allow for ligation during the circularization reaction, all oRTNs are 5'-phosphorylated.

Before PCR amplification, the circularized cDNA molecules require linearization by cutting between the two adapter regions. To guide this enzymatic reaction, the DNA oligonucleotide oAN is annealed to the cDNAs covering the *Bam* HI site and flanking regions. To prevent oAN from serving as a primer during the subsequent PCR amplification, it contains four non-complementary adenosines at its 3' end.

PCR amplification is performed with the oligonucleotides oP3 (61 nt) and oP5 (58 nt). Their 3' ends are complementary to the two adapter regions introduced by oRNA and oRTNs. oP3 and oP5 contain additional 41 and 39 nt, respectively, required for HITS with the Illumina Genome Analyzer.

40.2.4

Troubleshooting

Since the iCLIP protocol is very long and contains a diverse range of enzymatic reactions and purification steps, it is not always easy to identify a problem when an experiment fails. Therefore, this section contains a few general suggestions, while more specific comments are given as short notes throughout the protocol. Each step has to be performed with high accuracy to obtain proper results. To ensure high quality of the material, buffers, primer, or enzyme should be aliquoted so that each member of the laboratory working with iCLIP has their own set. In this way, problems with the material will not affect all experiments and can be identified more quickly. Precautions should also be taken to avoid contamination with PCR products from previous experiments. The best way to minimize this problem is to spatially separate pre- and post-PCR steps. Ideally, the analysis of the PCR products and all subsequent steps should be performed in a separate room. Finally, even though tricky at the beginning, iCLIP can be a lot of fun.

Protocol

UV CrossLinking (Figure 40.1, step 1)

Tissues

Note 1

We usually start with 20 mg of tissue, but this can be adjusted to the efficiency of protein–RNA crosslinking and immunoprecipitation.

- 1) Harvest tissue. Add 10 volumes of ice-cold PBS.
- 2) Here is the procedure we use to partially triturate brain tissue. Other tissue may require different dissociation protocols.
Sequentially pass the tissue several times through the following:
 - a. 10 ml pipette.
 - b. 10 ml pipette with cut p1000 tip on end (cut off a bit from the tip with a blade).
 - c. 10 ml pipette with uncut p1000 tip on end.
 - d. 10 ml pipette with p10 tip on end.

Note 2

UV light can penetrate a few cell layers, and so triturating to a single-cell suspension is unnecessary.

- 3) UV crosslinking.
 - a. Use 8 ml per 10 cm tissue culture plate.
 - b. Place tray on ice.
 - c. Irradiate suspension 4 x with 100 mJ cm^{-2} in Strat linker 2400 (Stratagene), mixing between each irradiation.

Note 3

The length of crosslinking should be optimized for each protein examined, as each RNA-binding domain crosslinks with different efficiency depending on its content of aromatic amino acids. Try 100, 200, and 400 mJ cm^{-2} , and then use the mildest condition that gives >70% of the maximum signal. This is important to avoid secondary crosslinking to other protein, which would blur the resolution of the technique.

- 4) Add 2 ml suspension to each microtube, spin at top speed for 10 s at 4°C to pellet cells, and then remove supernatant.
- 5) Snap-freeze pellets on dry ice and store at -80°C until use.

Tissue Culture Cells

- 1) Add 6 ml ice-cold PBS to cells growing in a 10-cm plate. Remove lid and place on ice. If cells are in suspension, follow the same protocol as for dissociated tissue.

- 2) Irradiate once with 150 mJ cm^{-2} .

Note 4

Cells grown in a monolayer are equally exposed to the UV light and hence only require a single irradiation to crosslink equally.

- 3) Remove PBS and harvest cells by scraping into a further 6 ml ice-cold PBS. Use cell lifters from Fisher.
- 4) Add 2 ml suspension to each microtube, spin at top speed for 10 s at 4°C to pellet cells, and then remove supernatant.
- 5) Snap freeze pellets on dry ice and store at -80°C until use.

Immunoprecipitation (Figure 40.1, steps 2–4)

Solutions

- Lysis Buffer
 - 50 mM Tris-HCl, pH 7.4
 - 100 mM NaCl
 - 1% NP-40
 - 0.1% SDS
 - 0.5% Sodium deoxycholate.

On the day of experiment, add 1/100 volume of protease inhibitor cocktail (Calbiochem) to the amount of buffer required for lysis (but not washing).

Note 5

Store all buffers in the fridge and perform the procedure on ice. If you are working with a tissue with high RNase A activity, adding 1/1000 volume of ANTI-RNase (AM2692, Ambion) will control the RNase conditions, without affecting the activity of RNase I.

- High-salt Wash
 - 50 mM Tris-HCl, pH 7.4
 - 1 M NaCl
 - 1 mM EDTA
 - 1% NP-40
 - 0.1% SDS
 - 0.5% Sodium deoxycholate.
- PNK Buffer
 - 20 mM Tris-HCl, pH 7.4
 - 10 mM MgCl_2
 - 0.2% Tween-20.

Bead Preparation

Note 6

The amounts given below are meant for the library preparation. For preliminary experiments, or the high-RNase control, less can be used.

- 1) Add 100 μ l of protein A Dynabeads (Dyna, 100.02) per experiment to a microtube.

Note 7

Use protein G Dynabeads for a mouse or goat antibody. These can sometimes work better for rabbit antibodies, too.

- 2) Wash beads 2 x with lysis buffer.
- 3) Resuspend beads in 200 μ l lysis buffer with 2–10 μ g antibody per experiment.

Note 8

The amount of antibody required depends on its quality and purity. This should be optimized in preliminary experiments.

- 4) Rotate tubes at RT for 30–60 min.
- 5) Wash 3 x with lysis buffer. Leave the beads in the last wash until the lysate is ready.

Cell Lysis, Partial RNA Digestion, and Centrifugation (Figure 40.1, steps 2 and 3)

- 1) Resuspend pellet in 1 ml lysis buffer (with added protease inhibitors).
- 2) Sonicate sample on ice. The sonication probe should be approximately 0.5 cm from the bottom of the tube and not touching the tube sides to avoid foaming. Sonicate 2 x with 10 s bursts at 5 dB. Sonicate with water before and after to clean the probe.

Note 9

We are aiming for a protein concentration of 10 mg ml^{-1} in the raw extract. Mouse brain pellets have approximately 50 mg, and cell culture pellets approximately 20 mg. Weighing pellets before freezing can help one to be more precise in the volume of lysis buffer used.

Note 10

Sonication helps when using cell culture as undigested viscous DNA can sometimes cause problems with the immunoprecipitation (IP). It can also alleviate problems caused by mild lysis buffers or hard-to-lyse tissues.

Note 11

Optionally, one can preclear with protein A Sepharose (this does not hurt, but usually makes little difference). Prepare 30% protein A Sepharose slurry in water. Add 100 μ l protein A Sepharose slurry to 1.5 ml lysate and rotate for 10 min in the cold room before spinning.

- 3) Spin 15 000 rpm for 10 min at 4 °C and collect supernatant.
- 4) Make RNase I (Ambion, AM2295) dilutions.
 - a. For low-RNase condition: 1/1000; 1 ml lysis buffer + 1 μ l RNase.
 - b. For high-RNase condition: 1/50; 49 μ l lysis buffer + 1 μ l RNase.

Note 12

High RNase is required for the initial experiments to optimize conditions. The size of the radioactive band viewed by SDS-PAGE should change in response to different RNase concentrations. This confirms that the signal corresponds to a protein–RNA complex. Furthermore, this experiment helps to determine the size of the immunoprecipitated RBP, as the protein will be bound to short RNAs and thus will migrate as a less diffuse band ~5 kDa above the expected molecular weight. Other recommended controls include a control where the RBP is absent from the original material (such as a knockout animal or knockdown cells) or if that is not possible, a control where the protein is overexpressed, a control where no crosslinking is done, and a control where no antibody is used during IP.

Note 13

Unlike other RNases, RNase I has no base preference, and therefore cleaves after all 4 nt. Only the low RNase condition is used for library preparation. The low RNase dilution depends on the batch of RNase, the concentration of protein, and the lysate tissue/cell type used, and so in the first experiment, several dilutions should be tested to find a condition that creates an approximate 10–50 kDa increase in the apparent size of the complex on the gel. This may not apply to RBPs that bind short non-coding RNAs (i.e., Argonaute proteins).

- 5) To each 1 ml of lysate add 10 μ l of RNase dilution and 5 μ l of Turbo DNase (Ambion, AM2238; 2 U μ l⁻¹).

Note 14

Unlike other DNases, Turbo DNase is active in conditions of up to 200 mM NaCl.

- 6) Incubate for 3 min at 37 °C, shaking at 1100 rpm. Transfer to ice for >3 min.

Note 15

It is important to digest for exactly 3 min. Shaking is important to equilibrate the warm-up process.

Note 16

Optional: To test a new antibody, collect 15 μ l at this step for analysis of protein depletion after IP using Western blotting (Note 17).

Immunoprecipitation (Figure 40.1, step 4)

- 1) Remove wash buffer from the beads, then add the extract to the bead aliquots from the section “Bead Preparation” (5).
- 2) Rotate beads/lysate mix for 1 h or overnight at 4 °C.

Note 17

Optional: Collect 15 µl of the supernatant at this step for analysis of protein depletion after IP using Western blotting (compare to samples from Note 16).

- 3) Discard the supernatant, wash 2 x with high-salt wash (rotate the second wash >1 min in the cold room).
- 4) Wash 2 x with PNK buffer and then resuspend in 1 ml PNK buffer.
- 5) Split low-RNase sample – use 900 µl for the section “3’ RNA Dephosphorylation” and “RNA Adapter Ligation” and 100 µl for the section “5’ Labeling.”

3’ RNA Dephosphorylation (Figure 40.1, step 5)

- 1) Discard supernatant. Resuspend beads in 20 µl of:

16 µl	Water
2 µl	10 x Shrimp alkaline phosphatase buffer
2 µl	Shrimp alkaline phosphatase (Promega, M820A)

- 2) Incubate for 20 min at 37 °C.
- 3) Wash 2 x with high-salt wash (rotate second wash >1 min in the cold room).

Note 18

Stringent washing is necessary to remove all phosphatase before proceeding to the ligation step.

- 4) Wash 2 x with PNK buffer (rotate second wash >1 min in the cold room).

RNA Adapter Ligation (Figure 40.1, step 6)

- 1) Remove supernatant carefully and resuspend the beads in 14 µl of the following mix:

6.8 µl	Water
1.5 µl	10 x RNA ligase buffer (NEB)
5 µl	PEG400 (81170, Sigma)
0.5 µl	RNA ligase (NEB)
0.2 µl	RNasin (NEB)

- 2) Add 6 µl of the oRNA adapter (20 µM).
- 3) Transfer into a 0.2 ml PCR tube and incubate overnight in the PCR block at 16 °C.
- 4) Add 100 µl of PNK buffer and continue to the section “5’ Labeling” (step 4).

5’ Labeling (Figure 40.1, step 7)

- 1) Collect 100 µl of beads (see the section “Immunoprecipitation” (step 5)).

- 2) Remove supernatant and add 4 μ l of hot PNK mix:

0.2 μ l	PNK (NEB)
0.4 μ l	[γ - ^{32} P]ATP
0.4 μ l	10 x PNK buffer
3 μ l	Water

- 3) Incubate in Thermomixer R (Eppendorf) for 5 min at 37 °C.
- 4) Combine with the beads from the section “RNA Adapter Ligation” (step 4) and transfer to 1.5 ml tubes.
- 5) Remove supernatant using a magnet ($\Rightarrow$ solid radioactive waste). Add 20 μ l 1 x NuPAGE loading buffer (Invitrogen, prepared by mixing 4 x stock with water) to the beads.
- 6) Incubate on thermomixer at 70 °C for 5 min.
- 7) Place on magnet and load the supernatant on the gel.

SDS-PAGE and Nitrocellulose Transfer (Figure 40.1, step 8)

- 1) Prepare gel: Wash gel with milliQ water, remove white plastic strip, and place into apparatus with the comb on the inner side. Once under buffer, remove the comb and wash the wells with a p1000 pipette set to 500 μ l.
- 2) Load sample into a 4–12% (usually better) or 10% NuPAGE Bis-Tris gel (Novex, Invitrogen). Use 0.5l of 1 x MOPS running buffer (20 x stock). Load 4 μ l of marker (polyacrylamide gel electrophoresis (PAGE) ruler plus, prestained protein ladder, Fermentas #SM1811).
- 3) Run 50 min at 180 V.

Note 19

The Novex NuPAGE gels are critical. A pour-your-own SDS-PAGE gel (Laemmli) has a pH during the run that can get to $\sim$ 9.5 and can lead to alkaline hydrolysis of the RNA. The Novex NuPAGE buffer system is close to pH 7. We use MOPS NuPAGE running buffer.

- 4) After the gel has run, remove the gel front and discard as solid waste (contains free radioactive ATP).
- 5) Transfer gel to Protan BA85 Nitrocellulose Membrane (Whatman) using the Novex wet transfer apparatus (1 h at 30 V):

Thoroughly soak all sponges/Whatman paper/nitrocellulose in transfer buffer (make 500 ml, remember to add methanol to 10%) before assembly.

Assembly: two to three sponges, use Whatman paper to remove gel from plastic and add on top of sponges with gel side up, add nitrocellulose carefully to avoid air bubbles, add second bit of Whatman paper, roll with tube to flatten, and add two to three more sponges.

Note 20

This pure nitrocellulose is a little fragile, but it works better for the RNA extraction step than PVDF membranes.

- 6) After transfer, rinse the nitrocellulose membrane in 1 x PBS and remove excess with kimwipes. Wrap filter in saranwrap and expose to X-ray film in cassette at -80°C (perform 30 min, 1 h, and overnight exposures).

Note 21

Most free RNA will have passed out of the gel or through the membrane, and so because of the loss of the free RNA, the membrane will be 10–100 times less radioactive than the sample loaded on the gel.

Note 22

Use a luminescent sticker inside the cassette so that one can later align the filter back to the autoradiograph.

RNA Isolation (Figure 40.1, step 9)

- 1) Use the high-RNase condition to examine the specificity of the RNA–protein complexes.

Note 23

If CLIP is being performed for the first time, use the following criteria to check that a specific RNA–protein crosslink and pulldown have been performed.

1. Is there a radioactive band ~ 5 kDa above the molecular weight of the protein in the high-RNase experiment?
2. Does the band disappear in the control experiments? These might include the following: no UV crosslink, pulldown with no antibody (beads only or preimmune serum), samples from a knockout organism or knockdown cells, or an appropriate control for overexpressed tagged proteins.
3. Does the band shift up and become more diffuse in the low-RNase condition? Because the RNA digestion is random, the RNA sizes vary more in the low-RNase condition and thus the RNA–protein complexes are more heterogeneous in size and the band appears more diffuse.

On this basis, if one is convinced of the veracity of results, proceed to RNA isolation and amplification. Note the following guidelines.

1. About 70 nt of RNA adds ~ 20 kDa to the apparent migration of the protein–RNA complex on an SDS-PAGE gel. As the tags contain an RNA adapter of 22 nt (oRNA), the ideal position of RNA–protein complexes that will generate CLIP tags of sufficient length is ~ 10 – 60 kDa above the expected molecular weight of the protein.
2. The width of the excised band depends on any other RNA–protein complexes present in the vicinity (this is determined by the high RNase experiment). If

none are apparent, then cut a wide band of approximately 10–60 kDa above the molecular weight of the protein. However, if other contaminant bands are present above the size of the protein, cut only up to the size of those bands. If the contaminating bands are below the RNA–protein complex, one could consider cutting an additional band of 10 kDa below to 5 kDa above the molecular weight of the protein – the RNA will come from the contaminating protein, and can later be used to identify the contaminating sequences in the set cut at ~10–60 kDa above the molecular weight of the protein.

- 2) For the low-RNase experiments, cut according to Note 23 using the autoradiograph as a mask and place the membrane piece into a 1.5 ml tube. Crush and fold the membrane piece with a p1000 tip.
- 3) Prepare the following buffers:
 - a. PK buffer
 - i. 100 mM Tris-HCl, pH 7.4
 - ii. 50 mM NaCl
 - iii. 10 mM EDTA.
 - b. PK buffer + 7 M urea
 - i. 100 mM Tris-HCl, pH 7.4
 - ii. 50 mM NaCl
 - iii. 10 mM EDTA
 - iv. 7 M urea.
- 4) Add 10 µl proteinase K (Roche, 03115828001) in 200 µl PK buffer to the nitrocellulose pieces (all should be submerged). Incubate shaking at 1000 rpm for 20 min at 55 °C.
- 5) Add 200 µl of PK buffer + 7 M urea and incubate for a further 20 min at 55 °C.
- 6) Add the solution (leaving behind the membrane) and 400 µl of RNA phenol/CHCl₃ (Ambion, 9722) to a 2 ml Phase Lock Gel Heavy tube (713-2536, VWR).
- 7) Incubate for 5 min at 30 °C, shaking at 1000 rpm (do not vortex). Separate phases by spinning for 10 min at 13 000 rpm at RT.

Note 24

RNA phenol can also be prepared by equilibrating pure phenol with 0.15 M NaOAc, pH 5.2; CHCl₃ is chloroform 49 : 1 with isoamyl alcohol.

- 8) Transfer the aqueous layer into a new tube (be careful not to touch the gel with the pipette). Precipitate by the addition of 0.5 µl glycoblue (Ambion, 9510), 40 µl 3 M NaOAc pH 5.5, then mix and add 1 ml 100% EtOH, mix again, and place at –20 °C overnight or for 1 h.

Note 25

Glycogen is necessary to efficiently precipitate the small quantity of RNA and to visualize the pellet.

- 9) Spin for 30 min at 15 000 rpm at 4 °C. Remove supernatant and wash pellet with 0.5 ml 80% EtOH (spin 5 min). Resuspend pellet in 6.25 µl water.

Note 26

Remove the wash first with a p1000 pipette and then with a p200 pipette or smaller. Try not to disturb the pellet, but if one does, spin it down again. Leave on the bench for 3 min, but no longer, with cap open to dry. When resuspending, make sure to pipette along the back area of the tube.

Reverse Transcription (Figure 40.1, step 10)

- 1) Prepare RNA/primer mix

6.25 µl	RNA
0.5 µl	One of the oRTN primers (0.5 pmol µl ⁻¹)
0.5 µl	dNTP mix (10 mM)

- 2) RT thermal program

70 °C	5 min
25 °C	Hold until RT mix added, mix by pipetting.
25 °C	5 min
42 °C	20 min
50 °C	40 min
4 °C	Hold

- a. RT mix

2 µl	5 x RT buffer
0.5 µl	0.1 M DTT
0.25 µl	Superscript III RT (Invitrogen, 200 U µl ⁻¹)

- 3) Mix samples to be multiplexed at this point.

Note 27

All samples with different barcodes can be multiplexed at this point. However, it can be also an advantage to keep certain samples separate to allow for visual inspection of the individual PCR products before deep sequencing. For example, if control experiments are not be multiplexed with the rest, they should not generate any PCR products on their own (alternatively, if they are

multiplexed, then sequencing will show that the corresponding barcodes do not generate many sequences).

- 4) Fill up with TE buffer to 100 μ l, add 0.5 μ l glycoblu and mix. Add 10 μ l NaOAc pH 5.5, mix, and then add 250 μ l 100% EtOH. Mix again and precipitate overnight or 1 h at -20°C .

Gel Purification (Figure 40.1, step 11)

- 1) Spin down and wash as before, and resuspend in 6 μ l of water.
- 2) Add 6 μ l 2 x TBE–urea loading buffer (Invitrogen) to the 6 μ l DNA. It is recommended, at least in initial experiments, to add loading buffer to 5 μ l marker in the last lane (NEB low molecular weight marker, N3233S) and RT primer in the next lane. Heat samples at 70°C for 3 min directly before loading. Leave one lane between each sample to facilitate cutting.
- 3) Prepare 0.8 l of 1 x TBE running buffer and fill the upper chamber with 0.2 l and the lower chamber with 0.6 l. Use p1000 pipette to flush wells before loading 12 μ l of each sample.
- 4) Run 6% TBE–urea gel for 40 min at 180 V until the lower (dark blue) dye is close to the bottom.
- 5) Open gel and cut two portions for each condition: high (H, 130–300 nt), upward from the light blue dye front, and low (L, 75–130 nt), downward from the light blue dye front. Transfer gel pieces to new 1.5 ml tubes.
- 6) Cut off the last lane containing the DNA size marker and stain the gel. Incubate for 10 min shaking in 50 ml TBE buffer with 5 μ l SYBR green II stock (Invitrogen, S-7586). Wash 1 x with TBE and visualize by UV transillumination.

Note 28

Record the sizes for the high and low cut using the gel picture from the section “Gel Purification” (step 6). Avoid purifying the RT primer and RT products smaller than 75 nt. The oRTN primers will migrate at 39 nt. Together with the full oRNA sequence, the primer sequence accounts for 52 nt of the CLIP sequence. The upper (lighter blue) dye runs at ± 130 nt and can be used to guide excision: cut the lower band from ± 1 cm below to the center of the dye (about 75–130 nt), and the top band from the center of the dye up to ± 2 cm above the dye (about 130–300 nt).

- 7) Add 400 μ l of TE and crush the gel piece into small pieces with a 1 ml syringe plunger. Incubate shaking at 1100 rpm for 2 h at 37°C .
- 8) Transfer liquid portion of the supernatant into a Costar SpinX column (Corning Incorporated, 8161) into which one has placed two 1 cm glass prefilters (Whatman 1823010).
- 9) Spin at 13 000 rpm for 1 min into a 1.5 ml tube. Add 0.5 μ l glycoblu and 40 μ l NaOAc pH 5.5, mix add 1 ml 100% EtOH. Mix again and precipitate overnight or for 1 h at -20°C .

Ligation of Primer to the 5' End of the cDNA by Circularization (Figure 40.1, steps 12–14)

- 1) Spin down and wash as previously, resuspend cDNA in 6.5 μ l of water.
- 2) Add 1.5 μ l of ligation mix and incubate for 1 h at 60 °C.

Ligation mix	
0.8 μ l	10 x CircLigase Buffer II
0.4 μ l	50 mM MnCl ₂
0.3 μ l	CircLigase II (Epicentre)

- 3) Add 30 μ l of Oligo annealing mix and anneal oligonucleotides with the following program:

95 °C	2 min
70 cycles of 1 min starting with 95 °C, then decreasing temperature with each cycle by 1 °C	
25 °C	Hold

Oligo annealing mix	
25 μ l	Water
4 μ l	FastDigest Buffer (Fermentas)
1 μ l	10 μ M oligonucleotide oAN

- 4) Add 2 μ l *Bam* HI (Fast Fermentas), and incubate for 30 min at 37 °C.
- 5) Add 50 μ l TE and 0.5 μ l glycobblue and mix. Add 10 μ l NaOAc pH 5.5, mix then add 250 μ l 100% EtOH. Mix again and precipitate overnight or for 1 h at –20 °C.

PCR Amplification (Figure 40.1, steps 15 and 16)

Spin down and wash DNA as described previously, resuspend in 11 μ l of water.

Optimize PCR Amplification

- 1) Set up PCR

0.5 μ l	CircDNA
0.5 μ l	Primer mix oP5/oP3, 10 μ M each
10 μ l	Immomix (Bioline)
9 μ l	Water

- 2) Run following PCR with 25 and 35 cycles: 95 °C 10 min (95 °C 10 s, 65 °C 10 s, 72 °C 20 s), 25–35, 72 °C 3 min, 4 °C ∞ .

Note 29

All work done post-PCR must be carried out on a specially designated bench. This DNA must never be taken to an area where work with CLIP RNA is done.

- 3) Desalt the samples using MicroSpin G-25 columns (Illustra, GE Healthcare).

Note 30

Desalting is necessary since the Immomix buffer interferes with TBE gel electrophoresis.

- 4) Combine 5 µl PCR with 5 µl 2 x TBE–urea loading buffer. Run on 6% TBE–urea gel and stain with SYBR green II as described before.

Preparative PCR

- 1) From the results in the section “Optimize PCR Amplification”, estimate the minimum number of PCR cycles to use to amplify the whole library, such that it will give a band on a gel. Consider that one will now be amplifying 20 times more material than they did in the preliminary PCR.
- 2) Set up PCR

9.5 µl	CircDNA
0.5 µl	Primer mix oP5/oP3, 10 µM each
10 µl	Immomix (Bioline)

- 3) Run the following PCR:
95 °C 10 min (95 °C 10 s, 65 °C 10 s, 72 °C 20 s), 25–35, 72 °C 3 min, 4 °C ∞.
- 4) Desalt the samples using MicroSpin G-25 columns (Illustra, GE Healthcare).
- 5) Examine 5 µl of PCR on TBE–urea gel as described before. The size of the cDNA insert to be mapped to the genome will be the size of product minus the combined length of oP3/oP5 and barcodes (128 nt total). Ideally, PCR products should be larger than 150 nt.
- 6) Subject the samples to HITS with the Illumina Genome Analyzer.

Linker and Primer Sequences

RNA	
oRNA	P-UGAGAUCCGAAGAGCGGUUCAG-Puromycin

Note 31

We order RNAs from Dharmacon. After receiving the RNA, deprotect it, dissolve to 200 µM, and then make aliquots of 20 µM. We do not gel-purify the linkers.

DNA

oRT1	P-NNNNAACAGATCGGAAGAGCGTCGTGgataCTGAACCGC
oRT2	P-NNNNGGCAGATCGGAAGAGCGTCGTGgataCTGAACCGC
oRT3	P-NNNNGATAGATCGGAAGAGCGTCGTGgataCTGAACCGC
oRT4	P-NNNNAGTAGATCGGAAGAGCGTCGTGgataCTGAACCGC
oRT5	P-NNNNCCCAGATCGGAAGAGCGTCGTGgataCTGAACCGC
oRT6	P-NNNNTTCAGATCGGAAGAGCGTCGTGgataCTGAACCGC
oRT7	P-NNNNTCTAGATCGGAAGAGCGTCGTGgataCTGAACCGC
oRT8	P-NNNNCTTAGATCGGAAGAGCGTCGTGgataCTGAACCGC

(P indicates a 5' phosphate and small letters indicate nucleotides added between the two adapter regions to generate a *Bam* HI restriction site.)

oAN	GTTCAGGATCCACGACGCTCTTCaaa
oP5	AATGATACGGCGACACCGAGATCTACACTCT TTCCTACACGACGCTCTTCGATCT
oP3	CAAGCAGAAGACGGCATAACGATCGGTCTCG GATTCTGCTGAACCGCTCTTCGATCT

Note 32

We order DNAs from Sigma Aldrich. There is no need to gel-purify the primers.

Acknowledgments

The authors thank all members of the Ule laboratory and Dr Kathi Zarnack for advice and discussion. This work was supported by the European Research Council grant 206726-CLIP to J.U. and a Long-term Human Frontiers Science Program fellowship to J.K.

References

- Moore, M.J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science*, **309**, 1514–1518.
- Keene, J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.*, **8**, 533–543.
- Wang, Z. and Burge, C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, **14**, 802–813.
- Denman, R.B. (2006) mRNPs take shape by CLIPPING and PAIRING. *Bioessays*, **28**, 1132–1143.
- Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
- Sengupta, D.J., Zhang, B., Kraemer, B., Pochart, P., Fields, S., and Wickens, M. (1996) A three-hybrid system to detect RNA-protein interactions *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 8496–8501.
- Tenenbaum, S.A., Carson, C.C., Lager, P.J., and Keene, J.D. (2000) Identifying mRNA subsets in messenger

- ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 14085–14090.
8. Mili, S. and Steitz, J.A. (2004) Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA*, **10**, 1692–1694.
 9. Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science*, **302**, 1212–1215.
 10. Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005) CLIP: A method for identifying protein–RNA interaction sites in living cells. *Methods*, **37**, 376–386.
 11. Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X. *et al.* (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, **456**, 464–469.
 12. Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA–protein interactions in stem cells. *Nat. Struct. Mol. Biol.*, **16**, 130–137.
 13. Ule, J. and Darnell, R.B. (2006) RNA binding proteins and the regulation of neuronal synaptic plasticity. *Curr. Opin. Neurobiol.*, **16**, 102–110.
 14. Xue, Y., Zhou, Y., Wu, T., Zhu, T., Ji, X., Kwon, Y.S., Zhang, C., Yeo, G., Black, D.L., Sun, H. *et al.* (2009) Genome-wide analysis of PTB–RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol. Cell*, **36**, 996–1006.
 15. Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009) Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature*, **460**, 479–486.
 16. Zisoulis, D.G., Lovci, M.T., Wilbert, M.L., Hutt, K.R., Liang, T.Y., Pasquinelli, A.E., and Yeo, G.W. (2010) Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nat. Struct. Mol. Biol.*, **17**, 173–179.
 17. König, J., Baumann, S., Koepke, J., Pohlmann, T., Zarnack, K., Feldbrügge, M. (2009) The fungal RNA-binding protein Rrm4 mediates long-distance transport of *ubi1* and *rho3* mRNAs. *EMBO J.*, **28**, 1855–1866.
 18. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M. Jr., Jungkamp, A.C., Munschauer, M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
 19. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.C., Munschauer, M. *et al.* (2010) PAR-CLIP—a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J. Vis. Exp.*, **41**. doi: 10.3791/2034
 20. Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV crosslinking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 9613–9618.
 21. Urlaub, H., Hartmuth, K., and Lührmann, R. (2002) A two-tracked approach to analyze RNA–protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods*, **26**, 170–181.
 22. König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.*, **17**, 909–915.
 23. Wang, Z., Kayikci, M., Briesse, M., Zarnack, K., Luscombe, N.M., Rot, G., Zupan, B., Curk, T., and Ule, J. (2010) iCLIP predicts the dual splicing effects of TIA–RNA interactions. *PLoS Biol.*, **8**. doi: 10.1371/journal.pbio.1000530

Part IV

RNA Function, RNP Analysis, SELEX, RNAi

41

Use of RNA Affinity Matrices for the Isolation of RNA Binding Proteins

Markus Englert, Bettina Späth, Steffen Schiffer, Sylvia Rösch,
Hildburg Beier, and Anita Marchfelder

41.1

Introduction

All living organisms use a variety of RNA–protein (RNP) interactions either in stable RNP complexes (e.g., the ribosome or the spliceosome) or in transient RNP interactions. Transcripts have to be capped, modified, edited, spliced, processed, transported, and finally degraded. Thus, RNA binding proteins play an essential role in key processes in the cell. To understand their function and to investigate them in detail, it is often required to purify and isolate these proteins. For the isolation of RNases, several different approaches can be employed. Purifications generally start with the isolation of a soluble protein extract, which is initially fractionated using precipitation and ion-exchange chromatography. Subsequently, group-specific affinity matrices (e.g., Cibacron Blue or heparin) are applied, which result in good purification and yield. The final and most efficient purification step can be achieved for RNA binding proteins using ligand-specific affinity matrices. Since affinity chromatography exploits the highly specific interaction between molecules, it is highly selective and offers high yield and purity. Sequence-specific RNAs as well as tRNAs and rRNAs have been successfully used to purify RNA binding proteins [1–16]. Taken together, these examples show that a ligand-specific chromatographic step is often feasible, which dramatically facilitates the purification scheme, when a new RNA-binding activity is investigated from extracts [8].

Since specific affinity columns are usually difficult to prepare, they are smaller and do not have such a high capacity as ion-exchange columns or group-specific affinity columns, in addition they are often difficult to prepare. In contrast to DNA affinity columns [17], RNA affinity columns are always sensitive to degradation especially from RNases and are therefore less stable. Thus, only extracts that already have been purified through several step and are highly enriched for the desired protein should be loaded onto such columns. The preceding purification step must be sufficient to remove ribonucleases that can degrade the RNA target attached to the column. Since specific RNA affinity matrices usually do not have high capacities, the prior purification step should in addition remove the proteins

binding unspecifically to the RNA matrix and thus blocking binding sites for the protein being purified.

The *N*-hydroxysuccinimide NHS-activated Sepharose was originally developed for the coupling of proteins *via* freely accessible primary amino groups. Since in highly structured RNA molecules, such as transfer RNAs, only a few amino groups are accessible, the coupling efficiency is low. Since the coupled tRNAs are presumably attached by only one covalent linkage, the free accessibility of the ligand to the RNA binding protein is ensured.

Recently, new methods to prepare RNA columns with higher capacity and better stability have been developed, which might circumvent some of the above-mentioned problems [18]. Updated methods for RNA affinity columns are published frequently, describing the development of specific RNA affinity tags or aptamers with high affinity [5, 16, 19–22].

Here, the application of an RNA affinity column to purify two tRNA-processing enzymes is described. tRNA molecules are transcribed as precursors containing 5' and 3' additional sequences and in some cases as introns that have to be removed to yield functional tRNAs [2, 23] (Figures 41.1 and 41.2).

Accurate tRNA 3'-end maturation is essential for aminoacylation and therefore for protein synthesis in all organisms (Figure 41.1). In archaea and in the majority of eukaryotes, the enzyme catalyzing the removal of the 3' trailer sequence is the endonuclease tRNase Z [24–27]. Only recently, it was shown that in *Bacillus subtilis* a tRNase Z homolog is responsible for the 3'-end maturation of a certain set of tRNAs [28]. Knockdown of the tRNase Z homolog in *Drosophila melanogaster* using RNAi resulted in accumulation of nuclear and mitochondrial pre-tRNAs containing 3' trailers [29].

Some nuclear-encoded tRNA genes are interrupted by intervening sequences at a highly conserved position 3' to the anticodon. The tRNA-splicing mechanism is not related to the spliceosome-mediated mRNA splicing and to the self-splicing of group I or II introns. Instead, pre-tRNA introns are removed by the stepwise reaction of the three enzymes (Figure 41.2).

To purify the tRNA ligase as well as the tRNase Z from wheat germ, we used a tRNA affinity column. Although the tRNA used for the column is not a substrate but the product of the reaction catalyzed by tRNase Z, this enzyme shows a high

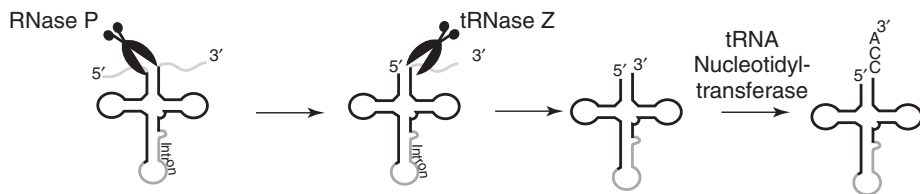


Figure 41.1 Maturation of precursor tRNAs I: RNase P, tRNase Z, and nucleotidyl transferase. tRNAs are transcribed as precursor molecules containing additional sequences at the 5' end, the 3' end, and in

some cases introns. The mature 5' end is generated by the enzyme RNase P, whereas the 3' trailer sequence is cleaved off by tRNase Z. Subsequently, the CCA triplet is added by the tRNA nucleotidyl transferase.

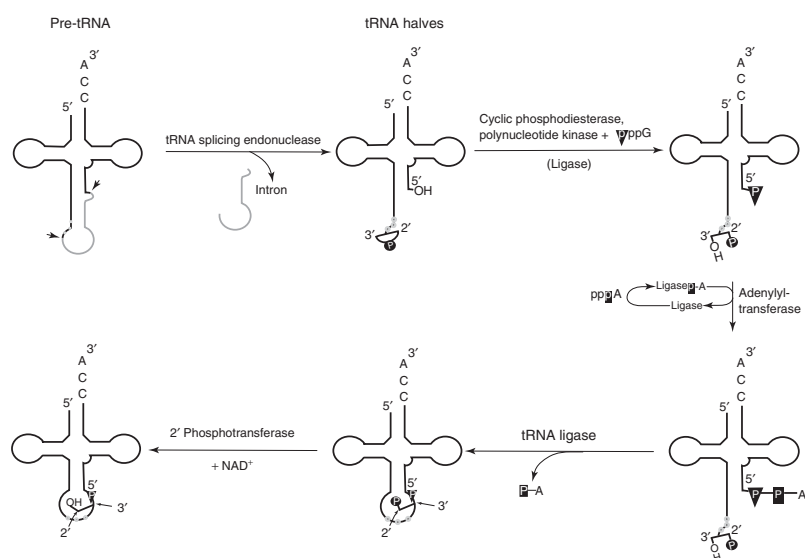


Figure 41.2 Maturation of precursor tRNAs II: mechanism of tRNA splicing in plants and yeast. Intron-containing pre-tRNAs are first cleaved by a tRNA-splicing endonuclease at the 5' and 3' boundaries of the intervening sequence, producing paired tRNA halves with 2', 3'-cyclic phosphate, 5'-OH ends, and a linear intron. Second, the halves are joined by a complex reaction requiring GTP and ATP and a multifunctional tRNA ligase that contains three intrinsic activities: an N-terminal adenylyltransferase, a central polynucleotide kinase, and a C-terminal cyclic phosphodiesterase domain. Finally, the 2' phosphate at the splice junction is removed by a NAD⁺-dependent 2' phosphotransferase. Nucleotides of the anticodon are shown as gray dots.

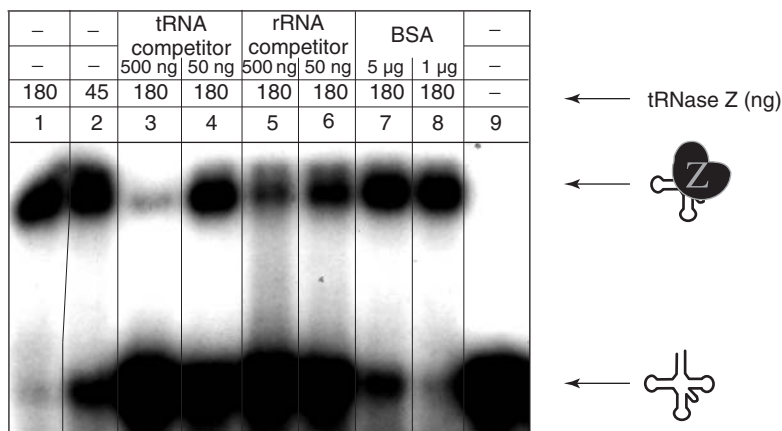


Figure 41.3 The tRNA 3' processing endonuclease trNase Z has a high affinity for tRNAs. Although the 3' end processed tRNA is a product of the reaction catalyzed by the endonuclease trNase Z, trNase Z still has a high affinity for tRNA molecules, as shown by electrophoretic mobility shift assays (EMSA). Recombinant trNase Z from *Arabidopsis thaliana* (rnuz) [26] was incubated with wheat tRNA, which was 3'-end labeled with [^{32} P]pCp (in lanes 1 and 3–8 180 ng rnuz were used, in lane 2 only 45 ng rnuz were used). In addition, either unlabeled wheat tRNA (lanes 3 and 4) or unlabeled rRNA (lanes 5 and 6) was added as competitor RNA (50 or 500 ng) to the reaction. Lanes 7 and 8 show reactions to which bovine serum albumin (BSA) was

added (5 or 1 µg). Lane 9 shows a control incubation of labeled wheat tRNA without the addition of protein. The tRNA and the tRNA–trNase Z complex are shown at the right schematically. Incubation of 180 ng rnuz with the labeled wheat tRNA resulted in an almost complete shift of the tRNA (lane 1). Addition of 500 ng unlabeled wheat tRNA as competitor RNA resulted in a drastic reduction of shifted tRNA molecules (lane 3). The addition of unlabeled rRNA as competitor reduced only slightly the amount of shifted tRNA (lane 5), showing that trNase Z binds specifically to wheat tRNAs. Addition of 1 µg BSA did not interfere with the shift, addition of 5 µg BSA slightly reduced the RNA–protein complex formation.

affinity to tRNAs, as can be shown by electrophoretic mobility shift assay (EMSA) (Figure 41.3). The use of an NHS–tRNA affinity column subsequently to a number of other chromatographic steps likewise turned out to be a powerful tool for the purification of tRNA ligase from wheat germ [30]. In addition, tRNA affinity columns have previously been successfully used to purify the tRNA 5' processing enzyme RNase P [31–33].

Materials

CNBr-Sepharose 4B Affinity Column

Cyanogen bromide (CNBr)-Sepharose 4B (GE Healthcare), 1 mM HCl, coupling buffer 1 (0.5 M NaCl, 0.1 M NaHCO₃ pH 8.3), 0.2 M glycine pH 8.0, 0.1 M NaOAc pH 4.5 containing 0.5 M NaCl, 0.1 M Tris-HCl buffer, pH 8.0, containing 0.5 M NaCl, sintered glass filter, tRNA (e.g., wheat tRNA V (Sigma)), and an empty column (e.g., C10/10 (GE Healthcare)).

NHS-Activated HiTrap Columns

Five milliliters of NHS-activated HiTrap Column (GE Healthcare), coupling buffer 2 (0.5 M NaCl, 0.2 M NaHCO₃, pH 8.3), 1 mM HCl, buffer A (0.5 M ethanolamine, 0.5 M NaCl, pH 8.3), buffer B (0.1 M acetate, 0.5 M NaCl, pH 4.0), tRNA (e.g., wheat tRNA V(Sigma)), and a 5- or 10-ml syringe.

Methods

For the purification of the wheat tRNase Z, we initially used the CNBr–tRNA affinity column, which yielded good purification. Subsequently, we used an NHS–tRNA affinity column, which resulted in an even better purification of the wheat tRNase Z. That could be due to the fact that in the CNBr–tRNA affinity column the ligand (tRNA) is only separated from the gel matrix by a single C-atom, whereas in the NHS–tRNA affinity column the tRNA is separated from the gel matrix by 10 C-atoms, giving the tRNase Z more space to bind to the tRNA (Figure 41.4).

Coupling of tRNAs to CNBr-Sepharose 4B*Preparing the Gel*

- 1) Put 2.5 g CNBr-activated Sepharose 4B (1 g powder gives about 3.5 ml final volume of gel) in 25 ml 1 mM HCl into a sintered glass filter and mix carefully using a spatula. (Do not let the CNBr-Sepharose 4B powder stand at room temperature for too long since it takes up air moisture.) Let the powder swell at room temperature for 2 min.
- 2) Wash with 4 × 100 ml 1 mM HCl, each wash for about 4 min, and do not let the powder run dry.
- 3) Wash with 100 ml coupling buffer 1 (0.5 M NaCl, 0.1 M NaHCO₃, pH 8.3).

Coupling the Ligand

- 1) Dissolve 100 mg tRNA in 40 ml coupling buffer 1 (0.5 M NaCl, 0.1 M NaHCO₃, pH 8.3). Cool to 4 °C.
- 2) All solutions should be cold (4 °C) and this step should be carried out at 4 °C to avoid degradation of the tRNA and to prevent the coupling reaction from progressing too fast. This would result in binding of only a single tRNA to several CNBr molecules. Mix the coupling solution containing the ligand with the gel in a stoppered vessel. Rotate the mixture end-over-end overnight (16–20 h) at 4 °C. Other gentle stirring methods may be employed alternatively. Do not use a magnetic stirrer.
- 3) Pour gel onto a sintered glass filter, suck the solution off (but do not let the gel run dry, keep it moist). Wash away excess ligand with 50 ml coupling buffer.
- 4) Block any remaining active groups by adding 40 ml 0.2 M glycine pH 8.0. Incubate for 2 h at room temperature and overnight at 4 °C.

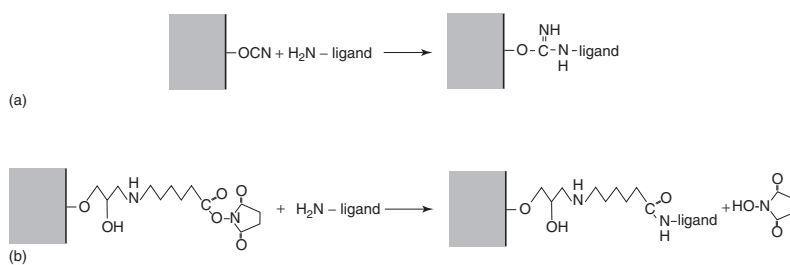


Figure 41.4 Coupling of ligands to CNBr-activated Sepharose and to NHS-activated Sepharose. Reaction of tRNA molecules with the activated matrices. Ligands bound to the CNBr-activated Sepharose are separated by a single C-atom from the gel matrix (a), whereas ligands bound to the NHS-activated column are separated by 10 C-atoms from the gel matrix (b), thus allowing better access of interacting molecules.

- 5) Remove the solution above the gel matrix (but do not let the gel run dry, keep it moist). Wash the gel with 50 ml 0.1 M NaOAc, pH 4.5 containing 0.5 M NaCl.
- 6) Wash the gel with 50 ml 0.1 M Tris-HCl buffer, pH 8.0, containing 0.5 M NaCl.
- 7) Repeat step 6 twice.
- 8) The gel is ready to be packed into a column (e.g., a C10/10 (GE Healthcare)) yielding a 5 ml tRNA affinity column.

Coupling of tRNAs to a 5 ml NHS-Activated HiTrap Column

- 1) Dissolve 20 mg tRNA in 5 ml coupling buffer 2 (0.5 M NaCl, 0.2 M NaHCO₃, pH 8.3). Cool to 4 °C.
- 2) Remove the top cap of the column and apply a drop of ice-cold 1 mM HCl to the top of the column to avoid air bubbles.
- 3) Connect the HiTrap Luer adaptor to the top of the column. Remove the twist-off end.

Coupling the Ligand

- 1) Wash out the isopropanol using a 5 or 10 ml syringe with ice-cold 1 mM HCl (3 × 10 ml) at a flow rate of 1 ml min⁻¹ (half a drop per second). Do not exceed the flow rate, or the gel can be irreversibly compressed.
- 2) Immediately inject 5 ml of ligand solution into the column.
- 3) Seal the column and let it stand for 4 h at 4 °C.
- 4) Inject 3 × 10 ml buffer A (0.5 M ethanolamine, 0.5 M NaCl, pH 8.3).
- 5) Inject 3 × 10 ml buffer B (0.1 M acetate, 0.5 M NaCl, pH 4.0)
- 6) Repeat step 4. Let the column stand for 30 min.
- 7) Inject 3 × 10 ml buffer B.
- 8) Inject 3 × 10 ml buffer A.
- 9) Repeat step 7.
- 10) Inject a buffer with neutral pH to adjust the pH. The column is now ready to use.
- 11) We usually load 29 mg (10.7 mg ml⁻¹) wheat extract purified through five steps onto the 5 ml NHS-tRNA affinity column.

41.2 Applications

41.2.1

Purification of the Nuclear tRNase Z from Wheat Germ

Earlier experiments had shown that the tRNA 3' processing activity in wheat germ is present in very low amounts, similar to other tRNA-processing enzymes such as RNase P and tRNA-splicing endonuclease [33, 34]. For the isolation of the

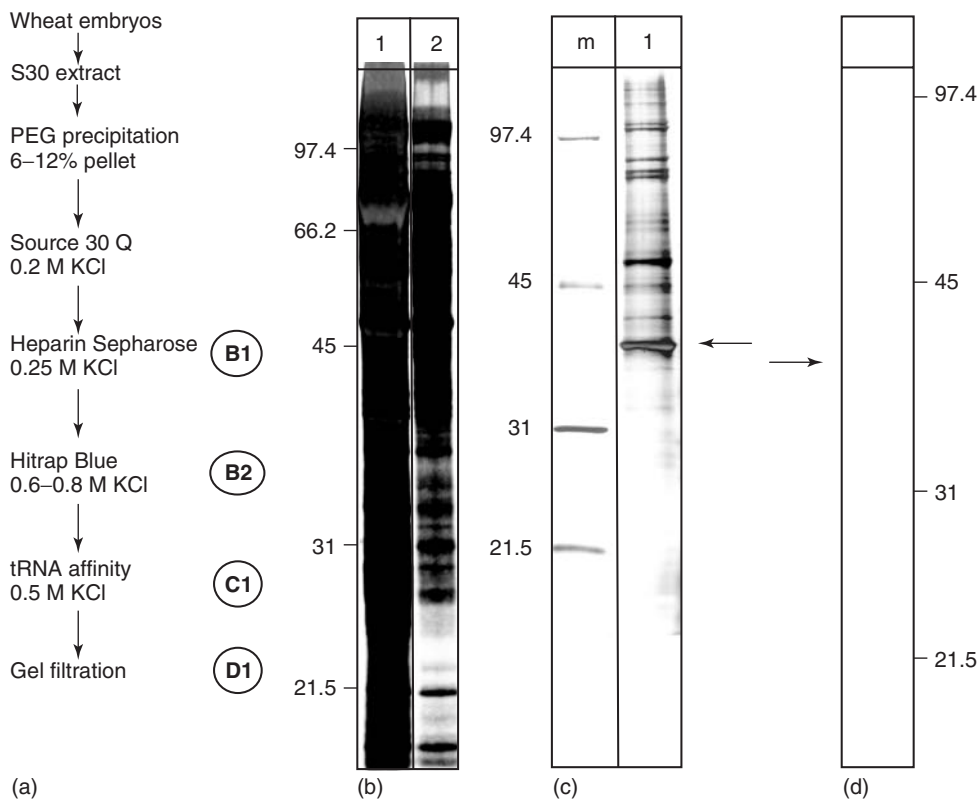


Figure 41.5 Isolation of the nuclear tRNase Z from wheat. (a) Purification scheme. tRNase Z was purified from the soluble protein fraction from wheat germ (S30) in six fractionation steps. (b) SDS-PAGE of tRNase Z active fractions from two purification steps. Aliquots of tRNase Z active fractions from the two purification steps A1 and A2 (see purification scheme in panel (a)) were loaded onto SDS gels; lane 1: 57 μ g of the 0.25 M heparin fraction, lane 2: 10 μ g of the 0.6 M KCl Blue fraction. Protein molecular weight markers (in kDa) at the left. (c) Since

little protein was left after the last purification step, protein concentrations could not be determined and therefore 10% of the tRNA affinity fraction was loaded (lane 1). At the left, a protein size standard is given (in kDa; lane m). At the right, an arrow indicates the tRNase Z. (d) SDS-PAGE of the tRNase Z active fraction after the last purification step. Again 10% of the tRNase Z-active gel filtration fraction was loaded onto the gel (lane 1). A protein size standard is given at the right (in kDa). An arrow at the left indicates the tRNase Z.

wheat tRNA 3' endonuclease a lot of material was thus required, although highly efficient fractionation step had been worked out in earlier experiments [35]. Briefly, a soluble protein fraction (S30) was extracted from 2.9 kg of wheat germ and purified through six purification step (Figure 41.5), the most efficient purification step provided by a tRNA affinity column to which tRNase Z bound tightly. Initially, we used a CNBr–tRNA affinity column [35], which already resulted in a high

purification factor. Since an NHS–tRNA affinity column showed an even higher purification than the CNBr–tRNA affinity column [26], it was subsequently used. The tRNA molecules are separated from the gel matrix in the CNBr–tRNA affinity column by only one C-atom, leaving only a little space between the gel matrix and the tRNA, whereas in the NHS–tRNA affinity column the tRNAs are separated by 10 C-atoms, giving the tRNase Z enzyme ample space to bind to the tRNA (Figure 41.4).

The 5 ml NHS–tRNA affinity column was made with wheat tRNA (Sigma), which was coupled to HiTrap NHS-activated Sepharose (GE Healthcare), as described in the Section “Coupling of tRNAs to a 5 ml NHS-Activated HiTrap Column.” The column was equilibrated with buffer A (40 mM Tris-HCl pH 8, 5 mM MgCl₂, 5% glycerol, 2 mM dithiothreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride (PMSF)), and loaded with 29 mg protein of the tRNase Z active fraction (0.6–0.8 M KCl) from the preceding Blue column, using a flow rate of 200 $\mu\text{l min}^{-1}$. Proteins were eluted with a step gradient (0.2, 0.5, 1.0, and 2.0 M KCl) using a flow rate of 500 $\mu\text{l min}^{-1}$. One milliliter fractions were collected and fractions from one gradient step were pooled, concentrated, and dialyzed using Amicon Ultra-15 (10 kDa molecular weight cutoff) concentrators (Millipore). Pooled fractions were subsequently analyzed for tRNA 3' processing activity and tRNase Z activity was recovered in the 0.5 M KCl fraction. tRNase Z purified with this tRNA affinity column showed an apparent molecular mass of 64 kDa on the subsequent gel filtration column (data not shown). After the gel filtration column, tRNase Z active fractions were analyzed by SDS polyacrylamide gel electrophoresis (PAGE), where only a 43 kDa protein correlated with the activity (Figure 41.5d). Gel filtration analysis indicated a molecular mass of 64 kDa for the tRNA 3' processing activity, and we could subsequently show that the enzyme is active as a homodimer [36].

Approximately 0.5 μg of the 43 kDa protein was excised from SDS-PAGE and subjected to tryptic digestion and tandem mass spectrometry (MS/MS) analysis. Subsequent database searches using the algorithm SEQUEST [37] and programs developed at the Harvard Microchemistry Facility [38] did not identify the corresponding gene. Therefore, peptides were separated by high-performance liquid chromatography (HPLC) and sequenced using Edman degradation. Four peptide sequences were obtained and database searches revealed a wheat cDNA sequence and two open reading frames in the *Arabidopsis thaliana* genome with high sequence similarity to these peptides. The wheat cDNA sequence translates into a protein, 100% identical to one of the sequenced wheat peptides. Since the wheat sequence is only a partial cDNA sequence, the other three peptides are outside of this sequence and consequently show no match to the cDNA. The corresponding wheat genomic sequence could not be identified in the few wheat sequences available in public databases. However, the full-length cDNA and genomic sequence were found in *Arabidopsis* and the encoded proteins were termed AthTrzS1 for the nuclear tRNase Z and AthTrzS2 for the chloroplast tRNase Z. These plant protein

sequences were the source sequence to identify homologs of this protein in all kingdoms [26].

41.2.2

Purification of the tRNA-Splicing Ligase from Wheat Germ

An RNA ligase activity from wheat germ was described over 30 years ago [39], and soon after the discovery several groups attempted to purify the activity by column chromatography [2, 40, 41]. Here, we describe a chromatographic purification procedure that employs classical ion-exchange and gel filtration chromatography as well as group-specific (Heparin and Cibacron Blue) and ligand-specific (tRNA-Sepharose) affinity chromatography. This purification procedure has finally led to the identification of the plant tRNA ligase gene [30].

A big challenge in improving the chromatographic purification scheme for the wheat RNA ligase activity is to avoid low-salt buffers as much as possible because the enzyme is not stable under these conditions. In addition, chromatographic media that exhibit a high affinity for the RNA ligase protein are preferred. The RNA ligase activity binds to Heparin Sepharose even with 200 mM salt in the binding buffer. Under these conditions, a very high selectivity for nucleic-acid-binding proteins is ensured, which enables to process large amounts of protein in the chromatographic start fraction. Hence, Heparin Sepharose is ideal for the capture step. Because the S30 extract contains a lot of endogenous ribonucleic acids that can compete with the heparin for binding, they were precipitated with 0.1% polyethylene imine and removed by centrifugation. Intermediate purification was achieved by Cibacron Blue, Source S, and Superdex 200 gel filtration chromatography, which highly enriched the RNA ligase activity (Figure 41.6b,c). For the polishing step, wheat tRNA was coupled to NHS-activated Sepharose (GE Healthcare) in a 1 ml HiTrap column according to Section “Coupling of tRNAs to a 5 ml NHS-Activated HiTrap Column.” The RNA ligase activity binds in a start buffer with 50 mM salt and elutes with 180–240 mM salt (Figure 41.6a).

In the final purification step, the tRNA ligase protein has been identified by its reaction with [α - ^{32} P]ATP, resulting in a covalently bound [^{32} P]AMP (Figure 41.6c). The specifically labeled chromatographic fraction was separated by SDS-PAGE and the single band correlating to a protein with a molecular weight of 125 kDa was identified by silver staining and autoradiography (Figure 41.6d). About 5 μ g of Coomassie-Blue-stained RNA ligase protein was excised from an SDS-PAGE for subsequent gel tryptic digestion followed by microcapillary reverse-phase HPLC nanoelectrospray tandem mass spectrometry (μ LC/MS/MS) on a Finnigan LCQ hundred and twenty sequenced tryptic fragments were run against the translated National Center for Biotechnology Information (NCBI EST (expressed sequence tags) database. Seven nonredundant EST sequences were identified by SEQUEST analysis, which all encode one annotated *Arabidopsis* protein now called *AtRNL*.

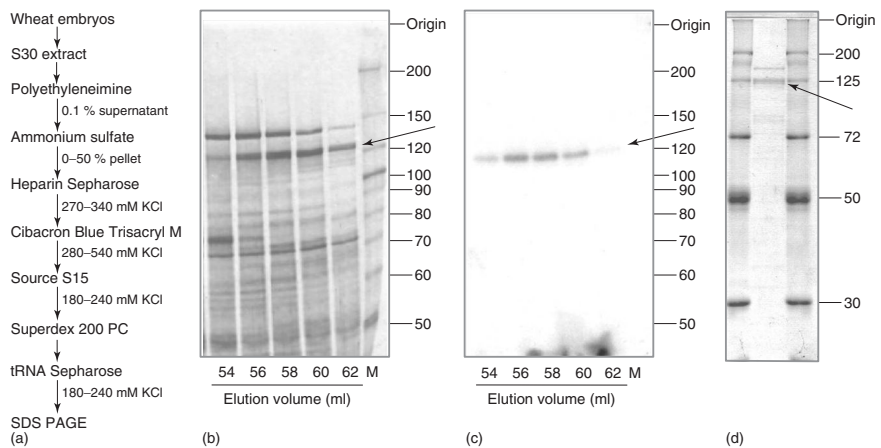


Figure 41.6 Isolation of the nuclear tRNA ligase from wheat. (a) Purification scheme. tRNA ligase was purified from the soluble fraction of wheat embryos (S30 extract) by six consecutive steps [30]. (b) Partially purified tRNA ligase from the Source S15 column was subjected to gel filtration on SuperdexTM 200. Fractions of 2 ml were collected. Aliquots of the elution fractions were analyzed on a 7.5% polyacrylamide/0.1% SDS gel. The proteins were visualized by silver staining. The arrow indicates the tRNA ligase. (c) The indicated fractions were assayed for adenyltransferase activity of wheat tRNA ligase. Appropriate aliquots were incubated in the presence of $[\alpha\text{-}^{32}\text{P}]\text{ATP}$ for 15 min at 37 °C. The ligase- $[\text{P}^{32}]\text{AMP}$ adduct was detected by autoradiography of the dried gel. The arrow indicates the tRNA ligase. (d) The peak fractions from the tRNA Sepharose column were concentrated by ultrafiltration and 1/20 of this material was applied onto a 10% polyacrylamide/0.1% SDS gel and stained with Coomassie Blue. The arrows point to the position of the RNA ligase protein with an approximate molecular weight of 125 kDa. Protein size standards in kilodaltons are indicated on the right.

41.3

Notes

RNA affinity columns are a powerful method to purify RNA binding proteins. For example, a 1000-fold increase in the purification factor has been achieved by a single chromatographic purification step on an RNA affinity matrix [1]. Here, the ammonium-sulfate fractionated *Escherichia coli* extract has been directly applied on the tRNA Sepharose column.

The ultimate goal – purification of the activity from an extract to homogeneity – can rarely be achieved by a single RNA affinity chromatographic step. Additional classical chromatographic step are usually necessary. Ion-exchange chromatography is especially suitable to process crude extracts and thereby deplete the chromatographic fractions from endogenous RNA, which might compete with the RNA affinity matrix for binding.

References

1. Aschhoff, H.J., Elten, H., Arnold, H.H., Mahal, G., Kersten, W., and Kersten, H. (1976) 7-Methylguanine specific tRNA-methyltransferase from *Escherichia coli*. *Nucleic Acids Res.*, **3**, 3109–3122.
2. Gegenheimer, P., Gabius, H.J., Peebles, C.L., and Abelson, J. (1983) An RNA ligase from wheat germ which participates in transfer RNA splicing *in vitro*. *J. Biol. Chem.*, **258**, 8365–8373.
3. Gerber, C.A., Relich, A., and Driscoll, D.M. (2004) Isolation of an mRNA-binding protein involved in C-to-U editing. *Methods Mol. Biol.*, **265**, 239–249.
4. Heo, I., Joo, C., Kim, Y.K., Ha, M., Yoon, M.J., Cho, J., Yeom, K.H., Han, J., and Kim, V.N. (2009) TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell*, **138**, 696–708.
5. Hogg, J.R. and Collins, K. (2007) RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. *RNA*, **13**, 868–880.
6. Hovhannisyan, R. and Carstens, R. (2009) Affinity chromatography using 2' fluoro-substituted RNAs for detection of RNA-protein interactions in RNase-rich or RNase-treated extracts. *Biotechniques*, **46**, 95–98.
7. Li, H. and Zassenhaus, H.P. (1999) Purification and characterization of an RNA dodecamer sequence binding protein from mitochondria of *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.*, **261**, 740–745.
8. Michlewski, G. and Cáceres, J.F. (2010) RNase-assisted RNA chromatography. *RNA*, **16**, 1673–1678.
9. Minich, W.B., Balasta, M.L., Goss, D.J., and Rhoads, R.E. (1994) Chromatographic resolution of *in vivo* phosphorylated and nonphosphorylated eukaryotic translation initiation factor eIF-4E: increased cap affinity of the phosphorylated form. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 7668–7672.
10. Remy, P., Birmele, C., and Ebel, J.P. (1972) Purification of yeast phenylalanyl-tRNA synthetase by affinity chromatography, on a tRNA(Phe)-sepharose column. *FEBS Lett.*, **27**, 134–138.
11. Schofield, P. and Williams, K.R. (1977) Purification and some properties of *Escherichia coli* tRNA nucleotidyltransferase. *J. Biol. Chem.*, **252**, 5584–5588.
12. Sharma, S. (2008) Isolation of a sequence-specific RNA binding protein, polypyrimidine tract binding protein, using RNA affinity chromatography. *Methods Mol. Biol.*, **488**, 1–8.
13. Soeno, Y., Taya, Y., Stasyk, T., Huber, L.A., Aoba, T., and Hüttenhofer, A. (1993) Identification of novel

- ribonucleo-protein complexes from the brain-specific snoRNA MBII-52. *RNA*, **16**, 1293–1300.
14. Trippe, R., Richly, H., and Benecke, B.J. (2003) Biochemical characterization of a U6 small nuclear RNA-specific terminal uridylyltransferase. *Eur. J. Biochem.*, **270**, 971–980.
 15. Ulbrich, N., Wool, I.G., Ackerman, E., and Sigler, P.B. (1980) The identification by affinity chromatography of the rat liver ribosomal proteins that bind to elongator and initiator transfer ribonucleic acids. *J. Biol. Chem.*, **255**, 7010–7019.
 16. Walker, S.C., Scott, F.H., Srisawat, C., and Engelke, D.R. (2008) RNA affinity tags for the rapid purification and investigation of RNAs and RNA-protein complexes. *Methods Mol. Biol.*, **488**, 23–40.
 17. Gadgil, H., Oak, S.A., and Jarrett, H.W. (2001) Affinity purification of DNA-binding proteins. *J. Biochem. Biophys. Methods*, **49**, 607–624.
 18. Allerson, C.R., Martinez, A., Yikilmaz, E., and Rouault, T.A. (2003) A high-capacity RNA affinity column for the purification of human IRP1 and IRP2 overexpressed in *Pichia pastoris*. *RNA*, **9**, 364–374.
 19. Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2003) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16719–16724.
 20. Srisawat, C. and Engelke, D.R. (2002) RNA affinity tags for purification of RNAs and ribonucleoprotein complexes. *Methods*, **26**, 156–161.
 21. Vasudevan, S. and Steitz, J.A. (2007) AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell*, **128**, 1105–1118.
 22. Windbichler, N. and Schroeder, R. (2006) Isolation of specific RNA-binding proteins using the streptomycin-binding RNA aptamer. *Nat. Protoc.*, **1**, 637–640.
 23. Mörl, M. and Marchfelder, A. (2001) The final cut. The importance of tRNA 3'-processing. *EMBO Rep.*, **2**, 17–20.
 24. Kruszka, K., Barneche, F., Guyot, R., Ailhas, J., Meneau, I., Schiffer, S., Marchfelder, A., and Echeverria, M. (2003) Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J.*, **22**, 621–632.
 25. Schierling, K., Rösch, S., Rupprecht, R., Schiffer, S., and Marchfelder, A. (2002) tRNA 3' end maturation in Archaea has eukaryotic features: the RNase Z from *Haloferax volcanii*. *J. Mol. Biol.*, **316**, 895–902.
 26. Schiffer, S., Rösch, S., and Marchfelder, A. (2002) Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes. *EMBO J.*, **21**, 2769–2777.
 27. Vogel, A., Schilling, O., Späth, B., and Marchfelder, A. (2005) The tRNase Z family of proteins. Physiological functions, substrate specificity and structural properties. *Biol. Chem.*, **386**, 1253–1264.
 28. Pellegrini, O., Nezzar, J., Marchfelder, A., Putzer, H., and Condon, C. (2003) Endonucleolytic processing of CCA-less tRNA precursors by RNase Z in *Bacillus subtilis*. *EMBO J.*, **22**, 4534–4543.
 29. Dubrovsky, E.B., Dubrovskaya, V.A., Levinger, L., Schiffer, S., and Marchfelder, A. (2004) Drosophila RNase Z processes mitochondrial and nuclear pre-tRNA 3' ends *in vivo*. *Nucleic Acids Res.*, **32**, 255–262.
 30. Englert, M. and Beier, H. (2005) Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other phylogenetic origins. *Nucleic Acids Res.*, **33**, 388–399.
 31. Lee, Y.C., Lee, B.J., Hwang, D.S., and Kang, H.S. (1996) Purification and characterization of mitochondrial ribonuclease P from *Aspergillus nidulans*. *Eur. J. Biochem.*, **235**, 289–296.
 32. True, H.L. and Celander, D.W. (1996) Ribonuclease P of *Tetrahymena thermophila*. *J. Biol. Chem.*, **271**, 16559–16566.
 33. Zimmerly, S., Drinas, D., Sylvers, L.A., and Söll, D. (1993) Identification of a 100-kDa protein associated with nuclear

- ribonuclease P activity in *Schizosaccharomyces pombe*. *Eur. J. Biochem.*, **217**, 501–507.
34. Kleman-Leyer, K., Armbruster, D.W., and Daniels, C.J. (1997) Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. *Cell*, **89**, 839–847.
35. Mayer, M., Schiffer, S., and Marchfelder, A. (2000) tRNA 3' processing in plants: nuclear and mitochondrial activities differ. *Biochemistry*, **39**, 2096–2105.
36. Späth, B., Kirchner, S., Vogel, A., Schubert, S., Meinlschmidt, P., Aymanns, S., Nezzar, J., and Marchfelder, A. (2005) Analysis of the functional modules of the tRNA 3' endonuclease (tRNase Z). *J. Biol. Chem.*, **280**, 35440–35447.
37. Eng, J.K., McCormack, A.L., and Yates, J.R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.*, **5**, 976–989.
38. Chittum, H.S., Lane, W.S., Carlson, B.A., Roller, P.P., Lung, F.D., Lee, B.J., and Hatfield, D.L. (1998) Rabbit beta-globin is extended beyond its UGA stop codon by multiple suppressions and translational reading gaps. *Biochemistry*, **37**, 10866–10870.
39. Konarska, M., Filipowicz, W., Domdey, H., and Gross, H.J. (1981) Formation of a 2'-phosphomonoester, 3',5'-phosphodiester linkage by a novel RNA ligase in wheat germ. *Nature*, **293**, 112–116.
40. Furneaux, H., Pick, L., and Hurwitz, J. (1983) Isolation and characterization of RNA ligase from wheat germ. *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 3933–3937.
41. Pick, L. and Hurwitz, J. (1986) Purification of wheat germ RNA ligase. I. Characterization of a ligase-associated 5'-hydroxyl polynucleotide kinase activity. *J. Biol. Chem.*, **261**, 6684–6693.

42

Biotin-Based Affinity Purification of RNA–Protein Complexes

Marco Preußner, Silke Schreiner, Inna Grishina, Zsolt Palfi,
Jingyi Hui, and Albrecht Bindereif

42.1

Introduction

Many cellular functions are carried out by proteins that are components of larger complexes. The identification of proteins present in biological complexes requires their purification, which can be achieved through affinity procedures. Often, the protein components are then analyzed and identified by a combination of mass spectrometry and database search. This article focuses on ribonucleoprotein (RNP) complexes, for which affinity purification techniques are particularly suitable since either a protein- or an RNA-specific affinity tag can be incorporated into the complex. The use of protein affinity tags has become routine for the purification and detection of protein complexes. A variety of affinity tags are available for labeling proteins of interest, for example, for tandem affinity purification strategies [1, 2]. On the other hand, RNA–protein complexes can also be affinity-purified through their RNA component, using RNA aptamers as affinity tags, for example, the streptavidin-binding S1 RNA aptamer [3], the D8 Sephadex-binding RNA minimal motif [4], or the tobramycin (Tob) aptamer [5, 6]. Alternatively, a well-known RNA–protein interaction can be exploited for this purpose, by incorporation of binding sites for the MS2 coat protein into the RNA of interest and using an MS2 coat protein–maltose-binding protein (MS2-MBP) fusion for affinity purification on an amylose resin. The bound RNP complex can subsequently be eluted from the resin under native conditions with free maltose [7, 8]. A recent example of this approach was the isolation of human spliceosomes in a functional form, resulting in their comprehensive proteomic analysis [9, 10]. In addition, a so-called RNA Affinity in Tandem (RAT) method has been developed and successfully applied to the isolation of endogenous RNPs: The optimized RAT tag added to the RNA of interest consists of two hairpin stem-loops that bind the PP7 coat protein (PP7) and Tob [11].

Recently, a new affinity purification strategy named *TagIt* has been developed. The *TagIt*-based approach uses a novel epitope derived from the N-terminal part of the human survival of motor neuron (SMN) protein. This tag is bound by a

monoclonal antibody with high specificity and affinity. The covalent coupling of the synthetic 7B10 epitope to an RNA of interest may permit the purification of RNA binding proteins [12].

In the following sections, we concentrate on the affinity purification of RNA–protein complexes through their RNA component, using biotin-based methods. Biotin can be introduced into the RNA (i) at the 5'-terminal or an internal site *via* biotinylated nucleotides during *in vitro* SP6 or T7 transcription, (ii) by chemical 3' biotinylation (see Chapter 6: 3' terminal attachment of fluorescent dyes and biotin), or (iii) indirectly by annealing a chemically synthesized biotinylated oligonucleotide antisense to the RNA of interest. Subsequently, affinity selections make use of the tight recognition of biotin residues by avidin or streptavidin.

Examples for the first approach come from the initial identification of spliceosomes, when biotinylated nucleotides were randomly incorporated into pre-mRNA substrates during *in vitro* transcription. Spliceosomes were then assembled on such biotinylated splicing substrates in extract, followed by binding to the streptavidin resin under native conditions [13–15]. In this way, the set of spliceosomal snRNAs U1, U2, U4, U5, and U6 were first identified. Alternatively, spliceosomes assembled on unmodified pre-mRNAs can be captured on a streptavidin resin with biotinylated antisense oligonucleotides [16].

The use of biotinylated antisense 2'-O-methyl RNA (2'OMe) oligonucleotides for the specific isolation of RNP complexes was first introduced by Lamond, Sproat, and coworkers [16–20, reviewed by Ref. 19]. They found that 2'OMe RNA oligonucleotides have several advantages over other types of oligonucleotides, which can be exploited for the affinity selection procedure. Most importantly, 2'OMe RNA oligonucleotides are nuclease resistant, and – in comparison to DNA oligonucleotides – have further advantages, including a higher affinity for RNA, increased specificity, faster hybridization kinetics, and a superior ability to bind to structured RNA targets [20–22]. The antisense affinity technology using biotinylated 2'OMe RNA oligonucleotides is very versatile and allows the rapid purification of any RNA–protein complex, in which the RNA component is accessible to oligonucleotide binding [23–27]. Purified complexes can be directly processed for protein identification by mass spectrometric analysis, and for analysis of their RNA components. Both protein and RNA analyses can be performed on the same samples (see Section 42.3.1.3).

However, because of the extremely stable biotin–avidin interaction, this standard procedure requires denaturation of the affinity-purified complex. To overcome this limitation, an important variation of the original protocol was suggested by Lingner and Cech [28], allowing the elution of purified RNP complexes from the streptavidin matrix and their further functional analysis. This so-called displacement strategy results in the release from the affinity matrix under native conditions and is based on disrupting the 2'OMe RNA/target RNA hybrid by an excess of a DNA oligonucleotide directed against the 2'OMe RNA oligonucleotide. This strategy has been applied to the affinity purification and functional analysis of different RNPs [29–32].

Here, we describe in detail experimental procedures of how RNPs can be affinity-purified on the basis of the biotin–streptavidin interaction, focusing on two

major strategies, the first involving biotinylated 2'OMe RNA oligonucleotides that are antisense to the RNA component of the RNP and the second using biotin-labeled RNAs. Specific examples are presented for either of the two strategies, and potential problems are discussed.

42.2

Materials

42.2.1

Biotinylated Probes

The following types of biotinylated probes are used for affinity selections: (i) 2'OMe RNA oligonucleotides complementary in sequence to the targeted RNA [16, 24], (ii) chimeric 2'OMe RNA/DNA oligonucleotides in combination with displacement DNA oligonucleotides [28], and (iii) biotinylated RNAs [13, 33, 34]. For oligonucleotide design, a good starting point is the secondary structure model of the target RNA. As a general rule, single-stranded regions are chosen, and their accessibility within intact RNPs can be experimentally confirmed by oligodeoxynucleotide-directed RNase H digestion [35]. Most importantly, the affinity oligonucleotide should be specific for the target RNA. The oligonucleotide used for selection generally comprises at least 10–12 residues and contains one or more biotin moieties for immobilization on affinity matrices. Accessibility of the biotin moiety to the affinity matrix is important for the efficiency of selection and can be influenced by its position on the oligonucleotide. Incorporation of biotin at both termini might improve binding efficiency [36]. All biotin reagents used for biotinylation of oligonucleotides or RNAs should contain a spacer arm, at least six C-atoms in length, to reduce steric hindrance.

42.2.2

Affinity Matrices

Different types of affinity matrices are available for the immobilization of biotinylated oligonucleotides. The most frequently used ones are *streptavidin agarose* (SAG; Sigma) or *streptavidin agarose resin* (Pierce). Streptavidin is a ~60 kDa protein composed of four identical subunits, each of which has a binding site for biotin. Streptavidin is immobilized on agarose beads, supplied as a 1 : 1 suspension in aqueous solution. It shows relatively low levels of non-specific binding. *Neutravidin agarose* (NAG; Pierce), a chemically modified (deglycosylated) version of avidin, shows very low non-specific binding compared to streptavidin and high binding capacity for biotinylated molecules. The binding affinity of *immobilized monomeric avidin* (Pierce) is several orders of magnitude lower than that of tetrameric avidin. Immobilized monomeric avidin can be used for reversible binding of biotinylated probes: Bound biotinylated molecules can be specifically eluted by ligand competition using a biotin-containing buffer. *Streptavidin-coated magnetic beads*

(e.g., Streptavidin-coupled Dynabeads, Invitrogen, or Streptavidin MagneSphere Paramagnetic Particles, Promega) are paramagnetic polystyrene beads with streptavidin covalently attached to the beads' surface. They have a lower binding capacity for biotinylated molecules than SAg or NAg, but have negligible non-specific binding for negatively charged molecules, and show less aggregation in high-salt buffers. They are handled fast and conveniently by a magnetic particle concentrator. Such magnetic beads are not recommended for selection of abundant RNPs or for unfractionated cell lysates [19]. Another disadvantage is that heating causes release of iron, which may degrade RNA.

42.2.3

Cell Extracts

Extracts from cells, tissues, or organisms can be prepared by a variety of procedures. The choice of the appropriate procedure will depend on the RNA–protein complex to be purified and on prior experience. In general, it will be necessary to check whether the applied extraction procedure is efficient and if the targeted RNP particle is stable under those conditions. Particularly relevant to the success of the affinity selection is the ionic strength of the extracts, the presence of non-ionic detergents (NP-40 or Triton X-100), and of RNase and protease inhibitors. Cell lysates are best used freshly for affinity selections. In case of long-term storage, extracts are supplemented with 10% glycerol, aliquoted, shock-frozen in liquid nitrogen, and kept at -80°C . In the example given below, total cell lysates were prepared from *Trypanosoma brucei* cells as described [37]. Cell fractionation can facilitate the affinity purification by enriching for the RNP. The preparation of nuclear extracts from human cells has been described by Dignam *et al.* [38]; modifications toward more efficient affinity selections have been introduced by Blencowe and Lamond [19]; extract preparation from small quantities of cells has been developed by Lee *et al.* [39]. In case of selecting RNPs of low abundance, chromatographic fractionation methods can also be applied to enrich for certain RNP particles (for example, diethylaminoethyl (DEAE)-Sephacrose chromatography in the case of snRNPs, see Refs. [31, 33]).

42.2.4

Buffers and Solutions

Buffers should be prepared freshly in dimethyl pyrocarbonate (DMPC)-treated water and kept at 4°C (except when otherwise stated). Note that Tris buffers cannot be DMPC-treated, because Tris reacts with DMPC. Therefore, to make a Tris buffer, the water should be DMPC-treated first and autoclaved twice before adding Tris. After addition of Tris, the solution should be autoclaved again. Dithiothreitol (DTT) and phenylmethylsulfonyl fluoride (PMSF) or complete protease inhibitor cocktail tablets (e.g., Roche) are added to the buffers just before use.

- Streptavidin agarose blocking buffer (SAgBB): 1 mg of RNase-free bovine serum albumin (BSA), 200 μ g glycogen, and 200 μ g yeast tRNA per milliliter low-salt wash buffer (LS-WB)
- LS-WB: 20 mM HEPES, pH 8.0, 100 mM KCl, 10 mM MgCl₂, 0.01% Nonidet P-40 (NP-40), 1 mM DTT
- High-salt wash buffer (HS-WB): The same as LS-WB, except that the KCl concentration is 300 mM or higher (as stated in the text)
- 6 and 9 M urea solutions: 6 or 9 M urea dissolved in DMPC-treated water containing 1 mM DTT and 0.01% NP-40
- 2 x proteinase K buffer: 0.2 M Tris, pH 7.5, 0.3 M NaCl, 2% SDS, 25 mM EDTA
- 2 x Laemmli gel loading buffer: 0.1 M Tris-Cl, pH 6.8, 4% SDS, 20% glycerol, 0.2% bromophenol blue, 0.2 M DTT
- WB400: 20 mM HEPES, pH 8.0; 400 mM KCl, 10 mM MgCl₂, 0.01% Nonidet P-40 (NP-40), 1 mM DTT
- Buffer D: 20 mM HEPES, pH 8.0, 100 mM KCl, 0.2 mM EDTA, 20% (v/v) glycerol, 1 mM DTT, 1 mM PMSF
- 10% Nonidet P-40 in DMPC-treated water
- Phenol-chloroform-isoamylalcohol (25 : 24 : 1) saturated with TE buffer (10 mM Tris, pH 8.0, 1 mM EDTA)
- 70 and 100% ethanol, 100% acetone.

42.2.5

Additional Materials

- Sterile, RNase-free 1.5 ml Eppendorf tubes, 15 and 50-ml Falcon tubes, pipette tips
- Microcentrifuge
- Rotating wheel
- Thermoshaker for Eppendorf tubes
- Magnetic particle concentrator (when using streptavidin magnetic particles).

42.3

Methods

In the following sections, we first describe how biotinylated antisense 2'OMe oligoribonucleotides are used for affinity purification of RNP complexes and, second, how specific RNA binding proteins can be isolated through biotinylated RNAs from cell extracts.

42.3.1

Affinity Purification of RNA-Protein Complexes (RNPs)

To select the optimal sequence for the biotinylated 2'OMe RNA oligonucleotide, regions of the target RNP should be determined that are accessible to oligonucleotide binding. This can be achieved, for example, by RNase H cleavage with oligonucleotides directed to different regions of the target RNA. When an RNP

particle should be purified under native conditions for functional assays, chimeric 2'OMe RNA/DNA oligonucleotide in combination with a displacement DNA oligonucleotide can be used (see under Section 42.3.1.4). In case of low-abundance RNPs, it might be important to enrich the RNPs of interest by chromatographic fractionation or by glycerol gradient centrifugation. As a control, mock selection (without oligonucleotide or with an unrelated oligonucleotide) must always be run in parallel. In case the mock control shows background binding, preclearing of the cell extract might be helpful (see Section 42.3.1.1). Streptavidin agarose should be preblocked in blocking buffer before each selection to saturate non-specific binding sites. As a specific example the affinity purification of snRNPs from *T. brucei* cell lysates is described in detail in the following section.

42.3.1.1 Depletion of Total Cell Lysate from SAg-Binding Material (Preclearing)

All steps should be carried out at 4 °C. SAg should be handled with care: Never use high-speed centrifugation to collect the beads, since this may damage the structure of the agarose particles, and do not vortex SAg-containing suspensions. For mixing, use a rotating wheel or a shaker adjusted to low speed.

- 1) To preclear 10 ml extract (protein concentration $\sim 10 \text{ mg ml}^{-1}$), resuspend SAg in the commercial bottle and transfer 200 μl slurry into two 1.5 ml Eppendorf tubes. Centrifuge the suspension at 1500 g for 1 min to remove the buffer (contains preservative). This gives 100 μl packed beads in each tube. (The amount of affinity matrix used to preclear crude cell lysates should be determined experimentally, but as a rough guide, we recommend to use $\sim 10\text{--}20 \mu\text{l}$ packed SAg beads per 10 mg of protein in the lysate.)
- 2) Add 1 ml LS-WB to the pelleted beads, and mix the suspension on a rotating wheel for 5 min.
- 3) Centrifuge the beads to remove the buffer (1500 g, 1 min).
- 4) Repeat washing the beads in HS-WB (steps 2–3).
- 5) Take the necessary amount of fresh cell extract (depending on whether analytical or preparative-scale affinity selections are planned) and supplement with NP-40 (final concentration: 0.01%). Frozen aliquots of cell lysates should be thawed quickly by warming in a 30 °C water bath, cooled on ice, and then centrifuged at 10 000 g for 10 min to remove large aggregates or precipitates (which would otherwise contaminate everything). Transfer the supernatant into fresh tubes.
- 6) Transfer one aliquot of the prewashed, pelleted SAg beads into the cell extract by resuspending the beads in a small volume ($\sim 500 \mu\text{l}$) of extract.
- 7) Rotate the tube slowly on a turning wheel for 1 h at 4 °C.
- 8) Centrifuge the suspension for 2 min at 1500 g.
- 9) Transfer the supernatant to a new tube and repeat steps 6–8 with the second aliquot of prewashed SAg beads.
- 10) Repeat centrifugation of the supernatant once more to be sure that all SAg beads are removed from the extract.

42.3.1.2 Preblocking Streptavidin Agarose Beads

- 1) Transfer the required amount of SAg slurry into a 1.5 ml Eppendorf tube. Centrifuge the suspension at 1500 g for 1 min; discard the supernatant.
- 2) Resuspend the SAg (100 μ l packed beads) in 1 ml cold LS-WB and rotate the tube on a turning wheel for 5 min.
- 3) Collect the beads by centrifugation (1500 g, 1 min).
- 4) Repeat steps 2–3.
- 5) Resuspend the beads in SAgBB, using 1 ml buffer for blocking 100 μ l packed beads.
- 6) Rotate the tube slowly on a turning wheel for 1 h at 4 °C.
- 7) Collect the beads by a 1 min centrifugation at 1500 g and discard the supernatant.
- 8) Mix the beads in 1 ml cold HS-WB containing 300 mM KCl (5 min on a turning wheel).
- 9) Collect the beads again by centrifugation (1500 g, 1 min).
- 10) Repeat steps 8–9 two more times. The preblocked SAg beads can be stored for one to two days at 4 °C.

42.3.1.3 Affinity Selection of RNPs for Biochemical Studies

In principle, this can be done in two different ways: (i) by prebinding the biotinylated antisense 2'OMe RNA oligonucleotide to SAg beads and then hybridizing the immobilized oligonucleotide to the target RNP in the extract and (ii) by binding the oligonucleotide to the target RNP in solution (extract), followed by immobilizing the complex on SAg beads. Which version is better? To decide, first carry out small-scale selections. The efficiency of selection may be influenced by the order of these steps. Below, we describe a specific example for affinity selection of snRNPs, demonstrating how low-abundance RNP particles can be purified.

Example: Affinity Purification of *T. brucei* U2, U4/U6, and spliced leader (SL) RNPs

In this case, the cell lysate is first enriched for snRNPs by DEAE Sepharose chromatography [37]. The biotinylated antisense 2'OMe RNA oligonucleotides are immobilized first on SAg and then incubated with the DEAE fraction. The following 2'OMe RNA oligonucleotides were used for affinity selection: Tb-U2-5', Tb-U4-int, and Tb-SL-int2 [24]. An unrelated oligonucleotide, HuU4-5' [40] served as a mock control. The result of this affinity selection procedure is shown in Figure 42.1.

Procedure

- 1) Take 200 μ l preblocked SAg beads (1 : 1 slurry in HS-WB) for each selection into a 1.5 ml Eppendorf tube (for four selections: 4 $\times$ 200 μ l).
- 2) Add 800 μ l HS-WB containing 10 μ g of biotinylated 2'OMe RNA oligonucleotide (~2 nM of a 15-mer oligonucleotide) to each tube. For each affinity selection, use the corresponding oligonucleotide. Mock selection can also be

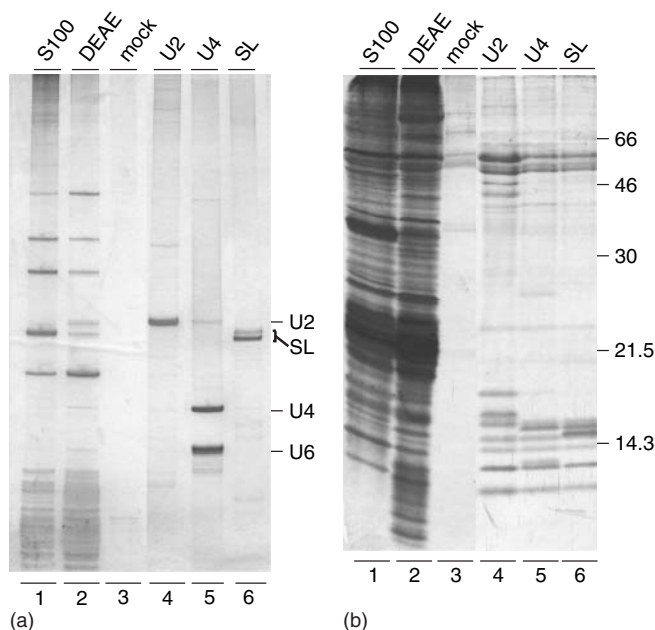


Figure 42.1 Affinity purification of *T. brucei* U2, U4/U6, and SL RNPs using biotinylated 2'OMe RNA oligonucleotides. (a) RNA analysis of affinity-selected snRNPs. After affinity selections from 2 ml of an snRNP-enriched DEAE fraction with 2'OMe RNA oligonucleotides specific for U2, U4, or SL RNAs (lanes 4–6) and subsequent protein dissociation, bound RNA was released from the affinity matrix, separated by 10% denaturing PAGE, and visualized by silver staining. As a mock control, an unrelated oligonucleotide was used in the affinity selection step (lane 3). For comparison, total RNA purified from 25 μ l of *T. brucei* S100 (lane 1) and 25 μ l of the DEAE fraction (lane 2)

were loaded as well. The positions of U2, SL, U4, and U6 RNAs are indicated on the right. (b) Protein analysis of affinity-selected snRNPs. Following affinity selection with U2, U4, or SL RNA-specific 2'OMe RNA oligonucleotides (lanes 4–6), proteins were released (see Section 42.3.1.3), separated by a 15% SDS-PAGE, and visualized by silver staining. For the mock control, an unrelated oligonucleotide was used in the affinity selection step (lane 3). In addition, proteins present in 10 μ l of *T. brucei* S100 (lane 1) and in 10 μ l of the DEAE fraction (lane 2) were analyzed. Protein markers are indicated in kilodaltons on the right.

done by leaving out the oligonucleotide. Bind the oligonucleotides to SAg during a 2 h incubation at 4 °C, slowly rotating the tubes on a turning wheel.

- 3) Collect the SAg beads by a short centrifugation (1500 g, 1 min); discard the supernatants.
- 4) Resuspend the pelleted beads in 1 ml HS-WB and incubate the samples for 5 min at 4 °C by slow rotation.
- 5) Spin down the beads (1500 g, 1 min) and repeat washing in 1 ml HS-WB three times (steps 4–5).
- 6) Transfer the 100 μ l packed SAg beads each carrying 10 μ g oligonucleotide into 2 ml Eppendorf tubes by using 200 μ l HS-WB for the transfer.

- 7) Add 2 ml pooled DEAE fraction (1 ml DEAE fraction corresponds to $\sim 5 \times 10^{10}$ cell equivalents) into each tube (for four selections: 4×2 ml) and mix the suspension on a turning wheel for 1 h by slowly rotating the tubes.
- 8) Collect the SAg beads by a short spin (1500 g, 1 min). Save the supernatant and resuspend the pellet in 1 ml HS-WB. Transfer the suspension into a fresh 1.5 ml Eppendorf tube. The supernatant can be used to check for the efficiency of depletion of the targeted RNP.
- 9) Centrifuge the SAg in a microcentrifuge at 1500 g for 1 min. Remove the supernatant.
- 10) Add 1 ml HS-WB to the pellet and rotate the samples for 5 min.
- 11) Repeat steps 10 and 11 five times.

Analysis of the Affinity-Selected Material: Protein and RNA Components

Both proteins and RNAs of the selected RNPs are recovered from the same sample. First, *proteins* are recovered by the following procedure:

- 1) Resuspend the beads in 100 μ l 9 M urea solution and incubate the samples at room temperature (RT) for 30 min by gentle shaking; then centrifuge at 1500 g for 1 min. Transfer the supernatant to a fresh tube and save it.
- 2) Repeat elution of the proteins by adding 100 μ l 6 M urea solution to the beads, incubating for 30 min at RT (by shaking the sample) and a short spin at 1500 g. Collect the supernatant and mix it with the first urea eluate.
- 3) Spin down the pooled fractions to remove all beads ($\geq 20\,000$ g, 1 min). Transfer the supernatant into a new Eppendorf tube.
- 4) Precipitate proteins by adding 5 volumes of cold acetone, and place the samples at -20°C . Efficient precipitation of proteins takes a minimum of 1 h.
- 5) Warm up the samples to RT. Collect the pelleted proteins by centrifugation for 30 min at $20\,000$ g in an Eppendorf microcentrifuge at RT (urea might precipitate at 4°C).
- 6) Wash the pellet with 80% ethanol at RT (add 1 ml ethanol to the pellet, invert the tubes a few times and centrifuge the samples for 5–10 min at $20\,000$ g in a microcentrifuge at RT).
- 7) Repeat step 6 once more to remove all urea.
- 8) Dissolve the protein pellet in a small volume of Laemmli gel loading buffer. To be able to load all isolated proteins into one well of a protein gel, dissolve the samples in a total volume of 20 μ l loading buffer.
- 9) Analyze the protein composition of the selected samples by SDS-PAGE. Proteins can be detected by Coomassie or silver staining, or sequential staining with Coomassie and silver (double staining). After silver staining, a minor amount of released RNA might also be visible on the protein gel. If antibodies are available, the identity of proteins can be checked by Western blotting. Other proteins may be identified by mass spectrometric analysis.

Next, RNAs are released from the affinity matrix by the following procedure:

- 1) After releasing the proteins, wash the 100 μ l packed beads in 1 ml HS-WB (rotating the suspension on a wheel for 5 min).

- 2) Pellet the beads by a short centrifugation (1500 g, 1 min) and discard the supernatant.
- 3) Mix the packed SAg beads with 100 μ l 2 x proteinase K buffer. Incubate the slurry at 80 °C on a thermoshaker for 10 min, then centrifuge the beads at 1500 g for 1 min.
- 4) Remove and save the supernatant and repeat the RNA elution with 100 μ l of 1 x proteinase K buffer.
- 5) Pool the two eluted fractions and centrifuge at 20 000 g for 1 min to remove all beads.
- 6) Then treat the eluted RNA samples with proteinase K: Add 1 mg ml⁻¹ proteinase K to the solutions and incubate at 50 °C for 1 h.
- 7) Phenolize the samples: Add 200 μ l phenol-chloroform-isoamylalcohol into the tubes and incubate by strong shaking for 10 min at RT.
- 8) Centrifuge the samples at 20 000 g for 10 min. Transfer the aqueous phase into fresh tubes and ethanol-precipitate the RNAs in the presence of 20 μ g glycogen. Store the samples at -20 °C for 1 h.
- 9) Collect the precipitated RNAs by a 30-min centrifugation at 20 000 g at 4 °C, wash the pellets in cold 70% ethanol, and dry and dissolve them in a small volume (10 μ l) of RNA gel sample buffer.
- 10) Separate the RNAs by denaturing PAGE and visualize them by silver staining (for low amounts of RNAs, use Northern blotting with specific DNA oligonucleotide probes).

Notes

- 1) How much of the oligonucleotides and of SAg beads should be used for affinity selections?
- 2) It is advisable to first do small-scale selections and to optimize the ratio of oligonucleotide to beads and extract. Also, vary the amount of oligonucleotide and determine the efficiency of selections (for example, by primer extension). The minimal amount of SAg that can be handled is 10–20 μ l of packed beads. The capacity for binding free biotin and biotinylated molecules is indicated on the product's data sheet. Do not use excess amounts of SAg since this might cause background problems.
- 3) How can one prove that the selected proteins are specific for the targeted RNP?
- 4) It is important to include a mock selection with an unrelated oligonucleotide (or without oligonucleotide) to be able to distinguish between specifically bound and contaminating proteins. When the background is too high, the stringency of washing can be increased by raising the salt concentration or by increasing the non-ionic detergent concentration in the washing buffer (HS-WB; NP-40, or Triton X-100 up to 1%). The oligonucleotide itself or the antisense oligonucleotide–RNA hybrid can bind proteins in the extract as well. Higher oligonucleotide amounts may lead to extensive contamination with various nucleic acid binding proteins. To exclude this, the extract may be

treated with an RNase before affinity selection to remove the target RNA (as a negative control). It is very useful to select the same particle with alternative antisense oligonucleotides complementary to different regions of the targeted RNA and to compare the resulting protein patterns.

- 5) If the affinity purification has to be performed on a large scale, it is sometimes easier to do several small-scale purifications in parallel, instead of scaling up to a single large-scale purification. Furthermore, the background of selection is strongly increased when more than 100 μ l of packed SAg beads are used in the same tube.
- 6) It is possible to purify different RNPs from the same extract sequentially [41].
- 7) We have used DEAE chromatography to enrich the total cell extract for spliceosomal snRNPs. However, DEAE chromatography imposes some stringency on the RNA–protein interactions, and many RNPs are reduced to stable core complexes during the high-salt elution step (>300 mM salt).

42.3.1.4 Elution of Affinity-Selected RNPs for Functional Studies by a Displacement Oligonucleotide

Three different strategies have been developed to release affinity-selected RNP particles under native conditions for functional studies: (i) elution with free biotin from immobilized monomeric avidin [42, 43], (ii) selection by a short biotinylated 2'OMe oligonucleotide in high-salt buffer at 4 °C and release at 37 °C in low-salt buffer [41], and (iii) use of a chimeric 2'OMe RNA/DNA oligonucleotide partially complementary to the RNA target for selection and a displacement oligonucleotide for elution [28, 29, 32]. This displacement strategy, which is generally applicable, is presented in detail below.

The affinity oligonucleotide is made of 2'OMe RNA (at least 10 nucleotides in length), perfectly complementary to the target RNA, and extended with a DNA tail (any length), which is not complementary. The DNA tail can be positioned 3' or 5' relative to the 2'OMe RNA part of the oligonucleotide, depending on which orientation is optimal in terms of avoiding steric hindrance effects.

Through a biotin residue, usually placed within the DNA portion, this chimeric oligonucleotide can be immobilized to an affinity matrix. Cell extracts are incubated with the chimeric oligonucleotide as described earlier (see Section 42.3.1.3). Following binding and extensive washing, the selected RNPs are eluted from the affinity matrix by addition of the so-called displacement oligonucleotide, complementary to the affinity oligonucleotide over its entire length. Since the displacement oligonucleotide can form a thermodynamically more stable duplex with the affinity oligonucleotide than the target RNA, the entire RNP is released from the affinity matrix (for a schematic, see Figure 42.2a). Efficient elution requires at least a twofold molar excess [28] of the displacement oligonucleotide over the affinity oligonucleotide. In some cases, displacement works without the DNA extension of the affinity oligonucleotide [30, 31].

Example: Affinity Selection and Release of Functional *T. brucei* U1 snRNPs

In this case, crude cell lysates have been used directly for affinity purification without prior enrichment of snRNPs. First, the biotinylated chimeric 2'OMe

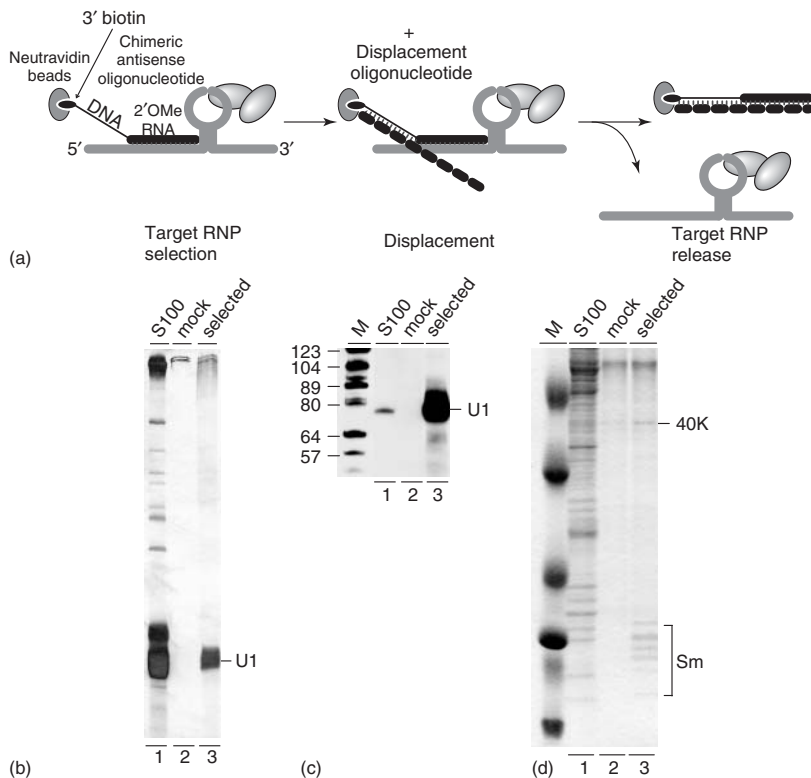


Figure 42.2 Affinity selection of *T. brucei* U1 snRNP by the displacement strategy. (a) Schematic illustration of the displacement strategy (target RNP selection, displacement, and target RNP release). For affinity selection, a specific antisense 2'OMe RNA oligonucleotide is used, which is extended by an unpaired 3'-terminal DNA tail. This oligonucleotide is immobilized through its biotin group to an affinity matrix (e.g., Neutravidin beads) and used for selection of the target RNP. Binding of the chimeric oligonucleotide to the target RNP is reversed by a displacement DNA oligonucleotide, which pairs with the chimeric oligonucleotide over its entire length. Thereby, the affinity oligonucleotide is displaced from the target to release the selected RNP from the affinity matrix. (b–d) Following this strategy, the U1 snRNP was purified from 50 ml of *T. brucei* S100 extract (lanes 1, 20 μ l samples) by affinity selection with Neutravidin beads and a biotinylated chimeric 2'OMe RNA/DNA oligonucleotide directed against

the 5'-terminal sequence of U1 snRNA. The affinity-selected material was released by an antisense displacement oligonucleotide (lanes 3), or, as a mock control, by an unrelated oligonucleotide (lanes 2). Following affinity purification, RNA, and protein components were analyzed separately. (b) The detection of RNA by silver staining. The position of U1 snRNA is indicated on the right. Note that the *T. brucei* U1 snRNA ((b), lane 3 runs in the tRNA region of total RNA on the gel. In (c) the *T. brucei* U1 snRNA was identified by Northern blot analysis, using a U1-specific digoxigenin (DIG)-labeled oligonucleotide probe. M, DIG-labeled DNA marker V (Roche 85598025; sizes of marker fragments: 57, 64, 80, 89, 104, and 123 nt). In (d), the affinity-selected proteins were detected by Coomassie staining. The Sm proteins and a U1-specific 40 kDa protein are labeled on the right [44]. M, Rainbow marker (Amersham Life Science RPN 755; sizes of marker proteins: 6.5, 14.3, 21.5, 30, 46 kDa).

RNA/DNA oligonucleotide was bound to its target RNPs within the extract, followed by immobilizing the formed complexes on the affinity matrix and release with the displacement oligonucleotide. Neutravidin agarose beads were used instead of streptavidin because of their lower non-specific binding properties in crude cell lysates. In our specific example, the affinity oligonucleotide was a chimeric 2'OMe RNA/DNA oligonucleotide (Tb-U1-5' chimeric) biotinylated at the 3' end [32]. The displacement oligonucleotide was "Tb-U1 displace" [32], complementary to the Tb-U1-5' chimeric oligonucleotide. As a mock control, the displacement oligonucleotide was replaced with another oligonucleotide of unrelated sequence (see Figure 42.2b–d).

Procedure

- 1) For preparative-scale U1 snRNP affinity selection, briefly centrifuge (10 000 g, 10 min) 50 ml of *T. brucei* total cell lysate (corresponding to $\sim 1 \times 10^{11}$ cells) to remove aggregates. Transfer the supernatant (10 ml aliquots) to five fresh tubes.
- 2) Dissolve 15 μ g biotinylated chimeric oligonucleotide in 500 μ l HS-WB. Add a 100 μ l aliquot of this solution to each aliquot of cell extract.
- 3) Incubate and slowly rotate the tubes at 30 °C for 1 h.
- 4) Transfer 100 μ l aliquots of a Neutravidin agarose suspension to five 1.5 ml Eppendorf tubes. Collect the beads by a short spin (1500 g, 1 min).
- 5) Preblock the NAg beads as described earlier (see Section 42.3.1.2).
- 6) Wash the beads three times in HS-WB by resuspending 50 μ l packed beads in 0.5 ml wash buffer.
- 7) Spin down the extract after the 1 h incubation (step 3) to remove the aggregates (10 min at 10 000 g). Collect the supernatant in new tubes and place them on ice.
- 8) Transfer each aliquot of the blocked and washed NAg beads to an aliquot of cell extract by using a small volume of the extract (~ 500 μ l) for the transfer. The mixture is slowly rotated at 4 °C for 2 h.
- 9) Collect the NAg beads by a short centrifugation (1500 g, 1 min). Remove and save the supernatant (for checking the efficiency of U1 RNP depletion). Transfer the pelleted NAg beads into 1.5 ml Eppendorf tubes using 0.5 ml HS-WB.
- 10) Incubate the beads in HS-WB for 5 min by slowly rotating at 4 °C.
- 11) Spin down the beads (1500 g, 1 min) and remove the supernatant.
- 12) Repeat washing the beads in 0.5 ml HS-WB three times (steps 10 and 11).
- 13) Resuspend the 50 μ l packed NAg in each tube in 100 μ l HS-WB and collect the five suspensions in a single Eppendorf tube. Spin down the beads again. The packed volume is now 250 μ l in total.
- 14) Add 25 μ g displacement oligonucleotide to 250 μ l HS-WB and mix the solution with the pelleted NAg beads by slowly rotating on a turning wheel for 30 min at 30 °C.

- 15) Collect the beads by centrifugation (1500 g, 1 min) and transfer the supernatant (containing the eluted U1 snRNP) to a new Eppendorf tube.
- 16) Repeat the elution step once: Resuspend the beads in 250 µl HS-WB containing 25 µg of the displacement oligonucleotide, and incubate the suspension at 30 °C for 30 min by slowly rotating the tube on a wheel.
- 17) Spin down the beads (1500 g, 1 min) and combine the supernatant (= second eluate) with the first eluate.
- 18) Spin down the pooled fractions (20 000 g, 1 min) to get rid of all beads. The pooled fractions containing the eluted U1 snRNP particles can be used directly for functional or structural studies.

Notes

- 1) The salt concentration of the cell lysate and the HS-WB used for RNP purification should be adjusted according to the stability of the targeted particle. At least core snRNP particles are usually stable under high-salt conditions.
- 2) In some cases, the presence of large amounts of displacement oligonucleotide may interfere with the functional assay (for example, by interacting with a 5' splice site to which a purified U1 snRNP is intended to bind). In such cases, the displacement oligonucleotide can be depleted from the eluted material through a biotinylated DNA oligonucleotide complementary to the displacement oligonucleotide ("anti-displacement oligo"). This step is best done on a small NAg column containing the immobilized antidisplacement oligonucleotide in excess to the displacement oligonucleotide [32].
- 3) Under the relatively mild conditions of the affinity purification by displacement, the integrity of the RNP particles should be preserved; only some loosely associated protein components might be lost. To check for integrity of a particle after affinity purification, the eluted samples should be analyzed by native gel electrophoresis or glycerol gradient centrifugation.

42.3.2

Affinity Purification of Specific RNA Binding Proteins by Biotinylated RNAs

In this section, we present a fast and simple one-step affinity purification method for the isolation of specific RNA binding proteins. A biotinylated RNA with a putative protein binding sequence is immobilized on an affinity matrix and incubated with a crude cell lysate. Following extensive washing, the specifically bound proteins are recovered from the affinity matrix by elution with urea (for a schematic, see Figure 42.3a). For RNA–protein complex formation, we recommend to use HPLC-purified, 5'- or 3'-biotinylated RNA oligonucleotides or internally biotinylated transcripts, which can be generated by *in vitro* transcription with T7 or SP6 polymerase in the presence of biotinylated nucleotides (for example, biotin-16-UTP from Roche). For internal biotinylation during *in vitro* transcription, one has to find empirically the optimal incorporation level, since too many internal

biotin residues may introduce steric hindrance, resulting in lower protein binding capacity.

Example: Isolation of a CA-Repeat RNA Binding Factor from HeLa Cell Nuclear Extract using a Biotinylated RNA Oligonucleotide

In this case, we have used a 5'-biotinylated RNA oligonucleotide, 5'-biotin-(CA)₃₂, which was first immobilized on streptavidin agarose beads [34]. As a mock control, the affinity selection was performed in the absence of bound RNA oligonucleotide. RNA-protein complexes were formed during a 1 h incubation at 30 °C. Then the non-specifically bound proteins were removed by washing in HS-WB with progressively increasing KCl concentration (up to 2 M). Affinity-selected proteins were eluted with 6 M urea and analyzed by SDS-PAGE (see Figure 42.3b).

Procedure

- 1) Prepare a crude nuclear extract from HeLa cells in the presence of protease inhibitors [35].
- 2) Add 20 µl preblocked SAg beads (1 : 1 slurry in HS-WB) to a 1.5 ml Eppendorf tube.
- 3) Add 200 µl HS-WB containing 6 µg of the 5'-biotinylated RNA oligonucleotide to the SAg suspension. Bind the oligonucleotide during a 5 h incubation at 4 °C with rotation.
- 4) Collect the SAg beads by centrifugation (1500 g, 1 min), discard the supernatant.
- 5) Resuspend the pelleted beads in 1 ml HS-WB containing 400 mM KCl and incubate the sample for 5 min at 4 °C with rotation.
- 6) Spin down the beads (1500 g, 1 min) and repeat washing in 1 ml HS-WB three times (steps 4–5).
- 7) Incubate the SAg beads with 150 µl HeLa cell nuclear extract (precleared by a 1 min centrifugation at 20 000 g) at 30 °C for 1 h by slow rotation.
- 8) Collect the SAg beads by a short spin (1500 g, 1 min). Remove the supernatant.
- 9) Add 1 ml HS-WB containing 400 mM KCl to the pelleted beads and rotate the sample for 5 min at 4 °C.
- 10) Centrifuge the SAg suspension at 1500 g for 1 min. Remove the supernatant.
- 11) Repeat steps 9–10 four times with HS-WB containing increasing amounts of KCl (0.6, 0.8, 1.2, and 2 M) to reduce background.
- 12) Elute the specifically bound proteins with 15 µl of 6 M urea, as described (see Section 42.3.1.3: Analysis of the affinity-selected material for protein components).
- 13) Collect washing fractions as well as the eluate and use 5 µl of each for analysis by SDS-PAGE.

Notes

Binding affinities of RNA-protein interactions vary within a wide range, and RNA-protein complexes may be disrupted during conventional purification

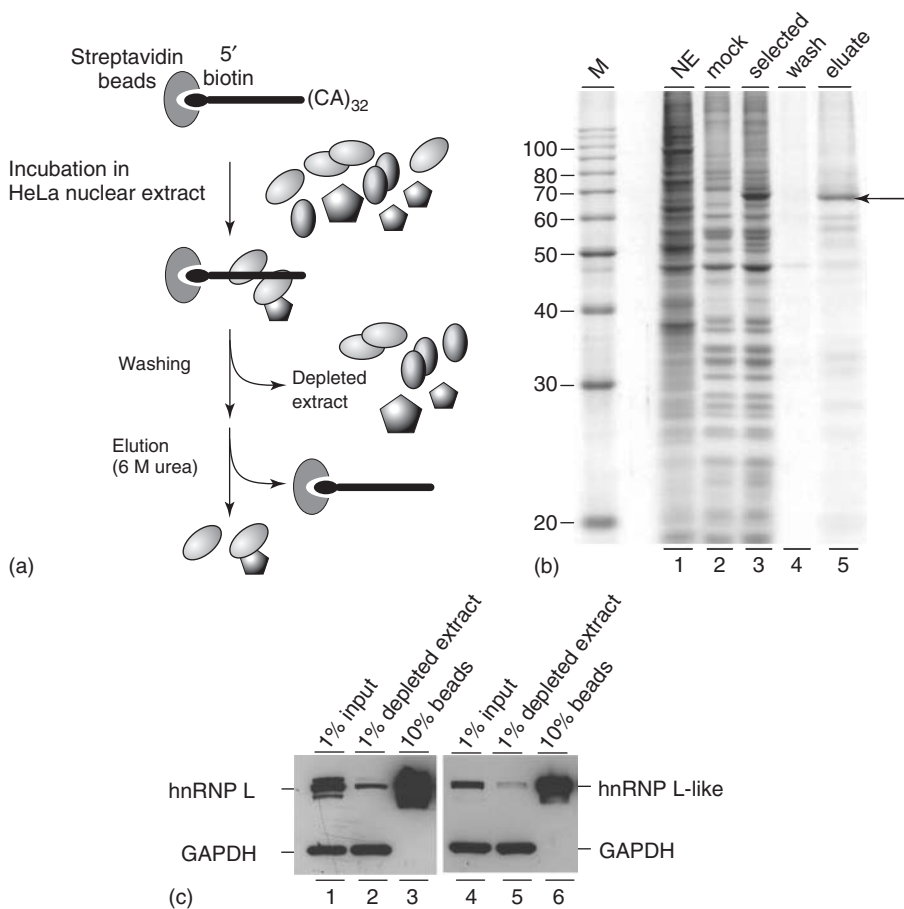


Figure 42.3 Affinity purification and depletion of CA-repeat RNA binding proteins. (a) Schematic representation of the affinity purification and depletion strategy. 5'-biotinylated (CA)₃₂ RNA is prebound on SAg beads and incubated with HeLa cell nuclear extract to deplete the extract of specific RNA binding proteins, in this example, hnRNP L and hnRNP L-like. Stringent washing removes unbound and loosely bound material, and tightly bound proteins are eluted from the matrix under denaturing conditions (6 M urea). (b) The protein composition of total nuclear extract (lane 1), mock-selected material (lane 2), (CA)₃₂ RNA affinity-selected material (lane 3), a 2 M KCl wash fraction after affinity purification (lane 4), and the 6 M urea eluate (lane 5) was analyzed by 12.5% SDS-PAGE and Coomassie

staining. The arrow points to the specifically selected 65 kDa CA-repeat RNA binding protein hnRNP L [34]. Marker proteins and their sizes (in kilodaltons) are shown on the left. (c) Specific depletion of CA-repeat RNA binding proteins hnRNP L and L-like. HeLa nuclear extract were depleted of hnRNP L and L-like by incubation with biotinylated (CA)₃₂ RNA, followed by Neutravidin agarose selection. The extent of depletion was investigated by Western blot analysis using anti-hnRNP L (lanes 1–3) and anti-hnRNP L-like antibodies (lanes 4–6). Total HeLa nuclear extract (1% input; lanes 1, 4), depleted extract (1%; lanes 2, 5), and 10% of the material bound to the beads (lanes 3, 6) were loaded onto the gel, with GAPDH serving as loading control.

procedures. A critical parameter is therefore the salt concentration of the binding/washing buffers. Make sure that the composition of these buffers is compatible with the application. For example, high-salt concentrations can lead to disruption of protein binding; on the other hand, low-salt conditions may increase background binding of proteins.

42.3.3

Depletion of Nuclear Extract with Biotinylated RNA

Functional testing of a particular RNA binding protein *in vitro* often requires to deplete the protein from extracts. If a high-affinity protein binding motif is known, this can be done efficiently by using a biotinylated RNA carrying this binding motif. For example, such depleted nuclear extracts were successfully applied for studying the function of hairpin binding factor (HBF) and stem-loop binding protein (SLBP) in histone pre-mRNA processing [45, 46], or the involvement of *trans*-acting protein factors (p54nrb, PTB, PSF, and U1A) in alternative polyadenylation of COX-2 mRNA [47].

Example: Depleting HeLa Cell Nuclear Extracts of hnRNP L and hnRNP L-Like with Biotinylated CA-Repeat RNAs

The endogenous hnRNP L and hnRNP L-like proteins were removed from HeLa nuclear extracts with a 5'-biotinylated RNA containing a CA-repeat sequence, 5'-biotin-(CA)₃₂, immobilized on Neutravidin agarose beads (Pierce; Ref. [34]). A control extract was incubated in parallel with agarose beads only. The depletion was done for 30 min at 27°C, and unspecific binding of other proteins was suppressed by increasing the KCl concentration up to 600 mM. Depletion was verified by Western blotting with specific anti-hnRNP L (4D11) and anti-hnRNP L-like antibodies (Ref. [48]; Figure 42.3c).

Procedure

- 1) Take 100 µl of packed Neutravidin agarose beads and block them in SAgBB for 2 h at 4°C followed by washing with WB400.
- 2) Incubate 30 µl preblocked Neutravidin beads (1 : 1 slurry in WB400) with 6 µg 5'-biotin-(CA)₃₂ RNA in 200 µl WB400 at 4°C overnight with rotation. Include mock incubation (Neutravidin agarose beads without RNA).
- 3) Wash the beads four times with WB400 and once with Buffer D and collect the Neutravidin agarose beads by centrifugation (1500 g; 1 min). Discard the supernatant.
- 4) Add 200 µl HeLa cell nuclear extract to the washed beads and incubate at 27°C for 30 min with rotation.
- 5) Increase KCl up to 600 mM and continue incubation for 20 min at 4°C with slow rotation.
- 6) Spin down the beads (1500 g, 1 min) and save the supernatant (depleted nuclear extract).

- 7) Dialyze the depleted HeLa nuclear extract against buffer D for 2 h at 4 °C.
- 8) The centrifuge-depleted HeLa nuclear extract (20 000 g, 1 min) is aliquoted, shock-frozen in liquid nitrogen, and kept at –80 °C.
- 9) Control the efficiency of depletion of hnRNP L and hnRNP L-like proteins by Western blot with specific antibodies.

42.4

Troubleshooting

The biotin-based affinity selection methods allow the rapid purification of RNA–protein complexes for protein identification or structural and functional assays. An intrinsic problem using biotinylated 2'OMe RNA oligonucleotides is that the target RNP might not be sufficiently exposed to the affinity oligonucleotide to allow efficient binding. Another obvious limitation to quantitative RNP affinity selection is that RNP particles might be heterogenous in nature such that only certain subtypes of complexes will be accessible to oligonucleotide hybridization.

To optimize the experimental setup for efficient affinity purification, we refer to the Notes under Sections 42.3.1.3, 42.3.1.4, and 42.3.2. In the following section, we discuss a few general checkpoints and give some additional recommendations.

42.4.1

Biotinylated 2'OMe RNA Oligonucleotides

- **Purity:** The synthesized biotinylated RNA oligonucleotide should be purified to avoid non-specific binding.
- **Length:** The oligonucleotide should be long enough for stable binding to the target RNA and should target a sequence as unique as possible.
- **Sequence:** The oligonucleotide sequence should be complementary to regions of the target RNA that are devoid of stable secondary structures and that are not protected by proteins.
- **Biotinylation:** The number and location of biotin residues can be variable, but the biotin groups should be accessible for binding to affinity matrices.

42.4.2

Extracts and Buffers

To avoid degradation of target RNA–protein complexes, the cell lysates should be prepared in the presence of RNase and protease inhibitors. The composition of all buffers should be adjusted according to the specific requirements for the formation of the RNA–protein complex of interest.

42.4.3

Optimization of the Experimental Conditions, When Yields Are Low

- Ensure that the affinity oligonucleotide is biotinylated (see Chapter 6), or that the internally biotinylated RNA used for selection of specific RNA binding proteins does not contain too many biotin residues that may impair protein binding.

- The affinity oligonucleotide might be inadequate. Try affinity purification with another biotinylated oligonucleotide complementary to a different region of the target RNA.
- It is useful to test binding/washing buffers with different stringencies to determine the optimal purification strategy for the target RNP.
- Addition of appropriate cofactors (such as Mg^{2+}) may support RNA–protein interactions.
- Check the cell lysate for the presence of the RNA–protein complex of interest (for example, by primer extension on the RNA; by Western blotting for the proteins). Ensure that the target RNP is not degraded in the extract. Add specific inhibitors to all buffers (proteins: protease inhibitors such as PMSF, Leupeptin, Aprotinin; RNA: RNase inhibitors such as RNasin; phosphorylated proteins: phosphatase inhibitors such as sodium fluoride).
- Make sure that the analysis method for RNAs and proteins is adequate. Use more sensitive methods such as primer extension for RNAs, and silver staining or Western blotting for proteins.

42.4.4

Optimization of the Experimental Conditions in the Case of High Background

- Check for the specificity of the RNA oligonucleotide used for affinity purification.
- Increase the stringency of binding/washing buffers by increasing the salt and/or detergent concentration.
- Add blocking reagents (such as BSA or tRNA) to binding reactions, which may help to reduce background binding.
- Do not use unnecessarily large amounts of beads and oligonucleotide for affinity selections.

References

1. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, **24**, 218–229.
2. Nilsson, J., Stahl, S., Lundeberg, J., Uhlén, M., and Nygren, P.-A. (1997) Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Expr. Purif.*, **11**, 1–16.
3. Srisawat, C. and Engelke, D.R. (2001) Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA*, **7**, 632–641.
4. Srisawat, C. and Engelke, D.R. (2002) RNA affinity tags for purification of RNAs and ribonucleoprotein complexes. *Methods*, **26**, 156–161.
5. Hartmuth, K., Urlaub, H., Vornlocher, H.-P., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16719–16724.
6. Behzadnia, N., Hartmuth, K., Will, C.L., and Lührmann, R. (2006) Functional spliceosomal A complexes can be assembled *in vitro* in the absence of a penta-snRNP. *RNA*, **12**, 1738–1746.

7. Das, R., Zhou, Z., and Reed, R. (2000) Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol. Cell.*, **5**, 779–787.
8. Zhou, Z., Sim, J., Griffith, J., and Reed, R. (2002) Purification and electron microscopic visualization of functional human spliceosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12203–12207.
9. Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419**, 182–185.
10. Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.*, **26**, 5528–5543.
11. Hogg, J.R. and Collins, K. (2007) RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. *RNA*, **13**, 868–880.
12. Kroiss, M., Brünner, K.M., Wiesner, J., Grimm, M., Sickmann, A., and Fischer, U. (2009) Native purification of protein and RNA-protein complexes using a novel affinity procedure. *Fly*, **3**, 221–228.
13. Grabowski, P.J. and Sharp, P.A. (1986) Affinity chromatography of splicing complexes: U2, U5, and U4 + U6 small nuclear ribonucleoprotein particles in the spliceosome. *Science*, **233**, 1294–1299.
14. Bindereif, A. and Green, M.R. (1987) An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *EMBO J.*, **6**, 2415–2424.
15. Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res.*, **12**, 1231–1245.
16. Ryder, U., Sproat, B.S., and Lamond, A.I. (1990) Sequence-specific affinity selection of mammalian splicing complexes. *Nucleic Acids Res.*, **18**, 7373–7379.
17. Barabino, S.M., Sproat, B.S., Ryder, U., Blencowe, B.J., and Lamond, A.I. (1989) Mapping U2 snRNP-pre-mRNA interactions using biotinylated oligonucleotides made of 2'-OMe RNA. *EMBO J.*, **8**, 4171–4178.
18. Blencowe, B.J., Sproat, B.S., Ryder, U., Barabino, S., and Lamond, A.I. (1989) Antisense probing of the human U4/U6 snRNP with biotinylated 2'-OMe RNA oligonucleotides. *Cell*, **59**, 531–539.
19. Blencowe, B.J. and Lamond, A.I. (1999) in *RNA-Protein Interactions*, Methods in Molecular Biology, Vol. 118 (ed. S. Haynes), Humana Press Inc., Totowa, NJ, pp. 275–287.
20. Sproat, B.S., Lamond, A.I., Beijer, B., Neuner, P., and Ryder, U. (1989) Highly efficient chemical synthesis of 2'-O-methyloligoribonucleotides and tetrabiotinylated derivatives; novel probes that are resistant to degradation by RNA or DNA specific nucleases. *Nucleic Acids Res.*, **17**, 3373–3386.
21. Iribarren, A.M., Sproat, B.S., Neuner, P., Sulston, I., Ryder, U., and Lamond, A.I. (1990) 2'-O-alkyl oligoribonucleotides as antisense probes. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 7747–7751.
22. Majlessi, M., Nelson, N.C., and Becker, M.M. (1998) Advantages of 2'-O-methyl oligoribonucleotide probes for detecting RNA targets. *Nucleic Acids Res.*, **26**, 2224–2229.
23. Wassarman, D.A. and Steitz, J.A. (1991) Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol. Cell. Biol.*, **11**, 3432–3445.
24. Palfi, Z., Günzl, A., Cross, M., and Bindereif, A. (1991) Affinity purification of *Trypanosoma brucei* small nuclear ribonucleoproteins reveals common and specific protein components. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 9097–9101.
25. Lee, B., Matera, G., Ward, D.C., and Craft, J. (1996) Association of RNase mitochondrial RNA processing enzyme with ribonuclease P in higher ordered structures in the nucleolus: a possible coordinate role in ribosome biogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 11471–11476.
26. Will, C.L., Schneider, C., Reed, R., and Lührmann, R. (1999) Identification of both shared and distinct proteins in the

- major and minor spliceosomes. *Science*, **284**, 2003–2005.
27. Pillai, R.S., Will, C.L., Lührmann, R., Schümperli, D., and Müller, B. (2001) Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein. *EMBO J.*, **20**, 5470–5479.
 28. Lingner, J. and Cech, T.R. (1996) Purification of telomerase from *Euplotes aediculatus*: requirement of a primer 3' overhang. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 10712–10717.
 29. Schnapp, G., Rodi, H.-P., Rettig, W.J., Schnapp, A., and Damm, K. (1998) One-step affinity purification protocol for human telomerase. *Nucleic Acids Res.*, **26**, 3311–3313.
 30. Fong, Y.W. and Zhou, Q. (2001) Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, **414**, 929–933.
 31. Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001) The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature*, **414**, 317–322.
 32. Palfi, Z., Lane, W.S., and Bindereif, A. (2002) Biochemical and functional characterization of the *cis*-spliceosomal U1 small nuclear RNP from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **121**, 233–243.
 33. Reed, R. (1990) Protein composition of mammalian spliceosomes assembled *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 8031–8035.
 34. Hui, J., Stangl, K., Lane, W.S., and Bindereif, A. (2003) HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat. Struct. Biol.*, **10**, 33–37.
 35. Günzl, A., Palfi, Z., and Bindereif, A. (2002) Analysis of RNA-protein complexes by oligonucleotide-targeted RNase H digestion. *Methods*, **26**, 162–169.
 36. Lamm, G.M., Blencowe, B.J., Sproat, B.S., Iribarren, A.M., Ryder, U., and Lamond, A.I. (1991) Antisense probes containing 2-aminoadenosine allow efficient depletion of U5 snRNP from HeLa splicing extracts. *Nucleic Acids Res.*, **19**, 3193–3198.
 37. Cross, M., Günzl, A., Palfi, Z., and Bindereif, A. (1991) Analysis of small nuclear ribonucleoproteins (RNPs) in *Trypanosoma brucei*: structural organization and protein components of the spliced leader RNP. *Mol. Cell. Biol.*, **11**, 5516–5526.
 38. Dignam, J.D., Lebowitz, R.M., and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.
 39. Lee, K.A., Bindereif, A., and Green, M.R. (1988) A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. *Gene Anal. Tech.*, **5**, 22–31.
 40. Wersig, C. and Bindereif, A. (1992) Reconstitution of functional mammalian U4 small nuclear ribonucleoprotein: Sm protein binding is not essential for splicing *in vitro*. *Mol. Cell. Biol.*, **12**, 1460–1468.
 41. Smith, H.O., Tabiti, K., Schaffner, G., Soldati, D., Albrecht, U., and Birnstiel, M.L. (1991) Two-step affinity purification of U7 small nuclear ribonucleoprotein particles using complementary biotinylated 2'-O-methyl oligoribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 9784–9788.
 42. Glover, B.P. and McHenry, C.S. (2001) The DNA polymerase III holoenzyme: an asymmetric dimeric replicative complex with leading and lagging strand polymerases. *Cell*, **105**, 925–934.
 43. Kim, D.R. and McHenry, C.S. (1996) Identification of the beta-binding domain of the alpha subunit of *Escherichia coli* polymerase III holoenzyme. *J. Biol. Chem.*, **271**, 20690–20698.
 44. Palfi, Z., Schimanski, B., Günzl, A., Lücke, S., and Bindereif, A. (2005) U1 small nuclear RNP from *Trypanosoma brucei*: a minimal U1 snRNA with unusual protein components. *Nucleic Acids Res.*, **33**, 2493–2503.
 45. Dominski, Z., Sumerel, J., Hanson, R.J., and Marzluff, W.F. (1995) The polyribosomal protein bound to the 3' end of histone mRNA can function in

- histone pre-mRNA processing. *RNA*, **1**, 915–923.
46. Dominski, Z., Zheng, L.X., Sanchez, R., and Marzluff, W.F. (1999) Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. *Mol. Cell. Biol.*, **19**, 3561–3570.
47. Hall-Pogar, T., Liang, S., Hague, L.K., and Lutz, C.S. (2007) Specific *trans*-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR. *RNA*, **13**, 1103–1115.
48. Hung, L.H., Heiner, M., Hui, J., Schreiner, S., Benes, V., and Bindereif, A. (2008) Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA*, **14**, 284–296.

43

Affinity Purification of Spliceosomal and Small Nuclear Ribonucleoprotein Complexes

Julia Dannenberg, Patrizia Fabrizio, Cindy L. Will, and Reinhard Lührmann

43.1

Introduction

The ability to isolate ribonucleoprotein (RNP) complexes under native conditions is often a prerequisite for subsequent functional and ultrastructural studies. A number of affinity-selection techniques have been employed to purify human and yeast snRNPs (small nuclear ribonucleoproteins) and/or spliceosomal complexes at distinct stages of assembly from extracts. These techniques include immunoaffinity chromatography [1, 2], affinity selection using biotinylated antisense oligonucleotides [3], selection of pre-mRNAs with randomly incorporated biotinylated nucleotides [4, 5], or binding of aptamer-tagged pre-mRNA by either viral MS2 protein fused to the maltose-binding protein (MBP) [6–11] or tobramycin [12]. Immunoaffinity chromatography with anti-peptide antibodies has proved to be highly effective for the isolation of native snRNPs and spliceosomal complexes. The RNP complexes of interest are bound by an immobilized antibody against a protein component of the complex and the matrix is washed to remove contaminating, unbound material. Highly purified complexes are then eluted under native conditions by the addition of an excess of the cognate peptide. The procedure is exemplified here for the purification of the human 17S U2 snRNP [1]. Currently, the best method for purification of spliceosomal complexes is affinity selection using a molecular tag on the pre-mRNA substrate. This affinity-purification procedure entails the use of a pre-mRNA tagged with three MS2 RNA aptamers. This RNA is incubated with the MS2–MBP fusion protein, which interacts (i) with the pre-mRNA by binding strongly to the MS2 hairpins and (ii) with an amylose affinity matrix through the MBP portion of the protein. The latter interaction is fully reversible by competition with maltose, thus allowing the elution of highly purified complexes under native conditions. The procedure is exemplified here for the purification of spliceosomal B, B^{act}, and C complexes from the yeast *Saccharomyces cerevisiae* (see under Section 43.3).

43.2

Immunoaffinity Purification

43.2.1

Generation of Antipeptide Antibodies: Peptide Selection Criteria

To choose a peptide for the production of antipeptide antibodies, several criteria should be taken into consideration. First, the peptide should be 15–18 amino acids in length and should not contain an internal cysteine (see below). When choosing the sequence of the peptide, conserved motifs (e.g., RNA recognition motifs (RRMs), zinc fingers, etc.) and regions rich in a particular amino acid should be avoided (e.g., leucine-rich, proline-rich, etc.) as the generated antibodies will likely crossreact with many other proteins containing similar motifs. Likewise, regions known to interact with other proteins or RNA in your RNP complexes should be avoided, as they likely will not be accessible for interaction with an antibody. Optimal regions are those found on the surface of the protein. If the atomic structure of your protein is known, then surface amino acids can be chosen with some certainty. An alternative approach is to use a computer program (e.g., Protean from DNA Star) that provides information about amino acid hydrophobicity, and thus the probability that a stretch of amino acids lies on the surface of your protein. These programs also provide information about the antigenicity of a particular region. Typically, regions containing prolines have a high antigenicity index and the presence of one or two internal prolines is thus desirable. Amino acids on the surface of a protein are typically charged and thus most candidate peptides contain multiple charged residues. However, one should avoid peptides that are extremely positively or negatively charged, as they may be difficult to synthesize and/or purify. The termini of a protein, in particular the C-terminus, often prove to be very effective for immunization purposes, providing they meet the criteria described above. After selecting the amino acid sequence, a cysteine residue should be added to the N or C-terminus. Before immunization, the peptide is normally coupled to a carrier protein such as KLH (keyhole limpet hemocyanin) or ovalbumin (Protocol 1) *via* the cysteine's sulfhydryl group using, for example, m-Maleimidobenzoyl-N-hydroxysuccinimide ester (MBS). If the chosen peptide sequence is more C-terminal in your protein, the cysteine should be added to the N-terminus of your peptide and vice versa. As the success rate for the production of antibodies suitable for immunoaffinity purification is only moderate, it is advisable to carry out immunizations with more than one peptide sequence.

Protocol 1: Coupling of Peptides to Ovalbumin with MBS

- Phosphate-buffered saline (PBS) (pH 7.0): 20 mM potassium phosphate (pH 7.0) and 130 mM NaCl
- Protein sample buffer: 60 mM Tris-HCl (pH 6.8), 1 mM EDTA, 16% glycerol, 2.0% SDS (w/v), 0.1% Bromophenol Blue, and 50 mM DTE. Add DTE directly before use.

- 1) Dissolve 25 mg (sufficient to couple six peptides) of ovalbumin (Sigma) in 3.75 ml PBS (pH 7.0) and pipette 3.0 ml into a 12 ml plastic, capped tube.
- 2) Directly before use, dissolve 50 mg of fresh MBS (Pierce) in 1.0 ml of dimethyl sulfoxide (DMSO). Combine 20 μ l of the MBS solution with 280 μ l DMSO and add immediately to the ovalbumin solution. Incubate for 30 min at room temperature.
- 3) Apply the ovalbumin solution to a PD10 column (Amersham Biosciences) equilibrated with PBS (pH 7.0). Collect ten 1.0 ml fractions and perform a Bradford protein concentration determination (or a similar method) with 4 μ l of each fraction to determine which contains ovalbumin. Typically, ovalbumin elutes in fractions 3–5. Combine the three most highly concentrated fractions.
- 4) Pipette 0.5 ml of the MBS-treated ovalbumin solution into a microfuge tube and add 150 μ l of peptide (20 mg ml⁻¹ in H₂O). Incubate for 30 min at room temperature.
- 5) Add 350 μ l of PBS (pH 7.0) to the peptide/ovalbumin solution and load onto a PD10 column equilibrated with PBS (pH 7.0). Collect six 1.0 ml fractions.
- 6) Combine 4 μ l of each fraction with 8 μ l protein sample buffer and heat at 85 °C for 5 min. Analyze on a 13% polyacrylamide/SDS gel and visualize by staining with Coomassie. As a control, analyze 1 μ l of the MBS-treated ovalbumin solution. If the peptide has been successfully coupled, multiple bands that run just above the MBS-treated ovalbumin will be observed. Peptide–ovalbumin conjugate typically peaks in fractions four and five. Pool the two most highly concentrated fractions and use for immunization. The end concentration of the peptide–ovalbumin conjugate solution is ~0.5–0.75 mg ml⁻¹.

Protocol 2: Affinity Purification of Anti-peptide Antibodies

- Peptide beads: 1.0 mg of peptide was covalently coupled with 1.0 ml of Sulfolink Coupling Gel exactly as described by the manufacturer (Pierce).
 - PBS (pH 8.0): 20 mM potassium phosphate (pH 8.0) and 130 mM NaCl.
- 1) Combine 8.0 ml of rabbit serum containing anti-peptide antibodies with 800 μ l 10 x PBS (pH 8.0) and filter by passing through a 0.45 μ m membrane. Wash 1.0 ml (bed volume) peptide beads twice with 10.0 ml PBS in a 15 ml capped tube, add the rabbit serum and incubate 2–3 h or overnight at 4 °C with head-over-tail rotation.
 - 2) Pellet beads by centrifuging for 1 min at 200 g (e.g., at 1000 rpm in a Heraeus Megafuge 1.0R centrifuge). Wash the beads 5 x with 10.0 ml PBS. After the last wash, transfer the beads to a plastic 10 ml column (e.g., a Poly-Prep Chromatography Column from BIO-RAD) and allow the buffer to drain from the column.
 - 3) To elute bound antibodies, apply 0.5 ml of 100 mM glycine (pH 2.5) to the column and collect the flowthrough in a microfuge tube containing 30 μ l of 1 M Tris (pH 9.5). Repeat 4 x, collecting each eluate in a separate tube.

Mix well to ensure that the pH of the eluates has been neutralized. Check for the presence of antibody in each fraction by Bradford assay (or a similar method) and pool the two most highly concentrated fractions (typically eluates 2 and 3).

- 4) Transfer the eluate to dialysis membrane (molecular weight cutoff (MWCO) 6000–8000) and dialyze at 4 °C against 1 l of PBS (pH 8.0) for 4 h, changing the buffer once after 2 h. Typically, 0.1–0.3 mg of affinity-purified antibody are recovered from 8 ml of serum, but this amount may vary considerably depending on the titer of the antibody. The activity of the eluted antibodies should subsequently be tested by Western blotting. Note that some antibodies lose significant activity when briefly exposed to extremely low pH. In those cases, an alternative method for the release of antibody from the peptide beads (such as incubating briefly with 3.5 M MgCl₂) should be used. To avoid loss of antibody activity, steps 3 and 4 should be performed as quickly as possible.
- 5) Peptide beads, which typically can be used multiple times, should be washed 3 x with PBS (pH 8.0) and stored at 4 °C in H₂O containing 0.02% sodium azide.

Example: Immunoaffinity Purification of 17S U2 snRNPs

To purify 17S U2 snRNPs from HeLa nuclear extract, antibodies were raised against a region near the C-terminus (aa 444–458; CMLRPPLPSEGPGNIP) of the 17S U2 SF3a66 protein [1]. Antibodies were affinity purified (Protocol 2), bound to Protein A Sepharose (PAS) beads, and covalently attached to the beads using dimethylpylimidate (DMP) as previously described [13]. 17S U2 snRNPs begin to dissociate at salt concentrations above 200 mM and thus HeLa nuclear extract was first dialyzed to reduce the concentration of KCl to 150 mM and then incubated with the anti-SF3a66-charged PAS beads. After extensive washing, bound snRNPs were eluted with an excess of the SF3a66 peptide used for immunization. To separate the 17S U2 snRNPs from contaminating U1 snRNPs and nuclear proteins, the eluate was fractionated on a linear glycerol gradient (Figure 43.1). U2 snRNA, as well as known 17S U2-associated proteins such as subunits of SF3b and SF3a, peaked in fractions 13–15, demonstrating the isolation of an intact U2 particle. On average, ~100 µg (100 pmol) of 17S U2 snRNPs can be recovered from a 12.5 ml HeLa nuclear extract. The anti-SF3a66-charged PAS beads can be regenerated (i.e., freed of bound peptide) and used repeatedly.

Protocol 3: Immunoaffinity Purification of 17S U2 snRNPs

- G150 Buffer: 20 mM Hepes (pH 7.9), 150 mM KCl, 1.5 mM MgCl₂, 5% glycerol, 0.5 mM DTE, 0.5 mM phenylmethylsulfonyl fluoride (PMSF).
- Anti-SF3a66 antibodies coupled to PAS: antibodies were affinity-purified as described in Protocol 2 and covalently coupled to PAS beads with DMP as previously described [13].
- HeLa nuclear extract: extract was prepared according to [14].

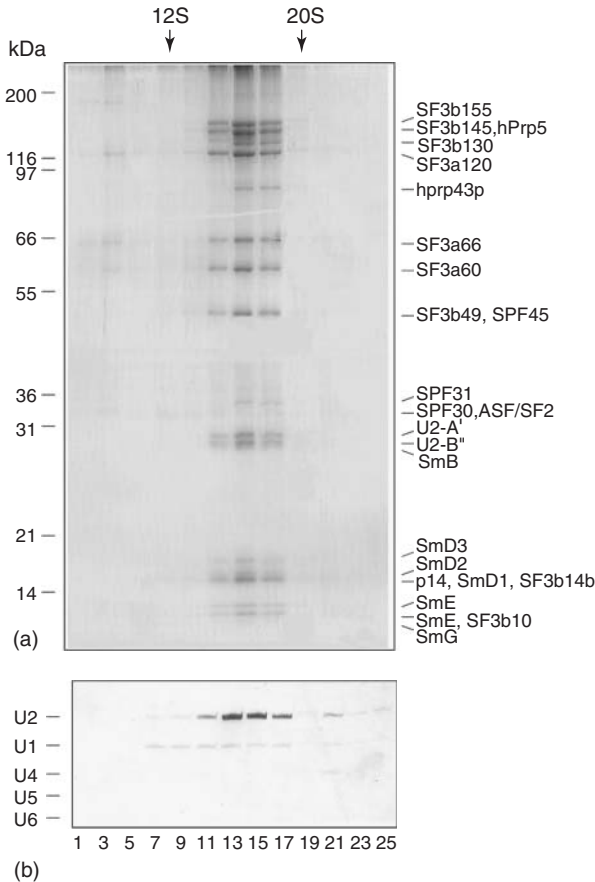


Figure 43.1 Immunoaffinity-purified 17S U2 snRNPs. Human 17S U2 snRNPs were immunoaffinity purified from HeLa nuclear extract with anti-peptide antibodies against the SF3a66 protein and subjected to 10–30% glycerol gradient centrifugation. Distribution of (a) protein and (b) snRNA across the gradient. Proteins were separated by SDS-PAGE

and visualized by staining with Coomassie. The identity of the major 17S U2 proteins is indicated on the right (a). RNA was fractionated by denaturing PAGE and stained with silver (b). Gradient fraction numbers are indicated at the bottom and the peak positions of 12S U1 and 20S U5 snRNPs, run in parallel, are indicated at the top.

- Protein sample buffer: 60 mM Tris-HCl (pH 6.8), 1 mM EDTA, 16% glycerol, 2.0% SDS (w/v), 0.1% Bromophenol Blue, and 50 mM DTE. Add DTE directly before use.
 - RNA loading buffer: 95% formamide, 0.02% Xylene cyanol, and 0.02% Bromophenol Blue.
- 1) Dialyze 12.5 ml of freshly prepared HeLa nuclear extract against 1.0 l G150 buffer for 4 h at 4°C, changing the buffer once after 2 h. Dilute with 1

volume of G150 buffer and centrifuge at 4 °C in a 30 ml corex tube at 10 000 g (e.g., 9500 rpm in a Sorvall SS34 rotor) for 10 min to pellet any precipitates.

- 2) All of the following steps should be carried out at 4 °C. Pipette 500 µl PAS coupled with 250 µg of affinity-purified anti-SF3a66 antibodies into a small glass column and wash with 10 ml of G150 buffer. Using a peristaltic pump (e.g., Pharmacia P1 pump), pass the nuclear extract over the anti-SF3a66 column at a flow rate of 1–2 ml h⁻¹ (i.e., overnight). Alternatively, combine the anti-SF3a66-charged PAS beads and nuclear extract in a 50 ml capped plastic tube and incubate overnight with head-over-tail rotation.
- 3) Wash the PAS beads three times with 10 ml of G150 buffer and transfer to a 1.5 ml microfuge tube. Wash two times with 1.0 ml G150 buffer and remove as much liquid as possible.
- 4) To elute bound snRNPs, incubate with 0.5 ml G150 buffer containing 0.4 mg ml⁻¹ SF3a66 peptide for 30 min with head-over-tail rotation. Microfuge at 3000 rpm for 1.0 min and transfer the supernatant to a fresh tube. Repeat the elution twice with a fresh batch of peptide and pool the eluates. Microfuge 1 min at 13 000 rpm and transfer to a new microfuge tube to ensure that the eluate is free of PAS beads.
- 5) Prepare six 4 ml, linear 10–30% glycerol gradients containing G150 buffer in 11 × 60 mm polyallomer centrifuge tubes. Load 200 µl of the eluate onto each gradient and centrifuge in a Sorvall TH660 rotor (or the equivalent) for 17 h at 27 000 rpm (75 000 g). By hand, fractionate the gradient from the top into twenty-seven 150 µl aliquots.
- 6) To determine the distribution of 17S U2 snRNPs across the gradient, analyze the RNA and/or protein content of each odd-numbered fraction of one of the gradients. Extract 150 µl of each fraction with an equal volume of phenol/chloroform and transfer the supernatant to a new tube. Add 5 volumes of acetone to the phenol phase (to precipitate protein) and 2.5 volumes of 100% ethanol plus 1/10 volume 3 M sodium acetate (pH 5.2) to the aqueous phase (to precipitate RNA). Mix well, incubate for 2 h at –20 °C or 30 min at –80 °C and microfuge for 10 min at 13 000 rpm. Wash the pellets 1 x with 80% ethanol and dry in a vacuum dessicator. Dissolve each RNA pellet in 6 µl RNA sample buffer and heat for 3 min at 95 °C. Analyze RNA on a 10% polyacrylamide/7 M urea gel and visualize by staining with silver [15]. Dissolve each protein pellet in 10 µl protein sample buffer and heat for 5 min at 85 °C. Analyze the proteins on a 10/13% polyacrylamide/SDS gel and visualize by staining with Coomassie. Under these conditions, 17S U2 snRNPs typically peak in fractions 13–15 of the gradient.
- 7) To regenerate the anti-SF3a66 antibodies coupled to PAS, transfer the beads to a glass column and wash with 20 ml of 10 mM NaPO₄ (pH 7.2). To avoid loss of activity, antibodies should be incubated only briefly with 3.5 M MgCl₂. Elute bound peptide by washing for 5 min with 5.0 ml of 3.5 M MgCl₂ in 10 mM NaPO₄ (pH 7.2). Wash three times with 10 ml PBS containing 0.02% sodium azide and store at 4 °C.

43.3

RNA Aptamer-Based Affinity Purification

43.3.1

Approaches for the Isolation of Native Spliceosomal Complexes

Spliceosomes are composed of the U1, U2, U4, U5, and U6 snRNPs and a large number of non-snRNP splicing factors. The spliceosomal snRNPs and non-snRNP splicing factors associate with the pre-mRNA in an ordered manner [16]. First, the U1 snRNP interacts, followed by the stable association of U2 snRNP, to form spliceosomal complex A. The preassembled 25S U4/U6.U5 tri-snRNP is then recruited to form complex B, which is structurally rearranged to form the activated spliceosome (complex B^{act}) and then the catalytically activated spliceosome (complex B*). The latter catalyzes the first transesterification reaction, generating complex C. After the second step of splicing, the mRNA is released, the postspliceosomal complex disassembles, and the snRNPs are recycled for new rounds of splicing. Thus, during their formation/catalytic activity, spliceosomes go through many intermediate assembly/functional stages (i.e., E, A, B, B^{act}, B*, and C complex) (Figure 43.2a).

Two general methods have been employed to isolate native spliceosomal complexes. The first is to target a protein component of the spliceosome. To isolate a specific spliceosomal complex, ideally such a component should transiently interact with the spliceosome at a specific stage of its assembly/function. Antipeptide antibodies raised against such a protein would allow immunoprecipitation and subsequent peptide-induced elution of only those spliceosomal complexes that contain the targeted protein (Section 43.2). Using antipeptide antibodies directed against spliceosomal proteins that are transiently but stably associated with the spliceosome, we have been able to isolate activated spliceosomes [2], as well as spliceosomal complex B lacking the U1 snRNP [23].

A second approach for the isolation of spliceosomal complexes is targeting the pre-mRNA by adding an aptamer sequence to its 5' or 3' end (e.g., the hairpin structures bound by tobramycin or the MS2 protein). Either before or after allowing spliceosomal complexes to form under splicing conditions in HeLa nuclear extract or in yeast *S. cerevisiae* whole cell extract, the aptamer is bound by its ligand (e.g., tobramycin or MS2 protein fused to the MBP), which binds or is covalently attached to a solid support. After washing, complexes can then be eluted under native conditions by the addition of an excess of the aptamer or ligand. This approach has been successfully used to isolate a mixture of spliceosomal complexes [6], as well as spliceosomes at a more defined stage of assembly and function such as the A, B, B^{act}, B*, or C complexes from human, *Drosophila*, and yeast cells under mild, physiological conditions [7–12, 24, 25].

A general problem of *in vitro* splicing is that spliceosome assembly is not synchronized; at a given incubation time point a heterogeneous population of spliceosomal complexes are assembled on the pre-mRNA. Thus, if a specific spliceosomal complex is to be isolated, additional measures must be undertaken

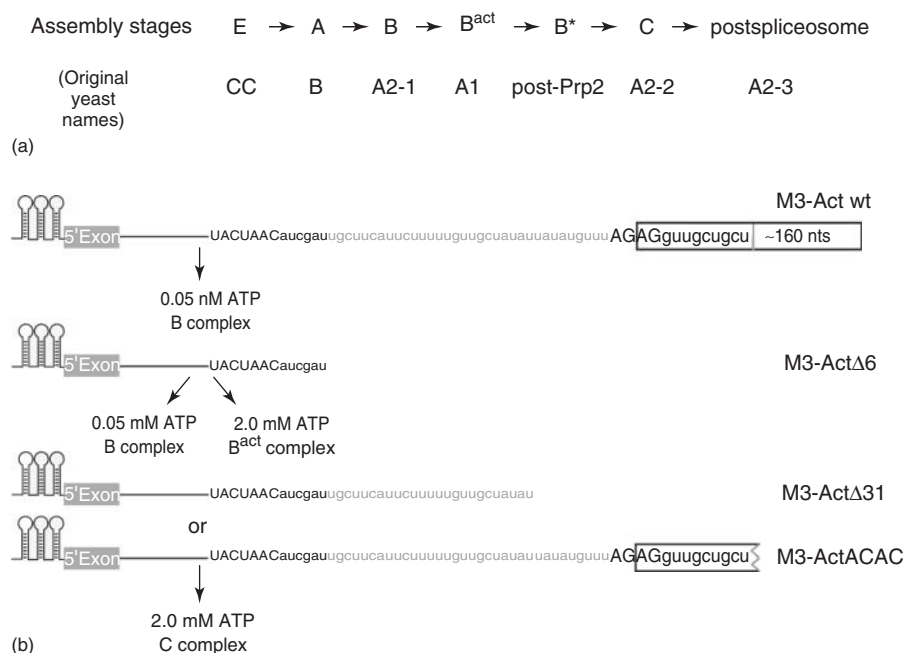


Figure 43.2 Constructs and conditions used for the isolation of yeast spliceosomes. (a) Spliceosomal complexes are named according to the standard human nomenclature and correspond to the yeast complexes shown [17, 18]. (b) Complex B was assembled on wt actin or a truncated actin substrate retaining six nucleotides downstream of the branch-site UAUAAC

sequence (M3-ActΔ6), and stalled by limiting the ATP concentration in the splicing reaction [19]. Complex B^{act} was stalled by using M3-ActΔ6, but increasing the ATP concentration to 2.0 mM [20]. Complex C was stalled by using substrates with a deleted or mutated 3' splice site (M3-ActΔ31 and M3-ActACAC, respectively) [21, 22].

to isolate a more homogeneous population. One approach is to stall spliceosome assembly at a given point. For example, an accumulation of human spliceosomal complex C can be achieved by removing the 3' splice site and the 3' exon of the pre-mRNA [9]. Affinity-purified C complexes formed on such a pre-mRNA substrate can even be chased through the catalytic steps of splicing and catalyze exon ligation in the absence of added factors. This provides a means to verify the functional integrity of the purified complexes [9]. Alternatively, splicing can be carried out for only very short periods of time, so that predominantly only early spliceosomal complexes such as E, A, and B will have time to form [10, 12].

Example: Isolation of Spliceosomal Complexes from the Yeast *S. cerevisiae* by Glycerol Gradient Centrifugation and Affinity Selection Using a Molecular Tag on the Pre-mRNA Substrate

The protein composition of human and *Drosophila melanogaster* spliceosomes [8–10, 24], as well as the protein composition of affinity-purified, *in vitro* assembled

S. cerevisiae spliceosomal complexes has been determined by mass spectrometry [11]. Analysis of yeast B, B^{act}, and C complexes revealed that the number of proteins identified in each complex is much lower than that in the corresponding metazoan complex. Altogether ~90 proteins were identified in yeast spliceosomes, nearly all of which have homologs in higher eukaryotes [11]. Thus, the yeast splicing machinery probably contains the evolutionarily conserved core set of spliceosomal proteins required for constitutive splicing. Indeed, most of the additional ~80 proteins found in human and *D. melanogaster* spliceosomes do not have counterparts in yeast and most of them play a major role in alternative splicing, a process that is essentially absent in yeast. The less complex protein composition of yeast spliceosomes provides a significant advantage for two-dimensional (2D) structure analyses by electron microscopy. In addition, purified yeast spliceosomes exhibit an unprecedented level of quality based on criteria such as structural integrity and homogeneity [11].

Pre-mRNA Constructs and Conditions Used for the Isolation of Yeast Spliceosomal Complexes

To isolate biochemically homogenous B, B^{act}, and C complexes, we employ an actin pre-mRNA substrate and modifications thereof. Each stalled complex is isolated by centrifugation, followed by MS2 affinity selection [11]. For MS2 affinity purification, three MS2 binding sites are introduced at the 5' end of various actin pre-mRNA constructs by overlapping PCR using the pAdML-M3 construct [6]. We term the resulting pre-mRNA "M3-Act." Previous studies showed that truncation/modification of the actin pre-mRNA and adjustment of the ATP concentration in the splicing reaction leads to stalling of the spliceosome cycle (Figure 43.2a,b) [20–22]. The intron in the actin pre-mRNA contains the yeast consensus branch-site sequence UACUAAAC. M3-ActΔ6 and M3-ActΔ31 are truncated to contain only 6 and 31 bases, respectively, after this sequence. Both RNAs therefore lack the 3' splice site and the 3' exon. The variant M3-ActACAC contains a mutation in the intron's 3' splice site (AGAG to ACAC), which is followed by a wild type (wt) but shortened 3' exon. The inclusion of only 0.05 mM ATP in the splicing reaction leads to stalling at the stage of complex B, both for M3-Act and M3-ActΔ6 (Figure 43.2b) [11]. Raising the ATP concentration to 2.0 mM allows splicing to proceed to completion for wt M3-Act, while for M3-ActΔ6, the B complex is converted to the activated B^{act} complex only (Figure 43.2b), without any detectable splicing intermediates or spliced mRNA (Figure 43.4b). Under the same conditions, the truncated M3-ActΔ31 pre-mRNA behaves differently from M3-ActΔ6; with M3-ActΔ31, but not M3-ActΔ6, spliceosome assembly proceeds to the C complex (Figure 43.2b), as demonstrated by the appearance of the cleaved 5' exon and intron-3' exon intermediates, but not the splicing products (Figure 43.4c). The behavior of the mutated pre-mRNA M3-ActACAC (Figure 43.2b) is qualitatively similar to that of M3-ActΔ31 but the yield of the C complex is generally higher with M3-ActACAC.

Purification of Yeast Spliceosomes

For the preparative isolation of B, B^{act}, and C complexes, the various M3-Act substrates described above, are transcribed *in vitro* using SP6 or T7 polymerase in the presence of [³²P]-UTP. The radiolabeled pre-mRNA is first preincubated with MS2-MBP fusion protein and then subjected to splicing conditions at 23 °C for 50–60 min. The strategy for the isolation of yeast spliceosomes is depicted in Figure 43.3 for the B^{act} complex. Subsequently, the splicing reactions are subjected to glycerol gradient centrifugation to separate B, B^{act}, or C from the majority of other spliceosomal complexes (such as complex A), and from excess MS2–MBP fusion protein. The distribution of spliceosomal complexes in the gradient is determined by measuring the amount of radiolabeled RNA in each

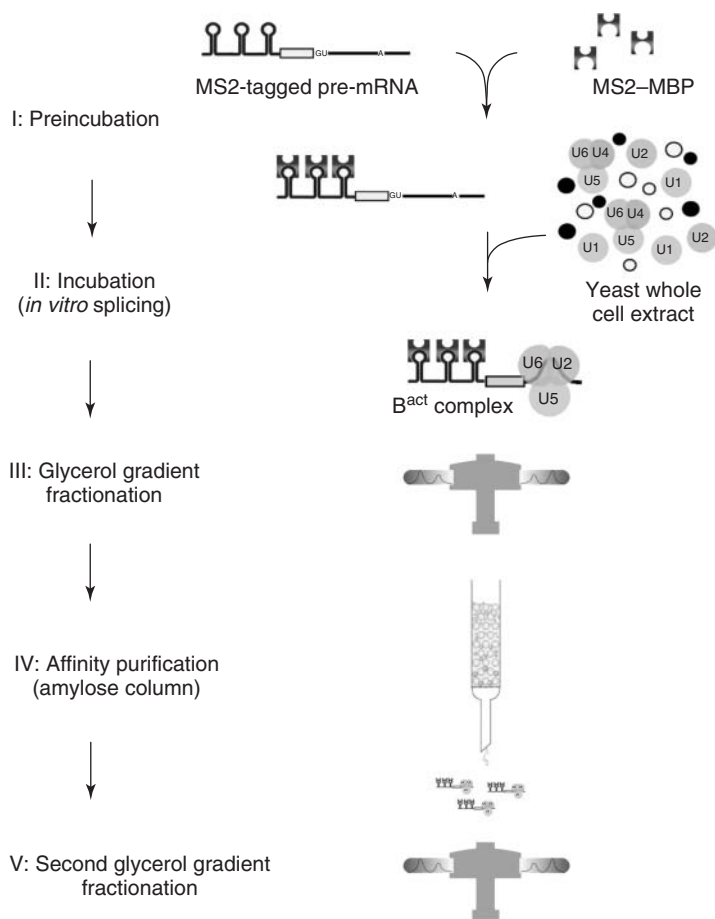


Figure 43.3 Schematic of the purification protocol for the isolation of yeast spliceosomes. Purification of the B^{act} spliceosome is depicted.

fraction, and the peak fractions in the 40S–45S region of the gradient (containing the bulk of B, B^{act}, or C) [11] are pooled and applied to amylose resin, to which the spliceosomal complexes bind selectively through the MBP. After thorough washing, spliceosomal complexes are eluted under native conditions with excess maltose and are subsequently fractionated on a second glycerol gradient. Their distribution in the gradient is determined *via* measuring the amount of ³²P-labeled RNA and their RNA and protein composition are subsequently analyzed by denaturing polyacrylamide gel electrophoresis (PAGE).

Complex B is assembled on M3-ActΔ6 pre-mRNA using splicing conditions with 0.05 mM ATP. The conditions chosen allow the formation of mainly B complexes. Indeed, because of the low ATP concentration, U1 and U4 cannot be released from the spliceosome and as a consequence the first step of splicing cannot occur. Purified B complex exhibits a sedimentation coefficient of ~40S (Figure 43.4a) and contains nearly equimolar amounts of M3-ActΔ6 pre-mRNA and U2, U4, U5, and U6 snRNAs, while U1 is slightly underrepresented (Figure 43.4a, RNA). The autoradiograph also confirms that no splicing intermediates are present. Thus, precatalytic B complexes of high homogeneity can be obtained. The distribution of radioactivity across the glycerol gradient shows, in addition to the main peak at 40S, a second peak with a sedimentation coefficient of ~20–25S. RNA analysis of this peak revealed the presence of unspliced pre-mRNA and also degradation products of the snRNAs. Similar extra peaks are generally observed for all of the complexes.

Complex B^{act} is prepared in the same way, using M3-ActΔ6 but with 2.0 mM ATP in the splicing reaction. It exhibits a slightly higher S value (~45S) than complex B (Figure 43.4b). Nearly stoichiometric amounts of uncleaved pre-mRNA and U2, U5, and U6 snRNAs are detected in these complexes after silver staining (RNA), while U1 and U4 snRNAs are essentially absent from the B^{act} complex peak (Figure 43.4b). The absence of U1 and U4 indicates that the predominant complex isolated is the activated B^{act} and not the precatalytic B complex.

Complex C is assembled on the M3-ActACAC pre-mRNA substrates (Figure 43.4c). DNA-oligonucleotide-directed RNase H cleavage is subsequently performed to remove earlier splicing complexes containing unspliced pre-mRNA. For the RNase cleavage, a DNA oligonucleotide complementary to nucleotides +13 to +24 (see below) downstream of the 5' splice site is added and the splicing reaction is incubated for an additional 20 min at 30 °C. This oligonucleotide directs the excision of the 5' exon by RNase H (present in the extract), but only in unspliced pre-mRNA, and not after the pre-mRNA has undergone step 1 of splicing. After affinity purification, complex C exhibits an S value of ~40 and contains similar amounts of U2, U5, and U6 snRNAs (RNA), excised 5' exon and intron-3' exon (see also autoradiograph). In general, less than 5–10% of unspliced pre-mRNA is detected in purified C complexes after oligonucleotide-directed RNase H digestion of unspliced pre-mRNAs (Figure 43.4c). An estimation of the recovery of B, B^{act}, and C complexes indicates that B^{act} (MW 3.2 Md) is the most abundant, 7.5 pmol (~30 µg) of B^{act} spliceosomes can be recovered from 15 to 30 ml of yeast splicing extract (a 36–72 ml splicing reaction), which is obtained from 35 to 70 g of yeast cells (~4–8 l yeast culture).

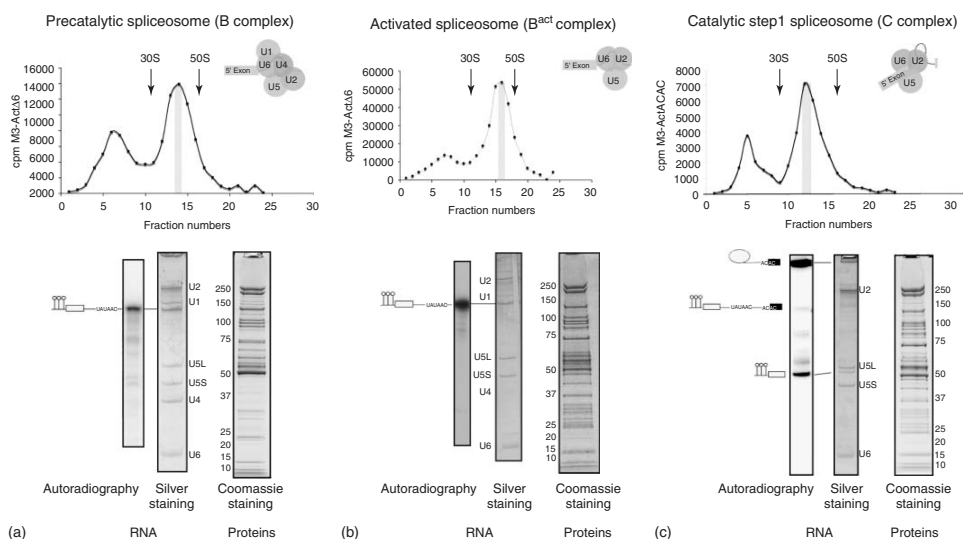


Figure 43.4 Affinity purification of B, activated B (B^{act}), and C *S. cerevisiae* spliceosomal complexes. Spliceosomes were affinity purified from yeast splicing extract with MS2-MBP. Profiles of purified B complex (a), B^{act} complex (b), and C complex (c) separated on a glycerol gradient. The radioactivity contained in each gradient fraction was determined by Cherenkov counting. Sedimentation coefficients were determined by analyzing the UV absorbance of a reference gradient containing prokaryotic ribosomal subunits. RNA from the gradient fraction indicated in (a–c) was recovered, separated by denaturing PAGE, and visualized by silver staining (RNA), and autoradiography. RNA identities are indicated on the right or left. Typical Coomassie-stained SDS-polyacrylamide gel (4–12% NuPAGE Novex Bis-Tris Mini gels, run in MOPS buffer; Invitrogen) of the same fractions are also shown.

Preparation of MS2–MBP and MS2-Tagged Pre-mRNA

MS2–MBP fusion protein is expressed in *Escherichia coli* and prepared essentially as described elsewhere [7, 26], except that the protein is eluted from the final heparin column with a buffer containing 100 mM KCl. Uncapped pre-mRNA, containing three MS2 hairpins, uniformly ^{32}P -labeled at its uridines, is then synthesized *in vitro* by SP6 or T7 runoff transcription. Following gel purification of the pre-mRNA, the amount of MS2–MBP required to saturate its MS2-binding sites completely should be determined empirically [27] at least once for each MS2–MBP preparation. To this end, increasing amounts of MS2–MBP protein are incubated with the tagged pre-mRNA – untagged pre-mRNA is used as a control – and the resulting complexes are analyzed on a native gel. Usually, a 20- to 35-fold molar excess of MS2–MBP is sufficient to shift all of the pre-mRNA on the gel, indicating that the tagged pre-mRNA is quantitatively bound by the MS2–MBP protein.

Protocol 4: Affinity Purification of *S. cerevisiae* Spliceosomal Complexes

- Yeast whole cell extract prepared according to [28, 29] and dialyzed in buffer D: 20 mM Hepes-KOH (pH 7.9), 50 mM KCl, 0.2 mM EDTA (pH 8.0), 20% (v/v) glycerol, 0.5 mM dithiothreitol (DTT), 0.5 mM PMSF, and 0.5 mM benzamidine.
 - Gradient buffer G-150 and G-75: 20 mM Hepes-KOH (pH 7.9), 1.5 mM MgCl_2 , 30 or 10% glycerol, 150 or 75 mM KCl, 0.5 mM DTT, and 0.5 mM PMSF.
 - Splicing cocktail concentrated 5 x: 300 mM potassium phosphate (pH 7.4), 15% polyethylene glycol 8000, 10 mM ATP (or 0.25 mM for B complex preparation), 12.5 mM MgCl_2 , and 10 mM spermidine.
 - Amylose resin (New England Biolabs) washing buffer G-150: 20 mM Hepes-KOH (pH 7.9), 1.5 mM MgCl_2 , 150 mM KCl, 0.01% NP40, 5% glycerol, 0.5 mM DTT, and 0.5 mM PMSF.
 - Amylose elution buffer G-150: amylose washing buffer plus 12 mM maltose.
 - Protein sample buffer: 60 mM Tris-HCl (pH 6.8), 1 mM EDTA, 16% glycerol, 2.0% SDS (w/v), 0.1% Bromophenol Blue, and 50 mM DTT. Add DTT directly before use. Alternatively, use NuPAGE LDS Sample Buffer (4X) (Invitrogen).
 - RNA loading buffer: 80% formamide, 0.5 x Tris/Borate/EDTA (TBE buffer), 0.1% Xylene cyanol, 0.1% Bromophenol Blue.
- 1) Bind MS2–MBP protein to the desired amount of pre-mRNA by incubating the two components on ice for a minimum of 30 min in 20 mM HEPES-KOH (pH 7.9) in a total volume of 1.5 ml. To facilitate tracking, a fraction of the pre-mRNA is cotranscriptionally ^{32}P -labeled (specific activity: 7000–10 000 cpm fmol $^{-1}$) and mixed with unlabeled pre-mRNA to give a specific activity of 150–250 cpm fmol $^{-1}$.
 - 2) To isolate 4–8 pmol of spliceosomal complexes (i.e., an amount sufficient to visualize proteins by Coomassie staining after SDS-PAGE), prepare, on ice, a 36.0 ml splicing reaction containing 40% yeast whole cell extract in dialysis

buffer, 1 x splicing cocktail and 2–2.4 nM ^{32}P -labeled pre-mRNA (specific activity: 150–250 cpm fmol $^{-1}$) bound by MS2–MBP (see above). Pipette the splicing reaction into three 50 ml plastic capped tubes such that each contains 12 ml. Incubate for 50–60 min at 23 °C in a water bath.

- 3) To prepare C complexes, RNase H digestions are additionally performed to remove pre-mRNA not assembled into C complexes. To this end, after incubating the splicing reaction in a water bath for 60 min at 23 °C, add a 125-fold molar excess of a DNA oligonucleotide complementary to nucleotides +13 to +24 downstream of the 5' splice site (5'–GATGGTGCAAGC–3'), and continue the incubation at 30 °C for an additional 20 min.
- 4) After incubation, pool all three reactions and store on ice. All subsequent procedures should be carried out at 4 °C.
- 5) Prepare six 36 ml linear, 10–30% glycerol gradients containing G-150 buffer in 25 × 89 mm polyallomer centrifuge tubes (use a Sorvall SureSpin 630 rotor or the equivalent). We form the gradient with an automatic Gradient Master (see www.biocompinstruments.com), which forms gradients with very reproducible, predefined parameters. Gradients are prepared at room temperature and then equilibrated to the desired running temperature, usually 4 °C, by placing them in a refrigerator or cold room for a minimum of 1 h.
- 6) Load 6 ml of the splicing reaction on top of each gradient. Before loading, an amount of liquid corresponding to the sample volume to be applied is removed from the top of the gradient. It is important to ensure that the tubes are balanced before the actual centrifugation run.
- 7) Centrifuge the gradients at 22 100 rpm (approximately 91 000 g) for 18 h at 4 °C in a Sorvall SureSpin 630 rotor. Under these conditions, the main fraction of the spliceosomal complexes sediments in fractions 15–20. By hand, fractionate the gradient from the top into twenty-three 1.5 ml aliquots.
- 8) Determine the distribution of the ^{32}P -labeled pre-mRNA in the gradient fractions by Cherenkov counting. Pool the six peak radioactive fractions (typically fractions 15–20 that correspond to the 40–45S region) from the six gradients.
- 9) Prepare two plastic 10 ml polypropylene columns (e.g., conical 0.8 × 4 cm Poly-Prep Chromatography Column from BIO-RAD), each with 300 µl bed volume of amylose resin (New England Biolabs (NEB), 100 µl amylose resin binds 60 µg MBP). Wash each column once with 10 ml of amylose washing buffer G-150 (see above).
- 10) Pipette half of the pooled fractions from the gradients (~27 ml), onto each amylose column. Allow the sample to pass into the matrix by gravity flow.
- 11) Wash the columns 4 x with 10 ml each of amylose washing buffer.
- 12) Elute the bound complexes from each amylose column with 10 times 100 µl amylose elution buffer containing 12 mM maltose. Measure the amount of the eluted radiolabeled RNA by Cherenkov counting. Pool the peak fractions.
- 13) Prepare two 4.4 ml linear 10–30% glycerol gradient containing G-75 buffer in 11 × 60 mm polyallomer centrifuge tubes. Load 500 µl of the eluate onto each gradient and centrifuge in a Sorvall TH660 (or equivalent) rotor at

- 60 000 rpm (490 000 g) for 107 min at 4 °C. Under these conditions, 50S and 30S *E. coli* ribosomal subunits (used as sedimentation coefficient markers) migrate in the bottom half of the gradient. By hand, fractionate the gradient from the top into twenty-three 175- μ l aliquots at 4 °C.
- 14) Determine the distribution of the ^{32} P-labeled pre-mRNA by Cherenkov counting. Examples of gradient profiles typically observed are shown in Figure 43.4.
 - 15) To determine the distribution of spliceosomal complexes across the gradient, analyze the RNA and protein content of each gradient fraction. Add 200 μ l stop buffer (50 mM NaOAc (pH 5.2), 0.3 mM EDTA (pH 8.0), and 0.1% SDS) to each 175 μ l fraction of interest. Extract RNA from each fraction with an equal volume of phenol/chloroform/isoamylalcohol (50:50:1, PCI) and transfer the supernatant to a new microfuge tube. Precipitate by adding 1.5 μ l of glycoblue 15 mg ml⁻¹ (Ambion), 1/10 volume 3 M sodium acetate (pH 5.2), and 4 volumes 100% ethanol. Mix well and then incubate overnight at -20 °C or 1 h at -80 °C.
 - 16) Precipitate the samples by microfuging at 13 000 rpm at 4 °C for 20 min. Wash the pellet with 70% ethanol, dry in a vacuum desiccator and dissolve the pellet in 3–5 μ l of formamide RNA sample buffer. After heating at 65 °C for 5 min, analyze on a 8 or 10% polyacrylamide–bisacrylamide (29:1), 8 M urea gel run for 90 min, and visualize by silver staining [13] and autoradiography (Figure 43.4).
 - 17) For protein analysis, pool the peak fractions (~4–8 pmol of spliceosomes) and concentrate them down to a volume of ~300 μ l under vacuum in protein Lobind Eppendorf tubes. Add to the sample 1/100 volume of glycoblue (15 mg ml⁻¹) and 4 volumes of 100% ethanol. Mix well as the samples are very viscous! Leave at -20 °C overnight or at -80 °C for 1 h.
 - 18) Precipitate the samples by microfuging at 13 000 rpm at 4 °C for 20 min. Wash with 70% ethanol. Dry the pellet in a vacuum desiccator. Resuspend the pellet in 50 μ l (total volume) 1 x SDS loading buffer or 1 x NuPAGE LDS sample buffer (Invitrogen).
 - 19) Analyze on a 10/13% polyacrylamide–SDS gel and visualize the proteins by staining with G-colloidal Coomassie brilliant blue (Sigma). Alternatively, analyze the sample on a 4–12% NuPAGE Novex Bis-Tris Mini gel, run in MOPS SDS running buffer (Invitrogen) at 200 V for ~45 min. Typical RNA and protein analyses of the purified spliceosomal complexes are shown in Figure 43.4.

Acknowledgments

We thank Reinhard Rauhut for helpful discussions and Josep Vilardell (IBMB, Spain) for providing the plasmid for the expression of the MS2-MBP fusion protein. This work was supported by a grant from the European Commission (EURASNET-518238) to Reinhard Lührmann.

References

- Will, C.L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Lührmann, R. (2002) Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J.*, **21**, 4978–4988.
- Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Lührmann, R. (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, **298**, 2205–2208.
- Ryder, U., Sproat, B.S., and Lamond, A.I. (1990) Sequence-specific affinity selection of mammalian splicing complexes. *Nucleic Acids Res.*, **18**, 7373–7379.
- Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.*, **20**, 46–50.
- Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res.*, **12**, 1231–1245.
- Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419**, 182–185.
- Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N., and Moore, M.J. (2002) Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*, **8**, 426–439.
- Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Lührmann, R. (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.*, **26**, 1737–1748.
- Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Lührmann, R. (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature*, **452**, 846–850.
- Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.*, **26**, 5528–5543.
- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell*, **36**, 593–608.
- Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16719–16724.
- Will, C.L., Kastner, B., and Lührmann, R. (1994) in *RNA Processing I* (eds S.J. Higgins and B.D. Hames), Oxford University Press, Oxford, pp. 141–177.
- Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.
- Merril, C.R. (1990) Gel-staining techniques. *Methods Enzymol.*, **182**, 477–488.
- Wahl, M.C., Will, C.L., and Lührmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701–718.
- Cheng, S.C. and Abelson, J. (1987) Spliceosome assembly in yeast. *Genes Dev.*, **1**, 1014–1027.
- Kim, S.H. and Lin, R.J. (1996) Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol. Cell. Biol.*, **16**, 6810–6819.
- Tarn, W.Y., Lee, K.R., and Cheng, S.C. (1993) The yeast PRP19 protein is not tightly associated with small nuclear RNAs, but appears to associate with the spliceosome after binding of U2 to the

- pre-mRNA and prior to formation of the functional spliceosome. *Mol. Cell. Biol.*, **13**, 1883–1891.
20. Cheng, S.C. (1994) Formation of the yeast splicing complex A1 and association of the splicing factor PRP19 with the pre-mRNA are independent of the 3' region of the intron. *Nucleic Acids Res.*, **22**, 1548–1554.
 21. Vijayraghavan, U., Parker, R., Tamm, J., Iimura, Y., Rossi, J., Abelson, J., and Guthrie, C. (1986) Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. *EMBO J.*, **5**, 1683–1695.
 22. Rymond, B.C. and Rosbash, M. (1985) Cleavage of 5' splice site and lariat formation are independent of 3' splice site in yeast mRNA splicing. *Nature*, **317**, 735–737.
 23. Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher, H.P., and Lührmann, R. (2002) Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J.*, **21**, 1148–1157.
 24. Herold, N., Will, C.L., Wolf, E., Kastner, B., Urlaub, H., and Lührmann, R. (2009) Conservation of the protein composition and electron microscopy structure of *Drosophila melanogaster* and human spliceosomal complexes. *Mol. Cell. Biol.*, **29**, 281–301.
 25. Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Lührmann, R. (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA*, **16**, 2384–2403.
 26. Jurica, M.S. and Moore, M.J. (2002) Capturing splicing complexes to study structure and mechanism. *Methods*, **28**, 336–345.
 27. Zhou, Z. and Reed, R. (2003) Purification of functional RNA-protein complexes using MS2-MBP. *Curr. Protoc. Mol. Biol.*, Chapter 27, Unit 27.23.
 28. Lin, R.J., Newman, A.J., Cheng, S.C., and Abelson, J. (1985) Yeast mRNA splicing *in vitro*. *J. Biol. Chem.*, **260**, 14780–14792.
 29. Umen, J.G. and Guthrie, C. (1995) A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev.*, **9**, 855–868.

44

Study of RNA–Protein Interactions and RNA Structure in Ribonucleoprotein Particles (RNPs)

Virginie Marchand, Annie Mougin, Agnès Méreau, Isabelle Behm-Ansmant, Yuri Motorin, and Christiane Branlant

44.1

Introduction

In cells, RNAs almost invariably function in association with proteins and form ribonucleoprotein particles (RNPs). In most of the characterized RNPs, one or more proteins with RNA-binding properties first associate with the RNA. Subsequently, other protein components may associate with the complex by protein–protein interactions or both protein–protein and RNA–protein interactions. Cellular RNA molecules can be classified into various groups according to their function or localization and different classes of proteins are associated with each of these groups [1–3]. The RNA-associated proteins have diverse functions. They can package and stabilize RNAs, participate in RNA transport and subcellular localization, mediate RNA interactions with other macromolecules, or modulate RNA activity [4, 5].

In contrast to DNA, RNA can adopt a large variety of three-dimensional (3D) structures. RNA–protein interactions may involve a defined nucleotide sequence and/or a specific 2- or 3D RNA motif. Several protein domains were selected in the course of evolution for their ability to bind peculiar RNA motifs with either a narrow or a broad specificity. One of the best-studied examples is the RNA recognition motif (RRM) [6] that was first discovered in spliceosomal particles. It turned out to be a very general motif present in many different proteins [7]. Members of the RRM family include proteins that bind mRNAs, snRNAs, or rRNAs [8]. Another, frequent RNA binding domain in proteins [1], the KH domain, was first identified in human hnRNP K protein [9]. The KH domains have well-defined and conserved 3D structures [2, 10–14]. Another RNA binding motif with a defined 3D structure (the L30 motif) was first described in the ribosomal L30 protein [15]. It was later found to bind RNAs that form peculiar “K-turn” or “K-loop” structures [16, 17]. Several “K-turn” structures were discovered in rRNAs [18, 19] and also in small nuclear RNAs and small nucleolar RNAs (snoRNAs) [16, 17], and they were found to bind “L30 type” protein domains [16, 17, 20–23]. Another RNA binding domain, which was first found in *Escherichia coli* RNase III [24], specifically binds the double-stranded RNA binding domain (dsRBD). This domain is limited to

interactions with the A-form RNA helix. Proteins carrying this domain play various functions and are especially important for miRNA- and siRNA-mediated regulatory processes [25, 26]. More generally, new RNA binding domains were discovered and characterized upon dissecting the silencing and miRNA machineries [3, 13, 27]. Other conserved RNA binding domains are frequently encountered in proteins that bind RNAs, but their 3D structures are not well characterized. This is the case for the RGG motif that was initially identified in the hnRNP U protein and is often found in combination with RRM motifs [28]. Other types of RNA binding motifs have also been described such as zinc fingers, arginine rich, and cold-shock domains [2, 13, 29–32].

There are numerous approaches to characterize RNA–protein interactions. First of all, one has to identify the proteins that are associated with the studied RNA. When possible, the best starting experiment is to purify the RNP complex to homogeneity. Classical immunoselection approaches have been used for a long time as a first step in RNP purifications. They were based on the use of antibodies directed against one of the protein components of the complex or against the specific cap structure of the RNA [33, 34]. This first immunoselection step was followed by monoQ/fast performance liquid chromatography (FPLC) and/or fractionation by glycerol gradient centrifugation [33, 34]. In the last decade, new approaches were developed for purification of both cellular RNPs and *in vitro* reconstituted RNPs by two successive immunoselection steps [16, 35, 36]. The methods dedicated to the purification of cellular RNPs are based on the expression of tagged components of the RNP: two proteins or one protein and the RNA are tagged, allowing RNP purification by two successive affinity chromatography steps [16, 37, 38]. The methods dedicated to the purification of RNPs reconstituted *in vitro* most generally contain a first affinity chromatography step based on immobilized RNA. Depending on the kind of reconstituted RNP, this first step is followed by a second purification step that may consist in glycerol gradient centrifugation or antibody-based affinity chromatography [36, 39]. RNA immobilization can be achieved by several means: first, covalent attachment to the chromatography matrix [40, 41]; second, non-covalent binding *via* specific aptamers added at one extremity of the RNA to RNA binding proteins coated on beads [39, 42–45]; third, affinity chromatography on streptavidin-coated beads coupled to biotinylated oligonucleotides [46]. Among the various methods proposed, recombinant MS2 coat protein-maltose-binding protein (MBP) fusion is now commonly used to achieve RNP isolation from cellular extracts. Once the RNPs or mixture of RNPs have been purified, their protein contents can be identified by mass spectrometry analysis.

The next question is which protein(s) bind(s) directly to the RNA and which ones are associated only by protein–protein interactions or by RNA–protein interactions. One way to identify the primary binding proteins is to produce them in a recombinant form in *E. coli*, yeast, or animal cells, and then to test their *in vitro* capacity to bind RNA by electrophoretic mobility shift assays (EMSAs). This approach can produce information on the affinity of the RNA and protein partners, and also on the conditions that favor the interaction. Sometimes, it is difficult to produce recombinant proteins because of solubility problems. In this case, another

approach can be used if specific antibodies are available. This approach is based on the formation of covalent bonds between ^{32}P -labeled RNA and proteins that are in close contact with the RNA by UV irradiation at 254 nm. Crosslinking is followed by RNA digestion. The free and crosslinked proteins are immunoselected by using antibodies coated on Sepharose beads. Only the proteins that were covalently linked to the RNA are labeled by the residual crosslinked nucleotides. Therefore, they can be visualized by autoradiography after SDS polyacrylamide gel electrophoresis (PAGE) analysis. Measurement of the radioactivity in the gel can give an indication on the affinity of the two partners. Crosslinking experiments may be performed either with cellular extracts or with recombinant proteins [47]. Crosslinking experiments in cellular extract are particularly informative because under these conditions, all the proteins in the extract are in competition with each other, as is the case in cells. Sometimes, one suspects that two proteins compete with each other for the same RNA site. To verify this hypothesis, crosslinking experiments can be performed with the labeled RNA and various relative amounts of the two proteins produced in a recombinant form, yielding interesting information on the relative affinities of the two proteins for the RNA site [48].

After the identification of the proteins that directly interact with the RNA, a further step consists in the precise mapping of their binding sites. One possibility is to produce different fragments of the RNA and to test their ability to bind to the proteins. However, pieces of the RNA may fold into structures different from those present in the entire molecule. One more direct method consists in mapping the protein binding sites by the use of chemical and enzymatic probes in solution. The bound RNA regions are protected against the action of the probes and are identified by comparison of the cleavages and modifications obtained in the same conditions in the RNP and the naked RNA. Such footprinting experiments can be performed on authentic native RNP and on complexes reconstituted from recombinant proteins and an *in vitro* transcribed RNA, or on complexes formed upon incubation of an *in vitro* transcribed RNA in extract [47, 48]. However, as described below, if the experiment is performed in an extract, special digestion and modification conditions have to be used because of the presence of a large excess of other proteins and RNAs in the extract. As sites of cleavages and modifications are identified by primer extension analysis, the footprinting analyses can be carried out in the presence of the endogenous RNAs from the extract. Thus, this approach is extremely powerful, as purification of the reconstituted complexes is not required.

Footprinting analysis of RNA–protein complexes formed in extracts can be performed without knowing the identity of the bound proteins. These data allow mapping of the RNA regions that are free or protein bound in the extract. Then, for rapid identification of the proteins bound to the RNA in extracts, supershift experiments can be done without further purification steps [47]. The principle is to form complexes between the RNA and proteins from the extract and to incubate the mixture with antibodies specific for a protein expected to be bound to the RNA of interest. An EMSA is performed with or without incubation with the antibody. If the antibody binds to the protein without dissociation of the complex, the mobility

of the RNP complex in the gel is decreased (supershift). If binding of the antibody to the protein dissociates the complex, the RNP band disappears.

If the secondary structure of an RNA target is not known, we recommend its study in parallel with the footprinting experiments. Indeed, the same series of chemical and enzymatic probes are used for secondary structure analysis and footprinting assays. Knowledge on the RNA secondary structure and, if possible, RNA tertiary structure allows a better delineation of the RNA binding domain. However, the fact that protein binding may alter the RNA structure should be taken into consideration for interpretation of these data.

When the binding site of a given protein has been delineated by this approach, its functional role can be confirmed *in vitro* and *in vivo*: *in vitro*, the absence of protein binding after disruption of the RNA binding site can be tested by reconstitution and EMSA experiments after site-directed mutagenesis of the RNA [22]. In addition to testing the capability of the protein to shift the variant RNAs by EMSA, we also recommend the use of EMSA to perform competition assays between the variant and wild-type (WT) RNAs [49]. The effect on RNP activity of the disruption of the RNA binding site can be also studied *in vivo* [22].

More generally, the biological relevance of the *in vitro* probing data may be tested *in vivo*, as one of the chemical probes, dimethylsulfate (DMS), can be used *in vivo* [50, 51]. As only a limited number of probes can be used *in vivo*, the best strategy is to perform a thorough analysis of the RNP *in vitro*, and then by using DMS as the probe, to verify that both the RNA secondary structure and protected areas are identical *in vivo* and *in vitro*.

Once an RNA protein binding site has been identified, the details of the RNA–protein interaction and the mechanism of its formation can be studied at the molecular level. To this end, the 3D structure of the free RNA and proteins partners and of the complex that they form have to be determined by X-ray or NMR analysis [17, 52–56].

Finally, for a more precise definition of the RNA-binding specificity of a protein, Systematic Evolution of Ligands by EXponential enrichment (SELEX) experiments can also be performed [57–60].

44.2

Methods

44.2.1

RNP Reconstitution

44.2.1.1 Equipment, Materials, and Reagents

Equipment Electrophoresis instruments for small vertical slab gels. Localization of the RNP complexes in gels is performed either by autoradiography using X-Ray films (Fuji or Kodak) processed in an X-ray film developer or by use of a PhosphorImager.

Temperature-controlled baths (96, 65, 30, 20 °C).

Materials Eppendorf tubes, tips, buffers, and MilliQ water should be sterilized before use. Wearing gloves is strongly recommended to avoid contamination of the samples by RNases.

Reagents

• RNP

Nuclear or cytoplasmic extracts from HeLa cells or other cell lines can be purchased from CilBiotech, Belgium ($\sim 13 \text{ mg ml}^{-1}$ of total protein) or prepared according to the method developed by Dignam *et al.* [61]. Before use, the extracts are incubated for 10 min at 30 °C in order to consume the endogenous ATP and kept on ice.

• Proteins

RNP proteins can be either extracted from the purified native RNP particles [62] or produced as recombinant proteins in *E. coli* or using other expression systems [48, 63, 64].

• Antibodies

Some of the primary antibodies used in the described examples can be purchased from ImmunoQuest, Abcam, or Santa Cruz.

• Chemicals and enzymes

SP6 RNA polymerase, T4 Polynucleotide kinase, and T4 RNA ligase are purchased from MBI Fermentas (Lithuania); T7 RNA polymerase is from Ambion; calf intestine phosphatase or shrimp alkaline phosphatase, glycogen (10 mg ml^{-1}), and RNase-free DNase I from MBI Fermentas; yeast total tRNA (20 mg ml^{-1}) from Roche Diagnostics; heparin sodium salt from porcine intestinal mucosa (ref H3393) from Sigma; Hybond C nitrocellulose membrane and ECL detection system are purchased from GE Healthcare, formerly Amersham Pharmacia Biotech.

• Radiochemicals

$[^{32}\text{P}]\text{pCp}$ ($3000 \text{ Ci mmol}^{-1}$), $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ ($3000 \text{ Ci mmol}^{-1}$), and $[\alpha\text{-}^{32}\text{P}]\text{UTP}$ (800 Ci mmol^{-1}) are purchased from Perkin Elmer.

• Buffers

- **Buffer D:** 20 mM HEPES-KOH (pH 7.9), 0.2 mM EDTA, 100 mM KCl, 20% glycerol (w/v). Add freshly prepared 0.5 mM dithiothreitol (DTT) and 0.5 mM phenyl methyl sulfonyl fluoride (PMSF) (dissolved in 96% ethanol).
- **Tris buffer:** 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 2.5 mM MgCl_2
- **Binding buffer A:** 20 mM HEPES-KOH (pH 7.9), 0.2 mM EDTA, 150 mM KCl, 10% glycerol (w/v), 1.5 mM MgCl_2
- **Binding buffer B:** 20 mM HEPES-KOH (pH 7.9), 0.2 mM EDTA, 100 mM KCl, 20% glycerol (w/v), 3 mM MgCl_2
- **CSB loading buffer:** 20 mM HEPES-KOH (pH 7.9), 40% glycerol (w/v), 0.05% bromophenol blue, 0.03% xylene cyanol

- **DNA loading buffer:** 0.02% bromophenol blue, 0.02% xylene cyanol in formamide
- **SDS-PAGE loading buffer:** 62.5 mM Tris-HCl (pH 6.8), 2% SDS, 100 mM β -mercaptoethanol, 10% glycerol, 0.01% bromophenol blue
- **TBE buffer:** 90 mM Tris-borate (pH 8.0), 2 mM Na₂EDTA
- **Elution buffer:** 500 mM Sodium acetate (pH 5.2), 1 mM EDTA
- **PBS-TM:** PBS containing 0.1% Tween 20 and 5% dry nonfat milk powder.

44.2.1.2 RNA Preparation and Renaturation Step

Production of Labeled and Unlabeled RNAs by *In vitro* Transcription RNAs are generated by run-off transcription from a DNA template (usually 0.5–2 pmol of a linearized plasmid or PCR product) using the SP6 or T7 RNA polymerase [48, 65–68]. Efficiency of transcription is usually higher for T7 RNA polymerase than for SP6 RNA polymerase. However, efficient transcription with T7 RNA polymerase requires the presence of at least one G residue at the initiation site. The presence of the GGG, GAG, or GGA sequence strongly enhances the transcription yield [69]. However, addition of these residues at the 5'-extremity of the RNA may alter its secondary structure and/or protein-binding capacity. For instance, the presence of a UAGGGA/U sequence at the 5'-extremity of the transcript often generates an hnRNP A1 binding site. Noticeably, it is not easy to get small RNA transcripts (less than 50 nt) in high yield. For the synthesis of small RNAs, we recommend the use of the MEGAscript® or MEGAshortscript™ kit provided by Ambion (catalog # 1330, 1333, or 1354). Several factors affecting the transcription yield must also be taken into account, such as the quantity of DNA template (generally 0.5–2 pmol), the incubation time (2–4 h), the Mg²⁺/NTP ratio (usually 1/1.75), the pH of NTP stocks, the preparation of the transcription reaction mixture in a defined order and at room temperature.

Uniformly labeled transcripts are produced by incorporation of an [α -³²P]NTP during transcription. After transcription, the DNA template is degraded by RNase free DNase I (10 U). The RNA transcript is purified from shorter RNA fragments and excess of NTPs by polyacrylamide–urea gel electrophoresis (PAGE) in 1 x TBE buffer. RNAs are eluted from the gel slices in 100 μ l of 1 x elution buffer and are precipitated by the addition of three volumes of 96% EtOH, in the presence of 1 μ g of glycogen. After centrifugation, the RNA pellet is washed with 70% EtOH, dried, and dissolved in MilliQ water.

For 5'-end-labeling, RNA (10–100 pmol) is first dephosphorylated at the 5'-end with the calf intestine phosphatase or shrimp alkaline phosphatase, and then labeled with [γ -³²P]ATP (3000 Ci mmol⁻¹) and T4 polynucleotide kinase [70]. For 3'-end-labeling, [³²P]pCp (3000 Ci mmol⁻¹) is ligated to the RNA transcript in the presence of the T4 RNA ligase [71]. Labeled RNAs are purified by denaturing PAGE and eluted as described above.

Unlabeled RNA transcripts used for RNP reconstitution and 2D structure analyses are produced by similar methods, except that, in order to improve the

transcription efficiency, the concentration of NTPs and Mg^{2+} is higher (up to 5 mM of each NTP).

RNA Transcript Renaturation A renaturation process is required to produce a homogeneous population of RNA molecules in terms of RNA secondary structure. Before probing of naked RNA or reconstitution of RNP complexes, the RNA transcript dissolved in buffer D or 1 x Tris buffer is incubated for 10 min at 65 °C and then slowly cooled down to room temperature. Then, addition of divalent ions such as Mg^{2+} (at a concentration between 1.5 and 10 mM) is required to favor RNA 2- and 3D structure formation and its stabilization during the probing and reconstitution experiments. After Mg^{2+} addition, the RNA is incubated for 10 min at room temperature. RNA should not be heated at 65 °C for too long, as phosphodiester bonds can be cleaved in these conditions, especially, when Mg^{2+} ions are present in the incubation buffer (1 x Tris buffer) [72]. In spite the frequent use of 1 x Tris buffer described in literature, we recommend the use of buffer D and the addition of Mg^{2+} ions after it has cooled down to room temperature.

44.2.2

EMSA

EMSA can be used for several purposes. EMSA experiments can be performed with an *in vitro* transcribed RNA and purified proteins or an appropriate cellular extract. For determination of dissociation constant values (K_d), a fixed concentration of labeled RNA and increasing concentrations of the protein or cell extract are used. For estimation of the RNA/protein ratio, which has to be used to form RNP complexes for footprinting analysis, unlabeled RNA is added to the labeled RNA. Under these conditions the same RNA concentration can be used in the EMSA and probing assays.

To perform EMSA with nuclear extract (NE), the buffer, previously defined for *in vitro* pre-mRNA splicing assays [73], is generally used, except that ATP and creatine phosphate are omitted [67, 74, 75]. We recommend the following incubation mixture: HeLa cell NE 40–50% of total assay volume (dialyzed against buffer D), 2.5 mM MgCl_2 , and buffer D. Other conditions can also be used for snRNP reconstitution; however, for the biological relevance of the data, it is important to select *in vitro* conditions as close as possible to the *in vivo* conditions.

44.2.2.1 EMSA Method

The reactions are performed in a total volume of 10–20 μL .

An amount of labeled RNA between 1 and 50 fmol, with or without 1–10 pmol of cold RNA, can be used for the assay. The use of 3'- or 5'-end-labeled RNA is recommended for K_d determination, as uniformly labeled RNAs give less defined bands on the EMSA gels, especially in the case of RNAs longer than 60 nt. However, uniformly labeled RNAs (0.2 fmol) mixed with cold RNAs (2–10 pmol) are convenient for optimization and control of RNP reconstitution. Competitor tRNAs (usually, total tRNAs from the yeast *Saccharomyces cerevisiae*) (10- to 1000-fold mass

excess, relative to the tested RNA) can be added to the mixture in order to prevent non-specific RNA–protein interactions.

RNPs are generally reconstituted in the binding buffers (1 x Tris buffer or 1 x buffer D) in the presence of Mg^{2+} ions at concentrations varying from 1 to 10 mM. NE, or another appropriate extract, or purified proteins are used as protein sources. A large range of protein concentrations can be used: generally from 10 nM to 10 μ M for purified proteins and from 1 to 100 μ g of total proteins for NE in a total volume of 11 μ l. Preincubation of the protein or NE with the yeast total tRNA is recommended in order to reduce the formation of non-specific RNA–protein interactions. A control experiment is performed in the absence of protein extract or purified protein (replaced by 1 x buffer D).

At the end of the incubation, 7 μ l of CSB loading buffer is added. Note that in order to limit the non-specific electrostatic interactions between RNA and proteins, heparin, a negatively charged polysaccharide, can be added (at a concentration of 5 μ g μ l⁻¹). Heparin addition is followed by a 10 min incubation at room temperature. The presence of heparin usually improves the selection of specific RNP complexes. Addition of heparin is recommended in the case of NEs that contain a large amount of positively charged proteins. Depending on the electrostatic properties of the protein, heparin may be omitted when purified proteins are used.

For all types of EMSA, CSB loading buffer is used. Electrophoresis is performed on 5–10% non-denaturing polyacrylamide gels containing 0.5 x TBE buffer and 5% (v/v) of glycerol (acrylamide: bisacrylamide ratio, 19 : 1). In order to limit RNP disruption and to obtain reproducible results, electrophoresis is generally performed at 4 °C and 10 V cm⁻¹. Conditions of electrophoresis, the acrylamide: bisacrylamide ratio as well as the type of running buffer may be modified, if required [76].

At the end of the electrophoresis, the gel is transferred on a sheet of Whatman 3MM paper and dried. Exposure is overnight at –80 °C, using intensifying screens, or at room temperature in a PhosphorImager cassette. Free and bound RNAs are visualized by autoradiography.

For K_d determination, the amounts of radioactivity in the bands of gel are estimated by PhosphorImager measurement and the SigmaPlot Software (SPSS Science Software) can be used for K_d calculations using the measured radioactivity [22].

Example 44.1: Experimental protocol used for K_d determination (Figure 44.1)

Figure 44.1 illustrates the determination of the K_d values for the complexes formed between the recombinant *S. cerevisiae* Snu13 protein and one of its WT or mutated target RNA motif. This target motif is one of the “K-turn” structures present in the nucleolar snoRNA U3 (U3B/C WT). In the variant RNA designated as U3B/C G.C → G.G, one base pair of the motif has been disrupted [54] (Figure 44.1a). RNA–protein complexes were performed under the following conditions: about 4.5 nM of uniformly labeled RNA, mixed with 1 μ g of yeast tRNA, were renatured in 6.5 μ l of 1 x binding buffer A. The recombinant Snu13 protein was added at various concentrations ranging from 50 to 2000 nM and the mixtures

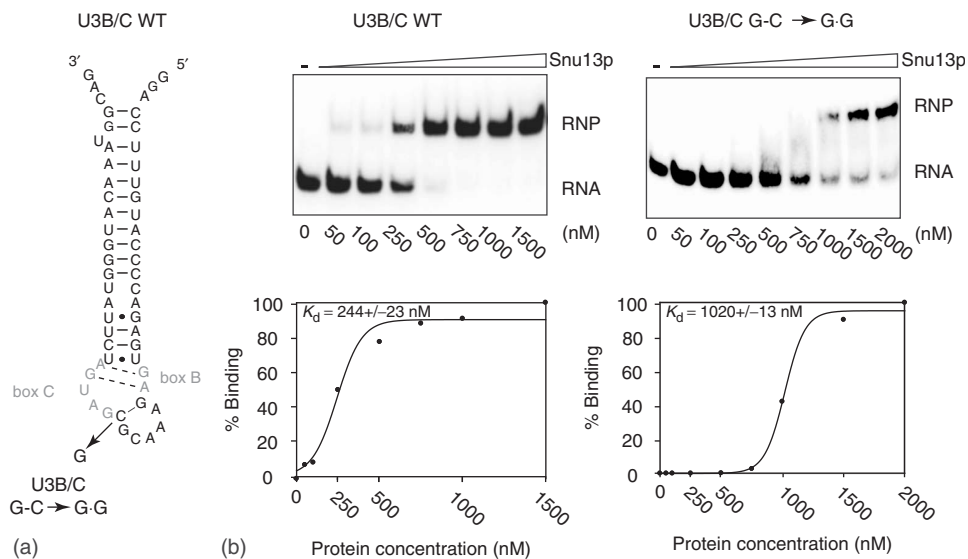


Figure 44.1 Use of EMSA to study the binding of the recombinant protein Snu13p to the WT or mutated B/C motif of yeast U3 snoRNA [22]. Secondary structure of the WT U3B/C RNA and position of the C → G mutation in the variant RNA are shown (a). The 3'-end-labeled RNA was incubated in the presence of increasing amounts of the Snu13 protein, as described above.

The autoradiograms obtained after gel electrophoresis are shown (b). The protein concentration (in nanomoles) is given below each lane. The K_d values of the RNA–protein complexes (protein concentration for which 50% of the input RNA is shifted to an RNP complex), were calculated for each experiment with the SigmaPlot Software (SPSS Science Software).

were incubated for 30 min at 4 °C. After addition of 7 µl of CSB loading buffer, the RNA–protein complexes formed were fractionated by electrophoresis on a 6% (19 : 1) non-denaturing polyacrylamide gel, as described above (Figure 44.1b).

44.2.2.2 Supershift Method

Supershift experiments are a variant of EMSA. RNA–protein complex formation is performed in the presence or absence of antibodies. Binding of antibodies to their target protein in the RNP complex increases the size of the RNP complex and, thus, lowers its electrophoretic mobility on the gel (so-called “supershift”). In some cases, the binding of the antibodies may lead to RNP complex disruption and loss of the shifted RNA band.

For supershift experiments, 0.5–2 µl of antibodies of interest are added before the heparin treatment and incubation is continued for 10 min at 4 °C. The amount of antibody added depends upon the antibody specificity and concentration; both monoclonal and polyclonal antibodies can be used for this type of experiments.

RNP complexes formed with or without antibodies are fractionated in parallel on a non-denaturing polyacrylamide gel (5–10% concentration can be used) and the

electrophoresis is performed in 0.5 x TBE buffer containing 5% of glycerol at 4 °C. The gel is dried, and RNP complexes are detected by autoradiography.

Example 2: “Supershift” experimental protocol (Figure 44.2b) RNP reconstitution experiments are performed with two different fragments of the HIV-1 BRU isolate RNA. The biological components used for the assays are: the SLS2,3 WT RNA (positions 7970–8068) described above, the C3 WT HIV RNA fragment (positions 5359–5408 in the entire molecule) [47], HeLa cell NE, and specific anti-hnRNP A1 and hnRNP H antibodies provided by G. Dreyfuss (University of Pennsylvania School of Medicine, Philadelphia, USA) and D. Black (University of California, Los Angeles, USA), respectively. The RNA fragments transcribed with SP6 RNA polymerase were 3′-end-labeled. With the HIV A7 SLS2,3 WT RNA, a supershift is obtained, demonstrating the presence of protein hnRNP A1 in the complex (Figure 44.2a1). The RNP complex formed with the C3 WT RNA is dissociated by binding of the anti-hnRNP H antibody (Figure 44.2a2). This suggests that interaction of the specific antibody directed against hnRNP H destroys the complex, which argues for the presence of hnRNP H protein in the complex and an essential role for complex stability. In addition, destruction of the complex probably exposes the RNA to RNase in the extract, leading to a partial RNA degradation.

The following experimental conditions were used: the HIV C3 and A7 SLS2,3 RNAs (2.5 and 6 pmol, respectively) were 3′-end-labeled by using [³²P]pCp. Labeled RNAs were incubated at 4 °C in the presence of 5 µg of yeast total tRNA, with 0.5 and 4 µl of NE, respectively. After a 15 min incubation, 0.5 µl of anti-hnRNP A1 antibodies [77] (Figure 44.2a1) or 1 µl of the monoclonal anti-hnRNP H antibodies [78] (Figure 44.2a2) were added, respectively. The incubation was continued for 10 min on ice and was followed by gel electrophoresis. As a control, the two RNAs were incubated under the same conditions in the absence of antibodies.

44.2.2.3 Identification of Proteins Contained in RNP by EMSA Experiments Coupled to a Second Gel Electrophoresis and Western Blot Analysis

This method also allows the identification of proteins present in RNP complexes fractionated on EMSA gels. The EMSA are performed as described above and the band containing RNP complexes are cut out from the gel, soaked in an SDS-PAGE loading buffer (which dissociates the RNP complex) and included in a 5–10% SDS-PAGE. The proteins from the complexes migrate in the gel at the same position as the corresponding proteins present in NE, which are used as a reference. After electrophoresis, they are electrotransferred on a nitrocellulose membrane. The search for the presence of defined proteins is done by immunoblotting using specific antibodies directed against these proteins [47].

Example 3: Protocol for an EMSA experiment coupled to a second gel electrophoresis (Figure 44.2b) Figure 44.2b illustrates the search for the presence of proteins hnRNP H and A1 in RNP complexes I and II, by Western blot analysis (Figure 44.2b). Complexes I and II were formed by incubation in HeLa cell NE of the HIV C3 WT and C3-A mutant RNAs (Figure 44.2b2), respectively. The HIV C3-A

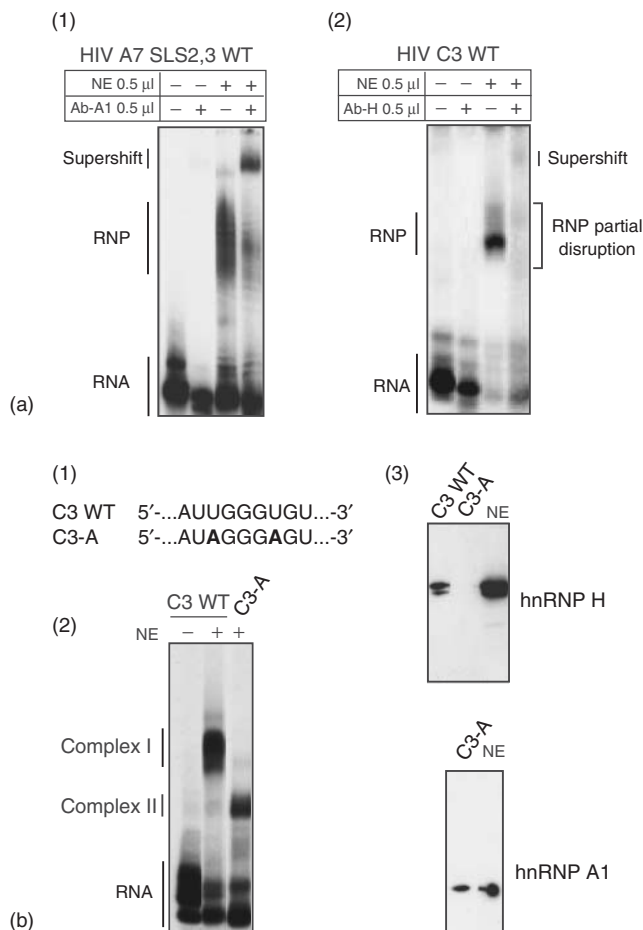


Figure 44.2 Supershift assays performed on the RNP complexes formed by incubation in a HeLa cell nuclear extract (NE) of the A7 SLS2,3 WT (a1) and C3 WT (a2) RNAs [47, 48]. (a) RNAs were 3'-end-labeled and anti-hnRNP A1 (Ab-A1) or anti-hnRNP H (Ab-H) antibodies were used in (a1) and (a2), respectively. Positions of the RNAs, RNP and the supershifted RNP are indicated on the left of the autoradiogram. (b) EMSA experiments coupled with a second gel electrophoresis and Western blot analysis, performed on complexes formed by the HIV WT or mutated C3 RNAs (C3 WT and C3-A) in HeLa cell nuclear extract [47]. The mutations present in the C3-A RNA are shown in (b1), the EMSA experiment

is illustrated in (b2). The positions of the free RNAs and RNP complexes I and II obtained with the HIV C3 WT and C3A RNAs, respectively, are indicated on the left of the autoradiogram. (b3) The bands containing complexes I (C3 WT) and II (C3 A) were included in an SDS-polyacrylamide gel. After electrophoresis proteins were transferred on a nitrocellulose membrane, that was used for Western blot analysis first with hnRNP H antibodies (upper panel) and then with anti-hnRNP A1 antibodies (lower panel). Proteins from nuclear extract (NE) were loaded on the gel for a control. Complex I contained hnRNP H protein and complex II contained hnRNP A1 protein.

RNA is a variant of the HIV WT C3 RNA (two uridines were substituted by two adenosines) (Figure 44.2b1) [47]. The bands of gel containing the RNP complex I formed with the WT RNA and the band of gel at the same level in the lane corresponding to the variant C3-A RNA (control experiment) were cut out. The band of gel corresponding to complex II formed with the mutated RNA was also sliced out. These bands of gel were soaked in 10 μ l of SDS-PAGE loading buffer for 1.5 h at 37 °C, and heated in the buffer for 5 min at 96 °C before their inclusion at the top of a 10% SDS-polyacrylamide gel (1.5 mm thickness). A sample of NE was fractionated in parallel in the lane on the right side of the gel. This allowed the identification of the migration position of the hnRNP H and hnRNP A1 proteins in the gel. After electrophoresis for 2 h at 20 V cm⁻¹, the fractionated proteins were electrotransferred onto a Hybond C nitrocellulose membrane (Pharmacia Amersham Biotech) (for 1 h at 100 V). Then, the membrane was blocked with 20 ml of PBS-TM buffer overnight at 4 °C or 2 h at room temperature with gentle shaking. It was then probed with the anti-hnRNP A1 (0.5 μ l) or anti-hnRNP H (2 μ l) antibodies [47]. The bound antibodies were detected with peroxidase-conjugated anti-mouse and anti-rabbit IgG antibodies, respectively, and visualized by the ECL detection system (Figure 44.2b3) [47].

44.2.3

Purification of RNPs Reconstituted in Complex Cellular Extracts

RNP complexes contained in cytoplasmic or nuclear cell extracts are usually purified through immunoaffinity chromatography methods. The specific antibodies used can be directed against endogenous proteins or tagged proteins expressed from a modified gene (tandem affinity chromatography, TAP-tag technique; for experimental details, see [35, 38, 79–81]). The presence of a particular cap structure at the RNA 5'-end, for instance, the presence of a m³G cap structure at the 5'-extremity of the spliceosomal UsnRNAs and of some snoRNAs was largely used for purification of the spliceosomal UsnRNP and snoRNP using immobilized anti-m³G cap antibodies [34, 62, 82, 83]. The spliceosomal 25S [U4/U6.U5] tri-snRNP, 20S U5 snRNP, 17S U2 snRNP, and 12S U1 snRNP contained in the RNP mixture that is retained on the anti-m³G cap antibodies can then be separated by glycerol gradient centrifugation. The importance of the salt concentration in these purification steps is evidenced by the fact that at KCl concentrations above 250 mM, the [U4/U6.U5] tri-snRNP is disrupted into 12S U4/U6 and 20S U5 snRNP and the 17S U2 snRNP is converted into a 15S or a 12S particle [84]. The insertion of an RNA tag sequence, which binds a protein recognized by a specific antibody, within the RNA expressed was also successfully used for purification of RNPs extracted from cells [16, 85, 86].

More recently, to counteract the difficulty in purifying authentic RNP complexes, very powerful RNA affinity chromatography techniques were developed to purify RNP complexes formed by incubation of *in vitro* transcribed RNA in cellular extracts (total cellular extract, cytoplasmic, or NE). They were successfully applied to the purification of spliceosomal complexes [36, 39, 42, 43, 46, 87–89] and other RNPs [45, 90]. These methods take advantage of the possibility to add a tag

sequence at one of the RNA termini. The most frequently used tag sequences are the tobramycin J6f1 aptamer [88], the StreptoTag aptamer [91, 92], or RNA binding sites for the phage MS2 coat protein [39, 43, 93]. We give here detailed protocols for the methods based on the addition of the tobramycin J6f1 aptamer to the RNA or binding sites for the MS2 coat protein. The principles of these two methods, which have been largely used to purify various kinds of reconstituted splicing complexes, are illustrated in Figure 44.3. It may be noted that if one uses the tobramycin-based method one would need to perform all the purification steps at pH 9.1 to avoid protonation of tobramycin, leading to unspecific binding of proteins to the RNA and unspecific retention of proteins on the beads. It may also be noted that in both methods, the RNPs formed with the RNA linked to the aptamer (or MS2 binding sites) as well as the RNPs formed on the aptamer itself (or MS2 binding sites) should be compared for their protein composition.

44.2.4

Methods for RNP Purification Using Tobramycin–Sephacrose or MS2-MBP Affinity Chromatography

44.2.4.1 Equipment and Materials Common to the Two Approaches

Equipment

Rotating wheel for Eppendorf tubes at 4 °C

Materials

Eppendorf tubes, tips, buffers, and MilliQ water should be sterilized before use. Wearing gloves is strongly recommended to avoid contamination of the samples by RNases.

44.2.4.2 RNP Purification Using Tobramycin–Sephacrose

Reagents

N-hydroxysuccinimide-activated Sepharose 4 Fast Flow (GE Healthcare # 17-0906-01), tobramycin (Sigma # T4014). RNase-free BSA (Sigma # B8894).

Buffer

Buffer for Tobramycin–Sephacrose chromatography (Tob buffer): (20 mM Tris-HCl (pH 9.1) at 4 °C, 1 mM CaCl₂, 1 mM MgCl₂, 0.2 mM DTT).

Tobramycin Immobilization on Activated NHS–Sephacrose Beads N-Hydroxysuccinimide-activated Sepharose 4 Fast Flow (GE Healthcare) is coupled with tobramycin, using the following protocol [94]. NHS–Sephacrose is extensively washed with 1 mM HCl and incubated with 5 mM tobramycin in coupling buffer (0.2 M NaHCO₃, 0.5 M NaCl, adjusted to pH 8.3 with NaOH) overnight at constant head-over-tail rotation. After removal of unreacted tobramycin, the beads are blocked for 2 h in 0.2 M NaHCO₃, 0.1 M NaCl, 1 M ethanolamine, adjusted to pH 8.0 with HCl, and washed three times with 100 mM Tris-HCl (pH 8.5), followed by three successive washes with 100 mM sodium acetate (pH 4.0). The tobramycin–Sephacrose can be stored in 20% EtOH at 4 °C for two to three months.

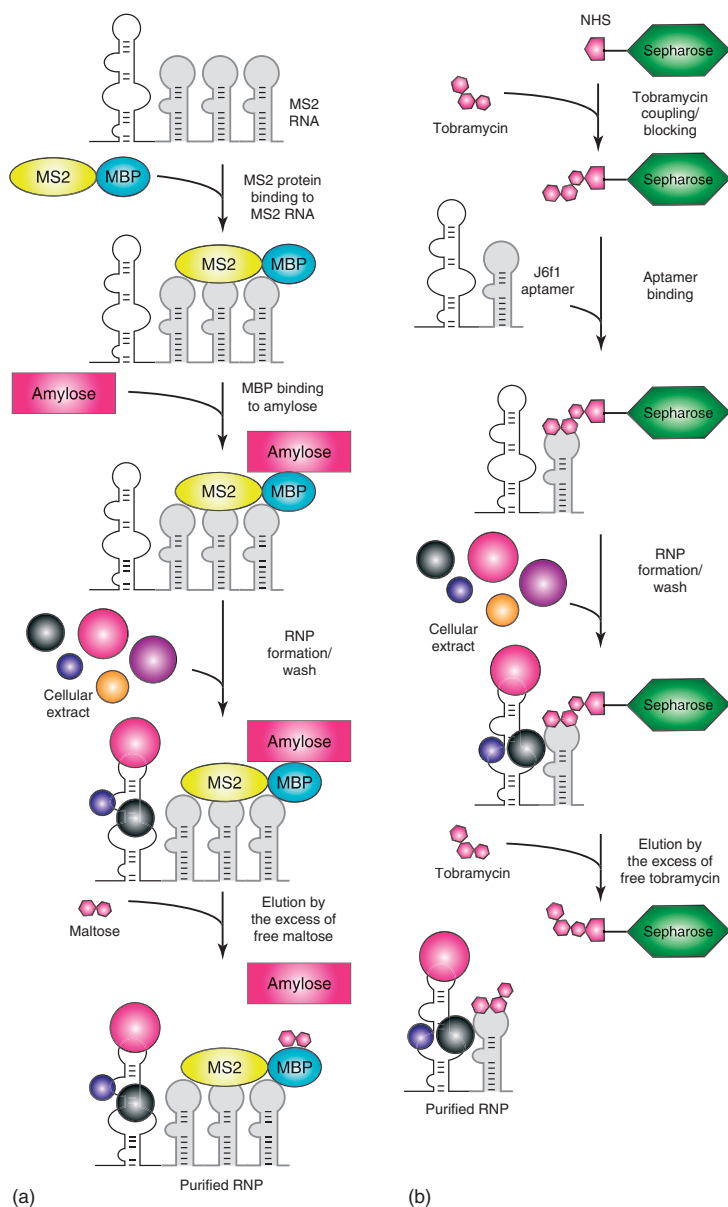


Figure 44.3 Schematic workflow of affinity chromatography purification of RNP complexes based on MS2-MBP (left) and tobramycin-Sepharose (right). RNA is represented by thin black lines, the structures of the J6f1 aptamer and the three MS2 coat

protein binding sites are shaded in gray. Cell extract proteins are represented by colored spheres. Purified RNPs can be biochemically analyzed for their composition or used for RNA structural studies.

RNA Immobilization on Tobramycin–Sephacrose Beads Genetic constructs allowing the production of a control J6f1 aptamer and of the studied RNA linked to the J6f1 aptamer should be prepared and RNAs are transcribed from these constructs as described in Section 44.2.1.2.

In parallel, the tobramycin matrix is blocked with 400 μl of blocking buffer (1 x Tob buffer supplemented with 145 mM KCl, 0.1 mg ml⁻¹ tRNA, 0.5 mg ml⁻¹ BSA, 0.01% Nonidet P-40) per 25 μl of beads suspension by overnight head-over-tail rotation at 4 °C. The matrix is collected by centrifugation. Then, 40 μg of tRNA are mixed with 500 pmol of J6f1 RNA (for the control experiment) or J6f1-tagged RNA (for the assay) in 100 μl of 1 x Tob buffer with 145 mM KCl, and this RNA mixture is subjected to a denaturation–renaturation step including 10 min incubation at 65 °C, followed by slow cooling to room temperature. Then, the 100 μl of RNA solution can be added to 25 μl of matrix. Incubation is done for 2 h at 4 °C. Then, the matrix is washed three times (750 μl each) with 1 x Tob buffer containing 0.1% Nonidet P-40.

44.2.4.3 Formation of RNPs in the Cellular Extract

The HeLa cell NE (500 μl , about 5 mg of total protein, CilBiotech, Mons, Belgium) should be dialyzed against 20 mM HEPES-KOH pH 9.1, 100 mM KCl, 3 mM MgCl₂, 0.2 mM EDTA pH 8.0, 0.5 mM DTT, 0.5 mM PMSF, 20% (vol/vol) glycerol. A yeast tRNA mixture at a 5–20 μM is added and the solution is incubated with 25 μl of tobramycin–Sephacrose matrix preloaded with the J6f1-tagged RNA or the J6f1 RNA for 15 min at 4 °C with head-over-tail rotation.

44.2.4.4 Elution of Purified RNPs under Native Conditions

After incubation with the cell extract, the matrix is collected by centrifugation and washed three times with 500–750 μl of washing buffer (1 x Tob buffer supplemented with 75 mM KCl and 0.1% Nonidet P-40). If necessary, washing steps may be performed at higher ionic strength, but no more than 300 mM salt should be used, to avoid destabilization of RNA–protein complexes.

RNP complexes are eluted by head-over-tail rotation for 15 min at 4 °C or room temperature, if only the protein content is studied, using 100 μl of elution buffer (1 x Tob buffer, 5 mM tobramycin, 75 mM KCl, 2 mM MgCl₂).

44.2.4.5 MS2-MBP Affinity Chromatography

Reagents for MS2-MBP Affinity Chromatography Amylose resin is from New England Biolabs (# E8021S), UK, the recombinant MS2 coat RNA binding domain–Maltose Binding Protein fusion protein (MS2-MBP protein) is expressed in *E. coli* and purified by amylose affinity chromatography [39], followed by an additional cleaning step using Heparin-Hi-Trap HP column (GE Healthcare, France, # 17-0406-01).

The plasmid containing the coding sequence for three MS2 coat protein binding sites: plasmid pMJM273 [87].

Buffer for MS2–MBP affinity purification (MS2–MBP buffer): 20 mM HEPES-KOH (pH 7.9), 100 mM KCl, 3 mM MgCl₂, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 20% (vol/vol) glycerol.

Immobilization of RNA-MS2–MBP Complexes on Amylose Beads Amylose beads (50 µl of 50% slurry, New England Biolabs) are washed three times with MS2–MBP buffer.

Genetic constructs to produce the control MS2 RNA and the studied RNA linked to the three MS2 binding sites should be prepared and their *in vitro* transcription is achieved as described in Section 44.2.1.2.

To form RNA–MS2–MBP protein complex, about 500 pmol of control MS2 RNA (control) or of MS2-tagged RNA (assay) are denatured by a 10 min incubation at 65 °C and renatured by slow cooling to room temperature before incubation at 4 °C for 15 min with a fivefold molar excess of purified MS2–MBP fusion protein prepared according to [39].

RNA–MS2–MBP complexes are then bound to amylose beads by incubation for 1 h at 4 °C with head-over-tail rotation to avoid sedimentation of the matrix. The beads are then washed three times (500 µl each) with MS2–MBP buffer.

Formation of RNPs on the RNA-MS2–MBP Complexes Bound to Amylose Beads HeLa cell NE (350–500 µl, about 5 mg of total protein, CilBiotech, Mons, Belgium) is dialyzed against buffer D equilibrated at pH 8 and total yeast tRNA is added to a final concentration of 5 µM. The concentration of yeast total tRNA may be varied from 1 to 25 µM, depending on the nature of cell extract and the required stringency of RNP purification.

The mixture is added to the amylose beads preloaded with the RNA–MS2–MBP complex and incubation is performed for 15 min at 4 °C with head-over-tail rotation.

44.2.4.6 Elution and Analysis of Purified RNPs

After incubation with the cell extract the matrix is collected by centrifugation and washed three times with 500–750 µl of MS2–MBP buffer. If necessary, washing steps may be performed at higher ionic strength. RNP complexes are then eluted by head-over-tail rotation for 15 min at 4 °C or room temperature with 100 µl of MS2–MBP buffer containing 10 mM maltose.

44.2.4.7 Analysis of the Purified RNP Protein Content

For both affinity chromatography approaches, the protein profile can then be analyzed by SDS-PAGE depending on the amount of eluted complex, 20–80% of the eluted fraction is loaded onto a 10% SDS-polyacrylamide gel and stained with colloidal Coomassie Blue. Then, the bands of gel can be cut out and the proteins in gel digested for further identification by mass spectroscopy. Note that the protein composition of the studied RNPs and the control RNPs should always be studied in parallel by SDS-PAGE and by mass spectrometry in order to identify the proteins that do interact with the studied RNA. Verification of the presence of given proteins

in the purified RNPs can then be achieved by Western blot analysis after SDS-PAGE purification or by immunoselection of the purified RNPs on protein G or protein A Sepharose beads coated with specific antibodies (see [45] for details).

Example 4: Experimental protocol used for purification of RNP complexes using Tobramycin–Sepharose and MS2–MBP affinity chromatography (Figure 44.4)

SLS2 RNA encompassing the A7 3'-splice site of HIV-1 pre-mRNA was fused with either the J6f1 tobramycin-binding aptamer (Figure 44.4a) or with 3 MS2-binding motifs (Figure 44.4b). RNP complexes were formed in HeLa cell NE with the aptamers alone or with the chimeric SLS2 RNAs (SLS2-J6f1 RNA or SLS2-MS2 RNA). After elution as described above, the purified RNPs were separated by 10% SDS-PAGE along with molecular mass markers.

Figure 44.4c illustrates the effect of the addition of various concentrations (5, 25, and 125 μ M) of yeast competitor tRNA in the HeLa cell NE on SLS2-J6f1 RNP formation and purification by the tobramycin-based approach. The RNPs were purified using the standard tobramycin-based protocol described above and analyzed by 10% SDS-PAGE and stained by colloidal Coomassie Blue. The molecular mass markers are shown on the left. The results indicate that excessive concentrations of tRNA competitor lead to a considerable loss of RNA-associated proteins.

44.2.5

Probing of RNA Structure

44.2.5.1 Properties of the Probes Used

Conditions used for RNP probing with chemical reagents or enzymes are chosen to be as mild as possible in order to preserve the structural integrity of RNP particles. If possible, probing reactions are performed in the buffer used for the RNP purification or reconstitution. Incubations are performed at moderate temperature for short times. A defined amount of yeast total tRNA is added in order to control the [RNA]/[probe] ratio. Modification and enzymatic digestion conditions should be selected so that statistically less than one modification or cleavage occurs per RNA molecule. The chemical and enzymatic probes used for footprinting are also used for determination of RNA secondary structure in solution. When probing experiments are performed on purified natural complexes or complexes reconstituted from RNA transcripts and recombinant proteins, almost all the probes used for RNA secondary structure analysis can be used. Only the S1 nuclease, which has an optimum pH of action of 4.5, cannot be used. However, when cellular extracts are used for RNP formation without further purification of the complexes, some of the chemical probes, especially 1-cyclohexyl-3(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate (CMCT), cannot be used. The method employed for the identification of cleavage and modification positions depends on the labeled state of the RNA. For unlabeled RNAs, that is, with natural purified RNPs or RNPs reconstituted from unlabeled RNAs, primer extension analyses with reverse transcriptase are performed: stops of extension occur at the cleavage site or one residue before the cleaved (depending on the enzyme used) or modified nucleotide. When 3'- or 5'-end-labeled

Figure 44.4 Structure of SLS2 RNAs used for affinity chromatography and SDS-PAGE analysis of proteins contained in affinity-purified complexes (a) Left: Secondary structure of the SLS2-J6f1 RNA used for RNP purification. The secondary structure of this HIV-1 RNA fragment was previously established experimentally [48]. The A7 3' splice site is indicated by an arrow. The ESE2 and ESE3 *cis*-regulatory elements are shaded. Right: Proteins from the purified RNP complexes. The control J6f1 RNPs and the SLS2-J6f1 RNPs were separated by 10% SDS-PAGE and stained with colloidal Coomassie Blue. Protein markers were fractionated in the two lanes on the left and right sides, and their molecular weights in kDa are given. (b) Left: Secondary structure of the SLS2-MS2 RNA used for RNP purification. The 5' part of the RNA folds into the SLS2-A7 stem-loop structure as in panel A. The three binding sites for the MS2 coat protein added at the 3'-extremity of the molecule can form three small stem-loops structures. The A7 3' splice site (ss) is indicated by an arrow. ESE2 and ESE3 *cis*-regulatory elements are shaded. Additional nucleotides at the 5'-end

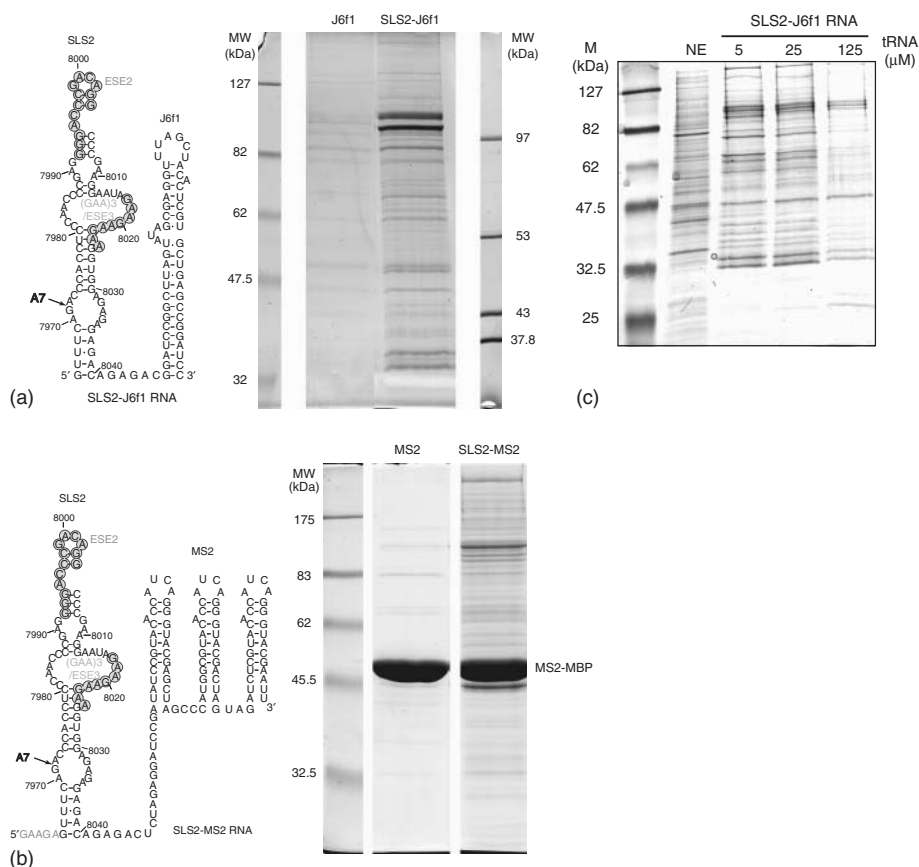
of SLS2-A7-MS2 RNA (gray characters) were included to improve SP6 RNA polymerase transcription. Right: Proteins from the purified RNP complexes, the control MS2 RNPs and the SLS2-MS2 RNPs were separated by 10% SDS-PAGE and stained with colloidal Coomassie Blue. Protein markers were fractionated in the lane on the left and their molecular weights in kDa are given. The position of the MS2-MBP fusion protein in the gel is indicated. (c) The effect of the addition of various concentrations of competitor tRNAs in the nuclear extract on the protein profile obtained after purification of SLS2-J6f1 RNPs by the tobramycin-based method. Complexes were assembled on the tobramycin–Sephacrose beads preloaded with the SLS2-J6f1 RNA in the presence of increasing concentrations of total yeast tRNA (5–125 μ M) added into the HeLa nuclear extract. The eluted complexes were analyzed by 10% SDS-PAGE and stained by colloidal Coomassie Blue. The molecular mass marker is shown on the left. Addition of 125 μ M tRNA strongly decreased the yield of RNP assembly. Therefore, we recommend the use of tRNA concentration inferior to 25 μ M.

RNAs are used for RNP reconstitution, only enzymatic probes are used and cleavages are mapped directly by gel electrophoresis without a reverse transcriptase step.

Chemical Probes DMS methylates RNAs at the N⁷-G, N¹-A, and N³-C positions of the bases. CMCT modifies RNAs at the N³-U and N¹-G positions, and kethoxal reacts at the N¹-G and N²-G positions. Only N⁷-G methylation by DMS can occur in double-stranded RNAs – all the other modifications are impaired.

Enzymes RNase T1 cleaves the phosphodiester bonds 3' to G residues, whereas RNase T2 cleaves after any residue. Both enzymes are used in conditions such that they preferentially cleave single-stranded RNA regions. RNase V1 is used to cleave double-stranded or stacked RNA regions. To identify the positions protected by proteins in an RNP, the naked RNA and the RNP are subjected to the same chemical and enzymatic treatments, and the reactive positions in RNA and RNP are compared. It should be noticed that in addition to RNA protection, RNP probing may detect some RNA conformational changes occurring upon protein binding [95].

Safety Rules DMS is a potential carcinogen and special care should be taken when using it.



44.2.5.2 Equipment, Material, and Reagents

Equipment

Sequencing gels for primer extension analysis.

Exposure with X-Ray films (Fuji or Kodak); visualization with a PhosphorImager equipment can also be used.

Temperature-controlled baths (96, 65, 30, and 20 °C).

Reagents

Probes: RNase T1 is purchased from Roche Diagnostics, RNase T2 from Mobitech (Germany), RNase V1 from Ambion, DMS from Aldrich, CMCT from Fluka and Kethoxal from Amersham.

Chemicals and enzymes: The avian myeloblastosis virus (AMV) reverse transcriptase is purchased from MP Biomedicals; glycogen (10 mg ml⁻¹), dNTPs

(100 mM of each) and ddNTPs are from MBI Fermentas. The yeast total tRNA (20 mg ml⁻¹) is from Roche Diagnostics; cacodylic acid and boric acid from Sigma.

Radiochemicals: [γ -³²P]ATP (3000 Ci mmol⁻¹) is purchased from Perkin Elmer.

Buffers

Buffer D: 20 mM HEPES-KOH (pH 7.9), 0.2 mM EDTA, 100 mM KCl, 20% glycerol (w/v). Add freshly prepared 0.5 mM DTT and 0.5 mM PMSF (dissolved in 96% ethanol).

Tris buffer: 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 2.5 mM MgCl₂

TBE buffer: 90 mM Tris-borate (pH 8.0), 2 mM Na₂EDTA

DMS/Ke buffer: mM sodium cacodylate (pH 7.5), 100 mM KCl, 2.5 mM MgCl₂

CMCT buffer: 50 mM sodium borate (pH 8.0), 100 mM KCl, 2.5 mM MgCl₂

DMS stop buffer: 1 M Tris-acetate (pH 7.5), 1.5 M sodium acetate, 1 M β -mercaptoethanol

Buffer A: 10 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 3 mM CaCl₂, 250 mM sucrose, 0.7 M β -mercaptoethanol

0.7 M ice cold β -mercaptoethanol

1 M EDTA (pH 8.0)

0.5 M potassium borate (pH 7.0)

0.5 M sodium cacodylate (pH 7.0)

3 M sodium acetate (pH 5.2)

10 x RT buffer: : 500 mM Tris-HCl (pH 8.3), 60 mM MgCl₂, 400 mM KCl (provided with the reverse transcriptase purchased from Q-BioGene).

44.2.5.3 Probing Method

Enzymatic and Chemical Probing of Native Purified RNP Particles Modifications and cleavages are performed in the purification or storage buffer.

To ensure statistical modifications and cleavages, all the reactions are performed in the presence of 1.25 μ g of yeast total tRNA in a total volume of 10 μ l.

Chemical Modifications

DMS

DMS modifications are performed for 6 min at 20 °C in 50 μ l of 1 x DMS/Ke buffer with 1 μ l of a DMS/EtOH solution (1/1, v/v).

Kethoxal

Same protocol as for DMS, except that modifications are performed for 10 min at 0 °C at a kethoxal concentration of 10 mg ml⁻¹.

CMCT

CMCT modifications are performed for 6 min at 20 °C in 50 μ l of 1 x CMCT buffer and at CMCT concentrations of 30–60 mM.

Enzymatic Cleavages RNase T1, T2, and V1 cleavages are performed for 6 min at 20 °C in 40 µl of 1 x Tris buffer with 0.025–2 U, 0.5–3 U, and 0.0005–0.05 U, respectively, of RNase.

Reaction Stop DMS modification is quenched by addition of 10 µl of DMS stop buffer (20% of the reaction mixture), followed by phenol extraction. CMCT modification is stopped by phenol extraction, followed by ethanol precipitation. Kethoxal modification is stopped by addition of 0.5 M potassium borate (pH 7.0) to stabilize the kethoxal-guanine adduct (25% of the reaction mixture volume), followed by phenol extraction and ethanol precipitation. RNase V1 digestion is stopped by the addition of 5 µl of 100 mM EDTA (pH 8.0) before phenol extraction. RNase T1 and T2 digestions are stopped by the addition of an excess of yeast total tRNA (10 µg), followed by rapid phenol extraction. To avoid reaction of the enzymes or chemical on the free RNA, phenol extractions should be quickly performed on ice.

Modification and digestion products are ethanol-precipitated, washed with 70% (v/v) ethanol, dried, and dissolved in milliQ water (except Ke-modified RNA pellets, which are dissolved in potassium borate 25 mM (pH 7.0) in order to stabilize the chemical adducts [96]).

Primer Extension Analysis For primer extension analysis, 5'-end-labeled primers are annealed to chemically modified or digested RNAs. As the RNA length that can be examined with one primer ranges between 100 and 200 nt, different primers (generally 12–20 nt) complementary to regions that are spaced by 100–200 nt have to be used. As each primer has its own efficiency for reverse priming, preliminary assays should be performed for each primer to define the amount of RNA, which is suitable for the analysis. 5'-end-labeling (with [γ -³²P]ATP, 3000 Ci mmol⁻¹) is described in [97]. Extension is achieved with the AMV reverse transcriptase in the presence of the four dNTPs, the conditions are described in [98].

Hybridization The RNA sample (1–10 pmol in 1 µl) is mixed with the 5'-end-labeled primer (100 000 cpm) and 10 x RT buffer, in a total volume of 2.5 µl. The mixture is incubated for 10 min at 65 °C and quickly cooled on ice for 10 min.

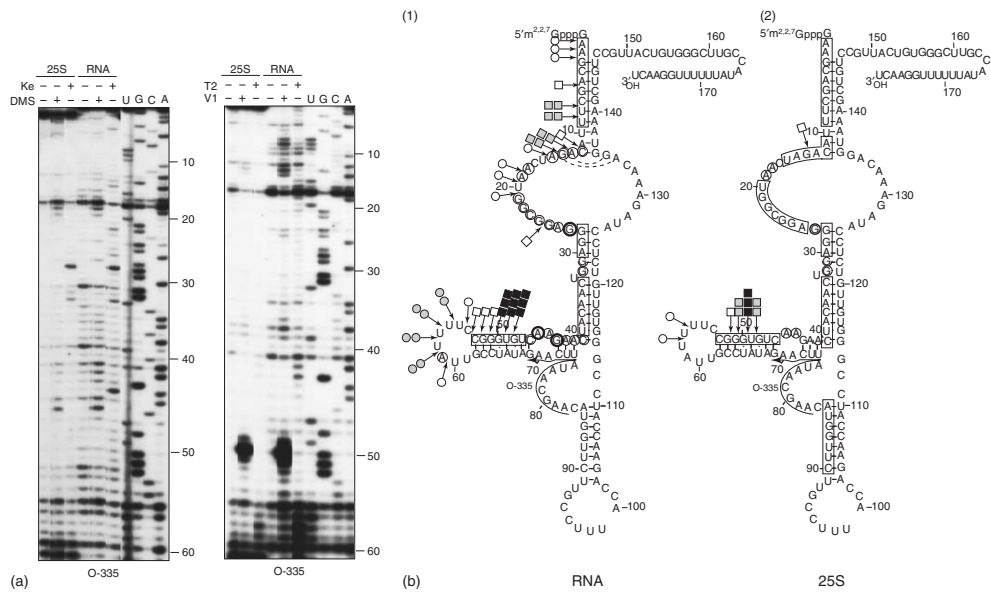
Primer Extension The primer extension reaction is performed in a final volume of 5 µl. The elongation mixture is prepared as follows: 0.1 µl of dNTP mixture (5 mM of each dNTP), 0.25 µl of 10 x RT buffer, 0.25 µl of reverse transcriptase (2 U µl⁻¹ dilution, freshly prepared before use), and 1.9 µl of H₂O. The hybridization mixture is mixed with 2.5 µl of elongation mixture and samples are incubated for 45 min at 42 °C. The primer extension is stopped by addition of 4 µl of DNA loading buffer. To prepare the sequencing ladder, the unmodified RNA is used as a template. The elongation mixture contains a ddNTP at a concentration of 0.5 mM and dNTPs with a dNTP: ddNTP ratio of 2 : 1.

Gel Fractionation The elongation mixture is denatured for 2 min at 96 °C and 2 µl aliquots are fractionated on a 7% denaturing (8 M urea) polyacrylamide (19 : 1 ratio acrylamide: bisacrylamide) sequencing gel in 1 x TBE. The gel is preheated for 30 min at 50 V cm⁻¹, and electrophoresis is performed at 50 °C using the same voltage. The migration time is adjusted to the sequence length to be read. After migration, gels are transferred on sheets of Whatman 3MM paper, dried and autoradiographed overnight at -80 °C with X-ray films (from Kodak or Fuji) and an intensifying screen (Amersham Biosciences).

Example 5: Native RNP probing protocol (Figure 44.5) The yeast U5 snRNA was probed as free RNA (RNA) and in the yeast [U4/U6.U5] tri-snRNP (25S) (Figure 44.5a). About 200 ng of RNP or 100 ng of free renatured RNA were incubated for 10 min at 20 °C in the presence of 1.25 µg of yeast total tRNA, in 50 µl of the digestion or modification buffer (1 x DMS/Ke buffer or 1 x Tris buffer). For kethoxal, the mixture was then put on ice for 10 min, as the reaction was performed at 0 °C. The reactions with DMS, Ke, RNases T2 and V1 were performed and stopped in the conditions described above. Aliquots of the treated RNAs were reverse transcribed with the 5'-labeled specific oligodeoxynucleotide primer O-335, complementary to positions 69–81 of the *S. cerevisiae* U5 snRNA (Figure 44.5b). The cDNA fragments obtained were fractionated on a sequencing gel, using as a reference a sequencing ladder performed with the unmodified RNA and the same 5'-end-labeled oligodeoxynucleotide (Figure 44.5a).

Enzymatic and Chemical Probing of RNP Particles Formed in Nuclear Extract or with Purified Proteins As for EMSA experiments, the RNA is renatured in 1 x buffer D, before complex formation, as described in paragraph 44.2.1.2.

Figure 44.5 Probing of U5 snRNA in the spliceosomal [U4/U6.U5] tri-snRNP purified from *S. cerevisiae* by use of chemical and enzymatic probes [68]. The yeast [U4/U6.U5] 25S tri-snRNP (25S, a) and the natural free U5 snRNA (RNA, a) were probed with kethoxal (Ke), DMS, RNase T2, and RNase V1 under conditions described in paragraph 44.2.5.3. (a) The primer extension analyses performed with the primer O-335. For each probe, a control experiment in the absence of the probe was performed (—). Lanes U, G, C, A correspond to the sequencing ladder. Positions of nucleotides in U5 snRNA are indicated on the right side of the panels. In (b), the probing data illustrated in (a) are schematically represented on the secondary structure of U5 snRNA (left 2D structure results for the naked RNA, right 2D structure results for the 25S tri-snRNP). Nucleotides modified by DMS or Ke are circled, the thickness of the circles indicate the levels of modification (weak, medium, and strong). RNase V1 or T2 cleavages are indicated by arrows linked to squares or circles, respectively. The intensity of gray scale and number of symbols indicate the yield of cleavage. Boxed nucleotides are not modified. In the U5 snRNA region that is analyzed with primer O-335, the tri-snRNP components generate a strong protection, except for the lateral stem-loop structure formed by residues 40–75. This stem-loop structure is specific to the yeast U5 snRNA and absent in vertebrate U5 snRNA [68].



*Enzymatic Reactions***RNase T1**

RNA (200 ng, 1 pmol) is mixed with 1 μ l of 62.5 mM MgCl_2 , 3.6 μ l of 1 x buffer D, 5 μ g of total tRNA, and the final volume is adjusted to 14 μ l with water. An adequate amount (based on EMSA) of HeLa cell NE is added and the reaction mixture is incubated at 30 °C for 10 min. The cleavage reaction is performed with 1 μ l of RNase T1 (0.025–2 U) at 30 °C for 10 min.

RNase T2

Same protocol as for RNase T1, except that the reaction is performed with 1 μ l of RNase T2 (1–3 U) at 30 °C for 10 min.

RNase V1

Same protocol as for RNase T1, except that the reaction is performed with 1 μ l of RNase V1 (0.02–0.2 U) at 30 °C for 10 min.

Chemical Modification by DMS The RNA (200 ng, 1 pmol) is mixed with 1 μ l of 62.5 mM MgCl_2 , 3.6 μ l of 1 x buffer D, 5 μ g of total tRNA are added and the final volume is adjusted with water to 14 μ l. An adequate amount of HeLa cell NE dialyzed against buffer D, as deduced from the EMSA experiment, is added and the reaction mixture incubated at 30 °C for 10 min. To improve the efficiency of DMS modification, sodium cacodylate at a final concentration of 50 mM (pH 7.5) is added in buffer D before the reaction. The modification reaction is performed at 30 °C for 10 min with quantities comprised between 1 to 5 μ l of a 1/1 (v/v) DMS/EtOH solution.

Reaction Stop Same protocol as described in Section 44.2.5.3. The hydrolysis reactions by RNases T1, T2, and V1 are stopped by the addition of 20 μ g of total yeast tRNA; in addition, 1 μ l of 100 mM EDTA is added for RNase V1. This is followed by phenol extraction on ice. DMS modifications are stopped by addition of 10 μ l of DMS stop buffer before phenol extraction and ethanol precipitation.

Ethanol precipitation of all samples is done by addition of 10 μ l of 3 M sodium acetate (pH 5.2), 1 μ g of glycogen, and at least three volumes of 96% EtOH, followed by a 15 min incubation at –80 °C. After centrifugation at 13 000 rpm for 15 min and at 4 °C, the supernatants are discarded and the RNA pellets are washed with 200 μ l of 70% EtOH. A second centrifugation is performed for 5 min at 13 000 rpm and 4 °C. The RNA pellets are dried for 2 min, and dissolved in 4 μ l of MilliQ water before primer extension analysis (see Section 44.2.5.3).

Example 6: Protocol used for probing of a reconstituted RNP complex (Figure 44.6) The complexes formed upon incubation of the HIV-1 A7 SLS1,2,3 RNA fragment (positions 7903–8170) in HeLa cell NE or with purified hnRNP A1 protein were analyzed by chemical and enzymatic probing of the RNA structure and accessibility (Figure 44.6). The following conditions were used: 1.12 pmol of cold HIV-1 A7 SLS1,2,3 RNA was incubated in the presence of 5 μ g of yeast total tRNA with 1 μ l of 62.5 mM MgCl_2 and 3.6 μ l of 1 x buffer D in a total volume of 14 μ l. Assays were performed in the presence (+) of 4 μ l of NE ([Protein]/[RNA]

(P/R) = 46) (Figure 44.5a) or 50 fmol of purified hnRNP A1 protein (P/R = 45) (Figure 44.5b) or in the absence of added extract or protein (–) (Figure 44.5a,b). After a 10 min incubation at 30 °C, 0.025 U of RNase T1 were added. For identification of reverse transcriptase pauses that are not because of RNA cleavage, a control primer extension was performed on the naked RNA incubated in buffer D without RNase T1. The reactions were stopped by addition of 20 µg of tRNAs, followed by phenol extraction as described in Section 44.2.5.3.

Probing of Yeast RNAs Modified *In vivo* by DMS Treatment DMS is able to penetrate efficiently in bacterial, yeast, and animal cells and can thus be used to probe RNA structure and accessibility in these cells. However, to get interpretable results the experiments should be performed on RNAs that form well-defined homogeneous particles in cells. Otherwise, information on RNA molecules having different structures and accessibilities would be superimposed in the primer extension analysis. The *in vivo* analysis of the *S. cerevisiae* U3 snoRNA structure and accessibility in the U3 snoRNP illustrated in Figure 44.8 was a very successful example of an RNP analysis *in vivo*. The methylation sites detected *in vivo* were compared to those detected *in vitro* by treatment of a partially purified U3 snoRNP and the naked *in vitro* transcribed U3 snoRNA [50]. The results obtained validated the protein binding sites identified by *in vitro* analysis and demonstrated the interaction of U3 snoRNA with the pre-ribosomal RNA [50].

Example 7: Protocol used for *S. cerevisiae* U3 snoRNP probing by DMS *in vivo* (Figure 44.7) About 15 ml aliquots of an *S. cerevisiae* culture, grown in YEPD medium at 30 °C to an A_{600} between 0.5 and 1.0, were gently rocked at room temperature for 2 min in the presence of DMS at concentrations between 40 and 160 mM. The reaction was quenched by addition of 0.7 M ice cold β -mercaptoethanol and 5 ml of cold water-saturated isoamyl alcohol, followed by shaking and centrifugation. Cell pellets were washed with buffer A (10 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 3 mM $CaCl_2$, containing 250 mM sucrose and 0.7 M β -mercaptoethanol). Cells were centrifuged for 10 min at 2500 rpm and 4 °C. Total RNA was extracted by the method described by Domdey *et al.* [99]. Sites of RNA methylation were mapped by primer extension using the 5'-end-labeled oligodeoxynucleotides complementary to two distinct regions of the yeast U3 snoRNA [50], and 10 µg of total RNA as the template.

44.2.6

UV Crosslinking and Immunoselection

Formation of covalent bonds between RNA and proteins can be established by incorporation of photoactivable residues such as 4-thiouridine in the course of reverse transcription [100]. However, the easiest way to test for a very near proximity between RNA and proteins is to use UV irradiation at 254 nm [47, 48]. Although the yield of crosslinking is low, it is sufficient to detect RNA–protein contacts with RNA molecules labeled at high specific activity.

Figure 44.6 Probing of the RNA structure and accessibility in RNP formed by the HIV-1 A7 SLS2,3 RNA and a HeLa cell nuclear extract (NE) or the purified hnRNP A1 protein (A1) [48]. Primer extension analyses of the A7 SLS1,2,3 RNA cleaved by RNase T1 in buffer D in the absence (–) or presence (+) of nuclear extract (a1) or in the presence of the recombinant hnRNP A1 protein (b1) are shown. As a control, a primer extension was performed using the intact RNA transcript incubated without RNase T1 as the template (left lane of the autoradiogram). Lanes U, G, C, A correspond to the sequencing ladder. Positions of nucleotides in the HIV-1BRU RNA and of the RNA secondary structure elements described for the HIV-1 A7 RNA region are indicated on the right side of the autoradiograms. (a2) and (b2) are schematic representation of the probing data illustrated in (a1) and (b1). Positions of RNase T1 cleavages are represented on the RNA secondary structure

established for the A7 SLS1,2,3 RNA [48]. Cleavages 3'- to the G residues are indicated by thin lines. They are surmounted with circles when the cleavages occurred in the naked RNA (a2 and b2). Cleavages occurring in the presence of nuclear extract (a2) or hnRNP A1 (b2) are indicated by the presence of a square. The colors of circles and squares indicate the level of cleavage observed in the naked RNA and the RNP complex, respectively (gray, dark gray, and black represent increasing intensities of cleavages, respectively). The G residues protected either in the presence of nuclear extract (a2) or with hnRNP A1 protein (b2) are circled and the intensity of the circle corresponds to the yield of the protection. The hybridization site of the oligodeoxynucleotide O-2382 is indicated. The *cis*-regulatory elements of splicing, acting at site A7 (ESE2, ESE3/(GAA)3, ESS3) are delimited by two opposite broken arrows and the name of the element is given.

44.2.6.1 Equipment, Materials, and Reagents

Equipment Electrophoresis instruments for small vertical slab gels. Exposure of gels can be done with X-Ray films (Fuji or Kodak) or a PhosphorImager.

Temperature-controlled baths (65, 50, 30, 20 °C).

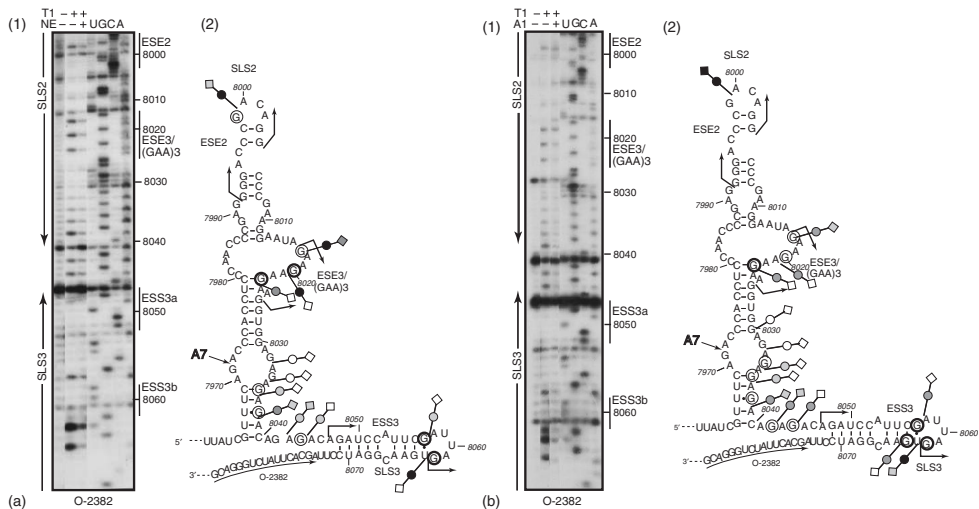
Materials Stratalink apparatus (Stratagene) and 96-well U- or V-shaped non-adsorbing plates (VWR).

Reagents

- Antibodies: Some of the primary antibodies used in the described examples can be purchased from Immuguest.
- Chemicals and enzymes: Protein G–Sepharose, protein A–Sepharose, Hybond C nitrocellulose membrane, and ECL detection system are purchased from GE Healthcare; Bovine serum albumin (B8894) is from Sigma; RNase A is from Roche Diagnostics.

Buffers

- Buffer D: 20 mM HEPES-KOH (pH 7.9), 0.2 mM EDTA, 100 mM KCl, 20% glycerol (w/v). Add freshly prepared 0.5 mM DTT and 0.5 mM PMSF (dissolved in 96% ethanol).
- SDS-PAGE loading buffer: 62.5 mM Tris-HCl (pH 6.8), 2% SDS, 100 mM 2-β-mercaptoethanol, 10% glycerol, 0.01% bromophenol blue



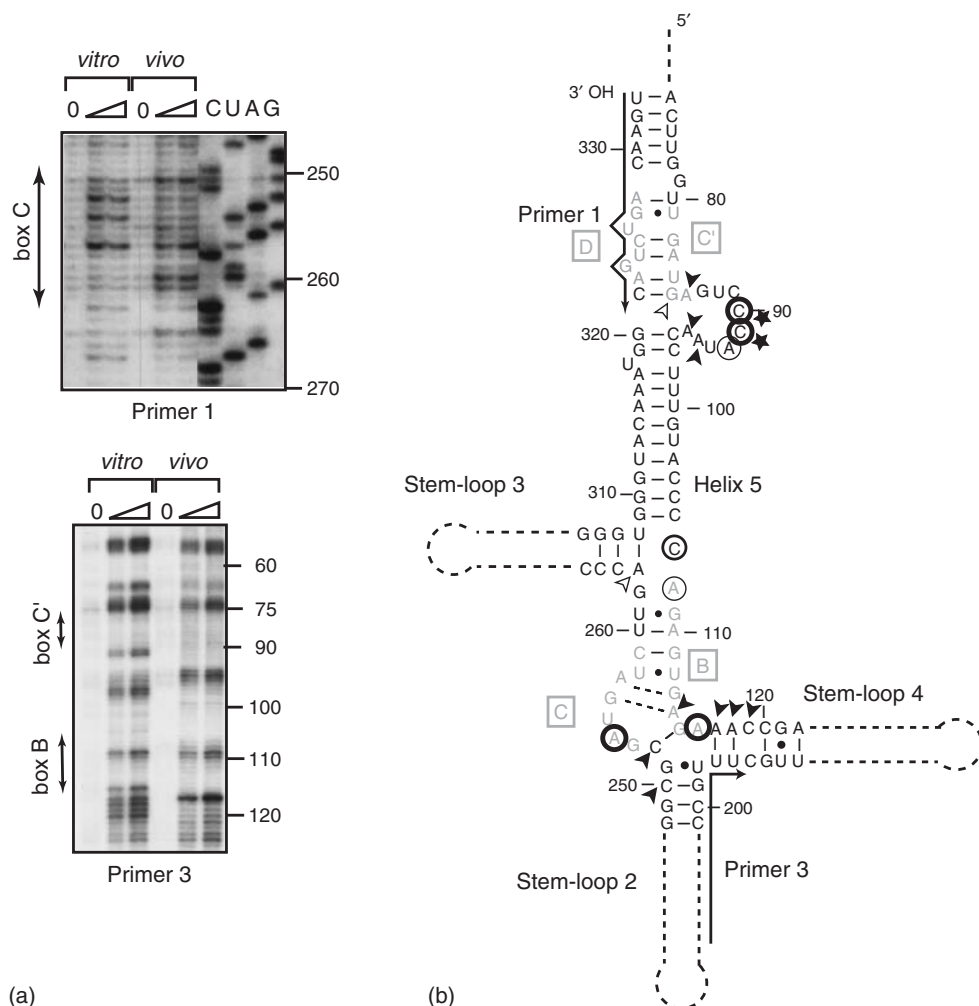


Figure 44.7 Probing of the structure and the accessibility of the *S. cerevisiae* U3 snoRNA *in vivo* [50]. Primer extension analyses of the U3 snoRNA modified by DMS *in vitro* or *in vivo* are presented in (a). As a control, primer extensions were performed with an untreated U3 snoRNA transcript (lanes 0). Lanes U, G, C, A correspond to the sequencing ladders. Positions of nucleotides in the *S. cerevisiae* U3 snoRNA sequence and the phylogenetically conserved RNA segments of U3 snoRNA [50] are shown on the right and left sides of the

autoradiogram, respectively. In (b) the experimental data shown in (a) are represented on the secondary structure proposed for the *S. cerevisiae* U3 snoRNA in interaction with the pre-rRNA [22, 50]. Nucleotides modified *in vitro* and protected *in vivo* are indicated by black triangles, nucleotides modified *in vitro* and protected *in vivo* are circled, the thickness of the circles reflect the yield of modification *in vivo*. Asterisks mark the nucleotides with an increased reactivity *in vivo* compared to *in vitro*.

- IP buffer: 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.1% (v/v) Nonidet P-40
- PBS-TM: PBS containing 0.1% Tween 20 and 5% dry nonfat milk powder.

44.2.6.2 UV-Crosslinking Method

Coating 96-Well Plates In order to avoid non-specific interactions, each well is coated with 20 μ g of bovine serum albumin diluted in 1 x buffer D in a total volume of 25 μ l, for 30 min at room temperature. This mixture is then discarded and each well is washed three times with 50 μ l of 1 x buffer D.

UV Crosslinking Formation of RNP complexes is performed as described in Section 44.2.2. Note that only RNA transcripts, which are uniformly labeled by incorporation of an [α - ^{32} P]NTP, can be used. Selection of the labeled nucleotide is decided after inspection of the nucleotide sequence of the RNA segment expected to interact with the proteins. For instance, if this fragment contains several uridines, labeling will be done with [α - ^{32} P]UTP; if numerous adenines are present, [α - ^{32} P]ATP will be used.

Reaction mixtures containing RNP complexes are transferred to a 96-well plate placed on ice, and irradiation is performed at 254 nm at an intensity of 2 J cm $^{-1}$ in a Stratalinker apparatus (Stratagene). After irradiation, the RNP complexes are transferred to microcentrifuge tubes, 15 μ g of RNase A is added for complete digestion, and incubation is done for 20 min at 37 °C. The free and crosslinked proteins can then be fractionated by 8–10% SDS-PAGE, or they can be purified by immunoselection before gel electrophoresis (see below). After electrophoresis, the gel is dried and autoradiographed as previously described.

Example 8: Crosslinking experimental protocol (Figure 44.8) About 50 fmol (about 500 000 cpm) of HIV-1 A7 SLS2 WT or C3 RNA (Figure 44.7a) uniformly labeled by incorporation of [α - ^{32} P]-UTP were incubated with 5 μ l of NE or with a mixture of the two recombinant ASF/SF2 and hnRNP A1 proteins, each at a final concentration of 3 μ M (1 : 1 ratio). The incubation was performed in conditions established for RNP formation (Figure 44.8). After a 10 min irradiation at 254 nm, the RNA was digested with 50 U of RNase T1 for 1 h at 37 °C, followed by a 30 min incubation at 50 °C. Then, 8 μ l of SDS-PAGE loading buffer were added, and the crosslinked products is fractionated on a 10% SDS-polyacrylamide gel for 1–2 h at 20 V cm $^{-1}$ and visualized by autoradiography. Before SDS-PAGE loading buffer addition, one half of the crosslinked products was immunoselected with antibodies bound to Sepharose beads.

Immunoselection RNase activities are often present in sera and media. It is thus necessary to bind the antibodies to protein (G or A)-Sepharose beads, and to eliminate the possible RNase contaminations by washing the beads. The washing solution contains a large amount of bovine serum albumin that coats the non-specific protein binding sites of the beads.

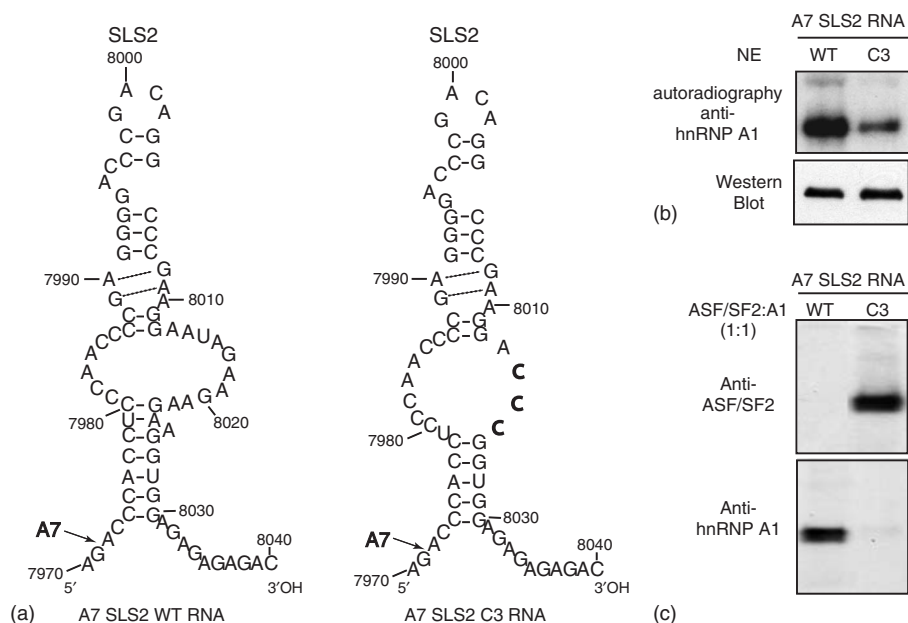


Figure 44.8 Crosslinking followed by immunoselection of the hnRNP A1 and ASF/SF2 proteins on the WT and C3 variant of the HIV-1 A7 SLS2 RNAs (positions 7971–8040 of HIV-BRU RNA) by irradiation at 254 nm [48]. (a) The secondary structure proposed for the A7 WT and C3 variant SLS2 RNAs are shown [48]. Bold characters indicate the CCC sequence, that is substituted for the AUA(GAA)3 sequence in the C3 variant. (b) The RNP complexes formed by the WT and C3 A7 SLS2 RNAs and proteins of a HeLa cell nuclear extract (NE) were UV irradiated at 254 nm. RNAs were digested with T1 RNase and the hnRNP A1 proteins were immunoselected on Sepharose beads coated with anti-hnRNP A1 antibodies. One half of the selected proteins were fractionated by electrophoresis on a 10% SDS-polyacrylamide gel (upper panel). The radiolabeled hnRNP A1 protein was detected by autoradiography. The second half of the

selected hnRNP A1 proteins was fractionated on another SDS-polyacrylamide gel used for Western blot analysis (lower panel). hnRNP A1 protein from HeLa cell nuclear extract was bound to the WT RNA and crosslinked at high efficiency; in contrast, crosslinking efficiency was very low for the C3 RNA. RNP complexes were formed by incubation of the WT and C3 variant SLS2 RNAs with equimolar amounts of proteins hnRNP A1 and ASF/SF2 and were UV irradiated at 254 nm. After RNA digestion, the hnRNP A1 and ASF/SF2 protein were immunoselected on Sepharose beads and fractionated by a 10% SDS-polyacrylamide gel. The radiolabeled hnRNP A1 and ASF/SF2 proteins were detected by autoradiography (c) hnRNP A1 protein crosslinked only to the WT RNA, but not to the C3 RNA, whereas vice versa, ASF/SF2 protein crosslinked only to the C3 RNA, but not to the WT RNA.

The protein (G or A)–Sepharose beads are coated with the antibodies. The digested crosslinked products are then incubated with the coated beads in the presence of the immunoselection buffer. Subsequently, the beads are washed with the immunoselection buffer containing 0.25% Nonidet P-40, suspended in the

SDS-PAGE loading buffer and heated for 5 min at 96 °C to elute bound proteins. The proteins are further fractionated by an SDS-PAGE (8–10% gel).

Each sample of eluted proteins is divided in two parts. One part is fractionated on a gel analyzed by autoradiography. The amount of radiolabeled protein is estimated by a PhosphorImager. The second half of the eluted proteins is fractionated on a gel used for immunoblotting with the antibody that was coated on the beads. This allows an estimation of the total amount of the protein retained on the beads. The ratio between the amount of protein estimated by immunoblotting and the radioactivity detected in the gel gives an indication on the level of crosslinking.

Example 9: Immunoselection protocol (Figure 44.8) About 20 µl of a protein (G or A)–Sephacrose beads suspension was coated for 2 h at 4 °C with 1 µl of anti-hnRNP A1 (4B10) or 2 µl anti-ASF/SF2 antibodies provided by G. Dreyfuss (University of Pennsylvania School of Medicine, Philadelphia, USA) and J. Stevenin (IGBMC, Strasbourg, France), respectively. The digested crosslinked products (Section 44.2.6.2) formed on A7 SLS2 WT and C3 RNAs (Figure 44.7a) were incubated with the coated beads by end-over-end rotation for 2 h at 4 °C in 400 µl of immunoselection buffer containing 0.1 mg ml⁻¹ bovine serum albumin. Subsequently, the beads were washed three times by incubation with 500 µl of the immunoselection buffer containing 0.25% Nonidet P-40, followed by centrifugation (2 min at 3000 rpm at room temperature). At this stage, the beads were suspended in 20 µl of SDS-PAGE loading buffer and boiled for 5 min to elute bound proteins. The proteins were further fractionated by SDS-PAGE (10% gel) for 1–2 h at 20 V cm⁻¹.

Each sample of the eluted proteins was divided in two parts: One part was fractionated by electrophoresis, and the amount of radiolabeled protein was estimated by PhosphorImager scanning (Figure 44.8b,c); the other part was fractionated on a gel and subjected to immunoblotting (Figure 44.8b).

44.3

Commentaries and Pitfalls

44.3.1

RNP Purification and Reconstitution

44.3.1.1 RNA Purification and Renaturation

In order to obtain an RNA transcript of homogeneous size, a step of purification based on denaturing PAGE (8 M urea) is advisable. Indeed, minor RNA degradation or abortive transcription products can be generated during transcription.

A homogeneous conformation of the RNA molecules is required. Thus, the RNA renaturation procedure before complex formation and probing experiment is a critical step. Alternative conformations lead to the simultaneous detection of single-stranded and double-stranded specific cleavages at some of the positions.

44.3.1.2 EMSA

Some naked RNAs are resolved in two bands in EMSA gels because of the occurrence of two distinct conformations or dimerization. If the RNA used corresponds

to a fragment of a larger RNA, it is advisable in this case to prepare another fragment by choosing other 5'- and 3'- extremities. Templates for production of small RNA may be produced by use of partially complementary DNA oligodeoxynucleotide primers and DNA polymerase to form a full-length double-stranded DNA.

For probing experiments, it is very important to get a complete and homogeneous formation of the RNPs. Otherwise, heterogeneous probing data will be obtained that will be difficult to interpret.

44.3.1.3 Tobramycin–Sephacrose Affinity Chromatography

This has to be performed at rather elevated pH as the specific binding of tobramycin aptamer J6f1 to the antibiotic relies on interaction with free amino groups of the tobramycin molecule. At lower pH, almost any RNA may be retained on the tobramycin–Sephacrose because of unspecific electrostatic interactions between negatively charged RNA phosphate groups and cationic NH_3^+ groups of the antibiotic. In addition, the protonated tobramycin can also non-specifically retain some proteins by electrostatic interactions. The reported pK_a values of tobramycin's amino groups are in the range of 6.2–8.6 [94], thus the retention of the J6f1 aptamer on tobramycin–Sephacrose is generally performed at pH values above 9.0.

In contrast, a wide range of conditions can be used for MS2–MBP affinity chromatography. MBP binds amylose beads very tightly at pH values between 6.5 and 8.5 and at ionic strength up to 1 M NaCl. Nonionic detergents such as Tween 20 or Triton X-100 may be used at low concentrations (0.05%) [101]. The interaction of MS2 RNA/MS2 coat protein is relatively stable at moderate salt concentrations (up to 300 mM NaCl).

44.3.2

Probing Conditions

The probing conditions (pH, ionic strength, Mg^{2+} concentration, temperature, probe concentration) have to be defined by several preliminary tests. The stability of the RNP in these various conditions has to be tested by electrophoresis on native gel [68].

44.3.2.1 Choice of the Probes Used

Some of the widely used probes for RNA 2D structure (such as CMCT, Pb^{2+} , Fe-EDTA(OH), and nuclease S1) cannot be used when the assays are directly performed in nuclear or cellular extracts. The reasons for this are given below.

- The presence of a NE inhibits RNA modification by CMCT, probably due to its accelerated hydrolysis or interactions with other components of the extract. Note that CMCT can be used on purified RNP complexes.
- $\text{Pb}(\text{II})$ (Pb^{2+}) cannot be used because of the presence of chloride ions in the RNP buffers, which induces precipitation and inhibits the reaction. However, it can be used for purified complexes using a suitable buffer and also *in vivo*, using defined conditions.

- Fe-EDTA(OH) cannot be used as hydroxylated compounds, such as glycerol present in the NE and buffer D, are known to be OH[•] scavengers and can inhibit OH[•]-mediated cleavage. In addition, Tris or HEPES buffers also reduce RNA cleavage by hydroxyl radicals, presumably by acting as free-radical scavengers. Again, Fe-EDTA(OH) can be used on purified RNPs in an appropriate buffer.
- Nuclease S1 cannot be used for any kind of RNP probing as its optimal pH of action is 4.5, which is generally deleterious for RNP complexes [67].

It should also be taken into account that the overall efficiency of a given enzymatic or chemical probe may be considerably diminished in extract, and this apparent decrease in efficiency of cleavage or modification does not necessarily represent RNA protection due to protein binding. For instance, DMS reacts with both RNA and proteins; thus, DMS has a large number of substrates when modifications are directly performed in extracts, which may decrease its activity on RNA.

To get informative data by RNP probing, the probes should be selected so that reactive residues or sensitive phosphodiester bonds are present all along the RNA molecule. It is sometimes difficult to fulfill these conditions in case of highly structured RNAs. The reactivity of the naked RNA with the selected probes should be sufficient to see clear variations upon protein binding. Hence, in order to define conditions suitable for RNP analyses, one first has to test different conditions of modifications and cleavages on the naked RNA. The effects of different parameters can be examined in these preliminary tests, such as the [RNA]/[probe] ratio, time of incubation, temperature of incubation, and Mg²⁺ ion concentration. The conditions given in this paper were found to be suitable for analyses of several RNP. However, if they turn out to be unsuitable, they can be used as a starting point to look for other more favorable conditions.

44.3.2.2 Ratio of RNA/Probes

The added exogenous tRNA (usually yeast total tRNA) minimizes the non-specific interactions between RNA and proteins, and is also used to get a defined ratio of [RNA]/[Probe].

Another difficulty that may be encountered in the course of RNP probing is a very strong protection generated by the proteins. They may mask very large parts of the RNA. This may be the case for probing with extracts containing a large number of proteins or with proteins such as hnRNP A1 that are able to multimerize along the RNA [48]. To get convincing probing data one needs to be sure that the protections observed are not simply due to global inhibition of the activity of the probes. To this end, it is necessary to obtain nearly unaltered modifications and cleavages, together with strongly diminished ones, in the same experiment. Here, again, conditions of reaction often have to be adapted to obtain such contrasted modifications of the reactivities along the RNA molecule. When proteins of the RNP generate very strong protections, we recommend the use of chemical probes, which are less sensitive to steric hindrance.

Sometimes, new cleavage sites are observed in RNP compared to naked RNA. This may reflect an RNA conformational change. However, one has to verify that

these additional or reinforced cleavages are not because of the presence of an RNase activity present in the extract or the purified protein.

44.3.3

UV Crosslinking

44.3.3.1 Photoreactivity of Individual Amino Acids and Nucleotide Bases

Crosslink formation depends on the photoreactivity of both individual amino acids and nucleotide bases. Pyrimidine residues form covalent bonds with protein more efficiently compared to purine residues. Thus, upon UV irradiation, a protein that binds a purine-rich sequence may be undetectable by UV crosslinking.

44.3.3.2 Labeled Nucleotide in RNA

Owing to the low level of UV-induced crosslinking, the specific activity of the labeled RNA must be high enough for detection by autoradiography of the crosslinked residues bound to the protein. The choice of the labeled nucleotide and the RNase used for the digestion should be made by taking into consideration the nucleotide sequence of the expected binding site.

44.3.4

Immunoprecipitations

The amount of serum, or antibodies, required for complete precipitation of a particular protein has to be determined for each individual batch of serum/antibody.

44.3.4.1 Efficiency of Immunoabsorbents for Antibody Binding

Antibodies from humans, rabbits, or guinea pigs bind better and have a stronger affinity for protein A than those from mouse or rat. Binding to protein G provides a convenient alternative for the use of mouse and rabbit antibodies [102]. Poor binding of antibodies to proteins A and G can be circumvented by the use of secondary antibodies (for instance, anti-mouse immunoglobulin raised in rabbits) that do bind to protein A. Alternatively, the secondary antibodies can be directly coupled to CNBr-activated Sepharose. These coupled secondary antibodies will then serve as efficient adsorbents.

44.4

Troubleshooting

44.4.1

RNP Purification by Tobramycin–Sepharose or MS2-MBP Affinity Chromatography

- RNA is not retained on the tobramycin–Sepharose beads: Check renaturation conditions for RNA and predicted folding of RNA-J6f1 aptamer fusion.
- Non-specific proteins are eluted from tobramycin–Sepharose beads: Check pH of working buffer under separation conditions. Insure that final pH is close or

slightly above of 9.0. Increase the ionic strength of wash solution or amount of RNA competitor (tRNA).

- MS2 RNA is not retained on amylose column: Check the integrity of fusion MS2–MBP protein; upon storage, this fusion may undergo a proteolytic cleavage in the linker between two functional domains.

44.4.2

RNP Reconstitution

- No RNP complex is observed: First modify the [RNA]/[protein] ratio used for the reconstitution assay; if there is still a problem, decrease the quantity of competitor tRNA used.
- If the RNP complex does not penetrate in the gel, use a lower polyacrylamide concentration.

44.4.3

RNA Probing

- High smearing in the gel: Decrease the amount of loaded material or digest the RNA with a DNase-free RNase.
- No elongation stop signal is detected in primer extension analysis of the modified or cleaved RNA: Decrease the amount of probe use.
- No protection is observed: Check if the RNP complex is formed and stable under the conditions use.
- Too much protections of the RNA: Decrease the quantity of purified protein or cellular extract added.

44.4.4

UV Crosslinking

- No crosslinked proteins are obtained: Verify that the UV wavelength is correct and that the UV light is still working.
- No crosslinked proteins are immunoselected: If the Western blot analyses indicate that immunoselection occurred, try to use another labeled nucleotide. If labeled proteins are still not immunoselected, you have to use multiple approaches to understand the parameters that govern binding of the protein to the RNA target.

44.4.5

Immunoprecipitations

- High background on the membrane: It is recommended to use less serum or antibodies.

- No signal is detected on the membrane: Make sure that primary and/or secondary antibodies are still active.

Acknowledgments

S. Jacquenet and A. Clery are acknowledged for providing materials for illustrations of EMSA experiments. V. Marchand was a fellow of the French Ministère de la Jeunesse, de l'Éducation Nationale et de la Recherche. C. Branlant, I. Behm-Ansmant, A. Méreau, and A. Mougin are permanent scientists from the Centre National de la Recherche Scientifique. I. Motorin is Professor at Lorraine University.

References

1. Burd, C.G. and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.
2. Hall, K.B. (2002) RNA-protein interactions. *Curr. Opin. Struct. Biol.*, **12**, 283–288.
3. Siomi, M.C., Mannen, T., and Siomi, H. (2010) How does the royal family of Tudor rule the PIWI-interacting RNA pathway?. *Genes Dev.*, **24**, 636–646.
4. Steitz, T.A. and Moore, P.B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.*, **28**, 411–418.
5. Hoogstraten, C.G. and Sumita, M. (2007) Structure-function relationships in RNA and RNP enzymes: recent advances. *Biopolymers*, **87**, 317–328.
6. Query, C.C., Bentley, R.C., and Keene, J.D. (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell*, **57**, 89–101.
7. Kenan, D.J., Query, C.C., and Keene, J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.*, **16**, 214–220.
8. Clery, A., Blatter, M., and Allain, F.H. (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.*, **18**, 290–298.
9. Siomi, H., Matunis, M.J., Michael, W.M., and Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.*, **21**, 1193–1198.
10. Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
11. Grishin, N.V. (2001) KH domain: one motif, two folds. *Nucleic Acids Res.*, **29**, 638–643.
12. Stefl, R., Skrisovska, L., and Allain, F.H. (2005) RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep.*, **6**, 33–38.
13. Lunde, B.M., Moore, C., and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell. Biol.*, **8**, 479–490.
14. Valverde, R., Edwards, L., and Regan, L. (2008) Structure and function of KH domains. *FEBS J.*, **275**, 2712–2726.
15. Mao, H., White, S.A., and Williamson, J.R. (1999) A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nat. Struct. Biol.*, **6**, 1139–1147.
16. Watkins, N.J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C., and Lührmann, R. (2000) A common core RNP structure shared between the small nucleolar box C/D RNPs and

- the spliceosomal U4 snRNP. *Cell*, **103**, 457–466.
17. Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R., and Ficner, R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell*, **6**, 1331–1342.
 18. Klein, D.J., Schmeing, T.M., Moore, P.B., and Steitz, T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214–4221.
 19. Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
 20. Tran, E.J., Zhang, X., and Maxwell, E.S. (2003) Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J.*, **22**, 3930–3940.
 21. Rozhdestvensky, T.S., Tang, T.H., Tchirkova, I.V., Brosius, J., Bachelier, J.P., and Huttenhofer, A. (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.*, **31**, 869–877.
 22. Marmier-Gourrier, N., Clery, A., Senty-Segault, V., Charpentier, B., Schlotter, F., Leclerc, F., Fournier, R., and Branlant, C. (2003) A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA*, **9**, 821–838.
 23. Boulon, S., Marmier-Gourrier, N., Pradet-Balade, B., Wurth, L., Verheggen, C., Jady, B.E., Rothe, B., Pescia, C., Robert, M.C., Kiss, T. *et al.* (2008) The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *J. Cell Biol.*, **180**, 579–595.
 24. Fierro-Monti, I. and Mathews, M.B. (2000) Proteins binding to duplexed RNA: one motif, multiple functions. *Trends Biochem. Sci.*, **25**, 241–246.
 25. Doyle, M. and Jantsch, M.F. (2003) Distinct *in vivo* roles for double-stranded RNA-binding domains of the *Xenopus* RNA-editing enzyme ADAR1 in chromosomal targeting. *J. Cell Biol.*, **161**, 309–319.
 26. Mueller, G.A., Miller, M.T., Deroose, E.F., Ghosh, M., London, R.E., and Hall, T.M. (2010) Solution structure of the Drosha double-stranded RNA-binding domain. *Silence*, **1**, 2.
 27. Lingel, A. and Sattler, M. (2005) Novel modes of protein-RNA recognition in the RNAi pathway. *Curr. Opin. Struct. Biol.*, **15**, 107–115.
 28. Kiledjian, M. and Dreyfuss, G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.*, **11**, 2655–2664.
 29. Theunissen, O., Rudt, F., Guddat, U., Mentzel, H., and Pieler, T. (1992) RNA and DNA binding zinc fingers in *Xenopus* TFIIIA. *Cell*, **71**, 679–690.
 30. Puglisi, J.D., Chen, L., Blanchard, S., and Frankel, A.D. (1995) Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science*, **270**, 1200–1203.
 31. Cusack, S. (1999) RNA-protein complexes. *Curr. Opin. Struct. Biol.*, **9**, 66–73.
 32. Ranjan, M., Tafuri, S.R., and Wolffe, A.P. (1993) Masking mRNA from translation in somatic cells. *Genes Dev.*, **7**, 1725–1736.
 33. Fabrizio, P., Esser, S., Kastner, B., and Lührmann, R. (1994) Isolation of *S. cerevisiae* snRNPs: comparison of U1 and U4/U6.U5 to their human counterparts. *Science*, **264**, 261–265.
 34. Bach, M., Bringmann, P., and Lührmann, R. (1990) Purification of small nuclear ribonucleoprotein particles with antibodies against modified nucleosides of small nuclear RNAs. *Methods Enzymol.*, **181**, 232–257.
 35. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, **24**, 218–229.
 36. Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2002) Protein

- composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16719–16724.
37. Neubauer, G., Gottschalk, A., Fabrizio, P., Seraphin, B., Lührmann, R., and Mann, M. (1997) Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 385–390.
 38. Dragon, F., Gallagher, J.E., Compagnone-Post, P.A., Mitchell, B.M., Porwancher, K.A., Wehner, K.A., Wormsley, S., Settlege, R.E., Shabanowitz, J., Osheim, Y. *et al.* (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature*, **417**, 967–970.
 39. Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.*, **26**, 5528–5543.
 40. Expert-Bezancon, A., Le Caer, J.P., and Marie, J. (2002) Heterogeneous nuclear ribonucleoprotein (hnRNP) K is a component of an intronic splicing enhancer complex that activates the splicing of the alternative exon 6A from chicken beta-tropomyosin pre-mRNA. *J. Biol. Chem.*, **277**, 16614–16623.
 41. Dreumont, N., Hardy, S., Behm-Ansmant, I., Kister, L., Branlant, C., Stevenin, J., and Bourgeois, C.F. (2010) Antagonistic factors control the unproductive splicing of SC35 terminal intron. *Nucleic Acids Res.*, **38**, 1353–1366.
 42. Zhou, Z., Sim, J., Griffith, J., and Reed, R. (2002) Purification and electron microscopic visualization of functional human spliceosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12203–12207.
 43. Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N., and Moore, M.J. (2002) Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*, **8**, 426–439.
 44. Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Lührmann, R. (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.*, **26**, 1737–1748.
 45. Maenner, S., Bland, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianferani, S., Van Dorsaelaer, A., Clerc, P., Avner, P. *et al.* (2010) 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol.*, **8**, e1000276.
 46. Reed, R. (1990) Protein composition of mammalian spliceosomes assembled *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 8031–8035.
 47. Jacquenet, S., Mereau, A., Bilodeau, P.S., Damier, L., Stoltzfus, C.M., and Branlant, C. (2001) A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP. *J. Biol. Chem.*, **276**, 40464–40475.
 48. Marchand, V., Mereau, A., Jacquenet, S., Thomas, D., Mougin, A., Gattoni, R., Stevenin, J., and Branlant, C. (2002) A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding. *J. Mol. Biol.*, **323**, 629–652.
 49. Hallay, H., Locker, N., Ayadi, L., Ropers, D., Guittet, E., and Branlant, C. (2006) Biochemical and NMR study on the competition between proteins SC35, SRp40, and heterogeneous nuclear ribonucleoprotein A1 at the HIV-1 Tat exon 2 splicing site. *J. Biol. Chem.*, **281**, 37159–37174.
 50. Mereau, A., Fournier, R., Gregoire, A., Mougin, A., Fabrizio, P., Lührmann, R., and Branlant, C. (1997) An *in vivo* and *in vitro* structure-function analysis of the *Saccharomyces cerevisiae* U3A snoRNP:

- protein-RNA contacts and base-pair interaction with the pre-ribosomal RNA. *J. Mol. Biol.*, **273**, 552–571.
51. Balzer, M. and Wagner, R. (1998) A chemical modification method for the structural analysis of RNA and RNA-protein complexes within living cells. *Anal. Biochem.*, **256**, 240–242.
 52. Hamma, T. and Ferre-D'Amare, A.R. (2004) Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. *Structure*, **12**, 893–903.
 53. Moore, T., Zhang, Y., Fenley, M.O., and Li, H. (2004) Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure*, **12**, 807–818.
 54. Charron, C., Manival, X., Clery, A., Senty-Segault, V., Charpentier, B., Marmier-Gourrier, N., Branlant, C., and Aubry, A. (2004) The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J. Mol. Biol.*, **342**, 757–773.
 55. Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Lührmann, R., Carlomagno, T., and Wahl, M.C. (2007) Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science*, **316**, 115–120.
 56. Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L., Pedersen, J.S., Seraphin, B., Le Hir, H., and Andersen, G.R. (2006) Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science*, **313**, 1968–1972.
 57. Irvine, D., Tuerk, C., and Gold, L. (1991) SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *J. Mol. Biol.*, **222**, 739–761.
 58. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
 59. Clery, A., Bourguignon-Igel, V., Allmang, C., Krol, A., and Branlant, C. (2007) An improved definition of the RNA-binding specificity of SECIS-binding protein 2, an essential component of the selenocysteine incorporation machinery. *Nucleic Acids Res.*, **35**, 1868–1884.
 60. Goers, E.S., Purcell, J., Voelker, R.B., Gates, D.P., and Berglund, J.A. (2010) MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res.*, **38**, 2467–2484.
 61. Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.
 62. Kastner, B. and Lührmann, R. (1999) Purification of U small nuclear ribonucleoprotein particles. *Methods Mol. Biol.*, **118**, 289–298.
 63. Baneyx, F. (1999) Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.*, **10**, 411–421.
 64. Griffiths, C.M. and Page, M.J. (1997) Production of heterologous proteins using the baculovirus/insect expression system. *Methods Mol. Biol.*, **75**, 427–440.
 65. Gurevich, V.V., Pokrovskaya, I.D., Obukhova, T.A., and Zozulya, S.A. (1991) Preparative *in vitro* mRNA synthesis using SP6 and T7 RNA polymerases. *Anal. Biochem.*, **195**, 207–213.
 66. Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.
 67. Jacquenet, S., Ropers, D., Bilodeau, P.S., Damier, L., Mouglin, A., Stoltzfus, C.M., and Branlant, C. (2001) Conserved stem-loop structures in the HIV-1 RNA region containing the A3 3' splice site and its *cis*-regulatory element: possible involvement in RNA splicing. *Nucleic Acids Res.*, **29**, 464–478.

68. Mougin, A., Gottschalk, A., Fabrizio, P., Lührmann, R., and Branlant, C. (2002) Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles. *J. Mol. Biol.*, **317**, 631–649.
69. Milligan, J.F. and Uhlenbeck, O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.*, **180**, 51–62.
70. Silberklang, M., Gillum, A.M., and RajBhandary, U.L. (1979) Use of *in vitro* ³²P labeling in the sequence analysis of nonradioactive tRNAs. *Methods Enzymol.*, **59**, 58–109.
71. England, T.E. and Uhlenbeck, O.C. (1978) 3'-terminal labelling of RNA with T4 RNA ligase. *Nature*, **275**, 560–561.
72. Pan, T., Long, D.M., and Uhlenbeck, O.C. (1993) Divalent metal ions in RNA folding and catalysis, in *The RNA World* (eds R.F. Gesteland and J.F. Atkins), Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory.
73. Gattoni, R., Schmitt, P., and Stevenin, J. (1988) *In vitro* splicing of adenovirus E1A transcripts: characterization of novel reactions and of multiple branch points abnormally far from the 3' splice site. *Nucleic Acids Res.*, **16**, 2389–2409.
74. Mayeda, A. and Krainer, A.R. (1999) Mammalian *in vitro* splicing assays. *Methods Mol. Biol.*, **118**, 315–321.
75. Zahler, A.M., Damgaard, C.K., Kjems, J., and Caputi, M. (2004) SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. *J. Biol. Chem.*, **279**, 10077–10084.
76. Konarska, M.M. (1989) Analysis of splicing complexes and small nuclear ribonucleoprotein particles by native gel electrophoresis. *Methods Enzymol.*, **180**, 442–453.
77. Pinol-Roma, S., Choi, Y.D., Matunis, M.J., and Dreyfuss, G. (1988) Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev.*, **2**, 215–227.
78. Chou, M.Y., Rooke, N., Turck, C.W., and Black, D.L. (1999) hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol. Cell. Biol.*, **19**, 69–77.
79. Burckstummer, T., Bennett, K.L., Preradovic, A., Schutze, G., Hantschel, O., Superti-Furga, G., and Bauch, A. (2006) An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat. Methods*, **3**, 1013–1019.
80. Shafey, D., Boyer, J.G., Bhanot, K., and Kothary, R. (2010) Identification of novel interacting protein partners of SMN using tandem affinity purification. *J. Proteome Res.*, **9**, 1659–1669.
81. Kroiss, M., Brunger, K.M., Wiesner, J., Grimm, M., Sickmann, A., and Fischer, U. (2009) Native purification of protein and RNA-protein complexes using a novel affinity procedure. *Fly (Austin)*, **3**, 221–228.
82. Will, C.L., Behrens, S.E., and Lührmann, R. (1993) Protein composition of mammalian spliceosomal snRNPs. *Mol. Biol. Rep.*, **18**, 121–126.
83. Urlaub, H., Hartmuth, K., and Lührmann, R. (2002) A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods*, **26**, 170–181.
84. Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Lührmann, R., and Fabrizio, P. (1999) Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J.*, **18**, 4535–4548.
85. Hogg, J.R. and Collins, K. (2007) RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. *RNA*, **13**, 868–880.
86. Said, N., Rieder, R., Hurwitz, R., Deckert, J., Urlaub, H., and Vogel, J. (2009) *In vivo* expression and purification of aptamer-tagged small RNA regulators. *Nucleic Acids Res.*, **37**, e133.
87. Jurica, M.S. and Moore, M.J. (2002) Capturing splicing complexes to study

- structure and mechanism. *Methods*, **28**, 336–345.
88. Hartmuth, K., Vornlocher, H.P., and Lührmann, R. (2004) Tobramycin affinity tag purification of spliceosomes. *Methods Mol. Biol.*, **257**, 47–64.
 89. Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Lührmann, R. (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature*, **452**, 846–850.
 90. Stevens, S.W. (2008) Purification of ribonucleoproteins using peptide-elutable antibodies and other affinity techniques. *Methods Mol. Biol.*, **488**, 65–84.
 91. Bachler, M., Schroeder, R., and von Ahsen, U. (1999) StreptoTag: a novel method for the isolation of RNA-binding proteins. *RNA*, **5**, 1509–1516.
 92. Windbichler, N. and Schroeder, R. (2006) Isolation of specific RNA-binding proteins using the streptomycin-binding RNA aptamer. *Nat. Protoc.*, **1**, 637–640.
 93. Zhou, Z. and Reed, R. (2003) Purification of functional RNA-protein complexes using MS2-MBP. *Curr. Protoc. Mol. Biol.*, **27**, 27.23.
 94. Hartmuth, K., Vornlocher, H.P., and Lührmann, R. (2004) Tobramycin affinity-tag purification of spliceosomes, in *mRNA Processing and Metabolism* (ed. D.R. Schoenberg), Humana Press, Totowa, NJ.
 95. Williamson, J.R. (2000) Induced fit in RNA-protein recognition. *Nat. Struct. Biol.*, **7**, 834–837.
 96. Moazed, D. and Noller, H.F. (1986) Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes. *Cell*, **47**, 985–994.
 97. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 98. Mougin, A., Gregoire, A., Banroques, J., Segault, V., Fournier, R., Brule, F., Chevrier-Miller, M., and Branlant, C. (1996) Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *RNA*, **2**, 1079–1093.
 99. Domdey, H., Apostol, B., Lin, R.J., Newman, A., Brody, E., and Abelson, J. (1984) Lariat structures are *in vivo* intermediates in yeast pre-mRNA splicing. *Cell*, **39**, 611–621.
 100. Branch, A.D., Benenfeld, B.J., Paul, C.P., and Robertson, H.D. (1989) Analysis of ultraviolet-induced RNA-RNA cross-links: a means for probing RNA structure-function relationships. *Methods Enzymol.*, **180**, 418–442.
 101. Riggs, P. (2000) Expression and purification of recombinant proteins by fusion to maltose-binding protein. *Mol. Biotechnol.*, **15**, 51–63.
 102. Lane, D. and Harlow, E. (1982) Two different viral transforming proteins bind the same host tumour antigen. *Nature*, **298**, 517.

45

Immunopurification of Endogenous RNAs Associated with RNA Binding Proteins *In vivo*

Minna-Liisa Änkö and Karla M. Neugebauer

45.1

Introduction

RNA binding proteins (RBPs) regulate gene expression at several steps from transcription to translation. RBPs can influence both the information content and abundance of RNAs through various regulatory processes, such as pre-mRNA splicing, RNA editing, polyadenylation, RNA stability, and export. The key role of RBPs in the regulation of gene expression is evidenced by many diseases caused by mutations in the RNA sequences bound by RBPs. However, in most cases only few endogenous genes regulated by individual RBPs have been identified. The identification of ribonucleoprotein (RNPs) complexes with which RBPs are associated can yield information on how different RBPs contribute to the regulation of gene expression in cells and tissues.

In this section, we describe the immunopurification of mRNPs from tissue culture cells. To systemically determine which mRNAs are associated with individual SR protein splicing factors, SR protein–RNA complexes were purified using a green fluorescent protein (GFP)-affinity tag [1]. The immunopurified mRNAs were analyzed by gene expression microarray and reverse transcription polymerase chain reaction (RT-PCR). Protocols using both uncrosslinked and formaldehyde-crosslinked cell extracts are provided, as well as alternative methods for the microarray as the downstream analysis platform.

45.2

Description of Methods

45.2.1

Overview

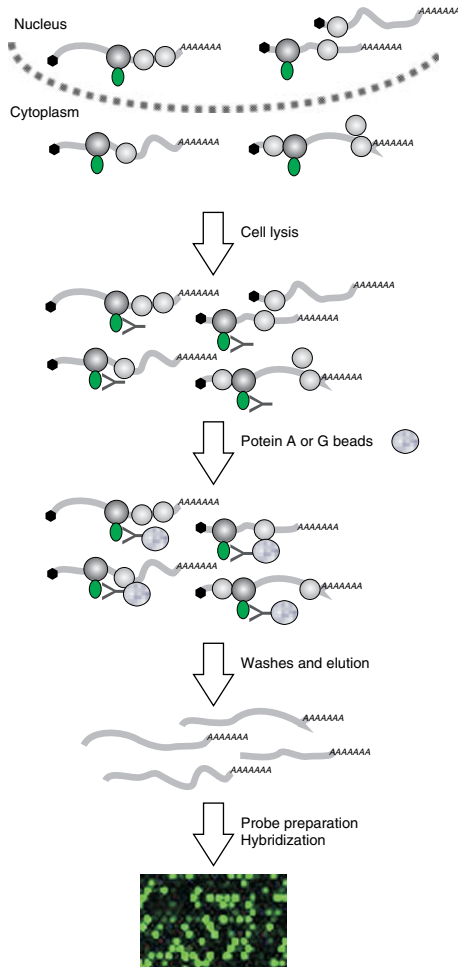
RNA coimmunoprecipitation followed by microarray analysis allows the global identification of RBP-associated RNAs and has been successfully used in several experimental systems [2–8]. In our work, we have used stable mouse P19 cells

expressing GFP-tagged versions of individual SR proteins at physiological levels from bacterial artificial chromosomes (BACs) as transgenes [1, 9]. Thus, an anti-GFP antibody could be used for the immunopurification. Antibodies specific to RBPs of interest can also be used, provided that they recognize the RBP in a complex with RNA; note that some antibodies may compete off the RNA bound and/or only bind the RBP in the unbound conformation. After cell lysis, the RNA–protein complexes are captured by an antibody and the immunocomplexes are recovered using protein-A- or protein-G-coated beads. Non-specifically bound complexes are removed by extensive washing. The lysis and wash buffer composition depends on the cell or tissue type and the antibody used. We have also used formaldehyde-crosslinked cell extracts for immunopurification, particularly at the validation step. If formaldehyde crosslinking is used, more stringent washing is necessary as well as an uncrosslinking step before RNA isolation. Alternatives for the immunopurification conditions described here can be found in the literature [2–8, 10]. The captured RNA is eluted from the beads by phenol:chloroform extraction. We analyzed mRNAs associated with SR proteins in uncrosslinked cell extracts by gene expression microarrays [1]. The immunopurification described yields sufficient amount of RNA for probe preparation and hybridization. Alternatively, when candidate genes are known, the RNA can be used for cDNA synthesis and conventional or quantitative PCR analysis. The isolated RNAs can also be used for the preparation of a sequencing library for next-generation sequencing approaches. An overview of the protocol is given in Figure 45.1.

Reagents and Materials

- Cell scrapers
- 50 ml conical tubes
- 10 ml conical tubes
- 1.5 ml microcentrifuge tubes
- 1 ml syringes
- 22 gauge needles
- 37% formaldehyde
- Phosphate-buffered saline (PBS)
- Phenol:chloroform:isoamylalcohol 125 : 24 : 1, pH 4.5
- Antibody against the protein of interest (we use polyclonal goat anti-GFP)
- Antibody against an unrelated protein or IgG from the same species as the specific antibody (we use purified non-immune goat IgG, Sigma-Aldrich)
- Protein A or protein G Sepharose beads (choice of beads depends on the antibody used; for goat antibodies, protein G beads should be used)
- Ethanol, 96%
- Glycogen
- 3 M Sodium acetate, pH 5.4
- Ethanol, 70%
- Sodium dodecyl sulfate (SDS), 20%

A. RNA immunopurification without crosslinking



B. RNA immunopurification with formaldehyde crosslinking

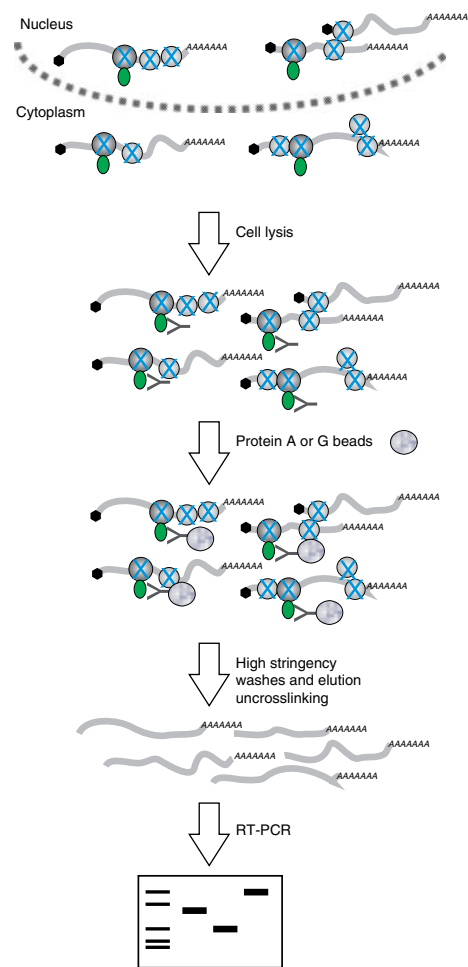


Figure 45.1 Overview of the RNA immunopurification protocol. (a) RNA immunopurification without crosslinking; (b) RNA immunopurification with formaldehyde crosslinking. The downstream analysis

methods may be replaced by methods discussed in Section 45.2.2. Light green ovals = GFP-tag, blue X = crosslinks, gray balls = RBPs, Y = antibody.

- Nuclease-free water
- Protease inhibitor cocktail (Complete Protease Inhibitor Cocktail, Roche)
- RNase inhibitor (RNaseOUT, Invitrogen)
- Bovine serum albumin (BSA, Sigma-Aldrich)
- Salmon sperm DNA (ssDNA, Invitrogen)
- Yeast tRNA (Sigma-Aldrich)

Lysis buffer	1% Nonidet P-40 0.2 M KCl 2 mM MgCl ₂ 20 mM Tris/HCl, pH 7.4 Add protease inhibitors and RNase inhibitor freshly right before use.
Wash buffer A	0.5% Nonidet P-40 0.2 M KCl 2 mM MgCl ₂ 20 mM Tris/HCl, pH 7.4
Wash buffer B1	0.1% SDS 1% Triton X-100 2 mM EDTA 20 mM Tris/HCl, pH 8.1
Wash buffer B2	150 mM NaCl 0.1% SDS 1% Triton X-100 2 mM EDTA 20 mM Tris/HCl, pH 8.1
Wash buffer B3	500 mM NaCl 1% Nonidet P-40 1% Deoxycholic acid (sodium salt) 0.25 M LiCl 1 mM EDTA 10 mM Tris/HCl, pH 8.1
NET-2 buffer	50 mM Tris/HCl, pH 7.5 150 mM NaCl 0.05% Nonidet P-40
TE buffer	10 mM Tris/HCl, pH 8.0 1 mM EDTA

RNA Coimmunoprecipitation

The immunopurification protocol described here is optimized for mouse P19 tissue culture cells. The amount of cells needed to yield sufficient amount of RNA depends on the cell type used, the abundance of the RBP, and the downstream application. We use $\sim 10^8$ P19 cells for the SR protein–RNA immunopurification. This corresponds to two 80% confluent $\varnothing 14$ cm cell culture dishes. All steps are performed on ice, unless otherwise mentioned.

RNA Immunoprecipitation with Uncrosslinked Cell Extracts

- 1) Wash the cells grown as a monolayer twice with cold PBS. Add 5 ml cold PBS-containing protease inhibitors per cell culture dish and harvest the cells by scraping them into the PBS solution. Pellet the cells in a conical 15 ml tube for 5 min at 300 g at $+4^\circ\text{C}$. If working with non-adherent cells, pellet the cells as above and wash them twice with cold PBS. Then proceed to cell lysis (step 2).

- 2) Resuspend the pellet into 1 ml lysis buffer containing protease inhibitors and RNase inhibitors. Incubate for 10 min on ice. To enhance the lysis, pass the extract through a syringe and 22 gauge needle 20 times. The lysis of cells and nuclei can be controlled under a light microscope by staining the cells with Trypan blue. Clear the extract by centrifugation for 10 min (20 000 g) at +4 °C.
- 3) Transfer the cleared extract into a 1.5 ml tube. Take a 50 µl aliquot of the extract to serve as an input sample. Store the input sample on ice until step 7. Divide the remaining extract into two aliquots. Increase the volume of each aliquot to 1 ml using wash buffer A. Add protease inhibitors and RNase inhibitors accordingly.
- 4) Add the primary antibody (10 µg goat anti-GFP; the amount of antibody has to be determined experimentally) or an equal amount of the control antibody (10 µg goat IgG). Incubate rotating end-over-end for 2 h at +4 °C.
- 5) Add an appropriate amount of blocked Sepharose beads; we use 40 µl of 1 : 1 slurry of blocked protein G GammaBind beads. Continue incubation for 1 h at +4 °C. The amount and type of beads have to be determined experimentally. The beads should be preblocked for 1 h with 0.2 mg ml⁻¹ BSA, 0.1 mg ml⁻¹ yeast tRNA, and 0.1 mg ml⁻¹ ssDNA and the preblocked beads can be stored in PBS at +4 °C for two weeks.
- 6) Pellet the beads (300 g) for 1 min at +4 °C. Wash the beads five times with wash buffer A. Between the washes, centrifuge at 300 g for 1 min at +4 °C. The amount of washes should be experimentally determined for each antibody and cell extract.
- 7) After the final wash, add 300 µl NET-2 buffer containing 1% SDS and an equal volume of phenol:chloroform:isoamylalcohol 125 : 24 : 1. Vortex vigorously and incubate for 1 h at 37 °C with mild shaking. Include (pH 4.5) the input samples at this step. Add 250 µl NET-2 buffer containing 1% SDS to the input sample and process it in parallel with the immunopurified samples.
- 8) Centrifuge (20 000 g) for 5 min at 20 °C. Carefully transfer the aqueous phase into a clean 1.5 ml tube and add 10 µg glycogen, 1/10 volume 3 M sodium acetate pH 5.4, and 3 volumes 96% cold ethanol. Precipitate for 1 h or overnight at -80 °C.
- 9) Pellet the RNA by centrifugation for 20 min (20 000 g) at +4 °C. Wash the pellet with cold 70% ethanol.
- 10) Carefully remove all ethanol and air-dry the pellet. Resuspend the pellet into RNase-free water.

RNA Immunoprecipitation with Formaldehyde-CrossLinked Cell Extracts

- 1) Add 37% formaldehyde to a final concentration of 0.1% on the cells growing as a monolayer. Incubate for 10 min at 25 °C. When working with suspension cells, perform the incubation in a shaker. The concentration of formaldehyde (normally 0.1–1%) and crosslinking time (5–60 min) depend on the cell type used and should be experimentally determined. After crosslinking, wash the

- cells as in the Section “RNA Immunoprecipitation with UncrossLinked Cell Extracts,” step 1.
- 2) Perform the lysis as in the Section “RNA Immunoprecipitation with UncrossLinked Cell Extracts,” step 2. The crosslinked extracts need to be sonicated to solubilize the crosslinked complexes. The sonication time and power depend on the cell line and the equipment used; we use a tip sonicator with 30% power and give three 10 s pulses in 20 s intervals on ice. Clear the lysate as in the Section “RNA Immunoprecipitation with UncrossLinked Cell Extracts,” step 2. Take a 50 μ l aliquot as an input and store it on ice. Proceed as in the Section “RNA Immunoprecipitation with UncrossLinked Cell Extracts,” steps 3–5.
 - 3) Wash the immunocomplexes sequentially with wash buffers B1–3 and finally, twice with TE buffer. Between the washes, pellet the beads by centrifugation (300 g) for 1 min at +4 °C.
 - 4) Add 300 μ l NET-2 buffer with 1% SDS and uncrosslink for 45 min at 70 °C. Include the input samples at this step. Add 250 μ l NET-2 buffer to the input sample and process it in parallel with the immunopurified containing 1% SDS samples.
 - 5) Isolate the RNA as in the Section “RNA Immunoprecipitation with UncrossLinked Cell Extracts,” steps 7–10.

45.2.2

Analysis of Coimmunoprecipitated RNA

The immunopurification protocol should be scaled according to the RNA yield required in the downstream application. Typically, 10^8 P19 cells used as starting material for the immunopurification yield 1–2 μ g in RNA. Determine the concentration of immunopurified RNA with Nanodrop or a similar small-scale spectrophotometer. Depending on the downstream application used, further quality control steps may be needed. The integrity of the RNA can be determined by Bioanalyzer (Agilent Technologies) or its equivalent. If associated RNAs are previously known, the analysis of the positive control RNAs by conventional RT-PCR should be performed before proceeding to genome-wide analysis.

45.2.2.1 Microarray Analysis of Immunopurified RNA

The immunopurified RNA is suitable for analysis with any microarray platform. When platforms other than the one described here are used, the recommendations of the supplier should be followed. Here, we describe the analysis with the Agilent one-color gene expression platform. The one-color platform is better suited for the analysis of immunoprecipitated RNA because it gives more flexibility in the data analysis. The specific immunoprecipitation can be compared to both input and control (mock) immunoprecipitation. In addition, the signal distribution of immunoprecipitated RNA may be different from the signal distribution of total gene expression (input), which may cause problems when using a two-color competitive hybridization platform. Detailed description of the labeling and hybridization

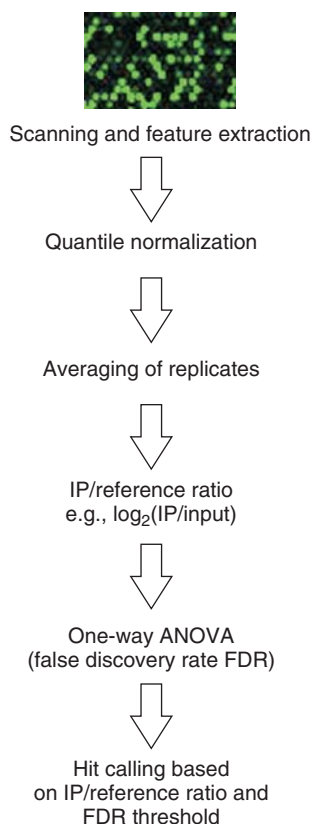


Figure 45.2 Simplified view of the analysis of microarray data. As discussed in Sections 45.2.2 and 45.3.3, each step of the analysis flow should be specifically considered for the given application.

protocols can be found in the manuals provided by the supplier. The data analysis is described shortly here and summarized in Figure 45.2, and additional comments can be found in Section 45.3.3.

- 1) Perform the required quality control steps described in the Section “RNA Immunoprecipitation with UncrossLinked Cell Extracts,” step 10.
- 2) Proceed to Agilent Low RNA Input Linear Amplification Kit Plus using the one-color platform. Include the specific immunopurification, mock immunopurification, and input sample for hybridization to separate microarrays. Three biological replicates are recommended.
- 3) Hybridize the labeled samples onto mouse Whole Genome Oligo Microarrays. Scan the arrays as recommended by the manufacturer.
- 4) Preprocess the data by Agilent feature extraction software without normalization.

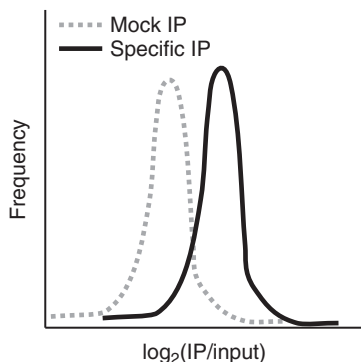


Figure 45.3 Typical distribution of RNA immunopurification signals when input is used as the reference sample.

- 5) Quantile-normalize the data within sample groups. If the distribution of the immunopurified RNA is similar to that of the input samples, all samples can be normalized together. However, in our experience, because of the experimental enrichment of RNA in the immunopurification, the signal distributions of the sample types differ.
- 6) Perform comparison between the specific immunopurification and the mock immunopurification and between the specific immunopurification and the input. The background level (signal from the non-specific control immunopurification) determines whether the mock immunopurification can be used as a reference sample. In our hands, the mock samples gave low, although reproducible, microarray signals because of the very low background of the immunopurification (Figure 45.3). The use of the mock-immunopurified RNA as reference would have resulted in arbitrarily high enrichment ratios. Therefore, mock-immunopurified RNA was not used as a reference sample. Instead, we computed an RNA immunoprecipitation (RIP) score by comparing the specific immunopurification to the expression level ($\text{RIP score} = \log_2(\text{IP}/\text{input})$). The candidate genes were determined by one-way ANOVA analysis, comparing specific immunopurification, mock immunopurification and input, and by a RIP score cutoff (fold enrichment over input).

45.2.2.2 RT-PCR Analysis of Immunopurified RNA

The RNA immunopurification can be used to analyze the enrichment of known RBP interaction partners. RT-PCR can also be used to validate the candidate genes identified by microarray analysis. Depending on the primers used for the reverse transcription, both polyadenylated and nascent transcripts can be analyzed and conventional and quantitative PCR analysis can be applied. Protocols for setting up a quantitative RT-PCR are widely available and are not discussed here. We have used conventional PCR to validate the candidate genes determined by microarray as described in Section 45.2.2.1 and find a good correlation between the methods.

45.2.2.3 Next-Generation Sequencing of Immunopurified RNA

The immunopurified RNA can be directly used for the preparation of a sequencing library for next-generation sequencing. Since the immunopurification enriches full-length RNA molecules, RNA is suitable for sequencing with a variety of platforms, including 454 Sequencing™, Illumina (Solexa)™, and SOLiD™. One should refer to the manufacturer's instructions for the specific requirements and protocols for the preparation of a sequencing library. The complexity of the RNA sample depends on the number of different RNA molecules associated with an RBP and should be taken into consideration when designing the library preparation. The mock immunopurification sample may be the preferred reference for next-generation sequencing analysis because of the similar complexity of the sample to the specific immunopurified RNA. The input, which essentially corresponds to the whole transcriptome, needs to be sequenced at considerably greater depth than the immunopurified samples.

45.3

Troubleshooting

45.3.1

Critical Points and Common Problems

The success of an immunopurification method relies on the efficiency and specificity of the antibody used. Other factors affecting the specificity and efficiency of the immunopurification include the abundance of the RBP and the cell or tissue type used. Therefore, it is important to optimize the immunopurification conditions for each experimental condition. It is advisable to test several antibodies and washing conditions. The enrichment of the RBP in the immunopurification should be tested by Western blot analysis before proceeding to the analysis of the RNA. After the immunopurification, the integrity of the RNA should be controlled to eliminate possible fragmentation or degradation of RNA during the purification. Some possible problems in the immunopurification are discussed in the following list.

- 1) **Low yield of RNA.** Check the efficiency of the immunopurification by Western blotting to ensure that sufficient amount of RBP is purified. If an antibody against an endogenous RBP is used for the purification, make sure that the antibody does not inhibit the association of the RBP with RNA and that the antibody recognizes the RNA-bound form of the RBP. Change the immunopurification buffers (salt and detergent) to less stringent conditions.
- 2) **No RNA detected after purification.** Do the same controls as in step 1. When no RNA is recovered, the most likely cause is RNase contamination. Change all the solutions and perform other precautions to eliminate all sources of RNase. Possible additional sources of RNase contamination include antibody preparations and beads used to collect the immunocomplexes. When working with a tissue extract with high RNase activity, increase the amount

of RNase inhibitors in the buffers. Make sure that all the steps are performed on ice.

- 3) **High background.** Increase the stringency of the washing conditions. If the RBP is abundant in the cell extract, lower the amount of starting material. This may increase the specificity of the immunopurification. Include a preclearing step before adding the antibody; incubate the cell extract with Sepharose beads for 30–60 min, spin down the beads, collect the supernatant, and then proceed to the incubation with the antibody.
- 4) **Good yield of RNA, but no signal in the downstream application.** Residual salt after RNA precipitation can inhibit enzymatic reactions. Do not store the RNA in ethanol for extended times. Likewise, residual phenol can inhibit enzymatic reactions. Make sure no carryover of the middle or lower phase occurred in the phenol:chloroform extraction step. Control for possible sources of RNase contamination during the downstream analysis of RNA.

45.3.2

Uncrosslinked or Crosslinked RNA Immunoprecipitation

In our experience, the uncrosslinked and formaldehyde-crosslinked RNA immunoprecipitation give almost identical results in RT-PCR experiments and we did not experience the reassociation of SR proteins after lysis as described for HuR [11]. However, this can be highly specific to the protein in question and cannot be generalized to all RBPs. In general, the yield of RNA is lower when crosslinking is used. The formaldehyde crosslinking may also affect the suitability of RNA to some downstream analysis methods, and formaldehyde crosslinking tends to increase the background signal. This is partially due to the fact that formaldehyde crosslinking results not only in protein–RNA complexes but also protein–protein and protein–DNA crosslinks. If crosslinking is desired but the indirect effect of formaldehyde needs to be avoided, UV crosslinking may be considered. UV treatment leads to covalent crosslinks between RBP and RNAs that interact directly; however, UV crosslinking is generally inefficient and biased toward uridine residues, and, thus, requires a very sensitive downstream analysis method such as next generation sequencing. Methods for the UV crosslinking approach are discussed elsewhere [12–14].

45.3.3

Microarray Data Analysis

We found most existing methods of analyzing gene expression microarray data unsuitable for the analysis of RNA immunopurification experiment. The immunopurified RNA is enriched and does not necessarily have similar signal distribution as all expressed genes. This affects mainly the normalization of array signals but should be considered in all steps of data processing. We established the analysis platform described in Section 45.2.2.1 to identify the SR-protein-associated mRNAs. We chose to use statistical measures, instead of ranking, to determine the candidate

RNAs. Therefore, for each protein and condition, the number of identified hits corresponds more closely to the number of true RBP–RNA associations. A ranking approach is essentially based on distribution of immunopurified RNA to control ratios; therefore, depending on the set threshold, a certain fraction of all genes is always recovered as hits.

45.4

Conclusions

RNA immunopurification, together with a genome-wide downstream analysis of the associated RNAs, is a powerful tool to discover new protein–RNA interactions. RNA depletion approaches can only provide indirect information about the functions of RBP, whereas the immunopurification approach gives direct information about the composition of RNA–protein complexes. The methods described here and in the references provided can be applied to a variety of experimental systems, ranging from tissue culture cells to complex model organisms. A more detailed picture of the subcellular functions of RBPs can be obtained by combining the RNA immunopurification with the subcellular fractionation. The identification of RNAs associated with individual RBPs will increase the understanding of how the regulation of gene expression is achieved in cells and tissues.

Acknowledgments

We thank Andreas Beyer and Lucia Morales for designing the analysis platform for RIP-chip data. We are grateful for the technical expertise of Julia Jarrells and Britta Schilling in performing the microarray experiments. The anti-GFP antibody was a kind gift from David Drechsel. The financial support was from Sigrid Juselius foundation (to M-LÄ), Helsingin Sanomain Foundation (to M-LÄ), the Max Planck Society (to KMN and M-LÄ), and the European Commission (EURASNET-518238 to KMN).

References

1. Änkö, M.-L., Morales, L., Henry, I., Beyer, A., and Neugebauer, K.M. (2010) Global analysis reveals SRp20- and SRp75-specific mRNPs in cycling and neural cells. *Nat. Struct. Mol. Biol.*, **17**, 962–970.
2. Hieronymus, H. and Silver, P.A. (2003) Genome-wide analysis of rna-protein interactions illustrates specificity of the mRNA export machinery. *Nat. Genet.*, **33**, 155–161.
3. Penalva, L.O., Tenenbaum, S.A., and Keene, J.D. (2004) Gene expression analysis of messenger RNP complexes. *Methods Mol. Biol.*, **257**, 125–134.
4. Gama-Carvalho, M., Barbosa-Morais, N., Brodsky, A., Silver, P., and Carmo-Fonseca, M. (2006) Genome-wide identification of functionally distinct subsets of cellular mRNAs associated with two nucleocytoplasmic-shuttling

- mammalian splicing factors. *Genome Biol.*, **7**, R113.
5. Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D., and Brown, P.O. (2008) Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol.*, **6**, e255.
 6. Kiesler, E., Hase, M.E., Brodin, D., and Visa, N. (2005) Hrp59, an hnRNP M protein in *Chironomus* and *Drosophila*, binds to exonic splicing enhancers and is required for expression of a subset of mRNAs. *J. Cell Biol.*, **168**, 1013–1025.
 7. Keene, J.D., Komisarow, J.M., and Friedersdorf, M.B. (2006) RIP-chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat. Protoc.*, **1**, 302–307.
 8. Gabut, M., Dejardin, J., Tazi, J., and Soret, J. (2007) The SR family proteins B52 and dASF/SF2 modulate development of the *Drosophila* visual system by regulating specific RNA targets. *Mol. Cell Biol.*, **27**, 3087–3097.
 9. Sapra, A.K., Änkö, M.-L., Grishina, I., Lorenz, M., Pabis, M., Poser, I., Rollins, J., Weiland, E.-M., and Neugebauer, K.M. (2009) SR protein family members display diverse activities in the formation of nascent and mature mRNPs *in vivo*. *Mol. Cell*, **34**, 179–190.
 10. Niranjanakumari, S., Lasda, E., Brazas, R., and Garcia-Blanco, M.A. (2002) Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions *in vivo*. *Methods*, **26**, 182–190.
 11. Mili, S. and Steitz, J.A. (2004) Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA*, **10**, 1692–1694.
 12. Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005) CLIP: a method for identifying protein–RNA interaction sites in living cells. *Methods*, **37**, 376–386.
 13. Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., Darnell, J.C., and Darnell, R.B. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, **456**, 464–469.
 14. Wang, Z., Tollervey, J., Briesse, M., Turner, D., and Ule, J. (2009) CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins *in vivo*. *Methods*, **48**, 287–293.

46

Protein–RNA Crosslinking in Native Ribonucleoprotein Particles

Olexandr Dybkov, Henning Urlaub, and Reinhard Lührmann

46.1

Introduction

Protein–RNA interactions lie at the structural and functional heart of ribonucleoprotein (RNP) particles. They govern such fundamental cellular processes as pre-mRNA processing, rRNA maturation, posttranscriptional control (mRNA stability), RNA export, translation, and translational control. In this chapter, we present a method we have developed in recent years, which allows us to characterize sites of direct protein–RNA contact in native particles, after the contacts have been made permanent by UV crosslinking [1–4].

Our method is especially suitable in situations where the objects of investigation are native RNP particles for which the RNA and the protein compositions are known, while little or no information is available on which proteins are in contact with RNA or where such contacts are located on the RNA. The method has further been proved to be of value for the identification of direct RNA–protein contact sites in RNP particles reconstituted *in vitro* in which several proteins interact with the RNA component. Complementary methods such as CLIP, CRAC, and their modifications have been developed recently [5–8] and employ cloning and (deep) sequencing of RNAs from UV-crosslinked RNPs instead of the direct sequencing of the RNA used in our method. Those methods are best suited for the analysis of RNA–protein contacts in RNP particles with unknown RNA composition or when RNA species are very long. Unlike our method, they are more time-consuming, often require sophisticated bioinformatic analysis, and do not always allow identification of crosslinking sites with a single-nucleotide resolution.

In the protocols listed, we refer to isolated U snRNP particles from HeLa cells [9, 10] involved in pre-mRNA processing (for review, see Refs. [11–13]) and to spliceosomal complexes assembled *in vitro* [14–18]. Importantly, we would like to note that the entire approach can be regarded as a general one, so that the protocols can be easily adapted to investigations of other native RNP particles or of RNP particles reconstituted *in vitro*.

46.2

Overall Strategy

The overall experimental strategy that we have used for the identification of protein–RNA contact sites in native RNP particles comprises crosslinking of RNP particles by UV irradiation at 254 nm, which fixes protein–RNA interactions covalently by generating a zero-length crosslink, followed by analytical procedures to identify the exact nucleotide(s) on the RNA, where the crosslink occurred and to identify the crosslinked polypeptide.

The approach is primarily a primer extension analysis of the crosslinked RNA derived from the UV-irradiated native particles. The correct assignment of putative protein–RNA crosslinking sites on the RNA requires parallel analysis of the RNA extracted from the non-irradiated particles as well as UV-irradiated “naked” (protein-free) RNA and of non-irradiated naked RNA. Comparison of the reverse transcriptase patterns obtained in these four experiments leads to the identification of the RNA base at which proteins are crosslinked. This first set of experiments gives an excellent overview if a certain protein – or several proteins of multiprotein complexes – is/are in direct contact with the RNA, but it yields no information about which protein of the RNA is crosslinked.

The identification of the corresponding crosslinked protein is achieved by immunoprecipitation combined with primer extension analysis. Thereby, one can define which protein of the RNP is crosslinked to the bases of the RNA, which have been identified in the first set of experiments. It is obvious that this type of identification depends on the availability of antibodies against the different proteins and on the efficiency with which each antibody precipitates its corresponding protein, especially under mild denaturing conditions (for details, see below).

Our method, however, is not restricted to the analysis of proteins against which specific antibodies are available. The immunoprecipitation step can be replaced by several other affinity purification techniques compatible with selection under mild denaturing conditions. Accordingly, tagging of proteins might be considered, as it can be more time- and cost-efficient when compared with raising protein-specific antibodies. In recent years, we have successfully used IgG – Protein A tag and Nickel – Histidine tag affinity systems instead of immunoprecipitation in order to enrich UV-induced protein–RNA conjugates and subsequently to locate crosslinking sites. The Protein A system is of particular value for researchers working with baker’s yeast as *S. cerevisiae* strains expressing TAP-tagged variants of nearly all individual proteins can readily be obtained from EUROSCARF [19].

However, a certain shortcoming of the method that will be described in detail in this chapter is due to the length of the RNAs to be analyzed. While the primer extension analysis of short RNAs (300 nt or less) is relatively fast and straightforward, analysis of crosslinks to long RNAs usually requires a prior knowledge of the interacting region. Such information can be obtained, for example, by northern blot analysis of a series of RNase H generated fragments of RNA, which remain

attached to the protein. Alternatively, other methods for sequencing of RNAs may be used (Section 46.1).

The main advantage of the method described here is that it can reveal multiple crosslinks between one protein and its cognate RNA in native particles with a single-nucleotide precision in a short time (i.e., a complete analysis is possible within a week). For example, we found in this manner that in U1 snRNP particles, the U1 70K protein is in contact with 2 nt in the loop of stem I of the U1 snRNA [1]. Another example is the U4/U6-specific protein 61K: this was found in contact with two distinct sites on the U4 snRNA in native tri-snRNP particles, that is, the loop in the 5' stem-loop of the U4 snRNA and nucleotides upstream of the 5' stem-loop [4].

46.3

UV Crosslinking

UV crosslinking of RNP particles is a straightforward technique. UV crosslinking at 254 nm generates a covalent bond between an amino acid side chain of a protein and a base of the RNA, whenever the relative orientation of the two components is favorable. In earlier studies, we found that UV irradiation of native complexes at 254 nm leads to crosslinking of the side chains of the following amino acids: methionine, tyrosine, histidine, leucine, phenylalanine, and cysteine ([1–4, 20–24], and our unpublished observations [Henning Urlaub, Katharina Kramer, Romina Hofele]). On the basis of work with halopyrimidine substituted RNAs, Koch *et al.* [25] suggested two possible mechanisms for UV-induced protein–RNA crosslinking events: (i) UV-induced electron transfer from the amino acid residue to the halopyrimidine followed by a loss of halide and subsequent radical combination or (ii) UV-induced homolysis of the carbon–halogen bond followed by radical addition to the aromatic ring of the amino acid residue.

The fact that we have also found highly specific amino acid RNA crosslinks in non-substituted RNAs of native complexes strongly supports the first mechanism. Our approach is highly specific, but it has some limitations. The crosslinking yield is relatively low when compared to that of crosslinking in particles reconstituted *in vitro* that carry an RNA species site-specifically labeled with a crosslinking moiety [26–34]. Furthermore, not all proteins that are tightly bound by RNA can also be directly crosslinked by UV irradiation. Examples of this are the U1A protein bound to the U1 snRNA particle ([35] and our unpublished observations [Henning Urlaub]) and the human 15.5K protein bound to the human U4 snRNA ([36] and our unpublished observations [Henning Urlaub]). On the other hand, every direct UV-induced protein–RNA crosslink found, in particular, in native RNP particles, must reflect a “real” interaction because of the short distance between the crosslinked entities. For that reason such crosslinks are referred to as “zero length.” Moreover, work with particles reconstituted *in vitro* that carry a site-specifically labeled RNA is dependent on the efficiency of reconstitution and

may even produce false-positive results if heterogeneous populations are generated as a result of incomplete assembly.

Protocol 1 describes the UV irradiation procedure. Some critical points have to be considered when one performs UV crosslinking experiments with native RNP particles:

- 1) **Concentration of the RNP particles:** For UV crosslinking, purified native RNP particles are typically adjusted to a concentration of not more than about 0.1 mg ml^{-1} . The final concentration of native particles in solution as such is not critical, as native particles are fully assembled, and interparticle crosslinking events in native RNP particles are highly unlikely. This item becomes much more of a problem when particles reconstituted *in vitro* are studied. Because an excess of protein over RNA has to be used for the efficient *in vitro* reconstitution, non-specific crosslinks due to the excess of protein may pose a problem (see Section 46.6, point (2) for details). For RNP particles reconstituted *in vitro*, an RNA concentration of $0.1 \text{ pmol } \mu\text{l}^{-1}$ is in general well sufficient.
- 2) **Choice of buffer:** First of all, the buffer should not contain a high concentration of reagents that are known to scavenge radicals, for example, glycerol. UV crosslinking is a UV-induced radical reaction that generates a new covalent bond between the side chain of an amino acid and a base of the RNA. Thus, radical scavengers drastically reduce the crosslinking yield, and, for example, glycerol concentrations should not exceed 5%. Furthermore, as samples are irradiated in small droplets (see below), any detergents in the buffer must be avoided, as the droplets will start to spread out over the sample plate. However, if detergent is, however, crucial for the complexes of interest, irradiation on Parafilm should be used. As the solubility of SDS decreases significantly in the presence of potassium ions, the latter should not be present in the buffer or, alternatively, the complexes should be precipitated before their denaturation with SDS.
- 3) **Crosslinking conditions:** These include the choice of UV lamp, the distance between the lamp(s) and the sample, and the irradiation time. Our laboratory uses a specially constructed device for UV irradiation at 254 nm (see Figure 46.1 and Protocol 1). Alternatively, other commercially available devices can be used (e.g., a UV Stratalinker 2400, Stratagene, La Jolla, USA); however, the conditions of UV irradiation, in particular, the irradiation time as a function of the power of the UV source and the distance of the lamp(s) to the sample have to be adjusted accordingly (see below). The samples can be irradiated in different ways, in droplets on a glass slide or Parafilm (Pechiney Plastic Packing, Menasha, USA) or in open plastic tubes. In cases of high sample volumes, Petri dishes, flat-bottomed microtiter plates, or custom-made larger glass dishes (4–12 cm in diameter) with a planar surface can be used. Strong as well as non-specific absorption to the surface must be avoided, as both of these lead to a drastic loss of sample material. Precooling of the samples (4°C) and the glassware is essential. In our hands, $25 \text{ } \mu\text{l}$ droplets on a precooled 10-well multitest slide (see Protocol 1) work best. For smaller volumes (e.g., $10 \text{ } \mu\text{l}$), we

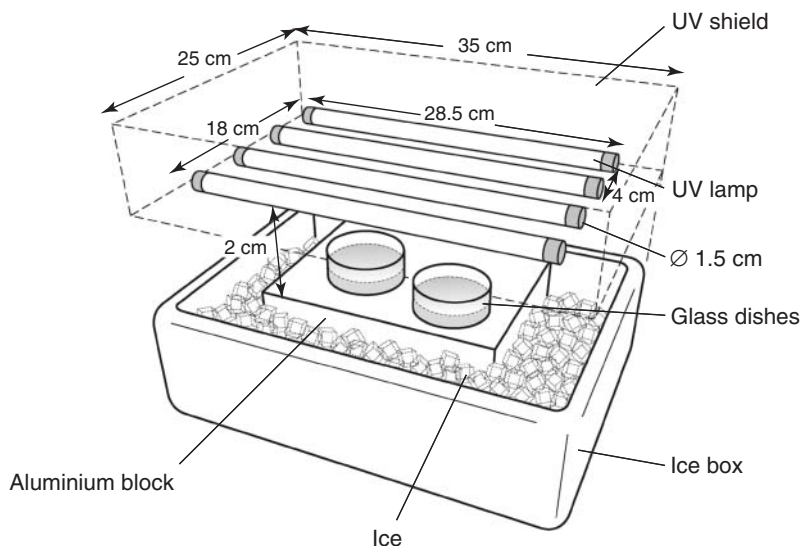


Figure 46.1 Schematic drawing of the custom-made UV crosslinking device.

use Eppendorf tubes (Eppendorf AG, Germany) mounted directly under the UV source. We irradiate the samples at a distance of 2–4 cm from the source (corresponding to the height of a tube mounted directly under an 8 W lamp). The glass slide with the samples is put on top of an aluminum block placed in ice. In addition to these items, the most critical point is the duration of direct UV irradiation. We have observed that the maximum yield of crosslinks under our conditions is obtained after 2 min. Longer irradiation (3 min) does not increase the crosslinking yield significantly, and further extended irradiation times lead to substantial loss of particles. On the other hand, when one is working with more rigid RNP particles such as ribosomes [20, 21], or *in vitro* reconstituted protein–RNA complexes with a very limited number of components, the UV irradiation may have to be prolonged. In any case, as a starting point, we recommend performing Protocol 1 with different durations of UV irradiation.

46.4

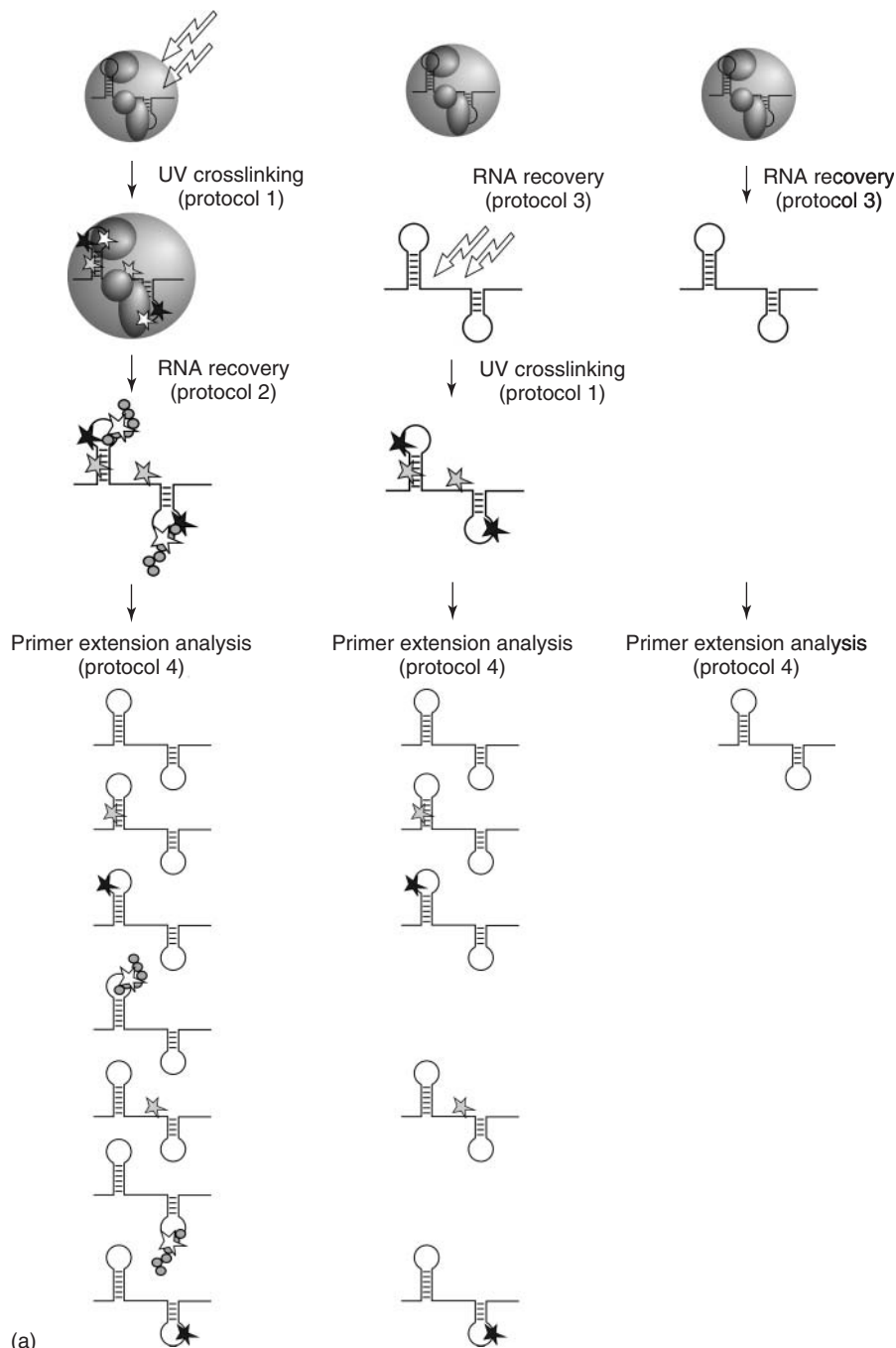
Identification of UV-Induced Protein–RNA Crosslinking Sites by Primer Extension Analysis

After exhaustive hydrolysis of the protein moiety of crosslinked RNP particles by proteinase K treatment, a few amino acids remain covalently attached to the RNA at the sites of crosslinking. In a primer extension reaction, RNA is primed with a 5′-³²P-labeled DNA oligonucleotide complementary to a chosen region on the RNA. The reverse transcriptase enzyme then adds dNTPs, which are complementary to

the nucleotides of the RNA, to the 3' end of the labeled DNA primer and thus generates radioactively labeled DNA molecules that are complementary to the entire RNA sequence. At a nucleotide on the RNA that has been covalently modified (i.e., by crosslinking), no complementary DNA nucleotide can be added by the reverse transcriptase, owing to either incomplete Watson–Crick base pairing or steric hindrance due to the presence of crosslinked amino acids. Thus, this nucleotide will cause a stop, or at least a “stuttering,” of the reverse transcriptase. It should be noted that the reverse transcriptase stops 1 nt before the actual crosslinking site. The complementary DNA generated has a certain length and the site of the stops (i.e., the length of the generated DNA) can be deduced from a sequencing gel when analyzed next to a marker of complementary DNA that has been generated with the help of dideoxynucleotides [37].

Crosslinking induced by UV irradiation at short wavelengths can also cause intra-RNA crosslinks or induce strand breaks in the RNA; both of these also lead to stops or stuttering of the reverse transcriptase. Therefore, the reverse transcriptase patterns from four RNAs must be compared: (i) RNA from UV-irradiated RNP particles, (ii) RNA from non-irradiated RNP particles, (iii) UV-irradiated naked RNA, and (iv) non-irradiated naked RNA. Comparison of the primer extension reaction from these sets of experiments on a high-resolution sequencing gel leads to the identification of putative protein–RNA crosslinking sites. Figure 46.2a illustrates the principle of the identification of protein–RNA crosslinking sites and Figure 46.2b gives an example of identified protein–RNA crosslinking sites in native UV-irradiated U1 snRNPs [1, 3].

Figure 46.2 (a) Schematic representation of the initial experiments necessary for the identification of putative protein–RNA crosslinking sites in native particles. RNA derived either from UV-irradiated particles, or RNA that was stripped of proteins before UV irradiation, or non-irradiated “naked” RNA is analyzed by primer extension analysis. White stars indicate nucleotides on the RNA that are covalently modified by crosslinked proteins, gray stars indicated intra-RNA crosslinks, and black stars indicate UV-induced strand breaks on the RNA. Gray balls indicate those amino acids that remain covalently attached to the sites of crosslinking after digestion of the RNP particles with Proteinase K. (b) Example of the analysis of protein–RNA crosslinking sites on the U1 snRNA in the native U1 snRNP. The primers A and B used in this experiment are complementary to nucleotides 134–152 and 63–77 of U1 snRNA, respectively. Lanes 1, 3, 5, and 7: controls with naked U1 snRNA isolated from non-irradiated U1 snRNP particles; lanes 2 and 6: UV-irradiated naked U1 snRNA (Protocol 3); lanes 4 and 8 are U1 snRNA derived from UV-irradiated U1 snRNP particles (Protocols 1 and 2). C, U, A, G: dideoxy sequence markers. Comparison of the reverse transcriptase pattern of the single lanes reveals protein–RNA crosslinking sites at nucleotides U128, G106, and A26 to C31. The position of the protein–RNA crosslinking sites on the U1 snRNA is indicated by black arrows. Black bars on the left indicate the Sm site and the stem-loop I sequence on U1 snRNA, respectively. A schematic diagram of U1 snRNA is shown below the panels. Arrows indicate the crosslinking sites, while the positions of the primers A and B are indicated by black lines.



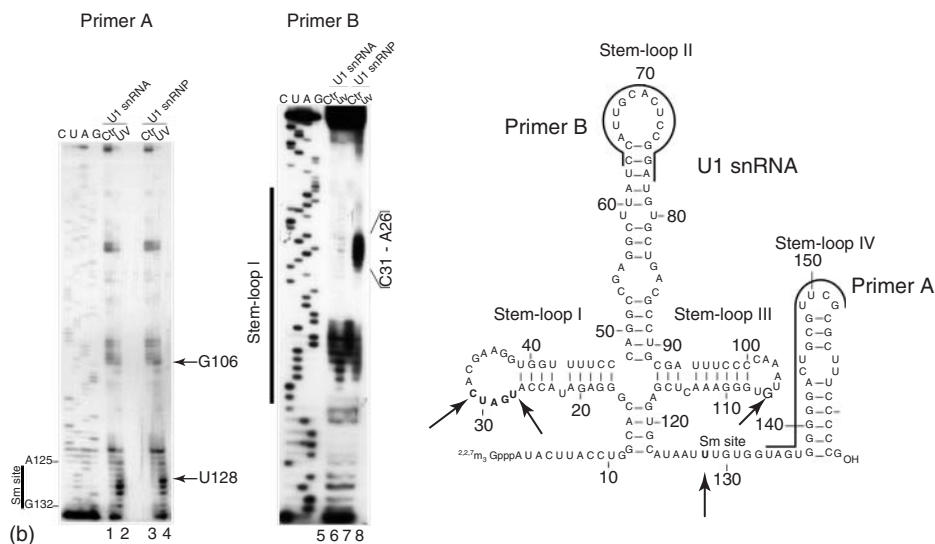


Figure 46.2 (continued)

Protocol 2 describes the purification of RNA derived from UV-irradiated RNP particles. Protocol 3 describes the experimental steps that are required in order to generate naked UV-irradiated and non-irradiated RNAs for controls. Protocol 4 gives a detailed description of the primer extension reaction (including DNA primer purification and labeling) and the subsequent gel electrophoresis that are needed for the reproducible visualization and identification of the crosslinking sites on the RNAs.

For a correct assignment of protein–RNA crosslinking sites in native particles by this method, we would like to emphasize several important points:

- 1) The first set of experiments probes putative crosslinking sites in samples that contain an excess of non-crosslinked (or unmodified) RNA and in which only a small percentage of the RNA is modified by the UV irradiation (depending on the crosslinking yield and on the degree of UV-induced damage). Therefore, signals corresponding to the full-length RNA will be strongest, as seen by autoradiography of a sequencing gel. To systematically compare crosslinks on RNA from UV-irradiated particles with those on UV-irradiated naked RNA, the amount of RNA probed by primer extension must be the same. Otherwise, weaker protein–RNA crosslinking sites may escape identification, in particular, when protein–RNA crosslinking events lead to an enhancement of reverse transcriptase stops compared with the background. In order to ensure that similar amounts of crosslinked material were loaded, the signals corresponding to full-length RNA should be of comparable intensity in both sample preparations. In those cases, where UV-induced crosslinks significantly reduce the signal intensity of the full-length transcript, the intensity of the

naturally occurring stops on the RNA before the crosslinks should be of comparable intensity.

Alternatively, the sum of the intensities from all stops occurring on the RNA must be the same in both experiments.

- 2) If possible, the RNA should be probed with different primers for a comprehensive analysis of all crosslinking sites. When crosslinking sites are located far upstream (5') of the primer binding site, the signals from the reverse transcriptase tend to be too weak. The detection of a putative crosslinking site is then no longer possible. As a rule of thumb, we use two primers (one matching at the extreme 3' end and one in the middle) for probing an RNA molecule with 120–150 nt (see also Figure 46.2b). It is furthermore obvious that putative crosslinking sites at the extreme 3' end of the RNA cannot be detected or require ligation of 3' linker to the RNA prior primer extension analysis that – however – not always will lead to positive identification. The detection of crosslinking sites close to the primer-binding region is difficult. In principle, it is possible to detect a reverse transcriptase stop 1 nt 5' to the binding site of the primer. In practice, the signal of the radioactively labeled primer is very strong and will thus mask a putative reverse transcriptase stop. This might be circumvented by empirically adjusting the conditions of electrophoresis such as to run the gel until the primer is about to migrate out of it and by extensive prerunning of the sequencing gel (Protocol 4.4).
- 3) For reproducible clean primer extension reactions, the commercially obtained primer should be gel-purified before use (Protocol 4.1). Furthermore, we recommend X-ray films for visualization of crosslinking sites (Protocol 4.4); on autoradiographs, the bands appear much sharper and less fuzzy when compared with phosphorimager scans. X-ray films thus facilitate the correct assignment of putative protein–RNA crosslinking sites, in particular, at weak crosslinking sites whose signals are only slightly above background.

46.5

Identification of Crosslinked Proteins

Once the overall protein–RNA crosslinking pattern in UV-irradiated RNP particles has been determined, the major challenge is to identify the corresponding crosslinked protein. This is achieved by performing immunoprecipitation of UV-irradiated RNP particles under conditions where protein–protein interactions within the particles are disrupted and only a single protein that is crosslinked to RNA(s) is affinity selected (Figure 46.3a).

This approach results in the precipitation of most of the non-crosslinked protein together with the portion of the same protein that is crosslinked to RNA. After digestion of the entire protein moiety with proteinase K and extraction of the RNA, the only RNA molecules isolated are those crosslinked to the precipitated protein. Importantly, non-crosslinked RNA molecules are not precipitated. Consequently, in a subsequent primer extension analysis, the only transcripts detected are

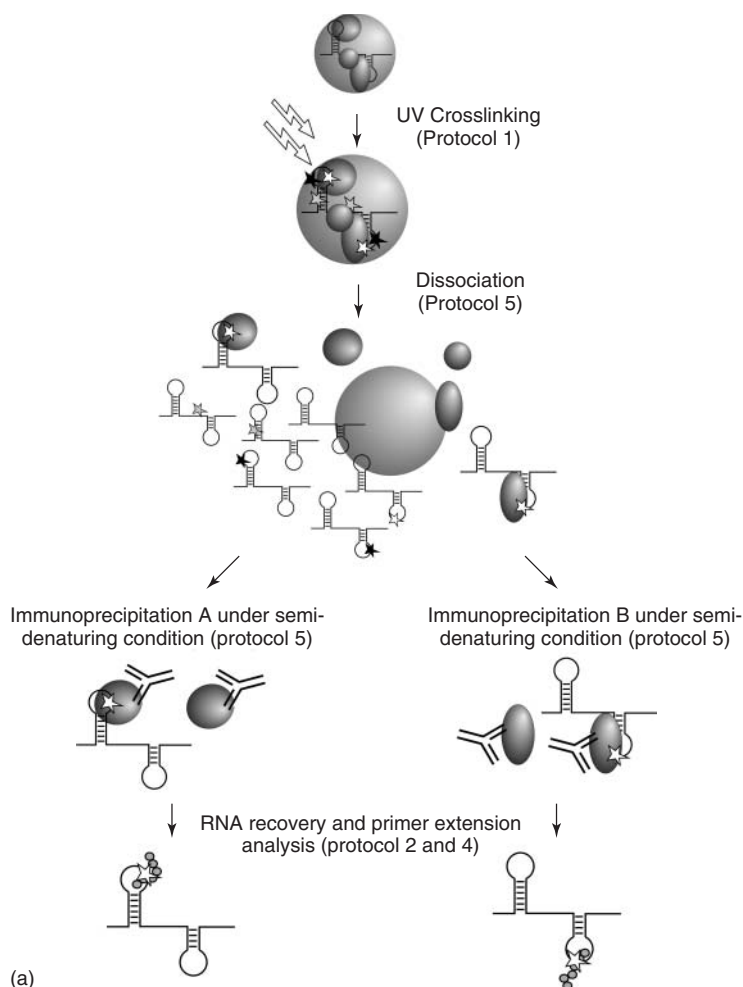


Figure 46.3 (a) Schematic representation of the identification of crosslinked proteins in native RNPs by immunoprecipitation with two different antibodies combined with primer extension analysis. See legend of Figure 46.2a and Protocol 5 for details. (b) Application of the procedure to the identification of protein–RNA crosslinks in U1 snRNPs. Immunoprecipitation was performed with antibodies against the SmF (α -SmF), SmG (α -SmG), and U1 70K (α -70K) proteins (Protocol 5). Primer extension analysis of the coprecipitated RNA was performed with the two primers A and B from Figure 46.2b (Protocol 4). Lanes 1, 3, and 5: non-irradiated U1 snRNPs subjected to the

immunoprecipitation and primer extension analysis with (controls). Lanes 2, 4, and 6: primer extension analysis of the coprecipitated U1 snRNAs after immunoprecipitation of UV-irradiated U1 snRNP particles. The immunoprecipitations with anti-SmG and anti-70K antibodies showed stops at U128 and A29/C31, respectively. Therefore, the SmG protein crosslinks to U127 and U1 70K protein crosslinks to G28 and U30. The weak signal observed for the SmF protein can be explained by strong protein–protein interactions between the Sm proteins SmG and SmF, which cannot be completely disrupted in 2% SDS. See legend of Figure 46.2 and text and Ref. [2] for further details.

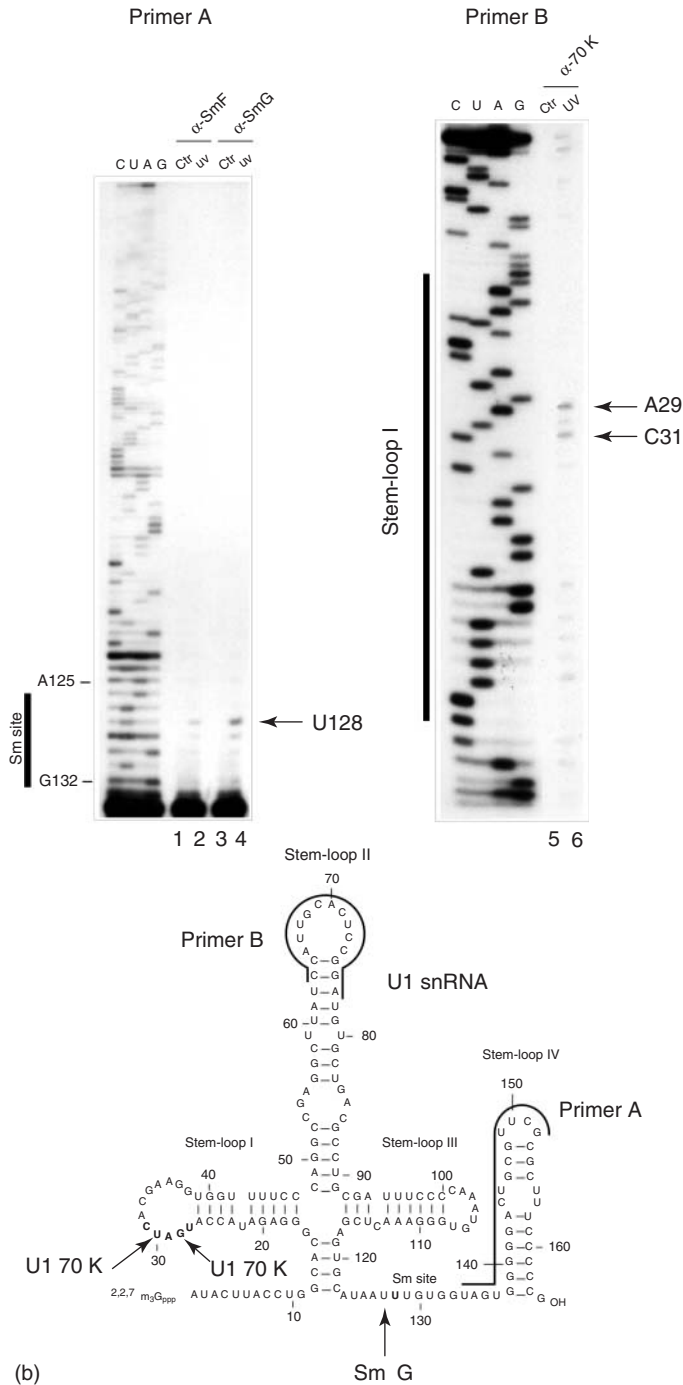


Figure 46.3 (continued)

those with the reverse transcriptase stop at the sites of protein–RNA crosslinking (Figure 46.3a). Full-length transcripts should not be visible. In practice, however, background stops in the full-length RNA or stops due to UV-induced RNA damage or intra-RNA crosslinking events (Figure 46.3b) are usually visible. However, the bands due to the stops at the crosslinking sites of the precipitated protein have significantly greater intensities. Figure 46.3a illustrates the principle of the immunoprecipitation combined with primer extension analysis of the coprecipitated RNA, and Figure 46.3b summarizes results that we have obtained from our crosslinking experiments with native U1 snRNPs [1–3].

This type of analysis is dependent on the number of antibodies available for the different proteins of the particle and on how efficiently each antibody precipitates its corresponding protein, in particular, under semidenaturing conditions that disrupt protein–protein interaction but still preserve the reactivity of the antibodies during immunoprecipitation. As a control, the performance of similar experiments with the preimmune sera of the corresponding antibodies is highly recommended or controls for the affinity purification, in which protein–RNA complexes with non-tagged proteins are used.

Protocol 5 describes in detail all the steps necessary for this analysis. The most critical steps during the analysis are the procedures for dissociation of the RNP particles before immunoprecipitation and for washing of the beads after immunoprecipitation to remove non-specifically bound material. In our hands, the dissociation of particles before immunoprecipitation usually works best in the presence of 1% SDS and is improved by subsequent heating of the samples to 70 °C.

However, in several cases, we observed that higher concentrations of SDS are required to dissociate the particles completely [2, 3]. Washing of the Protein A Sepharose beads after precipitation should include an additional washing step in a new tube. Furthermore, we observed that washing the samples with buffer containing detergent (e.g., Nonidet P-40) leads to a dramatic increase of non-specific background signals in the subsequent primer extension. We have no explanation for this.

46.6

Troubleshooting

- 1) The amount of material required for a single set of experiments (identification of the crosslinking site of one protein to an RNA using one primer) is in the range of at least 1–2 pmol and is strongly dependent on the efficiencies of crosslinking, affinity selection of crosslinked species, and reverse transcription, as well as the complexity of the RNP particles of interest. In general, strong crosslinks can be easily detected when RNA derived from 50 fmol of UV-irradiated RNPs is used as a template for primer extension.
- 2) The above procedures can be applied to the analysis of RNP particles assembled *in vitro* from a known number of defined components. The RNA is most easily

synthesized *in vitro* by phage RNA polymerases [37] or by chemical synthesis. The protein(s) can be produced in either *Escherichia coli* or insect cells, or they can be purified from a readily available biological source [37].

When embarking on such a project, a number of considerations should be borne in mind at the outset. The major problem is directly related to the efficiency of RNP assembly *in vitro*. It would be difficult to discern artificial crosslinking events because of incomplete or non-specific assembly from genuine crosslinks. Similar problems will arise if the protein preparation is contaminated by interfering proteins or by eubacterial RNA. This is, in particular, an obstacle when bacterially expressed RNA binding proteins are being studied.

- 3) Technically speaking, the most demanding aspect of the procedures outlined is that the interpretability of the final result depends heavily on the recovery of RNP and RNA in a large number of consecutive experimental manipulations. Great care has to be taken that the ethanol precipitations are quantitative. Similarly, recovery of RNA or RNP from the glass plate may pose a problem. Also, care has to be taken that all steps requiring the resuspension of a dry RNA pellet in buffer are performed with the necessary patience. For larger complexes, pelleting by ultracentrifugation is a good alternative to ethanol precipitation, as the pelleted material might be easier to resuspend in an appropriate buffer.
- 4) The numerous manipulations required to achieve the aims of the experiments are also possible entry points for contaminations by RNases. Standard precautions have to be taken at the outset. The most important ones are the following: gloves must be worn at all times; when preparing solutions, only double-distilled or Millipore Q water should be used; and all solutions should be sterilized by filtering through 0.2 mm nitrocellulose filters.
- 5) Irradiation conditions (in particular, irradiation time) must be individually adjusted in initial experiments. The nature of the particles and the buffer composition dramatically influence the efficiency of UV crosslinking. Reverse transcriptase analysis of RNPs irradiated for a time ranging from 20 s to 10 min (in small increments) should be performed. For further experiments, one should choose a duration of irradiation that gives a reasonable balance between increasing intensity of bands reflecting protein–RNA crosslinks and lowering of the signal from the full-length RNA.
- 6) The primer extension itself should be performed at least three times with RNA obtained from independent experiments, and it may be necessary to use the RNA from one particular experiment with two different primers.
- 7) The most critical point of the entire analysis is the immunoprecipitation combined with primer extension analysis. False-positive results are obtained if immunoprecipitation is performed under conditions where the RNP complexes are not fully disrupted. For example, the U5 snRNP-specific proteins 40K, 116K, 200K, and 220K form a remarkably stable heteromeric protein complex [38]. Using immunoprecipitation combined with primer extension analysis, we demonstrated that the U5 220K protein crosslinks to loop 1 of U5 snRNA

[1]. However, immunoprecipitation under less stringent conditions compared to Protocol 5 (i.e., dissociation in 0.05% (v/v) SDS; see Figure 46.4) resulted in coprecipitation of crosslinked RNA with antibodies against each of the proteins and thus a comparable reverse transcriptase pattern for the four proteins (data not shown). A similar situation was observed when crosslinking sites in the highly conserved Sm site of U1 snRNPs were analyzed. The seven U snRNP-specific Sm proteins form a highly stable heteromeric ringlike structure that interacts with the Sm site [39]. Immunoprecipitation combined with primer extension analysis under our standard conditions (1% SDS (v/v), 5% (v/v) Triton X-100) revealed a similar crosslinking pattern of the Sm F and Sm G protein to U127. To demonstrate that only Sm G is in contact with U127 in the U1 snRNA, we had to raise the SDS concentration to 2% (v/v) to allow complete dissociation of Sm F and Sm G proteins [2, 3]. The high specificity of our immunoprecipitation procedure is demonstrated in Figure 46.4 for the precipitation of the above-mentioned U5 snRNP-specific proteins 40K, 116K, and 220K. The silver-stained SDS–polyacrylamide gels of the precipitated proteins shows that a single protein can only be precipitated on denaturation in the presence of 1% SDS with subsequent addition of Triton X-100 to 5%.

- 8) Another important aspect is the reliability of the reactivity of the antibodies under these harsh conditions. To exclude the possibility that negative results (i.e., no crosslinking of a particular protein) are due to a poor reactivity of the antibodies, each antibody can be tested for its capability to precipitate a single protein from the RNPs (see above). Precipitated proteins can be visualized by silver-stained SDS–polyacrylamide gel using standard protocols [37, 48, 49]. Here, it is essential to couple the antibodies covalently to the beads in order to minimize the IgG background (Protocol 5.2).
- 9) Affinity purification of antibodies is required if antisera have low antibody titers and/or show cross-reactivity.

Protocols

Protocol 1: UV Irradiation of RNPs

- 1) Starting materials are purified snRNP particles [9, 10, 40, 41], spliceosomal complexes [14–18, 42–46], or reconstituted RNP particles [4, 24]. The concentration of particles that are used for analysis ranged from 1 nM (spliceosomes) to 1 μ M (purified snRNPs). More concentrated samples should be diluted to a protein concentration of approximately $0.1 \mu\text{g } \mu\text{l}^{-1}$. Any buffer is suitable, provided it conforms to the criteria as stated above (Section 46.3 and [2]). The particles that we analyze are in 20 mM Tris-HCl (pH 7.0), 370 mM NaCl, 1.5 mM MgCl_2 , 0.5 mM DTT (in case of U1 snRNPs); in 20 mM 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES)–KOH (pH 7.9), 1.5 mM MgCl_2 , 250 mM NaCl, 0.5 mM DTT, 0.2 mM EDTA

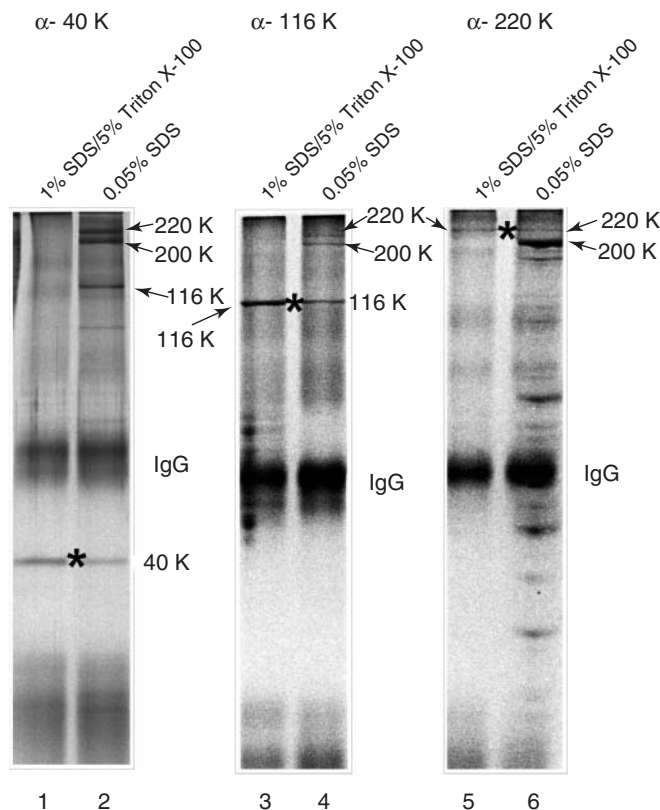


Figure 46.4 Immunoprecipitation of single U5 snRNP-specific proteins from the U5 snRNP through dissociation of the U5 snRNPs by SDS/Triton X-100. U5 snRNPs were dissociated in 1% SDS and 5% Triton X-100 (lanes 1, 3, 5, see Protocol 5.3) or in the presence of 0.05% (v/v) SDS only (lanes 2, 4, 6). The immunoprecipitation was

performed with covalently coupled antibodies (see Protocol 5.2) against the U5 snRNP specific proteins 40K (α -40K), 116K (α -116K), and 220K (α -220K). Proteins were visualized by silver staining. IgG: residual antibody released from the beads. Proteins precipitated under stringent conditions are marked with an asterisk.

(in case of 25S [U4/U6.U5] tri-snRNPs); in 20 mM Tris-HCl (pH 7.9), 250 mM NaCl, 1.5 mM MgCl₂, 0.5 mM DTT (for 12S U2 snRNPs), or in 20 mM HEPES-KOH (pH 7.9), 1.5 mM MgCl₂, 150 mM KCl, 5% glycerol, 12 mM maltose, 0.01% NP-40, 0.5 mM DTT, 0.5 mM PMSF (in case of baker's yeast B^{act} complexes). Below, we describe the protocols for purified snRNP particles with remarks of how to treat more dilute samples such as spliceosomes.

- 2) Divide the sample into two droplets of 25 μ l, and place the droplets carefully onto precooled 10-well multitest slides (ICN Biomedical, USA) or parafilm. Ensure that the droplets stay intact and do not spread over the slide.

- 3) Irradiate for 30–120 s at 254 nm at a distance of 4 cm from the UV source. We use a custom-made holder with four 8 W germicidal lamps (G8T5, Herolab, Germany) mounted in parallel, 4 cm apart (Figure 46.1).
- 4) Carefully pipette the droplets from the glass slide back into a new tube. It is essential that recovery of the sample from the slide is as complete as possible.
- 5) Recover the RNA from crosslinked complexes as outlined in Protocol 2 or perform the immunoprecipitation (Protocol 5) before.

Follow the same procedure to produce a negative control RNP sample, but omit irradiation in the step (3).

Protocol 2: RNA Recovery from UV-Irradiated RNPs

- 1) To 50 μ l of the pooled irradiated samples obtained in Protocol 1, add 40 μ l of the buffer in which the native particles were initially purified.
- 2) Add SDS to a final concentration of 1% (v/v), by adding 10 μ l 10% SDS (v/v) to the above volume.
- 3) Incubate for 10 min at 70 °C with gentle agitation and then allow the sample to cool down to room temperature over a period of 5 min.
- 4) First, add EDTA to 7.5 mM (1.5 μ l of 0.5 M EDTA to above volume), and then add proteinase K to around 1 mg ml⁻¹ (w/v) (10 μ l proteinase K (10 mg ml⁻¹; Roche, Germany) to above volume). Incubate the samples for a minimum of 30 min at 42 °C.
- 5) Extract the RNA by adding 100 μ l phenol/chloroform/isoamyl alcohol (PCI; Roth, Germany) and subsequent vigorous shaking for 2 min. Centrifuge (5 min, 13 000 rpm, 16 000 g) and transfer the aqueous phase containing the RNA to a new tube.
- 6) Add 20 μ g glycogen (Roche, Germany) and 1/10 of the sample volume 3 M sodium acetate, pH 5.3, and precipitate in three sample volumes of ethanol (p.a. grade; Merck, Germany) for 1–2 h at –20 °C (o/n precipitation should be avoided).
- 7) Collect the RNA by centrifugation (20 min, 13 000 rpm at 4 °C) and discard the supernatant. RNA recovery is monitored by inspection of the glycogen pellet, which must be clearly visible.
- 8) Dissolve the RNA in 100 μ l 0.3 M NaOAc, pH 5.3, and precipitate once more with 3 vol of ethanol for a minimum of 2 h at –20 °C. Collect the RNA by centrifugation (see step 7) and dry the sample for 3 min in a Speed Vac.
- 9) Dissolve the RNA in 6.5 μ l CE buffer (10 mM cacodylic acid–KOH, pH 7.0, 0.2 mM EDTA, pH 8.0) with shaking for 10 min. The RNA is stored at –20 °C.

Remark: For convenience, the RNP complexes from dilute samples can be precipitated with ethanol (as described in steps 6 and 7) before RNA extraction. Pellet containing RNPs is resuspended in 100 μ l Proteinase K mix (200 mM Tris-HCl, pH 7.9, 300 mM NaCl, 12.5 mM EDTA, 1% SDS, 1 mg ml⁻¹ Proteinase K). Perform incubation and RNA recovery as stated in steps 4–9.

Protocol 3: UV Irradiation of Naked RNA

- 1) As starting material, use twice as much naked RNA as that contained in the corresponding RNP particle employed in Protocol 1 (step 1) to compensate for loss of RNA on the glass slides during UV irradiation.
- 2) Perform the proteinase K digestion and RNA extraction essentially as described in Protocol 2, steps 2–8, except that glycogen is omitted.
- 3) Dissolve the RNA in 50 μl of the buffer used for the RNP (see Protocol 1, step 1).
- 4) Perform UV irradiation and sample recovery exactly as described in Protocol 1, steps 2–5.
- 5) Further processing of the samples is as described in Protocol 2, steps 6–9. In those cases where the starting material was not doubled, the RNA is resuspended in 3.5 μl instead of 6.5 μl CE buffer.
- 6) Primer extension is performed as outlined in Protocol 4.

Follow the same procedure to produce a negative control RNA sample by omitting irradiation in the step 4.

Protocol 4: Primer Extension Analysis**Protocol 4.1: Purification of the Primer**

Usually 16–22 mer DNA oligonucleotides serve as primers, which can be obtained from any commercial source. It must be gel purified for reproducibly clean primer extension reactions. Approximately, 5 nmol of the primer is first dissolved in 100 μl of 80% formamide, 0.5 x TBE, 0.001% xylene cyanol, and 0.001% bromophenol blue and denatured at 96 °C for 3 min. After cooling to room temperature, it is loaded onto a 25 mm wide slot of a 1 mm thick, 15–20 cm long, denaturing 20% polyacrylamide/8.3 M urea gel and electrophoresed at approximately 1.5 W cm^{-1} . Electrophoresis time depends on the primer length (2 h for a 24 mer). The region of the gel, which contains the primer, is identified by UV shadowing and excised. The gel slice is wetted with elution buffer (20 mM Tris-HCl, pH 7.5, 0.2 mM EDTA, pH 8, 0.15 M NaCl, 0.5% SDS) and cut into small cubes, which are subsequently transferred to a tube and overlaid with 300–500 μl elution buffer. Elution is performed by diffusion (16 h at 30 °C). The eluate is recovered, extracted with phenol/chloroform, and precipitated twice essentially as described for the RNA extraction in Protocol 2, steps 5–9, except that 300 μl PCI is used (step 5) and that the primer is dissolved at 5 pmol μl^{-1} CE buffer (step 9).

Protocol 4.2: 5'-³²P-Labeling of the Primer

- 1) For one 10 μl reaction, the following components are mixed: 2 μl (10 pmol) of purified DNA oligonucleotide, 1 μl 10 x T4 polynucleotide kinase (PNK)

buffer (0.7 M Tris-HCl, pH 7.6, 0.1 M MgCl₂, 0.05 M DTT), 1 µl T4 PNK (New England Biolabs, USA), and 6 µl [γ -³²P]ATP (6000 Ci mmol⁻¹; PerkinElmer, USA). Incubate for 40 min at 37 °C.

- 2) 40 µl CE buffer is added to the reaction and unincorporated nucleotides are removed by G-50 or G-25 Sephadex spin column chromatography (GE Healthcare, UK) according to the manufacturer's instructions. The volume is adjusted to 100 µl with CE buffer and the extent of incorporation is determined (usually about $0.8\text{--}1.5 \times 10^6$ cpm pmol⁻¹ of primer).
- 3) Residual protein and other impurities are removed essentially as described in Protocol 2, steps 5–9, with the following changes: (i) 10 µg glycogen is used in step 6 and (ii) the labeled primer is resuspended in 40 µl CE buffer (step 9).

Protocol 4.3: Primer Extension Reaction

The following different samples are probed by primer extension analysis: (i,ii) non-irradiated and UV-crosslinked naked RNAs from Protocol 3 (RNA Ctr and UV in Figure 46.2b); (iii,iv) RNA extracted from non-irradiated and UV-crosslinked RNPs, Protocol 2 (RNP Ctr and UV in Figure 46.2b); (v,vi) RNA isolated after immunoprecipitation from non- and UV-irradiated RNPs, Protocol 5.3 (Figure 46.3); (vii–x) template RNAs (either native RNA isolated according to Protocol 2 or RNA transcribed *in vitro* with bacteriophage RNA polymerases from an appropriate plasmid template) for the four sequencing reactions used as markers. Template RNAs for marker synthesis should have a concentration of 0.2 pmol µl⁻¹. The experimental procedure for the primer extension closely follows that described in [47].

- 1) For each RNA sample to be analyzed, 1.5 µl of a hybridization mix (HY) is required. It is composed of 0.25 µl 10 x hybridization buffer (0.5 M Tris-HCl, pH 8.4, 0.6 M NaCl, 0.1 M DTT), 0.5 µl ³²P-labeled DNA oligonucleotide, and 0.75 µl H₂O. Enough HY mix for the number of samples to be processed must be prepared.
- 2) To anneal the primer, 1 µl of the RNA is first mixed with 1.5 µl of the HY mix, then heated for 60 s at 96 °C, and allowed to cool at room temperature for 5 min. Samples are briefly centrifuged.
- 3) 0.5–1 µl of ddNTP is added to each of the four marker RNA samples (0.5 mM ddGTP, ddATP, ddTTP, or ddCTP; GE Healthcare, UK).
- 4) For each RNA sample, 2.5 µl of a reverse transcriptase mix (RT) is now required. It is composed of 0.25 µl 10 x reverse transcriptase buffer (0.5 M Tris-HCl, pH 8.4, 0.1 M MgCl₂, 0.6 M NaCl, 0.1 M DTT), 0.1 µl dNTPs (5 mM each dGTP, dATP, dTTP, and dCTP; GE Healthcare, UK), 0.1 µl AMV reverse transcriptase (15 U µl⁻¹, Affymetrix/USB, UK), and 2.05 µl H₂O. Enough RT mix for the number of samples to be processed must be prepared. 2.5 µl RT mix is added per sample, mixed, and incubated for 45 min at 42.5 °C. A hybridization oven is recommended to avoid condensation at the lid of the tube.

- 5) 6.5 μl loading buffer (8.3 M urea, 0.5 x TBE, 0.001% (w/v) bromophenol blue, 0.001% (w/v) xylene cyanol) is added to all samples, except for the markers, which receive 10 μl . Samples can be stored at -20°C for at least one week.

Protocol 4.4: Gel Electrophoresis

The transcribed cDNA products are analyzed on a 9.6% polyacrylamide (acrylamide:bisacrylamide, 19:1)/8.3 M urea gel in 1 x TBE in a Gibco/BRL Model S2 apparatus (0.4 mm thick gel) with 1 x TBE as electrophoresis buffer. Pre-electrophoresis is for 30 min at 65 W. Electrophoresis is at 65–72 W for a time depending on the length of the primer (approximately 2 h for a 20 mer). For autoradiography, the sequencing gels are first transferred to a used X-ray film for support and covered with kitchen wrapping film. Alternatively, sequencing gels can be fixed in 40% methanol/10% acetic acid, transferred to Whatman 3MM paper and dried under vacuum (Bio-Rad model 583 gel dryer). A BioMax film (Kodak) is exposed to the gel at -70°C for 1–10 days in the presence of intensifying screens. The long exposure times are required when performing the immunoprecipitation experiments combined with primer extension analysis, because of the inherently low yields of immunoprecipitation.

Protocol 5: Immunoprecipitation of the RNA–Protein Crosslinks

Protocol 5.1: Noncovalent Coupling of Antibodies to Protein A Sepharose

Immunoprecipitation was found to be optimal with per assay 15 μl packed matrix volume of Protein A Sepharose beads (GE Healthcare, UK) coupled with antiserum. Depending on the number of samples that are assayed, proportionally more bead slurry can be coupled with correspondingly increased amounts of antiserum. The coupled beads can be distributed afterward between the different tubes.

- 1) For coupling of the antibody to beads, an amount of slurry (30 μl) corresponding to 15 μl of beads (packed volume) is taken and washed three times with 500 μl aliquots of PBS (20 mM Na_2HPO_4 , pH 8.0, 130 mM NaCl).
- 2) The antiserum is diluted with PBS to 500 μl and added to the washed beads. Normally, 50 μl of antiserum is sufficient for one immunoprecipitation, but this volume may have to be adjusted, depending on the titer of the antiserum. Coupling is performed overnight by head-over-tail rotation at 4°C or 1–3 h at RT.
- 3) After coupling, beads are washed three times with 500 μl PBS. Tubes are changed by transferring the beads with the last washing aliquot to a new tube using a plastic pipette tip with a cut-off end. The washed beads with the coupled antibody are then overlaid with 15 μl PBS and kept on ice until use.

Remark: If affinity purified antibodies are used instead of antisera, blocking of the beads is required. For this, washed Protein A Sepharose beads are first incubated with 400 μl PBS supplemented with 50 $\mu\text{g ml}^{-1}$ tRNA, 50 $\mu\text{g ml}^{-1}$ glycogen, 0.5 mg ml^{-1} BSA by head-over-tail rotation for 1 h at RT or o/n at 4°C .

and then antibodies are added. Blocking might also help in case of antisera if high background is observed in lanes with non-irradiated immunoprecipitated RNP samples on a sequencing gel.

Protocol 5.2: Covalent Coupling of Antibodies to Protein A Sepharose

Covalent coupling of antibodies is recommended when the capability of the antibodies to selectively precipitate a single protein from an RNP under semidenaturing conditions (Protocol 5.3) is tested, in order to exclude the possibility that negative results are due to the poor reactivity of the antibodies under these conditions (see also Section 46.6, point (8)).

For covalent coupling, it is recommended to increase the total amount of Protein A Sepharose beads and the amount of antiserum is usually twice the volume of the beads.

- 1) 30 μ l Protein A Sepharose beads are washed with PBS (Protocol 5.1) and then incubated with 60 μ l of antiserum in a final volume of 500 μ l PBS, 0.05% (v/v) Nonidet P-40 (NP-40) with head-over-tail rotation overnight at 4 °C.
- 2) The beads are washed five times with 500 μ l PBS, 0.05% NP-40 at 4 °C.
- 3) The antibody-coupled beads are equilibrated two times with 300 μ l of 200 mM sodium borate (pH 9.0) at room temperature.
- 4) Crosslinking of the antibodies to the beads is achieved by incubation with 500 μ l DMP (dimethyl pimelimidate dihydrochloride; Sigma, USA) at a final concentration of 5.2 mg ml⁻¹ in 200 mM sodium borate for 1 h at room temperature with head-over-tail rotation. Note that the pH of the solution must be above 8.3.
- 5) The supernatant is removed as completely as possible, and the reaction is stopped by addition of 300 μ l of 0.2 M ethanolamine–HCl, pH 8.0, to the beads and further incubation with head-over-tail rotation for 1 h.
- 6) The beads are then washed with PBS (Protocol 5.1) and residual non-crosslinked antibodies are removed by three additional washes with 500 μ l of 0.1 M glycine–HCl, pH 2.7.
- 7) After a final wash with PBS containing 0.02% NaN₃, the slurry can be stored for at least six months at 4 °C.
- 8) The covalently coupled antibody beads are now used in the immunoprecipitation exactly as described in Protocol 5.3. After the final wash, beads are incubated with an appropriate volume of SDS sample buffer (125 mM Tris–HCl, pH 6.8, 1% (v/v) SDS, 5% (v/v) glycerol, 10 mM DTT, 0.005% (w/v) bromophenol blue) and heated for 5 min at 70 °C. Beads are spun down with maximum speed (10 000 g) and the supernatant is loaded onto an SDS–PAA gel [48].
- 9) Proteins are visualized by silver staining according to [49].

Protocol 5.3: Dissociation of RNP Particles and Immunoprecipitation

For the immunoprecipitation experiments, native or reconstituted RNP particles in a volume of 50 μ l of appropriate buffer (for buffer conditions and UV irradiation,

see Protocol 1, steps 1–4) are used. For the reliable assignment of crosslinks, it is essential to include a sample that was not UV irradiated, but otherwise treated in an identical manner. Usage of 0.5 ml tubes facilitates handling as well as decreases beads loss and unspecific background.

- 1) Add SDS to a final concentration of 1% (v/v) to the samples from Protocol 1 and incubate for 10 min at 70 °C on a shaker. Use 2% SDS (v/v) in those cases where the protein–protein interactions are known or were found to be extremely strong (see above).
- 2) Allow the samples to cool at room temperature for 5 min. Then, add Triton X-100 (density 1.06 g ml⁻¹, molecular biology grade; Sigma, USA) to a final concentration of 5% (v/v). Use of the concentrated Triton X-100 stock solution is necessary to keep the final volume as low as possible. Gentle mixing is necessary to completely dissolve the added Triton X-100, which initially forms a separate phase at the bottom of the tube.
- 3) Adjust the sample volume to 350 µl with PBS and add the mixture to the prepared antibody-coupled beads (Protocol 5.1). Incubate with head-over-tail rotation for 1–1.5 h at 4 °C.
- 4) Wash the samples four times with 500 µl aliquots of PBS and transfer the slurry into a new tube at the fourth washing step. Wash the beads once more with 500 µl PBS. Carefully check the recovery of the beads during the washing procedure by inspecting the amount of beads visible in the tube after each step; any loss of material must be avoided.
- 5) Remove the supernatant as completely as possible, then add to the beads 100 µl Proteinase K mix (see Remark to Protocol 2) and incubate at 42 °C for at least 30 min resuspending the beads time to time.
- 6) Spin the beads briefly in the centrifuge (6000 rpm for 10 s), transfer the supernatant to a new tube, and recover the RNA from it essentially as described in Protocol 2, steps 5–9, except that shaking is for 5 min (step 5) and that the RNA is dissolved in 3.5 µl (step 9).
- 7) Proceed with the primer extension as outlined in Protocol 4.

Remarks:

- 1) RNPs from dilute samples can be precipitated after UV irradiation (see Remark to Protocol 2; it is essential to avoid overdrying samples). Pellet containing RNPs is resuspended in 25 µl PBS supplemented with 3% SDS and incubated on a shaker for at least 10 min at RT and then at 70 °C for 10 min. After cooling to RT, 25 µl PBS containing 10% Triton X-100 is added and samples are shaken for further 10 min. Then proceed as described in steps 3–7.
- 2) As mentioned above, IgG beads (GE Healthcare, UK) can be used instead of protein-specific antibodies for affinity selection if RNP particles under study contain Protein A or TAP-tagged proteins. IgG beads must be blocked as described in Remark to Protocol 5.1 to decrease unspecific binding of proteins and RNA.

Protocol 5.4: Selection of Protein–RNA Conjugates Using Nickel Beads

Histidine-tag (His-tag) is commonly used as a fast and inexpensive tool for protein purification. Therefore, His-tagged proteins are often available for *in vitro* reconstitution of RNP complexes. Easiness of protein tagging in some organisms makes it also feasible to use His-tag for affinity selection of protein–RNA crosslinks from native RNPs or *in vitro* assembled complexes requiring whole cell or nuclear extracts.

- 1) Wash 40 μ l (packed volume) of nickel-nitrilotriacetic acid (Ni-NTA) Agarose beads (Qiagen, Germany) three times with 450 μ l PBS. Work with 0.5 ml tubes.
- 2) Block the beads with 400 μ l PBS supplemented with 50 μ g ml⁻¹ tRNA, 50 μ g ml⁻¹ glycogen, 0.5 mg ml⁻¹ BSA by head-over-tail rotation for 1 h at RT or o/n at 4 °C.
- 3) Wash the beads five times with 450 μ l PBS. Split the beads into two aliquots during the last wash.
- 4) Prepare the non- and UV-irradiated RNP samples and incubate them with the Nickel beads as described in steps 1–3 of Protocol 5.3 or Remark 1 to it except that the samples should additionally contain 7.5 mM imidazole.
- 5) Wash the beads six times with 450 μ l PBS supplemented with 0.1% NP-40 and 7.5 mM imidazole. Change the tube before the sixth wash.
- 6) Wash once more with 450 μ l PBS.
- 7) Remove supernatant and add 100 μ l proteinase K mix.
- 8) Incubate and recover RNA as in Protocol 2, steps 4–9.
- 9) Perform the primer extension according to Protocol 4.

Acknowledgments

We thank Reinhard Rauhut for helpful discussions and Josep Vilardell (IBMB, Spain) for providing the plasmid for the expression of the MS2-MBP fusion protein. This work was supported by a grant from the European Commission (EURASNET-518238) to Reinhard Lührmann.

References

1. Urlaub, H., Hartmuth, K., Kostka, S., Grelle, G., and Lührmann, R. (2000) A general approach for identification of RNA-protein cross-linking sites within native human spliceosomal small nuclear ribonucleoproteins (snRNPs). Analysis of RNA-protein contacts in native U1 and U4/U6.U5 snRNPs. *J. Biol. Chem.*, **275**, 41458–41468.
2. Urlaub, H., Hartmuth, K., and Lührmann, R. (2002) A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods*, **26**, 170–181.
3. Urlaub, H., Raker, V.A., Kostka, S., and Lührmann, R. (2001) Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J.*, **20**, 187–196.
4. Nottrott, S., Urlaub, H., and Lührmann, R. (2002) Hierarchical, clustered protein

- interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J.*, **21**, 5527–5538.
5. Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 9613–9618.
 6. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M. *et al.* (2010) PAR-CLIP – a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J. Vis. Exp.*, **41**. doi: 10.3791/2034.
 7. Zhang, C. and Darnell, R.B. (2011) Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.*, **29**, 607–614.
 8. Jensen, K.B. and Darnell, R.B. (2008) CLIP: crosslinking and immunoprecipitation of *in vivo* RNA targets of RNA-binding proteins. *Methods Mol. Biol.*, **488**, 85–98.
 9. Kastner, B. (1998) in *RNP Particles, Splicing and Autoimmune Diseases* (ed. J. Schenkel), Springer, Berlin, pp. 95–140.
 10. Will, C.L., Kastner, B., and Lührmann, R. (1994) in *RNA Processing: A Practical Approach* (ed. S.J. Higgins and B.D. Hames), IRL Press, Oxford, **1**, pp. 141–177.
 11. Wahl, M.C., Will, C.L., and Lührmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701–718.
 12. Burge, C.B., Tuschl, T., and Sharp, P.A. (1999) in *The RNA World* (eds R.F. Gesteland, T.R. Cech, and J.F. Atkins), 2nd edn, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, pp. 525–560.
 13. Will, C.L. and Lührmann, R. (2006) in *The RNA World*, 3rd edn (eds R.F. Gesteland, T.R. Cech, and J.F. Atkins), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 369–400.
 14. Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Lührmann, R. (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA*, **16**, 2384–2403.
 15. Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Lührmann, R. (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature*, **452**, 846–850.
 16. Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.*, **26**, 5528–5543.
 17. Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Lührmann, R. (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.*, **26**, 1737–1748.
 18. Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell*, **36**, 593–608.
 19. Puig, O., Casparly, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, **24**, 218–229.
 20. Urlaub, H., Kruft, V., Bischof, O., Müller, E.C., and Wittmann-Liebold, B. (1995) Protein-rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. *EMBO J.*, **14**, 4578–4588.
 21. Urlaub, H., Thiede, B., Müller, E.C., Brimacombe, R., and Wittmann-Liebold, B. (1997) Identification and sequence analysis of contact sites between ribosomal proteins and rRNA in *Escherichia coli* 30 S subunits by a new

- approach using matrix-assisted laser desorption/ionization-mass spectrometry combined with N-terminal microsequencing. *J. Biol. Chem.*, **272**, 14547–14555.
22. Kühn-Hölsken, E., Lenz, C., Dickmanns, A., Hsiao, H.-H., Richter, F.M., Kastner, B., Ficner, R., and Urlaub, H. (2010) Mapping the binding site of snurportin 1 on native U1 snRNP by cross-linking and mass spectrometry. *Nucleic Acids Res.*, **38**, 5581–5593.
 23. Kramer, K., Hummel, P., Hsiao, H.-H., Luo, X., Wahl, M., and Urlaub, H. (2011) Mass-spectrometric analysis of proteins cross-linked to 4-thio-uracil- and 5-bromo-uracil-substituted RNA. *Int. J. Mass Spectrom.*, **304**, 184–194.
 24. Luo, X., Hsiao, H.-H., Bubunenkov, M., Weber, G., Court, D.L., Gottesman, M.E., Urlaub, H., and Wahl, M.C. (2008) Structural and functional analysis of the *E. coli* NusB-S10 transcription antitermination complex. *Mol. Cell*, **32**, 791–802.
 25. Meisenheimer, K.M., Meisenheimer, P.L., and Koch, T.H. (2000) Nucleoprotein photo-cross-linking using halopyrimidine-substituted RNAs. *Methods Enzymol.*, **318**, 88–104.
 26. Brimacombe, R., Stiege, W., Kyriatsoulis, A., and Maly, P. (1988) Intra-RNA and RNA-protein cross-linking techniques in *Escherichia coli* ribosomes. *Methods Enzymol.*, **164**, 287–309.
 27. Cooperman, B.S. (1988) Affinity labeling of ribosomes. *Methods Enzymol.*, **164**, 341–361.
 28. Moore, M.J. and Sharp, P.A. (1992) Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science*, **256**, 992–997.
 29. Wyatt, J.R., Sontheimer, E.J., and Steitz, J.A. (1992) Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA. *Genes Dev.*, **6**, 2542–2553.
 30. Dix, I., Russell, C.S., O'Keefe, R.T., Newman, A.J., and Beggs, J.D. (1998) Protein-RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*. *RNA*, **4**, 1675–1686.
 31. Reed, R. and Chiara, M.D. (1999) Identification of RNA-protein contacts within functional ribonucleoprotein complexes by RNA site-specific labeling and UV crosslinking. *Methods*, **18**, 3–12.
 32. Konarska, M.M. (1999) Site-specific derivatization of RNA with photocrosslinkable groups. *Methods*, **18**, 22–28.
 33. Yu, Y.T. (2000) Site-specific 4-thiouridine incorporation into RNA molecules. *Methods Enzymol.*, **318**, 71–88.
 34. Rhode, B.M., Hartmuth, K., Urlaub, H., and Lührmann, R. (2003) Analysis of site-specific protein-RNA cross-links in isolated RNP complexes, combining affinity selection and mass spectrometry. *RNA*, **9**, 1542–1551.
 35. Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
 36. Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R., and Lührmann, R. (1999) Functional interaction of a novel 15.5kD[U4/U6 U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J.*, **18**, 6119–6133.
 37. Sambrook, J., Fritsch E.F., and Maniatis, T. (eds) (1989) *Molecular Cloning. A Laboratory Manual*, 2nd edn, Cold Spring Harbour Laboratory Press, Cold Spring Harbour.
 38. Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S., and Lührmann, R. (1998) The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol. Cell. Biol.*, **18**, 6756–6766.
 39. Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Lührmann, R., Li, J., and Nagai, K. (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*, **96**, 375–387.

40. Karaduman, R., Fabrizio, P., Hartmuth, K., Urlaub, H., and Lührmann, R. (2006) RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. *J. Mol. Biol.*, **356**, 1248–1262.
41. Dybkov, O., Will, C.L., Deckert, J., Behzadnia, N., Hartmuth, K., and Lührmann, R. (2006) U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. *Mol. Cell. Biol.*, **26**, 2803–2816.
42. Jurica, M.S. and Moore, M.J. (2002) Capturing splicing complexes to study structure and mechanism. *Methods*, **28**, 336–345.
43. Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Lührmann, R. (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, **298**, 2205–2208.
44. Hartmuth, K., Urlaub, H., Vornlocher, H.-P., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16719–16724.
45. Hartmuth, K., Vornlocher, H.-P., and Lührmann, R. (2004) Tobramycin affinity tag purification of spliceosomes. *Methods Mol. Biol.*, **257**, 47–64.
46. Reed, R. and Palandjian, L. (1997) in *Eukaryotic mRNA Processing* (ed. A.R. Krainer), IRL Press, Oxford, pp. 103–129.
47. Zaug, A.J. and Cech, T.R. (1995) Analysis of the structure of *Tetrahymena* nuclear RNAs *in vivo*: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. *RNA*, **1**, 363–374.
48. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
49. Blum, H., Beier, H., and Gross, H.J. (1987) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis*, **8**, 93–99.

47

Sedimentation Analysis of Ribonucleoprotein Complexes

*Tanja Rösel, Jan Medenbach, Andrey Damianov, Silke Schreiner,
and Albrecht Bindereif*

47.1

Introduction

Several cellular processes such as translation or pre-mRNA splicing are catalyzed by large multimeric complexes that contain essential RNA components. The ordered and stepwise assembly of these complexes often proceeds through complicated maturation pathways. Classical examples for this are provided by ribosomes, heterogeneous nuclear ribonucleoprotein (hnRNP) particles, and the spliceosomal small nuclear ribonucleoproteins (snRNPs) U1, U2, and U4/U6U5. The biogenesis of snRNPs, which involves, in addition, trafficking between different cellular compartments and the assembly of snRNPs to the active spliceosome are particularly well studied and are used in the following as a specific example (reviewed by Will and Lührmann [1]).

Different techniques such as affinity purification have been used to study the assembly and composition of the spliceosomal snRNPs. In particular, the fractionation of snRNPs by density gradient ultracentrifugation has proved to be a powerful tool for the separation and enrichment of individual snRNPs in their native state. Taking advantage of their stability under high-salt conditions in the presence of Mg^{2+} ions, snRNPs were purified from HeLa cell nuclei as early as 1981, applying only a series of different cesium chloride gradients [2]. Depending on the conditions, either intact snRNPs or core snRNPs containing the Sm proteins, but lacking the specific protein components, were obtained [3]. Earlier, the same approaches were applied to the study of ribosomes and hnRNP particles [4]. In addition, it has been shown for several complexes that catalytic activity is retained after ultracentrifugation, demonstrating that the particles remain in a native and functional state (as an example, see RNase P [5]).

A further advantage of the density gradient centrifugation is that it can be applied to crude cellular extracts, such as nuclear extracts [6], as well as to purified fractions or complexes, such as samples obtained by immunoaffinity selection. Density gradient ultracentrifugation has been combined with other fractionation

techniques to characterize spliceosomal subcomplexes, the exosome, and viral ribonucleoproteins (RNPs) [7–14].

In this chapter, we describe the techniques of glycerol and cesium chloride density gradient ultracentrifugation, using the fractionation of the spliceosomal snRNPs present in the HeLa cell nuclear extract as an example. The protocols given can also be used to fractionate other samples such as crude splicing reactions or eluates from affinity selections. The two ultracentrifugation methods separate the samples according to different physical properties. In the zonal glycerol gradient density centrifugation, the sample is separated in a preformed gradient of a viscous component (glycerol and sucrose) for a defined time, resulting in fractionation due to differences in the sedimentation constant (*Svedberg*, S), which depends on the weight, volume, density, and shape of the particle. In contrast, during isopycnic ultracentrifugation in cesium chloride or cesium sulfate, the gradient is formed during the run. During centrifugation, particles concentrate at their isopycnic (= equal density) position where their density matches that of the surrounding solution. Thus, different particles are separated based on their different densities.

47.2

Glycerol Gradient Centrifugation

As an example, we describe here the fractionation of nuclear extract in a linear 10–30% glycerol gradient. Depending on the particles to be separated, different glycerol gradients can be used. Instead of preparing RNA from the fractions (see Section “Preparation of RNA from Gradient Fractions”), they can also be subjected to immunoaffinity selections or other purification methods. In parallel to RNA, proteins can also be easily prepared from the fractions for further analysis (see Section “Simultaneous Preparation of RNA and Proteins”).

Equipment

- SW-40 Ti rotor with polyallomer centrifugation tubes
- TLS-55 rotor with polyallomer centrifugation tubes
- Gradient Mixer (Hoefer Scientific Instruments SG Series or Gradient Master, see “Notes and Troubleshooting” under Section 47.2)
- Sterile, RNase-free 1.5 ml Eppendorf tubes
- Cooling microcentrifuge
- Glass capillaries or disposable micropipettes (e.g., Ringcaps®, Hirschmann).

Reagents

- DMPC (dimethyl pyrocarbonate)–water (1 ml DMPC per 1 l dd-H₂O, stir for 1 h, autoclave twice)
- 1 M dithiothreitol (DTT) in DMPC–H₂O
- 0.1 M phenylmethyl sulfonyl fluoride (PMSF) in ethanol

- 3 M sodium acetate, pH 5.2, in DMPC–H₂O
- 20 mg ml⁻¹ of glycogen in DMPC–H₂O
- Ethanol, 100, 70, 80, and 50%
- Phenol–chloroform–isoamylalcohol (25 : 24 : 1), saturated with TE buffer (10 mM Tris/HCl, pH 8.0, 1 mM EDTA)
- Buffer D (20 mM HEPES/KOH, pH 8.0, 100 mM KCl, 0.2 mM EDTA pH 8.0, 1.5 mM MgCl₂, 20% glycerol, in DMPC–H₂O, add freshly 0.5 mM PMSF and 0.5 mM DTT)
- 10 x Buffer G (200 mM HEPES/KOH, pH 8.0, 1.5 M KCl, and 15 mM MgCl₂)
- 10% glycerol solution (freshly prepared; for 100 ml, combine 100 µl leupeptin, 50 µl 1 M DTT, 500 µl 0.1 M PMSF solution, adjust to 10% glycerol and 1 x buffer G, and add DMPC–H₂O to reach the final volume of 100 ml; filter solution through a 0.45 µm filter)
- Glycerol solutions of 15, 20, 25, and 30% (same as 10% glycerol solution, except for the glycerol concentration)
- SDS-polyacrylamide gel electrophoresis (SDS-PAGE) protein sample buffer (2% SDS, 10% glycerol, 50 mM Tris/HCl, pH 6.8, 0.005% bromophenolblue)
- 5 x Agarose gel loading buffer (5 x Tris/borate/EDTA (TBE), 10% glycerol, 0.025% bromophenolblue)
- Leupeptin (4 mg ml⁻¹ in DMPC–H₂O)
- 16S and 23S rRNAs from *Escherichia coli* (Roche)

Method

Preparation of the Glycerol Gradient

- 1) Clean the gradient mixer with 100 and 50% ethanol, close valves, and rinse the chambers with the 30% (left chamber in Figure 47.1) or 10% (right chamber in Figure 47.1) glycerol solution; make sure that no air bubbles get trapped in the connections.
- 2) Place the gradient mixer on a magnetic stirrer, fill the chambers (valves closed) with 5.5 ml of 10% (right chamber with the outlet pipe) and 30% glycerol solution without trapping air bubbles, and add a sterile magnetic stir bar to the chamber with the outlet pipe (Figure 47.1).
- 3) Place a centrifugation tube on ice, connect a sterile glass capillary with the outlet pipe of the mixer *via* a flexible silicone tube, and insert the capillary into the centrifugation tube such that it touches the bottom.
- 4) Start the magnetic stirrer and open the valves, such that the centrifugation tube is slowly filled with the glycerol solution, the 10% solution being underlayered with the denser solution.
- 5) Before pouring another gradient, rinse the mixer again with the 30 and 10% glycerol solution, respectively.
- 6) Place the gradients at 4 °C for 1 h for equilibration.

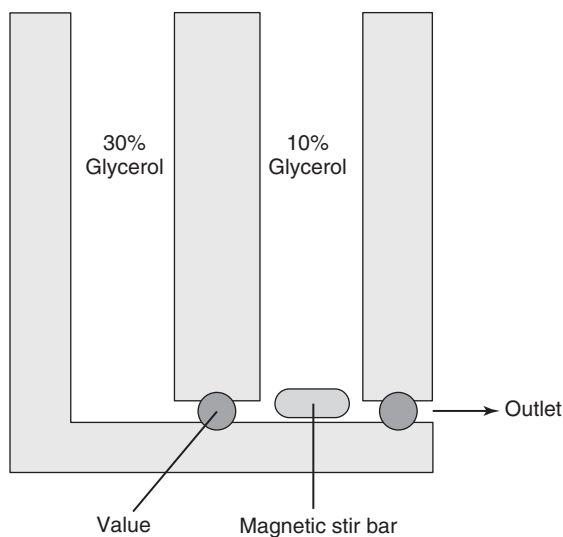


Figure 47.1 Schematic display of the Hoefer gradient mixer used to prepare glycerol gradients. For each chamber, the concentration of the glycerol stock solution for preparation of a linear 10–30% gradient is given; the outlet pipe is indicated by an arrow.

Preparation of a Small-Scale Glycerol Gradient (2 ml)

For applications on a smaller scale, it is possible to use 2 ml instead of 10 ml gradients, as described above. In principle, a step gradient is prepared and the linear gradient is achieved simply by diffusion.

- 1) Insert a short Pasteur pipette into a 2 ml polyallomer centrifugation tube, so that the tip of the pipette touches the wall of the tube just above the bottom.
- 2) Pipette the glycerol solutions (400 μ l each) one after another into the Pasteur pipette. Avoid air bubbles that stick to the bottom of the tube and might disturb separation during centrifugation.
- 3) Create a step gradient by underlayering the glycerol solutions starting with 10% glycerol followed by 15, 20, 25, and 30%.
- 4) Place the centrifugation tube at 4 °C for about 16 h or overnight to allow formation of a linear gradient through diffusion.
- 5) The sample preparation after centrifugation is carried out for the 10 ml gradient as described in the following paragraphs; note that only up to 120 μ l of sample volume should be loaded onto a 2 ml gradient.

Sample Preparation and Centrifugation

- 1) Thaw 400 μ l of HeLa nuclear extract slowly on ice, remove aggregates by a short centrifugation (14 000 rpm, 1 min), and mix it with 1.1 ml of 1 x buffer G freshly supplemented with 0.5 mM DTT and 0.5 mM PMSF.

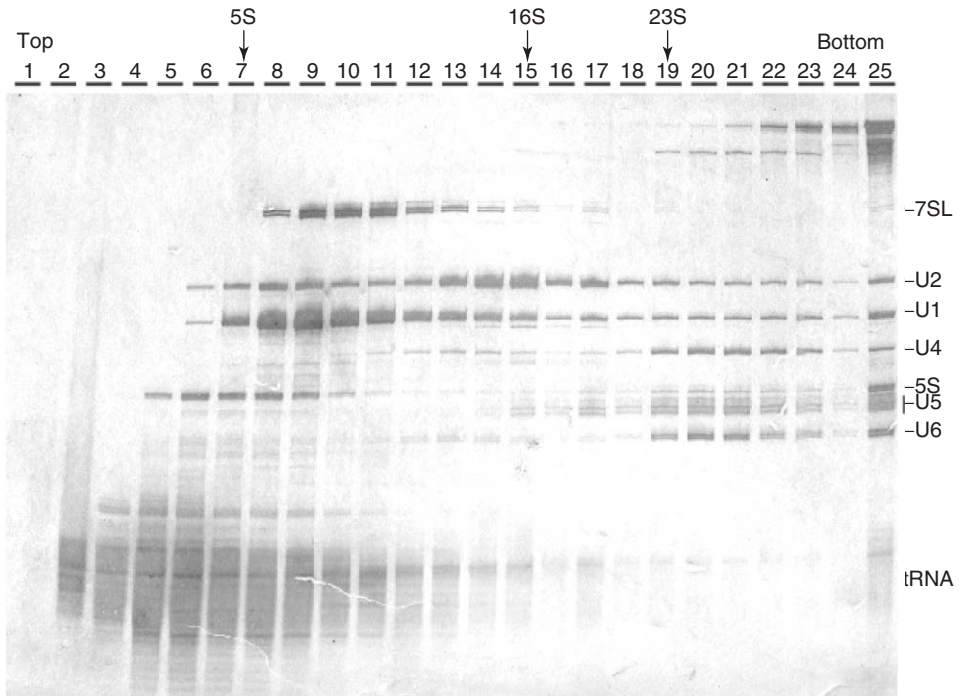


Figure 47.2 Fractionation of nuclear extract by glycerol gradient centrifugation: RNA analysis. After ultracentrifugation of 200 μ l of HeLa nuclear extract through an 11 ml linear 10–30% glycerol gradient (see Section “Sample Preparation and Centrifugation” for conditions), 25 fractions of 500 μ l each were withdrawn from the top to the bottom of

the gradient. RNA from 50 μ l of each fraction was isolated and analyzed by denaturing polyacrylamide gel electrophoresis and silver staining. The identities of the RNAs are indicated on the right (U1, U2, U4, U5, U6 snRNAs, 5S rRNA, 7SL RNA, and tRNAs). The positions of the 5S, 16S, and 23S sedimentation markers are given at the top.

- 2) Carefully overlay the prepared gradients with the diluted extract solution and balance the tubes together with the centrifugation buckets by adding 1 x buffer G.
- 3) Centrifuge for 17 h at 32 000 rpm (corresponding to 130 000 g) and at 4 °C in a precooled SW-40 rotor.
- 4) Carefully fractionate the gradient from top to bottom by withdrawing 25 aliquots of 500 μ l each; resuspend the pellet in the last aliquot (Figure 47.2).

Preparation of RNA from Gradient Fractions

- 1) Split each gradient fraction in two portions for easier handling during preparation of the RNA.
- 2) Add 250 μ l of phenol–chloroform–isoamylalcohol to each 250 μ l of the fractions, vortex thoroughly, and centrifuge for 10 min at 20 000 g.

- 3) Transfer each upper aqueous phase to a new Eppendorf tube and add 20 μl 3 M sodium acetate, 2 μl glycogen (20 mg ml^{-1}), and 550 μl absolute ethanol; mix by inverting the tube several times.
- 4) Put the tubes at -70°C for 10 min, centrifuge for 20 min at 20 000 g and 4°C , discard the supernatants, and wash each pellet at room temperature (RT) with 500 μl 70% ethanol.
- 5) Redissolve the pellets in 10 μl DMPC– H_2O each and combine the two aliquots from each fraction.
- 6) Analyze the RNA by denaturing PAGE.

Simultaneous Preparation of RNA and Proteins

In parallel to the isolation of RNA, proteins can be prepared from the same fractions. Because of the high concentration of proteins in some fractions derived from crude cellular extracts, an initial dilution step of the gradient fractions (1 : 2 with DMPC– H_2O) is recommended.

- 1) After phenol extraction and removal of the upper aqueous phase, add 5 volumes of acetone to the phenol phase for protein precipitation.
- 2) Mix and store samples for at least 1 h at -20°C .
- 3) Centrifuge for 30 min at 20 000 g, remove supernatant from pellet, and wash with 500 μl of 80% ethanol.
- 4) Dissolve the pellets in a small volume (10–20 μl) of SDS-PAGE protein sample buffer, boil for 10 min, and analyze by SDS-PAGE.

Control Gradient with Sedimentation Markers

Since 5S ribosomal RNA is no longer commercially available, it is recommended that a silver-stained denaturing polyacrylamide gel of a gradient containing nuclear extract to detect the peak of free 5S rRNA (Figure 47.2) be analyzed. For the 16S and 23S sedimentation markers, proceed as follows:

- 1) Prepare the gradient as described above
- 2) Mix 10 μl 16S/23S rRNA ($4 \mu\text{g } \mu\text{l}^{-1}$, Roche) with 390 μl buffer D (or the buffer of the sample loaded onto the analytical gradients).
- 3) Add 1.1 ml 1 x buffer G and mix.
- 4) Load the sample onto the gradient; centrifuge and fractionate as described in the Section “Sample Preparation and Centrifugation”.
- 5) Mix 12 μl of every second fraction with 3 μl agarose gel sample buffer and analyze on a 0.6% agarose gel in 0.5 x TBE.
- 6) Visualize RNAs by ethidium bromide staining.

Notes and Troubleshooting

- 1) Depending on the size, range, and properties of the RNPs to be separated, the running time (17 h), rotor type (SW-40), and conditions (32 000 rpm; 4°C) described in the example above (see Section “Sample Preparation and

- Centrifugation”) may be varied. Run the control gradient (see Section “Control Gradient with Sedimentation Markers”) under exactly the same conditions.
- 2) Instead of glycerol, as described here, gradient sedimentation can also be performed in sucrose gradients (e.g., 15–30% [4]; 15–45% [15]).
 - 3) Handle gradient fractions and prepared RNAs always on ice to minimize degradation. During RNA preparation make sure that only the aqueous phase is transferred to a fresh Eppendorf tube; do not withdraw material from the interphase or phenol phase to avoid contaminations that may perturb downstream RNA analyses. Before redissolving the RNA in DMPC–H₂O, air-dry the pellet and make sure that no residual ethanol is left inside the Eppendorf tube. On the other hand, do not overdry the pellet (3–5 min at RT is usually sufficient).
 - 4) Instead of using the Hoefer gradient mixer to prepare the gradients (see Section “Preparation of the Glycerol Gradient”), the programmable BioComp Gradient Master (Frederickton, NB, USA) is recommended for higher reproducibility and for faster preparation of the gradients.
 - 5) Make sure that the samples to be analyzed by glycerol gradient centrifugation do not contain too much glycerol – here more than 10% – to prevent submersion of the sample. If necessary, dilute or dialyze against a suitable buffer to reduce the glycerol concentration.
 - 6) If samples obtained from the glycerol gradient centrifugation are to be subjected to immunoaffinity purification or other methods, we also recommend dilution or dialysis (e.g., with NET-100 buffer: 50 mM Tris–HCl, pH 8.0, 100 mM NaCl, 0.05% Nonidet P-40, 0.5 mM DTT) in order to reduce the high glycerol concentration.
 - 7) For visualization of RNAs in agarose gels (see Section “Control Gradient with Sedimentation Markers”), add an aliquot of ethidium bromide stock solution (10 mg ml^{–1}) directly to the agarose solution (final concentration of 500 ng ml^{–1}) before casting the gel; if the gel is stained after electrophoresis, degradation of the RNA may occur.

47.3

Fractionation of Ribonucleoproteins (RNPs) by Cesium Chloride Density Gradient Centrifugation

When applying isopycnic ultracentrifugation for the analysis of proteins or RNA–protein complexes, one has to consider that the high ionic strength may destabilize and dissociate the complexes, resulting in denaturation and precipitation of the proteins. Spliceosomal core snRNPs, however, are stable under high-salt conditions in the presence of 15 mM MgCl₂. Thus, isopycnic ultracentrifugation can be applied to separate snRNPs from free proteins that stay at the top of the gradient and from free RNAs that are pelleted. The sedimentation behavior of individual snRNPs yields additional information on their protein–RNA ratio, which determines their buoyant density ρ (see Notes and Troubleshooting under

Section 47.3). As a typical example, we describe here the fractionation of nuclear extract by cesium chloride density gradient ultracentrifugation.

Equipment

- Beckman Optima TLX benchtop ultracentrifuge
- Beckman TLA 120.2 rotor with thick-walled polycarbonate centrifugation tubes (11 × 34 mm, Beckman # 343778)
- Sterile, RNase-free 1.5 ml Eppendorf tubes
- Cooling microcentrifuge

Reagents

- DMPC–H₂O (see “Reagents” in Section 47.2)
- 1 M MgCl₂ in DMPC–H₂O
- 3 M sodium acetate, pH 5.2, in DMPC–H₂O
- 20 mg ml^{−1} of glycogen in DMPC–H₂O
- 1 M DTT in DMPC–H₂O
- 0.1 M PMSF in ethanol
- Phenol–chloroform–isoamylalcohol (25 : 24 : 1) saturated with TE buffer (10 mM Tris/HCl, pH 8.0, 1 mM EDTA)
- 10% SDS
- 70% and absolute ethanol
- Buffer D/Mg: Buffer D (see “Reagents” in Section 47.2) containing 15 mM MgCl₂
- CsCl stock solution (dissolve CsCl in buffer D/Mg to a final density of 1.55 g ml^{−1}; the easiest way to achieve this is to mix two buffer D/Mg solutions: one containing no CsCl, the other containing approximately 1.8 g ml^{−1} CsCl; adjust the CsCl density of 1.55 g ml^{−1} precisely, since this is critical for the reproducibility of the gradients. The amount x of CsCl (expressed in g) needed to prepare 1 ml of a solution with the density ρ can also be calculated using the following equation [16]:

$$x = (\rho - 1)/0.92$$

where ρ indicates the numerical value of the density, expressed in grams per milliliter. Note that it is not recommended that PMSF and DTT be added to buffer D/Mg at this stage (see “Notes and Troubleshooting” under Section 47.3); store the CsCl solution at RT).

Method

Preparation of the Gradient and Ultracentrifugation

- 1) Thaw the nuclear extract carefully on ice, clear it of precipitates by a short spin (1 min at 20 000 g and 4 °C), and add MgCl₂ solution to a final concentration of 15 mM.
- 2) Take an aliquot of the extract as input control.

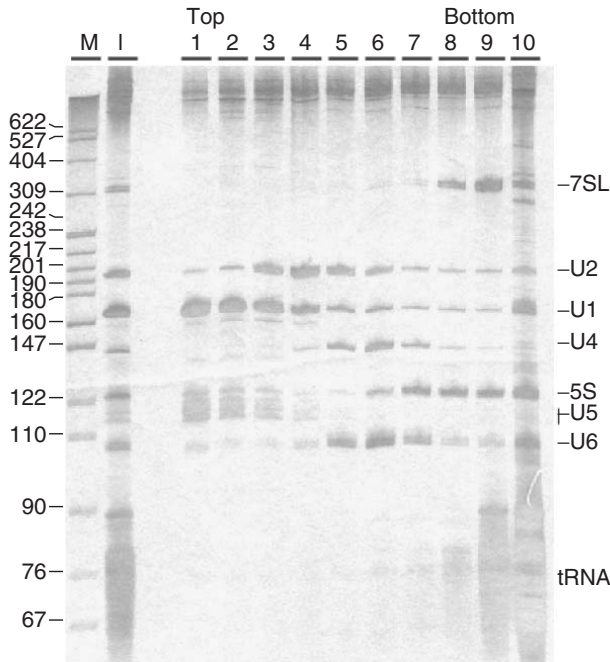


Figure 47.3 Fractionation of nuclear extract by CsCl density gradient ultracentrifugation: RNA analysis. After ultracentrifugation of 200 μ l of HeLa nuclear extract in a 1 ml CsCl density gradient (see Section “Preparation of the Gradient and Ultracentrifugation” for conditions), 10 fractions of 100 μ l each were taken from the top to the bottom of the gradient. RNA was prepared from each fraction (#1–10, #10 including the pellet) and

analyzed by denaturing PAGE, followed by silver staining. Marker sizes (pBR322 DNA digested with *Hpa*II) are shown on the left, and RNAs are indicated on the right. For example, the U4/U6 snRNP with base-paired U4 and U6 snRNAs peaks in fractions #5–7, free tRNAs are found in the pellet fraction (#10). For comparison, lane I contains RNA prepared from 10 μ l of HeLa nuclear extract, corresponding to 10% of the input material.

- 3) Freshly adjust 3 ml of the CsCl stock solution to 0.5 mM DTT and 0.5 mM PMSF.
- 4) Mix 200 μ l of the extract with 300 μ l of the prepared CsCl solution.
- 5) Pipette 500 μ l CsCl stock solution supplemented with PMSF and DTT into a precooled 1 ml polycarbonate tube (11 $\times$ 34 mm, Beckman, Part # 343778), and overlay it carefully with the prepared extract-CsCl-mixture.
- 6) Balance the tubes carefully with CsCl stock solution.
- 7) Centrifuge at 90 000 rpm (corresponding to $\sim$ 350 000 g) for 20 h at 4 $^{\circ}$ C in a Beckman TLX tabletop ultracentrifuge, using a precooled TLA 120.2 rotor.
- 8) Carefully fractionate the gradient from top to bottom by withdrawing nine 100 μ l fractions; resuspend the pellet in the residual gradient solution, representing the 10th fraction (Figure 47.3).

Preparation of RNA from the Gradient Fractions

- 1) Add 300 μl of DMPC–H₂O and 40 μl of 10% SDS to each fraction, then add 400 μl phenol–chloroform–isoamylalcohol, vortex thoroughly, and separate phases by centrifugation (10 min, 4 °C, 20 000 g).
- 2) Transfer the upper aqueous phase to a new Eppendorf tube and add 40 μl of 3 M sodium acetate, pH 5.2, 2 μl glycogen (20 mg ml^{−1}), and 1 ml ethanol.
- 3) Mix solution by inversion of the tube and incubate for 10 min at −70 °C.
- 4) Collect precipitated nucleic acids by centrifugation for 20 min at 20 000 g and 4 °C, discard the supernatant, and wash the pellet with 500 μl of 70% ethanol (at RT).
- 5) Analyze the RNA by denaturing PAGE.

Control Gradient for Density Calculation

- 1) Prepare a gradient as described, replacing the nuclear extract with buffer D.
- 2) Run the gradient and fractionate as described above.
- 3) Precisely weigh an aliquot of each fraction (e.g., 50 μl), which will give the density distribution across the gradient.

Notes and Troubleshooting

- 1) By measuring the buoyant density of a particle, the percentage of protein mass in the complex can be roughly calculated using the following empirical formula [16, 17]:

$$\% \text{Protein} = (1.85 - \rho) / 0.006$$

where ρ indicates the numerical value of the density, expressed in grams per milliliter. For example, densities of 1.36 and 1.51 g ml^{−1}, respectively, were observed for the *Trypanosoma brucei* U5 and U4/U6 core snRNPs; these values correspond to protein ratios of 82% (U5 core snRNP) and 57% (U4/U6 core snRNP). Taken the known masses of the RNA components into account, this results in total protein masses of 93 kDa per U5 core complex and 89 kDa per U4/U6 core complex [18].

- 2) Instead of the standard CsCl, Cs₂SO₄ has also been used in isopycnic density gradient centrifugation [4, 19, 20]; note that this appears to result in different stringencies for the RNPs (as discussed for the trypanosomal U2 snRNP [21]).
- 3) For the preparation of the CsCl stock solution, it is not recommended that PMSF and DTT be added to buffer D/Mg, since they are degraded during storage. Instead, PMSF and DTT should be added to the CsCl stock solution immediately before gradient preparation (see Section “Preparation of the Gradient and Ultracentrifugation”).
- 4) Owing to the high ionic strength of the fractions prepared from the gradient, it is recommended that the fractions be diluted or dialyzed (e.g., against buffer D) before subjecting them to immunoaffinity purification or other methods [22]. The high ionic strength also interferes with degradation of the RNAs; nevertheless, it is recommended that the obtained samples be handled on

ice, especially after removal of the salts by precipitation and washing. During phenol–chloroform extraction, make sure that only the upper, aqueous phase is transferred to a new Eppendorf tube without touching the interface or the phenol phase to avoid contaminations that may perturb downstream RNA analyses.

- 5) After ethanol precipitation, air-dry the pellet (3–5 min at RT) to remove residual ethanol. Avoid overdrying, since this may make it difficult to redissolve the RNA pellet in DMPC–H₂O.

References

1. Will, C.L. and Lührmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.*, **13**, 290–301.
2. Brunel, C., Widada, J.S., Lelay, M.N., Jeanteur, P., and Liautard, J.P. (1981) Purification and characterization of a simple ribonucleoprotein particle containing small nucleoplasmic RNAs (snRNP) as a subset of RNP containing heterogeneous nuclear RNA (hnRNP) from HeLa cells. *Nucleic Acids Res.*, **9**, 815–830.
3. Brunel, C. and Cathala, G. (1990) Purification and characterization of U small nuclear ribonucleoproteins in cesium chloride gradients. *Methods Enzymol.*, **181**, 264–273.
4. Pederson, T. (1974) Proteins associated with heterogeneous nuclear RNA in eukaryotic cells. *J. Mol. Biol.*, **83**, 163–183.
5. Heubeck, C. and Schön, A. (2001) Cyanelle ribonuclease P: isolation and structure-function studies of an organellar ribonucleoprotein enzyme. *Methods Enzymol.*, **342**, 118–134.
6. Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.
7. Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Lührmann, R. (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, **298**, 2205–2208.
8. Yoshimoto, R., Kataoka, N., Okawa, K., and Ohno, M. (2009) Isolation and characterization of post-splicing lariat-intron complexes. *Nucleic Acids Res.*, **37**, 891–902.
9. Kosowski, T.R., Keys, H.R., Quan, T.K., and Ruby, S.W. (2009) DExD/H-box Prp5 protein is in the spliceosome during most of the splicing cycle. *RNA*, **15**, 1345–1362.
10. Schneider, M., Will, C.L., Anokhina, M., Tazi, J., Urlaub, H., and Lührmann, R. (2010) Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol. Cell.*, **38**, 223–235.
11. Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Lührmann, R. (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA*, **16**, 2384–2403.
12. Bell, M., Schreiner, S., Damianov, A., Reddy, R., and Bindereif, A. (2002) p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J.*, **21**, 2724–2735.
13. Guo, X., Ma, J., Sun, J., and Gao, G. (2007) The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 151–156.
14. Wu, W.W., Weaver, L.L., and Panté, N. (2009) Purification and visualization of influenza A viral ribonucleoprotein complexes. *J. Vis. Exp.*, pii: 1105. doi: 10.3791/1105.

15. Ast, G., Goldblatt, D., Offen, D., Sperling, J., and Sperling, R. (1991) A novel splicing factor is an integral component of 200S large nuclear ribonucleoprotein (InRNP) particles. *EMBO J.*, **10**, 425–432.
16. Hamilton, M.G. (1971) Isodensity equilibrium centrifugation of ribosomal particles; the calculation of the protein content of ribosomes and other ribonucleo-proteins from buoyant density measurements. *Methods Enzymol.*, **20**, 512–521.
17. Spirin, A.S. (1969) The second Sir Hans Krebs lecture. Informosomes. *Eur. J. Biochem.*, **10**, 20–35.
18. Lücke, S., Klöckner, T., Palfi, Z., Boshart, M., and Bindereif, A. (1997) *Trans* mRNA splicing in trypanosomes: cloning and analysis of a PRP8-homologous gene from *Trypanosoma brucei* provides evidence for a U5-analogous RNP. *EMBO J.*, **16**, 4433–4440.
19. Michaeli, S., Roberts, T.G., Watkins, K.P., and Agabian, N. (1990) Isolation of distinct small ribonucleoprotein particles containing the spliced leader and U2 RNAs of *Trypanosoma brucei*. *J. Biol. Chem.*, **265**, 10582–10588.
20. Szybalski, W. (1968) Use of cesium sulfate for equilibrium density gradient centrifugation. *Methods Enzymol.*, **12**, 330–360.
21. Cross, M., Günzl, A., Palfi, Z., and Bindereif, A. (1991) Analysis of small nuclear ribonucleoproteins (RNPs) in *Trypanosoma brucei*: structural organization and protein components of the spliced leader RNP. *Mol. Cell. Biol.*, **11**, 5516–5526.
22. Palfi, Z., Xu, G.L., and Bindereif, A. (1994) Spliced leader-associated RNA of trypanosomes. Sequence conservation and association with protein components common to *trans*-spliceosomal ribonucleoproteins. *J. Biol. Chem.*, **269**, 30620–30625.

48

Identification and Characterization of RNA Binding Proteins through Three-Hybrid Analysis

Felicia Scott and David R. Engelke

48.1

Introduction

Macromolecular complexes between RNA and proteins play an integral role in many cellular processes. RNA–protein interactions have been studied using a variety of biochemical assays such as filter binding, electrophoretic mobility shift assays (EMSAs), and RNA footprinting. A disadvantage of these *in vitro* techniques is that they do not allow direct identification of the cognate genes encoding the proteins or RNAs of interest; moreover, studies of interactions formed *in vitro* are limited by how well the *in vitro* incubation recapitulates the cellular processes that influence the interaction *in vivo*. In addition, application of the *in vitro* biochemical techniques is often hindered by the relatively low abundance of many RNA binding proteins that make them difficult to detect, thereby making exogenous protein expression and purification mandatory for many of the *in vitro* biochemical analyses. In some instances, where an RNA and protein are known to interact *in vivo*, the RNA–protein complex cannot be detected by *in vitro* techniques due to its instability or folding issues *in vitro*. Many of these disadvantages have been overcome by the development of several methods using molecular genetics to study RNA–protein interactions.

We focus here on one such genetic method, the yeast three-hybrid system. The three-hybrid system is a modification of the widely employed genetic screen for detecting protein–protein interactions, the yeast two-hybrid system [1]. The three-hybrid system includes a chimeric RNA that fuses a “bait” RNA with an anchoring RNA. A specific RNA–protein interaction results in transcription of two independent reporter genes. The presence or absence of an interaction can be monitored by cell growth, colony color, or enzymatic activity. One potential drawback of the three-hybrid technique is that it does not yield quantitative information (K_d , complex stoichiometry) that can be gained from many biochemical approaches. One major advantage of this technique is that the RNA–protein interaction can be analyzed *in vivo* independent of its biological function. The three-hybrid system also has the added benefit in that a clone encoding the protein of interest is obtained directly when a library of cDNAs is used to screen for

RNA binding activities. Published examples of successful three-hybrid screens include the discovery of proteins that bind a known RNA sequence, confirmation of a putative RNA–protein interaction, mutational analysis of interacting RNAs and proteins, and the discovery and analysis of multiprotein complexes (see the following review for a list of examples [2–11]). We have successfully used the two- and three-hybrid systems in the study of the biomolecular interactions among the 10 subunits (1 RNA and 9 protein subunits) of nuclear RNase P from *Saccharomyces cerevisiae* [6, 12].

Beyond providing a general overview of the three-hybrid system including the most commonly used plasmids, strains and protocols, we discuss the strengths, limitations and pitfalls of the system. Many of the strengths and limitations of the two-hybrid system are also common in the three-hybrid system. For a detailed discussion of these aspects relative to the two-hybrid system, see Refs. [13–17].

48.2

Basic Strategy of the Method

Two different three-hybrid systems were developed and published independently. Although the two systems are based on the same strategy, they vary in the details [18, 19]. The differences include the DNA binding domain fusion protein, the bridge-forming RNA binding protein, and its RNA target site. The basic strategy of the three-hybrid method is illustrated in Figure 48.1. The yeast three-hybrid system is based on the fact that many *trans*-acting transcriptional regulators are made up of physically separable, functionally independent domains. These regulators behave as bifunctional units, often containing a DNA binding domain that binds to a specific promoter sequence and an activation domain that directs the RNA polymerase II complex to transcribe a downstream gene. Both domains are required to activate a gene. In principle, any activation domain can be associated with any DNA binding domain since it is the DNA binding domain that provides specificity. In the three-hybrid system, the DNA binding sites are located upstream of reporter genes in the yeast chromosome. The first fusion protein consists of a DNA binding domain linked to a known RNA binding domain. This RNA binding domain interacts with its RNA target site in a hybrid RNA molecule. The other portion of the RNA molecule is the “bait” that interacts with a second fusion protein composed of another RNA binding domain fused to a transcriptional activation domain. Thus, the hybrid RNA functions as a bridge between two fusion proteins. When this tripartite complex forms at the promoter of the reporter gene, transcription of the reporter gene is activated. Reporter gene expression can be monitored by growth phenotype or enzymatic activity.

The components of the three-hybrid system that are most commonly used were developed by Sengupta *et al.* [19] in the laboratory of Marvin Wickens at the University of Wisconsin. The DNA binding site consists of a 17-nt recognition site (operator) for the *Escherichia coli* LexA protein. Multiple LexA operators are located in the promoter regions of the two reporter genes, *lacZ* (eight operators)

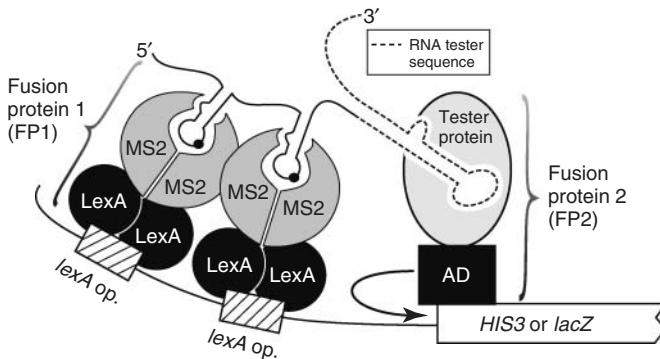


Figure 48.1 The basic strategy of the three-hybrid system and the components used most often. The first fusion protein (FP1) contains the full-length LexA protein fused to the MS2 coat binding protein. In the most commonly used strains, the gene encoding FP1 is integrated into the yeast genome. Hybrid RNAs contain two MS2 RNA sites (black dots indicate a point mutation in the tetraloop, 5'-AUUA to 5'-AUCA, which increases affinity to the MS2 protein) and the RNA tester sequence (dashed line). The MS2 coat protein and the LexA protein bind as dimers to their targets. Since binding of the MS2 protein to two adjacent stem-loop sites is cooperative, two MS2–LexA fusion dimers are shown to be bound simultaneously to the same hybrid RNA, with cooperativity indicated by the partial overlap of

MS2 domains between individual dimers. A single MS2–LexA fusion dimer bound to a single *lexA* operator (op.) and a single MS2 stem-loop would be the minimal requirement for mediating the bridging interaction. Since there are four and eight *lexA* operator copies upstream of the *HIS3* and *lacZ* promoters, respectively, multiple copies of the hybrid RNA may be recruited simultaneously to the promoter. The second fusion protein (FP2) contains the activation domain (AD) of Gal4 fused to the tester protein or cDNA library open reading frames. A productive interaction in the three-hybrid system leads to transcriptional activation of two reporter genes, *HIS3* and *lacZ*. The *Saccharomyces cerevisiae HIS3* gene encodes imidazoleglycerol-phosphate dehydratase involved in histidine biosynthesis. (Adapted from Zhang *et al.* [20])

and *HIS3* (four operators). The prokaryotic LexA protein binds as a dimer to the LexA operator. Fusion protein 1 (FP1) consists of LexA protein fused to the bacteriophage MS2 coat protein. In the most commonly used yeast host, the gene for the LexA–MS2 fusion has been integrated into the chromosome. The MS2 protein binds as a dimer to a stem-loop structure. The hybrid RNA contains two MS2 coat protein binding sites linked to the RNA tester sequence (RTS). Fusion protein 2 (FP2) consists of the transcriptional activation domain of the yeast Gal4 transcription factor fused to an RNA binding tester protein (TP).

Alternative versions of the three-hybrid system are also available. These include modifications where equivalents of the LexA–MS2 fusion protein and MS2 RNA binding sites are replaced with (i) the Rev responsive element (RRE) sequence in the *env* gene and the RRE-RNA binding protein RevM10 [18], (ii) the *nanos* regulatory element (NRE) in *hunchback* mRNA's 3' untranslated region and its protein binding partner, Pumilio [21], or (iii) hY5 RNA and its protein binding

partner Ro60 [22]. Similarly, the LexA DNA binding domain can be replaced with the Gal4 DNA binding domain [7, 18].

48.3

Detailed Components

The components of the three-hybrid system that are discussed in detail below can be obtained from the Wickens lab at the University of Wisconsin-Madison (<http://www.biochem.wisc.edu/faculty/>). To use the general procedure below, the researcher should be familiar with basic molecular biology techniques such as DNA ligations, *E. coli* transformations, and restriction enzyme analysis as well as basic yeast molecular biology and microbiological techniques. Suggested sources for additional information on these topics are *Current Protocols in Molecular Biology* [23], *Molecular Cloning: A Laboratory Manual* [24], and *Guide to Yeast Genetics and Molecular Biology* [25].

48.3.1

Yeast Reporter Strain

The genotype of the yeast reporter strain L40-coat is *MATa*, *ura3-52*, *leu2-3, 112*, *his3Δ200*, *trp1Δ1*, *ade2*, *LSY2::(lexA op)₄-HIS3*, *ura3::(lexA-op)₈-lacZ*, *LexA-MS2 coat (TRP1)*. The gene encoding the LexA-MS2 coat protein fusion has been integrated into the chromosome. The strain R40-coat is identical to the L40-coat except that it is of the opposite mating type (*MATα*). A canavanine-resistant derivative of the L40-coat, L40-coat-can, is also available. This strain carries a *can1* allele and becomes canavanine-sensitive when transformed with pACTII/CAN (see below). The strains are auxotrophic for uracil, histidine, adenine, and leucine. The strains are available from M. Wickens, University of Wisconsin. Two plasmids, one encoding the activation domain fusion protein (FP2) and the other encoding the hybrid RNA, need to be transformed into this strain for detection of the RNA-protein interaction.

48.3.2

Plasmids

Two interaction-specific constructs must be made by the user of the three-hybrid system, an RNA hybrid (containing the tester RNA sequence) and the activation domain FP2 (containing the protein of interest or a cDNA library fused to the Gal4 activation domain). The details of their construction along with technical considerations are discussed below. For maps and more detailed descriptions of the plasmids see <http://www.biochem.wisc.edu/faculty/wickens/lab/3h.aspx>.

48.3.3

Hybrid RNA

The four most common plasmids for producing hybrid RNAs are pIII/MS2-1, pIII/MS2-2, pIIIA/MS2-1, and pIIIA/MS2-2. Each is a multicopy plasmid that can be propagated in either yeast or bacteria. pIII/MS2-1 and pIII/MS2-2 are derived from pIIIEx426RPR [26] and, as the pIIIA plasmids, use the *S. cerevisiae* promoter of the RNase P RNA gene (*RPR1*) for RNA polymerase III (Pol III) transcription. Rpr1 RNA is normally retained in the nucleus [27, 28], and it is likely that this also holds true for most hybrid RNAs made from this promoter. The RTS is inserted at the *Sma* I or *Sph* I restriction sites in the pIIIA plasmids, but only at the *Sma* I site in the pIII plasmids where the *Sph* I site is not unique. The designation 2-1 or 2-2 indicates that the two plasmids differ only in the relative position of the restriction site(s) and the MS2 binding sites (Figure 48.2). Both RNA plasmid series carry the *URA3* selectable marker. pIIIA/MS2-1 and pIIIA/MS2-2 are similar to pIII/MS2-1 and pIII/MS2-2, but they additionally carry the *ADE2* gene. Screening for retention of the *ADE2* gene can be used to help eliminate false positives that are RNA-independent (see discussion below on false positives).

The hybrid RNA contains the following features: (i) the yeast 84-nt RNase P RNA leader sequence, (ii) two tandem MS2 coat protein binding sites, (iii) the RTS, and (iv) 41 nt of the 3' terminus of RNase P RNA. Two MS2 coat protein binding sites are used since binding of the protein to adjacent sites is cooperative [29, 30]. Each MS2 33-nt recognition site also contains a nucleotide exchange (the tetraloop is changed from 5'-AUUA to 5'-AUCA) that enhances binding of the coat protein to the MS2 stem-loop (K_d of 10^{-9} to 10^{-10} M) [31, 32]. The hybrid RNAs are transcribed from the Pol III promoter and terminate at the TTTT Pol III terminator. The expression of small RNA molecules from the *RPR1* promoter can produce several thousand copies of an RNA molecule per cell [26]. In addition, transcripts from the RNase P RNA promoter remain in the nucleus, which is where the three-hybrid interaction needs to take place.

48.3.3.1 Technical Considerations for the Hybrid RNA

The three-hybrid system is a sensitive method for detecting relatively weak and transient RNA–protein interactions. Such interactions may not be biochemically detectable, but may be critical for proper functioning of biological systems. The K_d values of RNA–protein interactions detected with the three-hybrid system have been in the range of 10^{-11} to 10^{-6} M. The minimal affinity that is required to yield a detectable transcriptional activation has not been determined. However, affinity is not the only determinant of whether an RNA–protein interaction is detectable. The abundance, the conformation, and the cellular location of the hybrid RNA and FP2 can also influence transcriptional activation.

There are several technical constraints on the production of hybrid RNAs. The use of the *RPR1* promoter limits the sequence of the RTS that can be analyzed. In general, the RTS should not contain more than four uridines (preferably not more than three) in succession since Pol III transcription can be terminated at

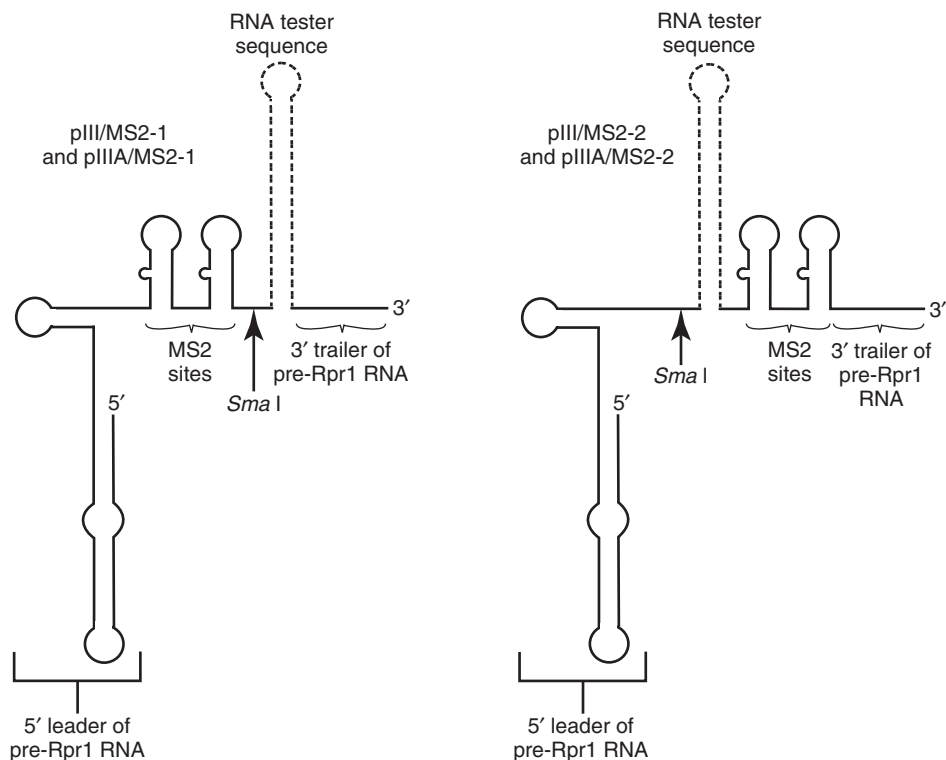


Figure 48.2 Design of hybrid RNAs. The orientation of the RNA tester sequence relative to the two MS2 RNA sites depends on the plasmid used for generating the hybrid RNA. The RNA tester sequence is inserted into the *Sma* I or *Sph* I restriction site in the pIII A plasmids; only the *Sma* I site is utilized

with the pIII plasmids because the *Sph* I site is not unique there. The designation 2-1 or 2-2 indicates that the two plasmids differ only in the relative position of the restriction site(s) and the MS2 binding sites; pre-Rpr1 RNA, precursor to Rpr1 RNA. (Adapted from Zhang *et al.* [20])

these sites. To circumvent this sequence limitation, an alternative system using an RNA polymerase II promoter has been developed [18] and used successfully in the analysis of hnRNP–C1/C2 interactions with synthetic and naturally occurring uridine-rich sequences [7].

The size of the RTS is also an important determinant of success when using the three-hybrid system. Typically, RNA inserts shorter than 200 nt yield higher signals than longer inserts. However, positive interactions have been detected for RNA tester sequences up to 1600 nt in length [33].

The MS2 and RNA tester sequences can be placed in either position relative to one another; however, the relative position of the RTS and the MS2 sites can make a difference in terms of the transcriptional activation. For example, in the case of the three-hybrid test using the iron-responsive element (IRE) and the iron-regulatory protein 1 (IRP), placing the IRE downstream of the MS2 sites resulted in two- to threefold less transcription than did the opposite arrangement [19]. Although

successful screens have been carried out with both arrangements of the MS2 sites, we have had the most success and experience with pIIIA/MS2-2 [6, 12]. The following aspects may be relevant to the relative placement of the MS2 sites: (i) four or more consecutive U residues may terminate Pol III transcription, although the efficiency of termination is context-dependent and could thus differ for MS2-1 and MS2-2 constructs; elimination of homouridine stretches by mutation may be considered if expression of the hybrid RNA is compromised by premature termination with both construct designs; (ii) the RTS may fold differently in the MS2-1 *versus* MS2-2 context, which might also be related to how the tester RNA is embedded in its native transcripts.

The MS2 coat protein binding site forms a stable stem-loop structure. The side-by-side placement of two of these structures should limit the formation of alternative structures owing to the presence of the RTS. However, in some instances, alternative structures can form that interfere with the interaction of the hybrid RNA with the LexA–MS2 fusion protein (FP1) or the activation domain fusion protein (FP2). Cassidy and Maher overcame this technical difficulty in their three-hybrid screen using NF- κ B and an RNA aptamer by inserting a 13-bp GC-rich clamp to arrest the correct conformation of their RNA aptamer [3, 34].

To aid in the optimization of a particular cloning strategy, it might be helpful to use RNA folding programs, such as *mfold* (www.bioinfo.rpi.edu/applications/mfold/) or *RNAfold* (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>), to predict the secondary structure of your RNA of interest fused to the MS2 recognition sites; however, these predictions often do not reflect the RNA's conformational equilibria *in vivo*. In any case, after clones with the desired sequence have been obtained, Northern blot analysis may be used to confirm the integrity and expression level of the hybrid RNA.

48.3.4

Activation Domain FP2

There are two fusion proteins in the three-hybrid system. FP1, the LexA–MS2 “bait” for the hybrid RNA, remains constant. FP2 varies on the basis of the interaction tested. The pACTII plasmid consists of the polypeptide of interest fused to the Gal4 activation domain to create FP2. pACTII, a multicopy plasmid, encodes the Gal4 activation domain followed by the hemagglutinin (HA) epitope tag {YPYDVPDYA} and a polylinker for cloning test polypeptide-coding sequences. The HA epitope tag permits immunodetection of the fusion protein by Western blotting using antibodies that are commercially available (Santa Cruz Biotechnology, Chemicon-Millipore, or Roche). Antibodies against the HA epitope tag can also be used for coimmunoprecipitation experiments intended to verify identified interactions. We have also used commercial antibodies against LexA and the Gal4 activation domain for immunodetection of the fusion proteins (Santa Cruz Biotechnology); however, many of these antibodies are not entirely robust and specific. The fusion protein is expressed from the constitutively expressed *ADH1* promoter and transcription is terminated by the *ADH1* terminator. The plasmid can be propagated in yeast (*LEU2* marker) and *E. coli* (ampicillin resistance).

Furthermore, the plasmid pACTII-CAN was constructed from plasmid pACTII [20] by insertion of the yeast *CAN1* gene into the unique *Sal* I site of pACTII. The *CAN1* gene encodes an arginine permease that causes cells to die in media containing the arginine analog canavanine. Thus, canavanine selection leads to the loss of the activation domain plasmid similar to the loss of the hybrid RNA plasmid in the presence of 5-fluoroorotic acid (5-FOA).

When designing a cloning strategy, remember that the gene of interest must be cloned in-frame with the sequence encoding the Gal4 activation domain and the HA epitope. Clones should be sequenced to confirm that the gene is cloned in the correct orientation and in-frame with the *GAL4* activation domain.

48.3.4.1 Technical Considerations for the Activation Domain of FP2

Sometimes RNA–protein interactions that occur normally *in vivo* are not detected by the three-hybrid system. This lack of detection might have several causes. First, if high expression levels of the activation domain fusion protein are toxic to the reporter strain, the transformants may not grow. To circumvent this problem, you may want to (i) reclon the protein gene of interest on a low-copy-number plasmid, (ii) put the gene of interest under an inducible promoter, and (iii) try using subfragments of the protein. There are a variety of plasmids (low-copy-number or inducible) available from the two-hybrid system that can easily be adapted to the three-hybrid system (see reviews [13–17]). It is important to make sure that the selectable marker is compatible with the three-hybrid system. As an alternative approach to alleviating the toxicity problem, it is possible to clone fragments of the gene of interest instead of the full-length gene into the pACTII plasmid and test the fragments for RNA binding activity.

Second, the expressed protein must be stable and the expression must be high enough to generate an interaction signal. Expression levels and stability of the proteins can be checked by Western blotting, using commercially available antibodies against the HA epitope tag. Nevertheless, it is important not to place too much emphasis on these tests. In some cases, detection of the epitope failed (the epitope may not be available because of the folding of the fusion protein or low expression may prevent detection), but the three-hybrid test was nonetheless successful. Also, successful detection of the expressed fusion protein does not imply that the interaction domain is accessible and properly folded; thus, the screen could fail in spite of the presence of the fusion protein. The best option is to utilize an antibody against your specific protein, if available.

A number of groups have successfully used the three-hybrid system to screen cDNA libraries to identify new interactions for their RNA of interest [35–37]. We have had success with a yeast cDNA library created in pACTII (a gift from Steve Elledge). Other cDNA libraries compatible with the three-hybrid system are available as well (for example, see Ref. [17]). One factor that needs to be taken into account when screening cDNA libraries is scaling up of the transformation protocol (see below) to cover the complete complexity of the library; this is particularly relevant for rare mRNAs. The degree of scale-up will depend on the library of choice.

The major technical challenge when using cDNA libraries is identifying those interactions that are biologically relevant. The strength of a positive signal is not always an indicator of relevance, since artificial interactions can be strong and specific, while biologically relevant RNA–protein interaction might be weak. It is essential to further verify true positives using an independent method such as *in vitro* binding assays, coimmunoprecipitation, colocalization, or genetic methods.

48.3.5

Positive Controls

The control plasmids for the three-hybrid system that are most commonly used are the pIIIA/IRE-MS2 and pAD-IRP plasmids. The plasmid pIIIA/IRE-MS2 expresses a hybrid RNA containing the rat ferritin light-chain IRE [38, 39] fused to MS2 RNA. The IRE RNA forms a hairpin loop structure that serves as the recognition site for the IRP [39]. Plasmid pAD-IRP expresses the rabbit IRP fused to the Gal4 activation domain. When pIIIA/IRE-MS2 and pAD-IRP are cotransformed into the L40-coat strain, robust transcriptional activation of *HIS3* and *lacZ* occurs. If you test for growth on histidine-minus media to assess *HIS3* activation, 3-aminotriazole (3-AT) should be included to eliminate low levels of basal expression (see discussion below on 3-AT). For the pIIIA/IRE-MS2 and pAD-IRP plasmids, transformants are usually plated directly on media containing 5 mM 3-AT [19].

Protocols

An overview of the experimental protocol is outlined in Figure 48.3. The protocol for testing a known or putative RNA–protein interaction is relatively straightforward. The RNA of interest and the gene encoding the putative RNA binding protein are cloned into the appropriate plasmids. The plasmid containing the hybrid RNA and the plasmid containing the FP2 are transformed into the yeast strain. Once transformants are obtained, the cells are assayed for transcriptional activity of the reporter genes: *HIS3* and/or *lacZ*.

We used this approach to identify proteins that bound directly to the RNA subunit of the nuclear RNase P from *S. cerevisiae* [6]. Similarly, an RNA–protein interaction between mouse telomerase RNA and a newly cloned telomerase protein (TP1) was confirmed by this approach [40].

The three-hybrid method can also be used to delineate the portions of the RNA or proteins that are required for binding. In our case, we were able to use this approach to demonstrate that Pop1p, the largest protein subunit of yeast nuclear RNase P, bound specifically and directly to the P3 subdomain of RNase P RNA, a 54-nt helix–internal loop–helix structure [12]. Specificity is best controlled by comparing RNA binding for the wild-type RNA *versus* variants containing point mutations that compromise *in vivo* function without necessarily affecting overall structure. Lee *et al.* [8] used this approach along with site-directed mutagenesis to identify the amino acids in the Rous sarcoma virus nucleocapsid that bind to the RNA-packaging signal.

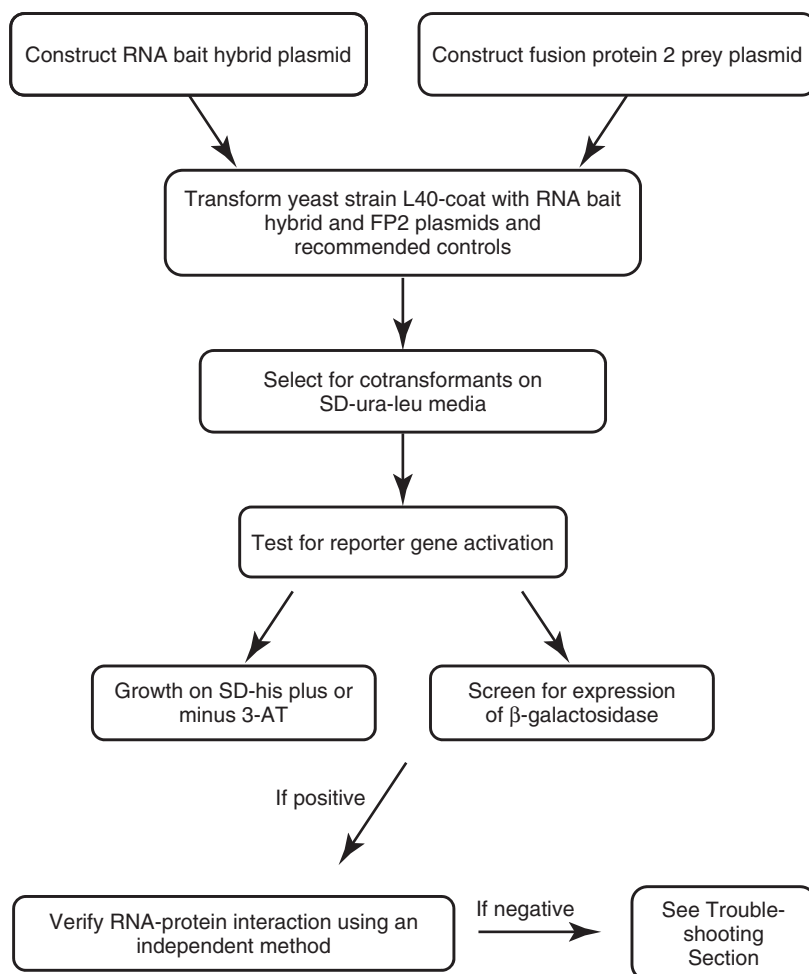


Figure 48.3 Experimental outline of the three-hybrid protocol.

Transformation of Yeast

Plasmids can be introduced into yeast by transformation of spheroplasts, transformation of chemically treated cells, or by electroporation. We use a modification of the lithium acetate method [23] for the transformation of the L40-coat reporter strain. Yeast cells are grown in rich media, treated with lithium acetate, and incubated in the presence of plasmid DNA. After transformation the cells are plated on minimal media that lack the appropriate nutrients for plasmid selection.

The steps are as follows:

- 1) Inoculate 5 ml of yeast extract/peptone/dextrose (YPD) medium (1% Bacto-yeast extract, 2% Bacto-peptone, 2% dextrose, pH 6.5) with a single yeast colony of the L40-coat strain. Grow overnight to saturation at 30 °C.

- 2) The night before transformation, inoculate a 2-liter sterile flask containing 300 ml yeast extract/peptone/adenine/dextrose (YPAD) medium (YPD + 0.003% adenine hemisulfate) with an appropriate amount of the saturated culture (25–100 μ l) and grow overnight at 30 °C to an OD₆₀₀ of 0.3–0.5 (1×10^7 cells ml⁻¹). For two to threefold higher transformation efficiency, dilute at this point to 2×10^6 cells ml⁻¹ in fresh YPAD medium and grow for 2–4 h. It is important to perform this step when high-efficiency transformation is needed for cDNA and genomic libraries.
- 3) Pellet the cells by centrifugation at 5000 rpm (Sorvall GSA rotor) for 5 min at room temperature.
- 4) Resuspend the pellet in 10 ml of sterile water and transfer to a 50 ml conical tube.
- 5) Repeat step 3.
- 6) Pour off the water and resuspend the cells in 1.5 ml freshly prepared buffered lithium acetate solution (10 mM Tris-Cl, 1 mM EDTA, 100 mM lithium acetate made from stocks of 100 mM Tris-Cl (pH 7.5), 10 mM EDTA, and 1 M lithium acetate (adjusted to pH 7.5 with dilute acetic acid). Incubate the cells for 15 min at 30 °C without agitation before adding the transforming DNA.
- 7) For each transformation, mix 50 μ g sheared, heat-denatured salmon sperm DNA (carrier DNA) with 1–5 μ g of each plasmid DNA in a sterile 1.5 ml microcentrifuge tube. Keep the total volume of DNA at or below 20 μ l. Maximum transformation efficiency is achieved by boiling and chilling the carrier DNA (not the plasmid) immediately prior to use. Heat for 5–10 min followed by cooling on ice.
- 8) Add 200 μ l of yeast suspension from step 6 to each microcentrifuge tube.
- 9) Add 1.2 ml polyethylene glycol (PEG, MW 3350) solution (40% PEG, 10 mM Tris-HCl, 1 mM EDTA, 100 mM lithium acetate made fresh from stocks of 50% PEG 3350, 100 mM Tris-HCl (pH 7.5), 100 mM EDTA, and 1 M lithium acetate (adjusted to pH 7.5 with dilute acetic acid) to each tube. Mix gently by inverting or vortexing.
- 10) Incubate at 30 °C for 30 min without agitation.
- 11) Add 150 μ l dimethyl sulfoxide (DMSO). Heat-shock for 20 min at 42 °C.
- 12) Pellet the cells by centrifugation for 5 s at high speed in a microcentrifuge. Remove supernatant and discard.
- 13) Resuspend the cell pellet in 200–1000 μ l of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). Plate 100 μ l of the cell suspension on SD-ura-leu plates, which stands for synthetic dropout (SD) medium lacking uracil and leucine (0.67% Bacto-yeast nitrogen base without amino acids, 2% dextrose, 2% Bacto-agar, 0.002% adenine sulfate, 0.002% L-tryptophan, 0.002% L-histidine HCl, 0.002% L-arginine HCl, 0.002% L-methionine, 0.003% L-Tyrosine, 0.003% L-isoleucine, 0.003% L-lysine HCl, 0.005% L-phenylalanine, 0.01% L-glutamic acid, 0.01% L-aspartic acid, 0.015% L-valine, 0.02% L-threonine, 0.04% L-serine), selecting for the hybrid RNA plasmid and the Gal4 activation domain fusion protein 2 plasmid.

- 14) Incubate at 30 °C for three to five days until colonies of transformants are visible.
- 15) Restreak colonies to obtain single-colony isolates on SD-ura-leu plates.

Transformation of the reporter strain with a single plasmid yields 10^4 – 10^6 transformants per microgram of plasmid DNA. If both the hybrid RNA plasmid and the pACTII fusion protein plasmid are introduced at the same time, the transformation efficiency can decrease by a factor of 10. Instead of introducing both plasmids at the same time, you may want to transform the RNA hybrid plasmid first, especially when using cDNA libraries. This strain can be selected on SD-ura media. Generating this strain first may be wise for several reasons. It allows you to transform the same clone not only with the plasmid for the fusion protein of interest, but also with control plasmids encoding the Gal4 activation domain only (no insert) or a fusion protein that is unrelated to your RNA of interest. Second, if no transformants are obtained in the case of the double transformation, you do not know whether it is due to the toxicity of your protein or to a technical failure of the transformation procedure.

Assaying for *HIS3* Expression

Once transformants containing the RNA bait and protein prey (FP2) plasmids have been confirmed, the tests for reporter gene activation are performed. The first test we routinely performed is the ability to grow on media lacking histidine plus or minus 3-amino-1,2,4-triazole (3-AT; Sigma–Aldrich). 3-AT is a competitive inhibitor of the *HIS3* gene product. Cells producing more His3p can survive higher concentrations of 3-AT. We typically test transformants on SD-his-leu-ura plates containing increasing concentrations of 3-AT (0, 1, 5, 10, and 20 mM) to select for stronger interactions and to eliminate weak activators. However, a less stringent selection is often preferable when the strength of the interaction is unknown. Higher concentrations of 3-AT have also been used successfully. For example, the stem-loop binding protein binds histone mRNA at 25 mM 3-AT [37].

Typically, single colonies that harbor the relevant plasmids (grown on SD-leu-ura plates) are streaked on SD-his-leu-ura plates and grown for two to three days. Single colonies from the SD-his-leu-ura plates are streaked again on SD-his-leu-ura plates minus or plus increasing concentrations of 3-AT. Cells are allowed to grow for three to five days at 30 °C. The ability to grow in the presence of 3-AT is judged by growth of individual colonies throughout the streak, rather than by the smear of yeast cells in the initial streaking area on the same plate. It is not unusual to see a few large colonies on 3-AT plates, which are likely revertants with mutations in the *HIS3* gene.

Assaying for β -Galactosidase Activity

β -Galactosidase activity can be assayed by measuring the conversion of a lactose analog to a chromogenic or luminescent product. The assay can be performed using cell lysates or colonies permeabilized on a filter. The filter assay yields qualitative results, whereas the liquid assay is more quantitative. Our laboratory routinely uses the filter assay and the protocol is presented below. The level of β -galactosidase can

be quantified in yeast using one of two different colorimetric assays. For a detailed protocol of these assays, see Zhang *et al.* [20] or Bartel and Fields [13].

Filter Assay

- 1) Place a fresh SD-leu-ura plate over a “replica patching grid”. Touch the top of a single colony grown on an SD-leu-ura plate with a sterile toothpick and draw a small x on the fresh SD-leu-ura replica plate. Do the same (“picking and patching”) with additional single colonies, using each toothpick only once. Grow the replica plate at 30 °C for two to three days.
- 2) Lay a dry nitrocellulose filter paper (Schleicher & Schuell Optitran BA-S 85) on top of the cells grown on the SD-leu-ura plate. Apply gentle pressure to the nitrocellulose filter to transfer cells to the filter.
- 3) Lift the filter from the plate and place the filter colony side up on an aluminum foil boat. Immerse the filter and boat in liquid nitrogen for 10–20 s.
- 4) Thaw the filter, colony side up, at room temperature for a few minutes to lyse the cells. Alternatively, instead of freezing in liquid nitrogen, place the filter in a plastic petri dish, cover it, and place the dish for 10 min in a –80 °C freezer, followed by thawing as above.
- 5) In the lid of a petri dish, place 1.5 ml of assay buffer (60 mM Na₂HPO₄, 40 mM NaH₂PO₄, 10 mM KCl, 1 mM MgSO₄, 50 mM 2-mercaptoethanol, pH 7.0) plus 30 µl of X-gal (50 mg ml^{–1} in *N,N*-dimethylformamide) onto a Whatman number #1 filter circle (diameter 90 mm).
- 6) Place the nitrocellulose filter containing the lysed cells (step 4) colony side up onto the Whatman filter circle soaked with assay buffer. Place the bottom of the petri dish over the nitrocellulose filter and Whatman filter circle. Seal the chamber with parafilm and incubate at 30 °C.
- 7) Monitor the color reaction for 30 min to overnight.
- 8) Stop the reaction by floating the nitrocellulose filter on 1–2 ml 100 mM EDTA on a Saran Wrap.
- 9) Dry nitrocellulose filter at room temperature in a new petri dish.

We normally include the control strain that contains the plasmids pIIIA/IRE and pAD-IRP in every experiment. An interaction as strong as in the IRE/IRP system usually results in the appearance of blue color within 30–60 min. With extended incubation times, weak interactions will also yield a blue color reaction.

48.4 Troubleshooting

Many of the problems that lead to an unsuccessful three-hybrid screen have been discussed under technical considerations for the hybrid RNA (Section 48.3.3.1) or activation domain fusion protein (Section 48.3.4.1). However, there are some problems that can be overcome with the proper controls and these will be discussed here.

In some instances, even though a protein and RNA are known to interact based on other biochemical or genetic data, a direct test of the RNA–protein interaction in the three-hybrid system may yield a negative result. The negative result might be due to improper folding of the protein or RNA and/or lack of expression of the protein or RNA.

There are several types of interactions that will yield a “false” positive result in the three-hybrid system. Most of these false positives are indicative of an RNA–protein interaction, although the interaction might not be biologically relevant. For example, some proteins can exhibit high affinity for binding RNAs in the three-hybrid context, yet may neither have sequence specificity for that particular RNA sequence nor mimic a physiologically relevant interaction. In addition, indirect bridging interactions can be registered as positive in the three-hybrid system. In such a case, the protein appears to interact with the RNA, but may, in fact, interact with a yeast cellular protein that binds directly to the RNA. This scenario tends to become more of a problem when the activation domain FP2 is a yeast protein. For example, yeast She3p gave a positive result when tested with a portion of the *ASH1* mRNA’s 3′ UTR in the three-hybrid system, but this positive result was due to its interaction with She2p that bound the hybrid RNA. If the Shep2p was removed from the cell genetically by gene deletion, the yeast She3p no longer gave a positive result in the three-hybrid assay [41].

Occasionally, RNA hybrid molecules will transactivate the reporter gene in the absence of a fusion protein. Furthermore, in cDNA library screening, the large majority of false positives can be RNA-independent. This is most likely due to direct protein–protein interactions with the LexA–MS2 coat protein. To circumvent this problem, a colony color assay was developed that helps eliminate this class of false positives [42]. Since the L40-coat strain is an *ade2* mutant, yeast colonies accumulate a red pigment when grown in the absence of adenine. After a few days of growth, the L40-coat cells turn pink to red because of this metabolic block in the adenine biosynthetic pathway. To take advantage of this colony color assay, the tester RNA should be cloned into the pIIIA/MS2-1 or pIIIA/MS2-2 RNA hybrid vectors. Because the pIIIA plasmids carry the wild-type *ADE2* gene, the yeast cells containing these plasmids remain white. The colony color difference enables one to distinguish between RNA-independent and RNA-dependent positives. After transforming the L40-coat strain with both the RNA plasmid and the cDNA library, plate the transformants on SD-his-leu plates selecting for the cDNA plasmid (*LEU2*) and for *HIS3* reporter expression but not for the RNA plasmid. If activation of the reporter gene is independent of the RNA, a small percentage of the cells will lose the RNA plasmid since it has not been selected for, and the colony will become pink/red or sectorized (white colonies with pink/red sectors). If the His⁺ phenotype is RNA-dependent, all colonies will remain white.

Moreover, the yeast strain YBZ1 was generated to improve screening of cDNA/AD libraries. The strain encodes a tandem (head-to-tail) dimer of a high-affinity variant of the MS2 coat protein, fused to a monomer of LexA. It was reported that this strain increases the number of genuine, sequence-specific positives in cDNA library screens, while reducing the background of RNA-independent false positives [43].

The RNA-dependent or RNA-independent interactions can also be confirmed by subsequent selections against the RNA-producing plasmid using 5-FOA [42]. Owing to the conversion of 5-FOA to the toxic 5-fluoro-deoxyuracil by the *URA3* gene product, yeast cells expressing the *URA3* gene are severely inhibited in growth on media containing 5-FOA. In contrast, cells lacking the *URA3* gene product can grow in the presence of 5-FOA if uracil is provided in the growth media. Since the plasmids expressing the hybrid RNA carry the *URA3* marker, cells that have lost the plasmid can be selected by plating the primary *HIS*⁺ transformants on media containing 0.1% 5-FOA. The 5-FOA screen can be used with either the pIII or pIIIA plasmids. In the case of an RNA-dependent interaction, cells lacking the RNA plasmid will no longer be able to activate the reporter.

Protein-independent false positives have also been observed in the three-hybrid system [20]. These include hybrid RNAs that activate transcription when bound to a promoter. The frequency of these activating RNAs in a genomic library can be quite high. The methods used to identify and eliminate these protein-independent false positives depend on the type of plasmid used to express the activation domain FP2. If being a pACTII derivative, the plasmid is cured by growing a transformant overnight in YPD medium (to allow for loss of the pACTII plasmid) followed by plating on SD-ura to select for the RNA plasmid. These *URA*⁺ colonies are replica-plated on SD-leu plates to identify the colonies that lack the *LEU2* marker on the activation domain plasmid. If the activation domain plasmid is derived from pACTII/CAN, the plasmid can be cured by streaking colonies on an SD-arg plus canavanine plate. Following curing of the plasmid by either method, colonies should be assayed for β -galactosidase activity and/or growth on histidine-deficient media to reveal clones that activate the respective reporter gene promoters in the absence of the activation domain plasmid.

Once an RNA–protein interaction has been confirmed, the system can be modified in various ways to test for mutations in the RNA or protein that eliminate binding. RNA deletion studies can be designed to identify the minimal RNA sequence needed for an RNA–protein interaction.

48.5

Additional Applications

Several studies suggested that the three-hybrid method could potentially be used to identify target RNAs of specific RNA binding proteins. Sengupta *et al.* [44] adapted the three-hybrid system to identify RNA ligands for the yeast Snp1 protein, a homolog of the human U1-70K protein, using an RNA library constructed from short fragments of yeast genomic DNA that were transcribed in yeast together with the RNA binding sites for the MS2 coat protein. Venables *et al.* [45] used a similar approach to identify tester RNA sequences that bind DazIp, an RNA binding protein encoded by a region on the Y chromosome implicated in infertility. The potential use of this application can be appreciated in light of genome sequencing projects that have identified many putative RNA binding proteins on the basis of

the presence of known RNA binding motifs. Furthermore, the three-hybrid system was used to confirm that several brain mRNAs of unknown function that bind human fragile X mental retardation protein *in vitro* are able to do so *in vivo* [5].

The three-hybrid system can also be applied as an *in vivo* method to identify and optimize artificial RNA ligands for proteins that might be useful as therapeutic targets. Cassidy *et al.* [3, 34] utilized the three-hybrid test to identify and optimize an RNA aptamer that functions as a decoy for the transcription factor NF- κ B. In this case, a combination of *in vitro* selected RNA aptamers and *in vivo* genetic selections was crucial for obtaining RNA aptamers that could inhibit transcriptional activation by NF- κ B.

The three-hybrid system may further permit the rapid screening for therapeutic inhibitors of a known RNA–protein interaction, such as those involved in viral replication, transcription, or assembly. So far, no published reports of the use of the three-hybrid system to screen chemical inhibitors of an RNA–protein interaction are available, but several parameters would have to be satisfied in order for the assay to work. First, the yeast cell must be permeable to the chemical inhibitor. Second, the chemical inhibitor must not be toxic to the yeast cell. Third, the chemical inhibitor must be specific to the RNA–protein interaction tested.

The efficacy of RNA decoys or small peptides as inhibitors of specific RNA–protein interactions can be easily tested in the three-hybrid system by the introduction of an additional plasmid expressing the RNA decoy or peptide. The use of RNA decoys to disrupt known viral RNA–protein interactions such as Tat–TAR [46] and Rev–RRE [47] has been published.

Multicomponent complexes have also been studied using the three-hybrid system. A practical limitation in analyzing complexes containing many proteins is the number of plasmids and markers required. There are several types of multicomponent interactions that have been analyzed successfully in the three-hybrid system: (i) independent interactions in which each protein binds to its own site on the RNA and is unaffected by the presence of the other proteins, and (ii) bridged interactions, in which one protein tethers another protein to the RNA; in this case, both proteins are required for the interaction. The interaction of the bl4 group I intron with both of its splicing partners, a tRNA synthetase and an intron-encoded maturase, is an example of a multicomponent complex in which the RNA has two independent protein binding partners [33]. In the case of yeast RNase P, we have shown that Pop1p and Pop4p can also bind independently to the RNase P RNA [6]. An example of a bridged interaction that has been analyzed using the three-hybrid system is the interaction of She2p and She3p with the 3' UTR of *ASH1* mRNA [41].

48.6

Summary

The three-hybrid assay is as useful and versatile for the study of RNA–protein interactions as is the two-hybrid system for the study of protein–protein interactions. It will be interesting to see the further adaptations of the three-hybrid system beyond

its current applications. The perspective of using the three-hybrid system to identify RNA targets for putative RNA binding proteins predicted by genomic sequencing is intriguing. In addition, three-hybrid screens for therapeutic compounds that inhibit RNA–protein interactions involved in the viral life cycle might contribute to the development of novel antiviral agents.

Acknowledgments

This work was supported by National Institutes of Health grant GM34869 to D.R.E.

References

1. Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, **340**, 245–246.
2. Bernstein, D.S., Buter, N., Stumpf, C., and Wickens, M. (2002) Analyzing mRNA-protein complexes using a yeast three-hybrid system. *Methods*, **26**, 123–141.
3. Cassidy, L.A. and Maher, L.J. 3rd (2003) Yeast genetic selections to optimize RNA decoys for transcription factor NF-kappa B. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 3930–3935.
4. Cohen, A., Reiner, R., and Jarrous, N. (2003) Alterations in the intracellular level of a protein subunit of human RNase P affect processing of tRNA precursors. *Nucleic Acids Res.*, **31**, 4836–4846.
5. Dolzhanskaya, N., Sung, Y.J., Conti, J., Currie, J.R., and Denman, R.B. (2003) The fragile X mental retardation protein interacts with U-rich RNAs in a yeast three-hybrid system. *Biochem. Biophys. Res. Commun.*, **305**, 434–441.
6. Houser-Scott, F., Xiao, S., Millikin, C.E., Zengel, J.M., Lindahl, L., and Engelke, D.R. (2002) Interactions among the protein and RNA subunits of *Saccharomyces cerevisiae* nuclear RNase P. *Proc. Natl. Acad. U.S.A.*, **99**, 2684–2689.
7. Koloteva-Levine, N., Amichay, M., and Elroy-Stein, O. (2002) Interaction of hnRNP-C1/C2 proteins with RNA: analysis using the yeast three-hybrid system. *FEBS Lett.*, **523**, 73–78.
8. Lee, E.-G., Alidina, A., May, C., and Linial, M.L. (2003) Importance of basic residues in binding of rous sarcoma virus nucleocapsid to the RNA packaging signal. *J. Virol.*, **77**, 2010–2020.
9. Pudi, R., Abhiman, S., Srinivasan, N., and Das, S. (2003) Hepatitis C virus internal ribosome entry site-mediated translation is stimulated by specific interaction of independent regions of human La autoantigen. *J. Biol. Chem.*, **278**, 12231–12240.
10. Maniataki, E., Martinez de Alba, A.E., Sägesser, R., Tabler, M., and Tsagris, M. (2003) Viroid RNA systemic spread may depend on the interaction of a 71-nucleotide bulged hairpin with the host protein VirP1. *RNA*, **9**, 346–354.
11. Mazumder, B., Sampath, P., Seshadri, V., Maitra, R.K., DiCorleto, P.E., and Fox, P.L. (2003) Regulated release of L13a from the 60S ribosomal subunit as a mechanism of transcript-specific translational control. *Cell*, **115**, 187–198.
12. Ziehler, W.A., Morris, J., Houser-Scott, F., Millikin, C., and Engelke, D.R. (2001) An essential protein-binding domain of nuclear RNase P RNA. *RNA*, **7**, 565–575.
13. Bartel, P.L. and Fields, S. (1995) Analyzing protein-protein interactions using two-hybrid system. *Methods Enzymol.*, **254**, 241–263.
14. Bartel, P., Chien, C.-T., Sternglanz, R., and Fields, S. (1993) Elimination of false positives that arise in using the two-hybrid system. *Biotechniques*, **14**, 920–924.

15. Bartel, P.L., Chien, C.-T., Sternglanz, R., and Fields, S. (1993) in *Cellular Interactions in Development: A Practical Approach* (ed D.A. Hartley), Oxford University Press, Oxford, pp. 153–179.
16. Brent, R. and Finley, R.L. Jr. (1997) Understanding gene and allele function with two-hybrid methods. *Annu. Rev. Genet.*, **31**, 663–704.
17. Mac Donald, P.N. (ed.) (2001) *Methods in Molecular Biology*, Two-Hybrid Systems, Vol. 177, Humana Press, Totowa.
18. Putz, U., Skehel, P., and Kuhl, D. (1996) A tri-hybrid system for the analysis and detection of RNA–protein interactions. *Nucleic Acids Res.*, **24**, 4838–4840.
19. SenGupta, D.J., Zhang, B., Kraemer, B., Pochart, P., Fields, S., and Wickens, M. (1996) A three-hybrid system to detect RNA-protein interactions *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 8496–8501.
20. Zhang, B., Kraemer, B., SenGupta, D., Fields, S., and Wickens, M. (2000) Yeast three-hybrid system to detect and analyze RNA-protein interactions. *Methods Enzymol.*, **318**, 399–419.
21. Sonoda, J. and Wharton, R.P. (1999) Recruitment of Nanos to hunchback mRNA by Pumilio. *Genes Dev.*, **13**, 2704–2712.
22. Bouffard, P., Barbar, E., Brière, F., and Boire, G. (2000) Interaction cloning and characterization of RoBPI, a novel protein binding to human Ro ribonucleoproteins. *RNA*, **6**, 66–78.
23. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith J.A., and Struhl K. (eds) (1996) in *Current Protocols in Molecular Biology*, *Saccharomyces cerevisiae*, Greene Publishing Associates and Wiley-Interscience, New York.
24. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn, Cold Spring Harbor Laboratory Press, Plainview.
25. Guthrie, C. and Fink, G.R. (eds) (1991) *Guide to Yeast Genetics and Molecular Biology*. *Methods Enzymol.*, **194**, 3–792.
26. Good, P.D. and Engelke, D.R. (1994) Yeast expression vectors using RNA polymerase III promoters. *Gene*, **151**, 209–214.
27. Bertrand, E., Houser-Scott, F., Kendall, A., Singer, R.H., and Engelke, D.R. (1998) Nucleolar localization of early tRNA processing. *Genes Dev.*, **12**, 2463–2468.
28. Srisawat, C., Houser-Scott, F., Bertrand, E., Xiao, S., Singer, R.H., and Engelke, D.R. (2002) An active precursor in assembly of yeast nuclear ribonuclease P. *RNA*, **8**, 1348–1360.
29. Bardwell, V.J. and Wickens, M. (1990) Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res.*, **18**, 6587–6594.
30. Witherell, G.W., Wu, H.-N., and Uhlenbeck, O.C. (1990) Cooperative binding of R17 coat protein to RNA. *Biochemistry*, **29**, 11051–11057.
31. Lowary, P.T. and Uhlenbeck, O.C. (1987) An RNA mutation that increases the affinity of an RNA-protein interaction. *Nucleic Acids Res.*, **15**, 10483–10493.
32. Keryer-Bibens, C., Barreau, C., and Osborne, H.B. (2008) Tethering of proteins to RNAs by bacteriophage proteins. *Biol. Cell*, **100**, 125–138.
33. Rho, S.B. and Martinis, S.A. (2000) The b14 group I intron binds directly to both its protein splicing partners, a tRNA synthetase and maturase, to facilitate RNA splicing activity. *RNA*, **6**, 1882–1894.
34. Cassidy, L.A. and Maher, L.J. 3rd (2001) *In vivo* recognition of an RNA aptamer by its transcription factor target. *Biochemistry*, **40**, 2433–2438.
35. Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J., and Wickens, M.P. (1997) A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature*, **390**, 477–484.
36. Wang, Z.-F., Whitfield, M.L., Ingledue, T.C. 3rd, Dominiski, Z., and Marzluff, W.F. (1996) The protein that binds the 5' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. *Genes Dev.*, **10**, 3028–3040.
37. Martin, F., Schaller, A., Eglite, S., Schümperli, D., and Müller, B. (1997)

- The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein. *EMBO J.*, **16**, 769–778.
38. Aziz, N. and Munro, H.N. (1987) Iron regulates ferritin mRNA translation through a segment of its 5' untranslated region. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 8478–8482.
 39. Klausner, R.D., Rouault, T.A., and Harford, J.B. (1993) Regulating the fate of mRNA: the control of cellular iron metabolism. *Cell*, **72**, 19–28.
 40. Harrington, L. McPhail, T., Mar, V., Zhou, W., Oulton, R., Bass, M.B., Arruda, I., and Robinson, M.O. (1997) A mammalian telomerase-associated protein. *Science*, **275**, 973–977.
 41. Long, R.M., Gu, W., Lorimer, E., Singer, R.H., and Chartrand, P. (2000) She2p is a novel RNA-binding protein that recruits the Myo4p-She3p complex to *ASH1* mRNA. *EMBO J.*, **19**, 6592–6601.
 42. Park, Y.W., Tan, S.L., and Katze, M.G. (1999) Differential sensitivity to 5-fluoro-orotic acid as a screen for bait RNA-independent false positives in a yeast three-hybrid system. *Biotechniques*, **26**, 1102–1106.
 43. Hook, B., Bernstein, D., Zhang, B., and Wickens, M. (2005) RNA-protein interactions in the yeast three-hybrid system: affinity, sensitivity, and enhanced library screening. *RNA*, **11**, 227–233.
 44. Sengupta, D.J., Wickens, M., and Fields, S. (1999) Identification of RNAs that bind to a specific protein using the yeast three-hybrid system. *RNA*, **5**, 596–601.
 45. Venables, J.P., Ruggiu, M., and Cooke, H.J. (2001) The RNA-binding specificity of the mouse *Dazl* protein. *Nucleic Acids Res.*, **29**, 2479–2483.
 46. Bohjanen, P.R., Liu, Y., and Garcia-Blanco, M.A. (1997) TAR RNA decoys inhibit Tat-activated HIV-1 transcription after preinitiation complex formation. *Nucleic Acids Res.*, **25**, 4481–4486.
 47. Konopka, K., Lee, N.S., Rossi, J., and Düzgüneş, N. (2000) Rev-binding aptamer and CMV promoter act as decoys to inhibit HIV replication. *Gene*, **255**, 235–244.

49

Experimental Identification of MicroRNA Targets

Michaela Beitzinger and Gunter Meister

49.1

Introduction

MicroRNAs (miRNAs) are a highly conserved class of small non-coding RNAs, which, in most cases, negatively regulate gene expression posttranscriptionally [1–3]. miRNA-guided gene regulation is important for various cellular pathways including apoptosis, cell cycle, differentiation, or development [3, 4]. Since miRNAs constitute fundamental gene regulators, their expression affects several diseases including cancer [5, 6]. Therefore, it is important to understand how miRNAs function and what target genes they bind and regulate. Here we describe a biochemical protocol for the identification of miRNA targets.

miRNAs are transcribed as primary transcripts that are successively processed by the RNase III enzymes Drosha and Dicer to short double-stranded RNA (dsRNA) intermediates [7, 8]. From these short dsRNAs, one strand is selected that gives rise to the mature miRNA. The other strand, referred to as the *miRNA** (*microRNA-star*) is degraded. The mature miRNA is incorporated into mRNPs, also referred to as the *RISC* [9]. Within miRNPs, miRNAs bind to a member of the Argonaute (Ago) protein family [10, 11]. The miRNA serves as a guide and directs miRNPs to the target mRNAs where they mediate translational repression or mRNA degradation [12, 13]. miRNA target sites are typically located in the 3' UTR of the target genes. Owing to the low degree of complementarity between target mRNA and the miRNA, identification of miRNA targets continues to be challenging.

Computational tools have been developed for miRNA target prediction; these are usually based on sequence conservation and the so-called seed region of the miRNA, which is most important for target recognition. However, target prediction is often not very precise and targets are predicted irrespective of tissue- or cell-line-specific expression patterns of miRNAs or their target mRNAs. Therefore, it is important to develop biochemical methods for the identification of miRNA targets. Most methods are based on the direct interactions of the miRNA with Ago proteins and its target mRNAs. To date, several different approaches exist. One basic method that will be described in this protocol relies on the coimmunoprecipitation of Ago proteins with miRNAs and their target mRNAs. This technique has been used

successfully by several groups [14–19] and allows for the identification of mRNAs physically associated with Ago proteins. A further development of this method is the so-called CLIP assay, where, before coimmunoprecipitation, miRNAs and mRNAs are directly crosslinked with Ago proteins [20–22]. This method not only identifies RNAs that are associated with Ago proteins but also their specific binding sites. In addition to the coimmunoprecipitation of miRNA targets with anti-Ago antibodies the use of biotinylated synthetic miRNAs for target identification has been reported [23, 24]. In this approach, one specific biotinylated miRNA is pulled down using streptavidin beads and the associated mRNAs and proteins are analyzed.

In this protocol, we describe the identification of miRNA targets by coimmunoprecipitation with endogenous Ago2. Further analysis of these targets can be performed using gene-expression arrays or qRT-PCR. However, a large pool of mRNAs will be detected because, at a time, many different miRNA/mRNA pairs are associated with Ago. mRNAs with a high affinity to Ago proteins, which are usually strongly regulated miRNA targets, can be easily detected. For detection of targets that are less regulated by miRNAs, it might be necessary to further optimize the protocol regarding the amount of cells and/or antibody used. Data analysis may be performed *via* gene-expression arrays to obtain a global overview of all miRNA targets associated with Ago proteins, or *via* qRT-PCR to analyze specific targets. In addition to the standard protocol for detection of miRNA targets associated with Ago proteins, we describe an improvement to this method, which enables the identification of specific targets for one single miRNA. Therefore, we use specific locked nucleic acid (LNA) probes or 2'-O-methyl-modified miRNA inhibitors. In this protocol, we describe all necessary steps and control reactions that are important for efficient data generation.

Protocol

Use fresh cells or snap freeze the dry cell pellet in liquid nitrogen before storage at -80°C to avoid protein degradation. For gene-expression-array analysis we typically use 5–10 mg total lysate; if qRT-PCR analysis is preferred, the amount of cells can be scaled down. If standard cell lines (e.g., HEK293 or HeLa) are used, one confluent 15-cm dish usually results in about 1 mg total protein lysate. The amount of cells needed varies between different cell types. If no standard cell lines are used, this step might need further optimization. For the detection of targets specific to one miRNA, it is necessary to transfect the cells with miRNA inhibitors (LNA-probes or 2'-O-methyl miRNA inhibitors), which can be purchased from different companies.

We recommend using magnetic beads for identification of coimmunoprecipitated mRNAs and to incubate the antibody with the lysate before adding the beads. This strategy will help reduce unspecific binding of RNA to the beads. Nevertheless, it is also possible to use sepharose beads and/or to precouple the beads with antibody before immunoprecipitation, in particular, if hybridoma supernatant is used.

Cell Culture and Transfection

For miRNA target identification, all kind of cells or tissues may be used and cultured according to their normal conditions. Here we describe the protocol using adherent cells, for example, HEK293 or HeLa. For transfection we recommend the use of Lipofectamine RNAi/Max reagent, which results, in our hands, in the best transfection efficiencies. For transfection, use the reverse protocol according to the manufacturer's instructions and use the same amount of inhibitor as described for transfection of siRNAs. The immunoprecipitation could be performed 24–48 h after transfection. If you want to use special cell lines or primary cells, which are more difficult to transfect, the transfection protocol might be adapted to another reagent or even to another transfection method (e.g., electroporation).

Sample Preparation

For harvesting, the dishes should be washed in PBS first to remove residual cell culture medium. Use about 5 ml PBS per 15-cm plate, scrape the cells directly into PBS from the plate, and centrifuge at 200 g and 4 °C for 10 min. Remove the supernatant and resuspend the cell pellet in cold (4 °C) cell lysis buffer (use 1 ml per 15-cm plate) and incubate on ice for 30 min. Be sure to remove the entire supernatant before adding lysis buffer to the cell pellet to avoid dilution of the lysis buffer. This protocol could easily be customized to suspension cells, but the amount of cells needs to be adjusted. Centrifuge at maximum speed in a tabletop centrifuge for 30 min at 4 °C and transfer the supernatant into a new tube. The lysate should be clear and free of residual cell debris. If not, just centrifuge again or filter the lysate with a 0.45- μ m syringe filter.

Immunoprecipitation

Aliquot the lysate to equal volumes per immunoprecipitation (e.g., Ago2 and control) and fill the tube with cell lysis buffer to avoid larger volumes of air in the tube. For data analysis and quality control, it is important to save about 10% of the lysate volume, which is used for the immunoprecipitation as input control. This input sample could be stored on ice until the immunoprecipitation is ready for RNA precipitation and protein analysis. Add 2–5 μ g of antibody per immunoprecipitation and incubate for at least 3 h at 4 °C under rotation.

For efficient immunoprecipitation with reduced unspecific binding of RNA, use 30 μ l magnetic beads per reaction for qRT-PCR or 90 μ l for gene-expression arrays. Before use, the beads need to be washed once with cold PBS (4 °C). For washing, pipette the appropriate amount of protein G beads (Dynal) in a 1.5-ml reaction tube and place in a magnetic rack. Once the beads attach to the wall of the tube, remove the supernatant carefully. Remove the tube from the magnetic rack, resuspend the beads in 1 ml PBS, and repeat the procedure. Resuspend in PBS, add the appropriate amount of beads directly into the immunoprecipitation sample, and incubate for 1 h at 4 °C under rotation. After binding, the beads need to be washed four times with IP wash buffer. To remove residual detergent, perform one additional washing step in PBS and transfer the beads into a new tube to reduce

protein/RNA contamination due to RNA and/or Proteins might stick to the walls of the tube.

To test the efficiency of the immunoprecipitation, take a small aliquot of beads (about 10% from the last washing step) for further analysis. The beads from this aliquot should be resuspended in 2 x SDS-loading dye for Western blot analysis of efficient Ago2 immunoprecipitation. The samples could be stored at -20°C until use.

Western Blot Analysis of Ago2 Immunoprecipitation

For Western blot analysis of immunoprecipitated Ago proteins follow standard protocols with some specifications. Thaw samples stored in SDS-loading buffer on ice first. Incubate the thawed samples for 5 min at 95°C and load them on a 10% SDS-PAGE. Include one well for a prestained molecular weight marker. After the gel run, transfer the proteins onto a nitrocellulose membrane *via* semidry electroblotting.

For protein detection, we typically use anti-rat-Ago2 hybridoma supernatant diluted 1 : 50 in TBS supplemented with 0.1% Tween and 5% milk powder. Follow standard ECL protocol to detect the immunoprecipitated Ago protein signals on the Western blot membrane. An example is shown in Figure 49.1a.

RNA Precipitation

RNA precipitation starts immediately after the final wash. The beads are resuspended in 250 μl proteinase K buffer supplemented with 1 μl proteinase K (20 mg ml^{-1}). The input should be treated equally. Incubate the beads/input in proteinase K buffer for 15 min at 65°C with gentle agitation to keep the beads in motion. This helps in better protein digestion. After proteinase K digestion, add 300 μl phenol–chloroform–isoamylalcohol and mix by vortexing. Centrifuge the samples at 17 000 g (maximum speed) for 5 min at 4°C to separate the phenol

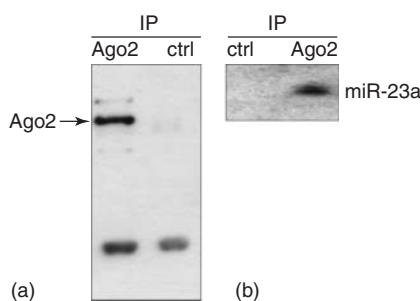


Figure 49.1 Ago2 coimmunoprecipitation of miR-23a. (a) Endogenous Ago2 was immunoprecipitated from 2 mg HeLa cell lysate, using anti-Ago2 antibody (lane Ago2) or an unspecific control antibody (lane ctrl). Efficient immunoprecipitation was confirmed

via Western blot analysis. (b) Coimmunoprecipitation of the abundant miRNA miR-23a after immunoprecipitation with anti-Ago2 (lane Ago2) or control antibody (lane ctrl) was confirmed by Northern blot analysis.

from the aqueous phase. The upper aqueous phase, which contains the RNA, needs to be transferred to a new 1.5-ml tube. Repeat the extraction once with chloroform–isoamylalcohol to remove residual traces of phenol, which might interfere with the following RNA analysis. For final RNA precipitation, add 1 μ g 20 mg/ml glycogen and 2.5 vol 100% EtOH, mix well and store at -20°C overnight (do not add additional salt for the precipitation). For final RNA isolation, centrifuge the samples at 4°C for 30 min at 17 000 g (maximum speed). Carefully remove the supernatant and wash once with 70% EtOH to remove residual salt from the pellet. Repeat the centrifugation for 10 min and remove the supernatant. Remove all of the supernatant, but be sure not to disturb the pellet. Dry the pellet until all residual EtOH is removed. Dissolve in appropriate volume of RNase-free water and store at -80°C until use. For detection of coimmunoprecipitated miRNAs take about 10% of the dissolved RNA (input and immunoprecipitation) for Northern blot analysis.

Northern Blot Analysis for miRNA Detection

To verify the efficiency of Ago immunoprecipitation regarding coimmunoprecipitated miRNAs and, therefore, the associated mRNAs, it is important to perform a miRNA analysis. We recommend standard Northern blot protocols and EDC (1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide crosslinking.

Before starting the Northern blot procedure, it is necessary to prepare a radioactively labeled probe for a miRNA, which is known to be expressed in the cells used (e.g., let-7a or miR-23a in HeLa cells). A DNA sequence that is complementary to the chosen miRNA is usually used as a probe. For the labeling reaction, use 1 μ l of 20 μ M DNA oligonucleotide and 1 μ l [γ - ^{32}P]ATP (3000 Ci/mmol) in a 20 μ l T4-polynucleotide kinase reaction and incubate at 37°C for 30 min. Add 30 μ l 30 mM EDTA and purify with a G25 spin column.

For detection of coimmunoprecipitated miRNA use a 12% denaturing polyacrylamide gel and let it sit for at least 2 h. Let the gel run in 1 x MOPS buffer for 10 min at 200 V first (prerun) before loading the samples. During this prerun, thaw stored samples on ice and add 2 x denaturing RNA sample buffer. Boil the samples for 5 min at 95°C . Immediately before loading, rinse the wells of the gel to remove urea, which diffuses from the gel and accumulates in the wells of the gel. Load the samples and run the gel at 300–400 V for 1–2 h. It is not necessary to run the dye front to the bottom of the gel. The miRNAs usually run between the bromophenol blue and the xylene cyanol dye front.

Prepare the blot sandwich as follows: three layers of 3MM Whatman paper, nylon membrane, gel, and again three layers of 3MM Whatman paper. Each component should be soaked in distilled water first. Use a semidry blotting chamber and blot the RNA onto the membrane for 30 min at 20 V. During blotting, prepare the prehybridization solution and prewarm it to 50°C . Immediately before use, saturate a 3MM Whatman paper with freshly prepared EDC crosslinking solution. Place the gel with the RNA side on top, wrap it in Saran wrap, and incubate at 50°C for 1 h. After crosslinking, wash the membrane in distilled water for a few seconds and place it in a hybridization bottle. Incubate the membrane in prehybridization solution for

1 h at 50 °C in a hybridization oven under rotation. After prehybridization add the [γ -³²P]ATP-labeled probe directly into the prehybridization solution and incubate overnight at 50 °C under rotation. To remove unbound probe, wash the membrane two times in wash buffer 1 and once in wash buffer 2 for 10 min each at 50 °C. The wash buffers need to be prewarmed to 50 °C before use. Finally wrap the membrane in Saran wrap and expose to X-ray film for several hours or days at –80 °C. The exposure time depends on the abundance of the miRNA and on the amount present in the coimmunoprecipitation. As an example, a Northern blot is shown in Figure 49.1b.

mRNA Analysis

Different possibilities exist for the analysis of mRNAs associated with Ago proteins. You can detect specific targets *via* qRT-PCR or use a more global approach to detect all mRNAs associated with Ago proteins *via* gene-expression arrays. If you are using specific miRNA inhibitors, both methods may be useful. Global gene expression can be analyzed either through one-color or two-color arrays. Two-color arrays will allow to probe the gene chip with two different samples at one time, which might be interesting, especially when using miRNA inhibitors.

qRT-PCR Analysis

For analysis of Ago2-associated mRNAs *via* qRT-PCR, perform a DNase I digestion first, to avoid DNA contamination and false positive results in the PCR reaction. Perform the DNase I digestion according to the manufacturer's protocol (e.g., Fermentas). After DNase I treatment, set up a cDNA synthesis reaction according to the supplier's protocol (Fermentas cDNA Synthesis Kit). The resulting cDNA reaction (usually 20 μ l) can be further diluted up to 50 μ l, depending on the amount of precipitated RNA.

For the qPCR reaction, usually one 1 μ l of diluted cDNA per well is sufficient, but this may be scaled up to 5 μ l per well. It is recommended to measure in triplicates.

Amounts of coimmunoprecipitated mRNA are often limited, especially for low-abundance target mRNAs and since unspecific binding of RNA to the protein G beads cannot be completely avoided. Therefore, we calculate the enrichment of a target mRNA relative to GAPDH mRNA or tRNA enrichment, since these RNAs do not seem to be targets of Ago proteins. Other mRNAs that are not miRNA targets may be used for normalization as well (for an example of qRT-PCR, see Figure 49.2). If you use miRNA inhibitors, the enrichment in this sample should be significantly lower than in a sample without an inhibitor.

Gene-Expression-Array Analysis

To perform a gene-expression-array analysis, it is first necessary to choose the system and the type of array one would like to analyze. Since many different systems and analysis software are available, this protocol does not go into detail here. For performing gene-expression-array analysis follow the instructions of the manufacturer. Data analysis depends on the software used. We recommend

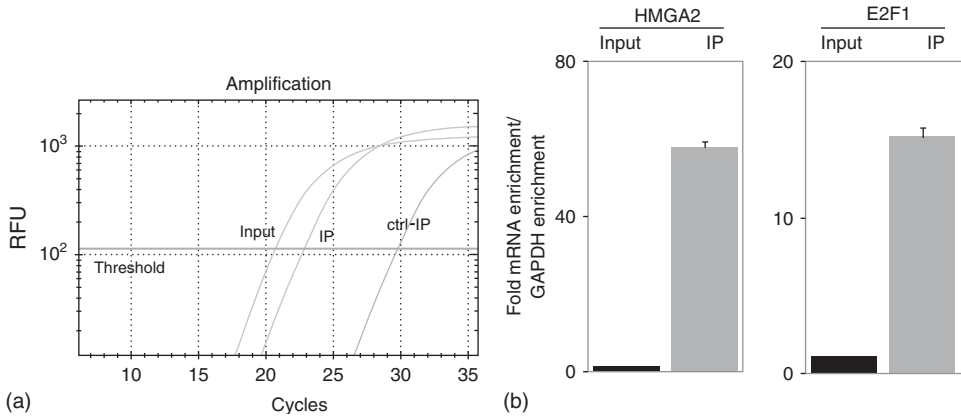


Figure 49.2 Ago2 coimmunoprecipitation of mRNAs: qRT-PCR analysis. (a) Endogenous Ago2 was immunoprecipitated from 10 mg HCT116 cell lysate. The coimmunoprecipitation of Ago2-associated mRNAs was confirmed by qRT-PCR of known miRNA targets HMGA2 and E2F1 mRNAs. As an example, the qRT-PCR amplification curve with E2F1 mRNA is shown. Sample identities

are depicted in the graph. Relative fluorescence intensity (RFU) is drawn against cycle number. The dark gray line in the middle depicts the threshold, which was used for Ct value determination. (b) The fold enrichment of HMGA2 and E2F1 mRNAs was calculated relative to the amount of input mRNA and to the unspecific enrichment of GAPDH mRNA levels.

calculation of the enrichment of the Ago immunoprecipitation on the levels of the input signals according to the qRT-PCR analysis.

49.2

Troubleshooting and Notes

- Be sure to use RNA-free reagents for each individual step since RNase contamination is very often the reason for complete loss of the RNA or unreproducible results.
- The amount of cells needed can vary depending on the type of analysis. If only qRT-PCR is performed, the amount of lysate can be reduced to 1–2 mg of total protein. For gene-expression analysis, more cells are needed because this analysis does not include amplification as PCR does.
- Be sure that the cells do not become too confluent to avoid cellular stress. Changing conditions of cell culture may result in stress-mediated gene regulation, which could also affect Ago mRNA association and thereby the reproducibility of the results.
- Be sure that the protein lysate is clear and free of cellular debris. This is a critical point for the success of the experiment.
- Avoid larger volumes of air or foam in the tube, because proteins start to aggregate preferentially at air–liquid interphases.

- Rinse the wells of the denaturing polyacrylamide gel immediately before loading the samples. Urea usually diffuses from the gel and accumulates on the bottom of the wells. This high urea concentration affects separation of the RNA during the gel run.
- The radioactively labeled probe for Northern blot analysis can be stored at -20°C until use but no longer than a few days, since ^{32}P has a half-life of only 14 days.
- If you use miRNA inhibitors, analyze the efficiency of miRNA inhibition *via* Northern blot analysis. Efficient miRNA inhibition would result in loss of miRNA binding to Ago proteins.

49.3

Buffers and Solutions

- Lysis buffer: 0.5% NP40, 150 mM KCl, 25 mM Tris pH 7.5, 2 mM EDTA, 0.5 mM DTT, and 1 mM AEBSF protease inhibitor (add both immediately before use)
- Phosphate-buffered saline pH 7.5
- IP wash buffer (300 mM NaCl, 50 mM Tris pH 7.5, 5 mM MgCl_2 , 0.05% NP40)
- TBS (150 mM NaCl, 10 mM Tris pH 8.0)
- Proteinase K digestion buffer (300 mM NaCl, 200 mM Tris pH 7.5, 25 mM EDTA, 2% SDS)
- EDC crosslinking solution (0.16 M EDC prepared in 0.13 M 1-methimidazole pH 8.0)
- 20 x SSC pH 7.0 (3 M NaCl, 0.3 M sodium citrate)
- 50 x Denhardt's solution (1% albumin fraction V, 1% polyvinyl-pyrrolidone K30, 1% Ficoll 400)
- Prehybridization solution (7.5 ml 20 x SSC, 600 μl 1 M Na_2HPO_4 pH 7.2, 21 ml 10% SDS, 300 μl 50 x Denhardt's solution)
- Wash buffer 1 (1 x SSC, 5% SDS)
- Wash buffer 2 (1 x SSC, 1% SDS)
- Phenol–chloroform–isoamylalcohol for RNA extraction (Roth).

Additional reagents:

- Anti-Ago2 antibody (11A9) available from Ascenion
- Unspecific control antibody, for example, anti-Flag
- Secondary antibodies for Western blot detection
- Protein-G magnetic beads (Dynal)
- G25 spin column (GE-Healthcare)
- $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (Perkin-Elmer)
- T4-polynucleotide kinase (NEB)
- Hybond-N+ Nylon membrane (GE Healthcare)
- Hybond ECL membrane (GE Healthcare)
- Cell scrapers
- DNase I (Fermentas)
- Ribolock (Fermentas)

- First strand cDNA synthesis Kit (Fermentas)
- qPCR Mix (Eurogentec 2 x Mesa Green with fluorescein for Biorad iCycler)
- Gene-expression array (Affymetrix human gene expression array U133 Plus and reagents).

References

- Kim, V.N., Han, J., and Siomi, M.C. (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, **10**, 126–139.
- Carthew, R.W. and Sontheimer, E.J. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell*, **136**, 642–655.
- Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
- Chen, P.Y. and Meister, G. (2005) microRNA-guided posttranscriptional gene regulation. *Agric. Biol. Chem.*, **386**, 1205–1218.
- Calin, G.A. and Croce, C.M. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, **6**, 857–866.
- Esquela-Kerscher, A. and Slack, F.J. (2006) Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer*, **6**, 259–269.
- Siomi, H. and Siomi, M.C. (2009) On the road to reading the RNA-interference code. *Nature*, **457**, 396–404.
- Kim, V.N. (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.*, **6**, 376–385.
- Meister, G. and Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature*, **431**, 343–349.
- Peters, L. and Meister, G. (2007) Argonaute proteins: mediators of RNA silencing. *Mol. Cell*, **26**, 611–623.
- Hutvagner, G. and Simard, M.J. (2008) Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.*, **9**, 22–32.
- Meister, G. (2007) miRNAs get an early start on translational silencing. *Cell*, **131**, 25–28.
- Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?. *Nat. Rev. Genet.*, **9**, 102–114.
- Weinmann, L., Hock, J., Ivacevic, T., Ohrt, T., Mutze, J., Schwill, P., Kremmer, E., Benes, V., Urlaub, H., and Meister, G. (2009) Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell*, **136**, 496–507.
- Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P.Y., Soll, S.J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M., and Tuschl, T. (2008) Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA*, **14**, 2580–2596.
- Hendrickson, D.G., Hogan, D.J., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2008) Systematic identification of mRNAs recruited to argonaute 2 by specific microRNAs and corresponding changes in transcript abundance. *PLoS ONE*, **3**, e2126.
- Karginov, F.V., Conaco, C., Xuan, Z., Schmidt, B.H., Parker, J.S., Mandel, G., and Hannon, G.J. (2007) A biochemical approach to identifying microRNA targets. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 19291–19296.
- Easow, G., Telesman, A.A., and Cohen, S.M. (2007) Isolation of microRNA targets by miRNP immunopurification. *RNA*, **13**, 1198–1204.
- Beitzinger, M., Peters, L., Zhu, J.Y., Kremmer, E., and Meister, G. (2007) Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol.*, **4**, 76–84.
- Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.

21. Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science*, **302**, 1212–1215.
22. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
23. Orom, U.A. and Lund, A.H. (2007) Isolation of microRNA targets using biotinylated synthetic microRNAs. *Methods*, **43**, 162–165.
24. Orom, U.A., Nielsen, F.C., and Lund, A.H. (2008) MicroRNA-10a binds the 5' UTR of ribosomal protein mRNAs and enhances their translation. *Mol. Cell*, **30**, 460–471.

50

Aptamer Selection against Biological Macromolecules: Proteins and Carbohydrates

Franziska Peter and C. Stefan Voertler

50.1

Introduction

Over the last almost two decades, combinatorial approaches have brought new solutions to chemical and biochemical problems due to the analysis of large sets of molecules in parallel (a library) rather than testing individual members one at a time. Aptamers are one product of such strategies, representing *de novo* generated macromolecules with the ability to bind a chosen target, as implied by the latin word root *aptus* = to fit. Generation of binding ability requires the formation of a binding pocket or surface with stabilizing ionic, H-bonding or hydrophobic interactions, features that can be provided by macromolecules. The two salient requirements of *in vitro* selection approaches are methods to separate active from inactive library members (the partitioning or selection step) and to subsequently regenerate the library, enriched in binders, for a next round of selection. The latter is based on replicability and the direct connection of pheno- with genotype, criteria that are fulfilled by nucleic acid polymers [1], making RNA and DNA the first and still most important molecules in the construction of complex libraries as a source for the isolation of aptamers with tailored properties [2–4]. In comparison, isolation of peptide aptamers was developed much later [5] and requires additional techniques such as phage-, ribosome-, or mRNA-display to overcome the missing direct linkage between pheno- and genotype, either by confining the DNA to the phage body presenting the phenotype (compartmentalization), formation of a stalling ribosome, or covalent attachment of peptide product to its mRNA template [6–10].

Targets for successful aptamer selections have included inorganic ions [11–13]; small organic molecules such as malachite green [14] or theophylline [15]; antibiotics [13, 16] and metabolites such as ATP [17, 18] or AdoMet [19]; amino acids [20, 21] and biological macromolecules such as peptides [22], proteins [3], or carbohydrates [23]; supramolecular structures such as viral particles [24] or ribosomes [25]; and whole cells, for example, trypanosomes [26] or red blood cell ghosts [27]. Aptamers often undergo a substantial conformational rearrangement upon binding to their target, a principle also providing the functional basis of natural riboswitches

[28, 29] and switch ribozymes [30] involved in the regulation of gene expression. Several reviews document the diversity of the field [131–36] while other articles focus on technical aspects [11, 37, 38] or introduce the reader to relevant online databases (Aptamer Database at <http://aptamer.freebase.com> and SELEX database at www.mgs.bionet.nsc.ru/mgs/systems/selex/ [39]). Since each selection is unique in its requirements, this chapter can just provide an overview of the technique, its potentials, and critical steps, based on two target types presented. Careful design by searching the literature and aptamer databases for targets related to the own selection aims remains essential.

What to expect from aptamers? First, aptamers bind their targets with high specificity. This is exemplified by differentiation among protein kinase C isoforms [40] or a single methyl group addition in caffeine compared to theophylline resulting in a 4 orders of magnitude weaker binding of an antitheophylline aptamer [15]. Secondly, they have high affinities for their targets, with K_d values in the range of low micromolar to low nanomolar, in exceptional cases even picomolar for protein targets, while aptamers against small organic molecules display usually higher K_d values in the micromolar range. Thirdly, aptamers allow for chemical modifications to stabilize against nucleases, for example, by incorporation of 2'-fluoro- or 2'-amino-nucleotide analogs during transcription [41, 42] or chemical synthesis [43]. Additionally, the so-called Spiegelmers composed of non-natural L-RNA as well as LNA (locked nucleic acid) and PNA (peptide nucleic acid) aptamers, which are stabilized against nucleases, have been selected [44–46]. In addition, selection efficiency may be improved by equipping the library molecules with photoactivatable bases that permit to crosslink target–aptamer complexes [47]. Fourthly, aptamers can be engineered to reduce their size, and affinity tags, fluorophores [48], or modules for allosteric regulation may be added. They work as chiral HPLC phases [49], sensors [50, 51], or signal transducers [52–54], and even in array-based technologies [55] and high-throughput assays [56, 57]. Fifthly, optimization or adaptation of binding can be performed by reselection using a partly randomized library based on the initial aptamer sequence (subtractive SELEX [21, 58, 59]). Sixthly, selection processes can be automated and further expedited by *in vitro* transcription/translation of target proteins with a biotin tag for immobilization [60, 61]. Finally, the selection process is often rapid compared with the generation of antibodies, and particularly aptamer production after selection stands out in terms of reproducibility and robustness of the process. Moreover, since selections are performed *in vitro*, any selection scheme and target can be used.

50.2

General Strategy

Combinatorial macromolecular libraries can be subjected to screening or selection/evolution processes. The first requires multitube (or multiwell) high-throughput assays, while the latter is a one-tube reaction, explaining why this approach is

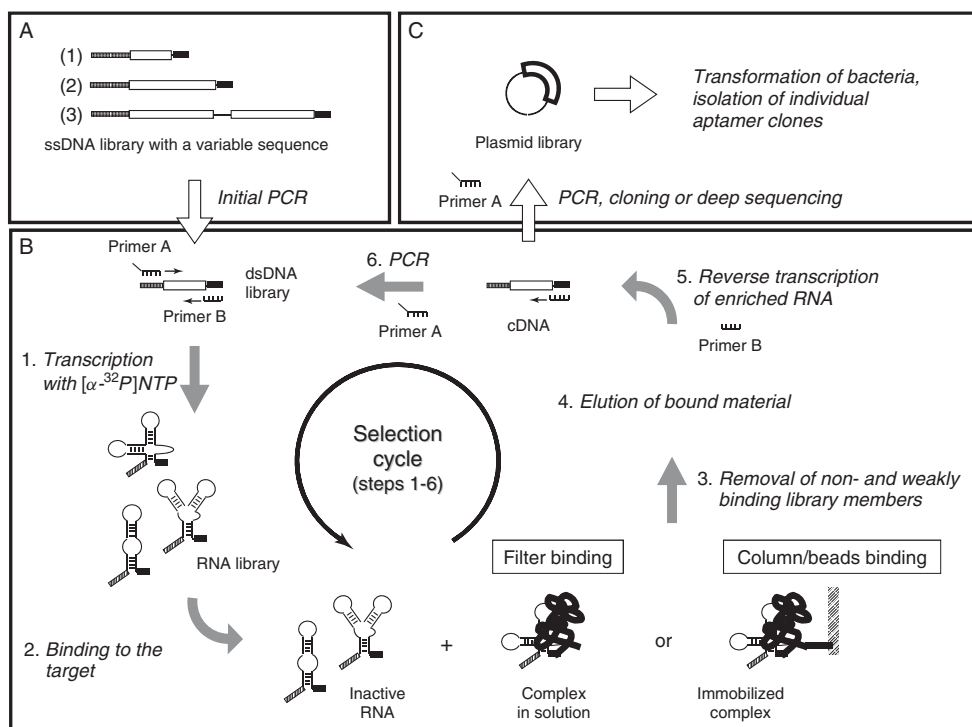


Figure 50.1 *In vitro* selection cycle. (a) The initial library is synthesized chemically as ssDNA including binding sites for primer A and B (striped and black boxes); primer A includes a T7 RNA promoter to initiate RNA transcription. The variable region (white boxes) can be a continuous stretch of varying length (1, 2) or a combination of two random regions interrupted by a fixed sequence (3). (b) Selection cycle including following steps: transcription of dsDNA pool

(1), binding of the RNA pool to the target (2), removal of non- and weakly binding RNA (3), elution of bound RNA (4), reverse transcription (5), and PCR (6) of enriched RNA sequences. For binding of RNA to target, a filter system or columns and magnetic beads can be used. (c) Cloning of library members to retrieve individual aptamer sequences for sequencing. Alternatively, deep sequencing can be performed.

suitable for any small laboratory with basic equipment. The library is passed through a sequence of steps, the selection cycle, aimed at separating active from inactive molecules (Figure 50.1). The cycle is repeated (counting in selection rounds) until the desired enrichment of functional variants is observed. With increasing cycle number, more and more functional molecules with low binding affinity are removed by performing the partitioning process under increasingly stringent conditions. “Systematic evolution of ligands by exponential enrichment” (SELEX) is one technical term describing this process, based on the Darwinian principle of the survival of the fittest. “*In vitro* evolution” implies more than simple “selection” since the principles of mutation and adaptation are concurrently

introduced by randomization techniques such as mutagenic PCR or DNA shuffling [62]. In contrast to *in vitro* selection, functional molecules finally isolated by *in vitro* evolution techniques are unlikely to have been present in the initial library since they were generated (evolved) in the course of the experiment. However, for most applications with the rather simple goal of isolating aptamers that bind a ligand with reasonable affinity and specificity, smaller variant libraries and an appropriate *in vitro* selection protocol will suffice.

50.2.1

Choosing a Suitable Target

Generally, targets with positively charged surface areas are expected to result in successful enrichment faster than those carrying negative or no charges. Among biological macromolecules, proteins are excellent targets based on their surface and charge distribution properties [63]. Carbohydrates seem to be more problematic, and much fewer cases of successful selections against carbohydrates have been reported. This is possibly a result of their uncharged nature and less rigid conformation resulting in the absence of distinct binding pockets and surface features. Nucleic acids as aptamer targets suffer from the likely selection of antisense binders since the strongest interaction is driven by Watson–Crick complementarity, as illustrated by the selection of aptamers that recognize their RNA targets predominantly *via* kissing loop interactions [64, 65]. Lipids are again a challenging target due to their hydrophobicity, although first examples of membrane-associating aptamers have been described [66, 67]. In the following, we focus on proteins and carbohydrates as targets and highlight specific problems and important aspects associated with these two target types.

50.2.1.1 Protein Targets

Nucleic acid binding proteins were among the first aptamer targets, with the goal to study the binding requirements of their highly adapted binding sites [3]. Meanwhile, selections against many more proteins, representing a variety of shapes and functions, have been successful, with more than 100 protein and peptide targets listed in the aptamer database. A very important aspect is the state or form of the protein used as target as well as the context in which it is presented during selection. For example, when the target is a cell surface receptor, intact cells [26], a membrane fraction containing the protein [27], isolated protein, or recombinant variants, proteolytic fragments, or just peptide stretches could be used. However, strategies have to be adapted accordingly. Complex assemblies such as cells presenting many targets give rise to different aptamers with a limited chance to find those specific for the protein of interest. Such approaches require 15–25 rounds of selection to remove less specific background binding, as well as specialized techniques such as deconvolution-SELEX. In the latter method, aptamers are photo-crosslinked to their protein targets and the complexes are resolved by SDS-PAGE and electroblotted onto nitrocellulose to remove non-covalently associated RNAs. The protein/RNA

complex band of interest is then excised from the membrane, followed by protein digestion and PCR amplification of the retained RNA [27]. In turn, aptamers successfully selected against targets presented in a complex context can be expected to work properly in their natural environment. Selection with a purified protein can never take this complexity into account, but has the advantage that a biochemically more defined target is used with a higher probability of successful selection in fewer cycles from an environment better to control. Most selections use this strategy, which requires rigorous purification of the target protein. In the case of protein degradation, expression, or solubility problems, isolated proteolytic fragments or derived peptides offer an alternative. Thus, aptamers may be directed against a specific domain or even a single loop, provided that such a structural element is part of an accessible epitope in the intact protein.

A cloned protein, overexpressed in soluble form with an affinity tag, is usually an optimal starting point for *in vitro* selection. Since tags may interfere with folding and function in a position-dependent manner, C- and N-terminally tagged protein variants should be prepared in parallel. In addition, different types of tags may be tested as well (Section 50.2.2). Affinity tags permit efficient purification in a short time and the tag can be used to immobilize the protein during selection. Otherwise, one is restricted in the selection design mainly to nitrocellulose filter binding, the method employed for the majority of proteins isolated from natural sources. The protein preparation has to be as pure as possible, since contaminants will affect the selection outcome. Possibly, a combination of affinity, ion exchange, and gel permeation chromatography is necessary, followed by analytical polyacrylamide gel electrophoresis (PAGE) with overloading the lanes to identify impurities. About 10 mg of pure protein suffice for selection and characterization. To reduce the possibility of artifacts, a single batch of the protein should be used throughout the selection procedure. Additionally, the selection process should be performed as fast as possible due to potential degradation of the target. Therefore, it is recommended to regularly control the integrity of the protein.

50.2.1.2 Carbohydrate Targets

Carbohydrates are the most ubiquitous and prominently exposed molecules on the surface of living cells. Cell–cell recognition and cell activation throughout development and maturation of a living organism depend on this class of macromolecules, and carbohydrate patterns frequently vary according to specific stages of cellular differentiation and development; likewise, alterations in carbohydrate are often related to diseases (e.g., cancer or retrovirus infection). Aptamers offer a versatile tool to analyze and interfere with carbohydrate-mediated recognition processes. However, carbohydrate recognition is generally characterized by exceptionally weak binding with K_d values in the millimolar range. Specialized natural proteins, the lectins, overcome this by clustering of carbohydrate recognition units, which results in micromolar K_d values. Binding to highly charged nucleic acids requires structural properties not common to carbohydrates. Moreover, as conventional biochemical methodology is adapted to charged macromolecules, characterization of aptamer–carbohydrate interaction requires adapted

techniques. Incorporation of positive charges reduces this problem, giving aminoglycosides a unique position among oligosaccharides by making them potent natural RNA binders that, for example, block RNA function on the ribosome and thus act as antibiotics. The spatial arrangement of the functional hydroxy and amino groups in aminoglycosides largely defines if specific rather than simply counter-charge-driven RNA binding can be achieved. The specific binding mode of natural aminoglycosides is illustrated by their blockage of prokaryotic ribosomes while leaving eukaryotic ribosomes largely unaffected [68–70]. Ionic interactions and even pseudo-basepairs between RNA bases and sugar rings apparently dominate aminoglycoside–RNA interactions, whereas hydrophobic intercalations seem to play only a marginal role, explaining why plain carbohydrates have a low RNA-binding capacity.

Details of target preparation are beyond the scope of this chapter. A basic problem is the high structural variability of oligosaccharides compared to proteins. Anomer, epimer, enantiomer, diastereomer, furanose/pyranose, α/β linkage, and branched forms exist. Isolation from natural sources, *de novo* chemical as well as combined chemical-enzymatic syntheses is feasible, but difficult, labor- and cost-intensive [71, 72]. Thus, aptamer selections published so far have concentrated on small fragments or commercially available polymers such as dextran, sephadex, and cellulose [23, 73–75].

50.2.2

Immobilization of the Target

For recombinant proteins, immobilization is easiest achieved by incorporation of a small terminal tag, such as (His)₆-, Strep-, or Nano-tag [76], which permits coupling to tag-specific substituted agarose or silica bead materials. Larger constructs, such as glutathione S-transferase (GST)-fusion proteins, carry the risk to select aptamers against the tag rather than the target. One should further bear in mind that any immobilization matrix represents a potential aptamer target itself (Section 50.3). An appropriate alternative is provided by biotinylation of the protein minimizing the risk of selecting non-functional aptamers. Biotinylation can be performed chemically using activated Biotin (e.g., EZ-Link[®] Sulfo-NHS-SS-Biotin, Thermo Fisher Scientific) leading to the addition of Biotin to the amino groups of Lys and Arg. To avoid the risk of protein denaturation through chemical biotinylation, it is recommended to biotinylate the target *in vivo* during protein expression [77]. As a prerequisite, the protein has to be tagged with an Avi-Tag[™] needed for site-specific recognition and biotinylation by the Biotin ligase (*BirA*). Vector systems are offered for cloning the protein target and recombinant coexpression of the ligase (GeneCopoeia).

It is good practice to assay activity of the immobilized protein before selection to avoid that aptamers against inactive or denatured targets may be isolated. Aptamers binding untagged proteins, such as those purified from natural sources, are selected by isolating the complex formed in solution, for example, by nitrocellulose filter binding.

Carbohydrates can be chemically coupled to succinimidyl-activated dextran or silica beads [78] after the introduction of a reactive amine during synthesis or postsynthetically. Often, *cis*-diol oxidation of sugar rings to aldehydes followed by coupling of amino functions is employed.

50.2.3

Selection Assays

Aptamers exploit surface features of the given target to achieve specific and tight binding. The required biochemistry is straightforward compared to the selection for catalytic function. However, proper experimental design must ensure that only target-specific library members are enriched. One strategy is self-immobilization of binders to a solid-phase-anchored target, followed by washing steps to remove bound molecules of lower affinity and elution of entire target–aptamer complexes or affinity elution of aptamers alone by the addition of free target (Protocol 4). More target material is needed for the latter elution variant, but it ensures higher specificity due to involving two successive binding events. Alternatively, target–aptamer complexes formed in solution are separated from unbound library members by adsorption to nitrocellulose (Protocol 5) or using an electrophoretic mobility shift assay (EMSA, Protocol 9). It is important to keep in mind that all steps of the selection procedure represent selection criteria contributing to its outcome, including target preparation and presentation, as well as reverse transcription (RT), PCR, and transcription reactions (Section 50.3).

50.2.4

Design and Preparation of the Library

Polynucleotides are ideal library molecules due to their replicability, inherent flexibility to adapt and bind to many surface topologies as well as the ease of ^{32}P -labeling for tracing purposes. Non-radioactive labeling (e.g., with fluorescent tags) may provide an alternative, but carries the danger to affect the binding event and bias the selection outcome. We focus here on RNA libraries, since they include all points relevant to dealing with DNA as well. An initial single-stranded (ss) DNA library is chemically synthesized, enzymatically amplified, and transcribed into the starting RNA pool (Protocol 1). Constant regions at the 5' and 3' ends are included for transcription and amplification, but may be omitted if methods for their postselectional addition are developed [79]. Follow PCR rules to optimize these primer binding sites. In the simplest design, a randomized region of up to 80 nt is introduced between flanking primer binding sites. *A priori*, it is difficult to estimate the size of the sought-after binding motif, although some guidelines have been summarized [11]. The diversity (or complexity) of the randomized region defines the library and depends on the total number of nucleotides randomized (n), and the probability p with which the four bases may occur at a given random position. Usually each nucleotide should have the same probability, allowing to calculate

the expected number of sequence variants within the library as $N = p^n = 4^n$ (the sequence space). Diversity can be further increased by including chemically modified nucleotide analogs during transcription [17, 18, 42]. Alternatively, a less complex library can be constructed originated from a given sequence by favoring incorporation of the original nucleotide at a position during chemical synthesis, with limited incorporation of the other three nucleotides (Protocol 1). Such biased or doped pools are useful to search for sequence variations around a natural RNA motif or to optimize aptamers in binding affinity by reselection. Libraries of low complexity can further be obtained by mutagenic PCR of a non-randomized template [80, 81].

An experimentally manageable library contains millimolar amounts of RNA, since at higher concentrations (approximately above 1 mM) RNA tends to precipitate as Mg^{2+} salts. This corresponds to 10^{12} – 10^{15} different sequences in the starting library, thus exceeding the number of variants obtained by other combinatorial methods such as phage display by 3–6 orders of magnitude, but still representing only a subset of the theoretically possible sequence space. Nevertheless, diversity usually suffices as a given RNA sequence can adopt a multitude of conformations, referred to as the *conformational hell* that causes general problems in RNA biochemistry [82]. In turn, however, this implies that a favorable binding conformation can be formed by many sequences, with a few of them likely to be present in the explored fraction of the theoretical sequence space.

50.3

Running the *In vitro* Selection Cycle

The addition of the prepared RNA library to the target starts the first round of selection (Protocol 4). Its outcome is determined by the type of selection assay (e.g., competing binding effects specific to the matrix material or filters used) and incubation conditions (stringency of the selection), which include absolute concentrations as well as the ratio of target to library, buffer composition, incubation temperature, and time.

The concentration of an individual sequence depends on the library complexity. One nanomole of RNA with one random position corresponds to 0.25 nmol (1.5×10^{14} copies) per individual sequence. Having 10 random positions, each of the $4^{10} = 1.05 \times 10^6$ library members is present roughly with 1 fmol or 6×10^8 copies. For a library of medium complexity (40 randomizations), at least 10 nmol of library as a millimolar RNA solution should be used to cover a substantial fraction of the sequence space. Target concentration should always be higher than library concentration to provide sufficient binding sites and to reduce matrix binding. A buffer should be chosen, which is low in divalent metal ion concentration, usually Mg^{2+} , to approximate physiological conditions and to minimize metal-ion-induced RNA hydrolysis. Nonetheless, divalent metal ions, such as Mg^{2+} , Ca^{2+} , or Mn^{2+} , are required for RNA structure formation. Aside

from its buffering effect, pH affects the net charge of a target protein. More positive surface charges lead to stronger, but rather less specific binding of the negatively charged library molecules. A similar argumentation holds for the ionic strength chosen, requiring monovalent ion concentrations of at least 100 mM to reduce unspecific binding. Temperature affects the binding kinetics as well as protein and library stability. Incubations to achieve complex formation are best performed at a temperature to be used for aptamer application. Unwanted binding, for example, to the immobilization matrix, can never be excluded as the matrix concentration is usually much higher than that of the immobilized target [12, 23]. This problem can be counteracted by preincubation (negative selection) with matrix material either unsubstituted (preselection) or derivatized with a protein or carbohydrate different from the target (counter-selection). A yet more powerful strategy is alternating between different selection assays, for example, performing even rounds on an agarose matrix and odd rounds on nitrocellulose filters.

Selections are generally started under conditions of lower stringency, which is then increased in later rounds. Thereby, a broad spectrum of binding molecules is first enriched to several copies per sequence, counteracting early loss of variety due to unspecific events, such as adsorption to tube walls before high-stringency rounds later permit to isolate the best binders. However, this approach in turn increases the chance to accumulate matrix binders in early low stringency rounds, illustrating the dilemma of selection experiments. Since association rates are likely to be fast for the majority of competing RNA variants, tight binding (low K_d values) will primarily depend on a low off-rate (k_{off}), and binding specificity is therefore determined by differences in k_{off} . Thus, high stringency is most efficiently reached by extensive washing rather than shortened incubation times. The latter is rather a measure of precaution to exclude RNAs that have to undergo slow rearrangements to reach a productive binding state. It is important to keep in mind that any change in stringency will affect non-target binding as well. For example, prolonged incubation times may increase the background of matrix binders or the fraction of lost sequences due to constant unspecific adsorption, for example, to tube walls. In addition, all enzymatic steps involved in the cycle potentially affect the outcome, as illustrated by RT-PCR-induced predominance of certain sequences [83].

In initial rounds, just a small fraction of the library, 2–3% or less, can be expected to bind to the target. Progress is monitored by measuring target-bound radioactivity, with increasing levels of retained material expected from round 2 to 3 onward, but taking into account that every change in stringency will change the selection course. In addition, prolongation of the selection process beyond a certain point will lead to progressive elimination of binding sequences. After recovery of the target-binding fraction in each round, the enriched library is regenerated by RT-PCR and T7 RNA polymerase transcription, and the next cycle is started. A fraction of the amplified DNA of each round must be kept, either for later analysis or to be able to repeat a cycle under identical or altered selection conditions.

After the final cycle, the library is ligated into a cloning vector, followed by transformation into bacteria to isolate individual members. However, cloning of the pool is time consuming and limits the number of sequences, which can be analyzed. This is problematic especially for enriched libraries with a large number of different sequences. High-throughput sequencing methods offer an appropriate alternative enabling a detailed analysis of the enriched sequence pool. Furthermore, the selection progress can be controlled simply by sequencing the library after each selection cycle. Therefore, the sequences of each selection round have to be tagged with a specific short sequence introduced by primers in the PCR [84].

50.4

Analysis of the Selection Outcome

Three levels of analysis are recommended. Since many of the suggested experiments will be covered in the chapters devoted to RNA structure probing (Chapters 9–25), not all protocols are detailed here. The starting point is a sequence comparison of the isolated clones. General strategies are described in the literature [85], and sequence analysis tools for fast alignments can be found on the internet (*www.expasy.org*; ClustalW2: <http://www.ebi.ac.uk/Tools/msa/clustalw2>; and MEME: <http://meme.sdsc.edu/meme/cgi-bin/meme.cgi>) or in commercial software packages (such as DNASTar, Vector NTI). Depending on the number of selection cycles performed, either a very diverse set of sequences or a few sequence families, each comprising several related sequences, can be expected. Most important are stretches of conserved positions found across several aptamers (a consensus motif), indicating likely target interaction sites and allowing arrangement into families. Sequence analysis is then followed by the prediction of secondary structure using Zuker's algorithms (<http://mfold.rna.albany.edu/?q=mfold> [86]) or other folding tools (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>; www.rna.uni-jena.de/rna.php). As different programs might result in different structure predictions, a comparison is recommended. The structure model needs to be validated experimentally by RNA secondary structure probing, footprinting of the target–aptamer complex, boundary experiments to identify minimal binding sequences and possibly two to four cycles of reselection after partial randomization of a defined aptamer sequence [87]. The latter is a powerful tool to prove the importance of a binding motif, but allows for optimization of the original motif as well. Frequently, more avidly binding aptamers are identified by reselection [88]. Finally, the third line of characterization concerns quantification of target–aptamer interaction by determining an apparent equilibrium binding constant (K_d) using filter binding, EMSAs, or surface plasmon resonance (SPR) measurements (Protocols 10–13). Additionally, K_d values can be determined using a number of additional methods such as quantitative affinity chromatography [89, 90] and in-line probing of the RNA–aptamer complex [91]. It is recommended to validate results by at least two different techniques.

50.5

Troubleshooting

It is in the nature of an *in vitro* selection experiment that its outcome can never be foreseen, being actually the hallmark of this approach, as new unexpected solutions are found. Still, selections can fail completely. This is usually recognized already during the selection process, when no enrichment occurs. However, enrichment can be very modest (below 10%) if the background of other sequences is high and merely a subfraction of binders reproducibly folds into their binding-competent conformation. For protein targets, the inability to retrieve an aptamer is most likely a technical problem rather than a general property of the target. The same holds for positively charged carbohydrates, while uncharged carbohydrates are at the front line of the technology. Relevant questions to be addressed during the selection process include the following: Is DNA produced in the RT–PCR step? Does it have the correct size and represent one clear band? Is the DNA transcribed into RNA, again of expected size without fragmentation? Does the RNA have a high affinity toward the matrix or filters used? Was the stringency too high in early rounds? Is the target correctly immobilized, active, and accessible?

Problems often encountered in individual steps are as follows: (i) No or multiple bands obtained by RT–PCR. Running a control RT–PCR without adding RT will exclude DNA contamination problems. Varying the cycle number is the first starting point to reduce PCR artifacts, followed by changes in the PCR extension time, template input, and primer concentration. Standard PCR optimization rules should be followed, particularly with respect to the design of primers, which should be at least 20 nt long and void of internal secondary structures and complementarities to minimize dimerization. (ii) No or additional smaller RNA fragments obtained in the transcription reaction: often T7 RNA polymerase encounters sequence-dependent pausing sites where it drops off the template, giving rise to smaller RNA products, which run below the main product. Changes in the transcript length pattern with increasing rounds of selection may indicate enrichment of truncated RNA species better adapted to the selection conditions. RNase contamination, often feared, is actually rather rare; backbone hydrolysis at fragile sequence stretches due to the presence of Mg^{2+} in the incubation mixture is far more likely.

For all steps, “good RNA working practice” should be followed. RNases are water soluble and contamination can be avoided even without diethylpyrocarbonate (DEPC) treatment: by wearing gloves, rinsing glassware extensively in warm water, followed by Millipore or double-distilled water before use, avoiding to speak while pipetting solutions, using freshly opened bags of plasticware that is usually RNase-free although not guaranteed, avoiding dirt and dust deposits on storage bags and working spaces, usage of pipettes exclusively assigned to RNA work, and possibly equipped with special filter-tips to avoid contamination.

For all selection processes, it is recommended to include a target that is known as a *good interaction partner for RNA*. Because of their positive charge and surface

characteristics, streptavidin and lysozyme are two suitable candidates leading to the selection of different nucleic acid aptamers [92–94].

Protocol 1: Preparing a Nucleic Acid Library

Design of the Template

Efficient and specific primer binding sites need to be coupled with one or more random sequence stretches and an RNA polymerase promoter, usually that for T7 RNA polymerase (5'-GGTAATACGACTCACTATA-3'; Figure 50.1). Start the coding sequence with at least two G residues for efficient transcription initiation. Such a pool of PCR products is then used as template for run-off transcription. Terminal restriction sites introduced by the PCR primers may later be used for cloning purposes, providing an alternative to the TA-cloning after selection (Protocol 8). Reduce the cross-contamination risk within the laboratory by altering primer sequences for each selection and include a unique restriction site next to the ones used for cloning to identify the origin of the selected molecules.

Synthesis of the Initial DNA Library

The initial ss DNA library needs to be synthesized chemically. Order from a commercial vendor or use an in-house DNA synthesis facility. A fifth port is needed for the synthesis of a complete random stretch to connect a vial with a premade amidite mixture in the molar ratio of dA:dC:dG:dT = 3 : 3 : 2 : 2, correcting for differences in coupling efficiencies [95]. If a partially biased library derived from one particular sequence is preferred, for example, to explore the sequence space around a given RNA motif for optimization of binding, two strategies can be followed: (i) dope the dA-port vial with small amounts of one, two, or all three of the other nucleotide amidites, proceed accordingly with the other ports, and run synthesis of the parental sequence [96]. This requires either an eight-port synthesizer or two changes of vials during synthesis, from the constant region to partial random and from the partial random to the second constant region. (ii) Alternatively, use a mutagenic PCR (Protocol 7). Purify the DNA product by denaturing polyacrylamide gel electrophoresis (dPAGE) (Protocol 10) or HPLC after complete deprotection. The concentration of the initial pool can be determined by UV spectroscopy on a NanoDrop or other spectrophotometer at 260 nm. The OD₂₆₀ value is multiplied with a conversion factor of 33 ng μl⁻¹, which is an average for ssDNA oligonucleotides [97]. For a 40 nt random pool, it is recommended to start chemical synthesis in a 1000 nmol scale to yield approximately 5 nmol of purified ssDNA (corresponding to 3×10^{15} different sequences).

Characterization of the Initial Pool

The complexity of the initial pool is verified by primer extension with a radio-labeled primer, following the RT protocol (Protocol 7.1) and including 0.5 mM of one ddNTP at a time in four parallel reactions, or by RNA sequencing with

RNase T1 (Chapter 9). An equal distribution of all four nucleotides in the random sequence tract is expected and serves as reference to monitor enrichment during selection rounds. Additionally, cloning members from the initial library (analyze about 20 clones) or deep sequencing should provide a set of completely unrelated sequences.

Large-Scale PCR to Prepare the Initial Pool

For transcription by T7 RNA polymerase, at least the promoter region of the ssDNA obtained by chemical synthesis needs to be converted to a double-stranded form [98]. A large-scale PCR will produce fully double-stranded DNA, eliminate sequences non-amplifiable due to synthesis artifacts or incomplete protection group removal, and increase copy number per sequence. For aptamer selections with a modest complexity of the library (about 40 nt random region), complete coverage of theoretical sequence space (sequence space $N = 4^n$ with n equal to the number of randomized positions; for 40 nt random region: 1.2×10^{24} possible sequences) is impossible, but not essential for the isolation of functional molecules. For most selections between 10^{13} and 10^{15} different sequences suffice. Accordingly, between 16 pmol and 1.6 nmol ssDNA should be amplified in the large-scale PCR. However, if specialized strategies aim at including all possible variants, which is feasible in the case of short randomized stretches, calculate the required amount of ssDNA using in PCR for full coverage of sequence variants by dividing the number of sequences within the library (sequence space $N = 4^n$ with n equal to the number of randomized positions) by Avogadro's number. Three to five times this amount should then be produced in the initial PCR to ensure that no variant gets lost. In some cases, PCR reactions were even up-scaled in such a way that manual cycling in water baths was required when reaction volumes exceeded thermocycler capacities. Additionally, high-throughput thermal cyclers can be used for large-scale reactions in 96-well plates [99].

- 1) For a standard library, follow Protocol 7.2 to set up 100 μ l PCR reactions in parallel with 1–8 pmol ssDNA, limited to a maximum of eight cycles. The optimal amount of PCR cycles can be determined using a cycle course reaction. It is also recommended to test different template as well as primer concentrations and annealing temperatures for efficient PCR. Analyze the outcome on an agarose gel. If insufficient amounts of DNA are produced, change PCR conditions or continue with more PCR reactions.
- 2) Combine reactions, mix and purify DNA by standard phenol/chloroform extraction and EtOH precipitation with 2.5–3 vol% ice-cold EtOH. If precipitation is insufficient, add 1:10 volume of 3 M NaOAc (higher salt concentration can result in inefficient T7 transcription) and glycogen (1 mg) as carrier. Pool all fractions and aliquot the PCR products in 20 μ l fractions for storage at -20°C .
- 3) Use approximately 5 μ l to test for RNA transcription (Protocol 3).

Protocol 2: Preparing the Target

Preparation of a Protein Target

The reader is referred here to excellent monographs addressing the problems and pitfalls of protein purification in much detail [100]. In short,

- 1) Clone the target protein into an expression vector providing a C- or N-terminal tag for purification, and transform an expression strain such as *E. coli* BL21(DE3) with the plasmid.
- 2) For testing optimal expression and purification conditions monitor growth of two 50 ml Luria-Bertani (LB) cultures until an OD_{600nm} of about 0.6 is reached and induce target protein expression in one flask by adding isopropyl-beta-D-thiogalactopyranoside (IPTG) to 1 mM. Continue growth monitoring over 3–4 h by taking 5 ml aliquots of induced and non-induced cultures at different time points for protein expression analysis. Harvest cells by centrifugation at 3000 g, remove medium, and store pellets at –80 °C until use.
- 3) Analyze expression by SDS-PAGE after boiling cells directly in SDS sample buffer or by using a small-scale cell disruption protocol (see comment 2), which allows to determine protein localization in the soluble fraction and, after SDS extraction, also in the pelleted material.
- 4) Test purification with the small-scale protein preparation. For a His₆-tagged target, wash 50 µl Ni-NTA-Agarose (Qiagen) slurry three times with 500 µl protein buffer, add 50 µl protein solution, incubate 60 min on ice with slow agitation, and recover the supernatant. Then wash three times with 500 µl protein buffer containing 20 mM imidazole and elute twice with 50 µl protein buffer containing 250 mM imidazole. Use 20 µl for SDS-PAGE.
- 5) Apply the identified optimal growth conditions to prepare target protein in large scale from a 2 to 10 l culture. If protein purification using the affinity tag is inefficient, do an anion-exchange chromatography step before affinity purification, followed by a final chromatographic step such as gel filtration. But note that when the more purification steps are performed, the more protein gets lost.
- 6) Determine concentration and, if possible, activity of the recombinant protein, which should be unaltered in the presence of the tag. Purity should be as high as possible and can be monitored by overloading an SDS-PAA (polyacrylamide) gel or by using sensitive staining methods such as silver staining [101]. Dialyze the purified protein into a suitable protein storage buffer, ideally the selection buffer, and supplement with glycerol to 50% (v/v) for storage at –20 °C.

Preparation of a Carbohydrate Target

- 1) Test the purity of targets, obtained from a commercial supplier, isolated from natural sources, or synthesized chemically, using for example, HPLC and/or ESI- or Maldi-MS. Further purify if needed, as each contaminant increases the risk to drive selection in an unwanted direction.

- 2) Prepare the carbohydrate for immobilization, for example, by tethering it to a spacer molecule containing an activated thiol, which in turn is used for disulfide coupling of thiopropyl–sepharose (Protocol 6). Spacer attachment is rather straightforward, particularly if primary amino functions are available as in the case of aminoglycosides. Detailed chemistry is beyond the scope of this chapter, but is well documented in the literature [78].

Solutions

- 1 x SDS sample buffer: 65 mM Tris/HCl pH 6.8, 10% glycerol, 2.3% (w/v) SDS, 5% (v/v) β -mercaptoethanol, 0.23% (w/v) bromophenol blue.
- 1 x Protein buffer: 50 mM Tris/HCl pH 8.1, 5 mM EDTA, 5% glycerol, 1 mM β -mercapto-ethanol, 20 μ M proteinase inhibitor phenylmethylsulfonyl fluoride (PMSF); the latter might modify amino acid side chains and is required only sometimes in early steps to irreversibly block serine-type proteases.

Comments

- 1) Generally, the target protein, when overexpressed in *E. coli* strains, should be visible as a strong or even predominant band in crude lysates. If not, try protein enrichment by affinity chromatography *via* its tag, although binding to the matrix might be impaired due to the complex composition and high total protein concentration in crude lysates. Alternatively, modify the expression protocol by reduction of growth temperature, variation of inducer concentration, use another strain for overexpression (e.g., *E. coli* BL21(DE3)pLys or Arctic Express), by changing the tag position, or even by employing a different expression construct. Note that recombinant target proteins can be detected with tag-specific antibodies, such as anti-His₆-antibodies (Sigma, Qiagen).
- 2) Cell disruption is a critical step since the target protein may be damaged during this process. Work on ice or in a coldroom and as fast as possible. For small-scale protein isolation, dissolve the cell pellet in 50 μ l Protein buffer, transfer to an Eppendorf tube, and add step-by-step 10 μ l 100 mg ml⁻¹ lysozyme, 20 μ l 50 mg ml⁻¹ sodium-deoxycholate, 10 μ l 100 mg ml⁻¹ DNase, and 20 μ l 1 M MgCl₂, followed by 10 min of incubation at 20 °C (room temperature), during which viscosity decreases, and final centrifugation for 15 min at 4000 g to remove cell debris. The supernatant is used directly for functional tests or analysis by SDS-PAGE. For large-scale preparations, cell disruption by nitrogen decompression (manufactured by Parr bomb; www.parrinst.com) is recommended, since the technique is simple, fast, and avoids oxidation as well as warming of the sample. Alternatives include the French press, microfluidizer, or ultrasound treatment often combined with lysozyme treatment, each followed by centrifugation steps at 30 000 and 100 000 g.

Protocol 3: *In vitro* Transcription Using T7 RNA Polymerase

Setup

		Final concentration
10 μ l	800 mM Tris/HCl pH 7.5	80 mM
10 μ l	220 mM $MgCl_2$	22 mM
4 μ l	100 mM ATP	4 mM
4 μ l	100 mM CTP	4 mM
4 μ l	100 mM GTP	4 mM
4 μ l	100 mM UTP	4 mM
10 μ l	10 mM Spermidine	1 mM
1.5 μ l	100 mM DTT	1.5 mM
10 μ l	1.2 mg/ml BSA	0.12 mg ml ⁻¹
0.5 μ l	40 U μ l ⁻¹ RNasin (Promega)	20 U
2 μ l	T7 RNA polymerase (NEB)	100 U
2 μ l	Thermostable inorganic pyrophosphatase (NEB)	4 U
5–10 μ l	DNA template	~0.125 μ M
to 100 μ l	With double-distilled H ₂ O	

For radioactive labeling, supplement the above reaction with 1 μ Ci α -[³²P]ATP (one of the other three NTPs can also be used for labeling), but maintain the “cold” ATP concentration at 4 mM for the sake of high transcription yields.

Procedure

- 1) Incubate 1.5–6 h at 37 °C; prolonged incubations may result in RNA degradation. Withdrawal of aliquots at different time points permits to analyze the course of RNA production and to adapt reaction conditions accordingly.
- 2) Add 1–2 μ l DNase I (molecular biology quality, RNase-free; Roche or NEB, etc.) together with 10 μ l 10 x DNase I buffer and continue incubation for 10–20 min at 37 °C to avoid amplification of contaminating DNA template rather than functional RNA in subsequent RT–PCR reactions.
- 3) Add 11 μ l 0.5 M EDTA, pH 7.7 (final concentration 5 mM) to chelate Mg^{2+} and dissolve precipitated Mg^{2+} salts. For inactivation of DNase I, incubate the mixture 10 min at 75 °C.
- 4) At this point, samples may be either shock-frozen in liquid nitrogen or a mixture of dry ice/isopropanol for storage at –20 °C, or purified directly by 8–10% dPAGE (Protocol 10). For large-scale transcription reactions and before dPAGE purification, extract once with 100 μ l phenol to remove protein and three times with 100 μ l ether or chloroform, followed by ethanol precipitation to concentrate the RNA (Protocol 4, step 6). Ether is preferred over chloroform due

to its higher volatility and since it partitions above the aqueous phase, allowing for easy removal without changing tubes. Alternatively, for concentration of RNA ultrafiltration may be employed, for example, using a microcon device (Millipore).

Comments

- 1) It is recommended to test conditions for RT–PCR and transcription before the first round of selection. Prepare a master mix of all components besides the enzyme for PCR and transcription each and aliquot the mixture in appropriate volumes.
- 2) Note that the real concentration of the dsDNA template in the PCR mixture could not be determined because it still contains NTPs and primer. Therefore, NTPs and primer have to be removed if the exact concentration of the template DNA has to be measured.
- 3) For transcription, the Mg^{2+} concentration should be 2–4 mM above the total NTP concentration to provide enough Mg^{2+} ions for RNA structure formation, NTP binding, and polymerase catalysis, while remaining below inhibitory levels, requirements usually fulfilled for Mg^{2+} concentrations of 14–22 mM.
- 4) An inorganic pyrophosphatase should be added to remove pyrophosphates resulting from transcription, which have an inhibitory effect on T7 RNA polymerase.
- 5) Minor 3'-end heterogeneities generated by T7 RNA polymerase (Chapters 1 and 2) are generated especially during long incubation, but usually are not critical for aptamer selections, as termini lie within the constant primer binding sites that are restored at the end of each selection round. For more details on T7 transcription see Chapter 1.
- 6) The T7 RNA polymerase can be overexpressed in *E. coli* and purified in the laboratory itself. Note that the enzyme preparation can be contaminated with RNases carried over from the purification process. Therefore, it is strongly recommended to test the preparation before use or to employ a commercial RNase inhibitor.

Protocol 4: Selection against Immobilized Protein Targets

Procedure

- 1) Prepare 300 μ l Ni-NTA-Agarose slurry (Qiagen) in a reaction tube by washing twice with 1000 μ l and resuspension in 300 μ l selection buffer.
- 2) Transfer one-third of the slurry to a fresh tube, add sufficient protein to saturate binding sites and incubate 1 h on ice. Keep the supernatant as well as the solutions of three washes with 500 μ l selection buffer to determine the fraction of bound protein. If the supernatant contains no protein, repeat the procedure to ensure saturation of the affinity material. Keep the matrix with bound protein on ice until use.

- 3) Denature 1.5 nmol of the ^{32}P -labeled initial RNA library (100 000–500 000 cpm) in 80 μl 10 mM NaCl by incubation for 3 min at 80 °C, place the tube in a styrofoam rack at room temperature (or any other temperature) for controlled refolding over a period of 10 min, and add 20 μl 5 x selection buffer; for preselection to remove matrix binders, combine this solution with the two-third of washed agarose slurry and incubate for 10 min. Rotate or shake the tubes to avoid sedimentation of the agarose.
- 4) Transfer the supernatant or flow-through (when using columns) obtained from the preselection step to the one-third of agarose slurry with bound target protein for affinity binding, incubate for 10 min, remove the supernatant or flow-through, and wash 3 x with 100 μl selection buffer. Keep all fractions for analysis to monitor enrichment.
- 5) Elute bound RNA by incubation with 100 μl 250 mM imidazole pH 7 in selection buffer, displacing the entire RNA–protein complex; alternatively, perform affinity elution with a more than 10-fold excess of free over immobilized protein, which increases the stringency since only molecules binding to immobilized as well as free target will be eluted.
- 6) Prepare the eluted RNA for RT-PCR reactions and storage as follows: extract once with phenol and three times with ether, add 1/10 vol 3 M NaAc pH 5.6 and 2.5 vol 100% EtOH to the aqueous phase, and place the tube at –20 °C for at least 2 h or overnight. For high-salt and low-volume samples, add one to two volumes of double-distilled H_2O (dd H_2O) before extraction to avoid phase separation problems. Remove the supernatant after centrifugation for 30 min at 4 °C and 10 000–15 000 g, briefly dry the pellet for 5 min in a speed-vac or vacuum desiccator, and redissolve in 40 μl dd H_2O by leaving the sample for 10 min at room temperature. Avoid more extensive RNA drying, since dehydrated RNA cannot be redissolved without disintegration. Shock-freeze the RNA solution in liquid nitrogen or in a dry ice/ isopropanol bath and store at –20 °C.

Buffers

- 1 x Selection buffer: 50 mM Tris/HCl pH 7.5, 100 mM NaCl, 7 mM MgCl_2 .

Monitoring Enrichment

Quantify the RNA contained in supernatant, eluted, and wash fractions, as well as that retained on pre- and selection matrices; measure cpm prior to preselection as reference to calculate the expected 100% level. This is achieved by Cerenkov counting of the entire reaction tube in a liquid scintillation counter or by spotting an aliquot onto a 3MM Whatman filter, followed by liquid scintillation counting. In the beginning, the fraction of target-bound RNA may even fall below 1%, but depending on the target a steady increase should be visible after rounds 3–7. Adjust the selection strategy accordingly, for example, by including a second preselection step if matrix rather than target binding increases. Likewise, performing more extensive washing during selection or reducing the concentration of target for

elution would increase stringency and forces enrichment of high-affinity binders. Monitor the decrease in library complexity during enrichment as mentioned in Protocol 1.

Comments

- 1) Handle agarose matrices with care. Use yellow or blue tips for pipetting, cut 5 mm above the tip with a sterile scalpel to widen the opening for the reduction of shearing forces. Centrifuge with low speed to avoid damage of the matrix resulting in leakage of bound material. Use of low-binding tubes reduces surface adsorption. Spin columns work equally well and are commercially available for self-packing. Alternatively, a simple RNase-free two-part system is rapidly assembled and works equally well: with a thin razor blade, make a small vertical cut into the bottom of a small polypropylene reaction tube (e.g., 400 μ l tube #7518.1 Roth, Germany); fill in the agarose slurry and insert the tube into a conventional 1.5 ml Eppendorf tube whose closed lid has been briefly touched with the heated end of a Pasteur pipette to create an opening just fitting the small tube in diameter. By using a needle, jab a second hole in the 1.5 ml tube's lid to allow air circulation and collect all liquid at the bottom of the large tube after gentle centrifugation at 1000 *g*.
- 2) Selection conditions must be adapted to each target and selection strategy. The buffer should provide enough ionic strength to reduce non-specific adsorption and contain Mg^{2+} ions for RNA structure formation. Moreover, varying salt and Mg^{2+} concentration offers a possibility to adjust selection stringency. Lowering the Mg^{2+} concentration may disrupt more labile three-dimensional RNA structures. Inclusion of unspecific competitors (any other natural RNA such as tRNA) should drive the selection to more specific aptamers. Likewise, specific competitors (natural ligands or binding partners of the target) are expected to introduce a bias toward high-affinity binders or may have the effect of restricting the number of aptamer-accessible epitopes on the protein. High pool to target concentration ratios carry the danger of eliminating many binders in early rounds and increase the risk to select matrix binders. Better start with a low ratio, and increase stringency by gradually reducing the number of available target molecules from one selection round to the next.
- 3) It is recommended to determine the amount of RNA binding unspecifically to agarose and tube material before selection. Furthermore check for precipitation of RNA and target under chosen buffer conditions because especially at high concentrations RNA tends to precipitate. Therefore, Prepare two probes: one contains RNA and the target molecule in selection buffer, the second includes RNA without target as control. Incubate both under selection conditions, centrifuge for 10 min at 10 000–14 000 *g* and take off exactly one-half of the total volume of each tube immediately and transfer it into a new tube. Measure the amount of RNA in all tubes by scintillation counting. If no precipitation occurs, the cpm values of both tubes of one probe have to

be nearly identical. If precipitation or high adsorption of RNA are observed, adjust the buffer composition. Try to increase the salt concentration first.

- 4) Note that, on the one hand, agarose or sepharose as affinity matrices exhibit high immobilization levels, but, on the other hand, they possess a high dead volume and volume to surface ratio causing dilution effects of RNA during washing and increasing the probability to select matrix binders. *In vitro* selection with magnetic beads [102] and filters (Protocol 5) offer a convenient alternative with the advantage that preselection usually is not necessary.

Protocol 5: Selection by Filter Binding of Protein–Aptamer Complexes

Procedure

- 1) Follow Protocol 4, step 3, for denaturation and refolding of the library. Incubate library and target to form target–protein complexes.
- 2) Filter the target–library mixture over nitrocellulose (NC) filters as described in Protocol 12.
- 3) Elute retained material by submerging the filter in 500–1000 μl elution buffer in a reaction tube, place 5 min at 95 °C, and remove the supernatant; add another 500–1000 μl elution buffer and repeat the heating step. Combine both elution fractions, add glycogen to 60 $\text{ng } \mu\text{l}^{-1}$, extract with phenol/ether, and recover the RNA by ethanol precipitation (Protocol 4.1). Continue with RT-PCR (Protocol 7).

Buffers

- 1 x Selection buffer: 50 mM Tris/HCl pH 7.5, 100 mM NaCl, 7 mM MgCl_2 .
- 1 x Elution buffer: 7 M urea, 50 mM Hepes/NaOH pH 7.5, 10 mM EDTA.

Comments

- 1) Ensure that the filter has the capacity to bind all complexes by filtering the same amount of target protein as used in the selection step and quantifying the amounts of protein retained on the filter and in the flow-through. This control experiment will also reveal if the target protein is only inefficiently retained on NC filters despite sufficient binding capacity of the membrane. In such cases, NC filter binding would be inappropriate for the selection assay.
- 2) Conceivable alternative elution and recovery methods are (i) incubation of the filter in 4 M guanidinium thiocyanate for 15 min at 75 °C, (ii) proteolytic degradation of the target protein, or (iii) dissolving the entire filter in a suitable organic solvent such as acetone, followed by the addition of an aqueous phase and phenol/ether extraction.

Protocol 6: Selecting against Carbohydrate Targets

Target immobilization is achieved, for example, by coupling an activated carbohydrate to thiopropyl-sepharose *via* a disulfide bond. This coupling is position

specific and allows reversible elution of RNA–target complexes by reducing agents, for example, DTT (1,4-dithiothreitol) or 2-mercaptoethanol. Both features are of advantage compared with coupling to epoxy-activated sepharose, for example, employed in selections against tobramycin and neomycin [103]. However, note that the disulfide bond between target and matrix is less stable than a covalent linkage, which is performed using epoxy-activated or EAH sepharose. As described for protein selection, a variety of chemically modified magnetic beads are available, which can be used alternatively.

Procedure

- 1) Swell 2 g of activated thiopropyl-sepharose 6B matrix (Amersham Biosciences) by incubation for 1 h in degassed 1 x activation buffer supplemented with 20 mM DTT. Monitor the release of 1-thiopyridone *via* absorption at 342 nm as the reaction progresses, which correlates with the formation of free SH groups (the binding capacity of the sepharose). Wash the matrix extensively over a sintered glass filter until absorption between 190 and –500 nm is no longer detectable. Split the matrix in three aliquots.
- 2) Couple the thiopyridyl-activated carbohydrate target to one-third of the matrix by overnight incubation at 4 °C in 1 x activation buffer. Monitor again the coupling process and thus substitution level (amount of immobilized target) *via* release of 1-thiopyridone (change in OD_{342nm}). Wash as described above. Use the remaining two-third of the matrix to prepare the preselection material, either by coupling the spacer employed to activate the carbohydrate target (washing procedure as above) or by straight deactivation. Deactivation is achieved by reacting the free SH groups with 500 mM iodoacetate in 100% EtOH at room temperature for 1 h. Wash extensively with 100% EtOH. Store reduced and modified sepharose material under 5% (v/v) EtOH in H₂O or buffer at 4 °C. No loss of activity is observable over 10 weeks or even longer storage periods.
- 3) Use the preselection and carbohydrate-coupled matrices as described for the protein selection according to Protocol 4. Bound RNA is recovered by affinity elution with free carbohydrate, by reduction and release of the carbohydrate-RNA with 20 mM DTT in selection buffer, or by adding the entire column material directly to the RT reaction (Protocol 7).

Buffer

- 1 x Activation buffer: 100 mM Tris/HCl pH 7.5 at 25 °C, 250 mM NaCl, 1 mM EDTA.

Comments

- 1) Handle agarose matrices with care (Protocol 4).
- 2) Embedding a carbohydrate moiety within a network of modified agarose requires stringent pre- and counter-selection techniques to enrich aptamers against the target. Alternate use of this and other immobilization surfaces or

an entire switch to other matrices, such as glass beads or microtiter plates, may mitigate the problem of matrix binder enrichment. However, available data are too limited to recommend such alternative surfaces.

Protocol 7: Regeneration of the Enriched Library

Reverse Transcription of the Selected RNA

Set up a 20 μl reverse transcription reaction:

		Final concentration
10 μl	10 μM Primer A	2.5 μM
23 μl	Selected RNA library in H_2O denaturing: 5 min, 65 $^{\circ}\text{C}$ reannealing	
10 μl	5 x first strand buffer (Invitrogen)	1 x
2 μl	100 mM DTT	4 mM
1 μl	dNTP mix (4 mM each) 2 min, 42 $^{\circ}\text{C}$	0.5 mM each
1 μl	Superscript II reverse transcriptase (Invitrogen) 60 min, 42 $^{\circ}\text{C}$	200 U

- 1) Add primer A (Figure 50.1) to the selected RNA library, denature for 5 min at 65 $^{\circ}\text{C}$, and anneal the primer by placing the reaction tubes for 15 min on ice.
- 2) Supplement the additional components and incubate for 60 min at 42 $^{\circ}\text{C}$. As in contrast to PCR, only one round of elongation occurs, and optimal primer design and annealing conditions are of particular importance.

PCR to Amplify the DNA Library

Set up a 100 μl PCR reaction

		Final concentration
5 μl	10 μM Primer A	0.5 μM
5 μl	10 μM Primer B	0.5 μM
$\times$ μl	DNA library as template or cDNA	≈ 1 fM
2 μl	dNTP mix, each dNTP 10 mM, pH 7.0	0.2 mM
10 μl	10 x PCR buffer	
0.4 μl	5 U μl^{-1} Taq polymerase with dd H_2O to 100 μl	2 mU

- 1) Set up the reaction, always adding template last to avoid cross-contamination. For RT-PCR of early selection rounds, use the half of the RT reaction mix as template for PCR and omit primer A as it is provided in sufficient amounts

with the RT reaction mix. In later selection cycles, the RT reaction mix may make up one-fifth of the PCR reaction volume.

- 2) Start cycling as exemplified for a library with 40 randomized positions we have used for protein and carbohydrate selections. The cycling program must be adjusted according to template and PCR performance in each selection round.

Step 1: 95 °C	1 min (denaturation, use 5–10 min in case of plasmid DNA)
Step 2: 95 °C	30 s (denaturation)
Step 3: 55 °C	30 s (annealing)
Step 4: 72 °C	20 s (polymerization) – go back to step 2, run 5–20 cycles
Step 5: 72 °C	5 min
Step 6: 4 °C	Storage until further processing

- 3) Analyze PCR progress on 2% agarose gels by taking a 5 µl aliquot every fifth PCR cycle to avoid the formation of artifacts indicated by multiple products. Before starting a new selection round, at least a 10 µl aliquot is stored at –20 °C as backup.
- 4) Recover amplified DNA by 1 x phenol/ 3 x ether extraction and ethanol precipitation as described in Protocol 4 (step 6) or use a membrane device (e.g., mini spin column, NucleoSpin Extract II, and Macherey-Nagel).

Buffer System

Usually buffers are supplied with purchased enzymes. Basic compositions that work well are as follows:

- 1 x RT buffer: 50 mM Tris/HCl pH 8.8 (25 °C), 50 mM KCl, 6 mM MgCl₂.
- 1 x PCR buffer: 10 mM Tris/HCl pH 8.8 (25 °C), 50 mM KCl, 1.5 mM MgCl₂.

Comments

- 1) To avoid contaminations, use filter-containing tips for pipetting and work at a freshly cleaned place or, if available, under a laminar flow cabinet or other sterile work bench. Include a negative control (omission of template), as well as a positive control using a different template resulting in a product of similar size as that expected for amplification of the library. Native 8% acrylamide gels provide better resolution than agarose in the case of blurred and diffuse PCR bands.
- 2) To test whether eluted RNA is contaminated with DNA (caused by inefficient DNase I treatment after transcription), omit RT and use a small aliquot of eluted RNA directly for PCR.
- 3) We observed that high amounts of DTT inhibit RT-PCR. If so, purify eluted RNA by performing ultrafiltration using Microcons, for example.

- 4) For the selection process, we prefer Taq Polymerase over high-fidelity polymerases. Alternatively, DAp Goldstar® Polymerase (Eurogentec) can be used, which provides high yield and specificity. Accordingly, SuperScript Reverse Transcriptase (Invitrogen) offers an alternative to Avian Myeloblastosis Virus (AMV) reverse transcriptase.
- 5) It is recommended to avoid overamplification of the DNA pool not to accumulate side-products and distinct polymerase-preferred sequences. Therefore, determine the number of cycles needed to yield a visible band in the agarose gel previously to PCR in a larger scale. Alternatively, an emulsion PCR can be performed.
- 6) Mutagenic PCR offers a fast possibility to screen sequence space around a given sequence (theoretical background discussed by Knight and Yarus [87]). Either adjust the PCR reaction to special salt conditions that reduce the fidelity of the polymerase [104, 105] or dope the dNTP mixture with mutagenic nucleotides that have an extended base-pairing potential [81]. Run three to six rounds of reselection with this mutagenized library using the initial selection setup.

Protocol 8: Determination of Pool Enrichment during Selection

The success of a selection is often not predictable due to the enormous quantity of parameters impairing the outcome. Hence, it is important to control the progress of a selection providing the opportunity to influence its direction. To track the enrichment of molecules, the amount of eluted RNA may be measured in each selection round. Another possibility to control the selection progress is the continuous determination of the diversity of the pool. If specific sequences get enriched during selection, the probability that these sequences form again double strands after denaturation will be higher than within a complete diverse pool. Therefore, the S1-Nuclease that specifically cleaves ss DNA is used to discriminate between double- and single-stranded DNA. For experimental procedure, the RT-PCR products yielded after each selection cycle are denatured, reannealed, and subsequently treated with S1 nuclease. After gel electrophoresis, one can determine the fraction of DNA that was not cleaved, suggesting that this DNA must be double-stranded. If selection results in an enrichment of distinct sequences, the amount of uncleaved double-stranded DNA must increase.

Procedure

- 1) Remove nucleotides and primer from PCR product by performing gel extraction (Protocol 9) or ultrafiltration using microcon devices (Millipore). Alternatively, PCR can be purified using magnetic beads (Agova).
- 2) Use 200 ng DNA in 1 x S1-buffer and incubate for 3 min at 98 °C for complete denaturation. Reduce temperature for 5 min to approximately 65 °C for reannealing. Add 1 U μl^{-1} S1 nuclease and incubate 30 min at 65 °C. Add 50 mM EDTA for the termination of S1 digest. Finally, analyze the

products on an agarose gel using exactly the same amount of DNA in each slot.

Comments

- 1) Purification of the PCR products using ultrafiltration and magnetic beads is less time consuming than employing gel extraction, and higher amounts of pure DNA are yielded.
- 2) The temperature for reannealing has to be adjusted according to one's own pool depending on pool length and base composition.
- 3) It is important to use exactly equal amounts of DNA for gel analysis to compare band intensities of the PCR products after each selection round providing information about the fraction of double-stranded DNA. Therefore, it is recommended to run the diversity assay after performing a few selection cycles for direct comparison of band intensity and diversity.
- 4) It is also suggested to include a control with DNA that is not treated with S1-nuclease and exhibits a band with completely intact DNA, simplifying quantitation and comparison with enriched DNA pools.
- 5) An alternative to the S1-nuclease treatment for the determination of pool diversity is offered by the real-time PCR technique. Using this method, it is recommended to additionally use a diversity standard consisting of sequences with different numbers of randomized positions.

Protocol 9: Isolation of Individual Aptamers from Enriched Pools

This is the only step in the whole process that requires *in vivo* techniques.

Cloning the Library into a Vector

- 1) Load RT-PCR-amplified pool DNA (Protocol 7) onto a 2% agarose gel, excise the corresponding PCR product after staining, and extract it by passive diffusion or by using a gel extraction kit such as QIAEX II (Qiagen). This step is not necessary if one have one distinct band in the agarose gel corresponding to the PCR product of the enriched pool or if shorter and longer sequences should additionally be cloned (comment 2).
- 2) The purified PCR products are easiest cloned into a linearized vector with single 3' thymidine overhangs (e.g., the pGEM-T vector, Promega or TOPO TA Cloning®, Invitrogen), since Taq polymerase adds a single deoxyadenosine, in a template-independent manner, to the 3' ends of amplified fragments. Incubate about 2 ng of the eluted DNA, 16 ng vector, and 400 U T4 DNA Ligase in a 10 µl reaction for 16 h at 4 °C. Alternatively, use the restriction sites included at the end of the primers for conventional cloning (Protocol 1).
- 3) Add the 10 µl reaction to 300 µl competent cells in a reaction tube, incubate for 30 min on ice, then for 1 min at 42 °C, place immediately on ice, add 1 ml LB medium without antibiotic, and incubate 60 min at 37 °C. Harvest bacteria

by gentle centrifugation (1000 g), resuspend in 100 µl LB, and plate on a LB agar plate with suitable antibiotic for overnight growth at 37 °C.

Isolation of Individual Aptamer Sequences

- 1) Ideally, apply blue-white screening to identify colonies with inserts, using an appropriate cloning vector such as pGEM-T or pCR2.1-TOPO. Additionally, colony PCR allows to determine insert size as outlined below. Pick 40 white colonies from the transformation plate; for this purpose, dip with a sterile yellow tip into a colony, then touch with the tip a new LB-antibiotic plate placed on a numbered square grid for replica plate growth and finally inoculate 3 ml LB-antibiotic broth by ejecting the tip into the medium. Incubate agar plates and cultures overnight at 37 °C.
- 2) The replica plate is used the next morning to run colony PCRs; to collect rather constant amounts of cell material, stab a pipette tip vertically into the colony, and wash the bacteria directly with 30 µl ddH₂O by pipetting up and down. After incubation at 95 °C for 15 min, use 5–15 µl of lysed cell material to run the PCR as described in Protocol 7.2.
- 3) If clones contain an insert of expected size, 300 µl of the corresponding overnight culture (step 1) is used to prepare glycerol stocks by adding the same volume of 75% (v/v) sterile glycerol for storage at –80 °C, while the remaining 2.7 ml is used for plasmid preparation and subsequent sequencing. Keep 100 ng of plasmid DNA as backup for later transformations. Alternatively, the purified colony PCR products can be sequenced directly without performing a plasmid preparation. To remove salts, nucleotides and primer high-throughput PCR purification are offered in 96-well formats (Qiagen). However, note that in this case not the plasmid but the bacterial colony serves as backup.

Comments

- 1) Always perform control transformations with intact pGEM-T or other high copy vectors such as pUC18 (>2000 colonies expected) to check if cells are competent.
- 2) Insert length may vary, as internal deletions will produce shorter sequences. These clones likely contain active aptamers and may provide important information about binding motifs and their length.
- 3) Since cloned inserts include the T7 promoter, prepared plasmids can be used directly for *in vitro* run-off transcription after linearization at a 3' terminal restriction site, either within the insert as part of the primer sequence or provided by the vector; however, ensure that the plasmid itself does not contain a T7 promoter to avoid side-products during transcription. Alternatively, the insert DNA may be reamplified by PCR with about 1 ng of plasmid as template following Protocol 7.2.
- 4) Note that deep sequencing offers an appropriate alternative with the advantage that cloning and isolation of individual aptamers is not necessary.

Protocol 10: Polyacrylamide Gel Electrophoresis (PAGE)**Procedure**

Depending on the experimental question, different setups are recommended:

	Size	Acrylamide (%)	Electrophoresis
Analytic dPAGE ^a	20 × 20 × 0.02 cm	8–20	1000 V/1–1.5 h
Preparative dPAGE	20 × 20 × 0.1 cm	8	500 V/2–4 h
Native PAGE	20 × 20 × 0.02 cm	8	250 V/2–4 h ^b

^adPAGE: denaturing polyacrylamide gel electrophoresis.

^bNative PAGE requires additional cooling during electrophoresis (Protocol 11).

- 1) Prepare solutions of the desired acrylamide concentration by mixing acrylamide with urea stock solutions or use premixed stock solutions (Rotiphorese® Sequenzierringel, Roth).
- 2) For analytic gels, assemble the sandwich of glass plates, clamp together with ear-plate up, tilt slightly, and allow the gel solution to enter slowly across the entire width between the glass plates by capillary forces. Insert comb and place horizontally until polymerized. Thicker preparative gels require a bottom spacer and are prone to leak at the corners, which can be prevented by polymerizing a few milliliters of the acrylamide solution barely visible at the bottom and adding the rest once the bottom layer has solidified.
- 3) Run a pre-electrophoresis to remove charged molecules left from the polymerization reaction, such as ammonium peroxodisulfate. For dPAGE, at least 30 min at 30 W is recommended to also warm up the gel.
- 4) Rinse the pockets well, especially for dPAGE, add loading buffer to samples, and run electrophoresis until the dye markers indicate sufficient resolution in the separation range of interest. For dPAGE, the gel temperature should not exceed 60 °C, since glass plates might break at higher temperatures.
- 5) Preparative amounts of RNA can be visualized directly by UV shadowing (254 nm) as described in Chapter 3. Briefly cut the edges of the band under UV light with a sterile razor blade or scalpel to mark their position, but fully excise the gel piece in the absence of UV light to avoid damage of RNA. Store as such at –20 °C or start elution by crushing the gel piece with a yellow tip, adding 0.3 M NaOAc pH 5.6 until the gel material is covered, incubate 30 min on ice, shortly centrifuge in a desktop centrifuge, recover the supernatant, and add another fresh aliquot of 0.3 M NaOAc as above. After three repetitions, usually 80% of the RNA has been eluted. Alternatively, elution can be carried out over night at 4 °C and with a second elution step afterward for 30 min at 37 °C under rotation. Precipitate RNA by adding 2.5 vol EtOH.

Solutions

- 20% Acrylamide stock solution: 19% (w/v) acrylamide, 1% (w/v) bisacrylamide, 8 M urea in 1 x TBE.
- 10 x Electrophoresis buffer (10 x TBE): 900 mM Tris, 900 mM boric acid, 20 mM EDTA (pH is 8.3).
- Denaturing loading buffer: 8 M urea, 50 mM EDTA, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol.
- Native loading buffer: any incubation buffer plus 50% (v/v) glycerol.

Comments

- 1) Use premade acrylamide solutions, such as 38% (w/v) acrylamide mixed with 2% (w/v) bisacrylamide (Roth, Germany) to avoid handling of acrylamide powder. Solutions should be used up within 3–4 months, since acrylamide can deaminate to acrylic acid. Check conductivity if unsure.
- 2) Use conditions for nPAGE as described in the EMSA Protocol 11.

Protocol 11: Electrophoretic Mobility Shift Assay (EMSA)

Setup

The electrophoretic mobility shift assay is a common method to determine the binding affinity of the RNA to the target (mostly quantified by the dissociation constant K_d), but it is also used for selection especially for large nucleic acid binding proteins [106, 107]. Native PAGE is best performed in a small (such as the Hoefer or BioRad mini systems) or medium-size (10 × 20 cm) gel system using 6–10% acrylamide gels (Protocol 10) for good separation, depending on size and charge of the complex. It is further recommended to use a buffer system without monovalent ions to reduce current and heat production and to employ a setup with temperature control to avoid irreversible complex dissociation due to overheating. The concentration of ^{32}P -labeled RNA is kept constant (low nanomolar range; 200–1000 Cerenkov cpm suffice in most cases), while the target concentration is varied in a first experiment to cover 3 orders of magnitude up to micromolar concentrations (for example, 10 nM, 100 nM, 1 μM , and 10 μM). A sharp transition should be visible, indicating the concentration range that needs to be assayed in more detail.

Sample Preparation and Assay

- 1) Denature 10 000 cpm of ^{32}P -labeled aptamer RNA in 10 mM NaCl for 1 min at 90 °C and transfer the tube to a styrofoam rack kept at room temperature; after 3 min, add incubation buffer to aid structure formation and permit refolding for another 7 min.
- 2) While renaturation is in progress, prepare for each target concentration a reaction tube on ice with buffer, protein, and additional components. Include control reactions, such as RNA without target, as well as target plus RNA

denatured before gel loading by adding 20 mM EDTA and heating for 1 min at 90 °C. Add RNA last, mix gently, and incubate for 15–60 min on ice or at any other temperature of choice.

- 3) Prerun the gel for at least 30 min. For electrophoresis at 4 °C, the use of a suitable electrophoresis apparatus directly connected to cryostat cooling the buffer reservoirs or at least an aluminum plate clamped to the front of the gel is recommended as it works better than simply performing the experiment in a coldroom. Add no more than one-fifth reaction volume of loading buffer to the samples, and rinse the gel pockets with a syringe immediately before gel loading.
- 4) Electrophoresis conditions depend on acrylamide concentration and size of the target–RNA complex. In the case of a 100 nt aptamer and a 40 kDa protein, electrophoresis for 2 h at 100 V in an 8% native polyacrylamide gel works well (Figure 50.2). After electrophoresis, expose a phosphorimager screen to the gel, provided that enough radioactivity is present to allow for a short-term exposition (2–10 h, depending on complex nature) to limit diffusion and to avoid image blurring; otherwise, dry the gel in a gel dryer for 45 min at 80 °C before phosphorimaging.

Buffers

- Gel shift buffer: 25 mM Tris-acetate pH 7.4, 5 mM magnesium acetate; well working system that is used for complex formation and electrophoresis; the pH can be raised to 8.4 to increase gel mobility, although this enhances the risk of RNA degradation. A common selection buffer can also be used (Protocols 4 and 5), especially if EMSA is employed for selection.
- Native loading buffer: Gel shift buffer plus 30% glycerol (v/v), 0.25% (w/v) each of bromophenol blue and xylene cyanol (see comments below).

Comments

- 1) EMSA experiments exploit the so-called cage effect that prevents irreversible dissociation during electrophoresis [108]. Organic dyes such as bromophenol blue have been suspected to affect complex association/dissociation equilibria. Omission of dyes from the sample buffer may improve complex stability, while the progress of electrophoresis can be monitored by loading an adjacent lane with a dye-containing buffer.
- 2) EMSA and nitrocellulose filter binding assays are best performed in parallel, using 6 µl of a 20 µl reaction setup for the native PAGE experiment and the remainder for filter binding (Protocol 12).
- 3) The EMSA technique may also be used to isolate aptamer–target complexes during selections. However, the approximate position of the shifted complex needs to be known to excise and elute the appropriate region of the gel. Therefore, EMSA is best suited as an alternate selection assay in later selection rounds when shifted bands become more prominent and discrete due to

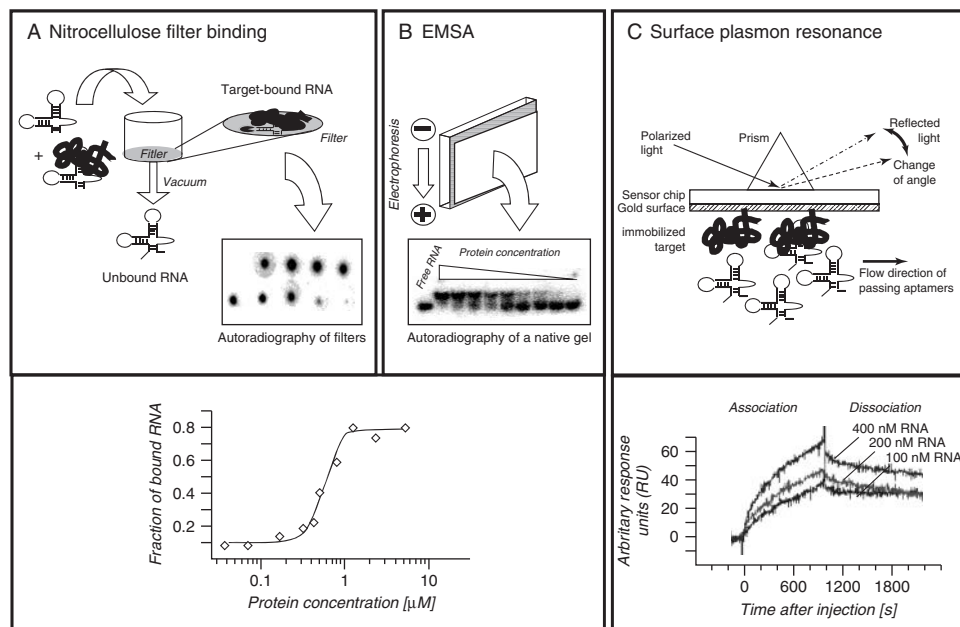


Figure 50.2 Characterization of aptamer binding. (a) In nitrocellulose filter binding assays, radioactively labeled aptamer RNA is retained on the filter *via* binding to its target protein. Trace amounts of RNA are incubated with increasing amounts of protein. Bound RNA can be quantified by phosphoimaging or liquid scintillation counting. (b) Electrophoretic mobility shift assay (EMSA) exploiting slower migration of aptamer–target protein complexes compared with free aptamer during native PAGE; decreasing amounts of protein (from left to right) are incubated with trace amounts of

radiolabeled aptamer; first lane on the left: RNA without protein. Quantitative analysis of data derived from (a) and (b) should produce binding curves of the type shown at the bottom. (c) Surface plasmon resonance (SPR) uses angular changes of reflected light on the chip surface due to mass changes, resulting in an ascending phase of the response curve during binding and a descending curve phase during washing. Binding constants can be determined based on the association and dissociation rates after injecting increased concentrations of the RNA aptamer.

enrichment of affine binders. Therefore, protein-bound RNA for the next selection round has to be recovered by elution out of the gel as described in Protocol 10.

Protocol 12: Nitrocellulose (NC) Filter Binding

Procedure

As described for EMSA, filter binding can not only be used for the measurement of binding parameters (K_d value) but also for the selection to separate unbound from target-bound RNA (Protocol 5). Various RNA aptamers were successfully selected using filter binding, summarized in [109].

Different assay formats exist, from 96-well microtiter plates to simple vacuum manifolds with round filter holders, usually 2.5 cm in diameter. The latter simple setup is used in the protocol described below:

- 1) Wet 0.45 μm nitrocellulose filters (Schleicher and Schuell) in the incubation buffer of choice for at least 5 min.
- 2) Follow EMSA Protocol 11 to prepare 10 μl samples. Ideally, perform EMSA in parallel, which allows for comparison of results.
- 3) Place a filter on the filter holder, ensuring proper and constant suction (around 300 torr using a water jet pump from Brand), prewash the filter with 100 μl buffer, then immediately pipette 4 μl of the reaction sample onto the center of the filter, and wash again with 100 μl buffer.
- 4) Remove filter carefully with flat-tip tweezers, repeat filter binding with a second 4 μl aliquot of the same sample to obtain two data points, allow filters to dry 20 min under red-light, and count by liquid scintillation or exposition of a phosphoimager plate.
- 5) Keep incubation tubes on ice until the last filter has dried, spot 1 μl of the remaining complex on a fresh nitrocellulose filter, dry without filtration, and count directly. This value is used to calculate total radioactivity in the assay.

Buffers

- Use buffers as described for EMSA Protocol 11.

Comments

- 1) Nitrocellulose is brittle, should not be touched with hands, and marked solely at the edges to avoid artifacts due to changes in binding capacity. The material will also age, particularly if exposed to sunlight, resulting in reduced binding affinity. Although in use since the 1960s and commonly applied to characterize RNA–protein interactions [110], the exact nature of the molecular principles underlying binding remains unclear. However, binding is dependent on hydrophobic interactions with the protein. RNA alone should usually bind to a much lesser extent than proteins, although nucleic acids can form hydrophobic interactions, which explains why nitrocellulose binders are often enriched during selections. It should further be noted that some proteins bind inefficiently to nitrocellulose, which may be a possible explanation in cases of failure to retain a protein–RNA complex on nitrocellulose filters.
- 2) For exact quantitation, a double-filter experiment is suggested [111], which places a second diethylaminoethyl (DEAE) filter disc (e.g., Whatman DE81 or Amersham Hybond-N+) beneath the nitrocellulose filter. By this means, all RNA material that is not retained on the top filter is trapped on the second filter, permitting to directly determine the ratio of bound to unbound RNA.
- 3) Use high precision pipettes for small assay volumes or double the assay volume to achieve more accurate pipetting.

Protocol 13: Surface Plasmon Resonance (SPR) for K_d Determination**Procedure**

SPR is an elegant technique to directly follow association and dissociation kinetics during binding events over a wide range ($k_{\text{on}} = 10^3 - 10^7 \text{ M}^{-1} \text{ s}^{-1}$, $k_{\text{off}} = 10^{-6} - 10^{-1} \text{ s}^{-1}$, $K_d = 1 \text{ mM}$ to 10 pM [112]), particularly if approaches such as NC filter binding (no protein involved, no filter retention) or EMSA (lack of charge, small size of the complex) fail. Instruments, such as the basic Biacore X and its more advanced successors [113], have increasingly come into use and detailed descriptions exist [114, 115]. SPR is even applied to select aptamers against proteins [116], although the potential for performing selection is limited. More recent application surveys are an excellent starting point to develop own experimental setups and problem solutions [117, 118].

- 1) Immobilize the target on a commercially available sensor chip, if possible, by using a coupling chemistry identical to that employed in the selection, such as a cyclodextrane-derivatized gold surface presenting nitrilotriacetic acid functionalities (Biacore). The CM5 chip is the most common sensor chip because molecules can be coupled covalently *via* amine, thiol, aldehyde, and carboxyl groups. For immobilization of His-tagged proteins, a nitrilotriacetic acid (NTA) sensor chip may be used. Apply slow flow rates ($5 \mu\text{l min}^{-1}$) for long contact times. After mounting the chip in the instrument and before sample injection, prewash the chip for 20 min with 350 mM EDTA in SPR buffer to remove interfering metals, and then 20 min with SPR buffer until a stable base line is reached. When using the NTA sensor chip for coupling His-tagged proteins, finally wash for 20 min with 500 μM NiCl_2 in SPR buffer.
- 2) Bind His-tagged protein by flushing the measurement cell only with 200 nM protein in SPR buffer until a reasonable level of arbitrary response units is reached (not more than 1500 RU for K_d determination to avoid repeated association and dissociation of the analyte); then wash with SPR buffer until a stable baseline is observed. As unspecifically adsorbed protein will be removed, about one-third of the initial response units will remain on the chip.
- 3) Prepare RNA analyte with a concentration of 0.1 and 1 μM in SPR buffer as a starting point for aptamer analysis. The smallest recommended sample size is 45 μl including dead volumes in valves and tubings. If possible, use larger sample volumes and prolong association phases.
- 4) Load sample loop using the two-bubble technique (Comment 12.3) against the direction of flow and inject by maintaining the flow rate ($5 \mu\text{l min}^{-1}$); use KINJECT program thereby. Follow association and dissociation kinetics until a stable signal is observed.
- 5) Analyze the data by fitting association and dissociation curves with a program such as BIAevaluation (Biacore, Sweden) or Clamp [119], paying special attention to the binding model chosen as starting condition of the fit [120].
- 6) For quantitative kinetic analysis, repeat injection with multiple samples covering a broad concentration range (roughly between 0.1 and up to 100 times the expected K_d).

Buffers

- 1 x SPR buffer: 10 mM HEPES/NaOH pH 7.4, 100 mM NaCl, 7 mM MgCl₂ (HBS-EP buffer).

Comments

- 1) The technique is based on the interaction of the binder with the target, affecting the refractive index of the surface. Binding events are measured at a surface, while a steady-state equilibrium is achieved, which can result in effects not observed for the unbound protein in solution. Chips can be stored overnight at 4 °C in SPR buffer, but it is best to complete the measurement (particularly K_d determination) in one sweep as immobilized ligand could be lost over time.
- 2) The NTA sensor chip allows fast and easy immobilization of His-tagged proteins, but the interaction to the nickel ions is not very stable resulting in a continuous loss of protein on the surface associated with reduced response units. The CM5 chip offers an alternative due to covalent attachment of the target onto the surface.
- 3) At least for K_d measurements, it is recommended to use a reference cell without immobilized target (but activated and deactivated afterwards if using the CM5 chip) for subsequent subtraction of the background signal.
- 4) SPR is very sensitive and thus prone to artifacts. Running tests for each individual binding reaction to be assayed is essential. Particular attention should be given to flow rate, temperature, concentrations of target and binder, measurement time, general buffer composition, addition of detergents to avoid precipitation, and system clogging [94]. For RNA work, the concentration of Mg²⁺ as well as the pH is particularly critical. Include controls, such as a non-specifically binding protein, nucleic acid, or carbohydrate.
- 5) The Biacore instrument is very sensitive, which might result in measuring artifacts. Therefore, it is important to filter and degas all buffers immediately before use. Additionally, clean the instrument and tubings regularly with the program provided by the manufacturer to prevent microbial growth and artifacts due to denatured deposits. However, to avoid stripping or inactivation of immobilized target, use a Maintenance chip.
- 6) Two-bubble technique: Set pipet to 45 µl, take up sample, remove pipet with tip from the liquid, adjust pipette volume to 48 µl, thereby soaking up a bit of air producing a bubble, dip pipet tip back into the liquid and adjust pipet volume to 55 µl, thus taking up more sample, again remove tip from the liquid and adjust pipet volume to 58 µl, soaking up a second air bubble. In the yellow tip should appear, from top to bottom: major sample amount – air bubble – small sample amount – air bubble. The tandem air bubbles will effectively prevent dilution of the major sample aliquot with buffer while passing through the instrument; in addition, the two bubbles will cause a twin signal indicating chip loading.

- 7) Alternatively to sample loading with a loop, particularly if no plateau and thus equilibrium is reached in the response curve due to limited sample volume, add the analyte to the flow buffer [121].

Protocol 14: Structural Characterization of Aptamers and Their Binding Motifs

Aptamer characterization requires RNA end-labeling to high specific activity, which is achieved by labeling transcribed and gel-purified aptamers 5'- or 3' with [γ - ^{32} P]ATP and T4 polynucleotide kinase (PNK) or [^{32} P]pCp and T4 RNA ligase, respectively (Chapter 3 and [122]). As a start, determine the size of the minimal binding motif and gather information on the tertiary structure of the free and target-bound aptamer. Reliable, fast, and direct probing techniques, such as cleavage by RNases V1, S1, and T1 as well as lead-induced hydrolysis, are recommended for the initial characterization. For structure probing with nucleases, lead ions, and chemical reagents, the reader is referred to the chapters in Part II and Chapter 44.

Boundary Experiment to Identify Minimal Binding Motifs

- 1) Produce an alkaline hydrolysis ladder by the addition of 4 μ l containing 50 000–100 000 Cerenkov cpm of 5'-end-labeled RNA to 36 μ l of 50 mM NaHCO₃ and incubation at 90 °C. The exact amount of time needs to be determined for each RNA. Recover by ethanol precipitation with glycogen as coprecipitant.
- 2) Use the selection assay set up, for example, target protein immobilization on a Ni-NTA matrix, to analyze complex formation between RNA and target; incubate the partially hydrolyzed RNA pool with the matrix-bound target protein as described for the selection in Protocol 4.1. Unbound material is removed by three to six washing steps, and bound fragments are specifically eluted with imidazole or free target protein.
- 3) After phenol/ether extraction (Protocol 4, step 6), RNA is precipitated with ethanol, redissolved in 8 μ l H₂O, and analyzed by 12% dPAGE. Boundaries defining the essential binding regions are identified by a sharp drop or, at best, the absence of ladder intensity below a certain fragment size in the elution fraction. Quality of the picture depends on the stringency of washing steps.
- 4) Repeat analysis with 3'-end-labeled material to have results from both ends, making the interpretation more meaningful.

Comments

- 1) As low concentrations of RNA are used in boundary and probing experiments, include molecular biology grade glycogen (Merck) as coprecipitant and, if possible, unlabeled, unspecific RNA to reduce surface adsorptions that can also be reduced by using siliconized reaction tubes.
- 2) Incubation conditions (buffer, temperature) should be closely adapted to those used during selection and should further take into account the later application

context of the aptamer as well as requirements for RNA structure formation (e.g., sufficient amounts of Mg^{2+}).

- 3) Of particular interest are the contact sites between target and aptamer, which require comparing aptamer structure in the presence and absence of target.

Acknowledgments

We thank Prof. Dr. Mario Mörl for ongoing support, and Prof. Dr. Roland Hartmann for excellent comments and editorial support.

References

1. Hermann, T. and Patel, D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820–825.
2. Ellington, A.D. and Szostak, J.W. (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
3. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
4. Robertson, D.L. and Joyce, G.F. (1990) Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, **344**, 467–468.
5. Blum, J.H., Dove, S.L., Hochschild, A., and Mekalanos, J.J. (2000) Isolation of peptide aptamers that inhibit intracellular processes. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 2241–2246.
6. Hoogenboom, H.R. (2002) Overview of antibody phage-display technology and its applications. *Methods Mol. Biol.*, **178**, 1–37.
7. Hoess, R.H. (2001) Protein design and phage display. *Chem. Rev.*, **101**, 3205–3218.
8. Liu, R., Barrick, J.E., Szostak, J.W., and Roberts, R.W. (2000) Optimized synthesis of RNA-protein fusions for *in vitro* protein selection. *Methods Enzymol.*, **318**, 268–293.
9. Takahashi, T.T., Austin, R.J., and Roberts, R.W. (2003) mRNA display: ligand discovery, interaction analysis and beyond. *Trends Biochem. Sci.*, **28**, 159–165.
10. Schaffitzel, C. and Plückthun, A. (2001) Protein-fold evolution in the test tube. *Trends Biochem. Sci.*, **26**, 577–579.
11. Ciesiolka, J., Illangasekare, M., Majerfeld, I., Nickles, T., Welch, M., Yarus, M., and Zinnen, S. (1996) Affinity selection-amplification from randomized ribooligonucleotide pools. *Methods Enzymol.*, **267**, 315–335.
12. Hofmann, H.P., Limmer, S., Hornung, V., and Sprinzl, M. (1997) Ni^{2+} -binding RNA motifs with an asymmetric purine-rich internal loop and a G-A base pair. *RNA*, **3**, 1289–1300.
13. Wang, Y., Killian, J., Hamasaki, K., and Rando, R.R. (1996) RNA molecules that specifically and stoichiometrically bind aminoglycoside antibiotics with high affinities. *Biochemistry*, **35**, 12338–12346.
14. Baugh, C., Grate, D., and Wilson, C. (2000) 2.8 Å crystal structure of the malachite green aptamer. *J. Mol. Biol.*, **301**, 117–128.
15. Jenison, R.D., Gill, S.C., Pardi, A., and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425–1429.
16. Derbyshire, N., White, S.J., Bunka, D.H.J., Song, L., Stead, S., Tarbin, J., Sharman, M., Zhou, D., and Stockley, P.G. (2012) Toggled RNA aptamers against aminoglycosides allowing facile detection of antibiotics using gold

- nanoparticle assays. *Anal. Chem.*, **84**, 6595–6602.
17. Huang, Z. and Szostak, J.W. (2003) Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer. *RNA*, **9**, 1456–1463.
18. Vaish, N.K., Larralde, R., Fraley, A.W., Szostak, J.W., and McLaughlin, L.W. (2003) A novel, modification-dependent ATP-binding aptamer selected from an RNA library incorporating a cationic functionality. *Biochemistry*, **42**, 8842–8851.
19. Burke, D.H. and Gold, L. (1997) RNA aptamers to the adenosine moiety of S-adenosyl methionine: structural inferences from variations on a theme and the reproducibility of SELEX. *Nucleic Acids Res.*, **25**, 2020–2024.
20. Majerfeld, I., Chocholousova, J., Malaiya, V., Widmann, J., McDonald, D., Reeder, J., Iyer, M., Illangasekare, M., Yarus, M., and Knight, R. (2010) Nucleotides that are essential but not conserved; a sufficient L-tryptophan site in RNA. *RNA*, **16**, 1915–1924.
21. Geiger, A., Burgstaller, P., von der Eltz, H., Roeder, A., and Famulok, M. (1996) RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity. *Nucleic Acids Res.*, **24**, 1029–1036.
22. Faulhammer, D., Eschgfäller, B., Stark, S., Burgstaller, P., Englberger, W., Erfurth, J., Kleinjung, F., Rupp, J., Dan Vulcu, S., Schröder, W. *et al.* (2004) Biostable aptamers with antagonistic properties to the neuropeptide nociceptin/orphanin FQ. *RNA*, **10**, 516–527.
23. Srisawat, C., Goldstein, I.J., and Engelke, D.R. (2001) Sephadex-binding RNA ligands: rapid affinity purification of RNA from complex RNA mixtures. *Nucleic Acids Res.*, **29**, E4.
24. Pan, W., Craven, R.C., Qiu, Q., Wilson, C.B., Wills, J.W., Golovine, S., and Wang, J.F. (1995) Isolation of virus-neutralizing RNAs from a large pool of random sequences. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 11509–11513.
25. Ringquist, S., Jones, T., Snyder, E.E., Gibson, T., Boni, I., and Gold, L. (1995) High-affinity RNA ligands to *Escherichia coli* ribosomes and ribosomal protein S1: comparison of natural and unnatural binding sites. *Biochemistry*, **34**, 3640–3648.
26. Homann, M. and Göringer, H.U. (1999) Combinatorial selection of high affinity RNA ligands to live African trypanosomes. *Nucleic Acids Res.*, **27**, 2006–2014.
27. Morris, K.N., Jensen, K.B., Julin, C.M., Weil, M., and Gold, L. (1998) High affinity ligands from *in vitro* selection: complex targets. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 2902–2907.
28. Winkler, W.C. and Breaker, R.R. (2003) Genetic control by metabolite-binding riboswitches. *ChemBioChem*, **4**, 1024–1032.
29. Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C., and Breaker, R.R. (2003) Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell*, **113**, 577–586.
30. Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A., and Breaker, R.R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, **428**, 281–286.
31. Osborne, S.E., Matsumura, I., and Ellington, A.D. (1997) Control of gene expression by a natural metabolite-responsive ribozyme. *Curr. Opin. Chem. Biol.*, **1**, 5–9.
32. Brody, E.N. and Gold, L. (2000) Aptamers as therapeutic and diagnostic agents. *J. Biotechnol.*, **74**, 5–13.
33. Burgstaller, P., Jenne, A., and Blind, M. (2002) Aptamers and aptazymes: accelerating small molecule drug discovery. *Curr. Opin. Drug Discovery Dev.*, **5**, 690–700.
34. Toulmé, J.J., Darfeuille, F., Kolb, G., Chabas, S., and Staedel, C. (2003) Modulating viral gene expression by aptamers to RNA structures. *Biol. Cell*, **95**, 229–238.
35. Eaton, B.E. (1997) The joys of *in vitro* selection: chemically dressing oligonucleotides to satiate protein targets. *Curr. Opin. Chem. Biol.*, **1**, 10–16.

36. Famulok, M. and Verma, S. (2002) *In vivo*-applied functional RNAs as tools in proteomics and genomics research. *Trends Biotechnol.*, **20**, 462–466.
37. Conrad, R.C., Giver, L., Tian, Y., and Ellington, A.D. (1996) *In vitro* selection of nucleic acid aptamers that bind proteins. *Methods Enzymol.*, **267**, 336–367.
38. Marshall, K.A. and Ellington, A.D. (2000) *In vitro* selection of RNA aptamers. *Methods Enzymol.*, **318**, 193–214.
39. Ponomarenko, J.V., Orlova, G.V., Frolov, A.S., Gelfand, M.S., and Ponomarenko, M.P. (2002) SELEX_DB: a database on *in vitro* selected oligomers adapted for recognizing natural sites and for analyzing both SNPs and site-directed mutagenesis data. *Nucleic Acids Res.*, **30**, 195–199.
40. Conrad, R., Keranen, L.M., Ellington, A.D., and Newton, A.C. (1994) Isozyme-specific inhibition of protein kinase C by RNA aptamers. *J. Biol. Chem.*, **269**, 32051–32054.
41. Aurup, H., Williams, D.M., and Eckstein, F. (1992) 2'-Fluoro- and 2'-amino-2'-deoxynucleoside 5'-triphosphates as substrates for T7 RNA polymerase. *Biochemistry*, **31**, 9636–9641.
42. Fitzwater, T. and Polisky, B. (1996) A SELEX primer. *Methods Enzymol.*, **267**, 275–301.
43. Kusser, W. (2000) Chemically modified nucleic acid aptamers for *in vitro* selections: evolving evolution. *J. Biotechnol.*, **74**, 27–38.
44. Klussmann, S., Nolte, A., Bald, R., Erdmann, V.A., and Furste, J.P. (1996) Mirror-image RNA that binds D-adenosine. *Nat. Biotech.*, **14**, 1112–1115.
45. Barciszewski, J., Medgaard, M., Koch, T., Kurreck, J., and Erdmann, V.A. (2009) Locked nucleic acid aptamers. *Methods Mol. Biol.*, **535**, 165–186.
46. Mascini, M., Palchetti, I., and Tombelli, S. (2011) Nucleic acid and peptide aptamers: fundamentals and bioanalytical aspects. *Angew. Chem. Int. Ed.*, **51**, 1316–1332.
47. Smith, D., Collins, B.D., Heil, J., and Koch, T.H. (2003) Sensitivity and specificity of photoaptamer probes. *Mol. Cell Proteomics*, **2**, 11–18.
48. Li, J.-J., Fang, X., and Tan, W. (2002) Molecular aptamer beacons for real-time protein recognition. *Biochem. Biophys. Res. Commun.*, **292**, 31–40.
49. Michaud, M., Jourdan, E., Villet, A., Ravel, A., Grosset, C., and Peyrin, E. (2003) A DNA aptamer as a new target-specific chiral selector for HPLC. *J. Am. Chem. Soc.*, **125**, 8672–8679.
50. Potyrai, R.A., Conrad, R.C., Ellington, A.D., and Hieftje, G.M. (1998) Adapting selected nucleic acid ligands (aptamers) to biosensors. *Anal. Chem.*, **70**, 3419–3425.
51. Nutiu, R. and Li, Y. (2003) Structure-switching signaling aptamers. *J. Am. Chem. Soc.*, **125**, 4771–4778.
52. Hamaguchi, N., Ellington, A., and Stanton, M. (2001) Aptamer beacons for the direct detection of proteins. *Anal. Biochem.*, **294**, 126–131.
53. Jhaveri, S., Rajendran, M., and Ellington, A.D. (2000) *In vitro* selection of signaling aptamers. *Nat. Biotechnol.*, **18**, 1293–1297.
54. Rajendran, M. and Ellington, A.D. (2003) *In vitro* selection of molecular beacons. *Nucleic Acids Res.*, **31**, 5700–5713.
55. Brody, E.N., Willis, M.C., Smith, J.D., Jayasena, S., Zichi, D., and Gold, L. (1999) The use of aptamers in large arrays for molecular diagnostics. *Mol. Diagn.*, **4**, 381–388.
56. Fang, X., Sen, A., Vicens, M., and Tan, W. (2003) Synthetic DNA aptamers to detect protein molecular variants in a high-throughput fluorescence quenching assay. *Chembiochem.*, **4**, 829–834.
57. Hong, P., Li, W., and Li, J. (2012) Applications of aptasensors in clinical diagnostics. *Sensors*, **12**, 1181–1193.
58. Wang, C., Zhang, M., Yang, G., Zhang, D., Ding, H., Wang, H., Fan, M., Shen, B., and Shao, N. (2003) Single-stranded

- DNA aptamers that bind differentiated but not parental cells: subtractive systematic evolution of ligands by exponential enrichment. *J. Biotechnol.*, **102**, 15–22.
59. Lozupone, C., Changayil, S., Majerfeld, I., and Yarus, M. (2003) A more complex isoleucine aptamer with a cognate triplet. *RNA*, **9**, 1315–1322.
 60. Cox, J.C., Hayhurst, A., Hesselberth, J., Bayer, T.S., Georgiou, G., and Ellington, A.D. (2002) Automated selection of aptamers against protein targets translated *in vitro*: from gene to aptamer. *Nucleic Acids Res.*, **30**, E108–E108.
 61. Glöckler, J., Schütze, T., and Konthur, Z. (2010) Automation in the high-throughput selection of random combinatorial libraries—different approaches for select applications. *Molecules*, **15**, 2478–2490.
 62. Ikebukuro, K., Okumura, Y., Sumikura, K., and Karube, I. (2005) A novel method of screening thrombin-inhibiting DNA aptamers using an evolution-mimicking algorithm. *Nucleic Acids Res.*, **33**, e108.
 63. Nimjee, S.M., Rusconi, C.P., and Sullenger, B.A. (2006) in *The Aptamer Handbook* (ed. S. Klussmann), Wiley-VCH Verlag GmbH & Co. KGaA, pp. 131–166.
 64. Tok, J.B., Cho, J., and Rando, R.R. (2000) RNA aptamers that specifically bind to a 16S ribosomal RNA decoding region construct. *Nucleic Acids Res.*, **28**, 2902–2910.
 65. Toulme, J.J., Darfeuille, F., Kolb, G., Chabas, S., and Staedel, C. (2003) Modulating viral gene expression by aptamers to RNA structures. *Biol. Cell*, **95**, 229–238.
 66. Khvorova, A., Kwak, Y.G., Tamkun, M., Majerfeld, I., and Yarus, M. (1999) RNAs that bind and change the permeability of phospholipid membranes. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 10649–10654.
 67. Janas, T. and Yarus, M. (2003) Visualization of membrane RNAs. *RNA*, **9**, 1353–1361.
 68. Pfister, P., Hobbie, S., Vicens, Q., Bottger, E.C., and Westhof, E. (2003) The molecular basis for A-site mutations conferring aminoglycoside resistance: relationship between ribosomal susceptibility and X-ray crystal structures. *ChemBioChem*, **4**, 1078–1088.
 69. Vicens, Q. and Westhof, E. (2003) Molecular recognition of aminoglycoside antibiotics by ribosomal RNA and resistance enzymes: an analysis of x-ray crystal structures. *Biopolymers*, **70**, 42–57.
 70. Vicens, Q. and Westhof, E. (2003) RNA as a drug target: the case of aminoglycosides. *ChemBioChem*, **4**, 1018–1023.
 71. Seeberger, P.H. and Haase, W.C. (2000) Solid-phase oligosaccharide synthesis and combinatorial carbohydrate libraries. *Chem. Rev.*, **100**, 4349–4394.
 72. Koeller, K.M. and Wong, C.H. (2000) Synthesis of complex carbohydrates and glycoconjugates: enzyme-based and programmable one-pot strategies. *Chem. Rev.*, **100**, 4465–4494.
 73. Jeong, S., Eom, T., Kim, S., Lee, S., and Yu, J. (2001) *In vitro* selection of the RNA aptamer against the Sialyl Lewis X and its inhibition of the cell adhesion. *Biochem. Biophys. Res. Commun.*, **281**, 237–243.
 74. Masud, M.M., Kuwahara, M., Ozaki, H., and Sawai, H. (2004) Sialyllactose-binding modified DNA aptamer bearing additional functionality by SELEX. *Bioorg. Med. Chem.*, **12**, 1111–1120.
 75. Boese, B.J. and Breaker, R.R. (2007) *In vitro* selection and characterization of cellulose-binding DNA aptamers. *Nucleic Acids Res.*, **35**, 6378–6388.
 76. Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.*, **60**, 523–533.
 77. Tatsumi, H., Fukuda, S., Kikuchi, M., and Koyama, Y. (1996) Construction of biotinylated firefly luciferases using biotin acceptor peptides. *Anal. Biochem.*, **243**, 176–180.

78. Feizi, T., Fazio, F., Chai, W., and Wong, C.H. (2003) Carbohydrate microarrays - a new set of technologies at the frontiers of glycomics. *Curr. Opin. Struct. Biol.*, **13**, 637–645.
79. Vater, A., Jarosch, F., Buchner, K., and Klussmann, S. (2003) Short bioactive Spiegelmers to migraine-associated calcitonin gene-related peptide rapidly identified by a novel approach: tailored-SELEX. *Nucleic Acids Res.*, **31**, e130.
80. Kore, A.R., Vaish, N.K., Morris, J.A., and Eckstein, F. (2000) *In vitro* evolution of the hammerhead ribozyme to a purine-specific ribozyme using mutagenic PCR with two nucleotide analogues. *J. Mol. Biol.*, **301**, 1113–1121.
81. Zaccolo, M., Williams, D.M., Brown, D.M., and Gherardi, E. (1996) An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.*, **255**, 589–603.
82. Uhlenbeck, O.C. (1995) Keeping RNA happy. *RNA*, **1**, 4–6.
83. Vaish, N.K., Heaton, P.A., and Eckstein, F. (1997) Isolation of hammerhead ribozymes with altered core sequences by *in vitro* selection. *Biochemistry*, **36**, 6495–6501.
84. Schütze, T., Wilhelm, B., Greiner, N., Braun, H., Peter, F., Mörl, M., Erdmann, V.A., Lehrach, H., Konthur, Z., Menger, M. *et al.* (2011) Probing the SELEX process with next-generation sequencing. *PLoS ONE*, **6**, e29604.
85. Davis, J.P., Janjic, N., Javornik, B.E., and Zichi, D.A. (1996) Identifying consensus patterns and secondary structure in SELEX sequence sets. *Methods Enzymol.*, **267**, 302–314.
86. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
87. Knight, R. and Yarus, M. (2003) Analyzing partially randomized nucleic acid pools: straight dope on doping. *Nucleic Acids Res.*, **31**, e30.
88. Hirao, I., Harada, Y., Nojima, T., Osawa, Y., Masaki, H., and Yokoyama, S. (2004) *In vitro* selection of RNA aptamers that bind to colicin E3 and structurally resemble the decoding site of 16S ribosomal RNA. *Biochemistry*, **43**, 3214–3221.
89. Arnold, F.H. and Blanch, H.W. (1986) Analytical affinity chromatography. II. Rate theory and the measurement of biological binding kinetics. *J. Chromatogr. A*, **355**, 13–27.
90. Arnold, F.H., Schofield, S.A., and Blanch, H.W. (1986) Analytical affinity chromatography: II. Rate theory and the measurement of biological binding kinetics. *J. Chromatogr. A*, **355**, 1–12.
91. Blount, K.F., Wang, J.X., Lim, J., Sudarsan, N., and Breaker, R.R. (2007) Antibacterial lysine analogs that target lysine riboswitches. *Nat. Chem. Biol.*, **3**, 44–49.
92. Srisawat, C. and Engelke, D.R. (2001) Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA*, **7**, 632–641.
93. Tahiri-Alaoui, A., Frigotto, L., Manville, N., Ibrahim, J., Romby, P., and James, W. (2002) High affinity nucleic acid aptamers for streptavidin incorporated into bi-specific capture ligands. *Nucleic Acids Res.*, **30**, e45.
94. Cox, J.C. and Ellington, A.D. (2001) Automated selection of anti-protein aptamers. *Bioorg. Med. Chem.*, **9**, 2525–2531.
95. Bartel, D.P. and Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science*, **261**, 1411–1418.
96. Peterson, E.T., Blank, J., Sprinzl, M., and Uhlenbeck, O.C. (1993) Selection for active *E. coli* tRNA(Phe) variants from a randomized library using two proteins. *EMBO J.*, **12**, 2959–2967.
97. Green, M.R. and Sambrook, J. (2012) *Molecular Cloning: A Laboratory Manual*, 4th edn, Cold Spring Harbor Laboratory.
98. Milligan, J.F. and Uhlenbeck, O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.*, **180**, 51–62.

99. Piasecki, S.K., Hall, B., and Ellington, A.D. (2009) Nucleic acid pool preparation and characterization, in *Nucleic Acid and Peptide Aptamers* (ed. G. Mayer).
100. Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R. *et al.* (2008) Protein production and purification. *Nat. Methods*, **5**, 135–146.
101. Yan, J.X., Wait, R., Berkelman, T., Harry, R.A., Westbrook, J.A., Wheeler, C.H., and Dunn, M.J. (2000) A modified silver staining protocol for visualization of proteins compatible with matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry. *Electrophoresis*, **21**, 3666–3672.
102. Meyer, C., Eydeler, K., Magbanua, E., Zivkovic, T., Piganeau, N., Lorenzen, I., Grötzinger, J., Mayer, G., Rose-John, S., and Hahn, U. (2012) Interleukin-6 receptor specific RNA aptamers for cargo delivery into target cells. *RNA Biol.*, **9**, 67–80.
103. Wallace, S.T. and Schroeder, R. (2000) *In vitro* selection and characterization of RNAs with high affinity to antibiotics. *Methods Enzymol.*, **318**, 214–229.
104. Cadwell, R.C. and Joyce, G.F. (1994) Mutagenic PCR. *PCR Methods Appl.*, **3**, S136–S140.
105. Cadwell, R.C. and Joyce, G.F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl.*, **2**, 28–33.
106. Sterba, K.M., Mackintosh, S.G., Blevins, J.S., Hurlburt, B.K., and Smeltzer, M.S. (2003) Characterization of *Staphylococcus aureus* SarA binding sites. *J. Bacteriol.*, **185**, 4410–4417.
107. Beinoraviciute-Kellner, R., Lipps, G., and Krauss, G. (2005) *In vitro* selection of DNA binding sites for ABF1 protein from *Saccharomyces cerevisiae*. *FEBS Lett.*, **579**, 4535–4540.
108. Carey, J. (1991) Gel retardation. *Methods Enzymol.*, **208**, 103–117.
109. Gopinath, S.C.B. (2007) Methods developed for SELEX. *Anal. Bioanal. Chem.*, **387**, 171–182.
110. Carey, J., Lowary, P.T., and Uhlenbeck, O.C. (1983) Interaction of R17 coat protein with synthetic variants of its ribonucleic acid binding site. *Biochemistry*, **22**, 4723–4730.
111. Wong, I. and Lohman, T.M. (1993) A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 5428–5432.
112. Johnsson, B., Lofas, S., and Lindquist, G. (1991) Immobilization of proteins to a carboxymethyl-dextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Anal. Biochem.*, **198**, 268–277.
113. Baird, C.L. and Myszka, D.G. (2001) Current and emerging commercial optical biosensors. *J. Mol. Recognit.*, **14**, 261–268.
114. Morton, T.A. and Myszka, D.G. (1998) Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Methods Enzymol.*, **295**, 268–294.
115. Myszka, D.G. (2000) Kinetic, equilibrium, and thermodynamic analysis of macromolecular interactions with BIACORE. *Methods Enzymol.*, **323**, 325–340.
116. Misono, T.S. and Kumar, P.K.R. (2005) Selection of RNA aptamers against human influenza virus hemagglutinin using surface plasmon resonance. *Anal. Biochem.*, **342**, 312–317.
117. Rich, R.L. and Myszka, D.G. (2003) A survey of the year 2002 commercial optical biosensor literature. *J. Mol. Recognit.*, **16**, 351–382.
118. Rich, R.L. and Myszka, D.G. (2001) Survey of the year 2000 commercial optical biosensor literature. *J. Mol. Recognit.*, **14**, 273–294.
119. Myszka, D.G. and Morton, T.A. (1998) CLAMP: a biosensor kinetic data analysis program. *Trends Biochem. Sci.*, **23**, 149–150.
120. Myszka, D.G. (1999) Improving biosensor analysis. *J. Mol. Recognit.*, **12**, 279–284.

121. Myszka, D.G., Jonsen, M.D., and Graves, B.J. (1998) Equilibrium analysis of high affinity interactions using BIACORE. *Anal. Biochem.*, **265**, 326–330.
122. Brunel, C. and Romby, P. (2000) Probing RNA structure and RNA-ligand complexes with chemical probes. *Methods in Enzymol.*, **318**, 3–21.

51

In Vitro Selection against Small Targets

Dirk Eulberg, Christian Maasch, Werner G. Purschke, and Sven Klussmann

51.1

Introduction

Aptamers have been raised against a plethora of small targets, for example, organic dyes, by so-called Systematic Evolution of Ligands by EXponential enrichment (SELEX) experiments [1, 2]. In the early days of *in vitro* selection, one important driver to select against a target was to study the ability of nucleic acids to form stable, binding pocket- or crevice-like three-dimensional structures in a system that was as simple as possible. In addition, the theory about an “RNA world” origin of life fueled many selection experiments; every new, functional nucleic acid structure corroborated the versatility of nucleic acids – which consist of only four different building blocks! The fact that aptamers alone could even be readily isolated against targets that do not naturally bind nucleic acids or polyanions – for example, hydrophobic molecules such as the amino acids isoleucine or valine – triggered many SELEX experiments [3, 4].

The molecular principles by which small molecules inhibit RNA function or interact with RNA was addressed by performing *in vitro* selections against antibiotics with the objective to understand how RNAs recognize small molecules, information that is of crucial importance for attempts to develop novel drugs targeted against pathogen-specific RNAs [5, 6].

The comparable properties of aptamers and antibodies regarding affinity and specificity suggested the use of aptamers as monitoring or typing tools. Two examples are (i) the isolation of an aptamer against 7,8-dihydro-8-hydroxy-2'-deoxyguanosine (oxodG), one of the most common mutagenic lesions in oxidative DNA damage, which can be applied to identify and quantify this modification in DNA [7], and (ii) a cyclic AMP (cAMP)-binding aptamer, which can be used to quantify this second messenger in solution [8].

Intracellular processes may be elucidated using aptamers that bind fluorophores such as sulforhodamine B or fluorescein [9]. For example, the expression and intracellular localization of transcripts fused to the aptamer may be identified through fluorophore binding to the aptamer domain.

Another fascinating field of nucleic acid biochemistry is the area of catalytically active nucleic acids, the ribozymes or deoxyribozymes. Binding of the substrate molecule is the first step in any macromolecule-catalyzed reaction; when aiming to design specific ribozyme activities, the experimental strategy consequently is to initially generate an aptamer for the desired substrate. Small targets such as arginine, ATP, or *S*-adenosylmethionine can be used to produce aptamers as highly affine substrate-binding building blocks, which then serve as the starting material to evolve catalytic nucleic acids [10–12]. In the same way, allosteric domains can be generated using aptamers that were selected against an allosteric effector of interest. Such a nucleic acid effector module may readily be conjugated to a catalytic nucleic acid domain [13] to give an allosteric ribozyme, also known as *aptazyme*.

Finally, nucleic acid ligands with therapeutic and/or diagnostic potential, which bind peptide targets, have been isolated. It could be shown that a nucleic acid antagonist for the pharmacologically relevant peptide gonadotropin-releasing hormone I (GnRH) effectively binds and neutralizes GnRH *in vivo* [14].

Compared to selection experiments against proteins, a major difference when selecting oligonucleotides specific to small targets is that the resulting aptamers have to bind and clamp the small target in a receptor-like manner; the required size of the isolated minimal binding motifs may therefore be larger in comparison to protein binders. Moreover, the potential contact surface to an aptamer is small. It can be hypothesized that the binding of peptides, for example, may require a rigid three-dimensional scaffold, since small peptides are, in general, intrinsically flexible structures and usually exist in solution as an equilibrium of multiple conformers [15]. For all these reasons, the resulting binding affinities may be negatively influenced by the conformational entropy lost upon binding of a small flexible molecule.

Perhaps the most important experimental detail affecting the success of *in vitro* selections against small targets is the mode of target immobilization. The smaller the target, the greater is the danger of isolating aptamers that require not only the target itself for binding but also structures involved in immobilization, that is the linker and/or the matrix.

The key step of every *in vitro* selection experiment is the process of partitioning (Figure 51.1) – the phase of the selection where the properties of the binders to be selected are defined. The input RNA pool, which may contain 10^{15} or even more different species, is brought into contact with the target of interest, often immobilized on an affinity chromatography column. The rare RNA species that interact with the immobilized target will remain bound to the column; the largest part of the RNA population, on the other hand, will simply pass through. As aptamers with the highest affinities possible are desired, weak binders are removed from the column by repeated washing. In theory, only specific binders should survive the process, to then be eluted in the final stage, either by applying denaturing conditions to the column or by applying a buffer containing free target in excess concentration to competitively displace the bound species from the

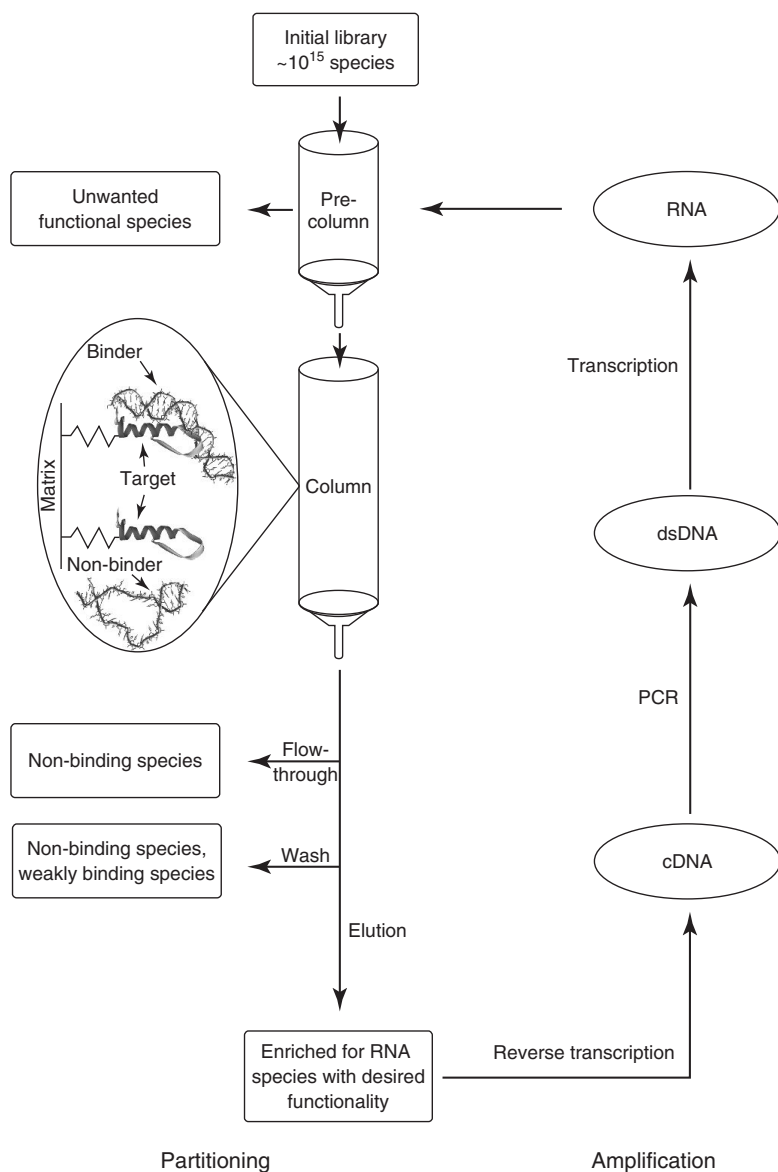


Figure 51.1 A schematic representation of the *in vitro* selection cycle.

immobilized target. In practice, however, the situation is often more complex and unpredictable.

Target molecules must be immobilized for partitioning, and the introduction of chromatography material brings surfaces into the system, which will bind RNA molecules either specifically or non-specifically. Specific matrix binders can be

eliminated by preselecting the RNA pool with non-derivatized column material and by discarding the column-bound material, using only the flow-through. Non-specifically retained RNA molecules represent the background binding, which is always present and should remain on a constant low level in a smoothly running *in vitro* selection until specific binders become enriched and predominate after several rounds of selection. In the following sections, all steps of an *in vitro* selection experiment are discussed in detail, with particular emphasis on selection against small targets. Target immobilization, partitioning, and assessment of binding constants are the focus of this chapter.

51.2

Target Immobilization

The aptamer species finally selected are those that derive the maximum benefit from the experimental setup. This does not necessarily mean that the selected aptamers are high-affinity binders of the target: unwanted or unforeseen selection pressure may lead to unwanted or unforeseen results. Since the selection process requires immobilization of the target to a surface, the danger of selecting for species that bind to the immobilization matrix or recognize target and matrix cooperatively is always present. Therefore, the most important parameter for the design of a successful *in vitro* selection experiment, especially when selecting against small targets, is the optimal presentation of the target during the binding reaction. Thus, key parameters are the type of immobilization matrix and the coupling chemistry.

Two partitioning methods that have been successfully employed for selecting protein-binding aptamers are unfortunately not suited for small molecules:

- Passive, hydrophobic adsorption on plastic surfaces: most small targets do not exhibit hydrophobic regions to the extent required, and even if they had such hydrophobic regions, adsorption on plastic surfaces would likely render small targets inaccessible to an aptamer.
- Filtration through nitrocellulose filter membranes: complexes between aptamers and small targets may not be retained on the filters.

Nevertheless, a cornucopia of useful media for the immobilization of small ligands exists. Usually, these media are commercially available as non-magnetic or paramagnetic particles composed of many different polymers. Such particles are offered with a wide variety of activated surfaces, which may influence binding between target and aptamer.

Length and nature of the spacer connecting the target and the solid support are critical parameters, especially when working with very small targets. The binding site of aptamers may be located deeply within the molecule, and matrices prepared by coupling small ligands directly to the support may suffer from problems of steric hindrance. On the one hand, it is desirable to attach the target *via* a long spacer to ensure optimal accessibility; on the other hand, the longer the spacer,

the more it will change the overall characteristics of the immobilized target with respect to charge or polarity and may thus reduce the specificity of the separation. Unfortunately, no general rule exists as to which matrix and coupling technique will eventually lead to a successful *in vitro* selection experiment.

The key step of each *in vitro* selection experiment is the separation of binding from non-binding species. The more efficiently this separation works, the better the prospects for isolation of the desired binders. Also, low-background binding of the nucleic acid starting pool to the matrix is a prerequisite for success. These parameters have to be checked before starting a selection experiment. Another important aspect is the binding capacity of the matrix, which can be determined in the course of target coupling. The immobilized target concentration per matrix volume can be assessed by measuring the difference between the target concentrations in the coupling solution before and after the immobilization process *via* high-performance liquid chromatography (HPLC), photometry, or other suitable methods.

Separation of ligand-binding from non-binding species during *in vitro* selection is comparable to a standard affinity chromatography. The preparation of some useful matrices is described in detail in this section. This includes covalent immobilization procedures on three different, widely used matrix types and the non-covalent immobilization *via* the biotin–avidin interaction.

51.2.1

Covalent Immobilization

51.2.1.1 Epoxy-Activated Matrices

Epoxy-activated matrices, such as Epoxy-activated Sepharose 6B (GE Healthcare or Sigma-Aldrich), Epoxy-activated Agarose (Sigma-Aldrich), or paramagnetic Dynabeads® M-270 Epoxy (Life Technologies Invitrogen), can be used for covalent immobilization of ligands *via* amino groups at pH 9–11, thiol groups at pH 7.5–8.5, or hydroxyl groups at pH 11–13 (Figure 51.2a). For Epoxy-activated Sepharose 6B, the functional epoxy group is linked to the polysaccharide support *via* a 12-atom hydrophilic spacer, which makes it suitable for immobilization of small molecules; the active group density is in the range of 30 $\mu\text{mol ml}^{-1}$ of drained coupling material.

Protocol: Epoxy-activated Sepharose 6B (GE Healthcare)

Gel Preparation (around 3.5 ml Matrix)

- 1) Suspend 1 g of lyophilized matrix in 10 ml RNase-free (e.g., double-distilled) water and incubate for 10 min at room temperature. Do not stir the suspension with a magnetic stirring bar, as this may grind the agarose to fine particles. The powder will swell to a volume of approximately 3.5 ml.
- 2) Wash the swollen gel in a sintered glass filter by adding at least 200 ml of RNase-free water over 1 h.

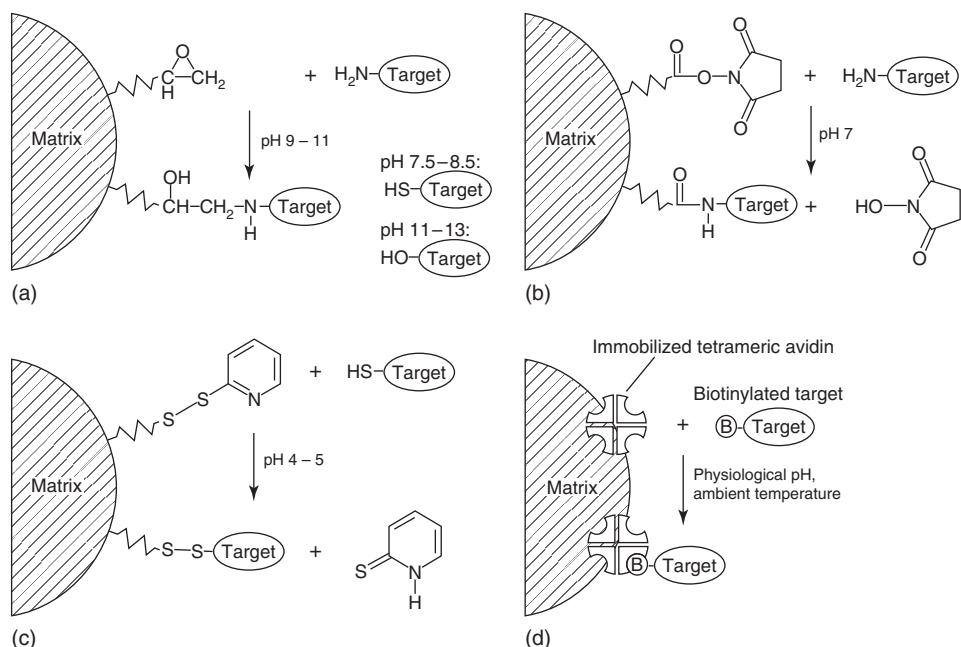


Figure 51.2 Immobilization of targets on solid supports. (a) Epoxy-activated support, (b) NHS-activated support, (c) pyridyl disulfide-activated support, and (d) support with immobilized tetrameric avidin.

Ligand Coupling

- 1) Prepare 1 ml of 200 mM ligand solution in 100 mM coupling buffer (carbonate, phosphate, or borate buffered). Amino group-containing buffers such as Tris or other nucleophiles, such as thiols, hydroxyls, or phenols, must not be present. The reaction proceeds most efficiently at pH 9–13. In case of coupling hydroxyl groups and if the ligand is alkali-stable, couple at pH 13 in NaOH solution. If low concentrations of ligand on the matrix are desired, ligand concentrations of 2 mM or less may be used.
- 2) Add the ligand solution to 1 ml of swollen, drained gel in a closed reaction tube and mix for 16 h at 37 °C or room temperature. Mix gently, for example, on a rotating wheel. Do not use a magnetic stirrer.
- 3) Wash the matrix twice with at least 5 volumes of coupling buffer to remove excess ligand. Save an aliquot of the ligand solution before and after the coupling reaction to determine the coupling efficiency.
- 4) Suspend the matrix in 1 volume of 1 M ethanolamine, pH 8, to block the remaining active groups.
- 5) Incubate at room temperature for 4–16 h.
- 6) Wash the matrix with 5 volumes of 0.1 M acetate buffer/0.5 M NaCl, pH 4, and then with 5 volumes of 0.1 M Tris-HCl buffer/0.5 M NaCl, pH 8.
- 7) Repeat step 6 at least twice.

The matrix material for preselection can be prepared by using 200 mM glycine solution in coupling buffer instead of ligand solution or by simply starting at step 4 of the coupling protocol with pre-swollen gel.

51.2.1.2 NHS-Activated Matrices

N-hydroxysuccinimide (NHS)-activated matrices allow for rapid, gentle, and stable coupling and are highly selective for primary amino groups under physiological conditions. Such media – Affi-Gel 10 (Bio-Rad) or NHS-activated Sepharose™ 4 Fast Flow (GE Healthcare) – are *N*-hydroxysuccinimide esters of a derivatized polymer support (Figure 51.2b). Both media contain a 10-atom spacer arm. The active ester content is $\geq 10 \mu\text{mol ml}^{-1}$ gel for Affi-Gel and around $18 \mu\text{mol ml}^{-1}$ drained gel for NHS-activated Sepharose. The media are provided fully swollen and solvent-stabilized in isopropyl alcohol.

Protocol: Affi-Gel 10 (Bio-Rad)

Ligand Coupling

- 1) Shake the vial and transfer the desired amount of resuspended slurry to a glass-fritted funnel. Drain the supernatant solvent.
- 2) Wash the gel with 3 volumes of ice-cold, RNase-free water. Watch out to avoid drying of the gel bed, particularly when working under vacuum. The time between solvent removal and addition of ligand solution should not exceed 20 min.
- 3) Transfer the gel to a tube and add 0.5–1 gel volume of ice-cold ligand in coupling buffer. The ligand concentration can be 20 mM or less, depending on the desired ligand concentration on the matrix. As coupling buffer, an acetate, carbonate, MOPS, or HEPES-buffered solution, pH 6.5–8.5, is recommended. The coupling buffer should be 10–100 mM to ensure pH stability. Amino-group-containing buffers such as Tris must not be used.
- 4) Incubate the suspension under gentle agitation on a rotating wheel or in a shaker for 1 h at room temperature or 4 h at 4 °C. Save an aliquot of the ligand solution before and after the coupling reaction to determine the coupling efficiency.
- 5) Block unreacted active esters by adding 0.1 volume of 1 M ethanolamine, pH 8.
- 6) Incubate for 1 h at room temperature.
- 7) Wash the gel five times with 5 volumes of coupling buffer or RNase-free water.
- 8) Wash the gel three times with 3 volumes of the buffer that will be used during *in vitro* selection.

Matrix material for preselection can be prepared using 100 mM ethanolamine solution (pH 8) instead of ligand solution.

51.2.1.3 Pyridyl Disulfide-Activated Matrices

Pyridyl disulfide-activated matrices such as Thiopropyl Sepharose 6B (GE Healthcare) or Thiopropyl-activated Agarose (Sigma-Aldrich) react with ligands that contain thiol groups under mild conditions to form mixed disulfides (Figure 51.2c). The active group content for Thiopropyl Sepharose 6B is 18–31 $\mu\text{mol ml}^{-1}$ drained gel; the medium contains a 4-atom spacer arm. Such matrices are especially suited for the immobilization of peptides, which can be easily synthesized with an additional cysteine residue at the N- or C-terminus as coupling site.

In contrast to other immobilization methods, coupling of ligands *via* disulfide bridges is reversible. These matrices therefore facilitate the elution of any binding RNA molecule together with the ligand by reduction of disulfide bonds with dithiothreitol (DTT) or β -mercaptoethanol.

Protocol: Thiopropyl Sepharose 6B (GE Healthcare)

Gel Preparation (around 3 ml Matrix)

- 1) Suspend 1 g of lyophilized matrix in 10 ml RNase-free water and incubate for 10 min at room temperature. Do not stir the suspension with a magnetic stirring bar as this may grind the agarose to fine particles.
- 2) Wash the swollen gel on a sintered glass filter by adding approximately 200 ml of RNase-free water over 15 min.

Ligand Coupling

- 1) Prepare a 100 mM Tris-, acetate-, or phosphate-based coupling buffer (e.g., Tris-HCl pH 7.5) containing 100–500 mM NaCl. EDTA may be added to a final concentration of 1 mM to remove trace amounts of heavy metal ions, which may catalyze oxidation of thiols.
- 2) Degas the coupling buffer to avoid oxidation of thiols.
- 3) Add 1 volume of 20 mM or lower concentrated ligand solution in coupling buffer to the swollen, drained gel in a closed reaction tube and mix gently for 1 h at room temperature, for example, on a rotating wheel or in a shaker. Do not use a magnetic stirrer.
- 4) Transfer the material to a chromatography column and drain the ligand solution. Save an aliquot of the ligand solution before and after the coupling reaction to determine the coupling efficiency. If monitoring unbound substances photometrically, the contribution of released thiopyridone to the absorbance value has to be taken into account.
- 5) Wash with 30 volumes of 100 mM sodium acetate, pH 6.
- 6) Drain the gel and wash with 30 volumes of the buffer that will be used during *in vitro* selection.

Matrix material for preselection can be prepared as described above, but the ligand must be omitted.

51.2.2

Non-covalent Immobilization

Covalent immobilization of ligands on solid supports is easily achieved using commercially available activated chromatography media as described above. Using such ligand-derivatized materials for partitioning is a straightforward process. However, a selection parameter that can only be controlled with great difficulty is the concentration of ligand immobilized on the matrix. An alternative to covalent coupling is the very effective non-covalent (strept)avidin:biotin system. The only prerequisite is the attachment of biotin to the target of interest. This can easily be achieved with any target molecule bearing a primary amino group by reaction with commercially available NHS-activated and linker-equipped biotin such as Sulfo-NHS-LC-Biotin (Thermo Scientific Pierce) or Biotin-NHS (EMD Biosciences Calbiochem). The (strept)avidin:biotin approach allows one to adjust a binding equilibrium between the biotinylated selection target and the RNA pool in solution in the absence of an affinity matrix. Thus, the researcher is completely free in the choice of target and RNA pool concentrations. Capturing biotinylated molecules by immobilized streptavidin, a tetrameric 60 kDa protein, proceeds rapidly and efficiently under a wide range of pH, temperature, and solvent conditions as well as in the presence of denaturants (Figure 51.2d). “Fishing” of target-binding RNA species can therefore be performed in the selection buffer (SB) used in the preceding binding process.

If working with immobilized streptavidin, the binding capacity toward the biotinylated selection target must be determined initially by incubating a constant concentration of biotinylated target with different amounts of streptavidin matrix in the buffer used during the selection. In a physiological buffer, binding will be complete after 15 min at room temperature or 37°C. The measurement of unbound target in the supernatant allows one to determine the binding capacity. Among the suppliers of immobilized streptavidin are Roche (Streptavidin Magnetic Particles), Thermo Scientific Pierce (ImmunoPure Immobilized Streptavidin Gel), and Sigma-Aldrich (Streptavidin-Agarose).

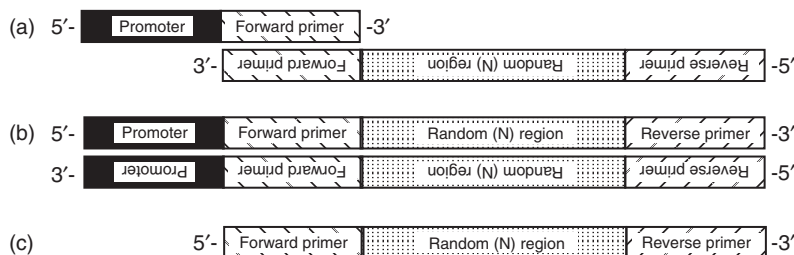


Figure 51.3 Structure of RNA library intermediates. (a,b) Annealed promoter-bearing forward primer and chemically synthesized DNA library (--) strand before (a) and after (b) fill-in or PCR reaction. (c) RNA resulting from *in vitro* transcription using the template shown in (b).

51.3

Nucleic Acid Libraries

In general, combinatorial chemistry makes it possible to generate a vast number of different molecules by the repeated execution of a limited number of synthetic steps. SELEX is the *in vitro* selection of oligonucleotides from combinatorial nucleic acid libraries. The complexity of these libraries is defined by one or more randomized regions within the oligonucleotides. Randomization can easily be accomplished by offering a mixture of all four building blocks during the coupling steps of the solid-phase synthesis. Usually, the random regions are flanked by constant sequences, rendering the molecules accessible to enzymatic amplification procedures.

51.3.1

Library Design

Nucleic acid libraries for the generation of RNA pools have three distinct functional features (Figure 51.3):

- *A 5'-terminal promoter region to initiate in vitro transcription.* The commonly used RNA polymerase for *in vitro* selection is that of phage T7, requiring the double-stranded 17 bp T7 promoter core region; other commercially available RNA polymerases are SP6 or T3 RNA polymerase. The promoter part is generally not synthesized as part of the library, but is added during the fill-in reaction or polymerase chain reaction (PCR) amplification of the chemically synthesized template DNA (Figure 51.3; see Section 51.3.2).
- *Forward and reverse primer regions for PCR amplification.* These regions usually comprise 15–20 nt and are designed to ensure efficient and specific PCR amplification, with a low potential for primer dimer formation and mispriming within the fixed regions. General rules for PCR primer design should be applied; an optimal annealing temperature of $>50^{\circ}\text{C}$ is recommended.
- *Central randomized region.* This region is the source of the vast complexity, which is necessary to start any *in vitro* selection. When designing the length of this region, the following aspects should be considered. The practical limits of molecular biology (approximately 10^{15} molecules is the upper limit of different molecules that can be adequately handled in an *in vitro* selection experiment) have already been reached when the four bases are randomly incorporated at 25 or 26 positions and each variant is represented once ($4^{25} - 4^{26} \approx (1.1-4.5) \times 10^{15}$ molecules, corresponding to roughly 1 nmol). However, solutions for the binding problem to be solved (that is, structures capable of specific interaction with the target) may not be contained in an RNA pool with only 25 randomized positions. Only larger RNAs may be capable of shaping the required stable tertiary structure. Commonly used randomized regions therefore comprise between 30 and more than 100 nt, even if only a minute fraction of the theoretically possible complexity is represented in practice. As the selection against small targets results in receptor mimics that often depend on a complex 3D structure, longer randomized stretches are preferable if the resulting aptamer length is not a critical limitation.

The successfully used pool AL.60 [16] may serve as an example of library design; the T7 promoter core sequence is underlined:

Pool AL.60 (–)strand	5'-TCAGCTGGACGTCTTCAAT-N ₆₀ -TGTCAGGAGCTCGAATT CCC-3'
RNA AL.60 (+)strand	5'-GGGAATTCGAGCTCCTGACA-N ₆₀ -ATTCGAAGACGTCCAGC TGA-3'
AL.T7F (forward primer)	5'-TTC <u>TAATACGACTCACTATAGG</u> GAATTCGAGCTCCTGACA-3'
AL.R (reverse primer)	5'-TCAGCTGGACGTCTTCAAT-3'

51.3.2

Starting Pool Preparation

The input material for *in vitro* selection is prepared by generating dsDNA from the chemically synthesized (–) ssDNA, followed by *in vitro* transcription into RNA. Transcription-ready dsDNA may be produced either by performing one to three PCR cycles (Protocol 1) or by a one-step fill-in reaction of the ssDNA, using DNA polymerases like *Taq* or the Large (Klenow) Fragment of *Escherichia coli* DNA polymerase I (Protocol 2).

Protocol 1: Large-Scale PCR

ssDNA template	1 nmol
10 x PCR buffer	100 µl
MgCl ₂	2.5 mM
dNTPs	200 µM each
Forward primer	5 µM
Reverse primer	5 µM
Betaine	1–2 M
<i>Taq</i> DNA polymerase (Invitrogen; 5 U µl ^{–1})	50 U
Final volume	1 ml

10 x PCR buffer: 200 mM Tris-HCl (pH 8.3), 500 mM KCl
(Invitrogen)

- 1) Divide the reaction mix into 100 µl aliquots.
- 2) Perform one to three PCR cycles. Incubation for 1 min each at appropriate denaturing, annealing, and extension temperatures is generally sufficient, with a final polymerization period of 5 min.
- 3) Ethanol-precipitate, wash with 70% ethanol, and air-dry the generated dsDNA.
- 4) Use the dsDNA pellet as template in a 1 ml *in vitro* transcription reaction (Section 51.4.3). If radiolabeled RNA is used in the *in vitro* selection procedure, an appropriate amount of an α-³²P-labeled NTP may be included in the transcription mix.

Protocol 2: Fill-In Reaction with Klenow Polymerase

ssDNA template	1 nmol
10 x Klenow buffer	100 µl
dNTPs	200 µM each
Forward primer	3 µM
Betaine	1–2 M
<i>Final volume</i>	1 ml

10 x Klenow buffer: 100 mM Tris-HCl (pH 7.5), 50 mM MgCl₂

- 1) Denature the reaction mix for 5 min at 95 °C.
- 2) Place on ice for 5 min.
- 3) Add DTT to a final concentration of 2.5 mM.
- 4) Add 100 U DNA polymerase I, Large Fragment (Klenow; New England Biolabs).
- 5) Incubate at 37 °C for 1 h.
- 6) Ethanol-precipitate, wash with 70% ethanol, and air-dry the generated dsDNA.
- 7) Use the dsDNA pellet as template in a 1 ml *in vitro* transcription reaction (Section 51.4.3). If radiolabeled RNA is used in the *in vitro* selection procedure, an appropriate amount of an α -³²P-labeled NTP may be included in the reaction.

51.4

Enzymatics

In this section, the enzymatic part of *in vitro* selection experiments is discussed, that is, the generation of input material for the following rounds of selection. For RNA selections, amplification is achieved by a series of enzymatic reactions including reverse transcription (RT), PCR, and *in vitro* transcription (Figure 51.1). These reactions convert the selected RNA species into cDNA, into dsDNA, and back again into RNA.

Although the exclusive goal of the amplification procedure is to generate sufficient material for the next round of selection, one should be aware that in practice every enzymatic step introduces an unwanted selection pressure beside the intended selection constraints. Care has to be taken to keep this pressure at a minimum. Depending on their nucleotide sequence, different molecules are amplified more or less efficiently by a given enzymatic system. It can be assumed that molecules that form complex and rigid 3D structures may serve as poor templates during the enzymatic procedures. These restrictions may limit or even reduce the library complexity and can hardly be eliminated. One possibility to counteract, however, is to include nucleic acid melting-point-lowering agents in the reactions, for example, dimethyl sulfoxide, betaine, or commercially available solutions such as the MasterAmp PCR Enhancer (Epicentre Biotechnologies) or the Qiagen Q-Solution. Such reagents help to destabilize secondary structures, thus reducing

the amount of break-off products. For the same reason, the highest possible temperature should be applied during annealing in the PCR and RT reactions.

Exemplary protocols that usually give satisfactory results for RT, PCR, and *in vitro* transcription are presented in this section. Various aspects of these protocols have undergone optimization and should therefore serve as a good starting point, although they should, by no means, be taken as a final protocol. Different pools require different conditions for optimal amplification. Likewise, different binders within a given pool may vary in their amplification efficiency, which makes it necessary to find conditions under which as many pool variants as possible are successfully amplified. Thus, the reaction conditions for every step in the amplification protocol should be adjusted to support the maximum level of diversity in the pool.

51.4.1

Reverse Transcription

RT is the first enzymatic step that follows the separation in bound and unbound RNA fractions. The most commonly used types of reverse transcriptase are avian myeloblastosis virus (AMV) and Moloney murine leukemia virus (M-MLV) RTase, both lacking 3'→5' exonuclease activity. The exemplary protocol given below uses SuperScript® II RTase (Life Technologies Invitrogen), an engineered version of M-MLV RTase with reduced RNase H activity and increased thermal stability.

Protocol: Reverse Transcription

The eluted and purified RNA is annealed to the reverse primer as follows:

RNA	0.1–5 pmol
Reverse primer	100 pmol
Betaine	2–3 M
<i>Final volume</i>	10 µl

- 1) Denature at 95 °C for 5 min.
- 2) Place on ice for 5 min.
- 3) Add the following components:

5 x First strand buffer	4 µl
dNTPs	500 µM each
DTT	10 mM
SuperScript® II RTase (Invitrogen; 200 U µl ⁻¹)	200 U
<i>Final volume</i>	20 µl

5 x First strand buffer: 250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM MgCl₂ (Invitrogen)

- 4) Mix and incubate at 42 °C for 30 min.

- 5) Inactivate the RTase by incubation at 65 °C for 10 min.
- 6) Ethanol-precipitate; wash with 70% ethanol, and air-dry the pellet.
- 7) Dissolve the nucleic acid pellet in 10 µl RNase-free H₂O.

51.4.2

Polymerase Chain Reaction

The purpose of the PCR step is the production of dsDNA as a transcription template for the next selection round. The T7 promoter is attached to the selected sequences during this step. PCR protocols that are widely used for *in vitro* selections do not differ significantly from standard protocols. However, since selections involve many cycles of amplification from the first to the last selection round, it is imperative to keep the number of PCR cycles per round as low as possible, thus avoiding the generation of amplification artifacts. As for every PCR experiment, the primers have to be carefully chosen to avoid the formation of potential primer dimers. Melting temperatures of the reverse (3') primer and the hybridizing part of the forward (5') primer should match and permit an annealing temperature of at least 50 °C.

Protocol: PCR

cDNA	0.05–1 pmol
10 x <i>Taq</i> buffer	10 µl
MgCl ₂	2.5 mM
5' Primer	1–3 µM
3' Primer	1–3 µM
dNTPs	200 µM each
Betaine	1–2 M
<i>Taq</i> DNA polymerase (Invitrogen; 5 U µl ⁻¹)	5 U
<i>Final volume</i>	100 µl

10 x *Taq* buffer: 200 mM Tris-HCl (pH 8.4), 500 mM KCl (Invitrogen)

As input cDNA, 20–100% of the reverse-transcribed RNA should be used (Section 51.4.1). Optimally, not more than 1 pmol cDNA should be amplified per 100 µl reaction volume. As already mentioned, the cycling parameters will depend on the pool and should be optimized with the initial pool. The following protocol may serve as a starting point for many pools:

1	94 °C	2 min	Denaturing
2	94 °C	1 min	Denaturing
3	50–72 °C	1 min	Annealing
4	72 °C	1 min	Polymerization
5	72 °C	5 min	Polymerization

Repeat steps 2–4 for 6–20 cycles. Steps 1 and 5 are performed only once, at the beginning (step 1) and end (step 5) of the PCR.

The number of cycles needed to generate enough product for the following *in vitro* transcription will mainly depend on the amount of template. To avoid performing more cycles than necessary, and thus risking the introduction of artifacts, it is better to start with a small number of cycles and only perform additional cycles if required (to be checked by polyacrylamide gel electrophoresis (PAGE)). A more systematic approach is performing multiple analytical PCR reactions with a small cDNA aliquot to determine the number of cycles required to obtain a lower limit amount of product, followed by the preparative PCR under those conditions.

51.4.3

***In Vitro* Transcription**

In vitro transcription of the dsDNA templates produced in the PCR step finishes the amplification part of the selection round. T7 RNA polymerase can produce up to 100 RNA copies from each dsDNA template. Therefore, the transcription step contributes significantly to the amplification efficiency. A 100 μl reaction containing 100 pmol dsDNA template typically yields 2–4 nmol RNA, with even higher amplification factors for lower template concentrations. The reaction is relatively time-consuming and is therefore most conveniently incubated overnight; however, yields obtained after 3–6 h will be sufficient in most cases. The following protocol works reliably and can be applied to most pools.

Protocol: *In Vitro* Transcription

10 x Reaction buffer	10 μl
dsDNA template	10–100 pmol
DTT	10 mM
NTPs	4 mM each
BSA	120 $\mu\text{g ml}^{-1}$
Betaine	1–2 M
(α - ^{32}P)ATP	As required
T7 RNA polymerase (Promega; 50 U μl^{-1})	100 U
RNase out (Invitrogen; 40 U μl^{-1})	40 U
<i>Final volume</i>	100 μl

10 x Reaction buffer: 800 mM HEPES-KOH (pH 7.5), 220 mM MgCl_2 ,
10 mM spermidine

- 1) Incubate at 37 °C for 3–16 h.
- 2) Add 20 U RNase-free DNase I (Sigma-Aldrich) to digest the dsDNA template.
- 3) Incubate for 15 min at 37 °C.
- 4) Add EDTA to 20 mM and incubate for 10 min at 65 °C to terminate the reaction.
- 5) Purify the RNA by denaturing 8% PAGE.

51.5

Partitioning

The process of separating desired from undesired molecules is the central part of an *in vitro* selection experiment. As depicted in Figure 51.1, the input population is split up in different subpopulations:

- Unwanted functional species that are removed by pre- or counterselection
- Non-binding and weakly binding
- Selected species that bind tightly to the target and can be amplified for the next selection round after elution

Every selection is started by denaturation and refolding of the RNA pool under buffer and temperature conditions to be used during the selection. This step permits the individual RNA molecules to fold into equilibrium structures. The RNA population is then loaded onto the pre-column. In the case of a simple preselection, the RNA is passed through a column containing non-derivatized matrix material, which retains matrix-binding species. Counterselection can be applied to ensure that the desired binders recognize the target, but not a structurally related molecule. The column used for counterselection will be derivatized with the structurally related molecule instead of the target and will only let non-binding species pass through, which are then subjected to selection for target binding.

Separation of RNA–target complexes from non-binding RNAs is difficult in selection experiments against small targets: the physicochemical differences between the free RNA molecules (typically around 30 kDa) and a complex of RNA and target are marginal in case of targets at least one order of magnitude smaller than the RNA. Therefore, partitioning is often carried out “on-column” using pre-immobilized target molecules.

Another approach takes advantage of the (strept)avidin/biotin interaction (Section 51.2.2), using biotinylated target molecules that are fished out of the mixture with immobilized streptavidin after the binding equilibrium between RNA pool and biotinylated target has been reached. A benefit of this strategy is that the concentration of the binding partners can be freely adjusted, thus enabling the researcher to establish more stringent conditions to favor selection of binders with high affinity. Another potential advantage is the possibility of extending the interaction time between pool and target, while keeping the contact time to the matrix short. An extended incubation time is necessary to reach binding equilibrium when the concentration of target binders is low. This usually pertains to the first selection rounds and possibly also to later rounds when target concentrations are intentionally lowered to favor the selection of binders with highest affinity.

A short contact time between pool and column material lessens the chance to enrich unwanted matrix binders. Ideally, target-binding species are retained on the affinity chromatography material, whereas non-binding RNAs are found in the unbound fraction. By washing under selection conditions, non-binding and weakly

binding species are removed; slightly more stringent washing may help reduce background binding.

Finally, the bound RNAs must be eluted from the immobilized target. Basically, two principles can be exploited to achieve this: (i) application of denaturing conditions and (ii) competitive affinity elution by introducing free (non-immobilized) target molecules. For the latter, the matrix is incubated with an excess of free target, which displaces the immobilized target and thus competitively elutes the bound nucleic acids, leaving the 3D structure of the RNA intact. Since the binding of an aptamer to its target depends on the aptamer's 3D structure and on the non-covalent interactions between both binding partners, denaturing agents such as urea or guanidinium thiocyanate can be alternatively used to elute even those molecules that bind to the target with very high affinity. Such molecules might get lost when applying competitive affinity elution because the off-rate for binders of said quality can be so slow that equilibrium may not be reached under practical experimental conditions.

Stringency is a critical selection parameter. In order to obtain highly affine aptamers, it is necessary to continuously adjust the conditions during the consecutive selection rounds by starting with relatively high target concentrations and decreasing the target level in the course of the selection. This supports the enrichment of highly affine binders, whereas the chance for weak binders to survive the selection process is minimized. Modifying the washing parameters (volume, time, and ionic strength) is an additional possibility to increase stringency. All these parameters must be adjusted according to the proportion of RNA specifically bound to the target. The bound fraction can be measured by using radioactively labeled pools. The general rule is that as soon as the amount of the bound molecules increases relative to the previous round, the stringency in the next round is increased by lowering the target concentration and/or intensifying the washing procedure.

Special attention must also be given to the conditions under which binding is performed; since the structure of aptamers is temperature-dependent and/or ion-dependent, it is essential to perform the selection process at the relevant temperature and in a buffer mimicking the conditions under which the aptamers shall be eventually applied. If functionality at or near physiological conditions is required, phosphate, Tris, or HEPES buffers at around 20 mM and pH 7.4 with 150 mM NaCl and 5 mM KCl can be used. The divalent cations Mg^{2+} and Ca^{2+} , which usually are included at about 1 mM in a physiological selection buffer, are of special relevance for nucleic acid structure formation. Finally, a detergent may be included at concentrations between 0.005 and 0.5%. The presence of detergents in the selection process is beneficial as they reduce unwanted hydrophobic interactions between RNA molecules and the selection matrix.

In the following, three exemplary protocols for the selection process are given. In protocol A, the concept of pre- or counterselection is exemplified; protocol B describes the selection against preimmobilized target and competitive affinity elution; and finally, protocol C details the steps of a selection against a target in solution combined with a denaturing elution. The composition of solutions we have used is specified in Table 51.1.

Table 51.1 Buffers and solutions used for partitioning.

Selection buffer (SB)	20 mM Tris-HCl, pH 7.4; 150 mM NaCl; 5 mM KCl; 1 mM MgCl ₂ ; 1 mM CaCl ₂ ; 0.005% Triton X-100
10 x SB without Mg ²⁺ /Ca ²⁺	200 mM Tris-HCl pH 7.4; 1.5 M NaCl; 50 mM KCl
10 x Mg ²⁺ /Ca ²⁺	10 mM MgCl ₂ ; 10 mM CaCl ₂
Affinity elution solution (AES)	0.1–1 mM target in SB
Denaturing elution solution (DES)	4 M Guanidinium thiocyanate; 25 mM sodium citrate, pH 7; 0.5% sarcosyl; 100 mM β -mercaptoethanol

Protocol 1: Pre- or Counterselection

- 1) Fill pure chromatography matrix (Section 51.2.1) without a target into a disposable column. Use the same matrix amount as in the main selection step. A simple preselection is carried out with non-derivatized column material to remove unwanted matrix-binding RNA species; for a counterselection, the target-related molecule that should not be recognized by the desired aptamers is coupled to the matrix.
- 2) Rinse the column with 10 column volumes of SB.
- 3) Denature the nucleic acid pool in SB without Mg²⁺/Ca²⁺ and without Triton X-100 for 1–5 min at 70–95 °C and place on ice.
- 4) Bring the pool to SB conditions with 10 x Mg²⁺/Ca²⁺ and 0.5% Triton X-100 and refold for >5 min at the selection temperature.
- 5) Load the pool onto the column; collect and pool the flow-through and two column volumes of SB (yielding the input RNA for the selection step). A distinct loss of material during this step is tolerated to minimize the risk of including matrix binders.

Protocol 2: Covalently Immobilized Target/Affinity Elution

- 1) Fill an appropriate amount of the chromatographic affinity matrix (Section 51.2.1), derivatized with the target at the desired concentration, into a small expendable column. See Table 51.2 for an example of a typical selection course.
- 2) Equilibrate the affinity matrix with 10 column volumes of SB.
- 3) Denature and refold the RNA pool as described in Protocol 1, steps 3–4, and apply it to the column. See Table 51.2 as a guideline for appropriate amounts of target and pool. Beginning with round 3, the pool is preselected against the pure matrix without target before every selection step (Protocol 1); as soon as the pool is pre- or counterselected, the denaturing/refolding procedure will only be carried out before the pre- or counterselection step.
- 4) Wash the affinity column with SB until only around 1% of the pool remaining on the matrix. Yields of 1% or less can be expected in the initial rounds; more RNA will bind as soon as target-binding sequences become enriched during

Table 51.2 Exemplary selection: matrix with pre-immobilized target.

Round	c_{Target} on matrix (μM)	Matrix volume (μl)	n_{Target} (nmol)	n_{RNA} (nmol)	Ratio target/RNA	RNA bound (%)
1	100	800	80	2	40	1
2	100	400	40	2	20	1.3
3	100	200	20	2	10	4
4	100	50	5	1	5	8
5	10	100	1	1	1	4
6	10	50	0.5	1	0.5	12

the course of the selection (Table 51.2). Washing conditions may be changed toward increased stringency using high-salt conditions (e.g., 3 M NaCl in SB) or buffered 1 M urea solution.

- 5) Pass three column volumes of affinity elution solution (AES) (Table 51.1) through the column and collect the eluate. If an elevated proportion of pool molecules remains bound to the matrix, the column can be incubated with AES for a period of time, instead of simple gravity flow-through, to increase elution efficiency. Alternatively, the AES with an at least 10 x higher concentration of free target can be used.
- 6) Add 10 μg yeast RNA as carrier and extract the eluate with phenol:chloroform.
- 7) Add 0.1 volume of 3 M sodium acetate, pH 5.2, and 1 volume of ice-cold isopropanol. Incubate at -20°C for 30 min and collect the RNA by centrifugation. Wash the pellet with 70% ethanol and air-dry before redissolving the RNA for enzymatic amplification.

Protocol 3: Biotinylated Target/Denaturing Elution

- 1) Denature and refold the RNA pool as described in Protocol 1, steps 3–4. Beginning with round 3, the pool is preselected against streptavidin beads before every selection step (Protocol 1); as soon as the pool is pre- or counterselected, the denaturing/refolding procedure will only be carried out before the pre- or counterselection step. Furthermore, a binding reaction without target as background control is introduced into the selection scheme from round 3 onwards. Determination of the ratio of RNA binding to the streptavidin matrix for samples with and without biotinylated target (i.e., the signal/control ratio) is helpful for monitoring the progress during the selection and adjustment of the stringency of the process.
- 2) Combine the biotinylated target and the RNA pool in a defined volume of SB to reach appropriate concentrations for pool and target. See Table 51.3 for an example of a typical selection course.
- 3) Let the binding process proceed for 2–16 h at the desired temperature.

Table 51.3 Exemplary selection: target in solution.

Round	n_{Target} ; c_{Target} (nmol); (μM)	Reaction volume (μl)	n_{RNA} ; c_{RNA} (nmol); (μM)	Ratio target/RNA	RNA bound with target (%)	RNA bound without target (%)
1	8; 50	160	4; 25	2	0.9	1.3
2	4; 25	160	2; 12.5	2	1.1	1.1
3	4; 10	400	2; 5	2	1.4	0.8
4	2; 10	200	1; 5	2	4	0.5
5	1; 1	1000	1; 1	1	2	0.5
6	0.5; 0.25	2000	1; 0.5	0.5	3	0.7
7	0.25; 0.25	1000	0.5; 0.5	0.5	4	0.4
8	0.1; 0.1	1000	0.5; 0.5	0.2	2	0.6
9	0.1; 0.1	1000	0.5; 0.5	0.2	9	0.6

- 4) Wash and equilibrate streptavidin agarose beads with 10 matrix volumes of SB.
- 5) Add the preequilibrated streptavidin agarose beads to the binding reaction. For efficient immobilization of RNA–target complexes, use three times the amount of matrix that is necessary to completely immobilize the amount of target present during the binding reaction.
- 6) Incubate for 30 min with agitation to keep the beads in suspension for efficient capture of RNA–target complexes.
- 7) Briefly centrifuge the reaction tube to pellet the beads.
- 8) Withdraw the supernatant and repeatedly wash the beads by resuspending in 10 matrix volumes of SB until around 1% of RNA remains bound to the matrix. Yields of 1% or less can be expected in the initial rounds; more RNA will bind as soon as target-binding sequences become enriched during the course of the selection (Table 51.3).
- 9) Elute the bound RNA by shaking the beads in 5 volumes of denaturing elution solution (DES) (Table 51.1) for 10 min at 37 °C in the presence of 10 μg yeast total RNA.
- 10) Briefly centrifuge to pellet the beads and collect the supernatant containing the eluted RNA; repeat the elution at 55 °C.
- 11) Extract the combined eluates with phenol:chloroform.
- 12) Add 0.1 volume of 3 M sodium acetate, pH 5.2, and 1 volume ice-cold isopropanol. Incubate at –20 °C for 30 min and collect the RNA by centrifugation. Wash the pellet with 70% ethanol and air-dry before redissolving the RNA for enzymatic amplification. The use of isopropanol is imperative here, as guanidinium thiocyanate will coprecipitate when employing ethanol, which can inhibit the following enzymatic reactions.

51.6

Binding Assays

Methods for the determination of the binding characteristics are important for tracking the selection progress, as well as for a detailed analysis of individual binders identified by sequencing of enriched pool clones. Most conveniently, commercial cloning and sequencing services are engaged for this purpose. For detailed information on cloning procedures, the researcher is referred to [17].

In the following, three procedures for the determination of equilibrium dissociation constants of RNA–small molecule complexes are described in detail.

51.6.1

Equilibrium Dialysis

Equilibrium dialysis is a specific application of dialysis, used to study small molecule–macromolecule interactions [18, 19]. The availability of a labeled ligand that is small enough to be dialyzed through a membrane, while the RNA aptamer is excluded, is the prerequisite for the determination of a dissociation constant by equilibrium dialysis. The amount of ligand bound to the macromolecule of interest is then determined by dialyzing the free ligand (loaded into one chamber) through the membrane, while the macromolecules (loaded into the second chamber) as well as macromolecule–ligand complexes are retained on their side of the membrane. It is essential that the concentration of the labeled binding partner is much lower than the RNA concentration, to ensure that the free RNA concentration is not significantly affected by ligand binding: $[RNA] \gg [\text{labeled ligand}]$. An equilibrium dialysis apparatus consists of a pair of chambers with volumes of 50 μl or more, separated by a dialysis membrane with the required molecular weight cutoff. When working with aptamers with a length of more than 50 nt and ligands with less than 2 kDa, a 6–8 kDa cutoff membrane is a good choice.

Protocol

- 1) Denature the RNA in the binding buffer without $\text{Mg}^{2+}/\text{Ca}^{2+}$ and without Triton X-100 for 1–5 min at 70–95 °C and place on ice. The binding buffer used for the experiments should contain at least 100 mM salt to compensate for the Gibbs–Donnan effect, which can perturb the equilibrium.
- 2) Bring the pool to binding buffer conditions with 10 x $\text{Mg}^{2+}/\text{Ca}^{2+}$ and 0.5% Triton X-100 and refold for at least 5 min at the selection temperature.
- 3) Prepare a dilution series of the RNA in binding buffer. The range of the dilution series depends on the expected dissociation constant of the examined RNA; for low micromolar binding, 0.05–50 μM represents a suitable range.
- 4) Prepare a solution of radioactively labeled ligand (not exceeding 50 nM) in binding buffer. The required ligand concentration depends on the specific radioactivity; choose a ligand concentration that is as low as possible.

- 5) Set up the dialysis units, load RNA dilution series into chambers on one side, and labeled ligand into the second chambers. Include a non-RNA control to correct for background (= small, non-diffusing fraction of labeled ligand).
- 6) Incubate the dialysis units for 24 h at the desired temperature. Rotating the chambers at around 10 rpm shortens the time required to reach equilibrium.
- 7) Measure radioactivity in the ligand (r_{LC}) and RNA (r_{RC}) chambers of each dialysis unit by scintillation counting.

The fraction of bound ligand is calculated using $F = (r_{RC} - r_{LC}) / (r_{RC} + r_{LC})$. Assuming that (i) the RNA–ligand dissociation constant K_d is identical for a labeled and an unlabeled ligand and (ii) complex stoichiometry is 1 : 1, the generated data can be fitted by a non-linear least squares regression analysis to a standard binding equation to obtain the K_d , using a software such as GraFit (Erithacus Software Ltd) [20].

51.6.2

Equilibrium Filtration Analysis

Equilibrium filtration analysis is a rapid alternative to equilibrium dialysis and takes advantage of the same physical phenomenon [21, 22]. It is even more convenient, as disposable, commercially available ultrafiltration devices can be used. Microcon centrifugal filter devices with Ultracel[®] YM-10 membranes (nominal molecular weight cutoff: 10 kDa; Millipore) are appropriate for this purpose. As for equilibrium dialysis, the concentration of the labeled binding partner must be much lower than the RNA concentration.

Protocol

- 1) Denature the RNA in the binding buffer without Mg^{2+}/Ca^{2+} and without Triton X-100 for 1–5 min at 70–95 °C and place on ice.
- 2) Bring the pool to binding buffer conditions with 10 x Mg^{2+}/Ca^{2+} and 0.5% Triton X-100 and refold for at least 5 min at the selection temperature.
- 3) Prepare a dilution series of the RNA in binding buffer. The range of the dilution series depends on the expected dissociation constant of the examined RNA–ligand pair; for low micromolar binding, 0.05–50 μM represents a suitable range.
- 4) Mix 225 μl of the RNA dilutions, including a non-RNA control, with 25 μl of radiolabeled ligand in binding buffer. The final concentration of ligand should not exceed 50 nM.
- 5) Equilibrate for 2 h at the desired temperature.
- 6) Load 200 μl onto a Microcon YM-10 spin filter.
- 7) Centrifuge at 13 000 g for 30 s to saturate the membranes.
- 8) Transfer filter unit into a fresh collection tube and discard the old tube.

- 9) Centrifuge at 13 000 g for 90 s. In order to skew the aptamer and ligand–aptamer complex concentration in the retentate not more than necessary, the filter units must not be centrifuged longer than required to obtain enough filtrate for determination of radioactivity (about 25–50 μ l).
- 10) Determine the radioactivity in 25 μ l aliquots derived from the top (retentate; r_T) and bottom (filtrate; r_B).

Only unbound ligand molecules pass through the membrane, while ligand–aptamer complexes, as well as free aptamer molecules, are retained. Hence, bottom counts correspond to the free ligand concentration, whereas the top counts represent total (bound plus free) ligand concentration. The fraction of bound ligand is calculated using $F = (r_T - r_B)/(r_T + r_B)$. A K_d can be calculated from the generated data under the same assumptions and with the same software as described in Section 51.6.1.

51.6.3

Isocratic Competitive Affinity Chromatography

The method of isocratic affinity chromatography allows one to determine dissociation constants based on the interaction of an aptamer with an affinity column in the presence *versus* absence of free ligand. The aptamer of interest is loaded onto the affinity column in the absence and presence of free ligand, followed by elution of the aptamer with pure binding buffer or with binding buffer containing free ligand, respectively (Figure 51.4). From the volumes of buffer with and without free ligand necessary to elute the aptamer, the dissociation constant of the aptamer–ligand complex can be calculated [10, 23, 24].

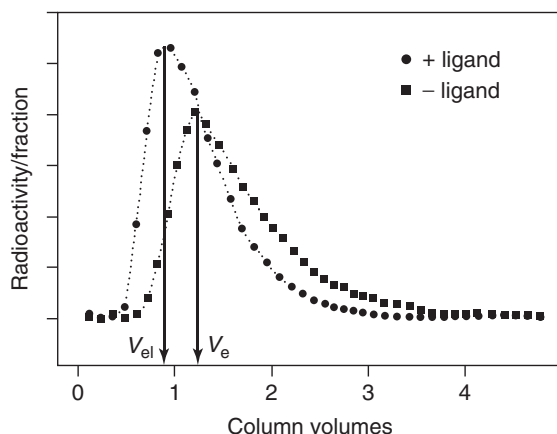


Figure 51.4 Isocratic competitive affinity chromatography. Exemplary elution profile of labeled aptamer from an affinity column. V_{el} , median elution volume in the presence of free ligand; V_e , median elution profile in the absence of free ligand.

Protocol

- 1) Denature and refold two aliquots of 1–10 pmol ^{32}P -labeled aptamer and a similar amount of unselected, ^{32}P -labeled pool RNA in the binding buffer without $\text{Mg}^{2+}/\text{Ca}^{2+}$ and without Triton X-100 for 1–5 min at 70–95 °C and place on ice.
- 2) Bring the three solutions to binding buffer conditions with 10 x $\text{Mg}^{2+}/\text{Ca}^{2+}$ and 0.5% Triton X-100 and refold for at least 5 min at the selection temperature.
- 3) Add ligand for elution (concentration: plus/minus one order of magnitude around the expected K_d) to the first aptamer aliquot and an identical volume of binding buffer to the second aptamer aliquot and the pool RNA.
- 4) Incubate at room temperature for 2 h to obtain binding equilibrium before loading onto an affinity column equilibrated with binding buffer.
- 5) Apply the sample containing ligand and aptamer to the column and elute the RNA isocratically with several column volumes of binding buffer containing the ligand in the chosen concentration.
- 6) Collect the flow-through in fractions and measure radioactivity in the fractions to determine the elution volume (V_{el}).
- 7) Equilibrate the column and repeat the procedure with the second aptamer aliquot (without ligand). Elute isocratically with pure binding buffer (V_e).
- 8) Equilibrate the column and repeat the procedure with the unselected pool RNA. Elute isocratically with pure binding buffer (V_n).

The dissociation constant K_d can be calculated as follows:

$$\text{Affinity for immobilized ligand: } K_d = L_c[V_n/(V_e - V_n)]$$

$$\text{Affinity for ligand in solution: } K_d = L[(V_{\text{el}} - V_n)/(V_e - V_{\text{el}})]$$

where

L_c = the concentration of immobilized affinity ligand within the column bed,

L = the concentration of free affinity ligand used to elute the RNA,

V_{el} = the median elution volume of RNA eluted in the continuous presence of free ligand (Figure 51.4),

V_e = the median elution volume measured in the absence of free ligand in the buffer (Figure 51.4),

V_n = the volume at which a random RNA population of similar molecular size, having no column interaction, would elute (void volume).

References

1. Ellington, A.D. and Szostak, J.W. (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
2. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
3. Majerfeld, I. and Yarus, M. (1994) An RNA pocket for an aliphatic hydrophobe. *Nat. Struct. Biol.*, **1**, 287–292.

4. Majerfeld, I. and Yarus, M. (1998) Isoleucine: RNA sites with associated coding sequences. *RNA*, **4**, 471–478.
5. Wallace, S.T. and Schroeder, R. (1998) *In vitro* selection and characterization of streptomycin-binding RNAs: recognition discrimination between antibiotics. *RNA*, **4**, 112–123.
6. Burke, D.H., Hoffman, D.C., Brown, A., Hansen, M., Pardi, A., and Gold, L. (1997) RNA aptamers to the peptidyl transferase inhibitor chloramphenicol. *Chem. Biol.*, **4**, 833–843.
7. Rink, S.M., Shen, J.-C., and Loeb, L.A. (1998) Creation of RNA molecules that recognize the oxidative lesion 7,8-dihydro-8-hydroxy-2'-deoxyguanosine (8-oxodG) in DNA. *Proc. Nat. Acad. Sci. U.S.A.*, **95**, 11619–11624.
8. Koizumi, M. and Breaker, R.R. (2000) Molecular recognition of cAMP by an RNA aptamer. *Biochemistry*, **39**, 8983–8992.
9. Holeman, L.A., Robinson, S.L., Szostak, J.W., and Wilson, C. (1998) Isolation and characterization of fluorophore-binding RNA aptamers. *Folding Des.*, **3**, 423–431.
10. Connell, G.J., Illangsekare, M., and Yarus, M. (1993) Three small ribooligonucleotides with specific arginine sites. *Biochemistry*, **32**, 5497–5502.
11. Sassanfar, M. and Szostak, J.W. (1993) An RNA motif that binds ATP. *Nature*, **364**, 550–553.
12. Burke, D.H. and Gold, L. (1997) Aptamers to the adenosine moiety of S-adenosyl methionine: structural inferences from variations on a theme and the reproducibility of SELEX. *Nucleic Acids Res.*, **25**, 2020–2024.
13. Tang, J. and Breaker, R.R. (1997) Rational design of allosteric ribozymes. *Chem. Biol.*, **4**, 453–459.
14. Wlotzka, B., Leva, S., Eschgfäller, B., Burmeister, J., Kleinjung, F., Kaduk, C., Muhn, P., Hess-Stumpp, H., and Klussmann, S. (2002) *In vivo* properties of an anti-GnRH Spiegelmer: an example of an oligonucleotide-based therapeutic substance class. *Proc. Nat. Acad. Sci. U.S.A.*, **99**, 8898–8902.
15. Rizo, J. and Gierasch, L.M. (1992) Constrained peptides: models of bioactive peptides and protein substructures. *Annu. Rev. Biochem.*, **61**, 387–418.
16. Leva, S., Lichte, A., Burmeister, J., Muhn, P., Jahnke, B., Fesser, D., Erfurth, J., Burgstaller, P., and Klussmann, S. (2002) GnRH binding RNA and DNA Spiegelmers: a novel approach toward GnRH antagonism. *Chem. Biol.*, **9**, 351–359.
17. Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning – A Laboratory Manual*, 3rd edn, Cold Spring Harbor Laboratory Press, New York.
18. Williams, K.P., Liu, X.-H., Schumacher, T.N.M., Lin, H.Y., Ausiello, D.A., Kim, P.S., and Bartel, D.P. (1997) Bioactive and nuclease-resistant L-DNA ligand of vasopressin. *Proc. Nat. Acad. Sci. U.S.A.*, **94**, 11285–11290.
19. Nieuwlandt, D., Wecker, M., and Gold, L. (1995) *In vitro* selection of RNA ligands to substance P. *Biochemistry*, **34**, 5651–5659.
20. Connors, K.A. (1987) *Binding Constants*, John Wiley & Sons, Inc., New York.
21. Jenison, R.D., Gill, S.C., Pardi, A., and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425–1429.
22. Davis, J.H. and Szostak, J.W. (2002) Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc. Nat. Acad. Sci. U.S.A.*, **99**, 11616–11621.
23. Scarabino, D., Crisari, A., Lorenzini, S., Williams, K., and Tocchini-Valentini, G.P. (1999) tRNA prefers to kiss. *EMBO J.*, **18**, 4571–4578.
24. Ciesiolka, J., Gorski, J., and Yarus, M. (1995) Selection of an RNA domain that binds Zn^{2+} . *RNA*, **1**, 538–550.

52

SELEX Strategies to Identify Antisense and Protein Target Sites in RNA or hnRNP Complexes

Martin Lützelberger, Martin R. Jakobsen, and Jørgen Kjems

52.1

Introduction

The recognition of RNA elements in the complex environment of the cell is a key issue, both for studying the basic molecular biology of the cell and for the development of therapeutic strategies that will interfere with gene expression. A cornerstone in this line of research is the Systematic Evolution of Ligands by EXponential enrichment (SELEX) procedure where ribonucleic acids that bind tightly to a ligand of interest are identified through successive rounds of binding, fractionation, and amplification of an RNA library. In SELEX, as originally developed, the library consists of 10^{14} – 10^{15} random sequences [1]. An extension of this method, named genomic SELEX, where natural DNAs are used as input sequences, was originally invented by Singer *et al.* [2]. Both methods involve a PCR amplification step which requires that the amplified sequences are flanked by primer annealing sites of fixed sequence. A T7 RNA polymerase promoter is included in one of the flanking sequences in order to produce RNA by *in vitro* T7 transcription. The protocol described here resembles the genomic SELEX protocol described by Singer *et al.* [2], but with a notable difference that the libraries produced here contain RNAs with a fixed size of 20 nt.

The RNA library may contain sequences derived from whole genomic DNA, cDNA, or smaller genetic entities such as plasmids or viruses. In addition to the library construction, we include protocols for two central applications:

- 1) The characterization of accessible target sites for antisense oligonucleotide annealing in highly structured RNA or heterogeneous nuclear ribonucleoprotein (hnRNP) complexes.
- 2) The characterization of binding sites for a specific protein within RNA derived from a single gene or complete genomes.

However, the RNA library generated by this approach may directly or in a modified form have many other potential applications.

52.1.1

Applications for Antisense

The application of various types of antisense technologies, such as antisense oligonucleotides, ribozymes, DNAzymes and RNAi, may be severely compromised by RNA structure and RNA binding proteins in the cell. In the absence of any detailed structural information on the target mRNA, antisense design has traditionally been based on the principle of trial and error. Finding the best target can therefore be a cumbersome and expensive exercise. Several theoretical and practical approaches to determine accessible regions in RNA have been described (reviewed in [3, 4]). The most popular approach involves the use of chemicals and enzymes reactive to RNA to elucidate the exposed regions (described in Chapters 10, 12, 13 and 15), but this method has proven inadequate in terms of predicting efficient target sites for antisense oligonucleotides. Approaches to measure the susceptibility to oligonucleotide annealing include the mapping of hybrids formed between the mRNA and random DNA oligonucleotides using RNase H [5–7] or primer extension [8]. Alternatively, a method has been described where labeled mRNA is annealed to an immobilized array of a complete set of antisense oligonucleotides, but this is costly and not practical if multiple mRNAs are to be investigated simultaneously [9]. In the protocol below, we describe the selection of 20 nt long sense or antisense oligoribonucleotides from a library derived from plasmids or whole genomes that are able to bind efficiently to a particular target RNA.

The advantages of this approach are that the length of the RNAs in the library is similar to that of therapeutically relevant antisense oligonucleotides and that the selection can be performed under any given condition. Moreover, the use of natural sequences as input DNA lowers the complexity of libraries significantly, enabling the selection to occur in a few rounds and in relatively small volumes.

52.1.2

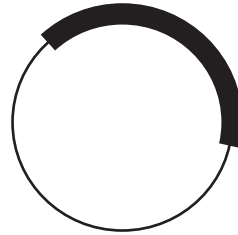
Selecting Protein Binding Sites

The realization that the expression of many mammalian genes is highly regulated at the level of RNA processing by a large number of proteins and hnRNP complexes has called for improved high-throughput methods to identify the RNA sequences involved. The genomic SELEX method previously described [2] is useful for this purpose [10], but the larger and variable size of the RNA fragments (100–300 nt) in these libraries makes it harder to interpret the result. For most proteins that recognize a primary RNA sequence motif, the 20-mer sequences in our library are sufficient to cover the entire or a substantial portion of the binding site.

52.2

Construction of the Library

The key step in generation of the library, outlined in Figure 52.1, are as follows:



Random fragmentation of genomic or plasmid DNA

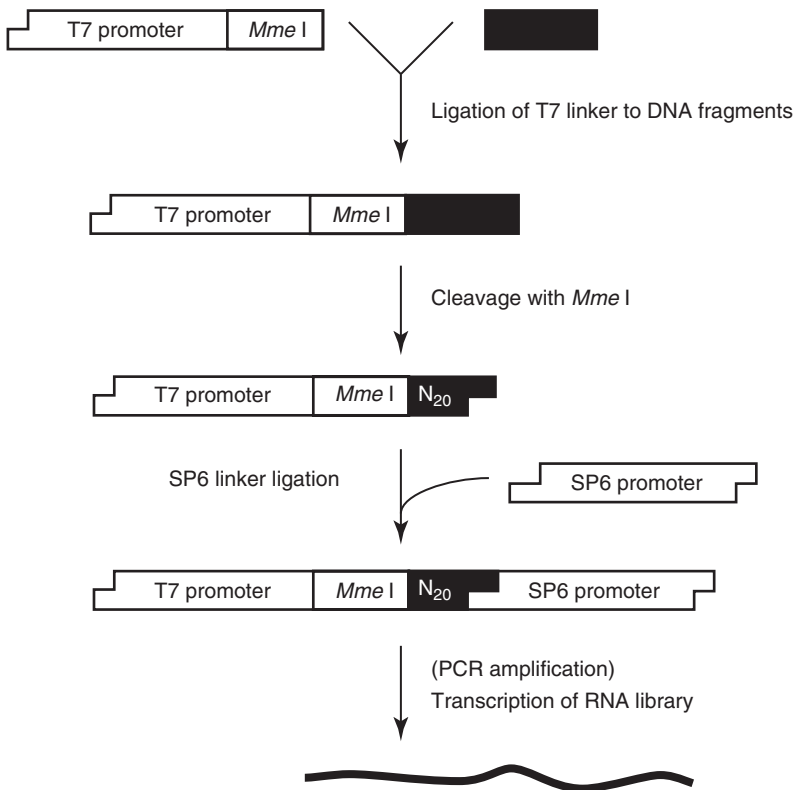


Figure 52.1 Flowchart diagram illustrating the key steps of library construction. Fragments of genomic or plasmid DNA are generated using ultrasound or DNase I treatment. The fragments are ligated to a T7 linker, consisting of the T7 promoter sequence and the recognition site for the restriction enzyme *Mme* I. After digestion

of the ligated products with *Mme* I, which cuts 20/18 bp downstream of its recognition site, fragments are ligated to an SP6 linker containing the SP6 promoter sequence. The library can be amplified by PCR or directly used for T7 transcription, producing transcripts of 66 nt.

- 1) Generation of 200- to 1000-bp fragments by random degradation of the input DNA, which may be either plasmid DNA or genomic DNA.
- 2) Ligation of an upstream linker to the DNA fragments, containing a T7 promoter sequence and an *Mme* I site.
- 3) Cleavage by the restriction enzyme *Mme* I, leaving 20 nt of genomic DNA attached to the T7 linker.
- 4) Ligation of a downstream linker, in our case containing an SP6 promoter sequence.
- 5) *In vitro* transcription, using the ligated product as template.

52.2.1

Generation of Random DNA Fragments from Genomic or Plasmid DNA

To break internucleotide bonds, the DNA can be treated with ultrasound or DNase I. The advantage of ultrasound is that it breaks internucleotide bonds more randomly than DNase I which binds to the minor groove with a slight preference to cleave within A/T-rich sequences [11]. Thus, we recommend sonication as the method of choice for the generation of DNA fragments. However, DNase I treatment may sometimes be preferred for small plasmid DNA, which is usually more difficult to break by ultrasound treatment. If a DNase I reaction buffer containing Mn^{2+} is used, blunt ends are produced and subsequent Klenow polymerase treatment is unnecessary.

If the library is constructed from a DNA fragment inserted into a plasmid, it is usually not critical to generate a pool of molecules that represent all possible 20 nt stretches of the insert. Taking into account that some material will be lost during enzymatic reactions, phenol extraction, and precipitation, about 10–20 μ g of plasmid DNA are a good starting point.

If the library is made from cDNA or genomic DNA, the complete coverage of the genome must be ensured. Hence, the pool of molecules should contain a sufficient set of overlapping subfragments, so that each nucleotide in the genome is represented by more than one molecule. Thus, for a library made from human genomic DNA, at least 20 μ g should be used as starting material: Given that the human genome has about 3×10^9 bp, not less than 3×10^9 fragments should be generated to have each nucleotide of the genome represented at least once at a 5'-terminal position. If the fragments produced by sonication have an average size of 600 bp, 10 μ g of them would equal 26 pmol or 1.6×10^{13} molecules, which is a 10 000-fold excess of ends over the total number of base pairs in the human genome. We suggest using 50 μ g of genomic DNA for the construction of a human SELEX library.

52.2.2

Preparing RNA Libraries from Plasmid, cDNA, or Genomic DNA

When a sufficient amount of DNA fragments have been produced, they are ligated to an upstream linker that contains a T7 promoter sequence and a recognition

site for the restriction enzyme *Mme* I (Figure 52.1). Since the ligation efficiency of blunt-end fragments largely depends on their concentration, we suggest to adjust the ratio between linker DNA and genomic fragments in such a way that for each linker molecule at least two genomic DNA fragments are present. This will “drive” the reaction and decrease the formation of linker dimers.

Mme I belongs to the class-II restriction endonucleases and has the unique feature to cleave DNA 20/18 nt downstream of its non-palindromic recognition sequence (TCCRAC), thereby generating a 2-nt (NN) 3'-overhang [12, 13]. The fragments cleaved with *Mme* I are gel-purified and subsequently ligated to a downstream linker containing an SP6 promoter sequence. The SP6 promoter is included to be able to produce RNA from the antisense strand. However, for the applications described here, it is not used as a promoter.

Both the T7 and SP6 linkers have a 1-nt 3'-overhang (see Protocols) to avoid the formation of linker concatemers. This enhances ligation efficiency and prevents the inclusion of linker sequences into the library.

Depending on the concentration, the purified ligation products may be directly used for T7 transcription. If the library is amplified by PCR, conditions must be adjusted carefully, that is, the amplification should be done with as few cycles as possible to avoid creating a bias. The inclusion of enhancing agents such as dimethyl sulfoxide (DMSO), which facilitates strand separation, might be advantageous for the amplification of G/C-rich sequences. In Figure 52.2 (lane 5), an example of a PCR-amplified library is given. It is usually not necessary to purify the PCR products before T7 transcription, provided that products of discrete size have been obtained.

With Protocol 3, about 20–40 pmol RNA can be produced from a single 50 μ l PCR reaction. If a higher yield is required, the reaction may be scaled up accordingly. It is convenient to include [α - ^{32} P]UTP in the transcription reaction to quantitate the yield and to trace the RNA in the subsequent selections.

52.3

Identification of Optimal Antisense Annealing Sites in RNAs

To characterize sites in RNA or RNP complexes that are susceptible to RNA annealing, an N_{20} RNA library is prepared starting from a plasmid containing the sequence encoding the target RNA as input material. The library is subsequently mixed with the radioactively labeled target RNA or RNP complex under physiological conditions to maintain the native structure in the RNA or RNP target. After a short annealing step, the target RNA or RNP complex is separated on a native gel from the unbound fraction of the library and eluted from the gel. The bound oligoribonucleotides are selectively amplified by RT-PCR using primers complementary to the flanking T7 and SP6 linker sequences.

Since a T7 RNA polymerase promoter sequence is included in the upstream linker, the PCR products can be used directly as templates for the next round of transcription. After three to five successive rounds of selection, the RT-PCR

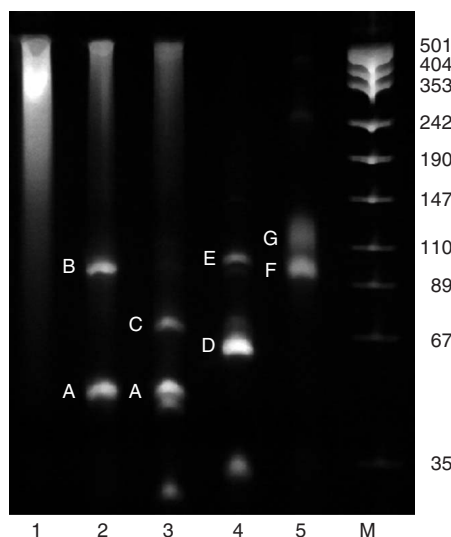


Figure 52.2 An example of a gel showing products of the individual steps in the construction of a human genome-wide library of 20-mer RNAs. Small aliquots of each reaction were taken and separated on a native 6% PAA gel ($1 \times$ TBE). Lane 1, sonicated human genomic DNA. Lane 2, products of the T7 linker ligation reaction; note that the intended ligation products (T7 linker-DNA fragment) are heterogeneous in size and thus do not appear as discrete bands. Lane 3, *Mme* I digestion after T7 linker ligation. Lane 4, products of the SP6 linker ligation reaction. Lane 5, PCR-amplified library.

Lane M, pUC18 vector digested with *Hpa* II. The bands marked with capital letters correspond to: A, T7 linker; B, T7 linker dimer; C, *Mme* I-digested fragments; D, SP6 linker dimer; E, final ligation product (library); F, PCR-amplified library; and G, products formed by reannealing (after the denaturation step) of single-stranded library molecules that differ in their variable parts and therefore form internal loops. The T7 and SP6 linker sequences of these molecules are double-stranded, which is sufficient for *in vitro* transcription.

products are cloned into a plasmid and 50–100 inserts are sequenced. In our hands, three rounds of selection are usually sufficient when using a plasmid insert to generate the input RNA library.

In the case of highly structured RNA, for instance, the HIV-1 5'-untranslated region (5'-UTR), we observed that only 1–2% of the potential antisense oligonucleotides are selected when using the conditions described below [14]. In addition, different sequences enriched in the final pool were found to be equally represented, suggesting that they bound to the target RNA in a non-competitive fashion. To select for oligoribonucleotides with fast binding kinetics, we generally shorten the library-annealing step from 30 min in the initial selection to only 5 min in the final round.

By varying the setup, it is possible to select for oligoribonucleotides that bind to structured or non-structured RNA, RNP complexes, or inhibit protein binding or dimerization of RNA molecules. In Protocol 4, the selection of oligoribonucleotides that bind to structured RNAs of 300–1000 nt is described.

52.4

Identification of Natural RNA Substrates for Proteins and Other Ligands

The selection of optimal RNA binding sites for a particular ligand (e.g., protein) can be performed by mixing the RNA library (derived from plasmid DNA, a cDNA library, or genomic DNA) with purified protein and separating the complexes from the non-bound RNA by filter binding, immunoprecipitation, protein tag-purification, or native gel electrophoresis. Protocol 5 uses the latter approach, which permits one to distinguish between monomeric and multimeric protein complexes. Native gel electrophoresis also has the advantage that an enrichment of RNA molecules with affinity to the purification matrix can be avoided.

In contrast to SELEX procedures using random libraries, our approach requires only three to five rounds to select RNA variants that bind to the ligand with high affinity. For natural RNA binding proteins, the binding buffer and stringency of the selection conditions must be optimized. Addition of competitor RNA during the initial round of selection is usually not required, since the majority of library molecules do not bind to the ligand. However, addition of a 10-fold excess of competitor RNA in the last selection round(s) may increase the stringency of the selection. Reamplification of the RNA pool follows the same scheme as for antisense RNA selection.

It is important to be aware of the limitations associated with an N_{20} library for the selection of RNAs that bind to a protein or RNP. The approach is essentially only applicable to proteins and RNPs that recognize primary RNA sequence motifs not exceeding 20 nt. For ligands that recognize more complex RNA secondary and tertiary structural elements, it is better to use the genomic SELEX protocol described by Singer *et al.* [2].

52.5

Cloning, Sequencing, and Validating the Selected Inserts

To evaluate the library, we recommend sequencing of at least 50 randomly picked clones from the unselected pool. It should be verified that the inserts have the correct length and represent different sequence regions of the input DNA in a balanced manner. Furthermore, the G/C-content of the 20-mers should be evaluated to rule out that a bias has been created by PCR amplification of the library.

Cloning of the RNA inserts can be performed by using the *Pst* I and *Eco* RI sites in the flanking sequences (see Materials). Alternatively, one can use the TA cloning procedure (e.g., using a TOPO TA Cloning Kit from Life Technologies Invitrogen) to clone the PCR fragments directly. If a large number of 20-mers have to be sequenced, it is recommended to blunt-end the RT-PCR products using Klenow polymerase, ligate them into concatemers, and ligate these into a plasmid. This approach will cut down the number of sequencing reactions up to 10-fold.

To ease the sequence analysis, we recommend using the EMBOSS suite [15], which is freely available for UNIX platforms from <http://emboss.sourceforge.net/>.

It contains tools, such as the vectorstrip program, to extract the 20-mers from raw sequence data, which is particularly useful for the analysis of a large number of sequences. For the conversion of the sequence information into aligned sequences of the selected inserts, we recommend the ClustalX toolkit, available from <ftp://ftp.ebi.ac.uk/pub/software/unix/clustalx>.

52.6

Troubleshooting

52.6.1

Sonication of Plasmid DNA does not Yield Shorter Fragments

The sonication of plasmid DNA usually requires more power than that of chromosomal DNA of high molecular weight, because the plasmids are small and covalently closed circular molecules that are less vulnerable to hydrodynamic shearing. If you do not obtain the desired fragment length with ultrasound, use DNase I treatment instead.

52.6.2

Inefficient Ligation

For efficient ligation of the SP6 linker to the *Mme* I fragments, it is crucial that it has a truly random 3'-NN overhang. Most custom-made oligonucleotides are synthesized by solid phase chemistry in the 3'–5' direction, starting with a column containing the 3' nucleotide temporarily immobilized on a solid support. On the basis of our experience, some companies have difficulties in producing oligonucleotides with a true "N" at the 3' end. To avoid problems, make sure that your distributor is able to synthesize such oligonucleotides, otherwise order a set of four primers, each ending with a different nucleotide, and mix them.

The 10 x T4 ligation buffer from New England Biolabs (NEB) already contains 10 mM ATP. Addition of ATP is usually not necessary, but recommended if the 10 x reaction buffer went through a couple of freeze–thaw cycles. To avoid such problems, store the buffer in small aliquots.

The addition of crowding agents, such as polyethylene glycol, to increase the effective concentration of the reactants might also improve the ligation efficiency, especially if the amount of DNA fragments is limited.

52.6.3

Inefficient *Mme* I Digestion

We observed that *Mme* I cleavage is inefficient if the enzyme is used in excessive amounts. Do not use more units than described in Protocol 2. The enzymatic activity of *Mme* I is completely inhibited in the presence of 100 mM KCl [12]. It is therefore important to ethanol-precipitate the DNA after the first ligation step and to get rid of residual salt by washing the DNA pellet with 70% ethanol.

52.6.4

The Amplification of the Unselected Library is Inefficient

It is recommended to use only the gel-purified ligated products as templates. We observed that the carry-over of non-ligated SP6 linker inhibits the amplification of the unselected library, since the antisense oligonucleotide of the SP6 linker is able to anneal to any library molecule containing an SP6 promoter sequence. However, because of its “NN”-end most of the annealed molecules will have an unpaired 3' end which cannot be extended by *Taq* DNA polymerase.

52.6.5

The Library Appears to be Non-Random in the Unselected Pool

It is important not to cleave DNA with any restriction enzymes before sonication. Otherwise, sequences flanking the restriction sites will become overrepresented. If linker sequences appear in the variable part of the RNA, the ratio of linker to DNA fragments should be lowered.

52.6.6

The Selected RNAs do not Bind to Native Protein

There is always a risk in SELEX to select RNAs that bind to the matrix or purification tag. Although this seems to be a minor problem when using RNA libraries derived from genomic sequences, it is preferable to alternate between different tags or matrices. For instance, when using a protein to select for RNA binders, we recommend using two different tags (e.g., His- and Glutathione S-transferase (GST)-tagged proteins) in alternating selection step.

Materials

Prepare all solutions using double-distilled and diethyl pyrocarbonate (DEPC)-treated RNase-free water (ddH₂O) (exception: deionized H₂O to dilute 5 x Tris/Borate/EDTA (TBE) to 1 x TBE electrophoresis buffer) and analytical grade reagents. All stock solutions except for electrophoresis buffers were subjected to sterile filtration (Filtropur S 0.2 µm; Sarstedt) and stored at −20 °C if not stated otherwise.

Oligonucleotides

- T7 linker containing a *Pst* I site, T7 promoter sequence (underlined), and an *Mme* I site (small letters)
 - Sense: 5'-AGCCTGCAGTTAATACGACTCACTATAGGGATCGCTTAGtccgac-3'
 - Antisense: 3'-GTCGGACGTCATTATGCTGAGTGATATCCCTAGCGAATCag gctg-5'
- SP6 linker, containing the SP6 promoter sequence (underlined) and an *Eco* RI site (small letters)

- Sense: 5'-CTTCTATAGTGTACCTAAAT gaattcG-3'
- Antisense: 3'-NNGAAGATATCACAGTGGATTActtaag-5'
- SP6 primer
 - 5'-CgaattcATTTAGGTGACACTATAG-3'

Enzymes

- DNA polymerase I (Klenow fragment), 5 U μl^{-1} , NEB
- DNase I, RNase-free, 2 U μl^{-1} , Life Technologies Ambion (Ambion® TURBO™ DNase, #AM2238)
- T4 polynucleotide kinase (T4 PNK), 10 U μl^{-1} , NEB
- T4 DNA ligase, 6 U μl^{-1} , NEB
- *Mme* I restriction endonuclease, 2 U μl^{-1} , NEB
- *Taq* DNA polymerase, 5 U μl^{-1} , Life Technologies Invitrogen
- T7 RNA polymerase, 50 U μl^{-1} , NEB
- Avian myeloblastosis virus (AMV) Reverse transcriptase, 10 U μl^{-1} , Promega

Special Reagents

- 3 mM S-adenosylmethionine (SAM) hydrochloride, NEB
- RNasin, RNase inhibitor, 30 U μl^{-1} , Promega

Buffers

Buffer pH values are those at 25 °C

- 10 $\times$ DNA Pol. I (Klenow) buffer (NEB): 100 mM Tris-HCl pH 7.5, 50 mM MgCl_2 , 75 mM dithiothreitol (DTT)
- 2 $\times$ DNase I buffer: 250 mM Tris-HCl pH 7.6, 2.5 mM DTT, 125 mM NaCl, 5 mM CaCl_2 , 50 mM MnCl_2 . Add MnCl_2 just before use.
- 10 $\times$ PNK buffer (NEB): 700 mM Tris-HCl pH 7.6, 100 mM MgCl_2 , 50 mM DTT
- 10 $\times$ T4 DNA ligase buffer (NEB): 500 mM Tris-HCl, 100 mM MgCl_2 , 100 mM DTT, 10 mM ATP, 250 $\mu\text{g ml}^{-1}$ bovine serum albumin (BSA), pH 7.5
- 10 $\times$ *Mme* I reaction buffer (NEB buffer 4): 500 mM KOAc, 200 mM Tris-acetate, 100 mM $\text{Mg}(\text{OAc})_2$, 10 mM DTT, pH 7.9
- 1 $\times$ TE buffer: 10 mM Tris-HCl, 1 mM Na_2EDTA , pH 7.5
- 10 $\times$ *Taq* DNA polymerase reaction buffer (Life Technologies Invitrogen): 200 mM Tris-HCl pH 8.4, 500 mM KCl
- 10 $\times$ T7 transcription buffer (NEB): 400 mM Tris-HCl pH 7.9, 60 mM MgCl_2 , 20 mM spermidine, 100 mM DTT
- Denaturing RNA loading buffer: 50 mM Tris-HCl pH 7.9, 10 mM EDTA, 0.025% xylene cyanol (w/v), 0.025% bromophenol blue (w/v), 80% formamide

- *Gel extraction buffer* (store in the dark at room temperature): 0.75 mM NH_4OAc , 0.1% sodium dodecyl sulfate (SDS), 10 mM $\text{Mg}(\text{OAc})_2$, 0.1 mM EDTA, pH 7.0
- *2 x RNA binding buffer*: 50 mM sodium cacodylate pH 7.5, 250 mM KCl, 5 mM MgCl_2
- *2 x Native loading buffer*: 10% glycerol, 0.025% (w/v) xylene cyanol, 0.025% bromophenol blue (w/v), 10 mM MgCl_2
- % polyacrylamide (PAA) (acrylamide : bisacrylamide = 19 : 1), 8 M urea in 1 x TBE
- 6% PAA in 1 x TBE
- 6% PAA in 1 x TBM
- *10 x TBM buffer*: 1 M Tris-HCl, 0.9 M boric acid, 50 mM MgCl_2
- *2 x hnRNP A1 binding buffer*: HEPES-KOH pH 7.6, 200 mM KCl, 4 mM MgCl_2 , 1 mM EDTA, 2 mM DTT, 20% glycerol
- *hnRNP A1 dilution buffer*: 20 mM HEPES-KOH pH 7.9, 150 mM KCl, 1 mM DTT, 0.2 mM EDTA, 20% glycerol
- *5 x AMV reverse transcriptase reaction buffer*: 250 mM Tris-HCl pH 8.3, 250 mM KCl, 50 mM MgCl_2 , 2.5 mM spermidine, 50 mM DTT

Protocols

Protocol 1: Generation of Random DNA Fragments from Genomic or Plasmid DNA

Sonication

- 1) Adjust the volume of the DNA solution to at least 500 μl . For example, dilute 50 μg DNA in 500 μl H_2O .
- 2) Sonicate the genomic DNA for 30 min. Chill the DNA on ice during the whole process to prevent heating of the sample. Using a Branson Ultrasonic Sonifier II, Model W-250, equipped with a microtip, we set the output to 10 W combined with a duty cycle of 10%, so that the total time of ultrasound treatment is about 3 min. As a rule of thumb, it is better to sonicate for a long time with a low output than for a short time with high intensity. For sonication of plasmid DNA, increase the power to 15 W and sonicate for 10 min with a 50% duty cycle. An example of sonicated human genomic DNA is shown in Figure 52.2, lane 1.
- 3) Take 5 μl aliquots in intervals of 5 min to check the efficiency of sonication.
- 4) Load the aliquots on a 6% PAA gel (1 x TBE). Run the gel until the xylene cyanol has migrated 2/3 of the gel length. Stain the gel with ethidium bromide.
- 5) If the average fragment length is about 600 bp, ethanol-precipitate the DNA and proceed to the next step. The range of fragment sizes should be between 200 and 1000 bp.
- 6) The sonicated DNA fragments are treated with Klenow polymerase to ensure that they have blunt ends.

<i>Klenow reaction, 40 µl</i>	
5 µl	Sonicated DNA, 50 µg
4 µl	10 mM dNTP, 2.5 mM each
4 µl	10 x DNA Pol. I (Klenow) buffer
2 µl	DNA polymerase I (Klenow fragment), 10 U
25 µl	H ₂ O

- 7) Incubate at 30 °C for 20 min.
- 8) Fill up with water to 100 µl and extract with 100 µl phenol/chloroform (50/50%). Centrifuge briefly and transfer the aqueous supernatant into a fresh tube leaving interphase and organic phase behind.
- 9) Repeat the extraction with 100 µl chloroform.
- 10) Precipitate the DNA with 2.5 volumes of 96% ethanol and 1/10 volume of 3 M sodium acetate, pH 6.5.
- 11) Centrifuge for 20 min with 10 000 g at 4 °C. Wash the DNA pellet with 70% ethanol and dry in a vacuum concentrator. Dissolve the DNA in 20 µl H₂O.

DNase I Treatment

It is strongly recommended to optimize the incubation time for DNase I treatment. This can be achieved by taking out small aliquots at different time points and terminating the reaction by addition of 50 mM EDTA. DNase I-digested fragments in a range between 200 and 1000 bp are suitable for library construction.

- 1) *DNase I reaction, 100 µl*

10 µg	Plasmid DNA
50 µl	2 x DNase I buffer
39 µl	H ₂ O
1 µl	DNase I, 2 U

- 2) Incubate at 37 °C until the desired range of fragments is produced (1–30 min time course). The reaction is terminated by addition of 10 µl 50 mM EDTA, followed by phenol extraction and ethanol precipitation (as in step 8–11 of sonication), and finally gel purification as described in the section “*Mme* I cleavage”. Dissolve the DNA in 20 µl H₂O.

Protocol 2: Construction of the RNA Library

Phosphorylating the 5’ Ends of DNA Fragments

- 1) All DNA fragments must have a 5’ phosphate end in order to be ligated to the T7 linker.

T4 PNK reaction, 30 µl

20 µl	DNA fragments from Protocol 1
3 µl	10 x PNK buffer
3 µl	10 mM ATP
2 µl	H ₂ O
2 µl	T4 PNK, 20 U

- 2) Incubate at 37 °C for 1 h.
- 3) Fill up with H₂O to 100 µl and extract with 100 µl phenol/chloroform (50/50%).
- 4) Ethanol-precipitate the DNA as described above and dissolve the DNA in 20 µl H₂O.

Preparing the Double-Stranded T7 Linker

- 1) Before the two primers are annealed, the antisense primer must be phosphorylated at its 5' end.

T4 PNK reaction, 10 µl

1 µl	10 x PNK buffer
1 µl	10 mM ATP
1 µl	Antisense primer (100 pmol/µl)
6 µl	H ₂ O
1 µl	T4 PNK, 10 U

- 2) Incubate for 1 h at 37 °C.
- 3) Fill up with water to 100 µl and extract with 100 µl phenol/chloroform (50/50%). Precipitate as described above and dissolve the DNA in 7 µl H₂O.
- 4) To anneal the two T7 linker oligonucleotides, add to the phosphorylated antisense oligonucleotide:

2 µl	Sense oligonucleotide (100 pmol/µl)
1 µl	1 M KCl

- 5) Heat the mixture to 90 °C in a heating block for 2 min. Let it cool down to room temperature for 10 min, but do not place on ice.

Ligation of the T7 Linker to the DNA Fragments

1) *T4* DNA ligation, 50 µl

20 µl	5'-Phosphorylated blunt-ended DNA fragments
10 µl	Double-stranded T7 linker
5 µl	10 x <i>T4</i> DNA ligase buffer
12 µl	H ₂ O
3 µl	<i>T4</i> DNA ligase, 18 U

- 2) Incubate the reaction overnight at 14–16 °C.
- 3) Fill up with water to 100 µl and extract with 100 µl phenol/chloroform (50/50%).
- 4) Precipitate as described above and dissolve the DNA in 20 µl H₂O.

***Mme* I Cleavage**

- 1) Before starting the *Mme* I cleavage reaction, save 2 µl of the undigested material to load it next to the preparative *Mme* I reaction products onto the gel. This will help to identify the cleavage products and to verify the ligation and digestion efficiencies. An example of the cleavage products generated by *Mme* I is shown in Figure 52.2 (compare lanes 2 and 3).

<i>Mme</i> I restriction reaction, 30 µl	
0.5 µl	3 mM SAM hydrochloride
3.0 µl	10 x <i>Mme</i> I reaction buffer (NEB buffer 4)
18.0 µl	Ligated products
7.0 µl	H ₂ O
1.5 µl	<i>Mme</i> I, 3 U

- 2) Incubate for 1 h at 37 °C.
- 3) Separate the reaction products on a 6% PAA gel (1 x TBE).
- 4) Cut out the band of interest and elute the fragments from the gel piece overnight in 300 µl TE buffer. For a more efficient extraction of the fragments, we strongly recommend electroelution.
- 5) Extract with 300 µl phenol/chloroform (50/50%) and ethanol-precipitate as described above.
- 6) Dissolve the DNA in 10–20 µl H₂O.

Preparation and Ligation of the SP6 Linker

- 1) 5'-Phosphorylate the sense primer of the SP6 linker and anneal it to the SP6 antisense primer as described for the T7 linker.
- 2) Ligate the eluted *Mme* I fragments to the SP6 linker using the same ligation conditions as described above. Depending on the amount of *Mme* I fragments

obtained, the reaction volume may be reduced to 25 μ l or less to increase the ligation efficiency.

- 3) Gel-purify the ligated products on a 6% PAA gel (1 x TBE) and precipitate the DNA as described above. An example of an SP6 linker ligation is shown in Figure 52.2, lane 4.

PCR Amplification of the Library

- 1) *PCR reaction*, 50 μ l

2.0 μ l	T7 Linker, sense primer (25 pmol/ μ l)
2.0 μ l	SP6 primer (25 pmol/ μ l)
4.0 μ l	10 mM dNTP mix (2.5 mM each)
2.0 μ l	50 mM MgCl ₂
5.0 μ l	10 x <i>Taq</i> DNA polymerase reaction buffer
5.0 μ l	DMSO
5.0 μ l	Eluted library DNA
24.5 μ l	H ₂ O
0.5 μ l	<i>Taq</i> DNA polymerase, 2.5 U

- 2) Amplify the DNA with: 92 °C for 3 min; 25–30 cycles of 92 °C for 30 s, 49 °C for 30 s, 72 °C for 30 s; 72 °C for 7 min.
- 3) Ethanol-precipitate the PCR product. Dissolve the DNA in 10 μ l H₂O.

Protocol 3: T7 Transcription of the Library

- 1) Mix all components of the reaction at room temperature to avoid precipitation.

T7 transcription reaction, 25 μ l

9.0 μ l	H ₂ O
2.5 μ l	10 x T7 transcription buffer
1.0 μ l	300 mM DTT
0.5 μ l	RNasin RNase inhibitor, 15 U
2.5 μ l	NTP mix (2 mM UTP, 5 mM each ATP, CTP, and GTP)
5.0 μ l	DNA (PCR product from Protocol 2)
2.5 μ l	[α - ³² P]UTP, 3000 Ci/mmol, 20 mCi/ml
2.0 μ l	T7 RNA polymerase, 100 U

- 2) Incubate for 2 h at 37 °C.
- 3) Add 2 μ l RNase-free DNase I (4 U) and continue the incubation for 30 min.
- 4) Save 1 μ l to determine the total amount (*cpm*) of Cerenkov counts per minute.
- 5) Add 10 μ l denaturing RNA loading buffer and heat the mixture for 3 min to 90 °C. Chill on ice and load the sample on a denaturing 4% PAA gel with 8 M urea in 1 x TBE.

- 6) When the bromophenol blue has migrated approximately two-thirds of the gel length, disassemble the gel and wrap it into plastic foil to prevent dehydration. Place tracking tape in the corners (to be able to align the film to the gel after development) and expose an X-ray film.
- 7) Cut out the desired band and immerse the gel piece in 400 µl (approximately 5 volumes) gel extraction buffer. Elute the RNA for at least 4 h under vigorous shaking at room temperature.
- 8) Add 400 µl isopropanol to the supernatant and precipitate the RNA overnight at -20°C . Spin for 30 min with 10 000 g at 4°C . Wash the pellet with 70% ethanol and dry it in a vacuum concentrator.
- 9) Redissolve the RNA in 20–30 µl H_2O , measure the cpm of 1 µl solution, and calculate the total cpm (*cpm2*). With the conditions above, the amount of RNA can be calculated as follows:

$$\begin{aligned}\text{RNA}(\text{pmol}) &= \frac{\text{cpm2} \times \text{pmol UTP in 25 } \mu\text{l transcription mixture}}{\text{cpm1} \times \text{number of U residues per transcript}} \\ &= \frac{\text{cpm2} \times 5000}{\text{cpm1} \times 19}\end{aligned}$$

Transcripts (66 nt, if PCR products are used directly; see above) include 4 U residues encoded by the T7 linker (including the *Mme* I recognition site), 5 U residues on average in the variant region, and 10 U residues encoded in the SP6 linker sequence, which gives a total number of 19 U residues.

- 10) Dilute the RNA to the desired concentration.

Protocol 4: Identification of Accessible Sites in a Structured RNA for Optimal Antisense Annealing

SELEX Assay

- 1) Prepare a solution containing 1 pmol of radioactively labeled target RNA (internally- or end-labeled) in 10 µl 1 x RNA binding buffer.
- 2) Denature the RNA for 2 min at 85°C and cool the reaction slowly to room temperature over 20 min. Chill the sample on ice for 5 min. Save 2 µl for native gel analysis.
- 3) *RNA binding reaction*, 20 µl

8 µl	^{32}P -Labeled target RNA
5 pmol	RNA library

- 4) Adjust the volume to 20 µl using 1 x RNA binding buffer.
- 5) Incubate the reaction at 37°C for 30 min (the incubation time may be lowered to 5 min in the final round). Add 20 µl 2 x native loading buffer and separate the RNA on a native 6% PAA gel containing 5 mM MgCl_2 (1 x TBM buffer).

- 6) Disassemble the gel as described in Protocol 3 and expose an X-ray film at 4 °C. Cut out the band corresponding to target RNA–antisense RNA complexes. Depending on the size of the target RNA, complexes may migrate with slightly slower mobility than the naked RNA.
- 7) Elute and precipitate the RNA with isopropanol as described in Protocol 3 and dissolve it in 10 µl H₂O. Proceed to the RT-PCR amplification step outlined below.

Protocol 5: SELEX Selection of Protein Binding Sites

The selection of RNA sequences that are specifically recognized by a ligand may be performed with any type of RNA binding ligand. Binding sites for the hnRNP A1 protein were selected in the example below.

Binding Reaction

- 1) *Binding reaction*, 50 µl

0.5 µl	RNasin, RNase inhibitor, 15 U
5 pmol	RNA library
5 pmol	hnRNP A1

- 2) Adjust the volume to 25 µl with H₂O and 2 x hnRNP A1 binding buffer to a final concentration of 1 x hnRNP A1 binding buffer. Dilute the protein in hnRNP A1 dilution buffer, if necessary.
- 3) Mix and incubate for 15 min at room temperature.
- 4) Load the mixture on a native 6% PAA gel (in 1 x TBE) using 0.5 x TBE as running buffer. Do not mix the samples with loading buffer, but put an aliquot of the latter into an empty well next to the samples. Run the gel at 4 °C with 300 V for 90 min until the bromophenol blue has migrated approximately two-thirds of the gel length.
- 5) Disassemble the gel as described in Protocol 3 and expose an X-ray film at 4 °C. Cut out the desired band and elute the RNA under vigorous shaking in 400 µl (approximately 5 volumes) gel extraction buffer for at least 4 h at room temperature.
- 6) Extract the supernatant with 400 µl phenol/chloroform (50/50%).
- 7) Precipitate the RNA with isopropanol as described in Protocol 3. Redissolve the RNA in 10 µl H₂O and proceed to the RT-PCR amplification step outlined below.

RT-PCR

- 1) Add 2 pmol of SP6 primer to 2–5 µl eluted RNA. Fill up with water to 8 µl and heat the mixture in a heating block to 70 °C for 5 min. Chill on ice immediately for 5 min.

2) Add to the annealed RNA/primer mixture:

5.0 µl	5 x AMV reverse transcriptase reaction buffer
2.5 µl	dNTP mix (2.5 mM each)
1.0 µl	RNasin, RNase inhibitor (30 U/µl)
2.5 µl	Sodium pyrophosphate, 40 mM (pre-warmed to 42 °C)
1.0 µl	AMV reverse transcriptase (10 U/µl)
5.0 µl	H ₂ O

- 3) Incubate at 42 °C for 60 min.
- 4) Use 2–5 µl as template for the PCR applying the same conditions as for amplification of the library.
- 5) The PCR fragments may either be used for a new round of transcription or for cloning and sequencing.

References

1. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
2. Singer, B.S., Shtatland, T., Brown, D., and Gold, L. (1997) Libraries for genomic SELEX. *Nucleic Acids Res.*, **25**, 781–786.
3. Ho, S.P., Britton, D.H., Bao, Y., and Scully, M.S. (2000) RNA mapping: selection of potent oligonucleotide sequences for antisense experiments. *Methods Enzymol.*, **314**, 168–183.
4. Sohail, M. and Southern, E.M. (2000) Selecting optimal antisense reagents. *Adv. Drug Deliv. Rev.*, **44**, 23–34.
5. Lima, W.F., Brown-Driver, V., Fox, M., Hanecak, R., and Bruice, T.W. (1997) Combinatorial screening and rational optimization for hybridization to folded hepatitis C virus RNA of oligonucleotides with biological antisense activity. *J. Biol. Chem.*, **272**, 626–638.
6. Birikh, K.R., Berlin, Y.A., Soreq, H., and Eckstein, F. (1997) Probing accessible sites for ribozymes on human acetylcholinesterase RNA. *RNA*, **3**, 429–437.
7. Ho, S.P., Bao, Y., Leshner, T., Malhotra, R., and Ma, L.Y. (1998) Mapping of RNA accessible sites for antisense experiments with oligonucleotide libraries. *Nat. Biotechnol.*, **16**, 59–63.
8. Allawi, H.T., Dong, F., Ip, H.S., Neri, B.P., and Lyamichev, V.I. (2001) Mapping of RNA accessible sites by extension of random oligonucleotide libraries with reverse transcriptase. *RNA*, **7**, 314–327.
9. Milner, N., Mir, K.U., and Southern, E.M. (1997) Selecting effective antisense reagents on combinatorial oligonucleotide arrays. *Nat. Biotechnol.*, **15**, 537–541.
10. Lindtner, S., Felber, B.K., and Kjems, J. (2002) An element in the 3' untranslated region of human LINE-1 retrotransposon mRNA binds NXF1(TAP) and can function as a nuclear export element. *RNA*, **8**, 345–356.
11. Herrera, J.E. and Chaires, J.B. (1994) Characterization of preferred deoxyribonuclease I cleavage sites. *J. Mol. Biol.*, **236**, 405–411.
12. Tucholski, J., Skowron, P.M., and Podhajska, A.J. (1995) *Mme* I, a class-IIIS restriction endonuclease: purification and characterization. *Gene*, **157**, 87–92.
13. Tucholski, J., Zmijewski, J.W., and Podhajska, A.J. (1998) Two intertwined methylation activities of the *Mme* I

- restriction-modification class-II system from *Methylophilus methylotrophus*. *Gene*, **223**, 293–302.
14. Jakobsen, M.R., Damgaard, C.K., Andersen, E.S., Podhajska, A., and Kjems, J. (2004) A genomic selection strategy to identify accessible and dimerization blocking targets in the 5'-UTR of HIV-1 RNA. *Nucleic Acids Res.*, **32**, e67.
 15. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.

53

Genomic SELEX

Jennifer L. Boots, Katarzyna Matylla-Kulinska, Marek Zywicki, Bob Zimmermann, and Renée Schroeder

53.1

Introduction

SELEX (Systematic Evolution of Ligands by EXponential enrichment) was developed to screen large libraries for sequences that bind with high affinity to ligands of choice [1, 2]. This is achieved by binding the library to a target, separating the bound from unbound sequences, and finally amplifying the selected sequences for further rounds of selection [2, 3]. The libraries were initially obtained by randomly synthesizing nucleic acids, leading to pools of highly complex sequence space. In the recent years, with the onset of tools to sequence entire pools in parallel and with so many sequenced genomes, the SELEX procedure has been adapted to screen genomes for functional DNA or RNA motifs that bind to interesting targets. When the libraries are derived from genomic DNA instead of random sequences, the procedure is referred to as *genomic SELEX*.

Genomic SELEX is an important tool for the discovery of genome-encoded aptamers and regulatory sequences that interact with proteins or other ligands. For example, genomic SELEX has been particularly useful in finding DNA targets for transcription factors [4–6], RNA targets for splicing factors [7], and novel RNA–RNA loop–loop interactions [8]. In genomic SELEX, the initial library is not random but is composed of varying lengths of genomic DNA that represent the entire genome [9]. The DNA or RNA molecules selected from genomic SELEX experiments are referred to as *genomic aptamers* [10]. The main advantages of genomic SELEX over the classical one are that it uses a significantly reduced allowable sequence space and increases the likelihood that a biologically relevant target is selected. Moreover, since the initial library originates from genomic DNA, it screens for aptamers regardless of RNA expression levels. Thus it is possible to select RNAs that are expressed at very low levels, or at a specific cell cycle point or developmental stage. However, a limitation of genomic SELEX is that the selected genomic aptamers are derived from RNAs that may or may not be expressed. Classical methods, such as massive sequencing of RNAs bound to a target, guarantee to result in expressed RNAs.

In this chapter, we give a comprehensive introduction into the genomic SELEX method. We discuss in detail how to construct a genomic RNA library starting with any available genomic DNA, then how to select for RNAs that bind to a target of interest (the bait), and finally how to evaluate the sequences obtained from the selection. Depending on the bait, a very large number of aptamers might be expected. In that case, high-throughput sequencing is essential. If the source genome is large and contains highly repetitive elements, finding the original genomic location of the selected sequences and statistical evaluation of the data requires special care. The major goal of the computational analysis is to create a basis for selecting candidates for further biochemical analysis. Since genomic aptamers discovered with genomic SELEX only represent the binding domain within an encoded RNA, and not necessarily the full transcript *in vivo*, further characterization of the transcript may be necessary to understand the biological relevance of the RNA–ligand interaction. We discuss methods to characterize these transcripts and ways to evaluate the potential biological function of the interaction.

53.2

Description of the Methods

53.2.1

Library Construction

The initial library for the genomic SELEX procedure is created from the genomic DNA pool of an organism of interest, which is randomly primed and transcribed into RNA [9, 11]. As a result, the library entirely covers the genome of interest, so every potential genomic aptamer is represented in the starting pool. The advantage of constructing a library from genomic DNA is to screen for genomic aptamers irrespective of their expression profile in general and as it relates to phases of the cell cycle or developmental stages.

After isolating or purchasing high-quality genomic DNA, the first strand is synthesized with the Klenow fragment of DNA polymerase. The hybREV primer (both hyb primers consist of a specific sequence followed by randomized nucleotides at the 3' end) is annealed to the genomic DNA at 25 °C and then extended. Before second strand synthesis with the hybFOR, the excess hybREV should be thoroughly removed (for example, with a microconcentrator) in order to reduce the formation of fragments flanked with the same sequence on both sides. After synthesis of both strands, size selection follows. The lengths of fragments should correspond to the size of potential aptamers being targeted. Klenow reaction products are usually resolved on denaturing polyacrylamide gel, and fragments of desired size are excised. After DNA elution from a gel piece, a T7 promoter sequence is introduced by a subsequent polymerase chain reaction (PCR) amplification with primers fixFOR/fixREV (see Figure 53.1a).

We have previously used the sequence 5'-CCAAGTAATACGACTCACTATAGG GGAATTCGAGCGGGCAGC-3' (T7 promoter sequence underlined) for the

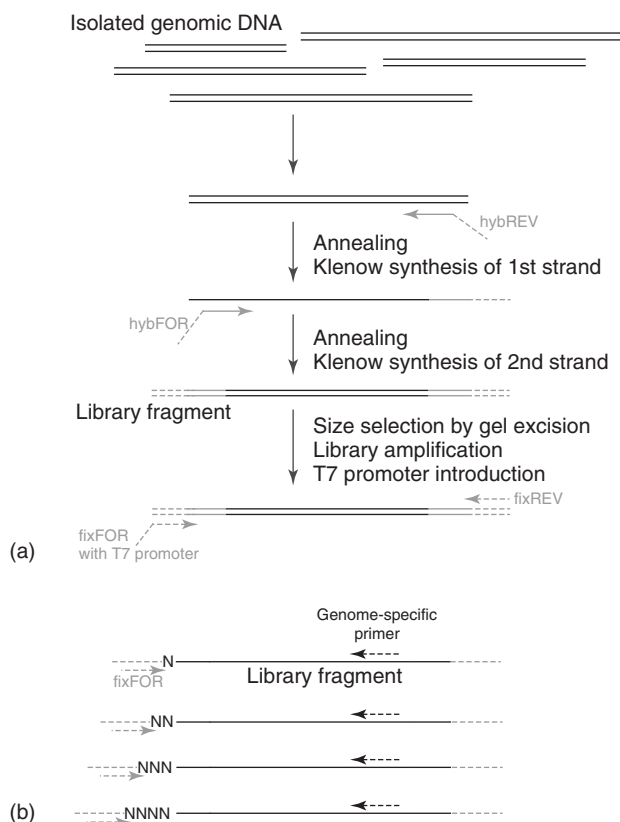


Figure 53.1 (a) Scheme for a genomic SELEX library construction. High-quality genomic DNA is used to construct the genomic SELEX library. First, the *hybREV* primer is annealed to genomic DNA at 25 °C and extended by the Klenow fragment of DNA polymerase. Then, before the second strand synthesis, the excess *hybREV* should be carefully removed. After second strand synthesis with *hybFOR*, fragments of desired size are selected in the denaturing gel electrophoresis and eluted from the polyacrylamide gel. Next, a T7 promoter sequence is introduced by a subsequent PCR

amplification with primers *fixFOR*/*fixREV*. Gray parts of *hybFOR*/*hybREV* represent nine random nucleotides. The dashed gray sections represent specific *fixFOR*/*fixREV* sequence. The T7 promoter sequence in the *fixFOR* is depicted by the dotted gray line. (b) Library quality control by analysis of the distribution of end points. If the genomic library represents the entire genome, PCR amplification with the combination of one genome-specific primer and one library-end-point-specific (*fixFOR*/*fixREV*) primer should result in amplicons that differ by a single nucleotide.

fixFOR-T7 primer [11]. The downstream sequence has the advantage that it is not present in current assemblies of human, yeast, and *Escherichia. coli* genomes. It also functions well under the recommended PCR annealing temperature (55 °C) of the reverse primer, *fixREV*, 5'-CGGGATCCTCGGGGCTGGGATG-3', which is also non-genomic. Restriction sites are also included at the 3' ends of *fixFOR*

(*Bbv* I) and fixREV (*Fok* I). This can be useful in swapping fixed primers between selection rounds to guarantee that the selected motifs do not include the fixed primer sequences. Since the T7 promoter is not needed in the Klenow phase, hybFOR consists of only the sequence 5'-AGGGAATTCTGGAGCGGGCAGC-3' followed by nine random nucleotides. The sequence of hybREV is the same as fixREV with nine random nucleotides at the 3' end.

Before the first SELEX round, it is advisable to test the quality of the genomic library. To ensure that the genome coverage is reasonable, PCRs with several arbitrarily chosen primer pairs are performed to give amplicons corresponding to the length of fragments in the library. For easier evaluation, it is recommended to test amplicons from single-copy regions. As an additional control, a PCR with a gene-specific primer in combination with fixREV or fixFOR can be performed to ensure that the obtained products have sizes that vary in the desired range (Figure 53.1b).

53.2.2

Choice of Bait

The choice of bait is an essential step. In principle, aptamers against any ligand can be obtained as long as the ligand is soluble under conditions in which the RNA is stable. SELEX has been performed against a very long list of small molecules, ranging from primary metabolites, coenzymes, antibiotics, to synthetic drugs [3]. When searching for genomic aptamers against small molecules, it is advisable to design a procedure that does not require a linker for immobilization. This is because it became apparent from the X-ray structures of ligand–aptamer complexes of riboswitches [12] that the ligand is entirely embedded within the aptamer, leaving no space for a linker. The most common baits for genomic SELEX are proteins that are chosen because they are involved in the regulation of RNA expression, folding, or activity. Many proteins contain predicted RNA-recognition motifs where the target RNA is not known. RNA binding proteins involved in transcription, processing, stability, and degradation are good candidates for genomic SELEX. However, proteins that have been shown to bind RNA non-specifically or only transiently would not be good candidates since they will most likely not enrich any specific aptamers [13]. Finally, it is important for genomic SELEX that the protein of interest be highly pure and stable *in vitro*.

53.2.3

SELEX Procedure

A summary of the SELEX procedure is shown in Figure 53.2. The specific details of the procedure are discussed in this section. This procedure was adapted from Ref. [11].

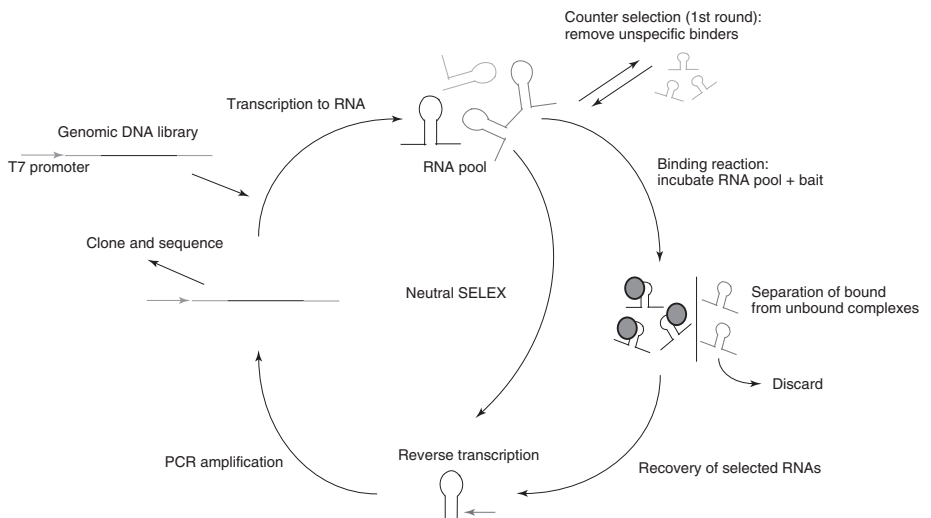


Figure 53.2 The SELEX procedure overview. The genomic DNA library is transcribed into RNA using T7 polymerase. In the first round of selection, a counter selection is used to remove unspecific binding RNAs that bind to the apparatus for separation. The cleared RNA library is incubated with the target of interest (the bait). The RNAs that bind to the target are then separated from the unbound using a variety of methods. Next, bound RNAs are recovered and amplified using reverse transcription and PCR. Finally, the RNA pool is either subjected to further rounds of selection or cloned and sequenced.

53.2.3.1 Transcription of Genomic Library into RNA Library

Genomic SELEX can be performed directly with the DNA library if DNA aptamers are the targets of interest. However, in this chapter, we focus on how to perform genomic SELEX with an RNA library. The first step in the SELEX procedure is to transcribe the DNA library into RNA. To do this, incubate 10 µg of the genomic library DNA with 20 µl of 25 mM rNTP mix, 1 µl of RNase inhibitor (Promega), T7 polymerase, 10 µl 10 x buffer, and trace amounts of [α - ^{32} P]GTP in a 100 µl reaction volume for 4 h at 37 °C. The amount of template, MgCl₂, rNTPs, T7 polymerase, and dithiothreitol can be varied to optimize the transcription reaction. The radioactivity is used to follow the RNA pool throughout the selection procedure and to estimate the enrichment after each round of selection.

After transcription, DNase I is added to degrade the template DNA library, and then the library is incubated at 37 °C for 30 min. It is important to eliminate the template DNA library when performing RNA genomic SELEX as certain DNA sequences may become enriched if the bait can also bind double-stranded DNA. To stop the reaction, heat-inactivate the DNase I at 65 °C for 10 min in the presence of EDTA (or according to manufacturer's guidelines). It is recommended to check the quality and size of the RNA fragments on an agarose gel or low-percentage (4%) polyacrylamide gel. Finally, dilute the RNA library to 500 µl in a binding buffer suitable for the RNA–bait interaction.

53.2.3.2 Counter Selection

To avoid enriching non-specific RNA aptamers in the library, which can bind to the apparatus used for separation of bound from unbound complexes, a counter selection must be performed. For example, the diluted pool can be precleared by incubation with the membrane or the column matrix. The precleared library will flow through the membrane and can be recovered and purified by ethanol precipitation. For column separation, the RNA library is incubated with the beads, and then the RNAs that do not bind are recovered by either centrifugation or gravity flow in the column apparatus and ethanol precipitation. In addition, if the bait protein has a tag for column purification, the library can be incubated with beads containing the tag to eliminate RNAs that bind specifically to the tag. Counter selection may be performed with an inactive form of the bait protein to assure that the selected RNA sequences are specific for the active bait.

53.2.3.3 Positive Selection

The positive selection of RNA–bait complexes involves first an incubation step, to allow for complex formation, and then the separation of bound from unbound complexes. The concentration of the RNA library should be measured with UV spectroscopy and a scintillation counter to determine the counts per mole of RNA. This will be important for the calculation of the concentration of recovered RNA from each round of selection. Next, decide on the ratio of the RNA to ligand concentration. If the bait is a protein, a good starting point is a 10 : 1 molar ratio of RNA library : protein. It is critical that the RNA be in molar excess of the protein to establish an environment of competition for binding different species of RNA

molecules in the library. For a protein with a known activity, it is advisable to use buffer conditions and time of incubation in which the protein is known to be active. However, if the protein function is unknown or if the bait is a small molecule, it is recommended to begin with near physiological buffer conditions. A good starting point for binding is room temperature (23 °C), where RNA secondary structure is stable and there is minimal denaturation of the protein [11].

Initial rounds of selection are normally carried out under moderate conditions, and then, during later rounds, the conditions are more stringent to increase the specificity of the selected complexes. For example, increasing the salt concentration, changing the RNA–bait ratio, or adding non-specific competitors in later rounds of selection may increase the stringency of binding. Before binding, denature the RNA library for 1 min at 95 °C and then let it slowly cool to room temperature for ~10 min to ensure refolding of the RNA. Next, incubate the RNA with the bait in the desired binding buffer for the amount of time required for the interaction. UV crosslinking may also be used to stabilize the RNA–protein complex [14].

After incubation of the RNA library with the bait, the bound and unbound complexes must be separated. This can be done using a variety of techniques. For RNA–protein complexes, the most convenient methods are membrane filtration [1, 13–16] and affinity chromatography [2, 17, 18]. For example, if the bait protein was purified with a fused tag, the RNA–protein complexes can be incubated with the appropriate affinity column. In addition, if there is an antibody that recognizes the target protein, the RNA–protein complexes can be immunoprecipitated. However, it must be kept in mind that the RNA may block binding to the antibody, so a control must be done to determine if this is a feasible method for separation.

If the protein is small, size exclusion methods, such as gel electrophoresis or size selection chromatography, are useful for separation [19]. Other methods include fluorescence-activated cell sorting [20] and surface plasmon resonance [21]. It is helpful to perform multiple selections in parallel using different binding conditions, by varying ligand concentrations, and using mutants of the bait protein to increase the specificity of the selection [10]. In order to confirm the selection results, it is also advisable to perform parallel selections using different immobilization methods, as well as to perform technical replicates.

Although genomic SELEX is usually performed *in vitro*, it is also possible to confirm a direct interaction between the RNA and protein by performing one or two cycles *in vivo*. In this case, the enriched library and the bait need to be fused to reporter molecules for a three-hybrid system readout [13]. Alternatively, the protein–RNA complex can be crosslinked and immunoprecipitated with an antibody.

53.2.3.4 Recovery and Amplification of Selected Sequences

After the positive selection step, the RNA sequences need to be recovered and amplified for further rounds of selection. After each round of selection, the membrane (or column) is counted on a scintillation counter to determine the approximate amount of selected RNA. This is useful for later estimation of the enrichment of the RNA pools during each round of selection.

The recovery of the selected RNAs depends on the method of separation. For membrane filtration, the RNA sequences can be recovered by incubating the membrane with 7 M urea, 20 mM sodium citrate, pH 5.0, 1 mM EDTA, and phenol (pH 5.2) and then shaking at 1400 rpm at room temperature for 10 min. This is followed by ethanol precipitation. For recovery of selected RNAs from a column, the protein–RNA complexes can be removed from the column by competition with a small molecule that binds the tag, the same as would be done for the purification of the tagged protein. Alternatively, the selected RNAs can be eluted from the column by digesting the protein with proteinase K. In both cases, it is advisable to then phenol, chloroform, isoamyl alcohol (PCI)-extract and ethanol-precipitate the RNA pool. For gel electrophoresis, the RNA is recovered through crushing and soaking in elution buffer (10 mM Tris-HCl, pH 8, 0.3 M NaOAc pH 5.4, 2 mM EDTA, 0.1% SDS) and ethanol precipitation.

After recovery of genomic aptamers, the pool is reverse transcribed into DNA. We recommend using an enzyme that is active at elevated temperatures, such as 50–60 °C, to allow reverse transcription (RT) of highly structured RNAs. Alternatively, an RNA helicase can be used in combination with the reverse transcriptase. The RT step is followed by PCR amplification with a polymerase with high fidelity (to minimize sequence artifacts during amplification) and the fixed primers (see Section 11.2.1). It is recommended to perform 7–10 cycles in the PCR to avoid dimerization of incomplete products, which can be extended when primer–template ratio decreases, resulting in chimeric products (see Table 53.1). After PCR, the DNA pool is phenol-extracted and ethanol-precipitated. The selected DNA pool is then transcribed, as previously discussed (see Section 11.2.3.1), and further rounds of selection can be performed. The number of cycles required to enrich the library depends on its initial complexity and on the desired affinity of the RNAs to the bait. About 7–12 rounds are typical for a genomic SELEX experiment. Usually, depending on the RNA–protein ratio used for complex formation, 30–60% of the input RNA will bind to the bait. Then the DNA fragments can be cloned and sequenced immediately (see Section 11.2.3.6), otherwise further rounds of selection are carried out.

53.2.3.5 Neutral SELEX

The amplification steps (PCR, *in vitro* transcription, RT) of genomic SELEX may introduce some bias in which sequences are ultimately selected. We previously developed a parallel control to SELEX in order to evaluate the effect of these steps on the initial library as SELEX proceeds [22]. To do this, the selection steps are omitted from the SELEX cycle (see Figure 53.2), and the results of each round can be sequenced. When we performed this, each round of the so-called neutral SELEX was sequenced. The average sequence had a less stable structure as the cycles progressed. We hypothesized that this is caused by the difficulty of the reverse transcriptase in denaturing highly structured RNAs. However, sometimes the selective pressures of binding can still lead to the enrichment of a highly structured RNA. For example, the SELEX-derived streptomycin aptamer was crystallized and shown to have a stable structure [23]. Other characteristic biases that were

Table 53.1 Troubleshooting Genomic SELEX. A list of common problems that arise with genomic SELEX and their solutions.

Step	Problem	Possible cause	Solution
Library construction	Uneven coverage of the genome	Genome is not amenable to random priming	As long as every region is represented, selection should be possible. Alternatively, whole genome amplification (WGA) kits can be used
	Fragments become shorter with subsequent rounds of genomic SELEX	Shorter products are amplified more efficiently	If the effect is strong, try increasing the elongation time in PCR. Also, relaxed stringency of selection can result in neutral selection effects; therefore, varying the stringency could help
	No PCR products during quality control	The library does not represent the entire genome	Generate more material during Klenow reaction, perhaps by varying the primer amounts
			Ensure that the isolated genomic DNA is highly pure
		Fragments are too long for amplification	WGA kits can be used (above)
		PCR conditions are not optimal	Pick shorter fragments, preferably within the size range of the library
Amplification	PCR products do not vary in expected size during quality control	Mispriming of designed primers	Vary MgSO ₄ , dNTPs, annealing temperature, or elongation times
	Sequences become longer	Chimeric products formed because of low primer to template ratio	Repeat size selection step
	Not enough material obtained	Suboptimal reaction conditions	Check the annealing conditions and sequence of the primers being used
	Structured RNAs not recovered	Reverse transcriptase does not denature RNA	Decrease the number of PCR cycles
Transcription			Increase initial primer concentration
			Vary annealing temperature
Reverse transcription			Vary MgCl ₂ , rNTPs, and template ratios
			Include an initial denaturation step
Positive selection	RNA was not completely removed from membrane/column	Protein was not fully denatured on membrane	Use a high-temperature reverse transcriptase
		Protein was not efficiently competed off column by small molecule	Use a helicase during the reverse transcription step
			Repeat filter-elution steps
	No RNAs were enriched/selected	Protein is nonspecific	Adjust levels of small molecule; repeat filter elution
		Binding conditions are not optimal	—
			Start with a decreased stringency by increasing the molarity of both RNA and protein in the solution

analyzed from the neutral SELEX sequences such as length, nucleotide content, and divergence from the initial library were only mildly affected, but this can vary depending on the features of the initial library. Thus we advise performing a parallel neutral SELEX control to any genomic SELEX in order to analyze the background signal.

53.2.3.6 Cloning and Sequencing

At any point during the initial rounds of selection, it is possible to clone and sequence the selected RNA pool to determine if any sequence is being enriched, and furthermore, if there are artifacts of DNA contamination or PCR chimeras (see Table 53.1). The pool can be cloned into any commercially available T/A cloning vector according to the manufacturer's instructions.

For baits that may have a large number of RNA targets, it is essential to use high-throughput sequencing. Current technologies are advancing at a staggering rate, and the specifics of these technologies are outside of the scope of this chapter. In any case, since the length of the sequences is constrained, no current technology would require sequence fragmentation. Without fragmentation, the fix primers on either end of the aptamers can be used to gain information about which genomic strand the aptamer lies. Additionally, both ends of the aptamer, or the whole aptamer, should be sequenced. Since the lengths are varying, this is essential to elucidate the enrichment patterns, binding motifs, and any potential structural elements encoded in the aptamer.

53.2.4

Troubleshooting

In Table 53.1, we describe common problems that could be faced during the genomic SELEX procedure. We suggest possible causes of these problems and recommend solutions.

53.3

Evaluation of Obtained Sequences

53.3.1

Computational Analysis of SELEX-Derived Sequences

The analysis of sequencing data obtained from the genomic SELEX experiment is usually focused on identification of genomic aptamers that have been enriched during the selection process. The sequenced data are referred to as *reads*. The typical procedure is to first perform an assembly of the reads into “contigs” based on sequence similarity to the reference genome, and then to identify high coverage peaks as putative binding motifs (Figure 53.3). All the activities can be performed step by step using a variety of available software or in a single run by using an automated pipeline named APART (Automated Pipeline for Analysis of

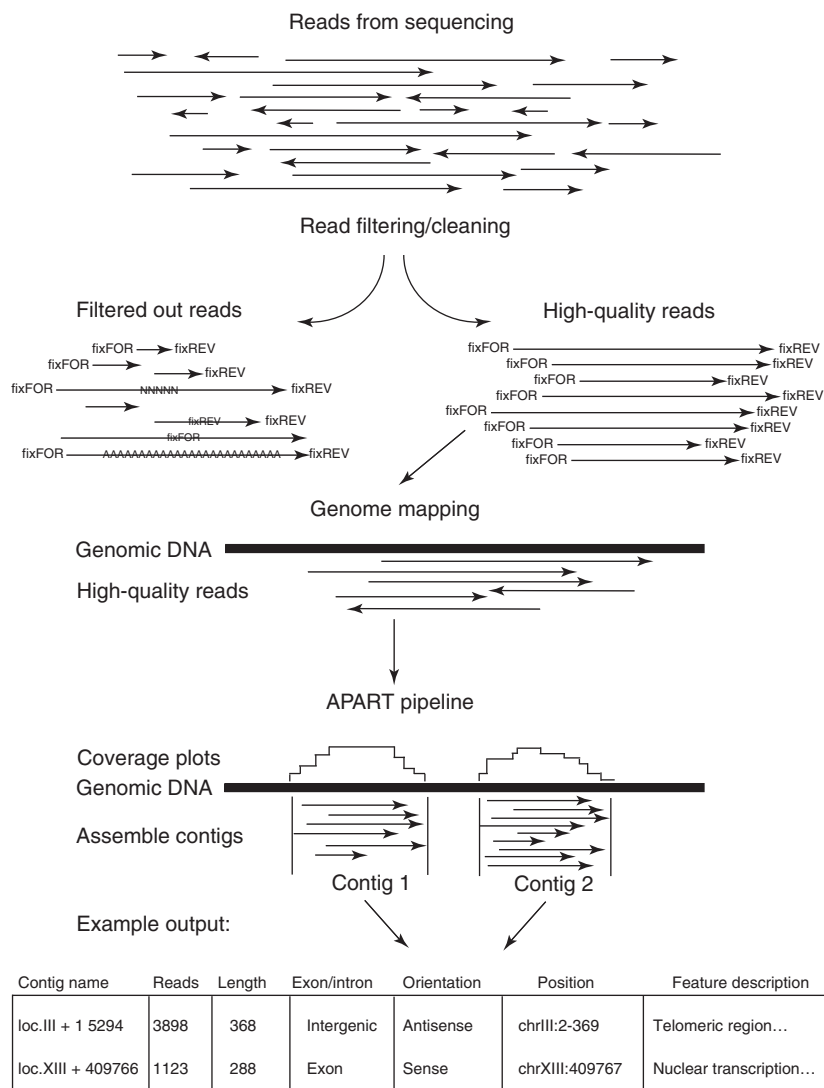


Figure 53.3 Computational analysis flowchart. The reads obtained from sequencing must first be filtered and cleaned in order to proceed with only the highest quality reads. Reads that are too short, contain none or only one of the fixed primers, have fixed primers in the middle of the sequence, are made up of more than 50% homopolymer, or contain unknown nucleotides (N) from sequencing must be filtered out. The

high-quality reads are then mapped to the genome using a variety of programs. The APART pipeline can then be used to group reads that map to the same location in the genome into contigs. Finally, the APART pipeline gives a table containing information such as number of reads, contig length, and location in the genome and a description of what is known about sequences from this region of the genome.

RNA Transcripts) [24]. We recommend using the APART pipeline for most of the steps because it has been well optimized for handling non-unique reads and includes the identification of highly abundant regions within the assembled contigs. The initial steps, including read preparation and genomic alignment, are highly dependent on the quality and content of the library. Thus it is recommended that these steps be performed before the automated APART workflow in order to adjust and optimize the parameters set.

53.3.1.1 Read Filtering and Cleaning

In the first step, low-quality sequence reads have to be removed from the pool. The threshold depends on the sequencing technology and should be weighted according to the quality score distribution within the library so the vast majority of the reads pass through to the subsequent steps. When no quality criteria are already given for selecting the reads, removing the sequence of the bottom 5% quality scores would be a good starting point. Optional filters remove reads with low information content (containing more than 50% of homopolymer) or reads containing unresolved bases ("N"). The next task is to locate the adaptor sequences (including the fixed primers, see Section 11.2.1) surrounding the genomic aptamer. This can be achieved using any pattern-matching program. Our recommendation is *patmatch* [25]. Its major benefit is the possibility to separately control insertions, deletions, and substitutions. For the first pass, we recommend using a value of 3 for every type of change. However, depending on library quality, it is worth to test mismatch allowance values between 2 and 4. Usually, the 3' ends of the reads are of lower quality than the 5', thus an increase of allowed mismatches by 1 or 2 for the 3' adaptor is usually a good solution. For downstream analysis, only the reads that contain both adapters should be used. When using the APART pipeline, all the above tasks can be performed by calling up a single automated script.

53.3.1.2 Genome Mapping

In order to map the reads to the reference genome, we recommend using the *bowtie* aligner [26] due to its speed, ability of reporting all matches for non-unique reads, and the extensive possibilities for control over the output. When using *bowtie*, we recommend the value of 1 up to 2 for libraries sequenced with high accuracy methods or 2 up to 3 for libraries obtained with technologies of lower base call accuracy.

For setting the alignments that *bowtie* should be allowed to report, we recommend the combined use of the `"-a,"` `"--best,"` and `"--strata"` options. This forces *bowtie* to report all matches for the particular read (`-a` option) sorted from the best to the worst (`--best` option) and limits the list to those within the best "stratum" (`--strata` option). "Stratum" refers to a level of alignment score (e.g., perfect match, one mismatch). The reasoning is that if a read is mapped to a repetitive or multicopy sequence, it will be mapped to all places where it could have originated, based on the sequencing data, and not be mapped to places that it is less likely to have originated. Reporting all best alignments maximizes the coverage of any given feature, allowing for higher copy features to show enrichment patterns.

Other options are mostly related to the speed performance of the aligner and should be adjusted according to the needs of the user. For compatibility with the APART pipeline, it is required to print the output in SAM (Sequence Alignment/Map, see [27]) format using the -S option and report the reads matching more than the maximum allowed number of times with -M option. The APART pipeline can automatically perform a genomic alignment, using the bowtie aligner with parameters set for the highest quality output. However, this will take longer than running bowtie manually.

53.3.1.3 Assembly and Annotation

The next step is to group reads into contigs, or regions of the genome where overlapping reads are found. The APART pipeline can automatically assemble the reads into contigs and generate genomic browser-compatible tracks in bed and wig formats. It will also utilize reads that map to multiple regions of the genome and group contigs together if they contain identical sets of reads. APART will provide a comprehensive functional annotation of the contigs based on a genome annotation and sequence similarity, including identification of all known non-coding RNAs and repeat units.

Running APART is straightforward. However, the default parameter set can be optimized for RNA-seq projects. When using it for genomic SELEX analysis, we recommend a couple of deviations from the default. The minimum number of reads per contig should be set to 1 in order to include all the reads in the statistics. Additionally, the contig clustering method should be set to use read name sets, instead of contig sequence. However, for libraries of low quality or derived from organisms with high genetic variability, which contain substantial number of mismatches in genomic alignment, sequence-based clustering may perform better.

53.3.1.4 Enrichment Analysis

The major aim of the SELEX procedure is to enrich the initial RNA pool with molecules that bind the bait. Thus, the investigation of the global enrichment of the output library is the primary analysis that indicates the success of the experiment. The first look should be focused on verification of the read distribution among the contigs. A successful experiment will have a highly stratified distribution. That is, the top few contigs should contain a large percentage of the total reads recovered from the experiment. For example, in our Hfq genomic SELEX experiment, over one-third of the sequenced reads belonged to the top 15 contigs [13]. However, there was a total of 1522 contigs, indicating a high stratification of enrichment. In case of the opposite situation, the efficiency of the selection procedure should be reconsidered (especially the ionic conditions for binding and stringency of the washing steps).

Another recommended analysis focuses on enrichment of the specific functional genomic regions. In some cases, it can result in conclusive statements about possible functions of the molecule used as bait. The basic analysis includes two steps. First, we recommend performing a calculation of the enrichment of genomic features (introns, exons, protein-coding genes, ncRNA genes, etc.) by comparing

the read number statistics generated by APART with the size of the respective feature in the genome. This step can also be used as a quality control for SELEX procedure if one knows what kind of sequences are expected to bind the bait (e.g., if one uses the intronic splicing enhancer as a bait, introns should be enriched). Second, the identification of enriched Gene Ontology (GO) categories within annotated contigs by using GeneTrail [28] or other GO enrichment analysis tool can be helpful in estimating the cellular functions of the molecule used as bait.

53.3.1.5 Benefits of Sequencing the Initial Library

For computational analysis of the genomic-SELEX-derived sequences, it is also recommended to sequence the initial library. There are two major benefits to sequencing the library. First, the presence of the selected read sequences in the initial library confirms its genomic origin. Thus, reads from a selected pool, which are confirmed to be in the initial library, can be utilized even if they are not matching the sequenced part of the reference genome. Second, after the assembly, the initial pool can be used as an exact background distribution for the enrichment analysis. It is possible that the initial library will vary from random genomic distribution because of differential accessibility of certain genomic regions for random priming, and therefore, it is useful to determine the possible artificial enrichment of certain sequences in the initial library.

53.3.1.6 Identification of the Binding Motif

The RNA motif responsible for binding to a target is usually determined by both the sequence and structure of the RNA. Unfortunately, the software available at present for *de novo* identification of complex RNA motifs is based on the assumption that all supplied sequences contain a unique motif. Since this is not always the case for genomic-SELEX-derived contigs, we suggest performing the motif search in the following several steps:

- 1) Cluster all contigs obtained from APART pipeline with sequence similarity threshold set to 70% using cd-hit [29] or any other clustering program of choice.
- 2) For each identified cluster, calculate the joint number of reads as the sum of reads for member contigs.
- 3) Depending on clustering results, identify a couple of clusters with the highest joint read number and perform a sequence motif search for contig sequences within the clusters. We recommend the use of Glam2 [30] program, which allows for the identification of gapped motifs. Check if there are similarities between the sequence motifs identified for individual clusters.
- 4) In parallel, perform the secondary-structure-based clustering of the contigs using either RNA Forester [31] or RNACluster [32] software. Compare clusters obtained with sequence-based clustering results.
- 5) Identify or refine consensus secondary structures for clusters of interest with Alifold [33] and RNA Consensus Shapes [34]. It is worthwhile to compare results from both tools, since they are based on different approaches.

53.3.2

Biochemical Analysis of the Genomic Aptamers

In this section, we discuss biochemical approaches suggested for the characterization of RNA molecules containing genomic aptamers mapped to the genome.

53.3.2.1 Validation of the RNA–Protein Interaction

The most evident control for the validation of selected genomic aptamers is to confirm the interaction between the enriched RNAs and the bait. Electrophoretic mobility shift assay (EMSA) [7, 35, 36] and filter binding assays [37, 38] are straightforward methods to check the interaction *in vitro* and to assess the binding strength. Alternatively, affinity between the transcript and the bait can be analyzed by surface plasmon resonance analysis [39] or fluorescence anisotropy [39]. It is also important to be aware that the entire RNA molecule may fold differently than the short genomic aptamer. In the context of the entire transcript, the selected domain may be involved in intramolecular interactions or be sterically inaccessible and therefore unable to bind the protein partner. Hence is advisable to repeat the binding assays once the full-length transcript is determined (see Section 11.3.2.3).

It is important to test whether the RNA interacts with the bait in the cellular environment. *In vivo* binding analysis is greatly facilitated when a specific antibody for the bait is available, so coimmunoprecipitation methods can be used (for example, see [40]). We recommend coimmunoprecipitation coupled with *in vivo* crosslinking, called *CLIP* [41], where the interaction is captured within the cell before cell lysis and therefore the amount of non-specific contaminating RNA is reduced. The precipitated RNA pool is then analyzed by Northern blot, RNase protection assay (RPA), or strand-specific RT-PCR for the presence of the genomic aptamer of interest.

Genomic SELEX provides an RNA-binding domain, but not necessarily the minimal binding site. Among the well-established methods to determine the exact contact sites, there are boundary determination analysis [36] and RNA footprinting [39]. In addition, if the bait protein has a metal-ion-binding pocket and the RNA binds in proximity to it, an iron-directed cleavage assay may be chosen [42].

53.3.2.2 Expression Analysis of Genomic Aptamers

Given that the selected genomic aptamers are merely randomly transcribed binding domains, it is crucial to confirm their expression in a target cell at a specific time point. Northern blot analysis is a convenient tool for detection of abundant genomic aptamers. Moreover, it provides information about the size of the entire RNA molecule comprising the selected genomic aptamer. However, the method lacks sensitivity and requires large amount of RNA material. Therefore, for analysis of genomic aptamers that are not abundant, RPA and strand-specific RT-PCR are the methods of choice. Nevertheless, the accuracy of RT-PCR results has been recently questioned because experimental artifacts are suspected. The main source of error is primer-independent cDNA synthesis caused primarily by RNA self-priming

[43, 44] or priming by other short RNAs or residual DNA after DNaseI treatment during RNA preparation. For that reason, it is essential to provide appropriate negative controls. It is recommended to perform the RT step in the absence of primer and then compare PCR products with those carried out with specific primer. It has also been reported that use of actinomycin D in the RT step blocks spurious synthesis of the cDNA [43].

53.3.2.3 Reconstruction of the Whole-Transcript-Comprising Genomic Aptamer

For studying the biological significance of genomic aptamers, it is important to know the size of the native transcript comprising the selected binding domain. As mentioned in Section 11.3.2.2, a Northern blot serves as a good method for both verification of the cellular expression and size determination of the native transcript. However, it lacks sensitivity. To assess the length of the entire RNA molecule, we suggest performing 3'- and 5' rapid amplification of cDNA ends [45–47] or RNA self-circularization followed by RT-PCR [48].

Alternatively, primer walking, an RT-PCR-based method, may be used (Figure 53.4). The genomic aptamer is first reverse transcribed from the desired RNA pool with a sequence-specific reverse primer and then amplified by PCR. In consecutive amplification steps, the reverse primer (used in the RT reaction) remains fixed, whereas the forward primer is placed several dozen nucleotides upstream. In subsequent PCR amplifications, the forward primer walks along the transcript by being shifted toward the 5' end of the cDNA, as long as the amplicon is detectable. On the other hand, in the 3' end mapping, the forward primer (used in the RT reaction) stays constant and the reverse primers are continuously placed downstream toward the 3' terminus. When both extreme ends are reached, the amplification with the outermost primers is performed to prove that the detected transcript spans the full length. Once the characterization of the genomic aptamer–protein interaction and the analysis of the RNA transcript containing the genomic aptamer are completed, it is possible to speculate about the function of the protein–RNA complex.

53.3.2.4 Determining the Function of the RNA–Protein Interaction

Finally, it is important to demonstrate the biological relevance of the protein–RNA interaction *in vivo*. Since RNA binding to a protein is not synonymous with function, the relevance of the protein–RNA interaction must be determined. It is difficult to generalize a strategy for all genomic aptamers because it is inherently connected to the nature of the bait. In cases where the function of a protein is unknown, the identity of the target RNA may give insight into the function [37]. Moreover, the RNA–protein interaction may be disrupted *in vivo* by mutating or knocking down the protein, and then, by analyzing the effect it has on the RNA, one may gain insight into the function [7]. If the protein of interest is important for the RNA localization in the cell, a knockdown of the protein may show a mislocalization of the RNA. The third strategy is specific for genomic SELEX against enzymes. The most obvious functional assay for these is to see if the RNA inhibits or accelerates the enzymatic activity of the protein [38]. These are only a few examples of the

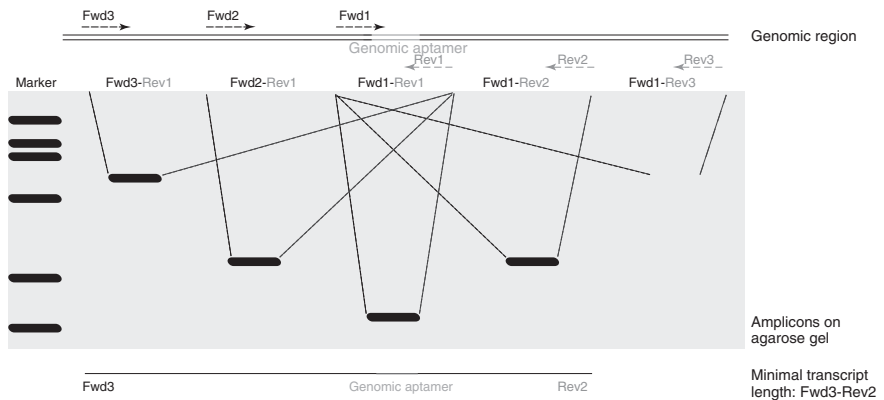


Figure 53.4 Reconstruction of the full-length transcript using a primer-walking method. Primer walking is a useful method to determine the full size of a transcript comprising the genomic aptamer selected in the genomic SELEX. The desired RNA pool is first reverse transcribed with a specific Rev1 primer, and then the genomic aptamer is amplified with PCR using Rev1 and Fwd1 primers. In consecutive amplification reactions, the 5' end of the putative transcript is mapped: the reverse primer Rev1 remains fixed but the forward primer is placed further upstream (Fwd2 and Fwd3) as long as the amplicon is detectable on agarose gel. In the 3' end mapping, the Fwd1 primer is used in the reverse transcription reaction and stays constant in the subsequent amplification reactions, whereas the reverse primers (Rev2 and Rev3) are shifted toward the 3' end of the transcript. Lack of Fwd1-Rev3 amplicon on agarose gel suggests that the 3' end of the transcript lies between sequences Rev2 and Rev3. When both extreme ends are reached, amplification with the most outer primers (Fwd3 and Rev2) is performed to prove that the detected transcript spans the full length.

many strategies that can be utilized to understand the importance of a particular protein–RNA interaction.

53.4

Conclusions and Outlook

Whole transcriptome analyses are delivering an unexpected high number of diverse transcripts, leading to the idea that probably every region of a genome is transcribed into RNA at some point of the organism's life cycle. Furthermore, the identity of a cell can be defined by its transcriptome. With this in mind, we need to find approaches to detect and functionally characterize those transcripts that are expressed rarely and at a low level. Because genomic SELEX is performed with libraries derived from the total DNA of an organism, every single part of the genome should be represented in the initial pool. We envision that in the near future, all genomic aptamers encoded within a genome, which interact with cellular proteins, RNAs, and metabolites, will have to be identified in order to describe the RNA regulon. To reach this goal, all available approaches will be necessary, and genomic SELEX will be a valuable approach to detect the low-abundance regulatory aptamers that otherwise might escape our attention.

Acknowledgments

Work in our laboratory is funded by the Austrian Genome Research Program GEN-AU “ncRNAs” and the Austrian Science Fund FWF grant no.^o F4308 to R.S.M.Z. was funded by a grant from the Austrian Ministry of Science and Research (GenAU project consortium “non-coding RNAs” D-110420-012-012 to N. Polacek)

References

1. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
2. Ellington, A.D. and Szostak, J.W. (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
3. Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. (1995) Diversity of oligonucleotide functions. *Annu. Rev. Biochem.*, **64**, 763–797.
4. Umezawa, Y., Shimada, T., Kori, A., Yamada, K., and Ishihama, A. (2008) The uncharacterized transcription factor YdhM is the regulator of the *nemA* gene, encoding N-ethylmaleimide reductase. *J. Bacteriol.*, **190**, 5890–5897.
5. Ogasawara, H., Hasegawa, A., Kanda, E., Miki, T., Yamamoto, K., and Ishihama, A. (2007) Genomic SELEX search for target promoters under the control of the PhoQP-RstBA signal relay cascade. *J. Bacteriol.*, **189**, 4791–4799.
6. Shimada, T., Fujita, N., Maeda, M., and Ishihama, A. (2005) Systematic search for the Cra-binding promoters using genomic SELEX system. *Genes Cells*, **10**, 907–918.

7. Kim, S., Shi, H., Lee, D.K., and Lis, J.T. (2003) Specific SR protein-dependent splicing substrates identified through genomic SELEX. *Nucleic Acids Res.*, **31**, 1955–1961.
8. Watrin, M., Von Pelchrzim, F., Dausse, E., Schroeder, R., and Toulmé, J.J. (2009) *In vitro* selection of RNA aptamers derived from a genomic human library against the TAR RNA element of HIV-1. *Biochemistry*, **48**, 6278–6284.
9. Singer, B.S., Shtatland, T., Brown, D., and Gold, L. (1997) Libraries for genomic SELEX. *Nucleic Acids Res.*, **25**, 781–786.
10. Zimmermann, B., Bilusic, I., Lorenz, C., and Schroeder, R. (2010) Genomic SELEX: a discovery tool for genomic aptamers. *Methods*, **52**, 125–132.
11. Lorenz, C., Von Pelchrzim, F., and Schroeder, R. (2006) Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat. Protoc.*, **1**, 2204–2212.
12. Montange, R.K. and Batey, R.T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.*, **37**, 117–133.
13. Lorenz, C., Gesell, T., Zimmermann, B., Schoeberl, U., Bilusic, I., Rajkowitsch, L., Waldsich, C., Von Haeseler, A., and Schroeder, R. (2010) Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Res.*, **38**, 3794–3808.
14. Jensen, K.B., Atkinson, B.L., Willis, M.C., Koch, T.H., and Gold, L. (1995) Using *in vitro* selection to direct the covalent attachment of human immunodeficiency virus type 1 Rev protein to high-affinity RNA ligands. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 12220–12224.
15. Gopinath, S.C., Misono, T.S., Kawasaki, K., Mizuno, T., Imai, M., Odagiri, T., and Kumar, P.K. (2006) An RNA aptamer that distinguishes between closely related human influenza viruses and inhibits haemagglutinin-mediated membrane fusion. *J. Gen. Virol.*, **87**, 479–487.
16. Hirao, I., Harada, Y., Nojima, T., Osawa, Y., Masaki, H., and Yokoyama, S. (2004) *In vitro* selection of RNA aptamers that bind to colicin E3 and structurally resemble the decoding site of 16S ribosomal RNA. *Biochemistry*, **43**, 3214–3221.
17. Tang, J., Xie, J., Shao, N., and Yan, Y. (2006) DNA aptamers that specifically recognize ricin toxin are selected by two *in vitro* selection methods. *Electrophoresis*, **27**, 1303–1311.
18. Lozupone, C., Changayil, S., Majerfeld, I., and Yarus, M. (2003) A more complex isoleucine aptamer with a cognate triplet. *RNA*, **9**, 1315–1322.
19. Yao, W., Adelman, K., and Bruenn, J.A. (1997) *In vitro* selection of packaging sites in a double-stranded RNA virus. *J. Virol.*, **71**, 2157–2162.
20. Mayer, G., Ahmed, M.S., Dolf, A., Endl, E., Knolle, P.A., and Famulok, M. (2010) Fluorescence-activated cell sorting for aptamer SELEX with cell mixtures. *Nat. Protoc.*, **5**, 1993–2004.
21. Tsuji, S., Hirabayashi, N., Kato, S., Akitomi, J., Egashira, H., Tanaka, T., Waga, I., and Ohtsu, T. (2009) Effective isolation of RNA aptamer through suppression of PCR bias. *Biochem. Biophys. Res. Commun.*, **386**, 223–226.
22. Zimmermann, B., Gesell, T., Chen, D., Lorenz, C., and Schroeder, R. (2010) Monitoring genomic sequences during SELEX using high-throughput sequencing: neutral SELEX. *PLoS ONE*, **5**, e9169.
23. Tereshko, V., Skripkin, E., and Patel, D.J. (2003) Encapsulating streptomycin within a small 40-mer RNA. *Chem. Biol.*, **10**, 175–187.
24. Zywicki, M., Bakowska-Zywicki, K., and Polacek, N. (2012) Revealing stable processing products from ribosome-associated small RNAs by deep-sequencing data analysis. *Nucleic Acids Res.*, **40**, 4013–4024.
25. Yan, T., Yoo, D., Berardini, T.Z., Mueller, L.A., Weems, D.C., Weng, S., Cherry, J.M., and Rhee, S.Y. (2005) PatMatch: a program for finding patterns

- in peptide and nucleotide sequences. *Nucleic Acids Res.*, **33**, W262–W266.
26. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
 27. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., (2009) The Sequence Alignment/Map format and SAM tools. *Bioinformatics*, **25**, 2078–2079.
 28. Backes, C., Keller, A., Kuentzer, J., Kneissl, B., Comtesse, N., Elnakady, Y.A., Müller, R., Meese, E., and Lenhof, H.P. (2007) GeneTrail – advanced gene set enrichment analysis. *Nucleic Acids Res.*, **35**, W186–W192.
 29. Li, W. and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
 30. Frith, M.C., Saunders, N.F., Kobe, B., and Bailey, T.L. (2008) Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput. Biol.*, **4**, e1000071.
 31. Höchsmann, M., Voss, B., and Giegerich, R. (2004) Pure multiple RNA secondary structure alignments: a progressive profile approach. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **1**, 53–62.
 32. Liu, Q., Olman, V., Liu, H., Ye, X., Qiu, S., and Xu, Y. (2008) RNACluster: An integrated tool for RNA secondary structure comparison and clustering. *J. Comput. Chem.*, **29**, 1517–1526.
 33. Hofacker, I.L., Fekete, M., and Stadler, P.F. (2002) Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, **319**, 1059–1066.
 34. Steffen, P., Voss, B., Rehmsmeier, M., Reeder, J., and Giegerich, R. (2006) RNASHapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics*, **22**, 500–503.
 35. Zolotukhin, A.S., Michalowski, D., Smulevitch, S., and Felber, B.K. (2001) Retroviral constitutive transport element evolved from cellular TAP(NXF1)-binding sequences. *J. Virol.*, **75**, 5567–5575.
 36. Kim, K.S., Ryoo, H., Lee, J.H., Kim, M., Kim, T., Kim, Y., Han, K., Lee, S.H., and Lee, Y. (2006) Selection and analysis of genomic sequence-derived RNA motifs binding to C5 protein. *Bull. Korean Chem. Soc.*, **27**, 699–704.
 37. Shtatland, T., Gill, S.C., Javornik, B.E., Johansson, H.E., Singer, B.S., Uhlenbeck, O.C., Zichi, D.A., and Gold, L. (2000) Interactions of *Escherichia coli* RNA with bacteriophage MS2 coat protein: genomic SELEX. *Nucleic Acids Res.*, **28**, E93.
 38. Windbichler, N., von Pelchrzim, F., Mayer, O., Csaszar, E., and Schroeder, R. (2008) Isolation of small RNA-binding proteins from *E. coli*: evidence for frequent interaction of RNAs with RNA polymerase. *RNA Biol.*, **5**, 30–40.
 39. Kim, H.J., Kwon, M., and Yu, J. (2007) Elucidation of the RNA target of linezolid by using a linezolid-neomycin B heteroconjugate and genomic SELEX. *Bioorg. Med. Chem.*, **15**, 7688–7695.
 40. Peritz, T., Zeng, F., Kannanayakal, T.J., Kilk, K., Eiríksdóttir, E., Langel, U., and Eberwine, J. (2006) Immunoprecipitation of mRNA-protein complexes. *Nat. Protoc.*, **1**, 577–580.
 41. Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005) CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods*, **37**, 376–386.
 42. Wassarman, K.M. and Saecker, R.M. (2006) Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science*, **314**, 1601–1603.
 43. Perocchi, F., Xu, Z., Clauder-Münster, S., and Steinmetz, L.M. (2007) Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res.*, **35**, e128.
 44. Haddad, F., Qin, A.X., Giger, J.M., Guo, H., and Baldwin, K.M. (2007) Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR. *BMC Biotechnol.*, **7**, 21.

45. Scotto-Lavino, E., Du, G., and Frohman, M.A. (2006) 3' end cDNA amplification using classic RACE. *Nat. Protoc.*, **1**, 2742–2745.
46. Scotto-Lavino, E., Du, G., and Frohman, M.A. (2006) 5' end cDNA amplification using classic RACE. *Nat. Protoc.*, **1**, 2555–2562.
47. Scotto-Lavino, E., Du, G., and Frohman, M.A. (2006) Amplification of 5' end cDNA with 'new RACE'. *Nat. Protoc.*, **1**, 3056–3061.
48. Meyer, B.J. and Southern, P.J. (1993) Concurrent sequence analysis of 5' and 3' RNA termini by intramolecular circularization reveals 5' nontemplated bases and 3' terminal heterogeneity for lymphocytic choriomeningitis virus mRNAs. *J. Virol.*, **67**, 2621–2627.

54

***In vivo* SELEX Strategies**

Thomas A. Cooper

54.1

Introduction

SELEX (Systematic Evolution of Ligands by EXponential enrichment) uses iterative selection strategies to enrich for specific RNA molecules from randomized pools based on binding or an RNA-mediated activity [1, 2]. The procedure is typically performed using recombinant protein in a cell-free system to identify optimal sequences bound by RNA binding proteins that have been associated with RNA-processing functions. Other uses include identification of RNA aptamers that small molecules or proteins of interest to human disease [3]. It is of interest to expand the application of a SELEX approach to include identification of RNA sequences that promote RNA-processing events *in vivo*.

The focus of this chapter is the use of a functional SELEX strategy to identify exonic elements that enhance pre-mRNA splicing *in vivo*. Exonic elements that enhance splicing are called exonic splicing enhancers (ESEs) and elements that repress splicing are exonic splicing silencers (ESSs). The selection strategy described in this chapter is particularly well suited to identify exonic sequences that enhance splicing, since they are “captured” in an mRNA by splicing. Initially, it was thought that ESEs and ESSs were in a small minority of exons and that only a few sequence classes of elements existed. It is now clear from bioinformatic and functional analyses that the variety and prevalence of exonic elements are much greater than previously realized [4]. It has also become clear that mutations that disrupt the function of ESEs and lead to aberrant splicing are a frequent cause of human disease [5, 6].

Several laboratories have used functional SELEX strategies to identify sequences that promote splicing in cell-free (*in vitro*) splicing assays that utilize HeLa nuclear extracts [7–10]. *In vitro* transcribed RNAs containing two exons separated by an intron are readily spliced *in vitro*. In these approaches, a poorly spliced downstream exon contains a randomized region of up to 20 nt. An RT-PCR-based strategy is devised to selectively amplify the exonic sequences of the spliced mRNAs, and then amplify and clone the selected sequences back into the pre-mRNA context and continue additional rounds of splicing and RT-PCR amplification. Two groups

used this approach to identify both purine-rich and non-purine-rich sequences that enhanced splicing in HeLa nuclear extracts [7, 10]. Among the sequences identified were binding sites for members of the SR family of splicing factors. Several SR protein family members have been shown to bind to ESEs and enhance splicing [11].

Cytoplasmic extracts prepared from HeLa cells are not competent for *in vitro* splicing and addition of individual SR proteins as recombinant proteins restore splicing activity [12]. This complementation assay was combined with a functional SELEX approach to identify the preferred ESE targets for individual SR proteins. Recombinant proteins for four SR proteins were used to complement cytoplasmic extracts for splicing of transcripts containing randomized nucleotides in the position of a required ESE of the downstream exon [7, 8]. Consensus ESE sequences required for enhanced splicing by each of the four SR proteins were derived and have been used to develop an ESE prediction program [13, 14].

54.2

Procedure Overview

This chapter describes an extension of the approaches described above to identify ESEs that function *in vivo*. A minigene plasmid expressing a pre-mRNA containing a poorly spliced internal alternative exon with a randomized sequence is transiently transfected and selective RT-PCR is used to amplify only those mRNAs that include the exon. Multiple cycles of transfection and selective RT-PCR enrich for exon sequences that enhance inclusion of the alternative exon. This approach has been described previously [15, 16] and has been used to identify exonic nucleotides within SMN exon 7 critical for exon inclusion [17]. More recently, *in vivo* iterative selection has been used to identify exonic elements that respond to PKA and CaMKIV signaling activity [18]. In an interesting twist, *in vivo* selection approaches have also been used to isolate elements within introns that affect splicing [19] and ESSs that inhibit splicing of the resident exon [20]. While these approaches were not iterative as described here, they demonstrate the flexibility and wide utility of *in vivo* selection.

The selection scheme is outlined in Figure 54.1. A synthetic single-stranded DNA template containing the randomized region ($n = 13$ nt for N_n in Figure 54.1) flanked by restriction sites is made double-stranded and amplified by up to five cycles of PCR. The PCR-amplified cassette is digested with the flanking restriction enzymes and directionally ligated into the alternative exon of the plasmid minigene. The exon is constructed to be recognized poorly in the absence of an ESE, leading to predominant exon skipping. This can be accomplished by either modifying the exon size (smaller is weaker) or by modifying the splice site sequences away from consensus [21]. The ligation reaction is then transiently transfected directly into cultured cells and total RNA is extracted after 40–48 h. RT-PCR is used to selectively amplify spliced mRNA that includes the randomized exon using oligos that prime across the exon junctions (Figure 54.2). The PCR product is digested

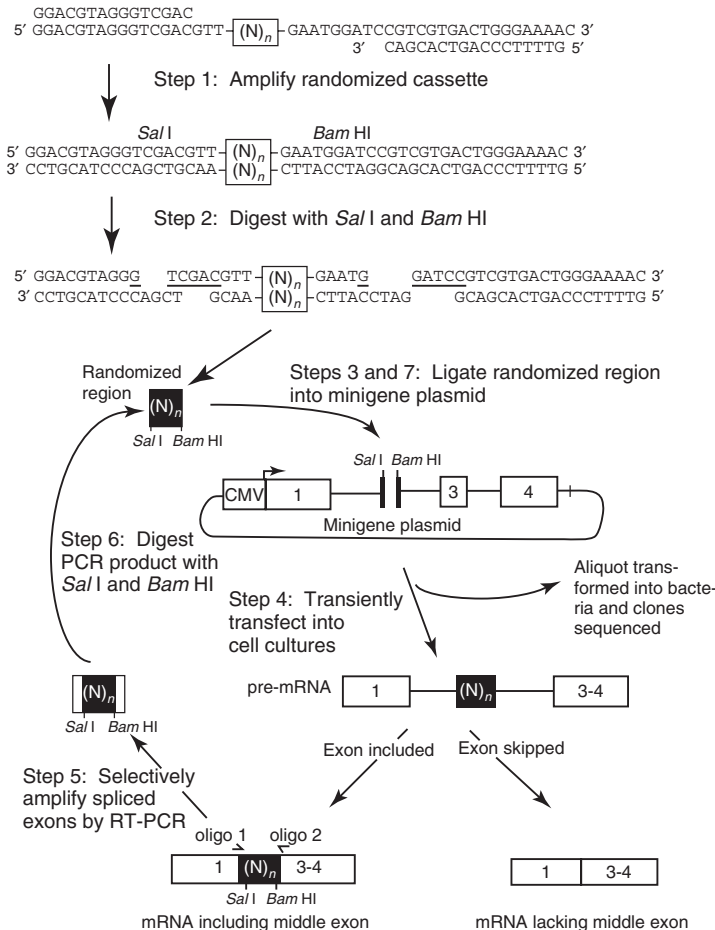


Figure 54.1 Iterative procedure to enrich for exon sequences that enhance splicing. The randomized region is represented as $(N)_n$.

by the restriction enzymes whose sites are incorporated into the single-stranded template to excise the selected randomized cassette. This digestion product contains a population of sequences that has been selected for enhanced exon inclusion as well as non-enhancing sequences spliced because of low levels of background splicing inherent in most minigenes. The excised randomized cassette is then ligated back into the minigene exon for additional cycles of selection.

The cassette exons obtained during the first three to four rounds of selection are evaluated with regard to: (i) sequence and (ii) enhancer activity. To identify sequence motifs that are enriched during selection, an aliquot of the ligation reaction from each round is transformed into bacteria; plasmid DNA is prepared from 20 to 30 individual colonies for sequencing. Plasmid DNA from individual clones is transiently transfected and the level of exon inclusion is assayed by

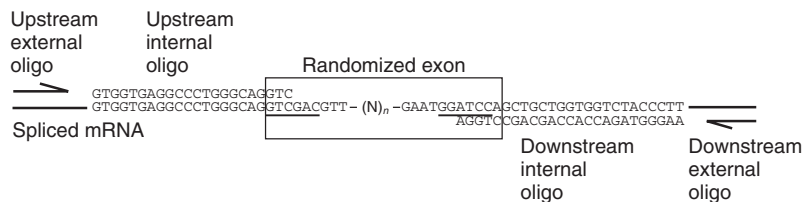


Figure 54.2 Amplification of the randomized exon included in the mRNA using double-nested RT-PCR. The mRNA sequence is represented by the sequence and thick lines on either side. The randomized exon is boxed. Following reverse transcription, PCR 1 is performed using the external primers. An aliquot of PCR 1 is then amplified using

the internal primers. Note that the internal primers are designed such that the last 2–4 nt anneal within the randomized exon. These oligonucleotides will prime DNA synthesis only on correctly spliced mRNAs and not on plasmid DNA, unspliced pre-mRNA, or mRNAs with a skipped randomized exon. Restriction sites are underlined.

RT-PCR to determine the splicing enhancer activity of individual cassettes. The level of inclusion of selected exons is compared to that of non-selected exons to determine whether the procedure enriched for splicing enhancers. The ability of selected cassettes to enhance splicing of a different alternative exon should be tested to determine whether the selected sequence has intrinsic enhancer activity, independent of the minigene used for its selection.

We have used this procedure in a fibroblast cell line to identify an AC-rich motif that enhances splicing [16]. Now that the feasibility of the approach has been established, it would be of interest to perform *in vivo* selection in different cell types to reveal the presence of cell-specific ESEs, in cells that overexpress a protein known to mediate enhancer activity to identify enhancer sequences that are preferred by individual proteins *in vivo* or, as published [18], under conditions of activated signaling to identify exonic elements that respond to specific signaling activity. It is also possible to use the procedure in cells deficient in particular proteins (either by RNAi or using cultures made null by homologous recombination). The latter could be useful to identify the roles of individual proteins that act cooperatively within exons to regulate exon inclusion.

54.2.1

Design of the Randomized Exon Cassette

The double-stranded cassette containing the randomized region is generated from low-cycle PCR using oligonucleotides complementary to the constant regions (Figure 54.1). The oligo containing the randomized region must contain several features (illustrated in Figure 54.1). First, priming sites for oligonucleotides located upstream and downstream of both restriction sites must allow PCR amplification of a double-stranded cassette (step 1, Figure 54.1). Second, it must contain two different restriction endonuclease sites on either side of the randomized region for directional cloning (Figure 54.1, *Sal* I and *Bam* HI sites). These must be unique to the minigene plasmid (see below), have incompatible overhangs (for directional

cloning and to prevent recircularization of the minigene plasmid), and should cut and ligate efficiently. It is most convenient if both enzymes are optimally active in the same buffer. Position the restriction sites within the oligonucleotide such that the fragment containing the randomized region is a different size than the other two fragments generated from digestion of the double-stranded cassette with both restriction enzymes. The restriction endonuclease digestion product containing the randomized region is gel-isolated and quantitated. Third, make certain that the constant regions within the selected exon do not contain sequences that can affect splicing such as known splicing enhancers, potential cryptic splice sites, or in-frame translation stop codons.

When designing the randomized region, keep in mind that as the number of randomized nucleotides increases, the number of molecules containing a particular sequence decreases in a constant amount of DNA. In our initial studies, we used a relatively low number of random positions because of the concern that a large variable region would result in insufficient copies of any one sequence to allow detection by RT-PCR. Now that the feasibility of the approach has been demonstrated and transfection efficiencies have greatly improved, it would be worthwhile testing larger randomized sequences. A second consideration is the total size of the exon. As the size of a cassette exon increases, the efficiency of exon recognition increases resulting in higher levels of exon inclusion that is less dependent on exon sequence [21]. Efficient exon inclusion can be compensated by decreasing the match of the 3' and/or 5' splice sites to the consensus. Another solution is to include the randomized region near the 3' splice site of a terminal exon that is inefficiently spliced (see Figure 54.3d) similar to the approaches used *in vitro* [7–10].

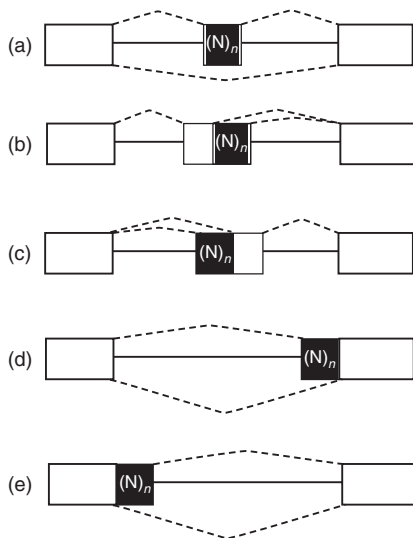


Figure 54.3 (a–e) Different pre-mRNA architectures used for selection. The randomized region is indicated by $(N)_n$. Splicing patterns are indicated by dashed lines.

54.2.2

Design of the Minigene

This section is geared to an internal cassette alternative exon as shown in Figures 54.1 and 54.3a. An advantage of this exon is that it selects for sequences that mediate exon inclusion either *via* the 3' or 5' splice sites or *via* enhancing interactions across the exon [22]. Different architectures of alternative exons can be utilized, which will theoretically select for sequences that enhance either 3' splice sites (Figure 54.3c,d) or 5' splice sites (Figure 54.3b,e).

The exon that is to receive the randomized region should have the appropriate restriction sites for directional cloning of the randomized cassette into an exon. It is useful to clone a large stuffer fragment into these sites so that plasmid cut with both enzymes can be distinguished from single cut and uncut plasmid. Low levels of contaminating plasmid containing a potentially “spliceable” sequence will generate a major contaminant of the selected sequences.

The stringency of selection can be adjusted by adjusting the basal level of exon inclusion in the absence of enhancer sequences. An exon that is completely skipped will theoretically select for stronger exonic enhancers than one that has a low basal level of inclusion. Note, however, that it is difficult to induce exon inclusion of particularly weak exons, no matter how strong the ESE. We prefer to have a low background level of exon inclusion to ensure that the exon is able to be spliced and to allow for selection of weaker ESE motifs. Establishing the desired balance of exon inclusion and exclusion may require modifying features such as exon size and strength of the 5' and 3' splice sites [21].

To determine the background level of exon inclusion in the absence of selected sequences, it is best to use pooled non-selected exons. This is accomplished by transforming an aliquot of the initial ligation (Figure 54.1) into high-efficiency competent bacteria according to standard procedures to the point of adding SOC and incubating at 37 °C. Instead of plating out the transformed cells on agar plates, use these bacteria to inoculate 100 ml of media including the appropriate antibiotic to prepare plasmid DNA for transfection. It is advisable to plate the transformed bacteria, pick a set of 10–15, and sequence to confirm that the majority contains the complete randomized cassette. Another approach to determine the background level of exon inclusion is to randomly pick colonies, prepare plasmids for transfection, transfect them individually, and average the level of exon inclusion. It is not unusual to obtain 1 or 2 out of 20 randomly picked non-selected plasmids that exhibit relatively high levels of exon inclusion demonstrating the strength of the effects of random sequence on the level of inclusion of an exon that is balanced between skipping and inclusion.

The presence of premature stop codons has dramatic effects on mRNA half-life and on splicing of the resident exon [23]. To limit analysis to the effects of ESEs, it is best not to have a natural open reading frame in the minigene mRNA.

Ubiquitously active transcription enhancers (such as RSV (respiratory syncytial virus) or CMV) are most useful as they allow the use of the minigene in almost any cell type. Cell-specific promoters are useful to ensure expression in the desired

cell of a mixed cell population such as primary cultures or a cell line induced to differentiate.

54.2.3

RT-PCR Amplification

The RT-PCR procedure performs two functions in this protocol: (i) it detects low amounts of RNA with very little background and (ii) selectively amplifies only those mRNAs that contain the alternative exon. These goals are accomplished by nested PCR (Figure 54.2). First, cDNA is synthesized using random hexamers (an alternative is to use an mRNA-specific primer that primes cDNA synthesis from a site located downstream from the external PCR primer pair; see Figure 54.2). Then mRNAs that include and exclude the alternative exon are amplified using oligonucleotides that anneal to the upstream and downstream exons (upstream and downstream external oligos, respectively; Figure 54.2). Finally, a double-nested PCR reaction is performed using the upstream and downstream internal oligos. Both the internal oligos will anneal only to correctly spliced mRNAs and prime DNA synthesis from within the alternative exon (see Figure 54.2), thereby selectively amplifying exons that are spliced into the mRNA. PCR products from mRNAs that skip the randomized exon or from unspliced pre-mRNA are not amplified. In addition, the specificity of nested primers means that a large number of cycles can be used, which is useful to amplify DNA from small amounts of RNA. Because the restriction sites are maintained, the PCR product is then digested and cycled through multiple rounds of ligation, transfection, and amplification.

54.2.4

Monitoring for Enrichment of Exon Sequences That Function as Splicing Enhancers

There are several approaches to monitor the success of the selection. The first is to determine whether identifiable sequence motifs are enriched after several rounds of selection. For example, we found clear enrichment of two different motifs after two and three rounds [16]. To obtain individual clones from each round of selection, 5 μ l of the 100 μ l ligation reaction containing randomized exon and minigene vector is transformed into competent bacteria (Figure 54.1). Miniprep DNA is prepared from 20 to 30 individual colonies and sequenced.

To determine whether the procedure is enriching for bona fide splicing enhancers, individual clones are tested by transient transfection and the level of exon inclusion is determined by a quantitative RT-PCR assay [16], primer extension, or RNase protection.

Alternatively, the levels of exon inclusion can be determined directly on the pools of RNA from each round using a quantitative RT-PCR assay. However, since cells express a high background level of mRNAs that lack the exon even in the absence of insert, this assay does not give reliable results. The reason for this is that transfected linear plasmid molecules (that fail to take up insert) become blunt-ended and circularized in cultured cells. This is demonstrated by

transfection of only vector DNA without ligase that results in the expression of a high level of mRNAs that lack the middle exon. While pools from sequential rounds of selection will show a relative enhanced exon inclusion, the background from recircularized plasmid makes it difficult to reliably quantitate the level of exon inclusion.

54.2.5

Troubleshooting

One major hurdle is to obtain high expression levels of spliced mRNA in transfected cultures. This requires high ligation and transfection efficiencies. The ligation efficiency of the minigene vector and insert preparations should be tested at a small scale before use. Ligation efficiencies can be optimized using a bacterial transformation assay that contains 50 ng of vector and a one- to fivefold molar excess of insert in a 10 μ l reaction. *Ligation efficiency* is defined as the number of colonies obtained per nanogram vector DNA. Important are low background in the absence of insert and efficient ligation when insert is included. A particularly labile component in the ligation reaction is ATP. Ligase buffer supplied by companies often includes ATP and it is best to store aliquots at -20°C . Each aliquot is used for no longer than one month.

If indicated, more vector DNA can be used for the ligation to increase the RNA expressed and the number of sequences available for selection. The amount of DNA used is limited primarily by the ability to isolate large amounts of cut minigene vector DNA.

Several alternative transfection reagents are commercially available and can give strikingly different efficiencies in different cell lines. It is often useful to talk to the suppliers to determine optimal conditions for individual cell lines.

The number of PCR cycles necessary to detect spliced mRNA is high. For example, our nested RT-PCR procedure required a total of 80 cycles [16]. The extremely high sensitivity of PCR makes it necessary to establish the rules listed below.

- 1) Use dedicated equipment for PCR, including a pipetman with aerosol-resistant tips, tube racks, microfuges, and vortexers. Do not use these for plasmid DNA.
- 2) Include “no RNA,” “no DNA,” and “minus reverse transcriptase” controls for each experiment to detect contamination.
- 3) Water is a common source of plasmid contamination. Purchase bottled water.
- 4) When using aliquot reagents, mark one tube as “in use” and use only that tube until it is finished.
- 5) If oligos are suspected to be contaminated with plasmid DNA, they can be reordered or purified by gel isolation on 10% non-denaturing acrylamide gel. Visualize the oligos by UV shadowing, cut out the piece of acrylamide containing the band, place it in a 1.5 ml Eppendorf tube and grind it using a blue tip. Add 1 ml of water and incubate with shaking overnight at 37°C . Filter the solution through a 0.2 mm Millex filter (Millipore) using a 3 ml syringe.

Protocols

Protocol 1: Preparation of the Randomized Cassette

- 1) One hundred nanograms of the single-stranded oligonucleotide containing the randomized region is amplified and converted to double-stranded DNA in a standard PCR reaction using oligos that flank the randomized region (Figure 54.1, step 1). Note that the amount of starting material is flexible and determines the number of random sequences represented. Prepare a 100 μ l reaction containing 500 ng of each flanking oligo, 1 $\times$ Vent DNA polymerase buffer (New England Biolabs: 10 mM KCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 20 mM Tris-HCl (pH 8.8 at 25 $^\circ\text{C}$), 2 mM MgSO_4 , 0.1% Triton X-100), 0.2 mM of each of the four dNTPs, and 1 unit Vent DNA polymerase (New England Biolabs). The PCR conditions depend on the T_m of the oligonucleotides and should be adjusted based on the quality of the final product as assayed on a non-denaturing acrylamide gel. Typically, we use no more than five cycles. The last step of PCR is 5 min at 72 $^\circ\text{C}$ to complete synthesis of the second strand.
- 2) Following the PCR reaction, add EDTA to 2 mM and NaCl to 0.2 M, and extract the PCR product once with an equal volume of phenol/chloroform/isoamyl alcohol (25 : 25 : 1) and then once with an equal volume of chloroform/isoamyl alcohol (25 : 1). Precipitate with 2.5 volumes of ethanol and wash the pellet once in 70% ethanol. The pellet is vacuum-dried and dissolved in 20 μ l of TE (10 mM Tris, pH 7.5, 0.1 mM EDTA).
- 3) Digest the PCR product with the appropriate restriction enzymes. Plan on using approximately 0.5 pmol of the randomized cassette DNA (7.4 ng of a 24 bp fragment) for every 1 μ g of vector (of 6.5 kb) in the ligation and 1 μ g of vector per round of selection. The restriction digest is loaded directly onto a 7% non-denaturing polyacrylamide gel (20 : 1 acryl:bis). Visualize the digestion products by staining in ethidium bromide and then gel-isolated as follows. Use a razor blade to cut a piece of acrylamide containing the band. Transfer to a 1.5 ml Eppendorf tube and cut into pieces of about 1 mm. Add 400 μ l of acrylamide elution buffer (0.5 M NH_4OAc , 10 mM 0.5 M EDTA) and incubate at 37 $^\circ\text{C}$ with shaking overnight. Spin 12 000 g for 5 min to pellet gel fragments and transfer the supernatant (completely free of gel fragments) to a clean tube and add 16 μ l of 5 M NaCl and 1 ml 100% ethanol, mix well and centrifuge (12 000 g) for 15 min. Residual soluble acrylamide from the gel precipitates in ethanol and acts as a carrier, so there is no need to add glycogen. Dissolve the pellet in 10–20 μ l TE and run an aliquot on an acrylamide gel alongside known amounts of size marker to estimate the DNA concentration.

Protocol 2: Vector Preparation

- 1) Cut 20 μ g of plasmid DNA with the appropriate enzymes. Load the restriction digest directly onto a 0.9% agarose gel. Use a single well comb or make a large

well from a standard comb by placing a tape over several teeth. Do not overload the gel since this can lead to smearing of the bands and contamination of the desired double-cut DNA with uncut plasmid DNA.

- 2) Isolate the DNA using QIAEX II Gel Extraction Kit (Qiagen).
- 3) Check the recovery and approximate concentration of the isolated DNA by running an aliquot on a minigel beside known amounts of marker DNA.

Protocol 3: Ligation of Randomized Exon Cassette into the Minigene Vector and Transfection into Cultured Cells

- 1) Ligate 1 μg of gel-isolated minigene plasmid vector and a twofold molar excess of insert at 15 °C overnight in 100 μl ligation reaction using 1 x ligation buffer and 400 units of T4 DNA ligase (New England Biolabs); 10 x ligation buffer (New England Biolabs): 500 mM Tris-HCl, pH 7.8, 100 mM MgCl_2 , 100 mM DTT, 10 mM ATP, 500 mg ml^{-1} BSA. Store 200 μl aliquots at -20 °C.
- 2) Transfect the ligation reaction directly into two 60 mm plates of QT35 cells in which cells were plated 4–18 h earlier at a density of 10^6 cells per plate. Conditions for transfection vary depending on the cell line. It is important to express as many pre-mRNA molecules as possible, so that a large pool of sequences are available for selection. The number of expressed RNAs is determined primarily by the ligation efficiency (the fraction of isolated vector that recircularizes with insert) and the transfection efficiency (the fraction of ligated molecules that make it to the nucleus and are transcribed). The conditions for ligation and transfection should be optimized as outlined in Section 53.2.5.
- 3) The transfection protocol will depend on the cell type used. In our original protocol, transfection was by the calcium phosphate protocol [15, 16]. We have since switched to FuGENE 6 (Roche).
- 4) Harvest cells 40–48 h after transfecting the DNA ligation reaction.

Protocol 4: RNA Extraction and DNase Treatment

- 1) Harvest RNA 40–48 h following transfection. Wash the plates once with 2 ml of cold 1 x PBS (136.9 mM NaCl, 2.7 mM KCl, 10.0 mM Na_2HPO_4 , 1.4 mM KH_2PO_4 ; bring pH to 7.4 with HCl; dilute 1 : 10 with ddH_2O and filter sterilize to make 1 x working stock). Let the plates drain at a 45° angle for 30 s and aspirate the remaining liquid at the bottom edge of the plate.
- 2) Add 650 μl of solution A (see below) to each of the two plates, then scrape off cells with a policeman and pool both plates into one 1.5 ml Eppendorf tube. Add 210 μl of chloroform/isoamyl alcohol (25 : 1). Vortex for at least 20 s with great agitation making sure that the phases mix. (RNA extraction solution A (52.5 ml): 25 ml of phenol (H_2O saturated), 25 ml of solution B, 2.5 ml of 2 M NaOAc (pH 4.0) and 180 μl β -mercaptoethanol. Can be stored at 4 °C for more than four months. RNA extraction solution B (50 ml): 4 M guanidinium thiocyanate, 25 mM sodium citrate, and then add ddH_2O to 45 ml. Adjust pH

- to 7.0 with 1 N HCl. Add ddH₂O to 50 ml. Can be stored at 4 °C for more than four months.)
- 3) Place the tubes on ice for 20 min, then spin (12 000 g) for 20 min at 4 °C.
 - 4) Transfer the upper (aqueous) phase to a new 1.5 ml Eppendorf tube containing 870 µl of isopropanol. Vortex to mix well and store at –20 °C for at least 1 h.
 - 5) Spin tubes (12 000 g) for 20 min at 4 °C. Pour off the supernatant and wash the pellet by vortexing with 1 ml of cold 75% ethanol (stored at –20 °C). Spin for 10 min. Pour off the supernatant immediately after the centrifuge stops; otherwise, the pellet may dislodge from the tube and pour out. Repeat this washing procedure a second time, then vacuum-dry the pellet.
 - 6) Redissolve the pellet in 50 µl of the DNase cocktail by gently vortexing for about 30 s, being careful not to introduce bubbles. Incubate at 37 °C for 30 min. For each sample, the DNase cocktail contains: 10 µl of 5 x *in vitro* transcription buffer (supplied by Promega: 200 mM Tris–HCl, pH 7.5 at 37 °C, 30 mM MgCl₂, 10 mM spermidine, 50 mM NaCl), 2.5 µl of 100 mM DTT, 0.2 µl of RNasin (40 unit ml^{–1}, Promega), 0.4 µl of DNase (Worthington DPFF DNase at 1.0 mg ml^{–1} (2 unit ml^{–1}) in 10 mM Tris–HCl, pH 7.5 in 10 µl aliquots stored at –80 °C) and 36.8 µl of ddH₂O.
 - 7) Make up a cocktail of stop solution to add following the incubation. For each reaction, add 2 µl of 0.5 M EDTA, 4 µl of 5 M NaCl, and 44 µl of ddH₂O. Extract once with phenol/chloroform/isoamyl alcohol and once with chloroform/isoamyl alcohol. Ethanol-precipitate the aqueous layer using 260 µl of ethanol. Redissolve the pellet in 384 µl ddH₂O, add 16 µl of 5 M NaCl and 1 ml of 100% ethanol, and vortex to mix. Store this precipitated RNA at –20 °C as a suspension in 70% ethanol; do not pellet the RNA.

Protocol 5: RT-PCR

- 1) Vortex each RNA sample to resuspend the RNA precipitate and remove 40 µl. This works out to be about one-fifth of a 60 mm plate or 10–15 µg of total RNA (from QT35 cell cultures prepared as described above). Spin down the RNA and vacuum-dry the pellet.
- 2) Make up the reverse transcriptase cocktail and use 20 µl aliquots to dissolve each RNA pellet. Then, for each reaction combine 2.0 µl of 10 x magnesium-free Taq PCR buffer (Promega: 50 mM KCl, 10 mM Tris–HCl (pH 9.0 at 25 °C), 0.1% Triton X-100), 2.4 µl of 25 mM MgCl₂, 2.0 µl of 1 mg ml^{–1} BSA, 2.0 µl of 10 mM of each of the four dNTPs, 1.0 µl of 100 pmol µl^{–1} hexamers, 1.0 µl of 100 mM DTT, 5 units of RNasin (Promega), 2 units of AMV (Avian Myeloblastosis Virus) reverse transcriptase (Life Sciences), and H₂O to 20 µl.
- 3) Once the RNA is dissolved, incubate at room temperature for 10 min and then transfer the tubes to 42 °C for 1 h. Following the reverse transcriptase reaction, heat the tubes at 95 °C for 5 min in a heating block and then immediately plunge into an ice/water slurry.

PCR 1

- 1) Make up a cocktail for 80 μ l PCR reactions (per reaction): 8.0 μ l of 10 x magnesium-free Taq buffer, 6.4 μ l of 25 mM MgCl_2 , 1.6 μ l of BSA (1 mg ml^{-1}), 0.5 μ l of Taq polymerase (Promega), 200 ng of upstream external oligo, 200 ng of downstream external oligo, and H_2O to 80 μ l. Note that additional deoxyribonucleotides is not necessary, since the final concentration is 0.2 mM from the reverse transcriptase reaction.
- 2) Mix and add 80 μ l of the cocktail to each of the 20 μ l reverse transcriptase reactions using a fresh pipette tip for each tube; transfer the reactions to 500 μ l tubes. Add two drops of light mineral oil and run a PCR program using appropriate conditions. RT-PCR conditions must be optimized empirically for sensitivity with low background by varying the number of cycles, the annealing and reaction temperatures, and the MgCl_2 concentration. PCR 1 is typically for 20 cycles.

PCR 2 (Nested PCR Reaction)

- 1) Make up a cocktail for 80 μ l PCR reactions (per reaction): 8.0 μ l of 10 x Taq PCR buffer, 6.4 μ l of 25 mM MgCl_2 , 0.64 μ l of 25 mM of each of the four dNTPs, 1.6 μ l of 1 mg ml^{-1} BSA stock, 200 ng of upstream internal oligo, 200 ng of downstream internal oligo, 2.5 U of Taq polymerase, and H_2O to 79 μ l.
- 2) Mix and add 79 μ l of the cocktail to 500 μ l tubes. Add 1 μ l of PCR 1, two drops of light mineral oil (if needed, depending on the PCR machine), and run a PCR program using appropriate temperatures and number of cycles. Note that for our nested PCR [16], it was necessary to use 79 °C for annealing and polymerization, since at 76 °C the reaction generated significant background and low amounts of the correct product after 60 cycles. At 81 °C, no PCR product was formed, probably because of an absence of annealing. At 79 °C, the reaction produced large amounts of only the correct product.
- 3) Add EDTA to 2 mM and NaCl to 0.2 M and extract the PCR product once with an equal volume of phenol : chloroform : isoamyl alcohol (25 : 25 : 1) and once with an equal volume of chloroform/isoamyl alcohol (25 : 1). Precipitate with 2.5 volumes of ethanol and wash the pellet once in 70% ethanol.
- 4) Digest the PCR product with the appropriate restriction enzymes and isolate the fragment from a 7% non-denaturing polyacrylamide gel as in step 3 of Protocol 1.5. Repeat for the desired number of rounds of selection.

Acknowledgments

This work was initiated by Mark Landree and completed by Lydia Coulter, and was supported by the National Institutes of Health.

References

1. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
2. Chen, X., Denison, L., Levy, M., and Ellington, A.D. (2009) Direct selection for ribozyme cleavage activity in cells. *RNA*, **15**, 2035–2045.
3. Yan, A.C., Bell, K.M., Breeden, M.M., and Ellington, A.D. (2005) Aptamers: prospects in therapeutics and biomedicine. *Front. Biosci.*, **10**, 1802–1827.
4. Wang, Z. and Burge, C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, **14**, 802–813.
5. Cartegni, L., Chew, S.L., and Krainer, A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.*, **3**, 285–298.
6. Wang, G.S. and Cooper, T.A. (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.*, **8**, 749–761.
7. Liu, H.X., Chew, S.L., Cartegni, L., Zhang, M.Q., and Krainer, A.R. (2000) Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol. Cell. Biol.*, **20**, 1063–1071.
8. Liu, H.X., Zhang, M., and Krainer, A.R. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.*, **12**, 1998–2012.
9. Tian, H.C. and Kole, R. (1995) Selection of novel exon recognition elements from a pool of random sequences. *Mol. Cell. Biol.*, **15**, 6291–6298.
10. Schaal, T.D. and Maniatis, T. (1999) Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.*, **19**, 1705–1719.
11. Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA*, **6**, 1197–1211.
12. Fu, X.D. (1995) The superfamily of arginine/serine-rich splicing factors. *RNA*, **1**, 663–680.
13. Smith, P.J., Zhang, C., Wang, J., Chew, S.L., Zhang, M.Q., and Krainer, A.R. (2006) An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum. Mol. Genet.*, **15**, 2490–2508.
14. Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. (2003) ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res.*, **31**, 3568–3571.
15. Cooper, T.A. (1999) *In vivo* SELEX in vertebrate cells, in *Methods in Molecular Biology*, vol. **118** (ed. S.R. Haynes), Humana Press, Totowa, NJ, pp. 405–417.
16. Coulter, L.R., Landree, M.A., and Cooper, T.A. (1997) Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol. Cell. Biol.*, **17**, 2143–2150.
17. Singh, N.N., Androphy, E.J., and Singh, R.N. (2004) *In vivo* selection reveals combinatorial controls that define a critical exon in the spinal muscular atrophy genes. *RNA*, **10**, 1291–1305.
18. Li, H., Liu, G., Yu, J., Cao, W., Lobo, V.G., and Xie, J. (2009) *In vivo* selection of kinase-responsive RNA elements controlling alternative splicing. *J. Biol. Chem.*, **284**, 16191–16201.
19. Bouck, J., Litwin, S., Skalka, A.M., and Katz, R.A. (1998) *In vivo* selection for intronic splicing signals from a randomized pool. *Nucleic Acids Res.*, **26**, 4516–4523.
20. Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. (2004) Systematic identification and analysis of exonic splicing silencers. *Cell*, **119**, 831–845.
21. Cooper, T.A. (1999) in *Methods in Molecular Biology*, vol. **118** (ed. S.R.

- Haynes), Humana Press, Totowa, NJ, pp. 391–403.
22. Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.*, **270**, 2411–2414.
23. Isken, O. and Maquat, L.E. (2007) Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.*, **21**, 1833–1856.

55

Gene Silencing Methods Using Vector-Encoded siRNAs or miRNAs

Ying Poi Liu and Ben Berkhout

55.1

Introduction

The discovery made in *Caenorhabditis elegans* that double-stranded RNA (dsRNA) triggers the degradation of complementary mRNA by means of a mechanism called RNA interference (RNAi) [1] has caught the attention of many investigators to use RNAi as a tool for gene knockdown in basic research as well as in therapeutic approaches. The introduced dsRNA is processed into ~22 bp small interfering RNAs (siRNAs) that are incorporated into a multi-protein complex called RNA-induced silencing complex (RISC) to cleave complementary target sequences. Initially, gene knockdown was obtained by transfection of synthetic siRNAs into the cell. Long-term knockdown became possible with the development of gene constructs that express siRNAs or transcripts that mimic cellular microRNAs (miRNAs).

For some research or clinical applications, an intensified RNAi protocol is critical to obtain a nearly complete gene knockdown, and other situations call for a combinatorial RNAi approach that uses multiple RNAi inducers. Several approaches can be used to induce combinatorial RNAi by expression of multiple siRNAs or miRNAs. In this chapter we provide practical advice and additional information for the construction of single RNAi vectors and also multi-siRNA/miRNA vectors. Furthermore, we describe the advantages and limitations of the most relevant viral vectors that can be used for delivery of the RNAi-inducing payload.

55.2

Background Information

RNAi is a conserved gene silencing mechanism in eukaryotes that triggers potent and sequence-specific inhibition of complementary mRNAs, which are either degraded or translationally suppressed [1–8]. In plants, insects, fungi, and nematodes, the prime role of RNAi is to defend the cell against invading viruses and to block transposon activity [7–10]. The RNAi pathway is induced by dsRNA, which can be derived from different sources such as infection by RNA viruses, inverted-repeat containing transgenes, aberrant transcription products, or

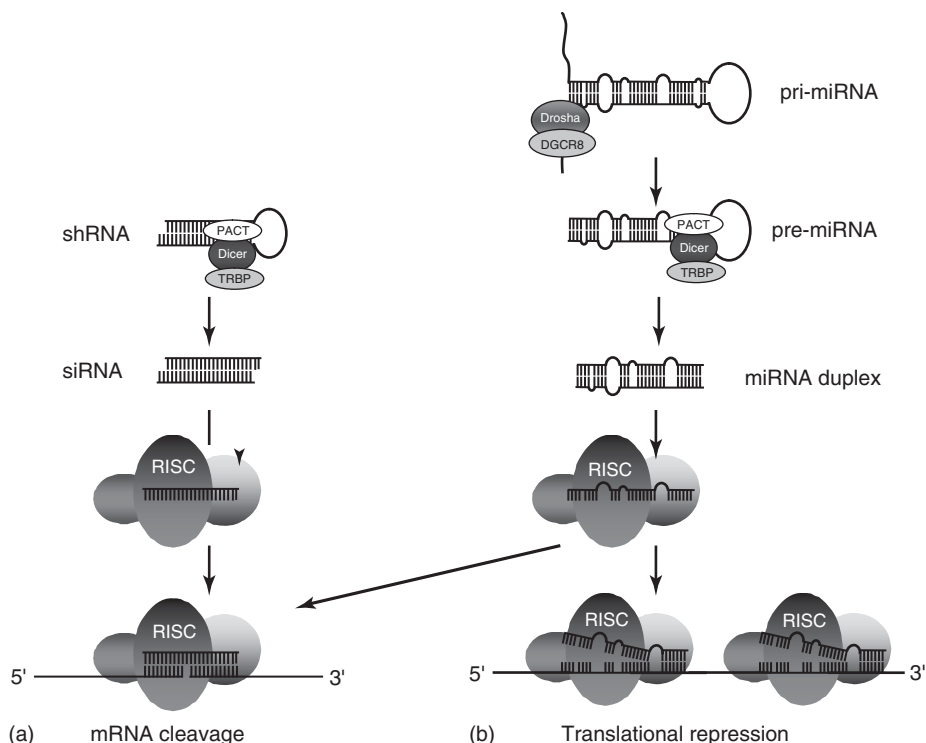


Figure 55.1 Vector-mediated RNAi induction using shRNA or miRNA constructs. (a) The expressed shRNA is processed by the Dicer/TRBP/PACT complex into siRNA duplexes. The guide strand of the siRNA duplex instructs RISC to cleave the complementary mRNA target. (b) The expressed pri-miRNA is first

processed in the nucleus by Drosha–DGCR into pre-miRNA. After nuclear export, this pre-miRNA is further processed by the Dicer/TRBP/PACT complex into a miRNA duplex. The mature miRNA strand guides RISC to translationally repress the mRNA target or to cleave it in the case of full complementarity.

complementary transcripts. These dsRNA precursors are cleaved by the cytoplasmic Dicer/TAR RNA binding protein (TRBP)/PKR activator (PACT) endonuclease complex to ~22 bp siRNAs with 2 nt 3'-overhangs [11–14]. The siRNA duplex is subsequently loaded into the multi-subunit ribonucleoprotein RISC complex. The “passenger” strand of the siRNA is cleaved and degraded [15–17], while the “guide” strand instructs RISC to cleave a perfectly complementary mRNA between position 10 and 11 within the base-paired duplex [18]. The efficacy of the RNAi response can be affected by reduced accessibility of the target mRNA sequence to RISC, for example, by a repressive RNA secondary structure or hindrance by RNA-bound proteins [19–23].

Some findings suggest that RNAi can also play an antiviral role in mammals [24, 25], although this issue is still a topic of debate [26–29]. In mammals, RNAi plays a major role in the regulation of cellular gene expression at the

post-transcriptional level *via* the extensive class of miRNA molecules (Figure 55.1b) [30]. The miRNAs are typically transcribed in the nucleus as primary microRNA (pri-miRNA) transcripts that fold into characteristic hairpin structures. These pri-miRNAs are processed into precursor microRNAs (pre-miRNAs) of ~70 nt by the nuclear RNase III-like endonuclease Drosha in complex with its cofactor DiGeorge syndrome critical region 8 (DGCR8) [31]. The pre-miRNAs are exported to the cytoplasm by Exportin-5 and further processed by the Dicer/TRBP/PACT endonuclease complex into an imperfect ~22 bp miRNA duplex [32–34]. The single-stranded mature miRNA directs RISC to specific mRNAs, which usually have multiple target sequences in their 3'-untranslated region (3'-UTR) [35, 36]. Imperfect base pairing of the miRNA with the target mRNA leads to translational repression, whereas a perfect duplex results in cleavage-mediated inactivation of the target mRNA [37–39]. In both cases, the targeted mRNAs are translocated to cytoplasmic compartments known as processing bodies (P-bodies), where storage, deadenylation, decapping, and finally degradation takes place [40, 41].

55.3

Construction of shRNA Vectors

Since its discovery, RNAi-based technology has been developed that exploits this cellular gene silencing mechanism. Transient knockdown of gene expression can be obtained by transfection of synthetic siRNAs into cells. Soon thereafter a variety of vectors that encode siRNAs – the so-called short hairpin RNAs (shRNAs) – were developed, which allow constitutive RNAi and stable gene silencing. The shRNA of 19–29 bp is modeled according to a pre-miRNA hairpin with a base-paired stem, a small apical loop, and a 3'-terminal UU-overhang (Figure 55.1a) [42–44]. Expression of shRNA constructs is mostly driven by RNA polymerase III promoters, such as those of the U6 RNA, H1 RNA, or tRNA genes, because of their natural function in the production of small transcripts with precise 5'- and 3'-ends [42, 44, 45]. Furthermore, these promoters produce high shRNA levels in all cell types. The shRNA is exported from the nucleus to the cytoplasm by Exportin-5 and processed by Dicer into an effective siRNA (Figure 55.1a).

The development of vector-based RNAi inducers enabled the design of gene therapy approaches against chronic viral infections caused by, for example, human immunodeficiency virus (HIV-1), hepatitis C virus (HCV), and herpes simplex virus (HSV) [46]. Although the initial results of HIV-1 inhibition by RNAi seemed very successful, it soon became apparent that the virus is a difficult and in fact a moving target because of its high mutation rate [47]. Targeting of HIV-1 with a single shRNA resulted in the emergence of RNAi-escape variants that acquired substitutions or deletions in the target sequence on prolonged culturing [21, 48, 49]. Furthermore, HIV-1 can also escape *via* alteration of the local RNA structure of the target sequence, which thereby becomes inaccessible to RISC [21]. To counteract viral escape, it is necessary to use multiple inhibitors that target different genome regions simultaneously in a combinatorial RNAi approach [50, 51]. For this purpose, multiple shRNA-expression cassettes could be combined in a single vector

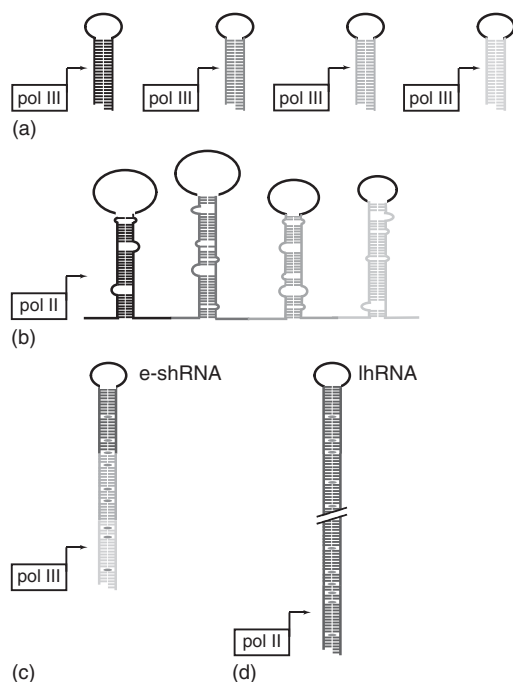


Figure 55.2 Expression strategies for multiple RNAi inducers. (a) A combinatorial RNAi approach using multiple shRNA cassettes driven by RNA polymerase III promoters in a single vector. (b) A miRNA polycistron driven by an RNA polymerase II promoter

that yields multiple miRNAs from a single transcript. (c) Combinatorial RNAi can also be obtained using extended shRNAs (e-shRNA). (d) The long hairpin RNA design (lhRNAs) is based on antiviral molecules that were successfully used in plant systems.

(Figure 55.2a) [52, 53]. Besides therapeutic applications, these vectors are also useful for basic research purposes, for example, to knock down the expression of multiple proteins in a single pathway or to enhance the knockdown of a specific gene.

Materials

Enzymes

- T4 DNA ligase (Life Technologies (LT) Invitrogen, Carlsbad, CA, USA)
- Ligase buffer (LT Invitrogen)
- Restriction enzymes
- Reddymix Master mix (Thermo Scientific Abgene, Epsom, UK)

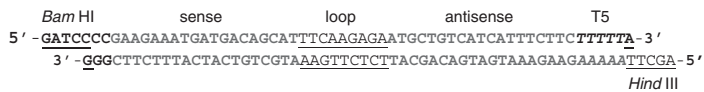
Buffers, Vectors, Bacterial Cells, and Other Requirements

- Annealing buffer (100 mM NaCl and 50 mM HEPES pH 7.4)
- pSUPER vector (OligoEngine, Seattle, WA, USA)
- BigDye terminator cycle sequencing kit (Life Technologies Applied Biosystems)

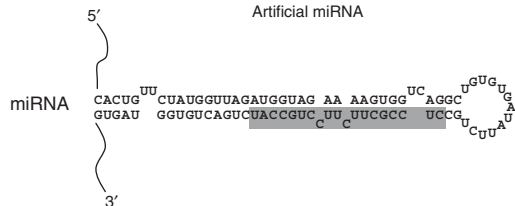
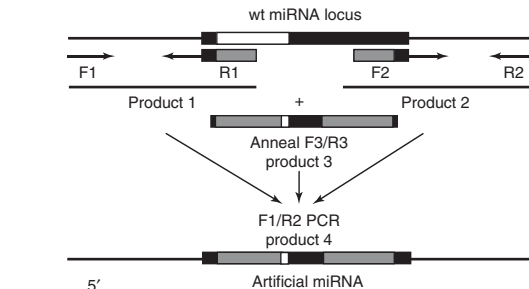
- pSilencer 2.0-U6 (Ambion Inc.; now commercially available: pSilencer 2.1-U6 series from Life Technologies Ambion)
- psiRNA-h7SKhygro (InvivoGen, San Diego, CA, USA)
- pGeneClip-BasicVector (Promega Corp., Madison, WI, USA)
- Gene Pulser II, cell electroporator (Bio-Rad, Hercules, CA, USA)
- 2 mm electroporation cuvettes (Eurogentec, San Diego, CA, USA)
- GT116 *Escherichia coli* cells (Δdcm , $sbcCD$) that are specifically designed for cloning and propagation of shRNA-expressing plasmids with hairpin structures (InvivoGen, San Diego, CA, USA)
- Luria Broth (LB) medium
- Eppendorf tubes (1.5 ml polypropylene tubes)
- Agar plates
- SmartLadder (Eurogentec, San Diego, CA, USA)
- Betaine (Sigma-Aldrich, St Louis, MO, USA)

Protocol and Notes

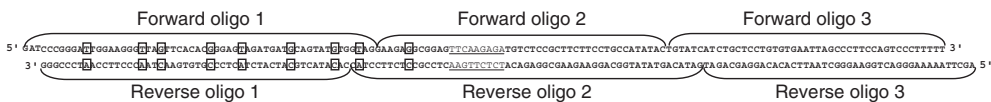
- 1) Anneal the forward and reverse DNA oligonucleotides encoding the shRNA sequence by combining 1 μ l of each oligonucleotide (3 μ g μ l⁻¹) in 48 μ l annealing buffer. The annealed double-stranded oligonucleotides have *Bam*HI- and *Hind* III-compatible ends for vector insertion (Figure 55.3a). We use the standard shRNA design that includes perfect complementary 19 nt sense and antisense strands connected with a 9 nt pSUPER loop (Figure 55.3a) [42]. Other shRNA designs have also been shown to efficiently induce RNAi [44].
- 2) Heat the sample to 94 °C for 5 min in a heating block or in a beaker of hot water and cool down to room temperature. The annealed oligonucleotides can be used immediately in a ligation reaction or can be stored at -20 °C.
- 3) Ligate 1 μ l of the annealed oligonucleotides into the *Bgl* II/*Hind* III-digested pSUPER vector (~40 ng) backbone overnight at 16 °C. Vectors that contain the human U6 or 7SK polymerase II promoter and the human U1 RNA polymerase II promoter can also be used to drive shRNA expression (see the section “Buffers, Vectors, Bacterial Cells, and Other Requirements”).
- 4) Before transformation, ligation mixes should be treated with *Bgl* II to reduce the level of background in the transformation. The *Bgl* II site is destroyed upon successful cloning of the annealed oligo pair, so only empty vector molecules will be cut by the enzyme. Add 0.5 μ l *Bgl* II to each ligation reaction and incubate for 30 min at 37 °C.
- 5) Transform 1 μ l of the ligation mix into 50 μ l of *E. coli* GT116 cells by electroporation (25 μ F, 200 Ω , and 2.5 kV).
- 6) Add 1 ml of LB medium to the cuvettes and incubate the cells for 45 min at 37 °C with shaking for recovery.
- 7) Transfer the cells to Eppendorf tubes and centrifuge for 5 min at 1485 g; remove ~800 μ l of the supernatant, resuspend the cells in the remaining supernatant and plate different volumes of cell suspension on two agar plates to increase the chance to obtain individual colonies.



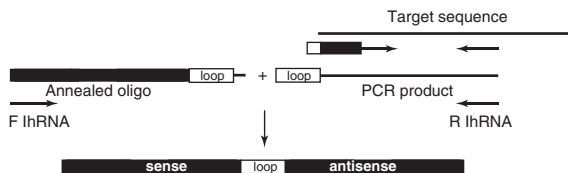
(a)



(b)



(c)



(d)

- 8) We use colony PCR to screen for correct clones. First, inoculate the colony in 50 μ l of LB medium and use 2.5 μ l of this suspension as template material for the PCR reaction. Include a negative (empty vector) and positive (construct that yields a similarly sized PCR product) control for the PCR.
- 9) The sequence of recombinant DNA constructs was verified using the BigDye terminator cycle sequencing kit. For sequencing of shRNA constructs with a perfect sequence repeat, we use a modified protocol in which Betaine is added to the reaction to a final concentration of 1 M and the sample is denatured at 98 °C.

Figure 55.3 Generation of siRNA, miRNA, e-shRNA, or lhrRNA expression cassettes. (a) Schematic of the annealed oligonucleotides that encode the shRNA transcript. The annealed oligonucleotide pair is inserted into the H1 RNA polymerase III expression cassette of pSUPER. Appropriate overhangs are created to enable immediate ligation into the vector. The 19 nt sense and antisense stretches are connected by a hairpin loop of 9 nt, which may be confined to only 5 nt if two additional base pairs form. The transcriptional termination signal consists of five consecutive thymidines (T5). The shRNA transcript folds into a small hairpin with a 2 nt UU 3'-overhang. The shRNA is processed by Dicer into an siRNA. (b) Construction of an artificial miRNA. The wild-type (wt) miRNA locus is amplified from cellular genomic DNA. The white box indicates the mature miRNA sequence, whereas the black parts indicate the remaining sequences of the pre-miRNA. Step 1: The 5'-part of the pri-miRNA was amplified using a forward primer F1 encoding a *Bam* HI site and a reverse primer R1 that partially encodes the HIV-1 sequence (gray). Step 2: the same was done for the 3'-part of the pri-miRNA using forward primer F2 that partially encodes HIV-1 sequences and a reverse primer R2 encoding *Bgl* II and *Xho* I sites. Step 3: two complementary oligonucleotides F3 and R3, which encode the stem-loop part of the new HIV-specific miRNA, were annealed. Step 4: the products from steps 1–3 are fused together by overlap extension in a PCR reaction that also includes the outer forward F1 and reverse R2 primers for amplification of the full-length product encoding the artificial miRNA. The final product is expressed as a pri-miRNA (bottom drawing of panel b) with

typical flanking sequences (here depicted as lines). The mature miRNA strand is boxed in gray. (c) Generation of an e-shRNA. Multiple oligonucleotides are designed that can anneal to each other and make up the forward and reverse strands. The appropriate overhangs are created upon annealing for cloning into the pSUPER vector. To facilitate cloning and sequencing, G–U base pairs can be introduced (A to G or C to T mutations, boxed nucleotides) in the hairpin stem. The e-shRNA folds into an extended hairpin with a 2 nt UU 3'-overhang. This e-shRNA encodes three different active siRNAs (marked in gray). (d) Construction of a lhrRNA. The lhrRNA construct is made by PCR amplification of the antisense strand of the hairpin RNA with a primer that encodes the loop sequences of the pSUPER system and DNA encoding the target sequence as a template. The sense strand of the hairpin RNA is made by annealing oligonucleotides containing the loop sequences and some additional nucleotides that are complementary to the antisense strand of the hairpin. The sense and the antisense strands of the hairpin were combined by a fusion PCR using primers (F and R lhrRNA) that encode the appropriate restriction sites. The lhrRNA folds into a long hairpin with a 2 nt 3' UU-overhang. In the example shown here, the lhrRNA encodes an active siRNA at the base of the hairpin (marked in gray) that is extended by consecutive sequences expanding the complementarity to the same target RNA. Thus, the lhrRNA design allows production of multiple siRNAs that target adjacent regions within the same mRNA, while the e-shRNA design targets different regions. Similar to the e-shRNA design, mutations (A to G or C to T) can be introduced to yield G–U base pairs in the hairpin stem (boxed nucleotides).

- 10) To combine identical or different shRNA-expression cassettes, standard cloning techniques can be used to insert them into a single vector.

55.4

Construction of miRNA Vectors

The original shRNA design has been optimized by inclusion of structural motifs from pri-miRNAs [54], that is, extended transcripts that fold into imperfect hairpin structures with mismatches, bulged nucleotides, and internal loops. These artificial miRNAs have been reported to induce RNAi more efficiently than the simple shRNA hairpins [55–57]. As an siRNA, a miRNA sequence that is perfectly complementary to the target mRNA sequence can be incorporated into the pre-miRNA hairpin to elicit mRNA cleavage. Similar to endogenous miRNAs, artificial miRNAs are processed in the nucleus by Drosha/DGCR and in the cytoplasm by Dicer/TRBP/PACT to yield a mature miRNA (Figure 55.1b). MiRNA constructs are usually expressed from RNA polymerase II promoters because this mimics the expression mode of most miRNA genes *in vivo* [58], although some endogenous miRNAs are transcribed from RNA polymerase III promoters [59]. For inhibition of multiple mRNAs or intensified knockdown of a single mRNA, a miRNA-like polycistron can be redesigned to encode multiple antivirals (Figure 55.2b) [57, 60–62].

Materials

- Most of the materials are identical to those listed in the section “Materials” under “Construction of shRNA Vectors”
- pcDNA6.2-GW/EmGFP-miR (LT Invitrogen)
- Cellular genomic DNA to amplify the wild-type miRNA locus as a template to construct the artificial miRNA

Protocol and Notes

- 1) Construction of a miRNA cassette consists of four steps (Figure 55.3b).
- 2) First, PCR-amplify the 5'-part of the pri-miRNA with a forward primer (F1) encoding a *Bam* HI site and a reverse primer (R1) with the HIV-1 sequence using the natural miRNA locus as the template. Subsequently, purify the PCR fragment from an agarose gel after electrophoresis. Similarly, PCR-amplify the 3'-part of the pri-miRNA with a forward primer (F2), containing HIV-1 sequences and a reverse primer (R2) introducing *Bgl* II and *Xho* I sites using again the natural miRNA locus as template and purify the PCR product from an agarose gel after electrophoresis.
- 3) Anneal two complementary oligonucleotides that encode the stem-loop structure of the antiviral miRNA as described for the shRNA vector construction (see section “Protocol and Notes”, steps 1 and 2, under “Construction of shRNA Vectors”).

- 4) The two PCR products and the annealed oligonucleotides contain some sequence overlaps that allow fusion by overlap extension PCR. Use the PCR products from step 2 and the annealed oligonucleotides from step 3 as combined templates with the forward (F1) and reverse (R2) primers to amplify the full-length product encoding the pri-miRNA (Figure 55.3b).
- 5) Purify the PCR product from an agarose gel after electrophoresis, digest with *Bam* HI and *Xho* I and ligate into the *Bam* HI- and *Xho* I-digested vector (pcDNA6.2-GW/EmGFP-miR).
- 6) Transform 1 μ l of the ligation mix into *E. coli* GT116 cells by electroporation (25 μ F, 200 Ω , and 2.5 kV) and plate different volumes or dilutions of cell suspension on two agar plates to increase the chance to obtain individual colonies.
- 7) For colony PCR, inoculate an individual colony in 50 μ l of LB medium and use 2.5 μ l of this suspension as template material for the PCR reaction. Include a negative and positive control for the PCR, for example, the empty vector and a construct that yields a PCR product of similar size.
- 8) To combine multiple miRNAs in a miRNA polycistron (Figure 55.2b), digest a miRNA hairpin construct with *Bgl* II and *Xho* I and purify the vector. This double digestion will open the vector at the 3'-flank of the miRNA insert. Digest the additional miRNA construct that is chained to the first miRNA cassette with *Bam* HI and *Xho* I (insert) and purify this fragment. Religate the insert into the *Bgl* II/*Xho* I sites of the vector. By repeating this procedure, multiple hairpins can be concatenated, resulting in vectors that direct the expression of the desired number of antiviral miRNA molecules.

55.5

Construction of Extended shRNAs and lhRNAs

Another possibility to implement combinatorial RNAi is the use of extended short hairpin RNAs (e-shRNAs), which are composed of multiple active siRNAs that are stacked on top of each other in a single hairpin design [63–67] (Figure 55.2c). We previously showed that the onset of viral replication could be significantly delayed with e-shRNAs compared to the use of a single shRNA [64]. Another design is the use of truly extended long hairpin RNAs (lhRNAs) that were used successfully in plants that lack the interferon system. These hairpins are generally much longer than e-shRNAs and their extended stem consists of contiguous viral sequences [65, 68–72] (Figure 55.2d). The lhRNA design has the theoretical advantage that numerous siRNAs can be produced from a single transcript. The disadvantage is that not all produced siRNAs are active inhibitors. We recently reported that there are strict size limitations to the e-shRNA and lhRNA designs [64], raising doubts about their usefulness in mammalian systems.

Materials

For materials, see the section “Materials” under “Construction of shRNA Vectors.”

Protocol and Notes

- 1) Compared with shRNA cassettes, the construction of e-shRNA expression cassettes differs in that multiple forward and reverse oligonucleotides are annealed to each other (Figure 55.3c). To improve plasmid stability in *E. coli* and to facilitate sequencing of the e-shRNA insert, G–U base pairs could be introduced into these hairpins by modification of the passenger strand of the siRNA, thus maintaining the active antisense strand sequence.
- 2) For annealing, combine 1 μl of each oligonucleotide ($3\text{ }\mu\text{g }\mu\text{l}^{-1}$) and add annealing buffer to a total volume of 50 μl . The number of oligonucleotides differs depending on the e-shRNA design. In the example provided in Figure 55.3c, we used six oligonucleotides to generate an e3-shRNA construct.
- 3) The annealed double-stranded oligonucleotide cassette with *Bam* HI- and *Hind* III-compatible ends is cloned into the *Bgl* II/*Hind* III-digested pSUPER vector.
- 4) For construction of lhRNAs, we first assemble the 5'-part of the expression cassette by annealing oligonucleotides that encode the passenger strand, the loop sequence, and some additional nucleotides encoding a small portion of the antisense strand (Figure 55.3d). The annealing is performed as described above.
- 5) The antisense strand was made by PCR amplification with a forward primer including the loop sequences of the pSUPER system and a reverse primer matching the 3'-end of the hairpin (Figure 55.3d).
- 6) The sense and antisense strands of the hairpin are combined by an overlap extension PCR with primers (F lhRNA and R lhRNA) that have terminal *Bam* HI and *Hind* III restriction sites. The final product is purified and can be used for cloning into the *Bgl* II and *Hind* III sites of the pSUPER vector.
- 7) Further cloning and sequencing of the e-shRNA and lhRNA constructs is performed as described in the section "Protocol and Notes" under "Construction of shRNA Vectors" (steps 3–9).

55.6

Production of Lentiviral Vectors Encoding Anti-HIV-1 shRNAs or e-shRNAs

Viral vectors have emerged as the vehicles of choice for delivery of the desired gene to the target cells. Different viral vectors have distinct characteristics, which make one vector more suitable for a certain purpose than others. To date, many viral vectors have been developed for gene therapy approaches. In Table 55.1, we summarize the application characteristics of five most frequently used vector systems that enable efficient transgene expression. These vectors are based on adeno-associated viruses, adenoviruses, retroviruses, lentiviruses, and simian virus 40. In our laboratory, we use the lentiviral vector to deliver anti-HIV-1 shRNAs or e-shRNAs to the target cells. As the lentiviral vector is based on HIV-1, antiviral shRNAs or e-shRNAs can possibly target HIV-1 sequences in the vector and cause low vector titers (vector targeting). Here, we describe a modified virus production method for such vector-targeting shRNAs or e-shRNAs.

Table 55.1 Characteristics of different viral vector types

Viral vector	Capacity (kb)	Advantages	Limitations	References
Adeno-associated virus	≤5	Lack of pathogenicity, broad target cell specificity	Low probability of integration, risk of insertional oncogenesis, small packaging capacity	[73–78]
Adenovirus	≤37	Broad target cell specificity, large packaging capacity	Immunogenicity, difficult production procedure	[79–84]
Retrovirus	≤8	Persistent gene expression due to chromosomal integration	Risk of insertional oncogenesis, unable to infect non-dividing cells, transcriptional silencing	[85–88]
Lentivirus	≤9	Persistent gene expression due to chromosomal integration, broad target cell specificity, ability to generate transgenic animals	Risk of insertional oncogenesis, transcriptional silencing	[88–94]
Simian Virus 40	≤5	Not immunogenic, broad target cell specificity	Small packaging capacity	[95, 96]

Materials

Cell Culture

- Human embryonic kidney (HEK) 293T cell line (ATCC, CRL-11268)
- SupT1 T cell line (ATCC, CRL-1942)
- Dulbecco's Modified Eagle's Medium (DMEM) (LT Invitrogen; brand: GIBCO®) (Catalog No 41966029) supplemented with 10% fetal calf serum (FCS; Thermo Scientific HyClone, South Logan, UT, USA), penicillin (100 U/ ml), streptomycin (100 µg/ ml), and minimal essential medium nonessential amino acids (MEM, Powder, NEAA) (DMEM/10% FCS)
- Advanced Roswell Park Memorial Institute (RPMI) 1640 medium (LT Invitrogen) (Catalog No 12633012) supplemented with 1% FCS (Thermo Scientific HyClone), 2 mM L-glutamine, 40 U/ml penicillin, and 40 µg/ ml streptomycin
- Dulbecco's phosphate buffered saline (D-PBS) solution, pH 7.4 (LT Invitrogen; brand: GIBCO) (Catalog No 14190086)
- 0.05% Trypsin with EDTA 4 Na solution (LT Invitrogen; brand: GIBCO) (Catalog No 25300054)
- Opti-MEM® Reduced Serum Medium, no Phenol Red (LT Invitrogen; brand: GIBCO) (Catalog No 11058021)

Constructs, Reagents, and Other Requirements for Lentiviral Vector Production

- 70- μ M Nylon cell strainers (BD Biosciences, Catalog No 352350; available from Thomas Scientific)
- Lipofectamine-2000 reagent (LT Invitrogen)
- Lentiviral vector constructs
- *Lentiviral vector packaging constructs*: pSYNGP [97] for the expression of a human codon-optimized gag-pol sequence without Rev Responsive Element (RRE); pRSV-Rev [98] for expression of the Rev protein, which interacts with the RRE on the vector genome and/or wild-type HIV-1 gag-pol transcripts. We used pVSV-g [98] for pseudotyping the vector with the VSV-g envelope protein (Figure 55.4a)
- *RNAi pathway competitor/inhibitor constructs*: the pSUPER construct p5xshRNA [52] was used that contains five repeated H1 promoters that express shRNAs against Gag-5, Pol-1, Pol-6, Pol-9, and Pol-47 to provide excess shRNAs to saturate the RNAi machinery (these shRNAs do not target the lentiviral vector). An siRNA against Dicer [44] was used to knock down Dicer function and to saturate RISC. A plasmid encoding a shRNA against Drosha was used to knock down Drosha function (a kind gift from Bryan Cullen, Duke University)
- 0.45 μ M cellulose acetate filter (GE Healthcare Whatman, Clifton, NJ, USA)
- Amicon Ultra-15 centrifugal filter units (Millipore, Billerica, MA, USA) (Catalog No UFC910096)
- Cryovials (Greiner Bio One, Kremsmuenster, Austria)
- BD FACSCanto II Flow Cytometer (BD Biosciences, San Jose, CA, USA)

Protocol and Notes

- 1) On day 1, plate 6.0×10^5 HEK 293T cells per well in a 6-well plate in 2 ml of DMEM/10% FCS without antibiotics. Optionally, the cells can be passed through a cell strainer to avoid clumping of cells.
- 2) On day 2, transfect the cells using Lipofectamine. Prepare the DNA–Lipofectamine 2000 complexes in Opti-MEM medium for each transfection as follows:
Dilute 0.95 μ g lentiviral vector plasmid, 0.6 μ g pSYNGP, 0.33 μ g pVSV-G, and 0.25 μ g pRSV-Rev in 250 μ l Opti-MEM medium. Add 2.9 μ g of RNAi competitor or inhibitor plasmid (see section “Constructs, Reagents, and Other Requirements for Lentiviral Vector Production”) to the co-transfection or 100 nM of siRNA against Dicer. These components will saturate the RNAi pathway, such that the anti-HIV-1 siRNA could not act on the lentiviral vector backbone during vector production. Thus, the vector RNA is spared from RISC attack and can be packaged and released as transducing unit.
- 3) For each transfection well, dilute 10 μ l Lipofectamine in 250 μ l Opti-MEM medium and incubate for 5 min at room temperature.
- 4) Combine the DNA and Lipofectamine solutions and mix gently. Incubate for 20 min at room temperature to allow formation of DNA–Lipofectamine complexes.

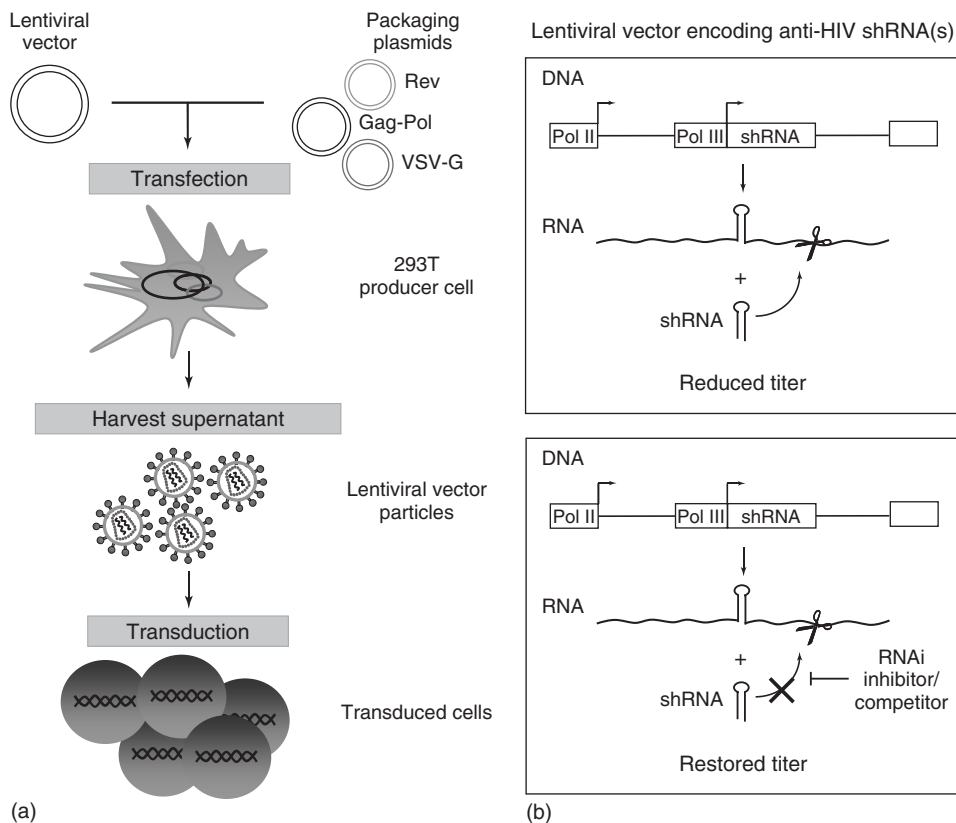


Figure 55.4 Production of lentiviral vectors encoding anti-HIV-1 shRNAs. (a) Schematic of the lentiviral vector production and transduction of target cells. For lentiviral vector production, the lentiviral vector plasmid is co-transfected with the packaging plasmids (constructs encoding Gag–Pol, Rev, and VSV-G envelope protein) into the 293T producer cells. The supernatant can be harvested, which contains the lentiviral vector particles. Subsequently, the particles can be used for transduction of the target cells. (b) Production of lentiviral vectors encoding anti-HIV shRNAs

raises the danger of targeting the lentiviral vector genome. Upper panel: anti-HIV shRNAs that could target sequences of the HIV-1-based lentiviral vector genome have reduced vector titers because of shRNA expression during vector production. Lower panel: the problem of a reduced vector titer due to vector targeting can be solved by using an RNAi inhibitor or competitor during vector production. Lentiviral vectors encoding e-shRNAs may have the same vector targeting problem that can be solved using the same strategy as for shRNAs.

- 5) Meanwhile, replace the medium on the cells with 500 μ l DMEM/10% FCS without antibiotics.
- 6) Add the DNA–Lipofectamine complexes dropwise to the cells.
- 7) On day 3, remove the transfection medium and add 2 ml of Opti-MEM medium supplemented with penicillin (30 U/ml), streptomycin (30 μ g/ml), and CaCl_2 (100 μ g/ml). Vector production in Opti-MEM medium improves

the vector titer about twofold compared to standard DMEM/10% FCS. Furthermore, the absence of serum improves filtration and concentration of the vectors.

- 8) On day 4, harvest the produced lentiviral vectors. To remove the cellular debris, centrifuge at 1305 g for 10 min and remove the viral supernatant. Filter the supernatant through a 0.45 μ m filter. The harvest can also be performed at day 5; however, the yield at day 4 is generally much higher (~10-fold).
- 9) The viral vectors can be concentrated using an Amicon Ultra-15 centrifugal filter spun at 4000 g. For a 15 ml vector preparation, we usually centrifuge for around 20–30 min to obtain a 250 μ l end volume.
- 10) Aliquot the viral vectors in cryovials and store the stocks at -80°C .
- 11) Determine the quantity of lentiviral vector particle production by ELISA (CA-p24) as described previously [99]. Transduction titers of the vectors are determined as follows. Transduce SupT1 T cells with a dilution series of lentiviral vectors that encode the enhanced Green Fluorescent Protein (eGFP) marker. Three days post-transduction, determine the transduction efficiency by detecting eGFP expression by FACS.

55.6.1

Troubleshooting

- 1) Annealing of the oligonucleotides for the construction of RNAi effector molecule expression cassettes may sometimes be unsuccessful due to self-complementarity. To prevent this, repeat the annealing step and add multiple temperature steps: 94°C for 5 min, 85°C for 4 min, 82°C for 4 min, 80°C for 4 min, 78°C for 4 min, and so on. This annealing protocol can be conveniently carried out in a PCR machine.
- 2) It is possible to treat the digested vectors with calf intestinal phosphatase (CIP); however, this is usually not necessary since the ends are not compatible. In case of CIP treatment of the vector, the oligonucleotides should be phosphorylated by kinase treatment.
- 3) Sequencing of hairpin RNA constructs may be problematic because of their tight structure. To improve the quality of sequencing, 1 M betaine is added to the reaction [100].
- 4) Be sure not to add antibiotics to the medium during transfection using lipofectamine because this will cause cell death. Cover the plates containing cells with saran wrap to diminish the chance of contamination.
- 5) A low lentiviral vector titer may be due to low quality of the plasmids, but also the cloned transgene can affect lentiviral vector titers. To find out, prepare fresh plasmids and repeat the procedure. The DNA concentration and purity can be assessed by UV spectrophotometry and on agarose gels. The ratio of absorbance at 260 *versus* 280 nm of pure DNA is generally between 1.8 and 2.0. Lower ratios indicate a contamination with proteins. Additional information on DNA purity can be inferred from the ratio of absorbance at 260 and

- 230 nm. This ratio should be between 2.0 and 2.2 for pure DNA. Lower ratios are obtained when the DNA is contaminated with organic chemicals.
- 6) When producing multiple different lentiviral vectors, a master mix of the packaging constructs can be prepared to simplify handling. In addition, for preparation of the Lipofectamine-Opti-MEM mix, a master mix can be prepared for all samples.
 - 7) When handling infectious material, be cautious and always use gloves and filter tips.
 - 8) For inhibition of the RNAi pathway, we have used several competitor/inhibitor constructs. Expression of excess amounts of shRNAs, an shRNA against Droscha, or an siRNA against Dicer have shown to boost the titer of vector with vector-targeting (e-)shRNAs (Figure 55.4b) [101]. The following constructs were also tested, but did not show an increase in titers: luciferase reporters that contain the corresponding target sequence to provide an RNAi target decoy, a plasmid encoding Adenovirus VA RNA that is a Dicer inhibitor [102], or a plasmid encoding the Ebola RNAi suppressor protein VP35 [103, 104].
 - 9) To test whether inefficient nuclear transport of viral genomic RNA was the cause of the titer reduction, we also tried overexpression of the chromosome region maintenance 1 (CRM1) protein, which is involved in the nuclear export of spliced and unspliced RNA. However, CRM1 overexpression did not increase the titer, indicating that reduced RNA export is not the cause of a reduced titer.
 - 10) Determine the titer of the lentiviral vectors for the actual target cell line because the transduction efficiency may differ between different cell types.

References

1. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.
2. Waterhouse, P.M., Wang, M.B., and Lough, T. (2001) Gene silencing as an adaptive defence against viruses. *Nature*, **411**, 834–842.
3. Voinnet, O. (2001) RNA silencing as a plant immune system against viruses. *Trends Genet.*, **17**, 449–459.
4. Wilkins, C., Dishongh, R., Moore, S.C., Whitt, M.A., Chow, M., and Machaca, K. (2005) RNA interference is an antiviral defence mechanism in *Caenorhabditis elegans*. *Nature*, **436**, 1044–1047.
5. Wang, X.H., Aliyari, R., Li, W.X., Li, H.W., Kim, K., Carthew, R., Atkinson, P., and Ding, S.W. (2006) RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science*, **312**, 452–454.
6. Segers, G.C., Zhang, X., Deng, F., Sun, Q., and Nuss, D.L. (2007) Evidence that RNA silencing functions as an antiviral defense mechanism in fungi. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 12902–12906.
7. Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z., and Zamore, P.D. (2008) Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science*, **320**, 1077–1081.
8. Sijen, T. and Plasterk, R.H. (2003) Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature*, **426**, 310–314.

9. Malone, C.D. and Hannon, G.J. (2009) Small RNAs as guardians of the genome. *Cell*, **136**, 656–668.
10. Yang, N. and Kazazian, H.H. Jr (2006) L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.*, **13**, 763–771.
11. Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, **101**, 25–33.
12. Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.*, **15**, 188–200.
13. Haase, A.D., Jaskiewicz, L., Zhang, H., Lainé, S., Sack, R., Gatignol, A., and Filipowicz, W. (2005) TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.*, **6**, 961–967.
14. Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002) Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.*, **21**, 5875–5885.
15. Leuschner, P.J., Ameres, S.L., Kueng, S., and Martinez, J. (2006) Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.*, **7**, 314–320.
16. Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell*, **123**, 607–620.
17. Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*, **123**, 621–629.
18. Hammond, S.M., Caudy, A.A., and Hannon, G.J. (2001) Post-transcriptional gene silencing by double-stranded RNA. *Nat. Rev. Genet.*, **2**, 110–119.
19. Tafer, H., Ameres, S.L., Obernosterer, G., Gebeshuber, C.A., Schroeder, R., Martinez, J., and Hofacker, I.L. (2008) The impact of target site accessibility on the design of effective siRNAs. *Nat. Biotechnol.*, **26**, 578–583.
20. Obernosterer, G., Tafer, H., and Martinez, J. (2008) Target site effects in the RNA interference and microRNA pathways. *Biochem. Soc. Trans.*, **36**, 1216–1219.
21. Westerhout, E.M., Ooms, M., Vink, M., Das, A.T., and Berkhout, B. (2005) HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome. *Nucleic Acids Res.*, **33**, 796–804.
22. Ameres, S.L., Martinez, J., and Schroeder, R. (2007) Molecular basis for target RNA recognition and cleavage by human RISC. *Cell*, **130**, 101–112.
23. Brown, K.M., Chu, C.Y., and Rana, T.M. (2005) Target accessibility dictates the potency of human RISC. *Nat. Struct. Mol. Biol.*, **12**, 469–470.
24. Bennasser, Y., Le, S.Y., Benkirane, M., and Jeang, K.T. (2005) Evidence that HIV-1 encodes an siRNA and a suppressor of RNA silencing. *Immunity*, **22**, 607–619.
25. Schnettler, E., de Vries, W., Hemmes, H., Haasnoot, J., Kormelink, R., Goldbach, R., and Berkhout, B. (2009) The NS3 protein of rice hoja blanca virus complements the RNAi suppressor function of HIV-1 Tat. *EMBO Rep.*, **10**, 258–263.
26. Umbach, J.L. and Cullen, B.R. (2009) The role of RNAi and microRNAs in animal virus replication and antiviral immunity. *Genes Dev.*, **23**, 1151–1164.
27. Cullen, B.R. (2006) Is RNA interference involved in intrinsic antiviral immunity in mammals? *Nat. Immunol.*, **7**, 563–567.
28. Berkhout, B. and Jeang, K.T. (2007) RISCy business: microRNAs, pathogenesis, and viruses. *J. Biol. Chem.*, **282**, 26641–26645.
29. Haasnoot, J., Westerhout, E.M., and Berkhout, B. (2007) RNA interference against viruses: strike and counter-strike. *Nat. Biotechnol.*, **25**, 1435–1443.

30. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
31. Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.*, **18**, 3016–3027.
32. Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.*, **17**, 3011–3016.
33. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004) Nuclear export of microRNA precursors. *Science*, **303**, 95–98.
34. Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, **436**, 740–744.
35. Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, **293**, 834–838.
36. Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.*, **16**, 720–728.
37. Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005) Principles of microRNA-target recognition. *PLoS Biol.*, **3**, e85.
38. Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
39. Yekta, S., Shih, I.H., and Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, **304**, 594–596.
40. Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R. (2005) MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat. Cell Biol.*, **7**, 719–723.
41. Sen, G.L. and Blau, H.M. (2005) Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nat. Cell Biol.*, **7**, 633–636.
42. Brummelkamp, T.R., Bernards, R., and Agami, R. (2002) A system for stable expression of short interfering RNAs in mammalian cells. *Science*, **296**, 550–553.
43. Siolas, D., Lerner, C., Burchard, J., Ge, W., Linsley, P.S., Paddison, P.J., Hannon, G.J., and Cleary, M.A. (2005) Synthetic shRNAs as potent RNAi triggers. *Nat. Biotechnol.*, **23**, 227–231.
44. Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002) Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.*, **16**, 948–958.
45. Kawasaki, H. and Taira, K. (2003) Short hairpin type of dsRNAs that are controlled by tRNA(Val) promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells. *Nucleic Acids Res.*, **31**, 700–707.
46. Haasnoot, P.C.J. and Berkhout, B. (2006) *Handbook of Experimental Pharmacology*, Springer-Verlag, Berlin, Heidelberg, pp. 117–150.
47. Goudsmit, J., Back, N.K.T., and Nara, P.L. (1991) Genomic diversity and antigenic variation of HIV-1: links between pathogenesis, epidemiology and vaccine development. *FASEB J.*, **5**, 2427–2436.
48. Das, A.T., Brummelkamp, T.R., Westerhout, E.M., Vink, M., Madiredjo, M., Bernards, R., and Berkhout, B. (2004) Human immunodeficiency virus type 1 escapes from RNA interference-mediated inhibition. *J. Virol.*, **78**, 2601–2605.
49. Boden, D., Pusch, O., Lee, F., Tucker, L., and Ramratnam, B. (2003) Human immunodeficiency virus type 1 escape from RNA interference. *J. Virol.*, **77**, 11531–11535.
50. Grimm, D. and Kay, M.A. (2007) Combinatorial RNAi: a winning strategy for

- the race against evolving targets? *Mol. Ther.*, **15**, 878–888.
51. Liu, Y.P. and Berkhout, B. (2008) Combinatorial RNAi strategies against HIV-1 and other escape-prone viruses. *Int. J. Biosci. Technol.*, **1**, 1–10.
 52. Ter Brake, O., Konstantinova, P., Ceylan, M., and Berkhout, B. (2006) Silencing of HIV-1 with RNA interference: a multiple shRNA approach. *Mol. Ther.*, **14**, 883–892.
 53. Ter Brake, O., 't Hooft, K., Liu, Y.P., Centlivre, M., von Eije, K.J., and Berkhout, B. (2008) Lentiviral vector design for multiple shRNA expression and durable HIV-1 inhibition. *Mol. Ther.*, **16**, 557–564.
 54. Zeng, Y., Wagner, E.J., and Cullen, B.R. (2002) Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell.*, **9**, 1327–1333.
 55. Silva, J.M., Li, M.Z., Chang, K., Ge, W., Golding, M.C., Rickles, R.J., Siolas, D., Hu, G., Paddison, P.J., Schlabach, M.R. *et al.* (2005) Second-generation shRNA libraries covering the mouse and human genomes. *Nat. Genet.*, **37**, 1281–1288.
 56. Boden, D., Pusch, O., Silberman, R., Lee, F., Tucker, L., and Ramratnam, B. (2004) Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. *Nucleic Acids Res.*, **32**, 1154–1158.
 57. Liu, Y.P., Haasnoot, J., Ter Brake, O., Berkhout, B., and Konstantinova, P. (2008) Inhibition of HIV-1 by multiple siRNAs expressed from a single microRNA polycistron. *Nucleic Acids Res.*, **36**, 2811–2824.
 58. Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.*, **23**, 4051–4060.
 59. Borchert, G.M., Lanier, W., and Davidson, B.L. (2006) RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.*, **13**, 1097–1101.
 60. Snyder, L.L., Esser, J.M., Pachuk, C.J., and Steel, L.F. (2008) Vector design for liver-specific expression of multiple interfering RNAs that target hepatitis B virus transcripts. *Antiviral Res.*, **80**, 36–44.
 61. Snyder, L.L., Ahmed, I., and Steel, L.F. (2009) RNA polymerase III can drive polycistronic expression of functional interfering RNAs designed to resemble microRNAs. *Nucleic Acids Res.*, **37**, e127.
 62. Aagaard, L.A., Zhang, J., von Eije, K.J., Li, H., Saetrom, P., Amarzguoui, M., and Rossi, J.J. (2008) Engineering and optimization of the miR-106b cluster for ectopic expression of multiplexed anti-HIV RNAs. *Gene Ther.*, **15**, 1536–1549.
 63. Liu, Y.P., Haasnoot, J., and Berkhout, B. (2007) Design of extended short hairpin RNAs for HIV-1 inhibition. *Nucleic Acids Res.*, **35**, 5683–5693.
 64. Liu, Y.P., von Eije, K.J., Schopman, N.C., Westerink, J.T., Ter Brake, O., Haasnoot, J., and Berkhout, B. (2009) Combinatorial RNAi against HIV-1 using extended short hairpin RNAs. *Mol. Ther.*, **17**, 1712–1723.
 65. Weinberg, M.S., Ely, A., Barichiev, S., Crowther, C., Mufamadi, S., Carmona, S., and Arbuthnot, P. (2007) Specific inhibition of HBV replication *in vitro* and *in vivo* with expressed long hairpin RNA. *Mol. Ther.*, **15**, 534–541.
 66. Saayman, S., Barichiev, S., Capovilla, A., Morris, K.V., Arbuthnot, P., and Weinberg, M.S. (2008) The efficacy of generating three independent anti-HIV-1 siRNAs from a single U6 RNA Pol III-expressed long hairpin RNA. *PLoS ONE*, **3**, e2602.
 67. Sano, M., Li, H., Nakanishi, M., and Rossi, J.J. (2008) Expression of long anti-HIV-1 hairpin RNAs for the generation of multiple siRNAs: advantages and limitations. *Mol. Ther.*, **16**, 170–177.
 68. Konstantinova, P., de Vries, W., Haasnoot, J., Ter Brake, O., de Haan, P., and Berkhout, B. (2006) Inhibition of human immunodeficiency virus type 1 by RNA interference using

- long-hairpin RNA. *Gene Ther.*, **13**, 1403–1413.
69. Nishitsuji, H., Kohara, M., Kannagi, M., and Masuda, T. (2006) Effective suppression of human immunodeficiency virus type 1 through a combination of short- or long-hairpin RNAs targeting essential sequences for retroviral integration. *J. Virol.*, **80**, 7658–7666.
 70. Barichiev, S., Saayman, S., von Eije, K.J., Morris, K.V., Arbuthnot, P., and Weinberg, M.S. (2007) The inhibitory efficacy of RNA POL III-expressed long hairpin RNAs targeted to untranslated regions of the HIV-1 5' long terminal repeat. *Oligonucleotides*, **17**, 419–431.
 71. Akashi, H., Miyagishi, M., Yokota, T., Watanabe, H., Hino, T., Nishina, K., Kohara, M., and Taira, K. (2005) Escape from the interferon response associated with RNA interference using vectors that encode long modified hairpin-RNA. *Mol. Biosyst.*, **1**, 382–390.
 72. Watanabe, T., Sudoh, M., Miyagishi, M., Akashi, H., Arai, M., Inoue, K., Taira, K., Yoshida, M., and Kohara, M. (2006) Intracellular-diced dsRNA has enhanced efficacy for silencing HCV RNA and overcomes variation in the viral genotype. *Gene Ther.*, **13**, 883–892.
 73. Wu, Z., Asokan, A., and Samulski, R.J. (2006) Adeno-associated virus serotypes: vector toolkit for human gene therapy. *Mol. Ther.*, **14**, 316–327.
 74. Kotin, R.M., Siniscalco, M., Samulski, R.J., Zhu, X.D., Hunter, L., Laughlin, C.A., McLaughlin, S., Muzyczka, N., Rocchi, M., and Berns, K.I. (1990) Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 2211–2215.
 75. Nakai, H., Montini, E., Fuess, S., Storm, T.A., Grompe, M., and Kay, M.A. (2003) AAV serotype 2 vectors preferentially integrate into active genes in mice. *Nat. Genet.*, **34**, 297–302.
 76. Russell, D.W. (2003) AAV loves an active genome. *Nat. Genet.*, **34**, 241–242.
 77. Smith, R.H. (2008) Adeno-associated virus integration: virus versus vector. *Gene Ther.*, **15**, 817–822.
 78. Donsante, A., Miller, D.G., Li, Y., Vogler, C., Brunt, E.M., Russell, D.W., and Sands, M.S. (2007) AAV vector integration sites in mouse hepatocellular carcinoma. *Science*, **317**, 477.
 79. Cao, H., Koehler, D.R., and Hu, J. (2004) Adenoviral vectors for gene replacement therapy. *Viral Immunol.*, **17**, 327–333.
 80. Buchbinder, S.P., Mehrotra, D.V., Duerr, A., Fitzgerald, D.W., Mogg, R., Li, D., Gilbert, P.B., Lama, J.R., Marmor, M., Del Rio, C. *et al.* (2008) Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *Lancet*, **372**, 1881–1893.
 81. McElrath, M.J., De Rosa, S.C., Moodie, Z., Dubey, S., Kierstead, L., Janes, H., Defawe, O.D., Carter, D.K., Hural, J., Akondy, R. *et al.* (2008) HIV-1 vaccine-induced immunity in the test-of-concept Step Study: a case-cohort analysis. *Lancet*, **372**, 1894–1905.
 82. O'Brien, K.L., Liu, J., King, S.L., Sun, Y.H., Schmitz, J.E., Lifton, M.A., Hutnick, N.A., Betts, M.R., Dubey, S.A., Goudsmit, J. *et al.* (2009) Adenovirus-specific immunity after immunization with an Ad5 HIV-1 vaccine candidate in humans. *Nat. Med.*, **15**, 873–875.
 83. Lasaro, M.O. and Ertl, H.C. (2009) New insights on adenovirus as vaccine vectors. *Mol. Ther.*, **17**, 1333–1339.
 84. Hartman, Z.C., Appledorn, D.M., and Amalfitano, A. (2008) Adenovirus vector induced innate immune responses: impact upon efficacy and toxicity in gene therapy and vaccine applications. *Virus Res.*, **132**, 1–14.
 85. Baum, C., Schambach, A., Bohne, J., and Galla, M. (2006) Retrovirus vectors: toward the plentivirus? *Mol. Ther.*, **13**, 1050–1063.

86. Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K. *et al.* (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.*, **118**, 3132–3142.
87. Mohamedali, A., Moreau-Gaudry, F., Richard, E., Xia, P., Nolta, J., and Malik, P. (2004) Self-inactivating lentiviral vectors resist proviral methylation but do not confer position-independent expression in hematopoietic stem cells. *Mol. Ther.*, **10**, 249–259.
88. Ellis, J. (2005) Silencing and variegation of gammaretrovirus and lentivirus vectors. *Hum. Gene Ther.*, **16**, 1241–1246.
89. Buchschacher, G.L. and Wong-Staal, F. Jr (2000) Development of lentiviral vectors for gene therapy for human diseases. *Blood*, **95**, 2499–2504.
90. Laufs, S., Guenechea, G., Gonzalez-Murillo, A., Zsuzsanna, N.K., Luz, L.M., del, V.C., Jonnakuty, S., Hotz-Wagenblatt, A., Jens Zeller, W., Bueren, J.A., and Fruehauf, S. (2006) Lentiviral vector integration sites in human NOD/SCID repopulating cells. *J. Gene Med.*, **8**, 1197–1207.
91. Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Ponzoni, M., Bartholomae, C., Sergi, L., Benedicenti, F., Ambrosi, A., Di Serio, C. *et al.* (2006) Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat. Biotechnol.*, **24**, 687–696.
92. Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Bartholomae, C.C., Ranzani, M., Benedicenti, F., Sergi, L.S., Ambrosi, A., Ponzoni, M. *et al.* (2009) The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. *J. Clin. Invest.*, **119**, 964–975.
93. Lois, C., Hong, E.J., Pease, S., Brown, E.J., and Baltimore, D. (2002) Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science*, **295**, 868–872.
94. Pfeifer, A. (2004) Lentiviral transgenesis. *Transgenic Res.*, **13**, 513–522.
95. Strayer, D.S. (1999) Gene therapy using SV40-derived vectors: what does the future hold? *J. Cell. Physiol.*, **181**, 375–384.
96. Vera, M. and Fortes, P. (2004) Simian virus-40 as a gene therapy vector. *DNA Cell Biol.*, **23**, 271–282.
97. Kotsopoulou, E., Kim, V.N., Kingsman, A.J., Kingsman, S.M., and Mitrophanous, K.A. (2000) A Rev-independent human immunodeficiency virus type 1 (HIV-1)-based vector that exploits a codon-optimized HIV-1 gag-pol gene. *J. Virol.*, **74**, 4839–4852.
98. Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998) A third-generation lentivirus vector with a conditional packaging system. *J. Virol.*, **72**, 8463–8471.
99. Jeeninga, R.E., Hoogenkamp, M., Armand-Ugon, M., de Baar, M., Verhoef, K., and Berkhout, B. (2000) Functional differences between the long terminal repeat transcriptional promoters of HIV-1 subtypes A through G. *J. Virol.*, **74**, 3740–3751.
100. Haqqi, T., Zhao, X., Panciu, A., and Yadav, S.P. (2002) Sequencing in the presence of betaine: improvement in sequencing of the localized repeat sequence regions. *J. Biomol. Tech.*, **13**, 265–271.
101. Liu, Y.P., Vink, M.A., Westerink, J.T., Ramirez de Arellano, E., Konstantinova, P., Ter Brake, O., and Berkhout, B. (2010) Titers of lentiviral vectors encoding shRNAs and miRNAs are reduced by different mechanisms that require distinct repair strategies. *RNA*, **16**, 1328–1339.
102. Andersson, M.G., Haasnoot, P.C., Xu, N., Berenjian, S., Berkhout, B., and Akusjärvi, G. (2005) Suppression of RNA interference by adenovirus virus-associated RNA. *J. Virol.*, **79**, 9556–9565.

103. de Vries, W., Haasnoot, J., van der Velden, J., van Montfort, T., Zorgdrager, F., Paxton, W., Cornelissen, M., van Kuppeveld, F., de Haan, P., and Berkhout, B. (2008) Increased virus replication in mammalian cells by blocking intracellular innate defense responses. *Gene Ther.*, **15**, 545–552.
104. Haasnoot, J., de Vries, W., Geutjes, E.J., Prins, M., de Haan, P., and Berkhout, B. (2007) The Ebola virus VP35 protein is a suppressor of RNA silencing. *PLoS Pathog.*, **3**, e86.

56

Using Chemical Modification to Enhance siRNA Performance

Jesper B. Bramsen, Arnold Grünweller, Roland K. Hartmann, and Jørgen Kjems

56.1

Introduction

The use of small interfering RNAs (siRNAs) to knock down gene function has truly revolutionized functional genomics in mammalian cell culture and holds great promise in therapeutics. However, limitations in potency, duration, delivery, and specificity of the gene knockdown (KD) are still major obstacles that need further addressing. Recent studies have successfully improved most aspects of siRNA performance by the introduction of numerous types of chemical modifications into discrete base positions of the siRNA during its chemical synthesis. Indeed, chemical modification of siRNAs can positively affect the pharmacokinetic properties (cellular uptake, biodistribution, metabolic stability, body clearance), can reduce off-target effects (OTEs) and unwanted immune responses, and may increase the duration of a KD. Here, we provide concepts and guidelines as to how the performance of siRNAs can be tailored by combining chemical modifications to suit specific applications *in vitro* and *in vivo*. This chapter is designed to advise researchers who have already identified one or a few siRNAs that specifically and effectively knock down the target of interest in cell culture and want to proceed to primary cells or *in vivo* applications.

56.2

Numerous siRNA Designs: What siRNA Architecture to Choose?

The canonical siRNA design used today mimics natural Dicer cleavage products and comprises a 21 nt guide strand antisense to a given RNA target and a complementary passenger strand annealed to form an siRNA duplex with a 19 bp dsRNA stem and 2 nt 3'-overhangs at both ends [1, 2]. The guide strand is also termed *antisense strand* (AS) and the passenger strand is named *sense strand* (SS). Another popular class of siRNA designs instead mimic Dicer substrates to enhance incorporation into RNAi pathways and thereby siRNA potency; synthetic 27 bp siRNAs [3–6] and short hairpin RNAs (shRNAs; Figure 56.1, see also Chapter 55) typically with RNA stems ≥ 21 bp [7] are such Dicer-substrate siRNAs (DsiRNAs) that were indeed reported to be more potent than corresponding conventional 19 bp siRNAs [8].

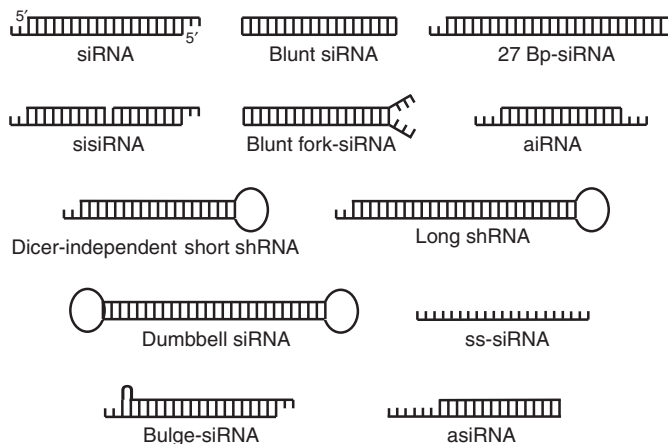


Figure 56.1 Examples of functional siRNA designs. The classical siRNA design (19 bp with 2 nt overhangs at the 3'-ends) can be modified in terms of duplex length, end structure, and variation in the length of

the AS or SS (see text for details); upper strands: SS; lower strands: AS. The 27 bp siRNA, long shRNA, and dumbbell siRNA are Dicer substrates and are thus termed *Dicer-substrate siRNAs* (DsiRNAs).

Yet, a study reporting cellular interferon (IFN) induction by 27 bp siRNAs has posed serious concerns [9], as have observations of inconsistent performance of early symmetric 27 bp siRNA designs [3]. As a precaution, most researches still choose to use the “safe” canonical siRNA despite recent improvements of the 27 bp siRNAs [5, 6, 10]. Other efficient, yet less widely used siRNA designs, are the Dicer-independent short shRNAs with RNA stems ≤ 19 bp [11–13], blunt 19 bp siRNAs [14, 15], blunt fork-siRNAs [16, 17], single-stranded siRNAs (ss-siRNAs) [18–20], dumbbell-shaped circular siRNAs [21], asymmetrical siRNAs (aiRNAs), and asymmetric shorter duplex siRNA (asiRNA) harboring a shortened SS [13, 22, 23], bulge-siRNA [24], siRNAs with a unilateral 2 nt 3'-overhang [25], and siRNAs composed of segmented passenger strands (sisiRNAs) [26]. A schematic representation of the various siRNA designs is illustrated in Figure 56.1.

56.3

siRNA Tolerance Toward Modification

As a general trend, the entire SS as well as the 3'-proximal part and 3' overhang of the AS are fairly tolerant to chemical modification (Figure 56.2), whereas the 5'-phosphate, the 5'-proximal part, and central positions of the AS are sensitive, especially to multiple or bulky modifications [26, 27]. These observations are in good agreement with solved structures of Ago2/nucleic acid complexes. The 5'-phosphate of the guide strand is consistently found in the Ago MID domain [28], an essential interaction, which predicts that modification at this location will in most cases compromise binding by Ago2 [29, 30]. Furthermore, Ago-mediated

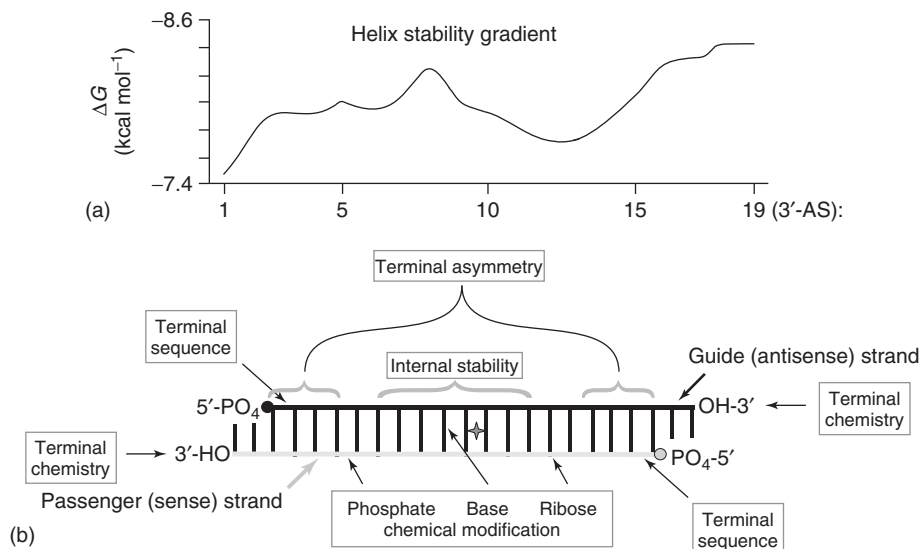


Figure 56.2 siRNA duplex architecture and modification. (a) Energy profile of helix stability asymmetry between siRNA duplex ends to favor guide strand selection by RISC (adapted from Ref. [35]). (b) Architecture and design options for siRNA duplexes (adapted from Ref. [36]). The entire passenger strand (in gray) as well as the 3'-proximal part (~ nt 17–21) and 3'-overhang of the guide strand (in black) are fairly tolerant to chemical modification. In contrast, the 5'-phosphate, 5'-proximal

part, and central positions of the guide strand are sensitive to chemical modification, especially to multiple or bulky modifications. Ago-mediated target RNA cleavage requires base pairing, or at least an A-helical duplex structure, between the guide strand and the target spanning both the seed region (positions 2–8) and the immediate environment of the cleavage site (opposite to guide strand position 10/11, indicated by the gray star), making these regions sensitive to modification.

target RNA cleavage requires base pairing between the AS and the target, spanning both the seed region (positions 2–8) and the immediate environment of the cleavage site (opposite to guide strand position 10/11) [31], making these regions sensitive to bulky modifications [32]. The fact that only 2–16 nt of the AS form base pairs with the target [32] is consistent with the observation that the 3'-proximal part of the AS is more tolerant to modification. In addition, the AS 3'-overhangs are highly tolerant to chemical modification [33] as long as such modifications do not interfere with anchoring of the AS 3'-overhang in the hydrophobic PAZ domain of Ago2 [33] during RISC activation; the AS 3' overhang is released from the PAZ domain binding pocket upon AS nucleation with the target RNA [32, 34].

56.4

Tools for Chemical Modification of siRNAs

In principle, three classes of chemical modifications have been used: (i) modification of the phosphodiester backbone to primarily enhance siRNA stability or to affect

its biodistribution and cellular uptake; (ii) modifications at the ribose 2'-OH group originally used to enhance siRNA stability, and now widely used to modulate siRNA potency, specificity of silencing and to reduce siRNA immunogenicity (see Figure 56.3 for prominent types of modification); and (iii) modifications of the ribose ring and nucleoside base have been introduced to modulate base-pairing properties in order to enhance siRNA potency. Finally, conjugation of lipophilic and cationic molecules to the termini of siRNAs has been employed primarily to facilitate cellular uptake and to alter the biodistribution of siRNAs *in vivo*.

56.4.1

siRNA Backbone Modifications

The most widely used phosphate backbone alteration is the phosphorothioate (PS) modification where one of the non-bridging phosphate oxygens is replaced with sulfur (Figure 56.3). This modification has been popular in antisense oligonucleotide (AON) designs to improve nuclease resistance and pharmacokinetic properties [37], and was early employed in siRNA design with the hope to achieve similar

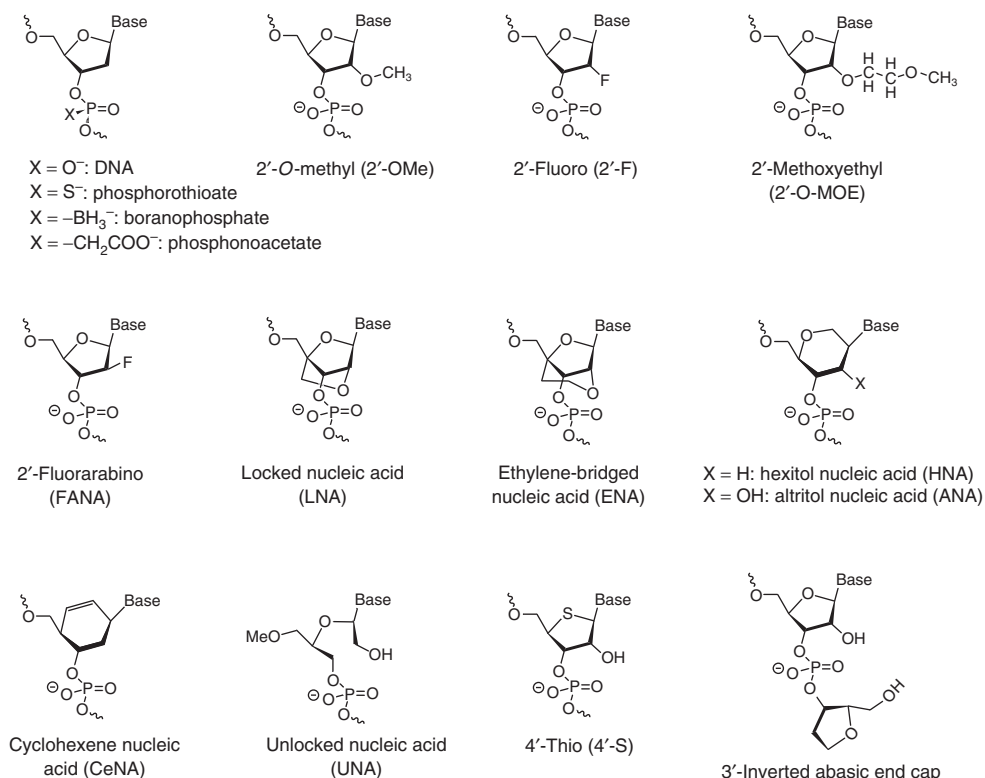


Figure 56.3 Chemical siRNA modifications discussed in the text. In thiophosphonates, the remaining non-bridging oxygen is replaced with sulfur.

effects. The PS modification lowers the melting temperature (T_m) of RNA duplexes ($\sim 0.5^\circ\text{C}$ per modification) [38–40] because PS diesters, in contrast to phosphate diesters, are fully negatively charged at physiological pH [38]. Although moderately PS-modified siRNAs support efficient RNAi [39–44], extensive PS modification reduces silencing [41, 45] and has toxic side-effects [39, 40] (Tables 56.1 and 56.2). Still, moderate PS modification has been rather widely used in combination with ribose 2'-OH modifications in AON and siRNA design (Figure 56.4) to achieve an efficient gene KD *in vivo* in settings suitable for therapeutic interventions [46, 47].

Furthermore, substitutions of the native phosphodiester linkage with a boranophosphate (PB) linkage have been explored in gene silencing (Figure 56.3; Table 56.1) using a canonical siRNA [45] or ss-siRNA design [20]. PB linkages decrease the T_m of RNA duplexes by approximately $0.5\text{--}0.8^\circ\text{C}$ per modification [57], are reportedly more than twice as nuclease resistant as their PS counterparts, less toxic, and also more lipophilic than phosphate linkages owing to an altered charge distribution of the PB [45]. Similarly, the phosphonoacetate (PAC) or thiophosphonoacetate backbone linkages (Figure 56.3) have potentially interesting

Table 56.1 Examples of phosphate backbone, sugar, and base modifications in siRNA design and some of their tested properties.

siRNA modifications	T_m	Nuclease resistance	Potency	GSS
PS	–	+	p/n dep.	+
PB	–	++	+	n.a.
PAC (neutral)	--	+++	n.a.	n.a.
2'-OMe	+	+	Unclear/p.d.	+
2'-F	++	++	p.d.	n.a.
2'-O-MOE	–	+	Unclear/–	n.a.
LNA	+++	++	p.d.	+
ENA	+	+	n.a.	n.a.
DNA (2'-H)	Unclear	+	Unchanged	n.a.
ANA	Unclear	Unchanged	Unchanged	n.a.
HNA	Unclear	+	+	n.a.
CeNA	Unclear	+	+	n.a.
UNA	--	n.a.	+	+
4'-S	+	+++	+	n.a.
2-Thiouracil	+	n.a.	+	+
PS/2'-OMe	n.a.	+	+	n.a.
2'-F/2'-OMe	+	++	++	n.a.
2'-F/2'-H	+	+	+	n.a.
FANA	+	+	+	n.a.

GSS: guide strand selection; + : increased; – : decreased; n.a.: not analyzed; p/n dep.: dependent on position and number of modifications; p.d.: position-dependent; PS: phosphorothioate; PB: boranophosphate; PAC: phosphonoacetate; 2'-OMe: 2'-O-methyl; 2'-O-MOE: 2'-methoxyethyl; LNA: locked nucleic acid; ENA: ethylene-bridged nucleic acid; ANA: altritol nucleic acid; HNA: hexitol nucleic acid; CeNA: cyclohexene nucleic acid; UNA: unlocked nucleic acid; and 4'-S: 4'-thio; FANA: 2'-fluorarabino. For details, see text.

Table 56.2 siRNA modifications with beneficial effects *in vivo*.

Modification	Potency	OTES	GSS	Toxicity	Biostability	Nuclease resistance	Immunogenicity
PS	+	Unclear	+	++	n.a.	+	n.a.
2'-OMe	Unclear	–	+	n.a.	+	+	–
2'-F	+	n.a.	n.a.	n.a.	+	++	–
DNA	Unchanged	–	n.a.	–	+	+	–
LNA	+	–	+	–	+	+++	–
UNA	+	–	+	n.a.	+	n.a.	n.a.

OTES: off-target effects; GSS: guide strand selection; +: increased; –: decreased; n.a. not analyzed. For details, see text.

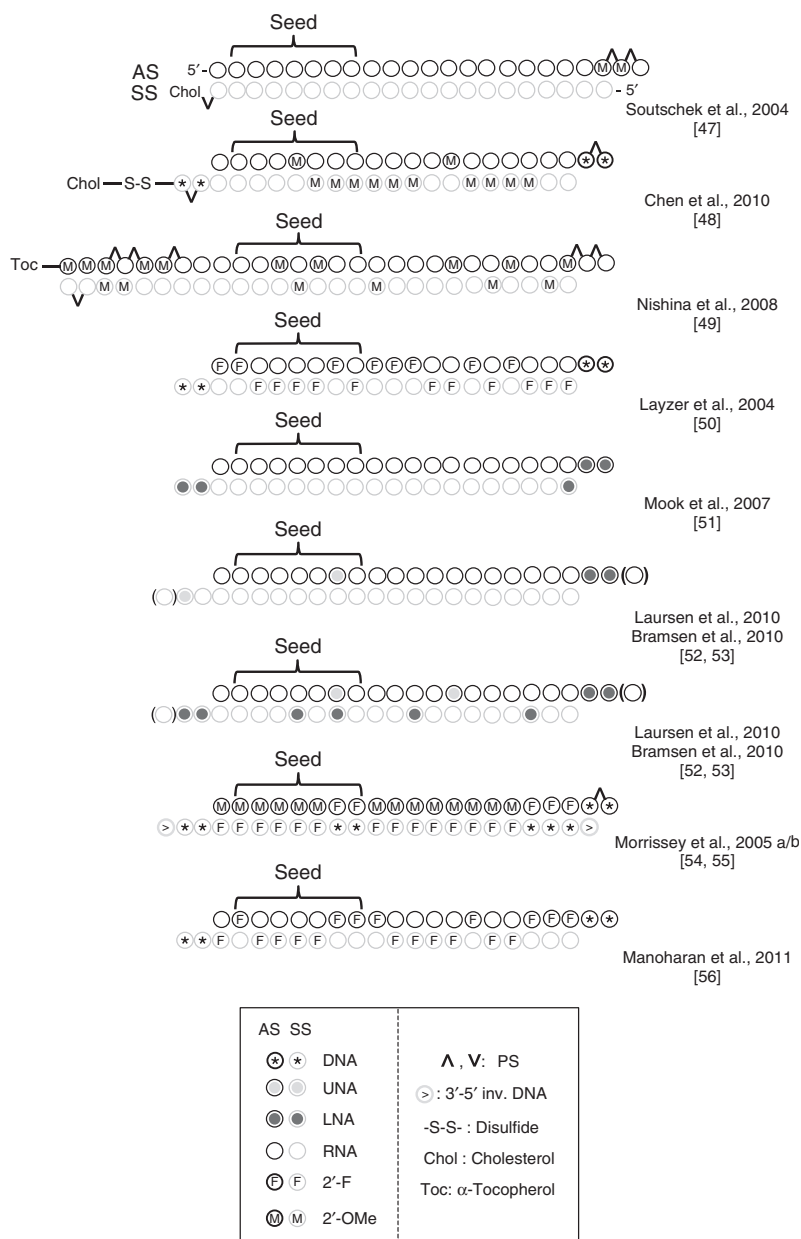
properties in siRNA design, since they are completely resistant to degradation (Table 56.1), and (thio)phosphonacetates esterified with, for example, methyl groups are electrochemically neutral [58], allowing modified oligonucleotides to be taken up efficiently by cells in the absence of delivery reagents [59]. Finally, amide linkages [60] and 2',5'-linkages [61] have been found to enhance nuclease resistance of siRNAs, yet are little used so far.

56.4.2

Ribose 2'-OH Substitutions

The ribose 2'-OH group is not required for siRNA function [41] and has been most extensively modified in siRNA design either by substitution with, for example, 2'-O-methyl (2'-OMe), 2'-fluoro (2'-F), 2'-methoxyethyl (2'-O-MOE), or by locking the 2'-oxygen *via* intramolecular linkages in bridged nucleic acids (BNAs), such as locked nucleic acid (LNA) and ethylene-bridged nucleic acid (ENA) (Figure 56.3). 2'-modifications are widely used to enhance nuclease resistance as RNA hydrolysis by ribonucleases is initiated through activation of the 2'-oxygen as the nucleophile for attack at the vicinal phosphorus atom. Yet, 2'-ribose modifications generally also affect the hybridization properties of the modified oligonucleotide (albeit slightly sequence-dependently) and can therefore be used to position-specifically modulate the thermodynamic stability of siRNAs (Figure 56.2a); small electronegative 2'-substituents, such as 2'-F and 2'-OMe, increase duplex stability by shifting the equilibrium of the sugar moiety toward the RNA-helical C3'-*endo* conformation, whereas larger 2'-O-alkyl substitutions such as 2'-O-MOE can destabilize the siRNA likely by steric hindrance with other functional groups of the duplex [62] (Figure 56.3 and Table 56.1).

The naturally occurring 2'-OMe modification has been among the first and most extensively tested 2'-substitutions [14, 15, 40–42, 44, 63]. For example, a partially 2'-OMe/PS-modified siRNA conjugated with cholesterol was the first to be shown to successfully silence an endogenous gene *in vivo* using a systemic delivery



strategy suitable for therapeutics [47] (Figure 56.4). 2'-OMe modifications increase nuclease resistance, slightly enhance the binding affinity toward RNA (T_m increase of 0.5–0.7 °C per modification) [62, 64], and are well tolerated at most duplex positions. However, extensive or full modification (particularly of the AS) can reduce siRNA potency [1, 15, 41, 43], albeit conflicting results have been reported [42, 64]. 2'-OMe modifications can be successfully combined with other 2'-modifications, for example, 2'-F, to generate fully substituted, nuclease-resistant, yet potent siRNAs [65] (Section 56.5). Interestingly, 2'-OMe modifications may also reduce the immunogenic potential of siRNAs (Section 56.8) and can position-dependently reduce siRNA OTEs (Section 56.9).

Fluorine substitution (2'-F) of the 2'-OH is another popular modification in siRNA design (Tables 56.1 and 56.2). It is among the most thermostabilizing 2'-substitutions (~ 1 °C increase in T_m per modification) [62, 65], is generally very well tolerated, and has proven to be beneficial in a mouse model for silencing of the factor VII gene (Figure 56.4) as directly compared to other 2'-modifications [56]. Moderate 2'-F modification is well tolerated at most positions within both strands of the siRNA duplex [14, 39, 41–43], and substitution of all siRNA pyrimidines was reported to greatly enhance serum stability and to support effective silencing *in vitro* and *in vivo* [50, 54, 66] (Figure 56.4). Alternating modifications of 2'-F and DNA [67] or 2'-OMe [65] have also been used to generate potent siRNAs with increased nuclease resistance (Table 56.1). Changing the stereochemistry of the 2'-F substituent by replacing the ribose unit with arabinose results in 2'-deoxy-2'-fluoroarabinonucleic acids (FANA), which are structurally similar to DNA (C2'-*endo* conformation) and increase duplex stability (~ 0.5 °C increase in T_m per modification). In the guide strand, FANA modification is only well tolerated in the 3'-overhang, but additional full modification of the SS can lead to significant enhancements in potency and nuclease resistance [68]. Moreover, heavily modified siRNAs that contain combinations of FANA with 2'-F-RNA or LNA show superior properties compared to native duplexes or duplexes with only one type of these modifications [69].

From early on, siRNAs have been synthesized with DNA overhangs, typically dTdT, to reduce cost and confer nuclease resistance [1] while having a minimal negative impact on siRNA potency [41, 43]. However, a recent study has suggested that dTdT modification of overhangs can decrease the silencing duration *in vitro* and *in vivo* [70]. Otherwise, the SS is relatively tolerant to 2'-deoxy modifications as complete 2'-deoxy modification of the SS had little effect on siRNA activity [71, 72]. Only partial substitution of the guide strand is tolerated [41, 73]; however, alternating modification with 2'-F has created fully substituted, active guide strands [41]. Notably, DNA substitutions in the guide strand seed region may even reduce OTEs caused by some siRNA sequences [74] (discussed in Section 56.9).

More bulky 2'-modifications such as 2'-O-MOE (Figure 56.3) and 2'-O-allyl were found, as other 2'-modifications, to enhance siRNA nuclease resistance. Yet, bulky 2'-O-alkyl substitutions are only tolerated at certain positions within the siRNA duplex, likely owing to their distortion of RNA helix structure [14, 27, 40, 75], thus complicating their use in siRNA design. As a rule, the most

bulky 2'-modifications are only tolerated at the 3'-ends of siRNA strands [75]. The decrease in thermodynamic stability imposed by large 2'-substituents can, however, be intentionally used to enhance siRNA potency: for example, a 2'-aminoethyl modification had such a beneficial effect when incorporated at position 17 close to the passenger strand's 3'-end, likely by altering the thermodynamic asymmetry of the siRNA duplex [27] (Figure 56.2). In conclusion, the majority of bulky 2'-modifications are not widely used and applications are mainly restricted to siRNA overhangs to enhance exonuclease resistance [40, 65].

56.4.3

Alteration of the Ribose Backbone

A more radical approach to modify 2'-OH groups is the use of BNAs that are conformationally locked and where the 2'-oxygen is connected to the 4'-carbon *via* a methylene bridge as in LNA [76] (Figure 56.3) and carbocyclic-LNA [27, 77], *via* an ethylene bridge as in ENA [78] and carbocyclic-ENA [27, 77], or to the 1'-carbon as in oxetane (OXE) [27, 79]. Among those, LNA has been most extensively used to enhance the performance of both AONs [46, 80, 81] and siRNAs [26, 27, 43, 44, 82]. The intermolecular methylene bridge locks the furanose ring of LNA in a C3'-*endo* RNA-helical conformation, which causes an additive increase in thermal stability of 2–10 °C per LNA monomer upon incorporation into RNA duplexes [83]. This significant increase in thermostability limits the number of LNA residues that are tolerated in siRNA design [43, 44, 82]. However, limited LNA modification makes it possible to modulate the local thermodynamic profile of the siRNA duplex in order to increase nuclease resistance *in vitro* [27, 43] and *in vivo* [51, 84], to reduce the immunogenic properties of certain siRNA sequences [85], to enhance siRNA potency and specificity by suppressing passenger strand selection during RISC loading [26, 27, 82], and has even allowed the development of the novel sisiRNA design [26] (Figure 56.1).

Substituting the ribose with other five- or six-carbon ring moieties has also been applied to siRNA design, such as in ANA (altritol nucleic acid), HNA (hexitol nucleic acid), FANA, and CeNA (cyclohexene nucleic acid) nucleotide analogs, which are characterized by altritol, hexitol, arabinose and cyclohexenyl moieties, respectively [27, 68, 86–89]. ANA modifications can enhance siRNA activity (up to twofold) and silencing duration moderately (up to 50%) when incorporated at both 3'-ends (AS and SS) [88]. Likewise, incorporation of HNA at both 3'-ends enhanced silencing potency, serum stability and silencing duration of an siRNA against B-Raf [87]. Similarly, CeNA modification of SS and AS 3'-ends increased siRNA potency and nuclease resistance [89], and a fully FANA-modified SS in combination with a FANA modification at the 3'-end of the AS was fourfold more potent and had a longer half-life in serum as compared to unmodified siRNA [86] (Table 56.1).

Another substantial alteration of the ribose is found in unlocked nucleic acids (UNAs), which are acyclic derivatives of RNA lacking the C2'–C3' bond of the RNA ribose ring (Figure 56.3) but still structurally mimic unmodified RNA upon incorporation into RNA duplexes. Incorporation of UNA residues introduces

additive instability by 5–8 °C per UNA monomer, which can be exploited to locally destabilize the siRNA duplex [90]. Extensive UNA modification will jeopardize annealing of siRNA strands [52], and UNA is not well tolerated at the 5'-terminal positions of the guide strand [91]. However, limited UNA modification can be strategically used in both siRNA strands to alter siRNA strand selection and to improve the potency of extensively (e.g., LNA-) modified siRNAs that are otherwise too stable or rigid to support RNAi [52]. Notably, siRNAs modified with UNA residues only in their 3'-overhangs have prolonged biostability *in vivo*, even compared to extensively LNA-modified siRNA, and elicit efficient gene KD in contrast to unmodified siRNAs [52]. Furthermore, the incorporation of a single UNA modification at position 7 of the guide strand can efficiently reduce siRNA off-targeting by weakening miRNA-like interactions between the guide strand and potential mRNA targets [53] (Figure 56.4 and Table 56.2). Finally, 4'-thio-modified nucleotides (4'-S RNA) contain a sulfur atom instead of the 4'-carbon of the ribose ring, which has been shown to enhance nuclease resistance, target affinity, and siRNA potency. However, modification of the guide strand is only tolerated at certain positions [92, 93], and sequence-specific effects of 4'-thio modifications have been reported [94]. 4'-Thio modifications were also combined with 2'-substitutions such as 2'-OMe, 2'-O-MOE, 2'-F or FANA [68, 92, 95]. 2'-OMe/4'-S modifications did not affect duplex stability, whereas 2'-F/4'-S and FANA/4'-S modifications increased or decreased the T_m by ~ 1 and approximately 1.0–1.4 °C, respectively. 2'-OMe/4'-S modifications rendered the siRNA more serum-resistant (approximately fourfold compared to 2'-OMe-modified counterparts), whereas 2'-F/4'-S modifications failed to confer nuclease resistance in human plasma [95].

56.4.4

Base Modifications

A number of modified nucleotide bases, such as 5-bromo-, 5-iodo-, 2-thio-, 4-thio-, dihydro- and pseudo-uracil, have been tested in siRNA design, yet are not widely used. Modified bases have been incorporated into siRNAs to stabilize base pairing. 5-Bromo- and 5-iodo-uracil slightly reduced siRNA potency [41], whereas 2-thio- and pseudo-uracil were reported to enhance siRNA potency [96] and to reduce cellular immune responses [97].

56.5

Improving siRNA Potency

Chemical modification may affect siRNA activity by modulating conformational features of the nucleoside sugar pucker and thus helix geometry, by altering the on- and/or off-rates of siRNA-target duplexes, or through effects on product release by RISC. Although chemical modification of siRNAs tends to negatively affect their activity, particularly when introduced into the guide strand, siRNA potency is reliably enhanced by chemical modification strategies that favor guide strand

selection during RISC loading. One such strategy is to alter the thermodynamic asymmetry of the siRNA duplex (Figure 56.2; Tables 56.1 and 56.2), for example, by incorporating stabilizing LNA residues into the 5'-proximal region of the passenger strand [82], by placing 2-thiouracil at the guide strand's 3'-end [96], or by introducing destabilizing modifications, such as OXE, ethylamino, UNA, dihydrouracil, or PS, into the 3'-proximal region of the passenger strand [27, 82, 98]. Similarly, chemical modification of the siRNA 3'-overhangs can modulate strand selection by RISC. Chemically modified overhangs that either favor or disfavor strand selection by RISC can be incorporated into the guide or passenger strands of the siRNA [27]. The PAZ domain of Ago2 (that recognizes the AS 3'-overhang) has been shown to force the nucleotides of the 3'-overhang into a DNA-like sugar pucker [99], suggesting that overhang flexibility may affect RISC incorporation. In addition, 5'-end modifications can modulate strand selection: placement of a single terminal 5'-OMe modification in the SS enhanced incorporation of the unmodified AS, even in the case of thermodynamically unfavorable siRNAs [100]. Interestingly, a few studies reported very significant improvements in siRNA potency despite extensive modification. Allerson *et al.* found an up to 500-fold increase in siRNA potency upon full modification of siRNAs with alternating 2'-OMe and 2'-F modifications [65]. However, such significant improvements appear to be sequence-specific [101], which is not fully understood yet. Nonetheless, 2'-OMe/2'-F-modified siRNAs are preferentially taken up into RISC relative to unmodified siRNA [101].

56.6

Enhancing siRNA Nuclease Resistance

56.6.1

siRNA Stability and Ribonucleases

Single-stranded RNA is highly labile in extracellular compartments due to degradation by ribonucleases [102], explaining why AONs need extensive, most often full, chemical modification to be efficient as therapeutics *in vivo* [103, 104]. Although double-stranded RNAs, such as siRNAs, are more resistant than their single-stranded counterparts [105], unmodified siRNAs are still degraded within minutes in mammalian serum [27, 41, 43, 47, 50, 106]. Thus, chemical modification of siRNAs was initially explored to enhance nuclease resistance. The 3'-exonuclease ERI-1 [107, 108] and RNase A-like, yet unidentified, endonuclease activities [109, 110] have been described as key mediators of siRNA degradation. However, RNase composition and activities, especially the relative activities of exonuclease(s) and pyrimidine-specific endonuclease(s), differ between biological fluids and species [111–115]. As a consequence, the chemical modifications applied to enhance nuclease resistance should be tailored to the particular environment to which the siRNA will be exposed. For example, siRNAs modified by UNA in their 3'-overhangs were quickly degraded in bovine serum, but were highly stable in mouse serum and produced an efficient gene KD in mice [52].

56.6.2

Strategies for siRNA Stabilization

The degradation of RNA by ribonucleases involves activation of the 2'-oxygen for nucleophilic attack on the adjacent intrastrand phosphodiester to yield a 2',3'-cyclic phosphate [116]. Therefore, the initial strategy for increasing siRNA nuclease resistance was chemical modification of internucleotide phosphate linkages or ribose 2'-OH groups (apart from physical separation of siRNA and nucleases by shielding delivery agents, which is not discussed here). As modification of siRNA overhangs is very well tolerated, modest enhancements of serum stability can be "safely" achieved by numerous types of overhang modifications that inhibit attacking 3'-exonucleases [27, 39–44] thought to be responsible for the bulk of nuclease activity in serum [115, 117]. This may explain why even blunt-ended unmodified siRNAs were reported to have slightly increased stability in fetal calf serum (FCS) [15]. Furthermore, siRNA stabilization is achieved by protecting siRNAs against endonuclease attack. Initial chemical modification strategies typically did not strive to identify the specific primary cleavage site, and therefore aimed at maximizing the number of modifications without compromising siRNA potency. Especially, combining moderate PS modification with various 2'-substitutions and end-conjugations has been very successful in creating highly stable and potent siRNAs for applications *in vivo* [47, 49, 118] (Figure 56.4). In addition, 4'-thio-ribose modification greatly increased siRNA stability [92, 93] (Table 56.1). Too extensive thermodynamic stabilization of the siRNA duplex as a consequence of comprehensive modification, for example, by LNA can be compensated for by the simultaneous incorporation of destabilizing modifications, such as UNA (Figure 56.4) [52] or by using the sisiRNA design (Figure 56.1) containing a segmented passenger strand [26].

Recent strategies have focused on identifying vulnerable positions of the siRNA, with the goal to greatly improve siRNA stability while preserving activity through minimizing the modification level [119, 120]. As most dsRNA-specific endoribonucleases are pyrimidine specific (and preferentially recognize UpA, UpG, and CpA dinucleotide motifs [113, 119–121]), such positions should be modified, for example, by 2'-OMe groups, whereas nuclease-insensitive sites should be left unmodified as they are not cleaved by serum endoribonucleases even during extended incubation [119, 120]. Alternatively, vulnerable nucleotide positions may be protected by lowering their accessibility to nucleases through embedment in stable duplex stretches. There is evidence that the thermodynamic stability of dsRNA greatly influences resistance toward endonuclease attack, also taking into account that even double strand-specific endoribonucleases access their dsRNA substrates by preferentially binding to short single-stranded regions that are transiently exposed due to spontaneous thermal fluctuations [122]. In this regard, the 5'-proximal region of the AS may be particularly nuclease-sensitive as it is thermodynamically unstable on purpose to favor AS selection during RISC loading, a problem that can be solved by 2'-OMe modification [123]. Also, slightly increasing siRNA thermostability, for example, by introducing LNA [27, 43, 44, 51, 82] or 4'-thio modifications

[92] at selected positions within the duplex can enhance siRNA stability while generally preserving siRNA potency.

As extensive chemical modification of siRNAs usually reduces their activity [27], it is important to adapt the level of modification to the particular siRNA application (e.g., target species/cells, delivery vehicle, and entry route). In cell culture, siRNAs are typically delivered using commercial transfection reagents that shield the siRNAs from serum RNases and ensure delivery of intact siRNA to the cell cytoplasm. This generally produces efficient silencing for 2–7 days depending on the cell type and rate of cell division [124], which is often sufficient for experiments *in vitro*. For such cell culture experiments, it is still unclear if and how siRNA nuclease resistance translates into enhanced potency or prolonged silencing duration upon delivery to the cell cytoplasm. *In vivo*, however, the advantages of nuclease-stabilized siRNAs primarily originate from higher siRNA bioavailability and more efficient siRNA delivery as well as from improved interaction with the intracellular RNAi machinery [124]. Several studies showed that chemical stabilization is indeed required for the delivery of “naked” siRNA using low-pressure intravenous injection, a strategy relevant to siRNA therapeutics. Dose-dependent gene silencing was reported for a fully modified siRNA consisting of a SS with 2'-F substitutions at all pyrimidine positions, DNA at all purine positions, and 5'- and 3'-inverted abasic end caps, and an AS with 2'-F substitutions at all pyrimidine positions, 2'-OMe substitutions of all purines, and a single PS modification at the 3'-terminal linkage [55] (Figure 56.4). A PS/2'-OMe-stabilized, cholesterol-conjugated siRNA (Figure 56.4) was reported to efficiently silence target gene expression in mouse livers, although an unmodified siRNA was not included in the study for comparison [47].

56.7

Enhancing siRNA Silencing Duration

While most chemical modification studies initially focused on maximizing siRNA potency *in vitro*, enhancing silencing duration is of great interest for therapeutic interventions. It remains to be fully established which factors determine siRNA silencing persistence. It is known that the siRNA sequence, although unclear how, can affect silencing duration [70] and that chemical modification can decrease/enhance silencing duration both *in vitro* [5, 41, 86, 87, 89, 119] and *in vivo* [70]. One line of studies has suggested that enhanced siRNA nuclease resistance leads to prolonged silencing duration, and silencing duration correlates with the levels of the exonuclease ERI-1 [107, 108]. Of note, silencing duration by unmodified siRNA is not improved by elevating the siRNA concentration above saturation levels [125], suggesting that siRNAs are rapidly degraded in the cells unless incorporated into siRNA pathways. SiRNAs modified with FANA, HNA or CeNA displayed enhanced silencing duration, an effect that was attributed to increased nuclease stability [86–89]. Similarly, 2'-F substitution of all U and C residues in the siRNA duplex enhanced its stability and significantly prolonged RNAi-mediated gene silencing [41]. 2'-OMe modification exclusively at nuclease-sensitive sites was reported to expand silencing duration, although at the cost of a marginal reduction of siRNA

potency [119]. In contrast, recent studies suggest that nuclease stability is primarily important to prevent siRNA degradation in extracellular environments, such as the bloodstream after systemic administration, yet is not directly influencing silencing duration upon siRNA internalization into the cell cytosol [124]. Likewise, siRNA stability had no effect on silencing duration after cationic lipid transfection in cell culture, or in mice after high-pressure hydrodynamic intravenous injection (HDI) of naked siRNAs [50, 124]. Moreover, silencing by unmodified siRNAs can last for several weeks in terminally differentiated non-dividing cells, such as macrophages [126], and up to one month in slowly dividing CCD-1074Sk cells [124]. Reports on long-lasting target inhibition in the livers of mice and non-human primates [118] suggest that dilution of intracellular siRNAs by cell division is the major factor limiting silencing duration, at least in dividing cells. Double-stranded siRNA, in contrast to single-stranded RNA, has also been found to be stable inside cells upon delivery by direct injection, suggesting that the siRNA AS does not need to be incorporated into RISC to survive intracellularly [127], hereby contrasting the proposed intracellular degradation reported by others [125]. Collectively, these results imply that elevated nuclease resistance is important during the delivery process of naked siRNA rather than affecting siRNA silencing duration. Nevertheless, in some cases, siRNA potency may include contributions from how the RNAi machinery handles an siRNA and thus affects silencing duration. For example, the more potent 27 bp siRNA design has been shown to extend silencing duration [5], and an ANA-modified siRNA with enhanced potency but similar serum stability as unmodified siRNA resulted in a prolonged silencing duration [88]. Furthermore, a recent study suggested that the siRNA sequence and guide strand 3'-overhang are determinants of silencing duration *in vitro* and *in vivo*; DNA residues in the guide strand overhang reduced silencing duration as compared to RNA and 2'-OMe [70]. Notably, different siRNA sequences were observed to result in different silencing durations [70] and affinities for RISC [101, 128]. Thus, silencing duration may well reflect the ability of a particular siRNA to enter RNAi pathways and to compete with endogenous miRNAs for the limiting amounts of RISC.

56.8

siRNA Immunogenicity

56.8.1

Cellular Response to siRNA

The innate immune system relies on a range of specialized pattern-recognition receptors (PRRs) [129]. Although it was initially anticipated that shorter exogenous siRNAs, mimicking endogenous RNA species, were non-immunogenic [1], subsequent studies have demonstrated that siRNAs do indeed trigger innate immune responses *in vivo* and in various cell types such as peripheral blood mononuclear cells (PBMCs) *in vitro*. In the cytoplasm of all cells, dsRNA is sequence-independently recognized by a dsRNA-responsive kinase (PKR) [130], as well as by the helicases retinoic acid inducible gene (RIG-I) [131, 132] and melanoma

differentiation-associated gene 5 (Mda5) [133]. Recent studies observed low levels of PKR activation by canonical (~19 bp) siRNAs [132, 134]. RIG-I in particular senses the nature of dsRNA ends and is activated by 5'-terminal triphosphates and blunt-ended dsRNA, including blunt-ended 27 bp DsiRNAs, whereas the 2 nt 3'-overhangs of standard ~19 bp siRNAs are not recognized [97, 132, 135]. The most prominent responses toward siRNAs, however, are mediated by the Toll-like receptors (TLRs) located in endosomes or exposed on the surface of cells primarily of the immune system; siRNAs are mainly detected by TLR3 [136], TLR7 [85, 137, 138], and TLR8 [137]. Human TLR3 primarily recognizes foreign dsRNAs including exogenous siRNAs in a sequence-unspecific manner, triggering the production of, for example, IFN and Interleukin 12 [139–142]. When tested in cell culture, canonical 19 bp (21mer) siRNAs appeared to be non-immunogenic, whereas longer 23 bp DsiRNAs (or longer species) were able to induce a concentration- and cell line-dependent, presumably TLR3-mediated IFN response [9]. However, all 19 bp siRNAs were found to activate TLR3 upon intraocular injection in mice regardless of siRNA sequence or modification, whereas siRNAs shorter than 19 bp (21 nt) did not activate TLR3 [140, 143]. TLR7 and 8 recognize specific sequence motifs within siRNAs, and particularly U- or G/U-rich ssRNAs can activate human TLR7 and TLR8 [85, 144–148].

The compartmentalization of the PRRs also implies that the host immunological responses toward siRNAs are very much dependent on the cellular entry route and hence delivery reagent. Delivery vehicles such as cationic lipids, polymers, and nanoparticles will expose siRNAs to potential TLR7/8 recognition in endosomes. In contrast, naked siRNAs usually do not activate TLR7/8 [85, 147, 149], but rather stimulate the surface-bound TLR3 [140]. Therefore, the dependence of immunogenicity on siRNA sequence, structure, and delivery mode calls for pre-evaluation in a relevant experimental setup using primary immune cells.

56.8.2

Chemical Modification Can Abrogate siRNA Immunogenicity

Until now, most efforts aiming at reducing siRNA immunogenicity have focused on abrogating the induction of TLR7/8 responses through modifying immunogenic sequence motifs within either siRNA strand. Examples include fully 2'-OMe/2'-F-modified siRNAs [54] or LNA modification at selected positions [85]. Moreover, it has been shown that a number of base modifications including 5-methylcytidine (m5C), 5-methyluracil (m5U), N6-methyladenosine (m6A), 2-thiouridine (s2U), or pseudouridine reduce TLR7/8 immunostimulation [150]. Notably, even modification of uridine residues only, either with 2'-F or 2'-OMe, abrogated any immunogenicity upon delivery to PBMCs *via* a cationic lipid [151]. Similarly, replacement of a few selected uridine or guanosine residues lowered siRNA immunogenicity significantly upon liposomal delivery in human PBMCs and mice [152]. Likewise, it was recently proposed that alternating 2'-OMe modification of the SS (every second residue modified) may be a universal approach to avoid TLR7 activation by siRNAs [153]. Interestingly, naturally occurring

2'-OMe-modified RNAs, such as endogenous ribosomal RNA, evade TLR7 activation [154], suggesting that 2'-OMe modification of endogenous RNAs allows immune receptors to distinguish pathogen-derived (unmodified) RNA from host cell RNA [150]. Immunogenicity of siRNAs may also be reduced by enhancing thermostability of siRNA duplexes, for example, by moderate LNA modification [155]. Mechanistically, siRNA duplex stabilization may render single-stranded immunogenic sequence motifs inaccessible to TLR7/8 recognition. The relevant observation of sequence-independent TLR3 activation by 21 bp siRNAs *in vivo* has so far not been successfully addressed by chemical modification strategies; at least 2'-OMe modification seems to be ineffective in preventing TLR3 activation. To mitigate potential immunogenic effects, researchers may preferentially choose the shorter 19 bp siRNA design albeit this is typically slightly less potent [140, 143]. The low RIG-I activity seen in the presence of standard (~19 bp) siRNAs with 2 nt 3'-overhangs has left this pathway less studied. However, blunt-ended DsiRNAs may be modified by 2'-OMe [6, 156] or DNA [132] to reduce RIG-I activation.

56.9

Reducing siRNA Off-Target Effects by Chemical Modification

56.9.1

Off-Target Effects Caused by miRNA-Like Activities

The RISC/Ago2-mediated cleavage of target RNAs by a matching siRNA is highly sequence specific allowing only few mismatches [157, 158]. However, the inherent miRNA-like behavior of all siRNAs may trigger unintended repression of hundreds of endogenous genes [159–162]. Beyond resulting in toxicity [163], this may lead to misinterpretation of an RNAi experiment. The resemblance to the miRNA interference pathway predicts that siRNA off-targeting is primarily mediated by interaction of the seed region of the RISC-associated guide strand (positions 2–8 counting from the 5'-end) with complementary sequences in mRNA 3'-UTRs [160, 164]. This may cause unintended translational inhibition and mRNA destabilization [165–167]. Significant OTEs have been noticed in large-scale siRNA library screens, and occasionally a more prominent phenotype is elicited by cumulated OTEs than by the intended siRNA target cleavage itself [164].

56.9.2

Reducing Off-Targeting by Chemical Modification of the siRNA Guide Strand Seed Region

The initial interactions between miRNA guide strand and target RNA are mediated by the miRNA seed region exposed by RISC [28]; mismatches in this region have a severe effect on the miRNA interference effect [74, 160]. In contrast, a fully

matching siRNA can usually accommodate a single mismatch at selected seed positions without a substantial loss of silencing efficacy. Accordingly, position-specific modifications in the seed region of the guide strand have been shown to reduce the off-targeting potential of siRNAs. Jackson and coworkers demonstrated that 2'-OMe modification at position 2 of the guide strand can site-specifically reduce off-targeting [63]. In another study, the seed–target interaction was modulated by substituting positions 1–8 with DNA [74]. Recently, we showed that incorporation of the strongly destabilizing UNA modification at position 7 very potently eliminated OTEs, with little negative impact on siRNA efficiency [53] (Figure 56.4). Although such strategies have the potential to solve toxicity issues associated with siRNAs *in vivo*, it remains to be fully explored how applicable these modification principles are to different siRNA designs.

56.9.3

Avoiding Passenger Strand Off-Targeting

Owing to the symmetry of the siRNA molecule, special action should be taken to make sure that the anticipated guide strand is incorporated into RISC. A number of strategies have been reported that abrogate potential silencing caused by passenger strand utilization. Usually, asymmetric thermostability is applied to the two siRNA duplex ends to ensure selective unwinding of the siRNA from the 5'-end of the guide strand (Figure 56.2), although partial incorporation of the passenger strand is still seen in most cases [26]. To further disfavor passenger strand incorporation, LNA modification in its 5'-proximal duplex region [82] or, for example, OXE, ethylamino, UNA, dihydrouracil or PS modifications at its 3'-end may be considered [27, 82, 98]. Moreover, a single UNA modification at the 5'-terminus of the passenger strand can block passenger strand utilization [168]. Another possibility to prevent passenger strand loading into RISC is the use of aiRNAs or asiRNAs because these designs involve a shortened passenger strand that cannot be effectively loaded into RISC [13, 22, 23]. Even more effective is the sisiRNA design (Figure 56.1) that precludes passenger strand incorporation by pre-cleaving it [26]. Alternatively, the 5'-phosphate of the passenger strand, required for strand function [29], may be chemically blocked, for example, by 5'-OMe modification [100].

56.10

Chemical Modifications Can Improve siRNA Pharmacokinetics

siRNAs have some inherent drawbacks for therapeutic applications: they are relatively big, labile in biological fluids, and the negatively charged phosphate backbone prevents them from penetrating cellular membranes by diffusion. Even phagocytic macrophages and dendritic cells do not immediately internalize siRNAs [126, 169]. Only inefficient cellular uptake has been observed in cell culture using high nucleic acids concentrations [170–173].

56.10.1

Enhancing Cellular Delivery by siRNA Conjugation

To facilitate siRNA delivery across the plasma membrane, a large number of delivery vehicles have been developed, including cationic lipids (such as lipofectamine, RNAifect, oligofectamine, DOTAP, and *Trans* IT-TKO), cationic polymers (such as polyethylenimine (PEI), chitosan, and cyclodextrin) and dendrimers, or combinations thereof. All these compounds electrostatically interact with anionic siRNAs, neutralize their negative charge, and thereby permit docking to negatively charged cell membranes to promote cellular uptake *via* adsorptive endocytosis. The application of these transfection vehicles interacting non-covalently with siRNAs has been reviewed in detail elsewhere [174]. An alternative route is to chemically modify the siRNA itself to enhance cellular uptake. For this purpose, the 3'-end of the guide strand and both ends of the passenger strand are available for modifications without jeopardizing the KD effect significantly. Reported transport-facilitating ligands include cationic cell-penetrating molecules or lipids typically attached *via* acid-labile or reducible linkages, such as thio-linkages. Cell-penetrating peptides (CPPs) in particular, such as penetratin [174–176], transportan [175], oligoarginine [177], and peptides derived from the HIV TAT protein [176, 178], have been reported to facilitate siRNA uptake by a mechanisms that is not fully understood [179]. Also, translating *in vitro* delivery systems into *in vivo* applications often causes problems. For example, TAT peptide conjugations were found to mediate a significant increase in cellular siRNA uptake and function *in vitro* [178], whereas no improvement in efficiency was seen *in vivo* for the same construct upon intratracheal delivery in mice [176]. siRNA conjugated to penetratin (derived from the antennapedia protein) or transportin (a fusion peptide derived from the neuropeptide galanin and mastoparan, a peptide toxin from wasp venom) has also been reported to confer a robust KD in a variety of cell lines [175]. However, penetratin activated the innate immune responses in mice upon intratracheal delivery [176]. Liver targeting resulting in 60% silencing of target *apoB* mRNA in mice has been accomplished *via* intravenous injection of chemically stabilized siRNAs conjugated with cholesterol at the 3'-end of the passenger strand [47]. In a similar study, siRNA conjugated with alfa-tocopherol at the 5'-end of the guide strand successfully reduced ApoB protein levels in mouse liver upon intravenous injection [49] (Figure 56.4). Concatamerization of siRNAs is an interesting possibility to enhance particle stability in conjunction with commercial transfection reagents. Strategies to implement this include the use of short complementary “sticky” overhangs [180] or tethering siRNA molecules to each other *via* reducible disulfide bridges [181].

Targeted delivery of siRNAs to specific tissues or cell types is desirable to minimize side-effects and to lower the required amount of siRNA. Cell-targeting ligands such as glycosylated molecules, lipids, peptides, antibodies, hormones, vitamins, and aptamers have been conjugated to various carrier systems (reviewed in [182]). Direct conjugation of siRNAs to cell-targeting ligands such as peptides, antibodies, aptamers, micelles, and nanoparticles has also been reported [183–188] and is detailed elsewhere [189].

56.10.2

Altering Biodistribution by siRNA Conjugation

Systemic delivery of non-formulated siRNAs faces the problem that oligonucleotides of this size are very rapidly cleared from the bloodstream by glomerular filtration and kidney excretion [47, 190], if not degraded by serum RNases [27, 41, 43, 47, 106]. The usual strategy is to increase siRNA size by formulation [54, 185, 188, 191–196] or by introducing chemically modified nucleotides that are known to interact with serum molecules, such as PSs [197]. Various chemical siRNA end-conjugations can improve siRNA pharmacokinetics as well. For example, cholesterol conjugation has been shown to increase siRNA half-life considerably *in vivo*, with detectable levels of siRNAs in liver, heart, lung, kidney and fat tissue after 24 h [47]. Conjugation to bile acids and various long-chain fatty acids also influenced siRNA tissue distribution presumably because these components mediate binding to various lipoproteins, lipoprotein receptors, and transmembrane proteins [198]. The pharmacokinetic behavior of siRNAs is naturally less of a concern if the siRNA is delivered locally rather than systemically to the target cells. Indeed, the first clinical trials relied on intraocular injection of siRNAs to treat macular degeneration and on pulmonary administration of siRNAs to cure respiratory syncytial virus (RSV) infections [199].

56.11

Chemical Modification of siRNAs – State of the Art

There is a growing list of promising nucleotide modifications to improve siRNA properties for *in vitro* and *in vivo* experiments. Modifications of the phosphate backbone, substitution of the ribose 2'-OH group, or more extensive changes of the sugar backbone are most widely used in siRNA design, whereas base modifications are less common. As listed in Table 56.1, several of these modifications confer very interesting properties like extremely high nuclease resistance (PAC or 4'-S). However their *in vivo* performance has not been investigated so far. Other modifications seem to increase siRNA potency (PB, HNA, CeNA, 4'-S, 2-thiouracil), again their *in vivo* properties have remained uncharacterized so far. Table 56.1 provides a current snapshot of the effects of the most commonly used nucleotide modifications, mainly inferred from *in vitro* experiments. The *in vivo* applicability of most of these promising modifications awaits further evaluation in future experiments.

56.12

A Guide for *In vivo* Studies

The variation of siRNA duplex architecture (Figure 56.1) and the growing number of chemical modifications and their combinations make it difficult to select the

most favorable design for *in vivo* experiments, taking into account that there is by far no gold standard of siRNA design. The scheme of Figure 56.5, detailed below, is thought to give the reader a guideline for the planning of RNAi experiments *in vivo*. We also refer the interested reader to Table 56.3, which guides to relevant original literature, sorted according to the functional parameter affected by the siRNA modification(s).

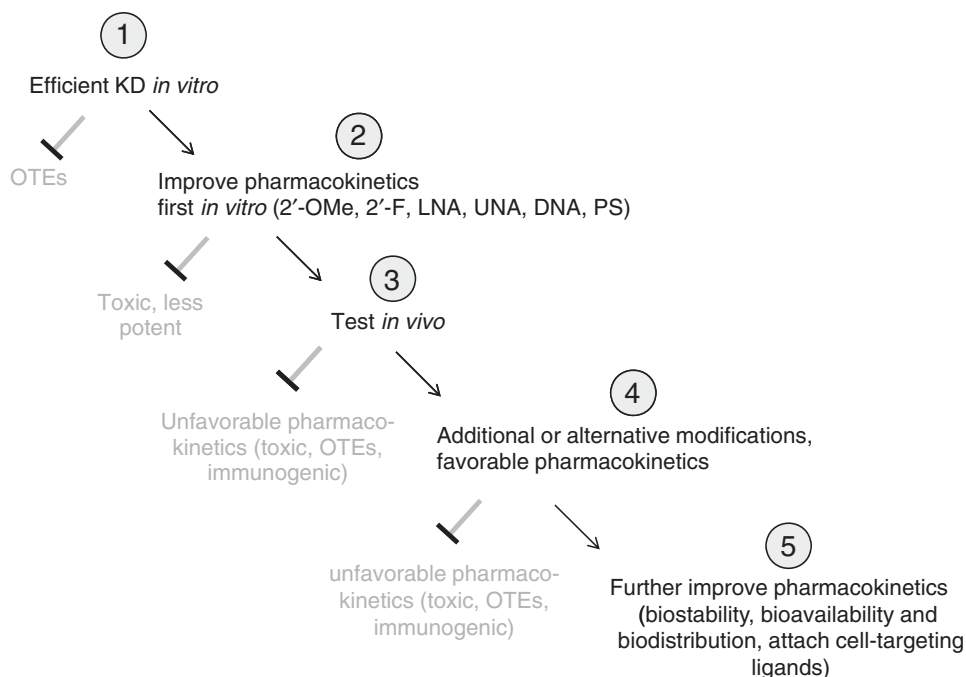


Figure 56.5 Decision guideline for *in vivo* applications of siRNAs. (1) Identify highly efficient siRNAs with little off-target effects (OTEs) in cell culture experiments; possibly test different architectural siRNA designs (Figure 56.1). Decide if chemical modification is required at all for the *in vivo* experiment; modifications that enhance nuclease resistance may be unnecessary when siRNAs are delivered as part of nanoparticles. (2) If chemical modification is anticipated to be crucial for *in vivo* performance, test *in vitro* siRNAs with modifications that confer favorable pharmacokinetic properties; for example, increased nuclease resistance and potency without compromising knockdown (KD) efficiency, low toxicity, and low OTEs (e.g., LNA and 2'-F modifications); potential immunogenicity can be reduced by 2'-OMe

modification at selected uridines and by choosing a <19 bp (21 nt) siRNA design. (3) If the *in vitro* results are promising, proceed to *in vivo*. If the siRNA is efficient *in vivo*, check for unwanted OTEs, toxicity profile and immunogenicity. (4) If unsatisfactory, go back to steps 2 and 3, reduce unwanted OTEs, toxicity, and immunogenicity, either by introducing additional modifications or by testing alternative modification patterns; the goal here is to increase guide strand selection, to minimize miRNA-like effects, to minimize toxicity, and to avoid adverse immunogenic effects. (5) Further improve pharmacokinetic properties, such as biostability, bioavailability, and biodistribution. Optional: improve cellular uptake, for example, by tethering the siRNA to a cholesterol- or tocopherol moiety.

Table 56.3 Examples of chemical modification strategies employed to enhance siRNA performance.

Enhancing siRNA potency	References
Favor AS selection by AS and SS 3'-overhang modification	[27]
Favor AS selection by stabilizing the duplex at the 5'-end of SS	[82, 96]
Favor AS selection by destabilizing the duplex near the 3'-end of SS	[27, 82, 98]
Favor AS selection by modifying the 5'-terminal phosphate of SS	[100]
Use of sisiRNA design to enhance the potency of extensively modified siRNAs	[26]
Limited modification at the 3'-ends of siRNA strands can lead to modest improvements in siRNA potency	[27, 68, 69, 86–89]
Depending on the sequence, full 2'-OMe/2'-F modification may lead to dramatic enhancements in siRNA activity	[65]
Enhancing siRNA stability	References
Modification of siRNA overhangs can enhance exonuclease resistance and may provide sufficient stability for some siRNA applications	[27, 39–44]
Modification of nuclease-sensitive sites within the siRNA duplex significantly enhances stability with minor impact on siRNA activity	[119, 120, 123]
Combining siRNA backbone, ribose ring, and 2'-modifications can significantly enhance nuclease resistance with minor loss of siRNA activity <i>in vivo</i>	[47, 52, 118, 49]
Full siRNA modification dramatically enhances stability yet usually reduces silencing activity	[27, 65, 54, 67]
Enhancing siRNA thermodynamic stability, for example, by moderate LNA modification can increase siRNA stability in biological fluids	[27, 43, 82, 51, 92]
Maximizing silencing duration	References
DNA modification of siRNA overhangs may reduce silencing duration	[70]
Certain Dicer-substrate siRNA designs may enhance silencing duration	[5]
Modification of nuclease-sensitive sites may enhance silencing duration	[41, 119]
Reducing siRNA immunogenicity	References
Global U and G modification, typically within the SS, can reduce siRNA immunogenicity	[151–153]
Selective modification of immunogenic sequence motifs reduces siRNA immunogenicity	[85]
Full siRNA modification abrogates siRNA immunogenicity	[55]
Enhancing siRNA thermodynamic stability, for example, by moderate LNA modification can reduce siRNA immunogenicity	[155]

(continued overleaf)

Table 56.3 (Continued)

Reducing siRNA off-target effects	References
Using siRNA pools can minimize off-targeting	[200, 201]
Eliminate SS selection by RISC through SS 5'-end modification	[63, 100, 168]
Eliminate SS selection by using a segmented strand design (sisiRNA)	[26]
Modify AS by UNA at position 7	[53]
Modify AS by 2'-OMe at position 2	[63]
Modify AS by DNA at positions 1–8	[74]

Refer to main text for further details.

- 1) The first question to be addressed is: What is the optimal design and sequence for my siRNA?
 - a. Find a highly accessible sequence region in the gene of interest with the help of public siRNA prediction tools (e.g., Dharmacon, Qiagen).
 - b. Test several classically designed siRNAs in cell culture and select at least two efficient siRNAs for further studies.
 - c. Optional: Test and compare efficient siRNAs in different designs (Figure 56.1).
- 2) Assess potential OTEs, for example, by using reporter assays to test for miRNA-like or passenger strand-related OTEs [53].
- 3) Think about positions to modify in the siRNA to reduce OTEs and to improve the pharmacokinetic properties:
 - a. Avoid modifying the 5'-phosphate of the AS.
 - b. The 5'-proximal part and central positions of the AS are sensitive especially to multiple or bulky modifications, which lead to a loss in siRNA potency.
 - c. The 3'-end of the AS and the entire SS are well suited for modification.
 - d. Bulky modifications in the AS strand should be positioned after nucleotide 16.
- 4) Modulate sequence properties:
 - a. Favor AS strand selection for the sake of reduced OTEs, for example, by incorporating 2'-OMe modifications.
 - b. Modulate the asymmetric duplex stability by LNA/UNA modifications.
 - c. Minimize unwanted miRNA-like effects of siRNAs by modifying position 7 of the AS with UNA.
 - d. Modification patterns we consider as promising are illustrated in Figure 56.4.
- 5) Check for toxicity of the modified siRNA, as inferred from marker enzyme activities in cell culture.
 - a. 2'-F and 2'-OMe modifications are generally well tolerated.
 - b. DNA and LNA are not toxic.
- 6) What is the pharmacokinetic profile of the siRNA?

- a. Beyond nucleotide modifications, consider to enhance cellular uptake of the siRNA by conjugation with, for example, CPPs, cholesterol, α -tocopherol, micelles, or by nanoparticle formulation (e.g., PEG-PEI).
 - b. Consider to improve cell-specific siRNA uptake by tethering to aptamers, antibodies, or peptides recognized by cell-specific receptors, either directly or as part of nanoparticles.
 - c. Slow down blood clearance and increase siRNA half-life (conjugation to cholesterol, bile acids, or various long-chain fatty acids).
- 7) Reduce immunogenicity of siRNAs
- a. Base modifications (m5C, m5U, m6A, s2U, or pseudouridine) reduce TLR7/8 immunostimulation.
 - b. 2'-OMe modification of the SS can be a universal approach to avoid TLR7 activation.
 - c. Enhancing thermostability of siRNA duplexes, for example, by moderate LNA modification, reduces immunogenicity of siRNAs.
 - d. The use of ≤ 19 bp siRNAs minimizes the potential to activate TLR3.

References

1. Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, **411**, 494–498.
2. Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001) Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.*, **20**, 6877–6888.
3. Kim, D.H., Behlke, M.A., Rose, S.D., Chang, M.S., Choi, S., and Rossi, J.J. (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat. Biotechnol.*, **23**, 222–226.
4. Amarzguioui, M., Lundberg, P., Cantin, E., Hagstrom, J., Behlke, M.A., and Rossi, J.J. (2006) Rational design and *in vitro* and *in vivo* delivery of Dicer substrate siRNA. *Nat. Protoc.*, **1**, 508–517.
5. Hefner, E., Clark, K., Whitman, C., Behlke, M.A., Rose, S.D., Peek, A.S., and Rubio, T. (2008) Increased potency and longevity of gene silencing using validated Dicer substrates. *J. Biomol. Technol.*, **19**, 231–237.
6. Collingwood, M.A., Rose, S.D., Huang, L., Hillier, C., Amarzguioui, M., Wiiger, M.T., Soifer, H.S., Rossi, J.J., and Behlke, M.A. (2008) Chemical modification patterns compatible with high potency dicer-substrate small interfering RNAs. *Oligonucleotides*, **18**, 187–200.
7. Siolas, D., Lerner, C., Burchard, J., Ge, W., Linsley, P.S., Paddison, P.J., Hannon, G.J., and Cleary, M.A. (2005) Synthetic shRNAs as potent RNAi triggers. *Nat. Biotechnol.*, **23**, 227–231.
8. Tanudji, M., Machalek, D., Arndt, G.M., and Rivory, L. (2010) Competition between siRNA duplexes: impact of RNA-induced silencing complex loading efficiency and comparison between conventional-21 bp and Dicer-substrate siRNAs. *Oligonucleotides*, **20**, 27–32.
9. Reynolds, A., Anderson, E.M., Vermeulen, A., Fedorov, Y., Robinson, K., Leake, D., Karpilow, J., Marshall, W.S., and Khvorova, A. (2006) Induction of the interferon response by siRNA is cell type- and duplex length-dependent. *RNA*, **12**, 988–993.
10. Rose, S.D., Kim, D.H., Amarzguioui, M., Heidel, J.D., Collingwood, M.A.,

- Davis, M.E., Rossi, J.J., and Behlke, M.A. (2005) Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic Acids Res.*, **33**, 4140–4156.
11. Ge, Q., Ilves, H., Dallas, A., Kumar, P., Shorenstein, J., Kazakov, S.A., and Johnston, B.H. (2010) Minimal-length short hairpin RNAs: The relationship of structure and RNAi activity. *RNA*, **16**, 106–117.
12. Ge, Q., Dallas, A., Ilves, H., Shorenstein, J., Behlke, M.A., and Johnston, B.H. (2009) Effects of chemical modification on the potency, serum stability, and immunostimulatory properties of short shRNAs. *RNA*, **16**, 118–130.
13. Chu, C.Y. and Rana, T.M. (2008) Potent RNAi by short RNA triggers. *RNA*, **14**, 1714–1719.
14. Prakash, T.P., Allerson, C.R., Dande, P., Vickers, T.A., Sioufi, N., Jarres, R., Baker, B.F., Swayze, E.E., Griffey, R.H., and Bhat, B. (2005) Positional effect of chemical modifications on short interference RNA activity in mammalian cells. *J. Med. Chem.*, **48**, 4247–4253.
15. Czauderna, F., Fechtner, M., Dames, S., Aygün, H., Klippel, A., Pronk, G.J., Giese, K., and Kaufmann, J. (2003) Structural variations and stabilising modifications of synthetic siRNAs in mammalian cells. *Nucleic Acids Res.*, **31**, 2705–2716.
16. Hohjoh, H. (2002) RNA interference (RNAi) induction with various types of synthetic oligonucleotide duplexes in cultured human cells. *FEBS Lett.*, **521**, 195–199.
17. Hohjoh, H. (2004) Enhancement of RNAi activity by improved siRNA duplexes. *FEBS Lett.*, **557**, 193–198.
18. Holen, T., Amarzguioui, M., Babaie, E., and Prydz, H. (2003) Similar behaviour of single-strand and double-strand siRNAs suggests they act through a common RNAi pathway. *Nucleic Acids Res.*, **31**, 2401–2407.
19. Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R., and Tuschl, T. (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell*, **110**, 563–574.
20. Hall, A.H., Wan, J., Spesock, A., Sergueeva, Z., Shaw, B.R., and Alexander, K.A. (2006) High potency silencing by single-stranded boranophosphate siRNA. *Nucleic Acids Res.*, **34**, 2773–2781.
21. Abe, N., Abe, H., and Ito, Y. (2007) Dumbbell-shaped nanocircular RNAs for RNA interference. *J. Am. Chem. Soc.*, **129**, 15108–15109.
22. Sun, X., Rogoff, H.A., and Li, C.J. (2008) Asymmetric RNA duplexes mediate RNA interference in mammalian cells. *Nat. Biotechnol.*, **26**, 1379–1382.
23. Chang, C.I., Yoo, J.W., Hong, S.W., Lee, S.E., Kang, H.S., Sun, X., Rogoff, H.A., Ban, C., Kim, S., Li, C.J. *et al.* (2009) Asymmetric shorter-duplex siRNA structures trigger efficient gene silencing with reduced nonspecific effects. *Mol. Ther.*, **17**, 725–732.
24. Dua, P., Yoo, J.W., Kim, S., and Lee, D.K. (2011) Modified siRNA structure with a single nucleotide bulge overcomes conventional siRNA-mediated off-target silencing. *Mol. Ther.*, **19**, 1676–1687.
25. Sano, M., Sierant, M., Miyagishi, M., Nakanishi, M., Takagi, Y., and Sutou, S. (2008) Effect of asymmetric terminal structures of short RNA duplexes on the RNA interference activity and strand selection. *Nucleic Acids Res.*, **36**, 5812–5821.
26. Bramsen, J.B., Laursen, M.B., Damgaard, C.K., Lena, S.W., Babu, B.R., Wengel, J., and Kjems, J. (2007) Improved silencing properties using small internally segmented interfering RNAs. *Nucleic Acids Res.*, **35**, 5886–5897.
27. Bramsen, J.B., Laursen, M.B., Nielsen, A.F., Hansen, T.B., Bus, C., Langkjær, N., Babu, B.R., Højland, T., Abramov, M., Van Aerschot, A. *et al.* (2009) A large-scale chemical modification screen identifies design rules to generate siRNAs with high activity, high stability and low toxicity. *Nucleic Acids Res.*, **37**, 2867–2881.
28. Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D.J. (2008) Structure of the guide-strand-containing

- argonaute silencing complex. *Nature*, **456**, 209–213.
29. Nykanen, A., Haley, B., and Zamore, P.D. (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*, **107**, 309–321.
 30. Lima, W.F., Wu, H., Nichols, J.G., Sun, H., Murray, H.M., and Crooke, S.T. (2009) Binding and cleavage specificities of human Argonaute2. *J. Biol. Chem.*, **284**, 26017–26028.
 31. Meister, G. and Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature*, **431**, 343–349.
 32. Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature*, **461**, 754–761.
 33. Ma, J.B., Ye, K., and Patel, D.J. (2004) Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, **429**, 318–322.
 34. Tomari, Y. and Zamore, P.D. (2005) Perspective: machines for RNAi. *Genes Dev.*, **19**, 517–529.
 35. Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209–216.
 36. Walton, S.P., Wu, M., Gredell, J.A., and Chan, C. (2010) Designing highly active siRNAs for therapeutic applications. *FEBS J.*, **277**, 4806–4813.
 37. Aboul-Fadl, T. (2005) Antisense oligonucleotides: the state of the art. *Curr. Med. Chem.*, **12**, 2193–2214.
 38. Eckstein, F. (2002) Developments in RNA chemistry, a personal view. *Biochimie*, **84**, 841–848.
 39. Harborth, J., Elbashir, S.M., Vandenburgh, K., Manninga, H., Scaringe, S.A., Weber, K., and Tuschl, T. (2003) Sequence, chemical, and structural variation of small interfering RNAs and short hairpin RNAs and the effect on mammalian gene silencing. *Antisense Nucleic Acid Drug Dev.*, **13**, 83–105.
 40. Amarzguoui, M., Holen, T., Babaie, E., and Prydz, H. (2003) Tolerance for mutations and chemical modifications in a siRNA. *Nucleic Acids Res.*, **31**, 589–595.
 41. Chiu, Y.L. and Rana, T.M. (2003) siRNA function in RNAi: a chemical modification analysis. *RNA*, **9**, 1034–1048.
 42. Choung, S., Kim, Y.J., Kim, S., Park, H.O., and Choi, Y.C. (2006) Chemical modification of siRNAs to improve serum stability without loss of efficacy. *Biochem. Biophys. Res. Commun.*, **342**, 919–927.
 43. Braasch, D.A., Jensen, S., Liu, Y., Kaur, K., Arar, K., White, M.A., and Corey, D.R. (2003) RNA interference in mammalian cells by chemically-modified RNA. *Biochemistry*, **42**, 7967–7975.
 44. Grünweller, A., Wyszko, E., Bieber, B., Jähnel, R., Erdmann, V.A., and Kurreck, J. (2003) Comparison of different antisense strategies in mammalian cells using locked nucleic acids, 2'-O-methyl RNA, phosphorothioates and small interfering RNA. *Nucleic Acids Res.*, **31**, 3185–3193.
 45. Hall, A.H., Wan, J., Shaughnessy, E.E., Ramsay Shaw, B., and Alexander, K.A. (2004) RNA interference using boranophosphate siRNAs: structure-activity relationships. *Nucleic Acids Res.*, **32**, 5991–6000.
 46. Elmén, J., Lindow, M., Schütz, S., Lawrence, M., Petri, A., Obad, S., Lindholm, M., Hedtjärn, M., Hansen, H.F., Berger, U. *et al.* (2008) LNA-mediated microRNA silencing in non-human primates. *Nature*, **452**, 896–899.
 47. Soutschek, J., Akinc, A., Bramlage, B., Charisse, K., Constien, R., Donoghue, M., Elbashir, S., Geick, A., Hadwiger, P., Harborth, J. *et al.* (2004) Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature*, **432**, 173–178.
 48. Chen, Q., Butler, D., Querbes, W., Pandey, R.K., Ge, P., Maier, M.A., Zhang, L., Rajeev, K.G., Nechev, L., Kotelianski, V. *et al.* (2010) Lipophilic siRNAs mediate efficient gene silencing in oligodendrocytes with direct

- CNS delivery. *J. Controlled Release*, **144**, 227–232.
49. Nishina, K., Unno, T., Uno, Y., Kubodera, T., Kanouchi, T., Mizusawa, H., and Yokota, T. (2008) Efficient *in vivo* delivery of siRNA to the liver by conjugation of alpha-tocopherol. *Mol. Ther.*, **16**, 734–740.
 50. Layzer, J.M., McCaffrey, A.P., Tanner, A.K., Huang, Z., Kay, M.A., and Sullenger, B.A. (2004) *In vivo* activity of nuclease-resistant siRNAs. *RNA*, **10**, 766–771.
 51. Mook, O.R., Baas, F., de Wissel, M.B., and Fluiter, K. (2007) Evaluation of locked nucleic acid-modified small interfering RNA *in vitro* and *in vivo*. *Mol. Cancer Ther.*, **6**, 833–843.
 52. Laursen, M.B., Pakula, M.M., Gao, S., Fluiter, K., Mook, O.R., Baas, F., Langkjaer, N., Wengel, S.L., Wengel, J., Kjems, J. *et al.* (2010) Utilization of unlocked nucleic acid (UNA) to enhance siRNA performance *in vitro* and *in vivo*. *Mol. Biosyst.*, **6**, 862–870.
 53. Bramsen, J.B., Pakula, M.M., Hansen, T.B., Bus, C., Langkjaer, N., Odadzic, D., Smicius, R., Wengel, S.L., Chattopadhyaya, J., Engels, J.W. *et al.* (2010) A screen of chemical modifications identifies position-specific modification by UNA to most potently reduce siRNA off-target effects. *Nucleic Acids Res.*, **38**, 5761–5773.
 54. Morrissey, D.V., Blanchard, K., Shaw, L., Jensen, K., Lockridge, J.A., Dickinson, B., McSwiggen, J.A., Vargeese, C., Bowman, K., Shaffer, C.S. *et al.* (2005) Activity of stabilized short interfering RNA in a mouse model of hepatitis B virus replication. *Hepatology*, **41**, 1349–1356.
 55. Morrissey, D.V., Lockridge, J.A., Shaw, L., Blanchard, K., Jensen, K., Breen, W., Hartsough, K., Machemer, L., Radka, S., Jadhav, V. *et al.* (2005) Potent and persistent *in vivo* anti-HBV activity of chemically modified siRNAs. *Nat. Biotechnol.*, **23**, 1002–1007.
 56. Manoharan, M., Akinc, A., Pandey, R.K., Qin, J., Hadwiger, P., John, M., Mills, K., Charisse, K., Maier, M.A., Nechev, L. *et al.* (2011) Unique gene-silencing and structural properties of 2'-fluoro-modified siRNAs. *Angew. Chem. Int. Ed. Engl.*, **50**, 2284–2288.
 57. Li, P., Sergueeva, Z.A., Dobrikov, M., and Shaw, B.R. (2007) Nucleoside and oligonucleoside boranophosphates: chemistry and properties. *Chem. Rev.*, **107**, 4746–4796.
 58. Sheehan, D., Lunstad, B., Yamada, C.M., Stell, B.G., Caruthers, M.H., and Dellinger, D.J. (2003) Biochemical properties of phosphonoacetate and thiophosphonoacetate oligodeoxyribonucleotides. *Nucleic Acids Res.*, **31**, 4109–4118.
 59. Yamada, C.M., Dellinger, D.J., and Caruthers, M.H. (2007) Synthesis and biological activity of phosphonocarbonylate DNA. *Nucleosides Nucleotides Nucleic Acids*, **26**, 539–546.
 60. Iwase, R., Toyama, T., and Nishimori, K. (2007) Solid-phase synthesis of modified RNAs containing amide-linked oligoribonucleosides at their 3'-end and their application to siRNA. *Nucleosides Nucleotides Nucleic Acids*, **26**, 1451–1454.
 61. Prakash, T.P., Kraynack, B., Baker, B.F., Swayze, E.E., and Bhat, B. (2006) RNA interference by 2',5'-linked nucleic acid duplexes in mammalian cells. *Bioorg. Med. Chem. Lett.*, **16**, 3238–3240.
 62. Freier, S.M. and Altmann, K.H. (1997) The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res.*, **25**, 4429–4443.
 63. Jackson, A.L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Johnson, J.M., Lim, L., Karpilow, J., Nichols, K. *et al.* (2006) Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA*, **12**, 1197–1205.
 64. Kraynack, B.A. and Baker, B.F. (2006) Small interfering RNAs containing full 2'-O-methylribonucleotide-modified sense strands display Argonaute2/eIF2C2-dependent activity. *RNA*, **12**, 163–176.
 65. Allerson, C.R., Sioufi, N., Jarres, R., Prakash, T.P., Naik, N., Berdeja, A., Wanders, L., Griffey, R.H., Swayze,

- E.E., and Bhat, B. (2005) Fully 2'-modified oligonucleotide duplexes with improved *in vitro* potency and stability compared to unmodified small interfering RNA. *J. Med. Chem.*, **48**, 901–904.
66. Capodici, J., Kariko, K., and Weissman, D. (2002) Inhibition of HIV-1 infection by small interfering RNA-mediated RNA interference. *J. Immunol.*, **169**, 5196–5201.
 67. Blidner, R.A., Hammer, R.P., Lopez, M.J., Robinson, S.O., and Monroe, W.T. (2007) Fully 2'-deoxy-2'-fluoro substituted nucleic acids induce RNA interference in mammalian cell culture. *Chem. Biol. Drug Des.*, **70**, 113–122.
 68. Watts, J.K., Choubdar, N., Sadalapure, K., Robert, F., Wahba, A.S., Pelletier, J., Pinto, B.M., and Damha, M.J. (2007) 2'-fluoro-4'-thioarabino-modified oligonucleotides: conformational switches linked to siRNA activity. *Nucleic Acids Res.*, **35**, 1441–1451.
 69. Deleavey, G.F., Watts, J.K., Alain, T., Robert, F., Kalota, A., Aishwarya, V., Pelletier, J., Gewirtz, A.M., Sonenberg, N., and Damha, M.J. (2010) Synergistic effects between analogs of DNA and RNA improve the potency of siRNA-mediated gene silencing. *Nucleic Acids Res.*, **38**, 4547–4557.
 70. Strapps, W.R., Pickering, V., Muir, G.T., Rice, J., Orsborn, S., Polisky, B.A., Sachs, A., and Bartz, S.R. (2010) The siRNA sequence and guide strand overhangs are determinants of *in vivo* duration of silencing. *Nucleic Acids Res.*, **38**, 4788–4797.
 71. Pirollo, K.F., Rait, A., Zhou, Q., Hwang, S.H., Dagata, J.A., Zon, G., Hogrefe, R.I., Palchik, G., and Chang, E.H. (2007) Materializing the potential of small interfering RNA via a tumor-targeting nanodelivery system. *Cancer Res.*, **67**, 2938–2943.
 72. Hogrefe, R.I., Lebedev, A.V., Zon, G., Pirollo, K.F., Rait, A., Zhou, Q., Yu, W., and Chang, E.H. (2006) Chemically modified short interfering hybrids (siHYBRIDS): nanoimmunoliposome delivery *in vitro* and *in vivo* for RNAi of HER-2. *Nucleosides Nucleotides Nucleic Acids*, **25**, 889–907.
 73. Parrish, S., Fleenor, J., Xu, S., Mello, C., and Fire, A. (2000) Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. *Mol. Cell*, **6**, 1077–1087.
 74. Ui-Tei, K., Naito, Y., Zenno, S., Nishi, K., Yamato, K., Takahashi, F., Juni, A., and Saigo, K. (2008) Functional dissection of siRNA sequence by systematic DNA substitution: modified siRNA with a DNA seed arm is a powerful tool for mammalian gene silencing with significantly reduced off-target effect. *Nucleic Acids Res.*, **36**, 2136–2151.
 75. Odadzic, D., Bramsen, J.B., Smicius, R., Bus, C., Kjems, J., and Engels, J.W. (2008) Synthesis of 2'-O-modified adenosine building blocks and application for RNA interference. *Bioorg. Med. Chem.*, **16**, 518–529.
 76. Wengel, J., Petersen, M., Nielsen, K.E., Jensen, G.A., Håkansson, A.E., Kumar, R., Sørensen, M.D., Rajwanshi, V.K., Bryld, T., and Jacobsen, J.P. (2001) LNA (locked nucleic acid) and the diastereoisomeric alpha-L-LNA: conformational tuning and high-affinity recognition of DNA/RNA targets. *Nucleosides Nucleotides Nucleic Acids*, **20**, 389–396.
 77. Srivastava, P., Barman, J., Pathmasiri, W., Plashkevych, O., Wenska, M., and Chattopadhyaya, J. (2007) Five- and six-membered conformationally locked 2',4'-carbocyclic ribo-thymidines: synthesis, structure, and biochemical studies. *J. Am. Chem. Soc.*, **129**, 8362–8379.
 78. Hamada, M., Ohtsuka, T., Kawaida, R., Koizumi, M., Morita, K., Furukawa, H., Imanishi, T., Miyagishi, M., and Taira, K. (2002) Effects on RNA interference in gene expression (RNAi) in cultured mammalian cells of mismatches and the introduction of chemical modifications at the 3'-ends of siRNAs. *Antisense Nucleic Acid Drug Dev.*, **12**, 301–309.
 79. Pradeepkumar, P.I., Amirkhanov, N.V., and Chattopadhyaya, J. (2003) Antisense oligonucleotides with oxetane-constrained cytidine enhance

- heteroduplex stability, and elicit satisfactory RNase H response as well as showing improved resistance to both exo and endonucleases. *Org. Biomol. Chem.*, **1**, 81–92.
80. Kurreck, J., Wyszko, E., Gillen, C., and Erdmann, V.A. (2002) Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res.*, **30**, 1911–1918.
 81. Fluiter, K., ten Asbroek, A.L., de Wissel, M.B., Jakobs, M.E., Wissenbach, M., Olsson, H., Olsen, O., Oerum, H., and Baas, F. (2003) *In vivo* tumor growth inhibition and biodistribution studies of locked nucleic acid (LNA) antisense oligonucleotides. *Nucleic Acids Res.*, **31**, 953–962.
 82. Elmén, J., Thonberg, H., Ljungberg, K., Frieden, M., Westergaard, M., Xu, Y., Wahren, B., Liang, Z., Ørum, H., Koch, T., and Wahlestedt, C. (2005) Locked nucleic acid (LNA) mediated improvements in siRNA stability and functionality. *Nucleic Acids Res.*, **33**, 439–447.
 83. Petersen, M. and Wengel, J. (2003) LNA: a versatile tool for therapeutics and genomics. *Trends Biotechnol.*, **21**, 74–81.
 84. Glud, S.Z., Bramsen, J.B., Dagnaes-Hansen, F., Wengel, J., Howard, K.A., Nyengaard, J.R., and Kjems, J. (2009) Naked siLNA-mediated gene silencing of lung bronchoepithelium EGFP expression after intravenous administration. *Oligonucleotides*, **19**, 163–168.
 85. Hornung, V., Guenther-Biller, M., Bourquin, C., Ablasser, A., Schlee, M., Uematsu, S., Noronha, A., Manoharan, M., Akira, S., de Fougerolles, A. *et al.* (2005) Sequence-specific potent induction of IFN- α by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nat. Med.*, **11**, 263–270.
 86. Dowler, T., Bergeron, D., Tedeschi, A.L., Paquet, L., Ferrari, N., and Damha, M.J. (2006) Improvements in siRNA properties mediated by 2'-deoxy-2'-fluoro-beta-D-arabinonucleic acid (FANA). *Nucleic Acids Res.*, **34**, 1669–1675.
 87. Fisher, M., Abramov, M., Van Aerschot, A., Rozenski, J., Dixit, V., Juliano, R.L., and Herdewijn, P. (2009) Biological effects of hexitol and altritol-modified siRNAs targeting B-Raf. *Eur. J. Pharmacol.*, **606**, 38–44.
 88. Fisher, M., Abramov, M., VanAerschot, A., Xu, D., Juliano, R.L., and Herdewijn, P. (2007) Inhibition of MDR1 expression with altritol-modified siRNAs. *Nucleic Acids Res.*, **35**, 1064–1074.
 89. Nauwelaerts, K., Fisher, M., Froeyen, M., Lescrinier, E., Aerschot, A.V., Xu, D., DeLong, R., Kang, H., Juliano, R.L., and Herdewijn, P. (2007) Structural characterization and biological evaluation of small interfering RNAs containing cyclohexenyl nucleosides. *J. Am. Chem. Soc.*, **129**, 9340–9348.
 90. Langkjær, N., Pasternak, A., and Wengel, J. (2009) UNA (unlocked nucleic acid): a flexible RNA mimic that allows engineering of nucleic acid duplex stability. *Bioorg. Med. Chem.*, **17**, 5420–5425.
 91. Kenski, D.M., Cooper, A.J., Li, J.J., Willingham, A.T., Haringsma, H.J., Young, T.A., Kuklin, N.A., Jones, J.J., Cancilla, M.T., McMasters, D.R. *et al.* (2009) Analysis of acyclic nucleoside modifications in siRNAs finds sensitivity at position 1 that is restored by 5'-terminal phosphorylation both *in vitro* and *in vivo*. *Nucleic Acids Res.*, **38**, 660–671.
 92. Dande, P., Prakash, T.P., Sioufi, N., Gaus, H., Jarres, R., Berdeja, A., Swayze, E.E., Griffey, R.H., and Bhat, B. (2006) Improving RNA interference in mammalian cells by 4'-thio-modified small interfering RNA (siRNA): effect on siRNA activity and nuclease stability when used in combination with 2'-O-alkyl modifications. *J. Med. Chem.*, **49**, 1624–1634.
 93. Hoshika, S., Minakawa, N., Kamiya, H., Harashima, H., and Matsuda, A. (2005) RNA interference induced by siRNAs modified with 4'-thioribonucleosides in cultured mammalian cells. *FEBS Lett.*, **579**, 3115–3118.

94. Hoshika, S., Minakawa, N., Shionoya, A., Imada, K., Ogawa, N., and Matsuda, A. (2007) Study of modification pattern-RNAi activity relationships by using siRNAs modified with 4'-thioribonucleosides. *ChemBioChem*, **8**, 2133–2138.
95. Takahashi, M., Minakawa, N., and Matsuda, A. (2009) Synthesis and characterization of 2'-modified-4'-thioRNA: a comprehensive comparison of nuclease stability. *Nucleic Acids Res.*, **37**, 1353–1362.
96. Sipa, K., Sochacka, E., Kazmierczak-Baranska, J., Maszewska, M., Janicka, M., Nowak, G., and Nawrot, B. (2007) Effect of base modifications on structure, thermodynamic stability, and gene silencing activity of short interfering RNA. *RNA*, **13**, 1301–1316.
97. Hornung, V., Ellegast, J., Kim, S., Brzózka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K.K., Schlee, M. *et al.* (2006) 5'-Triphosphate RNA is the ligand for RIG-I. *Science*, **314**, 994–997.
98. Li, Z.Y., Mao, H., Kallick, D.A., and Gorenstein, D.G. (2005) The effects of thiophosphate substitutions on native siRNA gene silencing. *Biochem. Biophys. Res. Commun.*, **329**, 1026–1030.
99. Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat. Struct. Mol. Biol.*, **11**, 576–577.
100. Chen, P.Y., Weinmann, L., Gaidatzis, D., Pei, Y., Zavolan, M., Tuschl, T., and Meister, G. (2008) Strand-specific 5'-O-methylation of siRNA duplexes controls guide strand selection and targeting specificity. *RNA*, **14**, 263–274.
101. Koller, E., Propp, S., Murray, H., Lima, W., Bhat, B., Prakash, T.P., Allerson, C.R., Swayze, E.E., Marcusson, E.G., and Dean, N.M. (2006) Competition for RISC binding predicts *in vitro* potency of siRNA. *Nucleic Acids Res.*, **34**, 4467–4476.
102. Tsui, N.B., Ng, E.K., and Lo, Y.M. (2002) Stability of endogenous and added RNA in blood specimens, serum, and plasma. *Clin. Chem.*, **48**, 1647–1653.
103. Lai, J.C., Tan, W., Benimetskaya, L., Miller, P., Colombini, M., and Stein, C.A. (2006) A pharmacologic target of G3139 in melanoma cells may be the mitochondrial VDAC. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 7494–7499.
104. Miner, P.B. Jr., Wedel, M.K., Xia, S., and Baker, B.F. (2006) Safety and efficacy of two dose formulations of alicaforsen enema compared with mesalazine enema for treatment of mild to moderate left-sided ulcerative colitis: a randomized, double-blind, active-controlled trial. *Aliment. Pharmacol. Ther.*, **23**, 1403–1413.
105. Bertrand, J.R., Pottier, M., Vekris, A., Opolon, P., Maksimenko, A., and Malvy, C. (2002) Comparison of antisense oligonucleotides and siRNAs in cell culture and *in vivo*. *Biochem. Biophys. Res. Commun.*, **296**, 1000–1004.
106. Chiu, Y.L. and Rana, T.M. (2002) RNAi in human cells: basic structural and functional features of small interfering RNA. *Mol. Cell*, **10**, 549–561.
107. Kennedy, S., Wang, D., and Ruvkun, G. (2004) A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature*, **427**, 645–649.
108. Takabatake, Y., Isaka, Y., Mizui, M., Kawachi, H., Takahara, S., and Imai, E. (2007) Chemically modified siRNA prolonged RNA interference in renal disease. *Biochem. Biophys. Res. Commun.*, **363**, 432–437.
109. Haupenthal, J., Baehr, C., Zeuzem, S., and Piiper, A. (2007) RNase A-like enzymes in serum inhibit the anti-neoplastic activity of siRNA targeting polo-like kinase 1. *Int. J. Cancer*, **121**, 206–210.
110. Haupenthal, J., Baehr, C., Kiermayer, S., Zeuzem, S., and Piiper, A. (2006) Inhibition of RNase A family enzymes prevents degradation and loss of silencing activity of siRNAs in serum. *Biochem. Pharmacol.*, **71**, 702–710.
111. Hickerson, R.P., Vlassov, A.V., Wang, Q., Leake, D., Ilves, H., Gonzalez-Gonzalez, E., Contag, C.H., Johnston, B.H., and Kaspar, R.L.

- (2008) Stability study of unmodified siRNA and relevance to clinical use. *Oligonucleotides*, **18**, 345–354.
112. Tourriere, H., Chebli, K., and Tazi, J. (2002) mRNA degradation machines in eukaryotic cells. *Biochimie*, **84**, 821–837.
 113. Sorrentino, S. (1998) Human extracellular ribonucleases: multiplicity, molecular diversity and catalytic properties of the major RNase types. *Cell. Mol. Life Sci.*, **54**, 785–794.
 114. Probst, J., Brechtel, S., Scheel, B., Hoerr, I., Jung, G., Rammensee, H.G., and Pascolo, S. (2006) Characterization of the ribonuclease activity on the skin surface. *Genet Vaccines Ther.*, **4**, 4.
 115. Zou, Y., Tiller, P., Chen, I.W., Beverly, M., and Hochman, J. (2008) Metabolite identification of small interfering RNA duplex by high-resolution accurate mass spectrometry. *Rapid Commun. Mass Spectrom.*, **22**, 1871–1881.
 116. Usher, D.A. (1969) On the mechanism of ribonuclease action. *Proc. Natl. Acad. Sci. U.S.A.*, **62**, 661–667.
 117. Eder, P.S., DeVine, R.J., Dagle, J.M., and Walder, J.A. (1991) Substrate specificity and kinetics of degradation of antisense oligonucleotides by a 3' exonuclease in plasma. *Antisense Res. Dev.*, **1**, 141–151.
 118. Zimmermann, T.S., Lee, A.C., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M.N., Harborth, J., Heyes, J.A., Jeffs, L.B., John, M. *et al.* (2006) RNAi-mediated gene silencing in non-human primates. *Nature*, **441**, 111–114.
 119. Volkov, A.A., Kruglova, N.S., Meschaninova, M.I., Venyaminova, A.G., Zenkova, M.A., Vlassov, V.V., and Chernolovskaya, E.L. (2009) Selective protection of nuclease-sensitive sites in siRNA prolongs silencing effect. *Oligonucleotides*, **19**, 191–202.
 120. Turner, J.J., Jones, S.W., Moschos, S.A., Lindsay, M.A., and Gait, M.J. (2007) MALDI-TOF mass spectral analysis of siRNA degradation in serum confirms an RNase A-like activity. *Mol. Biosyst.*, **3**, 43–50.
 121. Qiu, L., Moreira, A., Kaplan, G., Levitz, R., Wang, J.Y., Xu, C., and Drlica, K. (1998) Degradation of hammerhead ribozymes by human ribonucleases. *Mol. Gen. Genet.*, **258**, 352–362.
 122. Libonati, M. and Sorrentino, S. (1992) Revisiting the action of bovine ribonuclease A and pancreatic-type ribonucleases on double-stranded RNA. *Mol. Cell. Biochem.*, **117**, 139–151.
 123. Hoerter, J.A. and Walter, N.G. (2007) Chemical modification resolves the asymmetry of siRNA strand degradation in human blood serum. *RNA*, **13**, 1887–1893.
 124. Bartlett, D.W. and Davis, M.E. (2007) Effect of siRNA nuclease stability on the *in vitro* and *in vivo* kinetics of siRNA-mediated gene silencing. *Biotechnol. Bioeng.*, **97**, 909–921.
 125. Life Technologies Ambion® Technical Resources, RNAi/siRNA, Duration of siRNA induced silencing: your questions answered. *TechNotes* (online) 15.
 126. Song, E., Lee, S.K., Dykxhoorn, D.M., Novina, C., Zhang, D., Crawford, K., Cerny, J., Sharp, P.A., Lieberman, J., Manjunath, N., and Shankar, P. (2003) Sustained small interfering RNA-mediated human immunodeficiency virus type 1 inhibition in primary macrophages. *J. Virol.*, **77**, 7174–7181.
 127. Raemdonck, K., Remaut, K., Lucas, B., Sanders, N.N., Demeester, J., and De Smedt, S.C. (2006) *In situ* analysis of single-stranded and duplex siRNA integrity in living cells. *Biochemistry*, **45**, 10614–10623.
 128. Yoo, J.W., Kim, S., and Lee, D.K. (2008) Competition potency of siRNA is specified by the 5'-half sequence of the guide strand. *Biochem. Biophys. Res. Commun.*, **367**, 78–83.
 129. Sioud, M. (2009) Deciphering the code of innate immunity recognition of siRNAs. *Methods Mol. Biol.*, **487**, 41–59.
 130. Meurs, E., Chong, K., Galabru, J., Thomas, N.S., Kerr, I.M., Williams, B.R., and Hovanessian, A.G. (1990) Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. *Cell*, **62**, 379–390.

131. Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S., and Fujita, T. (2004) The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat. Immunol.*, **5**, 730–737.
132. Marques, J.T., Devosse, T., Wang, D., Zamanian-Daryoush, M., Serbinowski, P., Hartmann, R., Fujita, T., Behlke, M.A., and Williams, B.R. (2006) A structural basis for discriminating between self and nonself double-stranded RNAs in mammalian cells. *Nat. Biotechnol.*, **24**, 559–565.
133. Kang, D.C., Gopalkrishnan, R.V., Wu, Q., Jankowsky, E., Pyle, A.M., and Fisher, P.B. (2002) mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 637–642.
134. Puthenveetil, S., Whitby, L., Ren, J., Kelnar, K., Krebs, J.F., and Beal, P.A. (2006) Controlling activation of the RNA-dependent protein kinase by siRNAs using site-specific chemical modification. *Nucleic Acids Res.*, **34**, 4900–4911.
135. Kim, D.H., Longo, M., Han, Y., Lundberg, P., Cantin, E., and Rossi, J.J. (2004) Interferon induction by siRNAs and ssRNAs synthesized by phage polymerase. *Nat. Biotechnol.*, **22**, 321–325.
136. Kariko, K., Bhuyan, P., Capodici, J., Ni, H., Lubinski, J., Friedman, H., and Weissman, D. (2004) Exogenous siRNA mediates sequence-independent gene suppression by signaling through toll-like receptor 3. *Cells Tissues Organs*, **177**, 132–138.
137. Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., and Bauer, S. (2004) Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science*, **303**, 1526–1529.
138. Gantier, M.P., Tong, S., Behlke, M.A., Xu, D., Phipps, S., Foster, P.S., and Williams, B.R. (2008) TLR7 is involved in sequence-specific sensing of single-stranded RNAs in human macrophages. *J. Immunol.*, **180**, 2117–2124.
139. Muzio, M., Bosio, D., Polentarutti, N., D'Amico, G., Stoppacciaro, A., Mancinelli, R., van't Veer, C., Penton-Rol, G., Ruco, L.P., Allavena, P., and Mantovani, A. (2000) Differential expression and regulation of toll-like receptors (TLR) in human leukocytes: Selective expression of TLR3 in dendritic cells. *J. Immunol.*, **164**, 5998–6004.
140. Kleinman, M.E., Yamada, K., Takeda, A., Chandrasekaran, V., Nozaki, M., Baffi, J.Z., Albuquerque, R.J.C., Yamasaki, S., Itaya, M., Pan, Y.Z. *et al.* (2008) Sequence- and target-independent angiogenesis suppression by siRNA via TLR3. *Nature*, **452**, 591–U591.
141. Kariko, K., Bhuyan, P., Capodici, J., and Weissman, D. (2004) Small interfering RNAs mediate sequence-independent gene suppression and induce immune activation by signaling through toll-like receptor 3. *J. Immunol.*, **172**, 6545–6549.
142. Alexopoulou, L., Holt, A.C., Medzhitov, R., and Flavell, R.A. (2001) Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature*, **413**, 732–738.
143. Kleinman, M.E., Kaneko, H., Cho, W.G., Dridi, S., Fowler, B.J., Blandford, A.D., Albuquerque, R.J., Hirano, Y., Terasaki, H., Kondo, M. *et al.* (2012) Short-interfering RNAs induce retinal degeneration via TLR3 and IRF3. *Mol. Ther.*, **20**, 101–108.
144. Zarembler, K.A. and Godowski, P.J. (2002) Tissue expression of human Toll-like receptors and differential regulation of Toll-like receptor mRNAs in leukocytes in response to microbes, their products, and cytokines. *J. Immunol.*, **168**, 554–561.
145. Diebold, S.S., Massacrier, C., Akira, S., Paturel, C., Morel, Y., and Reis e Sousa, C. (2006) Nucleic acid agonists for Toll-like receptor 7 are defined by the presence of uridine

- ribonucleotides. *Eur. J. Immunol.*, **36**, 3256–3267.
146. Diebold, S.S., Kaisho, T., Hemmi, H., Akira, S., and Reis e Sousa, C. (2004) Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science*, **303**, 1529–1531.
 147. Judge, A.D., Sood, V., Shaw, J.R., Fang, D., McClintock, K., and MacLachlan, I. (2005) Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nat. Biotechnol.*, **23**, 457–462.
 148. Sioud, M. (2005) Induction of inflammatory cytokines and interferon responses by double-stranded and single-stranded siRNAs is sequence-dependent and requires endosomal localization. *J. Mol. Biol.*, **348**, 1079–1090.
 149. Heidel, J.D., Hu, S., Liu, X.F., Triche, T.J., and Davis, M.E. (2004) Lack of interferon response in animals to naked siRNAs. *Nat. Biotechnol.*, **22**, 1579–1582.
 150. Kariko, K., Buckstein, M., Ni, H., and Weissman, D. (2005) Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity*, **23**, 165–175.
 151. Cekaite, L., Furset, G., Hovig, E., and Sioud, M. (2007) Gene expression analysis in blood cells in response to unmodified and 2'-modified siRNAs reveals TLR-dependent and independent effects. *J. Mol. Biol.*, **365**, 90–108.
 152. Judge, A.D., Bola, G., Lee, A.C., and MacLachlan, I. (2006) Design of noninflammatory synthetic siRNA mediating potent gene silencing *in vivo*. *Mol. Ther.*, **13**, 494–505.
 153. Hamm, S., Latz, E., Hangel, D., Müller, T., Yu, P., Golenbock, D., Sparwasser, T., Wagner, H., and Bauer, S. (2009) Alternating 2'-O-ribose methylation is a universal approach for generating non-stimulatory siRNA by acting as TLR7 antagonist. *Immunobiology*, **215**, 559–569.
 154. Robbins, M., Judge, A., Liang, L., McClintock, K., Yaworski, E., and MacLachlan, I. (2007) 2'-O-methyl-modified RNAs act as TLR7 antagonists. *Mol. Ther.*, **15**, 1663–1669.
 155. Goodchild, A., Nopper, N., King, A., Doan, T., Tanudji, M., Arndt, G.M., Poidinger, M., Rivory, L.P., and Passioura, T. (2009) Sequence determinants of innate immune activation by short interfering RNAs. *BMC Immunol.*, **10**, 40.
 156. Zamanian-Daryoush, M., Marques, J.T., Gantier, M.P., Behlke, M.A., John, M., Rayman, P., Finke, J., and Williams, B.R. (2008) Determinants of cytokine induction by small interfering RNA in human peripheral blood mononuclear cells. *J. Interferon Cytokine Res.*, **28**, 221–233.
 157. Du, Q., Thonberg, H., Wang, J., Wahlestedt, C., and Liang, Z. (2005) A systematic analysis of the silencing effects of an active siRNA at all single-nucleotide mismatched target sites. *Nucleic Acids Res.*, **33**, 1671–1677.
 158. Dahlgren, C., Zhang, H.Y., Du, Q., Grahn, M., Norstedt, G., Wahlestedt, C., and Liang, Z. (2008) Analysis of siRNA specificity on targets with double-nucleotide mismatches. *Nucleic Acids Res.*, **36**, e53.
 159. Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.*, **21**, 635–637.
 160. Birmingham, A., Anderson, E.M., Reynolds, A., Ilesley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J. *et al.* (2006) 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods*, **3**, 199–204.
 161. Lim, L.P., Lau, N.C., Garrett-Engle, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769–773.

162. Doench, J.G., Petersen, C.P., and Sharp, P.A. (2003) siRNAs can function as miRNAs. *Genes Dev.*, **17**, 438–442.
163. Fedorov, Y., Anderson, E.M., Birmingham, A., Reynolds, A., Karpilow, J., Robinson, K., Leake, D., Marshall, W.S., and Khvorova, A. (2006) Off-target effects by siRNA can induce toxic phenotype. *RNA*, **12**, 1188–1196.
164. Lin, X., Ruan, X., Anderson, M.G., McDowell, J.A., Kroeger, P.E., Fesik, S.W., and Shen, Y. (2005) siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res.*, **33**, 4527–4535.
165. Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.*, **18**, 504–511.
166. Williams, A.E. (2008) Functional aspects of animal microRNAs. *Cell. Mol. Life Sci.*, **65**, 545–562.
167. Wu, L., Fan, J., and Belasco, J.G. (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 4034–4039.
168. Vaish, N., Chen, F., Seth, S., Fosnaugh, K., Liu, Y., Adami, R., Brown, T., Chen, Y., Harvie, P., Johns, R. *et al.* (2011) Improved specificity of gene silencing by siRNAs containing unlocked nucleobase analogs. *Nucleic Acids Res.*, **39**, 1823–1832.
169. Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A. *et al.* (2003) Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA*, **9**, 493–501.
170. Laktionov, P.P., Dazard, J.E., Vives, E., Rykova, E.Y., Piette, J., Vlassov, V.V., and Lebleu, B. (1999) Characterisation of membrane oligonucleotide-binding proteins and oligonucleotide uptake in keratinocytes. *Nucleic Acids Res.*, **27**, 2315–2324.
171. de Diesbach, P., Berens, C., N’Kuli, F., Monsigny, M., Sonveaux, E., Wattiez, R., and Courtoy, P.J. (2000) Identification, purification and partial characterisation of an oligonucleotide receptor in membranes of HepG2 cells. *Nucleic Acids Res.*, **28**, 868–874.
172. Lingor, P., Michel, U., Schöll, U., Bähr, M., and Kügler, S. (2004) Transfection of “naked” siRNA results in endosomal uptake and metabolic impairment in cultured neurons. *Biochem. Biophys. Res. Commun.*, **315**, 1126–1133.
173. Overhoff, M., Wünsche, W., and Sczakiel, G. (2004) Quantitative detection of siRNA and single-stranded oligonucleotides: relationship between uptake and biological activity of siRNA. *Nucleic Acids Res.*, **32**, e170.
174. Davidson, T.J., Harel, S., Arboleda, V.A., Prunell, G.F., Shelanski, M.L., Greene, L.A., and Troy, C.M. (2004) Highly efficient small interfering RNA delivery to primary mammalian neurons induces MicroRNA-like effects before mRNA degradation. *J. Neurosci.*, **24**, 10040–10046.
175. Muratovska, A. and Eccles, M.R. (2004) Conjugate for efficient delivery of short interfering RNA (siRNA) into mammalian cells. *FEBS Lett.*, **558**, 63–68.
176. Moschos, S.A., Jones, S.W., Perry, M.M., Williams, A.E., Erjefalt, J.S., Turner, J.J., Barnes, P.J., Sproat, B.S., Gait, M.J., and Lindsay, M.A. (2007) Lung delivery studies using siRNA conjugated to TAT(48-60) and penetratin reveal peptide induced reduction in gene expression and induction of innate immunity. *Bioconjug. Chem.*, **18**, 1450–1459.
177. Kim, W.J., Christensen, L.V., Jo, S., Yockman, J.W., Jeong, J.H., Kim, Y.H., and Kim, S.W. (2006) Cholesteryl oligoarginine delivering vascular endothelial growth factor siRNA effectively inhibits tumor growth in colon adenocarcinoma. *Mol. Ther.*, **14**, 343–350.
178. Chiu, Y.L., Ali, A., Chu, C.Y., Cao, H., and Rana, T.M. (2004) Visualizing a correlation between siRNA localization, cellular uptake, and RNAi in living cells. *Chem. Biol.*, **11**, 1165–1175.
179. Abes, S., Moulton, H., Turner, J., Clair, P., Richard, J.P., Iversen, P., Gait, M.J., and Lebleu, B. (2007) Peptide-based

- delivery of nucleic acids: design, mechanism of uptake and applications to splice-correcting oligonucleotides. *Biochem. Soc. Trans.*, **35**, 53–55.
180. Bolcato-Bellemin, A.L., Bonnet, M.E., Creusat, G., Erbacher, P., and Behr, J.P. (2007) Sticky overhangs enhance siRNA-mediated gene silencing. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 16050–16055.
 181. Lee, S.Y., Huh, M.S., Lee, S., Lee, S.J., Chung, H., Park, J.H., Oh, Y.K., Choi, K., Kim, K., and Kwon, I.C. (2009) Stability and cellular uptake of polymerized siRNA (poly-siRNA)/polyethylenimine (PEI) complexes for efficient gene silencing. *J. Controlled Release*, **141**, 339–346.
 182. Ikeda, Y. and Taira, K. (2006) Ligand-targeted delivery of therapeutic siRNA. *Pharm. Res.*, **23**, 1631–1640.
 183. Medarova, Z., Pham, W., Farrar, C., Petkova, V., and Moore, A. (2007) *In vivo* imaging of siRNA delivery and silencing in tumors. *Nat. Med.*, **13**, 372–377.
 184. Derfus, A.M., Chen, A.A., Min, D.H., Ruoslahti, E., and Bhatia, S.N. (2007) Targeted quantum dot conjugates for siRNA delivery. *Bioconjug. Chem.*, **18**, 1391–1396.
 185. Oishi, M., Nagasaki, Y., Nishiyama, N., Itaka, K., Takagi, M., Shimamoto, A., Furuichi, Y., and Kataoka, K. (2007) Enhanced growth inhibition of hepatic multicellular tumor spheroids by lactosylated poly(ethylene glycol)-siRNA conjugate formulated in PEGylated polyplexes. *ChemMedChem*, **2**, 1290–1297.
 186. Cesarone, G., Edupuganti, O.P., Chen, C.P., and Wickstrom, E. (2007) Insulin receptor substrate 1 knockdown in human MCF7 ER+ breast cancer cells by nuclease-resistant IRS1 siRNA conjugated to a disulfide-bridged D-peptide analogue of insulin-like growth factor 1. *Bioconjug. Chem.*, **18**, 1831–1840.
 187. Xia, C.F., Zhang, Y., Boado, R.J., and Pardridge, W.M. (2007) Intravenous siRNA of brain cancer with receptor targeting and avidin-biotin technology. *Pharm. Res.*, **24**, 2309–2316.
 188. Oishi, M., Nagasaki, Y., Itaka, K., Nishiyama, N., and Kataoka, K. (2005) Lactosylated poly(ethylene glycol)-siRNA conjugate through acid-labile beta-thiopropionate linkage to construct pH-sensitive polyion complex micelles achieving enhanced gene silencing in hepatoma cells. *J. Am. Chem. Soc.*, **127**, 1624–1625.
 189. Kim, S.S., Garg, H., Joshi, A., and Manjunath, N. (2009) Strategies for targeted nonviral delivery of siRNAs *in vivo*. *Trends Mol. Med.*, **15**, 491–500.
 190. Gao, S., Dagnaes-Hansen, F., Nielsen, E.J., Wengel, J., Besenbacher, F., Howard, K.A., and Kjems, J. (2009) The effect of chemical modification and nanoparticle formulation on stability and biodistribution of siRNA in mice. *Mol. Ther.*, **17**, 1225–1233.
 191. Bartlett, D.W. and Davis, M.E. (2008) Impact of tumor-specific targeting and dosing schedule on tumor growth inhibition after intravenous administration of siRNA-containing nanoparticles. *Biotechnol. Bioeng.*, **99**, 975–985.
 192. Urban-Klein, B., Werth, S., Abuharbeid, S., Czubayko, F., and Aigner, A. (2005) RNAi-mediated gene-targeting through systemic application of polyethylenimine (PEI)-complexed siRNA *in vivo*. *Gene Ther.*, **12**, 461–466.
 193. Schifferers, R.M., Ansari, A., Xu, J., Zhou, Q., Tang, Q., Storm, G., Molema, G., Lu, P.Y., Scaria, P.V., and Woodle, M.C. (2004) Cancer siRNA therapy by tumor selective delivery with ligand-targeted sterically stabilized nanoparticle. *Nucleic Acids Res.*, **32**, e149.
 194. Song, E., Zhu, P., Lee, S.K., Chowdhury, D., Kussman, S., Dykxhoorn, D.M., Feng, Y., Palliser, D., Weiner, D.B., Shankar, P. *et al.* (2005) Antibody mediated *in vivo* delivery of small interfering RNAs via cell-surface receptors. *Nat. Biotechnol.*, **23**, 709–717.
 195. Heidel, J.D., Yu, Z., Liu, J.Y., Rele, S.M., Liang, Y., Zeidan, R.K., Kornbrust, D.J., and Davis, M.E. (2007) Administration in non-human primates of escalating intravenous doses

- of targeted nanoparticles containing ribonucleotide reductase subunit M2 siRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 5715–5721.
196. Kim, S.H., Mok, H., Jeong, J.H., Kim, S.W., and Park, T.G. (2006) Comparative evaluation of target-specific GFP gene silencing efficiencies for antisense ODN, synthetic siRNA, and siRNA plasmid complexed with PEI-PEG-FOL conjugate. *Bioconjug. Chem.*, **17**, 241–244.
 197. Shen, J., Samul, R., Silva, R.L., Akiyama, H., Liu, H., Saishin, Y., Hackett, S.F., Zinnen, S., Kossen, K., Fosnaugh, K. *et al.* (2006) Suppression of ocular neovascularization with siRNA targeting VEGF receptor 1. *Gene Ther.*, **13**, 225–234.
 198. Wolfrum, C., Shi, S., Jayaprakash, K.N., Jayaraman, M., Wang, G., Pandey, R.K., Rajeev, K.G., Nakayama, T., Charrise, K., Ndungo, E.M. *et al.* (2007) Mechanisms and optimization of *in vivo* delivery of lipophilic siRNAs. *Nat. Biotechnol.*, **25**, 1149–1157.
 199. Bumcrot, D., Manoharan, M., Koteliensky, V., and Sah, D.W. (2006) RNAi therapeutics: a potential new class of pharmaceutical drugs. *Nat. Chem. Biol.*, **2**, 711–719.
 200. Myers, J.W., Chi, J.T., Gong, D., Schaner, M.E., Brown, P.O., and Ferrell, J.E. (2006) Minimizing off-target effects by using diced siRNAs for RNA interference. *J. RNAi Gene Silencing*, **2**, 181–194.
 201. Kittler, R., Surendranath, V., Heninger, A.K., Slabicki, M., Theis, M., Putz, G., Franke, K., Caldarelli, A., Grabner, H., Kozak, K. *et al.* (2007) Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nat. Methods*, **4**, 337–344.

Appendix: UV Spectroscopy for the Quantitation of RNA

UV spectroscopy provides the almost universal basis for nucleic acid quantitation. Reliable determination of RNA/DNA concentration is critical for many applications such as ribozyme kinetics or measurements of equilibrium constants in RNA-ligand interactions. A recent publication [1] has readdressed the “old” molar extinction (absorbance) coefficients determined by UV spectroscopy for nucleoside-5'-monophosphates ($\epsilon_{260 \text{ nm}}$ values are essentially identical for 5'- and 3'-NMPs). This study, utilizing NMR for accurate determination of concentration and detection of impurities, revealed that $\epsilon_{260 \text{ nm}}$ values, particularly those for pC and pDC, had previously been overestimated. Thus, concentrations of single-stranded RNA (ssRNA) and single-stranded DNA (ssDNA), determined by complete hydrolysis and using the “old” extinction coefficients, were somewhat underestimated – for mixed non-repetitive sequences by about 4 and 2–3%, respectively. The authors thus recommended revised rules of thumb for such sequences: 1 A_{260} unit = 38 $\mu\text{g/ml}$ for ssDNA and 1 A_{260} unit = 37 $\mu\text{g/ml}$ for ssRNA (free acid) due to the higher ϵ_{260} of pU versus pT. This correlation changes with the type of counter ion present (e.g., multiply 37 $\mu\text{g/ml}$ with a factor of 1.07 for the sodium salt of the ssRNA). Nonetheless, the value of 37 $\mu\text{g/ml}$ roughly matches information given in [2] (1 A_{260} unit = 38 $\mu\text{g/ml}$ for ssRNA), but slightly differs from the frequently recommended correlation 1 A_{260} unit = 40 $\mu\text{g/ml}$ (e.g., Ambion or Chapter 3).

Two classes of nucleic acids can deviate more substantially from the aforementioned rules of thumb: (i) oligonucleotides with strong sequence bias (e.g., purine clusters) and (ii) structured RNAs, as discussed in the following.

A central issue is hypochromicity owing to base stacking interactions. Hypochromicity effects increase with the number of neighboring purines. A strategy often applied to calculate ϵ_{260} values for oligonucleotides is subtraction of ϵ values of individual mononucleotides from those of dinucleotide neighbors along the sequence (nearest neighbor estimates). A concise guideline of this method, including a calculation example, is available from Integrated DNA Technologies (Coralville, IA, USA; PDF for download: “Oligonucleotide Yield Resuspension and Storage”; access: IDT homepage, Support, FAQs, Reading Room, Technical Reports). Using this algorithm and correcting by factors derived from the revised ϵ_{260} values of mononucleotides [1] should provide a more accurate basis for UV-based quantitation of oligonucleotides.

In the case of structured RNAs, the classical approach is alkaline or ribonuclease hydrolysis of the RNA to mononucleotides and subsequent calculation of RNA concentration on the basis of ϵ_{260} values for the individual mononucleotides. To obtain ϵ_{260} of the intact RNA, the sum of mononucleotide ϵ_{260} values is then corrected by a factor corresponding to the ratio of A_{260} values of the RNA sample before and after hydrolysis. Alternative but more rarely used RNA quantitation approaches are phosphate analysis [3] or a fluorometric assay [4]. To give an example for the hydrolysis approach, this method resulted in an ϵ_{260} value of $3.46 \times 10^6 \text{ M}^{-1} \text{ cm}^{-1}$ for a 409-nt transcript (116 A, 116 G, 83 C, and 94 U residues) of *Bacillus subtilis* RNase P RNA [5]. Calculating the molecular weight (M.W.) according to the Life Technologies Invitrogen guidelines ("DNA and RNA Molecular Weights and Conversions"), a value of 132.5 $\mu\text{g}/\text{nmol}$ is obtained ($\text{M.W.} = (n \times \text{A} \times 329.2) + (n \times \text{U} \times 306.2) + (n \times \text{C} \times 305.2) + (n \times \text{G} \times 345.2) + 159$ for a 5'-triphosphate), resulting in 38.7 $\mu\text{g}/\text{ml}$ per A_{260} unit for the free acid based on Lambert-Beer law and using the $\epsilon_{260} = 3.46 \times 10^6 \text{ M}^{-1} \text{ cm}^{-1}$. An experimental ϵ_{260} of $3.2 \times 10^6 \text{ M}^{-1} \text{ cm}^{-1}$ was obtained for the 388-nt L-21 *ScaI* transcript of the *Tetrahymena* ribozyme (about 39.6 $\mu\text{g}/\text{ml}$ per A_{260} unit for the free acid; [6]). Thus, the experimentally determined ϵ_{260} values for these structured RNAs are in the same range as those given for ssRNA (37–40 $\mu\text{g}/\text{ml}$). The reason why structured RNAs containing more purines than pyrimidines adhere to the rule $1 A_{260} = 37\text{--}40 \mu\text{g}/\text{ml}$ could be related to two counteracting phenomena: (i) The ϵ_{260} values for mononucleotides at neutral pH decrease in the order $\text{pA}: 15.02 > \text{pG}: 12.08 > \text{pU}: 9.66 > \text{pC}: 7.07$ (in $\text{mM}^{-1} \text{ cm}^{-1}$; [1]). Thus, overrepresentation of A (and G) residues should lead to overestimation of RNA concentration. (ii) This effect is at least partially compensated by enhanced hypochromicity due to increased stacking of purines.

Some RNAs may substantially exceed the correlation of $\sim 40 \mu\text{g}/\text{ml}$ per A_{260} unit. For example, about 51 $\mu\text{g}/\text{ml}$ per A_{260} unit (free acid) was obtained for a 23-meric RNA-stem loop structure based on ϵ_{260} determined by phosphorus quantitation [7]. It is therefore suggested to use one of the two straightforward experimental protocols outlined below to determine more precise ϵ_{260} values for structured RNA molecules of known sequence. The molecular weight of the RNA should be determined as that of the free acid (taking into account whether the RNA carries a 5'-PPP, 5'-P, or 5'-OH terminus), since RNAs are usually precipitated with ethanol in the presence of NaOAc, washed with 70% ethanol and are redissolved in water after air-drying. Residual Na^+ ions redissolved with the RNA, as well as the RNA's phosphate moieties, will be shielded by hydration shells such that these charged groups interact more weakly and diffusely. Thus, calculations based on the M.W. of the free acid of RNA will be appropriate and sufficient.

First Approach

- 1) Measure the A_{260} of your RNA sample at room temperature in a defined buffer, for example, 50 mM ammonium acetate (pH 7.0). Trace amounts of components that stimulate activity of hydrolysis enzymes (step 2), such as ZnSO_4 , may be added.

- 2) Add nuclease P1 (or a mixture of RNases A, T1, and T2); incubate until the absorbance at 260 nm no longer increases over time (at least 1 h); measure A_{260} at the endpoint of hydrolysis [6, 8].
- 3) Calculate ϵ_{260} based on the A_{260} value at the endpoint of hydrolysis using the extinction coefficients of the individual mononucleotides according to [1].
- 4) Multiply the ϵ_{260} value obtained in step 3 by the ratio of A_{260} for the intact (step 1) versus hydrolyzed (step 2) RNA solution. This yields the ϵ_{260} value for your folded RNA in this particular buffer.

Variation of the First Approach (Adapted from Open Net Ware, Protocols, RNA, Quantitating Nucleic Acids, Robert T. Sauer Lab; http://openwetware.org/wiki/Quantitating_nucleic_acids):

- 1) Mix, for example, 5 μ l of your RNA sample with 995 μ l 0.1 M phosphate buffer/1 mM EDTA (pH 7.0) and measure the A_{260} at room temperature.
- 2) Take another 5 μ l of your RNA sample, add 20 μ l of 1.0 N NaOH, incubate for 1 h at 37 °C, then add 20 μ l 1.0 N HCl to neutralize the solution (note that ϵ_{260} values for 5'-GMP and 5'-UMP are lower at pH 11 versus 7 [9]). Add 955 μ l 0.1 M phosphate buffer/1 mM EDTA (pH 7.0) and measure the A_{260} at room temperature.
- 3) Proceed as in steps 3 and 4 above.

Second Approach

The second approach [8] essentially differs from the first in step 2. Instead of digestion to mononucleotides, the RNA is approximately fully denatured at a high temperature at which A_{260} is measured (> 80 °C, in the absence of divalent metal ions and generally at low ionic strength, for example, at 20 mM Tris-HCl, pH 7.0).

Perform the simple A_{260} measurements before and after RNA hydrolysis in triplicate to get robust ϵ_{260} values. Finally, although seemingly trivial, it should be mentioned that a major source of erroneous RNA quantitation will be insufficient separation of nucleotides from RNA samples after transcription. Also, one should be aware that UV absorption of nucleic acids is pH-sensitive [9] and that reliable measurement requires an OD_{260} of at least 0.1, making cuvettes with small volumes rather attractive when amounts of RNA are limited. Ionic strength, nature of ions, presence of EDTA or denaturants such as urea and formamide, and temperature influence UV absorption by affecting the folding state of the RNA. In addition, the quality of the UV spectrophotometer will play a role, such as the accuracy of cuvette positioning.

References

1. Cavaluzzi, M.J. and Borer, P.N. (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.*, **32**, e13.
2. Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning. A Laboratory Manual*, 3rd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

3. Murphy, J.H. and Trapane, T.L. (1996) Concentration and extinction coefficient determination for oligonucleotides and analogs using a general phosphate analysis. *Anal. Biochem.*, **240**, 273–282.
4. Jones, L.J., Yue, S.T., Cheung, C.Y., and Singer, V.L. (1998) RNA quantitation by fluorescence-based solution assay: RiboGreen reagent characterization. *Anal Biochem.*, **265**, 368–374.
5. Fang, X.W., Yang, X.J., Littrell, K., Niranjankumari, S., Thiagarajan, P., Fierke, C.A., Sosnick, T.R., and Pan, T. (2001) The *Bacillus subtilis* RNase P holoenzyme contains two RNase P RNA and two RNase P protein subunits. *RNA*, **7**, 233–241.
6. Zaug, A.J., Grosshans, C.A., and Cech, T.R. (1988) Sequence-specific endoribonuclease activity of the *Tetrahymena* ribozyme: enhanced cleavage of certain oligonucleotide substrates that form mismatched ribozyme-substrate complexes. *Biochemistry*, **27**, 8924–8931.
7. Cowan, J.A., Ohyama, T., Wang, D., and Natarajan, K. (2000) Recognition of a cognate RNA aptamer by neomycin B: quantitative evaluation of hydrogen bonding and electrostatic interactions. *Nucleic Acids Res.*, **28**, 2935–2942.
8. Pan, T. and Sosnick, T.R. (1997) Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity. *Nat. Struct. Biol.*, **4**, 931–938.
9. Bock, R.M., Ling, N.S., Morell, S.A., and Lipton, S.H. (1956) Ultraviolet absorption spectra of adenosin-5'-triphosphate and related 5'-ribonucleotides. *Arch. Biochem. Biophys.*, **62**, 253–264.

Index

a

- abstract shape analysis
 - applications 583–584
 - comparative
 - from aligned sequences 587–588
 - from unaligned sequences 588–592
 - definition 580–582
 - functions 580
 - general caveats 582–583
 - probabilistic shape analysis 585–587
 - RNA folding space 579–580
 - RNAlishapes parameter 593
 - RNASHapes parameter 592–593
 - shape representative structures computing 581
- acetal levulinyl ester (ALE) 130
- ADE2 gene 1071, 1080
- ADH1 promoter 1073
- 3'-adapter
 - ligation 862–863, 889–890
 - preadenylated, for Illumina sequencing 864–865
- 5'-adapter
 - ligation 869, 890–891
 - oligoribonucleotide, compatible with Illumina sequencing 865
- Advanced Photon Source (APS) 409, 411–412
- affinity matrices, for RNA binding proteins
 - isolation 921–924, 948–952
 - applications
 - nuclear tRNase Z purification from wheat germ 927–930
 - tRNA-splicing ligase purification from wheat germ 930
 - materials 924–925
 - methods 925–927
- affinity purification. *See* chromatography
- affinity-selected material analysis 943–944
- Affymetrix Integrated Genome Browser 724
- AFM. *See* atomic force microscopy imaging
- Agilent eArray system 796
- Ago2 immunoprecipitation, Western blot analysis 1090
- ALIFOLDZ algorithm 727
- aliphatic isocyanates synthesis 160–161
- altritol 1251
- AMBER (Assisted Model Building with Energy Refinement) program, MD simulation 687, 695
- 2'-amino approach
 - with aromatic isothiocyanates 158–159
 - postsynthetic modification 156–158
 - with aliphatic isocyanates 159
 - with succinimidyl esters 158
- 2'-amino-modified RNA postsynthetic labeling 161–164
- aminoglycosides 493–494
 - as functional and structural probes 401–403
- amino-propyltriethoxy silane (APTES) 534
- analytical crosslinking 246–247
- anion-exchange. *See* chromatography
- antibodies testing, for immunoprecipitation 882
- antifading agents 510–520
- antipeptide antibodies generation 958–962
- antisense (AS)
 - accessible sites in target RNA/RNP, by SELEX 1165–1166, 1169–1170, 1180–1181
 - binders, in SELEX 1100
 - *cis*-AS RNAs 794
 - kinetics of AS RNA:target RNA pairing
 - low-abundance *cis*-antisense transcripts to protein-coding genes 728

- antisense (AS) (*contd.*)
 - ncRNA sequence alignment 789
 - Northern blot probes 729, 747
 - oligonucleotide, phosphorothioate-modified 1246
 - primer extension on RNA 774–776
 - primer PCR 38–39, 755, 757–758, 760–761, 763, 768, 772, 1173–1174, 1177–1178
 - regulation by sRNAs 739, 743
 - – RybB-*ompC* interaction 767
 - RNA elements favorable for AS pairing 740
 - rRNA capture by AS oligonucleotides 722
 - sRNA-mRNA interaction 719–720
 - strand, e-shRNA, 1230
 - strand, lhRNA 1226–1227, 1230
 - strand, shRNA 1225–1227
 - strand, siRNA 1243–1245, 1249–1251, 1253–1255, 1263–1264
 - strand of genome, microarrays 720
 - strand transcription 1169
 - to miRNAs, oligonucleotides 837
 - to protein-coding genes, long ncRNAs 814
 - to RNA/RNP, biotinylated oligonucleotides 178–180, 936–948, 957
 - – displacement DNA oligonucleotide 940, 945–948
 - true and false AS transcripts 795
- APART (Automated Pipeline for Analysis of RNA Transcripts) 1194–1196
- aptamers 735–736
- aptamer selection, against biological macromolecules 1097–1098
 - general strategy 1098–1100
 - – carbohydrate targets 1101–1102
 - – library design and preparation 1103–1104
 - – protein targets 1100–1101
 - – selection arrays 1103
 - – suitable target choosing 1100–1102
 - – target immobilization 1102–1103
 - *in vitro* selection cycle 1104–1106
 - protocols 1108–1131
 - selection outcome analysis 1106
- 3D architectures 667
 - S2S and Assemble2 668–673
 - – installation 673–684
 - – semi-automatic architectures 672–673
- aromatic isothiocyanates 158–159
- arylazide photocrosslinking agent attachment, to 5'-terminal phosphorothioate 236–238
 - 3'-addition of arylazide photocrosslinking agent 238–240
- ASH1 gene 1080, 1082
- Assemble2 671–672
 - S2S 668–671
 - – installation 673–684
 - – semi-automatic architectures 672–673
- atomic force microscopy (AFM) imaging 527
 - example protocol 537–538
 - force spectroscopy 540–543
 - of RNA structures
 - – experiment example for salt-dependent RNA folding 535–537
 - – general preconditions 528–531
 - – imaging in liquid 535
 - – surface preparation conditions 531, 533–535
- b**
- bacterial aging 743
- bait 1067–1068, 1073, 1078, 1188
- barcoded cDNA libraries for miRNA profiling, by next-generation sequencing 861–862
 - method overview 862–863
 - – materials 863–866
 - – procedure 866–872
- base-labile protecting groups deprotection 137
- BED sequence alignment format 823, 826
- bimolecular transitions 452
 - absorbance parameters temperature dependence 455
 - basic equations about melting curves 454
 - entropic considerations 452–453
 - fitting or shape method 456–457
 - half-width method 456
 - higher order transitions 455
 - results 457–458
 - slope method 455–456
- binning 505–506
- 3'-biotinylated deoxyoligonucleotide 178–179
- biochemical studies and RNP affinity selection 941–945
- biotin attachment 120–121
- biotin-based affinity purification of RNA-protein complexes 935–937
 - affinity matrices 937–938
 - biotinylated probes 937
 - buffers and solutions 938–939
 - cell extracts 938
 - methods
 - – nuclear extract depletion 951–952
 - – RNA-protein complexes affinity purification 939–948

- specific RNA binding proteins affinity purification 948–951
- BLASTN program 598, 600–602, 623–624, 736, 808
- BLAST program 596, 598, 620, 625, 627, 726, 788–789
- Boltzmann weighting 433
- BOWTIE sequence read alignment program 823, 826, 1196–1197
- breathing 449
- burst mode (smFRET) 500
- BWA sequence read alignment program 823, 826
- C**
- calf intestinal alkaline phosphatase (CIAP) 56–57, 172
- CAN1 gene 1074
- cantilever 528–529, 535, 538, 540–542
- capillary electrophoresis (CE) 206–207, 220–221, 265
- carbohydrate targets 1101–1102
- CASP (Critical Assessment of techniques to protein Structure Prediction) 702
- catalytic RNA 29, 33, 261, 304, 308
- cDNA library of ncRNAs. *See also* barcoded cDNA libraries for miRNA profiling, by nex-generation sequencing
- computational analysis of ncRNA sequences 811, 816
- construction and screening 804–808
- exclusion of cDNAs for abundant known RNA species 808
- Northern blot analysis 808
- sequence analysis 807–808
- specialized library 808–811
- RNA sample enriched by immunoprecipitation 809
- RT-PCR procedure with specialized primers 809–811
- PAR-CLIP 882–883
- preparation from recovered RNA fragments 889–893
- RNP-derived construction 814
- RNA preparation and construction of library 815–816
- RNP extracts and glycerol gradient centrifugation preparation 814–815
- cell lysate/extract
- affinity purification of RNA binding protein with immobilized RNA 735
- affinity purification of snRNPs and CA-repeat RNA binding factor with biotinylated antisense oligonucleotides 940–952
- β -galactosidase activity assay 1078
- crosslinking of RNA 241
- endogenous RNase activity 902
- EMSA 981
- immunoprecipitation/immunoaffinity chromatography of RNA binding proteins and RNPs 805, 882, 884, 986, 988–990, 1020–1021, 1090, 1093
- formaldehyde-crosslinked 1017–1018, 1021–1022
- isolation of proteins associated with *in vivo* expressed tagged RNAs 736, 963, 966, 969
- magnetic beads for RNP affinity purification 938
- preparation 938
- ribosome-free (S100) 176
- cell-penetrating peptides (CPPs) 1260
- CHARMM (Chemistry at HARvard Molecular Mechanics), MD simulation 688, 695
- chemical RNA synthesis, purification, and analysis 129–132
- deprotection 136–138
- purification 138–139
- anion-exchange HPLC purification 139–140
- desalting by HPLC 142–143
- trityl-on RNA detritylation 142
- trityl-on RNA reversed-phase HPLC purification 140–142
- purified RNA analysis 143–144
- RNA solid-phase synthesis 132–136
- automated RNA synthesis 136
- manual RNA synthesis 134–136
- chemical sequencing 187–189
- chromatography
- affinity 921, 1191
- biotinylated antisense 2'-O-methyl RNA oligonucleotides to select RNPs *via* immobilization to streptavidin agarose 941–943, 976
- 3'-biotinylated DNA oligonucleotides to affinity-purify RNAs *via* immobilization to streptavidin agarose 178–180
- Cibacron Blue, HighTrap Blue 927–931
- heparin 927–931
- immunoaffinity using anti-peptide antibodies specific to snRNPs 957–962, 976

- chromatography (*contd.*)
- – immunoaffinity using 5'-cap-specific antibodies 986
 - – Ni-NTA agarose, proteins 1110–1111
 - – purification of RNA binding proteins (depletion of cell extracts) *via* biotinylated RNA and streptavidin agarose 948–952
 - – RNA aptamer-based affinity purification 963–971, 976, 986–993, 1006, 1008
 - – SELEX, covalent target immobilization 1140–1146
 - – SELEX, non-covalent target immobilization, biotinylated target captured *via* (strept)avidin matrix 1147
 - – tRNA-sepharose 927–931
 - fast-performance liquid chromatography (FPLC) 21–23, 107, 112, 482, 976
 - flash column (CH₂Cl₂), synthesis of aromatic isothiocyanates and aliphatic isocyanates 160
 - gel (exclusion, permeation) chromatography, gel filtration 114–115, 121, 213, 320, 509, 927–929, 1110
 - – NAP columns 19, 80, 142, 146, 512
 - – Sephadex G-25 19, 142, 744, 1046, 1091
 - – Sephadex G-50 122, 323, 350, 403, 537, 1046
 - – Sephadex G-75 122
 - – spin column 122, 323, 350, 402–403, 744, 1046, 1091
 - – Superdex 200 930–931
 - ion-exchange 235, 921, 930, 932, 1110
 - – weak anion-exchange, diethylaminoethyl (DEAE) sepharose 18, 105, 107, 111–115
 - – DEAE, snRNP enrichment 938, 941, 945
 - – strong anion-exchange 138–140, 928
 - – strong cation exchange 22–23, 930–931
 - – MonoQ 213, 976
 - isocratic competitive affinity chromatography 1161–1162
 - high-performance liquid chromatography (HPLC) 18, 132, 138–147, 162–166, 195, 290–291, 361, 471, 948, 1098, 1108, 1110, 1143
 - – Dowex 50 cation-exchange, conversion of sodium to ammonium salt of RNA 146–147
 - – oligoribonucleotide desalting with Sephadex G-25 142–143
 - – separation of peptides 929–930
 - HPLC, anion-exchange 132, 138–140, 142–144, 146
 - – DEAE 235–237
 - – Dionex 471, 482–483
 - HPLC, reversed-phase 132, 138, 140–145, 147, 162, 164–165, 512
 - – analysis of ribonucleosides 893–896
 - – of trityl-on RNA 138, 140–142
 - – phosphorothioate stereoisomers 290–291
 - ion-pair reversed-phase liquid chromatography (IP-RPLC) 142–143, 213
 - thin-layer (TLC)
 - – cellulose acetate strips/DEAE cellulose 175, 190–194
 - – cellulose 195–197, 200–201
 - – polyethyleneimine (PEI) 196, 201, 236
 - ultra performance liquid chromatography (UPLC) 143–144
 - circular potato spindle tuber viroid (cPSTVd) 434–435
 - cleavage reactions by Tb³⁺ 256–258
 - click reaction 509
 - between alkyne-containing RNA oligonucleotide and azide-functionalized fluorophore 511–512
 - CLIP (crosslinking-immunoprecipitation) 899, 1088, 1199
 - cloverleaf structure 76
 - CLUSTALW program 598, 620
 - COLORES program (SAXS) 418–419
 - comparative genomics 836–837
 - concatenation method 737
 - concentration calculation, of RNA 60, 1279–1281
 - Cordycepin 182–183, 337–339. *See* radioactive labeling methods
 - Coulombic field 319, 322
 - coupling chemistries 119, 121–122
 - covalent coupling of antibodies, to protein A sepharose 1047–1048
 - covalent immobilization
 - epoxy-activated matrices 1143–1145
 - NHS-activated matrices 1145
 - pyridyl disulfide-activated matrices 1146
 - covariance model (CM). *See* INFERNAL
 - cross-correlation function (CCF) 419
 - crosslinking 8, 50, 1199. *See also* PAR-CLIP
 - RNA-RNA crosslinking
 - – disulfide crosslinking 156–160
 - – photocrosslinking for probing of RNA structure 231–233, 236, 238–244, 246–249
 - RNA-matrix crosslinking 1048

- EDC [1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide] crosslinking 89, 91–92, 95–97, 100, 1091, 1094
- RNA-protein crosslinking 736, 977
- formaldehyde crosslinking 1017–1019, 1021, 1026
- iCLIP 899–902, 904, 908
- UV 91, 363, 746, 883, 899–901, 905, 999, 1003–1005, 1008–1009, 1026, 1029, 1030–1038, 1040–1042, 1191
- crude transcription reaction 105, 107, 114
- cryo-electron microscopy (cryo-EM) structures 690
- crystallization of RNA 481–482
 - assays evaluation 488–489
 - complexes with organic ligands 493–494
 - optimization process 489, 491
 - purification 482
 - gel electrophoresis 483–484
 - HPLC 482–483
 - recovery 484
 - renaturation 485
 - RNA constructs designing, with improved crystallization capabilities 491–493
 - search for conditions 485–488
- CURVES+ program (conformational analysis of nucleic acids) 688, 703
- d**
- DAMAVAR/DAMFILT/DAMMIF/DAMMIN programs (SAXS analysis) 416–417
- DAP Goldstar® Polymerase (Eurogentec) 1120
- Dashboard 633, 636–639, 641, 650, 652, 657, 659, 661–663
- Debye–Hückel approximation 462
- decoy 636, 659–661, 1082, 1235
- deep sequencing. *See* next-generation sequencing (NGS)
- dephosphorylation protocols 180
 - DNA 5′-ends 110
 - ribonucleoside preparation for HPLC analysis 895
 - RNA 5′-ends 64, 181, 337–338, 350, 885, 980
 - RNA 3′-ends 39–40, 56–57, 121, 909
- deprotection 136–138
- desalting, by HPLC 142–143
- detritylation 132, 138, 141–142, 145
- Diels-Alderase ribozyme 501
- diethylaminoethyl (DEAE) 107, 112–114, 176, 190–192, 235, 938, 941–943, 945, 1127
- diethyl pyrocarbonate (DEPC). *See also* probing RNA structure
 - RNase-free water preparation 20, 59, 305, 309–310, 313, 744, 803, 1107, 1173
- differential melting curve (DMC) 448, 450, 451–452, 454, 456–457, 471–474
- differential scanning calorimetry (DSC) 453–454, 457–458, 462, 472
- diffusion elution 19
- 5′-digoxigenin (DIG)-labeled probes 89, 91, 946,
 - detection 95, 101–102
 - DIG Easy Hyb Granules 97, 101,
 - probe generation by T7 transcription using DIG-11-UTP 89, 92, 100–101
- dimethylsulfate (DMS) modification. *See* probing RNA structure
- divalent metal ion binding sites, in nucleic acid-metabolizing enzymes 397–398
 - probing 398
 - aminoglycosides as functional and structural probes 401–403
 - Fe(II)-mediated hydroxyl radical cleavage 398–399
 - mapping 399–401
- DNAMAN sequence analysis program 811
- DNase treatment 18, 53, 59–60, 67, 80, 83, 213, 278, 745, 752, 769, 908, 980, 1009, 1111–1112, 1153, 1167–1168, 1172, 1176, 1179, 1190, 1216–1217
- 3DNA program (nucleic acid structure analysis, rebuilding, and visualization) 688, 703
- DNA splint 49–52, 57–59, 66–70, 75, 78, 80, 82–83, 286, 291–293, 338, 362
- DNazyme 35, 380, 1166
- donor substrate. *See* ligation
- dot plot folding comparisons 565
- double-stranded RNA (dsRNA) 527, 530, 534, 537, 1221, 1254, 1256–1257
- dRNA-seq approach 722–724, 732
- dynamic programming 222, 552, 554–555, 572, 586, 604
- e**
- EDC crosslinking. *See* crosslinking
- Edman degradation 929
- electroelution 19, 41, 223, 886, 888, 897, 1178
- electron paramagnetic resonance (EPR) 319
 - active probes 156
 - spin labeling reagent 4-isocyanato TEMPO 157
 - spin probe/label TEMPO 159

- electrophoretic mobility shift assay (EMSA)
 - 924, 976–978, 981–985, 996, 998, 1005, 1010, 1067, 1103, 1106, 1124–1128, 1199
 - method 977–978, 981–986
 - protein identification in RNP 984–986
 - supershift method 983–984
- Elutrap[®] 19
- emacs (text editor system) 621, 625
- EMSA. *See* electrophoretic mobility shift assay
- end heterogeneity of RNA 6, 29, 45, 54, 57, 84, 121, 223, 338
- end-labeling of RNA 257, 346, 1130. *See also* radioactive labeling methods
 - at 3'-end 33, 73, 181–182, 213, 378–379, 382–383, 385, 389, 980
 - – with fluorescent dye 121, 509
 - – with biotin 119–121
 - at 5'-end 5–7, 10–11, 32, 92, 180–181, 213, 380–381, 383, 389, 774, 883, 886, 980, 995
 - – with fluorescent 2-aminopurine 73
 - – with 3'-amino ATP (for biotinylation) 73
 - – with 5'-biotin-GMP 7
 - – with GMPs 7, 233–234
 - – with 6-thio-GMP 7, 235, 242
- 3'-ends 6, 33, 53–54, 182–183
 - addition of arylazide photocrosslinking agent 238–240
 - generation by autocatalytic ribozyme cassettes
 - – cassettes construction 36–37
 - – *cis*-cleaving 30–31
 - – dephosphorylation protocols 39–40
 - – PCR protocols 37–39
 - – removal of 2',3' cyclic phosphate 32–33
 - – *trans*-cleaving 33
 - generation by RNase P 33–35
 - – protocol for RNase P cleavage 40–41
 - RNase H approach to generate homogeneous transcript 3'-ends 54–56
- 5' ends 5–6, 35, 183–184, 195, 199
 - generation by *cis*-cleaving autocatalytic ribozyme cassettes 30–31
 - – cassettes construction 36–37
 - modification by transcription priming 6–7, 11–12, 17–18, 51, 61, 233–234
 - – 5'-monophosphorylation of nucleosides 234–236
 - PCR protocols 37–39
 - radioactive labeling, of RNA termini 64–65, 180–183, 337–339, 909–910
- endogenous RNAs associated with RNA binding proteins, immunopurification 1017
 - methods 1017–1022
 - microarray analysis of immunopurified RNA 1022–1024
 - – next-generation sequencing of immunopurified RNA 1025
 - – RT-PCR analysis of immunopurified RNA 1024
 - – critical points and common problems 1025–1026
 - – microarray data analysis 1026–1027
 - – uncrosslinked or crosslinked RNA immunoprecipitation 1026
- endothermic process 451
- energy dot plot 558–561, 563, 567
- enhanced green fluorescent protein (eGFP) 1234
- ENSEMBL database and genome browser 620, 623–624, 629, 816, 851
 - bacterial genome browser 602
- entropic considerations 452–453
- enzymatic hydrolysis 184–187, 191, 224, 226, 895
- enzymatic probing 205–206, 214–215, 224, 243, 435, 462, 602, 634–635, 998
- enzymatics 1150–1151
 - *in vitro* transcription 1153
 - polymerase chain reaction 1152–1153
 - reverse transcription 1151–1152
- epoxy-activated matrices 1143–1145
- EPR. *See* electron paramagnetic resonance
- equilibrium dialysis 320, 323–325, 1159–1160
- equilibrium filtration analysis 1160–1161
- ERPIN program (RNA motif identification) 597–599, 609–616
- ethanol (EtOH) precipitation
 - protocols for DNA 1176, 1215, 1218
 - protocols for RNA 19, 61–62, 98, 120, 142, 162, 177, 213–214, 219–221, 223, 226, 248, 258, 278, 306, 309–310, 336, 350, 355, 381–382, 385, 752, 769, 775–776, 806, 810, 867–868, 871–872, 888, 890, 897, 944, 962, 971, 998, 1021, 1044, 1060–1061, 1064–1065, 1114, 1217
- EUROSCARF (*Saccharomyces cerevisiae* database, strain and plasmid collection) 1030
- EvoFold (RNA structure identification in multiple-sequence alignments) 572, 788–789
- exonic splicing enhancers (ESEs) 1207–1208, 1210, 1212
- exonic splicing silencers (ESSs) 1207–1208

exothermic process 451
 extended shRNA (e-shRNA) 1224,
 1226–1227, 1229–1230, 1233

f

FASTA (format) 556–557, 566, 598,
 600–602, 608, 611–612, 616, 623, 625, 650,
 674–675, 678–679, 681
 fast-performance liquid chromatography
 (FPLC). *See* chromatography
 Fe(II)-mediated hydroxyl radical cleavage
 301–312, 397–401. *See also* Fenton
 reaction/cleavage
 FeBABE reagent 302
 Fenton chemistry 308, 314
 Fenton cleavage
 – mapping divalent metal ion binding sites in
 proteins 397–401
 – – methods description 402–404
 – – aminoglycosides as functional and
 structural probes 401–402
 – of RNA 301, 303, 310, 312–313
 – – methods description 305–312
 – – drug-directed (tetracycline) 310
 Fenton footprinting 304
 Fenton reaction 301, 306, 398
 flow cytometry 738, 756–757, 764
 – protocol 764–766
 fluorescence resonance energy transfer
 (FRET) 50, 66, 68, 263, 499–503
 – dye pairs 504
 – – choice of 507–508
 – initial spectrometric characterization by
 bulk measurements 513–514
 – smFRET. *See* single molecule fluorescence
 resonance energy transfer
 fluorescent labeling 121, 123, 796–797
 5-fluoroorotic acid (5-FOA) 1074, 1081
 footprinting 207, 214, 219, 225, 255–257,
 259–261, 302, 304, 636, 660, 662, 739,
 977–978, 981, 991, 1067, 1106, 1199.
See also terbium(III) footprinting
 – semi-automated footprinting analysis
 (SAFA) software 221
 forced dialysis 320–324, 326
 force spectroscopy AFM, of RNA 527,
 540–543
 Förster radius 502–504, 507
 Freiburg RNA Tools 620
 FRET. *See* fluorescence resonance energy
 transfer
 Friedel law 472
 fusion protein (FP)
 – MS2-MBP 957, 966, 969, 989–990, 992

– in the yeast three-hybrid system 1067
 – – FP1 with DNA binding domain
 1068–1070
 – – FP2 with transcriptional activation
 domain 1068–1070
 – – FP1, LexA fused to MS2 coat protein
 1069, 1073
 – – FP2, activation domain of yeast Gal4 fused
 to RNA binding tester protein
 1069–1070, 1073–1074

g

GAL4 activation 1069–1070, 1073–1075,
 1077–1078
 β -galactosidase 738, 1076, 1078–1079, 1081
 – filter assay 1079
 gel electrophoresis, of DNA 220, 889, 902,
 916, 1108, 1120, 1153
 gel electrophoresis, of RNA 722, 1192
 – denaturing polyacrylamide (denaturing
 PAGE) 6, 15, 18–20, 29, 31, 39, 41, 53,
 59, 61–64, 67, 69–70, 80–84, 92, 105,
 113–114, 123–124, 142–143, 162–164,
 183, 207, 212–213, 216, 223, 226,
 244–245, 248, 257, 261, 292–294, 306,
 324, 372, 377, 379–384, 389, 392, 422,
 471, 483–484, 495, 520–521, 537, 740,
 768, 770, 774, 862, 899–901, 942, 944,
 961, 967–968, 980, 992, 1005, 1036, 1047,
 1059–1060, 1063–1064, 1112, 1123–1124,
 1130, 1153, 1187
 – native polyacrylamide
 (native/non-denaturing PAGE) 18, 31,
 99, 224, 344, 352, 481, 948, 982–986,
 1123–1126, 1181
 – temperature-gradient (TGGE) 427–433,
 462
 – two-dimensional 176, 192–193
 gel exclusion (permeation) chromatography,
 gel filtration 114–115, 121, 213, 320, 509,
 927–929, 1110
 – NAP columns 19, 80, 142, 146, 512
 – Sephadex G-25 19, 142, 744, 1046, 1091
 – Sephadex G-50 122, 323, 350, 403, 537,
 1046
 – Sephadex G-75 122
 – spin column 122, 323, 350, 402–403, 744,
 1046, 1091
 – Superdex 200 930–931
 gel fractionation 221, 996
 gel purification, of RNA 15, 21, 52, 56, 65,
 122–123, 242–243, 247, 257, 331, 361, 379,
 482, 740, 770, 914, 969, 1176
 – preparative 84

- gel purification, of RNA (*contd.*)
 - kit 751, 880
- gel running buffer
 - 1 x TBE 19
 - 5 x TBE 59
- GenBank 433, 556, 560, 563, 566, 569, 571, 595, 619, 622, 624, 808, 821
- gene expression profiling computational methods. *See* next-generation sequencing (NGS)
- general buffers and procedures (GBP) 59
- gene silencing methods 1221
 - background information 1221–1223
 - extended shRNA (e-shRNA) and long hairpin RNAs (lhRNAs), construction 1229–1230
 - lentiviral vectors, production 1230–1234
 - miRNA vectors, construction 1228–1229
 - shRNA vectors, construction 1223–1228
- genome(s) 625–628, 667, 719–720, 731, 736, 787–791, 794, 811, 822, 829, 1165–1166, 1168, 1185–1186, 1193, 1202
 - annotation 628, 787, 794, 1197, 1223
 - APART pipeline 1196–1198
 - *Arabidopsis thaliana* 929
 - assembly 623–624, 628, 829
 - avian 626, 628
 - bacterial 595, 600, 602, 616, 726–727, 787, 790–791, 1187
 - BLAST search 623
 - browser 602, 620, 623, 625, 628, 722–724, 823, 826, 829
 - – Affimetrix integrated 724
 - – UCSC 620, 625, 823, 826, 829
 - comparative analysis 721, 787–788
 - coverage 1188
 - coverage by tiling microarray probes 795–796
 - database 596, 599, 608, 614, 624, 726
 - encoded aptamers and regulatory sequences 1185, 1199
 - eukaryotic projects 620
 - gene copies 622
 - HIV-1 206
 - human, mammalian 623–625, 627–628, 801, 849, 864, 877, 1168, 1170, 1187
 - Illumina Genome Analyzer 902, 904, 916
 - invertebrate 623
 - lizard 625
 - low complexity region 598
 - mitochondrial 583
 - mRNA profiling on microarrays, genome-wide 729, 733
 - ncRNA prediction, genome-wide 787–791
 - NGS read alignment 826–827
 - of lentiviral vector 1232–1233
 - organellar 620
 - phage T4 46
 - phage T7 4
 - probe complementarity 93
 - PSTVd 434
 - (nc)RNA cDNA read mapping 816, 916, 1195–1197
 - SELEX (genomic) 1185
 - sequence(s) 175, 595, 604, 624, 726, 1185
 - sequencing 1081
 - shuffled database 609
 - specific primer 1187
 - transcriptome analysis 732–733, 825–826
 - vertebrate 620, 623
 - wallaby 627
 - whole genome oligo microarrays, mouse 1023
 - yeast 1069, 1187
- genome mapping, of cDNA reads 723, 1195–1197
- genome screening, for RNA genes/motifs 595–596
 - search for bacterial sRNA homologs 600
 - – alignment and structure prediction 602–604
 - – BLASTN 600–602
 - – ERPIN 609–614
 - – HMMER 604–605
 - – INFERNAL 614–615
 - – RNAMOTIF 606–609
 - RNA search procedure overview 597–598
 - search program choice 596–597
 - search specificity assessment 598–600
- genomic aptamer(s) 1185–1186, 1188, 1192, 1194, 1196
 - biochemical analysis 1199–1202
 - – RNA-protein interaction validation 1199
 - expression analysis 1199–1200
 - size of native transcript 1200
 - function of RNA-protein interaction 1200
- genomic SELEX 720, 728, 1165–1166, 1171, 1185–1186
 - bait choice 1188
 - computational analysis of sequences 1194–1198
 - library construction 1186–1188
 - procedure 1188–1194
 - specialized SELEX method to identify antisense and protein target sites in RNA or hnRNPs 1165–1166

- generation of RNA 20-mer library using *Mme* I 1166–1169
 - natural RNA substrates for proteins, identification 1171
 - optimal antisense sites identification 1169–1170
 - procedure and protocols 1171–1182
 - glyoxy-oxygen-scavenging enzyme stock 515–516
 - glycerol gradient (centrifugation) 802, 813–816, 940, 948, 960–962, 964, 966–968, 970, 976, 986, 1056–1061
 - graph theory 552
 - green fluorescent protein (GFP) 732, 1228–1229
 - affinity tag 1017–1019
 - anti-GFP antibody 768–769, 771, 1018, 1021
 - enhanced GFP (eGFP) 1234
 - flow cytometry 764–767
 - GFP-based reporter system for sRNA target validation 755–757
 - lentiviral vector encoding Emerald Green Fluorescent Protein (EmGFP) 1228–1229
 - read out approaches 763–764
 - translational GFP fusion, cloning 759
 - pXG-10 759–761
 - pXG-20 761–762
 - pXG-30 762–763
 - GROMACS package, MD simulation 688, 695
 - GROMOS (GROningen MOlecular Simulation computer program package) 688, 695
 - GSNAP sequence read alignment program 823, 826
 - guanosine 5'-monophosphorothioate (GMPS) 7–8, 233–238
 - guide strand. *See* small interfering RNA (siRNA)
 - Gwyddion software (AFM) 530
- h**
- hammerhead ribozyme 6, 30–33, 36–39, 75, 79–80, 121, 123–124, 152, 158, 264, 286–287, 337, 374, 500
 - HaMMY software package, smFRET analysis 507
 - hepatitis delta virus (HDV) ribozyme 6, 30–39, 261–264, 337
 - heterogeneous nuclear ribonucleoprotein (hnRNP) 900, 950–952, 975–976, 984–986, 998–1000, 1003–1007, 1055, 1072, 1165–1182
 - Hfq 719–720, 725, 728–729, 731, 734–738, 740–743, 770–771, 1197
 - Hfq coimmunoprecipitation. *See* immunoprecipitation
 - Hidden Markov Model (HMM) 507, 604–605
 - high-performance liquid chromatography (HPLC). *See* chromatography
 - high-throughput RNomics 794–799
 - high-throughput sequencing (HTS) 89, 206, 720–723, 742, 802, 812, 877, 899, 901–902, 904, 916, 1106, 1186, 1194. *See also* RNA-seq
 - HTSeq, high-throughput sequencing data analysis software 823, 829
 - Hill equation 297
 - HIS3* gene 1069, 1075, 1078, 1080
 - histidine-tag (His-tag) 56, 372–373, 865, 1030, 1050, 1128–1129
 - HMMER program (search for RNA homologs) 596–599, 604–605, 614–616
 - holoenzyme 289, 292, 294, 326, 375, 387, 492
 - reaction 294
 - homogeneous 5'- and 3'-ends. *See* T7 RNA polymerase
 - homology-based programs 596–597
 - homology search 596–597, 602, 604, 811
 - for small structured ncRNAs 619–621
 - mascRNA example 621–629
 - initial BLAST search 622–624
 - initial secondary structure model 624–625
 - structure-based searches 625–629
 - web-service-independent software 621
 - web services 620
 - RsaE RNA test case 600
 - alignment and secondary structure prediction using LOCARNA 602–604
 - BLASTN first training set 600–602
 - descriptor- vs. homology-based programs 596–597
 - RNA motif search specificity and protocol 598–600
 - RNA motif search with sequence and secondary structure constraints 595
 - searching with ERPIN 609–614
 - searching with INFERNAL, covariance models 614–615
 - searching with HMMER 604–605
 - searching with RNAMOTIF 606–609
 - program versions and download sites 616
 - Hooke's law 528

- hot phenol method 90, 94
 - RNA extraction from cells 98
- HQS. *See* 8-hydroxyquinoline-5-sulfonic acid
- HTSeq, high-throughput sequencing data analysis software 823, 829
- hybridization 117, 352, 527, 555, 648, 650, 836, 871, 936, 1000, 1018–1019, 1248
 - cross-hybridization 827, 861
 - experimental RNomics 803, 807–808, 816
 - identification of small non-coding RNAs 725, 729, 737, 744, 746–749, 768
 - immunopurification 1018–1019, 1022–1023
 - microarray hybridization 796–799
 - Northern blot 89–92, 94, 96–102, 794, 1091–1092, 1094
 - oligonucleotides 179, 427, 434–435, 437–438, 952
 - primer extension 220, 223, 226, 280, 310, 995, 1046
 - to DNA splint 58, 67–68
- hybrid RNA 1068–1075, 1077–1081
- hydroxyl radicals 206, 209, 225, 301–302, 304, 313–314, 398, 421, 740, 1007. *See also* Fenton chemistry/cleavage/footprinting/reaction
- 8-hydroxyquinoline-5-sulfonic acid (HQS) 320, 324–325, 327
- hyperchromism 446, 449
- hypochromism 446, 449, 464
- i**
- iCLIP (individual-nucleotide resolution crosslinking and immunoprecipitation. *See* crosslinking)
- ImageJ software 530, 536, 744, 747, 774
- ImageSXM software (AFM) 530
- immunoprecipitation 725, 805, 809, 840, 963, 1008–1009, 1020–1027, 1030, 1037–1038, 1040–1044, 1046–1048, 1171
 - coimmunoprecipitation 720, 1073, 1075, 1119
 - – Hfq coimmunoprecipitation 722–725, 728, 734
 - – RNA coimmunoprecipitation 1017, 1020–1027
 - iCLIP 899–902, 905–909
 - identification of miRNA targets 1087–1095
 - PAR-CLIP 877, 882, 884–887
 - RISC components 838–839
- INFERNAL program (structural RNA sequence alignment based on covariance models) 597–599, 614–616, 621, 626–628
- in-line probing 206, 210–211, 213, 216, 218, 225–226, 417, 1106
- inner-sphere coordination, cation : RNA 319, 370, 373
- interference suppression 330–332, 361, 363, 375, 378
- intergenic region (IGR) 719–720, 722, 724, 726–728, 748, 788, 790–791, 793–796, 1195
 - Intergenic Sequence Inspector (ISI) 726
- intermittent contact mode (tapping), AFM 529
- intron ribozyme activity (group II) 339–344
- in vitro* evolution 442, 1099–1100
- in vitro* probing. *See* probing
- in vitro* selection 16, 442, 1097, 1099–1101, 1104–1107, 1116. *See also* SELEX
 - against proteins and carbohydrates 1097–1131
 - against small targets: 1139–1162
 - selection cycle: 1099, 1104–1106, 1141
- in vivo* SELEX. *See* SELEX
- iodine cleavage (iodine-induced hydrolysis) 330–331, 341–342, 344, 346–347, 354–355, 385–386, 391–393
- ion-exchange chromatography. *See* chromatography
- isocratic competitive affinity chromatography. *See* chromatography
- isopycnic ultracentrifugation 1056, 1061. *See also* ultracentrifugation
- isopropanol precipitation
 - of plasmid DNA 110–111
 - of RNA 99, 752, 1157–1158, 1180, 1217
- isothermal titration calorimetry (ITC) 453
- k**
- Klenow polymerase (Klenow Pol/Klenow fragment/large fragment of *E. coli* DNA polymerase I)
 - 3′-³²P-end-labeling of RNA 337–338, 341
 - Fe(II)-mediated hydroxyl radical cleavage 397, 399–403
 - fill-in reaction, *in vitro* selection/SELEX 1149–1150, 1168, 1171, 1174–1176
 - genomic SELEX 1186–1188, 1193
- K-Means Clustering 661–662
- knockdown (KD) 837, 839, 851, 902, 908, 911, 922, 1200
- gene silencing 1221, 1223–1224, 1228
- chemical modifications to enhance siRNA performance 1243, 1247, 1252–1253, 1260, 1262
- “K-turn” (structure/motif) 501, 671, 975, 982

I

- 3'-labeling. *See* end-labeling of RNA
 - 5'-labeling. *See* end-labeling of RNA
 - lacZ* (gene) 735, 738–739, 756, 762, 1068–1070, 1075
 - lanthanide (ion) 255
 - lentivirus/lentiviral
 - vector production 1230–1235
 - LEU2* gene 1070, 1073, 1080–1081
 - Lewis
 - acid 270, 286, 370
 - base 286, 370
 - lhRNA. *See* long hairpin RNA
 - ligand coupling (*in vitro* selection) 1144–1146
 - ligation of RNA 6, 29, 33, 45–84, 152, 521, 801
 - acceptor substrate 46, 49, 70–73, 75–76, 78–80
 - donor substrate 46, 49, 70–73, 75, 78–80
 - ligation efficiency 46, 54, 70, 72, 75–76, 78, 82, 1169, 1172, 1179, 1214, 1216
 - ligation site accessibility 78
 - reaction mechanism 46, 48–49
 - simultaneous splint ligation of RNA fragments for FRET experiments 66–70
 - T4 DNA ligase (T4 Dnl) 47, 49–66. *See also* T4 DNA ligase
 - dephosphorylation and phosphorylation 56–57
 - homogeneous acceptor 3'-ends, generation 53–54
 - large-scale transcription and purification 53
 - RNA ligation 57–58
 - RNA ligation substrates generated with RNase H 54–57, 63–64
 - T4 polynucleotide ligase 46–47
 - T4 RNA ligase(s) 47, 70–84. *See also* T4 RNA ligase(s)
 - LOCARNA program (alignment and folding of RNA sequences) 573, 598–599, 602–604, 616, 620, 624–625
 - Locomotif tool (abstract shape analysis) 583
 - long hairpin RNA (lhRNA) 1224, 1226–1227, 1229–1230
- m**
- magnesium (Mg^{2+}) ions bound to RNA, quantification 319–320
 - Mg^{2+} complexes with 8-hydroxyquinoline-5-sulfonic acid (HQS) 320
 - separation of free Mg^{2+} from RNA-bound Mg^{2+} 320–324
 - equilibrium dialysis 323–324
 - forced dialysis 321–323
 - size exclusion chromatography 323–324
 - Mg^{2+} concentration determination using HQS 324–326
 - affinity and stoichiometry of Mg^{2+} binding 322–323
 - RNase P holoenzyme 326
 - magnetic bead(s)
 - aptamer selection 1099, 1116, 1120–1121
 - capture of rRNA using antisense oligonucleotides 722
 - immunoprecipitation
 - PAR-CLIP 882–886
 - iCLIP 900
 - miRNA targets 1088–1089, 1094
 - streptavidin-coated 937–938
 - MALAT-1 RNA/transcript 622, 625, 628
 - maltose-binding protein (MBP) 293, 736, 935, 957, 963, 966–970, 976, 987–993, 1006, 1008–1009
 - masCRNAs 621–629
 - BLAST search 623–624
 - secondary structure model 624–625
 - structure-based searches 625–629
 - mate-pair reads, next-generation sequencing 825
 - MC-Tools, RNA 2D and 3D structure prediction 633–634
 - equipment and input 634
 - MC-Cons 636, 651–654
 - MC-fold 635, 637–651
 - MC-Sym 636, 654, 657–661, 663
 - menRNA 627–629
 - metal ion coordination interactions with RNA
 - identification and characterization 285–286
 - thiophilic metal ion rescue of RNA phosphorothioate modifications 286–290
 - purification of phosphorothioate stereoisomers by RP-HPLC 290–291
 - phosphorothioates incorporation into RNA, techniques 291–293
 - kinetic analysis of thiophilic metal ion rescue 293–297
 - NAIM/NAIS analyses 329–332, 370–374
 - metal ion-induced cleavage/hydrolysis 256, 270–271, 275, 1104. *See also* probing
 - metalloenzyme 287, 397–398, 401
 - MICROBExpress kit 722

- Micro Bio-Spin[®] chromatography column 122
- microRNA (miRNA) targets 833–836.
 - See also* barcoded cDNA libraries for miRNA profiling
 - comparative genomics 836–837
 - experimental identification 1087–1088
 - Ago2 immunoprecipitation, Western blot analysis 1090
 - cell culture and transfection 1089
 - gene expression array analysis of Ago2-associated mRNAs 1092–1093
 - Northern blot analysis for miRNA detection 1091–1092
 - qRT-PCR of Ago2-associated mRNAs 1092–1093
 - RNA precipitation 1090–1091
 - sample preparation and immunoprecipitation 1089–1090
 - functional binding site properties
 - data set type, quality and reproducibility 839–840
 - evolutionary conservation 841
 - miRNA-mRNA duplex stability 841
 - seed binding criterion 840–841
 - sequence composition 842
 - spatial effects 842–843
 - structural accessibility 841–842
 - targeting prediction 843–847
 - perturbation and omics 837–838
 - RISC components, immunoprecipitation 838–839
 - translation repression inferred from polysome profiles 839
- microRNA (miRNA) expression vector for gene silencing 1221–1223
- miRNA polycistron 1224
- vector construction 1226–1229
- minimum free energy (MFE), definition 549–555
- Mme* I restriction enzyme/site 1167–1170, 1173–1174
 - cleavage protocol 1178
 - digestion 1170, 1172
- mobility shift analysis
 - terminal RNA sequence identification 184, 190–194
 - electrophoretic mobility shift assay (EMSA) 166, 540, 899, 924, 976, 1067, 1103, 1124–1126, 1199
- modified nucleotides
 - incorporation by phosphoramidite chemistry 153–154, 359, 361
 - T7 transcription 6–8, 16–18, 61, 231–234
 - in NAIM/NAIS 333–336, 371–372
- modified nucleotides determination, by postlabeling methods 194–201
- position and identity determination 198–201
- total nucleotide content analysis 195–198
- modified nucleotide incorporation techniques 232–233
- modified RNAs 151–152
 - 2'-amino approach 156–158
 - 2'-amino groups reaction
 - with aliphatic isocyanates 159
 - with aromatic isothiocyanates 158–159
 - with succinimidyl esters 158
 - experimental protocols 159–166
 - 2'-amino-modified RNA, postsynthetic labeling 161–164
 - 4-thiouridine-modified RNA, postsynthetic labeling 164
 - aromatic isothiocyanates and aliphatic isocyanates, synthesis 160–161
 - label incorporation verification 164–165
 - phosphoramidite method 152–154
 - postsynthetic labeling strategy 154–156
- molecular barcodes 825
- molecular dynamics (MD) simulations of RNA systems 304, 687–704
 - comparison with experimental data 702
 - conformational parameters 703
 - consistency checks 702
 - convergence issues 703
 - data analysis 704
 - validation based on structural databases 703
 - visualization 702
 - methods 689
 - setups
 - box size 693
 - clustering artifacts and ion parameters 696
 - conformational checks 690–691
 - divalent ions 694
 - duration of simulations 699–701
 - electrostatic interactions treatment 697
 - environment and ion types 693
 - equilibration 699
 - force fields 695
 - hydrogen atoms addition 693
 - initial solute and solvent configurations 694
 - minimal salt conditions 694
 - monovalent ions 693–694
 - multiple molecular dynamics (MMD) simulations 701

- parameterization of modified nucleotides, ligands, and ions 696
- programs 688, 695
- protonation issues 692
- rare non-covalent interactions 691–692
- shake, time steps, and non-bonded pair list update 698
- solvent 692–693
- starting structure selection 689–690
- temperature and pressure 698
- thermodynamic ensemble 697–698
- water models 696–697
- water molecules 694
- mRNA analysis
 - enzymatic and chemical probing 213–227
 - gene-expression-array analysis 1092–1093
 - qRT-PCR analysis 1092
- mRNAs (short) and dual-function RNAs 794
- MS2 aptamer/tag, for the phage MS2 coat protein 735–736, 1071–1073, 1075
- MS2-LexA fusion in three-hybrid analysis 1069–1070, 1080
- MS2-MBP affinity chromatography 736, 935, 957, 963–971, 976, 987–993, 1006, 1008–1009
- multi-FASTA format 598, 600
- multiple molecular dynamics (MMD) simulations 701
- n**
- NAIM. *See* nucleotide analog interference mapping
- NAIS. *See* nucleotide analog interference suppression
- NAMD program, MD simulation 418, 688, 695
- NcDAlign program (genome-wide alignments of non-protein-coding sequences) 788–789
- ncRNAs. *See* non-coding RNAs
- neutravidin 937, 946–947, 951
- next-generation sequencing (NGS). *See also* RNA-seq
 - cDNA library preparation 889–893
 - computational methods for gene expression profiling 821–822
 - association of reads with transcripts 827–828
 - data generation: experimental background information 823–825
 - expression determination and uncertainty 828
 - library generation 823–825
 - normalization 828
 - output and viewing 828–829
 - read alignment 826
 - sequencing 825
 - of immunopurified RNA 1025
 - associated with RBPs 1018
 - sequencing platforms
 - 454 Sequencing™ (Roche Diagnostics) 721–724, 812, 825, 1025
 - Solexa™ (Illumina) 721, 812, 822, 825, 861–862, 864–866, 872, 887, 889, 893, 902, 904, 916, 1025
 - SOLiD™ (ABI) 721, 807–808, 812, 1025
 - NF- κ B transcription factor 1073, 1082
 - N-hydroxysuccinimide (NHS)
 - acetal-PEG-NHS coupling to tip, AFM 542
 - activated sepharose matrices
 - tRNA coupling 922, 924–927, 929–930
 - tobramycin coupling 987–988
 - small ligand coupling for SELEX 1144–1145
 - biotin-NHS
 - for AFM 542
 - for SELEX experiments 1102, 1147
 - maleimide-PEG-NHS, AFM 543
 - esters, for RNA modification 67, 156, 508
 - nitrocellulose (NC)
 - filter assay 1079
 - filter binding 1101–1102, 1116, 1125–1127
 - membrane transfer 1090
 - iCLIP 900, 910–911
 - EMSA coupled to SDS-PAGE and Western blot 984–986
 - NMR (nuclear magnetic resonance spectroscopy)
 - isotopically labeled nucleotides, incorporation by *in vitro* transcription 34, 106
 - spectroscopic analysis 119, 164, 166, 286, 301, 407
 - ^{19}F NMR of nucleic acids 165
 - of mutant RNAs 439
 - RNA analysis and purity 29, 132, 143
 - RNA-metal ion interactions 304
 - structural dynamics of RNA 499
 - structure(s)
 - RNA 211, 263, 503
 - RNA/protein 409
 - in MD simulations of RNA, 688–690, 692
 - non-canonical base pairs 596
 - in NAIM analyses 340
 - in RNA structure prediction 560, 633–635, 645–648

- non-canonical base pairs (*contd.*)
 - in sRNA-mRNA interaction 736
 - in tRNAs 34
 - in UV spectroscopy 446
- non-coding RNAs (ncRNAs) 595, 667, 671, 1197. *See also* “small non-coding RNA (sRNA)” and “homology search”
 - characterization and prediction of miRNA targets 833–834
 - dual-function RNAs 794
 - experimental identification of miRNA targets 1087–1088
 - experimental RNomics 794–799, 801–802
 - cDNA library construction and analysis 804–811
 - computational analysis 811, 816
 - expression verification of cloned ncRNAs 808
 - immunoprecipitation of ncRNA-protein complexes 809
 - laboratory protocols 744–776, 797–799, 804–811, 813–816
 - identification and characterization in bacteria 719–720, 787–788
 - expression-based discovery 720–726
 - computational search/prediction 726–728, 787–794
 - computational searches for ncRNA targets 736–737
 - Hfq coimmunoprecipitation 724–726
 - high-throughput sequencing (RNA-Seq) 721–724
 - microarray 720–721, 794–797
 - mRNA target validation 737–742
 - laboratory protocols 744–776, 797–799
 - strategies to decipher the biological role 728–737
 - miRNA profiling, barcoded cDNA libraries for NGS 861–872
 - Pb²⁺-induced cleavage 275
 - purely computational approaches. *See also* homology search
 - computational methods for expression profiling based on RNA-seq data 821–823, 827
 - homology search for small structured ncRNAs 595–616, 619–621
 - mascRNA 621–629
 - RsaE RNA 600
 - shape analysis, ncRNA consensus structure prediction 583
 - screening for RNA genes and motifs 595–596
 - identification *via* APART 1197
- non-coding RNP (ncRNP) 801–802
 - ncRNP transcriptome 814
- non-radioactive labeling of RNA 89, 92, 100–101, 1103–1104
 - digoxigenin, internally 92, 100–101
 - biotin 8, 66, 117–125, 937
 - fluorescent dye 117–125, 152, 158, 238, 796–797
- phosphorothioates and iodine cleavage 329–364, 369–393
- Northern blot analysis 89–102, 744–747, 808, 816, 1199. *See also* small RNAs, Northern blot detection
 - aptamer detection 1199
 - laboratory protocols 98–102, 744–747, 808
 - mRNA targeted by miRNA 835
 - miRNA detection 1091–1092, 1094
 - ncRNA verification 789, 808, 816
 - of RNA fragments crosslinked to protein 1030–1031
 - probe with digoxigenin label 89, 95, 101–102
 - radioactively labeled probe 744–747, 1094
 - sRNA expression profile 729, 816
- nuclear extract depletion, with biotinylated RNA 951–952
- nucleation zone, RNA crystallization 485–486, 489
- nucleic acid libraries 1148–1150
 - design 1148–1149
 - pool preparation 1149–1150
- nucleic acid library preparation 1108–1109
- nucleoside analog. *See also* modified RNAs
 - photoreactive 900
 - 4-thiouridine 7, 151–166, 231–250, 878–897, 999
 - 6-thioguanosine 7–8, 231–250, 333–334, 878–882
 - thiotriphosphate 329, 333–334
- nucleotide analog interference mapping (NAIM) 207, 232, 285–286, 329–330, 369–371
 - experimental protocols 333–355, 380–386
- denaturing PAGE of analog-modified RNA pools 336, 381–383
 - nucleoside analog thiotriphosphates 333
 - incorporation by transcription 335–337, 371–372, 380–381
 - iodine cleavage 354–355, 385–386
 - radioactive labeling of RNA pool 337–339, 383–384

- method description 330–332, 369–373
 - NAIM results
 - analysis and interpretation 355–358, 373–375
 - data evaluation 386–387
 - influencing factors 372–373
 - selection step/assay
 - analysis of bacterial RNase P function 369–372
 - *cis*-cleaving RNase P RNA-tRNA conjugates 375–380, 384–385
 - group II intron transesterification 339–344
 - group II intron folding, Mg²⁺-induced 344–347
 - Rho-dependent transcription termination 347–351
 - Rho helicase unwinding 351–354
 - nucleotide analog interference suppression (NAIS) 329–332
 - applications 332–333
 - elucidating tertiary contacts in group I and II ribozymes 332–333
 - to the RNase P system 375, 378–379
 - data analysis and presentation 363–364
 - experimental protocols 358–364
 - functional analysis of RNA mutants 362
 - RNA mutants, design and construction 358–362
 - RNA molecules, preparation with single-atom substitutions 359, 361–362
 - selection step 362–363
 - selection assay, *trans*-branching 339–341
 - nucleotide blast. *See* BLASTN
- O**
- OligoCalc (oligonucleotide properties calculator) 60
 - oligonucleotide, DNA
 - biotinylated (bait) oligonucleotide 178–179, 735
 - biotinylated oligonucleotide for affinity purification of RNA-protein complexes
 - complementary biotinylated DNA oligonucleotide 948
 - chimeric 2'-*O*-methyl RNA/DNA oligonucleotide 940
 - displacement oligonucleotides 937, 940, 945–948
 - circularization of linear RNA 75
 - complementary DNA oligonucleotide for Tb(III)-footprinting 259–260
 - DNA splint for RNA ligation 51, 57–70, 77–79, 293, 362, 501
 - DNAzyme 35–37
 - microarray for detection of small ncRNAs 720
 - microarray for HT-RNomics 795, 802, 804–808, 809–812
 - microarrays for gene expression profiling using RNA-seq 821, 827–828
 - Northern blot probe 745–747, 944
 - for miRNA detection 1091
 - primer extension 207, 220, 278–279, 1033, 1045–1046
 - randomized, in SELEX 16, 1103, 1108–1109, 1148, 1172–1173, 1186, 1208–1209, 1215
 - RNA ligase substrate specificity 72
 - RNase H cleavage 54–56, 967, 1166
 - RNase H mapping of crosslinked RNA 248
 - splint labeling 183
 - template for T7 transcription 4–5, 10, 12
 - unwinding of RNA-DNA hybrids by Rho helicase 351–354
 - oligonucleotide LNA, LNA/DNA, LNA/RNA 153, 729, 1088, 1098, 1247–1254, 1258–1265
 - LNA-ATP/UTP, for T7 transcripts 7
 - LNA/DNA mixmers as probes for Northern blotting 92–94, 96–97, 100–101
 - 5'-digoxigenin (DIG)-labeled 89, 91–97
 - oligonucleotide, RNA
 - adaptor RNA 724
 - analysis of synthetic RNA 143
 - atomic force microscopy (AFM) 528
 - biotinylated antisense oligonucleotides for affinity purification of RNP complexes 957, 976
 - 2'-*O*-methyl RNA oligonucleotide 936–941, 945–947, 949, 952–953
 - chimeric 2'-*O*-methyl RNA/DNA oligonucleotide 940
 - calibrator oligoribonucleotide 863–864, 866–867
 - cDNA libraries for miRNA profiling 863–864, 871
 - chemical RNA synthesis 130, 137, 139–147
 - crystallization of RNA 482, 493, 495
 - enhancing siRNA performance 1248, 1261
 - homooligonucleotide stretch (NAIM) 372
 - iCLIP 900–904, 915
 - initiator oligonucleotides for T7 transcription 8

- oligonucleotide, RNA (*contd.*)
 - labeling of alkyne-containing RNA for smFRET 509–511
 - oligo-U stretch 789
 - PAR-CLIP 878, 890, 895
 - purification of phosphorothioate stereoisomers 290–293
 - RACE 747–750, 753
 - Rho helicase interaction (NAIM) 351–354
 - RNA ligation using T4 DNA ligase 45–70
 - RNA ligation using T4 RNA ligase 75–76
 - – RNA donor oligonucleotide 81
 - RNA molecules containing single atom substitutions 359
 - RNA sequencing 192–194
 - – oligonucleotides derived from alkaline hydrolysis or nuclease P1 digestion 191–192
 - separation of oligomeric and monomeric RNA 108, 115
 - smFRET experiments 509
 - synthesis of modified oligonucleotides 151–166
 - temperature-gradient gel electrophoresis (TGGE) 427, 430, 434–435, 438
 - vector-encoded siRNA and miRNA 1228–1234
- oligoribonucleotide. *See* oligonucleotide, RNA
- Open BABEL utility, MD simulation 702
- OPLS force field, MD simulation 695–696
- optical density 446–447, 765
- osmolytes 463
- outer-sphere coordination, cation : RNA 319
- overlapSelect algorithm, next-generation sequencing 827
- p**
 - Packmol package, MD simulation 694
 - PAGE (polyacrylamide gel electrophoresis). *See* gel electrophoresis
 - paired-end reads, next-generation sequencing 825–826
 - PAR-CLIP
 - (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) 877–897
 - cDNA library construction outline 882
 - cDNA library preparation, from recovered RNA fragments 889–893
 - crosslinking and immunoprecipitation procedure 882–888
 - – UV crosslinking of 4SU-labeled cells 883
 - determination of 4SU incorporation levels 893–897
 - materials 878–881
 - protein-binding sites on RNA, transcriptome-wide identification 877–878
 - particle mesh Ewald (PME) summation method 697
 - partition function 550, 555, 566–570
 - pattern-recognition receptors (PRRs) 1256
 - Pb²⁺-induced cleavage. *See* probing RNA structure
 - P_{BAD} promoter. *See* promoter
 - PCR. *See* polymerase chain reaction
 - PDB (Protein Data Bank) 688–693
 - imprecision and errors in PDB structure files 690–692
 - refinement error 691
 - NMR structures 690
 - nucleic acid structures 689
 - PDB_REDO database/web site 688
 - – automatic refinement of PDB structures 691
 - PDB2SAX program (SAXS) 419
 - PDB2VOL program (SAXS) 417–419
 - periodate oxidation
 - 3'-terminal ribose of RNA 49, 73, 117–119, 124, 863
 - nucleosides and nucleotide-5'-phosphates 118
 - nucleotides 124
 - peripheral blood mononuclear cells (PBMCs) 1256–1257
 - Pfu* (DNA) polymerase
 - 3'-5' proofreading activity 37
 - phenol, phenol/chloroform extraction. *See* RNA extraction
 - phosphoramidites 45, 129, 131–134, 152–154, 156, 359, 509
 - phosphorothioate 156, 231, 285–286, 333, 370
 - analogs and transcripts preparation 335–337
 - incorporation techniques, into RNA 291–293
 - metal ion rescue of phosphorothioate modifications 285
 - stereoisomers purification, by RP-HPLC 290–291
 - photocrosslinking/UV crosslinking 159, 231–249. *See also* crosslinking and PAR-CLIP
 - probing RNA solution structure 231–232
 - – techniques for modified nucleotide incorporation 232–233

- transcription priming with GMPS 233–234
- aryl azide photocrosslinking agent, attachment to 5'-terminal phosphorothioate 236–238
- 3'-addition of aryl azide photocrosslinking agent 238–240
- 4-thiouridine (4SU) and 6-thioguanosine (6SG) 240–243
- photocrosslinking of RNase P RNA and 6SG-modified precursor tRNA 244–247
- primer extension mapping of crosslinked nucleotides 247–249
- plasmid DNA sonication 1172
- PLEXY (prediction of box C/D snoRNAs) 811
- Pme* I restriction enzyme/site 864–865
 - digestion 871–872
- poly(A) polymerase 398, 401
 - 3'-³²P-end-labeling of RNA using [α -³²P]-Cordycepin-triphosphate 182–183, 337–339
 - C-tailing of ncRNAs for cDNA library construction in experimental RNomics and RNPomics 803–806, 812, 814–815
- poly(A)-specific ribonuclease (PARN) 397–403
- polyacrylamide gel electrophoresis (PAGE). *See* gel electrophoresis
- polymerase chain reaction (PCR). *See also* RACE
 - melting point lowering agents
 - betaine 1150
 - dimethyl sulfoxide (DMSO) 1150
 - MasterAmp PCR Enhancer 1150
 - Qiagen Q-Solution 1150
 - qRT-PCR 89, 861, 1092–1093
 - RT-PCR 809–811, 1107, 1118, 1181–1182
 - inhibition by DTT 1119
 - double-nested 1210, 1213–1214, 1217–1218
 - kit, one-step 305, 310
- polystyrene
 - chemical RNA synthesis 131, 133, 135–136, 145
 - paramagnetic beads 938
 - tubes in flow cytometry 764
- postsynthetic RNA labeling 151–152, 155–157, 164–166
 - fluorophore 509
 - introduction of EPR nitroxide spin label 152, 157, 159–160, 164–166
 - at 2'-amino groups 156–164
 - reaction with succinimidyl esters 158
 - reaction with aromatic isothiocyanates 158–161
 - reaction with aliphatic isocyanates 159–161
 - at 4-thiouridine 164
- potato spindle tuber viroid (PSTVd) 431–440, 532–537
- prey
 - (target) RNA 179
 - protein prey plasmid 1076, 1078
- primer extension 56, 206–207, 210, 212, 217, 220–224, 226, 231, 241–243, 245, 247–249, 259, 265, 276, 278–279, 304, 308–311, 635, 644, 747, 773, 900, 944, 953, 977, 991, 993, 995–996, 998–1000, 1002, 1009, 1030, 1033–1038, 1040–1042, 1045–1047, 1050, 1108, 1166, 1213
- primer walking 740, 1200–1201
- probabilistic shape analysis 585–587
- probability dot plot 567, 570
- probing divalent metal ion binding sites of proteins by hydroxyl radicals 397–404
- probing RNA structure. *See also* RNase (ribonuclease)
 - *in vitro* with enzymes and chemicals 205, 224, 422, 435, 462, 563–564, 572, 602–603, 634–635, 740–741
 - backbone-specific chemical probes 211
 - base-specific chemical probes 210
 - CMCT 208–210, 212, 217–218, 635, 740, 991–995, 1006
 - DEPC 188–189, 208, 210–212, 218–220, 225–226
 - dimethylsulfate (DMS) 188–189, 206, 208, 210–212, 217–218, 220, 224, 226, 310, 359, 635, 642, 644, 740, 992–999, 1002, 1007
 - ENU 209–210
 - enzymes 207, 992–993, 995–1000, 1130
 - in-line probing 206, 210–211, 213, 216, 218, 225–226, 417, 1106
 - kethoxal 212, 740, 992, 994–996
 - lead(II) = Pb²⁺ 206, 225, 277–278, 740, 1130
 - Pb²⁺-induced cleavage of RNA 269–271, 1006
 - Pb²⁺-induced cleavage to probe metal ion binding sites, RNA structure and ligand interactions 271–275
 - primer extension 220, 1037
 - probing RNA solution structure by photocrosslinking 231–232, 244–249

- probing RNA structure. *See also* RNase (ribonuclease) (*contd.*)
 - – probing RNA structure and ligand binding sites on RNA by Fenton cleavage 301–314
 - – ribose acetylation 219
 - – RNA helicase, NAIM 354
 - – RNA preparation 213
 - – RNase hydrolysis 214
 - – SHAPE, 2'-hydroxyl acylation 206, 635, 644
 - – structure-specific chemical and enzymatic probes 208–210
 - – Terbium(III)-footprinting 255–267
 - – tertiary contacts, NAIS 332
 - – 2'-hydroxyls involved in tertiary contacts, NAIM 370
 - *in vivo* 226–227, 314
 - – dimethylsulfate (DMS) 206, 208, 222–223, 225, 314, 978, 999, 1002
 - – kethoxal 206, 314
 - – Pb²⁺-induced cleavage of RNA 275–279
 - RNP probing 981, 991–992, 996–1000, 1005–1007
 - promoter 726–731, 761, 1072, 1074, 1081, 1147, 1212, 1228
 - *ADH1* 1073
 - *HIS3* 1069
 - *istR* sRNA 729
 - *lacZ* 1068–1069
 - λ P_L 733
 - *lexA* operator 1068–1069
 - miRNA vector 1228
 - *ompC* 731
 - P_{BAD} (*araC*) 731, 733, 735, 738–739, 755
 - P_{Lac} 733
 - P_{LacO}–1 731–732, 755–758
 - P_{LtetO}–1 755–757
 - P_{tac} 733
 - *RPR1* 1071
 - shRNA vector 1223–1225, 1232
 - SP6 4, 1167–1169, 1173
 - *sraC* 748
 - T7 promoter(s) 4–9. *See also* T7 RNA polymerase
 - T3 promoter 3–4
 - proteinase K digestion 888
 - PURESYS[®] 742
 - *in vitro* translation 766, 768–772
 - PURExpress[®]. *See* PURESYS[®]
 - PyMOL program 418, 657–659, 688, 693, 702
 - pyridyl disulfide-activated matrix/support 1144, 1146
- q**
- QPALMA sequence read alignment program 823, 826
 - QRNA software (comparative genome sequence analysis to detect conserved RNA secondary structures) 726, 788–789
 - qRT-PCR. *See* polymerase chain reaction.
 - quantification plots 311
 - QuB software package, smFRET analysis 507
- r**
- RACE (rapid amplification of cDNA ends) 75, 184, 731, 738, 759, 761
 - 5'-RACE protocol 747–755
 - radioactive labeling (radiolabeling) methods
 - 5'-³²P-end-labeling of RNA 64, 180–181, 198–199, 213, 337–338, 350, 383, 866–867, 883, 885–886, 909–910, 980
 - 5'-³²P-labeling of nucleoside 3'-phosphates 195–196
 - 3'-³²P-end-labeling of RNA 73, 181–183, 213, 337–339, 383–384
 - – with Klenow polymerase, [α -³²P]dNTP and DNA splint 337–338, 341
 - – with poly(A) polymerase and [α -³²P]-Cordycepin-triphosphate 182–183, 337–339
 - – with T7 DNA polymerase, [α -³²P]dATP and DNA splint 182–183
 - DNA primer/probe 5'-³²P-end-labeling 774, 1045–1046, 1091
 - iCLIP RNA 5'-labeling 909–910
 - internal ³²P labeling of RNA by *in vitro* transcription 14, 16, 980–981, 1003, 1112, 1153, 1179, 1190
 - NAIM 369
 - 5'-³²P-end-labeling of analog-modified pool RNA 383
 - – 3'-³²P-end-labeling of analog-modified pool RNA 383–384
 - NAIM/NAIS 329
 - – ³²P-labeling of an RNA pool 337–339
 - ³²P-labeling of recombinant polypeptide 403
 - probing RNA structure *in vitro* with enzymes and chemicals 213
 - radiolabeling of RNA segments crosslinked to immunoprecipitated proteins 885–886
 - radiolabeling of RNA size markers 866–867, 883

- radius of gyration 407, 416–417
- RALEE, RNA alignment editor 572, 621, 625
- randomized exon cassette
 - design 1210–1211
 - ligation 1216
 - preparation 1215
- Rapidshapes heuristics 586
- parameters 592–593
- reads. *See* RNA-Seq
- reads per kilobase per million (RPKM) units 828
- recombinant polypeptide, radioactive labeling 403
- renaturation 224, 312, 429, 431, 436, 980–981, 1005
- reporter gene 738–739, 769, 1067–1069, 1075–1076, 1078, 1080–1081
- respiratory syncytial virus 1212, 1261
- reversed-phase HPLC (RP-HPLC) 132
 - analysis of purified RNA 143–147
 - analysis of ribonucleosides 893–896
 - click reaction 512
 - ion-pair reversed-phase liquid chromatography (IP-RPLC) 142–143, 213
 - postsynthetic labeling of 2'-amino-modified RNA 162–165
 - purification of phosphorothioate stereoisomers 290–291
 - purification of trityl-on RNA 138, 140–142
- reverse transcription 211, 280, 313, 753, 772–773, 804–807, 809–811, 813–816, 869–870, 891–892, 900–904, 913–914, 1017, 1024, 1099, 1118, 1141, 1151–1152, 1189, 1192–1193, 1208–1210
 - SuperScript™ plasmid system 802–804, 806–807
 - SuperScript II reverse transcriptase 279, 753, 773, 775, 806, 811, 813–815, 1118, 1120, 1151
 - SuperScript III reverse transcriptase 224, 260–261, 305, 310, 750, 753, 865, 869, 879, 881, 892, 913, 1120
- Rfam database 583, 619, 629, 667, 789
- Rho
 - Rho helicase 351–354
 - Rho-independent transcription termination 347–351, 595, 604, 726–727, 789–791
- rhodamine 159
 - tetramethylrhodamine 504
 - sulforhodamine 1139
- ribonuclease. *See* RNase
- ribonucleoprotein (RNP). *See also* RNA binding proteins and RNA-protein interaction
 - analysis of RNA-protein interaction using iCLIP 899
 - affinity purification 957
 - – immunoaffinity purification 958–962
 - – RNA aptamer-based affinity purification 963–971
 - biotin-based affinity purification of RNA-protein complexes 935
 - experimental identification of microRNA targets 1087
 - identification and characterization through three-hybrid analysis 1067
 - immunopurification of endogenous RNAs associated with RNA binding proteins (RBPs) *in vivo* 1017
 - isolation of RNA binding proteins using RNA affinity matrices 921
 - PAR-CLIP 877
 - protein-RNA crosslinking in native ribonucleoprotein particles 1029
 - – UV crosslinking 1031–1033
 - – identification by primer extension analysis 1033–1037
 - – identification of crosslinked proteins 1037–1040
 - – protocols 1042–1050
 - RNA-protein interactions and RNA structure 975
 - – RNP reconstitution 978–981
 - – EMSA 981–986
 - – purification of RNPs reconstituted in complex cellular extracts 986–987
 - – tobramycin-sepharose or MS2-MBP affinity chromatography 987–991
 - – RNA structure probing 991–999
 - – UV crosslinking and immunoselection 999–1005
 - RNPomics 801–802, 812–817
 - sedimentation analysis of RNP complexes 1055
 - – glycerol gradient centrifugation 1056–1060
 - – fractionation of RNPs by cesium chloride density gradient centrifugation 1061–1064
 - specialized SELEX method to identify antisense and protein target sites in RNA or hnRNPs 1165–1166
 - – generation of RNA 20-mer library using *Mme* I 1166–1169

- ribonucleoprotein (RNP). *See also* RNA binding proteins and RNA-protein interaction (*contd.*)
 - – natural RNA substrates for proteins, identification 1171
 - – procedure and protocols 1171–1182
- ribosegulation, riboregulator 96, 724, 743
- ribosomal RNA (rRNA) 93, 95, 118, 175, 206, 220, 232, 302, 304, 308, 310, 481, 493, 501, 549, 677–678, 719–722, 724, 787, 791, 802, 807, 816, 823, 921, 975, 999, 1002, 1029, 1258
 - competitor RNA 924
 - rRNA extraction protocol 309
 - 5S rRNA 95, 269, 551, 571, 746, 1060
 - – temperature-gradient gel electrophoresis 432–435
 - – endogenous standard 89, 94–95, 732
 - – abstract shapes 583
 - – structure prediction 642–644
 - – glycerol gradient centrifugation 1059–1060
 - 16S rRNA 90, 95, 269, 278, 302, 304, 746, 791–794, 1057
 - – Fe²⁺-mediated hydroxyl radical cleavage 310, 311
 - – crystallization 494
 - – high-throughput sequencing 722
 - 18S rRNA 90, 95
 - 23S rRNA 90, 95, 278, 302, 746, 1057
 - – high-throughput sequencing 722
 - 28S rRNA 90, 95, 637, 640
 - – structure prediction 646, 657, 659
- ribozyme 6, 53–54, 75, 78–79, 106–107, 124, 129, 152, 158, 176, 223–224, 241, 244, 248, 416, 442, 492, 500–501, 527, 535, 549, 595, 1098, 1140, 1166
 - folding 245–246
- NAIM/NAIS: application to the RNase P system 372, 375, 378, 380–381, 384,
 - NAIM/NAIS: group I and group II ribozymes 332, 339, 344
- primer extension analysis of crosslinks between photoagent-modified precursor tRNA and RNase P 248–249
- probing RNA structure and ligand binding sites by Fenton cleavage 301
- protocol for RNase P cleavage 40–41
- ribozyme cassettes 30
 - – *cis*-cleaving 5'-cassette 30
 - – *cis*-cleaving 3'-cassette 30–31
 - – *trans*-cleaving for generation of homogeneous 3'-ends 33–35
 - – construction 36–39
 - terbium (III) footprinting of HDV ribozyme 261–264
 - thiophilic metal ion rescue of RNA phosphorothioate modifications 286
- RISC. *See* RNA-induced silencing complex
- rmprune utility program 607
- RNA Affinity in Tandem (RAT) method 935
- RNA backbone correction (RNABC) program 688, 691
- RNA binding proteins (RBPs) 244, 373, 440, 527, 579, 842, 852, 976, 1041, 1166, 1171, 1188, 1207, 1222. *See also* ribonucleoprotein and RNA-protein interaction
- analysis of RNA-protein interaction using iCLIP 899
- biotin-based affinity purification of RNA-protein complexes 935
- RNPomics 801–802, 812–817
- identification and characterization of small ncRNAs in bacteria 722
 - – RNASeq 721
 - – Hfq coimmunoprecipitation 724–726
 - – genomic SELEX 728
 - – finding sRNA-associated proteins 735–736
- identification and characterization through three-hybrid analysis 1067
 - – basic strategies 1068–1070
 - – components 1070–1075
 - – protocols 1075–1079
 - – additional applications 1081–1082
- immunopurification of endogenous RNAs associated with RBPs *in vivo* 1017
 - – RNA coimmunoprecipitation 1020–1025
- isolation of RNA binding proteins using RNA affinity matrices 921
 - – method 921–927
 - – applications 927–931
- PAR-CLIP 877
- RNABOB, fast RNA motif/pattern search software 93, 621, 626, 811
- RNA extraction 222, 225, 280, 724, 744, 901, 911, 1037, 1044–1045
 - phenol, chloroform, phenol/chloroform 5, 15, 18–19, 20, 31, 35, 61–66, 69, 80, 90, 94, 98–99, 105, 107, 112, 121, 125, 183, 213–214, 219, 223, 238, 247, 278, 305, 309, 338, 350, 381, 482, 537, 744–745, 750, 752–753, 768–770, 774–775, 804–806, 809–810, 812–813, 815–816, 866, 880–881, 888, 912, 939, 944, 962, 971, 995, 998–999, 1018, 1021, 1026, 1044, 1057, 1059–1062, 1064–1065,

- 1090–1091, 1094, 1112–1114, 1116, 1130, 1157–1158, 1181, 1192, 1216–1217, 1231
- (diethyl)ether 69, 1116, 1130
- elution from gels 62–63, 806, 1180–1181
- hot phenol method 90, 94
- *in vivo* SELEX 1216–1217
- kits 90
- recovery/extraction from membranes 901, 910–913, 1192
- rRNA extraction 309
- TRIzol reagent/method 90, 744, 750–751, 803, 806, 808, 812, 866, 894
- RNAfold 42, 84, 92, 438, 579, 620, 673, 675, 681, 740, 811, 1073, 1106
- secondary structure prediction 549
- – formal background 552
- – internet addresses 557
- – RNAfold 565
- – output from the RNAfold Server 570
- RNAforester, structure alignment program 572, 592, 624–625
- RNA fragments, simultaneous splint ligation, for FRET experiments 66–67
- construct design 68–69
- RNAhybrid algorithm, RNA:RNA interaction prediction 737, 759, 767
- RNA interference (RNAi) 845, 1166, 1210.
- See also small interfering RNA
- gene silencing using vector-encoded siRNAs or miRNAs 1221
- – background information 1221–1223
- – construction of shRNA vectors 1223–1228
- – construction of miRNA vectors 1228–1229
- – construction of extended shRNAs (e-shRNAs) and lhRNAs 1229–1230
- – production of lentiviral vectors encoding anti-HIV-1 shRNAs or e-shRNAs 1230–1234
- RNA-induced silencing complex (RISC) 833, 841–842, 843, 1087, 1221–1223, 1232
- – immunoprecipitation of RISC components 838–839
- using chemical modification to enhance siRNA performance 1243
- RNA isolation. See also RNA extraction, RNA preparation and RNA purification
- 5'-RACE 751–752
- crosslinked RNA 246–247
- iCLIP 911–913
- miRNA profiling by next-generation sequencing 866
- Northern blot detection of small RNAs 89
- PAR-CLIP 883–888, 894
- Pb²⁺-induced cleavage *in vivo* 277–279
- pure RNA species from biological material 176–180
- – single RNA species with partially known sequence 178–179
- – single unknown RNA species following a functional assay 176–178
- – size-fractionated RNA 176
- RNAlishapes program 580, 583, 587
- parameters 588, 593
- RNAMOTIF, RNA secondary structure definition and search algorithm 596–599, 606–609, 615, 811
- program versions and download site 616
- RNA preparation. See also RNA extraction, RNA isolation, RNA purification and T7 RNA polymerase
- Atomic Force Microscopy (AFM) 537–538
- *in vitro* transcription and purification of milligram quantities of RNA 12–13, 111–113
- – weak anion-exchange chromatographic purification 113–115
- Northern blot detection of small RNAs 89
- preparation from gradient fractions 1059–1060, 1064
- preparation for RT-PCR 1114
- RNA ligation 80–81
- RNA molecules containing single-atom substitutions 359–362
- RNA-protein interaction and RNA structure 980–981
- RNA sample enriched through immunoprecipitation 809
- RNA structure probing 213, 223
- RNomics and RNPomics 801
- – cDNA library construction 804–808
- – specialized cDNA library construction 808–811
- – RNP-derived cDNA library construction 814–816
- simultaneous preparation of RNA and protein 1060
- transcripts containing phosphorothioate analogs 335–336, 378–384
- RNA-protein crosslinking 975–978, 999–1005, 1008, 1009, 1191, 1199
- aptamer target selection 1100–1101
- iCLIP 900–902, 905–906
- *in vivo* 1021–1022
- native RNP particles 1031–1033, 1042–1044
- PAR-CLIP 877–878, 882–884

- RNA-protein interaction. *See also*
 - ribonucleoprotein and RNA binding proteins
 - affinity purification 935, 945, 949, 953
 - aptamer-tagged sRNA 736
 - atomic force microscopy (AFM) imaging and force spectroscopy 527, 536, 541
 - coimmunoprecipitation 724
 - deep sequencing 742
 - fluorescence techniques 8, 117, 501
 - genomic SELEX 1199–1200
 - NAIM/NAIS 333, 347–350, 351–354
 - RNA ligation methods 50
 - study by crosslinking 232
 - study in RNPs 975
 - temperature-gradient gel electrophoresis (TGGE) 439–441
 - three-hybrid analysis 1067–1068, 1070–1071, 1074–1076, 1080–1083
- RNA purification. *See also* RNA extraction, RNA isolation and RNA preparation
 - anion-exchange chromatography 111–115
 - HPLC purification 482–483
 - polyacrylamide gel electrophoresis (PAGE) 483–484
 - – denaturing (PAGE) 18, 61–63, 83–84, 381–383, 770
 - – diffusion elution 19, 484
 - – electroelution 19–20
 - – RNA downstream purifications 122
 - – gel chromatography 122
 - – purification on denaturing polyacrylamide (PAA) gels 122–123
 - chemical RNA synthesis 138–139
 - – anion-exchange HPLC purification 139–140
 - – reversed-phase HPLC purification of trityl-on RNA 140–142
 - – detritylation of trityl-on RNA 142
 - – desalting by HPLC 142–143
 - isolation of pure RNA species from biological material 176
 - – preparation of size-fractionated RNA 176
 - – isolation of a single unknown RNA species following a functional assay 176
 - – comments on electrophoretic purification 178
 - – purification of a single RNA species from 1 mg crude small RNAs 179–180
 - purification of phosphorothioate stereoisomers by RP-HPLC 290–291
 - immunopurification of endogenous RNAs associated with RBPs *in vivo* 1017
 - – RNA coimmunoprecipitation 1020–1022
- RNA recognition motif (RRM) 958, 975–976
- RNase (ribonuclease) 210, 224–225, 325, 921, 945, 984, 1008–1009, 1025, 1280
 - activity of T7 RNAP 9
 - cleavage susceptibility to derive base-pairing constraints (MC-Tools) 635
 - complete hydrolysis of RNA for UV spectroscopy 1281
 - – 3'-exonuclease ERI-1 1253
 - from plants 434
 - nuclease P1 190–192, 195–196, 198–199, 1281
 - – identification of terminal nucleotides 184, 190
 - poly(A)-specific ribonuclease (PARN) 397–402
 - ribonuclease recognition sites 738
 - RNase I
 - – iCLIP 900, 902, 906–909, 911–912
 - RNase III 975, 1087
 - – Dicer 833, 852, 1087, 1222–1223, 1227–1228, 1232, 1235, 1243–1244, 1249, 1263
 - – Droscha 833, 1087, 1222–1223, 1228, 1232, 1235
 - probing RNA structure 740
 - RNase A 1000, 1003, 1253, 1281
 - – analysis of nucleotide content 195–196, 198
 - – probing RNA structure 635, 740
 - – RNA 3'-ends 124
 - – tissue activity 906
 - RNase CL3
 - – enzymatic sequencing 185–187
 - RNase contamination 20, 59, 70, 84, 146, 175, 226, 280, 324, 327, 389, 495, 520, 536, 897, 921, 979, 987, 1003, 1025–1026, 1041, 1093, 1107, 1113
 - RNase D 397
 - RNase E 742–743
 - RNase H 750
 - – cDNA library construction 804, 807
 - – engogenous activity 59
 - – generation of homogeneous ends 35, 54
 - – generation of RNA ligation substrates 51–52, 57
 - site-directed cleavage 51–56, 64, 967, 970
 - – site-directed cleavage protocol 63
 - mapping 248, 937, 939, 1030, 1166
 - reverse transcription 753, 1151

- RNase inhibitor 10, 13–14, 17–18, 20, 60–61, 63–65, 70, 81–82, 234, 293, 305, 310, 336, 362, 536, 750, 752–753, 812, 815, 906, 938, 952–953, 1019, 1020–1021, 1026, 1113, 1153, 1174, 1179, 1181–1182, 1190
- RNase P (RNA) 80, 922, 924, 927, 1055, 1280
- bioinformatic studies 595, 619, 622, 628, 678
- cleavage protocol 40–41, 293–294
- crystallization 481, 492
- deletion of gene for RNA subunit 730
- generation of homogeneous 3'-ends 33–35
- *in vitro* T7 transcription 8, 13–14
- mascRNA processing 622
- menRNA processing 628
- metal ion coordination interactions 287–290, 292
- Mg²⁺-probing 277
- NAIM 369–393
- Pb²⁺-probing 271–273, 275–276
- RNA ligation 77
- SAXS 411, 417, 422
- single molecule FRET 500
- stoichiometry of Mg²⁺ ions bound to RNA 326
- structure probing by photoaffinity crosslinking 231–232, 241–242, 244–246, 248–249
- terbium (III) footprinting 264
- three-hybrid analysis 1068, 1071, 1075, 1082
- RNase protection assay 1199, 1213
- RNase S1
 - probing RNA structure 635, 1130
- RNase T1 1003–1004, 1281
- RNA 3'-ends 124
- enzymatic sequencing 185–187, 192, 1108–1109
- probing RNA structure 207–208, 212, 214–217, 224, 226, 635, 740, 992–993, 995, 998–1001, 1130
- terbium (III) footprinting 258, 263
- sequencing ladders 306–307
- PAR-CLIP cDNA library construction 879, 882, 884–887
- RNase T2 1281
- identification of terminal nucleotides 184, 190–191
- analysis of nucleotide content 195–196, 198
- probing RNA structure 207–208, 212, 214–215, 635, 740, 992–993, 995–998
- RNase U2
 - enzymatic sequencing 185–187
- probing RNA structure 635
- RNase V1
 - probing RNA structure 207–208, 210, 212, 214–215, 217, 224, 635, 740, 992–993, 995–998, 1130
- RNase Z (tRNase Z)
 - menRNA processing 628
- purification using RNA affinity matrices 922, 924–925, 927–929
- siRNA stability 1248, 1253–1255, 1261
- *Staphylococcus aureus* nuclease 185
- RNase-free water preparation 8, 59, 175, 280, 389, 512–513, 744, 803, 1143, 1173
- RNA-Seq (high-throughput sequencing)
 - aptamer selection 1099, 1106, 1109, 1122
- barcoded cDNA libraries for miRNAs 861–872
- cDNA reads
 - iCLIP 901
 - microRNAs 861–863
 - PAR-CLIP 878
- computational methods for gene expression profiling 821–826, 829
- expression-based discovery of bacterial sRNAs 720–726
- differential RNA sequencing (dRNA-seq) approach 722–724, 732
- genomic SELEX 1186, 1194–1195, 1197
- iCLIP 899–900, 901, 916
- immunopurification of endogenous RNAs associated with RBPs 1018, 1025–1026
- PAR-CLIP 877, 886–887, 889, 892–893
- protein-RNA crosslinking in native RNP particles 1029
- RNA sequencing
 - chemical 187–189
 - enzymatic 184–187
 - identification of modified nucleotides 194–201
- terminal RNA sequence determination 190–194
- RNashapes program 580–581, 584–586, 588, 590, 624
- parameters 585, 592–593
- RNA tester sequence (RTS) 1069, 1071–1073
- RNAz program, conserved secondary structure identification 620, 624, 788–790
- RNomics 794, 801
- library construction and analysis protocols 804–811

RNomics (*contd.*)

- materials 802–804
- ncRNA sequence, computational analysis 811

RNPomics

- library construction protocols 813–816
- materials 812–813
- RPR1 promoter 1071

S

S2S 572, 667–675, 677–681

- interactive RNA alignment viewer and editor 668–671
- Assemble2 667–668, 671–677, 680–684
- – interactive RNA 3D modeler 671–672
- – installation of S2S and Assemble2 673–674

SAM (S-adenosylmethionine) 1174, 1178

- SAM riboswitch 501

SAM format (computational tool for RNA-Seq) 823, 826, 1197

SAMs (self-assembled monolayers) of alkanethiols on gold surfaces 543

SAXS. *See* small-angle X-ray scattering

S-domain (RNase P RNA) 272, 388–389, 408, 411–416, 418–419, 421

- folding intermediate
- – global conformation 412–414
- – stable and extended conformation 414–416

SDS-PAGE. *See* SDS-polyacrylamide gel electrophoresis

SDS-polyacrylamide gel electrophoresis

- (SDS-PAGE) 23, 114, 399–404, 729–732, 881–882, 886–887, 897, 901, 908, 910–911, 928–931, 942–943, 949–950, 961, 968–969, 971, 977, 980, 984–986, 990–992, 1003–1005, 1042, 1057, 1060, 1090, 1100–1101, 1110–1111

- NuPAGE 880, 910, 968–969, 971

secondary structure prediction, of RNA 549

- mfold (UNAFold web server) 92, 549, 555–565

- – energy dot plot 560–561

- – p-num 561–564

- – ss-count 561–564

- thermodynamics 550–552

- RNAfold (Vienna RNA server) 92, 549, 565–571

- – mountain plot 571

- – probability dot plot 569–570

sedimentation analysis, of ribonucleoprotein complexes

- cesium chloride density gradient 1055–1056, 1061–1064

- glycerol gradient centrifugation 802, 813–816, 940, 948, 960–962, 966–968, 970–971, 976, 986, 1056–1059, 1061

- sucrose gradient 839, 1061

SELEX (Systematic Evolution of Ligands by EXponential enrichment)

- database 1098

- genomic 720, 728, 1165–1166, 1171, 1185–1186

- identification of antisense and protein target sites in RNA or hnRNP complexes 1165–1182

- *in vitro* selection 16, 442, 1097, 1099–1101, 1104–1107, 1116

- – against proteins and carbohydrates 1097–1131

- – against small targets 1139–1162

- – selection cycle 1099, 1104–1106, 1141

- *in vivo* strategies 1207–1218

- – minigene design 1212–1213

- – monitoring for enrichment of exon sequences that function as splicing enhancers 1213–1214

- – procedure overview 1208–1210

- – protocols 1215–1218

- – randomized exon cassette design 1210–1211

- – RT-PCR amplification 1213

- subtractive 1098

Semi-Automated Footprinting Analysis (SAFA) software 221

sense strand (SS)

- DNA cleavage by *Bse* RI and *Bsg* I 761

- genome

- – coverage in microarrays 720

- – ncRNA-coding 789

- lhRNA 1226–1227, 1230

- shRNA 1225–1227

- siRNA 1243–1245, 1249–1251, 1253, 1255, 1257, 1263–1265

Sephadex, Sepharose. *See* chromatography sequence reads. *See* RNA-seqsequencing. *See also* next-generation sequencing (NGS), RNA-seq and RNA sequencing

- direct RNA sequencing 175, 183–184, 1029

- – enzymatically 184–187, 725

- – chemically 187–189

- – determination of terminal RNA sequences by two-dimensional mobility shift 190–194

- identification of modified nucleotides 194–201
- phosphorothioate-modified RNA by iodine 341, 350
- DNA sequencing reaction/ladder, dideoxy 217, 221, 223, 226, 245, 248, 259–261, 311, 741, 774, 807, 995–997, 1000–1002, 1034, 1046–1047, 1108
- additive betaine 1234
- DNA sequencing kit 772, 804, 1224, 1227
- of PCR products 754–755
- of 5'-RACE products 748–749
- Sanger method 720–721, 802, 816, 833
- SEQUEST, tandem mass spectrometry data analysis program 929–930
- ShapeFinder software, analysis of nucleotide reactivity 221, 265
- short hairpin RNA (shRNA) 1222, 1243–1244. *See also* extended shRNA (e-shRNA)
- construction of vectors 1223–1228
- Dicer-dependent short shRNA 1244
- lentiviral vectors encoding anti-HIV-1 shRNAs 1230–1235
- long shRNA 1244
- shreps (shape representative structures) 581, 586–591
- computing 584–585
- shRNA. *See* short hairpin RNA
- shuffle program 599–600, 608, 616
- Simulaid package, MD simulation 694, 704
- single-end reads, next-generation sequencing 825
- single molecule fluorescence resonance energy transfer (smFRET) 499–500
- antifading agents 510–511, 515–516
- “click reaction” between RNA-alkyne and dye azide 509–510
- confocal microscopy 500
- “burst mode” vs. measurements with immobilized RNA molecules 500
- donor-only populations 521
- dye labeling of RNA construct 508–509
- dye pairs 504
- choice of pair 507–508
- Förster theory of FRET 502–503
- instrumental setup 505–507
- protocols 511–521
- bulk FRET measurements 513–515
- measurement chambers and sample preparation 516–520
- removal of unbound fluorophore 520–521
- RNA construct design 503–505
- RNA dynamics addressed by smFRET 500–502
- RNA immobilization 519
- biotin-streptavidin-biotin interaction 519
- surface coverage 521
- surface passivation 519
- RNase contamination 520
- sample drying 521
- single-stranded DNA-binding protein (SSB), in T7 transcription 21
- SIPHT web interface, prediction of bacterial intergenic loci 790–791
- siRNA. *See* small interfering RNA
- SITUS suite for SAXS analysis 417–419
- Sma* I restriction enzyme/site 532–533, 537, 1071–1072
- 7SK RNA 628, 827
- 7SK polymerase II promoter 1225
- small-angle X-ray scattering (SAXS) 407–409
- application example bacterial RNase P RNA 411
- folding intermediate of *Bacillus subtilis* RNase P RNA S-domain 412–416
- experimental setup 410–412
- RNA modeling, low-resolution real-space reconstruction 416–421
- small interfering RNA (siRNA) 129–130, 142, 301, 842, 976, 1221–1223, 1227–1228, 1230, 1232, 1235, 1243. *See also* gene silencing methods
- architectures, siRNA designs 1243–1244
- helix stability asymmetry 1245
- aiRNA, asiRNA, bulge-siRNA, blunt fork-siRNA, blunt siRNA, 27 bp-siRNA, Dicer-independent short shRNA, dumbbell siRNA, long shRNA, sisiRNA, ss-siRNA 1244
- chemical modification 1244–1245
- backbone modification 1246–1248
- base modifications 1252
- ribose 2'-OH substitutions 1248–1251
- ribose backbone alteration 1251–1252
- guide for *in vivo* studies 1261–1265
- guide strand = antisense strand (AS) 1243
- passenger strand = sense strand (SS) 1243
- potency improvement 1252–1253
- stability against ribonucleases 1253
- stabilization strategies 1254–1255
- silencing duration enhancement 1255–1256
- immunogenicity
- abrogation by chemical modification 1257–1258

- small interfering RNA (siRNA) (*contd.*)
 - – cellular responses 1256–1257
 - – off-target effects, reduction 1258–1259
 - pharmacokinetics, improvement by chemical modification 1259
 - – biodistribution, altering by conjugation 1261
 - – cellular delivery enhancement by conjugation 1260
 - tolerance toward modification 1244–1245
- small non-coding RNA (sRNA), bacterial 89, 301, 600, 719–720
 - biological role deciphering 728–729
 - – associated proteins, identification 735–736
 - – biocomputational approaches to find targets 736–737
 - – deletion of sRNA gene (knockout strain) 729–731
 - – expression profile 729
 - – libraries (collections) of sRNA deletion strains or overexpression plasmids 734–735
 - – overexpression 731–732
 - – pulse expression combined with transcriptome analysis 733–734
 - experimental target validation 737–738
 - – *in vitro* characterization of function 741–742
 - – *in vitro* RNA–RNA footprinting 739–741
 - – reporter gene fusions and sRNA chimera 738–739
 - expression-based discovery 720
 - – Hfq coimmunoprecipitation 724–726
 - – high-throughput sequencing and RNA-seq 721–724
 - – microarray 720–721
 - expression-independent searches 726
 - – biocomputational prediction, of sRNAs 726–728
 - – biocomputational prediction, of sRNA targets 736–737
 - – genomic SELEX 728
 - protocols
 - – 30S ribosome toeprinting assay 772–776
 - – flow cytometry 764–767
 - – green-fluorescent protein (GFP)-based reporter system for target validation 755–764
 - – Northern blot protocol 744–747
 - – PURESYSTEM *in vitro* translation 766, 768–772
 - – 5'-RACE protocol 747–755
- small RNAs, Northern blot detection 89
 - application example 6S RNA-derived pRNAs 96
 - DIG-labeled probes for detection 95, 101–102
 - hybridization with complementary probe 92
 - – DNA/LNA mixmer probe design 92–94
 - method limitations 96–97
 - native versus denaturing gels 90–91
 - Northern hybridization protocols 98–102
 - probe generation by T7 transcription using DIG-11-UTP 89, 92, 100–101
 - RNA isolation 89–90
 - RNA transfer and fixation to membranes 91
 - – 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) crosslinking 89, 91–92, 95–97, 100
- smFRET. *See* single molecule fluorescence resonance energy transfer
- SnoReport (computational identification of snoRNAs with unknown targets) 811
- snRNP (small nuclear ribonucleoprotein). *See* spliceosomal small nuclear ribonucleoproteins
- SOAP sequence read alignment program 823, 826
- solid-phase synthesis of RNA 132–136, 151–152
 - automated RNA synthesis 136
 - chemistry variants
 - – 2'-ACE method 129
 - – 2'-O-TBDMS method 129
 - – TOM method 129
 - deprotection protocol 136–137
 - desilylation protocols 137–138
 - detritylation of trityl-on RNA 142
 - HPLC purification 138–142
 - – anion-exchange 139–140
 - – desalting 142–143
 - – reversed-phase 140–142
 - manual RNA synthesis 134–136
 - modified nucleoside phosphoramidites 117, 152–155, 359, 509
 - – quality control of modified RNA 164–166
 - product quality and purity analysis 143–147
- SP6 linker 1169, 1173, 1180
 - – dimer 1170
 - – ligation 1167, 1170, 1172
 - – preparation and ligation protocol 1178–1179
- spatial resolution of AFM 528, 530–531

- Sph* I restriction enzyme/site 1071–1072
- spliceosomal and small nuclear ribonucleoprotein complexes, affinity purification 957
- anti-m₃G cap antibodies 986
 - biotin-based 935–937
 - immunoaffinity purification using antipeptide antibodies 958–962
 - RNA aptamer-based affinity purification 986–993
 - isolation of native spliceosomal complexes 963–971
- spliceosomal small nuclear ribonucleoproteins (snRNPs) 957, 986, 996, 1029, 1040–1043, 1055–1056, 1061, 1063–1064
- affinity purification using biotinylated antisense 2'-O-methyl RNA oligonucleotides and streptavidin agarose 940–948
 - enrichment by DEAE chromatography 938
 - immunoaffinity purification using antipeptide antibodies 958–962
 - reconstitution 981
 - U1 type 945–948, 960–961, 963, 986, 1031, 1034, 1036, 1038, 1040, 1042
 - U2 type 957, 960–963, 986, 1043, 1064
 - U4 type 963
 - U4/U6 941–942, 986, 1031, 1063–1064
 - U4/U6.U5 tri-snRNP 963, 986, 996–997, 1031, 1043, 1055
 - U5 type 961, 963, 986, 1041–1043, 1064
 - RNA aptamer-based affinity purification 963–971
- splint labeling, of RNA 3'-end 182–183
- spring constant, of cantilever in AFM 528, 535, 538, 541
- sRNA. *See* small non-coding RNA
- sRNATarget web server (bacterial sRNA target prediction) 737
- ss-count (mfold) 561–564
- staining
- cells with Trypan blue 1021
 - DNA 1121
 - ethidium bromide (EtBr) 807, 1215
 - proteins
 - Coomassie (Brilliant Blue) 114, 176, 943, 946, 950, 959, 961–962, 968–969, 971
 - silver 399–400, 402, 930–931, 942–943, 953, 1043, 1048, 1110
 - RNA 430, 768, 770
 - toluidine blue 15, 112, 114, 177–178
 - ethidium bromide (EtBr) 15, 24, 41, 59, 90, 95, 471, 1060
 - silver 178, 180, 427, 433–434, 437, 439, 942, 944, 946, 962, 967–968, 971, 1059, 1063
 - SYBR Gold 41, 768
- statistical RNA hydrolysate 190–192
- stem-loop RNA structure 595, 992–993
- binding to MS2 coat protein 992–993, 1069, 1071, 1073
 - binding to PP7 coat protein 935
 - 3'-extension recognized by *trans*-acting VS ribozymes 33
 - miRNA 1227–1228
 - qRT-PCR of miRNAs 89
 - Rho-independent transcription termination 347, 350–351, 588, 604
 - stem-loop binding protein (SLBP) 951, 1078
 - tRNA 76, 692
 - U1 snRNA 1034, 1036, 1039
 - U4 snRNA 1031
 - yeast U3 snoRNA 1002
 - yeast U5 snRNA 996–997
- streptavidin 117, 947
- alkaline phosphatase conjugate 118
 - aptamer 735, 935, 1108
 - gel retardation/shift of biotinylated RNA 123, 125
 - (agarose) beads/resin/matrix 178–179, 372, 936–937, 940–941, 949–950, 1147, 1154, 1157–1158
 - in smFRET 519–521
 - magnetic beads 937–939, 1147
 - RNP affinity selection/chromatography 936, 976, 1088
 - sensor/cantilever functionalized with, in AFM
 - neutravidin agarose 947
- structural RNA alignment 598–599, 667, 669
- INFERNAL 614
 - LOCARNA 602
 - S2S 668–669, 673, 678–680
 - visualization 573
- 2D and 3D structure prediction, over the internet 633–634
- equipment and input 634
 - MC-Tools 633–634
 - MC-Cons 636, 651–654
 - MC-Fold 635, 637–651
 - MC-Sym 636, 654, 657–661, 663
- 4SU, 6SG. *See* 4-thiouridine and nucleoside analog
- succinimidyl ester 156, 158. *See also* N-hydroxysuccinimide (NHS)

- SuperScript. *See also* reverse transcription
- SuperScript™ plasmid system 802–804, 806–807
 - surface and sensor functionalization, AFM 542–543
 - surface passivation, AFM 519
 - surface plasmon resonance (SPR) 1106, 1126, 1128–1130
 - Survival of Motor Neuron (SMN) protein 935
 - synchrotron 409
 - beamline 411
 - generation of hydroxyl radicals 304, 314
 - SAXS experiments 409–411, 421–422
 - X-ray beams 301
 - Systematic Evolution of Ligands by EXponential enrichment (SELEX). *See* SELEX
- t**
- T4 DNA ligase (T4 Dnl) 293, 361–362, 757, 1174, 1224
- advantages 49
 - DNA ligation 110, 759–760, 807, 1121, 1178, 1216
 - protocols 60–61, 65–66, 68–69
 - reaction mechanism 46, 48–49, 54
 - RNA ligation method 50–52, 57–58, 78
 - – simultaneous splint ligation of five RNA fragments 66–69
 - substrate specificity 47
- T4 polynucleotide kinase (PNK) 70, 180, 182, 772, 803, 865, 879, 979, 1094, 1174
- 5′-end-labeling with [γ - $^{32}/^{33}$ P]ATP 32–33, 57, 64–65, 68–69, 80–81, 181, 195, 198–199, 213, 337–338, 350, 383, 404, 774, 808, 810, 866–867, 883, 885–886, 902, 909–910, 923, 980, 1045–1046, 1091, 1130, 1176–1177
 - 5′-end-labeling by phosphate exchange 181
 - 5′- ^{32} P-labeling of nucleoside monophosphates 196
 - 3′-dephosphorylation, removal of 2′,3′ cyclic phosphate 33, 39–40, 121
- T4 polynucleotide ligase 46, 51. *See also* T4 DNA ligase and T4 RNA ligase
- T4 RNA ligase(s) 182, 212, 750, 757, 803, 813, 865, 879, 979
- adapter ligation 752–753, 804–805, 809–811, 814–815, 862–872, 886–887, 889–891, 901, 903–904, 909–910, 1037
 - applications 73–76, 79–80
 - comparison to T4 DNA ligase 49
 - 3′-labeling by ligation of 5′- ^{32} P-pCp 181–183, 213, 337–338, 383–384, 980, 1130
 - ligation of large RNA molecules 76–79
 - mechanism and substrate specificity 46–49, 71–73
 - properties 46
 - protocols 80–84, 817
 - site-specific 3′-terminal modification 233
 - T4 RNA ligase 1 (T4 Rnl1, RnlA) 46–50, 70–73, 75–76, 78–82, 862, 865, 869, 879
 - T4 RNA ligase 2 (T4 Rnl2) 46–47, 49, 68
 - T4 Rnl2 “Truncated K227Q” 46–47, 862–863, 865, 868, 872, 879, 889
- T7 DNA Polymerase
- 3′- ^{32} P-labeling with [α - ^{32} P]dATP 182–183
- T7 RNA Polymerase (T7 RNAP)
- bacteriophage promoters (T3, T7, SP6)
 - conformational heterogeneity
 - – analyzed by TGGE 436–437
 - – RNA renaturation to generate homogeneous conformers 213, 224, 412, 485, 980–981, 989–990, 1005, 1008, 1124
 - endogenous DNase and RNase activity 9
 - *in vitro* transcription
 - – annealed oligonucleotides as template 12
 - – for AFM studies 537
 - – for RNA probing 213
 - – for SAXS studies 412
 - – for SELEX studies 1099, 1108, 1112–1113, 1153, 1179
 - – homogeneous 5′-/3′-ends and 5′-end groups 5–6, 29–42
 - – homogeneous 5′-/3′-ends, RNase H approach 54–56, 63
 - – homogeneous 3′-ends by 2′-O-methyl residues at 5′-end of template strand 53–55
 - – incorporation of phosphorothioate analogs 335–337, 369–372, 380–381
 - – internally digoxigenin-labeled (DIG-11-UTP) 100–101
 - – internally ^{32} P-labeled 14, 16, 747, 980, 1179
 - – large RNA amounts for structural studies 12–13, 111–113
 - – low or no product yield 20–21, 1107
 - – nucleotide analogs for internal and 5′-terminal incorporation 6–8
 - – purification of transcripts 18–20, 61–63, 336, 381–382, 1179–1180

- RNAs with biased nucleotide composition 14–15
- transcripts initiated with guanosine, 5'-GMP or 5'-ApG 11–12
- with 5'-cap structures 17–18, 51–52, 61
- with 5'-terminal GMPS modification 233–234, 240
- with 2'-fluoro-modified nucleotides 16–17
- with unmodified nucleotides 9–15, 60–61, 980
- mutant T7 RNAP
- Y639F 7–8, 16–17, 21, 335–336, 348–349, 351, 371
- Y639F/H784A 7–8, 349, 371
- NAIM analysis of T7 RNAP termination 347–351
- rapid T7 RNAP preparation 21–25
 - procedure 22–24
- template types 4–5
- plasmid template preparation 108–111
- TagIt*-affinity purification 935
- Taq* (DNA) Polymerase 37, 750, 753–754, 768, 772, 803, 811, 813, 816–817, 865, 870, 879, 892, 1118, 1120–1121, 1149, 1152, 1173–1174, 1179, 1218
- TargetRNA2 algorithm (bacterial sRNA target identification) 737
- TBDMS (2'-O-) RNA synthesis chemistry 129–132, 136
- temperature-gradient gel electrophoresis (TGGE) 427, 462
 - application examples 433–443
 - analysis of RNA mixtures in a single TGGE experiment 432–435
 - mutant RNA analysis 438–440
 - protein-RNA complex detection 439–441
 - structure transition analysis 435–438
- experimental conditions optimization 430–431
 - electric field 430
 - gel matrix pore size 430
 - ionic strength and urea 431
 - handling 429–430
- instruments 429
- interpretation rules 431, 433
- principle 428–429
- template-stripped gold (TSG) method (force spectroscopy AFM) 543
- terbium(III) footprinting 255–256
 - application example HDV ribozyme 261–264
 - of long RNAs by primer extension 259–261
 - of short RNAs 256–259
 - 3'-terminal attachment of fluorescent dyes and biotin 117–125
 - background and chemistry 118–120
 - protocols
 - biotin attachment 120–121
 - fluorescence labeling 121
 - labeled RNA purification and quality control 122–123
 - RNA 3'-dephosphorylation 121
- terminal (deoxynucleotidyl) transferase (TdT) 76
- tertiary (3D) structures. *See also* crystallization of RNA, NAIM, NAIS and SAXS
- probing, RNA
 - enzymatic 205–206
 - Tb(III) 255–266
 - Pb²⁺ 269–280
 - Fenton cleavage 301–314
 - NAIM/NAIS 333, 375
- temperature-gradient gel electrophoresis 427–442
- UV melting 462, 466
- RNA crystallization 492
- FRET 501, 507
- AFM 531
- prediction/modeling 633, 667. *See also* 3D architectures, Assemble2, S2S and molecular dynamics (MD) simulations of RNA systems
- molecular dynamics 687
- RNA-protein interaction 975–976, 978, 981
- aptamers, *in vitro* selection 1130, 1148, 1150, 1155, 1171
- proteins, Fe(II)-mediated cleavage 400
- tetracycline, mapping of binding sites on RNA by drug-directed Fenton cleavage 310–312
- thin-layer chromatography (TLC) 62, 175, 184, 190, 192, 194–199, 201, 236
- thiophilic metal ion rescue
 - kinetic analysis 293–294
- RNA phosphorothioate modifications 286–290
- 6-thioguanosine. *See* nucleoside analog
- 4-thiouridine. *See also* nucleoside analog, PAR-CLIP and photocrosslinking/UV crosslinking
- postsynthetic labeling of modified RNA 152, 157, 160, 164

- 4-thiouridine. *See also* nucleoside analog, PAR-CLIP and photocrosslinking/UV crosslinking (*contd.*)
- UV crosslinking of 4SU-labeled cells 883–884
- three-way ligation 51–52, 58
- tiling (micro)arrays 720–721, 725, 787, 795–796
- T-jump technique 468
- 2'-thionocarbamate (TC) chemistry method 130
- TLC. *See* thin-layer chromatography
- tobramycin 957, 963, 987–989, 991–992, 1006, 1117
- aptamer 735, 935, 957, 963, 987–989, 1006, 1008
- tobramycin-sepharose 987–989, 991–992, 1006, 1008
- toeprint assay, 30S ribosomes 741–742, 772–776
- toll-like receptor (TLR) 1257–1258, 1265
- TOPHAT sequence read alignment program 823, 826
- total internal reflection fluorescence microscope, prism-based (prism-TIRF) 500, 511, 517
- transcription elongation complex (TEC) 3, 347, 349
- transcription priming 17, 61, 231, 233–234
- priming of reverse transcription 901, 903, 995
- transcriptome 789, 794, 853
- biocomputational analysis 727, 821, 823, 826, 830
- (d)RNA-seq 724, 821, 823, 826, 830, 1025
- experimental RNomics/RNPomics 801, 814
- genomic SELEX 1202
- meta-transcriptome data 727
- microarray 720, 796
- miRNA targeting 853
- sRNA pulse expression and transcriptome analysis 733–734
- transcriptome-wide identification of protein binding sites on RNA (PAR-CLIP) 877
- transduction 730, 1233–1235
- transfection 837, 844–845, 849, 1089, 1208, 1211–1214, 1216, 1221, 1223, 1232–1234, 1255–1256, 1260
- TransTermHP algorithm (prediction of bacterial transcription terminators) 791
- trimethylamine oxide (TMAO) 463
- tris-(hydroxypropyltriazolylmethyl)amine (THPTA) 509–510, 512
- trityl-off RNA 145, 361
- desilylation 137–138
- failure sequences 141
- trityl-on RNA 137, 139, 145–146
- desilylation 138
- detritylation 142
- reversed-phase HPLC purification 140–142, 144
- transfer RNA (tRNA) 355
- abstract shape analysis 579, 584, 588
- affinity purification of RNPs 1050
- AFM 532–533, 536–537
- 3'-biotinylation/tagging with fluorescent dyes 118, 123
- crystallization 492
- homogenous 5'/3'-ends 32–35, 38, 40–42
- homology search for small ncRNAs 619, 622, 627–628
- identification and characterization of small ncRNAs in bacteria 719
- – microarray 720
- – RNA-seq 721–722, 724
- – cell-free translation 741–742
- – toeprinting assay 741, 772–775
- immunopurification of RNPs 1021, 1047
- isolation of RNA binding proteins 921–932, 946
- metal ion coordination 287–290
- – incorporation of phosphorothioates 291–293
- – thiophilic metal ion rescue 293–294
- miRNA targets 1092
- molecular dynamics (MD) simulation 692, 700
- NAIM 355, 372–375, 377–381, 383–384, 387, 389–392
- quality control of total cellular RNA, Northern blotting 95
- RNA ligation, *in vitro* production of full-length tRNA 50, 70–71, 74–81, 84
- RNA-protein interactions and RNA structure in RNPs 981–982, 984, 989–999, 1007, 1009
- RNA sequence and modification determination by radiolabeling 175–176, 182, 185, 187–188, 200
- – alkaline hydrolysis 191
- – chemical sequencing 188–189
- – end group analysis 190–191
- – enzymatic hydrolysis 191–192
- – production of labeled random RNA fragments 199
- – radiolabeling 180–181
- – TLC 192, 196

- RNomics/RNPomics 802
 - screening of genomes for known RNA genes or motifs 595–596
 - sedimentation analysis of RNPs 1059, 1063
 - shRNA vectors 1223
 - smFRET 501, 504–505, 515
 - – dye labeling 508
 - structure and folding analysis by SAXS 409, 411, 416
 - structure prediction
 - – 2D 551
 - – 2D and 3D 644–645, 651–652, 655
 - structure probing by crosslinking 242, 244–246, 248–249
 - – 5'-end-labeling by transcription priming 233–234
 - structure probing by Fenton cleavage 302
 - structure probing by Pb²⁺ cleavage 269
 - – probing of metal ion binding sites 271, 273
 - structure probing by Tb³⁺ footprinting 264
 - structure probing *in vitro* 205, 211, 214, 216, 218–219
 - temperature-gradient gel electrophoresis 434, 439, 441
 - UV melting 446, 449, 461–462
 - tRNA processing. *See* RNase (RNase P, RNase Z), tRNase Z and tRNA nucleotidyltransferase
 - tRNA nucleotidyltransferase 182, 922
 - tRNA synthetase 1082
 - aminoacylation of tRNA 29, 33, 79, 176, 504–505, 922
 - aminoacyl-tRNA 501
 - aminoacyl-transferase ribozyme 264
 - tRNA splicing 922–923, 927, 930
 - adenylyltransferase 923, 931
 - 2' phosphotransferase 923
 - endonuclease 923, 927
 - ligase 922–924
 - – purification 930–931
 - tRNase Z (RNase Z)
 - bioinformatic studies 628
 - purification from wheat germ using RNA affinity matrices 922, 924–925, 927–929
 - tRNAdb (transfer RNA database) 816
 - Turnip Yellow Mosaic Virus (TYMV) RNA 531, 533
- u**
- UCSC Genome Browser 620, 625, 816, 822–823, 826, 829
 - ULSTM Fluorescent Labeling Kit, for Agilent arrays 796–797
 - ultracentrifugation 427, 1041, 1055
 - cesium chloride/sulfate 1056, 1061–1064
 - glycerol gradient 814–816, 1056, 1058–1059
 - sucrose gradient 1061
 - – polysome profiling 839
 - ultra performance liquid chromatography (UPLC) 143–144
 - ultraviolet (UV) crosslinking/irradiation
 - RNPs *in vivo* 877–878, 883–884, 899–901, 905–906
 - RNP structure analysis 977, 999, 1003–1005, 1008–1009, 1026, 1030–1035, 1042–1044
 - – naked RNA control 1045
 - RNA structure probing 240–247
 - RNA to nylon membrane 91, 746, 808
 - ultraviolet (UV) melting studies of RNA 439, 445
 - definitions and nomenclature 446–447
 - equilibrium and non-equilibrium 459–460
 - experimental solution to sloping baseline 468
 - – experimental data processing 472–473
 - – heating rate and data sampling 471
 - – pH variation and buffers 468–470
 - – RNA degradation 470–471
 - – softwares 473
 - experiments for thermodynamic studies 449
 - – bimolecular transitions 452–459
 - – melting temperature definition 450
 - – unimolecular transitions 451–452
 - melting curves of large RNAs, thermodynamic information 461–462
 - parameters influencing melting temperature 462–463
 - practical problems
 - – evaporation during heating 463–464
 - – sloping baseline 464, 466–467
 - self-complementary sequence pitfalls 460–461
 - two-state approximation 459
 - UV absorption
 - – characteristics of nucleic acid bases 447–449
 - – physical basis 445–446
 - ultraviolet (UV) shadowing 13, 19, 41, 53, 62–63, 84, 112, 163, 336, 382, 484, 1045, 1123, 1214
 - ultraviolet (UV) spectroscopy, for RNA
 - chemical reaction monitoring 164

- ultraviolet (UV) spectroscopy, for RNA (*contd.*)
 - duplex melting (thermal denaturation) 159, 164, 166
 - GMPS 236–237
 - RNA quantification 60, 1108, 1279–1281
- UNAFold web server (UNAFold program package) 549, 555–557
- unlocked nucleic acid (UNA) 1246–1249, 1251–1254, 1259, 1262, 1264
- URA3* gene (*ura3*) 1070–1071, 1081
- v**
- Varkud satellite (VS) ribozyme 33, 500
- vbFRET software package, smFRET analysis 507
- Vienna RNA package/Web Services/Websuite 549, 620–621, 673, 811
- w**
- Watson-Crick base pair 66, 91, 301, 463, 596, 606, 633, 650, 671, 684, 689, 1100
 - disruption of H bonds, UV melting 449
 - NAIM 371
 - non-Watson-Crick base pair 206, 255, 493, 551, 596
 - reverse transcription 1034
 - RNA probing 210–211, 217, 220
 - stacking 549
- Weblogo program 620
- wiggle track file format 826, 829
- Wincott deprotection protocol 359
- wobble base pair 549, 560, 633, 645, 671, 684, 692
- worm-like chain (WLC) model 534
- x**
- X-ray
 - crystallography/structure of RNA
 - – crystallization of RNA 481–495
 - – 2D and 3D RNA structure prediction 633, 642, 644–645, 652, 656
 - – ligand-aptamer complex 1188
 - – molecular dynamics (MD) simulations 688–690, 692–693
 - NAIM 374–375
 - – RNA-metal ion interaction 285, 287, 301, 308
 - – RNA structure probing 206, 211, 241, 264, 269, 301
 - – RNA synthesis 132, 143, 152
 - – RNP 978
 - – SAXS 416
 - – smFRET 502–503
 - film 63, 95, 194, 197, 201, 403, 437, 867, 911, 978, 993, 996, 1000, 1037, 1047, 1092, 1180–1181
 - generation of hydroxyl radicals 301, 421–422
 - small-angle X-ray scattering (SAXS) 407–423
 - – RNA damage 421–422
- y**
- yeast (*Saccharomyces cerevisiae*)
 - mitochondrial ai5 γ group II intron 344
 - poly(A) polymerase 182, 339
 - protein A system 1030
 - pyrophosphatase 10, 234
 - small nuclear ribonucleoproteins (snRNPs), spliceosomal complexes 957
 - – isolation of native complexes 963–971
 - Snu13 protein 982
 - three-hybrid system 899, 1067–1083
 - total RNA 279, 1157–1158
 - tRNA 38, 78, 123, 181, 185, 188–191, 199, 269, 271, 409, 644–645, 655, 700, 939, 979, 981–982, 984, 989–992, 994–996, 998, 1007, 1019, 1021
 - – intron-containing 923
 - U3 snoRNA/snoRNP 983, 999, 1002
 - U5 snRNA 996
 - [U4/U6.U5] tri-snRNP 996
- z**
- zero-length crosslink 1030–1031

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.