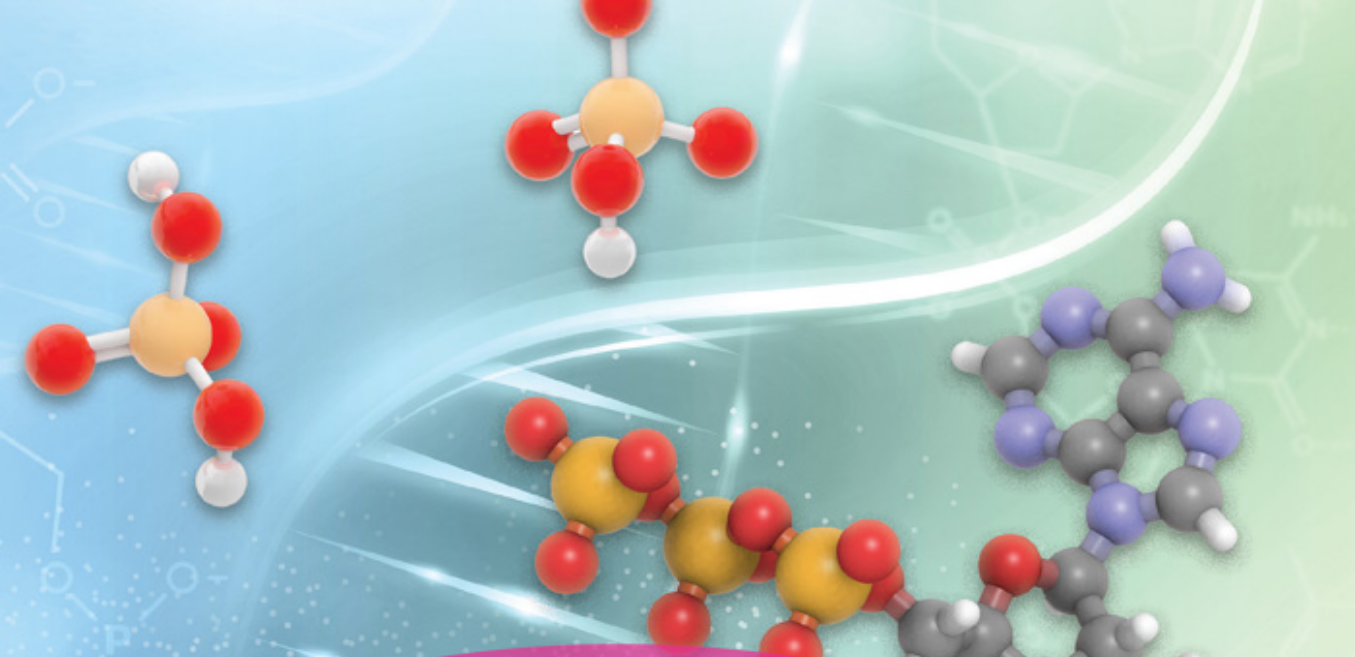


Chemical Biology

The Chemical Biology of Phosphorus

Christopher T. Walsh

ROYAL SOCIETY OF CHEMISTRY

The Chemical Biology of Phosphorus

Chemical Biology

Editor-in-chief:
Kira J. Weissman, *Lorraine University, France*

Titles in the series:
1: High Throughput Screening Methods: Evolution and Refinement
2: Chemical Biology of Glycoproteins
3: Computational Tools for Chemical Biology
4: Mass Spectrometry in Chemical Biology: Evolving Applications
5: Mechanisms of Primary Energy Transduction in Biology
6: Cyclic Peptides: From Bioorganic Synthesis to Applications
7: DNA-targeting Molecules as Therapeutic Agents
8: Protein Crystallography: Challenges and Practical Solutions
9: Oxidative Folding of Proteins: Principles, Biological Regulation and Design
10: Chemical and Biological Synthesis: Enabling Approaches for Understanding Biology
11: Synthetic Glycomes
12: The Chemical Biology of Sulfur
13: The Chemical Biology of Phosphorus

How to obtain future titles on publication:
A standing order plan is available for this series. A standing order will bring delivery of each new volume immediately on publication.

For further information please contact:
Book Sales Department, Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge, CB4 0WF, UK
Telephone: +44 (0)1223 420066, Fax: +44 (0)1223 420247,
Email: booksales@rsc.org
Visit our website at www.rsc.org/books

The Chemical Biology of Phosphorus

By

Christopher T. Walsh
Stanford University, USA
Email: cwalsh2@stanford.edu

ROYAL SOCIETY OF CHEMISTRY

Chemical Biology No. 13

Print ISBN: 978-1-83916-202-2
PDF ISBN: 978-1-83916-231-2
EPUB ISBN: 978-1-83916-232-9
Print ISSN: 2055-1975
Electronic ISSN: 2055-1983

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

The Royal Society of Chemistry is a charity, registered in England and Wales, Number 207890, and a company incorporated in England by Royal Charter (Registered No. RC000524), registered office: Burlington House, Piccadilly, London W1J 0BA, UK, Telephone: +44 (0) 20 7437 8656.

For further information see our web site at www.rsc.org

Printed in the United Kingdom by CPI Group (UK) Ltd, Croydon, CR0 4YY, UK

Dedication

Diana

Allison

Thomas

Sean

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

Preface

Adult humans contain about 2 pounds (approximately 0.8 kilograms) of phosphorus. While this is less than 1% of the total mass of an average 70 kg person, phosphorus is critical for all life on earth. Phosphorus in biology is almost exclusively in the P^V [+5] oxidation state and rarely changes oxidation states. Essentially all the phosphorus atoms in organisms are present as anions of the fully oxidized inorganic phosphoric acid (H_3PO_4), its phosphoric anhydrides, and phosphate monoesters and phosphate diesters. Some 85% of body phosphorus is present as the solid inorganic calcium phosphate salts of bone and teeth. The remaining 15% (about 120 grams) is distributed among soluble inorganic phosphate and organophosphate esters and anhydrides.

The most prevalent and perhaps most central is adenosine triphosphate (ATP), coming in at an inventory of about 75 grams in an adult human. The bis-phosphoric anhydride linkages in the triphosphate side chain are activated thermodynamically, but sufficiently stable kinetically as the tetra-anion at physiological pH values that ATP is the major cellular energy currency. Evidence of its metabolic centrality to power almost every mechanical and chemical process in living organisms is the observation that organisms make and use their body weight in ATP *every day*. For humans that could be up to 75 kilograms, representing about a thousand-fold turnover of the ATP pool daily, 99.9% of which is used to power all the nonequilibrium processes that keep cells and organisms alive. By the time a human reaches the age of 75 years, he/she will have biosynthesized and spent up to two million kilograms of ATP to stay alive. That is testament to phosphoric anhydride chemical biology.

Depending on whether one is energy-centric or information-centric, the other prime or more prime type of phosphate linkage in organisms is the phosphodiester linkage. Most notably, the only covalent linkage holding the monomer nucleotides together in both RNAs and DNAs is the internucleotide 3′,5′-phosphodiester bond. The fidelity of information from

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

generation to generation stored in DNA and the transcriptional fidelity of information transfer from DNA to RNA on the way to all the proteins of proteomes depends on the kinetic stability of the internucleotide phosphodiester bonds, each of which is a monoanionic in biological milieus.

Both RNA and DNA editing, from RNA splicing to restriction endonuclease cuts on double stranded DNA to clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR associated protein 9 (Cas9) manipulations of genomic DNA, involve enzyme-mediated internucleotide phosphodiester phosphoryl group transfers.

The importance of being anionic is at the center of phosphate chemical biology. The negative charge on inorganic phosphate, phosphoric anhydride, phosphodiesters, and phosphomonoesters, so essential to all their biological roles, stems from the three p*K*as of the parent phosphoric acid H_3PO_4. With pKa_1 at pH 2, and pKa_2 at pH 7.2, inorganic phosphate at physiological pH is approximately 50% the monoanion and approximately 50% the dianion. When glucose is enzymatically phosphorylated to glucose-6-phosphate it is trapped inside cells because it is a mix of monoanion and dianion and faces an energy barrier for diffusion out through the lipidic cell membrane. The internucleotide bonds of RNA and DNA bear one negative charge at every phosphorus atom, making them polyanions, which both affects their intrinsic stability as repositories of information and promotes electrostatic interactions with cations, large and small.

Phosphorus chemical biology thus underlies most of life's reactions and processes, from the covalent bonds that hold RNA and DNA together to the making and spending of 75 kg of ATP every day to run almost all metabolic and mechanical events in cells, from ion pumping to protein synthesis, to protein motors moving cargoes to microtubules and other protein polymers that give shape to cells from moment to moment.

Starting from inorganic phosphate, the most abundant source of phosphorus on the planet, in the ionization state of the phosphoric acid dianion, we examine the chemical consequences of two forms of dual or orthogonal reactivity. First, is that phosphate anions are electrophilic at the central phosphorus atom and nucleophilic at the peripheral oxy anions, giving polar opposite modes of reactivity in organismic metabolism. Second, the dehydrative condensation of two molecules of phosphoric acid yields the thermodynamically activated phosphoric anhydride linkage that is also kinetically stabilized at physiological pH, in part by the phalanx of surrounding oxyanion negative charges. This second duality, thermodynamic activation with unusual kinetic stability, allows ATP and congeners to power otherwise unfavorable equilibria as the side chain phosphoric anhydride linkages are hydrolyzed.

Phosphate chemistry is so pervasive in biology that it undergirds almost every contemporary topic in biochemistry, molecular biology, and cell biology. Phosphodiester chemistry enables RNA splicing and CRISPR–Cas9 editing of genomes. Phosphoryl transfers arising from the attack of nucleophilic substrates on Pγ of ATP (cleaving the Pβ–O–Pγ anhydride linkage) both dominate low molecular weight carbohydrate metabolism and carry the

flux of all 520 protein kinases of the human kinome. Attacks by cosubstrate nucleophiles instead at Pα of ATP, congeneric NTPs, and on Pα of 2′deoxy ATP and dNTP congeners to cleave the Pα–O–Pβ phosphoric anhydride side chain, constitute nucleotidyl transfers that are the dominant reactivity mode for building or replicating RNAs, DNAs, proteins, oligosaccharides, and membrane phospholipids. The preference for nucleotidyl transfers to build energetically unfavorable condensed biopolymers may arise from enzymatic hydrolysis of the phosphoric anhydride bond of inorganic pyrophosphate coproduct in every chain elongation step.

Some 30% of cellular ATP may be spent in pumping ions, in and out of cells or into and out of subcellular organelles (protons into vacuoles, calcium ions into the endoplasmic reticulum). ATP and guanosine triphosphate (GTP) are spent for assembly and disassembly of microtubules, for powering motors that transport protein cargoes from one part of the cell to another, for unwinding DNA and RNA by helicases, for ATP-dependent cleavage of short-lived proteins by proteasomes, and for a plethora of other cellular reactions. Cyclic nucleotides, from the classical 3′,5′ cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP), to dicyclic-GMP and dicyclic-AMP, to the mammalian hybrid cyclic nucleotide cyclic guanosine monophosphate–adenosine monophosphate (cGAMP) that signals foreign double-stranded DNA in host cell cytoplasm, form the basis of low molecular weight signaling molecules.

Over the past three decades phosphoproteomics, posttranslational modification of protein side chain residues, has taken center stage in phosphoryl group biology. Some 270 000 serine-, threonine-, and tyrosine-protein side chains are predicted as protein kinase target phosphorylation sites in the global human proteome. However, these three residues are joined by six others (aspartate, glutamate, cysteine, histidine, arginine and lysine) that are undercounted or missed entirely in the standard mass spectrometry workflows to identify acid-stable phosphopeptides. This is *not* a monograph centered on the canonical phosphoproteome and its myriad biological effects, but it does examine the noncanonical phosphoproteome of the remaining six protein residues. The chemistry of those six phosphorylated residues illustrates catalytic *vs.* regulatory roles, thermodynamic and kinetic stability, and the chemical logic for protein phosphatases to reverse phosphorylation states for only some of the nine amino acid side chains in full phosphoproteomes.

In a real sense the approach of this monograph follows the utterance of Alexander Todd, 1957 Nobel laureate in chemistry: "*where there is life, there is phosphorus*". From a few central precepts of reactivity of inorganic phosphate, phosphoric anhydride linkages, phospho-monoesters and diesters, the effects of natural substitution of one of the four oxygens of phosphate by carbon (phosphonates), nitrogen (phosphoramidates), or sulfur (phosphorothioates), the scope and logic of phosphorus chemical biology is developed to examine why life depends on phosphorus.

Christopher T. Walsh

Contents

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

Section II

SECTION I

Introduction to Phosphorus and Inorganic Phosphates

Section I contains two introductory chapters. The first summarizes some of the properties of elemental phosphorus relevant to its chemical biology. It also frames the themes and attributes of phosphorus in the inorganic phosphate oxidation state that permeates every aspect of cell metabolism. Phosphorus joins carbon, hydrogen, oxygen, nitrogen, and sulfur as one of the six most abundant elements requited for life.

Chapter 2 introduces both inorganic phosphate, inorganic pyrophosphate and inorganic polyphosphates and their roles in form and function of organisms. Some 85% of inorganic phosphates are deposited in bones and teeth enamel while the other 15% have myriad roles in metabolic biology. Pyrophosphate and the higher order linear polyphosphates introduce the thermodynamically activated but kinetically stable P–O–P anhydride bond as the key functional group in phosphate chemical biology.

There are four properties of phosphorus which form the backbone themes of its chemical biology. First, is the observation that phosphorus is largely redox inert in biological systems. Essentially all of its chemistry relevant to biological function is in the [+5] oxidation state. Second, over the eons of global oxygenation of the earth planetary phosphorus in surface rocks has accumulated as the most oxizided P^V form, inorganic phosphoric acid. The first two pK_a values of inorganic phosphoric acid are at pH2 and pH7, so phosphates can exist as divalent anions complexed with neutralizing cations.

Third, inorganic phosphate can react in two modes in biology. It can behave as an electrophile at phosphorus, or it can behave as an oxygen nucleophile, *e.g.* as the phosphate dianion. This orthogonal reactivity allows

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

great versatility in the roles of phosphate and to phosphoric anhydrides in biology. The fourth key property is evinced in the phosphoric anhydride linkages of inorganic pyrophosphate and inorganic triphosphate, and is carried over to the triphosphate side chains of ATP and congeneric nucleoside triphosphates. The P–O–P–O–P side chain in ATP is tetra-anionic at neutral pH and thereby kinetically stable. The two phosphoric anhydride [P–O–P] bonds are thermodynamically activated for transfer of each P as an electrophile. This duality of thermodynamic activation with kinetic stability enables life.

CHAPTER 1

Introduction to Phosphorus Chemical Biology

Alexander Lord Todd, 1982

"I would guess that if life exists anywhere else in the universe it will do so only on a planet on which phosphorus is readily available."[6]

1.1 The Element: Discovery, Abundance, Valence States

Adult humans contain about 2 pounds (0.8 kg) of phosphorus, almost exclusively in the form of the fully oxidized phosphate atoms. Some 85% is deposited in bones and teeth as calcium phosphate in the form of crystalline hydroxyapatite. The remainder is in soft tissues, with the serum concentration of soluble inorganic phosphate (HPO_4^{2-}) maintained within narrow limits at 1.1 to 1.4 mM. At these levels, phosphorus is slightly less abundant than calcium, and about half the levels of bodily nitrogen (1.8 kg). The other three abundant elements H, C, and O come in at 7 kg, 16 kg, and ~43 kg, respectively. The ~1% abundance of phosphorus by mass belies the essential roles of phosphates. Lord Alexander Todd, British chemist and Nobelist who carried out the first syntheses of nucleotides, remarked "Where there is life, there is phosphorus".[1]

Phosphorus sits just below nitrogen in the periodic table, as atomic number 15, with a molecular weight of 30.974 atomic mass units. While ^{31}P is the most abundant and stable isotope, the short lived radioactive ^{32}P has found extensive use in radiotracer studies (half-life ~14 days) and the stable heavy atom isotope ^{33}P also has its uses as a tracer for phosphorus metabolism.

The 15 electrons of phosphorus fill orbitals in the order $1s^2 2s^2 2p^6 3s^2 3p^3$. A simplistic description of the available valence electrons for bonding (Figure 1.1)

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

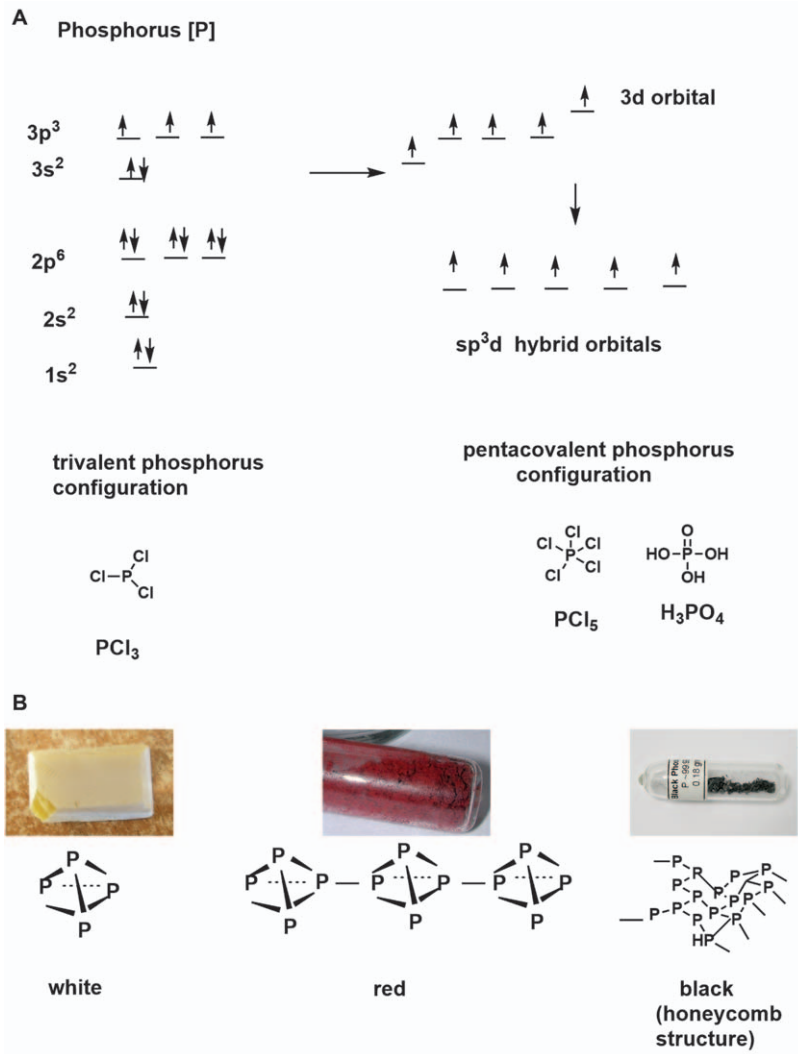


Figure 1.1 (A) Orbital filling diagram for phosphorus: configurations for trivalent phosphorus (P^{III}) or pentavalent phosphorus (P^{V}). (B) Three allotropic solid forms of elemental phosphorus (Left to Right: White, Red and Black phosphorus). B – Left image courtesy of BXXXD, https://en.wikipedia.org/wiki/File:Wei%C3%9Fer_Phosphor.JPG, under the terms of a CC BY-SA 3.0 license, https://creativecommons.org/licenses/by-sa/3.0/deed.en. B – Right image courtesy of https://en.wikipedia.org/wiki/File:Black_Phosphorus_Ampoule.jpg, under a CC By-SA 3.0 license, https://creativecommons.org/licenses/by-sa/3.0/deed.en.

indicates that trivalent compounds engage the 3p electrons while pentavalent phosphorus involves hybridized sp^3d orbitals. Phosphorus trichloride and pentachloride are classic abiotic compounds of phosphorus in the P = [+3] and

P = [+5] oxidation states, respectively. At the other end of the oxidation spectrum, phosphine, PH_3, is a fully reduced trivalent phosphorus compound, but now with P = [−3] oxidation state.

Phosphorus was first isolated as a pure element by Hennig Brand in Germany in 1669 in pursuit of alchemical studies.[2] Concentration of human urine followed by heating to high temperature led to distillation of white phosphorus. This was a luminescent and phosphorescent species. The name phosphorus is from the ancient Greek combination of phos (light) and phoros (bearer) There are some 10 known allotropes of elemental phosphorus, from the P_4 tetrahedron of white phosphorus to polymeric chains of red, black, and other phosphorus allotropes. The red phosphorus allotrope is not phosphorescent.[3] Three of those allotropes are depicted in Figure 1.1B

Phosphorus is thought to be formed only in the hot interiors of large stars such as red giants, by gain of a proton from ^{31}silicon. The relative under-abundance of phosphorus in the universe (and on earth) has raised questions about whether it was sufficiently available to be present in the earliest life forms, but only arsenic has been proposed as an incomplete alternative to fulfill its myriad roles and it fails the test of stability of arsenate diesters in aqueous solutions.

Unlike nitrogen, which can cycle between [−3] and [+5] oxidation states, and neighboring sulfur, with [−2] to [+6] oxidation states, phosphorus in biology is *almost* exclusively in the [+5] oxidation state as inorganic phosphate with four oxygens surrounding the central phosphorus atom. One of the P–O bonds is formally a P=O double bond. The other three are P–OH bonds in phosphoric acid (H_3PO_4) (Figure 1.2). The first pK_a is at pH 2.12 (50% ionized to the indicated resonance stabilized monoanion at that pH). The second pK_a is 7.12, so at neutral pH values in organisms there is a mix of monoanionic and dianionic inorganic phosphate species. The third pK_a of phosphoric acid, at pH ∼ 12, is outside the range of physiological pH values so there is no significant biological role for inorganic phosphate trianions. (In most microenvirons the undissociated phosphoric acid is also not relevant, although at stomach pH values of ∼2, the undissociated phosphoric acid species will be relevant.) We will see that the monoanionic and dianionic species of inorganic phosphate are essential features of the chemical biology of phosphorus. Most of the global cycle of phosphorus occurs in the movement of phosphate species.

Inorganic phosphoric acid and its physiologically relevant monoanions and dianions exhibit tetrahedral geometries (Figure 1.2B), with one formal P=O double bond and three P–O single bonds. The three single P–O bonds are chemically equivalent. Keeping in mind the tetrahedral geometry, we will typically not specify it explicitly in most of the ChemDraw structural formulas throughout the text except when stereochemical issues arise, as in phosphorothioates in Chapter 12. Stereochemistry is otherwise cryptic in transformations of phosphate with its three equivalent oxygens.

By far the most abundant form of phosphorus available in the earth's crust is inorganic phosphate-containing rock, the most oxidized form of pentavalent phosphorus, formed over eons from weathering in the oxygen-

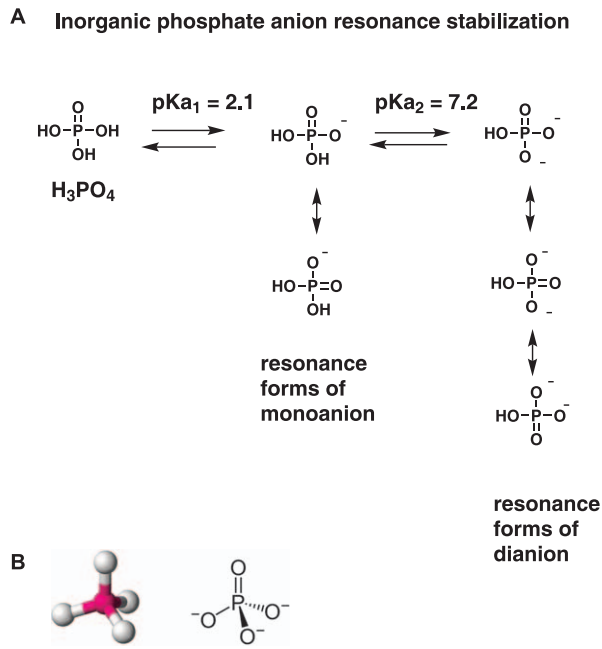


Figure 1.2 (A) pK_a values for the first and second pK_as (pH value at 50% ionization) for phosphoric acid. Note that monoanions and dianions are stabilized by resonance delocalization of the charges. (B). Phosphates have tetrahedral geometry.

containing planetary atmosphere. About 50–200 million tons of phosphate-rich rock are harvested each year and converted abiotically to pure phosphoric acid. The inorganic acid and its salts are used for fertilizers, animal feed, flame retardants, weed killers, detergents, and other industrial uses.

The daily dietary input of phosphate for a human adult can range from ~700 mg (estimated daily amount needed) to ~1500 mg per day. Excess phosphate is eliminated by kidney excretion.[4] Failure of the kidneys to keep up with phosphate loads leads to pathological hyperphosphatemia conditions (and eventual deposition as insoluble calcium phosphates).

1.2 Chemical Biology of Phosphorus

Inorganic phosphate, typically as fluorapatite [$Ca_5(PO_4)_3F$] or hydroxyapatite [$Ca_5(PO_4)_3OH$], in limestone and mudstone rocks can vary from 4 to 30% in those rocks.[5] That abundance, coupled with the observation that over 99% of phosphorus biology occurs in the P = [+5] oxidation state, makes inorganic phosphate and its compounds the center of phosphorus chemical biology. Throughout this volume, we take the simplifying view that the chemistry of inorganic phosphate, its anhydrides, monoesters and diesters with alcohol groups of metabolites, both of low molecular weight and as macromolecules, dominate and determine the myriad roles of phosphate in biology (Figure 1.3).

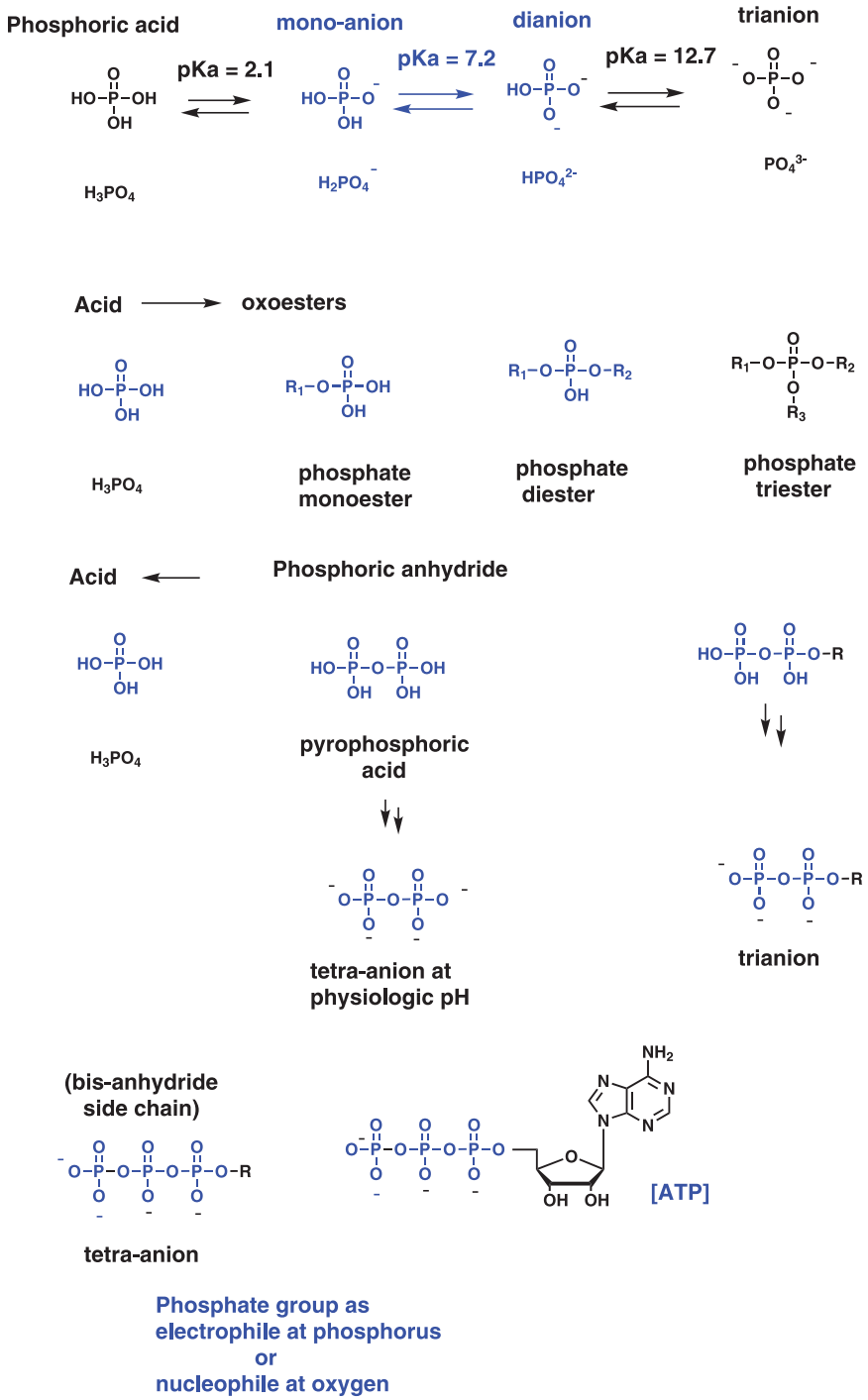


Figure 1.3 The dianionic form of inorganic phosphate and phosphate monoesters, the monoanionic form of phosphodiesters, and the tetra-anionic form of inorganic pyrophosphate are dominant forms in phosphorus chemical biology.

1.2.1 Where There is Life There is Phosphorus

The opening three-line quote at the start of this chapter is from Alexander Todd a British organic chemist, professor at Oxford in the mid twentieth century who pioneered synthetic methods for making phosphate-containing molecules central to biology. The citation is to a paper ('Where there's life there's phosphorus') delivered by Todd in 1981 at a conference in Japan.[1] In a review article in 2010 Bowler *et al.*[6] analyze Todd's arguments and provide additional perspective 30 years on. Among the properties embedded in phosphate chemistry essential to biology are the following:

1. DNA information retention from organism generation to generation requires *ultrastable phosphodiester internucleotide* bonds, that are nonetheless rapidly repairable when necessary.
2. If life started *via* an RNA world then *phosphodiesters* were key covalent linkages stitching together bits of chemical information.
3. Lipid membrane barriers define an inside and outside for cells and organisms and allow establishment of transmembrane electrochemical potentials. *Phospholipid diester head groups*, attached covalently to lipid tails, are the universal building blocks for biological membranes.
4. To build skeletons strong enough to support large organisms, most notably mammals, *calcium phosphate salts* comprise the building block for bones and teeth
5. Organisms constantly need energy supplies to survive. The energetics of life are built around *phosphoric anhydride* chemistry
6. Regulation of information flow and signaling over different time regimes is dominated by *two parallel signaling regimes involving phosphate chemistry*: (a) a set of low molecular weight second messengers: cyclic adenosine monophosphate (cAMP), cyclic guanosine monophosphate (cGMP), di-cyclic GMP, di-cyclic GMP–cAMP; (b) protein posttranslational phosphorylations.

We take up each of these facets of phosphate group chemistry in biological contexts through the subsequent chapters of this book. Life is certainly more than phosphoryl group chemistry but no other element in the periodic table exhibits chemical properties that would allow life to be constructed by a phosphorus replacement. (Pentavalent arsenic has some replacement properties, but its esters are chemically labile in aqueous physiological media).

1.2.2 Inorganic Phosphate, Phosphoric Anhydrides to Nucleoside Triphosphates (NTPs)

The chemical biology of inorganic phosphate is pervasive throughout all forms of life. A classic review article by Westheimer in 1987 extolled the properties of phosphates with the title '*Why Nature Chose Phosphates*' and is

still compellingly relevant more than 30 years later.[7] Furthermore, that review occurred just as the enormous reach of phosphoprotein biology was becoming apparent, reinforcing the extensive reliance on phosphate chemistry in both low molecular weight and high molecular weight metabolites. We will devote some time to the fact that the only covalent linkages in both DNA and RNA biopolymers are the internucleotide phosphodiester bonds, each bearing a net negative charge.

Inorganic phosphate plays three major roles in biology as we will elaborate in subsequent chapters. The first is that it is poorly soluble as the calcium salt and precipitates out of supersaturated solutions as hydroxyapatite during bone and teeth enamel formation.[8] Higher pH redissolves phosphate dianions and is the basis for the dynamics of bone resorption.

The second aspect of phosphate chemistry is that the central P atom in the [+5] oxidation state is electron deficient. It reacts as an electrophile with nucleophilic atoms that can penetrate the phalanx of the four surrounding oxygens of the phosphate group, especially when complexed with cations. The capture of inorganic phosphate at its central electrophilic phosphorus atom by an oxyanion of adenosine diphosphate (ADP) is the fundamental chemical reaction enacted by proton-conducting, membranous adenosine triphosphate (ATP) synthases. Humans make and then turn over their body weight in ATP (~70–80 kg) every day.[9] The electrophilicity of the P atom in inorganic phosphate is central to life.

The third attribute, alluded to in the previous paragraph is that inorganic phosphate at physiological, neutral pH values is a mix of monoanion and dianion. Those oxyanions are biological nucleophiles (as also noted for ADP just above). Perhaps the most visible biological manifestation of that phosphate oxyanion nucleophilicity is in the catalytic cycle of phosphorylase enzymes. Such phosphorylases transfer electrophilic fragments of substrates to inorganic phosphate (Chapter 9). The mobilization of glucosyl units from starch or glycogen storage polysaccharides by phosphorylases releases glucose-α-1-phosphate units.

1.3 Inorganic Pyrophosphate

In many ways the simple dehydrative condensation of two molecules of inorganic phosphate to the phosphoric anhydride bond in pyrophosphoric acid (PPi) illuminates the features of the most important phosphate functional group in all of biology (Figures 1.2 and 1.3). Because of the pK_a values of phosphoric acid, inorganic pyrophosphate at neutral pH exists in large part as the tetra-anion. That phalanx of negative charges helps shield the central phosphoric anhydride linkage from adventitious attack by water. Thus, although the equilibrium constant greatly favors pyrophosphate tetra anion hydrolysis to two molecules of inorganic phosphate, kinetically the phosphoric anhydride linkage is stable enough for pyrophosphate tetra-anion to be a diffusible metabolite in all cells.

The dichotomy of thermodynamic activation for phosphoryl group transfer to potential nucleophiles (including water) coupled to kinetic stability is *the central feature of phosphoric anhydride biology*, from inorganic pyrophosphate to ATP (Figure 1.3), and includes other nucleoside triphosphates (NTPs) as well as the 2′-deoxynucleoside triphosphates (dNTPs) building blocks for DNA. By comparison acetic anhydride, a simple carbon-based anhydride, is similarly activated thermodynamically, in this case as an acetyl transfer reagent, but it is not kinetically stable in neutral physiological aqueous media. It would not have existed intact long enough to serve as a carrier of activated acetyl groups for metabolic transfers (That turns out to be the role of the thermodynamically activated but kinetically stable acetyl thioester link in acetyl-*S*-coenzyme A[10]) The pK_as of the polyprotic phosphoric acid are essential to life.

While free pyrophosphate is rarely used by contemporary multicellular organisms as a phosphoryl donor, it is cleaved in almost every cell by the action of the ubiquitous hydrolase inorganic pyrophosphatase. The net concentration of [PPi] in cells is often below 10^{-6} M for this reason, ensuring that the thermodynamically activated P–O–P bond has been cleaved to draw off the equilibrium of any PPi-forming enzyme reactions in the forward direction (see Chapter 4).

1.4 Phosphoric Anhydride Side Chains in ATP, NTPs, and 2′deoxyNTPs

The apotheosis of phosphoric anhydride chemistry in biological systems is found in adenosine 5′-triphosphate (ATP) (Figure 1.4), related NTPs that are the building blocks for RNA biogenesis, and the corresponding 2′-deoxyribonucleoside triphosphate precursors to DNA. ATP and congeners have two P–O–P anhydride bonds in their triphosphate side chains, offering three electrophilic P atoms (Pα, Pβ, and Pγ) for thermodynamically favored attack by cellular nucleophiles. Yet, ATP is kinetically stable to 100 °C, so even thermophilic microbes use it as major cellular energy currency.[10]

The equilibrium constant (K_{eq}) values for cleaving either of the two P–O–P anhydride bonds in ATP and congeners is about 10^5. Thus, if a kinetic condition for rapid cleavage of one of the two phosphoric anhydride bonds in ATP or NTPs can be found, then ATP or NTP side chain cleavage can be used to drive otherwise unfavorable equilibria. This is the job of hundreds of enzymes in cells. Those catalysts lower the energy barriers for cosubstrate nucleophiles to attack any one of the three electrophilic phosphorus atoms in the ATP, NTP, or dNTP side chain. They thereby couple a highly favorable phosphoric anhydride bond fragmentation to a huge variety of other chemical transformations in metabolism.

We will take up the distinct regiochemistries of enzyme-directed ATP, NTP, and dNTP phosphate side chain cleavages to categorize large swaths of metabolism by simple chemical classification. Attack of an electron-rich

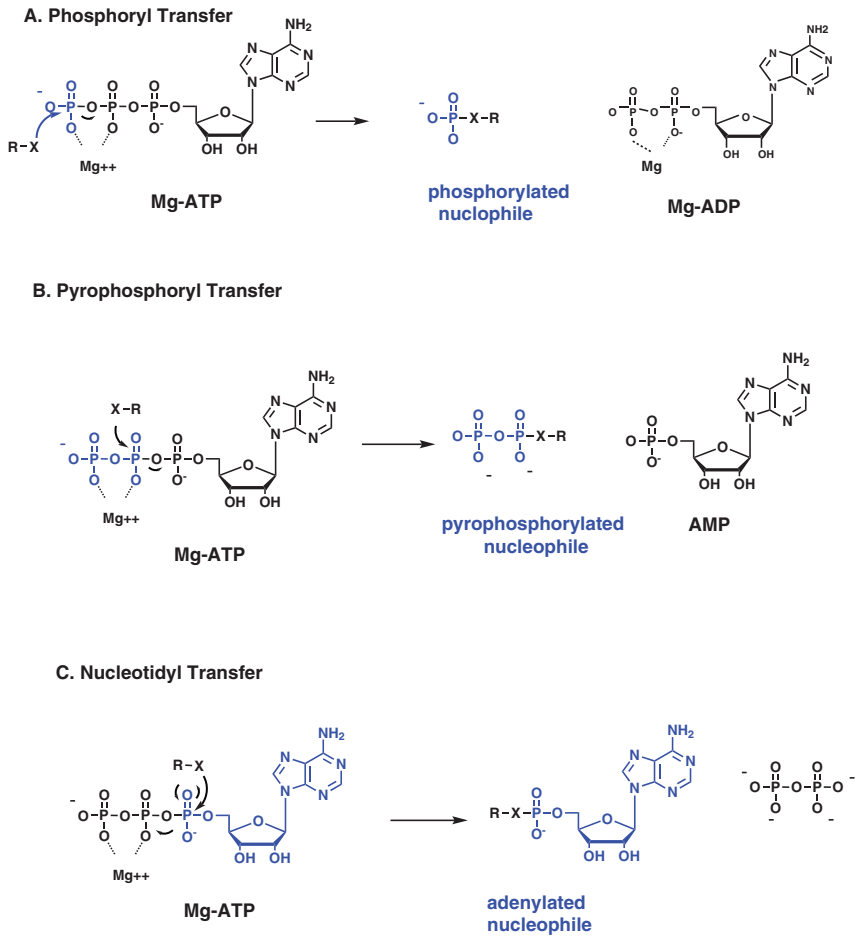


Figure 1.4 Three modes of attack of cosubstrate nucleophilic atoms on one of the electrophilic phosphorus atoms in the side chain of Mg–ATP. (A) phosphoryl transfer, (B) pyrophosphoryl transfer, and (C) nucleotidyl transfer involve attack at Pγ, Pβ, and Pα, respectively.

nucleophile at the outside Pγ of ATP, NTP, or dNTP constitutes a *phosphoryl* transfer (Figure 1.4). Attack on the central Pβ-atom constitutes a *pyrophosphoryl* group transfer to the incoming nucleophile (Figure 1.4). Finally, attack on Pα is generically a *nucleotidyl* transfer (for ATP it is an adenylyl transfer). We will note that nucleotidyl transfers generate PPi as coproducts that get further cleaved by inorganic pyrophosphatase action to pull further on forward equilibria.

These nucleotidyl transfers are the backbone strategies for monomer unit activation and/or incorporation into the growing chains of proteins, RNA, and DNA. Internal nucleotidyl transfers are the chemical mechanisms that enable the biology of cyclic nucleotides and cyclic di-nucleotides as intracellular messenger metabolites. Repair of DNA single strand and even

double strand breaks is an ongoing activity in every cell in humans and other organisms. Tandem nucleotidyl transfers are at the center of the DNA ligases that repair such strand breaks to protect the integrity of genomes.

The enzymes catalyzing ATP, NTP, or dNTP cleavages without coupling of the $-PO_3^{2-}$ group in flight to a substrate *other than water* are termed ATPases or more generally NTPases [*e.g.* guanosine triphosphatases (GTPases)]. At first glance they look to be a hydrolytic waste of an activated P–O–P anhydride bond. However, these ATPases are all coupled to some conformational changes in their protein substrates to drive component motions, marking of proteins, *e.g.* for accelerated degradation, or to drive mechanical forces in cells. ATP and its paired side chain phosphoric anhydrides are so central to cellular energy metabolism that humans take their daily inventory of ~75 g of ATP, make 1000 times that amount (~75 kg!) and spend it all, every day. In many ways this is the core of phosphorus chemical biology.

1.5 Families of Thermodynamically Activated Phosphoryl Derivatives

While phosphoric anhydride linkages in nucleoside triphosphates are the central phosphoryl group transfer currency in cells, other phosphoryl functional groups are also thermodynamically activated *and* sufficiently kinetically stable to serve as phosphoryl group donors.[10] Three of them are doubly activated for both phosphoryl group transfer and transfer of the other moiety as an activated electrophile. The classes of metabolites are mixed sulfuric–phosphoric anhydrides, represented by 3′-phospho-5-adenosine phosphosulfate (PAPS), acyl phosphates such as the simple acetyl phosphate, and phosphorylated enols of which phosphoenolpyruvate is the paradigm (Figure 1.5).

All three can be attacked by one of the terminal oxyanions of ADP in the presence of particular enzyme catalysts to make ATP, thus evincing phosphoryl group potential. More usefully, PAPS is the cellular donor of the activated sulfuryl group to oxygen and nitrogen cosubstrate nucleophiles. Acyl phosphates are donors of activated acyl groups in a variety of cellular acyl transferase active sites. While phosphoenolpyruvate (PEP) is thought of mainly as a phosphoryl group donor, the enolpyruvyl group is transferred intact in bacterial assembly of aromatic scaffolds and in cell wall peptidoglycan construction.

1.5.1 Routes to ATP Biosynthesis

Given the thousand-fold turnover of the existing bodily inventory of 75 g of ATP every day by humans, we will note there are only three common routes to make that thousand-fold increase in daily ATP pools. The membrane-embedded ATP synthase in mitochondria, acting as a conduit for protons running down their electrochemical potential, makes about 94–95% of the 75 kg of daily ATP under fully oxygenated conditions. This requires the access of pools of ADP and inorganic phosphate in the mitochondria into the ATP synthase active site for net

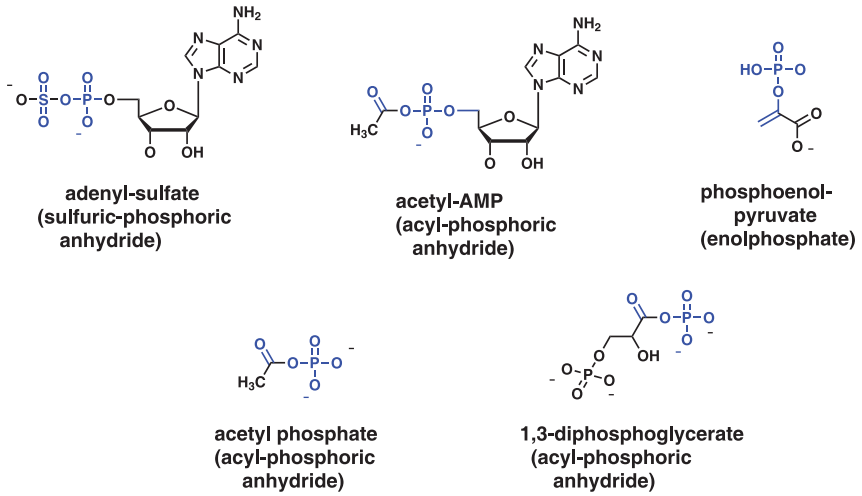


Figure 1.5 Additional phosphoryl-containing functional groups that are utilized in biological group transfer enzymatic reactions: sulfuric–phosphoric anhydrides, acyl–phosphoric anhydrides, and enolphosphates.

dehydrative coupling to make the Pβ–O–Pγ anhydride bond in ATP. About 6% of ATP (~4.5 kg per day) is generated by glycolysis, the ten-enzyme pathway converting intracellular glucose to two molecules of the three-carbon keto acid pyruvate. The two molecules in this pathway that lead to ATP formation are an acyl phosphate (1,3-diphosphoglycerate) and PEP, two of the activated metabolites noted in the preceding paragraph. These three molecules, ATP, PEP, and acyl phosphate, give insights into the relative phosphoryl group transfer potentials of these three classes of thermodynamically activated phosphoryl group scaffolds (Figure 1.5).

1.6 Phosphomonoesters: Alcohols as Nucleophiles in Phosphoryl Transfers

After consideration of the metabolic changes enabled by the thermodynamically activated phosphorylated frameworks, just noted, we turn to phosphate esters. Most discussions of biological phosphates would start with phosphomonoesters—*e.g.* glucose-6-phosphate—and then phosphodiesters—*e.g.* the 3′5′-internucleotide phosphodiester bonds in RNA and DNA—including their routes of formation and breakdown.

Indeed, the capture of phosphoryl groups by alcohol functionalities in cosubstrates dominates sugar metabolism. We believe strongly that a prior discussion of phosphoric anhydride chemical biology then gives the insights necessary to understand how ATP is a phosphorylating reagent for alcohol groups in the above metabolites. This is the essence of the strategy of over 100 low molecular weight kinase enzymes and more than 500 protein kinases in humans.

Phosphomonoester metabolites turn over in a dynamic fashion. The catalyzed hydrolysis of phosphomonoesters back to alcohol and inorganic phosphate are the province of phosphatase enzymes. These include low molecular weight selective phosphatases as well as many kinds of phosphoprotein phosphatases.

1.7 Phosphodiesters: Alcohols as Nucleophiles in Nucleotidyl Transfers

Phosphodiesters contain two alcohol groups with a bridging phosphoryl group. Two of the four oxygens in the sandwiched phosphoryl group are esterified to those alcohols: hence the terminology of phosphodiesters. Two key biologic classes of phosphodiesters are RNA and DNA molecules. The only covalent bonds holding the RNA nucleoside monophosphate units (NMPs) together and correspondingly the 2′deoxyNMP units in DNA chains are the internucleotide phosphodiester bonds (Figure 1.6A). These phosphodiesters span the 3′-alcohol of a ribose unit in one NMP or dNMP to the 5′-OH of the ribose in the next NMP or dNMP group. All of the three billion dXMP base pairs in the human DNA genomes and the billions of XMP base pairs in all the classes of cellular RNAs are formed enzymatically by nucleotidyl transfer enzymes. Those nucleotidyl transferases, known as RNA polymerases or DNA polymerases, respectively, offer up NTPs or dNTPs for attack at P*alpha* by the 3′-OH of the terminal XMP or dXMP residue in the growing RNA or DNA chain.

Analogously, the activation of each of the 20 proteinogenic amino acids to serve as building blocks in the ribosome-based protein biosynthetic nanomachinery starts with nucleotidyl transfers. All the amino acids are activated as mixed-acyl-AMP anhydrides as the first step in aminoacyl tRNA synthetase active sites (Figure 1.6B). The second step is transfer of the now activated aminoacyl groups to cognate tRNAs to be ferried to the ribosomes by chaperone proteins.

Thus, each step in in macromolecular information transfer biology–genome replication, transcription, and translation—involves nucleotidyl transfer strategies. RNA and DNA biogenesis make stable phosphodiester linkages, while protein biosynthesis makes kinetically labile aminoacyl-adenylates that get converted to kinetically stable aminoacyl tRNA oxoester links.

Other aspects of nucleotidyl transfers to be explored include the generation of membrane phospholipids (Figure 1.7A). Most of the $\sim 10^9$ lipid molecules in the plasma membranes of animal cells are such phospholipids, where the phosphorus is present as a phosphodiester. Given $\sim 10^{13}$ cells in a human adult times 10^9 phospholipids per cell crudely computes to $\sim 10^{22}$ phosphodiester linkages in membrane phospholipids in each person. This is a serious biosynthetic task and phospholipid remodeling goes on in all cells.

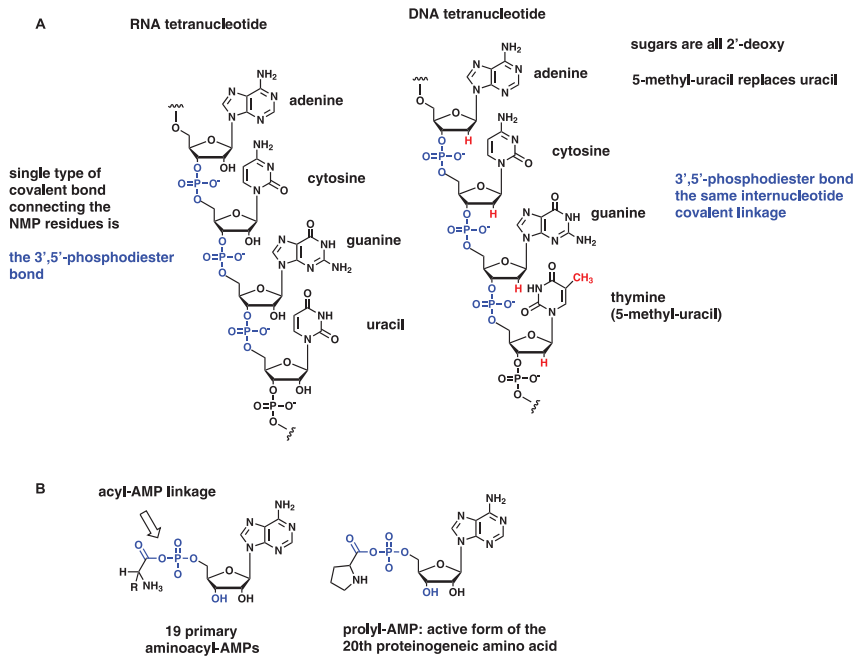


Figure 1.6 (A) 3′,5′-phosphodiesters are the only covalent linkages in both RNA and DNA chains; (B) Aminoacyl-AMPs are thermodynamically activated, kinetically stable forms of the 20 proteinogenic amino acids that serve as monomers for protein biosynthesis.

A second aspect of low molecular weight nucleotidyl transfers that we will dwell on are *intramolecular* nucleotidyl transfers. The enzyme-mediated formation of classic second messengers such as 3′,5′cyclic AMP and 3′,5′-cyclic GMP occur by action of such intramolecular nucleotidyl transferases (Figure 1.7 B). The more recently described di-cyclic nucleotides, including di-cyclic GMP (di-CGMP) and the mammalian cyclic guanosine-adenosine phosphodiesters (cGAMPs) are intramolecular and/or intermolecular variants of that catalytic logic.

As with phosphomonoester metabolites, phosphodiester bonds are cleavable enzymatically. Such phosphodiester hydrolases (phosphodiesterases) typically act in a controlled fashion. For the RNA and DNA polymers as substrates there are a plethora of phosphodiester hydrolases, generally termed nucleases. Some show strict endonuclease or exonuclease specificity. Sequence specific restriction endonucleases from bacteria number in the thousands, selective for recognition of particular oligonucleotide sequences before engaging in hydrolytic internucleotide phosphodiester cuts, in blunt end or staggered cuts on both strands of DNA.

Some exonucleases act in the 5′ to 3′ direction, others in the 3′ to 5′direction. Some are specific for double stranded regions of RNA or DNA substrates, some selective for single strand hydrolytic reactions. Some are

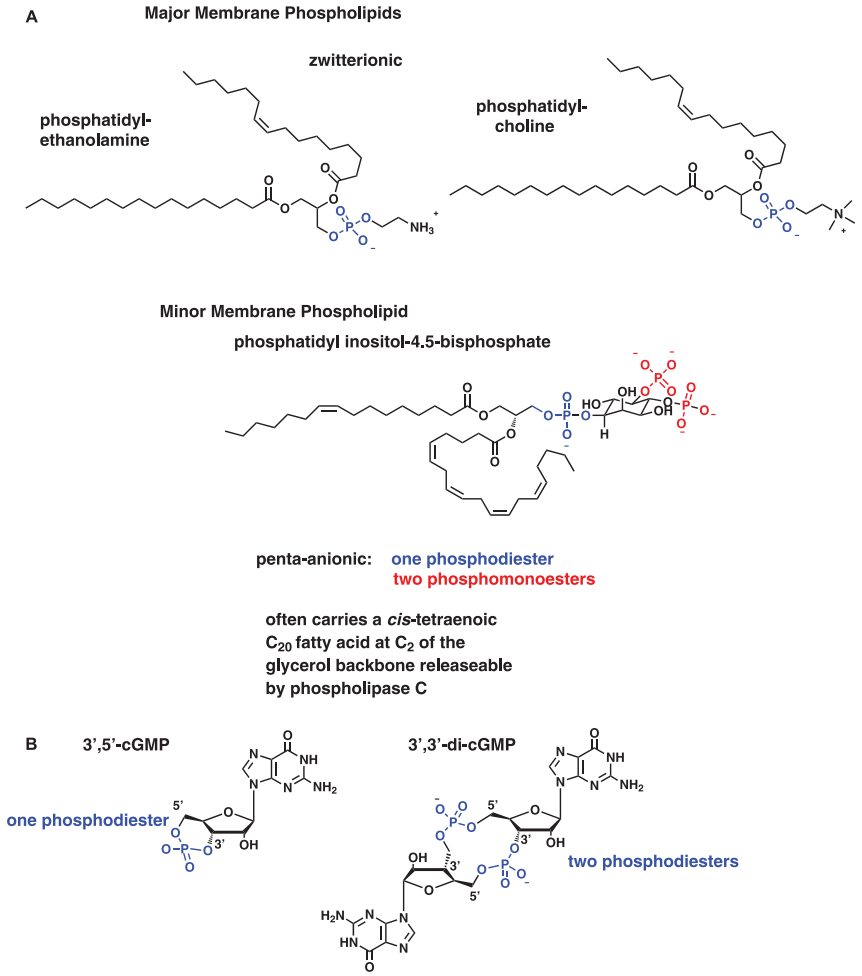


Figure 1.7 (A) Membrane phospholipids also contain phosphodiesters linkages, connecting the polar head groups to the hydrophobic diacylglycerol backbone. Shown are forms of phosphatidyl ethanolamine and phosphatidyl choline that are often bulk membrane phospholipids. Also shown is the minor phospholipid phosphatidylinositol-4.5-bisphosphate, a membrane lipid signaling metabolite. (B) Cyclic nucleotides that function as biological second messengers (hormones are the first messengers) have cyclic phosphodiester linkages. Shown are 3′,5′-cyclic GMP and 3′,3′-dicyclic GMP.

involved in repair or in formation of specific single strand or double strand nicks. Some are protective against foreign DNA; some appear to be purely degradative in function.

There are about a dozen cyclic nucleotide phosphodiesterases in human cells, to effect different physiologic and signaling responses to cAMP and cGMP of varying duration in different tissues. Thus, the universe of

phosphodiesterase enzymes has evolved to perform a wide variety of specific biological functions in both prokaryotic and eukaryotic organisms.

Turning away from nucleic acid substrates towards membrane phospholipids, phospholipase C and phospholipase D family members carry out hydrolytic cleavages on distinct sides of the sandwiching phosphodiester bond.

1.7.1 Three Nucleotidyl Transfers in a Catalytic Cycle of DNA Ligase

DNA ligases, crucial enzymes in DNA repair of both single strand gaps and double strand gaps with 3′-OH and 5′-phophomonoester ends, engage in tandem nucleotidyl transfer strategies. Starting with ATP, the first nucleotidyl transfer is to the ε-NH_2 of an active site lysyl side chain of the enzyme. The second nucleotidyl transfer of that AMP moiety is to the 5′-P on one side of a gapped DNA strand, creating an ADP adduct. The third nucleotidyl transfer is to water as the internucleotide phosphodiester bond is formed in the actual DNA repair step.[11]

The chemical biology strategy is beautifully exemplified in the catalytic logic of DNA ligases. The initial conversion of (1) the Pα–O–Pβ phosphoric anhydride bond in ATP precedes (2) the phosphoramidate N–P bond in the AMP–Lys–enzyme adduct, on the way to (3) the phosphoric anhydride bond in the ADP adduct on the 5′-side of the DNA gap. Finally, step (4) of the tandem nucleotidyl transfers is elimination of AMP in the internucleotide ligation step (Figure 1.8). In aggregate, these four steps show the chemical versatility of nucleotidyl transfers in DNA repair.

The great bulk of the myriad roles of phosphorus in biology occur in the form of phosphate, with four oxygens around the central P = [+5] oxidation state. The oxygens can be unsubstituted in inorganic phosphate, or in anhydride linkages. Those anhydrides can consist of P–O–P bonds in phosphoric anhydrides, mixed acyl phosphoric or mixed sulfuric–phosphoric anhydrides. Alternatively, the phosphate oxygens can be substituted by one or more alcohol groups. Esterification of phosphoric acid to one alcohol gives [R–OPO_3^{2-}] as phosphomonoesters. Esterification to two alcohols [RO–PO_2–OR′] gives the classes of phosphodiesters. (Figure 1.2). The mechanisms for enzyme-mediated phosphoryl group transfers in these different classes of phosphate metabolites have been scrutinized to decipher the underlying chemical logic and general strategy.

1.8 Phosphonates, Phosphoramidates, and Phosphorothioates

There are three additional classes of naturally occurring phosphate derivatives that reflect substitution of one of the four phosphate group oxygens by a carbon, a nitrogen, or a sulfur atom. (Figure 1.9), yielding phosphonate, phosphoramidate, or phosphorothioate scaffolds, respectively.[12]

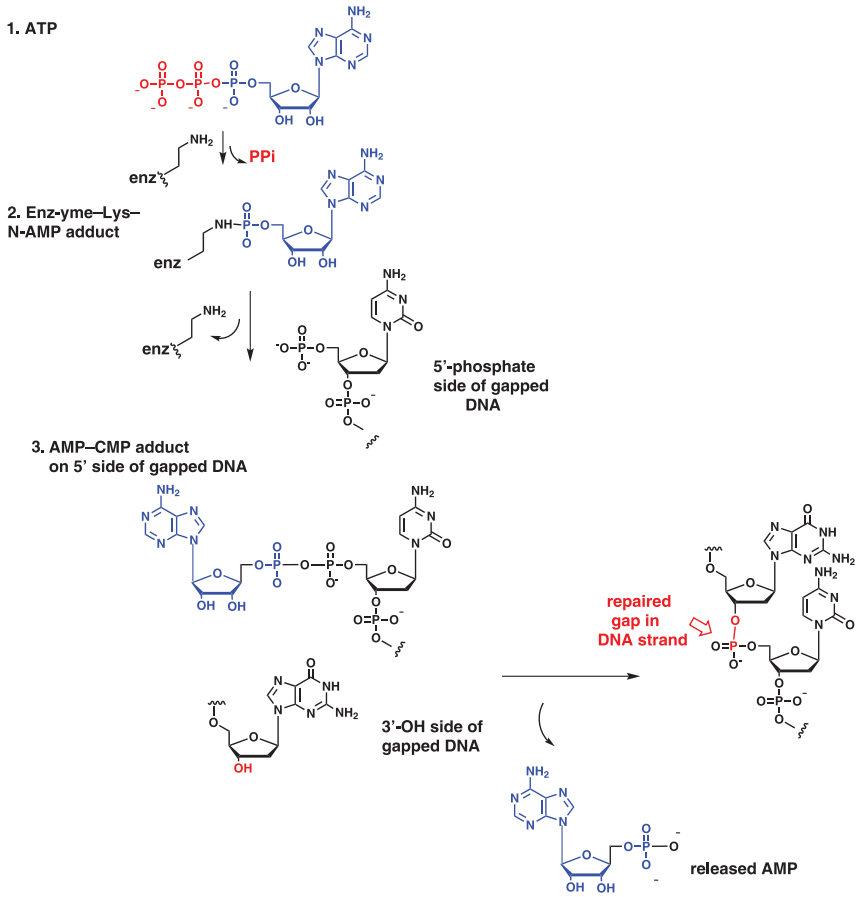


Figure 1.8 Schematic of the role of ATP in repair of DNA strand gaps by the enzyme DNA ligase. Gaps with a 5′-phosphate and 3′-OH are repaired by a successive set of transfers of the AMP moiety of ATP to an active site lysine side chain and then the 5′-side of the DNA gap, as prelude to internucleotide phosphodiester repair.

There are some 200 known microbial metabolites with the direct C–P bonds.[13] The C–P links are chemically stable to hydrolysis, resistant to phosphatase action, and offer producers a strategy for generating stable molecules in hostile microenvironments and for the formation of anti-metabolites as molecular weaponry. There is a single enzymatic route for C–P bond formation in biology, the enzyme PEP mutase, converting the enol phosphate PEP to the C–P direct linkage in phosphonopyruvate. This is the gateway enzyme to all downstream C–P metabolites, including some very rare phosphinates with two C–P bonds and two O–P bonds (Figure 1.9).

The $-N-PO_3^{2-}$ linkage is termed a phosphoramidate. There are only about 50 known phosphoramidate metabolites. They show up mainly in three distinct categories. Five of the twenty proteinogenic amino acids are known

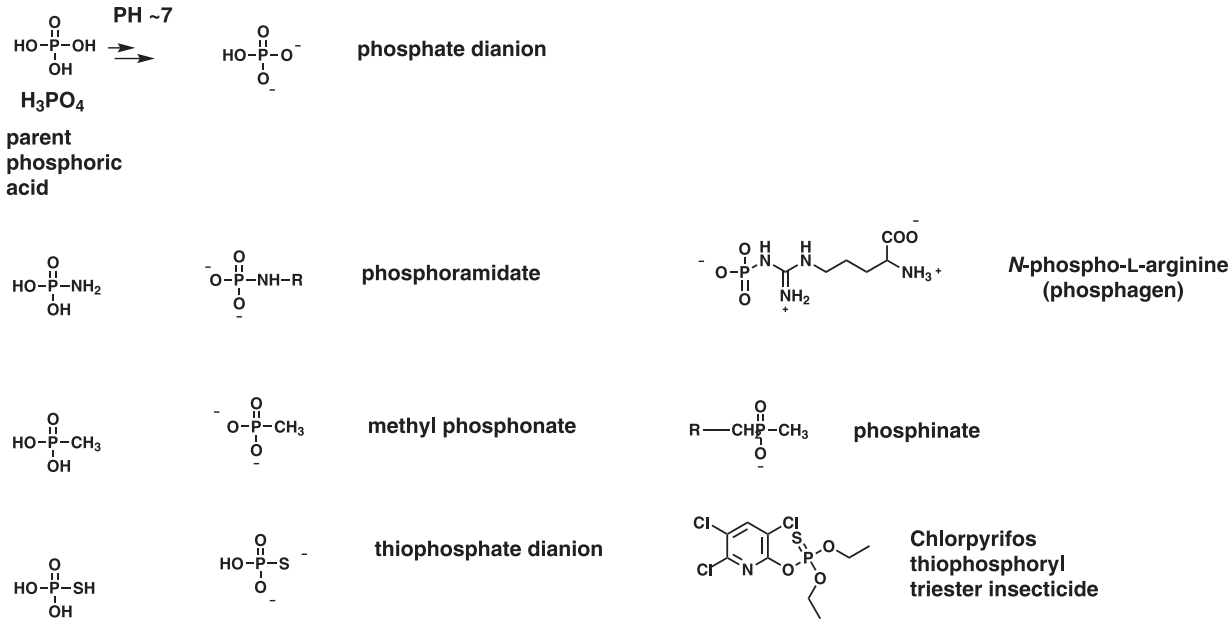


Figure 1.9 Examples of metabolites arising from the formal substitution of one of the four oxygens of phosphoric acid by an N, C, or S atom, generating phosphoramidates, phosphonates, and thiophosphates, respectively.

to undergo enzymatic *N*-phosphorylations on the way to end product metabolites, including phosphoramidate linkages in the capsular polysaccharide of some strains of the human gastrointestinal pathogen *Camphylobacter jejuni.*[14]

The second class of phosphoramidates are –N–PO_3^{2-} bonds in guanidinium scaffolds. Examples include arginine–N–PO_3^{2-} in invertebrates and creatine–N–PO_3^{2-} in vertebrates, including humans. These are so called phosphagens, reservoirs of molecules with high phosphoryl group transfer potential[15] (Figure 1.9). Enzymes can transfer these N–[PO_3^{2-}] groups to ADP to replenish ATP levels in tissues that are in metabolic states with high ATP expenditure (*e.g.* contracting skeletal muscles).

The third grouping of phosphoramidates are in the realm of phosphoproteomics. The transient formation of N–PO_3^{2-} bonds to the imidazole side chains of histidine residues in enzyme active sites is a feature of several enzyme catalytic cycles. In bacterial sugar phosphotransferase systems involved in transport of hexoses into cells with concomitant trapping by phosphorylation, up to three tandem phosphohistidinyl proteins may participate in a cascade of phosphoramidate-mediated phosphoryl transfers.

The last biological category of replacement of one of the four oxygen atoms in phosphate groups by a different atom is the replacement of oxygen by sulfur. As it happens, here are almost no examples of endogenous low molecular weight phosphorothioate compounds. P–S single bonds and P=S double bonds do show up in two kinds of xenobiotics: nerve gases and insecticides, often as phosphorothioate triesters (Figure 1.9). Phosphotriesterase detoxification catalysts have evolved to hydrolyze those P–S or P=S linkages. In bacteria some of the internucleotide phosophodiester bonds are substituted by naturally occurring phosphorothioate diester bonds as protective mechanisms to distinguish self from foreign DNA.

Synthetic phosphorothioate analogs of natural phosphate monoester and diester substrates proved particularly valuable as stereochemical probes of phosphoryl transfer enzymes. These included phosphoryl transfer enzymes from ATP (kinases) and phosphomonoester hydrolases (phosphatases). The phosphorothioates were also used to deconvolute stereochemistry and infer mechanisms of nucleotidyl transferases (phosphodiester-forming) and of phosphodieterases (phosphodiester-cleaving).

1.9 Phosphoproteomics

Proteomics is the systems-wide study of all the proteins in a given collection. It could be all the proteins in a bacterium such as *Escherichia coli* or could encompass all bacterial proteomics. It could be the proteomics of a given type of human cell, (say a neutrophil in the blood) or tissue. A proteome could encompass the global human proteome from transcription and translation of all ~26 000 human genes, including the products from all alternate mRNA splicings.

Proteomes could also encompass all the posttranslationally modified forms of proteins and that expands the output from 26 000 human genes to millions of modified proteins that change dynamically from moment to moment in different parts of each human organism.

In this volume we look somewhat superficially into phosphoproteomics, the sets of proteins that have been posttranslationally modified by enzymes that add and remove phosphoryl groups. The coverage is superficial, not for lack of interest in phosphoproteomics, but because there are dozens of volumes and tens of thousands of research papers that have provided information on this topic over the past three to four decades. There are ~520 protein kinases (ATP-dependent phosphoryl transferases) in the human kinome. While some may have a limited set of substrates, protein kinase A is reputed to have more than a hundred protein substrates.

The biology of cell signaling and metabolic regulation by protein posttranslational modifications constitutes a parallel and intersecting cellular signaling universe to the low molecular weight signaling networks (*e.g. via* cyclic nucleotides; Figure 1.7B). As noted in Figure 1.10A of the several kinds of covalent protein modifications, the reversible phosphorylation of proteins at serine, threonine, and tyrosine side chains predominates.

Our approach will stay focused on how the chemistry of specific phosphoprotein residues conditions the biological strategies. Nine of the twenty proteinogenic amino acid building blocks are observed to be phosphorylated once incorporated into proteins. These occur in vastly differing numbers and with distinct stabilities and functions.

Three of the residues are *N*-phosphorylated to yield phosphoramidates (Figure 1.10B) [histidine (His) to P-His, arginine (Arg) to P-Arg, lysine (Lys) to P-Lys]. These are acid labile, base stable, often go undetected in standard acid workups of protein modifications and may be thermodynamically activated.

Both of the amino acid residues with acidic, carboxylate side chains, glutamate and aspartate, can undergo posttranslational phosphorylation (Figure 1.10C). These are thermodynamically activated acyl phosphates, labile to hydrolysis and may have short half-lives. The short half-life of aspartylphosphate residues in a large family of bacterial DNA transcriptional regulators is tuned to the desired duration of gene expression responses.

The thiol side chain of ~250–300 000 cysteine residues of human proteomes, with average pK_as of around 8, offer the nucleophilic thiolate anions as potential attacking agents for the central phosphorus atom of phosphoryl groups. Surprisingly few classes of contemporary phosphoproteins use the cysteinyl side chains as nucleophiles. The hundred or so phosphotyrosyl protein phosphatases are the sole class where –S–PO_3^{2-} covalent adducts have been detected in any numbers.

Most of the ~250 000 phosphoprotein sites that have been tabulated and studied in detail for structure and function fall into the class of acid-stable phosphoserine (P-Ser), phosphothreonine (P-Thr), and phosphotyrosine (P-Tyr) phosphomonoester residues. These three phosphomonoester residues exemplify the broad sweep of these stable modifications that introduce

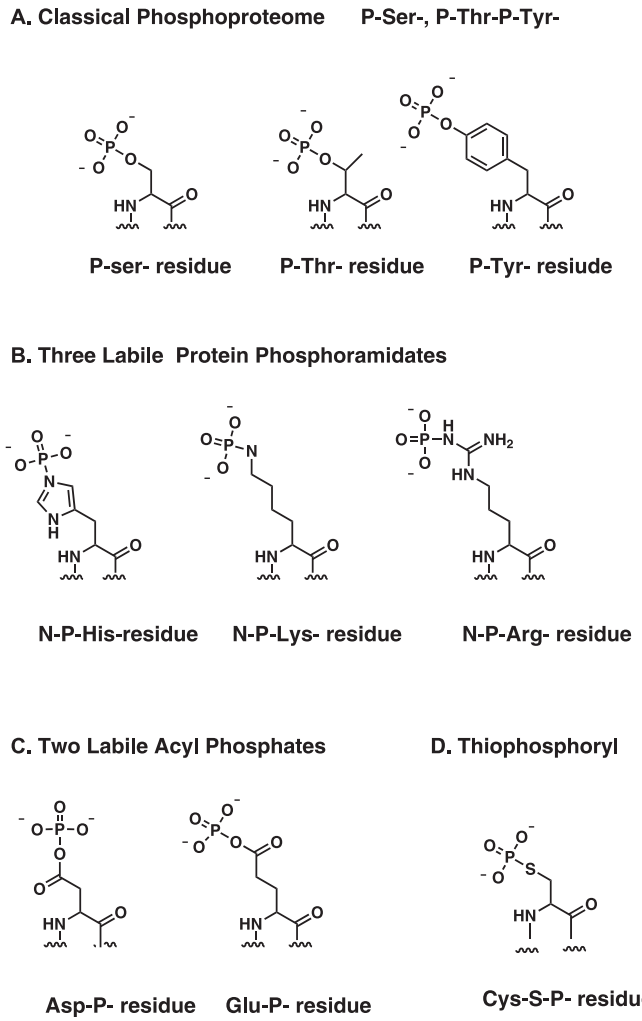


Figure 1.10 Nine protein side chains that can undergo enzymatic posttranslational phosphorylation. (A) Canonical phosphoproteomics follows the formation, removal, and biological effects of posttranslational enzymatic phosphorylation of three alcohol side chains to generate P-Ser, P-Thr, and P-Tyr residues. (B) Three *N*-phosphoramidate side chain linkages, N–P-His, N–P-Lys, and N–P-Arg residues, are labile to the usual mass spectrometry work up conditions and are often not counted. (C) Similarly, the two acyl phosphates Asp-P and Glu-P are likely to be transient intermediates in phosphoproteomic studies. (D) The *S*-phosphocysteinyl residues are intermediates in phosphotyrosine phosphatase catalytic cycles.

tetrahedral dianions into protein microenvironments. Those geometric and charge perturbations drive protein conformational changes and altered partner recognitions at the heart of phosphoproteomics-based signal transduction logic.

What goes on in *reversible* phosphoproteomics signaling must come off. The classes of $-N-PO_3^{2-}$, $S-PO_3^{2-}$, and $CO-OPO_3^{2-}$ groups may not need much enzymatic help in rapid dephosphorylation but the P-Ser, P-Thr, and P-Tyr residues do. The countervailing or balancing phosphoprotein phosphatases can be chemospecific, context selective, and subject to many kinds of spatial and temporal regulations to manage the dynamics of thousands of phosphoproteins in a eukaryotic cell.

1.10 Phosphoryl Groups: Kinetic *vs.* Thermodynamic Stability

Phosphoryl group transfers permeate every corner of cell and organismal metabolism. The default metabolic state of $-PO_3^{2-}$ groups is inorganic phosphate, perhaps the thermodynamically most stable form of phosphorus in biology. Many of the enzyme-mediated transfers of phosphoryl groups involve ultimate transfer to an attacking water molecule, activated as a nucleophile in a hydrolase or ATPase active site.

We take up metabolites with diphosphoric and triphosphoric anhydride side chains (ADP, and most notably ATP and its NTP and 2′deoxyNTP congeners), as phosphoryl group donors to cosubstrate nucleophiles. The high phosphoryl group ($-PO_3^{2-}$) transfer potential derives from the P–O–P phosphoric anhydride linkages. These anhydride linkages have the dual properties, essential for their biological roles, of kinetic stability but thermodynamic activation. The negative charges on the phosphate oxygens at physiological pH are attributes of kinetic stability.

The metabolic products of phosphoryl transfers of the γ-PO_3 of ATP and NTP congeners are phosphate monoesters, typically a mix of monoanionic and dianionic forms in cells. The complementary transfer of the α-phosphoryl group, still attached to adenosine (adenylyl or nucleotidyl) transfers instead generates phosphodiesters. They are monoanionic and this is a central property of the internucleotide bonds in all RNA and DNA molecules. The phosphomonoester and phosphodiester metabolites, with their one or two oxoester bonds to the phosphoryl group, are both kinetically stable and thermodynamically stable under biological conditions.

Thus, we divide phosphoryl group-containing metabolites into three classes that reflect the chemical features of each class as they play out in cell and organismal biology. The thermodynamically activated, kinetically stable group has ATP as its flagship molecule. The thermodynamic activation is due to the phosphoric anhydride linkages in the side chain. The kinetic stability derives from the tetra-anionic state of that triphosphate side chain. ATP and NTP congeners are the regents for $-PO_3^{2-}$ group transfers to nucleophiles, including water, to drive otherwise unfavorable equilibria in cells, from chemical coupling to mechanical movement.

The second group of phosphorus molecules in biology are the phosphomonoesters and phosphodiesters. They are chemically stable in both the

thermodynamic and kinetic senses. The monoesters arise largely by phosphoryl transfers from ATP, the phosphodiesters largely from nucleotidyl transfers from NTPs (RNA) or 2′dNTPs (DNA). The retention of one negative charge in phosphodiesters is a crucial factor in RNA and DNA internucleotide bond stability. The thermodynamic stability of these two ester classes requires enzymatic catalyses to return them to the constituent alcohols and inorganic phosphate. Alkaline phosphatase is estimated to increase hydrolysis rates of phosphomonoesters by factors up to 10^{27}–a measure both of the stupendous catalytic efficiency of the enzyme and the basal chemical stability of the phosphate esters.[16]

The third position is occupied by inorganic phosphate, the dissociated form of the parent inorganic phosphoric acid, H_3PO_4. The first two pK_a, at pH 2 and pH 6, are the parameters that allow inorganic phosphate at pH 7.2 to exist preferentially as the phosphate dianion (and phosphate monoesters as dianions and phosphodiesters as monoanions, see Chapter 2). The chemical versatility of inorganic phosphate as an electrophile at phosphorus (P^V) and a nucleophile at its oxyanions allows it to sit at multiple key metabolic intersections. It behaves as oxygen nucleophile in phosphorylase-mediated mobilization of glucose units from glycogen but behaves as an electrophile, attacked by ADP, in the ATP synthase reactions that power life. On complexation and crystallization with calcium ions, the sparingly soluble calcium phosphate is the inorganic material that forms the bony skeleton and teeth enamel so essential to the evolution of macrofauna.

1.11 The Structure of This Book

The volume begins with this introductory chapter then proceeds to the chemical properties of inorganic phosphate that enable its several biological roles. The dimeric phosphoric anhydride inorganic pyrophosphate is then addressed as the central functional group that enables the P–O–P anhydride linkage to function throughout biology *via* the dual properties of thermodynamic activation coupled to kinetic stability in aqueous medial at physiological pH values.

From inorganic pyrophosphate with one P–O–P anhydride bond it is a short chemical traverse to ATP with two P–O–P anhydride bonds in its kinetically stable tetra-anionic triphosphate side chain. Adenine (C_5N_5) may have been available prebiotically from condensation of HCN; and AMP from phosphoribosylation. AMP-binding proteins may have evolved early to give AMP and then ADP and ATP pride of place in phosphoric anhydride energy metabolism.

The two P–O–P anhydrides in ATP and other NTPs and 2′deoxyNTPs offer nucleophiles three choices of electrophilic phosphorus atoms to attack. Transfer of Pγ as the $-PO_3^{2-}$ group opens the biological world of phosphate monoesters brought about by phosphoryl transfer enzymes and includes all the kinases for low molecular weight metabolites and for proteins in their myriads.

The other major regiochemistry of nucleophile attack, at Pα, transfers the AMP (adenylyl group; NMP generically is nucleotidyl) to the incoming nucleophile. Nucleotidyl transfers to attacking alcohols generate phosphodiesters and this is the route to every phosphodiester bond in RNA, in DNA, and in membrane phospholipids.

From phosphoric anhydride linkages then we have proceeded to phosphate monoester metabolites by phosphoryl transfer and to phosphodiesters by nucleotidyl transfers. These are the small set of pervasive forms of phosphorus- containing metabolites in all organisms. Along the way we examine mechanisms of phosphoryl transfer enzymes to lower energy barriers dramatically and to analyze the interplay of thermodynamic activation and kinetic stability that is at the heart of phosphate aqueous chemistry. The phosphate monoester hydrolases and phosphodiester hydrolases that have become adapted to the substrates and distinct chemical tasks at hand are examined.

The last part on low molecular weight phosphate metabolites examines the occurrence and biological consequences of replacement of one (or more) of the four oxygens of phosphate by carbon, nitrogen, or sulfur atoms.

The final section of the book focuses on phosphoproteomics from the chemical point of view of the nine kinds of amino acid residues that can become enzymatically phosphorylated. The four different subclasses ($-N-PO_3^{2-}$; $-S-PO_3^{2-}$; $-COOPO_3^{2-}$, and $O-PO_3^{2-}$) of phosphoproteins offer different roles, different thermodynamics, and different stabilities that constrain biological function. By far the best investigated are the hundreds of thousands of stable phosphate monoester side chains of P-Ser, P-Thr, and P-Tyr residues in the broad sweep of their signaling roles in eukaryotic metabolism.

Versatile as phosphorus chemical biology is in its permeation of all aspects of metabolism, biological information transfer, and signal transduction mechanisms, all these roles are carried out in the P(v) oxidation state., largely as the fully oxidized phosphate group. [There is limited P(III) redox chemistry in some anaerobic microbes]. In this context, phosphorus differs from nitrogen, sulfur, and oxygen, which have rich redox roles in biology.

The citations to the literature are relatively light especially considering the multiply complex cell biology events, reflecting the volume's focus on phosphate chemical biology. Many of the references are to review articles, rather than original literature findings, to introduce readers to a body of literature on a given topic.

References

1. Todd A. R., *Where There's Life There's Phosphorus*, ed. K. Nakamura, M. Kageyama and T. Oshima, Japan Sci Soc Press, 1981, p. 275.
2. G. Blazeski, German chemist Hening Brand discovered phosphorus when he trieds to make gold from urine, *The Vintage News*, **2019**, 2016.
3. Chemistry of Phosphorus in *Chemistry Libre Texts*, **2019**, libretexts.org.

4. V. K. Bansai, in *Clinical Methods: The History, Physical, and Laboratory Examinations*, ed. W. W. Wllker, H. K. Hurst and J. Willis Hurst, Butterworths, Boston, 3rd edn, 1990.
5. Q. Liu, S. Huang, J. P. Matinlinna, Z. Chen and H. Pan, Insight into biological apatite: physicochemical properties and preparation approaches, *BioMed. Res. Int.*, 2013, 929748.
6. M. W. Bowler, M. Cliff, J. P. Waltho and G. M. Blackburn, Why did nature select phosphate for its dominant roles in biology ?, *New J. Chem.*, 2010, **34**, 784–794.
7. F. H. Westheimer, Why nature chose phosphate, *Science*, 1987, **235**, 1173–1178.
8. V. S. Kattimani, S. Kondaka and K. P. Lingamaneni, Hydroxyapatite-past, present and future in bone regeneration, *Bone Tissue Regener. Insights*, 2016, **7**, 9–19.
9. R. Milo and R. Phillips, *Cell Biology by the Numbers*, Taylor & Francis Books: Garland Science Press, 2015.
10. C. T. Walsh, B. Tu and Y. Tang, Eight thermodynamically activated but kinetically stable molecules power cell metabolismm, *Chem. Rev.*, 2018, **118**, 1460–1494.
11. S. Shuman, DNA ligases progress and prospects, *J. Biol. Chem.*, 2009, **284**, 17365–17369.
12. J. J. Petkowski, W. Bains and S. Seager, Natural products containing rare organophosphorus groups, *Molecules*, 2019, **24**, 866.
13. X. Yu, J. Doroghazi, S. C. Janga, J. K. Zhang, B. Circello, B. M. Griffin, D. P. Labeda and W. W. Metcalfe, Diversity and abundance of phosphonate biosynthetic genes in nature, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 20759–20764.
14. Z. W. Taylor, H. A. Brown, H. M. Holden and F. M. Raushel, Biosynthesis of nucleoisde diphosphoramidates in Camphylobacter jejuni, *Biochemistry*, 2017, **56**, 6079–6082.
15. M. J. McLeish and G. Kenyon, Relating structure to mechanism in creatine kinase, *Crit. Rev. Biochem. Mol. Biol.*, 2005, **40**, 1–20.
16. R. Wolfenden, Benchmark reaction rates, the stability of biological molecules in water, *Annu. Rev. Biochem.*, 2011, **80**, 645–667.

CHAPTER 2

Inorganic Phosphate, Pyrophosphate, and Polyphosphate

2.1 Inorganic Phosphate Bodily Levels

The typical content of inorganic phosphate in a 70 kg adult human is around 800 g. At a molecular weight of 80 g per mole, this corresponds to 10 moles of body phosphate. Some 85% of that inorganic phosphate is present as insoluble hydroxyapatite [$Ca_5(PO_4)_3(OH)$], typically written as the dimer [$Ca_{10}(PO_4)_6(OH)_2$] to reflect the two molecules in the crystal unit cell in bone, and in the enamel of teeth (Figure 2.1). The typical content of Pi (a mix of $H_2PO_4^-$ and HPO_4^{2-} at physiological pH) in serum is 1.1 to 1.4 mM[1] and maintained closely within those limits by kidney excretion.

One can compare the amount of phosphorus in human bone to the phosphorus in human DNA, where a covalent phosphodiester bond links every base in the 3×10^9 bases of DNA in every cell. Given that all but germ cells are diploid, then each of the human body's somatic cells will contain 6 billion base pairs, and thereby 6 billion atoms of phosphorus in the phosphodiester links. The weight of human genomic DNA in each cell corresponds to ~1 picogram. The number of cells in each human adult is estimated to be between $3–4\times10^{13}$ cells, so the total DNA content would be 30–40 g, with the phosphorus accounting for 6–8 g.

If 50–100 billion cells turn over every day, that would correspond to lots of DNA synthesis, but only require about 30–60 mg of newly available phosphate each day (see below). We will note subsequently in Chapter 3 that humans make and consume up to 75–80 kg of ATP per day to run the many

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

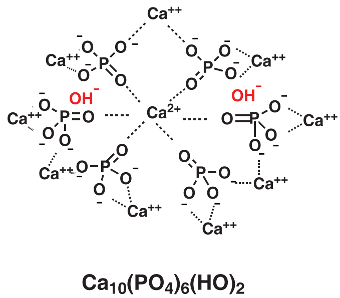


Figure 2.1 Schematic of a hydroxyapatite dimer [$Ca_{10}(PO_4)_6(OH)_2$].

metabolic pathways of cells and tissues, so that involves perhaps 1000-fold recycling of the soluble fraction of body inorganic phosphate.

Given the calcium content of the hydroxyapatite in bone, the corresponding amount of calcium in such an adult human would be ~1200 g, 99% in bone and only 1% circulating (Serum levels of [Ca^{++}] are kept at ~10^{-3} M while intracellular [Ca^{++}] in cells is ~10^{-6} M) (Americanbonehealth.org).

The recommended daily allowance for phosphate (and also calcium) in the diet is about 1 g per day (5- to 30-fold the amount required for new DNA synthesis, see above). Phosphate is present in high levels in vegetables, meat, fish, and milk such that dietary phosphate deficiencies are rare. In cases of diminished kidney function and lessened excretion of inorganic phosphate, serum phosphate levels increase and hyperphosphatemia can result. The chronic sequelae can lead to calcium phosphate deposits in organs and intracellular spaces, leading to pathogenic calcifications.

The epithelial cells of the intestine have inorganic phosphate transporter proteins on both apical and serosal surfaces so absorption efficiency can be high.[2] The 1–1.4 mM [Pi] serum level is maintained by action of the kidneys and the renal Pi transporters mostly in the proximal tubule, to excrete excess phosphate (600–1500 mg per day), not required for bones, teeth, and other metabolic functions.[3]

2.1.1 Phosphoric Acid pK_as: The Biological Importance of Being Ionized

The parent compound of inorganic phosphate salts in all of biology is phosphoric acid, H_3PO_4. It is an inorganic acid and stronger than common organic acid metabolites such as acetic acid, citric acid, succinic acid and all the proteinogenic amino acids (stronger acid = lower pK_a value for dissociation into a proton and acid anion). The triprotic H_3PO_4 can, in principle, sequentially give up each of its three protons, to give the monoanionic, dianionic, and trianionic forms of inorganic phosphate respectively.

Given that the three formal –OH substituents around the central phosphorus atom in phosphoric acid are chemically equivalent, one might have anticipated that the propensity of dissociation of an H^+ from each of those

three –OH groups would be equally probable. That is true among the three –OH groups to give up the first proton and form the phosphate monoanion, with 50% dissociation at pH2.2, for $pK_{a1} = 2.2$.

However, dissociation of the second H^+, from the phosphate monoanion to the phosphate dianion places two negative charges intramolecularly, within close proximity of the central phosphorus atom, an unfavorable electrostatic situation (by Coulomb's law that similar charges repel). The observed pK_{a2} of inorganic phosphate turns out to be 7.2; that is 50% dissociation at pH 7.2. This is in the physiological pH range. Thus, at pH 7.2 there would be 50% inorganic phosphate dianion, about 50% monoanion and a very small fraction (one part in 10^5) of the undissociated H_3PO_4. We write much of inorganic phosphate metabolism in this volume as involving the phosphate dianion. The perturbation of pK_{a2} from pK_{a1} of 10^5 ($pK_{a1} = 2.2$, $pK_{a2} = 7.2$) is an indication of the energy barrier of the unfavorable electrostatic repulsion of two negative charges around the phosphorus. One must go five pH units higher before the second proton dissociates.

The trend continues on examination of pK_{a3} for phosphoric acid: pK_{a3}=12.2. This pH value of 12.2 for 50% dianion and 50% trianion is five pH units away from physiological pH. The ΔpK_a value for pK_{a2} to pK_{a3} of 5 pK_a units (7.2 to 12.2) is the same as that from pK_{a1} to pK_{a2}, going from two negative charges to three negative charges around the central phosphorus. It is highly likely that the inorganic phosphate trianion plays no significant role in phosphate metabolism, given the presence of only one part in 10^5 of trianionic phosphate at pH 7.2.

Every phosphate-containing metabolite, from inorganic phosphate to phosphoric anhydrides pyrophosphate and ATP, to phosphate monoesters and diesters, carries at least one negative charge at physiological pH, with PPi and ATP carrying four negative charges. This inventory reflects the biological importance of being ionized. We will see this property of negatively charged phosphate-containing metabolites come up in many contexts, from keeping metabolites trapped inside cells, to stability of the anionic phosphodiester bonds linking RNA and DNA together, to the consequences of introduction of tetrahedral anionic phosphate ester groups in protein phosphorylations, to the reactivity of in organic phosphate as a nucleophile at oxygen in catalytic action of phosphorylase enzymes.

2.2 Inorganic Phosphate Transport

Figure 2.2 schematizes the movement of inorganic phosphate from the intestinal lumen into blood by way of two routes across the gastrointestinal (GI) epithelial barrier.[4] A minor, passive route is paracellular transport of inorganic phosphate through the tight junctions between cells. The quantitatively more consequential route is transcellular transport. At the luminal surface of polarized epithelial cells phosphate anions are cotransported with sodium cations by transmembrane phosphate transporter proteins.[5,6]

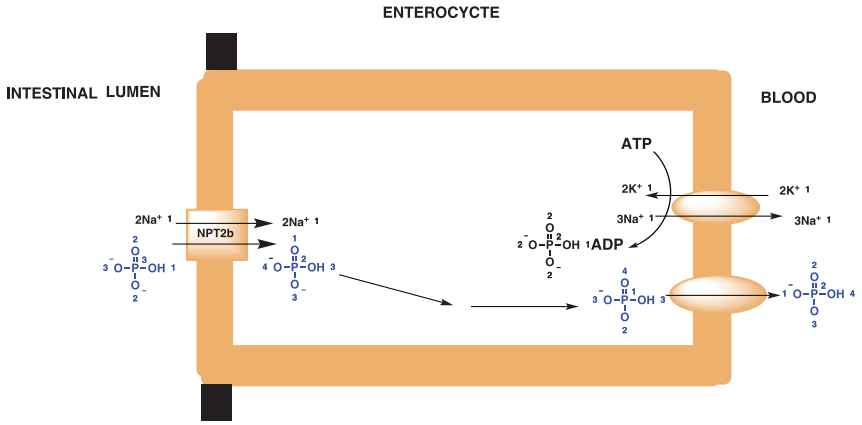


Figure 2.2 Schematic of inorganic phosphate transport in and out of cells. Transepithelial transport leading to ~1mM concentration of inorganic phosphate in blood.

Cytoplasmic concentration in these epithelial cells is around millimolar. This may represent active transport of Pi out of the GI lumen, often co-transported with sodium ions running down their concentration gradient. At the serosal side of the cell membrane there are also Pi transport proteins and they may function by counter transport of anions such as chloride ions as Pi is pumped out.

The steady state concentration of inorganic phosphate dianions in human serum is thus controlled by a balance of uptake from the GI tract (Figure 2.2), transport into other tissues, and excretion of excess phosphate by the kidneys.[2,4] This balance can also be mediated by two growth factors, fibroblast growth factor 23 (FGF23) and parathyroid hormone, that mainly affect phosphate transporter levels (gene expression and receptor transporter internalization) and thereby uptake capacity from the GI tract. Vitamin D3 as its biologically active trihydroxy metabolite also is a significant player in the regulation of calcium and phosphate metabolism. The other major reservoir for adjusting serum, and thereby cellular, levels of inorganic phosphate is bone mass. The balance between osteoblasts laying down new layers of crystalline calcium phosphate in bone (see below), and osteoclasts catalyzing the solubilization and resorption of bone calcium phosphate is a dynamic process[7] for control of soluble inorganic phosphate ions.

Bacteria also have an absolute requirement for scavenging phosphate from the external media and have evolved multigene systems to generate and import inorganic phosphate, *e.g.* from the periplasmic space of Gram-negatives such as *Escherichia coli* or *Salmonella typhimurium.* The *pho* genes encode similar transmembrane transporters as well as secreted alkaline phosphatase enzyme molecules to convert extracellular phosphomonoesters into the product alcohols and inorganic phosphate for dedicated uptake.[8]

The bacterial pho systems also include a sensor kinase PhoR and a response regulator PhoB that is a DNA binding protein that turns on a variety of genes that are expressed under limiting inorganic phosphate conditions.[9] We will note in section IV on phosphoproteomics, that this type of sensor kinase–response regulator pair is the major strategy by which bacteria sense external signals and nutrients. The sensor kinases are autophosphorylating histidine kinases that function as transmembrane sensors. This brings phosphoryl-histidine enzyme residues into play as will also be discussed in Chapter 14. The response regulators in turn receive the itinerant PO_3^{2-} group from the phosphohistidine residue in the cytoplasmic domain of the histidine kinases. The response regulator receiver is the side chain of an aspartate residue in a recognition domain. Phosphorylation of that aspartyl residue induces conformational changes in the neighboring DNA binding domain, thereby activating or repressing sets of responsive genes. We will examine both phosphohistidine and aspartyl phosphate chemical properties that control their biology in section IV.

2.3 Hydroxyapatite Formation and Deposition by Osteoblasts During Bone Formation

The intimate details of formation, maturation, remodeling, and growth of the 208 human bones and the enamel and dentine of teeth formation are far beyond the scope of phosphate chemical biology and are the subject of mineralization homeostasis physiology.[9a–c] We focus here only on the facts that the mineralization of those tissues is highly regulated both by the extracellular collagen (and other noncollagenous protein partners) laid down by the mineralizing cells and by secretion of phosphate and calcium ions into the local microenvironment.

At one level, this is inorganic phosphate physical chemistry where calcium phosphate crystallizes out of locally supersaturated solutions, oriented by the extracellular matrix of type I collagen fibrils (Figure 2.3). No special metabolic chemistry is occurring, except to control the timing and concentration of the calcium cations and phosphate anions in a given local extracellular matrix environs.[10]

The calcium phosphate crystals grow along the long axis of the collagen fibrils, at first in holes between the aligned fibrils and then across the mineralizing matrix. The intimate interaction of the collagen protein triple helices and minor partner proteins are crucial in axial growth of the nascent hydroxyapatite crystalline phase.[12,13] Sulfated regions of the matrix glycosaminoglycan chondroitin sulfate family[14] have also been reported to provide adsorption layer nucleation sites that mediate the axes of hydroxyapatite crystallization and growth axes (Figure 2.4).

The bone-forming cells are osteoblasts. The inorganic phosphate is secreted by those cells from intracellular vesicles that have accumulated high concentrations of inorganic phosphate. The counterions in the matrix are

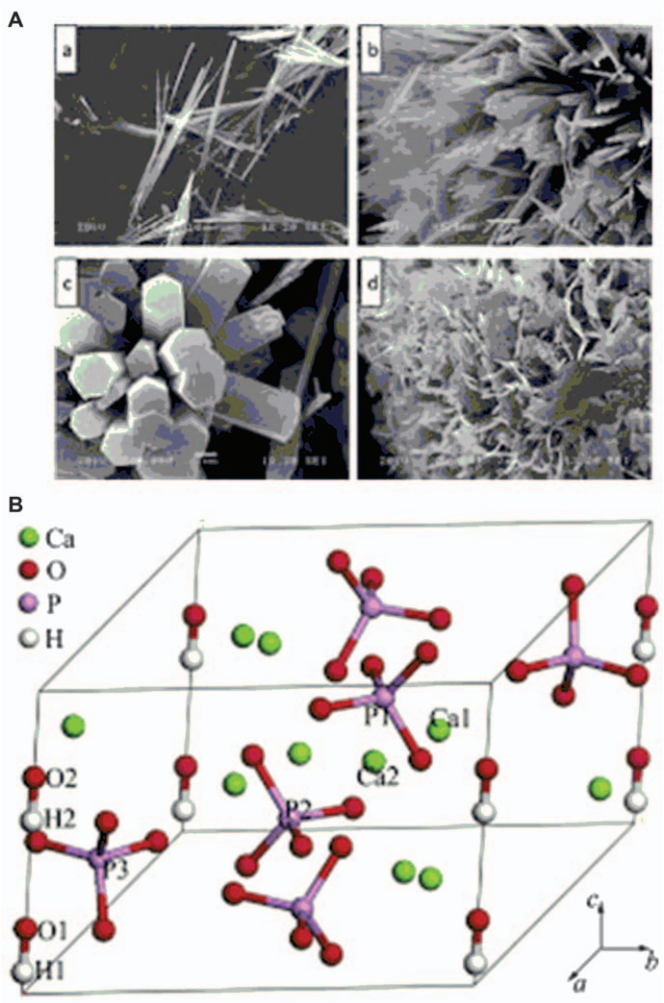


Figure 2.3 Hydroxyapatite crystals: (A) growth of crystals, (B) schematic of unit cell of carbonated hydroxyapatite.
(A) Reproduced from ref. 39 with permission from the Royal Society of Chemistry, (B) Reproduced from ref. 11 with permission from Elsevier, Copyright 2014.

mostly Ca^{2+}, generating the nascently insoluble hydroxapatite ion pairs. In different forms of bone and teeth enamel there are crystal defects, arising from varying amount of carbonate anions and deficits of calcium ions that make *in vivo* forms of hydroxyapatite distinct from abiotic, geologically formed crystalline hydroxyapatite.

As the osteoblasts become surrounded and encased by the mineralizing extracellular nascent hydroxyapatite, they become functional osteocytes, generating a cellular connectivity for bone tissue function.[16] Another key cell

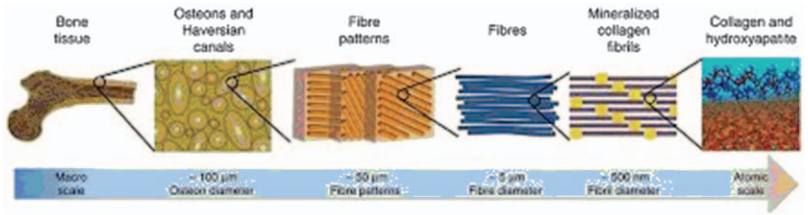


Figure 2.4 Schematic of hydroxyapatite crystals adsorbed to and mineralizing collagen fibrils in bone from nanoscale to macroscale. Reproduced from ref. 15 with permission from Springer Nature, Copyright 2013.

type in bone dynamics, remodeling, and pathological bone loss and resorption are osteoclasts that resolubilize the hydroxyapatite crystals by local pH changes. The interplay between osteoblasts and osteoclasts mediates bone dynamics and growth. During transient resolubilization of enamel layers in the mouth by lowered pH, the water molecule in the hydroxyapatite matrices can be replaced by fluoride. The fluorohydroxyapatite is a stronger material, leading to less decay. This replacement forms the basis for water and toothpaste fluoridation (https:www.nanowerk.com).

2.4 Inorganic Pyrophosphate: The Simplest Inorganic Phosphoric Anhydride

When inorganic phosphoric acid is heated, water is driven out as monomers condense, first to form pyrophosphoric acid, and then longer chains of polyphosphoric anhydrides (Figure 2.5). The prefix *pyro* refers to heat or fire, consistent with the observation that pyrophosphoric acid is derived from high temperature dehydrative condensation to create the P–O–P anhydride link in pyrophosphoric acid and congeners.

Pyrophosphoric acid is then the simplest of the phosphoric anhydrides structurally that participate throughout biological energy metabolism. At physiological pH values (~pH 7), pyrophosphoric acid is predominantly the tetra-anionic pyrophosphate (Figures 2.6–2.8). Analogously, the longer polyphosphoric anhydrides, typically referred to as *linear polyphosphates*,[17] are also anionic in physiological media.

The phalanx of negative charges may contribute to their persistence against nonenzymatic hydrolysis in physiological aqueous media inside and outside cells. Thus, the pyrophosphate tetra-anion is indefinitely stable in cells. [An estimate of the nonenzymatic rate constant for hydrolysis at 25 °C is 1×10^{-7} s^{-1} [18] (approximating to a half-life of months).] By comparison, the neutral acetic anhydride, shorn of any protective negative charges, is

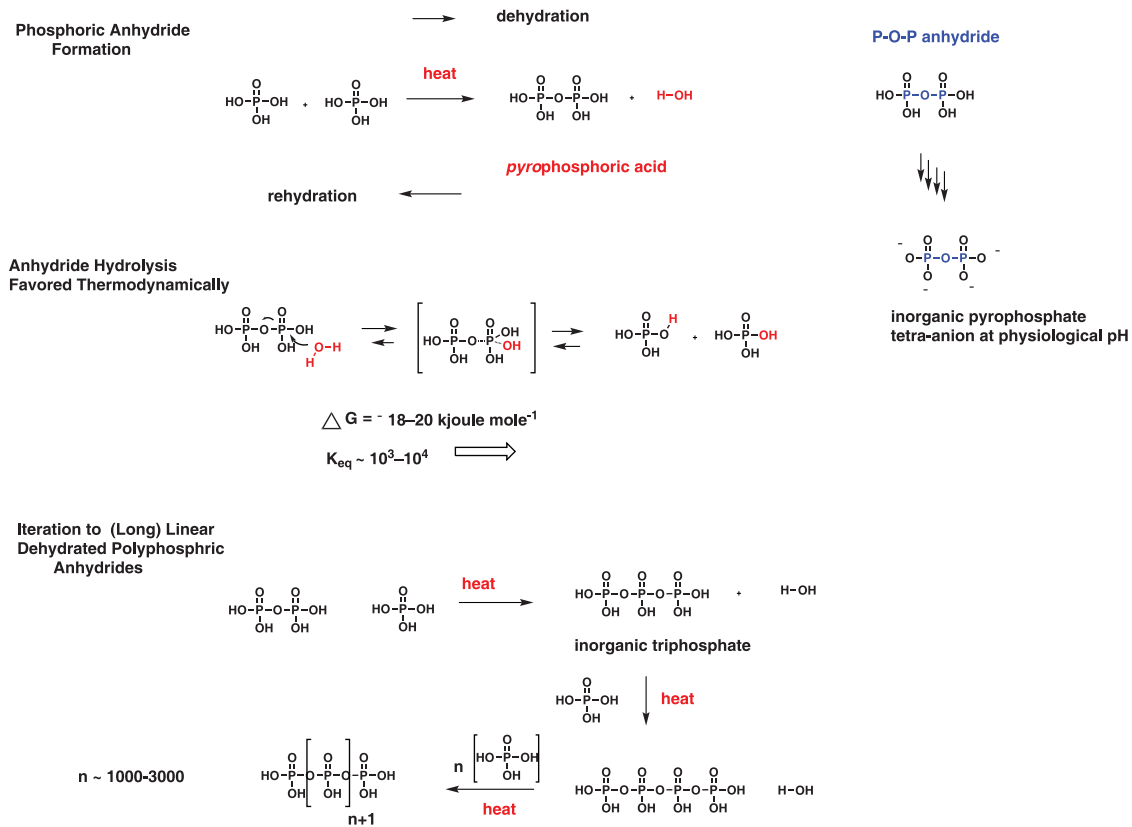


Figure 2.5 Inorganic pyrophosphate at neutral pH is the tetra-anion with one phosphoric anhydride (P–O–P) bond. Hydrolysis is thermodynamically favored: when coupled to other reactions, its hydrolysis can drive otherwise unfavorable equilibria.

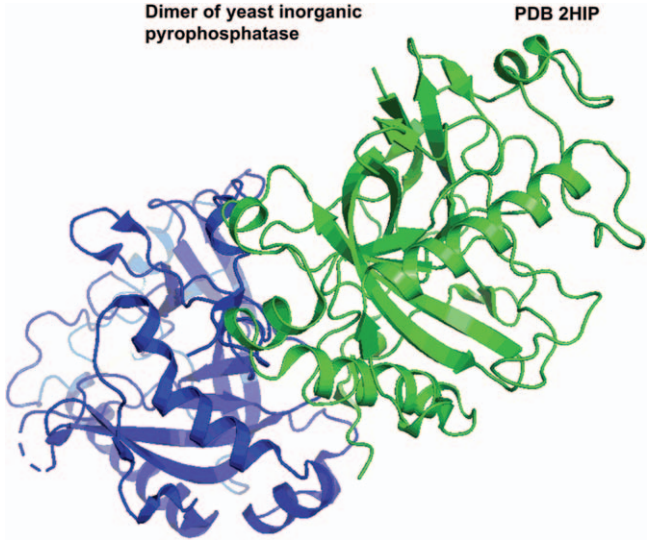
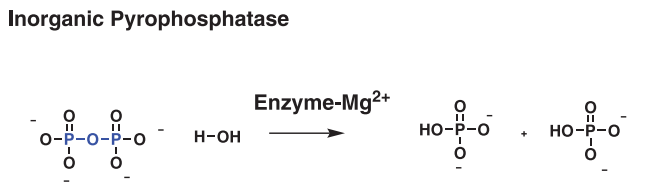


Figure 2.6 Soluble inorganic pyrophosphatase.



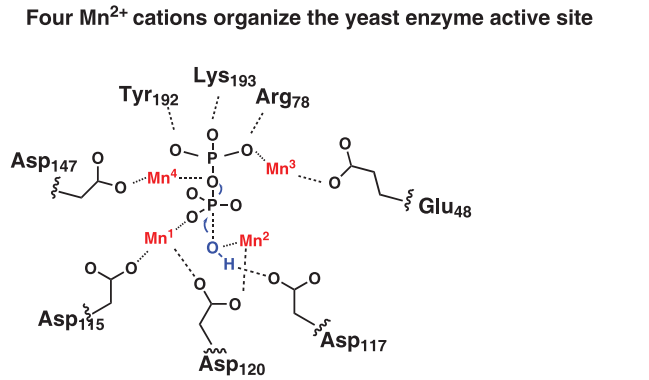


Figure 2.7 Orientation of pyrophosphate substrate in the active sites of inorganic pyrophosphatases.

hydrolyzed nonenzymatically to two molecules of acetic acid or acetate within minutes. The carboxylic acid anhydrides are not used as thermodynamically activated metabolites in cell metabolism because of this kinetic lability.

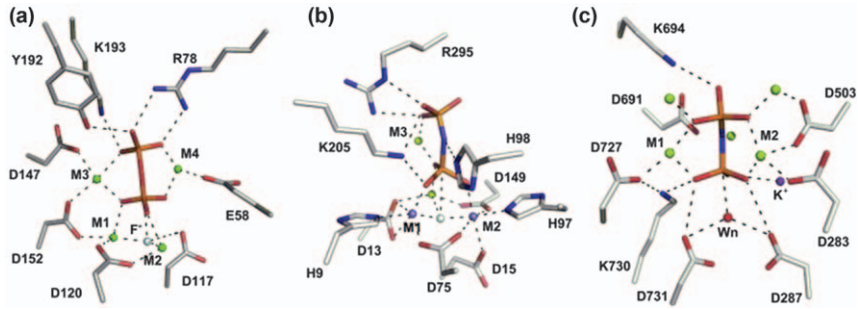


Figure 2.8 Substrate PPi orientation in the active site of two soluble PPiases (yeast and *B. subtilis*) and a transmembrane proton-pumping PPiase from mung bean. All subtypes use metal cations to orient the PPi and water substrates to lower barrier for water attack on an electrophilic phosphorus atom. (a) soluble yeast PPiase; (b) soluble *Bacillus subtilis* PPiase; (c) H proton pumping transmembrane PPiase from mung bean.
Reproduced from ref. 19 with permission from John Wiley and Sons, © 2015 Federation of European Biochemical Societies.

One practical measure of the stability of inorganic pyrophosphate and its next longer homolog, inorganic tripolyphosphate, is their historical and contemporary widespread use in detergents. These include detergents for automatic dishwashers (sodium tripolyphosphate), silver polish (tetrapotassium pyrophosphate) and for heavy duty cleaning (mixed sodium and potassium salts of pyrophosphate).

On the other hand, typical of other anhydrides, the P–O–P anhydride linkage is uphill energetically from the parent pair of acids (here two H_3PO_4). For pyrophosphoric acid, the free energy of hydrolysis is favorable, at the level of 18–20 kilojoules, corresponding to an equilibrium constant of ~5000 in favor of hydrolysis back to two molecules of inorganic phosphoric acid.

This $K_{eq.}$ value means that if PPi can be hydrolyzed by an enzyme at a specific time and place in cells, it could be used to couple otherwise unfavorable, or partially favorable, equilibria. Inorganic pyrophosphatases are such catalysts.[18,20]

Ubiquitously expressed and highly active (from 200 to 2000 hydrolytic events per second),[21] they serve to couple to any reaction producing nascent pyrophosphate as product. The coupled hydrolysis of PPi on to two molecules of Pi adds another factor of 5000 in the net $K_{eq.}$ for the tandem two stage reactions (Figure 2.5).

The energetically favored, coupled hydrolysis of PPi coproducts from upstream ATP-cleaving enzymes to two molecules of inorganic phosphate by pyrophosphatases turns out to be key sequelae of biosynthetic strategies for most classes of macromolecules. This is the strategy for: (1) amino acid activation for protein biosynthesis, (2) for the stepwise buildup of polysaccharides such as glycogen and starch, and (3) for the biogenesis of tens of millions of membrane phospholipids in every cell cycle. Perhaps most

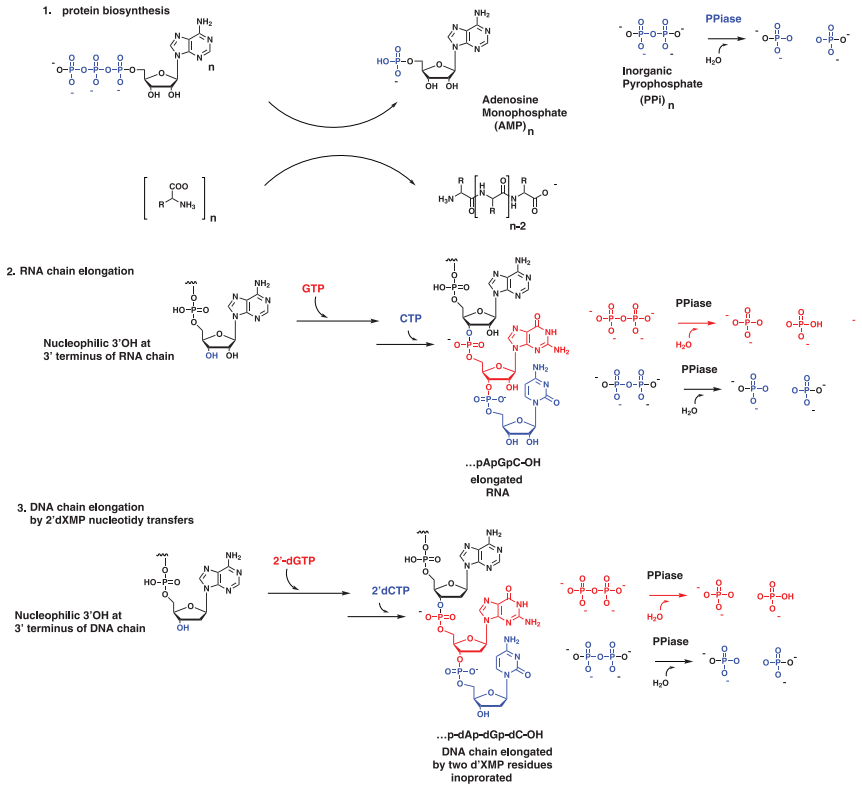


Figure 2.9 Schematic illustration of release of a molecule of inorganic pyrophosphate in each chain elongation step in protein biosynthesis, RNA chain elongation and DNA chain elongation. Coupled action of inorganic pyrophosphatase to hydrolyze coproduct PPi molecules draws the equilibria (further) in favor of polymer biosyntheses.

compellingly, this is also the strategy for: (4) every chain elongation step in both RNA and DNA biogenesis, by RNA polymerases and DNA polymerases, respectively (Figure 2.9).

2.5 Thermodynamic Activation Balanced by Useful Kinetic Stability

It is the double combination of properties of *kinetic stability and thermodynamic activation* that characterize phosphoric anhydride linkages under physiological conditions that have enabled the ubiquitous distribution of P–O–P-containing molecules to be the energy currencies of all cells.[22] This holds true not only for the inorganic pyrophosphate and longer polyphosphates, but also for such organic pyrophosphates as nucleoside triphosphates (*e.g.* ATP, GTP, 2′-deoxyATP), nucleoside diphospho sugars

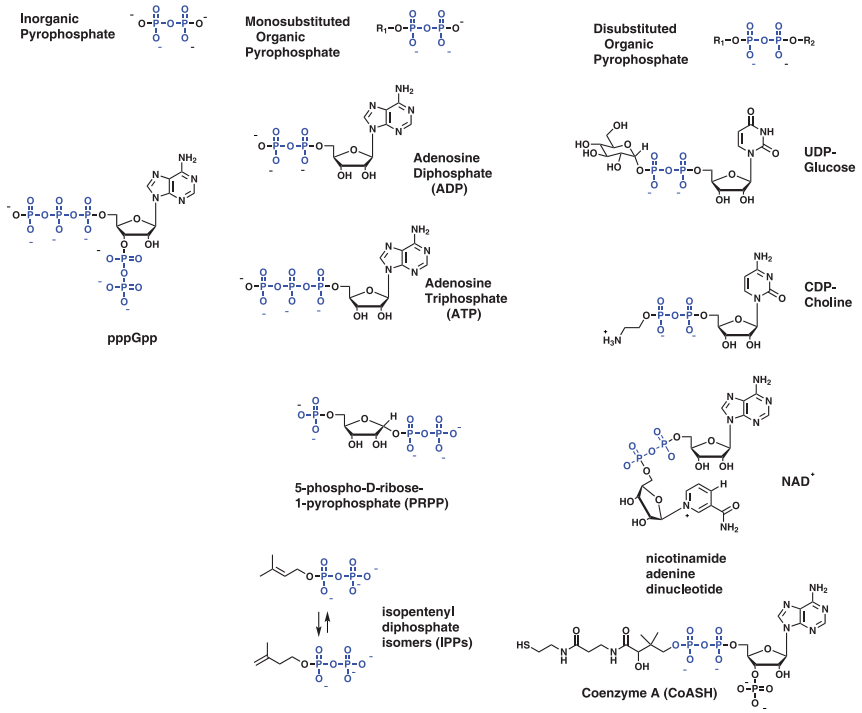


Figure 2.10 Common primary metabolites, including central coenzymes, containing kinetically stable phosphoric anhydride linkages in the form of substituted pyrophosphates.

[*e.g.* uridine diphosphate (UDP)-glucose, ADP-glucose, GDP-mannose], the signaling tetranucleotides and pentanucleotides ppGpp, pppGpp, and coenzymes including nicotinamide adenine dinucleotide (NAD^+) (Figure 2.10). The chemical biology of those pyrophosphate (phosphoric anhydride) linkages is explored in Chapter 3 and subsequent chapters.

2.6 Thermodynamically Favored Enzymatic Hydrolysis Drives Otherwise Unfavorable Cellular Equilibria

2.6.1 Two Types of Inorganic Pyrophosphatases

The utility of the kinetically stable, thermodynamically activated P–O–P anhydride bond in pyrophosphate is put to use for specific cellular work by two types of inorganic pyrophosphatases.

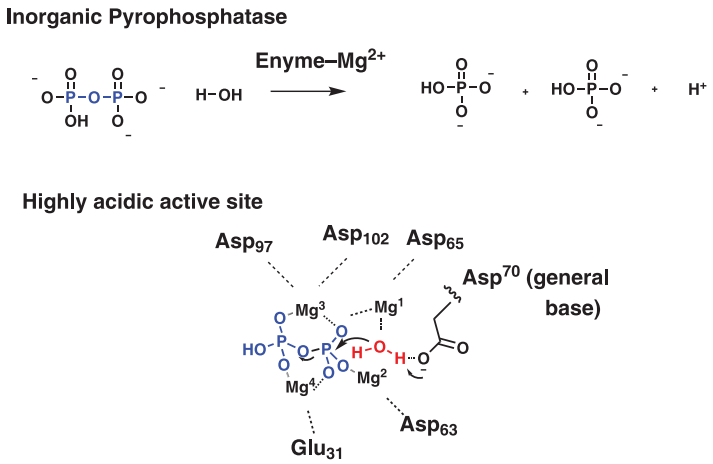


Figure 2.11 Schematic of the active site of *E. coli* inorganic pyrophosphatase with four magnesium ions coordinating bound inorganic pyrophosphate. The ionized side chain of an active site aspartate residue acts as a base to initiate a catalytic cycle of hydrolysis. The magnesium ions are shown as hexacoordinate ions with bound waters.[23]

2.6.1.1 Soluble Inorganic Pyrophosphatase

The most common form is represented by the soluble inorganic pyrophosphatase from yeast, a dimer of 32 kDa (285 amino acid residues) subunits.[18] Divalent cations, preferably Mg^{2+} (but also Mn^{2+} or Zn^{2+} will substitute), are required for activity. The cations act as chelators, and negative charge neutralizers, to the PPi oxyanions. X-ray analysis of the corresponding inorganic pyrophosphatase from the bacterium *E. coli* shown in Figure 2.11 depicts four magnesium cations and a molecule of PPi in the active site. The four magnesium cations, labeled Mg^1, Mg^2, Mg^3, and Mg^4, are coordinating the pyrophosphate polyanion through all three of the nonbridging oxygens at each phosphorus center. An active site aspartate residue carboxylate side chain acts as the base to activate a water molecule to attack one of the two P atoms in bound pyrophosphate. The catalytic rate constant (k_{cat}) for the pyrophosphatase is 10^3 s^{-1}; each catalytic cycle complete in a millisecond. This rapid rate allows pyrophosphatases to function effectively in the coupling role for tandem equilibria displacements in the forward direction. The catalytic advantage of the enzymatic–nonenzymatic hydrolysis rate for this enzyme is $\sim 10^{10}$, a powerful ten billion-fold acceleration.

2.6.1.2 Membrane-bound Proton-pumping Inorganic Pyrophosphatases

A second class of inorganic pyrophosphatases has evolved separately to couple the thermodynamically favored hydrolysis of PPi[21] to acidify vacuolar

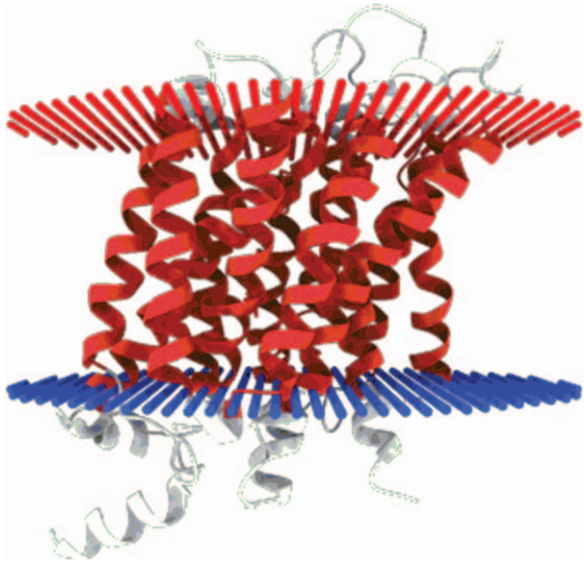


Figure 2.12 Membrane-bound proton-pumping pyrophosphatases.[25] The figure schematizes the membrane helical nature of a fungal multi-pass transmembrane PPiase.[26] (PDB structure 4A01).

compartments in plant cells. These are intrinsic membrane proteins that drive anisotropic pumping of protons into membrane vacuoles and isolated membrane vesicles, acidifying the internal compartment's pH, as PPi as hydrolyzed.[24] These transmembrane PPiases are used as marker enzymes for plant vacuolar membranes. In this case, the energy released by PPi hydrolysis is not used to couple some prior unfavorable biosynthetic step (*e.g.* RNA- or DNA-chain extension). Instead, the vectorial pumping of protons from cytoplasm to vesicle compartment sets up or further reinforces a transmembrane electrochemical potential that can perform cellular work.

The membrane PPiases from yeast and plants have been studied mechanistically and architecturally to examine how PPi hydrolysis in the active site drives conformational changes that open and close protein gating domains on the cytoplasmic and then vesicular side of the membrane (Figure 2.12). There are intriguing structural and logical similarities between these membrane-embedded proton pumping PPiases and the vacuolar ATPases of yeast and plants, discussed below and in Chapter 3. Both types of catalysts pump protons into vacuoles or endosomal compartments to acidify them (Figure 2.13).

Both enzyme types hydrolyze a phosphoric anhydride bond to drive those processes. The PPiases hydrolyze the single P–O–P anhydride bond in PPi while the vacuolar ATPases hydrolyze the P_{β}–O–P_{γ} anhydride bond in ATP. Figure 2.12 schematizes the extensive helical array of a transmembrane PPiase consistent with its role as a transmembrane proton pump. These

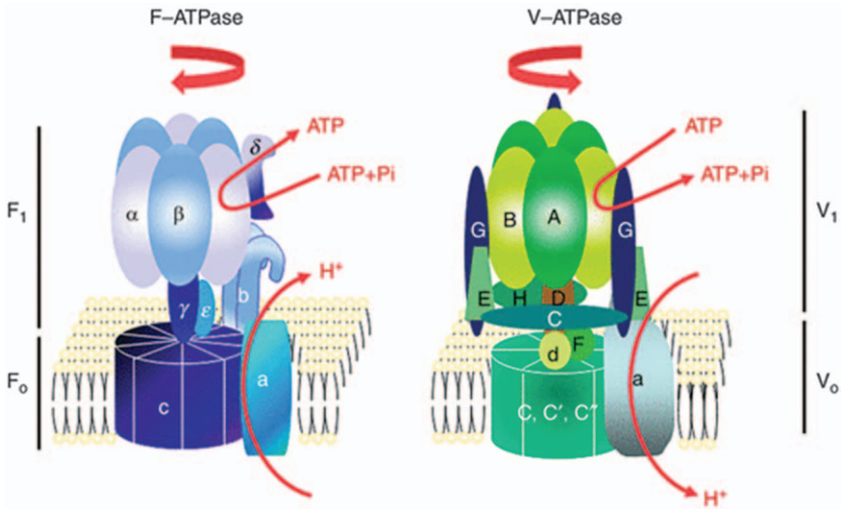


Figure 2.13 Two transmembrane energy machines with similar logic and architectures that run in opposite directions. Vacuolar-ATPases (V-ATPases) hydrolyze ATP and use the energy to drive proton uptake into vacuoles and endoplasmic reticulum, thereby acidifying those compartments.[29] F-Type ATP synthases sit in mitochondrial membranes and use the thermodynamically favorable flow of protons from the acidified intermembrane space of mitochondria back to the matrix to drive ATP synthesis.
Reproduced from ref. 29 with permission from Springer Nature, Copyright 2013.

are typical transmembrane proteins with 10–12 transmembrane helices. The X-ray structure of a fungal transport protein for inorganic phosphate has been solved and is schematized in Figure 2.12, with the outer membrane surface at the top and the inner membrane surface at the bottom. The multiple transmembrane helices dominate the structure.

2.6.2 Vacuolar ATPases and F-ATP Synthases

Figure 2.13 indicates that the hydrolysis reactions could be run backwards if the transmembrane proton potential would become high enough to drive the synthesis of PPi from two molecules of Pi or analogously of ATP from ADP and Pi. Regmi *et al.*[27] have indeed observed that plant vacuoles can use proton gradients to synthesize pyrophosphate under some conditions. Furthermore, Figure 2.13 depicts schematic similarities between the subunit composition and integrated architecture of the vacuolar ATPases that typically run in the ATP cleavage direction and the F-ATPase complexes. The F-type ATPases typically run in the opposite, ATP synthase direction, generating ATP from ADP and Pi as protons run back across the membrane (bacterial, chloroplast, mitochondrial) down their electrochemical gradient: a thermodynamically

favored flow strong enough to make essentially all the ATP generated during oxidative metabolism of prokaryotic and eukaryotic cells.[28]

These two examples of inorganic pyrophosphatases, in two distinct cellular compartments, coupling thermodynamically favored hydrolysis of the phosphoric anhydride linkage in tetra-anionic pyrophosphate, show how organisms have adapted the central utility of such anhydride linkages for distinct kinds of cellular work. This simple phosphoric anhydride linkage, likely to have been available abiotically from phosphorus-containing volcanic emissions, became the backbone strategy for essentially every macromolecular biopolymer biosynthetic strategy as will be elaborated in Chapter 3.

2.7 From Triphosphates to Long Inorganic Polyphosphates

Longer chains of inorganic phosphoric anhydrides, typically termed linear polyphosphates, are also constituents of microbial and eukaryotic cells. However, they have largely been ignored, even though they were first detected over a century ago, concentrated in certain bacterial granules (summarized in[30]). The Nobel laureate in physiology or medicine for 1959 for his characterization of DNA polymerase I, Arthur Kornberg, spent much of the last 15 years of his research career,[17] resurrecting the forgotten features of polyphosphates that had been relegating to the category of ''molecular fossils''. Kornberg[30] argued that polyphosphoric anhydrides, elongated to tens to hundreds of repeat units (Figure 2.14), were found in most microbial and animal cells. They serve as flexible anionic polymers, conformationally mobile, mediated by divalent cation chelations, and most significantly as phosphoric anhydride energy reserves. They may be notionally akin to glycogen and starch, polymers of glucose that are mobilizable to glucose monomers, to be metabolically hydrolyzed to power energy metabolism in cells.

Kornberg argued that such linear polyphosphates were products of ''volcanic condensates and deep ocean vents...likely a key agent in evolution from prebiotic time'' (Figure 2.14A). Brown and Kornberg[17] further argued that these polyphosphates were ready sources of monomeric inorganic phosphate whose intracellular concentration is narrowly maintained in the 5 mM range, despite large swings in environmentally available [Pi]. The premise was they had a role in inorganic phosphate homeostasis.

In keeping with his deep expertise in phosphoryl transfer enzymology, Kornberg and colleagues characterized suites of enzymes that transferred phosphoryl groups to and from polyphosphate substrates.[31] One was an ATP-dependent polyphosphate kinase,[32] transferring the terminal $-PO_3^{2-}$ from ATP as an electrophilic fragment, to one of the oxyanions of a terminal phosphate group of a polyphosphate chain in an elongation reaction that builds up P–O–P anhydride links (Figure 2.14B). This is presumed to be the main biological route to linear polyphosphates.

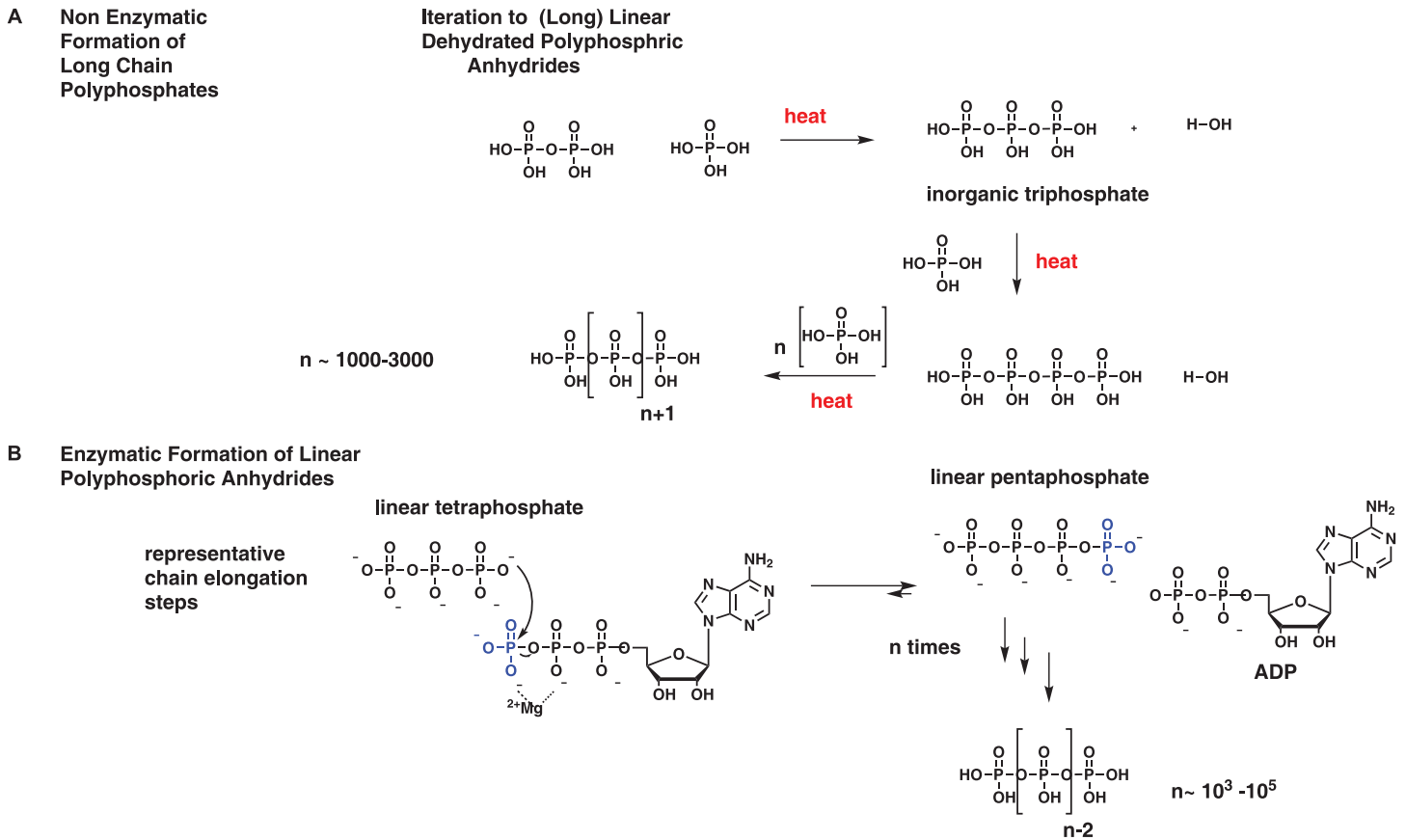


Figure 2.14 (Continued).

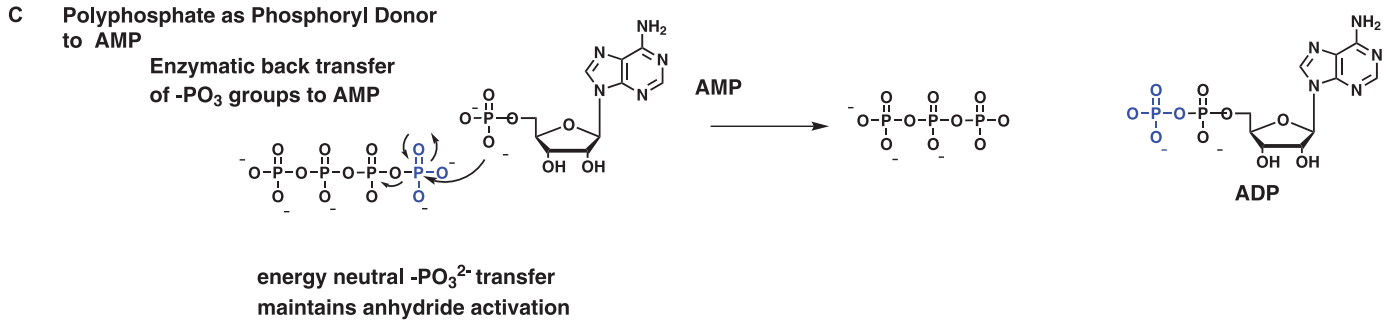


Figure 2.14 Linear inorganic polyphosphates: (A) Nonenzymatic formation of inorganic polyphosphates, perhaps by geological processes; (B) Enzymatic formation of linear polyphosphates by phosphoryl transfer from ATP to a growing polyphosphate chain by polyphosphate kinase. (C) Polyphosphate as an enzyme-catalyzed donor of phosphoryl groups to ADP.[31]

A second enzyme, working in the opposite direction, validated the thermodynamic activation of the P–O–P bonds in polyphosphate chains by transferring a $-PO_3^{2-}$ group from the poly-P donor to AMP as an acceptor. (Note in Figure 2.14C it is the AMP-oxyanion that is the nucleophile and an electrophilic phosphoryl group fragment undergoing transfer from the polyphosphate). Kornberg's group also characterized both exo- and endophosphatase activities (hydrolases) acting on poly-P chains (data not shown). Analogously, some bacteria can donate a PO_3^{2-} group from poly-P chains to GDP to make GTP.[33]

More recently, Xie and Jakob[34] have summarized subsequent findings that these linear inorganic polyphosphates are multifunctional polyanionic scaffolds for sets of partner proteins, in particular as nonprotein scaffolding and chaperoning agents to assist in protein folding. In some thrombocytes and mast cells the poly-P chains in dense granules represent up to [130 mM Pi] if totally hydrolyzed. Consistent with such localization, polyphosphates activate the protease factor XII in the blood coagulation cascade and also chelate calcium ions. The Jakob research group found that these polyanionic phosphoric anhydride chains at micromolar concentrations prevented protein aggregation under a set of otherwise proteotoxic cellular stress conditions. Xie and Jakob[34] posited "poly-P, which itself is redox-inert and thermostable, might have served as the primordial chaperone, and now functions in the first line of defense under extreme stress conditions that might inactivate protein-based chaperones."

It is still unclear at this point how a polyanion stabilizes or reroutes partially folded or unfolded protein species often thought to have hydrophobic regions exposed. It has been proposed that β-sheet secondary structure regions of proteins are most sensitive to poly-P protection. Osmotic and metal cation-chelating properties of the highly anionic poly-P chains may be relevant in such protections. Intriguingly, there is also evidence that long poly-P chains accelerate amyloidogenic rates, perhaps converting toxic intermediates into inactive fibrils, consistent with stabilizing pro-aggregatory β-sheet folding intermediates.

Among the current several unknowns of these presumably ancient polyanionic phosphoric anhydride-containing polyphosphate chains in animal cells is their mode of assembly. Figure 2.14C shows the structure of a growing inorganic polyphosphate chain in the active site of a eukaryotic PolyP polymerase.[31]

While there may be a parallel kinase, family yet to be discovered, it is perhaps intriguing to think that intracellular membrane compartments running proton-motive forces backwards from the proton-pumping PPiases noted in Section 2.6.1.2 above are responsible. This would be equivalent to the F-type ATPases that use proton gradients to make ATP rather than hydrolyze it (Figure 2.13). By analogy, the proton gradients in animal cell subcellular compartments could turn Pi into long chain polyphosphates. If so, that could be an evolutionary relic of early membrane or proton gradient routes to thermodynamically activated but kinetically stable inorganic phosphoric anhydride polymeric chains.

2.8 Organic Pyrophosphate Scaffolds. Monosubstituted Organic Pyrophosphate Metabolites: Incorporation of Inorganic Phosphate and Pyrophosphate Groups into Organic Scaffolds Enables a Wide Range of Chemistries in Energy Metabolism

From a structural and functional perspective, the phosphoric anhydride linkage embedded in pyrophosphate has proven to be such a successful chemical entity that it has spread to many organic metabolites in organismal biologies. Thus, the activated but kinetically stable pyrophosphate group has been embedded within organic frameworks of several classes of extant cellular metabolites to serve a variety of strategic roles (*e.g.* Figure 2.10).

To explore some of those diverse metabolic roles we look at monosubstituted and then disubstituted pyrophosphate groups. The monosubstituted pyrophosphates, attached at only one end, comprise several distinct structural types and sit in quite diverse metabolic pathways. The nucleoside diphosphates and triphosphates fit in this category and represent one introductory path to the structure and biological function of their pyrophosphate side chains. Also, in this category is the 5-phosphoribosyl-1-pyrophosphate building block for purine and pyrimidine precursors of RNA and DNA. So too is Δ^2-isopentenyl-pyrophosphate where enzymatic cleavage of the pyrophosphate moiety is the first essential step in isoprenoid chemical biology (Figure 2.10).

Then, we turn to classes of molecules where both ends of the inorganic pyrophosphate group have been tethered to organic functional groups. We note some of the different metabolic intersections where such disubstituted pyrophosphates are encountered.

2.8.1 Nucleotidyl Transfers to Cosubstrate Nucleophiles

In many instances, such as the four common nucleoside triphosphate building blocks for RNA and the congeneric four 2′deoxynucleoside triphosphates for DNA, one of the two side chain activated phosphoric anhydride bonds in the triphosphate side chains are cleaved to release nascent pyrophosphate (Figure 2.9). Then, the tandem inorganic pyrophosphatase cleavage strategy kicks in to draw the first biosynthetic step – in these cases RNA or DNA chain extensions – towards completion.

ADP, shown along with ATP in Figure 2.10, in its diphosphate side chain also possesses such an activated phosphoric anhydride moiety. One might have imagined organisms could have evolved to split that pyrophosphate chain in a host of phosphoryl transfers to nucleophilic cosubstrates. In the event, there are very few extant examples of enzymes using ADP as an electrophilic $-PO_3^{2-}$ group donor. That role seems to have been arrogated to ATP instead. We do note in passing that the key step converting ribonucleotide

RNA precursors to 2′deoxribonucleotide DNA precursors does occur at the diphosphate not the triphosphate levels.[35] That may (or may not) reflect an evolutionary history when nucleoside diphosphates were the more abundant nucleotides in early organisms.

2.8.2 C–O Bond Cleavage in PRPP and Δ^2-IPP to Generate Carbocationic Transition States

Two other monosubstituted pyrophosphate metabolites are depicted in Figure 2.10. The activated form of D-ribose for both pyrimidine and purine nucleotide biosynthesis is 5-phospho-D-ribose-1-pyrophosphate (PRPP). (We shall note in Chapter 5 that the pyrophosphate group in PRPP arises from direct enzymatic transfer of that phosphoryl group from ATP by 5-phosphoribose pyrophosphokinase). The second set of isomeric monosubstituted pyrophosphate metabolites in Figure 2.10 are the isomeric Δ^2- and Δ^3-isopentenyl pyrophosphates (IPPs), also termed isopentenyl diphosphates. We shall note that the Δ^2-isomer is formed initially in the biosynthetic pathway to all (55 000 known) isoprenoid metabolites and then isomerized enzymatically to the Δ^3-isomer.[36]

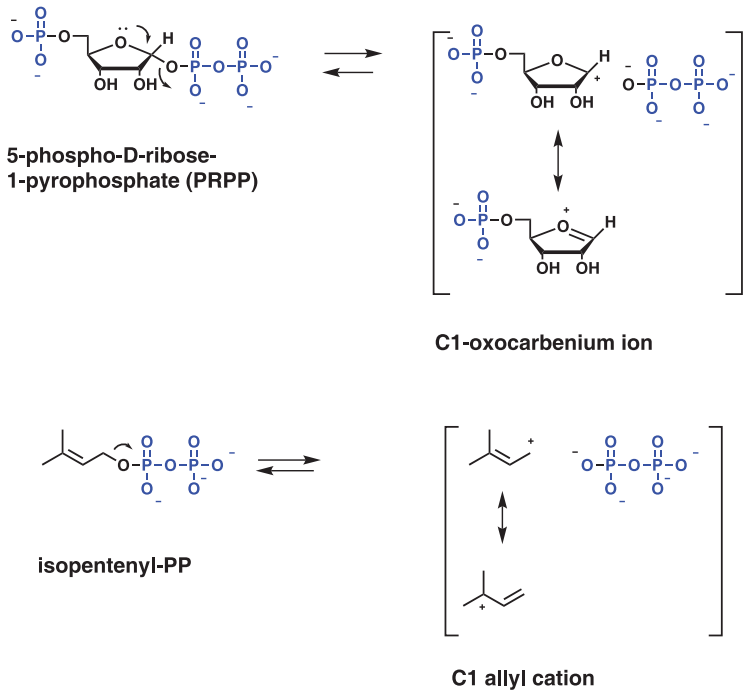


Figure 2.15 Two central metabolites that are precursors to carbenium ions in metabolic transformations are 5-phosphoribosyl-1-pyrophosphate and Δ^2-isopentenyl diphosphate.

The pyrophosphate (–PP) group in both PRPP and the Δ^2-IPP serves a similar role in their multiple enzymatic transformations. They are low energy leaving groups (the nascent pyrophosphate has several resonance contributors that lower its energy), enabling C–OPP bond cleavage. The result of a unimolecular C1–OPP bond cleavage in PRPP would be a 5-phosphoribosyl oxocarbenium ion at C1. In Δ^2-IPP the result would be a stabilized C1-allyl carbocation. Thus, the pyrophosphate groups in PRPP and Δ^2-IPP can be progenitors of carbocationic transition states (Figure 2.15). The tandem action of inorganic pyrophosphatase ensures forward pull on the equilibria of the first enzymes in the two-enzyme couple.

2.8.3 Pyrimidine and Purine *N*-phosphoribosylations of the Phosphoribosyl C1-oxocarbonum Ion Transition State

What are the fates of the incipient carbocations? Figure 2.16 A shows that in the third step of *de novo* pyrimidine biosynthesis in all cells, N_1 of orotic acid is phosphoribosylated by the carbocationic transition state from cosubstrate PRPP. The products are orotidine monophosphate (OMP)and PPi. PPi is then hydrolyzed to 2 Pi molecules, pulling the formation of OMP to completion. This key OMP pyrimidine nucleotide is next decarboxylated to UMP one of the four canonical bases in RNA. Later in the biosynthetic pathway UTP is converted to cytosine triphosphate (CTP), so both pyrimidine building blocks for RNA come from this common OMP.

Figure 2.16 A reveals similar logic for the C1-oxocarbenium ion from PRPP in purine biosynthesis, but this time, as the very first committed biosynthetic step. PRPP and glycine are reacted to give PPi and β-1-amino-ribose-5-phosphate. This step sets the β-ribosyl stereochemistry of all purine nucleotides and provides the amine group to systematically build up the full bicyclic purine scaffold on the way to IMP as first purine nucleotide, Note again that PPiase activity in tandem with the first enzyme draws the combined equilibrium to accumulate the amino-ribose-5-phosphate.

These coupled PPiase equilibria in both pyrimidine and purine biosynthetic pathways are particularly important because in every cell cycle some 1.5 billion molecules of each of the two purine and two pyrimidine nucleoside diphosphates must be made and converted to 2′dNDPs, then to 2′dNTPs for genome replication in the S (DNA Synthesis) phase of the cell cycle. Then, there are the additional tens to hundreds of millions of copies of the four NTPs for all the RNA synthesis also being generated in cells.

2.8.4 Iterative C–C Bond Forming Chemistry from the C1-allyl Carbonium Ion Transition State in Isoprenoid Biosynthetic Pathways

The allylic carbocation arising from C_1–OPP bond cleavage as Δ^2-IPP is processed might be an almost equally abundant carbocation metabolite in

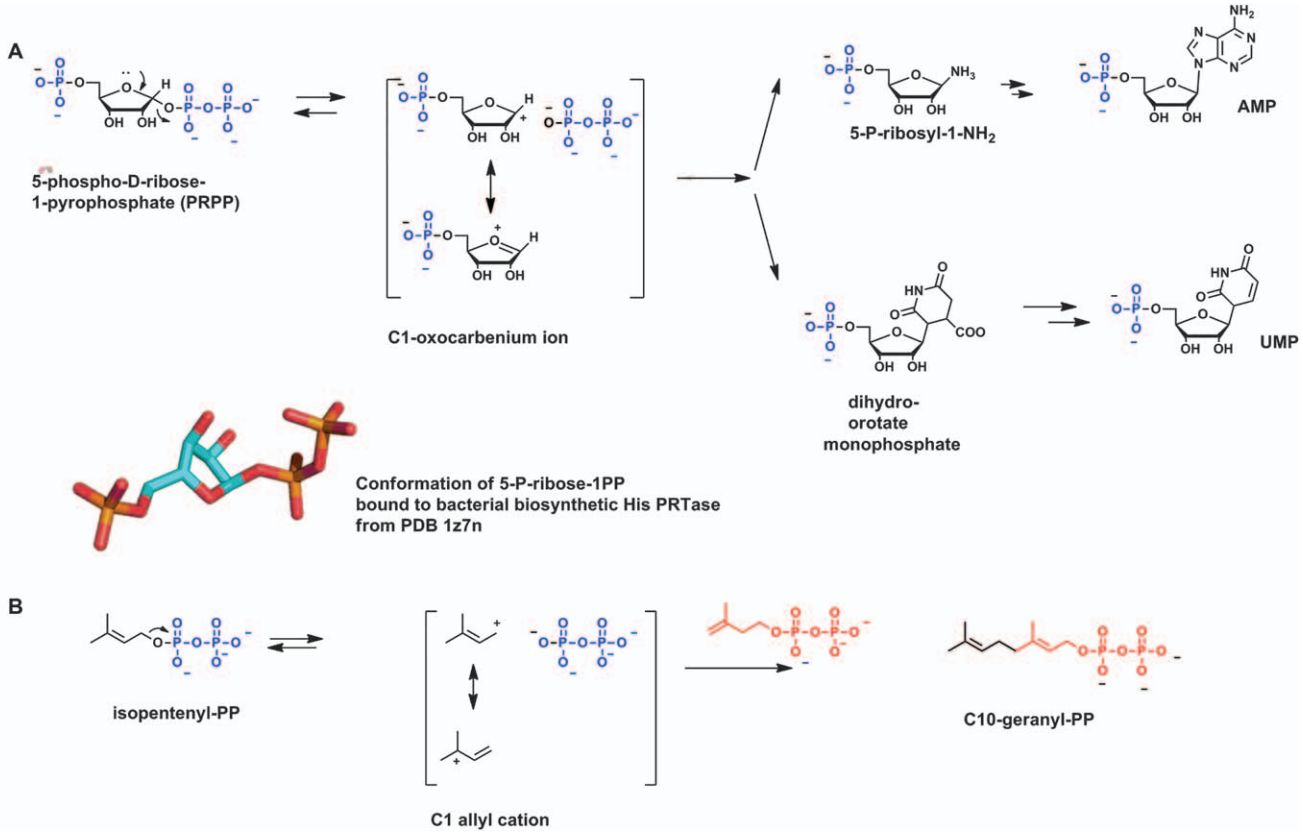


Figure 2.16 (A) PRPP is the donor of 5-phopho-ribosyl units in biosynthesis of all nucleotides for both RNA and DNA biosynthesis; (B) The Δ^2-isopentenyl diphosphate is the precursor to allyl carbocations in all isoprenoid and steroid biosynthetic pathways.

nature. Among its most frequent partner reactants is the Δ^3-IPP monomer, where the π electrons of the Δ^3-double bond act as a carbon nucleophile. The resulting C–C bond connects C3 of the nucleophile with C1 of the electrophilic five-carbon partner to yield the carbocation of the ten carbon geranyl diphosphate (Figure 2.16B).[37] Quenching of that cation by proton removal from C4 yields geranyl diphosphate, a Δ^2-isoprenyl diphosphate.

The genius of this enzyme-directed coupling is revealed when the next enzyme in various isoprenoid pathways iterates the first C–OPP bond cleavage, generating now the C10 allyl cationic transition state. Capture by the π electrons of another five carbon Δ^3-IPP acting as nucleophile leads to the C15 farnesyl-PP, an elongated Δ^2-IPP. This pairing of allyl cation and π electrons as carbon electrophile and carbon nucleophile in C–C bond formation can be iterated several more times. Menaquinone and ubiquinone coenzymes can have 4–10 prenyl units in their 20–50 carbon side chains. The mammalian *N*-glycoprotein machinery can link from 18 to 21 five carbon units, (90–105 carbon long prenyl chains). At the C15 level two molecules of the C15 farnesyl-PP can be condensed head to head to shunt carbon flux into squalene (C30) and then to cholesterol and the panoply of other cyclic sterol-like frameworks.

2.9 Overview

The monosubstituted pyrophosphate chains in these two categories of metabolites-ATP and other NTPs *vs.* PRPP and the Δ^2-IPP isomers reflect two distinct properties and uses of the monofunctionalized –PP units. In the ATP and many related cases of nucleotidyl transfer, a thermodynamically activated phosphoric anhydride P–O–P linkage is often cleaved and provides (much of) the driving force for the particular reaction.

In the PRPP and Δ^2-IPP examples, the P–O–P bond stays intact, the pyrophosphate group departs, prying out an oxygen atom that had initially been attached to the reactive carbon center (Figure 2.16). As the nascent pyrophosphate departs, accessible, stabilized carbocationic transition states in the residual organic scaffold are generated and captured by cosubstrate nucleophiles (for N–C and C–C bond formations) This latter strategy is in play in two of the most abundant and central biosynthetic pathways in all free living organisms: *de novo* nucleotide biosynthesis and *de novo* isoprenoid and steroid biosynthesis.

2.10 Disubstituted Organic Pyrophosphate Linkages

Figure 2.17 depicts five central metabolites in which an inorganic pyrophosphate group has been embedded with both ends connected to atoms from the organic portions of the molecular architecture. In all five cases the connecting atoms on both ends are alcoholic oxygen atoms, preserving the formalism of the embedded pyrophosphate group (and giving some

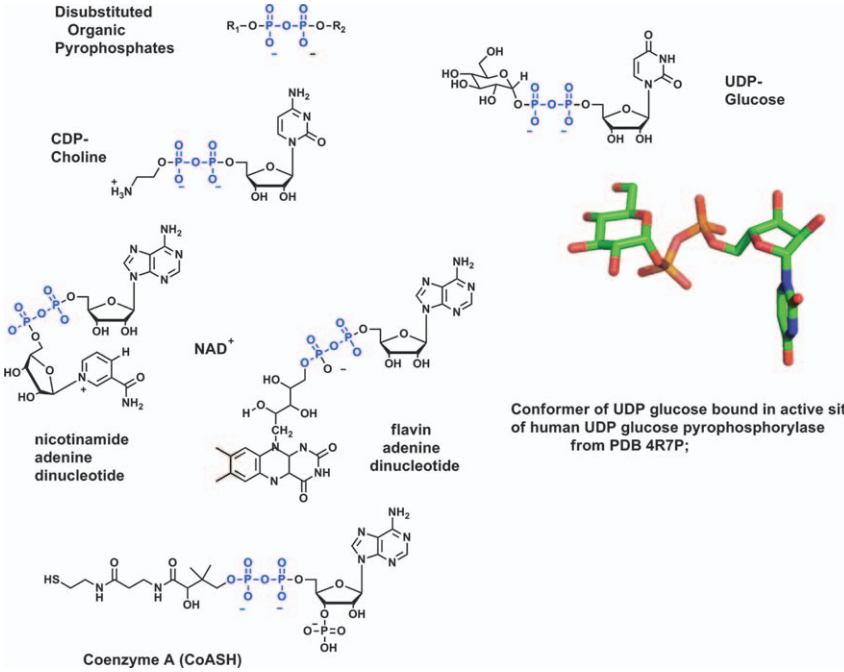


Figure 2.17 Five central metabolites with embedded disubstituted pyrophosphate bridges linking the two halves of the molecules.

preliminary insight into biosynthetic strategy for assembly of such embedded pyrophosphates)

UDP glucose is emblematic of many nucleoside diphosphosugars that are the thermodynamically activated forms of sugar monomers in both oligosaccharide-and polysaccharide biosynthesis as well as in *N*-glycoprotein assembly of carbohydrate chains. One can find, in different organisms and different pathways, similar ADP-glucose, and GDP-mannose activated metabolites.

CDP choline represents a similarly activated precursor to membrane phospholipids. This choline moiety can be captured by enzymes adding in different water-soluble head groups to create species such as phosphatidylethanolamine or phosphatidylserine that insert into newly forming membranes as cells divide or form new organelles (*e.g.* in secretory vesicles).

NAD^+ and flavin adenine dinucleotide (FAD) are the most abundant of the nicotinamide-based and flavin-based redox coenzymes respectively. NAD^+ leads a double metabolic life, as the major redox coenzyme and also has nonredox roles for both mono-ADP ribosylation and poly-ADP ribosylation of partner proteins in epigenetic functions.

Also depicted in Figure 2.17 is the structure of coenzyme A, abbreviated as CoASH to emphasize the chemical role of the terminal thiol in coenzymatic function. Coenzyme A and FAD do *not* use the embedded P–O–P

bond in the pyrophosphate moiety for any discernable chemical role. It may well be that these two molecules, (the active forms of vitamins B2, and B5)[36] participating in many central metabolic reactions, have evolved to join a nonreacting AMP moiety to the distinct business ends of the coenzymes. (On the other hand, metabolic cleavage of the P–O–P bond in the pyrophosphate bridge of NAD^+ has recently been detected as a catalytic property of an enzyme from the bacterium *Legionella pneumophila,* noted in Chapter 16.)

The AMP moiety may be brought along to provide binding energy to protein domains that evolved early to recognize AMP. Combining a nonreacting portion with a reacting portion in the coenzymes may provide sufficient binding energy for high affinity and specificity for coenzyme recognition and reaction.[38] In such instances the disubstituted pyrophosphate moiety may be considered a bridging structural element, not employed as a thermodynamically activated unit, and perhaps a vestigial structural moiety to impart specificity.

References

1. V. K. Bansai, in *Clinical Methods: the history, physical, and laboratory examnations*, ed. W. D. Hall, H. K. Walker and J. W. Hurst, Butterworths, Boston, 3rd edn, 1990.
2. H. N. Biber, J. Hernando and N. Forster, Phosphate transporters and their function I, *Annu. Rev. Physiol.*, 2013, **75**, 535–550.
3. H. Komaba and M. Fukagawa, Phosphate – a poison for humans? *Kidney Int.*, 2016, **90**, 753–763.
4. M. Levi, E. Gratton, I. C. Fortser, N. Hernando, C. A. Wagner, J. Biber, V. Sorribase and H. Murer, Mechanisms of phosophate transport, *Nat. Rev. Nephrol.*, 2019, **15**, 482–500.
5. B. P. Pedersen, H. Kumar, A. B. Waight, A. J. Risenmay, Z. Roe-Zurz, B. H. Chau, A. Schlessinger, M. Bonomi, W. Harries, A. Sali, A. K. Johri and R. M. Stroud, Crystal structure of a eukaryotic phosphate transporter, *Nature*, 2013, **496**(7446), 533–536.
6. B. Kasterbaum and T. Drueke, *Comprehensive Clinical Nephrology*, ed., 2010; bC. TO, in MD Text.com ed. A. B. Feingold and K. R. Boyce, 4th edn, 2000.
7. R. Florencia-Silva, G. R. da Silva Sasso, E. Sasso-Cerri, M. J. Simoe and P. S. Cerri, Biology of bone tissue and factors that influence bone cells, *Biomed. Res. Int.*, 2015, 421746.
8. F. Santos-Beneit, The Pho Regulon: a huge regulatory network in bacteria, *Front. Microbiol.*, 2015, **6**, 402.
9. E. Groisman, The pleiotropic two component regulatory system PhoP-PhoQ, *J. Bacteriol.*, 2001, **183**, 1835–1842.
10. (a) F. Barrere, C. van Blitterswijk and K. de Groot, Bone regeneration: molecular and cellular interactions with calcium phosphate, *Int. J. Nanomed.*, 2006, **1**, 317–332; (b) P. H. Bomans, A. Dey, F. A. Muller,

J. Will, P. M. Frederik, G. de With and N. A. Sommerdijk, *Nat. Mater.*, 2010, **9**, 1010–1014.
11. S. Ruijitanapanich, P. Kumpapapn and P. Wanjanoi, Synthess of hydroxyapatite from oyster shell by precipitation, *Energy Procedia*, 2014, **56**, 112–117.
12. F. Ren, X. Lu and Y. Leng, *Ab initio* simulation of the crystal structure and physical properies of carbonated apatite, *J. Mech. Behav. Biuomed. Mater.*, 2013, **26**, 59–67.
13. A. L. Boskey, Mineralization of bone and teeth, *Elements*, 2007, **3**, 385–391.
14. H. Jiang, X.-Y. Liu, G. Zhang and Y. Li, Kinetics and template nucleation of self-assembled hydroxyapatite nanocrystallites by chondroitin sulfate, *J. Biol. Chem.*, 2005, **280**, 42061–42066.
15. A. K. Nair, A. Gautieri, S.-W. Chang and M. J. Buehler, *Nat. Commun.*, 2013, **4**, 1724.
16. A. Haider, S. Haider, S. S. Han and I.-K. Kang, Recent advances in synthesis, functionalization and biomedical applications of hydroxyapatite: a review, *RSC Adv.*, 2017, **7**(13), 7442–7458.
17. M. R. W. Brown and A. Kornberg, Inorganic pyrophosphate in the origin and survival of species, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 116085–116087.
18. B. S. Cooperman, A. A. Baykov and R. Lahti, Evolutionary consideration of the active site of soluble inorganic pyrophosphatase, *Trends Biochem. Sci.*, 1992, **17**, 262–266.
19. T. Kajander, J. Kellosaio and Goldman, Inorganic pyrophosphatases: one substrate, three mechanisms A, *FEBS Lett.*, 2013, **587**, 1863–1869.
20. B. Springs, K. M. Welsh and B. S. Cooperman, Thermodynamics, kinetics and mechanism of inorganic pyrophosphatase, *Biochemistry*, 1981, **20**, 6384–6391.
21. R. B. Stockbridge and R. Wolfenden, Enhancement of the rate of pyrophosphate hydrolysis by nonenzymatic catalysts and by inorganic pyrophosphatase, *J. Biol. Chem.*, 2011, **286**, 18538–18546.
22. C. T. Walsh, B. Tu and Y. Tang, Eight kinetically stable but thermodynamically activated molecules that power cell metabolism, *Chem. Rev.*, 2018, **118**, 1460–1494.
23. L. Yang and R.-Z. Laio, DFT study on the mechanism of *E.coli* inorganic pyrophosphatase, *J. Phys. Chem. B*, 2009, **113**, 6505–6510.
24. (a) M. Maeshima, Tonoplast transporters: organizers and function, *Ann. Rev. Plant Biol.*, 2001, **52**, 469–497; (b) M. Maeshima and S. Yoshida, Purification and properties of vacuolar membrane proton-translocating inorganic pyrophosphatase from mung bean, *J. Biol. Chem.*, 1989, **264**, 20068–20073.
25. M. Maeshima, Vacuolar H + phosphatase, *BBA Biomembr.*, 2000, **1465**, 37–51.
26. S.-M. Li, J.-Y. Tsai, C. D. Hsiao, C.-L. Chiu, R.-L. Pan and Y.-L. Sun, Crystal structure of H^+-translocating pyrophosphatase, *Nature*, 2012, **484**, 399–403.

27. K. C. Regmi, G. A. Pizzio and R. A. Gaxiola, Structural basis for the reverse of proton pyrophosphatase, *Plant Signaling Behav.*, 2016, **11**, 257–268.
28. (a) P. D. Boyer, The ATP Synthase- a splendid molecular machine, *Ann. Rev. Biochem.*, 1997, **66**, 717–749; (b) D. Stock, A. Leslie and J. E. Walker, Moecular architecture of the rotary motor in ATP synthase, *Science*, 1999, **286**, 1700–1705; (c) D. Okuno, R. Kino and H. Noji, Rotation and Structure of F0F1 ATPase, *J. Biochem.*, 2011, **149**, 655–664.
29. G. C. K. Roberts, H. Okamoto and M. Futai, *Vacuolar ATPases in animal and Plant cells, Encyclopedia of Biophysics*, Springer, Berlin/Heidleberg, 2013.
30. A. Kornberg, Inorganic polyphosphate- toward making a forgotten polymer unforgettable, *J. Bacteriol.*, 1995, **177**, 491–496.
31. M. Hothon, H. Neumann, E. D. Lenherr, M. Wehner, V. Rybin, P. O. Hassa, A. Uttenweiler, M. Reinhardft, A. Schmitdt, J. Seiler, A. G. Ladurner, C. Hermann, K. Scheffzek and A. Mayer, Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase, *Science*, 2009, **324**, 513–516.
32. H. Zhang, K. Ishige and A. Kornberg, A polyphosphate kinase (PPK2) widely conserved in bacteria, *Proc. Nat;. Acad. Sci. U. S. A.*, 2002, **99**, 16678–16683.
33. A. Kuroda and A. Kornberg, Polyphosphate kinase as a nucleoside diphosphokinase in *E.coli* and *Pseudomonas aeruginosa*, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 39–442.
34. L. Xie and U. Jakob, Inorganic polyphosphate, a multifunctional polyanionic protein scaffold, *J. Biol. Chem.*, 2018, 2180–2190.
35. P. Nordlund and P. Reichard, Ribonucleotide reductases, *Ann. Rev. Biochem.*, 2006, **75**, 681–706.
36. C. T. Walsh and Y. Tang, *The Chemical Biology of Human Vitamins*, Royal Society of Chemistry, London, 2018.
37. C. T. Walsh and Y. Tang, *Natural Product Biosynthesis Chemical Logic and Enzymatic Machinery*, Royal Society of Cemistry, London, 2017.
38. W. P. Jencks, *Catalysis in Chemistry and Bilogy*, McGraw Hill, New York, 1979.
39. C. Chen, Z. Huang, W. Yuan, J. Li, X. Cheng and R.-A. Chi, *Cryst. Eng. Comm.*, 2011, **13**, 1632–1637.

SECTION II

ATP and Nucleoside Triphosphate Congeners: Substrates for Phosphoryl-, Pyrophosphoryl-, and Nucleotidyl Transferases

The progression from inorganic phosphate to inorganic pyrophosphate to inorganic polyphosphates in the preceding chapter centered on the P–O–P anhydride bond connections. The accessible pK_{a1} and pK_{a2} values for phosphoric acid dissociation constants hold for PPi and extended polyphosphates such that they are polyanions at physiological pH, contributing to their kinetic stability.

The biosynthetic attachment of the pyrophosphate and triphosphate side chains to nucleosides extends the phosphoric anhydride paradigm to nucleotide chemical biology. Adenosine diphosphate (ADP) and adenosine triphosphate (ATP) are first among other nucleoside diphosphate and nucleoside triphosphate equals in powering almost every aspect of energy metabolism in cells. Although RNA biosynthesis pairs ATP with GTP, CTP, and UTP, and DNA biosynthesis utilizes the 2'deoxy versions dATP, dGTP, and dCTP, along with 5-methyl dUTP (=dTTP), ATP dominates and exemplifies how P–O–P anhydride chemistry funnels into every aspect of primary metabolic pathways.

The triphosphate side chain of ATP offers three potential electrophilic phosphorus atoms, Pα, Pβ, and Pγ for attack by electron rich cosubstrate

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

nucleophiles. Chapters 3, 4, and 5 examine the scope and mechanism of the three regiochemical alternatives.

Attack at Pγ results in phosphorylation of the incoming nucleophile with the $-PO_3^{2-}$ group. This is phosphoryl transfer enzymology (Chapter 3). Attack at Pα leads to attachment of the AMP moiety to the nucleophile with release of PPi. This is the predominant mode of ATP and other NTP cleavages in the biogenesis of biopolymers (Chapter 4). The third mode of attack at Pβ, constituting pyrophosphoryl transfer to incoming nucleophiles, occurs less frequently but at key metabolic intersections (Chapter 5).

The fourth chapter of this section, Chapter 6, examines other scaffolds containing thermodynamically activated but kinetically stable types of phosphoryl groups as comparable donors of electrophilic $-PO_3^{2-}$ groups The interconversions between P–O–P anhydride linkages in ATP and congeners with the activated acyl phosphates, enol phosphates and sulfuryl-phosphoric anhydrides are examined. Finally, the three routes to ATP biosynthesis, from acyl phosphates, from enol phosphate and from coupling to proton gradients (ATP synthase) are compared.

CHAPTER 3

ATP as the Premier Biological Phosphoryl Transfer Reagent

3.1 Phosphoryl Group Transfers under Enzymatic Control

The two major types of phosphate-containing groups in biology are on the one hand the phosphoric anhydride linkages, found in pyrophosphate, ATP and FAD for example, and on the other hand phosphate monoesters and phosphate diesters, both anionic at physiological pHs (Figure 3.1). In contrast to the phosphoric anhydride linkages that are kinetically stable but thermodynamically susceptible to hydrolysis under physiological conditions, both phosphate monoesters and diesters are indefinitely stable in biological milieus. They require enzymatic catalysis for phosphoryl transfer to water or other cellular nucleophiles.

The enzymatic transfer of phosphoryl groups ($-PO_3^{2-}$) to cosubstrates can happen in two kinds of reactions. The first and most common is carried out by the family of kinases, more than 600 of them in humans (517 of them being protein kinases) where it is the γ-PO_3^{2-} group of Mg–ATP that is transferred to a cosubstrate nucleophile. These phosphoryl group transfers typically generate phosphomonoesters (Figure 3.2), such as glucose-6-phosphate or phosphoserine and phosphotyrosine side chains in protein posttranslational modifications. The second and complementary category is subsequent enzymatic hydrolysis of phosphate monoesters and phosphoryl group transfer to water (Figure 3.3).

An additional unique case involves inorganic phosphate itself as substrate. In the active site of the F-type ATP synthase, noted later in this chapter, one of the β-phosphoryl oxyanions of bound Mg–ADP acts as a nucleophile on the phosphorus central atom of inorganic phosphate.

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

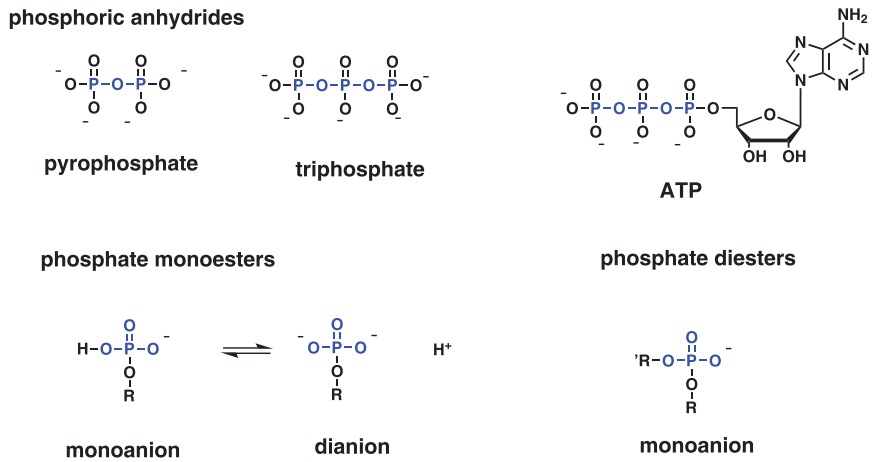


Figure 3.1 Three major classes of phosphate metabolites: phosphoric anhydrides, phosphate monoesters, and phosphate diesters.

In a formal sense organic phosphate is the phosphoryl donor despite the absence of any thermodynamic activation for $-PO_3^{2-}$ group transfer (Figure 3.2). One can view it as a formal dehydration of inorganic phosphate (!). This remarkable reaction in human adults happens at levels up to 70 kg per day! It is unique and yet ubiquitous as we note in a subsequent section of this chapter.

One other facet of inorganic phosphate reactivity exploited by enzymes is taken up in Chapter 9 in which inorganic phosphate is a substrate and sugar monophosphate esters are products. This involves inorganic phosphate as a *nucleophile at oxygen* rather than an *electrophile at phosphorus* as in the ATP synthase example just noted. This is not $-PO_3^{2-}$ group transfer to nucleophilic cosubstrates. Rather, it is glycosyl transfer to inorganic phosphate. All four oxygens of inorganic phosphate end up in the sugar-1-phosphate products (Figure 3.2). There are about three dozen validated enzymes in this class known as phosphorylases. They are the subject of Chapter 9 and emphasize the diversely and dual reactivity of inorganic phosphate chemical biology.

Phosphorylases are notionally akin to hydrolase enzymes. Hydrolases use water as a nucleophile in reaction with some electrophilic fragment of a substrate. Phosphorylases substitute inorganic phosphate as an oxygen nucleophile in place of water. All three types of biologically prevalent phosphates, phosphoric anhydride, phosphomonoesters, and phosphodiesters, undergo enzyme-mediated transfers of the phosphoryl groups to water (Figure 3.3). These transformations constitute the back half of phosphoryl transfer enzymology as noted above.

The ATP hydrolases are known generically as ATPases and the hundreds of them are categorized in Section 3.6. Phosphomonoesters are cleaved by phosphatase enzyme families and they are the subject of Chapter 7.

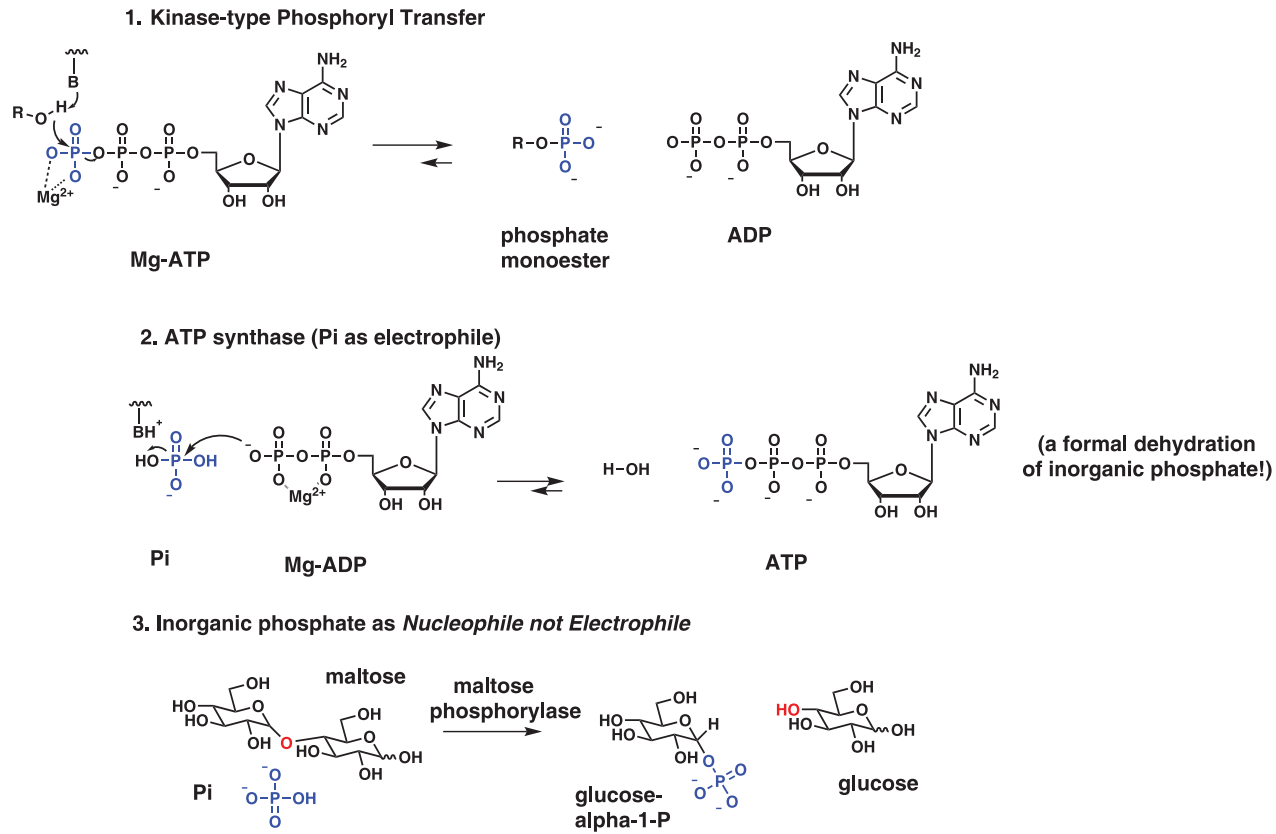


Figure 3.2 Kinase-type phosphoryl transfers; ATP synthase; inorganic phosphate as nucleophile not electrophile.

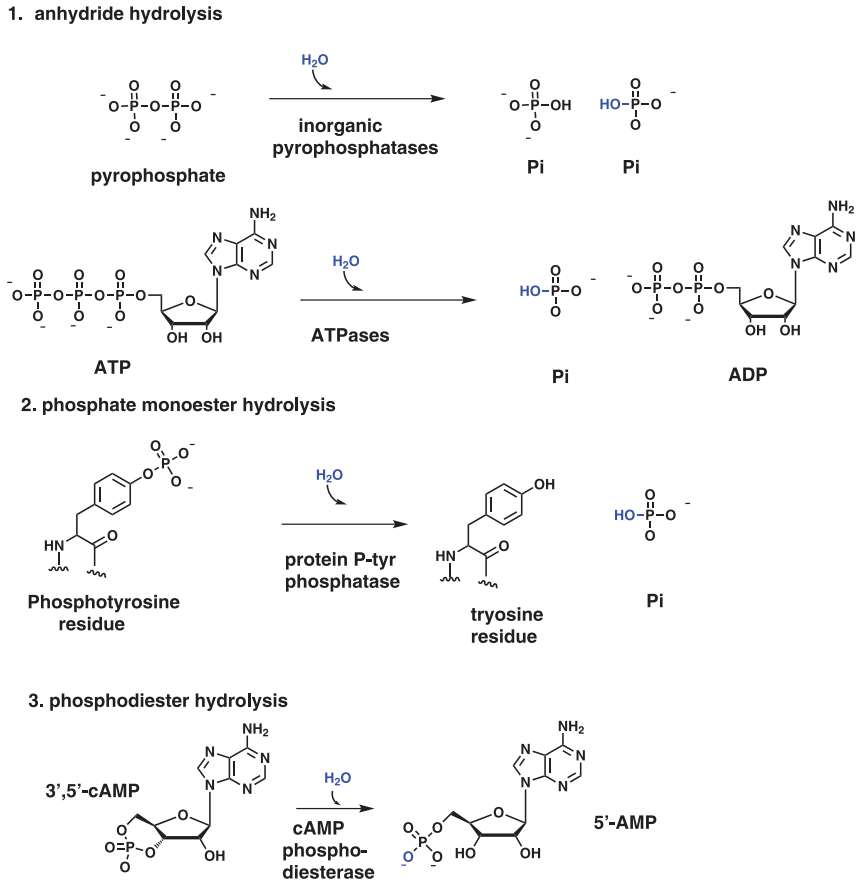


Figure 3.3 Enzyme catalyzed hydrolyses of the three major classes of phosphate metabolites.

Analogously, phosphodiesters are hydrolyzed by phosphodiesterase enzymes. Over 100 phosphodiesterases acting on the internucleotide phosphodiester bonds in RNA and DNA molecules have been described. They are termed nucleases and are discussed in Chapter 8.

3.2 Timing and Mechanism of Phosphoryl Group Transfers

Phosphoryl group transfers were intensively studied by Westheimer and colleagues leading up to his 1987 comprehensive review[1] (''Why. Nature chose phosphates'') and then by the Jenck research group[2] and many others to understand the basic phosphorus chemistry that would be in play in the diaspora of enzymatic catalyses. One can illustrate some of those issues in reaction of a generic phosphomonoester (Figure 3.4A) undergoing

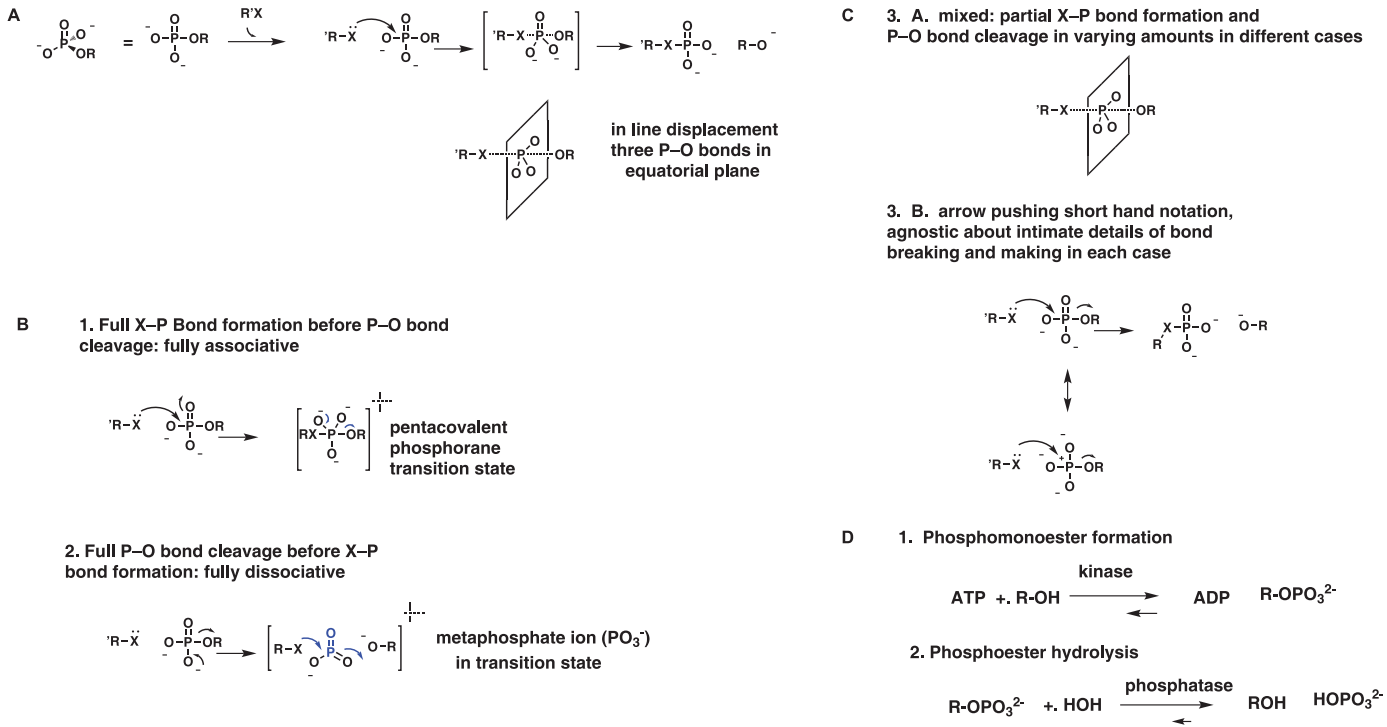


Figure 3.4 (A) Generic phosphoryl transfer of a phosphomonoester to an incoming nucleophile RX: in-line displacement geometry; (B) Two limits: associative *vs.* dissociative transition states; (C) Mixed transition state with partial X–P bond formation and P–O bond cleavage as indicated by dashed lines. (D). Two stages of phosphoryl transfers in metabolites: kinase-mediated formation and phosphatase-mediated breakdown.

$-PO^{2-}$ transfer to an incoming nucleophile RX (*e.g.* HOH). Among the conclusions reached were that ''in line'' geometry was favored in transition states where the incoming RX nucleophile and the departing RO alcohol were in apical positions, with the three other P–O bonds arranged equatorially.

The timing of X–P bond formation and P–O bond breakage could be variable. In one limit, an early transition state for X–P bond formation before O–P cleavage, an associative mechanism would apply, there would be five full bonds to phosphorus (a pentacovalent phosphorane), however transiently, before the O–P bond was broken (Figure 3.4). In the other limit, where the O–P bond breaks early before the X–P bond is formed, a dissociative mechanism, there would be a PO_3^- metaphosphate ion species, capturable by the RX nucleophile. In many cases the timing of (P–O) bond breakage and (X–P) bond formation would be closer to simultaneity (dashed X–P and P–O bonds in Figure 3.4).

The timing of enzymatic phosphoryl transfers that span hundreds of low molecular weight and protein kinase reactions, the F-type ATP synthases, and both phosphatases and phosphodiesterases have also been subjected to study. We will note in Chapters 7 and 12 that stereochemical studies indicate inversion of stereochemistry in every case examined, arguing against a freely rotating metaphosphate intermediate (which should give racemization at the itinerant phosphorus). On the other hand, we will note in Chapter 10 that phosphoenol pyruvate (PEP) mutase, the route to C–P-containing phosphonate metabolites is proposed to involve a dissociative transition state.

The results from nonenzymatic mechanistic analyses of phosphoryl transfers indicate that phosphate monoesters may proceed largely through dissociative transition states, phosphodiesters *via* mixed associative and dissociative transition states. Phosphotriesters may favor associative transition states but they play a very limited role in biological systems (we discuss these in Chapter 12 on phosphorothioates).

Because there is no direct evidence for pentacovalent phosphoranes as general species in enzyme-mediated phosphoryl transfers of any of the types noted above, we will write enzymatic phosphoryl transfers. (and pyrophosphoryl and nucleotidyl transfers) in the arrow pushing schemes of organic chemistry as depicted in Figure 3.4B. We eschew writing explicit pentacovalent phosphoranes where the P=O electrons would rehybridize in every transfer, given the absence of evidence of that participation. Indeed, given that one of the resonance contributors to phosphate monoesters bears a full positive charge on phosphorus in a $P^+–O^-$ contributor, we write direct arrow displacements as shorthand notations (Figure 3.4C). These arrow pushing diagrams are in the backdrop of in line geometry for displacements, tetrahedral geometries in the ground states of phosphate compounds (Figure 3.4C) and the implicit likelihood that most of the enzymatic reactions will have partial bond formation and bond cleavages in the transition states where phosphoryl groups (and pyrophosphoryl and nucleotidyl groups) are in flight (dashed or partial bond-making and breaking in a canonical phosphoryl group transfer: Figure 3.4C).

3.3 Phosphoryl Group Transfers from Mg–ATP: Electrophilic γ-PO_3^{2-} in Flight

In the preceding chapter we focused on inorganic pyrophosphate with its thermodynamically activated but kinetically stable P–O–P anhydride linkage as serving two essential purposes. The first is as donor of an electrophilic $-PO_3^{2-}$ fragment to cellular nucleophiles: phosphoryl group transfers. The second, and related attribute is that when water is the attacking nucleophile, in the presence of pyrophosphatase hydrolytic enzymes, the hydrolysis is accelerated 10^{10} to a k_{hydrol} of 10^3 per second *vs.* a nonenzymatic hydrolysis rate measured in $months^{-1}$.[3]

We also noted in Figures 2.10 and 2.17 a few representative structures of central metabolites that contain either monosubstituted pyrophosphate groups or disubstituted pyrophosphate moieties. Those metabolites retain the thermodynamic activation conferred by the P–O–P anhydride linkages. In this chapter we turn to the two most prominent of the metabolites with monosubstituted pyrophosphate side chains: ADP and ATP (Figure 3.5). ATP predominates over ADP in cells by a factor of ~200 : 1, with [ATP] in the millimolar range, indicating a greater utility of ATP, with its two P–O–P phosphoric anhydride moieties over ADP, with its single P–O–P linkage.

In principle, any of the three metabolites, inorganic pyrophosphate, ADP, and ATP, should be able to deliver an electrophilic phosphoryl group to a nucleophilic cosubstrate. Indeed, Figures 3.6 and 3.7 show three such enzymatic examples. First is the pyrophosphate–AMP phosphoryl transferase found in archaeal microbes and perhaps representing an early evolutionary vestige of how to form ADP from AMP.

Second is an enzyme initially thought to be of substantially restricted biological distribution but also detected in mammalian cells, a transferase where ADP is the $-PO_3^{2-}$ donor to the C6–OH of D-glucose to yield glucose-6-phosphate and AMP.[4] Glucose-6-phosphate is the canonical starting point for the energy-yielding glycolytic pathway in all prokaryotic and eukaryotic cells. This enzyme illustrates the chemical capacity of ADP to offer the terminal PO_3 of its phosphoric anhydride side chain to the most abundant cellular hexose, trapping it in cells as the impermeant glucose-6-phosphate dianion. More recently, a PPi-dependent glucokinase has been characterized from an *Anabena* cyanobacterium.[5] Analogously, *Entamoeba histolytica* has a phosphofructokinase with a million fold preference for PPi over ATP for phosphoryl transfer to the C1–OH of fructose-6-phosphate, raising the question of whether PPi rather than ATP may have been the primordial biological donor of $-PO_3^{2-}$ groups.[6]

There is also a validated ADP-utilizing *O*-phosphoserine synthetase in the biosynthetic pathway to cysteine in some microbes,[7] but those two ADP-dependent kinases are rare, compared with many hundreds of ATP-dependent ATP-consuming phosphoryl transferases.

In eukaryotic cells, the enzymatic production of glucose-6-phosphate from glucose transported into cells is mediated not by phosphoryl transfer from ADP, but instead by ATP-cleaving phosphoryl transferases, known for

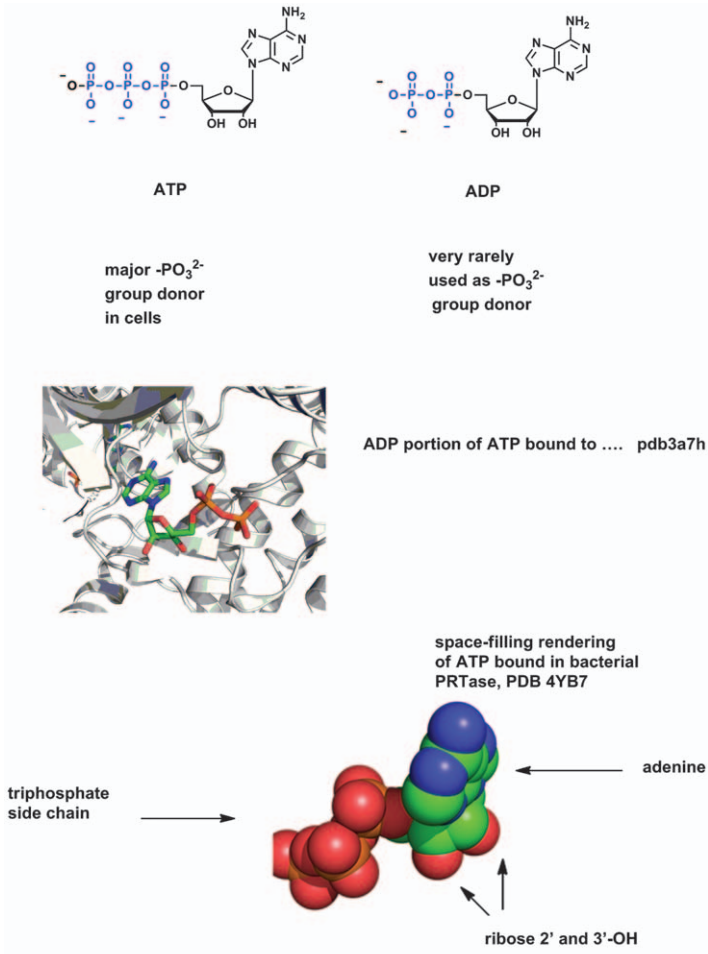


Figure 3.5 ATP *vs.* ADP: Both have phosphoric anhydride-containing side chains but while ATP is the prevalent cellular phosphoryl group donor to nucleophiles, ADP functions only very rarely as phosphoryl donor.

historical reasons as hexokinases[8] (Figure 3.4). Kinases are a generic term for ATP-dependent phosphoryl transfer enzymes. Among the possible reasons for ATP supplanting ADP as the principal cellular phosphoryl donor over evolutionary time may be that the second of the two P–O–P linkages in the triphosphate side chain offers two additional regiochemical modes of thermodynamically favored, phosphoric anhydride-cleaving outcomes. (We will note PEP-dependent phosphorylation of glucose in bacterial transport and trapping of the hexose as glucose-6-phosphate, in detail in Chapter 11.)

As we note in the next section and in Chapter 4, the most important of those additional cleavage routes, transfer of AMP to the incoming nucleophile, releases inorganic pyrophosphate as cosubstrate, and brings into play

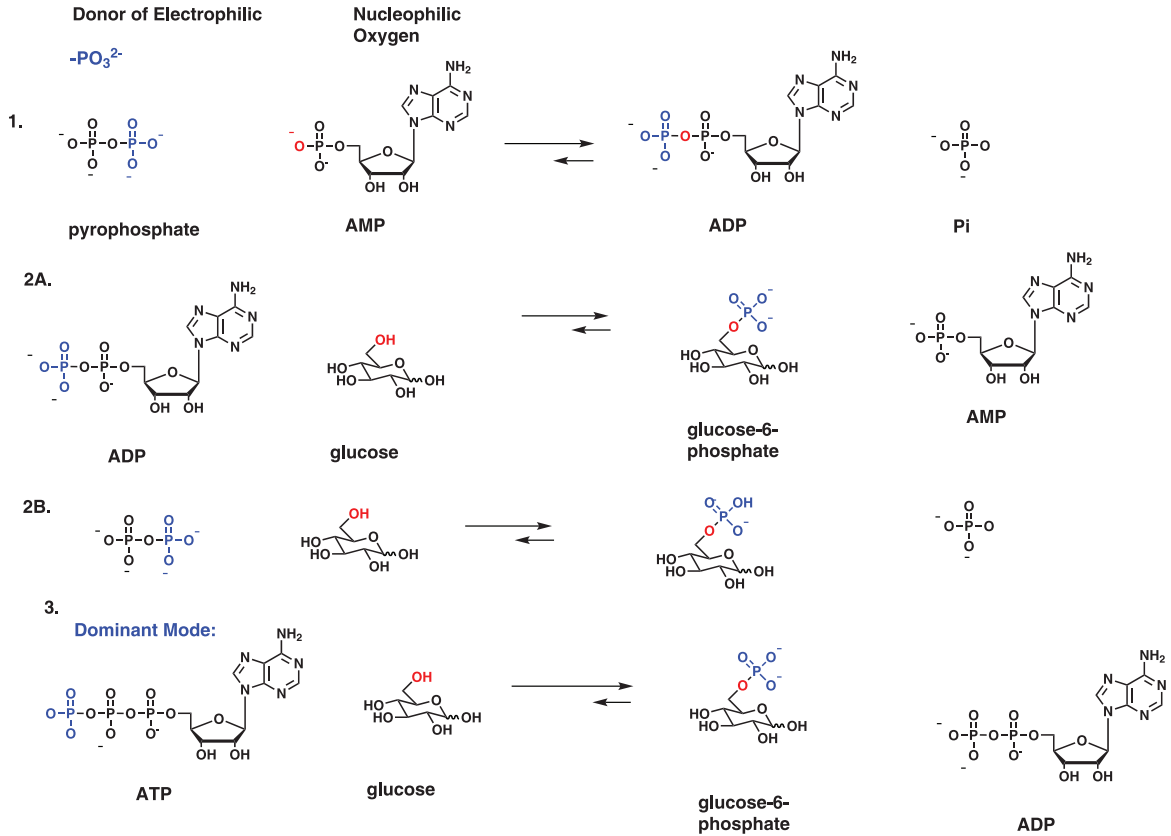


Figure 3.6 Examples of phosphoryl group transfers from PPI, from ADP, and from ATP.

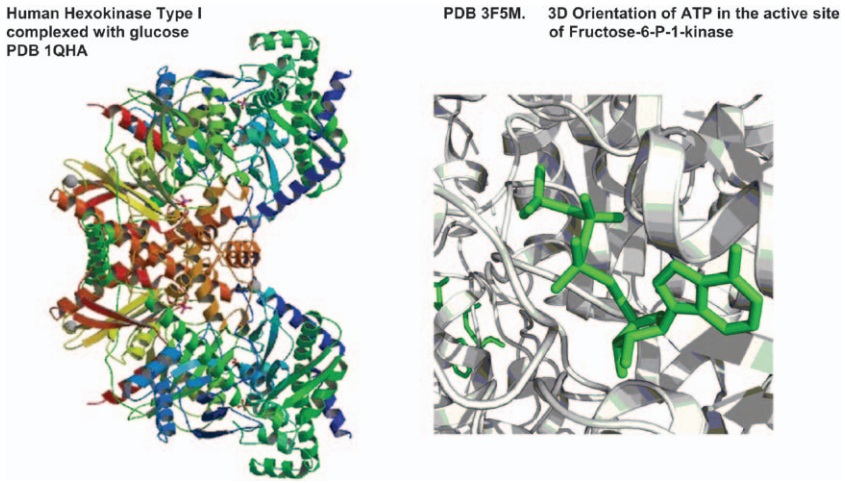


Figure 3.7 Structure of two kinases involved in hexose phosphorylations. Left: hexokinase, right the 3D conformation of ATP (phosphoryl donor) in the active site of 6-phospho-fructose-1 kinase. Structures drawn from pdb 1QHA and 3F5M, respectively.

the coupled action of inorganic pyrophosphatase, emphasized in the previous chapter as a key strategic second step in driving otherwise potentially unfavorable biosynthetic equilibria.

Before turning to the three enzyme-mediated cleavage modes of ATP, however, we note that its predominant role as cellular workhorse for phosphorus chemical biology can be enumerated in several ways. One factoid, taken up in Section IV, is that ATP serves as phosphoryl donor substrate for ~520 protein kinases in the human proteome. The enzyme class is so numerous and consequential for health and disease that those enzymes are collectively termed the human kinome.[9]

A second indicator is the large class of enzymes (~200 in humans) termed ATPases, for the hydrolysis of ATP to ADP and Pi.[10] The energy released in this hydrolysis is captured as protein conformational changes that drive an enormous range of events, from ion pumps across membranes, mechanical force generation in actomyosins, kinesins, and other cellular motors. ATPase action also powers ATP-dependent protein degradation by proteasomes, and many other coupled but otherwise energetically unfavorable actions in cells.

3.4 ATP: Daily Human Consumption Equal to Total Body Weight

Perhaps the single most powerful indicator of the preeminence of ATP in all facets of cellular energy metabolism is the estimate that a 70 kg adult human generates and consumes his or her weight (70 kg = 154 pounds) in ATP every day.[11] That is an astounding number, reveals a dynamic

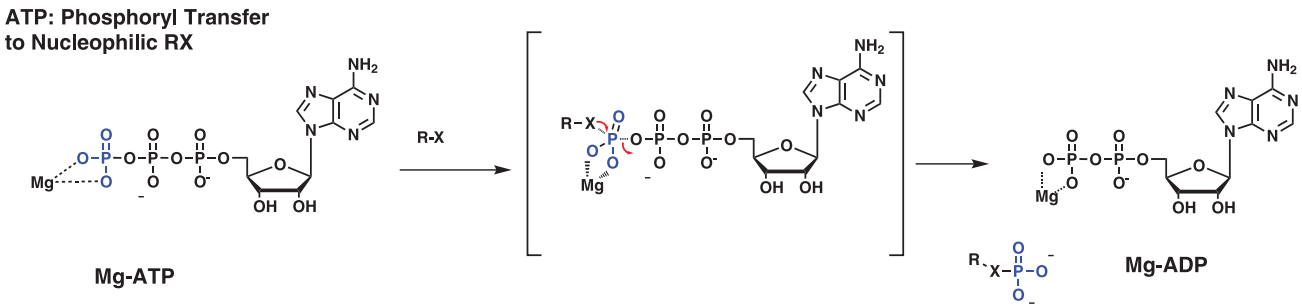
commitment of organisms to make ATP so it can be spent to power essentially all cellular events. By the time a human has reached the age of 75, he or she will have generated and consumed ~*two million kilograms of ATP* during their lifetime.

The steady state levels of ATP in an adult human are about 70–75 g. Thus, to make and utilize 75 kg per day, each person turns over the full 75 g inventory about 1000 times! The cellular concentration of ATP is in the range of 1–10 mM, varying in different cell compartments. Estimates of the lifetime of a given ATP molecule in a cell are in the range of one second. One other way of thinking about the flux of ATP is to note its molecular weight of 507 g per mole. Then, the daily throughput is 75/0.5 or about 150 moles of ATP. Recalling that a mole of a substance is defined as 6×10^{23} molecules, then humans make and utilize $\sim10^{26}$ molecules of ATP to stay alive every day. One could think of humans as biological machines dedicated largely to the formation of ATP and spending its phosphoric bond energies in myriad ways to keep us alive.

As summarized above and detailed in the next section, all these ATP molecules are cleaved by two major chemical routes, to make ADP by phosphoryl transfers or to make AMP by nucleotidyl (adenylyl) transfers.[12] To turn over the ATP pool a thousand times a day means converting all the AMP and ADP product molecules, plus the attendant inorganic phosphate anion coproducts, back to ATP. On examination, most ATP (>95%) is made directly from ADP not AMP, by the mitochondrial ATP synthase, so all that ~70 kg of ADP must be transported, from whatever subcellular compartments in which it is formed from ATP cleavage, back across the mitochondrial membranes to the active site of the ATP synthases.

3.5 Phosphoryl Transfers from ATP: Logic and Mechanisms

One of the two major modes of enzyme-accelerated side chain cleavage of Mg–ATP in several hundreds of cellular reactions that drive coupled equilibria is the transfer of the terminal electrophilic γ-PO_3^{2-} of ATP to a cosubstrate nucleophile. These constitute *phosphoryl group transfers.*[1] Most often those cosubstate nucleophiles are oxyanions of a cosubstrate. Figures 3.8 and 3.9 note five categories of common oxyanion nucleophiles: (1) the alcohol oxygen of many types of alcohol scaffolds, (2) carboxylate oxygens, (3) the peripheral oxyanions of phosphate groups, (4) enolate oxyanions, (5) water. Working backwards, when water is the cosubstrate, the net result is ATP hydrolysis, an ATPase reaction that on the surface looks wasteful in that the inorganic phosphate product has not maintained the thermodynamic activation of the starting β,γ-phosphoric anhydride linkage. However, we shall note that almost all such ATPases store the energy released transiently in driving a conformation change, typically in a partner protein: that constitutes cellular work.[10]



Common RX Substrates	$RX\text{-}PO_3$ Products	
ROH	$ROPO_3$	phosphate monoesters
HOH	$HOPO_3^-$	inorganic phosphate
$R\text{-}OPO_3^{2-}$	$ROPO_2PO_3$	pyrophosphates
$R\text{-}NH_2$	$RNH\text{-}PO_3$	phosphoramidates

Figure 3.8 Categories of oxygen nucleophiles acting as cellular substrates in phosphoryl group transfers from ATP.

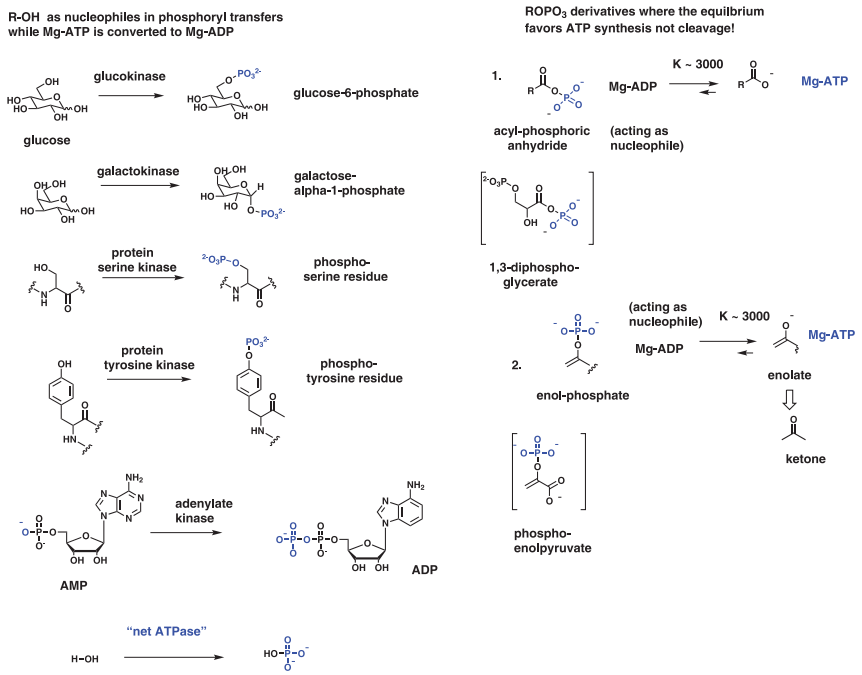


Figure 3.9 Specific enzymatic examples.

In addition to oxygen nucleophiles in cells there are also nitrogen, and sulfur atoms, which are potential reactants towards the electrophilic Pγ of ATP. Surprisingly, given the high concentrations of glutathione and coenzyme A as potent thiol nucleophiles in cells, there are no documented *S*-phosphoryl transfers from ATP to glutathione, coenzyme A or other cellular low molecular weight thiols (There are *S*-phosphoryl-cysteinyl enzyme intermediates—see Chapter 14).

Over the past half-century, the arena of kinase-mediated transfer of the ATP γ-PO_3^{2-} group to side chain-OH groups of serine, threonine, and tyrosine residues in protein substrates has grown exponentially to blot out most other aspects of enzymatic phosphoryl transfers. We will summarize several facets of the protein kinomes of humans in Section IV on phosphoproteomics. We stay focused on low molecular weight substrates for phosphoryl transfer in this section of the book.

3.5.1 Phosphoryl Transfers to Cosubstrate Alcohols

Phosphoryl transfer from ATP to oxygen nucleophiles is thought to proceed *via* associative mechanisms, with transition states in which there is at least partial O–P bond formation from the incoming oxygen nucleophile with at least partial P–O remaining bond character to the potential –O–ADP leaving group.[13] Different electronic and steric factors in the attacking substrate,[3]

regioselectivity of Mg^{++} coordination to the side chain phosphates, and placement of side chains of the enzyme catalyst in the active site, among others, can affect the degree of bond-making and bond-breaking in the proposed associative transition state in any given case.[2b] In the PEP mutase example in Chapter 11, the argument has been made instead for a dissociative transition state with early O–P bond cleavage.

Figure 3.9 depicts phosphoryl transferase enzymes, adenylate kinase (ADP as nucleophile), and ATPases (where HOH is the nucleophile). The first two are classic examples in sugar metabolism. Hexokinase is the first enzyme in glucose metabolism, trapping neutral glucose as the divalent glucose-6-phosphate dianion within cells and preventing diffusion back outside the anion-impermeant cell plasma membrane. The regiochemistry is absolute for the C6–OH, the only primary alcohol group in glucose. Glucose in physiological pH range aqueous solutions is a mix of rapidly equilibrating alpha and beta anomers of the cyclic hemiacetal forms with <1% as the acyclic free aldehyde. The convention of a squiggly line to the C1–OH group is to denote the equilibrating anomeric mixtures (Figure 3.10).

Analogously, the next most common hexose in metabolism, D-galactose, the C4-axial epimer of D-glucose, is also an equilibrating mix of α- and β-hemiacetals at C1. However, the galactokinase enzyme selects out of the mixture only the α-isomer (anomer) of the mix of hemiacetals and acyclic aldehyde forms, activating only that α-C1–OH as the kinetically competent nucleophile. The disaccharide lactose in nursing mother's milk is galactosyl-β-1,4-glucose. Its cleavage by the enzyme lactase gives infants equal mixtures of glucose and galactose as hexose nutrients. The two hexoses get phosphorylated regiospecifically to D-glucose-6-phospahte and D-galactose-1-phosphate, respectively. They only get interconverted after subsequent nucleotidyl transferase action on the galactose-1-phosphate metabolite. We examine the interconversion and two modes of thermodynamic activation of these sugar phosphate monoesters in Chapter 7.

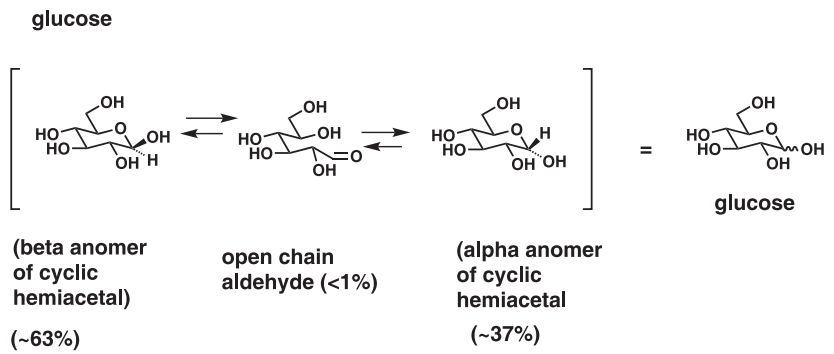


Figure 3.10 The most abundant hexose in metabolism, D-glucose, is a rapidly equilibrating mixture of two cyclic hemiacetal monomers at C1 that interconvert by way of the open chain aldehyde. The aldehyde form, though minor, is the one through which glycolysis occurs.

The other two alcohol kinase reactions shown in Figure 3.9 are for the protein serine–threonine kinases and protein tyrosine kinases, respectively. The mechanism in each case is again a direct phosphoryl transfer to the side chain alcohol groups of particular Ser and Thr, or Tyr residues in a given set of partner proteins recognized by one or more enzyme in the >500 member human protein kinome. We return to these reactions in part IV. Although not shown in this figure recall that Figure 3.6 shows *O*-phosphorylation of free L-serine in a microbial biosynthetic pathway to L-cysteine. In that figure the example was of an unusual ADP-dependent kinase β-PO_3^{2-} donor, rather than the anticipated ATP phosphoryl donor.

3.5.2 Phosphoryl Transfers to Cosubstrate Carboxylates: Mixed Acyl–Phosphoric Anhydrides

When cosubstrate carboxylates are attacking oxygen nucleophiles at the γ-P of Mg–ATP, the phosphoryl transfer step results in an *acyl phosphate* product. Inspection of the generic structure in Figure 3.11 shows that acyl phosphates are mixed anhydrides of carboxylic and phosphoric acids. Thus, they join the category of thermodynamically activated but kinetically stable molecules that are useful as diffusible intermediates in energy metabolism (analyzed in Chapter 6).[12]

Acyl phosphates are doubly activated, for transfer of either portion of the mixed anhydride to nucleophiles. Figure 3.11 shows that capture by thiols is the pathway of *acyl* group transfer to form acyl thioesters, most notably acetyl coenzyme A. The lower path, reaction with ADP to produce ATP, reveals the phosphoryl transfer step. Indeed, reaction of 1,3-diphosphoglycerate with ADP is one of the two reactions in glycolysis that produce ATP. The conversion of mixed carboxylic–phosphoric anhydride of substrate to phosphoric anhydride in an ATP product emphasizes that both tides of these anhydrides are thermodynamically activated.

The bottom half of Figure 3.11 displays the roles of three particular acyl phosphates as metabolic intermediates. Each is formed from attack of the parent carboxylate on the γ-PO_3 of Mg–ATP. The first example, acetyl phosphate, is produced by the bacterial enzyme acetate kinase and in turn is substrate for phosphotransacetylase that carries out acyl transfer to the thiolate anion of coenzyme A to yield acetyl coenzyme A. This acyl thioester is the preferred carrier and donor of acetyl groups in essentially all phases of metabolism.

The second example involves the acyl phosphate, aspartyl-β-phosphate, at the side chain carboxylate of the proteinogenic amino acid L-aspartate. This is then a substrate for double reduction of the acyl phosphate group, to the aldehyde and then the alcohol, producing the amino acid homoserine. Homoserine is subsequently a precursor to cystathionine and ultimately methionine.

The third example involves a tandem pair of aminoacyl- and dipeptidyl-carboxylic–phosphoric anhydrides in the pathway to glutathione. The first

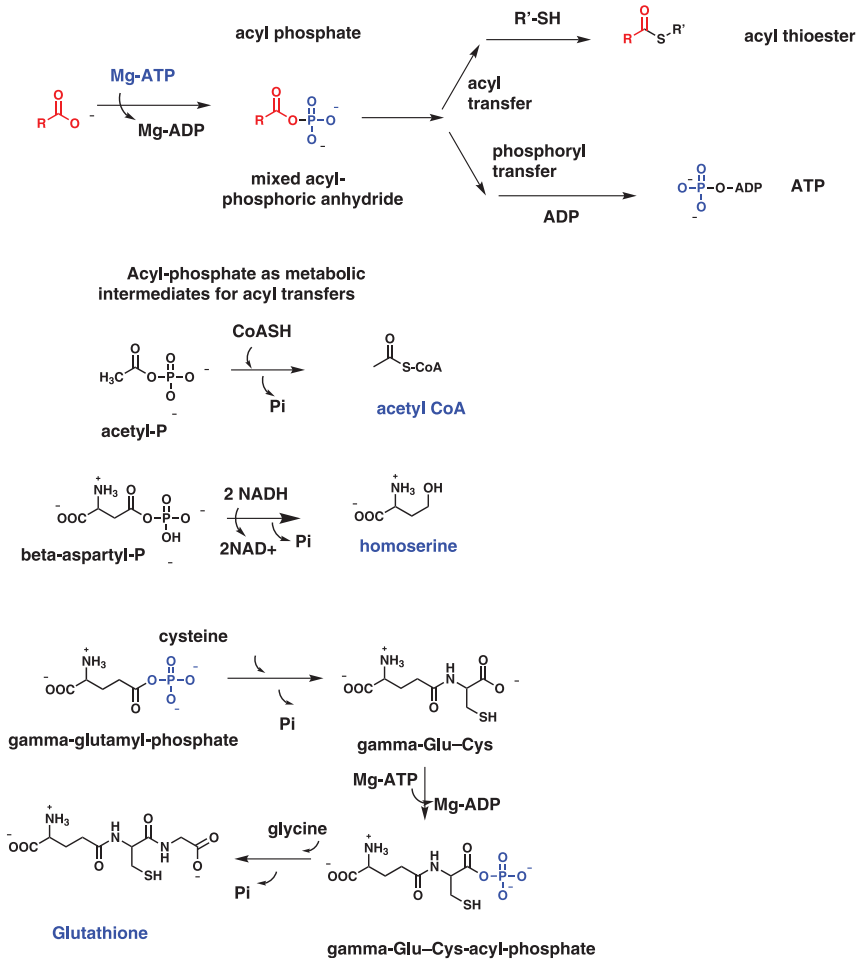


Figure 3.11 Acyl phosphates, arising from phosphoryl transfer from ATP to carboxylate oxygens are mixed anhydrides of phosphoric acid and the organic acid. Three specific acyl phosphate metabolites in primary metabolism are acetyl-phosphate, aspartyl-β-phosphate and glutamyl-γ-phosphate.

enzyme makes γ-glutamyl-phosphate and condenses it with L-cysteine. The amine of cysteine is the nucleophile to which the activated glutamyl group is transferred. The product is the dipeptide γ-glutamyl-cysteine. This dipeptide is then a subject for another kinase, this one making the indicated γ-glutamyl-cysteinyl-phosphate. It is captured by the amine group of glycine for acyl (dipeptidyl) transfer to make the tripeptide glutathione, the major scavenging agent of primary metabolism through its side chain thiol group. This latter enzyme, although clearly a kinase, is instead named glutathione synthetase.

These three examples of acyl phosphate formation and utilization show three different types of enzyme-mediated outcomes: thioester formation (acetyl CoA), acyl phosphate reduction (homoserine), and tandem amide bond formations (glutathione). One might also have listed succinyl phosphate in the citric acid cycle and citryl phosphate in the cytoplasmic ATP-dependent cleavage of citrate back to acetyl CoA for fatty acid biosynthesis, as two additional examples of the involvement of acyl phosphate in primary metabolic pathways.

3.5.3 Phosphoryl Transfer to Cosubstrate Phosphate Esters: Adenylate Kinase, Nucleoside Diphosphate Kinase

The third type of oxygen nucleophile used enzymatically to attack the electrophilic γ-phosphorus atom of ATP is one of the oxyanions of phosphate monoester substrates. Figures 3.12 and 3.13 illustrate that multiple kinase reactions involve nucleotide monophosphate conversions to diphosphates and triphosphates. For example, *de novo* purine biosynthesis produces inosine monophosphate (IMP) that undergoes two distinct aminations to yield AMP and GMP (Figure 3.12 upper). To get to the ADP and GDP level AMP kinase and GMP kinase, respectively, use the AMP and GMP oxyanions as nucleophiles. Analogously, in pyrimidine pathways, the emerging metabolite UMP undergoes two rounds of kinase reactions to get to UTP as one of the four building blocks for RNA chain elongations (Figure 3.12 middle).

The most well studied of the nucleotide kinase family members may be adenylate kinase, (adenylic acid is an old name for AMP; adenylates are the ionized forms) (Figure 3.9 and Figure 3.12 bottom). As the enzyme name implies, AMP (adenylate) functions as a nucleophile in the active site of the kinase and thereby undergoes phosphoryl transfer from ATP. The result is two product molecules of ADP. There are six distinct isoenzyme forms of human adenylate kinases.[14] Figure 3.14 shows the complex of adenylate kinase with two ADP molecules.

The $K_{eq.}$ is close to unity so the reaction can proceed in either direction, although it is thought that physiologically this enzyme is deployed to bring AMP molecules up to the level of ADP. The back reaction utilizes one ADP substrate molecule as *phosphorus electrophile* and one as *oxygen nucle*ophile, explicitly recognizing the versatility of the phosphate group in biological chemistry.

Given the generation and utilization of 75 kg of ATP per day per adult human, there will be a continuing pool of both AMP and ADP that have to be brought back up to the ATP level. We will note that for ADP most of the flux back to ATP is *via* action of transmembrane proton gradients to drive ATP synthase. However, AMP has to gain two side chain phosphoric anhydrides to get back to ATP. Adenylate kinase is the major catalytic agent for setting the first P–O–P anhydride bond as AMP is raised back to the ADP level.

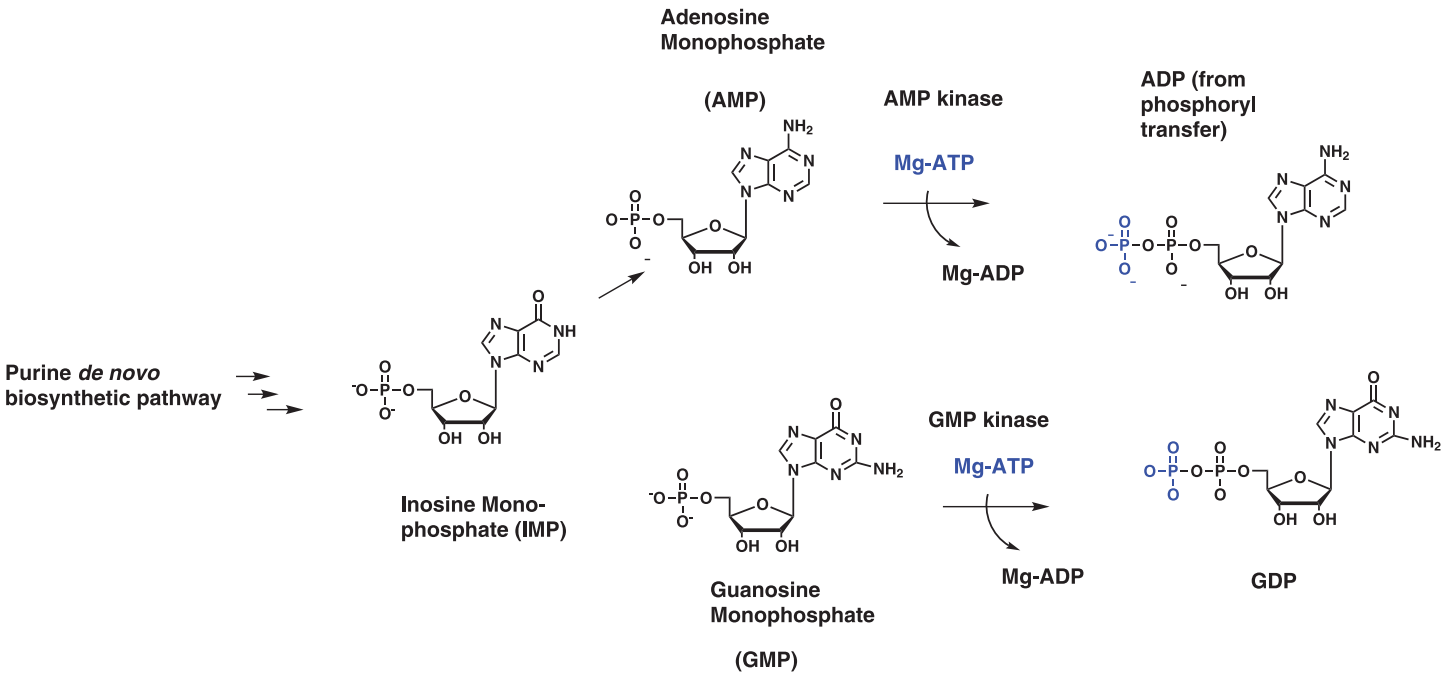
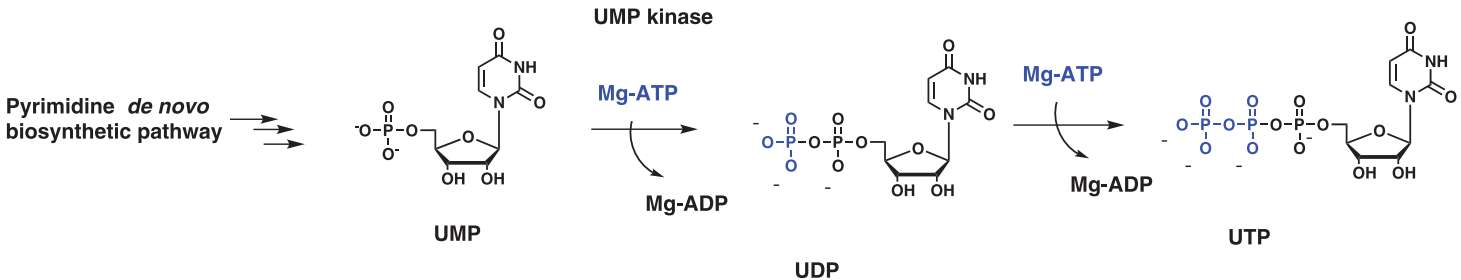


Figure 3.12 Kinases for interconverting nucleoside monophosphates, diphosphates and triphosphates. AMP kinase, GMP kinase, UMP kinase; Adenylate kinase.



Adenylate Kinase : may fuction in cells to retrieve AMP to ADP

AMP + ATP = ADP + ADP

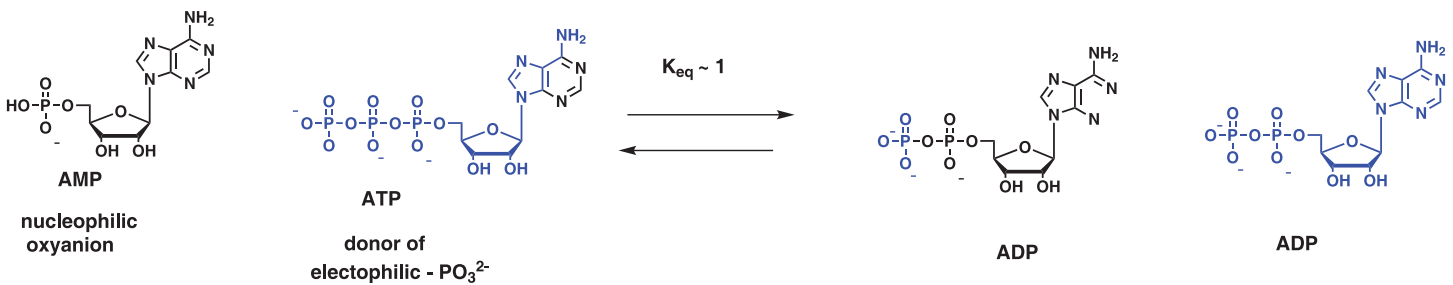


Figure 3.12 (Continued).

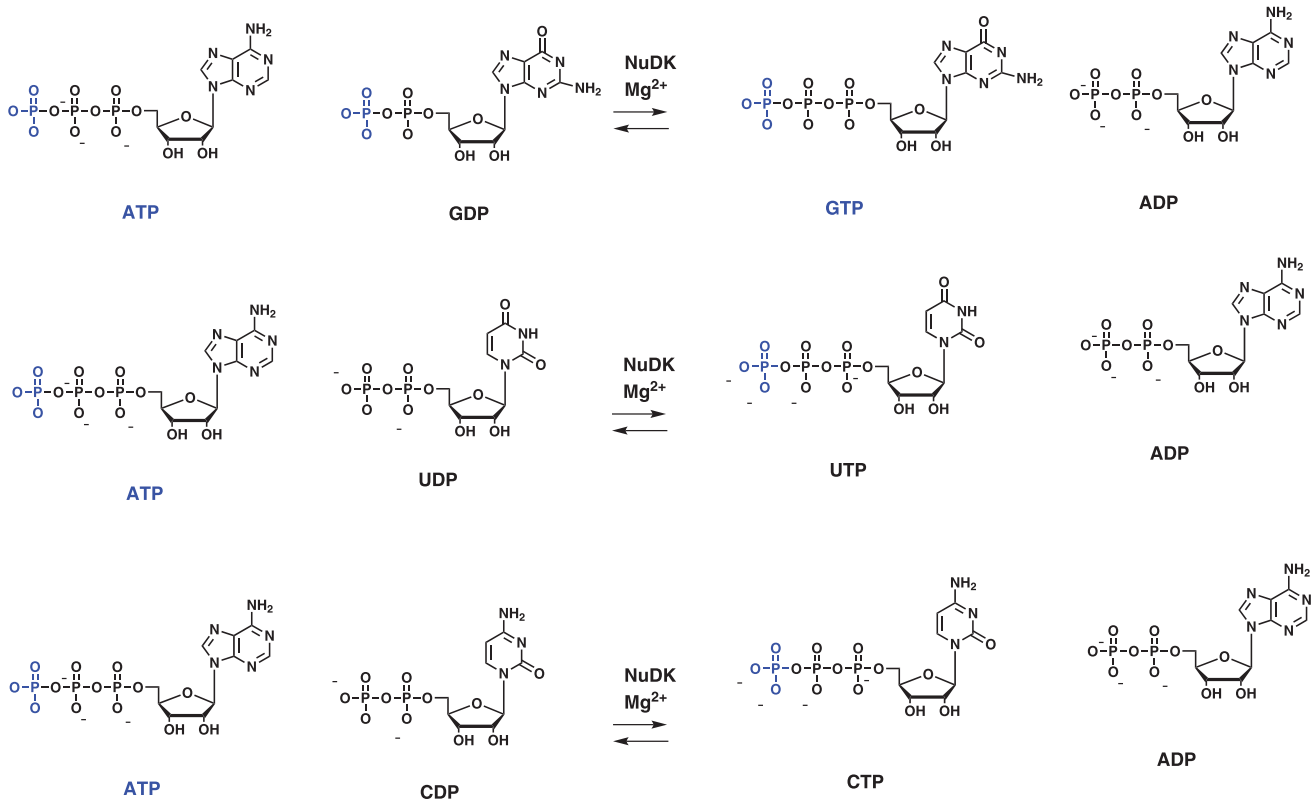


Figure 3.13 Kinases for interconverting nucleoside monophosphates, diphosphates and triphosphates. Nucleoside diphosphokinase acts on different NDPs as nucleophilic substrates.

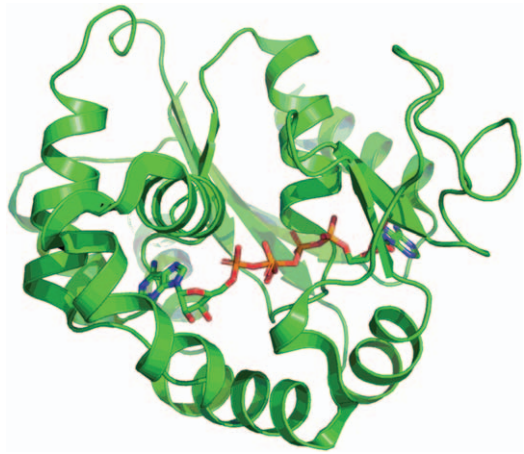


Figure 3.14 Kinases for interconverting nucleoside monophosphates, diphosphates and triphosphates. Structure of adenylate kinase with two molecules of bound ADP (PDB file 6HAM).

The F-type membrane-associated ATP synthase then converts ADP and Pi back to ATP.

A second enzyme in this category of phosphoryl transfer from ATP to a phosphate monoester is nucleoside diphosphokinase (*aka* nucleoside diphosphate kinase). Its generic stoichiometry is shown in Figure 3.13 in which the other three building blocks for RNA synthesis, CDP, UDP, and GDP, are the nucleophiles on ATP as phosphoryl donor.[15] The result is generation of CTP, UTP, and GTP, as three molecules of ATP are converted, one at a time, to three ADPs. All four nucleoside triphosphates need to be available, at billions of molecules per cell, for RNA polymerases to transcribe DNA chains into various species of RNAs (messenger RNAs, ribosomal RNAs, transfer RNAs, noncoding RNAs and other minor RNA species). Nucleoside diphosphokinase is a key enzyme in setting this inventory of NTPs. Figure 3.13 shows the outcome of three phosphoryl transfers: from ATP to GDP, from ATP to UDP, and from ATP to CDP.

3.5.4 Enolate Anions in Phosphoryl Transfers

An unusual category of oxygen nucleophile occurs during the reaction of pyruvate kinase.[16] This enzyme catalyzes one of two ATP-harvesting steps in glycolysis, the pathway breaking the six carbon aldohexose glucose to two molecules of the three-carbon α-ketoacid pyruvate. The $K_{eq.}$ lies far in the direction of pyruvate and ATP, ($K = 38\,900$ at pH 7.0), but may possibly be reversed in cardiac and skeletal muscle.[17] In that back direction, as shown in Figure 3.15, the attacking oxyanion is the enolate form of pyruvate. The enolate is generated on deprotonation of carbon 3 of pyruvate, with resonance contribution from the species with the C2 as ketone with the negative

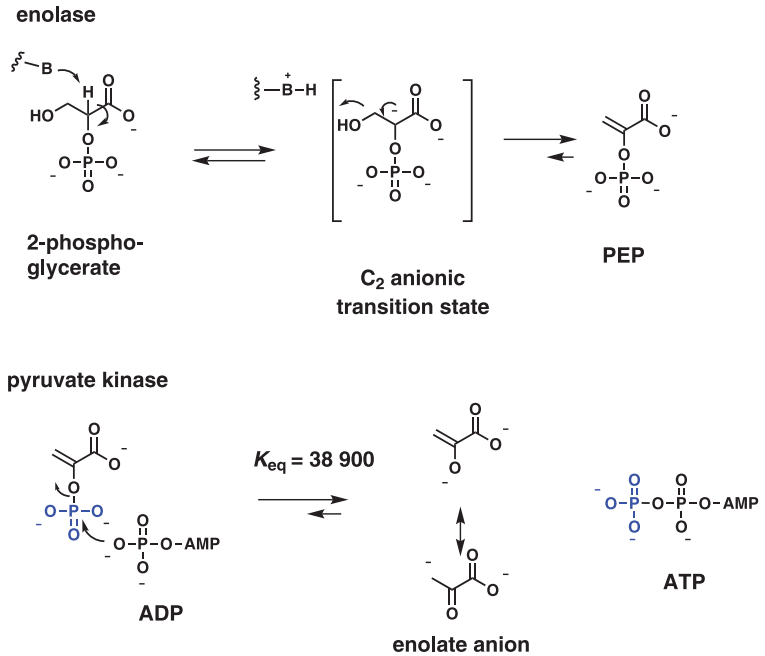


Figure 3.15 Tandem action of enolase and pyruvate kinase generates and utilizes the phosphorylated enol in PEP for phosphoryl group transfer. PEP has higher phosphoryl group transfer potential than ATP.

charge at C3. The product in the back direction is phosphoenolpyruvate (PEP), the substrate in the forward direction for phosphoryl transfer to ADP. The unfavorable $K_{eq.}$ for the back reaction indicates that the trapped enol phosphate, albeit kinetically stable, is substantially more activated thermodynamically as phosphoryl donor than ATP. The enol-phosphate group is more activated than the P–O–P phosphoric anhydride group.

The naming of the enzyme historically as pyruvate kinase reflects the equilibrium favoring the back reaction. However, neither of the two oxygen functionalities in the predominant, ground state form of pyruvate at physiological pH, the carboxylate and ketone oxygen, are reactants. It is the oxyanion of the very minor enolate form of pyruvate, available at very low concentrations free in solution.

Further indication that PEP is activated for phosphoryl group transfer occurs in the transmembrane bacterial phosphotransferase for hexose uptake into bacteria as will be explored in Chapter 14. Sugar passage across the membranes is accompanied by phosphorylation at the C6–OH (Figure 3.16) to trap the resultant anionic sugar phosphate inside. The phosphoryl donor is PEP but indirectly by way of a set of phosphorylated proteins, the first of which is the protein HPr, named for a histidine that gets phosphorylated.

The enzymatic strategy of formation of the highly energetic PEP during the glycolytic pathway is intriguing. The immediate precursor is

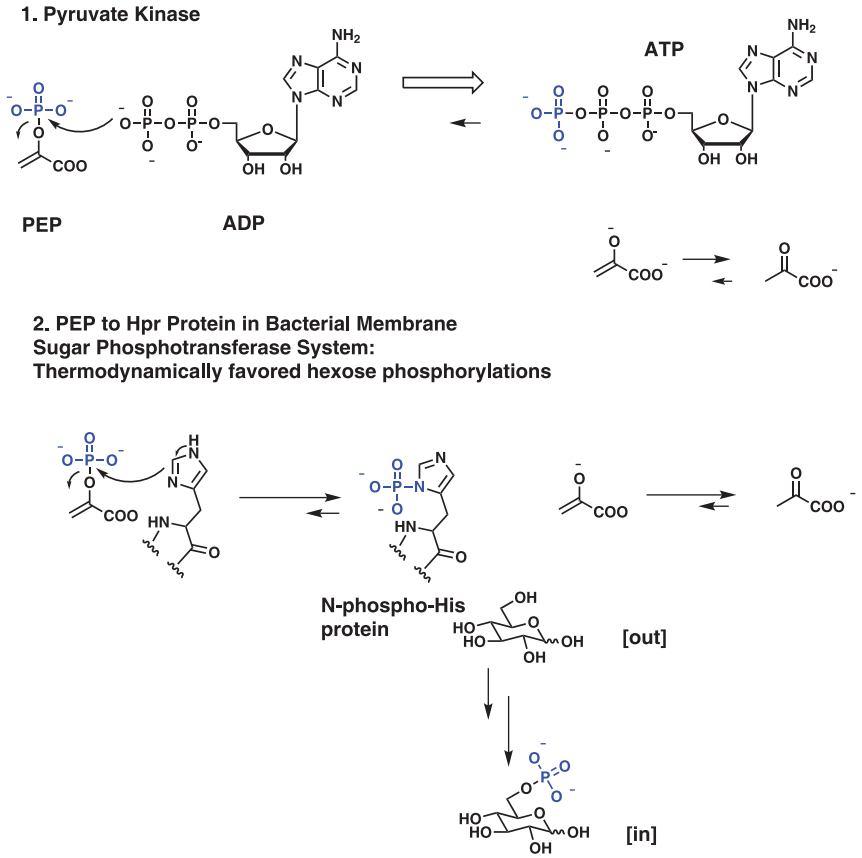


Figure 3.16 Two examples of enzymatic phosphoryl group transfer from PEP: in glycolysis and in bacterial transport and phosphorylation of hexoses.

2-phosphoglycerate. The enzyme enolase can abstract a C2-hydrogen as a proton to generate a carbanionic transition state. On elimination of the 3-OH substituent, water has been released and the phosphoenolpyruvate has emerged as product (Figure 3.15).

3.6 ATPases and Coupled Equilibria

3.6.1 ATPases: Phosphoryl Transfer to Water, Coupled to Cellular Work

A fifth category of oxygen nucleophile that enzymes use to attack the γ-phosphorus of Mg–ATP is water. The products are Mg–ADP and inorganic phosphate (Pi) (Figures 3.17 and 3.18). These phosphoryl transfers to water on initial consideration seem to have wasted the thermodynamic potential of ATP as a phosphoryl donor. Instead of driving some coupled chemical

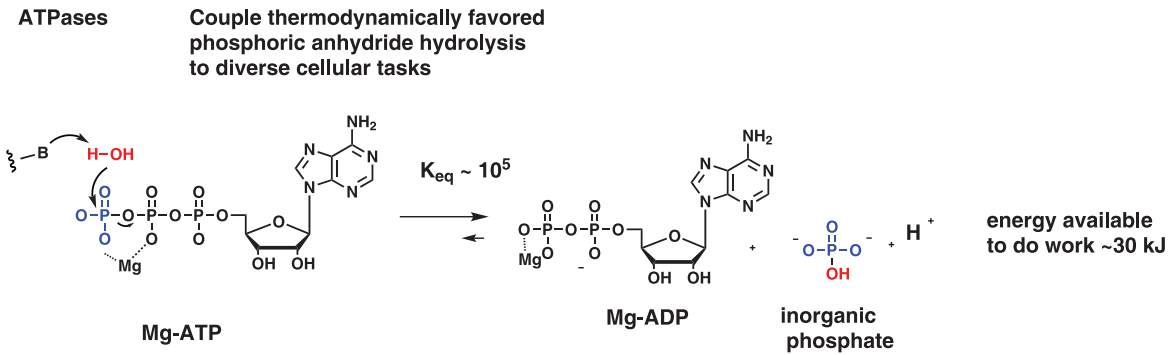


Figure 3.17 The broad range of cellular functions of enzymes that hydrolyze ATP (ATPases) and couple the energy released to perform distinct types of cellular work.

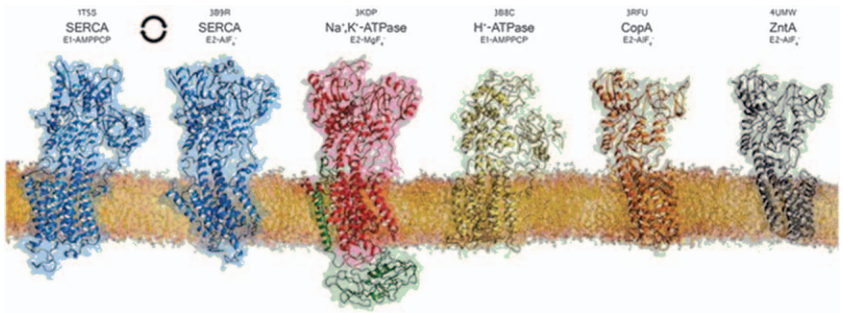


Figure 3.18 Collection of PDB structures of several transmembrane ATPases. Image courtesy of Dr Oleg Sitsel, used with permission.

equilibrium in the forward direction, the P–O–P anhydride linkage seems to have been dissipated unproductively.

There are close to 300 distinct ATPases in human proteomes[18] (https://en.wikipedia.org/wiki/HUGO_Gene_Nomenclature_Committee), indicating that this view of unproductive hydrolysis of ATP is too narrow. Indeed, many different types of cellular work are coupled to the hydrolysis of ATP by such ATPases. The $K_{eq.}$ for ATP hydrolytic cleavage to ADP and Pi of 10^5, corresponding to an energy yield of ~30 kJ per mole, is used to drive many cellular events crucial to life.

The coupled chemical work includes pumping of ions across both intracellular membrane compartments and across the plasma membrane to the outside microenvironment.[18a] Ion selectivity for such ATPase driven pumps includes H^+, Na^+, $K+$, Ca^{2+}, Mg^{2+}, and Zn^{2+} as well as many organic molecules. Metabolites pumped against their concentration gradients across various membranes include amino acids, sugars, organic acids, lipids including cholesterol, vitamins, peptides, oligosaccharides, and oligonucleotides. Internal cellular ATPase-coupled proton pumps include ones acidifying lysosomal compartments. Parietal cells lining the stomach and osteoclasts on the surface of bones both pump protons to the exterior of the cells: parietal cells to provide acid pH in the stomach, osteoclasts for dissolution of calcium phosphate in resorption and remodeling of bones. In some cells up to 30% of total ATP expended is utilized by such ion-pumping (P-type)-ATPases.

Although the ~300 ATPases have enormously diverse functions, they are united in converting the energy, released as ATP is hydrolyzed into protein conformations distinct from the resting or ground state, typically still containing one or both of the nascent products ADP or Pi. Alternatively, covalent phosphoenzyme intermediates in ion-pumping P-class ATPases (Chapter 14) can drive conformational change in the ATPase catalyst. The conformation change energized by ATP side chain cleavage can be transmitted to noncatalytic subunits and/or partner proteins. Such coupled equilibria transmit the conformational change information over substantial cellular distances as they drive a plethora of coupled processes.

The large ATPase family can be parsed into two major categories based on structural and mechanistic features. The first category can be termed transmembrane ATPases. These integral membrane proteins, often but not always consist of *many* subunits that span intracellular and/or cell surface membranes.[19] Typically, ATP hydrolysis is coupled to ion flow or organic metabolite flow across the membrane in which the ATPase is embedded. The ion or metabolite flow is usually against its concentration gradient and therefore driven by the energy available from ATP side chain fragmentation. We will note below multiple subcategories of the transmembrane ATPases.

The second category of ATPases are much more diverse structurally and functionally[20] and are termed *AAA*-proteins[21] where the three *A* letters stand for ATPase associated with cellular activities; a catchall phrase for diverse functions.[22] There are 239 *AAA*-proteins encoded in human genomes [*Human Genome Nomenclature Committee webpage: AAA ATPases* (*ATAD*)], based on bioinformatic predictions, indicating widespread evolutionary radiation to couple ATP hydrolytic energy to an enormous range of diverse cellular work at the molecular level. We take up the transmembrane ATPase types first and then the *AAA*-proteins, while noting at this outset that entire tomes can be written fruitfully on either ATPase category and those detailed narratives are beyond the scope of this volume.[23]

3.7 Transmembrane ATPases

Transmembrane ATPases have been further divided into four subcategories based on structure, mechanism and/or coupled pump function. We note (1) P-type ATPases,[18a] (2)[18b] V-type ATPases, (3) F-type ATPases (=ATP synthase),[24] and (4) ABC type ATPases[25] (Box 3.1).

Box 3.1 Four subtypes of transmembrane ATPases.

1. Transmembrane ATPases Function as Ion Transporters, Multiple Subunits, and Pump Cations

(a) F type ATPases: function in the ATP synthase direction driven by the flow of protons down their electrochemical gradients through the enzyme as transporter
(b) P type ATPases. *e.g.* membrane Na- or K-exchange across plasma membrane; Calcium pumps; parietal cell proton pumps acidify stomach contents
(c) V type ATPases acidify vacuoles and other endomembrane compartments; lysosomal proton pump; osteoclast proton pump secretes protons to solubilize and resorb bone
(d) ABC type transmembrane enzymes

3.7.1 P-type Transmembrane ATPases

The P-in P-type ATPases refers to a phosphoenzyme intermediate, unique to this enzyme subclass. The hydrolysis of ATP proceeds in two discrete steps. A water molecule does not participate in the first half reaction. Instead, the kinetically competent nucleophile is the β-carboxylate side chain of an aspartyl residue in the ATPase enzyme, yielding a covalent phosphoryl enzyme intermediate (see Chapter 14 for more context on the covalent aspartyl-P-enzyme intermediate). The covalent P-enzyme takes on a distinct conformation from resting enzyme, opening a transient channel among its 10 transmembrane helices to allow ion flow.

In the second half reaction, ATPase dephosphorylation, the aspartyl-P intermediate is attacked by water, releasing inorganic phosphate to go along with the ADP coproduct. The enzyme relaxes back to ground state and ion flow stops until the next catalytic cycle. Two well-studied P-type transmembrane ATPases in humans are the plasma membrane Na–K-ATPase and the sarcoplasmic reticulum muscle calcium ATPase (SERCA enzyme). The Na^+–K^+ ATPase pumps K^+ ions out while Na^+ ions flow in the counterbalancing direction. The calcium-ATPase in muscle is the agent for pumping Ca^{2+} ions back from cytoplasm to sarcoplasmic reticulum in muscle cells at the end of a contractile cycle.

The P-type ATPases are unique in using covalent catalysis, the intermediate acyl-phosphate *via* the active site aspartyl residue. That mixed anhydride is still thermodynamically activated and capable of imparting conformational energy. This category will be examined in Chapter 14, Section IV. Other ATPases induce such conformational information noncovalently, often *via* holding on tightly to the nascent inorganic phosphate product dianion.

3.7.2 V-type Transmembrane ATPases

The letter V-designation stands for vacuolar, reflecting extensive characterization of the vacuole-associated enzyme in yeast.[26] Proton pumping by V-ATPases can lower intravacuolar pH values to pH 1. These ATPases are generally specific for proton pumping and so are pH control agents. In higher eukaryotes the V-ATPases are found embedded in lysosomal membranes, primary endosomes, secretory vesicles, and, in cells such as osteoclasts, macrophages, and neutrophils, at the plasma membrane.

The V-ATPase in lysosomal membranes can acidify the lumen of lysosomes to ~pH 4.5 to permit enzymes with acidic pH optima to digest different categories of ingested substrates. In the secretory pathway for cargo export, the lumen of late Golgi vesicles can be pH 5, which fits the most active pH for proteolytic processing of preproteins to their cleaved mature forms for extracellular secretion.[27] Likewise, packaging of neurotransmtter amines into presynaptic vesicles of neurons is driven by the low pH created by V-type ATPases in those synaptic vesicles.

The yeast V-ATPase has been the prototype, with 13 subunits clustered in V1 and V0 assemblies.[26] The peripheral membrane-associated V1 domain

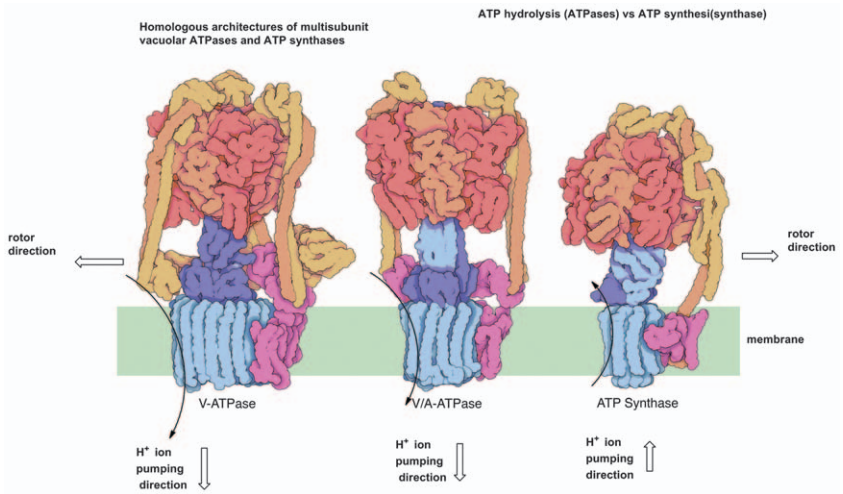


Figure 3.19 Similarity of V- and V/A-type transmembrane ATPases and F0/F1-ATP synthase, albeit with orthogonal directions of proton-pumping and ATP hydrolysis or synthesis.
Adapted from 10.2210/rcsb_pdb/mom_2005_12, courtesy of David Goodsell.

exhibits the ATPase activity while the V0 domain is the intrinsic transmembrane moiety that drives coupled proton translocation or extrusion across the membrane. The architecture of the V1 and V0 domains (Figure 3.19) and their coupling of ATP hydrolysis to proton translocation strongly resemble the F-type-ATPase described next. The ATPase domain acts as a stator and causes the membrane embedded V0 domain to act as a rotor, translocating protons. The roles of specific subunits have been studied by genetic manipulation. Stoichiometry of protons pumped per ATP hydrolyzed can vary but has been reported to be between two and four H^+ per ATP hydrolyzed for the yeast V-ATPase. Finally, lipid composition in the immediate membrane vicinity can control V-ATPase activity. Long chain sphingolipids and the minor lipid phosphatidylinositol-3,5-bisphosphate have been reported to be essential activators.

A cryoelectron microscopic analysis of the V/A-type ATPase from the thermophilic bacterium *Thermus thermophilus* has recently been reported.[28] It can function as ATPase or ATP synthase. Figure 3.19 shows separation of the proton pumping transmembrane subunits, in this case from periplasm to cytoplasm, from the rotor subunits that drive ATP synthesis.

3.7.3 F-type Transmembrane ATPase = ATP Synthase

The V-type and F-type ATPases are highly related in terms of domain architecture, membrane embedding, subunit composition and specificity for protons *versus* other cations (Figure 3.20). However, the F-type ATPase runs

in the opposite direction physiologically.[29] It is the ATP synthase nanomachine,[30] responsible for probably >95% of the ~70–75 kg of ATP made by adults each day.

Both the V-type ATPase and the F-type ATP synthase are described as rotary enzymes and the similarity of subunit composition and organization are described in Box 3.1.[31] The reaction of ATP with an H_2O molecule generates ADP, Pi and a proton (H^+). Proton concentrations at pH 7 are low enough for the equilibrium to proceed far in the direction of ATP hydrolysis for most ATPases. However, if the F-ATPase were to be located with the active site in a mitochondrial organellar membrane compartment with very high proton content the equilibrium would instead by pushed from right to left, in the ATP synthase direction.

The mitochondrial F-ATPase/ATP synthase is in such a microenvironment. In the inner membrane of mitochondria, as electrons flow down the electron transport chain from NADH to O_2, protons are pumped anisotropically at three stages (complexes I–III) along the chain from matrix to inner membrane space, acidifying it (Figure 3.21).

If the protons can run back down their electrochemical gradient they generate energy. The F-ATPase has two half-channels for protons. As protons reach the end of the first half-channel they protonate a glutamate reside, which then flips its conformation. This is proposed to occur in all ten of the F0 domain subunits. These sets of coordinated flips of glutamate anions to protonated glutamates constitute a clockwise rotor, opening the protonated glutamates to the half channel that will allow those glutamate-tethered protons to flow into the matrix space to a higher pH (lower $[H^+]$).

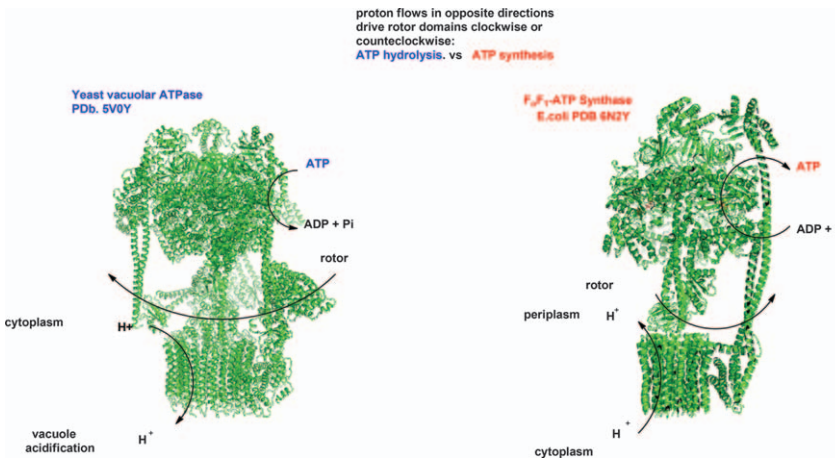


Figure 3.20 Similarity in subunit composition and architecture of transmembrane V-type ATPases and F-type ATP synthases. The yeast vacuolar ATPase is drawn from PDB 5V0Y and the *E. coli* ATP synthase from PDB 6N2Y.

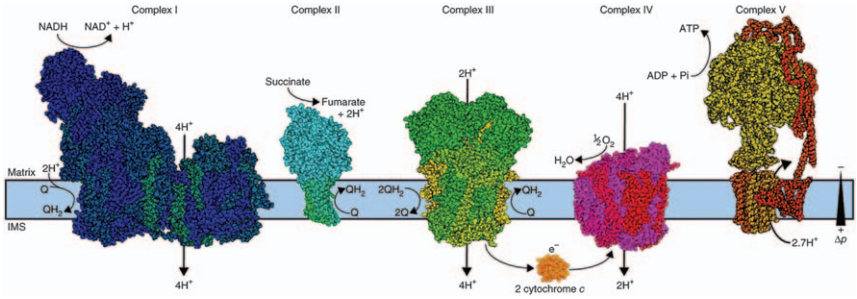


Figure 3.21 Electron flow down respiratory membranes in bacteria, and mitochondria are coordinated with anisotropic pumping of protons at three sites of protein complexes I, III, and IV.[32] In mitochondria, protons are pumped into the intermembrane space and accumulate to high enough concentrations that they flow back down their electrochemical gradient in the active site of the transmembrane F-type ATP synthase. That electron flow drives ATP synthesis. The figure schematizes the size and orientation of the proton pumping sites in the membrane, including the ATP synthase.
Reproduced from ref. 57 with permission from Springer Nature, Copyright 2017.

The clockwise motions of the membrane-embedded rotor subunits are transmitted to the F1 domain containing the nucleotide recognition domains. The bound ADP and inorganic phosphate molecules must be compressed closely enough to form the P–O–P anhydride bond in nascent ATP, the reverse of the ATPase reaction, splitting off a water molecule as the P–O–P anhydride forms (Figures 3.22 and 3.23) (https://pdb101.rcsb.org). The stoichiometry of F-ATPase acting as ATP synthases has been observed to be one ATP synthesized for every batch of three to four protons that flow down their electrochemical gradient, depending on whether bacterial membrane ATP synthase, chloroplast ATP synthase, or mitochondrial ATP synthase stoichiometries have been examined.[33]

Recently, Zhou and Sazanov[28] have succeeded in obtaining an atomic structure of the F-type ATP synthase from the cytoplasmic membrane of the thermophilic bacterium *Thermus thermophilus* at 3.5 angstroms resolution by cryoelectron microscopy (Figure 3.22). Readers can consult this primary literature to examine the detailed path of protons from periplasm to cytoplasm through the periplasmic and cytoplasmic half channels of the transmembrane subunits. Also, Figure 3.22 shows that the ATP synthase rotor portion is above the plane of the membrane, mechanically coupled to conformational changes driven by proton flow through the membrane subunits to the rotor domains.

Given that proton gradient-driven, transmembrane ATP synthases account for more than 90% of ATP generated by aerobic organisms (~70–80 kg per day), one could argue that the F-ATP synthase occupies a top place in the pantheon of life's crucial enzyme catalysts, along with DNA and RNA polymerases, and ribonucleotide reductase.

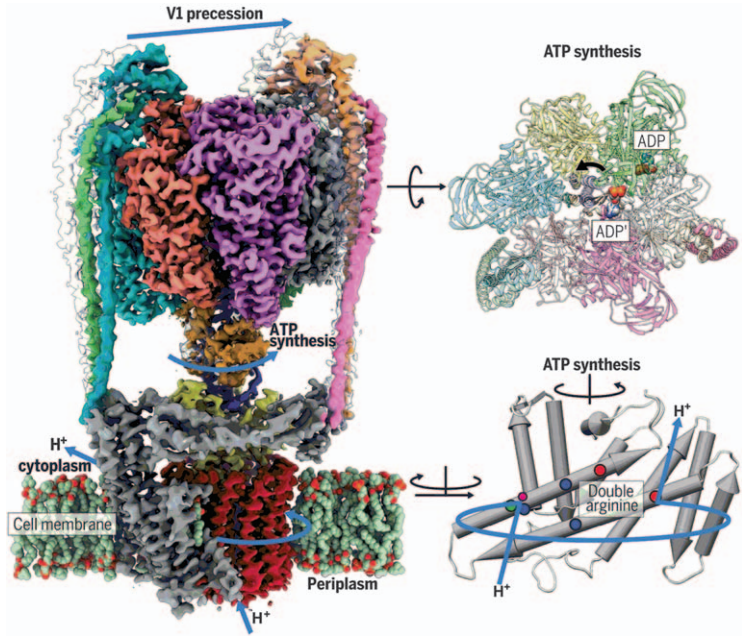


Figure 3.22 Cryoelectron microscopic structure of *T. thermophilus* F/A-type ATP synthase at 3.5 Å resolution.
Reproduced from ref. 28 with permission from AAAS, Copyright © 2019 The Authors.

3.7.4 ABC Type Transmembrane ATPases

The ABC nomenclature in this fourth family of transmembrane ATPases stands for ATP Binding Cassette. These are much simpler architecturally than the V- and F-type ATPase nanomachines.[25] Typically, there are two nucleotide triphosphate binding domains (NBD) and two transmembrane (TM) domains. On ATP binding to each NBD, they dimerize and affect the orientation of the transmembrane domains. In the activated, loaded ATPase, the transmembrane region has 12–20 helices which reorient to form a transient pore on hydrolysis of bound ATP[35] (Figure 3.24).

This form of ATPase is quite common in bacteria that mediate inside and outside milieus *via* expression of many solute-specific ABCs that can act either as import or export pumps. In humans there are some 48 predicted ABC type ATPases and transporters, with partner proteins that are thought to exhibit specificity for recognition of specific solutes to be transported.[37] As the ABC transporters have radiated and gained differential solute recognition, they have become categorized into seven subgroups. Among the molecules transported by individual ABC machinery across cell external or internal membranes are sterols, other lipids, retinoids, nucleosides, and various peptides.

The left-hand panel of Figure 3.24 shows the two component ModBC, an ATPase that functions to import sugars, amino acids and ions into bacteria.

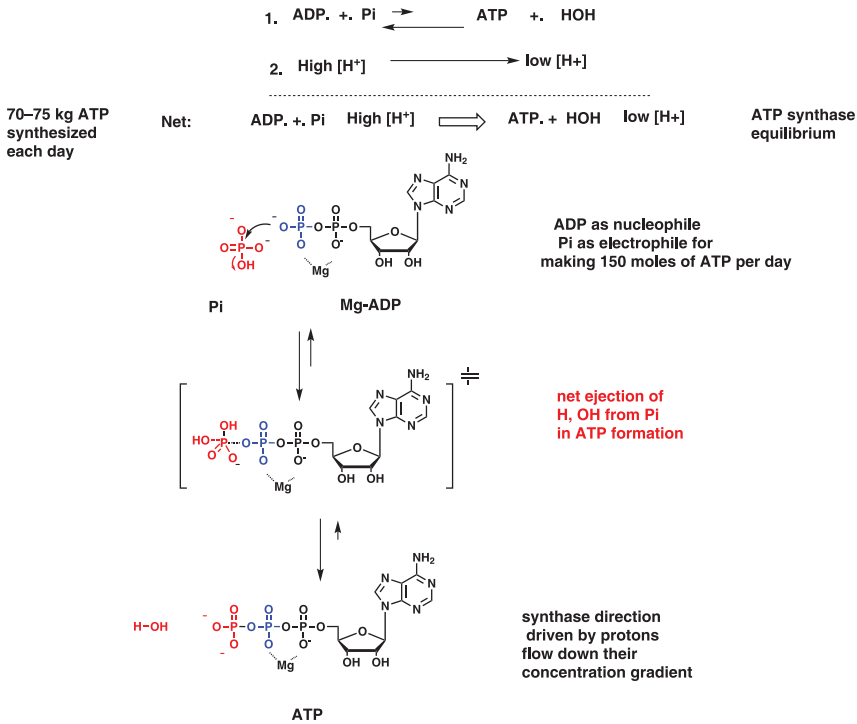


Figure 3.23 In ATP synthase active sites, ADP is the nucleophile at oxygen, inorganic phosphate is the electrophile at phosphorus. The favorable thermodynamics of proton flow through the ATPase ion channel drive the otherwise unfavorable condensation of ADP with Pi to form ATP.

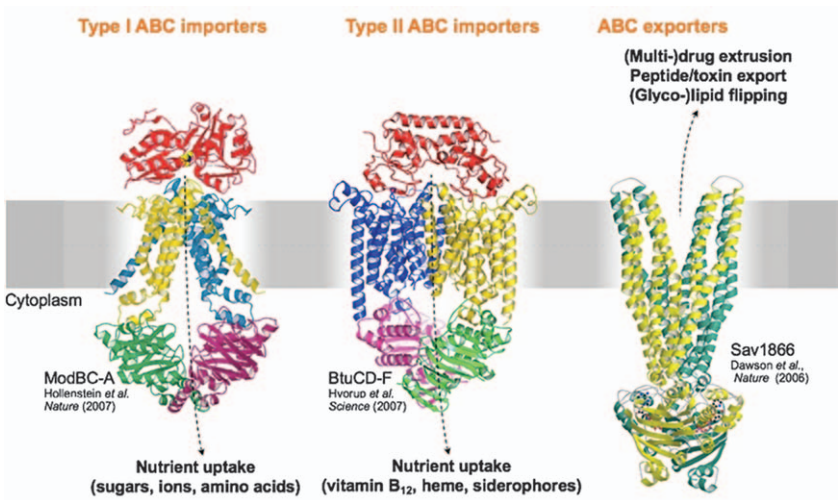


Figure 3.24 Three representative bacterial ABC transporter systems: two functioning as importers, one as an exporter.[36]
Adapted from ref. 36 with permission from Springer Nature, Copyright 2016.

The middle panel depicts the architecture of a type II importer in this superfamily, specifically for uptake of porphyrinoid macrocycles such as vitamin B12 and heme. Other members of the ABC ATPase transmembrane family act as export rather than import machineries. The Sav1866 protein system functions to export peptide toxins and also to flip glycosylated peptides from inner leaflet to outer leaflet during cell wall assembly.[34,35]

The cystic fibrosis transmembrane conductance regulator (CFTR) protein pump in humans falls into a variant subcategory where only one of the two nucleotide binding domains seems to be functional.[38]

The study of the structural biology of ABC transporters has progressed greatly in the past two decades and given functional and mechanistic insights into inward facing conformations *vs.* outward facing conformations of the ABC systems. ATP hydrolysis timing and the subsequent lifetime of bound nascent ADP and Pi products are proposed to control transmembrane conformations and active passage of specific partner solutes.

3.8 AAA Proteins: A Plethora of ATPase Domains Coupled to Diverse Functions

The second major category, *AAA* proteins, with 239 predicted ATPase members in the human proteome, are more diverse in structure and certainly in function. As the ATPases associated with cellular activities (*AAA*) name implies it is not readily possible to categorize these ATPase domain-containing protein machineries by some common functions. One author has written: *"The defining feature of AAA+ proteins is a structurally conserved ATP binding module that oligomerizes into active arrays" (???).* That begs the question of what constitutes active arrays. Nonetheless, the ATPases domains of *AAA* proteins are all thought to generate energy to drive conformation changes and/or remodeling within the *AAA* protein and onto the designated cellular target partner. The timing of ATP side chain hydrolysis, the life-time of protein conformations with bound ADP and/or Pi, are thought to enable allow ATPase domains within *AAA* proteins to act as molecular switches. Rhythmic hydrolysis of ATP in molecular *AAA* motors allows for quantile advancement of these processive motors, often with their associated protein cargoes.

Often the ATPase domain is N-terminal to one or more additional domains that specify recognition with a given cellular target. For comparison to the 239 predicted *AAA* proteins in humans, 364 are predicted in the plant *Arabidopsis thaliana*, but only 7 in *Escherichia coli*, perhaps indicating radiation of *AAA* proteins during eukaryotic evolution. The large AAA family can be subdivided structurally based on insertion of secondary structure elements at loci around the core fold,[39] allowing for diversification of coupled functions. Multimeric forms of *AAA* proteins, *e.g.* DNA helicases, can communicate conformational changes originating in one subunit, throughout the hexameric array to operate on cellular targets.

Figure 3.25 illustrates distinct subcellular locations for ATPases, including as essential components in organelle biogenesis and functions, for endoplasmic reticulum, Golgi and nuclear envelope formation during cell cycles, meiosis, mitosis, apoptosis, and underlying microtubule dynamics. Cellular motors are powered by ATPase or cognate GTPase hydrolysis. The range of coupled activities encompass informational macromolecule activity from DNA replication, recombination, and repair to transcriptional regulation and ribosomal steps in protein biogenesis. This cellular sketch begins to give granularity to why humans make and consume their body weight in ATP every day.

It is beyond the scope of this section to list the universe of known *AAA* functions but Figure 3.25 displays a cellular map with localization and function of 14 different ATPases. Some of these are ion pumps for mitochondria, lysosomes, endosomal compartments, and peroxisomes. Other ATPases show classic motor functions, from myosin in muscle contractions to kinesins mediating aspects of microtubule dynamics in chromosome condensation during meiosis and mitosis. Some carry cargo proteins into and through secretory compartments and/or from neuronal cell bodies into axon and dendrite termini. Some mediate steps in apoptotic cell death. Others are involved in DNA replication, recombination and repair, such as the ATPase activity of DNA helicases. ATP-dependent proteolysis *via* proteasome action involves six ATPases in the mutisubunit proteasomal cap that spend up to hundreds of ATPs per protein to unwind the folded

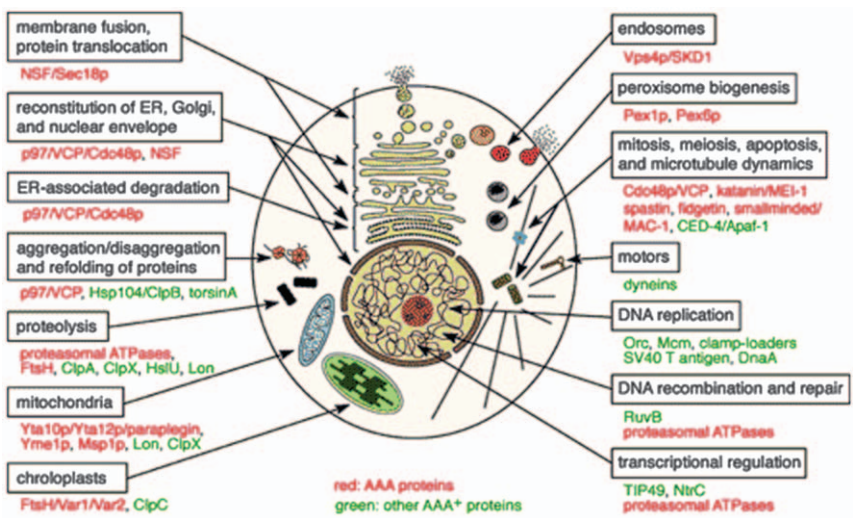


Figure 3.25 List of intracellular location and function of some of the >200 members of the AAA ATPase families.[40]
Reproduced from ref. 20 with permission from John Wiley and Sons, © The Molecular Biology Society of Japan/John Wiley & Sons Australia, Ltd.

conformations into unstructured regions that are then threaded into the protease chambers of the proteasomes. Other *AAA* proteins act as disaggregators of protein aggregates.

The two large groups of ATPases, transmembrane pumps of ions and organic metabolites *vs.* multimeric *AAA*-proteins with ATPase domains coupled to specific effector domains, total more than 300 members in the human proteome. They provide explicit examples of how the thermodynamic activation for phosphoryl group transfer to a cosubstrate water molecule has been harnessed in this enzyme class. The released energy, in the form of conformational equilibria, is used to drive an enormous range of events in cells, many of them mechanical.

3.9 Beyond ATPases: GTPases as Conditional Hydrolases Coupled to Multiple Forms of Cellular Work

The hundreds of ATPase enzyme-directed attacks of water on the terminal Pγ of ATP in thermodynamically favored hydrolysis coupled to driving some protein conformation change are complemented by families of dedicated GTPases. There are few if any comparable UTPases or CTPases that drive cellular work. Notably, organisms have evolved protein coupling logic for both of the purine nucleoside triphosphates ATP and GTP. In eukaryotic cells, where [ATP] ~3 mM, GTP is present at ~0.5 mM concentrations, while [CTP] and [UTP] are somewhat less than [GTP].

This GTPase subject is itself a large universe. Multiple volumes on distinct subclasses of GTPases and their biological roles have been written, particularly on the heterotrimeric G proteins,[41] the small G protein family headlined by the Ras GTPase[42] and microtubules.[43] Here we focus only on protein conformational lifetimes and slow GTP hydrolysis cycles.

Among GTPases there are different classes of proteins and enzymes that act in distinct cellular modes (Figure 3.26). They all exhibit conditionality in the hydrolysis of GTP such that uncoupled GTPase rates are low to negligible.

3.9.1 Heterotrimeric G Proteins

The largest mammalian family may be the thousand or so G protein-coupled membrane receptors (GPCRs), typically seven transmembrane helical proteins that serve as receptors for small molecule ligands, peptides, and proteins. The G in the term G protein stands for GTPase.[41] The G proteins that serve as transducers of ligand occupancy of the GPCRs on the external face of those receptors are bound on the cytoplasmic side of those receptors. They signal to cytoplasmic and ultimately nuclear responses as heterotrimeric G proteins. The heterotrimeric GTPases typically have one of 18 related Gα

catalytic subunits, combined with different members of Gβ and Gγ subunits that function as a hetrodimeric βγ complex.[44] Thus, many Gαβγ heterotrimeric complexes have varying subunit composition, receptor specificity, and activating or inhibitory signaling function from a released Gα–GTP species (see later in this chapter) to offer combinatorial versatility to signaling regimes.

In the absence of ligand occupancy of a GPCR, the cytoplasmically-anchored Gαβγ trimer is associated as a tightly bound GDP complex. On ligand binding to the external face of the GPCR, the transmembrane helices transmit that information and a cytoplasmic guanine exchange factor partner protein alters Gα conformation to pry out or release GDP. The Gα subunit then binds GTP and that Gα–GTP complex typically dissociates from the βγ heterodimer. The freed up βγ and Gγ–GTP subunits can each act on partners to initiate separate, parallel, intersecting cytoplasmic signaling chains of information. The Gα–GTP subunit will eventually (seconds to minutes) bind a water molecule productively in the active site and act as a GTPase, releasing Pi and reverting to the inactive Gα–GDP state that can recombine with the βγ dimer and go back to the hetrotrimeric Gαβγ–GDP inactive form.

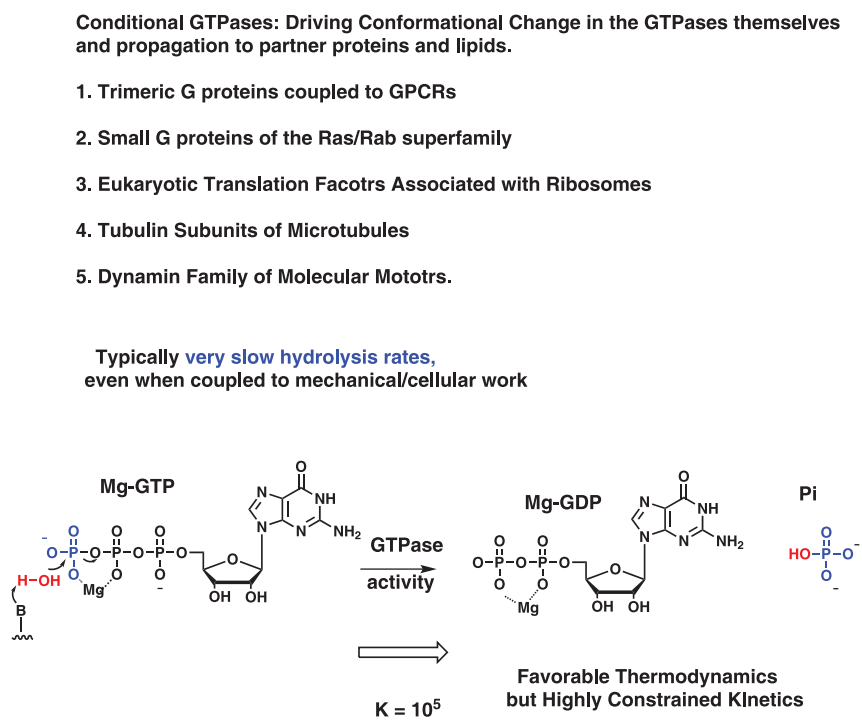


Figure 3.26 Five major families of conditional GTPases where slow catalytic turnover and tight binding of both substrate GTP and product GDP enable conformational coupling to drive mechanical cellular events.

The GTPase activity, conversion of bound GTP to bound GDP, alters up to three 'switch' regions in the GTPase Gα subunit. It is these distinct guanine nucleotide-bound enzyme conformers that interact differentially with partner proteins that control signaling of membrane receptor occupancy.

These are conditional enzymes: hydrolysis is slowed to allow Gα–GTP complexes to exist for a sustained period to signal. From the point of view of GTP hydrolysis the GTPases are miserably inefficient catalysts but their purpose is exactly that. These GTPases are not about indiscriminate hydrolytic consumption of thermodynamically activated GTP molecules, but rather getting the most out of every one of them when bound and creating an active signaling protein form for the lifetime of the GTP complex (before deactivating hydrolysis to bound GDP).

Note that unlike essentially all other enzyme catalysts, the product GDP is not released. It stays bound tightly. Among other effects, neither the heterotrimeric GTPase nor the released Gα subunit turn over and deplete cellular GTP stores by unwanted uncoupled hydrolysis.

3.9.2 Rab, Rac, Ran, Rho, Ras Family of Small GTPases

This tight binding, slow catalytic turnover logic holds for a second family of ''conditional'' GTPases, the so called ''small'' GTPases [low molecular weight (20–25 kDa) single subunits] families of GTPases. There are some 35 members of the RAS GTPase family. A broader set of ~160 members of the RAS-related superfamily of GTPases have also been catalogued.[45] The biology of these ~200 GTP hydrolases[46] defines a large swath of intracellular signal transduction biology. The superfamily is typically divided into subfamilies: Rab, Rac, Rho, Ras, Arf, and Ran, based on coupling proteins and roles in cell biology (Figure 3.27).

Many of them, such as the Rab family members are effectors for vesicle and cargo transport through endoplasmic reticulum and Golgi organelles in intracellular secretory membranes. Figure 3.28 depicts specialized roles in vesicle packaging and fusion events at synapses. They have been termed ''mechano-chemical enzymes''.[50] Again, these are extraordinarily slow enzymes (0.01–1.0 GTPase catalytic events per minute are the typical range of hydrolysis), consistent with catalysis as secondary to lifetimes of the active GTP bound forms to perform their coupled cellular work. As tightly bound GTP is hydrolyzed to tightly bound GDP, the signaling function is typically switched off as the small GTPase protein reverts to a ground state inactive conformation. Dramatic conformational changes at the switch regions of KRAS for the Ras family of GTPases[49] are shown in Figure 3.29 and also in Figures 3.30 and 3.31 for dynamin–GTP *vs.* dynamin–GDP.

3.9.3 Ribosome Translation Factors

A third family of GTPases function in ribosome-mediated protein synthesis where GTP rather than ATP appears to be spent as a gating function for

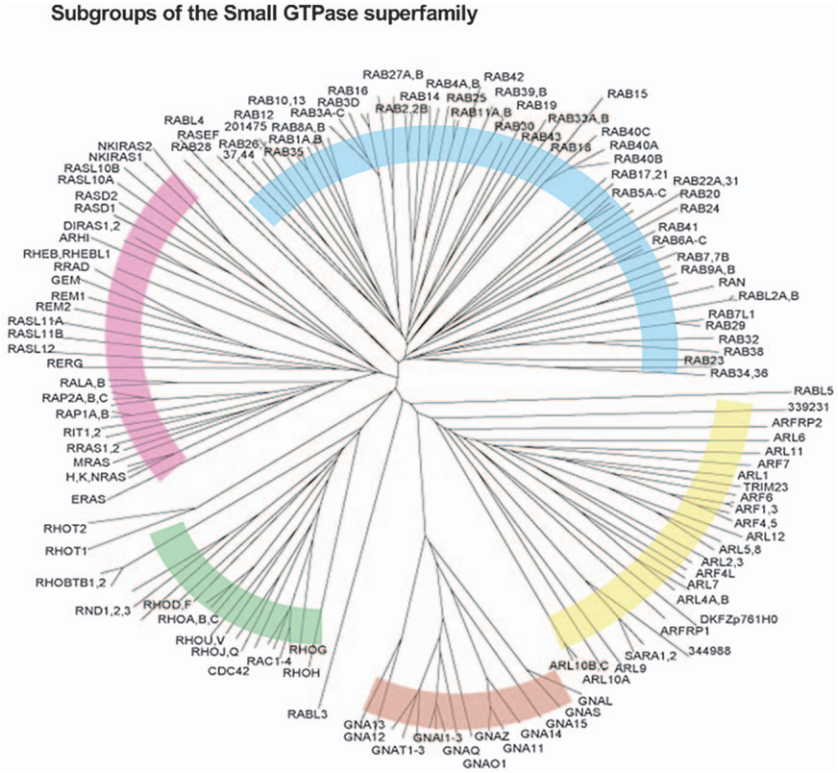


Figure 3.27 Circular diagram of small GTPase family members.
Reproduced from ref. 58 with permission from AAAS, Copyright 2004.

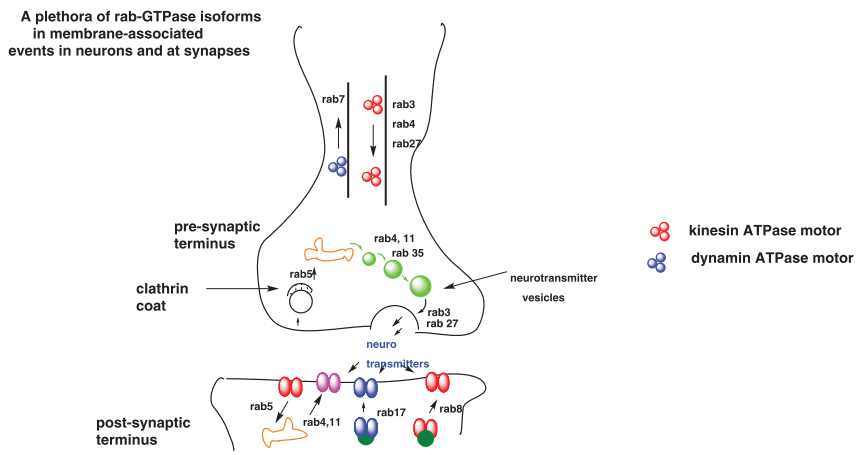


Figure 3.28 Subcellular distribution of Rab GTPase isoforms at pre and postsynaptic neuron termini.
Adapted from ref. 47, https://doi.org/10.3390/cells5010007, under the terms of a CC BY 4.0 license, https://creativecommons.org/licenses/by/4.0/.

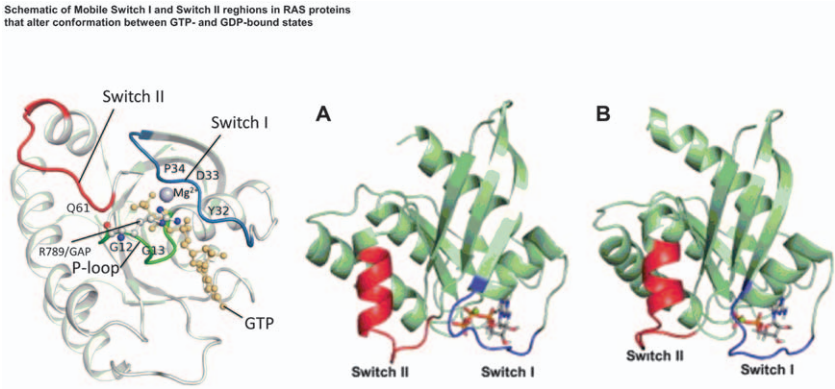


Figure 3.29 Schematic of mobile switch I and switch II in regions in GTP- *vs.* GDP-bound forms of Ras GTPase. Left hand side shows the two switch regions.[48] Right hand side shows the switch region conformations in (A) GTP-Ras and (B) GDP-Ras.[49]
Left reproduced from ref. 48, https://doi.org/10.1371/journal.pone.0055793, under the term of a CC BY 4.0 license, https://creativecommons.org/licenses/by/4.0/. Right reproduced from ref. 49 with permission from Elsevier, Copyright 2010.

specificity and movement of aminoacyl tRNAs as they approach the peptidyl transferase center. One such GTPase is a chaperone for aminoacylated tRNAs and GTP is only hydrolyzed when there is a correct codon–anticodon triplet base match. A second GTPase acts to use the energy of GTP hydrolysis to help move peptidyl tRNAs from the aminoacyl (A) site to the peptidyl (P) site in each round of peptide chain growth (at rates of $\sim$10 s^{-1}).[51] These eukaryotic protein translation factors are classic examples of coupling side chain phosphoric anhydride hydrolysis to drive otherwise unfavorable protein or nucleic acid conformational equilibria. In these two cases GTP, not ATP, is spent.

3.9.4 Tubulins

The two main types of cytoskeletal filaments are actin and microtubules, comprised of monomers (G-actin *vs.* α- and β-tubulin subunits) that can undergo polymerization in to filaments (F-actin or microtubules).[43] As in so many other areas of biology, the cytoskeleton literature is immense and largely beyond the scope of this phosphorus-centered inquiry. We focus here only on the conditional GTPase activity of the 50 kDa tubulin subunit isomers. Microtubules are famously dynamic scaffolding elements that periodically grow, shrink (in events termed catastrophes) and then regrow again (with time periods that may average a minute per cycle). The microtubule filaments are molecular highways to move cargoes to different parts of cells and to aid in chromosome segregations during cell division. Microtubular filaments are typically hollow cylinders with 13 protofilaments (although the number can vary).

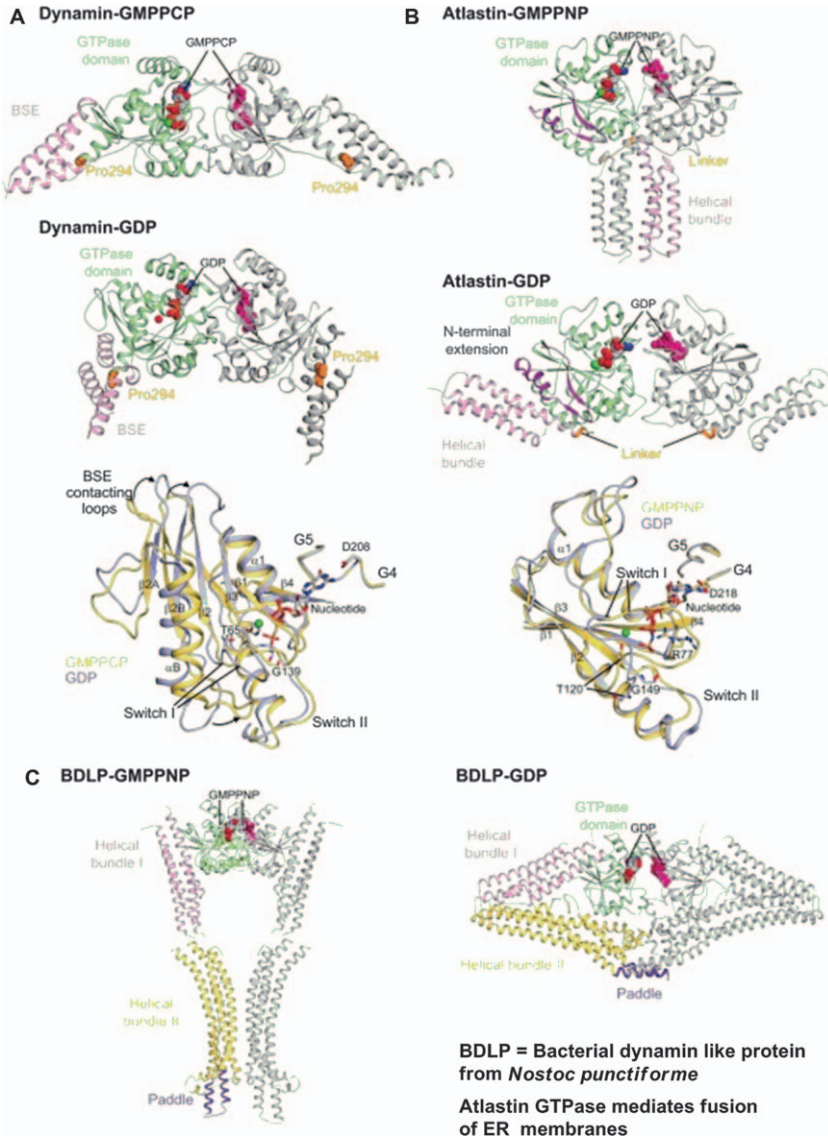


Figure 3.30 Structures of three dynamin family members, dynamin, atlastin, and bacterial dynamin like protein (BDLP), show dramatic architectural changes between GTP-bound and GDP-bound states. These changes are consistent with use of the GTPase-driven conformational changes to couple to cellular mechanical work of the specific dynamins.

Reproduced from ref. 50 with permission from John Wiley and Sons, © 2016 The Authors. Biopolymers Published by Wiley Periodicals, Inc.

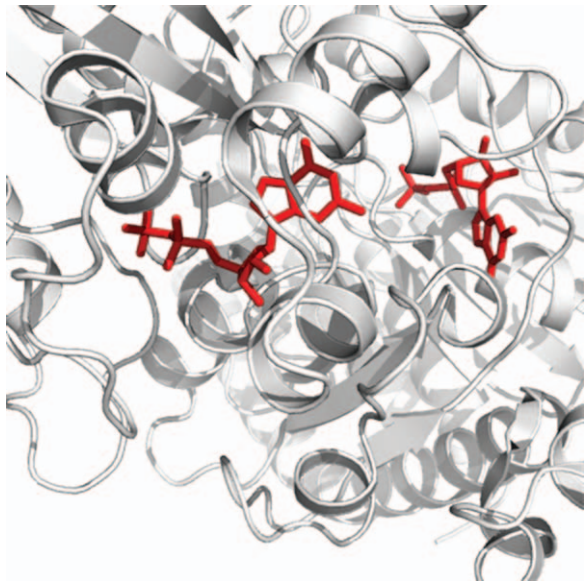


Figure 3.31 Conformation of bound GDP after the hydrolysis step in the active site of dynamin (from PDB 5D3Q).

Consistent with many of the other GTPase families, the tubulin αβ dimer can bind GTP tightly at both α- and β-subunits and present a conformational envelope that is different from the GDP state. In fact, formation of the dimer by protein chaperones[52] stimulate the almost negligible GTPase activity of the free nucleotide–tubulin subunit complexes. Generally, the GTPase activity of the tubulin dimer is only manifested after binding to growing microtubule filaments, thereby contributing to microtubule instability,[53] and results in *only one turnover* of bound GTP to bound GDP, with attendant shape change in the tubulin subunits. In this stoichiometric limit tubulins strain the definition of proteins and enzymes with catalytic activity. The logic is clearly about altered protein conformation.

About 100 tubulin–GTP monomers per second are added to the plus end of elongating microtubules. The growth and catastrophe events are conditioned by ancillary proteins but also by the rates of hydrolysis of bound GTP to GDP. In depolymerization mode, over the 13 protofilaments, up to 1000 monomers per second can be released as tubulin-GDP monomers.

As noted, from the point of view of stoichiometric rather than catalytic hydrolysis of bound GTP, tubulin subunits are at the low end of protein catalysts. The hydrolysis of only one GTP per subunit shows that the high group transfer potential of GTP as a cellular energy source is not being wasted and the logic is all about driving stoichiometric shape change in the protein, rather than multiple turnovers of the guanosine triphosphate, perhaps the clearest biological case of coupling a nucleoside triphosphate favorable hydrolysis to cellular work.

3.9.5 Dynamins

The last category of GTPases we note in passing are the dynamins, GTPases that function in the endocytic import of cargo into cells, in cytokinesis, and in organelle division in cells. Dynamins are ''large'' GTPases, in the ~100 kDa size range. In endocytosis dynamin molecules form constrictive rings, during GTP hydrolysis, around membrane invaginations that result in closure to endocytic vesicles.[50] In this aspect they resemble the function of Rab small GTPases that drive membrane fusion and fission events. Basal rates of GTPase activity in dynamins can be 100-fold up from the basal level of small GTPases, in the range of one to five hydrolysis events per minute. Membrane templates can drive the GTPase rate one to two orders of magnitude higher so while the dynamins are also conditional, activatable GTPases, they run at much higher ''burn rates''. They do not require dedicated ancillary guanine exchange factors or GTPase activating partners, as do most of the small GTPases.[54] Dynamins spend GTP to move as cargo-pulling molecular motors towards the plus ends of microtubules. An orthogonal set of microtubule motors are kinesins, dragging cargoes towards the minus direction, spending ATP rather than GTP, balancing ATPase- *vs.* GTPase-driven molecular motors.

Figure 3.30 shows the structures of dynamin in complex with a non- hydrolyzable GTP analog {guanosine 5′-[β,γ-imido]triphosphate trisodium salt hydrate (GMPPNP)} compared with a GDP complex, with dramatic conformational change, extending outside the GTPase domain, notably around proline 294. Figure 3.30 also shows similar dramatic architectural change between GMPPNP- and GDP-bound forms of atlastin, a dynamin family protein involved in membrane fusions in the endoplasmic reticulum. Panel C shows the dramatic conformational changes reach all the way back in evolution to a bacterial dynamin like protein (BDLP) from the cyanobacterium *Nostoc punctiforme*, thought to be descendants of the primordial photosynthetic microbes. Figure 3.31 shows a pair of bound product GDP molecules in two dynamin dimers after GTP hydrolysis and before dissociation.

3.9.6 GTPase Summary

The GTPase families, perhaps even more than the diaspora of cellular ATPases, illustrate a key tenet of integrative chemical and biological logic. The energy-consuming hydrolysis of the terminal phosphoric anhydride linkage as the γ-PO_3^{2-} group is transferred to water is all about driving protein conformations that are read and interpreted by cellular partners for setting in motion distinct forms of chemical and mechanical work in the cell economy. The coupled motions include protein–protein interactions (Ras), RNA–protein interactions (ribosomal translation factors), moving cellular cargoes (tubulins, dynamins), and providing directionality to membrane fusion events (Rabs).

3.10 Mg–ATP Bidentate Coordination Complexes are the Immediate Forms of ATP, GTP, and other NTP Substrates

A central feature of all ATP-dependent enzymes (and also the GTPases alluded to above) discussed in this and subsequent chapters is the use of Mg^{2+} (occasionally replaced by Mn^{2+}) to act as divalent counterion to the tetra-anionic triphosphate side chain of ATP. This partial, site-specific charge neutralization is crucial for effective dissipation of the negative cloud around any of the three α-, β-, and γ-phosphorus atoms at the center of the anionic phosphate groups. The dissociation of chelated Mg^{++} occurs rapidly in solution, at about 10^4 s^{-1}, so a mix of bidentate Mg–ATP coordination isomers are populated. Four such are shown in Figure 3.32. Presumably, one conformer can be selected for binding to a particular enzyme active site and may be immobilized there as part of the regiospecificity factors determining phosphoryl-, pyrophosphoryl-, or nucleotidyl transfer to appropriately placed cosubstrate nucleophile. Figure 3.33 depicts the position of the attacking water molecule relative to Pγ of the ATP triphosphate side chain in the active site of myosin as it prepares to act as an ATP hydrolase (ATPase). In Figure 3.34 the βγ-Mg–ATP coordination isomer is shown with the additional

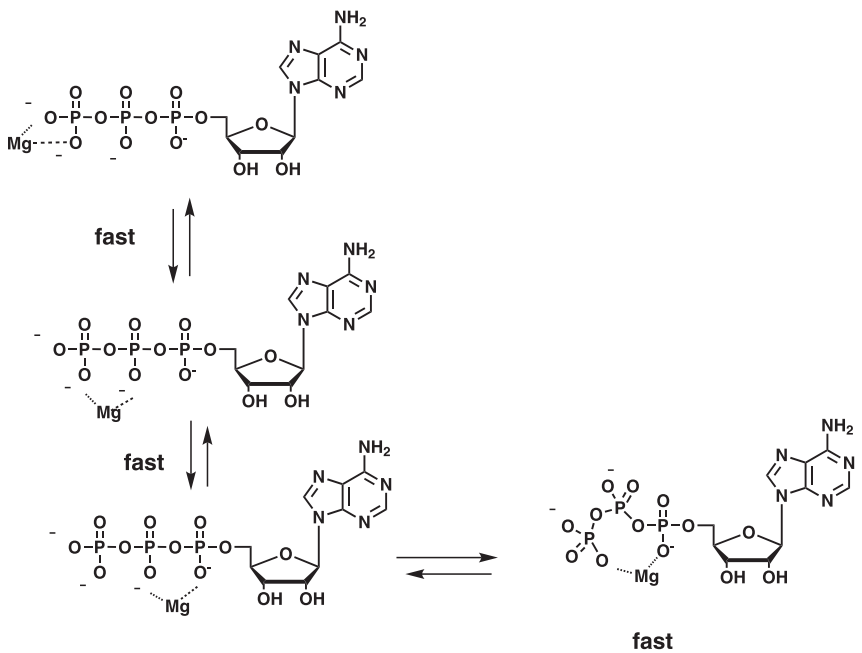


Figure 3.32 Rapidly equilibrating conformers of chelates of Mg^{2+} ions and ATP. Mg-ATP complexes are the universal substrates for all enzymes that use nucleoside triphosphates as substrates.

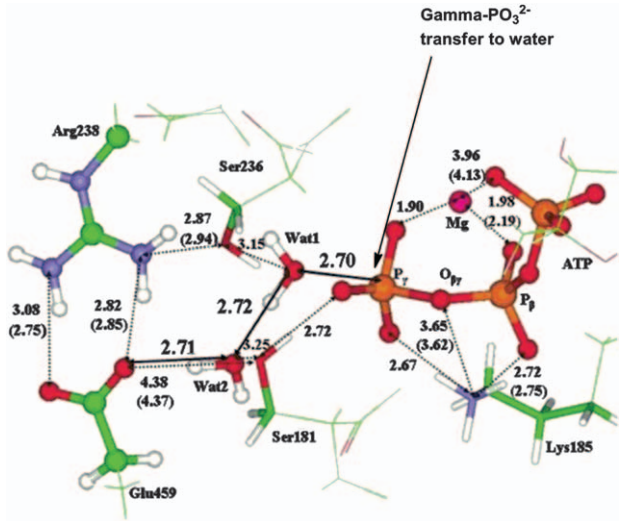


Figure 3.33 Coordination geometry of attacking water molecule and the triphosphate side chain of ATP in the myosin ATPase active site.[55] Reproduced from ref. 55 with permission from National Academy of Sciences, U.S.A., Copyright 2007.

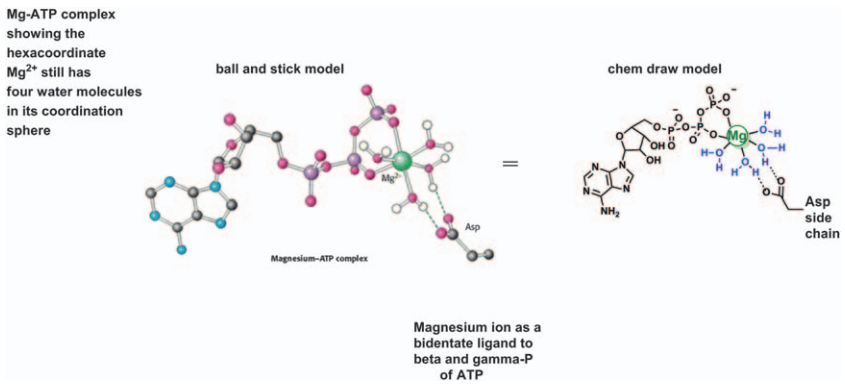


Figure 3.34 A Mg–ATP-βγ-chelate complex with four solvent water molecules[56] still coordinated to hexacoordinate Mg^{2+}.

four water ligands around the Mg^{2+} ion, indicating that Mg^{2+} is typically hexacoordinate, solvated with six water molecules in solution, and retaining four of them in this ATP chelate (Figure 3.35).

It is in this context that we turn in the next chapter to the other two main types of side chain cleavage reactions of ATP, including metabolic examples of phosphoryl transfers, pyrophosphoryl transfers, and nucleotidyl transfers.

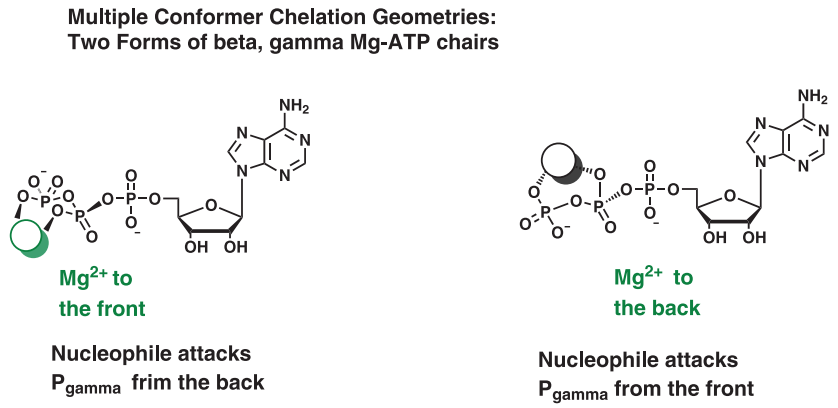


Figure 3.35 Conformation of ATP in the active site of myosin ATPase. The three phosphate groups of the triphosphate side chain.[55]

References

1. F. H. Westheimer, Why Nature Chose Phosphates, *Science*, 1987, **235**, 1173–1178.
2. (a) T. R. Cech, D. Herschlag, J. A. Piccirilli and A. M. Pyle, RNA catalysis by a group I ribozyme, *J. Biol. Chem.*, 1992, **267**, 17479–17482; (b) J. K. Lassila, J. G. Zalatan and D. Herschlag, Biological phosphoryl transfer reactions: understanding mechanism and catalysis, *Annu. Rev. Biochem.*, 2011, **80**, 669–792.
3. M. W. Bowler, M. J. Cliff, J. P. Waltho and G. M. Blackburn, Why did nature select phosphate for its dominant roles in biology ?, *New J. Chem.*, 2010, **34**, 784–794.
4. (a) S. W. Kengen, J. Tuininga, F. A. de Bok, A. F. Stams and W. M. de Vos, Purification and characterization of a novel ADP-dependent glucokinase from the hyerthermophilic archaeon Pyrococcus furiosus, *J. Biol. Chem.*, 1995, **270**, 30453–30457; (b) R. S. Ronimus and H. W. Morgan, Cloning and biochemical characterization of a novel mouse ADP-dependent glucokinase, *Biochem. Biophys. Res. Commun.*, 2004, **315**, 652–658.
5. F. Klemke, G. Beyer, L. Sawade, A. Saitov, T. Korte, I. Maldener, W. Lockau, D. J. Nurnberg and T. Volkmer, All1371 is a polyphosphate-dependent glucokinase in Anabena PCC 7120, *Microbiology*, 2014, **160**, 2807–2819.
6. A. Chi and R. G. Kemp, The primordial high energy compound: ATP or inorganic phosphate, *J. Biol. Chem.*, 2000, **275**, 35677–35679.
7. Y. Makino, T. Sato, H. Kawamura, S.-I. Hachisuka, R. Takeno, T. Imanaka and H. Atomi, An archael ADP-dependent serine kinase involved in cysteine biosynthesis and serine metabolism, *Nat. Commun.*, 2016, **7**, 1–12.

8. J. E. Wilson, Isozymes of mammalian hexokinase: structure, subcellular location and metabolic function, *J. Exp. Biol.*, 2003, **206**, 2049–2057.
9. N. Sugiyama, H. Immamura and Y. Ishihama, Large-scale discovery of substrates of the human kinome, *Sci. Rep.*, 2019, **9**, 1–12.
10. G. Gruber, H. Wieczorek, W. R. Harvey and V. Muller, Structure-function relationships of A-, F, and V-Type ATPases, *J. Exp. Med.*, 2001, **204**, 2597–2605.
11. R. Milo and R. Phillips, *Cell Biology by the Numbers*, Taylor & Francis Books: Garland SciencePress, 2015.
12. C. T. Walsh, B. Tu and Y. Tang, Eight thermodynamically activated but kinetically stable molecules power cellular metabolism, *Chem. Rev.*, 2018, **118**, 1460–1494.
13. J. R. Knowles, Enzyme catalyzed phosphoryl transfer reactions, *Annu. Rev. Biochem.*, 1980, **49**, 877–919.
14. H. Ren, L. Wang, M. Bennett, Y. Liang, X. Zheng, F. Liu, L. Li, J. Nan, M. Luo, S. Eriksson, C. Zhang and X.-D. Su, The crystal structure of human adenylate kinase 6: an adenylate kinase localized to the cell nucleus, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 303–308.
15. B. Schneider, M. Babolat, Y. W. Xu, J. Janin, M. Veron and D. Deville-Bonne, Mechanism of phosphoryl transfer by nucleoside diphosphokinase pH dependence and role of the active site Lys16 and Tyr56 residues, *Eur. J. Biochem.*, 2001, **268**, 1964–1971.
16. V. Gupta and R. N. Bamezai, Human pyruvate kinase M2 a multifunctional protein, *Protein Sci.*, 2010, **19**, 2013–2044.
17. G. P. Dobson, S. Hitchins and W. E. Teague Jr., Thermodynamics of the pyruvate kinase reaction and the reversal of glycolysis in heart and skeletal muscle, *J. Biol. Chem.*, 2002, **277**, 27176–27182.
18. (a) M. Bublitz, J. P. Morth and P. Nissen, P-type ATPases at a glance, *J Cell Sci.*, 2011, **124**, 2515–2519; (b) K. Cotter, L. Stransky, C. McGurie and M. Forgeac, Recent insights into the structure, regulation, and function of V-ATPases, *Trends Bicohem. Sci.*, 2015, **40**, 611–622.
19. R. L. Cross and V. Muller, The evolution of A-, F- and V-type ATP synthases and ATPases: reversals in function and changes in $H+$/ATP coupling ratio, *FEBS Lett.*, 2004, **576**, 1–4.
20. T. Ogura and A. J. Wilkinson, AAA^{+} superfamily ATPases: common structure-diverse function, *Genes Cells*, 2001, **6**, 575–597.
21. R. S. Yedidi, P. Wendler and C. Enenkel, AAA-ATPases in protein degradation, *Front. Mol. Biosci.*, 2017, **4**, 42.
22. M. Rappas, H. Niwa and X. Zhang, Mechanisms of ATPases-a multidisciplinary approach, *Curr. Protein Pept. Sci.*, 2004, **5**, 89–105.
23. J. Snider, G. Thibault and W. A. Houry, The AAA superfamily of functionally diverse proteins, *Genome Biol.*, 2008, **9**, 216.
24. W. Kuhlbrandt, Structure and mechanism of F-type ATP synthase, *Annu. Rev. Biochem.*, 2019, **88**, 515–549.
25. S. Wilkens, Structure and mechanism of ABC transporters, *F1000 Prime Rep.*, 2015, **7**.

26. M. Forgeac, Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology, *Nat. Rev. Mol. Cell. Biol.*, 2007, **8**, 917–929.
27. M. E. Maxson and S. E. Grinstein, The vacuolar type H+ -ATPase at a glance-more than a proton pump, *J. Cell Sci.*, 2014, **127**, 4987–4993.
28. L. Zhou and L. A. Sazanov, Structure and conformational plasticity of the intact Thermus thermophilus V/A-type ATPase, *Science*, 2019, **365**, eaaw9144.
29. R. K. Nakamoto, S. Baylis and M. K. Al-Shawi, The rotary mechanism of the ATP synthase, *Arch. Biochem. Biophys.*, 2008, **476**, 43–50.
30. G. Cingolani and T. M. Duncan, Structure of the ATP synthase catalytic complex (F1) from *E. coli* in an auto-inhibited conformation, *Nat. Struct. Mol. Biol.*, 2011, **8**, 701–707.
31. D. Stock, A. G. Leslei and J. E. Walker, Molecular architecture of the rotary motor in ATP synthase, *Science*, 1999, **286**, 1700–1705.
32. P. Neupane, S. Bhuju, N. Thapa and H. K. Bhattarain, in *ATP synthase: structure, function, inhibition*, ed. E. DiCara, Biomolecular Concepts, **vol. 10**, 2019.
33. P. D. Boyer, The ATP Synthase-a splendid molecular machine, *Annu. Rev. Biochem.*, 1997, **66**, 717–749.
34. I. B. Holland and M. A. Bright, *Journal of Molecular Biology*, 1999, **293**(2), 381–399.
35. K. Hollenstein, R. J. Dawson and K. P. Locher, Structure and Mechanism of ABC transporter proteins, *Curr. Opin. Struct. Biol.*, 2007, **17**, 412–418.
36. K. P. Locher, Mechanistic diversity in ATP-binding cassette (ABC) transporters, *Nat. Struct. Mol. Biol.*, 2016, **23**, 487–493.
37. M. Dean, *The Human ATP-binding cassette (ABC) transporter superfamily*, National Center for Biotechnology Information (US), Bethesda MD, 2002.
38. F. Liu, Z. Zhang, L. Csanady, D. C. Gadsby and J. Chen, Molecular structure of the Human CFTR ion channel, *Cell*, 2017, **169**, 85–95.
39. J. M. Miller and E. J. Enemark, Fundamental characteristics of AAA+ protein family structure and function, *Archaea*, 2016, 9294307.
40. K. Karata, T. Inagawa, A. J. Wilkinson, T. Tatsuta and T. Ogura, *J. Biol. Chem.*, 1999, **274**, 26225–26232.
41. A. G. Gilman, G proteins: transducers of receptor-generated signals, *Annu. Rev. Biochem.*, 1987, **56**, 615–649.
42. K. Wennerberg, K. L. Rossman and C. J. Der, The Ras superfamily at a glance, *J. Cell Sci.*, 2005, **18**, 843–846.
43. R. D. Vale, The molecular motor toolbox for intracellular transport, *Cell*, 2003, **112**, 467–480.
44. E. K. Hurowitz, J. M. Melnyk, Y. J. Chen, H. Kouros-Mehr, M. I. Simon and H. Shizuya, Genomic characterization of the human heterotrimeric G protein alpha, beta, and gamma subunit genes, *DNA Res.*, 2000, **7**, 111–120.
45. A. M. Rojas, G. Fuentes, A. Rausell and A. Valencia, The Ras protein superfamily: evolutionary tree and role of conserved amino acids, *J. Cell Biol.*, 2012, **196**, 189–201.

46. L. Qu, C. Pan, S. He, B. Lang, G.-D. Gao, X.-L. Wang and Y. Wang, The Ras superfamily of small GTPases in non-neoplastic cerebral diseases, *Front. Mol. Neurosci.*, 2019, **21**, 121.
47. B. Binotti, R. Jahn and J. J. E. Chua, Functions of rib proteins at presynaptic sites, *Cells*, 2016, **5**, 7.
48. T.-K. Er, C.-C. Chen, Y.-Y. Liu, J.-K. Hwang, M. J. Barrio, M. Rodrigo, E. Garcia-Toro and M. Herreros-Villanueva, *PLoS One*, 2013, **8**, e55793.
49. C. Kobayashi and S. SaIto, Relation between the conformational heterogeneity and reaction cycle of Ras: Molecular simulation of RAS, *Biophys. J.*, 2010, **99**, 3726.
50. O. Daumke and G. J. Praefke, Mechanisms of GTP hydrolysis and conformational transitions in the dynamin superfamily, *Biopolymers*, 2016, **105**, 580–593.
51. N. Clementi and N. Polacek, Ribosome-associated GTPases: the role of RNA for GTPase activation, *RNA Biol.*, 2010, **7**, 521–527.
52. G. Tian, A. Bhamdipati, N. J. Cowan and S. A. Lewis, Tubulin folding cofactors as GTPase-activating proteins, *J. Biol. Chem.*, 1999, **274**, 24054–24058.
53. C. A. Wang, C. Cormier, A. Gigant and B. Knossow, M, Insight into the GTPase activity of tubulin from complexes with stratmin-like domains, *Biochemistry*, 2007, **46**, 10595–10602.
54. K. Maeda, T. Nakata, Y. Noda, R. Sato-Yoshitake and N. Hirokawa, Interaction of dynamin with microtubules: its structure and GTPase activity investigated by using highly purified dynamin, *Mol. Biol. Cell*, 1992, **3**, 1181–1194.
55. B. L. Grigorenko, A. V. Rogov, I. A. Topol, S. K. Burt, H. M. Martinez and A. V. Nemukhin, Mechanism of the myosin catalyzed hydrolysis of ATP as rationalized by molecular modeling, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 7057–7061.
56. J. Berg, J. Timoczko and L. Stryer, *Biochemistry*, 5th edn, W. H. Freeman & Co., 2002.
57. J. Letts and L. Sazanov, *Nat. Struct. Mol. Biol.*, 2017, **24**, 800–808.
58. J. Colicelli, *Science's STKE*, 2004, **250**, re13.

CHAPTER 4

Nucleotidyl Transfers (ATP and NTPs)

4.1 Adenylyl and Nucleotidyl Transfers

The second major route of enzyme-accelerated cleavage of Mg–ATP involves attack by a cosubstrate nucleophilic atom at the α-phosphate of ATP (Figure 4.1). A generic reaction stoichiometry is shown, with variable bond formation and bond cleavage extents possible in different members of the large nucleotidyl transferase class. Figures 4.1 and 4.2 also show that the inorganic pyrophosphate product, in cells that express highly active inorganic pyrophosphatases, is rapidly hydrolyzed to two molecules of inorganic phosphate dianions. The low concentration of PPi in cells ($<10^{-6}$ M) means a significant pull on the equilibrium of the nucleotidyl transferase reaction in the ATP (NTP) cleavage direction.

There are mechanistic analogies and distinctions from the ATP-dependent phosphoryl transfers of the preceding chapter. In both phosphoryl and nucleotidyl transfers a cosubstrate nucleophile attacks one of the electrophilic side chain phosphorus atoms embedded in the triphosphate side chain. The terminal Pγ is *a priori* more accessible sterically, but the magnesium ion chelate geometry (Section 3.9) may favor a less shielded Pα in the active sites of nucleotidyl transferases. ADP is the coproduct in phosphoryl transferring kinases. Inorganic pyrophosphate is the coproduct in every nucleotidyl transferase. ADP and PPi each have an intact phosphoric anhydride linkage. However, while there are few if any dedicated ADP hydrolases (ADP to AMP and Pi), inorganic pyrophosphatases are widespread in organisms and may be a key determinant favoring the nucleotidyl transfer regiochemistry as elaborated in the examples of this chapter.

This two-enzyme combination, nucleotidyl transferase and pyrophosphatase accounts for cleavage of *both* of the two P–O–P anhydride links in

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

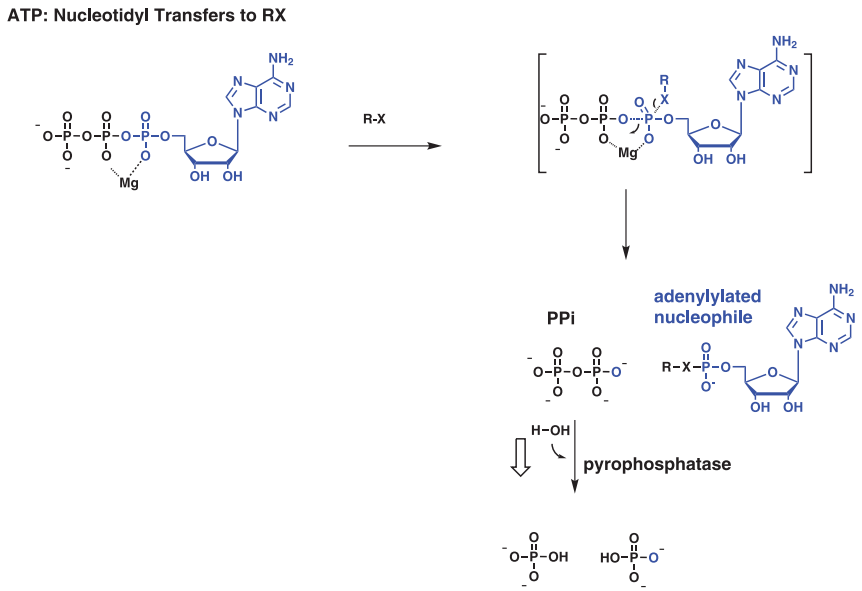


Figure 4.1 Generalized stoichiometry for adenylyl transfers: attack of a nucleophile (RX) on the alpha phosphorus of ATP, releasing adenylated nucleophile and coproduct inorganic pyrophosphate.

Common RX Substrates	RX–AMP Products	
ROH	ROPO–AMP	AMP-phosphodiester
HOH	AMP	adenylic acid: not generally observed
$RCOO^-$	RCOO–AMP	Acyl-AMP
$R\text{-}NH_2$	RNH–AMP	AMP-N_6-Lysine

Figure 4.2 Common oxygen-containing nucleophilic substrates for adenylyl transfers.

the ATP (and related NTP) substrate side chain. The favorable equilibrium allows full accumulation of the nucleotidyl–nucleophile product adduct. This state of affairs may account for the observation that *all three classes of bio-macromolecules*: proteins, nucleic acids (both RNA and DNA), and polysaccharides, as well as membrane phospholipids, are accumulated by nucleotidyl transfer enzyme chemistry (Figure 4.3). Each chain elongation event – each monomer addition – involves cleavage of a nucleoside triphosphate (NTP or 2′-dNTP) to PPi and RX–NMP or RX–2′-dNMP, that may be an intermediate or a stable metabolite. The PPi product from each elongation

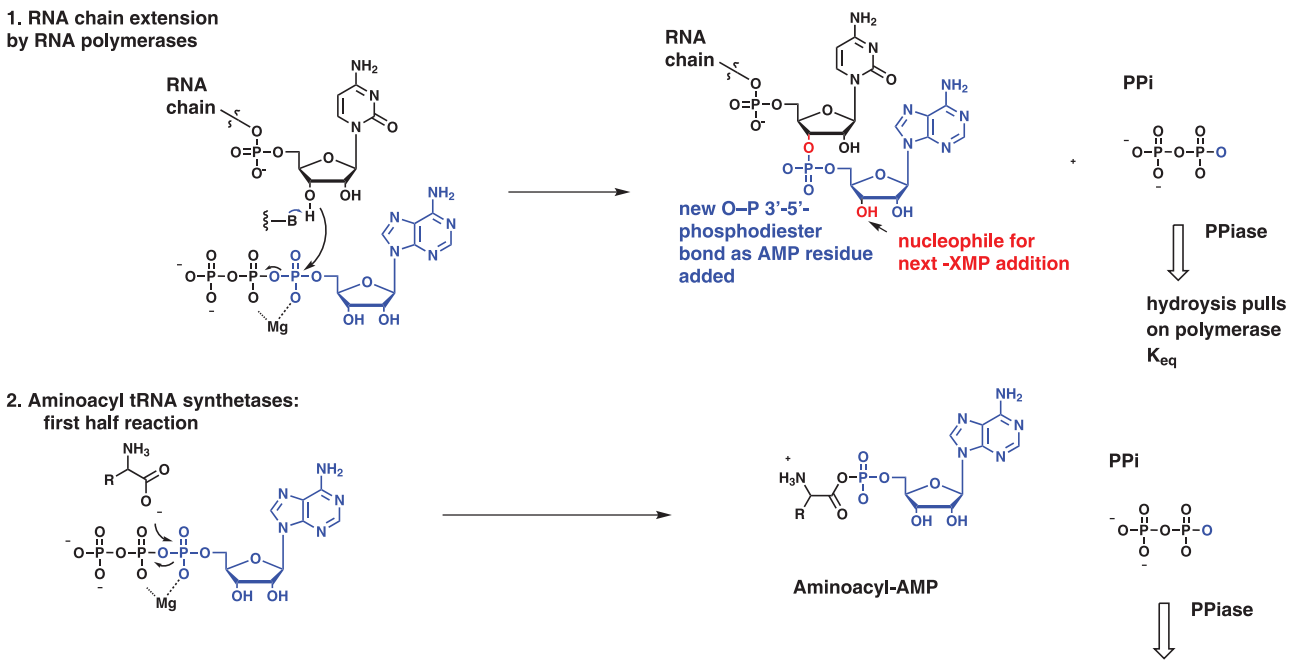


Figure 4.3 Four types of nucleotidyl transferase reactions in four different areas of metabolism: proteins; nucleic acids; oligosaccharides; phospholipids.

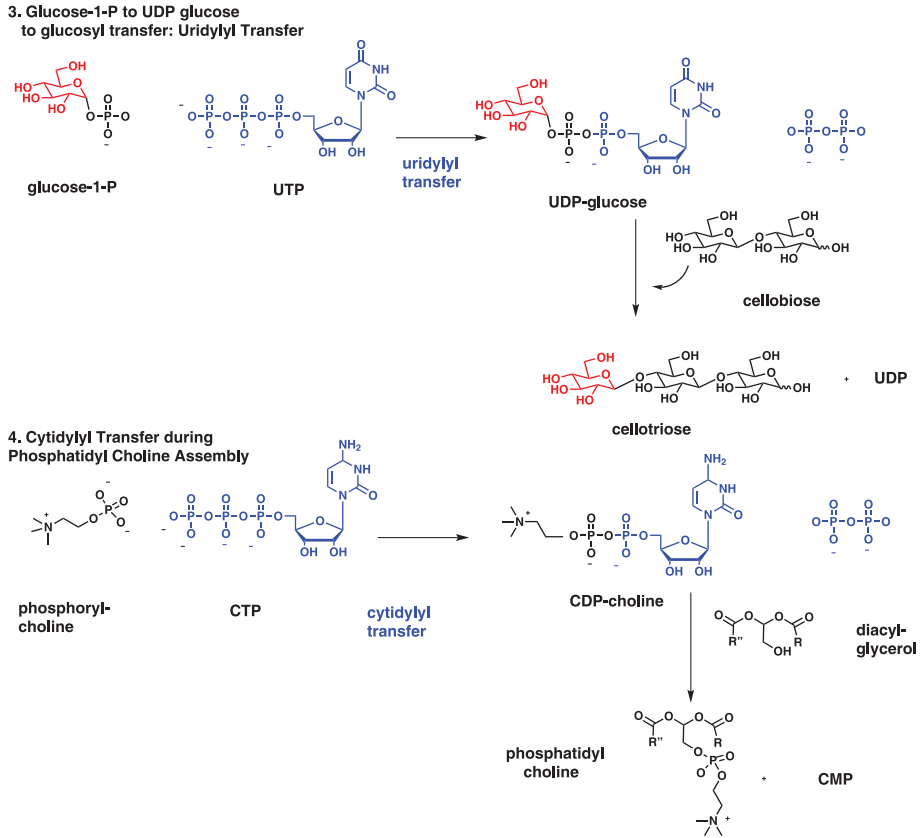


Figure 4.3 (Continued).

step is hydrolyzed to two Pi, to pull on the otherwise unfavorable equilibria for accumulating the entropically unfavorable long, linear biopolymer chains.

Parenthetically, perhaps because hydrolytic cleavage of both P–O–P bonds of ATP is potentially wasteful, there are few if any ATPases that cleave ATP hydrolytically to AMP and PP. This is in contrast to the 300 or so human ATPases that hydrolyze Mg–ATP regiospecifically to ADP and Pi noted in Chapter 3.

4.1.1 Classes of Nucleophiles in Adenylyl Transferases

Figure 4.2 shows oxygen and nitrogen nucleophiles that serve as the attacking species at Pα of ATP. Three classes of oxygen nucleophiles – alcohols, carboxylates, and phosphate oxyanions – are cosubstrates. When ATP is the nucleoside triphosphate substrate attacked, the generic nucleotidyl transfer is specifically termed adenylyl transfer (after AMP; adenylic acid, historically). Figure 4.3 shows representative adenylyl transfers for protein biosynthesis and RNA chain elongation.

4.2 Aminoacyl tRNA Synthetases

Each of the 20 amino acids that become incorporated into proteins (and so termed proteinogenic) are activated in two steps by the aminoacyl transfer RNA (tRNA) synthetases before the released aminoacyl tRNAs are chaperoned to the ribosome.[1] There, they undergo base pairing to the appropriate triplet codon of a messenger RNA transcript and then are covalently incorporated in amide linkage into the encoded growing protein chain. The first half reaction of all the aminoacyl tRNA synthetases in a cell is to select its cognate proteinogenic amino acid and then catalyze attack of that amino acid carboxylate oxygen on Pα of bound cosubstrate Mg–ATP (Figure 4.3, line 2). The resultant aminoacyl-AMP has chemical analogies to the aminoacyl-phosphates noted in the phosphoryl transfers in Chapter 3. Aminoacyl-AMPs are mixed carboxylic–phosphoric anhydrides, where the phosphoric moiety is not inorganic phosphate (as in acyl phosphate) but instead is AMP.

In general, the aminoacyl-AMP intermediates are tightly bound in the synthetase active sites and do not diffuse away into bulk aqueous solution because they are hydrolytically labile. Instead, the second half reaction ensues in which a cognate transfer RNA (tRNA) that has been selected and bound, uses either its 3′-OH, or less often the 2′-OH, of the ribose moiety of the terminal AMP residue in the ~75 residue transfer RNA terminal to transfer the aminoacyl moiety (Figure 4.4). In this second half-reaction the thermodynamically activated amino group is transferred to become a less activated oxoester covalently attached to the bound tRNA (thermodynamically favored direction).

Thus, alanyl tRNA synthetase, first makes alanyl-AMP and then transfers it to the 3′-OH of $tRNA^{ala}$ to produce the loaded form of $tRNA^{ala}$

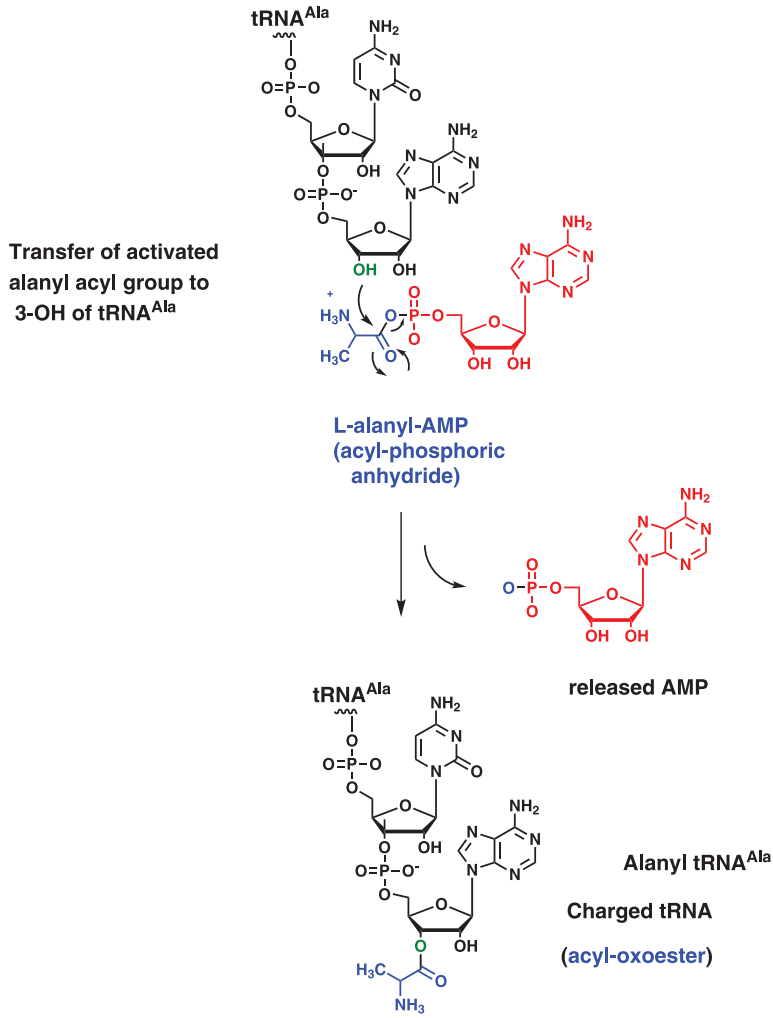


Figure 4.4 Second half-reaction of aminoacyl-tRNA synthetases. Shown here is capture of the bound alanyl-AMP by the 3′-OH of the 3′-terminus of a cognate $tRNA^{ala}$. Thermodynamically favored conversion of a mixed acyl-AMP anhydride to the oxoester of Ala-$tRNA^{ala}$, stable enough to carry the still activated aminoacyl group to ribosomes under the aegis of protein chaperones.

(alanyl-$tRNA^{ala}$) (Figure 4.4). Each of the other 19 aminoacyl-AMPs bound to their cognate aminoacyl tRNA synthetases likewise become covalently attached as oxoesters to their cognate tRNAs, *e.g.* Tyr-$tRNA^{tyr}$, Lys-$tRNA^{lys}$, and so on. Once at the ribosome, with the order of peptide chain growth determined by the base paring of the triplet codon in a messenger RNA (mRNA) and the complementary triplet anticodon in a given tRNA, amide bond formation occurs, as depicted in Figure 4.5 for attack of the alanyl-NH_2

group on a dipeptidyl *N*-formylmethionine (Fmet)-Ser to create the tripeptidyl-tRNA. Each aminoacyl addition constitutes a chain elongation step in protein biosynthesis.

Figure 4.6 schematizes subsequent addition of a glycyl tRNA, creating a growing tetrapeptidyl-tRNA in which the Ala residue is now embedded in peptide linkage at both its amino and carboxyl ends. At every step ATP is cleaved to an aminoacyl-AMP and then released as free AMP in the second half reaction as the aminoacyl group is transferred to its cognate tRNA. The chemical progression at the carboxylate of each amino acid activated and incorporated is: (1) *free carboxylate* to (2) *acyl-AMP* mixed carboxylic–phosphoric anhydride to (3) *aminoacyl-oxoester* when linked to tRNA to (4) *amide* on incorporation to the growing peptidyl chain.

Of these four acyl species, energy is spent (ATP cleavage) to convert the stable carboxylate anion to the aminoacyl-AMP mixed anhydride. From that high point of thermodynamic activation the oxoester is downhill energetically (and thus favored) and then the amide is even more stable than the oxoester.

Meanwhile, every aminoacyl tRNA molecule formed is accompanied by a molecule of PPi that gets hydrolyzed by PPiase activity, pulling on the forward equilibria of the aminoacyl tRNA synthetases to accumulate the activated proteinogenic amino acids needed to fuel protein synthesis.

There are also GTPases that are hydrolyzed to drive architectural changes at the ribosome to effect delivery of aminoacyl tRNAs to the complementary mRNA triplet and to move the growing peptidyl tRNA from the aminoacyl to the peptidyl site.[2] They are notionally analogous to the AAA-ATPase family of enzymes (Chapter 3) that couple ATP hydrolysis to some set of mechanical events in cells.

4.3 Nucleotidyl Transfers in Each Step of RNA Chain Elongation

There are many RNA molecules of different sequences, lengths, and functions in each prokaryotic and eukaryotic cell. In bacteria, the estimate of mRNA molecules per cell at any given time (*e.g.* in log phase) is ~1000 per cell. The larger animal cells may contain 10^5–10^6 messages at any time point of perhaps average length of ~1000 bases in 3′,5′-phosphodiester linkage.[3] Abut 3–5% of the total RNA in a cell is present as messenger RNAs. Another 10–20% is comprised of all the transfer RNAs required for activation and transport of amino acids to ribosomes for protein biosynthesis. The remaining RNA, up to 80%, is ribosomal RNA (rRNA) to enable a sufficient complement of ribosomes to enable translation of the full complement of mRNAs into a proteome. In prokaryotes there are three such structural rRNAs, while in eukaryotes there is an additional small rRNA component.[4]

Messenger RNAs are read at ribosomes. There may be ~10^6 to 10^7 ribosomes with their three copies of ribosomal structural RNAs in animal cells.[3]

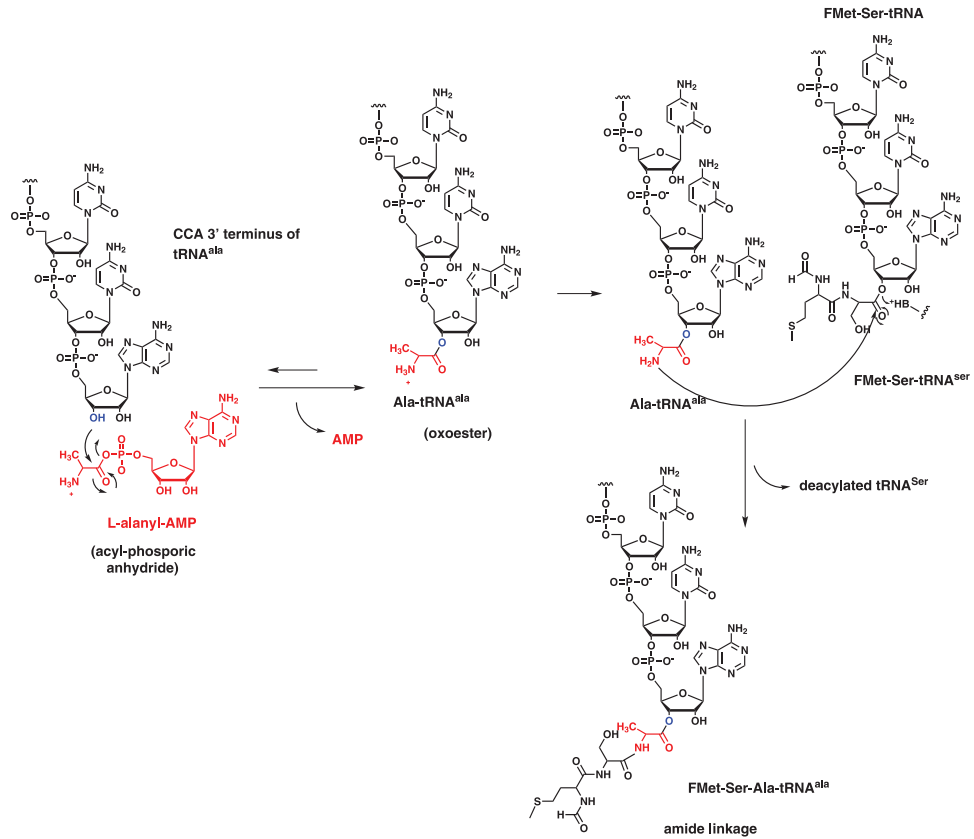


Figure 4.5 Schematic of transfer of alanyl residue from Ala-tRNA^{ala} to the dipeptidyl FMet-Ser-tRNA^{ser} at the peptidyl transferase center of the ribosome: thermodynamically favored conversion of oxoester to amide linkage in the FMet–Ser–Ala-tRNA^{ala} tripeptidyl tRNA product chain.

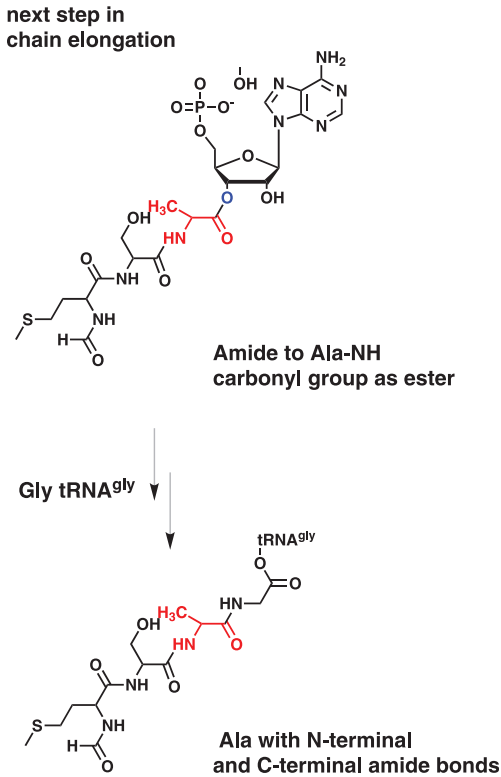


Figure 4.6 The next step in peptidyl chain growth adds a glycine residue (in this example) to form an *N*-formylated tetrapeptidyl-tRNA^{gly}.

There are about 500 genes in humans that code for tRNAs. Then there are other forms of RNA, including noncoding RNAs. It may be that a cell needs to incorporate several billions of the four nucleoside triphosphates – UTP, CTP, ATP and GTP – into RNA molecules in every cell cycle. This need requires high capacity and high fidelity RNA polymerases and a sufficiency of the nucleoside triphosphate building blocks.

We noted earlier in Chapter 3 that the purine nucleoside monophosphates, AMP and GMP, emerge from purine biosynthesis. UMP emerges from the pyrimidine biosynthetic pathway and undergoes two consecutive enzymatic phosphorylations to UTP. At that point UTP is aminated enzymatically to CTP with phosphorylation of the isoamide form of the uracil ring. (Figure 4.7).

Both the purine and pyrimidine nucleotide monophosphates undergo enzyme-catalyzed phosphoryl transfers by specific nucleoside monophosphate transferases at the expense of ATP cleavage to ADP. In Chapter 3 we noted that nucleoside diphosphokinase action can turn any NDP into NTP by another phosphoryl transfer from Mg–ATP (Figure 3.13). Thus, cells can generate sufficient ATP, UTP, GTP, and CTP for RNA synthesis (provided the

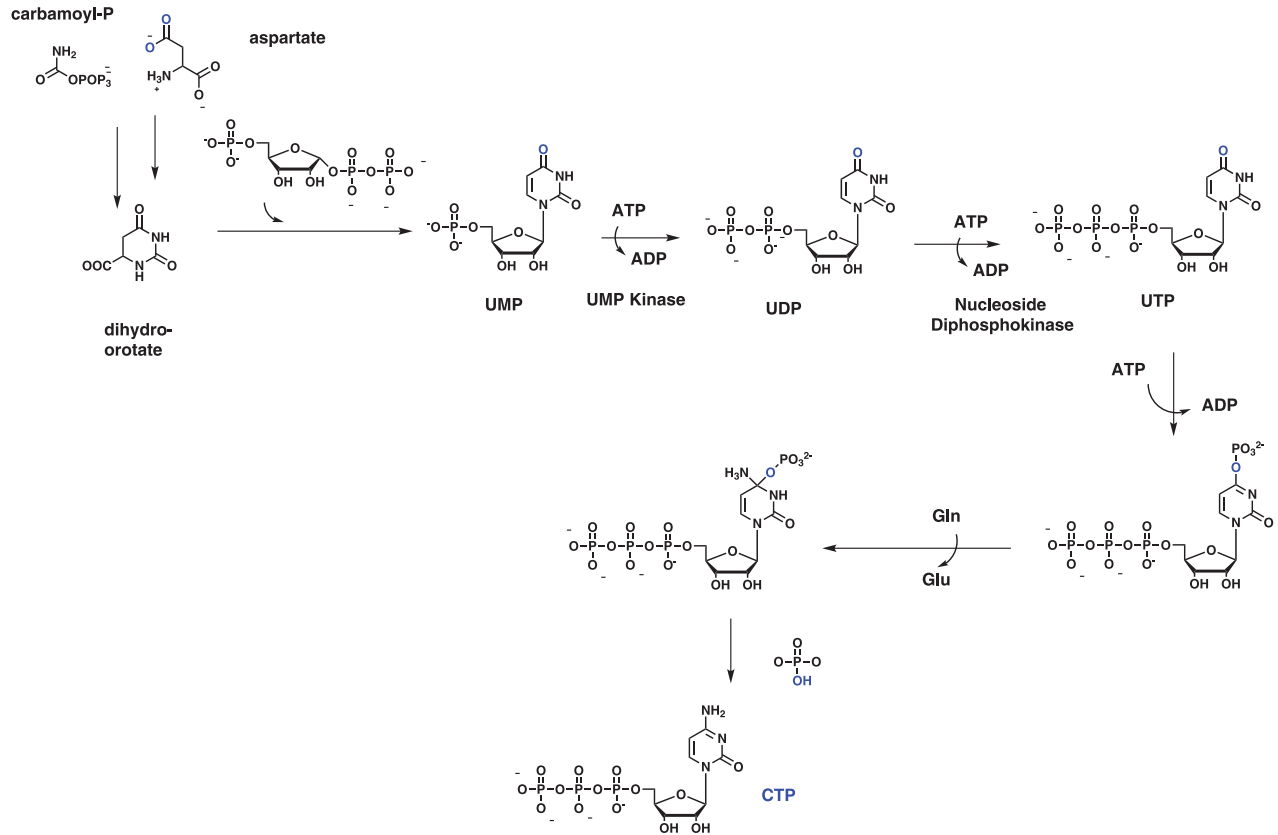


Figure 4.7 Biosynthetic logic to the two pyrimidine nucleoside triphosphates UTP and CTP that are building blocks for RNA biosynthesis. Enzyme-catalyzed amination of UTP at the pyrimidine ring C4 position to CTP occurs at the triphosphate level.

membrane F-type ATP synthase is churning out enough ATP to bring all the other nucleotides up to the triphosphate level).

The RNA polymerases impose specificity by either copying a DNA template strand or in certain viral genome replications, by copying RNA *via* RNA-dependent RNA polymerases. The particular NTP to be incorporated in a given chain elongation step is selected by Watson–Crick base pairing rules. CTP forms three H bonds with a GMP residue in a template strand; GTP pairs likewise with a CMP residue; ATP binds opposite a UMP in the primer strand; UTP comes in opposite an AMP residue with two hydrogen bonds.[5]

Once the base-pairing has been established and the cognate NTP bound at the polymerase active site, the covalent chemistry for chain growth is *nucleotidyl* transfer. As illustrated in Figure 4.8, when ATP is selected and bound noncovalently it becomes oriented in the active site such that the 3′-end of the growing RNA chain can use its terminal ribose-3′-OH as nucleophile. In the example a 3′-CMP residue attacks Mg–ATP at Pα. The transition state unravels in the forward direction with the AMP moiety installed at the attacking nucleophile. This entails formation of a new 3′,5′-phosphodiester bond between the CMP and AMP. The CMP is now the penultimate residue; the newly introduced AMP is the new terminal 3′ residue of the

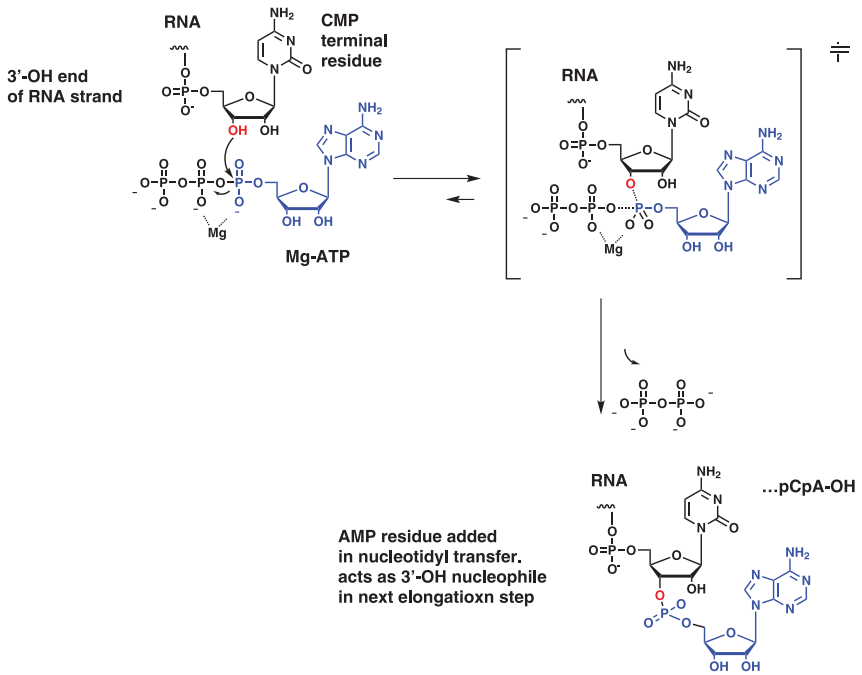


Figure 4.8 RNA chain elongation comprises a series of nucleotidyl transfers to the growing 3′-OH-end of the RNA chain. In this example ATP is the incoming NTP, and its AMP moiety is installed as the new 3′-end with release of PPi.

growing RNA chain. The 3′OH of that AMP residue will be the nucleophile in the next XMP-addition step in RNA chain growth. Pulling the equilibrium forward, to 3′,5′-phophodiester bond formation in each elongation step, will be the hydrolytic cleavage of nascent PPi by pyrophosphatases.

In the example in Figure 4.8 the RNA chain elongation step shown is an AMP (adenylyl) transfer. For CTP as substrate, it would be a net cytidylyl (CMP) transfer. Likewise, GTP gives a GMP moiety (guanylyl transfer) and UTP yields a UMP residue (uridylyl transfer). All four additions can be subsumed under the general term nucleotidyl transfer.

There are multiple RNA polymerases in eukaryotic cells with core and peripheral catalytic and regulatory subunits and complexes of partner proteins that separately transcribe ribosomal RNA, transfer RNAs, and messenger RNAs. The biology is complex for chain initiation, elongation, and termination events, even though the nucleotidyl transfer chemistry of elongation is common[6] for every elongation step by any polymerase.[5]

4.3.1 Two Additional Nucleotidyl Transfers at 5′ and 3′ Ends of Mature mRNAs

During the maturation of precursor forms of messenger RNAs to mature mRNAs a series of enzyme-catalyzed steps occur, including addition of a GMP-based cap at the 5′ end,[7] and a string of AMP residues as 3′-tail.[8,9] Most eukaryotic mRNAs also undergo splicing reactions to remove internal introns and join up exons to provide the final coding sequence in frame.[10]

At the 5′ end of an immature mRNA the first nucleotide retains its 5′-triphosphate. The initial step in subsequent capping is phosphatase action (phosphoryl group transfer to water, Figure 3.2) to convert the 5′-triphosphate to a 5′-diphosphate as shown in Figure 4.9. Then, a nucleotidyl transferase binds that 5′-diphosphate end of the processed mRNA and also a molecule of GTP. The GTP gets attacked at its α-P by the 5′-diphosphate terminus of the mRNA. This generates an unusual 5′–5′ triphosphate linkage between the newly added GMP and the original 5′ end of the mRNA. The third capping enzyme uses *S*-adenosylmethionine as methyl donor to N7 of the GMP portion of the cap to yield the fully mature 5′-capped RNA. Note that the unusual 5′–5′ triphosphate linkage is protective against hydrolytic degradation by 5′-exo-ribonucleases (see Chapter 8), ensuring a longer half-life for these capped messages.

At the opposite end of the nascent mRNA, the 3′-end, polyadenylation occurs by processive action of a polyadenylating enzyme that adds from 100 to 250 AMP residues by successive adenylyl transfers from ATP (Figure 4.10). The polyA polymerase is part of a multienzyme complex that binds to a specific sequence in the 3′-untranslated region of the mRNA, effects limited cleavage (RNA endonuclease cut) to reveal an elongatable sequence, and then adds successive AMP moieties from ATPs to create the poly A tail. That poly A tail is then bound by a poly A-binding protein and the complex

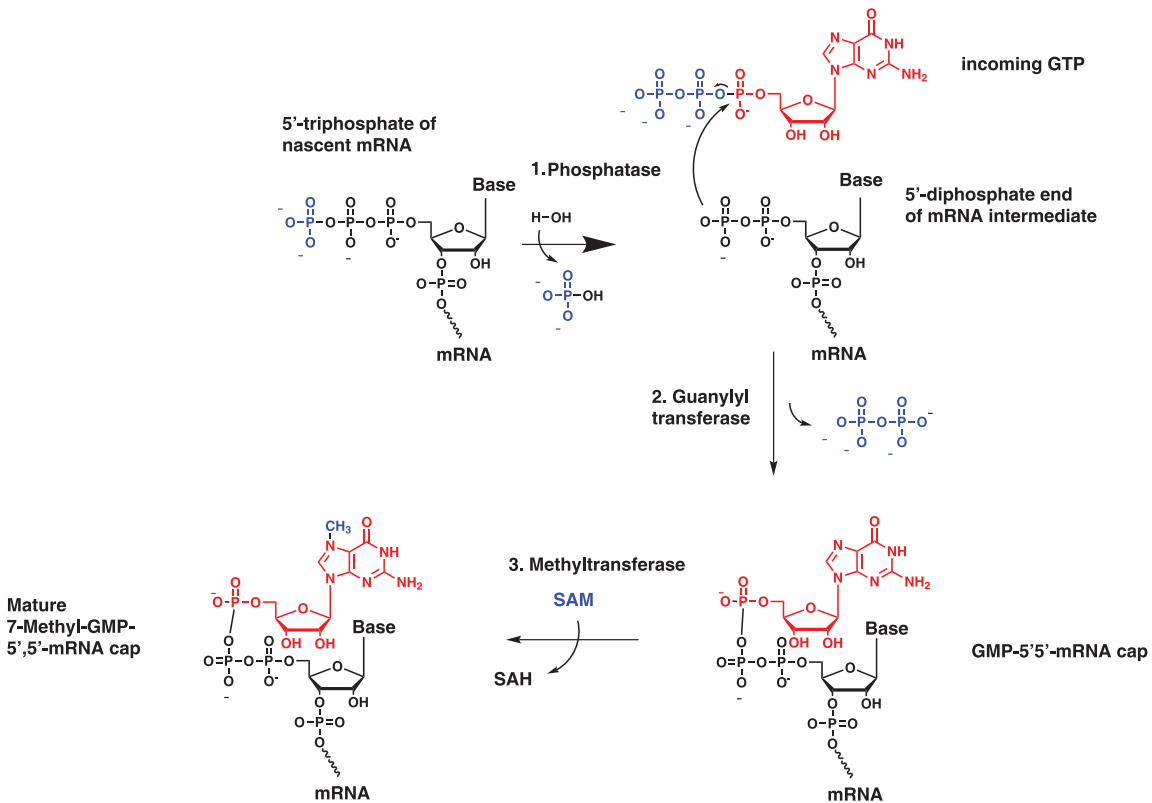


Figure 4.9 Nucleotidyl transfers: Formation of the 7-methyl GMP cap at the 5′ end of mRNAs.

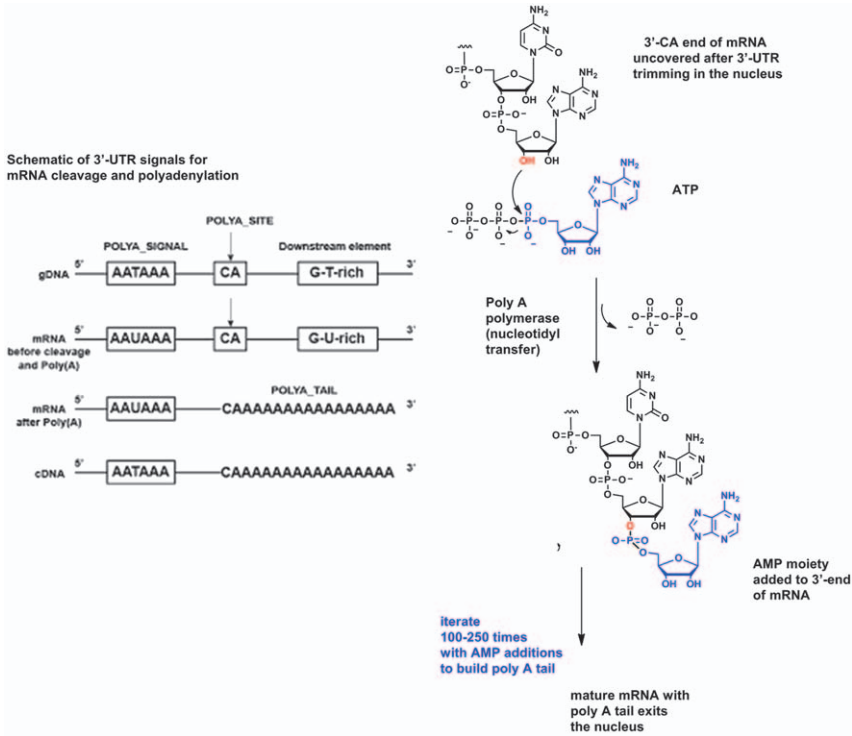


Figure 4.10 Nucleotidyl transfers: Polyadenylation (AMP_n residues) at the 3′ end of mature mRNAs occurs after nucleolytic trimming at the nascent 3′ untranslated region (UTR) end.

efficiently exits from the nucleus to the cytoplasm for translation by cytoplasmic ribosomes. The poly A tail also helps stabilize the mRNA against 3′-exonucleolytic RNAse-mediated hydrolytic degradation. In turn, this stabilizing effect leads to longer lifetimes, and enables higher levels of translation of those messenger RNA molecules. The poly A polymerase is a *template-independent, iterative nucleotidyl transferase*[12] (Figure 4.10). The limit of 100–250 AMP residues added may depend on distance from the original protein complex that bound to and cleaved the 3′ untranslated region (UTR) of the pre mRNA.

4.3.2 Uridylyl and Guanylyl Transfers to 3′-ends of Different Classes of RNAs

An additional set of template-independent nucleotidyl transfers occur at the 3′-OH ends of distinct sites of RNAs to control maturation and function. Humans encode seven noncanonical nucleotidyl transferases. They have structural homology to the DNA polymerase beta sub family, as do XMP transferases, but act at 3′-OH ends of RNAs. Two of the seven transferases

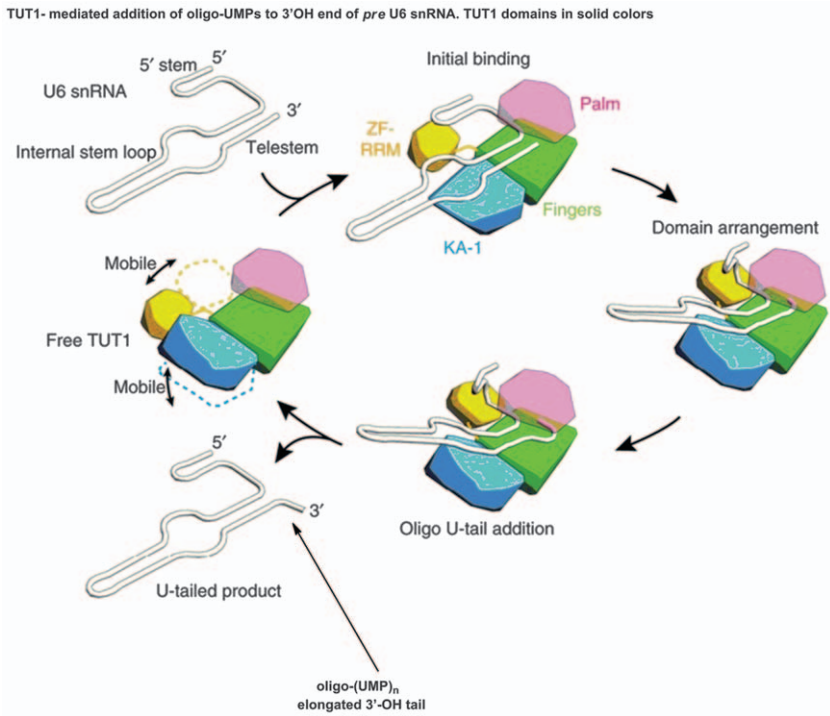


Figure 4.11 Schematic for oligo-UMP addition to 3′OH end of U6snRNA to yield RNA splicing-competent form.
Reproduced from ref. 11 with permission https://doi.org/10.1038/ncomms15788, under the terms of a CC By 4.0 license, https://creativecommons.org/licenses/by/4.0/.

carry out RNA guanylylation, but three of them, terminal uridylyl transferases (TUTs) 1, 4, 7, are selective uridylyl (UMP group) transferases.[13] TUT 1 catalyzes oligo-UMP additions to the U6 small nuclear RNA (snRNA) component of splicesomes (see below) required for functional maturation[11] (Figure 4.11). TUTs 4 and 7 catalyze both mono- and oligo-uridylation of the precursor form of microRNA lethal 7(*let7*), required for subsequent DICER RNAse processing to mature *let7.* The TUT enzymes are viewed as possible therapeutic targets in regulation of specific nodal points of RNA metabolism.

4.4 Equivalent 2′dXMP Transfers in Every Chain Elongation Step of DNA Polymerization

The nucleotidyl transfer strategy employed in RNA chain elongation is mirrored in DNA chain elongation. There are 14 DNA polymerases encoded in the human genome, reflecting a range of replication, recombination, and repair processes.[14,15] While that inventory of DNA polymerases indicates many biological operating contexts, the elongation chemistry logic is

identically nucleotidyl transfer for all of them. The distinct roles of the 14 human DNA polymerases in maintenance of genome integrity, recombination, repair, and replication are beyond the scope of this chemical biology focus and accessible *via* standard biochemistry and cell biology texts and reviews.

There are two exceptions in the building blocks between DNA and RNA, but otherwise the logic is identical. The first change in building blocks is the substitution of 2′-deoxyribonucleoside triphosphates (2′-dNTPs) for the ribonucleoside triphosphates in RNA chain elongation. The second change is replacement of uracil by 5-methyluracil. The 5-methyl-2′deoxy-UTP is known as thymidylate triphosphate, abbreviated dTTP (Figure 4.12A). The enzyme ribonucleotide reductase takes all four of the ribonucleoside diphosphates and reduces them at the 2′-ribose carbon to 2′-deoxyribonucleoside diphosphates. Nucleoside diphosphokinase action then takes them up to the triphosphate level as dNTPs. The uracil to 5-methyluracil transformation occurs at the dUMP level and is catalyzed by thymidylate synthase. The methyl group derives from C3 of serine by way of 5,10-methylene-tetrahydrofolate (*N*-5,10-methylene-THF) as proximal one carbon donor to C5 of the uracil ring (Figure 4.12B).

Figure 4.12A depicts a DNA polymerase-mediated incorporation of a dTMP residue, selected by base pairing to a 2′dAMP on the opposite DNA strand. The 3′ dNMP residue on the elongating chain in this diagram is a dCMP, specified in turn by a dGMP residue on the primer strand.

4.5 DNA Ligases

We note one additional family of nucleotidyl transfer enzymes in DNA repair and replication: DNA ligases.[16] These ligases act at both single strand nicks and also at double strand cuts in double strand DNA templates to repair the gaps between a 5′-phosphate and a 3′-OH on either side of the nick or gap (Figure 4.13). There are about 200 000 spontaneous nicks per day in the human chromosomes that need to be repaired by DNA ligases to maintain the integrity of the genome. This is also the nick configuration in the lagging strand Okazaki fragments in normal DNA polymerase action to replicate DNA. In human genome replication the estimate is about 20 million of the discontinuous Okazaki fragments accumulate that must be sealed by DNA ligase action to create an intact lagging DNA strand.

DNA ligases are also the catalysts for ligation of DNA double strand breaks, in the process of nonhomologous end rejoining[17] (NHEJ) that can occur up to 50 times a day per cell. It can be higher in T cells and B cells where purposeful double strand breaks are part of the immune cells' maturing repertoires. In NHEJ processes, a large assemblage of proteins come together to stabilize the ends, convert any single strand overhangs or gaps to blunt ends with 3′-OH and 5′-phosphates appropriate for action by ligases. Figures 4.14 and 4.15 schematize the timing for collection of proteins that bind to and stabilize the ends of DNA fragments, then process them to blunt ends such that a DNA

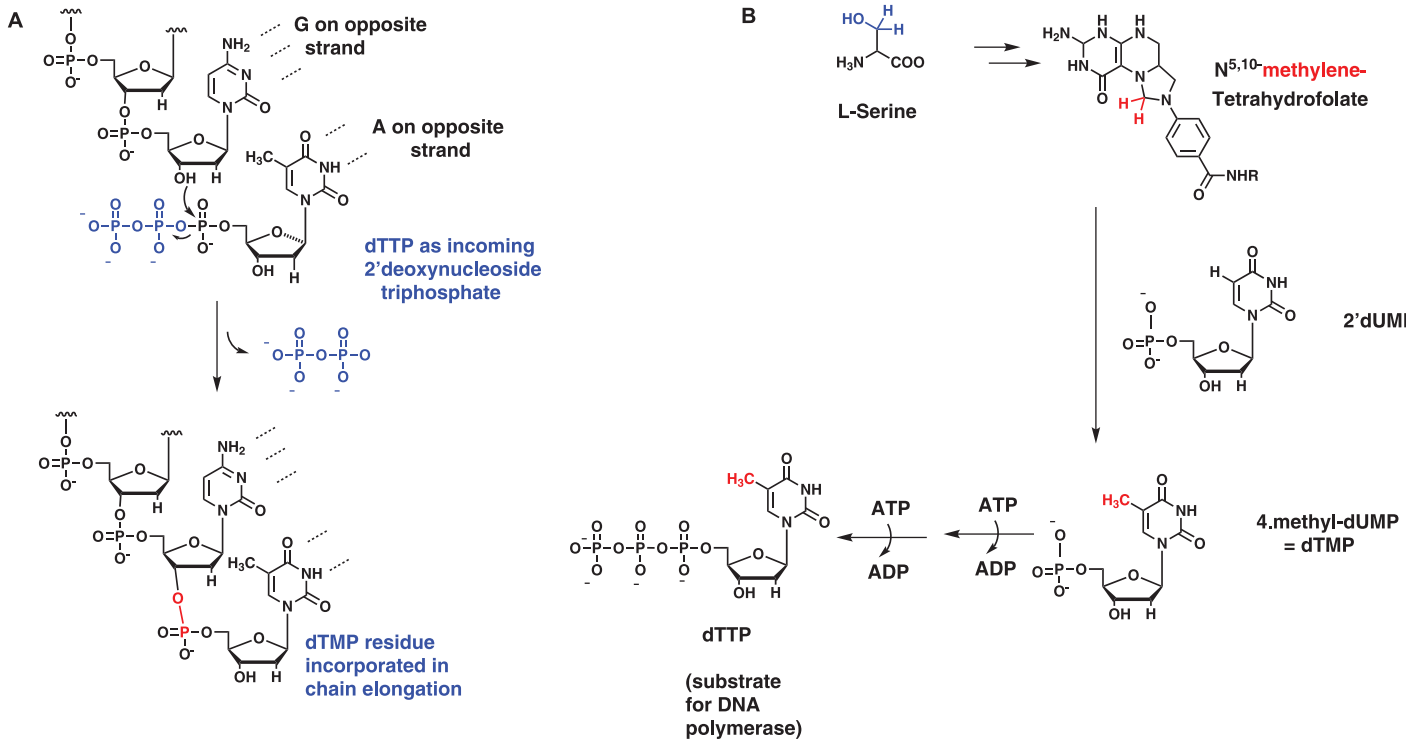


Figure 4.12 (A) DNA chain elongation by nucleotidyl transfer: incorporation of a 2′dTMP moiety from a 2′dTTP monomer to the 3′end-OH as nucleophile: (B) Schematic for conversion of the C3–CH_2–*OH* group of serine to the C5–CH_3 group in dTMP on the way to dTTP.

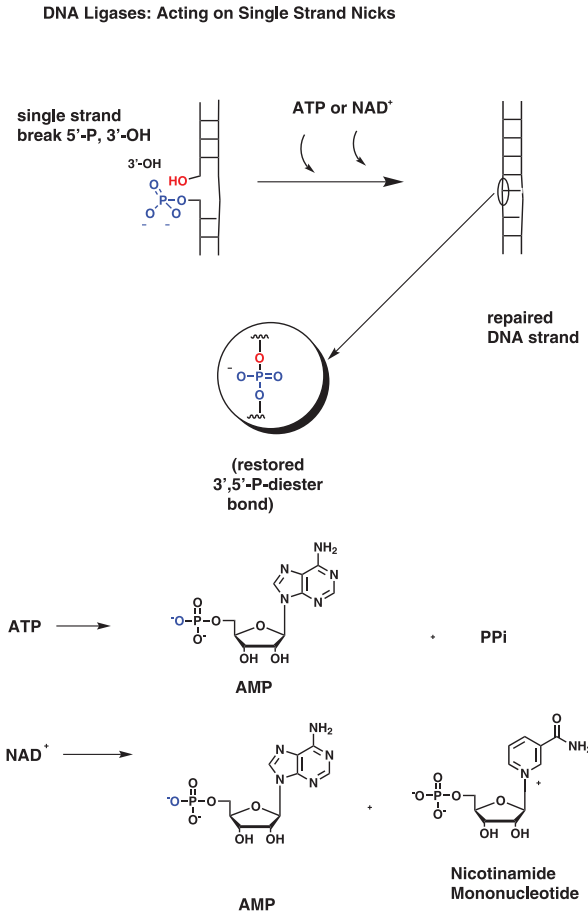


Figure 4.13 Schematic of repair of a single strand break or nick in DNA by ATP- or NAD^+-dependent DNA ligases. The repair occurs at gaps with a 5′-P and a 3′-OH and involves nucleotidyl transfers from either ATP or (rarely) NAD^+.

ligase can repair both strands by reforming the internucleotide phosphodiesters bonds. DNA ligase IV (of four encoded human DNA ligases) appears to be the major ligase in NHEJ double strand religations.[18,19]

Eukaryotic DNA ligases use ATP as a nucleotidyl donor as noted below. Some prokaryotic DNA ligases instead use NAD^+ as donor, a unique role for this ubiquitous coenzyme as an adenylyl group donor (Figure 4.13). The coproduct from ATP donation of its AMP moiety is the familiar inorganic pyrophosphate. From NAD^+ as AMP group donor, the coproduct is nicotinamide mononucleotide, an intermediate in NAD^+ biosynthesis (Figure 4.13). There is an open question if NAD^+ as AMP donor was an evolutionary *cul de sac*, or an experiment that was not as successful as using ATP as the AMP donor. There are reports of ADP as donor of the AMP moiety

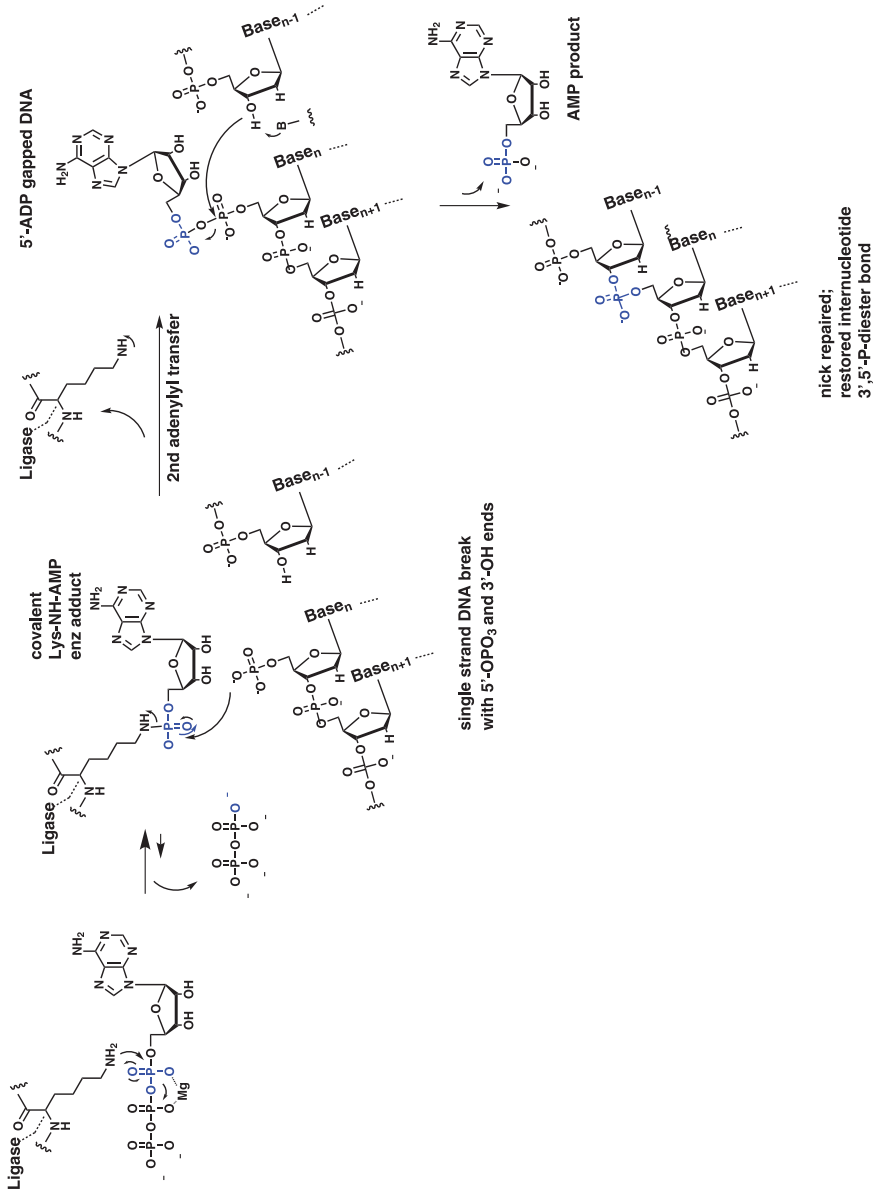


Figure 4.13 (Continued).

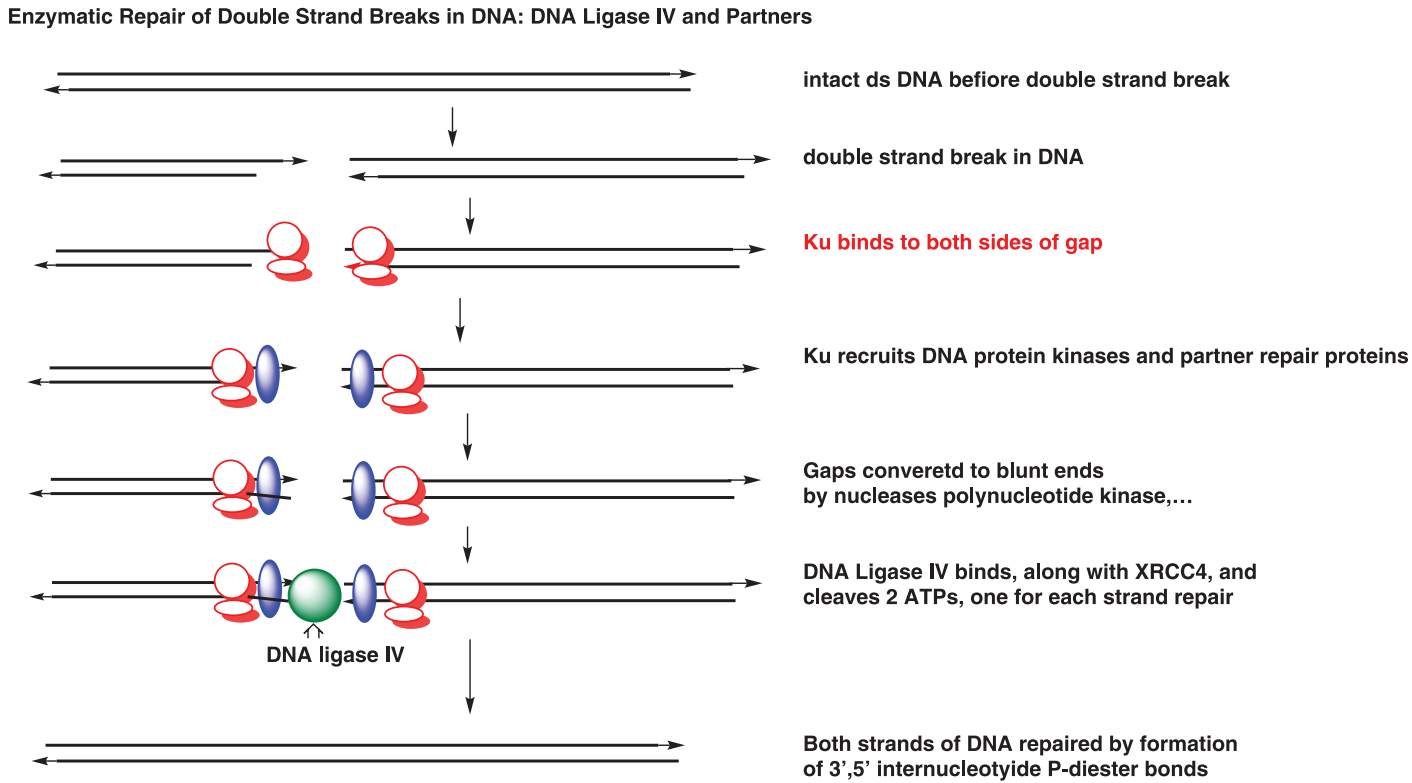


Figure 4.14 Double strand DNA break repair involves a complex of multiple enzymes to bind and protect the single strand ends, to trim or fill in overhangs, to set up blunt ends for DNA ligase action on each strand.[20]

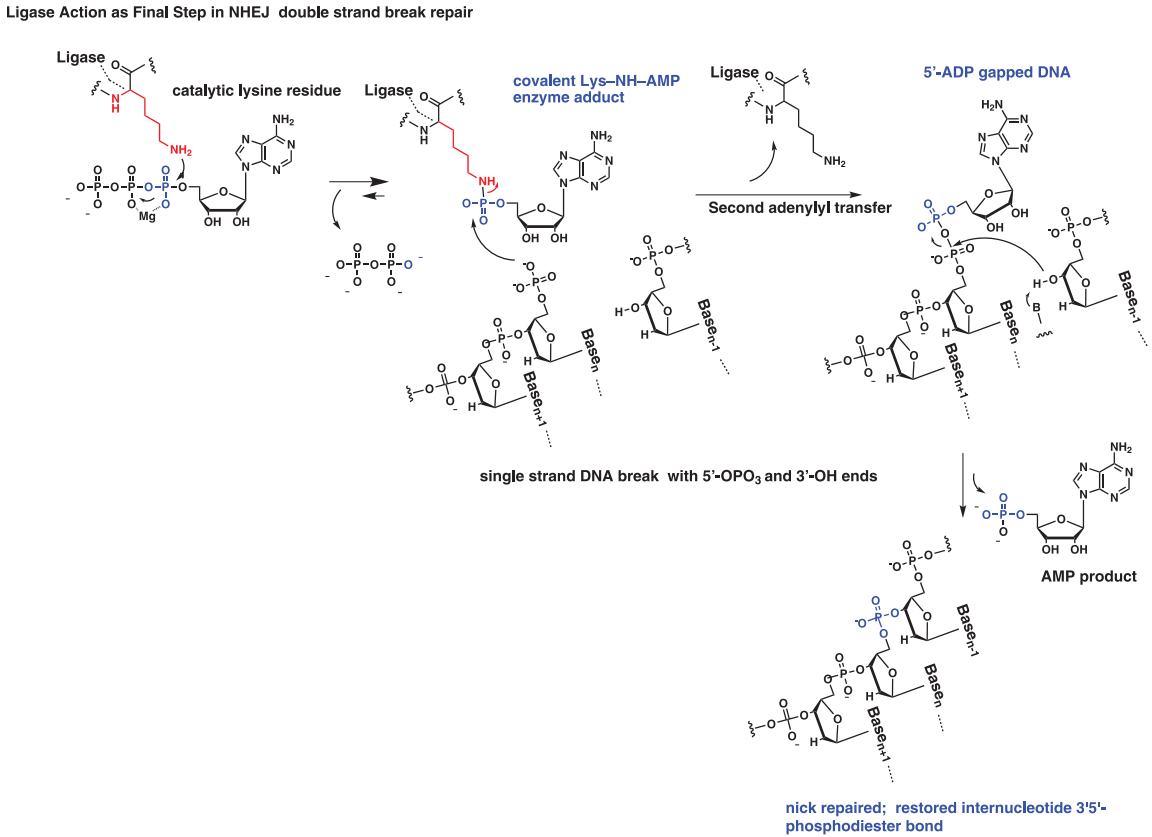


Figure 4.15 Mechanism of action of DNA ligase on one strand of a DNA break (repeated at opposite strand break) involves three sequential nucleotidyl transfer steps.

in a small ligase set, yielding Pi as the coproduct.[21] That could have been an early evolutionary step in DNA ligase evolution.

The DNA ligases are of unusual mechanistic interest because they carry out two consecutive nucleotidyl (adenylyl) transfers: one on the ligase itself, and the next on the gapped DNA[16] (Figure 4.15). Then, there is an expulsion of the AMP moiety with one of the original 5′-phosphate oxygen atom as the gap is resealed.

The first transfer is to one of the lysine side chains in the enzyme, a covalent *N*-AMPylation to generate the active adenylylating agent for the second step. The enzyme-lysine-AMP is donor to the 5′phosphate at the DNA break site. This is an unusual example of a phosphate oxyanion as a nucleophile in an adenylyl transfer. It yields an ADP group on the 5′ side of the break. Then, the 3′-OH of the DNA dXMP residue on the 3′ side of the break attacks the internal phosphate of the 5′-ADP moiety of the DNA. That step expels AMP–O^- as the product while the 3′,5′ phosphodiester bond forms and closes the gap or nick at that DNA site. In NHEJ double strand break repair, this process has to occur on *both DNA strands*.

From ATP to AMP–Lys–enzyme adduct to 5′-ADP intermediate, three aspects of phosphorus enzymatic chemistry are in play: the phosphoric anhydride between Pα and Pβ in ATP, the phosphoramidate linkage in the covalent enzyme–AMP adduct, and the phosphodiester in the ADP version of the 5′ side of the gap. The ligase catalytic strategy for DNA gap repair is a microcosm of the range of phosphorus chemical biology.

DNA ligases carry out hundreds of thousands to millions of DNA strand gap joinings in human cells per day. The enzyme–AMP intermediate is both an example of an amine nucleophile attacking Pα of ATP and an enzyme auto-AMPylation reaction (a proteomics event, see Section IV, Chapter 16). The second adenylyl transfer involves migration to the 5′ phosphate monoester of DNA at the 5′ side of the single strand nick, generating a covalent ADP adduct. This step has activated the 5′-phosphate of the DNA strand for capture by the 3′-OH across the gap. As the 3′,5′ internucleotide phophodiester is restored, an isoenergetic phosphodiester is cleaved as AMP departs with one of the original oxygens of the 5′ phosphate of the gapped DNA. The DNA ligase melds protein AMPylation, DNA AMPylation, and phosphodiester exchange in execution of the DNA repair strategy.

4.6 RNA Splicing

The major distinction between prokaryotic *vs.* eukaryotic messenger RNA, the mediator between the stable genetic information in DNA, and the encoded protein sequence specified by each gene, is the phenomenon of mRNA splicing in eukaryotes, first deciphered in the late 1970s.[10,22] An evolutionary debate has raged as to whether intervening sequences that get excised – ''introns early'' – were ''junk'' DNA that has been subsequently trimmed away in gene-packed, efficient prokaryotic genomes. Alternatively, there is

the ''intron late'' hypothesis that the extant introns in eukaryotes developed after the split between prokaryotes and eukaryotes. While there are arguments that alternate splicing of exons, skipping some, creates increased diversity in the human proteomes beyond what would be achievable with ''only'' 26 000 genes, the energy cost for keeping so much noncoding DNA within gene sequences is itself enormous.

The *intron* terminology is paired with *exon* terminology. Introns are intervening sequences in pre mRNAs that are spliced out as expressed regions – exons– are spliced together to form the in-frame coding sequences of mature mRNAs.[23] In humans some 95% of the ~26 000 genes are calculated to contain introns that are to be spliced out as mRNAs exit the nucleus and reach the cytoplasm and the ribosomal factories of protein biosynthesis.

The total number of computed introns and exons in human DNA are in the range of ~240 000, for an average of almost nine exons per gene and eight introns to be spliced out in each gene. While canonical splicing would remove every intron and splice each upstream exon to its downstream neighbor, there are many cases of alternative splicing where some exons are skipped. For the famous Drosophila gene Down syndrome cell adhesion molecule (*dscam*), up to 38 000 different mature mRNA forms could arise if all the alternate splicing patterns were enacted!

The sizes of introns vary dramatically, with a lower limit of ~30 base pairs for the human macrophage stimulating 1-like (*mst1l*) gene.[24] That may reflect some minimal distance or architecture for recognition or action by spliceosome machinery. Some 80% of human introns are <200 bp, while 10% exceed 11 000 base pairs. Three of the longest human genes, at about 2.2 to 2.3 million base pairs, are the dystrophin gene, contactin protein associated, like 2, and a receptor type protein tyrosine phosphatase delta (Quora web site) gene. The nebulin gene has 150 calculated exons while the voltage gated potassium (KV) channel interacting protein has seven introns, the longest of which is 1.04 million base pairs. In total, abut 25% of human DNA is found in introns, all of which get deleted before being translated into protein coding sequence.

The scope of intron biology far exceeds this volume's chemical biology focus on phosphorus and how its chemical properties both constrain and enable biology. Indeed, the complexity is so high, the scope of information transfer so broad, that even 40 years after RNA splicing was discovered, the structure, mechanism, specificity and regulation as well as the overall question of why introns are retained in eukaryotic DNA continue to be areas of deep contemporary inquiry.

Five small nuclear RNAs are involved at distinct regions of introns and exons in mRNAs to be spliced and at distinct stages of the spliceosome nanomachines.[25] Up to 150 proteins are also recruited and released at different stages of the process of splicing as the intron–exon boundaries are recognized and brought into physical apposition for the splicing and ligations.[26] It is difficult to comprehend how such complex, dynamic machinery evolved and to such high efficiency.[27]

4.6.1 Internucleotide Phosphodiester Interchange at an Exon–Intron Boundary: Step 1

Each time an intron is excised and two exons are ligated together, two types of nucleotide phosphodiester enzymology are in play (Figure 4.16). In the first step, a *net transesterification* of two phosphodiester bonds occurs. In the second step an upstream free 3′-OH of a 3′-XMP tail becomes ligated to the oligonucleotide at the downstream intron–exon boundary, as the existing 3′–5′ phosphodiester bond at that junction is cleaved. Both reactions are energy neutral transesterifications. One internucleotide bond forms as another breaks. No ATP is required for the transesterifications, albeit an ATP is spent to organize elements of the spliceosome protein–RNA complexes. mRNA splicing is clear evidence for RNA autocatalysis of phosphodiester exchanges.[28]

There is a distinction in the regiochemistry of the two ligations. In a pre-mRNA molecule there are no internal free 3′-OH groups on any of the

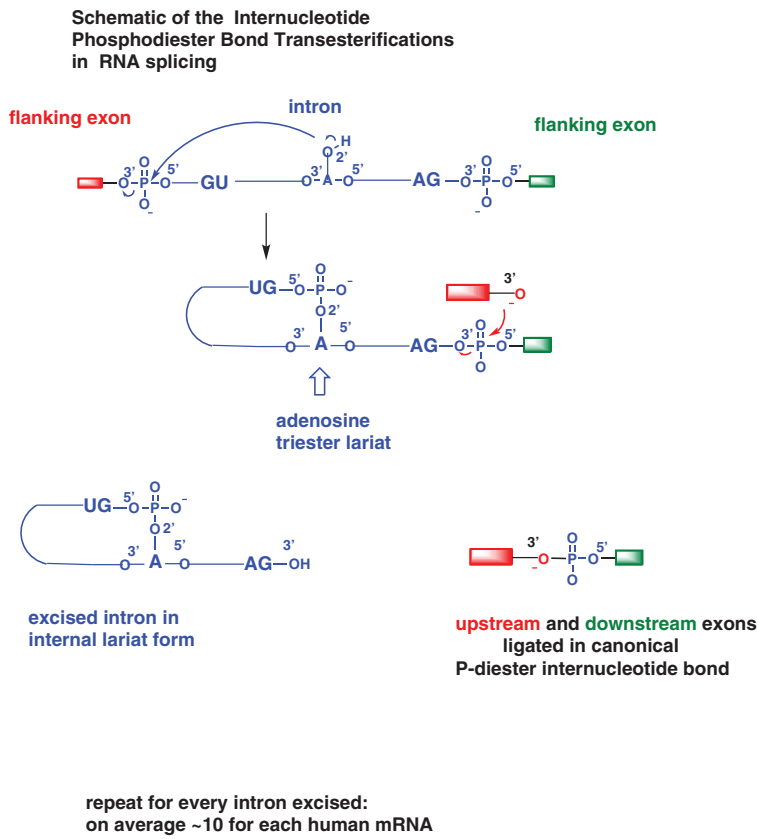


Figure 4.16 Schematic of the internucleotide phosphodiester *transesterifications* in RNA splicing. The first creates a lariat structure. The second is the in-frame ligation of the two exons with release of the intron in lariat form.

ribose units of the oligonucleotides. They are all tied up in covalent 3′, 5′-internucleotide phosphodiester bonds. Because splicing happens in RNA not DNA, there is a *free 2′-OH* on every one of those ribose units in each XMP nucleotide unit. It is in fact the 2′-OH of a specific adenosine 5′-monophosphate residue within an intronic stretch that acts as the initiating nucleophile in the first chemical step of splicing (Figure 4.16). Its target, guided by the snRNAs and the many proteins in spliceosome action, is the 3′,5′ internucleotide bond of a guanyl reside at the upstream intron–exon border, That G is part of a GU pair that together define an upstream-exon–intron boundary. At the other, downstream end of the intron-next exon boundary is an AG dinucleotide pair. Additionally, the nucleophilic A residue that will act as nucleophile in lariat formation is within the intron and upstream of a conserved polypyrimidine stretch.

Initial attack of the 2′-OH on the 5′-phosphodiester connecting the upstream exon to the intron boundary is presumed to follow an associative, phosphorus transition state mechanism. The decomposition in the forward direction is regioselective, as in so many cases of internucleotide cleavage (see Chapter 8), to liberate the free 3′-OH of the upstream oligonucleotide, the precise boundary of that upstream exon. The intron's 5′ terminus has now been captured in a 2′,5′-phosphodiester, linking that G to the attacking A's 2′-OH (Figure 4.17). This unusual regiochemistry, available to RNA but not DNA, is also seen in primary metabolism of NAD^+, where NAD^+ kinase phosphorylates the 2′-OH of NAD^+ to form $NADP^+$.

Inspection of that nucleophilic adenosine moiety reveals that it is a branch point with *three* phosphodiester binds. The two original phosphodiesters link the 5′-OH to the upstream XMP residue and the 3′-OH is likewise part of the normal 3′,5′-phosphodiester to the downstream X'MP residue in the intronic region of the RNA chain. It is the *just formed 2′5′-A to G phosphodiester bond* that is unusual. That is the bond that has led to or enabled cleavage of the pre mRNA chain at the exon–intron boundary. The intron's most upstream GMP moiety now provides the 5′-P in that new 2′,5′-phisphodiester bond (Figure 4.17).

Energetically, the transesterification is neutral, one phosphodiester bond replaced by a different one, albeit with novel regiochemistry. The hard part for spliceosome machinery is spotting every exon–intron junction in the first place *and* creating the RNA conformer that brings the nucleophilic internal A residue close enough for it to use its 2′-OH to attack the exon–intron boundary's internucleotide 3′,5′-phosphodiester bond. For a pre mRNA with a hundred or more introns, it is a logistical nightmare to excise them all correctly.

4.6.2 Internucleotide Phosphodiester Bond Interchange at an Intron–Exon Boundary: Step 2

At this stage the spliceosome machinery is in mid-cycle of removing an intron and ligating upstream exon to downstream exon That second chemical

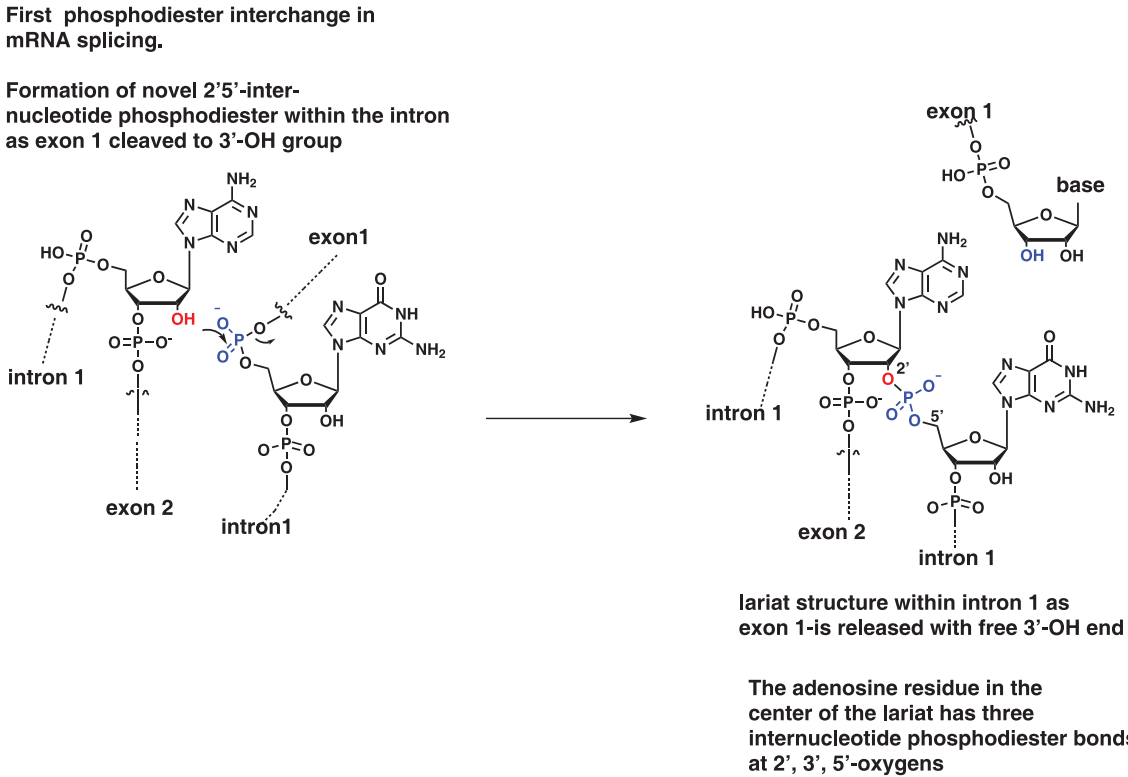


Figure 4.17 The *two phosphodiester interchange steps* in mRNA splicing, step 1. The 2′-OH of an AMP residue within intron 1 attacks a GMP residue at the exon 1–intron 1 boundary. Phosphodiester interchange releases the upstream exon 1 with a free 3′-OH as the unusual 2′,5′-phosphodiester creates a lariat structure in the intron.

step is another directed, highly specific phosphodiester interchange (Figure 4.18). The upstream exon at this juncture has a free 3′-OH. It is somehow guided to the downstream intron–exon boundary (the immediately downstream boundary in canonical splicing; to an alternate downstream intron–exon boundary in exon skipping, that is the process of alternate splicing). With the relevant regions in apposition, the 3′-OH acts as nucleophile at the internucleotide bond at the intron–exon boundary, almost always a pApGpX sequence. As in the first step of phosphodiester interchange, this should be energetically neutral. Again, its decomposition in the forward direction has the same regiochemistry, release of the free 3′-OH group of the upstream GMP moiety.

The result of this second directed phosphodiester interchange is net ligation of the upstream exon to the downstream exon without any change in reading frame. The intron has now been excised by the two phosphodiester exchange steps. The upstream end of the intron still has that adenosine residue with three phosphodiester bonds – a lariat structure, while the downstream end of the excised intron is the typical GMP moiety with a 3′-OH.

The free lariat structure is next hydrolyzed by the spliceosome machinery *via* 2′,5′-phosphodiesterase activity that linearizes the free intron(s) such that all the internucleotide bonds can be returned subsequently to the XMP pool by canonical 3′,5′-phosphodiesterase enzymes (Figure 4.19). Otherwise, tens to hundreds of millions of RNA base pairs would be sequestered in the lariat structures of excised, discarded introns.

These two steps of directed phosphodiester interchange first break, then remake an RNA internucleotide bond, sometimes as close as 30 nucleotides apart, sometimes more than a million bases away. The phosphorus-centered transesterification chemistry is straightforward, but the biological consequences of error are profound in terms of disrupting the coding sequence in the mature mRNAs. Given the ∼240 000 introns in the ∼26 000 human genes, splicing is a highly active process, happening in millions of pre-mRNA molecules in a cell cycle. For example, how the spliceosome machinery finds the two ends of a known 1.04 million base pair intron in a pre mRNA to splice it out is a staggering example of biological specificity and needle finding in a plethora of haystacks.

4.7 Cyclic Nucleotides: Intramolecular and Intermolecular Nucleotidyl Transfers

4.7.1 cAMP and cGMP

One additional set of nucleotidyl transfers occurs on both Mg–ATP and Mg–GTP that constitute *intramolecular* capture of Pα of the triphosphate side chain by the 3′-ribose-OH within the same molecule of ATP or GTP. Each of the cyclases is specific: adenylyl cyclase for ATP[29] and guanylyl cyclase[30] for GTP. The products are the second messenger molecules 3′5′-cyclic AMP (cAMP) and 3′,5′cyclic GMP (cGMP) (Figures 4.20 and 4.21). Among their

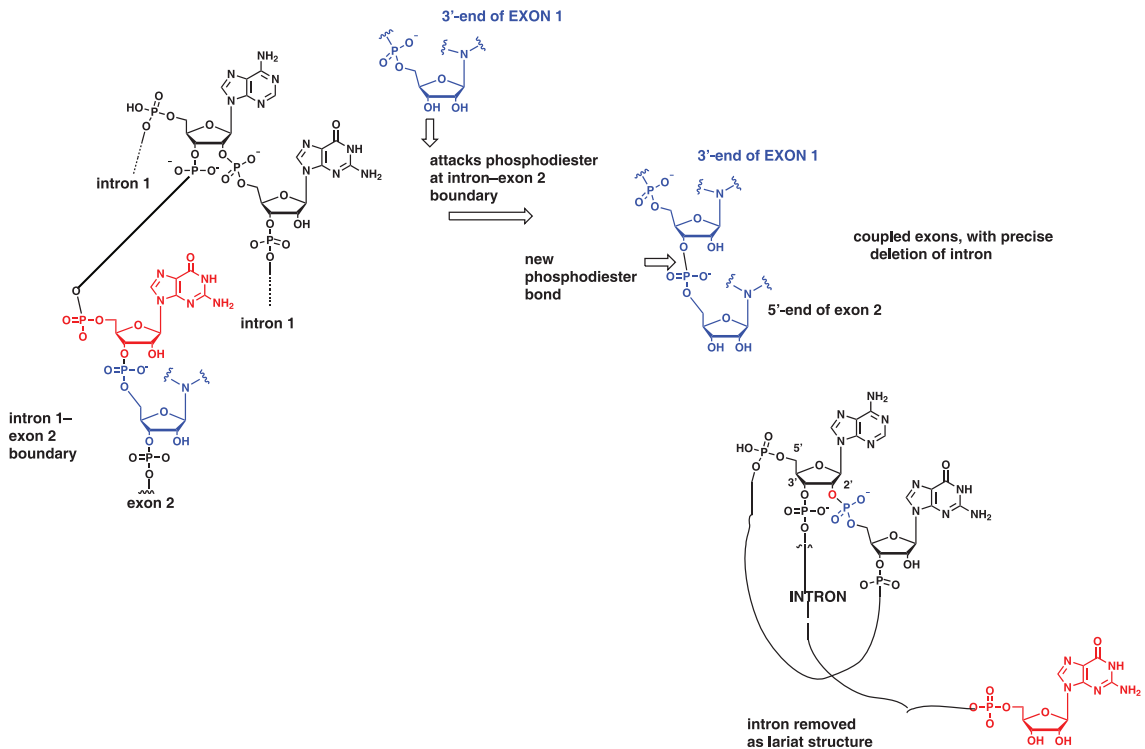


Figure 4.18 The *two phosphodiester interchange steps* in mRNA splicing, step 2. The free 3′-OH of exon 1 acts as nucleophile at the AMP residue at the intron 1–exon 2 boundary. This ligates exon 1 to exon 2 (splicing achieved) with a new 3′,5′-internucleotide bond as the intron is excised as the lariat structure.

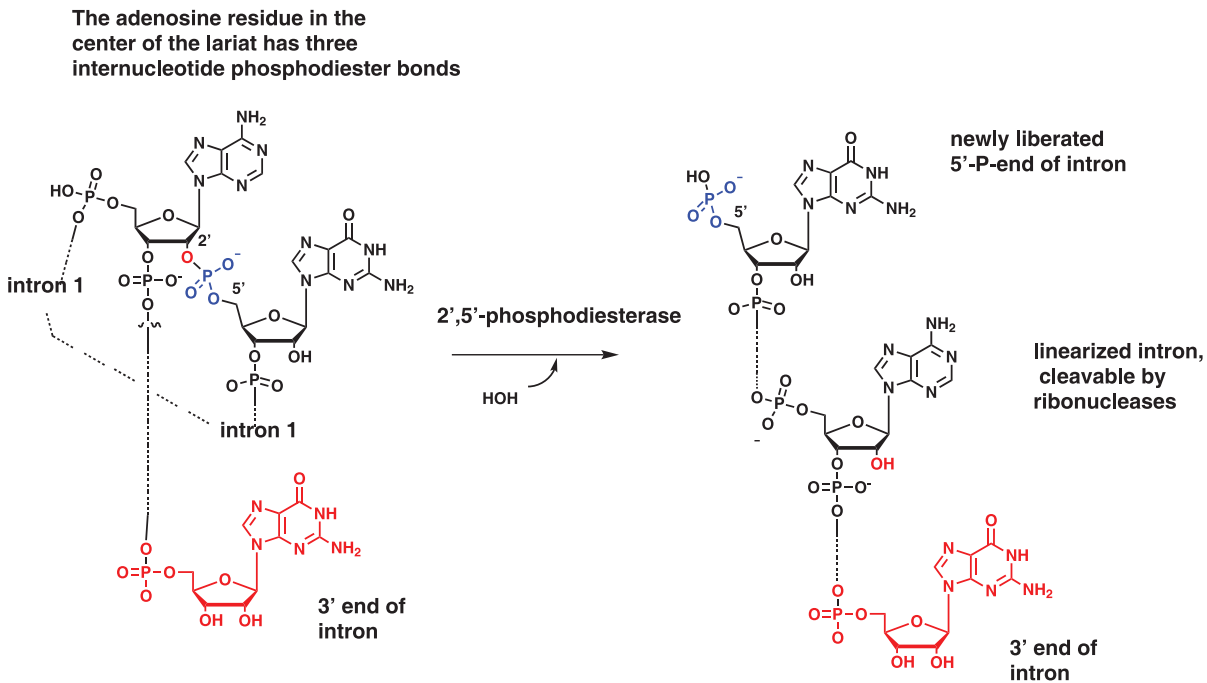


Figure 4.19 The 5′-P lariat structure of excised introns is resolved by hydrolysis *via* unusual 2′,5′-phosphodiesterase action. This creates a canonical 5′-phosphate end in a linear intron scaffold, readily hydrolysable to monomer NMP units by typical RNAses.

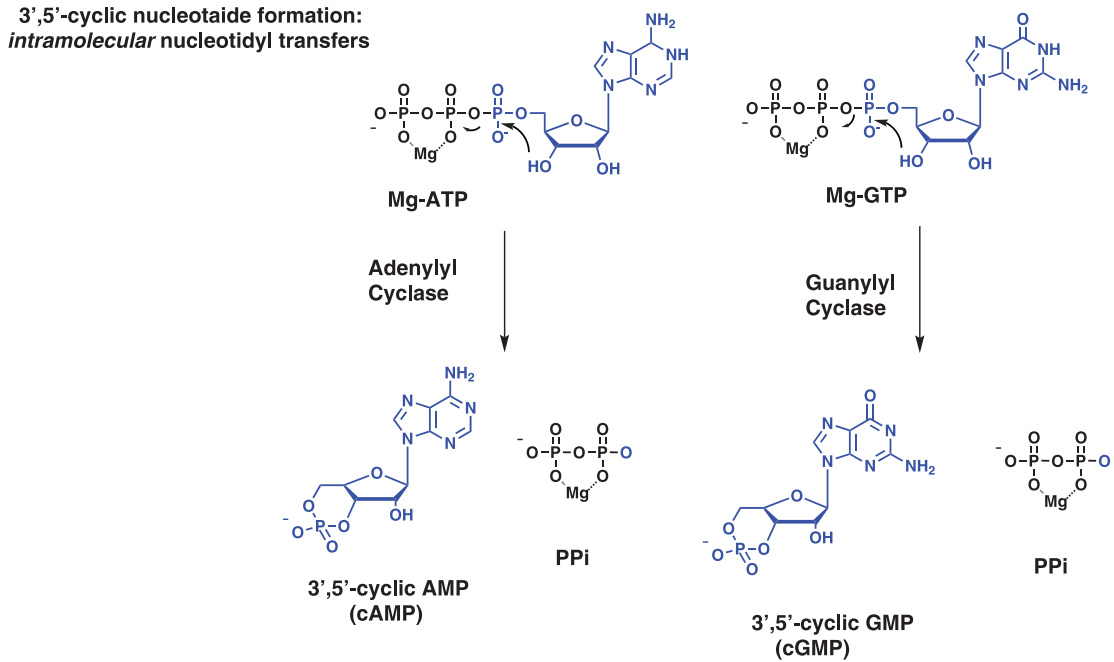


Figure 4.20 Adenylyl and guanylyl cyclases: Intramolecular nucleotidyl transfers as the 3′-OH attacks the side chain Pα in ATP or GTP.

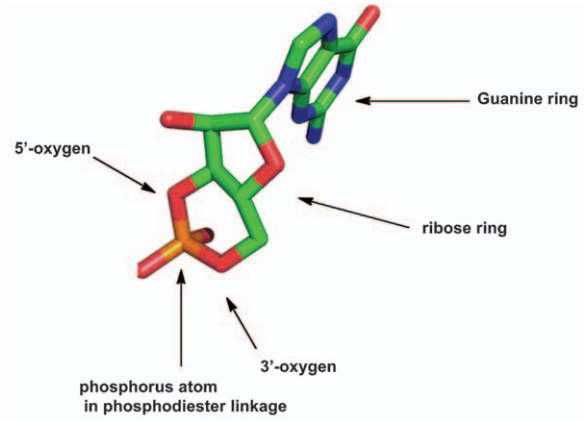


Figure 4.21 Orientation of cGMP bound as activator to protein kinase G (PDB 3OD0).

well-known downstream activities are service as activating ligands for protein kinases, discussed further in Section IV dealing with phosphoproteomics, as all the substrates for the cAMP- and cGMP-activated kinases are protein serine and threonine side chains. 3′,5′-Cyclic AMP is also a ligand for transcription factors and ion channels (as is the congeneric cGMP for ion channels in retinal visual biology).

Some half dozen distinct adenylate cyclases have evolved independently, with ten predicted subforms in humans. All are transmembrane proteins with up to 12 transmembrane helices.[31] The bound conformations of Mg–ATP and Mg–GTP in their respective cyclase active sites must recognize and/or rearrange conformations of the 5′-linear triphosphate side chains such that P_{alpha} becomes relocated within bonding distance of the same molecule's 3′-OH (Figures 4.20 and 4.21).

A now classic signaling role for the cyclase occurs in adrenaline receptor signaling.[31] That receptor interacts on its cytoplasmic face with a trimeric GTPase, inducing dissociation to Gα and the Gβγ dimer. The free Gα subunit activates adenylate cyclase to turn Mg–ATP into cAMP. In turn cAMP binds to the inactive form of protein kinase A and activates it. The fully active PKA molecules then phosphorylate up to 100 cellular protein targets and mediate the activated cell state in response to adrenaline extracellular binding (Figure 4.22).

4.7.2 Di-cyclic GMP and Di-cyclic AMP

A variant of the cyclic nucleotide synthetase strategy occurs when GTP, or less frequently ATP, undergoes both enzymatic dimerization *and* cyclization to form di-cyclic nucleotide scaffolds (Figure 4.23). The GTP cyclization to 3′,3′-dicyclic GMP has been best studied both for mechanism and biological function, first in bacteria and then in yeast.

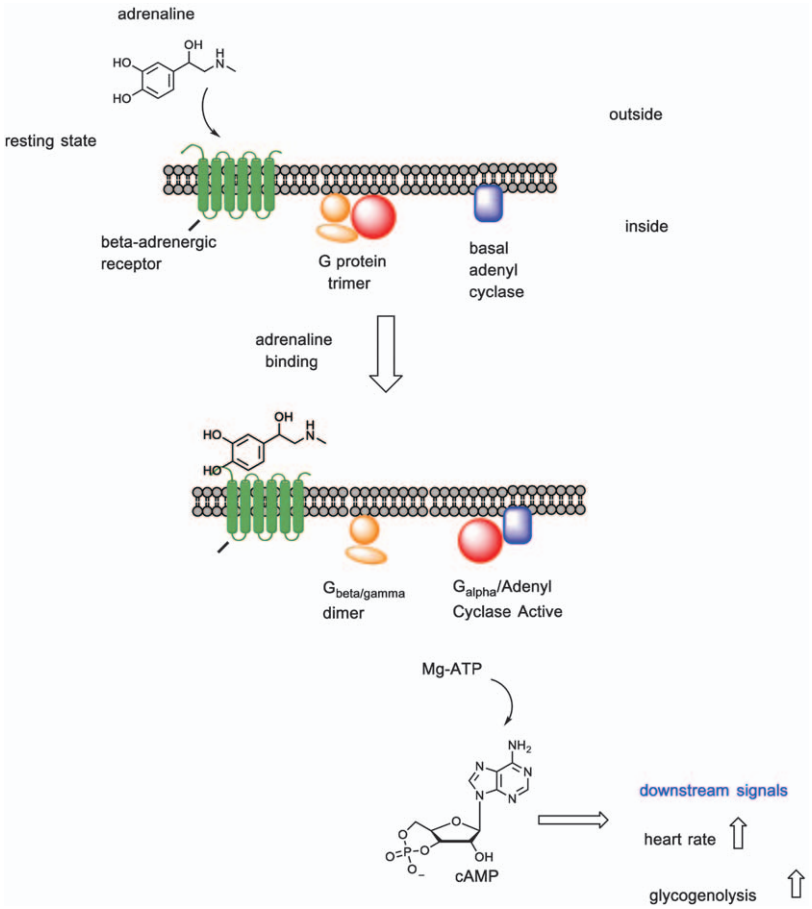


Figure 4.22 Cartoon schematic of signal transfer from occupancy of the adrenergic receptor on the external face of the cell membrane to Gα to adenylyl cyclase to protein kinase A.[32]

The seminal study on di-cyclic nucleotides was reported in 1987[33] when a stimulatory factor for cellulose synthase in the environmental bacterium *Acetobacter xylinum* was tracked down and determined to be a novel dimeric form of a cyclic diguanylic acid. That observation lay fallow, as a one-off finding, until twelve years later when only ten papers on di-cyclic-GMP had appeared. Then the literature exploded. As depicted in Figure 4.24 from Romling and Galperin in 2017,[34] there were a thousand publications by 2015, initially spurred by the finding of key roles for di-cyclic-GMP in biofilm formation by bacterial pathogens. A variety of proteins with high affinity for di-cyclic-GMP binding and subsequent conformational isomerization set off a multitude of bacterial cell signals from transcription to translation, to flagella expression, to biofilm formation (Figure 4.25).

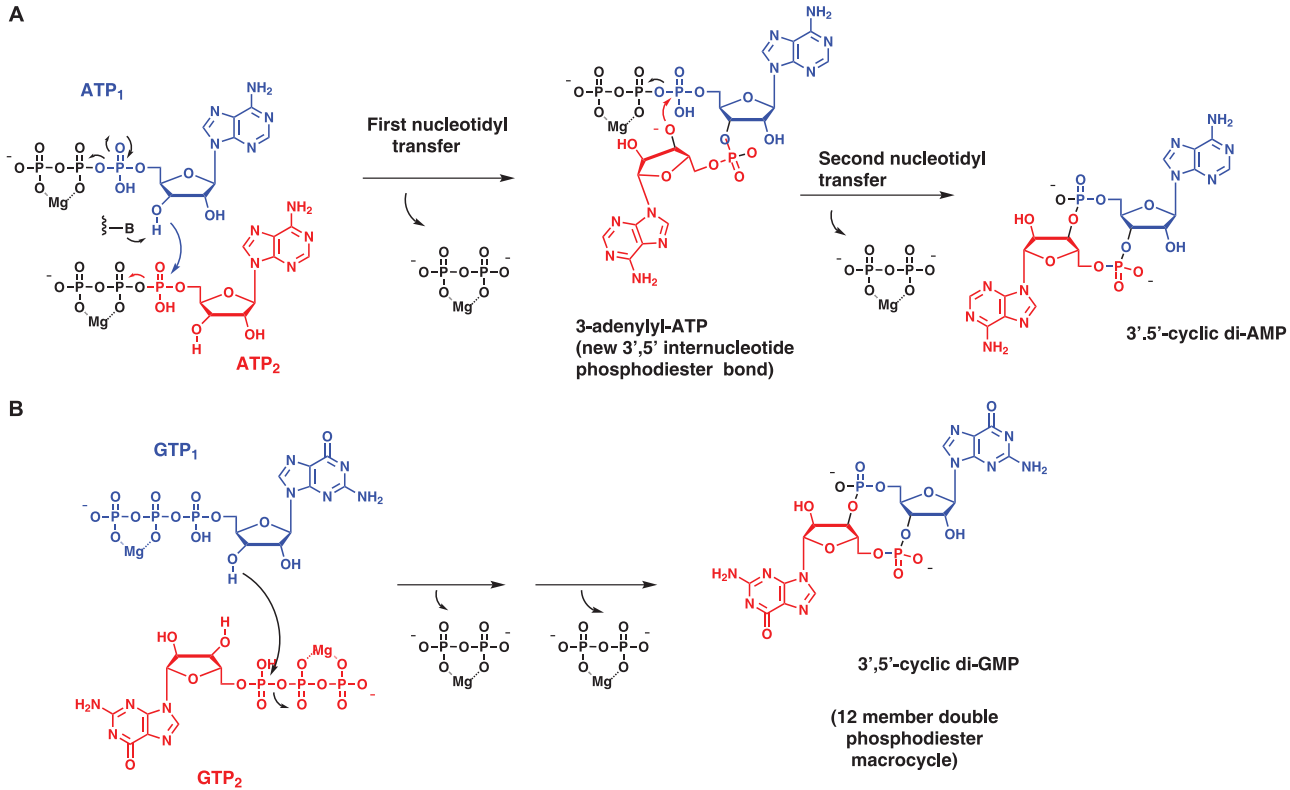


Figure 4.23 Di-cyclic-3′,3′- GMP and di-cyclic-3′,3′-AMP are formed from two molecules of GTP or ATP *via* two consecutive nucleotidyl transfers. The first is *intermolecular*, the second *intramolecular* to create 12-member cyclic-bis-di ester scaffolds. (A) Di-cyclic-AMP; (B) Di-cyclic-GMP; (C) Di-cyclic-UMP generated by other bacterial cyclases.

C

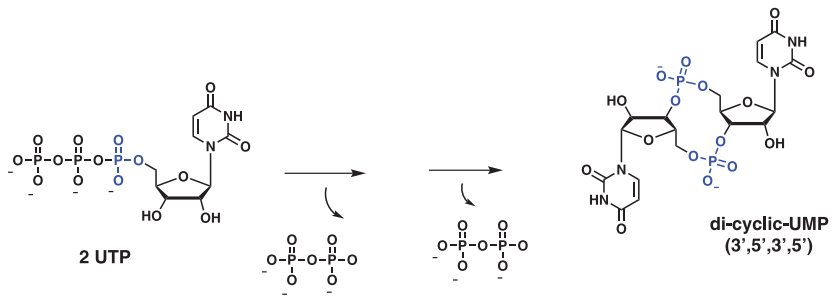


Figure 4.23 (Continued).

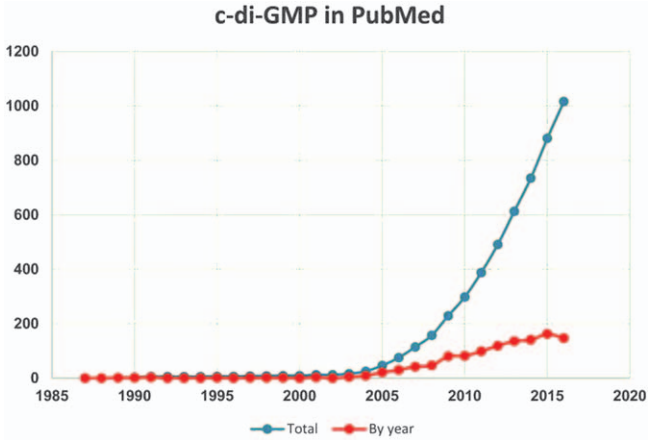


Figure 4.24 Exponential growth of publications on cyclic di-GMP for the decade 2005–2015.

Intriguingly, the GTP-specific and the ATP-specific enzymes catalyze *tandem double nucleotidyl transfers*, one each on the two GTP (or ATP) substrate molecules. The first nucleotidyl transfer is *intermolecular*. The 3′-OH of one Mg–NTP molecule attacking the Pα of the second Mg–NTP molecule (Figure 4.23). In the case of Mg–ATP the indicated linear dimer, 3-adenylyl-ATP is the intermediate, shown in Figure 4.23. Then, the second adenylyl transfer is *intramolecular*, mirroring the step in adenylate cyclase above. The free 3′-OH of the adenylyl moiety of adenylyl-ATP attacks Pα of the ATP side chain. The product is 3′,5′-di-cyclic-AMP. The strategy of intermolecular and then intramolecular nucleotidyl transfers yield a cyclic product scaffold with two phosphodiester linkages in a central 12-member ring, a unique shape for partner protein recognition. Analogously, the GTP cyclase performs intermolecular then intramolecular guanylyl transfers on the two Mg–GTP substrates to give the double phosphodiester macrocyclic core of di-cyclic GMP (Figure 4.23).

Recent bioinformatic studies have found more than 5000 homologs of the di-cyclic-GMP and di-cyclic AMP synthases, including a *Vibrio* bacterial synthase that takes two molecules of Mg–UTP and makes the corresponding di-cyclic-UMP, with the same tandem 3′,5′-phosphodiester bond chemistry and probably the equivalent mechanism for enzymatic assembly. The enzyme will also make cyclic trinucleotides such as c-(AMP–AMP–GMP)[35] (see the following sections). This indicates both that there are more dicyclic nucleotide variants to be found and more as yet unexamined biological signaling roles.

4.7.3 c-GAMP: 2,5, 3′5′-cyclic GMP–AMP

The third iteration of cyclic nucleotide strategy is the most recently described hybrid containing both AMP and GMP moieties, derived from ATP

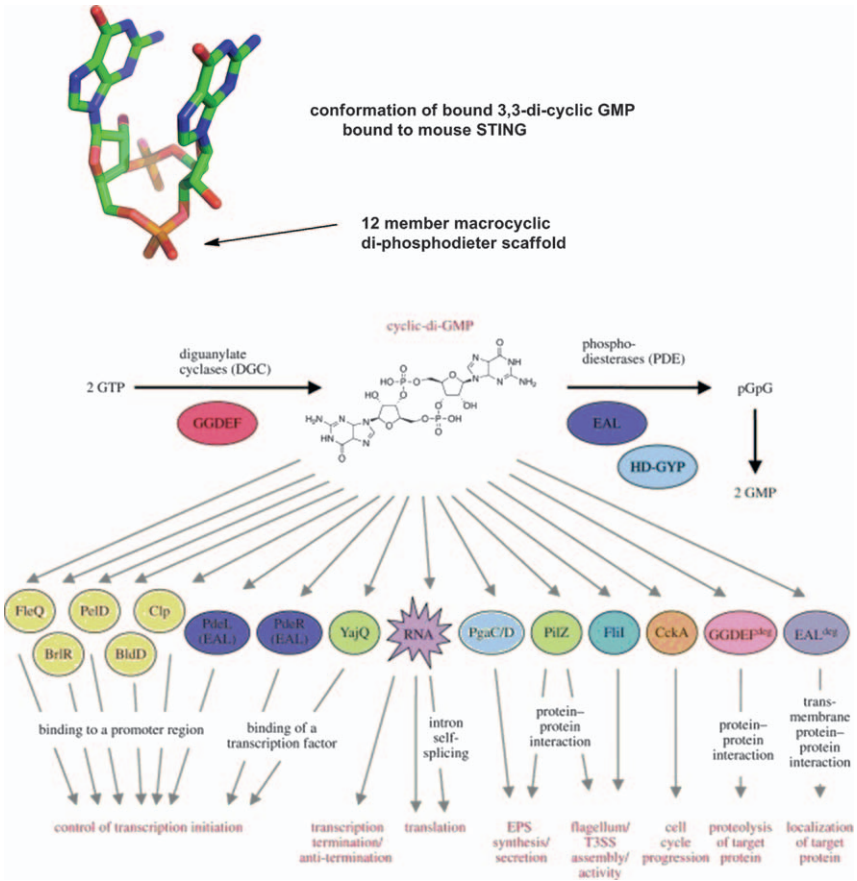


Figure 4.25 Diverse roles of cyclic-di-GMP in bacterial metabolism. Reproduced from ref. 46, https://doi.org/10.1098/rstb.2015.0498, under the terms of a CC BY 4.0 license, https://creativecommons.org/licenses/by/4.0/.

and GTP, respectively, by *intermolecular then intramolecular nucleotidyl transfers* (Figures 4.26–4.29). The enzyme cGMP–AMP synthase (cGAS) shows a distinct regiospecificity from the 3′,5′cyclic dinucleotides in the above subsection.[36] Instead, cGAS activates the *C2′-OH* of Mg–GTP as the first nucleophile, attacking Pα of ATP, *via* intermolecular nucleotidyl transfer. Then, the second, intramolecular nucleotidyl transfer is the conventional 3′-OH of the adenosyl group attacking Pα of the GTP moiety, splitting out the second molecule of inorganic pyrophosphate, PPi. The result is a cyclic G–A dinucleotide with two phosphodiester bonds. However, the regiochemistry is mixed. There is one 2′,5′ phosphodiester and one 3′,5′-phosphodiester linkage. Thus, the macrocycle is a 13-member macrocycle rather than the 12-member macrocycle above. Also, the sidechains are, differentially, a guanine and an adenine purine ring.

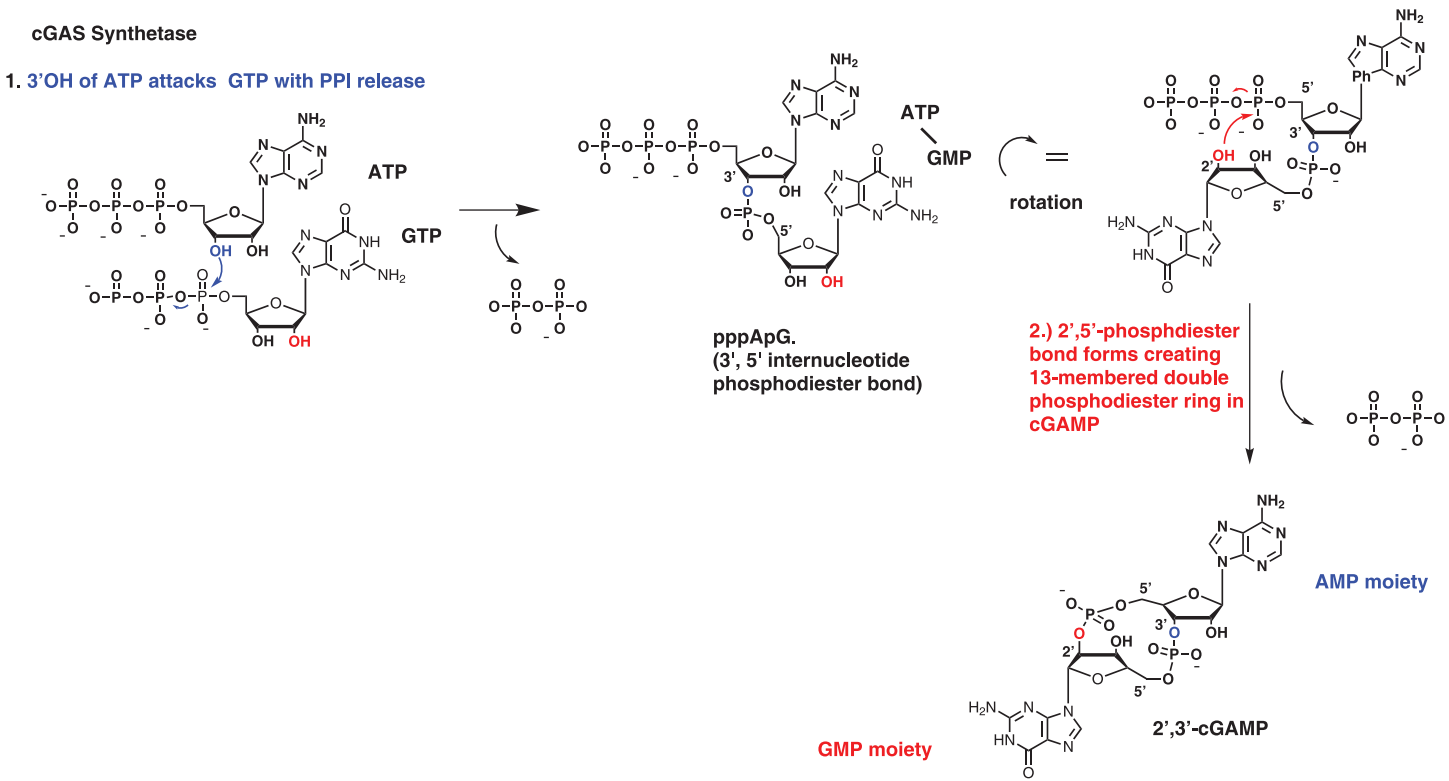


Figure 4.26 Enzymatic formation of cGAMP. The 2′-OH of GTP is initiating nucleophile on Pα of ATP in an intermolecular nucleotidyl transfer. In the second step the 3′-OH of the AMP moiety attacks Pα of the GTP side chain in an intramolecular nucleotidyl transfer. The cyclic diphosphodiester is a 13-member scaffold due to the 2′, 3′ mixed regiochemistry.

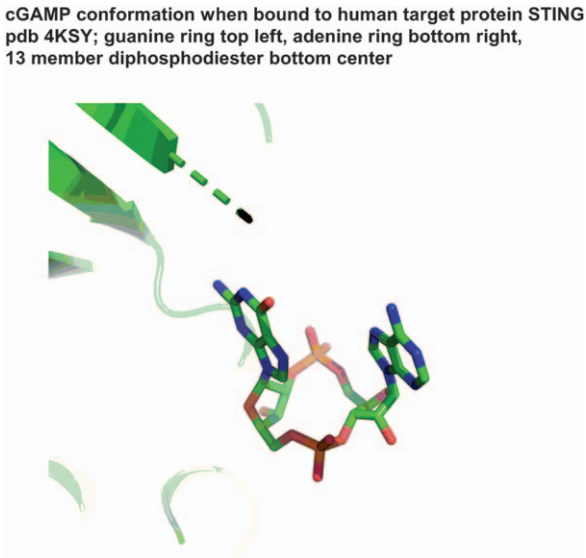


Figure 4.27 Conformation of cGAMP bound as activating ligand to target protein stimulator of interferon genes (STING).

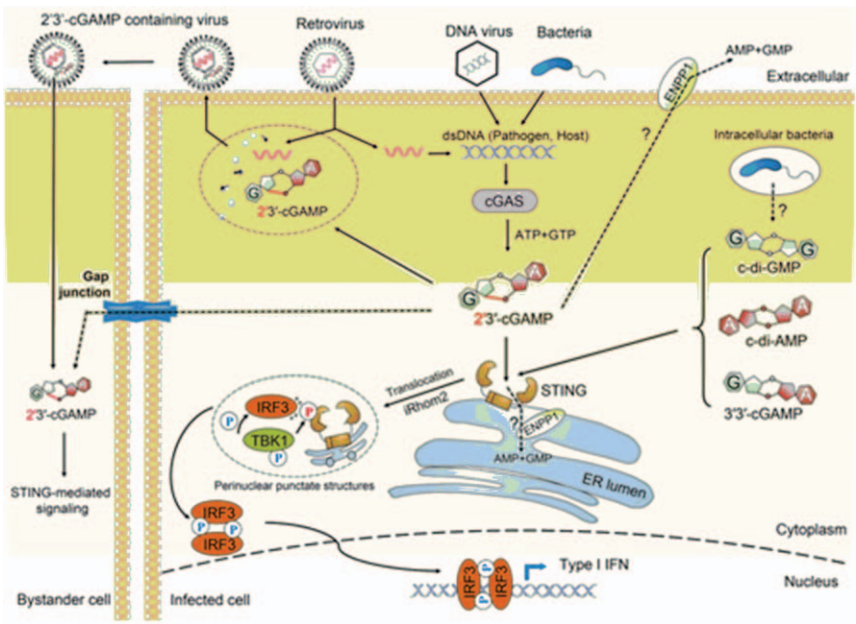


Figure 4.28 Schematic for the cGAMP pathway on viral DNA invasion of host cell cytoplasm; activation of STING leads to antiviral interferon gene expression.

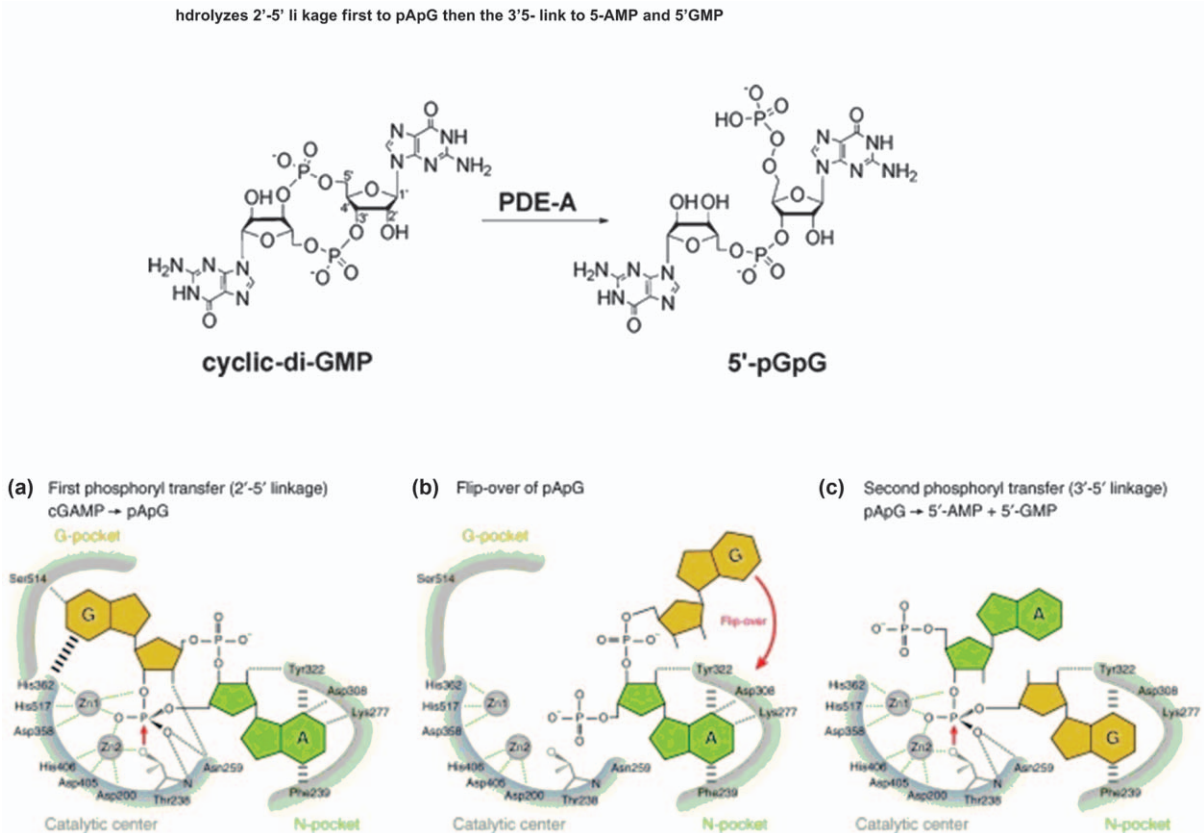


Figure 4.29 cGAMP is turned over by enzymatic hydrolysis by the enzyme ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1), schematized to cut both phosphodiester bonds to release 5′-AMP and 5′-GMP.

In addition to the structural novelty and hybrid dinucleotide composition of 2′, 3′-cGAMP as a novel eukaryotic regulatory metabolite is its role as a stimulatory factor for an arm of human innate immune response when foreign DNA is detected in cell cytoplasm. DNA is normally restricted to the nucleus.[36–40]

As schematized in Figure 4.28, double stranded DNA fragments, of bacterial or viral origin, in mammalian cell cytoplasm encounter the cGAS synthetase. Binding of such dsDNA activates cGAS to perform its double, cyclizing nucleotidyl transfers and elevate the cytoplasmic levels of cGAMP. cGAMP is then an activating ligand for an endoplasmic reticulum protein named stimulator of interferon genes (STING). Activated STING can be translocated to perinuclear vesicles where it activates genes that synthesize type 1 interferons as part of innate immune responses.

4.7.4 Larger Ring Cyclo-oligoadenylates in Type III CRISPR Bacterial Defense Systems

Remarkably, the enzymatic cyclization of nucleoside triphosphates by successive nucleotidyl transfers with final intramolecular capture to a cyclic scaffold is not finished at the level of cyclic trinucleotides noted above for the Vibrio cyclic nucleotidyl transferase.[35] We delay summary discussion of the clusters of regularly interspersed short palindromic repeats (CRISPR) and RNAse bacterial defense systems for hydrolytic interception of invading viral RNAs and DNAs till Chapter 8 in the context of discussion of nucleases, DNA- and RNA-cleaving phosphodiester hydrolases. However, we note here that one of the several types of CRISPR defense systems, type III, has an enzymatic component that makes cyclo-AMP tetramers and hexamers (Figure 4.30) as activators of one of the RNAse degrading catalysts for self-protection.[41,42] The structures of the $c(AMP)_4$ tetramer and $c(AMP)_6$ hexamer indicate the iterated logic of 3′,5′-phosphodiester bonds, reflecting sequential intermolecular nucleotidyl transfers up to linear tetramer ($pppApApApA_{OH}$) and hexamer scaffolds ($pppApApApApApA_{OH}$) before a final, ring closer intramolecular cyclizing enzymatic nucleotidyl transfer. How the tetramer and hexamer cyclase positions the linear chains to bring the ends together for nucleotidyl cyclization is not yet known, nor is the full distribution of products nor maximal size of a cyclic oligomeric NMP. The discovery of larger cyclic oligoadenylates indicates that the second messenger roles for such scaffolds may still to be fully worked out.[41]

4.8 Isoamide Anions as Nucleophiles in Nucleotidyl Transfers

Three enzymes that convert amide carbonyl groups to amidines, by net replacement of C=O by C–NH bonds, utilize the rare isoamide anion oxygen atom as a nucleophile for nucleotidyl transfers. Two of those enzymes are

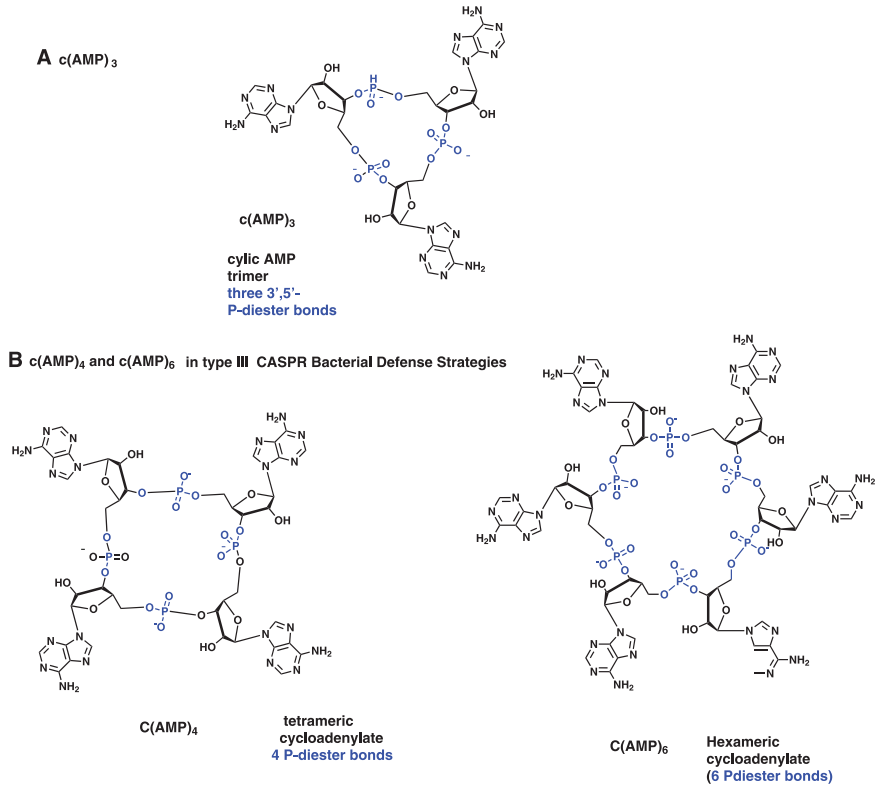


Figure 4.30 Structures of three larger 3′,5′-linked cyclic oligoadenylates. (A) $c(AMP)_3$ from the *Vibrio* dicyclic nucleotide transferase. (B) $c(AMP)_4$ and $c(AMP)_6$ structures generated when type III CRISPR systems sense invading foreign nucleic acids.

involved in pathways from IMP to AMP or to GMP, respectively. The first enzyme is adenylosuccinate synthase, the second is GMP synthase (Figures 4.31–4.33). The third enzyme catalyzes the penultimate step in arginine biosynthesis, generating argininosuccinate.

While the rare isoamide oxyanion forms of the amides are the nucleophiles that attack Pα of Mg–ATP to make the isoamide–phosphodiester links to the AMP moiety, the nitrogen sources for resolution of the tetrahedral adducts differ (Figures 4.32 and 4.33). In the adenylosuccinate and argininosuccinate synthetase cases, the nitrogen donor is the amino group of aspartate, carrying the rest of the four-carbon scaffold with it. Indeed, the next enzymes, adenylosuccinase or argininosuccinylase, are required to eliminate the four-carbon scaffold as fumarate. This tandem enzyme pair synthetase–lyase enables the deposition of aspartate's nitrogen as a primary NH_2 embedded in an amidine group.

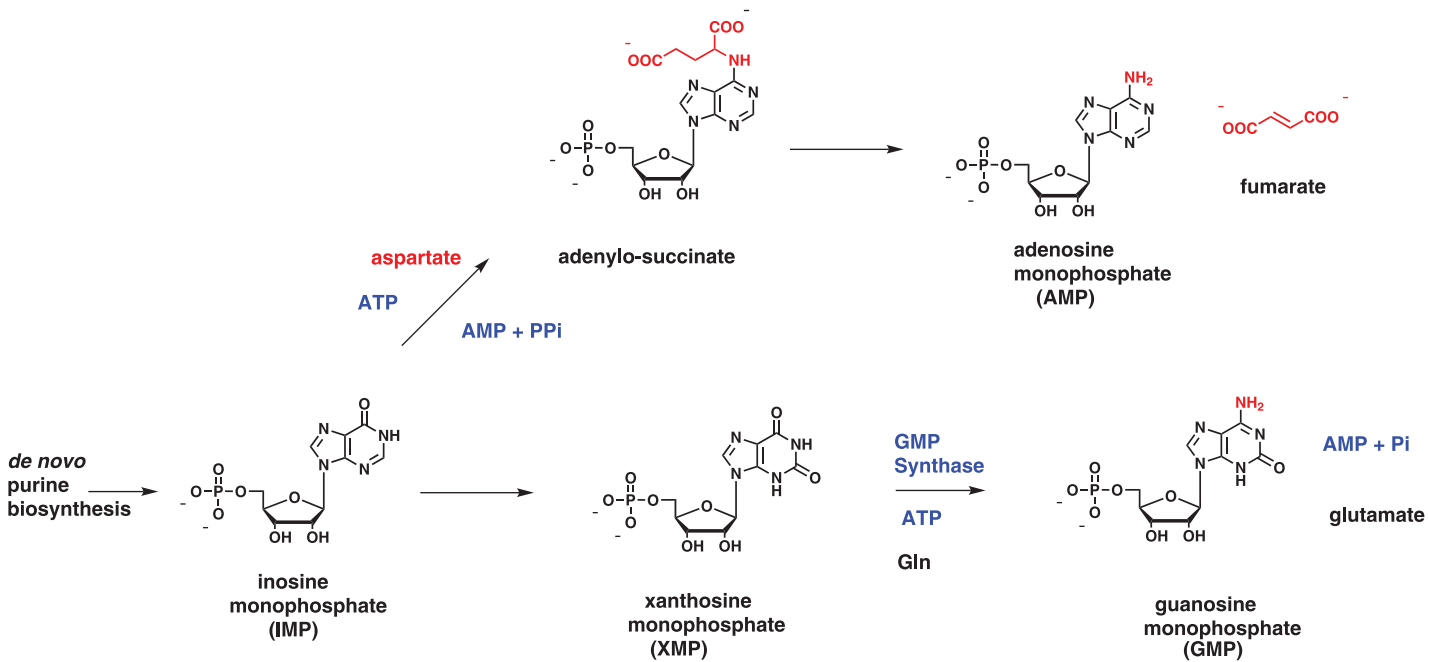


Figure 4.31 Enzymatic conversions of amides to amidines in nucleotide biosynthesis utilize nucleotidyl transfer mechanisms: Overview of branched pathway from IMP to AMP or GMP.

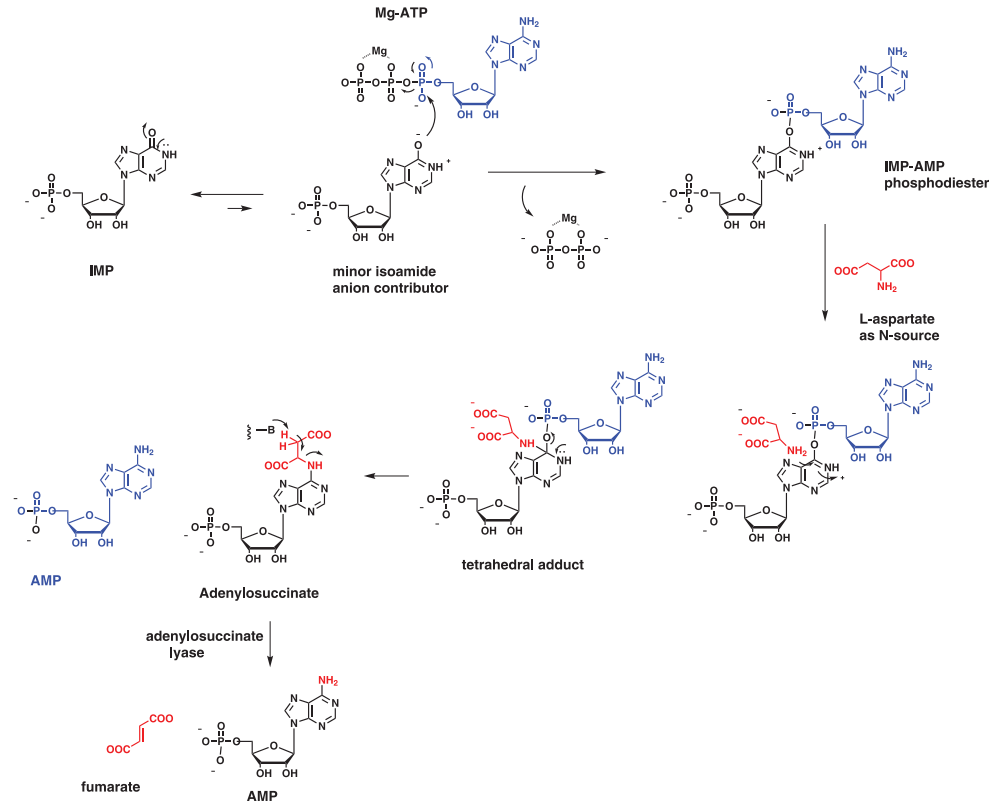


Figure 4.32 Enzymatic conversions of amides to amidines in nucleotide biosynthesis utilize nucleotidyl transfer mechanisms: The exocyclic 6-amino group of AMP derives from the α-amine of aspartate.

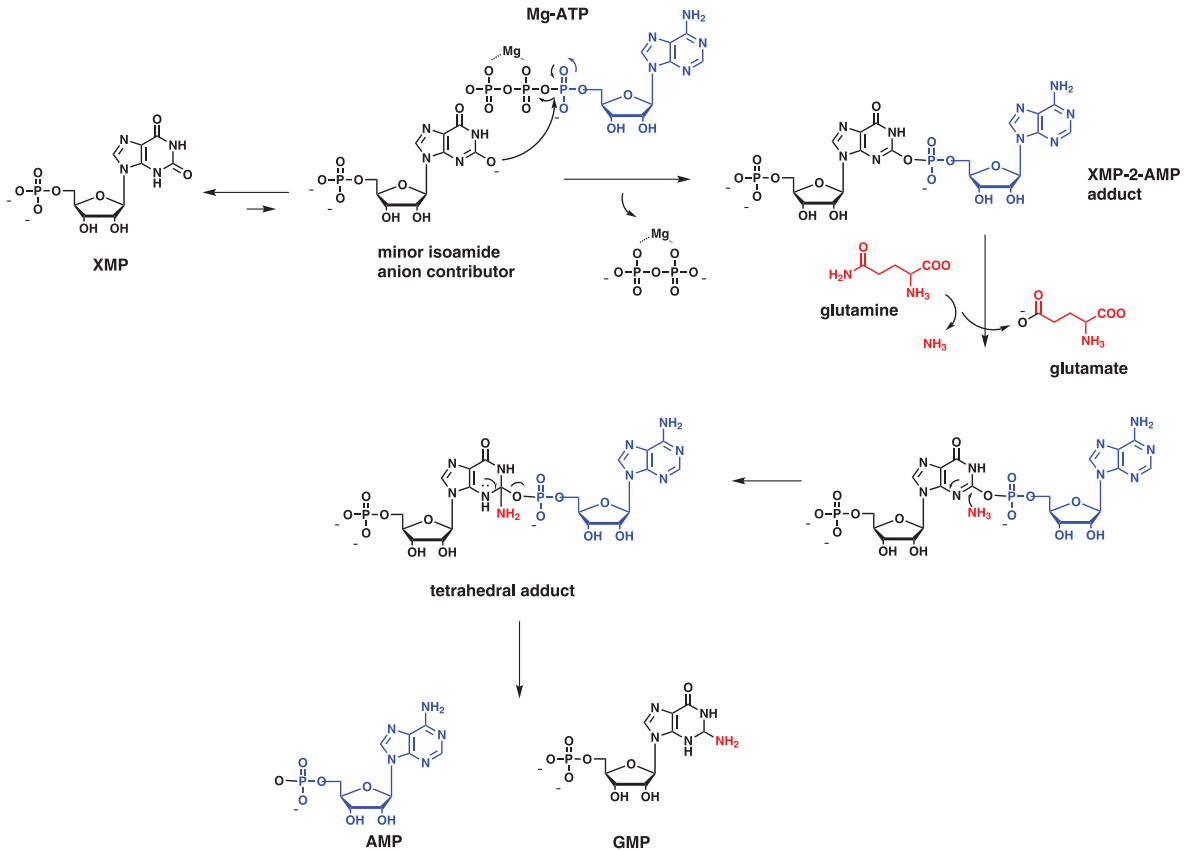


Figure 4.33 Enzymatic conversions of amides to amidines in nucleotide biosynthesis utilize nucleotidyl transfer mechanisms: The corresponding exocyclic 2-amino group of GMP derives from the carboxamide of glutamine.

A simpler scheme would be to use free NH_3 rather than the aspartate amino group, allowing a single enzyme rather than the synthetase–lyase pair to complete the net ketone to amidine conversion. This is indeed the strategy for GMP synthetase (Figure 4.33). This enzyme conscripts the more general strategy for generation of nascent ammonia from *in situ* hydrolysis of glutamine. Glutamine represents a mobile carrier of latent NH_3 in a non-nucleophilic form. Glutaminase activity on demand, in a glutaminase domain active site, generates the nascent NH_3 that gets captured by electrophilic cosubstrate before it can diffuse into free solution and engage in promiscuous reactivity. In the case of GMP synthase, nascent NH_3 is attacking the bound binary XMP–AMP adduct to create the ternary adduct that then decomposes in the forward direction to GMP and the AMP fragment, originally from ATP.

In passing, we note the enzyme converting uridine nucleotides to cytidine nucleotides, completing the pyrimidine pair required for RNA biosynthesis. This is CTP synthetase. The C4-amide carbonyl of UTP conversion to the C4-amidine of CTP uses the same initial strategy as the above three enzymes in trapping the rare isoamide oxyanion as a phosphate ester (Figure 4.34). In this case the isoamide oxyanion attacks the γ-P of Mg–ATP not the α-P: phosphoryl transfer rather than adenylyl transfer. The trapped isoamide is then the monophosphate rather than the AMP phosphate diester. The amine source is not aspartate but rather glutamine, delivering nascent NH_3 by glutaminase action. The resultant tetrahedral adduct, ejects inorganic phosphate on the way to CTP.

4.9 Tandem Phosphoryl and Nucleotidyl Transfers

The two most common modes of Mg–ATP (NTP) cleavage, phosphoryl transfers and nucleotidyl transfers, sometimes occur as tandem enzymatic reactions in pathways to create the common set of disubstituted pyrophosphate backbones in molecules.

One such example is in the conversion of vitamin B_2, riboflavin, into its two active coenzyme forms (Figure 4.35). Phosphoryl transfer at the ribityl 5′-OH of riboflavin by the enzyme flavokinase yields flavin mononucleotide (FMN). This is followed by the action of FAD synthetase, *via* adenylyl transfer, creating the pyrophosphate linkage in flavin adenine dinucleotide (FAD). The same tandem strategy is observed in the coenzyme form of vitamin B_5 (Coenzyme A) (not shown;[43]). A second example in Figure 4.35 of tandem phosphorylation and nucleotidylation is shown for conversion of choline to CDP choline.

4.9.1 Nucleotidyl Transfers in Activation of Sugar Monomers

Nucleoside diphosphosugars are the currency for activated sugar monomers, both for polysaccharide chain assemblies and for protein posttranslational glycosylations.[44] All four ribonucleoside triphosphate precursors of RNA can

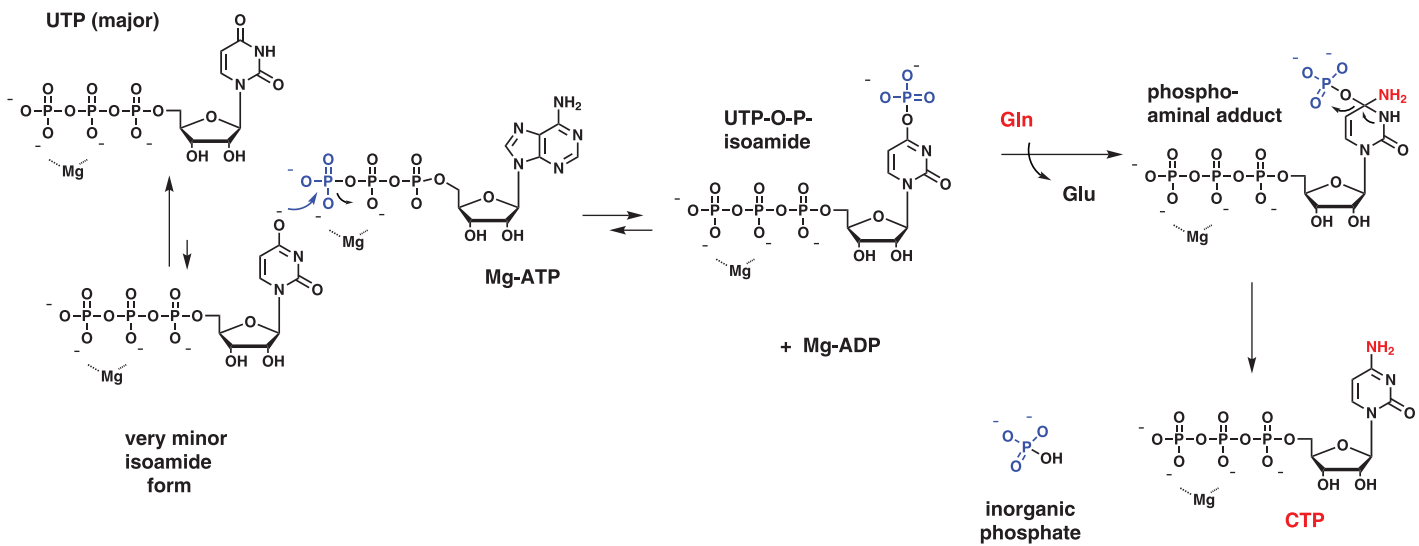


Figure 4.34 CTP synthetase uses the rare isoamide form at C4 of UTP as a nucleophile, directed to attack Pγ of Mg–ATP. The resultant phosphorylate isoamide is captured by nascent NH_3 produced from *in situ* glutamine hydrolysis by the associated glutaminase domain. This is phosphoryl rather than nucleotidyl transfer to set up C_4=O bond cleavage.

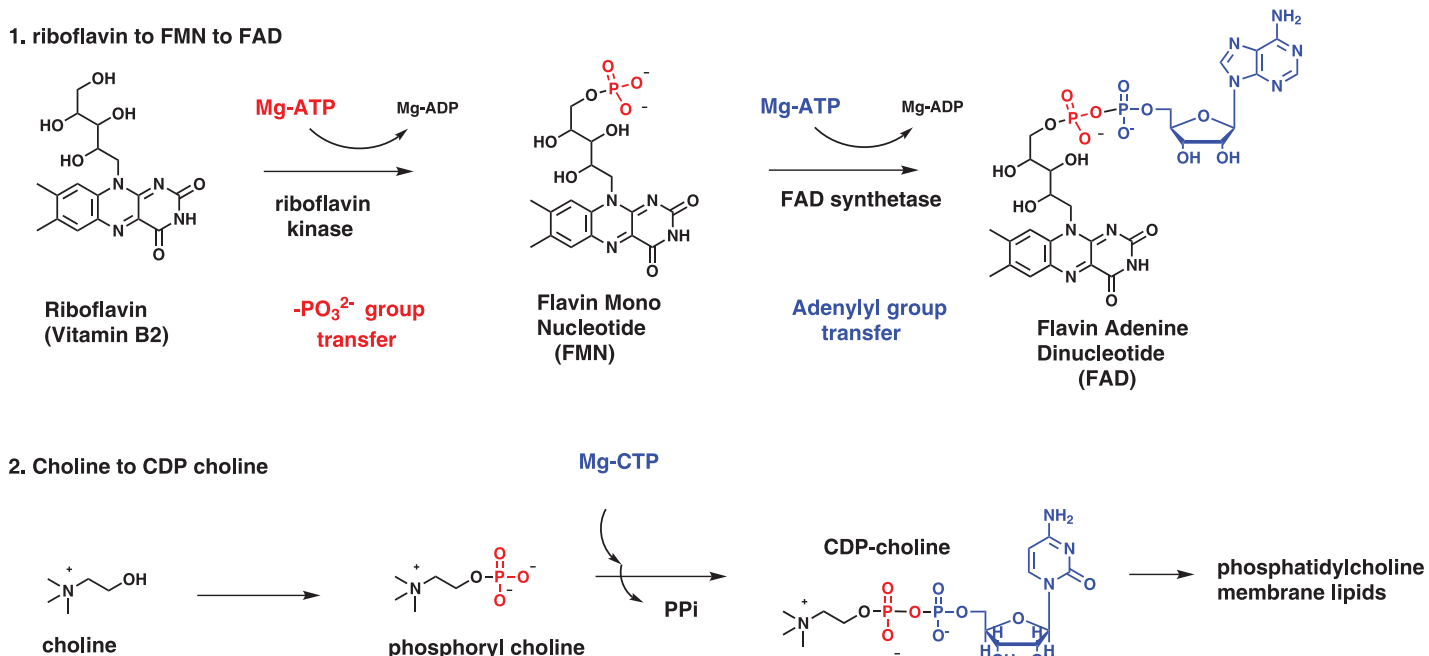


Figure 4.35 Tandem phosphoryl transfers then nucleotidyl transfers are the strategy for a number of key phosphorus-containing metabolite scaffolds. One example is tandem conversion of riboflavin to FAD. A second is the conversion of choline to CDP choline.

serve to activate distinct sugar monomers by nucleotidyl transfers to an oxygen nucleophile. Figures 4.36 and 4.37 show such an array of four nucleoside diphosphosugars. UDP-glucose is the building block for the homoglucans, starch, cellulose and glycogen. On the other hand, plants and microbes use ADP-glucose instead to make amylopectins, amylose in green algae, and bacterial glycogen. In yeast, for mannans, and also as provider of mannose residue in *N*-glycoproteins, GDP-mannose is the preferred nucleoside diphosphosugar donor. These NDP sugars arise from glucose-1-phosphate and mannose-1-phosphate undergoing enzyme-catalyzed attack at Pα of the designated NTP cosubstrates.

A fourth example in Figures 4.36 and 4.37 brings the cytidine nucleotides into the activated sugar monomer arena. In some mammalian glycoproteins the terminal sugar is a nine-carbon scaffold known as sialic acid, or alternatively as *N*-acetylneuraminic acid.[45] In this case the metabolite undergoing nucleotidyl transfer is not a sugar-1-phosphate. Instead, the carboxylate anion of the sialic acid is the attacking nucleophile. The activated sialic acid monomer is then CMP–sialic acid, not CDP–sialic acid. This structure emphasizes that the nucleoside diphosphosugar inventory is really an XMP-fragment linked to a sugar-1-phosphate element.

This is a general strategy for generation of NDP-metabolites and is exemplified further by examination of the biosynthetic origins of cytidine diphosphocholine (CDP-choline) (Figure 4.35). A phosphoryl transferase acts in tandem with a nucleotidyl transferase. Choline is converted by ATP cleavage to O-phosphoryl choline. The oxy anion of that phospho-monoester is directed to attack Pα of CTP, creating the disubstituted pyrophosphate linkage in CDP choline. CDP choline subsequently is condensed enzymatically with diacylgycerols (diglycerides) to release CMP and produce the membrane phospholipid constituent phosphatidyl choline.

4.9.2 Nicotinamide Coenzyme Assembly

A different mixture of phosphorus chemical biology logic is in play in the biosynthesis of the cellular electron transfer currency NAD^+ (Figure 4.38). The logic is hybrid, mixing a phosphoribosyl transfer and then a nucleotidyl transfer, rather than a phosphoryl and then a nucleotidyl transfer, showing some of the mix and match versatility of these phosphoric anhydride building blocks. The first committed step is attack of the heterocyclic nitrogen of nicotinate on C1′ of 5′-phosphoribosyl-1-pyrophosphate. We will note this logic in the next chapter as the first enzymatic step in purine ring assembly to activate C1′ for N–C bond formation with inversion.

The product of this phosphoribosylation is nicotinate mononucleotide that has a permanently quaternized cationic nitrogen in the heterocyclic ring. Then, the adenylyl transferase acts. The nucleophile is the phosphate anion of nicotinate mononucleotide, the electrophile is the Pα of ATP. The product is nicotinate adenine dinucleotide ($NAAD^+$). It is an amination step away from the side chain amide in the NAD^+ coenzyme. That amide is

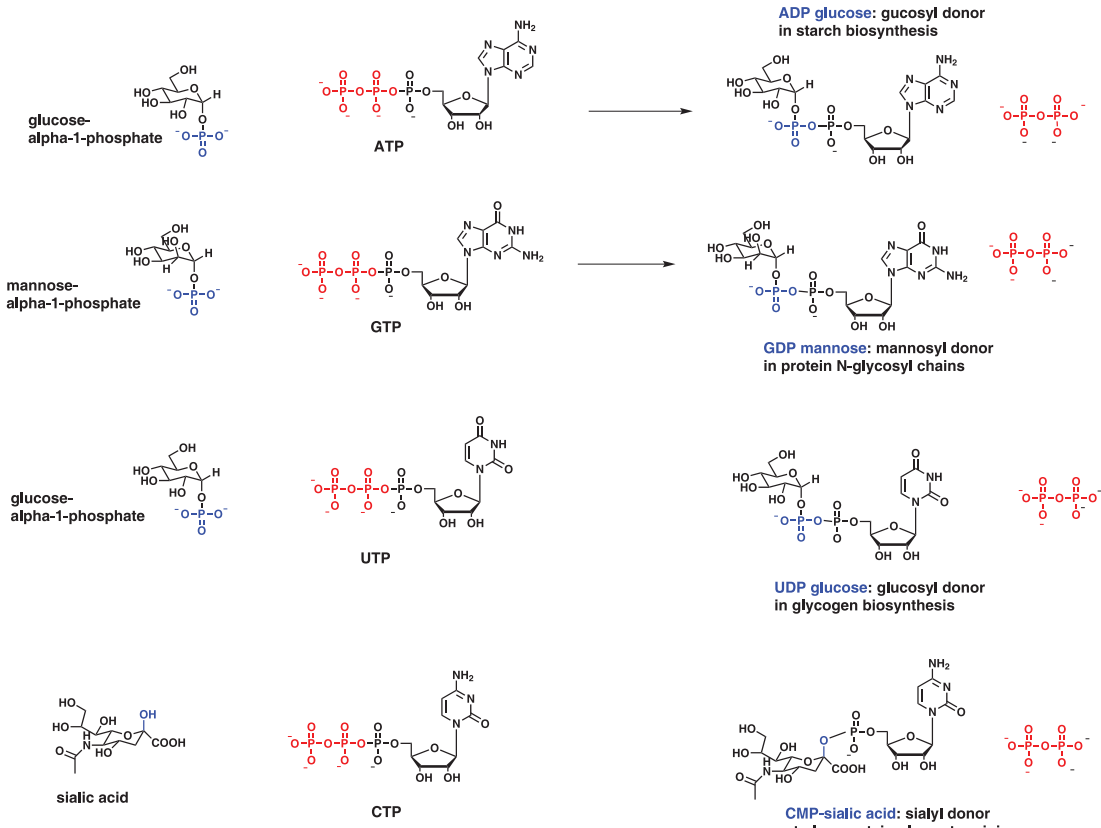


Figure 4.36 All four ribonucleoside triphosphates (building blocks for RNA) are also utilized for nucleotidyl transfers to activate sugar monomers for subsequent biosynthetic glycosyl transfers.

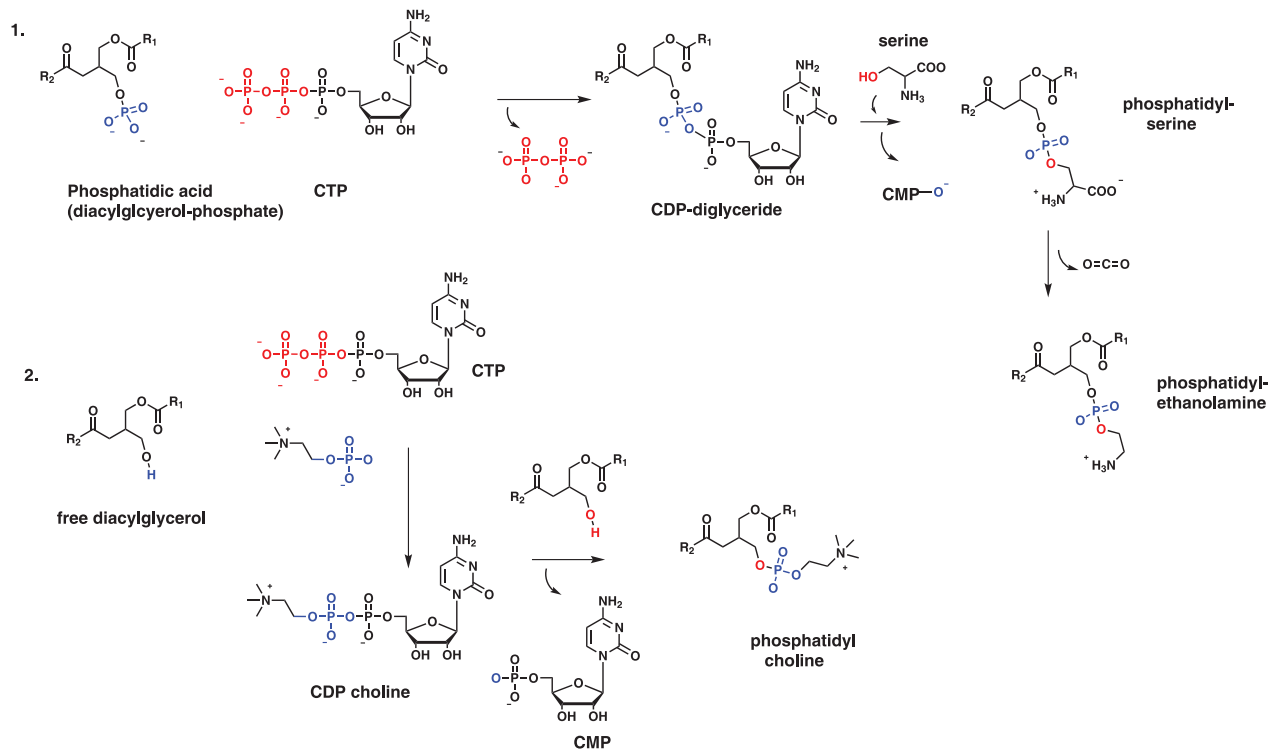


Figure 4.36 (Continued).

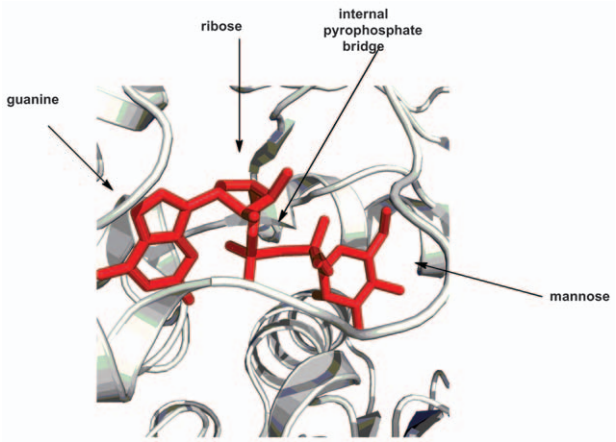


Figure 4.37 Three-dimensional view of GDP mannose bound to a dehydratase enzyme; pdb 6GPK.

installed by the side chain carboxylate of $NAAD^+$ attacking Pγ of another ATP, yielding the transient acyl phosphate shown. NH_3 attack, leads to displacement of Pi and formation of NAD^+.

Although NAD^+ functions in dozens of enzyme-catalyzed reactions as a two-electron sink in the form of NADH and also in both mono- and poly-ADP-ribosylations of proteins, there is one case where the pyrophosphate bridging element [between the nicotinamide mononucleotide (NMN) and AMP halves of NAD^+] is cleaved, in a subset of DNA ligases.[16] Both the NAD^+ versions and the more widespread ATP-dependent DNA ligases, Figure 4.39, transfer an AMP moiety to an active site lysine residue. The coproduct from that nucleotidyl transfer from ATP is the familiar PPi. On the other hand, AMP fragment transfer from NAD^+ releases NMN as the product fragment. It is a clear example that the pyrophosphate bridging element is a thermodynamically activated electrophile whatever the molecular context.

4.10 Summary

Attacks of substrate nucleophilic atoms on Pα in the side chain triphosphates of ATP, GTP, CTP, and UTP (and congener 2′-dNTPs) are pervasive in biosynthetic pathways. Accumulation of dehydrated linear biopolymers – DNA and RNA, proteins, and polysaccharides – is driven by the combined favorable thermodynamic cleavage of the Pα–O-Pβ anhydride linkage in the NTP or dNTP. This is coupled to the paired action of inorganic pyrophosphatase to cleave the P–O–P anhydride bond in the initial PPi product. The net result is cleavages of ATP, NTP, and dNTP into the monophosphates AMP, NMP, and 2′dNMP and two molecules of Pi. Two phosphoric anhydride binds have been

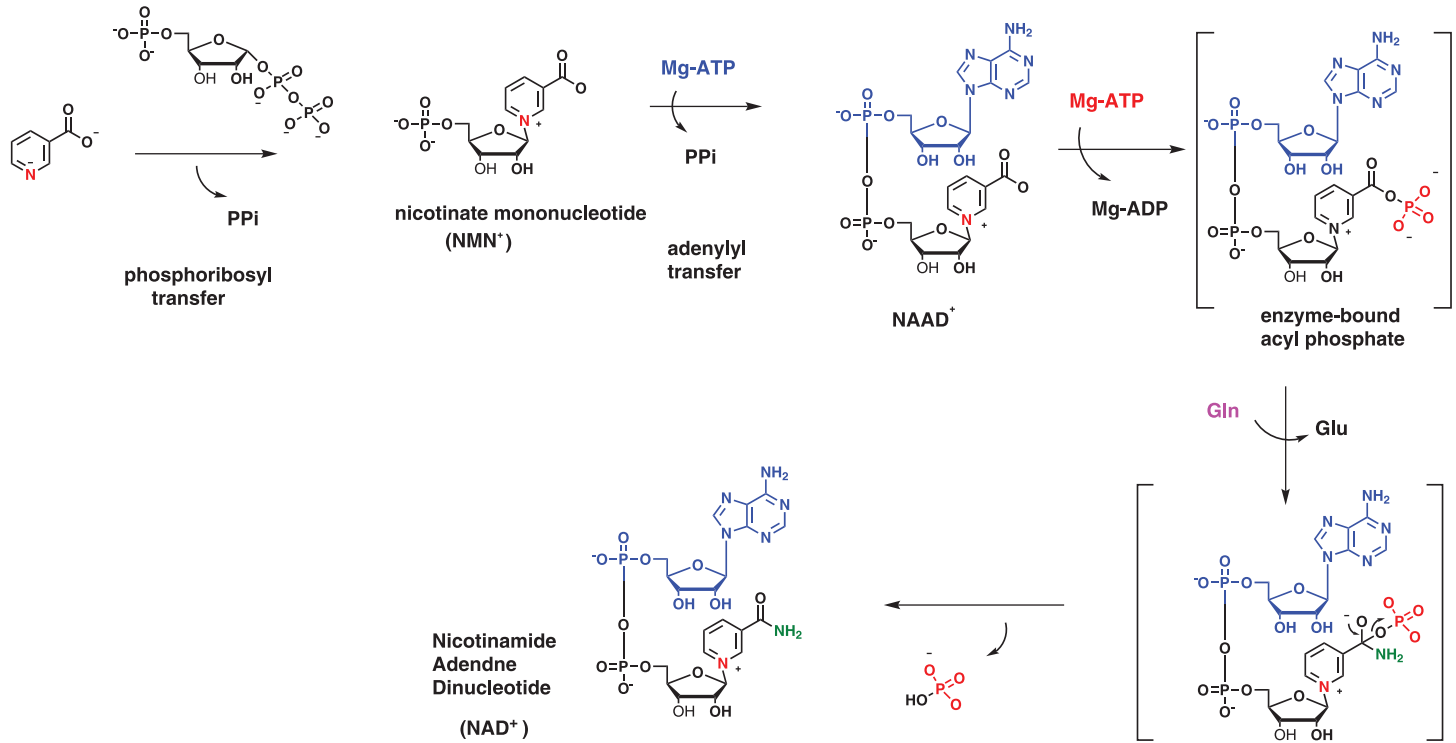


Figure 4.38 The biosynthetic pathway from nicotinate to NAD^+ employs three types of activated phosphorus logic. The first is C1–O bond cleavage of 5-phospho-ribose-1-pyrophosphate by displacement of PPi. The second is adenylyl transfer from quaternized NMN^+ to quaternized $NAAD^+$. The third is a reprise of the isoamide–phosphorylation strategy noted for CTP synthetase as a carboxylate is activated and converted to an amide, generating the end product coenzyme NAD^+.

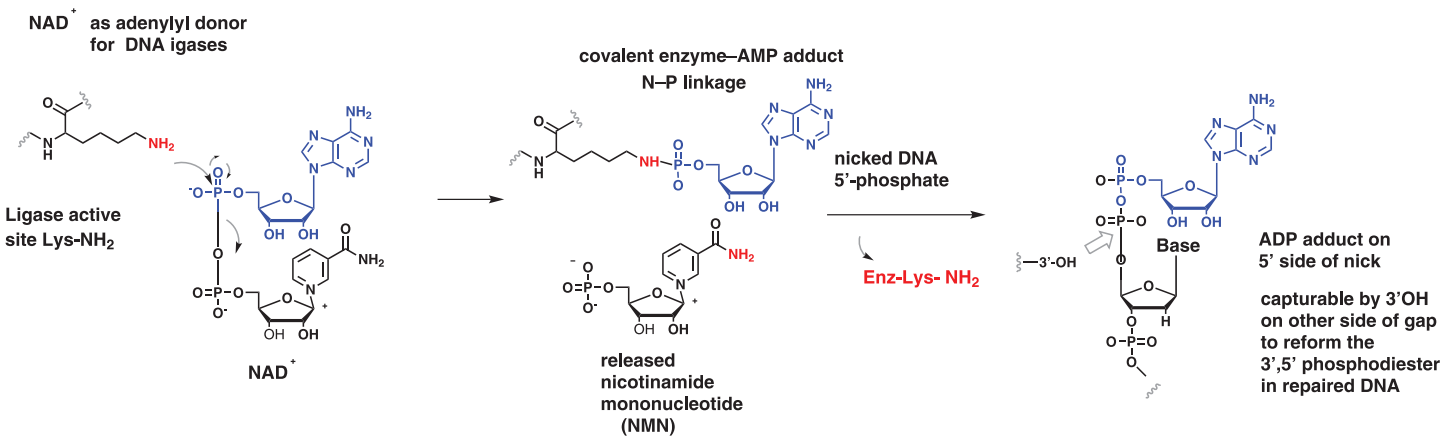


Figure 4.39 NAD^+ as adenylyl group donor in DNA ligase-mediated repair of DNA gaps with a 5′-PO_3^{2-} and a 3′-OH end.

cleaved to drive the entropically disfavored dehydrative accumulation of long chain biopolymers. Adenylyl transfers are also employed in the biosynthesis of many of the coenzymatically active forms of vitamins.

References

1. Y. I. Wolf, A. L. Grishin, N. V. Grishin and E. V. Koonin, Evolution of aminoacyl-tRNA synthetases-analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events, *Genome Res.*, 1999, **9**, 689–710.
2. H. Hartman and T. F. Smith, GTPases and the origin of ribosomes, *Biol. Direct*, 2010, **5**, 1–10.
3. R. Milo and R. Phillips, *Cell Biology by the Numbers*, Taylor & Francis Books: Garland SciencePress, 2015.
4. J. Wu, J. Xiao, Z. Zhang, X. Wang, S. Hu and J. Yu, Ribogenomics: the science and knowledge of RNA, *Genomics, Proteomics, Bioinf.*, 2014, **12**, 57–63.
5. R. Carter and G. Drouin, Structural differentiation of the three eukaryoitc RNA polymerases, *Genomics*, 2009, **94**, 388–396.
6. H. Ehara, T. Yokoyama, H. Shigematsu, S. Yokoyama, M. Shirouzo and S.-i. Sekine, Structure of the complete elongation complex of RNA polymerase II with basal factors, *Science*, 2017, **357**, 921–924.
7. A. Ramanathan, G. B. Robb and S.-H. Chan, mRNA capping: biological functions and applications, *Nucleic Acids Res.*, 2016, **44**, 7511–7526.
8. M. Dreyfus and P. Regnier, The poly (A) tail of mRNAs, *Cell*, 2002, **111**, 611–613.
9. A. L. Jalkanen, S. J. Coleman and J. Wilusz, Determinants and impications of mRNA poly (A) tail size- does this protein make my tail look too big?, *Semin. Cell Dev. Biol.*, 2014, **2014**, 24–32.
10. J. Han, J. Xiong, D. Wang and X.-D. Fu, pre-mRNA splicing: where and when in the nucleus, *Trends Cell Biol.*, 2011, **21**, 336–343.
11. S. Yamashita, Y. Takagi, T. Nagaike and K. Tomita, Crystal structures of U6 snRNA-specific terminal uridylyltransferase, *Nat. Commun.*, 2017, **8**, 1–11.
12. P. B. Balbo and A. Bohm, Mechanism of poly(A) polymerase: structure of the enzyme-Mg-ATP-RNA ternary complex and kinetic analysis, *Structure*, 2007, **15**, 1117–1131.
13. Y. Yashiro and K. Tomita, Function and Regulation of Human Terminal Uridylyltransferases, *Front. Genet.*, 2018, **9**, 538.
14. D. Goodsell, Molecule of the Month: DNA polymerase, in *PDB 101*, 2000.
15. L. Loeb and R. J. Monnat, DNA polymerases and human disease, *Nat. Rev. Genet.*, 2008, **9**, 594–604.
16. Shulman S, DNA ligases; progress and prospects, *J. Biol. Chem.*, 2009, **284**, 17365–17369.
17. N. R. Pannunzio, G. Watanabe and M. R. Lieber, Nonhomologous DNA end joining for repair of double-strand breaks, *J. Biol. Chem.*, 2017, **293**, 10512–10523.

18. U. Grawunder, D. Zimer, S. Fugmann, K. Schwarz and M. R. Lieber, D. N. A. Ligase IV is essential for V(D)J recobination and double strand break repair in human precursor lymphocytes, *Mol. Cell. Proteomics*, 1998, **2**, 477–484.
19. T. Altmann and A. R. Gennery, DNA Ligase IV syndrome: a review, *Orphanet J. Rare Dis.*, 2016, **11**, 1–7.
20. M. R. Lieber, The mechanism of DNA double strand break repair by the nonhomologous DNA end joining pathway, *Ann. Rev. Biochem.*, 2011, **79**, 181–211.
21. S. J. Jeon and K. Ishikawa, A novel ADP-dependent DNA ligase from *Aeropyrum pernix K1*, *FEBS Lett.*, 2003, **550**, 69–73.
22. A. J. Berk, Discovery of RNA splicing and genes in pieces, *Proc. Natl, Acad. Sci. U. S. A.*, 2016, **113**, 801–805.
23. M. K. Sakharkar, V. T. Chow and P. Kanguaene, Distribution of exons and introns in the human genome, *In Silico Biol.*, 2004, **4**, 387–393.
24. A. Piovesan, M. Caracacausi, M. Ricci, P. Strippoli and L. Vitale, Identification of minimal eukaryotic introns through GeneBase, a user-friendly tool for parsing the NCBI gene databank, *DNA Res.*, 2015, **22**, 495–503.
25. A. Nabih, J. Sobotka, M. Z. Wu, C. J. Wedeles and J. M. Claycomb, Examining the intersection between splicing, nuclear export and small RNA pathways, *Biochim. Biophys. Acta*, 2017, **1861**, 2948–2955.
26. G. Matera and Z. Wang, A day in the life of the spliceosome, *Nat. Rev. Mol. Cell Biol.*, 2014, 108–121.
27. A. G. L. Douglas and M. J. A. Wood, RNA splicing: disease and therapy, *Briefings Funct. Genomics*, 2011, **10**, 151–164.
28. S. M. Fica, N. Tuttle, T. Novak, N. S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley and J. Piccirilli, RNA catalyzes nuclear pre-mRNA splicing, *Nature*, 2013, **503**, 229–234.
29. J. H. Hurley, Structure, mechanism and regulation of mammalian adenylyl cyclase, *J. Biol. Chem.*, 1999, **274**, 7599–7602.
30. I. N. Maruyama, dceptor Guanylyl cyclases in sensory processing, *Front. Endocrinol.*, 2017, **7**, 173.
31. D. Cooper, Regulation and organization of adenylyl cyclases, *Biochem. J.*, 2003, **375**, 517–529.
32. K. Rogers, G protein-coupled receptors, in *Encyclopedia Britannica*, 2012.
33. P. Ross, H. Weinhouse, Y. Aloni, D. Michaeli, P. Weinberger-Ohana, R. Mayer, E. de Vroom, G. A. van der Marel, J. H. van Boom and M. Benziman, Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid, *Nature*, 1987, **325**, 279–281.
34. U. Romling and M. Y. Galperin, Discovery of the second messenger cyclic Di-GMP, *Methods Mol. Biol.*, 2017, **1657**, 1–8.
35. A. T. Whiteley, J. B. Eaglesham, C. C. de Olivera Marnn, B. Morehouse, B. Lowey, E. A. Nieminen, O. Danilchanka, D. S. King, A. S. Y. Lee, J. J. Mekalano and P. J. Kranusch, Bacterial cGAS-like enzymes synthesize diverse nucleotide signals, *Nature*, 2019, **567**, 194–199.

36. L. Sun, J. Wu, F. Du, X. Chen and Z. J. Chen, Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway, *Science*, 2012, **339**, 786–791.
37. J. Wu, L. Sun, X. Chen, F. Du, H. Shi, C. Chen and Z. J. Chen, Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA, *Science*, 2012, **339**, 826–830.
38. D. Gao, J. Wu, Y. T. Wu, F. Du, C. Aroh, N. Yan, L. Sun and Z. J. Chen, Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses, *Science*, 2013, **341**, 903–906.
39. P. Gao, M. Ascano, Y. Wu, W. Barchet, B. L. Gaffney, T. Zillinger, A. A. Serganov, Y. Liu, R. A. Jones, G. Hartmann, T. Tuschl and D. Patel, Cyclic[G(2′5′)pA(3′5′)p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase, *Cell*, 2013, **153**, 1094–1107.
40. X. Zhang, H. Shi, J. Wu, X. Zhang, L. Sun, C. Chen and Z. J. Chen, Cyclic GMP-AMP containing mixed phosphodiester linkage is an endogenous high affinity ligand for STING, *Mol. Cell*, 2013, **51**, 226–235.
41. O. Niewoehner, C. Garcia-Duval, J. T. Rostol, C. Berk, F. Schwede, L. Bigler, J. Hall, Marraffini and M. Jinek, Type III Crispr-Cas systems produce cyclic oligoadenylate second messengers, *Nature*, 2017, **548**, 543–548.
42. C. Rouillon, J. S. Athukoralage, S. Graham, S. Gruscgow and M. F. White, Control of oligoadenylate synthesis in a type III CRISR system, *eLife*, 2018, **7**, e36734.
43. C. T. Walsh and Y. Tang, *The Chemical Biology of Human Vitamins*, Royal Society of Chemistry,Cambridge, 2018.
44. C. T. Walsh, B. Tu and Y. Tang, Eight kinetically stable but thermodynamically activated molecules that power cell metabolism, *Chem. Rev.*, 2018, **118**, 1460–1494.
45. N. Sprenger and P. I. Duncan, Sialic Acid Utilization, *Adv. Nutr.*, 2012, **3**, 392S–397S.
46. R. Hengge, Trigger phosphodiesterases as a novel class of c-di-GMP effector proteins, *Philos. Trans. R. Soc., B*, 2016, **371**(1707), 20150498.

CHAPTER 5

Pyrophosphoryl and Adenosyl Transfers from Mg–ATP

All told, there are five enzyme-mediated fragmentation patterns of the ATP side chain (Figure 5.1). In Chapters 3 and 4 we have noted that phosphoryl transfers and adenylyl (generically nucleotidyl) transfers make up the vast majority of the hundreds of known enzyme catalyzed reactions that harvest the P–O–P phosphoric anhydride bond energy, locked in the kinetically stable triphosphate chain of ATP. One other mode of the triphosphate side chain cleavage is termed *pyrophosphoryl transfer* (by attack of cosubstrate nucleophile at Pβ of ATP) and is the subject of this chapter.

The other two modes of ATP fragmentation occur at C5′ of the ribose ring or at C1′ of the ribose ring. They are exceedingly rare and do not involve phosphoric anhydride bond cleavages. We will note in subsequent chapters that the two known examples of adenosyl transfers (attack of nucleophile on the 5′ CH_2 of the adenosine moiety's ribose group), albeit close to unique chemistry, are consequential both in terms of the chemical biology available to and experienced by ATP and for entry into certain metabolic regimes. The fifth fragmentation mode is involved in C–P bond cleavage as described in Chapter 11.

5.1 Pyrophosphoryl Transfers

Three pyrophosphoryl transferases operate in quite diverse biological realms (Figures 5.2 and 5.3). Ribose-5-phosphate pyrophosphokinase provides a central metabolite, 5′-phosphoribose-1-pyrophosphate (PRPP), for subsequent phosphoribosylations in both purine and pyrmidine nucleoside and nucleotide biosynthesis.[1] Those phosphoribosylations set up the nucleoside monophosphate tripartite structures of heterocyclic base, ribose or 2′-deoxyribose, and 5′-phosphoester. These are the key building blocks of both RNA and DNA biopolymers. Furthermore, we have just noted in

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

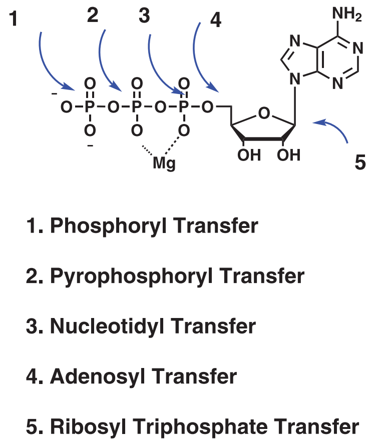


Figure 5.1 Five sites of attack on ATP by specific nucleophilic cosubstrates. Paths 1 and 3 are most common. Path 2, pyrophosphoryl transfers, is the subject of this chapter. Path 4 is adenosyl transfer, also examined in this chapter. Path 5 has one example in the C–P lyase cleavage of carbon–phosphorus bonds.

Chapter 4 Section 4.9 that the most abundant redox coenzymes in all cells, the nicotinamide coenzymes NAD^+–NADH (and thereby $NADP^+$–NADPH), are also constructed by enzymatic phosphoribosylation from PRPP.

5.1.1 5′-Phosphoribose-1-pyrophosphate (PRPP)

The 5′-phosphoribose pyrophosphokinase orients the Mg–ATP cosubstrate and the C1′–OH of the α-anomer population of 5′-phosphoribose such that that alcohol in the cyclic hemiacetal mixture becomes the kinetically competent nucleophile for attack at Pβ of Mg–ATP (Figures 5.2 and 5.3). The transition state resolves in the forward direction with net cleavage of the Pβ–O–Pα anhydride bond. The Keq is $\sim 10^5$ in favor of the forward direction. AMP is the coproduct. Note that the same Pβ–O–Pα bond is cleaved in nucleotidyl transfers so equivalent energetics apply. The distinction is the regiochemistry of which phosphorus atom, Pβ (pyrophosphoryl transfer) or Pα (nucleotidyl = adenylyl transfer) in the triphosphate side chain is attacked by the incoming nucleophile (Figures 5.1–5.3).

The phosphoribose pyrokinase action has converted and trapped the C1′–α-OH of the substrate ribose-5-phosphate hemiacetal as a chemically and configurationally stable *acetal* linkage at the C1′ oxygen, fixing its configuration in the PRPP product. At the transferring pyrophosphoryl moiety an unsymmetric, monosubstituted pyrophosphate derivative has been created.

The utility of this C1′–OPP linkage becomes evident in the subsequent metabolic fates of PRPP and emphasizes the *leaving group function of the (intact) pyrophosphate group* rather than cleavage of the embedded P–O–P anhydride. Thus, all three reactions noted in Figure 5.4,

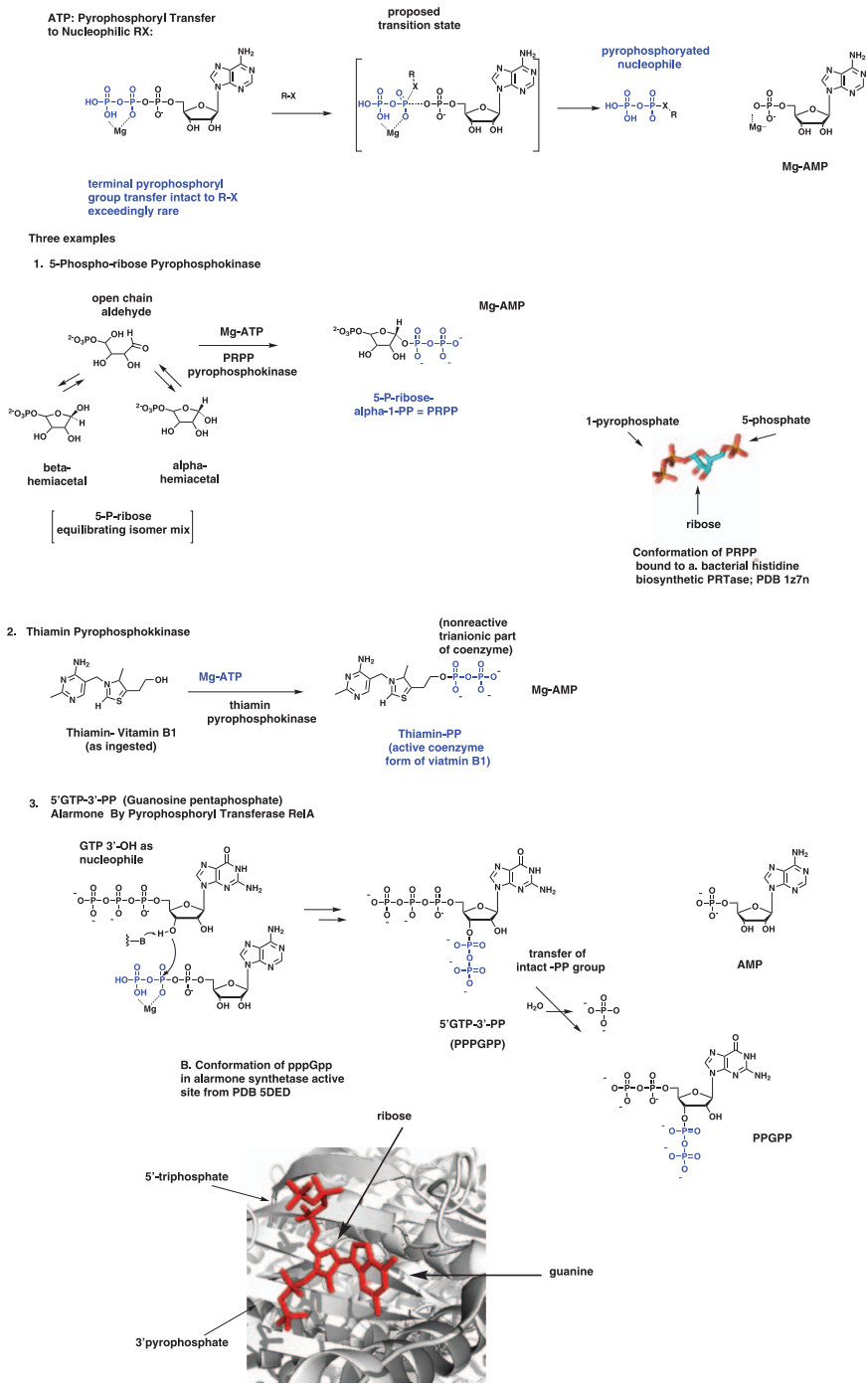


Figure 5.2 Cosubstrate nucleophiles attack Pβ of Mg–ATP for *pyrophosphoryl group* transfers: Examples include 5′-phosphoribose-1-pyrophosphate synthase, thiamin pyrophosphokinase, and bacterial conversion of GTP (5′-pppG) to guanosine pentaphosphate (5′-pppG-3′–pp).

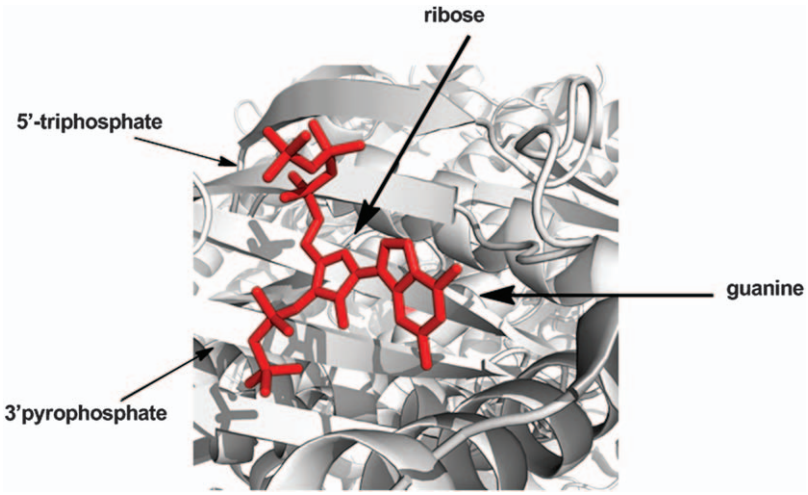


Figure 5.3 Conformation of pppGpp as nascent product in the alarmone synthetase active site (PDB 5DED).

phosphoribosylation of orotic acid to orotidine monophosphate in pyrimidine biosynthesis, amination of PRPP to form the β-regioisomer of 1′-amino-ribose-5′-phosphate at the very start of purine bicyclic ring construction, and substitution at C1′ by an incoming nicotinate in NAD^+ biosynthesis involve formal displacement of inorganic pyrophosphate by C1′–OPP bond cleavage.

The transition state in each of these three biosynthetic phosphoribosylations of amine nucleophiles is thought to be early. Significant C_1′–OPP (C_1–O bond) cleavage produces a stabilized C_1′-oxocarbenium ion-like transition state.[1,2] The partial positive charge at C_1′ is captured by the amine nucleophiles in the net substitution reactions. Looking back two enzymes to the 5′-phosphoribose substrate, cleavage of that underivatized C_1′–OH looks formidable and not likely to be low energy. The strategic derivatization of that nucleophilic oxygen as a pyrophosphoryl ester makes subsequent C1′–O bond cleavage lower in energy. Instead of a formal leaving group of OH^-, the oxygen atom departs as part of tetranionic pyrophosphate (Figure 5.4). The pK_a difference for protonating OH^- (p$K_a = 14$) *vs.* PPi $^{4-}$ to PPi^{3-} (pK_a ~6) is eight powers of ten: the more stable the leaving group anion, the lower the energy.

It is a general biological strategy to convert –C–OH groups to C–$OPO_3{}^{2-}$ groups to lower the energy for C–O bond cleavages. The choice here of conversion of C–OH to a C–$OP_2O_7{}^{3-}$ pyrophosphate monoester is unusual but highly effective in lowering the C–O bond cleavage energy barriers. Once the free PPi is released, inorganic pyrophosphatase action pulls the equilibrium forward by another factor of ~5000 or so. Making purine-, pyrimidine-and nicotinamide-nucleotides are key tasks for spending ''energy rich'' cellular reagents. [One could argue that conversion of hexose hemiacetals in two steps to UDP-sugars as glycosyl donors (Chapter 4) is similar chemical logic for removal of the C1′–OH as a resonance-delocalized PPi derivative.]

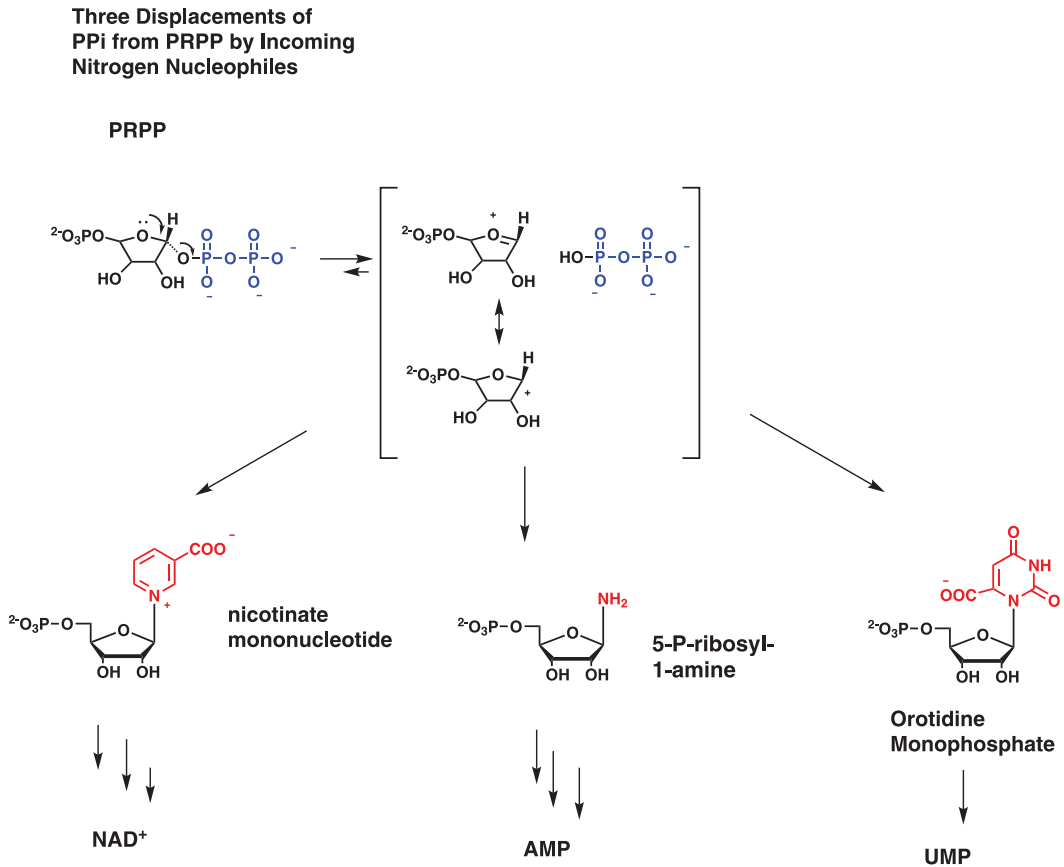


Figure 5.4 PRPP is a substrate for C1′–O bond cleavage in pyrimidine nucleotide assembly, in purine nucleotide assembly and in NAD^+ construction. PRPP thus sits at three central metabolic intersections in nitrogen metabolism.

5.1.2 pppGpp

A second example of enzymatic pyrophosphoryl transfer (Figures 5.2–5.3) involves formation of a third class of signaling nucleotide. We have noted in Chapter 4 Section 4.7 on nucleotidyl transfers that bacteria convert ATP and GTP into two classes of signaling nucleotides: cAMP or cGMP and di-cyclic-GMP (Figures 4.14 and 4.16). They are joined in bacterial metabolism by a third set of signaling nucleotides guanosine pentaphosphate and guanosine tetraphosphate (known as pppGpp and ppGpp respectively).[3] pppG is shorthand for 5′-triphosphate and Gpp is shorthand for 3′diphosphate. Thus, pppGpp has the structure 3′-pyrophosphoryl-GTP (Figures 5.2–5.3) and ppGpp is the corresponding 5′-pyrophosphoryl-guanosine-3′-diphosphate (pyrophosphate).

pppGpp and ppGpp were first detected as ^{32}P-labeled factors (from ^{32}Pi incorporation) on paper chromatograms in the stringent response of bacteria to amino acid starvation and were termed ''magic spots'' long before their structures were determined. These two guanine nucleotides bind to several targets including RNA polymerase and the ribosome to slow down metabolism. As GTP is converted to the penta- and tetra-GTP derivatives, the GTP concentration falls below the levels required for RNA synthesis and cell metabolism shuts down in starvation response.

In contrast to cGMP and dicyclic-GMP biosynthesis that reflect internal nucleotidyl transfers on GTP, pppGpp is assembled by a *pyrophosphory*l transfer (attack at Pβ) from Mg–ATP to the 3′-OH of GTP as a nucleophile (Figures 5.2 and 5.3). The 3′-pyrophophate group interferes with binding to any of the typical GTP targets in cells. ppGpp is then formed by action of a 5′-nucleotidase enzyme acting as a phospho-monoesterase on pppGpp. The lifetimes of the two ''magic spot'' metabolites are controlled by sequential hydrolytic removal of the 3′-PP group by phosphatases.

The attachment of a pyrophosphoryl group at the 3′-OH of GTP is an unusual modification of a nucleotide. Coenzyme A has a 3′-phosphate ester on its AMP moiety (creating a 3′5′-ADP group) and $NADP^+$ has a 2′-phosphate monoester scaffold. The utilization of a 3′-pyrophosphate arising by the rare pyrophosphoryl transfer regiochemistry creates a unique scaffold for the alarmone functions of the magic spot metabolites.

5.1.3 Thiamin-pyrophosphate

The third example of enzymatic pyrophosphoryl transfer from Pβ of Mg–ATP is, in a sense, both more specialized and more prosaic. Vitamin B1, thiamin, is made by microbes, and lower eukaryotes but not by humans.[4] Vitamins in the B series are grouped together because they each fulfill some coenzymatic function. For thiamin that involves both the oxidative α-keto acid decarboxylations, exemplified by pyruvate conversion to acetyl-CoA to fuel the citric acid cycle, and α-ketol transfers in the pentose phosphate cycle that is responsible for moving excess ribose-5-phosphate back to glucose-

6-phosphate. Life without ribose or 2′-deoxyribose nucleic acids is not possible on this planet.

Thus, thiamin is a vitamin essential to health and ultimately to life. However, dietarily ingested thiamin is not in the coenzymatically active form. Instead it is thiamin-pyrophosphate (TPP; Figures 5.2 and 5.3) that is required for coenzymatic function. This is the job of the enzyme thiamin pyrophosphokinase. The one step addition of both phosphorus atoms to the alcohol side chain of thiamin is an efficient strategy.

The pyrophosphate side chain is a bystander in the thiamin coenzyme chemical biology roles noted above. However, it is a key nonreactive element in providing tight binding of thiamin-pyrophosphate to its apoenzyme partners. One might have imagined generating thiamin monophosphate by classical kinase action but thiamin monophosphate has much lower affinity for its enzyme partners.

None of the two features of the pyrophosphate side chain we have noted in other biological contexts are involved in TPP chemical biology. Neither P–O–P cleavage of the thermodynamically activated, kinetically stable phosphoric anhydride linkage, nor elimination of the alcoholic oxygen as part of a pyrophosphate leaving group are utilized. This appears to be use of the trianionic pyrophosphate side chain merely as a tight binding element, a third attribute of its biological function to provide specificity.

5.2 Tandem Phosphorylations as An Alternative Strategy to Pyrophosphorylation

We note that cells could have made any one of these monosubstituted pyrophosphate metabolites by two tandem kinase actions, alcohol to phosphate monoester and then another phosphoryl transfer to generate the pyrophosphate side chain. Indeed, that is is precisely the strategy executed in the conversion of mevalonate to mevalonate-5-pyrophosphate in the early stages of isoprenoid pathways. As depicted in Figure 5.5 the six-carbon metabolite mevalonate has a carboxylate group and two alcohols derived initially from three acetyl groups. The primary alcohol is the target for tandem kinase action by mevalonate kinase to produce ADP and 5-phosphomevalonate.[5] In the next enzymatic step, a second phosphoryl transfer follows. A phosphate oxyanion of mevalonate-5-phosphate attacks Mg–ATP at the γ-PO_3^{2-} group to yield mevalonate-5-pyrophosphate and the second molecule of ADP as coproduct. A third successive phosphoryl transfer is directed not to the –PP group but to the tertiary alcohol group, setting up the indicated decarboxylation–phosphate elimination to give the Δ^2-isopentenyl diphosphate metabolite.[6]

The downstream isopentenyl diphosphates are then progenitors of allylic carbocationic transition states or intermediates in all the chain elongation reactions that can yield >50 000 known isoprenoid natural product scaffolds.[7] In each case of isoprenoid C–C bond formation or chain elongation,

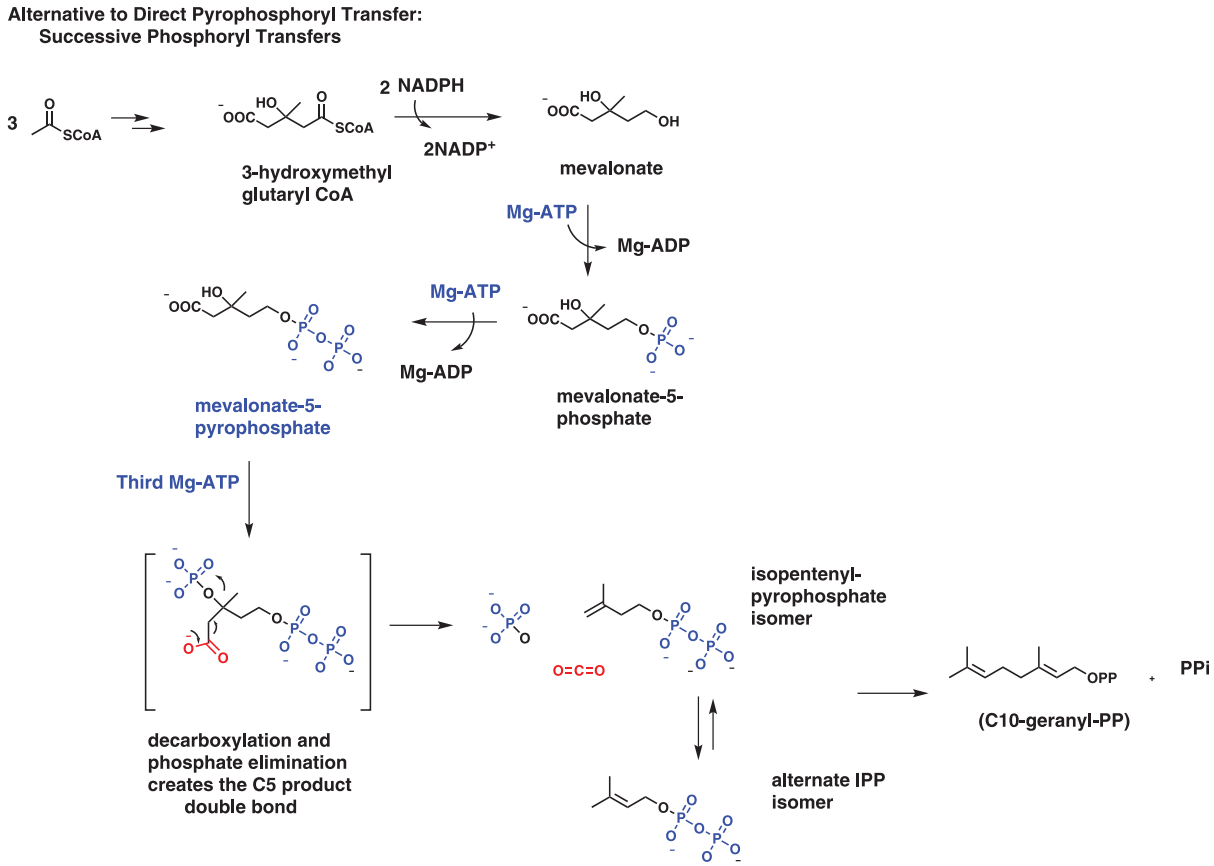


Figure 5.5 Early steps in the isoprenoid pathway: from 3 acetyl CoA molecules to isomeric pentenyl diphosphates involves three consecutive phosphoryl transfers from three Mg–ATPs at the level of the six-carbon metabolite mevalonate. The first two $-PO_3^{2-}$ transfers create mevalonate-pyrophosphate. The third $-PO_3^{2-}$ transfer induces decarboxylation to the five carbon Δ^{3-} isopentenyl-diphosphate.

the C_1–O bond cleavage energy is lowered by prior incorporation of that oxygen as part of the pyrophosphate, in close analogy to the role of the pyrophosphate group in the PRPP reaction strategies.

In a cellular energy accounting sense, tandem phophorylations spend two ATPs to install the substituted pyrophosphate group. Pyrophosphokinases achieve the same end by spending only one ATP molecule. Given the billions of purine and pyrimidine building blocks needed for RNA and DNA synthesis in every cell cycle, (*e.g.* from PRPP), the energy economics may have favored development and maintenance of pyrophosphokinases.

5.3 Adenosyl Transfers

There is a *rare* fourth mode of enzyme-catalyzed fragmentation of ATP that is employed in two well validated cases.[2] This cleavage pattern releases nascent triphosphate, as a cosubstrate nucleophile attacks C5′ of the ribose moiety of Mg-ATP, rather than any of the three phosphorus atoms in the triphosphate side chain. That ribose C5′ carbon is not a notably reactive alcohol but the C5′–O bond cleavage is facilitated by expulsion of the triphosphate hexaanion rather than OH^- (Figure 5.6A).

The two nucleophiles that attack ATP in these unusual adenosyl transfers are the sulfur atom in thioether linkage of the amino acid L-methionine for one, and the Cobalt (I) form of vitamin B12 for the second (Figure 5.6B).

5.3.1 *S*-adenosylmethionine Synthetase

The first enzyme is *S*-adenosylmethionine synthetase, creating a trigonalized sulfur atom in the form of a permanently charged sulfonium ion. The immediate coproduct PPPi is hydrolyzed regiospecifically between Pβ and Pα to yield Pi and PPi fragments that are then released into solution.[8] In the presence of the ubiquitous inorganic pyrophosphatase, the cell converts the triphosphate side chain of ATP into three molecules of Pi. Although *S*-adenosylmethionine (SAM) is uphill energetically from ATP, SAM is accumulated in cells because of the tandem action of the synthetase and pyrophosphatase, cleaving both P–O–P anhydride links in ATP. In terms of relative abundance, where steady state intracellular concentrations of ATP may be in the 5–8 mM range, SAM may be in the 50–180 micromolar range. Given a daily turnover of ~75 kg of ATP in humans, the amount of SAM made and consumed may be in the1–5 kg range.

The trivalent sulfur cation in *S*-adenosylmethionine activates all three of its immediate substituents for transfer as electrophilic fragments to cosubstrate nucleophiles[9] (Figure 5.7). By far the most common S^+–C bond cleaved is the S^+–CH_3 as the methyl group gets transferred to a huge range of both low molecular weight metabolites and macromolecular targets (DNA, RNA, lysine residues in histones and other proteins) (Figure 5.8).

One could write a complete volume on SAM chemical biology where the cationic sulfur is the key atom.[10] Here, we note only that while SAM

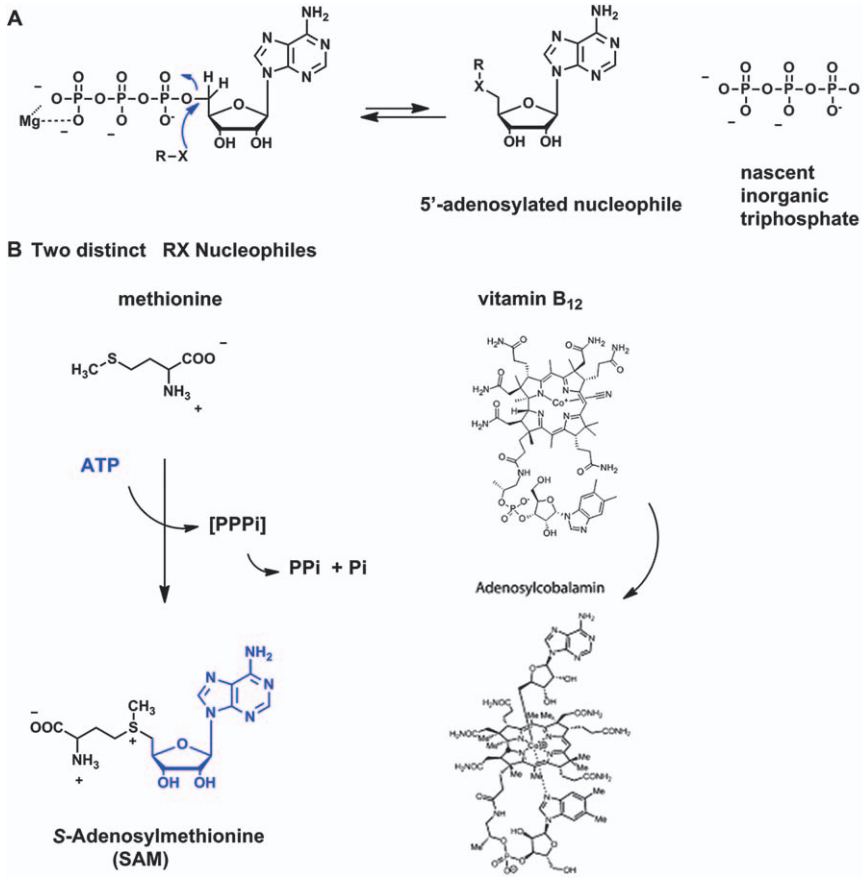


Figure 5.6 (A) Stoichiometry of enzyme-catalyzed adenosyl transfers to an attacking nucleophile. Cleavage of the C5′–O bond yields the adenosylated nucleophile and inorganic triphosphate. (B) The two known examples are *S*-adenosylmethionine synthetase, and adenosyl cobalamin synthetase.

synthetase is an almost singular adenosyl transfer catalyst, its product SAM is, after ATP, the second most widely used coenzyme or substrate in all of cellular metabolism.

5.3.2 Adenosyl-B_{12} Coenzyme

Vitamin B_{12}, is unique among the 13 human vitamins as a modified-porphyrin macrocycle (corrin ring system) with a cobalt ion coordinated equatorially between the four pyrrole nitrogen atoms as ligands[4] (Figure 5.6B). Its structural complexity made it the last of the vitamins characterized, by structure analysis including subsequent crystallography, in 1948. B_{12} is biosynthesized by microbes that reroute some of the heme

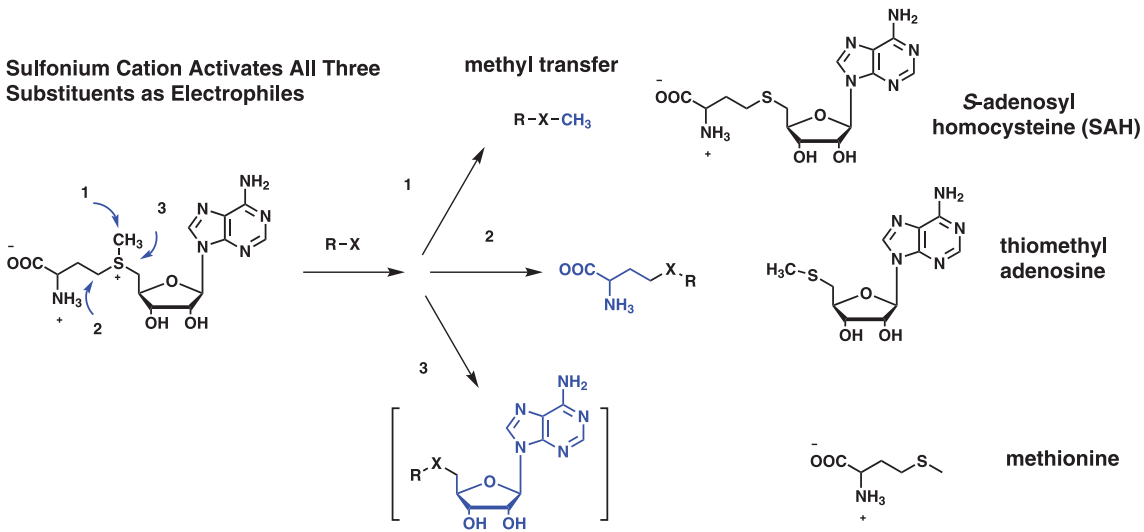


Figure 5.7 *S*-Adenosylmethionine: nothing goes to waste.[9] Each of the three substituents attached to the sulfonium cation are activated for transfer as electrophilic fragments.

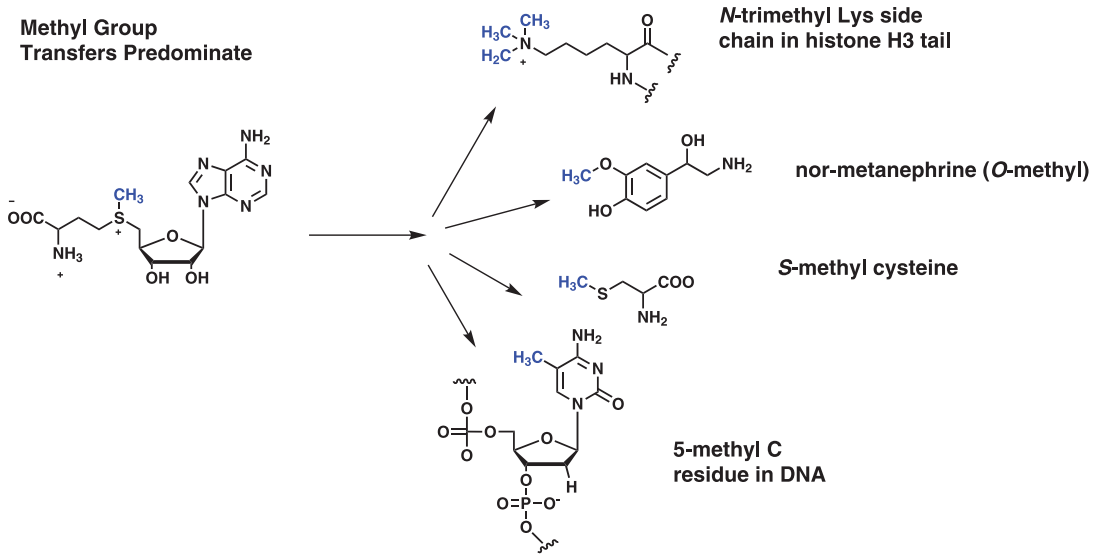


Figure 5.8 Methyl group transfers from SAM as [CH_3^+] equivalents can occur to nucleophilic nitrogen, oxygen, sulfur, and even carbon atoms in a vast array of cosubstrates.

biosynthetic intermediates to form the corrin ring system which is one atom contracted from the typical heme macrocyclic ring scaffold. The smaller internal volume fits cobalt preferentially over iron (in heme) or magnesium (in chlorophyll).

As ingested, the vitamin is found with the top axial ligand as a water molecule (aquo B_{12}) or often with cyanide as a tight binding axial ligand that was concentrated from trace levels by affinity for B_{12} (cyanocobalamin). Neither form of the vitamin is biologically active. Instead, there are two organometallic forms of B_{12} that operate in distinct metabolic niches (Figure 5.6B).

Methyl coenzyme B_{12} has a CH_3 group as the top axial ligand and is the immediate donor of that one-carbon unit [as a CH_3^+ equivalent] to the thiolate anion of homocysteine in the active site of methionine synthase. The second organometallic coenzyme form with a carbon–cobalt bond is adenosyl-B_{12} coenzyme[4] (Figure 5.9).

Adenosyl-B_{12} is generated by the second of the known ATP-dependent adenosyl transferases.[11] (Figure 5.6B) The nucleophilic form of the vitamin B_{12} substrate is the Co^{I} oxidation state. Vitamin B_{12} as isolated has cobalt in the air-stable, but inactive, Co^{III} oxidation state. Bacteria employ an anaerobic activase system to reduce the Co^{III} atom by two electrons to the supernucleophilic Co^{I} oxidation state.

The Co^{I}-vitamin B_{12} is the substrate form along with Mg–ATP utilized by adenosyl B_{12} synthetase. As depicted in Figure 5.9 the Co^{I} atom is the nucleophile attacking C5′ of ATP with C5′–O bond cleavage. The nascent triphosphate anion is hydrolyzed by phosphatase and pyrophosphatase action to three Pi molecules. Again, the cleavage of both phosphoric anhydride linkages in the ATP side chain provides the driving energy to accumulate adenosyl B_{12}. Note that in forming the cobalt–adenosyl carbon bond to C5′, the cobalt atom has been oxidized formally to the Co^{III} oxidation state.

The coenzymatic role of adenosyl B_{12} is to serve as a generator of the 5′-deoxyadenosyl radical.[12] Homolysis of the Co^{III}–C5′ bond produces the Co^{II} oxidation state (so Co^{I}, Co^{II}, and Co^{III} are all in play in B_{12} chemical biology) and the 5′-deoxyadenosyl radical (dA$^{\bullet}$). dA$^{\bullet}$ serves as initiator of radical reactions in specifically bound substrates by abstracting a hydrogen atom (H$^{\bullet}$) from them. That hydrogen atom transfer produces 5′dA-H and the substrate radical. Rearrangement to a product radical can occur before the H$^{\bullet}$ is transferred back from dA-H. This quenches the product radical as P–H while regenerating dA$^{\bullet}$. A catalytic cycle ends as dA$^{\bullet}$ and the Co^{II} atom each provide one electron in reforming the carbon–Co^{III} bond in adenosyl B_{12}.

A classic case of such a 1,2-radical-based rearrangement is in the adenosyl B_{12}-dependent propanediol to propionaldehyde rearrangement catalyzed by the enzyme known as diol dehydrase (Figure 5.10). The nascent product is the 1,1-diol (aldehyde hydrate) that is stereospecifically converted to the aldehyde.

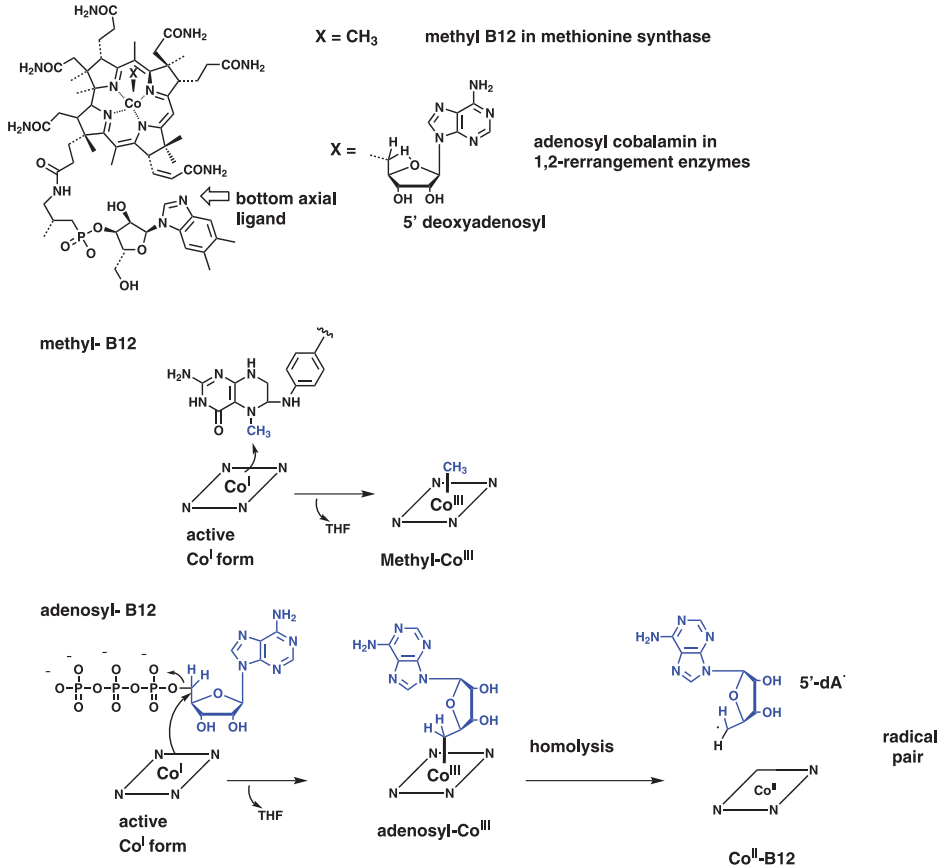


Figure 5.9 Two organometallic coenzyme forms of vitamin B_{12}. Methyl B_{12} with a CH_3 group as top axial ligand is formed from N^5–CH_3–tetrahydrofolic acid and transfers the methyl group to the sulfur of homocysteine in the active site of methionine synthase. Adenosyl B_{12} undergoes Co–C5′ homolysis in enzyme active sites that use adenosyl B_{12} as a radical generator of the 5′-deoxyadenosyl radical to initiate one-electron pathways in cosubstrates.

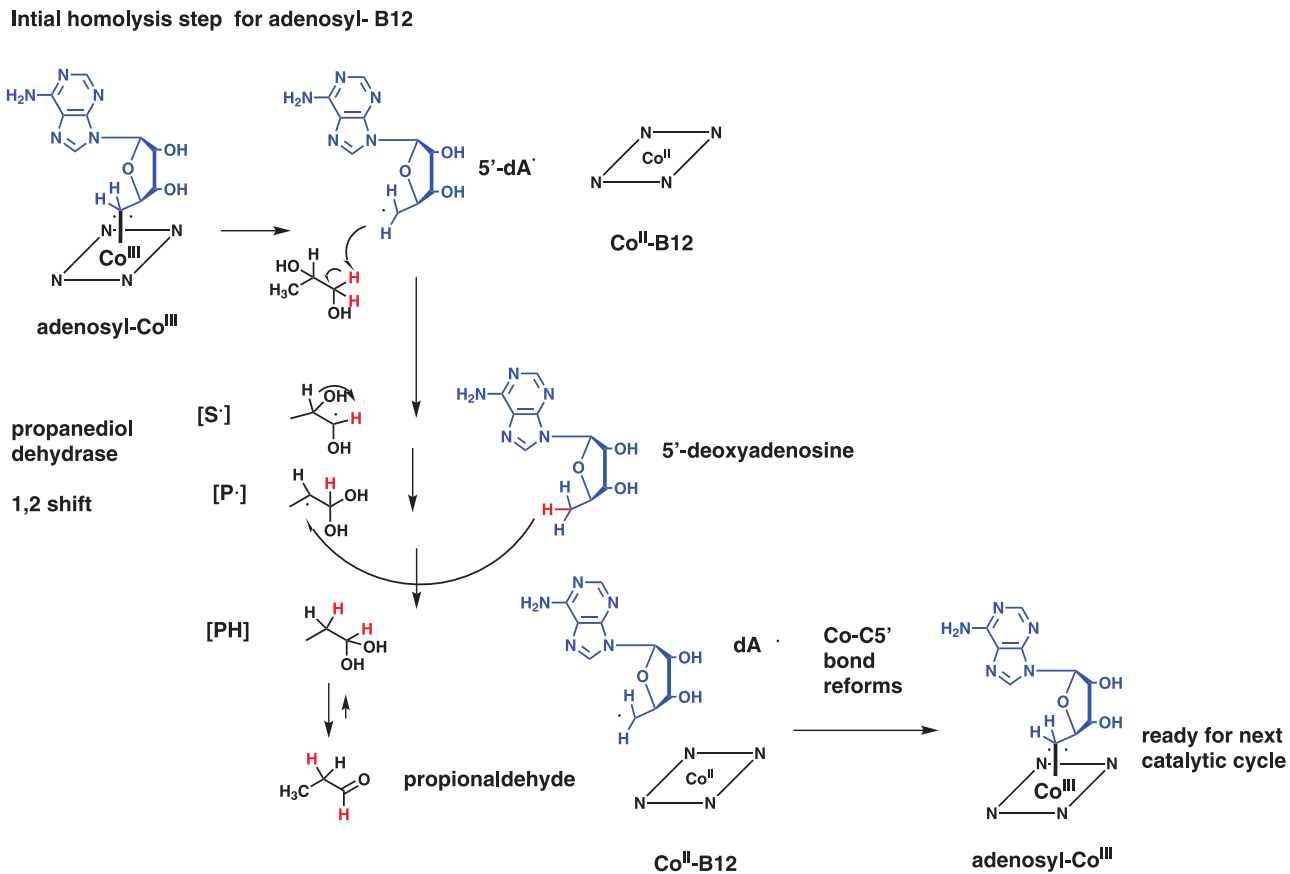


Figure 5.10 Bacterial propanediol dehydrase, rearranging propanediol to propionaldehyde, uses adenosyl-B12 as generator of the 5′-deoxyadenosyl radical *in situ* to form a carbon-centered radical on bound substrate propanediol.

The two adenosyl transferases that fragment ATP to nascent inorganic triphosphate hexa-anion and an adenosylated nucleophile sit in remarkably different but consequential biological niches. Adenosyl transfer to the thioether sulfur of methionine creates SAM, which is used billions of times per cell cycle as a methyl transfer agent. Adenosyl transfer to the Co^{I} form of vitamin B_{12} creates a remarkable organometallic carbon–cobalt bond. Homolysis of that Co^{III}–C5′ bond is the source of 5′-deoxyadenosyl radical initiators in rearrangement catalysis. These are two unanticipated uses of enzymatic ATP fragmentation during evolution of 5′deoxyadenosyl radical chemistry.

References

1. B. Hove-Jensen, K. Anderson, M. Kilstrup, J. Martinussen, R. L. Switzer and M. Willemoes, Phosphoribose diphosphate (PRPP) enzymology, utilization and metabolic significance, *Microbiol. Mol. Biol. Rev.*, 2016, **81**, e00040.
2. C. T. Walsh, B. Tu and Y. Tang, Eight kinetically stable but thermodynamically activated molecules that power cell metabolism, *Chem. Rev.*, 2018, **118**, 1460–1494.
3. V. Hauryliuk, G. Atkinson, K. S. Murakami, T. Tenson and K. Gerdes, Recent functional insights into the role of pppGpp in bacterial physiology, *Nat. Rev. Microbiol.*, 2015, **13**, 298–309.
4. C. T. Walsh and Y. Tang, *The Chemical Biology of Human Vitamins*, Royal Society of Chemistry, London, 2018.
5. D. Pilloff, K. Dabovic, M. J. Romanowski, J. B. Bonanno, M. Doherty, S. K. Burley and T. S. Leyh, The kinetic mechanism of phosphomevalonate kinase, *J. Biol. Chem.*, 2003, **278**, 4510–4515.
6. M. Chen, J. Dixon and G. Manning, Genomics and evolution of human phosphatases, *Sci. Signaling*, 2017, DOI: 10.1126/scisignal.aag1796.
7. C. T. Walsh and Y. Tang, *Natural Product Biosynthesis Chemical Logic and Enzymatic Machnery*, Royal Society of Cemistry, London, 2017.
8. F. Takusagawa, S. Kamitori and G. D. Markham, Structure and function of *S*-adenosylmethionine synthetase: crystal structures of SAM synthetase with ADP, BrADP and PPi at 2.8 Angstroms resolution, *Biochemistry*, 1996, **35**, 2586–2596.
9. M. Fontecave, M. Atta and E. Mulliez, S-adenoslmethionine: nothing goes to waste, *Trends Biochem. Sci.*, 2004, **29**, 243–249.
10. W. Loenen, *S*-adenosylmethionine: jack of all trades and master of everything, *Biochem. Soc. Trans.*, 2006, **34**, 330–333.
11. J. Escalante-Semerena, Conversion of cobinamide into adenosylcobinamide in Archae, *J. Bacteriol.*, 2007, **189**, 4855–4860.
12. E. N. Marsh and G. Melendez, Adenosylcobalamin enzymes: theory and experiment begin to converge, *Biochim. Biophys. Acta*, 2012, **1824**, 1154–1164.

CHAPTER 6

Activated Phosphoryl Groups and Biosynthetic Paths to ATP

6.1 The Spectrum of Biologically Activated Phosphoryl Groups

Thus far, we have focused on the two phosphoric anhydride bonds in the side chains of ATP, the other three NTPs (UTP, CTP, GTP) and the corresponding 2′dNTPs (with dTTP replacing dUTP as DNA building block). These are kinetically stable thermodynamically activated structural elements at neutral cellular pH values. Yet their transfer to oxygen nucleophiles including water is so favorable at equilibrium that their fragmentation, at both γ-P (phosphoryl transfers), but especially at α-P (nucleotidyl transfers) drives otherwise unfavorable equilibria (see Chapter 1 Section 1.10).

In this chapter we note a small number of other phosphorylated metabolic scaffolds that are similarly thermodynamically activated yet stable enough to be formed and then utilized as freely diffusible molecular species in cellular energy metabolism.[1] In principle each of the scaffolds described are doubly activated-for phosphoryl transfer to a nucleophile or for transfer of the other moiety (acyl, enolate, guanidinium, sulfuryl). It is often the other activated group that is transferred as an electrophilic fragment. This is the general case for instance in the acyl phosphates described next.

6.1.1 Acyl Phosphates *vs.* Acyl Adenylates

We have noted that two kinds of mixed acyl phosphoric anhydrides are formed as activated acyl species in primary metabolic pathways: acyl phosphates and acyl adenylates. They arise, respectively, from attack of substrate carboxylate anions on either Pγ or Pα of Mg–ATP side chains (Figure 6.1).

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

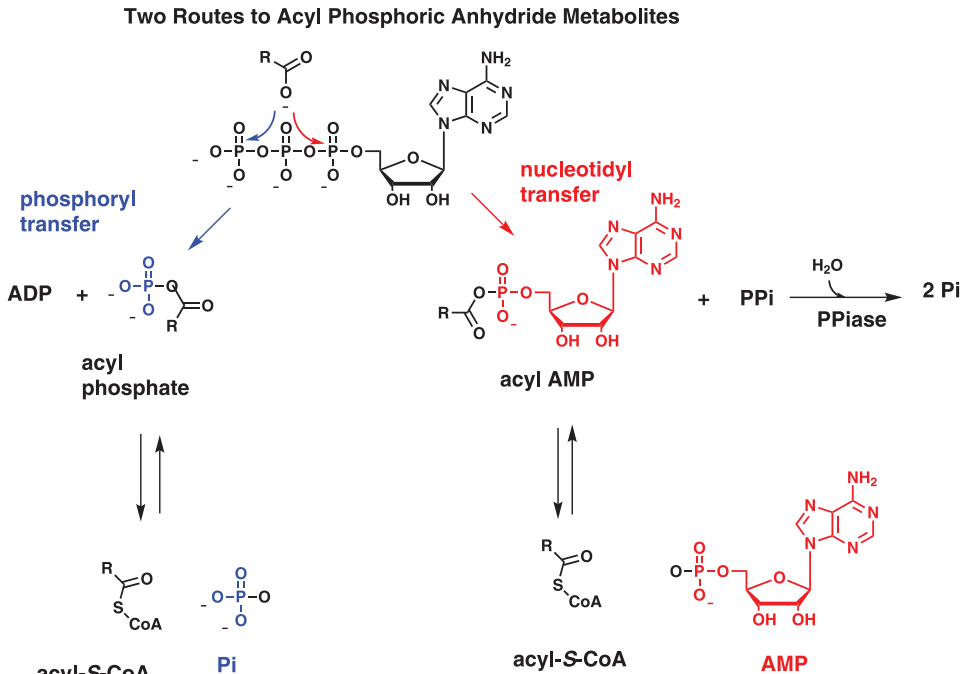


Figure 6.1 Two types of acyl phosphoric anhydride metabolites in cells. Acyl phosphates arise from carboxylate attack on Pγ of Mg–ATP: phosphoryl group transfers. Acyl-AMPs arise from carboxylate attacks on Pα of Mg–ATP: adenylyl transfers. Both variants of acyl–phosphoric anhydrides are doubly activated for transfer of the acyl or phosphoryl moiety. Both acetyl phosphate and acetyl AMP are enzymatically converted to the kinetically more stable acetyl-CoA thioester.

From the point of view of thermodynamic activation between the two mixed carboxylic–phosphoric anhydrides, both the acyl phosphates and the acyl-AMPs have achieved equivalent utility as *subsequent acyl donors*. Thus, acetyl-AMP and acetyl phosphate are about equally effective in subsequent enzymatic acetyl transfers to the thiolate of coenzyme A to generate acetyl CoA (Figure 6.2).

We have noted that nucleotidyl transfers generate PPi as a coproduct while phosphoryl transfers produce ADP. High activity pyrophosphatases in cells hydrolyze the PPi to two molecules of inorganic phosphate. This coupling makes the acyl-adenylate route more attractive in accumulating the acyl-AMP over acyl phosphate. This logic is consistent with nucleotidyl transfer regimes as the preferred route for monomer activation to build the thermodynamically unfavorable condensed linear polymers of DNA, RNA, and proteins. The adenosyl moiety, in the AMP portion of acyl-AMPs also presumably offers extended interactions with widely distributed AMP binding pockets in many dozens of proteins to provide binding energy and affinity.

Nonetheless, and perhaps because of the protein binding affinity from the adenosyl moiety, acyl-AMPs are almost never used as freely diffusing intermediates in cells. (Nucleoside diphosphosugars *are* diffusible building blocks.) This is in contrast to acyl phosphates where there are dozens of acyl activation examples in primary metabolic pathways. The acyl phosphates are utilized metabolically either for acyl or for phosphoryl transfer to cosubstrate nucleophiles.

The simplest example may be activation of carbonic acid (HCO_3^-), the hydrated form of CO_2. The biotin-dependent acetyl CoA carboxylase,[2] the first committed catalyst in fatty acid biosynthesis, generates carbonyl phosphate (Figure 6.2A) by phosphoryl transfer from ATP to carbonic acid. Analogously, the one carbon carbamic acid is likewise enzymatically phosphorylated to carbamoyl phosphate (Figure 6.2A). Carbamoyl phosphate then acts in two distinct subcellular compartments for two distinct functions: in the cytoplasm to initiate pyrimidine biogenesis *via* carbamoyl transfer to the amino group of aspartate, and in mitochondria to begin the urea cycle. A second one-carbon acid formic acid is phosphorylated in bacteria to activate it as acyl donor in pyrimidine ring construction (*N*-formyl amide formation) (Figure 6.2B).

The two proteinogenic amino acids with side chain carboxylates, aspartate and glutamate, are phosphorylated on the β- and γ-carboxylates to yield β-aspartyl phosphate or γ-glutamyl phosphate, respectively (Figure 6.2B). The activated aspartyl-group in the mixed anhydride is then the locus for consecutive hydride transfers to reduce the carboxylic–phosphoric anhydride to the alcohol in homoserine. Separately, the glutamyl phosphate is captured by ammonia as a nucleophile to form the carboxamido group of glutamine.

The tripeptide glutathione (γ-Glu–Cys–Gly), a key thiol protective metabolite at concentrations reaching 10–16 mM in cells is assembled from its three constituent amino acids in a short two-enzyme pathway.[3] The γ-Glu–Cys *isopeptide* bond is indicative that a ribosomal process is unlikely to be involved

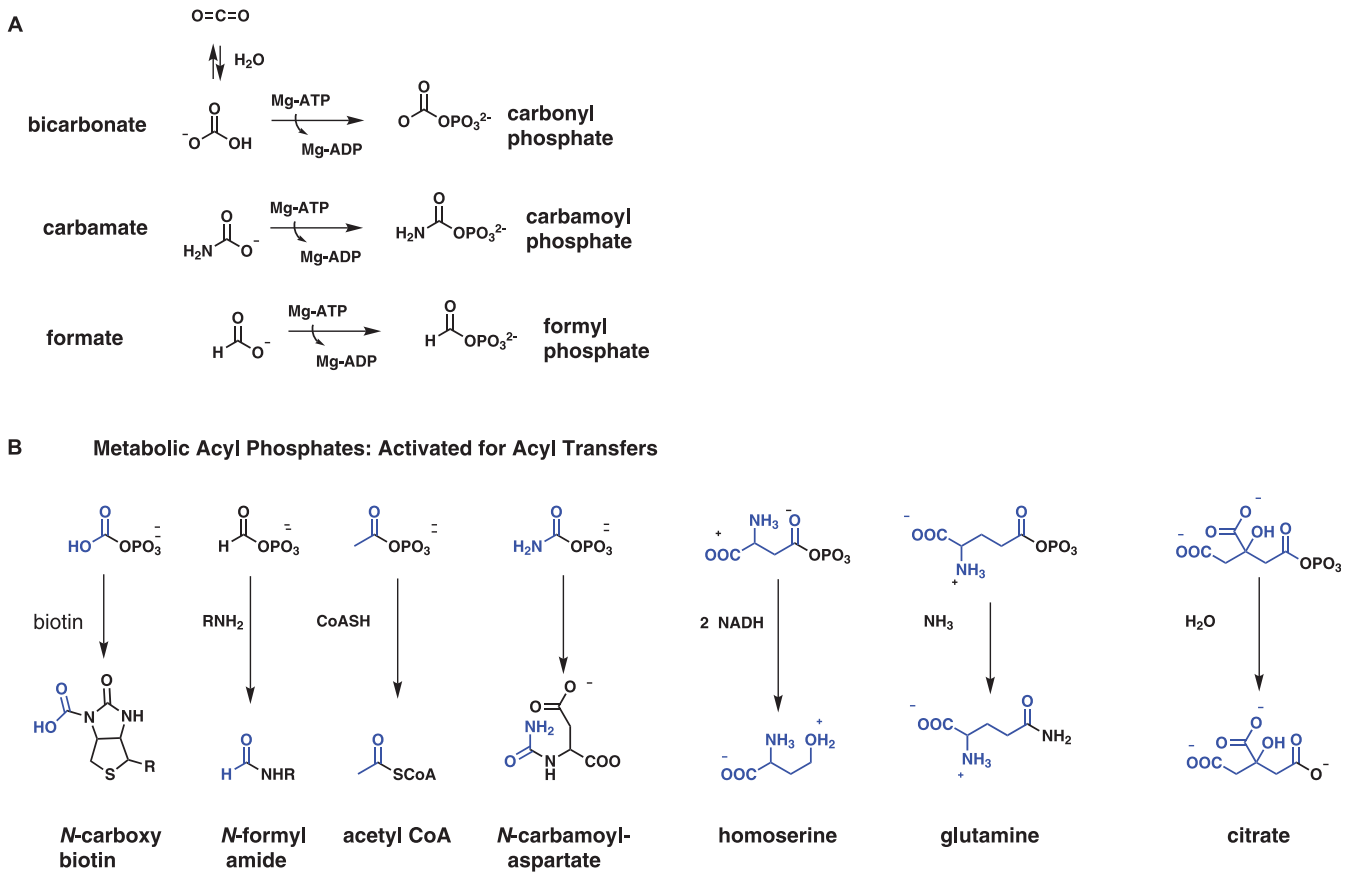


Figure 6.2 Acyl phosphate metabolites. (A) Three one-carbon acyl phosphates as cellular metabolites; (B) illustrative reactions of seven acyl phosphates.

in biosynthesis. Indeed, two enzymes of the ATP grasp family are responsible. They each cleave ATP to ADP (Figure 6.3), rather than the ATP to AMP and PPi pattern for protein synthesis at ribosomes. Accordingly, γ-Glu–Cys synthetase and then glutathione synthetase each generate bound aminoacyl phosphates rather than aminoacyl-AMPs. The first enzyme makes γ-glutamyl phosphate, activating the γ carboxylate to attack Pγ of Mg–ATP. In turn the bound

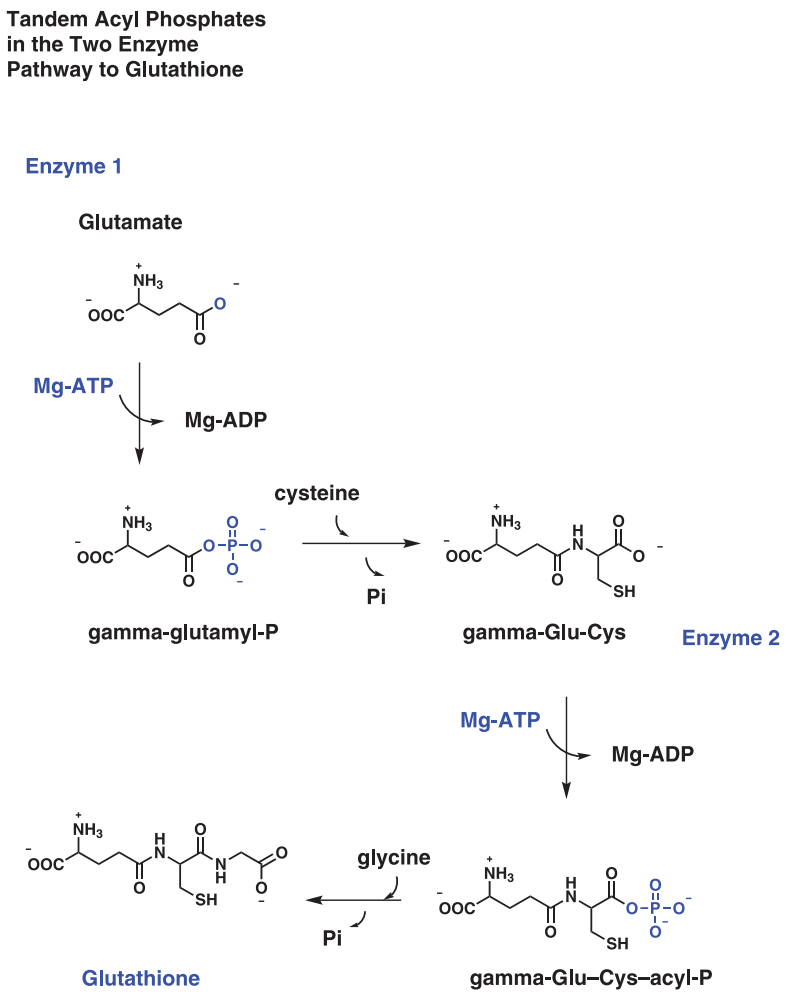


Figure 6.3 The two-enzyme pathway to convert glutamate, cysteine, and glycine to glutathione involves consecutive action of gamma-glutamyl-cysteine synthase followed by glutamine synthase. Acyl phosphate intermediates (gamma-glutamyl-phosphate and then gamma-glutamyl-cysteinyl-phosphate) are the activated forms of the glutamate gamma carboxylate and the dipeptidyl Glu–Cys for subsequent amide bond formations. The isopeptide bond between Glu and Cys indicates a nonribosomal origin for glutathione.

γ-glutamyl phosphate is captured at the activated γ-carbonyl by the deprotonated amine of cosubstrate cysteine. Correspondingly, in the glutathione synthetase active site a Glu–Cys–acyl phosphate is formed in, and captured for, amide bond formation by the deprotonated amine of cosubstrate glycine.

The double headed activation of acyl phosphates, for acyl or phosphoryl group transfer, is performed by three pairs of tandem enzymes (Figure 6.4). The first acts during glycolysis, the second in the tricarboxylate cycle, and the third by diverting citrate carbons to fatty acids. The interplay of (1) ATP and free carboxylate, (2) the activated acyl phosphate, and (3) the more kinetically stable acyl thioester is played out in these three sets of tandem enzyme reactions.

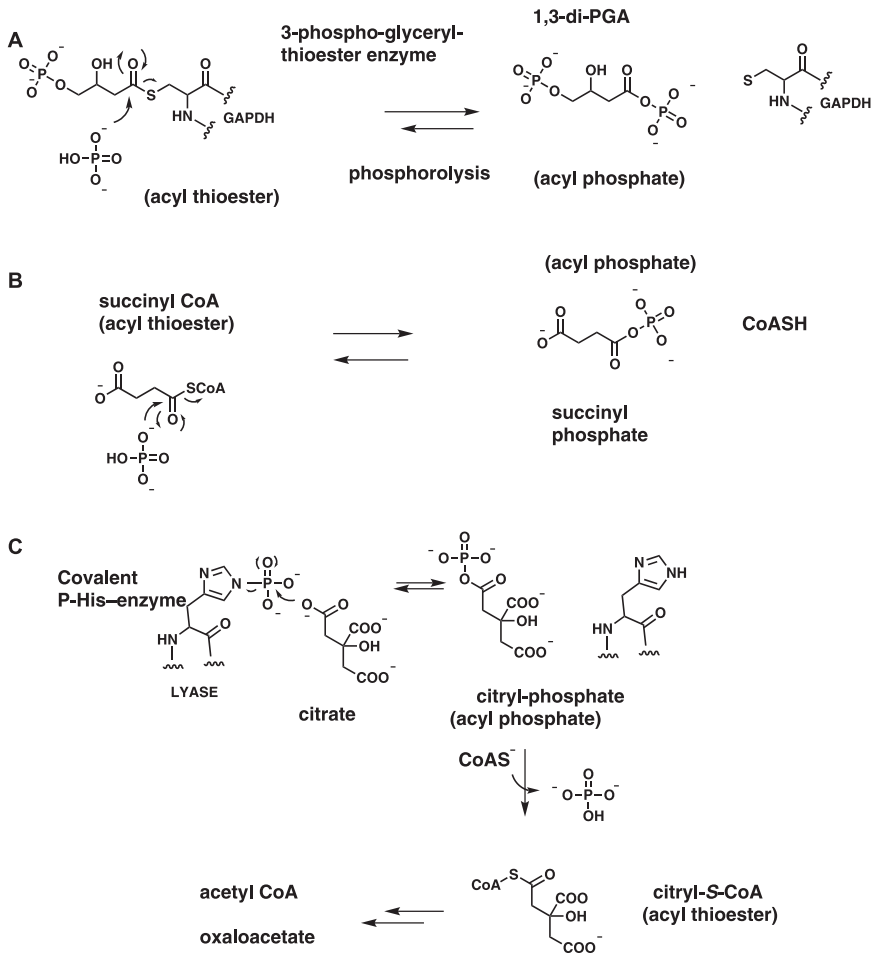


Figure 6.4 Three enzymes with acyl phosphate intermediates or substrates: (A) glyceraldehyde-3-phosphate dehydrogenase (glycolysis); (B) succinyl CoA synthetase (citrate cycle); (C) ATP citrate lyase (cytoplasmic fatty acid biosynthesis).

In glycolysis the phosphorolytic conversion of the 3-phosphoglyceryl-*S*-enzyme to 1,3-diphosphoglycerate (Figure 6.4) by glyceraldehyde-3-phosphate dehydrogenase is the preamble to phosphoryl transfer to ATP by the next enzyme phosphoglycerate kinase. CoASH is not a participant and the 3-phosphoglyceric acid (3-PGA) scaffold is carried deeper into glycolysis.

The two-enzyme flux goes the same way in the citric acid cycle but commences with nascent succinyl CoA from the 2-ketoglutarate dehydrogenase reaction. Succinyl CoA is phosphorolyzed enzymatically to the mixed acyl–phosphoric anhydride succinyl phosphate by the enzyme succinic thiokinase that then transfers the phosphoryl group to GDP to make GTP.

The third example is the enzyme ATP citrate lyase, the cytoplasmic enzyme that shunts two of the six carbons of citrate into fatty acyl chains. As citrate is transported out of the mitochondria (subcellular site of the citrate biosynthetic cycle) ATP citrate lyase transfers a phosphoryl group from a phosphohistidinyl–enzyme residue to one of the three citrate carboxyl groups, the proS primary carboxylate. The resultant acyl phosphate citryl phosphate is then converted to bound citryl CoA: a transfer of the activated citryl group to the thiolate of coenzyme A. Now the citryl CoA undergoes an enzyme-mediated retro-thioclaisen reaction, cleaving the citryl CoA to oxaloacetate and acetyl CoA. The acetyl units are then used by cytoplasmic fatty acyl synthases to store the two carbon units in saturated long chain fatty acids as energy reserves (Figure 6.4).

In general, most acyl phosphates in metabolism are generated from ATP and used for subsequent transfer of the activated acyl groups to nucleophiles (amines, thiols, hydride ion).[4] When they engage in phosphoryl transfers to make ATP or GTP, they reflect energy harvesting steps accompanying redox metabolism.

6.1.2 Phosphorylated Enols

The trapped enol phosphate phosphoenolpyruvate is uphill thermodynamically from ATP and is stable enough to serve as a freely diffusible intermediate in cell metabolism. Evidence that both portions, the phosphoryl group and the trapped enol are activated for capture by co-substrate nucleophiles is provided by two examples of each outcome (Figure 6.5).

Phosphoryl transfer from PEP to ADP to yield ATP and the nascent enolate of pyruvate is the signature reaction of pyruvate kinase,[5] with equilibrium in favor of ATP formation $>10^3$. In contrast, PEP is the energy source during uptake of a number of hexose sugars into bacteria. As the sugars, including glucose, mannose, and glucosamine, cross the bacterial membrane they are phosphorylated and deposited inside as the membrane-impermeant hexose-6-phosphates.[6] PEP is the initial phosphoryl group donor but the $-PO_3^{2-}$ groups may undergo another two to three transfers before ending up on the 6-OH of a transported sugar (see Chapter 14).

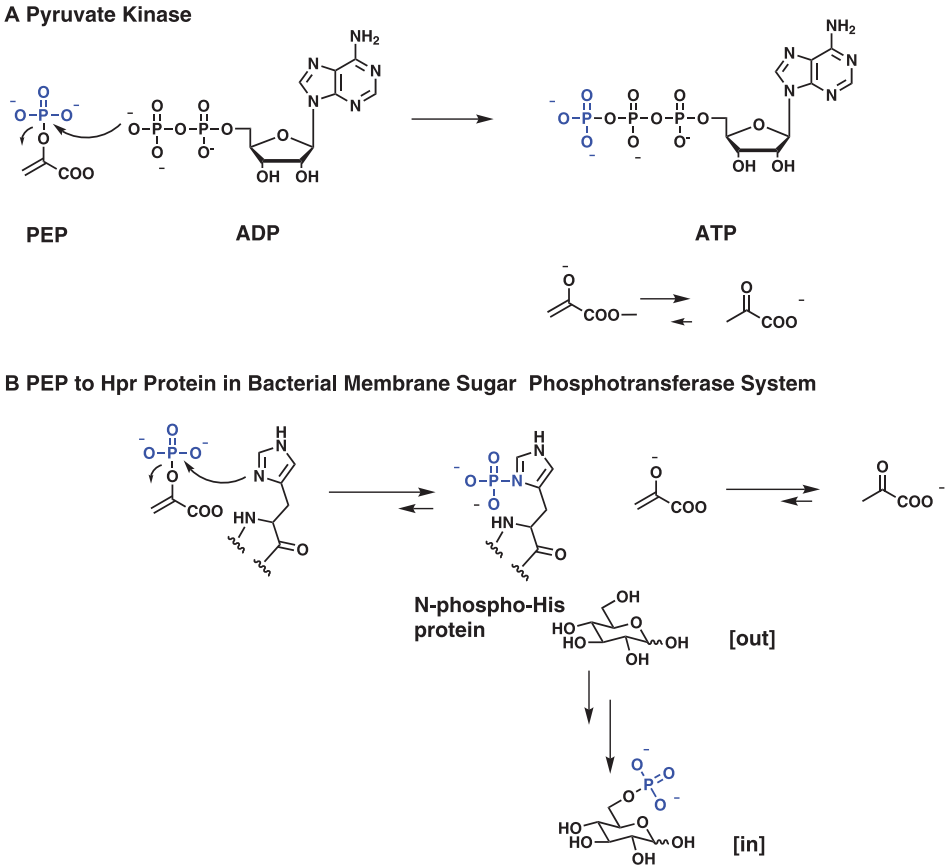


Figure 6.5 Two enzymes that catalyze phosphoryl group transfer from the trapped enol PEP. (A) Pyruvate kinase mediates a direct $-PO_3^{2-}$ transfer to ADP while (B) enzyme 1 of the bacterial hexose phosphate transferase system catalyzes transfer of the $-PO_3^{2-}$ group from PEP to an imidazole side chain of a protein histidine residue.

PEP is attacked on the phosphoryl group phosphorus atom by the N3 of the imidazole side chain of a histidine residue in enzyme-I. That $-PO_3^{2-}$ is then subsequently transferred to a His residue in a histidine rich protein (HPr), that acts as a phosphorylation hub to transfer the $-PO_3^{2-}$ out to a substrate-specific EII protein.[7] Typically, EIIs are likewise phosphorylated on a His residue but some are instead phosphorylated on a cysteine thiolate side chain. Finally, the EII–PO_3^{2-} group is captured by a hexose-6-OH to give the ''low energy'' alcoholic phosphomonoester.

The serial transfers of the $-PO_3^{2-}$ group my reflect evolution of a system where the EI has specificity for PEP but none of the hexose sugars. The EIIs may have evolved from specific sugar carrier proteins that gained phosphoryl transfer activity. The HPr protein is the phosphoryl transfer middleman.

Although the chemistry is distinct there are strategic analogies to the three enzymes of the eukaryotic ubiquitin systems where an information rich acyl group [ubiquitin (Ub)] is transferred first to an enzyme that only recognizes Ub and ATP, then to a set of EIIs as covalently acylated middlemen. The EIIs (akin to the PTS HPr middleman) are the hubs for the EIIIs that impart specificity to the client proteins that become covalently ubiquitinylated.[8]

The enolpyruvyl group can also be transferred from PEP to cosubstrates, demonstrating the activation of both pieces of PEP. In a formal sense, if the O–[PO_3^{2-}] bond of PEP was disconnected early in a reaction coordinate, the nascent enolate will develop negative charge density on both the enolate oxygen and C3 as the carbanion contributor to the enolate structure. Then it would behave as carbon nucleophile toward electrophilic centers of cosubstrates.

In fact, the two enzyme-catalyzed enolpyruvyl transfers noted below show the opposite polarity: cosubstrate nucleophiles attack C3 of the PEP substrate.[9] Resolution of the conundrum lies in timing of the O–[PO_3^{2-}] bond cleavage. It is in fact late, not early.

As shown for the murein A protein (MurA) reaction, the first step in bacterial peptidoglycan pentapeptide chain formation, the 3'OH of a UDP GlcNAc substrate attacks C3 of PEP to create a tetrahedral adduct (Figure 6.6A). That adduct then undergoes an enzyme-catalyzed internal elimination of the elements of inorganic phosphate. The cosubstrate 3′-hydroxyl has undergone a net alkylation by a three-carbon vinyl ether. In the end the O–[PO_3^{2-}] bond was not cleaved. Instead the C–[OPO_3^{2-}] bond has been cleaved from PEP.

This is the same pattern observed in the shikimate pathway for biosynthesis of the aromatic core of phenylalanine and tyrosine. (Figure 6.6B). The 5-OH of shikimate-3-P is nucleophile attacking carbon 3 of PEP. Net C–[OPO_3^{2-}] bond cleavage occurs in PEP. The 5-OH becomes derivatized as the enol-vinyl ether in the enzymatic product 5-enolpyruvyl-shikimate-3-phosphate (EPSP), on the way to chorismate.

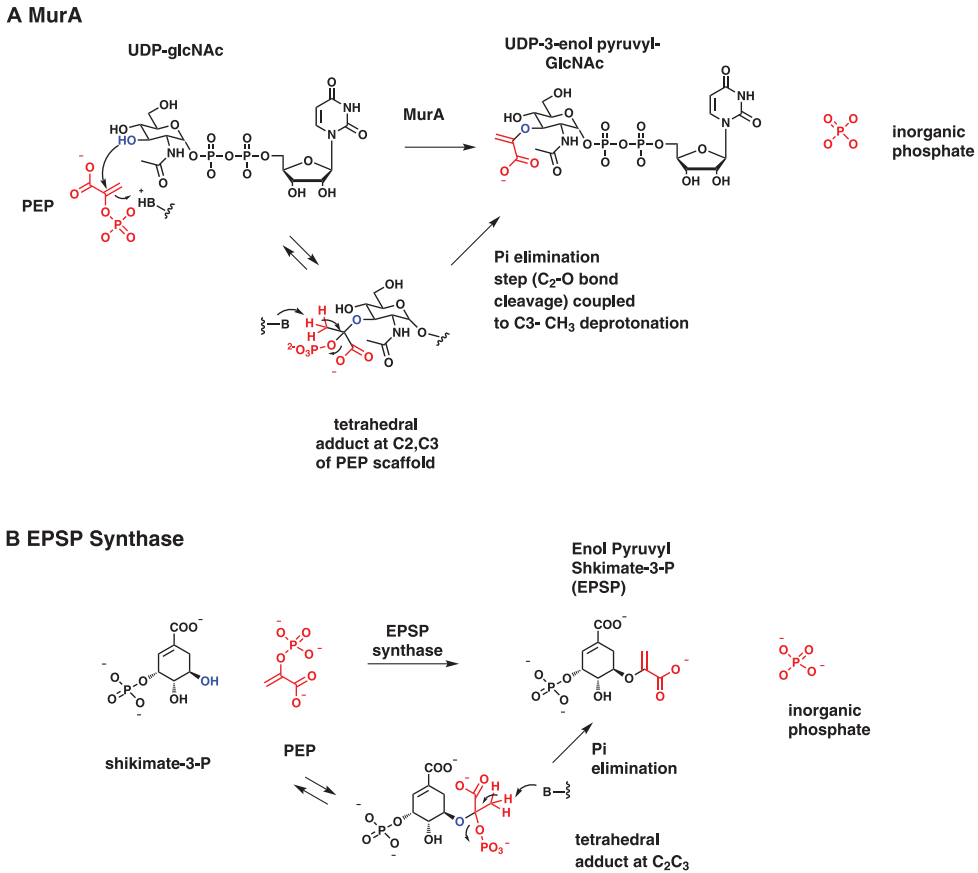


Figure 6.6 Two enolpyruvyl transfers from PEP in bacterial cell wall biosynthesis. (A) MurA transfers the enolpyruvyl group to the 3′OH of UDP-*N*-acetyl-muramic acid (UDP-MurNAc). The resultant tetrahedral adduct eliminates inorganic phosphate. (B) The enzyme enolpyruvylshikimate phosphate synthase transfers an enolpyruvyl moiety from PEP to the 5-OH of shikimate-3-phosphate. Again, a tetrahedral intermediate eliminates the elements of phosphoric acid to yield the product.

6.1.3 Sulfuric–Phosphoric Anhydrides

The closest formal analogy to the thermodynamically activated phosphoric anhydride linkages in ATP may be the mixed sulfuric–phosphoric anhydride bond found in adenylyl-sulfate (APS) and its 3′-phospho-congener 3′-phospho-5-adenosyl-5-phosphosulfate (PAPS). APS arises by enzyme-mediated attack of one of the oxygen anions of inorganic sulfate on Pα of ATP: a nucleotidyl transfer to inorganic sulfate. Sulfate acted as a *nucleophile* in the active site of ATP sulfurylase but in all subsequent reactions of the APS–PAPS mixed sulfuric–phosphoric anhydride, the $-SO_3^-$ group is *transferred as an electrophile*. While APS is a sulfuryl transfer reagent in prokaryotes, kinase action to create PAPS gives the predominant sulfuryl transfer metabolite in eukaryotes.[10]

The phosphoric sulfuric anhydride linkage in APS and PAPS is doubly activated for phosphoryl or sulfuryl transfer, but the inventory of reactions observed is heavily biased to sulfuryl transfer to nucleophilic cosubstrates (Figure 6.7). The most common nucleophiles are oxygen atoms or nitrogen atoms, in such diverse metabolites as estrogens, xenobiotic drugs, glycosylated proteins and tyrosyl residues in certain proteins. When nucleophilic nitrogen atoms are the attacking species, the product sulfamates have an N–S bond in the $N–SO_3^-$ grouping.

The bis-sulfurylation of the ion channel marine toxin saxitoxin at both nitrogen and then oxygen nucleophiles by attack on the sulfuric anhydride group of two molecules of PAPS[11] is depicted in Figure 6.8. The side chain anhydrides of APS and PAPS are trianionic at neutral pH values and the negative charges slow down adventitious hydrolysis. Thus, the sulfuric–phosphoric anhydrides are sufficiently kinetically stable to be freely diffusible metabolites in cells.

6.1.4 Guanidinium Phosphoramidates (Phosphagens)

The other class of low molecular weight phosphorylated metabolites that are sufficiently activated to act as reversible reservoirs for ATP are the *N*-phosphoryl-guanidinium scaffolds of phosphagens.[12] Arginine phosphate may be the ancestral metabolite of this class while creatine phosphate is the higher eukaryotic signature phosphagen. As noted in Chapter 10, the *N*-phosphoryl arginine, -creatine, -lombricine, and congeners are guanidinium phosphoramidates. At first blush, the guanidinium groups seem like unlikely nitrogen nucleophiles given the pK_a values of 10 and above. However, once formed they can undergo reversible transfer to Mg–ADP stores to replenish Mg–ATP levels in tissues that are rapidly consuming it, such as contracting muscles (Figure 6.7; also see Chapter 10).

The phosphagens act as phosphoryl transfer metabolites rather than guanidinium moiety transfer agents in cellular pathways.

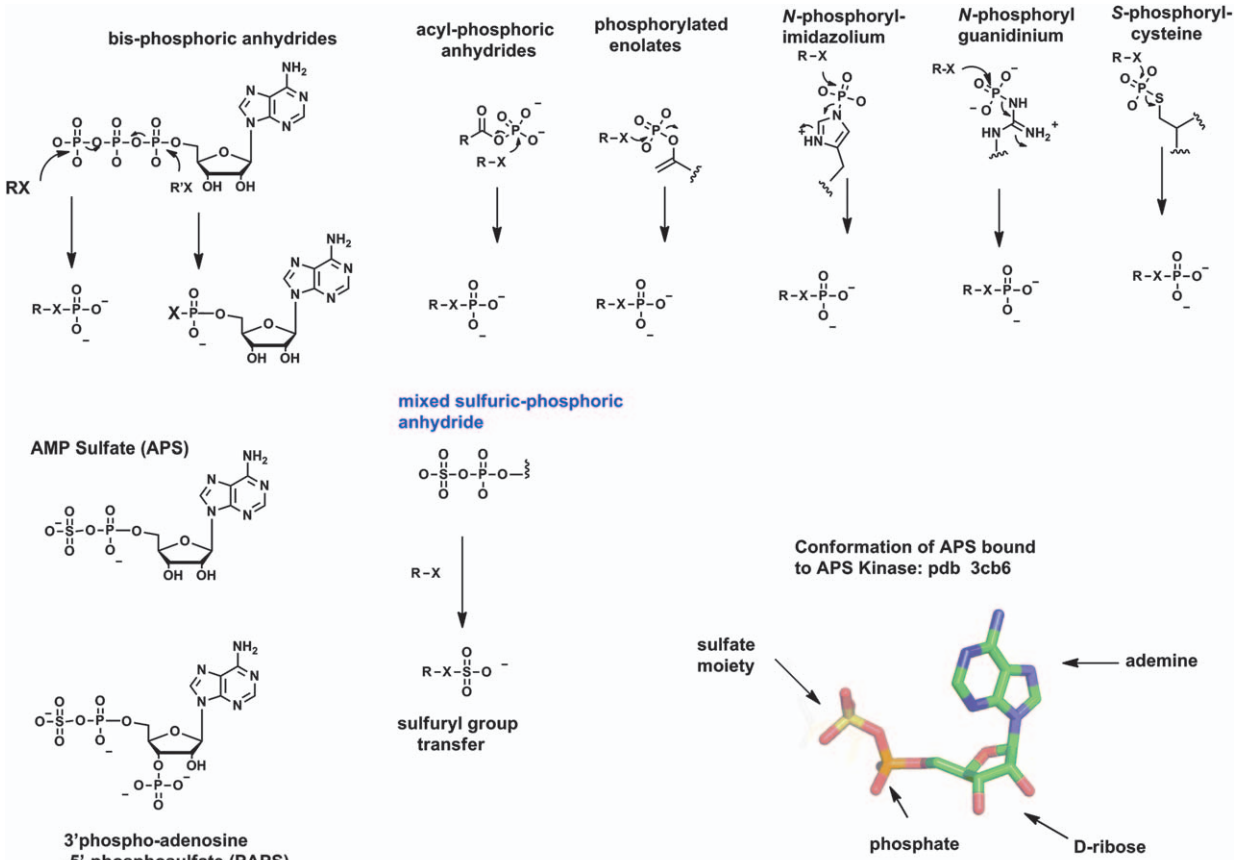


Figure 6.7 Several phosphorylated scaffolds with high group transfer potential are displayed. In particular the mixed sulfuric–phosphoric anhydride side chain of 5′adenylsulfate and 3′phospho-5-adenylsulfate (PAPS) are biological metabolites that function as sulfuryl group ($-SO_3^-$) group donors.

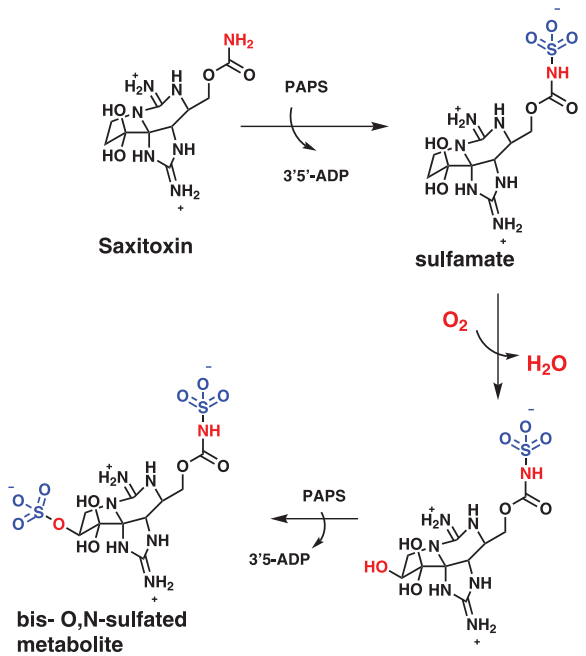


Figure 6.8 Marine dinoflagellate neurotoxins can be sulfated on both oxygen and nitrogen atoms. Saxitoxin is shown to undergo *N*-sulfamation and then *O*-sulfation by a pair of sulfuryl transferases using PAPS as $-SO_3^-$ donor.

6.1.5 Protein Histidinyl Phosphates, Aspartyl Phosphates, Cysteinyl Phosphates.

These three phosphorylated protein species maintain the ability to transfer the $-PO_3^{2-}$ group onward to nucleophiles. We will take up these reactivities in Chapter 14 on noncanonical phosphoproteomics. An *N*-phospho-imidazole side chain of a phosphohistidine residue is shown in Figure 6.5.

6.1.6 Protein Serinyl Phosphates, Threonyl Phosphates, Tyrosyl Phosphates

In contrast, these three protein phosphomonoesters that are the backbone of conventional phosphoproteomics studies are *not* considered to be thermodynamically activated. They do not generally transfer the $-PO_3^{2-}$ groups to substrates other than water, and then only under the catalytic aegis of specific protein phosphatases. We discuss this class of phosphoproteins in Chapter 15 of the phosphoproteomics section IV.

6.1.7 Low Molecular Weight Phosphomonoesters

We have noted in Chapter 1, Section 1.10 and also in passing in Chapter 3 that each of the specific low molecular weight phosphate mono-oxoesters formed in metabolic pathways are also *not* thermodynamically activated for

phosphoryl transfers. As discussed in Chapter 7, phosphatases remove the $-PO_3^{2-}$ groups by net hydrolysis, transfer to water. Low molecular weight phosphomonoesters that show low to negligible phosphoryl group transfer potential to cosubstrates include the common hexose-6-phophates and 1-phosphates, ribose-5-phosphate, inositol-phosphates, *O*-phosphorylcholine, *O*-phosphoserine and *O*-phosphothreonine.

6.2 Three Ways to Make ATP: Acyl Phosphates, Enol Phosphates, and Coupled Proton Gradients

Given the manifold roles of ATP and related NTPs derived from ATP, or 2′dATP by the actions of nucleoside diphosphokinases, it is worth a brief excursion for one more look at how cells and organisms make ATP molecules *in vivo*. One starting point is to note that cells and tissues get most of their energy from either glucose or stored fatty acids, oxidizing them all the way to CO_2. In those oxidations their associated electrons are captured and stored temporarily in reduced forms of nicotinamide or flavin coenzymes before those electrons are sent down a series of potential drops to reduce O_2 by four electrons to two water molecules (Figure 6.9).

The glucose substrate molecules may derive from glucose monomers transported into cells through multiple hexose carrier proteins or may come from polysaccharide stores, such as starch or glycogen glucose polymers, by enzymes acting as glycoside hydrolylases or glycoside phosphorylases (Chapter 9), clipping off of one monomer unit at a time.

Fatty acids are broken down by beta oxidation enzymes, releasing two carbon units at a time, as acetyl-CoA molecules, that get fed into the citric acid cycle. The workings of the eight enzymes of the citrate cycle in each turn, extract eight electrons from the acetyl unit (six from the methyl group, two from the acyl carbon), generating two molecules of CO_2 that diffuse out of cells, ultimately exhaled by the lungs of humans. The eight electrons are stored in four pairs, three pairs in NADH and one pair in $FADH_2$. These eight electrons are fed into the mitochondrial electron transport chain, one at a time, to pass from iron–sulfur clusters to quinones, to cytochromes, on the way to cytochrome oxidase. That oxidase can store up to four electrons and reduce O_2 by those four electrons to two H_2O molecules. During the passage of electrons down the respiratory chain, protons are pumped at three locations from the matrix across the mitochondrial inner membrane to the intermembrane space (Figure 6.9).

The pumping of protons against their concentration gradient, creating the proton component of the electrochemical potential, is driven by the favorable thermodynamics of the electrons flowing from NADH ($Eo' = -320$ mV) or $FADH_2$ ($Eo' = -200$ mV), down to O_2, ($Eo' = +820$ mV). The potential drop from NADH to O_2 of 1.1 volts corresponds to a Keq of $\sim 10^{38}$, in favor of the direction of electron flow. That is a large enough energy gradient to pump protons anisotropically across the inner mitochondrial membrane

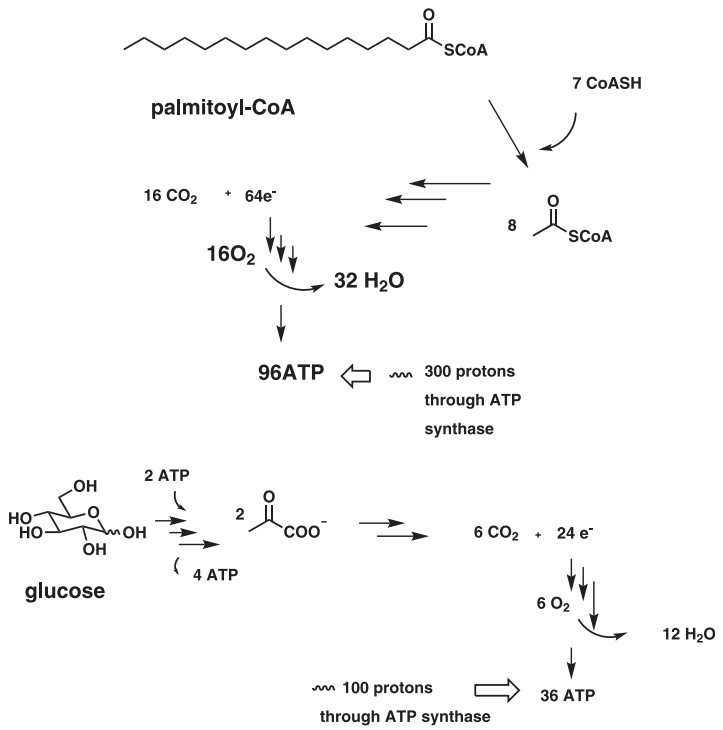


Figure 6.9 Glucose and palmitoyl CoA are two of the most common energy sources for cells, undergoing complete oxidation of every carbon to CO_2 (16 CO_2 from the sixteen-carbon fatty acid chain or six CO_2 from glucose, reflecting removal of 64 electrons and 24 electrons, respectively from the two substrates). The 64 electrons suffice to make 96 ATPs from flow of about 300 protons through the ATP synthase rotor. The 24 electrons from complete oxidation of glucose to six CO_2 similarly leads to ~100 protons flowing down their potential gradient to make 36 ATP.

into the intermembrane space during electron flow at strategic proton channels in transmembrane proteins (Figure 6.10A).

It is the buildup of protons in the intermembrane space that powers the F-ATPase to run in the ATP synthase direction. (Figure 6.10B) As the protons flow back into the matrix side down their concentration and potential gradient through the ATP synthase channel, they drive rotation of the membrane subunits and couple the otherwise unfavorable condensations of ADP and inorganic phosphate to form the Pβ–O–Pγ phosphoric anhydride bond as ADP goes to ATP.[15–18] This occurs probably at the level of ~70 kg per day in humans, $\sim 10^{26}$ molecules of ATP made by these proton flows.

Glucose, often the more immediate and quantitatively dominant energy source under many conditions, is likewise oxidized to six CO_2 molecules as the 24 electrons are extracted and stored as 12 pairs in 10 NADH and 2 $FADH_2$ equivalents. These are fed into the same electron transport chains in mitochondria, reducing six O_2 to 12 H_2O molecules (Figure 6.9). The yield of ATP

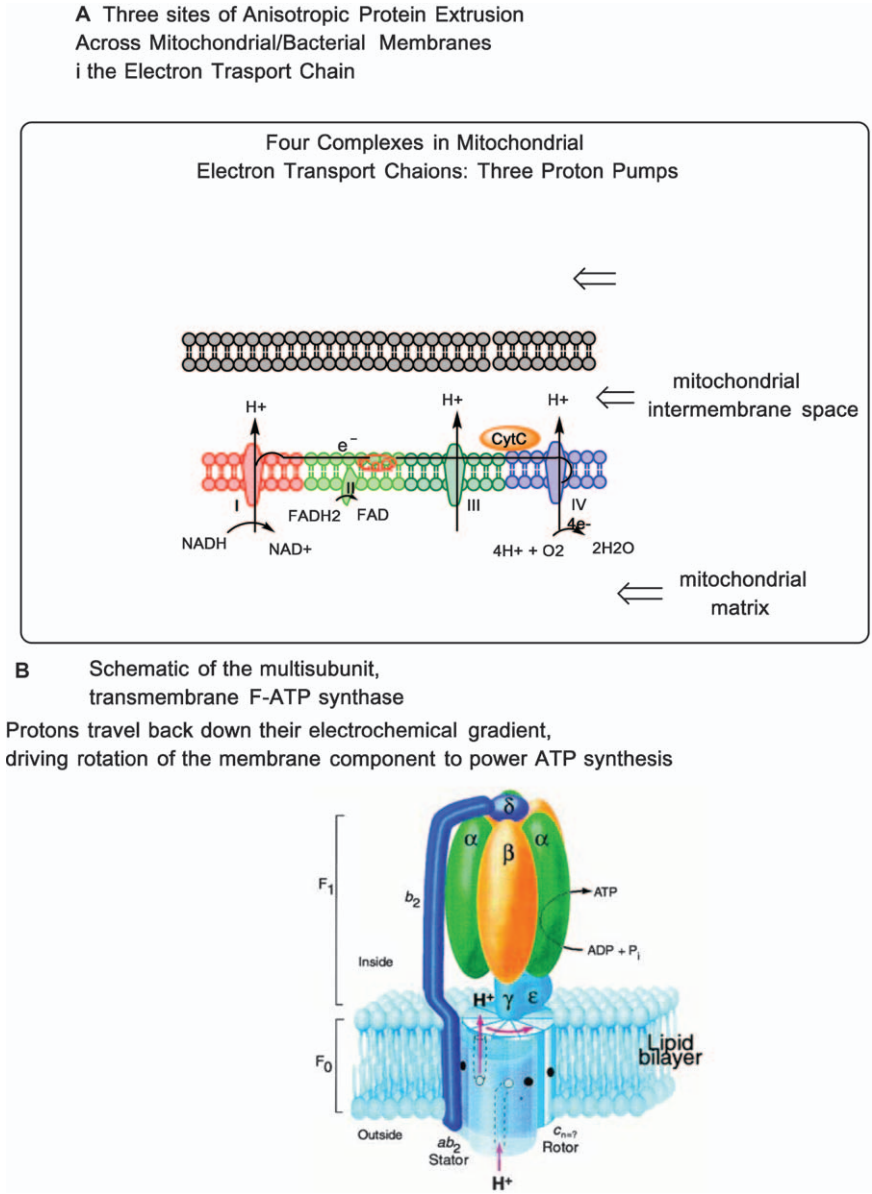


Figure 6.10 (A) Schematic of electron flow down the mitochondrial electron transport chain from NADH to O_2 in the mitochondrial matrix is accompanied by anisotropic pumping of protons into the intermembrane space at three transmembrane protein complexes (I, III, IV);[13] (B) Protons flow back from the intermembrane space to the lower concentration of protons in the mitochondrial matrix through a channel in the ATP synthase protein complex, powering the otherwise uphill condensation of ADP with Pi.[14] (A) Adapted from rref. 13, download for free at http://cnx.org/contents/185cbf87-c72e-48f5-b51e-f14f21b5eabd@10.8, under the terms of a CC BY 4.0 license, https://creativecommons.org/licenses/by/4.0/. (B) Reproduced from rref. 14 with permission from the AAAS, Copyright 1999.

varies from 34 to 38 molecules of ATP for every glucose molecule completely oxidized to six carbon dioxides, depending on rates of partially reduced oxygen species leakage, generally as one electron-reduced superoxide anions.

Figure 6.9 notes that complete oxidation of the fatty acid palmitate would yield 96 ATP molecules from letting ~300 protons flow through the F-ATP synthase channel. By comparison, the more oxidized hexose glucose could give 38 ATP if all six carbons are fully oxidized by four electrons each to six CO_2. That would occur by transit of about 100 protons through ATP synthase channels.

6.2.1 The Phosphorylation Logic of Glycolysis

The glucose oxidative metabolism logic is more revealing than fatty acid oxidation in terms of two phosphorylated intermediates in glycolysis that give rise to ATP. These are separate mechanisms from the proton gradient-driven F-ATP synthase. In order of appearance, they are the formation of an acyl phosphate (a mixed acyl–phosphoric anhydride) and then the formation of an enol phosphate, both thermodynamically activated phosphoryl transfer agents as denoted earlier in this chapter. The two enzymes phosphoglycerate kinase and pyruvate kinase harvest those activated phosphoryl groups by transfer to ADP to form ATP, with highly favorable equilibrium constants in the ATP-forming direction. (Figure 6.11).

It is worth examining the logic of phosphate group manipulations in the glycolysis pathway to see how these two ''high energy'' phosphate derivatives are produced and function as energy sources. Overall, glycolysis consumes two ATPs, within the first three steps of the pathway, in a sense as a priming operation (hexokinase, 6-phospho-fructose-1-kinase). Four ATPs then come back, for a net of two ATPs by the time the six-carbon glucose has been converted to two molecules of the three-carbon α-keto acid pyruvate (Figure 6.12).

The first enzymatic phosphorylation of free intracellular glucose occurs *via* hexokinase action at C6 of glucose, with Mg–ATP cleavage to ADP. The next enzyme isomerizes glucose-6-phosphate to fructose-6-phosphate, an aldo–keto isomerization that sets up the six-carbon ketose scaffold for subsequent fragmentation to two equal three-carbon pieces for most efficient downstream processing. The third enzyme, phosphofructokinase, spends the second ATP to generate fructose-1,6-bisphosphate. Both of the phosphoryl transfer steps are unexceptional transfers of the γ-PO_3^{2-} to oxygen nucleophiles, the first the C_6-primary alcohol of glucose, the second to a hemiacetal oxygen at C1 of fructose-6-phosphate (Figure 6.12).

The fructose-1,6-bisphosphate (FDP) is next subjected to enzymatic retroaldol cleavage by FDP aldolase. Aldol cleavages require a β-hydroxy-carbonyl part structure. That is provided by carbons 2–4 of FDP in the *acyclic ketose form (not in the prevalent hemiketal forms)* (Figure 6.12). The two three-carbon products are the aldehyde glyecraldehyde-3-phosphate and the ketose dihydroxyacetone-phosphate. These two are equilibrated by the next

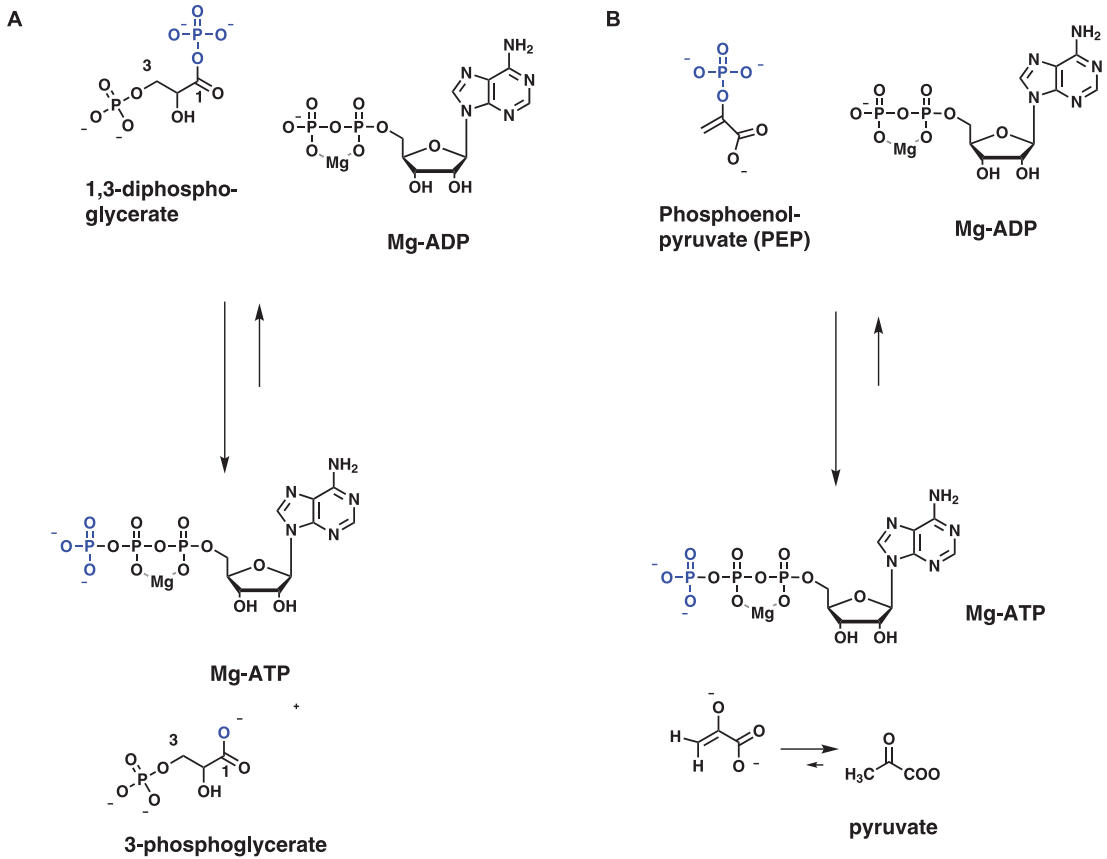


Figure 6.11 Two kinases in the glycolysis pathway from glucose to pyruvate run in the direction of ATP synthesis. (A) The first 3-phosphoglycerate kinase uses the acyl phosphate 1,3-diphosphoglycerate as a $-PO_3^{2-}$ donor to Mg–ADP. (B) The second kinase uses the trapped enol PEP as a $-PO_3^{2-}$ donor to Mg–ADP.

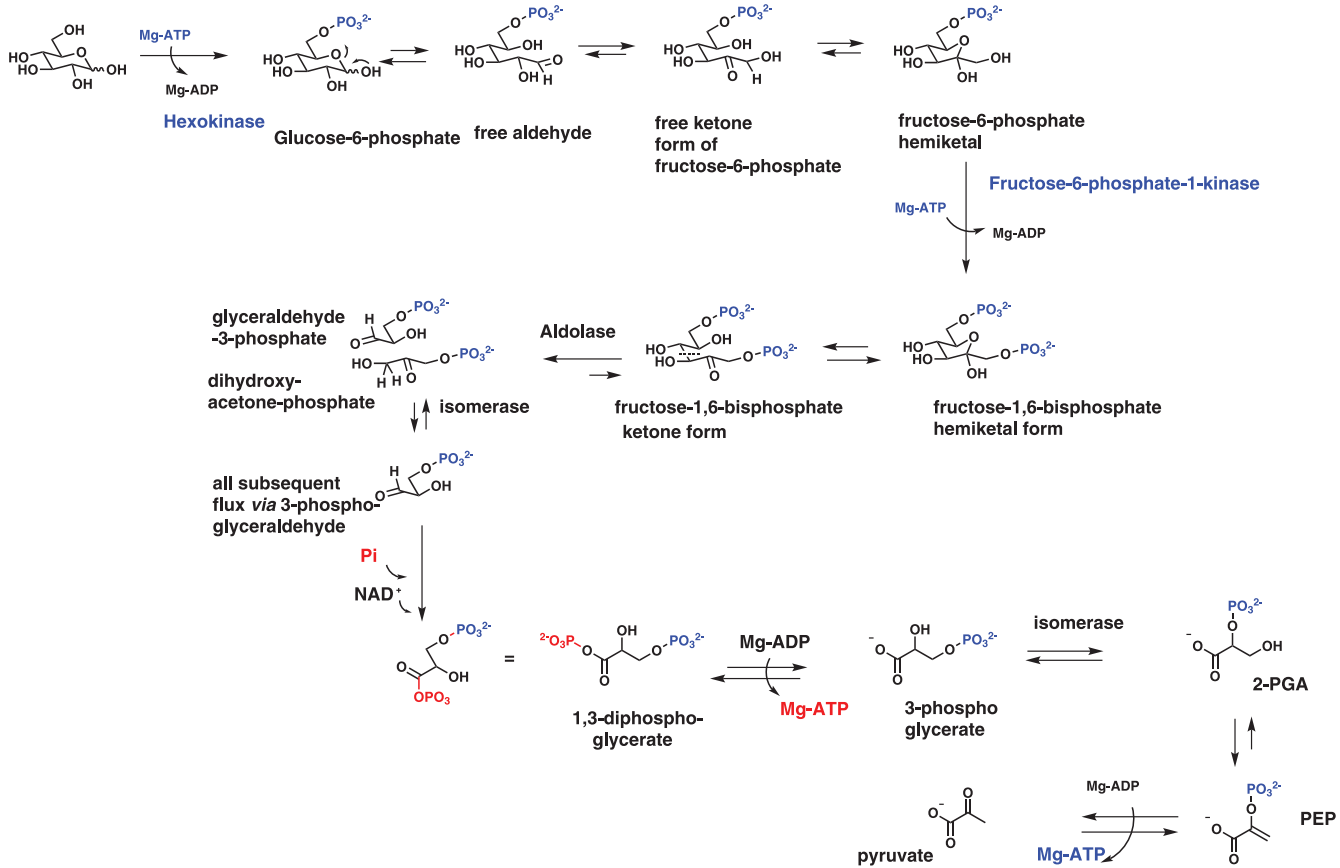


Figure 6.12 The ten-enzyme pathway of glycolysis converts glucose to pyruvate. Two ATPs are spent early, in the hexokinase and 6-phospho-fructokinase steps, to create the fructose diphosphate scaffold. Four ATPs come back, from PGA kinase and pyruvate kinase action, for a net of two ATPs produced.

enzyme, triose-phosphate isomerase, allowing subsequent enzymatic flux to go only through the glyceraldehyde-3-phosphate, while drawing off all the mass from the original glucose (Figure 6.12).

6.2.2 Glyceraldehyde-3-Phosphate Dehydrogenase as an Energy Capture Nanomachine

Now, the first of the ATP-generating enzymes comes into play and illustrates how the logic and mechanism of the sole redox step in the conversion of glucose to two molecules of pyruvate results in generation of a thermodynamically activated phosphate species. The enzyme glyceraldehyde-3-phosphate dehydrogenase uses NAD^+ as a cosubstrate and acceptor of a hydride ion (hydrogen with two electrons) as the aldehyde group of glyceraldehyde-3-phosphate gets oxidized by two electrons up to the acyl oxidation state.[1,19]

To capture some of the energy in a chemically useful form as the aldehyde oxidation proceeds, the enzyme acts to oxidize not the free aldehyde but instead a covalent enzyme–thiohemiacetal adduct. This is the key logic for energy harvesting (Figure 6.13).

An active site cysteine thiolate side chain adds reversibly into the aldehyde carbonyl of bound 3-phospho-glyceraldehyde to form a covalent tetrahedral adduct between substrate and enzyme. This hemithioacetal is the substrate for hydride transfer to bound NAD^+, constituting the oxidation step (Figure 6.13). NADH formation is coincident with conversion of the covalent thiohemiacetal adduct to a covalent acyl thioester: a 3-phospho-glyceryl-*S*-cysteinyl enzyme adduct.

We have noted in Figure 6.7 that both acyl thioesters and also acyl–phosphoric anhydrides are thermodynamically activated acyl groups. The Keq for hydrolyses are in the range of 10^5, a clear measure of acyl group activation. Thus, the strategy of glyceraldehyde-3-phosphate dehydrogenase to oxidize the thiohemiacetal to the acyl thioester is the energy capture step. If instead the aldehyde form of the bound substrate were oxidized, the result would be the resonance-stabilized 3-phospho-glycerate, the free acid. None of the acyl group activation energy would have been captured.

The problem the dehydrogenase now faces is that the activated three-carbon acyl group is covalently tethered to an active site cysteine residue. The enzyme cannot act catalytically until it gets back to a free cysteine residue in the active site. Separately, the energy-rich noncovalently bound NADH can dissociate stochastically to be replaced by the next molecule of NAD^+. The enzyme solves this catalysis-limiting issue by using inorganic phosphate as a third substrate. One of the oxyanions of a dianionic phosphate ion attacks the activated acyl group (Figure 6.13), with net transfer to Pi, to release 1,3-diphosphoglycerate. This is a *phosphorolytic* removal of the phosphoglyceryl acyl group from the enzyme. (see Chapter 9) The enzyme is now free to commence the next catalytic cycle.

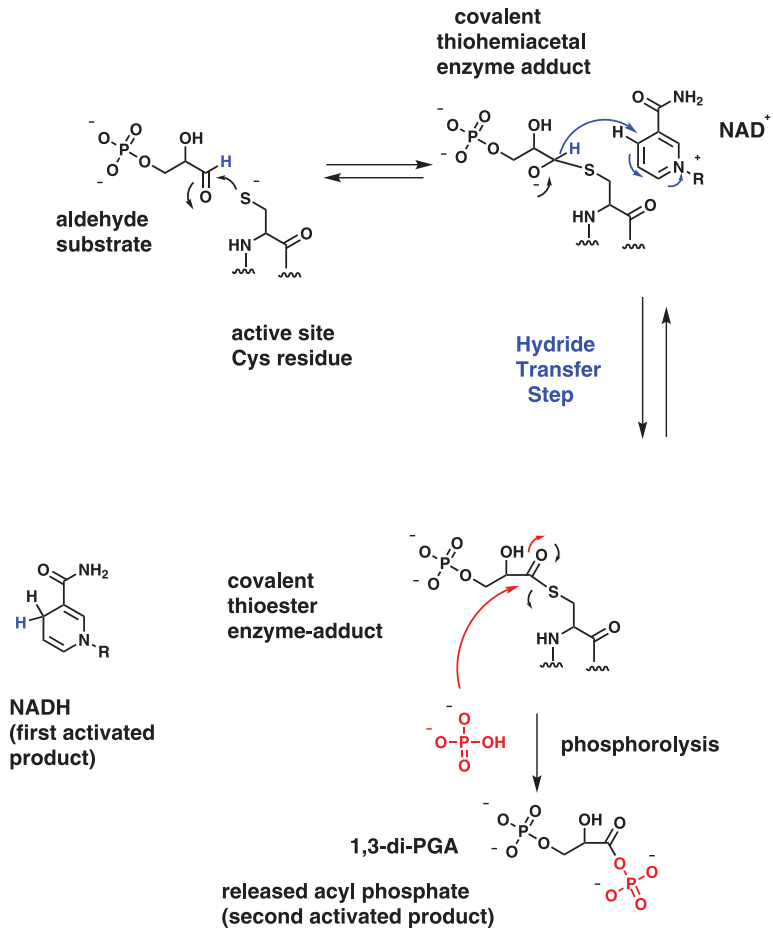


Figure 6.13 Glyceraldehyde-3-phosphate dehydrogenase is the first energy-capture catalyst in glycolysis. Oxidation of the thiohemiacetal substrate–enzyme adduct rather than free 3-phospho-glyceraldehyde, harvests energy in both products. NADH is a thermodynamically activated cellular redox currency. The 3-phospho-glyceryl-*S*-enzyme is an activated thioester, phosphorolyzed to yield the released 1,3-diphosphglycerate with acyl-phosphate linkage at C1.

Of the two phosphoryl groups in the 1,3-diphosphoglycerate product, the 3-phosphoryl group is a ''low energy'' alcohol–phosphate monoester. The newly generated 1-phosphate is a mixed acyl–phosphoric anhydride linkage. As an anhydride it is activated for transfer to nucleophiles. This is validated experimentally by the next (seventh) enzyme in the pathway. It is named 3-phosphoglycerate kinase but runs physiologically in the opposite direction: the formation of ATP from ADP and 3-PGA from the 1,3-diphospholycerate (Figure 6.14). This is an energy harvesting step (not an energy

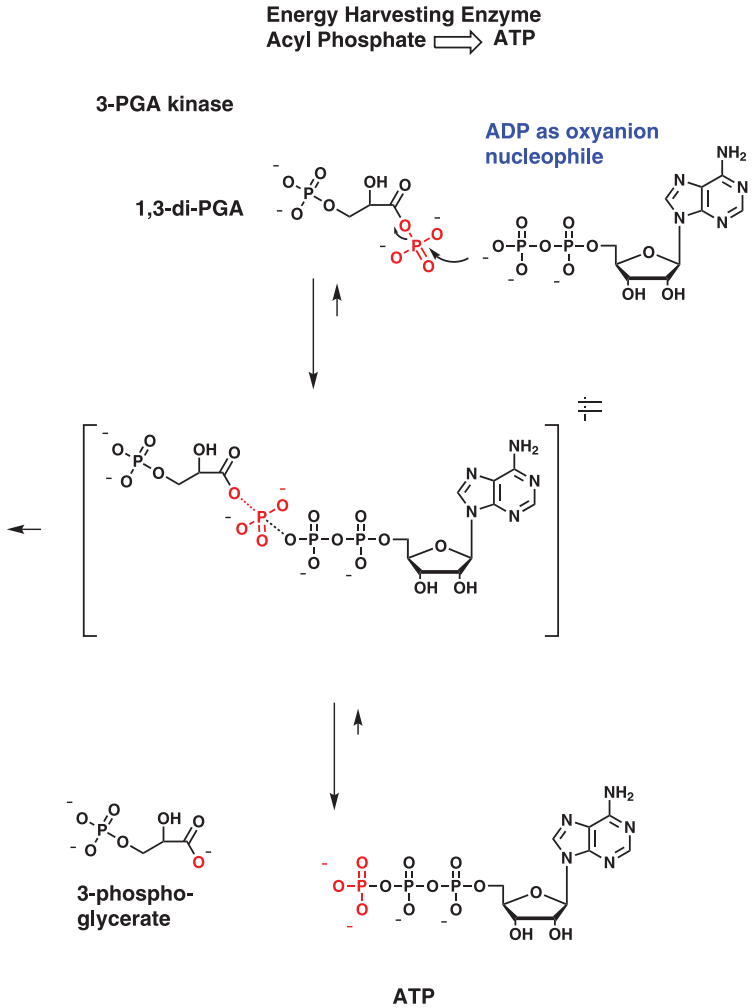


Figure 6.14 3-PGA kinase is misnamed in that it does not ordinarily run in the direction of ATP transfer of a $-PO_3^{2-}$ group to 3-phosphoglycerate. Instead, it runs in the opposite direction, reflecting the higher thermodynamic phosphoryl group transfer of the acyl-phosphate moiety compared with ATP. This enzyme converts the specialized acyl phosphate scaffold of 1,3-di-PGA to the common ATP as the cellular phosphoryl transfer currency.

capturing step), converting the kinetically labile acyl-phosphate to the kinetically stable workhorse energy supply ATP.

Of the several strategic lessons the 3-phosphoglycerate enzyme machinery displays for converting low energy phosphates to high energy phosphates for formation of ATP, four are worth explicit mention. First is the capture of the

two electrons removed from glyceraldehyde-3-phosphate in NADH, the common cellular electron transfer currency, set to send electrons down mitochondrial electron transport chains for the F-ATP synthase to function. Second is the oxidation, not of the 3-phospho-glyceraldehdye but instead the covalent thiohemiacetal to produce the activated acyl–thioester enzyme adduct. Third is the use of inorganic phosphate as an oxyanion nucleophile to remove the activated acyl group in an energy neutral acyl transfer (from thiolate to phosphate) to yield released 1,3-diphosphoglycerate.

Fourth, we note in passing in this transfer step the dehydrogenase is acting as a phosphorylase. That is, the activated acyl group is transferred out to inorganic phosphate. Had it been transferred to water, a hydrolase outcome, the thermodynamic activation would have been dissipated. Remarkably, the enzyme takes the resonance stabilized inorganic phosphate dianion–monoanion mixture and converts that phosphate to the phosphoryl group in the mixed acyl–phosphoric anhydride.

Finally, we note that both products, NADH and 1,3-di-PGA, are thermodynamically activated and represent capture of some of the oxidative energy in chemically useful cellular scaffolds, especially after 1,3-di-PGA is processed to 3-PGA and ATP. Since two molecules of the three-carbon 1,3-di-PGA scaffolds are created from each six-carbon starting glucose molecule, the return on ATP is two molecules in this step, balancing the two ATPs spent to phosphorylate glucose and then fructose-6-phosphate. The net energy balance of ATPs in *vs.* ATPs out to this point in glycolysis is zero.

The 3-phosphoglycerate is three enzymatic steps away from pyruvate, considered the end of the ten-enzyme cytoplasmic branch of glycolysis. Pyruvate subsequently enters mitochondria, is oxidatively decarboxylated to deliver both NADH and acetyl-CoA for its entry into the citric acid cycle.

The eighth enzyme in glycolysis, moves the phosphate group between the starting 3-position and the adjacent 2-OH of the glycerate scaffold, to equilibrate 3-PGA with 2-PGA. As in the earlier triose-phosphate isomerase step, all downstream flux proceeds from one isomer, the 2-PGA isomer.

6.2.3 Enolase as the Second Glycolytic Enzyme Creating a Thermodynamically Activated Phosphate Scaffold

The ninth enzyme of the glycolytic pathway is the enzyme enolase. As the name indicates it generates an enol product, in particular phosphoenolpyruvate (PEP) from the 2-PGA substrate.[20] (Figure 6.15) The results of mechanistic study have indicated that catalysis begins with proton abstraction from C_2 of 2-PGA, producing a carbanionic transition state. That incipient negative charge density at C_2 helps expel the adjacent C_3–OH. The net result is loss of an equivalent of water as a double bond is created across C_2–C_3. That new double bond is part of the phosphorylated C_2-enol scaffold in phosphoenolpyruvate.

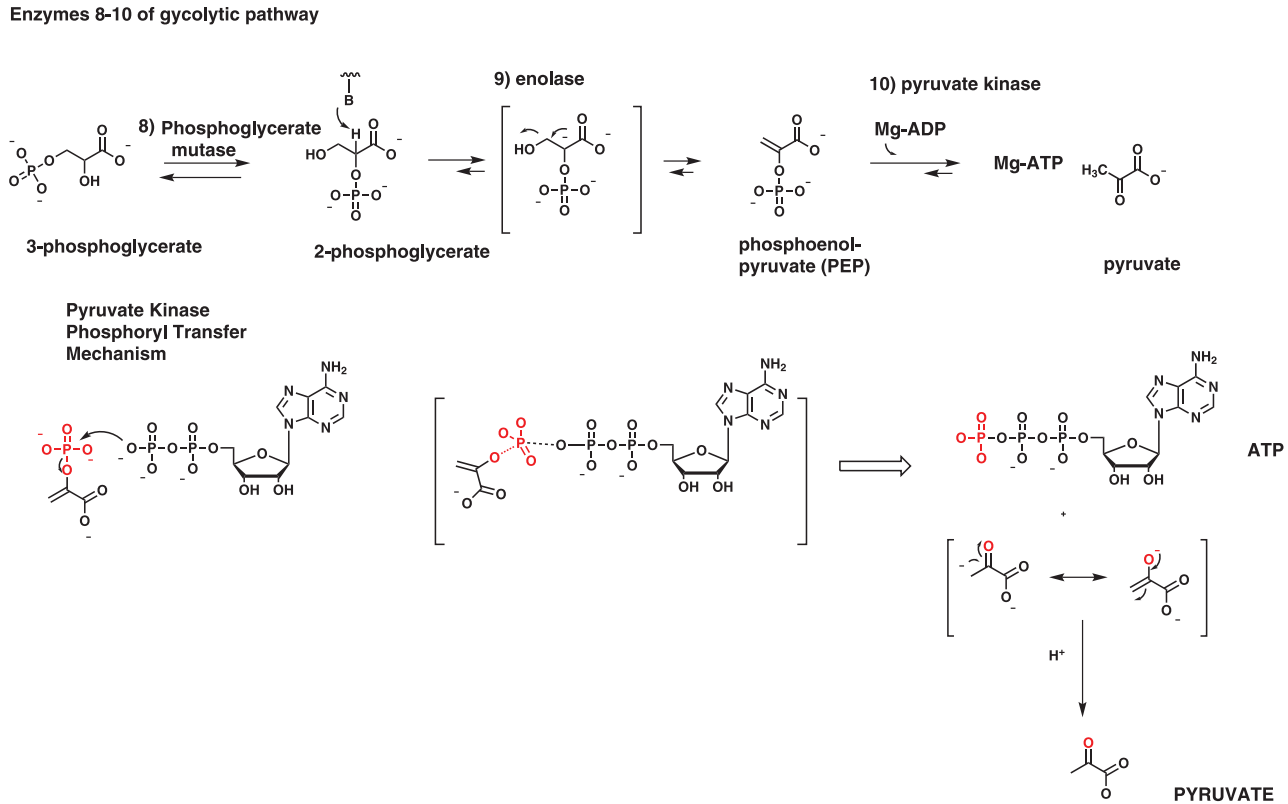


Figure 6.15 The final three enzymes of the ten-enzyme glycolysis pathway convert 3-PGA to 2-PGA, dehydrate 2-PGA to the trapped enol PEP, and then transfer the $-PO_3^{2-}$ group to ADP. The PGA mutase uses covalent phosphoryl enzyme logic to move the $-PO_3^{2-}$ group from the C2-OH to the C3-OH of PGA. Enolase effects a remarkable dehydration to convert a low energy 2-PGA scaffold to the trapped enol in PEP. The favorable conversion of PEP and ADP to pyruvate and ATP by pyruvate kinase again is an indication that the phosphoryl group transfer potential in PEP is greater than that of ATP.

The phosphoryl group represents a derivatization of the enolate in a trapped form of enolpyruvate. The enol is much less stable than the keto form. Thus, if the enolate were free it would isomerize to the more stable keto group in the molecule pyruvate such that the enolate would be at vanishingly low abundance. This is the structural basis for the phosphorylated, trapped enol in PEP as a thermodynamically activated phosphoryl transfer agent. The free keto acid pyruvate is much more stable.

Evidence for the high phosphoryl group transfer potential of the enol phosphate group in PEP is provided by the tenth (and last enzyme) of the glycolytic pathway. Pyruvate kinase is named for transfer of ATP to the enol form of pyruvate, but the enzyme runs physiologically in the direction of ATP synthesis, not ATP cleavage (Figure 6.15). One of the ADP β-phosphate oxyanions attacks the phosphoryl group of PEP, effecting a favorable $-PO_3^{2-}$ transfer and ATP formation. The immediate coproduct is the enolate anion of pyruvate. Its favored, nonenzymatic ketonization helps drive the forward equilibrium in the direction of ATP synthesis.

The net output of ATP formation: two ATPs from pyruvate kinase action and two ATPs from phosphoglycerate kinase action for a total of four ATPs produced for each six-carbon glucose converted by the ten enzymes of glycolysis to two pyruvates. Cells spend two ATPs to prime the first two kinases in the pathway, so the net output from glycolysis is two ATPs, from the four electrons removed from the glucose scaffold. The subsequent metabolic oxidation of each pyruvate molecule to three molecules of CO_2 yields the other 20 electrons originally in glucose. Those ten electron pairs can give 32–36 ATPs by F-ATP synthase action, as summarized above.

Thus, glycolysis produces about 1/17 to 1/19, or about 6–7%, of the ATP yield available from oxidative phosphorylation. This is a lean way to eke out a living, may have been a mainstay of some anaerobic ways of life, and shows the energy advantage in full aerobic oxidation of glucose to six CO_2.

In humans some stem cells in certain microanaerobic niches and some tumor cells practice glycolysis rather than respiration.[21] If there is enough flux down the glycolytic pathway, cells and tissues can survive. (As an aside: as pyruvate and NADH build up in glycolysis, the pathway becomes inhibited and electrons get dumped into pyruvate to produce lactate which is secreted from cells: think muscles that build up lactate as they go transiently anaerobic.)

In sum, cells have learned three ways of forming the P_β–O–P_γ anhydride bond in ATP, driving the unfavorable conversion of ADP and Pi to ATP. Two of them are chemical transfers of activated electrophilic phosphoryl groups from an acyl-phosphate (1,3-diphosphoglycerate) and from a trapped, phosphorylated enol (PEP). These are tandem energy-generating and energy-harvesting steps in glycolysis. These two strategies presumably were early success stories in the metabolic evolution of ATP generation and introduced the mixed acyl–phosphoric anhydrides and the phosphorylated enols as kinetically stable, thermodynamically activated forms of organic phosphate

in biological systems, similarly activated to the phosphoric anhydride side chain of ATP.

The third route to ATP synthesis dominates the energy calculus of aerobic metabolism and is likely to be the driving force for metabolic energy yields of aerobic organisms. The F-ATP synthase is clearly an evolutionarily related member of the transmembrane ion pumps, used in many subcellular compartments (Chapter 3). The ATP synthase directionality depends on running the ATPase-driven ion pumps in the opposite direction. The ability of the bacterial membranes, and the evolved mitochondria and chloroplasts to pump protons anisotropically into intermembrane spaces and then let them flow back down electrochemical gradients in the ATP synthase active site is the necessary precondition for providing enough energy to run ATP hydrolysis backwards. We have noted in Figures 3.1 and 3.2 that ATP synthase is unique in using an ADP-oxyanion to attack cosubstrate Pi and catalyze what amounts to a formal dehydration of that unactivated inorganic phosphate, an uphill process energetically. The downhill flow of protons is what drives the unfavorable coupling of Pi and ADP to yield ATP.

The thought that humans can run these proton pumps to make ~70 kg of ATP per day, replenishing the inventory of 75 g a thousand times is a stupendous advance over the substrate level production of ATP in glycolysis.

References

1. C. T. Walsh, B. Tu and Y. Tang, Eight kinetically stable but thermodynamically activated molecules that power cell metabolism, *Chem. Rev.*, 2018, **118**, 1460–1494.
2. C. T. Walsh and Y. Tang, *The Chemical Biology of Human Vitamins*, Royal Society of Chemistry, Cambridge, 2018.
3. C. C. Franklin, D. Backos, I. Mohar, C. C. White, H. J. HForman and T. J. Kavanagh, Structure, function, and post-translational modification of the catalytic and modifier subunits of glutamate cysteine ligase, *Mol. Aspects Med.*, 2009, **30**, 86–98.
4. M. Y. Galperin and E. Koonin, A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity, *Protein Sci.*, 1997, **6**, 2639–2643.
5. V. Gupta and R. Bamezai, Human pyruvate kinase M2 a multifunctonal protein, *Protein Sci.*, 2010, **19**, 2013–2044.
6. J. Deutscher, C. Francke and P. W. Postma, How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria, *Mircrobiol. Mol. Biol. Rev.*, 2006, **70**, 939–1031.
7. M. Saier Jr, The bacterial phosphotransferase system: new frontiers 50 years after its discovery, *J. Mol. Microbiol. Biol.*, 2016, **25**, 73–78.
8. K. K. Deol, S. Lorenz and E. R. Strieter, Enzymatic logic of ubiquitin chain assembly, *Front. Physiol.*, 2019, **10**, 835.

9. C. T. Walsh, T. Bugg, D. H. Kim and W. J. Lees, The versatility of phosphoenolpyruvate and its vinyl ether products in biosynthesis, *Chem. Biol.*, 1996, **3**, 83–91.
10. K. V. Venkatachalam, H. Akita and C. A. Strott, Molecular cloning, expression, and characterization of human bifunctional 3'-phosphoadenosine-5'-phosphosulfate synthase and its functional domains, *J. Biol. Chem.*, 1998, **273**, 19311–19320.
11. A. L. Lukowski, N. Denomme, M. E. Hinze, S. Hall, L. L. Isom and A. R. H. Narayan, Biocatalytic detoxification of paralytic shellfish toxins, *ACS Chem. Biol.*, 2019, **14**, 941–948.
12. W. R. Ellington, Evolution and physiological roles of phosphagen systems, *Annu. Rev. Physiol.*, 2001, **63**, 289–325.
13. (downloaded 11-01-2019): Biology for Majors 1; module 6 metabolic pathways. Figure 1-electron transport chain., lumenlearning.com/courses/biology-for-majors-i/ In Lumen.
14. H. Luecke, B. Schobert, H.-T. Richter, J.-P. Cartailler and J. K. Lanyi, *Science*, 1999, **286**(5438), 255–260.
15. P. D. Boyer, The ATP Synthase-a splendid molecular machine, *Annu. Rev. Biochem.*, 1997, **66**, 717–749.
16. D. Stock, A. Leslie and J. E. Walker, Molecular architecture of the rotary motor in ATP synthase, *Science*, 1999, **286**, 1700–1705.
17. R. L. Cross and V. Muller, The evolution of A-, F- and V-type ATP synthases and ATPases: reversals in function and changes in H + /ATP coupling ratio, *FEBS Lett.*, 2004, **576**, 1–4.
18. G. Cingolani and T. Duncan, Structure of the ATP synthase ctalytic complex (F1) from *E.coli* in an auto-inhibited conformation, *Nat. Struct. Mol. Biol.*, 2011, **8**, 701–707.
19. C. Tristan, N. Shahanin, T. W. Sedlak and A. Sawa, The diverse functions of GAPDH: views from different subcellular compartments, *Cell. Signal.*, 2011, **23**, 317–323.
20. G. H. Reed, R. Pvner, T. M. Larsen, J. E. Wedekind and I. Rayment, Structural and mechanistic studies of enolase, *Curr. Opin. Struct. Biol.*, 1996, **6**, 736–743.
21. N. S. Chang and H.-H. Ng, The metabolic programming of stem cells, *Genes Dev.*, 2017, **31**, 336–346.

CHAPTER 7

Phosphomonoesters: Enzymatic Formation and Decomposition

To fill out the chemical biology of inorganic phosphate, inorganic pyrophosphate, and the phosphoric anhydride backbones of the side chain of ATP and cognate nucleoside triphosphates, most of the rest of phosphate metabolism occurs as phosphoesters (see Chapter 1 Section 1.10). Phosphoesters form by enzymatic condensation of an alcohol oxygen atom with a phosphoryl group (PO_3^{2-}), most often from attack on Pγ of Mg–ATP (Figure 7.1).

In Chapter 3 we focused on ATP as the donor of phosphoryl groups to cosubstrate alcohols, the favorable thermodynamics for transfer of the $-PO_3^{2-}$ group, and its role as an electrophilic fragment in flight. We also noted that many phosphorylated metabolites undergo subsequent enzymatic transformations where the $-PO_3^{2-}$ is transferred on to water. These net dephosphorylations constitute the back half of phosphoryl group transfer metabolism, including (as presented in Section IV) up to ~250 000 dephosphorylations of P-Ser, P-Thr, and P-Tyr side chains in the human proteome. All these phosphoryl transfers to water molecules are catalyzed by the large class of enzymes known as *phosphatases*. They are a main focus of this chapter.

In keeping with the fact that phosphorus in the +5 oxidation state in inorganic phosphate has one double bonded oxygen and three –OH bonds, phosphomonesters, phosphodiesters, and phosphotriesters are all possible (Figure 7.1) from successive esterification with alcohols. We will note that the chemical biology of phosphomonoesters and phosphodiesters are prevalent, with only a small number of phosphotriester metabolites and most of these are of xenobiotic origin (*e.g.* pesticides and nerve gases).

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

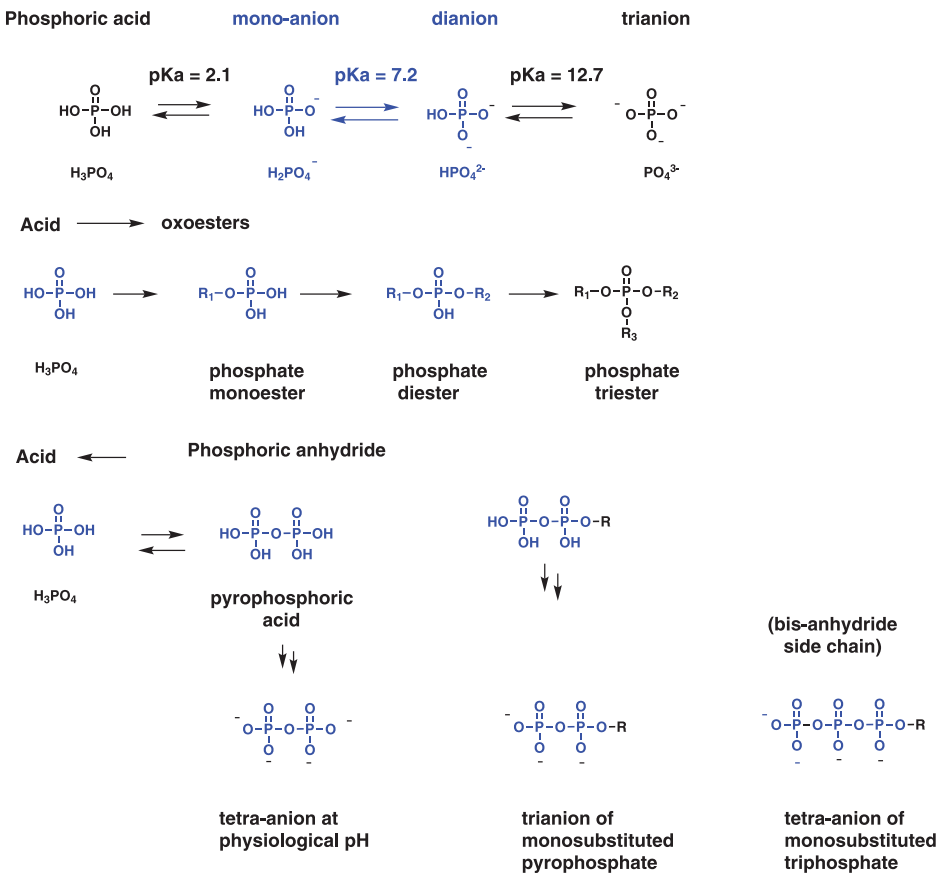


Figure 7.1 The tribasic inorganic phosphoric acid, H_3PO_4 can undergo three successive ionizations, from acidic to neutral to basic pH. Analogously, alcohols can react with phosphoric acid, once to generate phosphomonoesters, twice to phosphodiesters, or three times to phosphotriesters. The monoesters and diesters are common biological metabolites, the phosphotriesters are not.

Phosphomonoesters typically arise by kinase action: mediating an alcohol substrate attacking Pγ of Mg–ATP. (Some classes of substrates that can give accessible carbocations can be captured by inorganic phosphate in the active sites of phosphorylases—see Chapter 9).

Correspondingly, the two predominant classes of biological phosphodiesters – the internucleotide 3′5-phosphodiester bonds in RNA and DNA and the cyclic nucleotides such as cAMP and cGMP – typically arise by enzymatic nucleotidyl transfers. Those occur by attack of an alcohol oxygen nucleophile on Pα of the cosubstrate NTP or 2′deoxyNTP (Figures 7.2 and 7.3). We reprise and summarize the catalytic machinery for formation and breakdown of some specific phosphomonoesters in this chapter and then phosphodiesters in Chapter 8.

7.1 Phosphomonoesters: Formation, Function, Hydrolysis

Dozens of phosphate monoesters are formed in cellular metabolism of low molecular weight compounds and hundreds of thousands of phosphoprotein monoesters in phosphoproteomes. The latter topic is covered separately in Section IV. While the great majority of phosphomonoesters arise by kinase action from ATP, there are key phosphate esters generated by action of phosphorylase enzymes, these will be covered in Chapter 9.

Sugar metabolism may be the richest in phosphate ester deployment. We have noted in Chapter 6 how glucose-6-phosphate and then fructose-6-phosphate are phosphorylated by kinases in the two of the first three steps of glycolysis. These are often regarded as ionized handles to keep the otherwise neutral sugars trapped within cells. Furthermore, we have seen in Chapter 6 how four of the subsequent seven enzymes of the glycolytic pathway turn those ''low energy'' phosphomonoesters into ''high energy'' acyl phosphate or enol phosphate, capable of driving ATP synthesis. These are remarkable chemical transformations for which glyceraldehyde-3-phosphate dehydrogenase and enolase are energy-capturing nanomachines.

Two alternative metabolic branchings from the initial glucose-6-phosphate metabolite also reveal the utility of the phosphomonoester functionalities. One involves the conversion of glucose-6-phosphate to glucose-1-phosphate to vector hexose units into pathways that store the glucosyl units as energy reserve polymers such as starch (plants) or glycogen (animals) (Figure 7.4). The other branching is named the pentose phosphate pathway. It involves an enzymatic oxidative decarboxylation route to ribose-5-phosphate. We have noted in the pyrophosphoryl transferase section of Chapter 5 how 5-phospho-ribose is then converted to 5-phospho-ribose-1-pyrophosphate (PRPP) as the active form of ribose for biosynthesis of RNA and DNA nucleotide monomers (Figure 7.4).

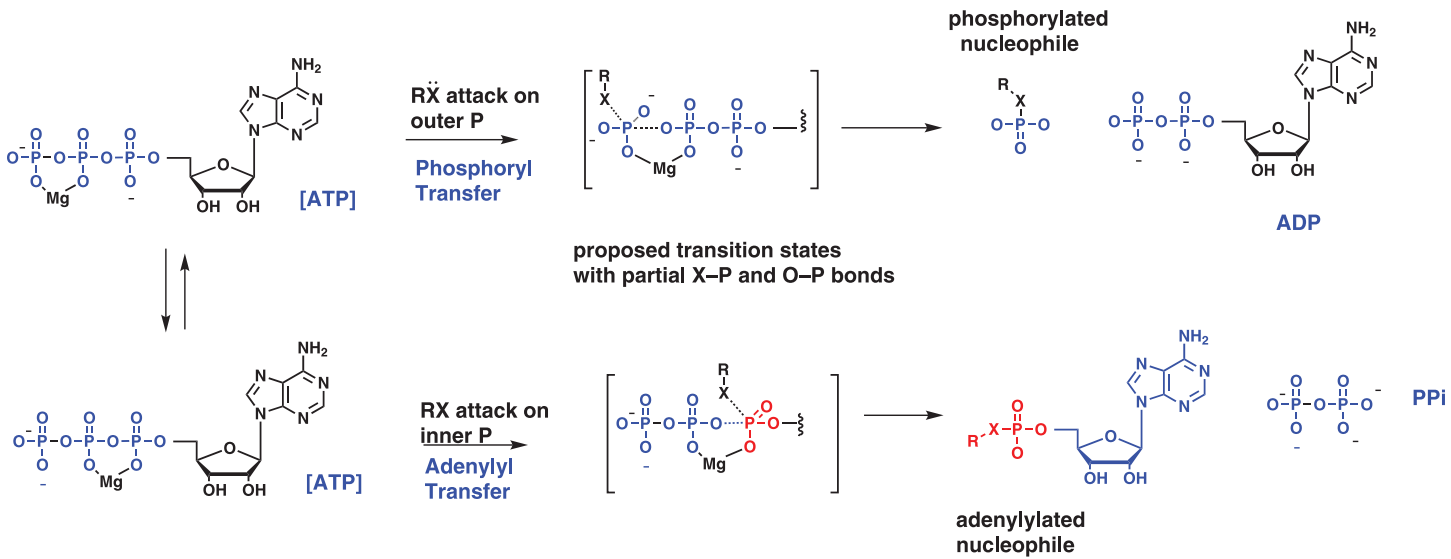


Figure 7.2 When alcohols are the cosubstrates, enzymatic phosphoryl transfers generate phosphomonoesters.

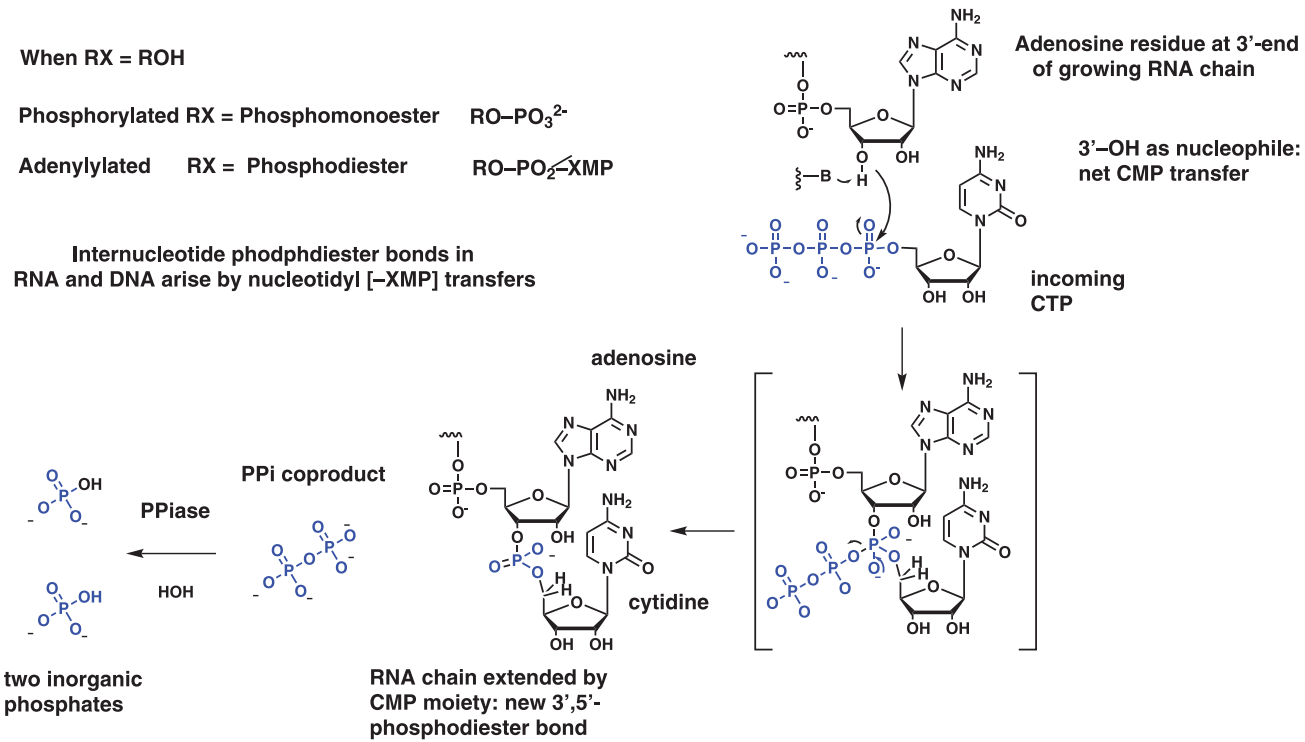


Figure 7.3 Adenylyl transfers instead generate phosphodiesters.

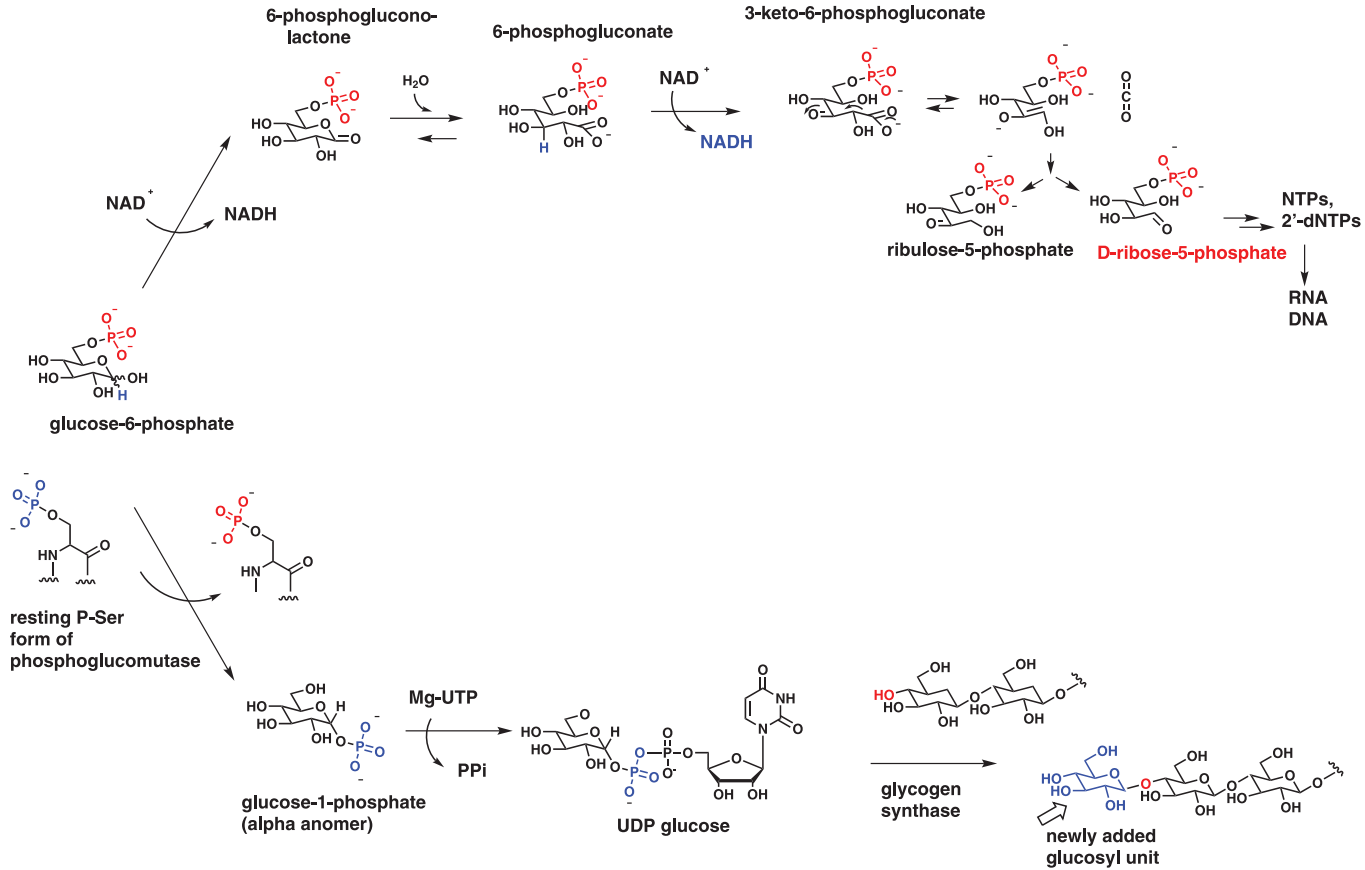


Figure 7.4 Two pathways for glucose-6-phosphate that compete with and partition the substrate away from glycolysis. The upper (pentose-phosphate) pathway is an oxidative decarboxylation route to convert the six-carbon aldohexose framework of glucose-6-phosphate to the five-carbon aldose scaffold of ribose-5-phosphate. The bottom path converts glucose-6-phosphate to glucose-1-phosphate on the way to UDP glucose and then oligosaccharide and glycoprotein biosynthesis.

Between the glycolytic pathways and the pentose phosphate pathways, the three-carbon aldose glyceraldehyde-3-phosphate, the four-carbon erythrose-4-phosphate, the five-carbon ribose-5-phosphate, and the six carbon glucose-6-phosphate are all featured as sugars with an aldehyde at C1 and a phosphomonoester at the other, primary alcohol terminus (Figure 7.5).

We have also discussed the fate of such phosphorylated hexoses as glucose-1-phosphate, galactose-1-phosphate, and mannose-1-phosphate as substrates for nucleotidyl transferases. These yield UDP–glucose, UDP–galactose and GDP–mannose, respectively, as the activated forms of hexose units electrophilic at the C1′ carbon for chain elongation reactions, *e.g.* for glycogen, starch, or cellulose synthesis (Chapter 4).

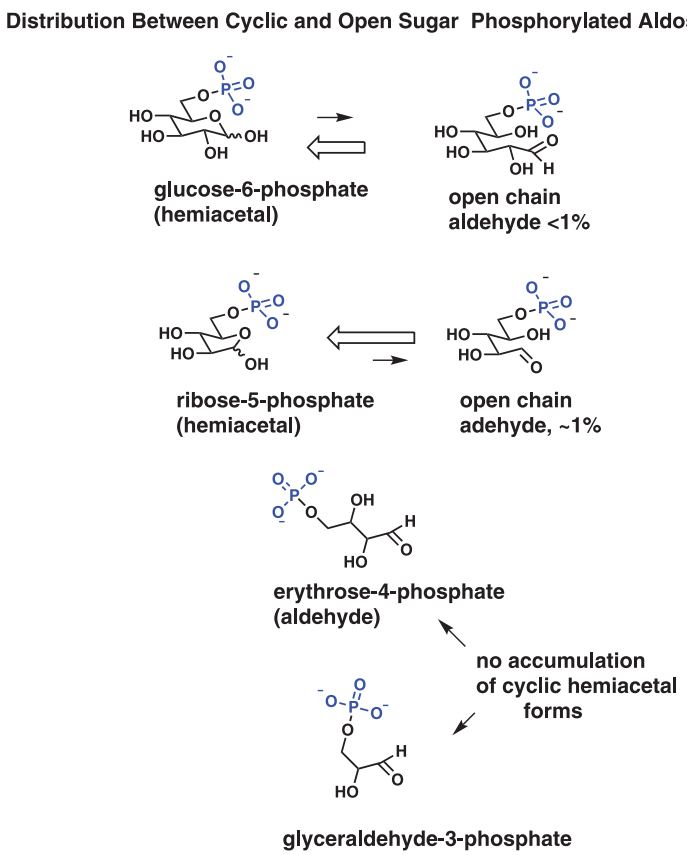


Figure 7.5 Primary carbohydrate metabolism involves sugar scaffolds from C3 to C6 that each have a phosphomonoester on the primary alcohol group of the sugar and an aldehyde group at C1. The two functional groups enable a wide range of metabolic scaffold transformations.

7.2 Phosphoglucomutase: Connecting Glycolysis to Hexose Oligomerizations

Phosphoglucomutase is not a *net* phosphorylating enzyme. The compositional phosphorylation state of the substrates glucose-6-phosphate and glucose-1-phosphate does not change. The regiochemistry does. On the surface the stoichiometry might suggest the $-PO_3^{2-}$ shuttles between C6–OH and C1–OH of the glucose framework. In reality the transformation is substantially more complex. The enzyme is a phosphoryl transferase but not a kinase (no ATP utilization).

The strategy for converting glucose-6-phosphate, the first trapped form of glucose once it enters cells, into glucose-1-phosphate as the hexose derivative committed to polysaccharide and protein glycosylations, is an intriguing one.[1] The enzyme phosphoglucomutase equilibrates glucose-6-phosphate with glucose-1-phosphate by starting each catalytic cycle with a $-PO_3^{2-}$ moiety covalently bound to an active serine side chain hydroxyl: a phosphoseryl enzyme as resting form of the catalyst (Figure 7.6). As glucose-6-phosphate binds, the C1′-OH of the α-anomer is within attacking distance of the phosphoserine group. As depicted in Figure 7.6, phosphoryl group transfer occurs to yield bound glucose-1,6-diphosphate transiently and the dephosphorylated serine residue. The trick from here is for the newly formed 1,6-diphosphoglucose then to rotate within the active site, perhaps driving an altered active site conformation, such that the 6-phosphate of glucose-1,6-diphosphate comes close enough for the seryl residue –OH side chain to act as a nucleophile. This constitutes a second phosphoryl group transfer, regenerates the enzyme–Ser–OPO_3^{2-} for the next catalytic cycle and yields bound gucose-1-phosphate. Its dissociation completes the phosphoglucomutase catalytic cycle. Presumably as the newly synthesized mutase enzyme molecules come off the ribosome and fold, they are in inactive dephospho forms. An initial priming reaction with glucose-6-phosphate would produce glucose and the active post-translationally modified enzyme–PO_3^{2-} (Figure 7.6).

Phosphoglucomutase thus utilizes an ingenious route to convert glucose-6-phosphate molecules to glucose-1-phosphate molecules. There is no direct transfer of a given $-PO_3^{2-}$ group from C6 to C1 or *vice-versa*. Geometric considerations may prevent such direct intramolecular transfer between those two hydroxyls. One could imagine evolution of an active site to bind two hexose-phosphate molecules and enable an intermolecular PO_3^{2-} transfer. The extant enzyme–PO_3^{2-} route, binding one molecule of either sugar–phosphate regioisomer, may be more economical. The enzyme carries out two phosphoryl group transfers in each catalytic cycle: from active site serine to glucose (at C1 or C6) and then back.

Vitamins undergo enzymatic alcohol phosphorylations from ATP; pyridoxine to pyridoxal-phosphate, riboflavin to FMN[2] (Figure 7.7). The amino acid serine and threonine are O-phosphorylated by ATP-dependent kinases

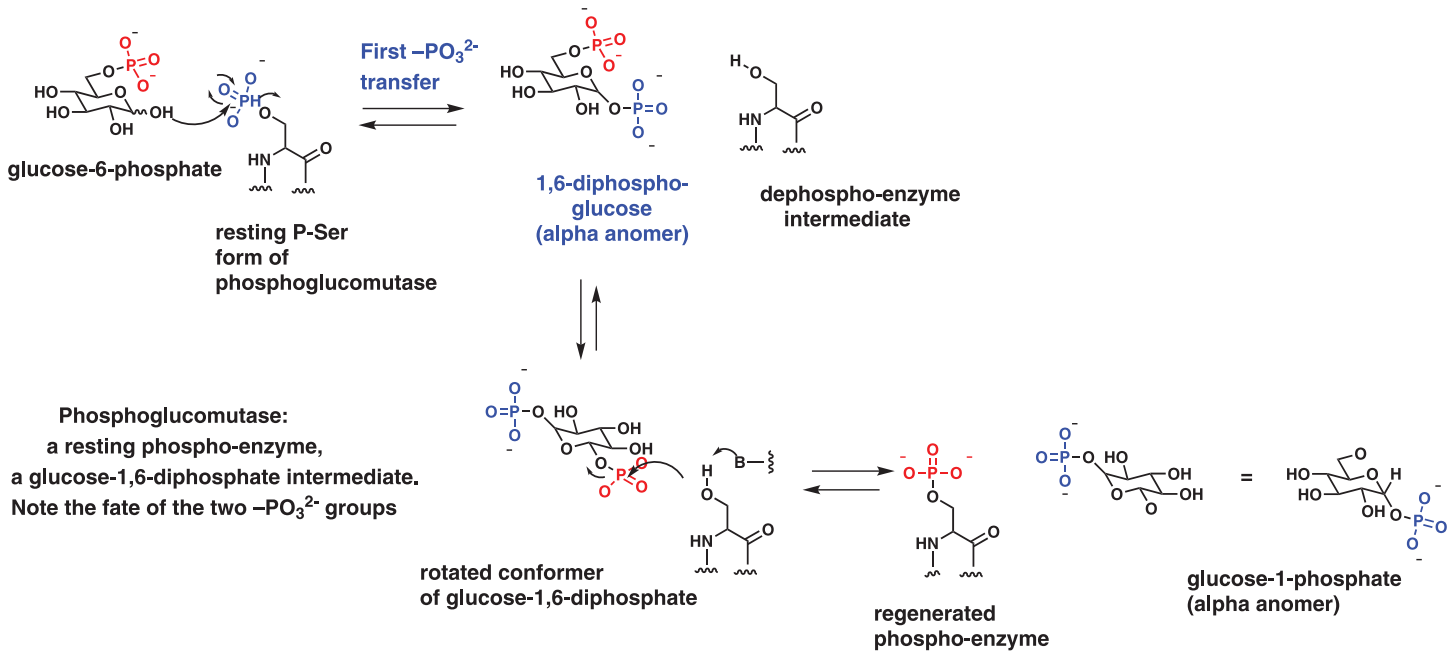


Figure 7.6 Phosphoglucomutase starts each catalytic cycle as a phosphoseryl enzyme species. When glucose-6-phosphate or glucose-1-phosphate binds to the active site, the enzyme–PO_3^{2-} group is transferred to substrate, to generate transiently the dephosphoenzyme and 1,6-diphosphoglucose. Attack of the active site Ser–OH then retrieves a phosphoryl group from the diphosphoglucose scaffold. There must be motion of the 1,6-diphosphoglucose intermediate in the active site to offer either the 1-phosphate or 6-phosphate in the $-PO_3^{2-}$ transfer that completes catalytic cycles.

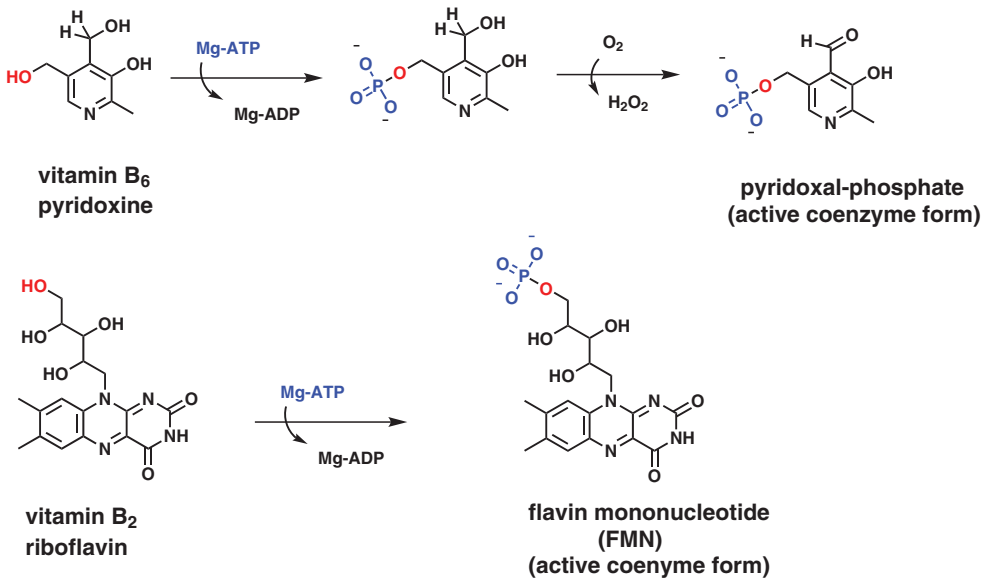


Figure 7.7 Vitamins such as B2 (riboflavin) and B6 (pyridoxine) are ingested as the neutral forms. The biologically active coenzyme forms require enzymatic phosphorylation to gain high affinity for their target enzymes. Pyridoxine kinase and riboflavin kinase are two such catalysts.

as preamble to subsequent eliminations of inorganic phosphate, thus achieving C–O bond cleavage in the starting serine or threonine scaffold. C–O Bond cleavage in mevalonate-5-pyrophosphate is effected after enzymatic phosphorylation of the tertiary 3-OH, concomitant with decarboxylation to generate Δ^3-isoentenyl-pyrophosphate[3] (Figure 7.8).

7.3 PEP Activated for P–O or C–O Bond Cleavage

The pyruvate kinase reaction is a classic example of how the monophosphate ester linkage in the trapped enol scaffold has set up PEP for P–O bond cleavage and phosphoryl transfer to ADP acting as oxygen nucleophile. A second mode of reactivity of PEP, net C–O bond cleavage with release of inorganic phosphate, is seen in two bacterial enzymes. As discussed in Chapter 6, one is involved in the central biosynthetic pathway to aromatic compounds including phenylalanine and tyrosine. The other is a key enzyme in forming the peptidoglycan layer of bacterial cell walls (Figures 6.5 and 6.6).

7.4 Phosphorylated Lipids

In Chapter 4 on formation of scaffolds with disubstituted pyrophosphate frameworks, we noted formation of phosphorylcholine from ATP and choline in a kinase reaction.[4] The phosphoryl choline then serves as nucleophile at the phosphoryl group oxyanion to attack Pα of CTP in a cytidylyl transfer enzyme active site (Figure 4.35). A lipid kinase acting on the hydroxyl group of sphingosine creates sphingosine-1-phosphate as a lipid messenger molecule in cellular signaling[5] (Figure 7.9).

7.4.1 Life Cycle of a Phosphatidylinositol Molecule

Although phosphatidylinositols are minor membrane phospholipid constituents (1–4%) in many cell membranes, they are enriched to ~10% abundance in brain membrane fractions. As will be noted in the next chapter, phosphatidylinositols are favored substrates for the action of phospholipase C, acting as a phosphodiesterase. This class of membrane phospholipid is also typically enriched in animals with the tetra-unsaturated C_{20} fatty acid arachidonate ($\Delta^{5,8,11,14}$-C_{20}) in ester linkage at the 2-positon of the glycerol phosphate backbone. On action by phospholipase A, the released arachidonate is precursor to the whole family of prostaglandins and prostacyclins, a class of lipids that act as autocrine and paracrine signaling agents[6] (Figures 7.10–7.12).

It is worth a brief detour to consider the enzymatic formation and breakdown of phosphatidylinositol and its phosphorylated downstream metabolites to see the full sweep of phosphorus chemical biology in play in this one molecular lipid class that sits at so many regulatory and signaling intersections.[7–9]

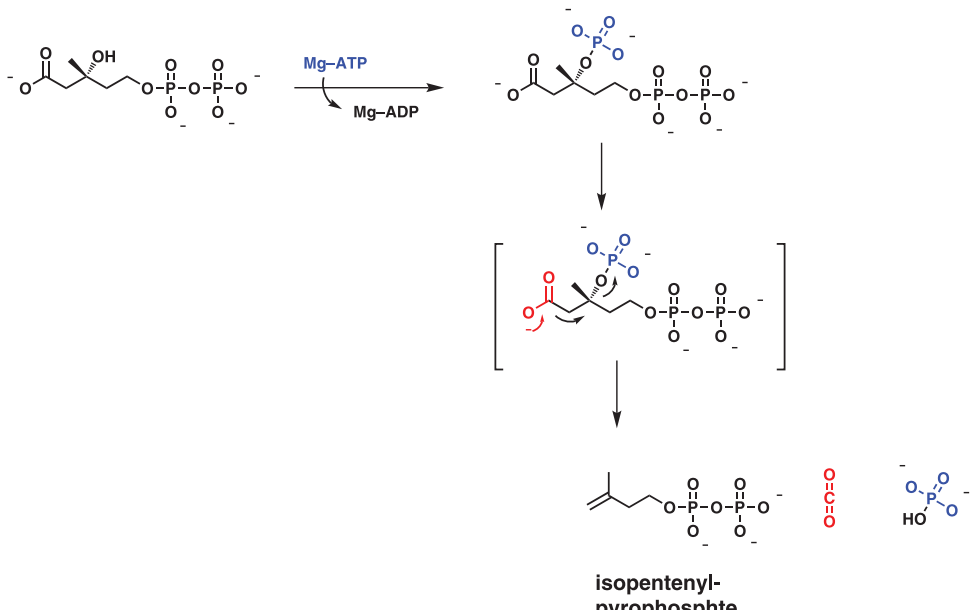


Figure 7.8 In the biosynthesis of more than 50 000 isoprenoid natural products, a key early step in the pathway involves decarboxylation of the six-carbon mevalonate-1-pyrophosphate metabolite to the five carbon Δ^3-isopentneyl-pyrophosphate. Along with the C–COO bond breakage is elimination of the C3-tertiary-OH to create the olefinic link in the isopentenyl-pyrophosphate product. ATP-dependent phosphorylation of that alcohol sets up a low energy barrier for concerted decarboxylation and Pi elimination.

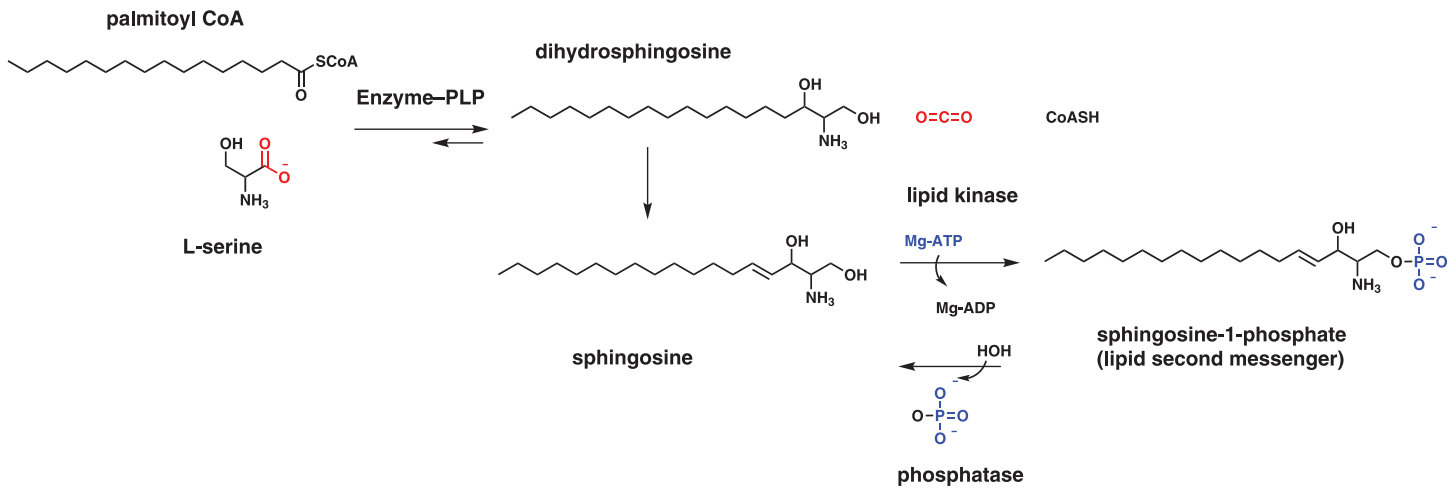


Figure 7.9 Enzymatic phosphorylation is key to the activity of the lipid signaling molecule sphingosine-1-phosphate. The lipid kinase acts on free sphingosine and its lifetime can be controlled by a lipid phosphatase.

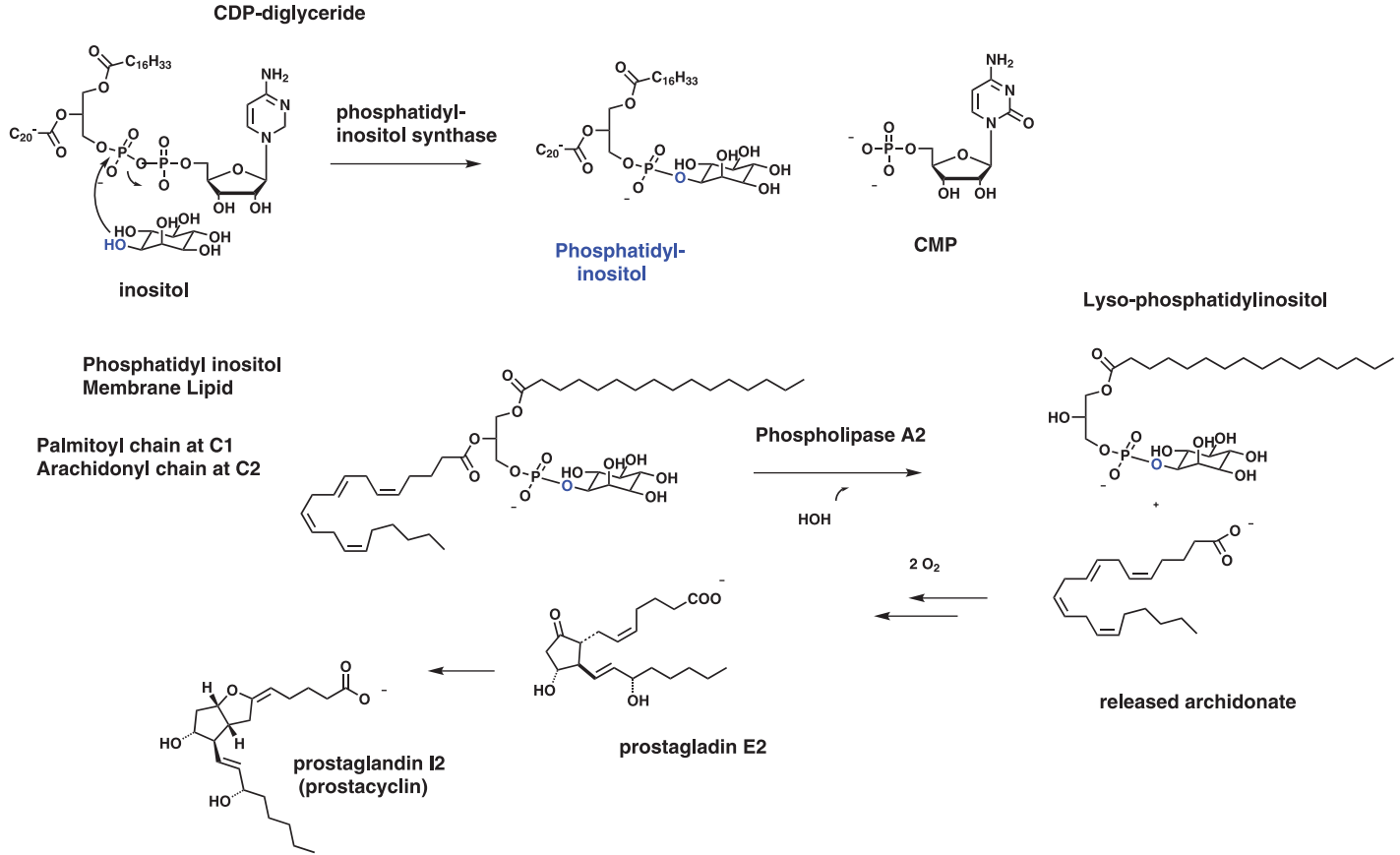


Figure 7.10 The C_{20} tetraenoic acid ($\Delta^{5,8,11,14}$) arachidonate is usually esterified at the 2-position of the minor membrane phospholipid phosphatidylinositol (PI). The arachidonyl ester can be cleaved by action of PI-specific phospholipase A2 to the free C_{20} acid. The released arachidonate serves as pathway-initiating substrate for a host of prostaglandins (including the bicyclic prostacyclins).

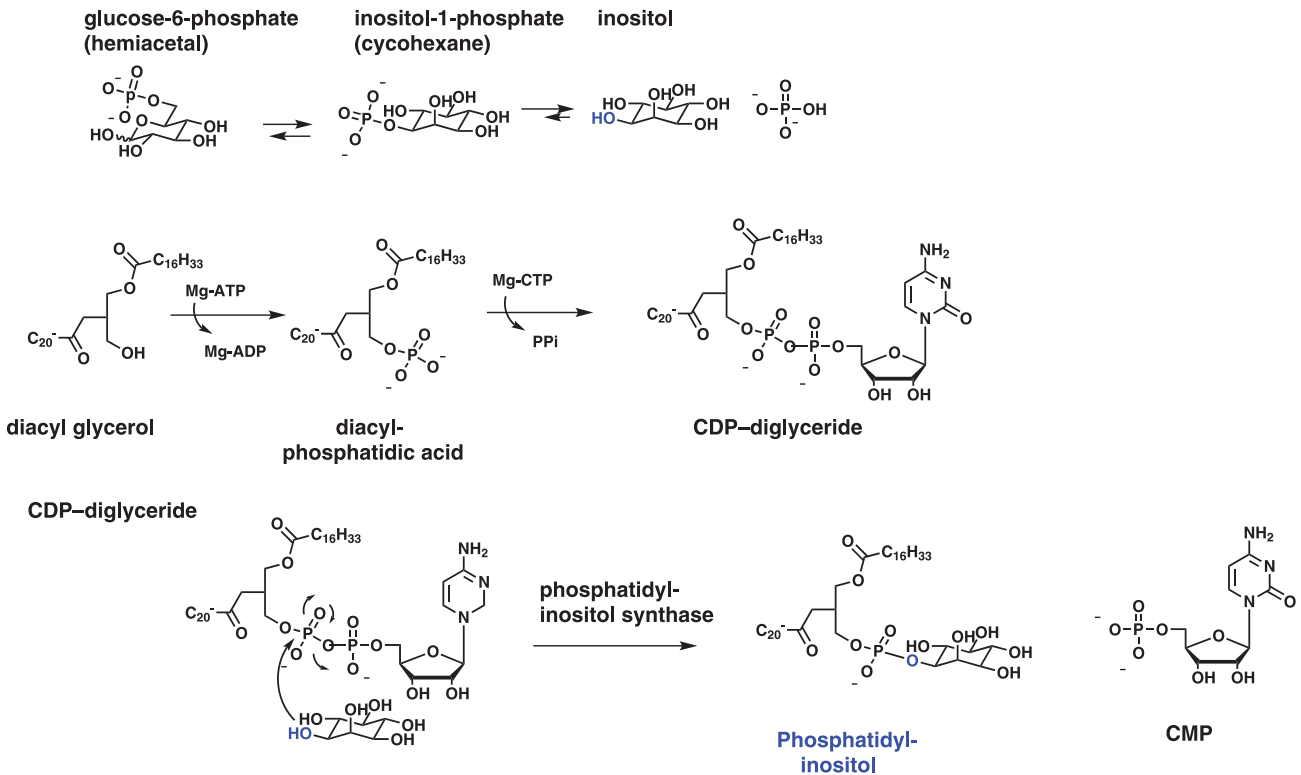


Figure 7.11 Life cycle of a phosphatidyl inositol membrane phospholipid. Biosynthesis of inositol proceeds from glucose-6-phosphate, *via* formation of the hydroxycyclohexane scaffold of inositol-1-phosphate and then phosphatase action to release free inositol.

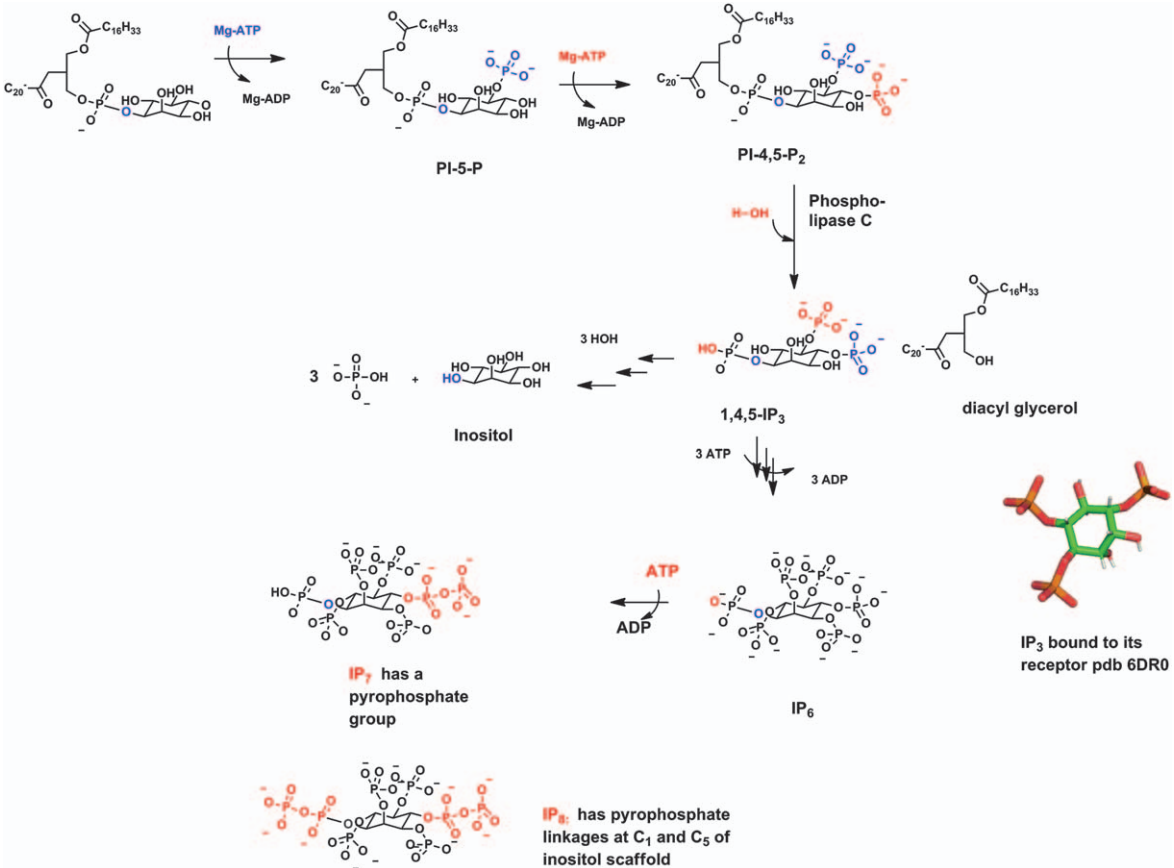


Figure 7.12 Life cycle of a phosphatidyl inositol membrane phospholipid. PI can undergo tandem phosphorylation by lipid kinases to PI-4,5-P_2. PI-4,5-P_2 is then a substrate for PI-selective phospholipase C isoenzymes to cleave that substrate to diacylglycerol and IP_3, each a signaling molecule. In turn IP_3 can be further phosphorylated, all the way to IP_6 and even IP_7 and IP_8 that bear unusual pyrophosphate groups at C5 or C1 and C5.

The inositol moiety is an unusual sugar in that it has a cyclohexane (all carbon) framework rather than a typical aldose or ketose scaffold. Biosynthesis proceeds from gucose-6-phosphate through an internal redox and aldol condensation sequence to yield myoinositol-1-phosphate. Phosphatase action then releases Pi and free inositol to serve as nucleophile towards CDP diglyceride in phosphatidylinositol synthase active sites.

In humans the conversion of free inositol to a form activated for transfer follows a typical strategy. The C_1–OH of inositol is selected as a kinetically competent nucleophile from among the five other –OH groups by phosphatidyinositol synthase to generate the PI product. The cosubstrate that is attacked is CDP–diacylglycerol so the product is phosphatidylinositol and CMP (Figure 7.11). CDP–diglyceride in turn arose from the phosphoryl group oxyanion of phosphatidic acid attacking Pα of CTP in a nucleotidyl transfer reaction. Thus, the involvement of phosphorylated metabolites in generation of one phosphatidylinositol molecule involves three kinds of enzymes acting on phosphorus groups: a phosphatase, a nucleotidyl transferase and a regiospecific variant of a nucleotidyl eliminase in the phosphatidyl transfer step carried out by phosphatidylinositol synthase.

Phosphatidylinositols are also the substrates for at least three lipid kinases, using either the 3-OH, 4-OH, or 5-OH as nucleophiles towards Pγ of Mg–ATP to produce three regioisomeric phosphoinositol phosphate (PIP) monoesters.[10] Tandem action by a second kinase leads to phosphoinositol diphosphate (PIP_2) regioisomers, of which the 4,5-PIP_2 is a major product (Figure 7.12).

Action of phospholipase C on 4,5-PIP_2 produces the diacylglycerol and triphosphorylated inositol (IP_3), both mediators of metabolic steps in lipid metabolism and/or calcium ion release from the endoplasmic reticulum, respectively.[11–13] IP_3 has a half-life of a few seconds in most cell cytoplasms as inositol phosphate phosphatases act to remove all three phosphates hydrolytically to get back to starting inositol (Figure 7.12).

Thus, this phase of PI conversion to PIP_2 formation and cleavage involves three kinases, some acting more than once, the phosphodiesterase activity of phospholipase C, and three phosphatase steps converting IP_3 to inositol and three Pi molecules.[14] A total of seven enzymes in this phase plus three enzymes in PI formation sums to *ten enzymes* dealing with phosphate groups in some form or context in the life cycle of this phosphatidylinositol membrane lipid molecule.

Remarkably, the IP_3 product can also serve as a substrate for three more kinase steps to create the hexaphosphorylated inositol,[15] IP_6, thought to be a protein ligand, perhaps in regulation of cellular homeostasis of inorganic phosphate levels (Figure 7.12). Finally, a kinase can phosphorylate IP_6 on the 5′-phosphate group to yield IP_7. IP_7 has five monophosphate ester groups and one *pyrophosphoryl* ester grouping[16] at C5. Whether the IP_7 acts as an alternative phosphorylating agent through this –P–O–P– anhydride linkage is not clearly established.

7.5 Phosphatases: Hydrolysis of Phosphate Monoesters

More than 100 low molecular weight metabolites are known to be phosphorylated and then acted on by intracellular enzymes acting as hydrolytic phosphoesterases, commonly abbreviated to the term *phosphatases*. The metabolic roles of phosphatases that act on low molecular weight phosphomonoesters are generally to release the inorganic phosphate for use in a variety of biosynthetic roles, most notably in bone and teeth formation, in ATP synthase action (~10 kg per day for this function), and as a substrate for phosphorylases (Chapter 9).

We commented in Chapter 2 that, in humans, serum [Pi] levels are closely regulated in the 1–1.5mM range. Prokaryotes often secrete phosphatases to enable hydrolysis of external phosphoesters and subsequent import of inorganic phosphate. In eukaryotes, lysosomes are digestive organelles that contain both highly active phosphomonoesterases and phosphodiesterases to cleave the alcoholic C–O–P bonds in both types of phosphoester substrates. Such catalysts liberate inorganic phosphate from monoesters and generate alcohol fragments and monophosphates from phosphodiesters.

The tandem action of phosphodiesterases and phosphomonoesterases, for example, will convert long chain DNA and RNA molecules to Pi and the phosphate-free nucleosides (Figure 7.13). The neutral, phosphate-free organic product scaffolds are often exported by solute transport proteins across the lysosomal membrane into the cytoplasm where they are phased into metabolic pathways (*e.g.* nucleoside phosphoribosylation from PRPP as salvage pathways). The charged dianionic inorganic phosphate coproduct passes out of the lysosomes *via* phosphate transport proteins. In Gram-negative bacteria, these phosphate ester hydrolases are often in the periplasmic space, allowing the neutral products from such cleavages to be transported readily into cells.

7.5.1 Phosphate Monoester Hydrolysis Mechanisms

In cells there is typically a balance of kinase and phosphatase action, both on the low molecular weight ~100 phosphorylated metabolites in primary metabolic pathways as well as the protein kinase and protein phosphatase universe of Section IV.

We noted in Chapter 1 that phosphate esters are chemically stable under physiological conditions. Thus, the energy barriers for spontaneous hydrolysis are so high that essentially every phosphoester bond hydrolyzed by organisms requires enzymatic catalysis. The negative charges (up to two on phosphomonoesters and one on phosphodiesters) contribute to the hydrolytic stabilities. Bowler *et al.*[17] noted that trimethyl phosphate is hydrolyzed 10^9-fold more rapidly than the monoanionic dimethyl phosphate while the monomethyl phosphate was some 10^{12}-fold more resistant to hydroxide ion compared with the trimethylester.[18] Analogously, the rate constant for

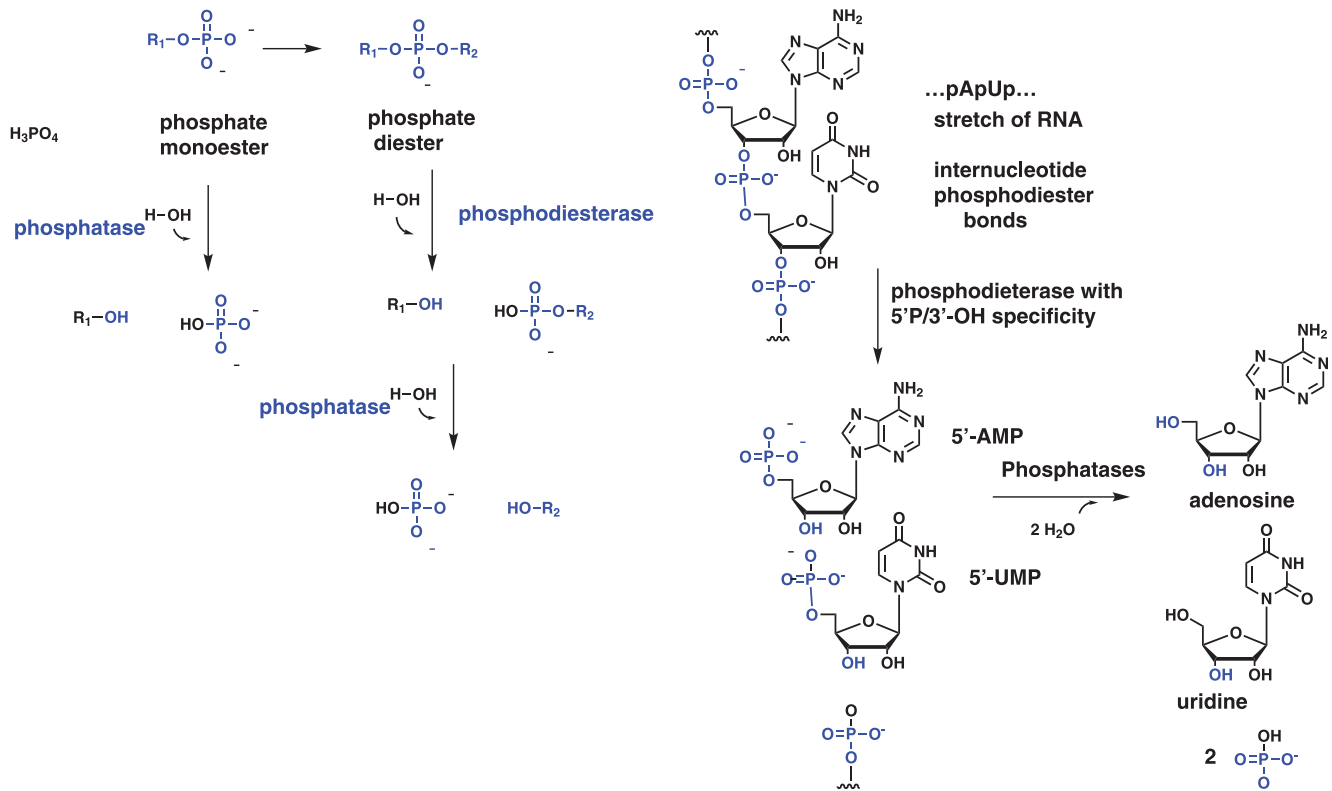


Figure 7.13 Phosphatases *vs.* phosphodiesterases. Enzymes that hydrolyze phosphate monoesters are colloquially termed phosphatases. Phosphodiester hydrolases are designated phosphodiesterases. For example, tandem action of phosphodieterases, then phosphatases on a –pApUp– stretch of RNA will yield the free neutral nucleosides and inorganic phosphate dianions.

hydrolysis of the di-neopentyl phosphate diester was estimated at 7×10^{-16} s^{-1} for an extrapolated half-life of anionic 3′,5′-phosphodiester bonds in DNA of ~31 million years.[18]

Despite the chemical stability that borders on almost infinite for biological time horizons, phosphate monoesters and diesters are hydrolyzed enzymatically with turnover rates that can occur in milliseconds. Catalytic rate accelerations over nonenzymatic hydrolysis rates[17] can range from 10^{14} to 10^{21}, spurring investigators to many detailed mechanistic efforts to parse the several orders of magnitude accelerations into particular subcategories.[19] The timing of water attack on the central electrophilic phosphorus has consequences for the various mechanistic possibilities among this large class of hydrolytic enzymes.

Three limiting possibilities for phosphomonoester hydrolysis by cellular phosphatases are displayed in Figures 7.14 and 7.15 and mimic the possibilities in Figure 3.4. The two external routes can be classified as associative *vs.* dissociative mechanisms. In the associative mechanism (top arrows) the transition state indicates bond forming before bond breaking. Full bond formation before any bond breakage would generate a pentacovalent phosphorane species, as noted previously at the outset of Chapter 3. The incoming water has formed a HO–P bond. The phosphorus atom remains in its [+5] oxidation state, but now has five oxygen substituents, two apical and three equatorial. The negative charge on one of the apical oxygen anions can act as an intramolecular assistant to ejection of RO^- (with protonation as it leaves from the other apical position in an *inline* displacement). This creates the inorganic phosphate product along with the displaced ROH.

In contrast the bottom arm of Figures 7.14–7.15 depicts a dissociative mechanism in which the RO^- bond to phosphorus breaks early, before any bond formation to the incoming –OH nucleophile. This dissociative route would generate the indicated phosphate species known as the *metaphosphate anion*. This is a known species in some phosphate ester nonenzymatic chemistry.[20,21] The metaphosphate anion transition state would then be captured by a late P–OH bond formation to give the observed product inorganic phosphate.

Within these two limits there can be a large number of intermediate cases with partial bond breaking and partial bond making at phosphorus in reaction coordinates and these hybrid mechanisms are probably most highly populated. As noted in Chapter 3 we have adopted the simplest formulation in arrow pushing of electrons attacking the central electrophilic phosphorus atom.

A third possible route for enzymatic phosphatase action is to create an intermediate where the initial kinetically competent nucleophile is a nucleophilic amino acid side chain in the active site if the enzyme. As shown in Figures 7.14 and 7.15 this would produce a covalent phosphoryl enzyme intermediate while ROH departs. Then, in a second step the enzyme catalyzes the attack of the cosubstrate water molecule on the phosphoenzyme intermediate, cleaving the P–X–Enzyme bond and generating inorganic

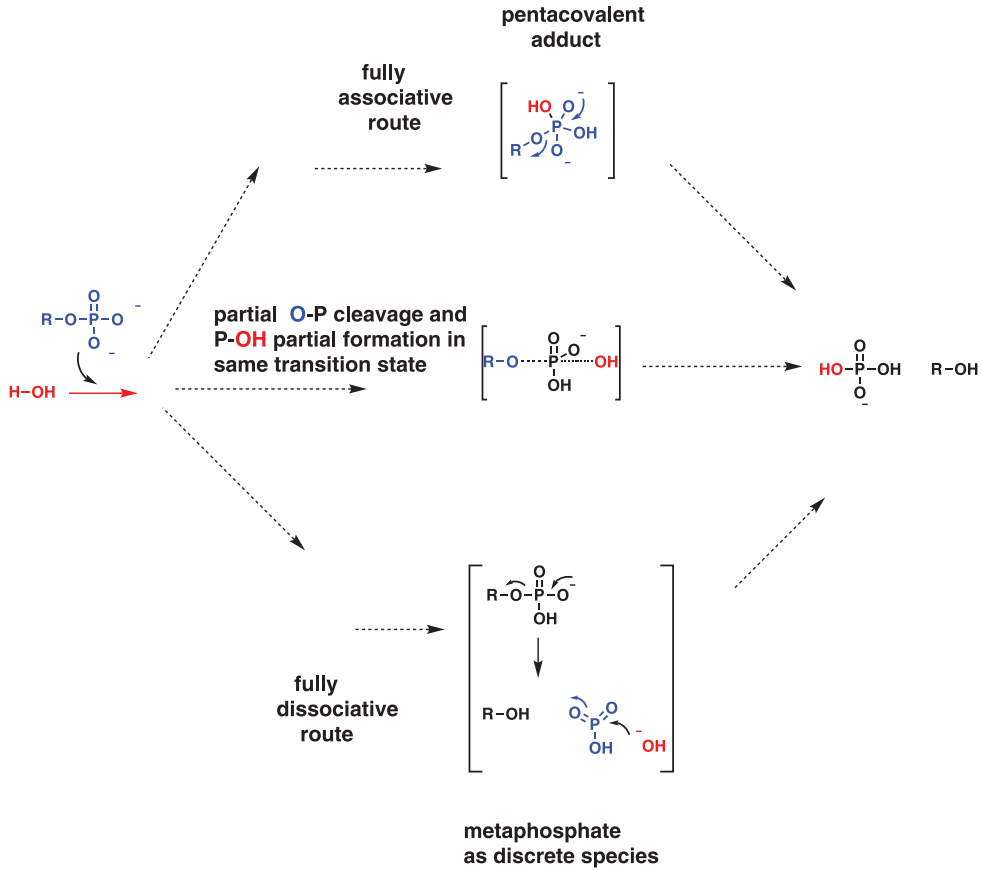


Figure 7.14 Two limiting mechanisms for direct phosphomonoester hydrolysis. The top path shows an associative route, with pentacovalent phosphorane transition state before the leaving group alcohol departs. The bottom route shows early cleavage of the P–OR phosphorus oxygen bond to create a metaphosphate ion that then adds water. There can be intermediate paths and transition states with partial P–O bond cleavage *vs.* P–O bond formation from the incoming water molecule (middle arrows).

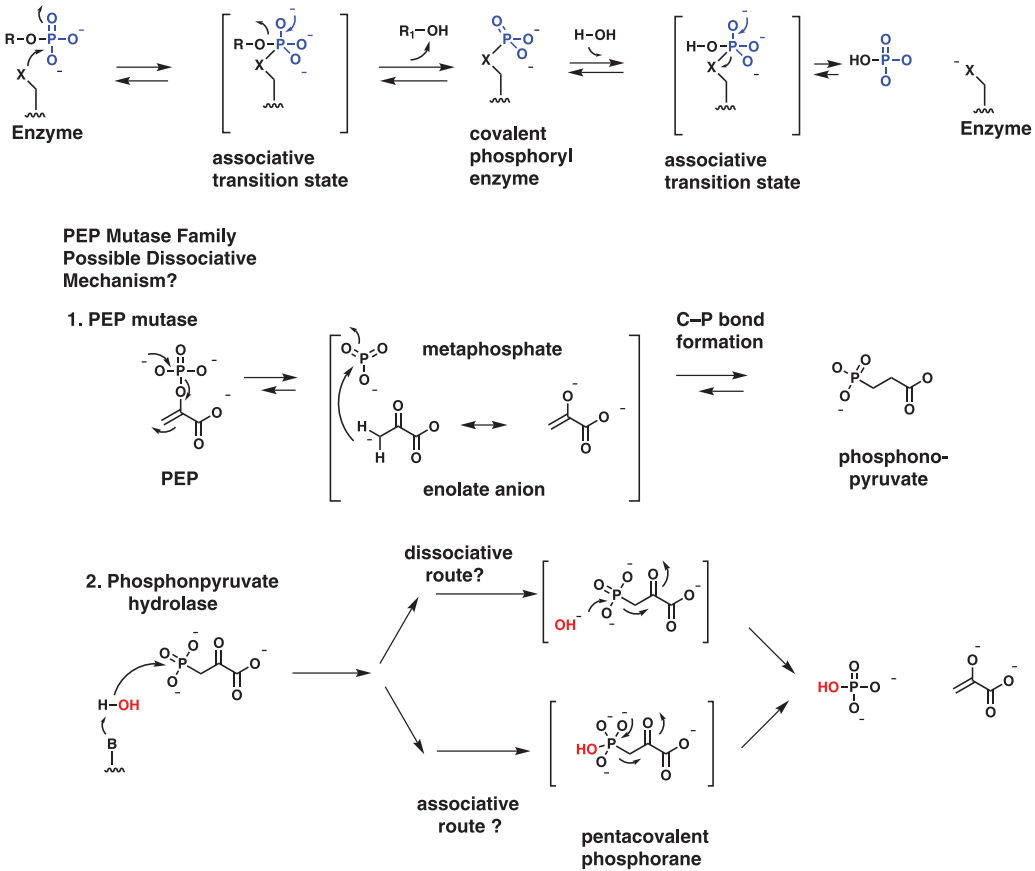


Figure 7.15 Indirect hydrolysis route *via* a covalent phosphoryl enzyme intermediate. In this case there are two kinetically significant nucleophiles. In the first step it is an enzyme side chain. In the second step it is a water molecule attacking the covalent phosphoryl–enzyme adduct.

phosphate product. We will note in Chapter 13 that nine side chains of amino acid residues in proteins can, in principle, serve as nucleophiles in protein phosphorylations by ATP.

An enormous amount of effort over the years has been applied to deconvolution of phosphatase enzyme mechanisms to get at the fundamental chemical biology parameters of phosphate groups in metabolic pathways. Two thoughtful reviews, written 30 years apart, summarize lessons from many of those mechanistic efforts. The first was a review by Jeremy Knowles[22] in 1980, the second from the Herschlag group[19] in 2011.

The Knowles review was written at the height of stereochemical studies designed to constrain mechanisms for phosphoryl transfer enzymes, from kinases, to phosphatases and phosphodiesterases. The stereochemical probes were either phosphoryl groups labeled chirally with three isotopes of oxygen (O^{16}, O^{17}, O^{18})[23,24] (Figure 7.16) or diastereomers of phosphorothioates. The use of the P–S–phosphorothioates as sterochemical probes[25] is discussed in Chapter 12.

The determination of chirality in the products required much ingenuity and inventive methodologies These included chiral syntheses, *e.g.* of the S_P isomer of ATP–Pγ-[^{16}O, ^{17}O, ^{18}O] (Figure 7.17), and then NMR analysis of rigidified products with the oxygen-isotope shifts on ^{31}P-NMR spectra as the analytical method for product chirality determinations.[22–24]

Figure 7.17 depicts three kinases that revealed phosphoryl group transfer with inversion of stereochemistry when examined by the oxygen isotope chirality approach.[22] Hexokinase was assessed with the *S*-isomer of –Pγ–[^{16}O, ^{17}O, ^{18}O]–ATP and the C6–OH of glucose as cosubstrate nucleophile. Pyruvate kinase was assessed in the favored direction of phosphoryl group transfer from the enol phosphate PEP to ADP as nucleophile. The third example was with chiral ATP and acetate kinase generating the mixed acyl phosphoric anhydride acetyl phosphate, again with inversion of stereochemistry at the $-PO_3^{2-}$ group in flight.

From these three and additional subsequent enzymatic phosphoryl transfers, the evidence accumulated to indicate that phosphoryl transfer always involves inversion of stereochemistry at phosphorus.[22,24] This rules against the dissociative (metaphosphate where racemization would be expected) and in favor of the associative, in-line mechanistic route. In the cases where a phosphoryl–enzyme intermediate was kinetically and chemically competent, the stereochemical result was retention at phosphorus, consistent with two steps, each involving an inversion of phosphorus stereochemistry.

The 2011 article by Herschlag's group[19] sums up much of the findings in the intervening 30 years from the Knowles analysis. In particular it is an in depth physical chemical approach to examine how well-studied phosphatases such as alkaline phosphatase achieve their stupendous rate accelerations. A well-studied set of alkaline phosphatases are prototypes of a phosphatase superfamily that has a bimetallic zinc ion center that orients the phosphoester substrate and also activates the cosubstrate water for

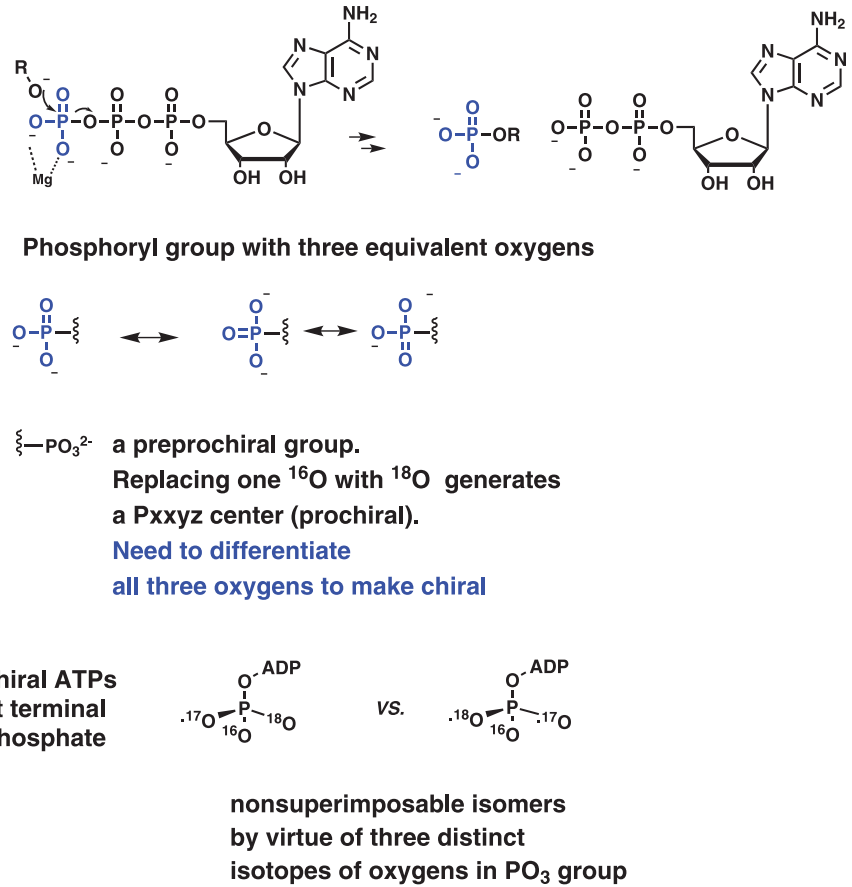


Figure 7.16 Chirality at phosphate groups: The terminal phosphate group of ATP has three equivalent oxygens. In stereochemical terms that phosphate is a preprochiral center. Stereospecific replacement of one ^{16}O by say an ^{18}O isotope would convert the phosphorus center only to a prochiral center. All three oxygens have to be differentiable. That condition is satisfied in a chirally formed [^{16}O,^{17}O,^{18}O]-phosphate monoester or substituted anhydride. The S_P isomer of [^{16}O,^{17}O,^{18}O]–Pγ–ATP was generated and used to assess enzymatic phosphoryl group stereochemistry.

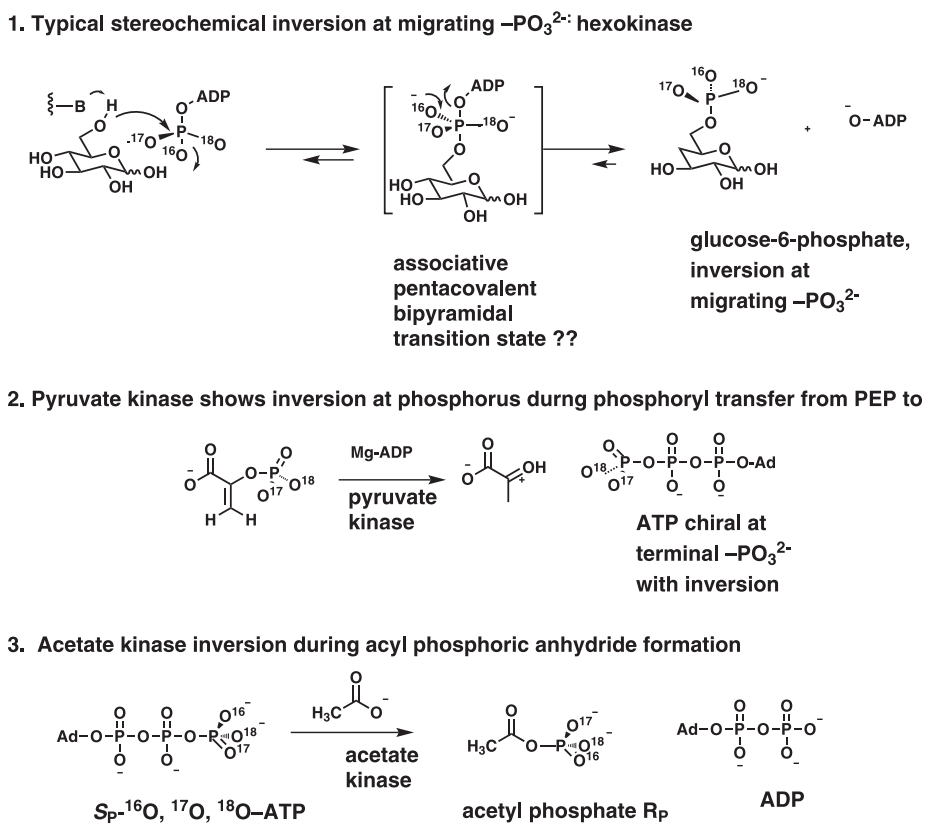


Figure 7.17 Chirality at phosphate groups: Hexokinase, pyruvate kinase (in the PEP to ATP direction), and acetate kinase were all shown to proceed with inversion of stereochemistry at the phosphoryl group in flight.

acceleration of the hydrolysis. In the alkaline phosphatase case the rate acceleration is10^{11} fold over the nonenzymatic rate.[19]

A plethora of structural and mechanistic information is available on this prototypic phosphomonoester hydrolase. Figure 7.18 sketches the location of the two essential zinc atoms. In the initial complex of bound phosphomonoester substrate, Zn_1 coordinates the Ser_{102} –OH catalytic nucleophile and one of the phosphate ester oxygens. Zn_2 coordinates that same phosphate oxyanion and is held in place by three side chain residues of the enzyme. The middle panel of the top line of Figure 7.18 indicates a relatively loose transition state (more dissociative than associative but with inversion of stereochemistry at the $-PO_3^{2-}$ in flight). The result of the first of two consecutive phosphoryl transfers is a covalent phosphoryl enzyme with the $-PO_3^{2-}$ group covalently bound to the Ser_{102} side chain oxygen atom. The alcohol leaving group is initially liganded to Zn_2 and then released to be replaced by a water molecule.

Then, the second half reaction transpires: dephosphorylation of the phosphoryl enzyme by transfer of the $-PO_3^{2-}$ group to the water molecule held in place and activated by Zn_2. This also goes with inversion at phosphorus. Two inversions equal a net retention.

Although alkaline phosphatase involves a covalent phosphoryl enzyme intermediate, many dozens of other members of the bimetallic phosphatase superfamily use a water molecule directly as the catalytically competent nucleophile. In those cases, the roles of Zn_1 and Zn_2, or their replacement metal cation surrogates (Mg^{2+}, Fe^{2+}), are equivalent to the alkaline phosphatase strategy. Zn_1 coordinates and activates the water molecule to attack as nucleophile. The phosphate ester oxygen is coordinated by one or both zincs and the leaving alcohol is typically coordinated in the transition state by metal cation$_2$.

The catalytic turnover rate constant (k_{cat}) of $\sim$100 s^{-1} contrasts with a half-life of about 100 years for the nonenzymatic phosphate ester P–O bond cleavage. Levels of alkaline phosphatase activity rise during active bone formation, to create high enough levels of inorganic phosphate (from phosphomonoester-organified forms of phosphate) to form the supersaturated calcium phosphate solutions secreted by osteoblasts.

The Lassilla, Zalatan and Herschlag[19] review is an excellent reference point for readers interested in how enzymes in general and phosphatases in particular achieve both specificity and rate accelerations, enumerating some of the forces that contribute to specificities and stupendous rate accelerations.

7.5.2 Glucose-6-Phosphatase and Fructose-1,6-Bis Phosphatase

In Chapter 6 we briefly traced the logic of the ten enzymes in glycolysis from glucose to two molecules of pyruvate to emphasize the steps that create and

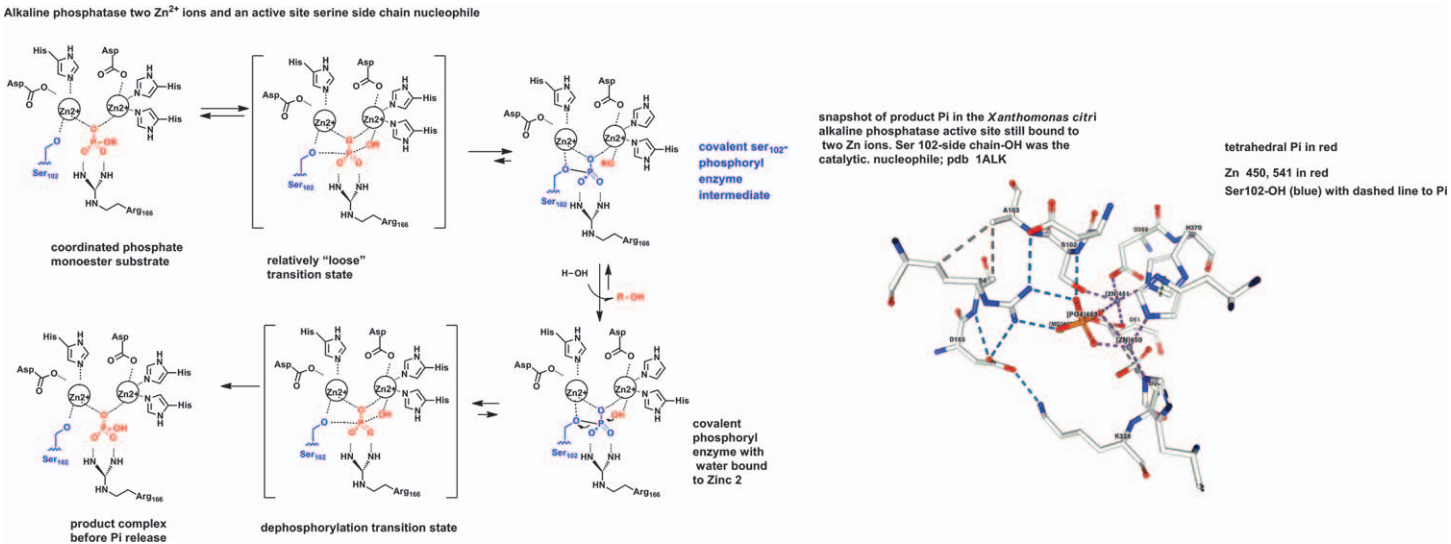


Figure 7.18 Alkaline phosphatase has been studied as a prototypic phosphoester hydrolytic catalyst with stupendous rate acceleration. It is part of a superfamily of phosphatases and phosphodiesterases that use two or more divalent cations to orient both the phosphate ester substrate and the incoming water molecule in the active site. Alkaline phosphatase also forms a covalent phosphoseryl enzyme intermediate, thus tandemly using two catalytic nucleophiles, the active site Ser–OH and then the solvent H_2O, as nucleophiles to effect phosphoester hydrolysis in two tandem steps.

then harvest thermodynamically activated phosphoryl groups. When there is sufficient glucose in a cell, glycolysis runs in the fragmentation direction. However, during fasting the liver and to a lesser extent the kidneys can run the pathway in reverse, converting two pyruvate molecules into glucose, creating up to 200 to 300 g of glucose to be dumped into the blood to travel to other organs for sustenance. This is new synthesis of glucose, termed gluconeogenesis[26] (Figures 7.19 and 7.20). (There is also concomitant mobilization of any stored glycogen by glycogen phosphorylase release of glucose-1-phosphate units).

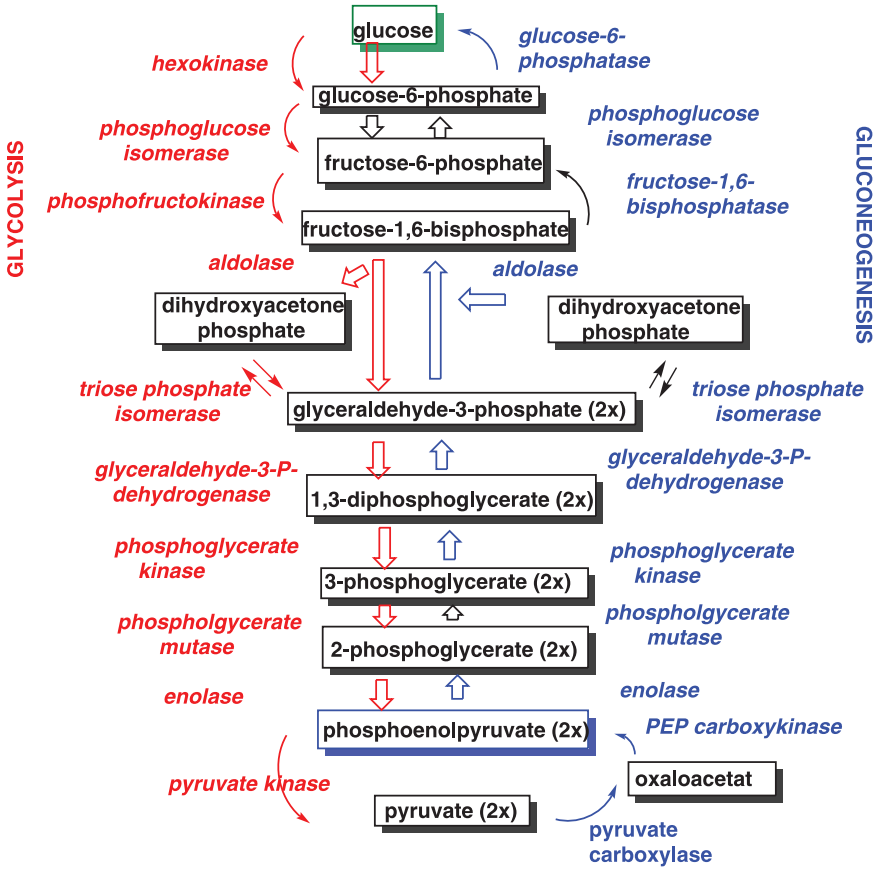
There are four named kinases in the glycolytic pathway: in order they are hexokinase, 6-phospho-fructose-1-kinase, 3-phosphoglycerate kinase and pyruvate kinase. The latter two enzymes can be reversed at the expense of ATP cleavage but to get around the 6-phospho-fructose-1 kinase and the hexokinase steps, the gluconeogenesis pathways employs two phosphatases.

The first is fructose-1,6-bis-phosphatase, regiospecifically hydrolyzing the 1-phosphoester linkage to yield fructose-6-phosphate.[28] The human enzyme is a homotetramer and requires a divalent cation (Mg^{2+}, Mn^{2+}, or Zn^{2+}) for catalysis. X-ray structures are available, indicating two Mg^{2+} sites, one of which coordinates the 1-phosphate dianion of substrate fructose-1,6-bis-phosphate, while the second Mg^{2+} is thought to orient and activate the catalytic water molecule as an incipient hydroxide equivalent. The mechanism is thought to be associative with the addition or elimination steps occurring in an ''in line'' geometry. However, the rabbit enzyme has been crystallized under conditions where metaphosphate is tightly bound, raising questions of slow cleavage of Pi *vs.* a possible dissociative pathway.[29]

The second phosphatase in gluconeogenesis that bypasses a kinase that runs in the opposite direction in glycolysis is glucose-6-phosphatase, the final enzyme in the formation of glucose.[30] This is an intrinsic transmembrane enzyme that resides not in the cytoplasm (where hexokinase isozymes are located) but in secretory compartment membranes. It is part of a complex with a glucose transporter. As glucose-6-phosphate is hydrolyzed in the secretory vacuoles, the free glucose product is transported out of the liver cells into the blood. The outpouring of glucose into the blood by nocturnal gluconeogenesis during fasting maintains blood glucose homeostasis and provides fuel to non-hepatic and non-renal tissues.

A comparison of the energy requirements and products of the two pairs of steps, the two kinase reactions in glycolysis and the two phosphatase reactions in gluconeogenesis, reveal the complementary strategies. The two ATPs spent to convert glucose to fructose-1,6-bisphosphate, eventually get converted to two high energy phosphoryl groups as the acyl phosphate group of 1,3-diphosphoglycerate and the trapped enol phosphate in PEP.

In the gluconeogenesis direction the reversal of the phosphofructokinase and hexokinase reactions are so uphill energetically, that an alternate strategy is used. Two phosphatases sequentially remove the 1-phosphate and then the 6-phosphate *via* $-PO_3^{2-}$ group transfer to two water molecules



Blue Route work arounds for 3 of the 4 kinases in glycolysis

1. Use pyruvte carboxylase and PEP carboxykinase to bypass pyruvate kinase

2. Use fructose-1,6-bis phosphatase to work around phosphofructokinase

3. Use glucose-6-phosphatase to work around hexokinase

Figure 7.19 Gluconeogenesis *vs.* glycolysis in chart form.

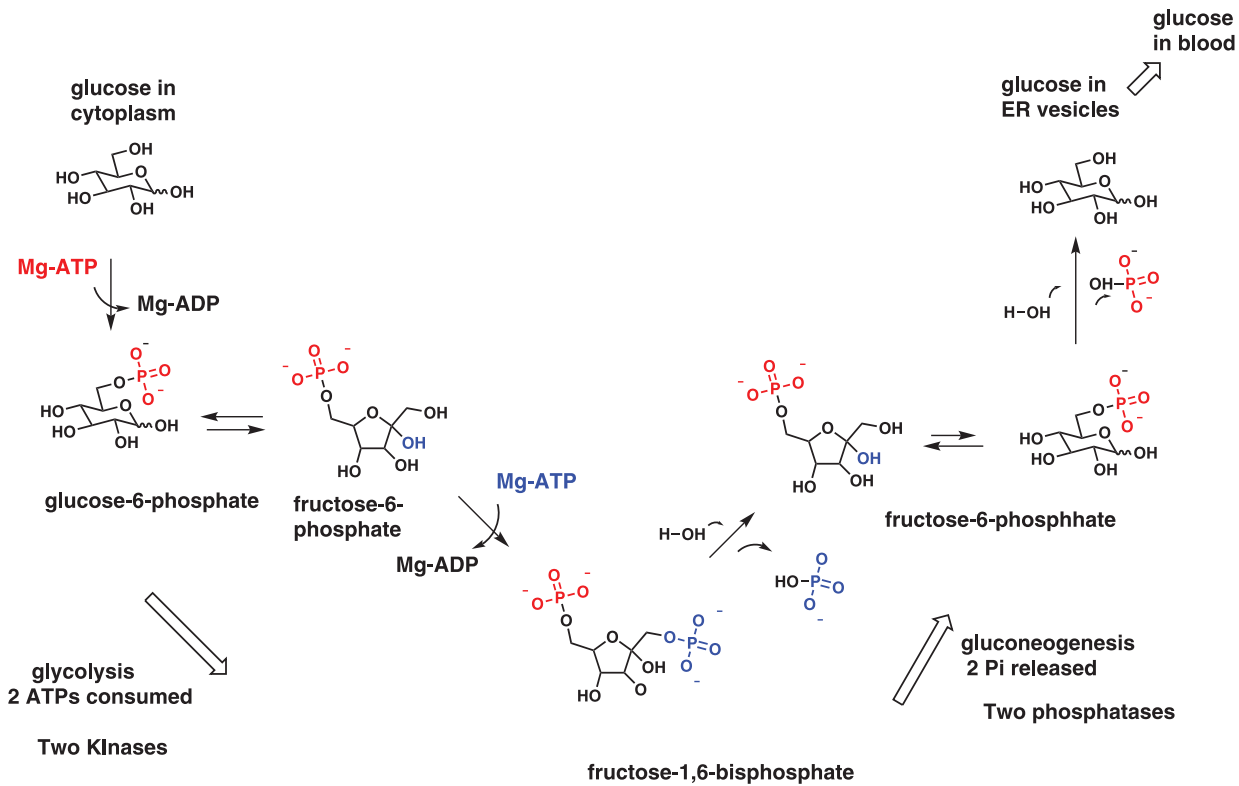


Figure 7.20 Of the four kinases in glycolysis, two are worked around. The third and fourth 3-PGA kinase will run backwards. At the level of fructose-1,6-bis-phosphate, gluconeogenesis uses a phosphatase to get around phosphofructokinase. At the level of glucose-6-phosphate, gluconeogenesis uses an ER-localized glucose-6-phosphatase to get around the hexokinase energetic barrier.[27]

activated as nucleophiles. The second phosphatase acts not in the cytoplasm but in a distinct subcellular compartment, the endoplasmic reticular membrane, allowing the nascent glucose to be pumped outside the cell, *e.g.* into blood to maintain blood glucose homeostasis.

While the bis-magnesium ion-requiring fructose-1,6-bisphosphatase is a typical two metal ion phosphatase that catalyzes direct attack of a cosubstrate water molecule, glucose-6-phosphatase engages in covalent catalysis. The kinetically competent nucleophile towards bound glucose-6-phosphate is one of the two nitrogens in the imidazole side chain of histidine$_{176}$, as shown in Figure 7.21A. The free glucose product departs the active site with P–O bond cleavage as N_1 of the imidazole ring forms a new N–P bond in the covalent phosphoryl enzyme.[31] This is a phosphoramidate linkage (Figure 7.21A). The second half reaction of a glucose-6-phosphatase catalytic cycle is catalyzed attack of a water molecule on the enzyme–His–N–PO_3^{2-} covalent intermediate. An inversion of stereochemistry at phosphorus is assumed in each of the two half reactions: phosphoenzyme formation and phosphoenzyme breakdown. Note that the enzyme nucleophile in alkaline phosphatase was an active site serine rather than the active site imidazole here (Figure 7.21B).

Although we have only examined in detail three of the more than100 phosphomonoesterases and phosphatases in detail here, they reflect two mechanistic possibilities: direct water attack in a metalloenzyme-catalyzed hydrolysis *vs.* a covalent phosphoenzyme intermediate.

7.5.3 Nocturnin: A Circadian Rhythm-Regulated Specific Phosphatase for NADPH and $NADP^+$

A recently characterized phosphatase example provides insight into the intersection of circadian rhythm and diurnal control of metabolism. It also highlights an example of a phosphatase with exquisite substrate scaffold specificity for dephosphorylating just one of the >100 phosphorylated low molecular weight metabolites in primary metabolic pathways.

Among the many genes whose expression is regulated by circadian clock biomachinery, the mRNA encoding a protein named nocturnin is greatly increased at the onset of dark.[32] Nocturnin is a predicted member of the superfamily of metal-requiring phosphoryl–adenylyl-transferases. Two isoforms are expressed in mammals, a constitutive form localizing to the endoplasmic reticulum and the circadian-regulated protein that transits to mitochondria.[32]

The substrate for nocturnin has recently been identified as nicotinamide adenine diphosphate, the form of nicotinamide adenine dinucleotide (NAD) that has undergone kinase-mediated biosynthetic phosphorylation at the 2′-OH of the ribose ring in the AMP moiety (generating a 2′,5′-ADP portion)[32,33] (Figure 7.22). NADH is the more common nicotinamide coenzyme in cells, at ~10-fold higher concentration than nicotinamide adenine diphosphate

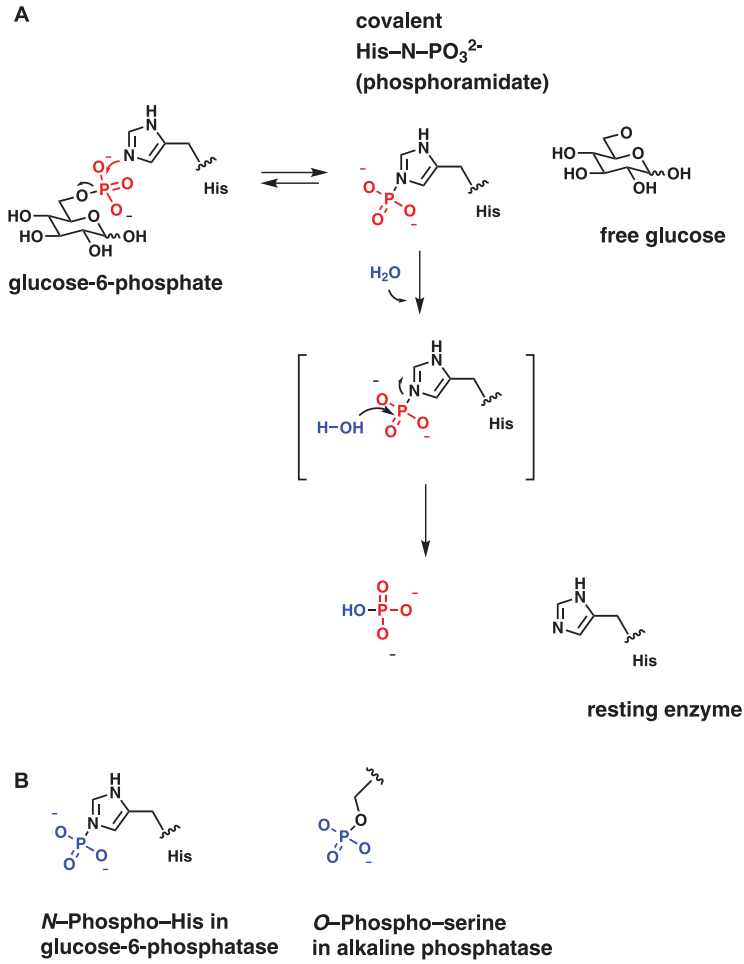


Figure 7.21 Glucose-6-phsphatase catalyzes the terminal step in gluconeogenesis. (A) The kinetically competent nucleophile at the start of a catalytic cycle is the imidazole side chain of an active site histidine, resulting in release of free glucose and formation of a covalent phosphoryl–histidinyl enzyme adduct. In the second step a water molecule is activated to attack the $-PO_3^{2-}$ group attached to the enzyme. (B) Two kinds of covalent phosphoryl enzymes in phosphatase catalysis: a phosphoseryl enzyme in alkaline phosphatase and a phosphohistidinyl enzyme in glucose-6-phosphatase.

(NADPH). NADH and NADPH cycle between these reduced forms, where the redox active nicotinamide ring is in the dihydro oxidation state a thermodynamically activated hydride donor, and the corresponding two electron-oxidized forms NAD^+ and $NADP^+$. The NADPH:$NADP^+$ ratios in cell approach 95 : 5 to 100 : 1 where NADPH, and not NADH, is the obligate coenzyme for fatty acid biosynthesis reductases and for glutathione reductase.

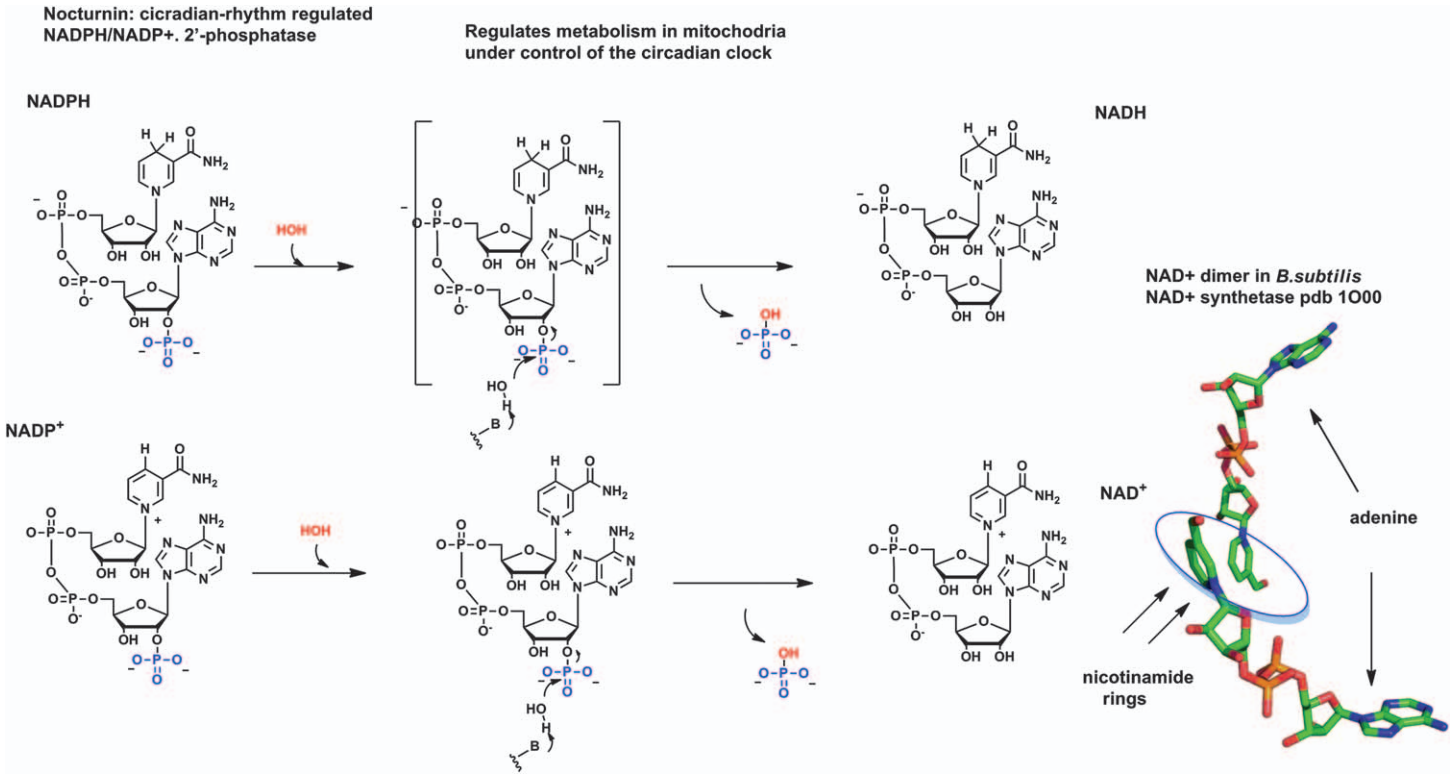


Figure 7.22 The circadian rhythm-controlled protein nocturnin is a specific 2′-phosphatase acting on NADPH or $NADP^+$ to hydrolyze the 2′-phosphoester linkage, generating NADH or NAD^+ and Pi, removing the substrate required for function of NADPH- or $NADP^+$-specific oxidoreductases in mitochondria.

However, within the mitochondrial compartment itself the two forms of the coenzyme approach equal levels, [NADH] ~ [NADPH].

Nocturnin turns out to be a selective phosphatase–phosphomonoesterase, transferring the 2′-phosphoryl group of both reduced NADPH and oxidized $NADP^+$ to water, generating inorganic phosphate and NADH or NAD^+, respectively[33] (Figure 7.22). It is likely that this depletion of reduced NADPH and the reducing power for NADPH-specific oxidoreductases is disrupted diurnally in mitochondrial metabolism. While nocturnin mRNA peaks and then decays during dark periods, the lifetime of the inducible form of the nocturnin protein has not yet been shown to follow that time course, which might indicate circadian control of accelerated proteolysis. Nocturnin does not act on the pyrophosphate bridge present in both NADPH/$NADP^+$ and NADH/NAD^+. Very few metabolites have a ribose-2′-phsophate moiety, allowing nocturnin to have narrow biological specificity in manipulation of the two common forms of the nicotinamide coenzymes.

References

1. G. I. Rhyu, W. Ray and J. L. Markley, Enzyme-bound intermediates in the conversion of glucose-1-phosphate to glucose-6-phosphate by phosphoglucomutase, *Biochemistry*, 1984, **23**, 252–260.
2. C. T. Walsh and Y. Tang, *The Chemical Biology of Human Vitamins*, Royal Society of Chemistry, London, 2018.
3. H. Miziorko, Enzymes of the mevalonate pathway of isoprenoid biosynthesis, *Arch. Biochem. Biophys.*, 2011, **505**, 131–143.
4. G. Wu and D. Vance, Choline kinase and its function, *Biochem. Cell Biol.*, 2010, **88**, 559–564.
5. A. Olivera and S. Spiegel, Sphingosine kinase: a mediator of vital cell functions, *Prostaglandins Other Lipid Mediators*, 2001, **64**, 123–134.
6. T. Balla, Z. Szentpetery and Y. J. Kim, Phosphoinositide signaling: new tools and insights, *Physiology*, 2009, **24**, 231–244.
7. A. L. Marat and V. Haucke, Phosphatidylinositol-3-phosphates-at the interface between cell signalling and membrane traffic, *EMBO J.*, 2016, **34**, 561–579.
8. S. J. Bulley, J. Clarke, A. Droubi, M. L. Giudici and R. F. Irvine, Exploring phosphatidylinositol-5-phosphate 4-kinase function, *Adv. Biol. Regul.*, 2015, **57**, 193–202.
9. G. R. V. Hammond, M. Fischer, K. E. Anderson, J. Holdich, A. Koteci, T. Balla and R. F. Irvine, PI4P and PI(4,4)P2 are essential but independent lipid determinants of membrane identitty, *Science*, 2012, **337**, 727–730.
10. D. A. Fruman, R. Meyers and L. C. Cantley, Phosphoinositide kinases, *Annu. Rev. Biochem.*, 1998, **67**, 481–507.
11. S. G. Rhee and BaeY, Regulation of phosphoinositide-specific phospholipase C isozyme, *J. Biol. Chem.*, 1997, **272**, 15045–15048.

12. S. B. Hansen, Lipid agonism: the PIP2 paradigm on ligand-gated ion channels, *Biochim. Biophys. Acta*, 2015, **1851**, 620–628.
13. D. W. Prole and C. Taylor, Structure and fuction of IP3 receptors, *Cold Spring Harbor Perspect. Biol.*, 2019, **11**, a035063.
14. J. M. Dyson, C. Fedele, E. M. Davies, J. Becanovic and C. A. Mitchell, Phosphoinositide phosphatases: just as important as the kinases, *Subcell. Biochem.*, 2012, **58**, 215–279.
15. A. J. Hatch and J. York, Snapshot: inositol phosphates, *Cell*, 2010, **143**, 1030–1031.
16. S. B. Shears, Inositol pyrophosphates: why so many phosphates ?, *Adv. Biol. Regul.*, 2015, **57**, 203–216.
17. M. W. Bowler, M. Cliff, J. P. Waltho and G. M. Blackburn, Why did nature select phosphate for its dominant roles in biology ?, *New J. Chem.*, 2010, **34**, 784–794.
18. G. K. Schroeder, C. Lad, P. Wyman, N. H. Williams and R. Wolfenden, The time required for water attack at phosphorus, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 4052–4055.
19. J. K. Lassila, J. Zalatan and D. Herschlag, Biological phosphoryl-transfer reactions: understanding mechanism and catalysis, *Annu. Rev. Biochem.*, 2011, **80**, 669–792.
20. A. Todd, Kyoto Lecture. Where there's life there's phosphorus, 1982.
21. F. Westheimer, Why nature chose phosphates, *Science*, 1987, **235**, 1173–1178.
22. J. R. Knowles, Enzyme-catayzed phosphoryl transfer reactions, *Annu. Rev. Biochem.*, 1980, **49**, 877–919.
23. S. R. Jones, L. Klindman and J. R. Knowles, Stereochemistry of phosphoryl group transfer using a chiral [^{16}O,^{17}O,^{18}O)] ATP to study the course of alkaline phosphatase,, *Nature*, 1978, **275**, 564–565.
24. G. Lowe, Chiral [^{16}O,^{17}O,^{18}O] Phosphate Esters, *Acc. Chem. Res.*, 1983, **16**, 244–251.
25. P. A. Frey, Sulfur as a mechanistic probe in enzymatic and nonenzymatic substitution at phosphorus, *New J. Chem.*, 2010, **34**, 820–828.
26. F. Q. Nuttall, A. Ngo and M. Gannon, Regulation of hepatic glucose production and the role of gluconeogenesis in humans: is the rate of gluconeogenesis constant ?, *Diabetes/Metab. Res. Rev.*, 2008, **24**, 438–458.
27. MEDBULLETS Team(downloaded 1201-2019) Glucoenogenesis.
28. D. J. Timson, Fructose-1,6-bisphosphatase: getting the message across, *Biosci. Rep.*, 2019, **39**, BSR20190124.
29. J.-Y. Choe, C. Ianeau, H. J. Fromm and R. B. Honzatko, Metaphosphate in the active site of fructose-1,6-bisphosphatase, *J. Biol. Chem.*, 2003, **278**, 16015–16020.
30. E. van Schaftingen and I. Gerin, The glucose-6-phosphatase system, *Biochem. J.*, 2002, **362**, 513–532.
31. A. Ghosh, J. Shieh, C. J. Pan, M. S. Sun and J. Y. Chou, The catalytic center of glucose-6-phosphatase. His176 is the catalytic nucleophile

forming the phosphohistidine-enzyme during cataysis, *J. Biol. Chem.*, 2002, **277**, 32837–32842.
32. I. Laothamatas, P. Gao, A. Wickramatne, C. G. Quintanilla, A. Dino, C. A. Khan, J. Liou and C. B. Green, Spatiotemporal regulation of NADP(H) phosphatase nocturnin and its role in oxidative stress response, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 993–999.
33. M. A. Estrella, J. Du, L. Chen, S. Rath, E. Prangley, A. Chitrakar, T. Aoki, P. Schedl, J. Rabinowitz and A. Korennykh, The metabolites NADP + and NADPH are the targets of the ciurcadian protein nocturnin (curled), *Nat. Commun.*, 2019, **10**, 1–10.

SECTION III

Types of Phosphorylated Metabolites and Metabolic Logic

Section II was focused on ATP and related NTPs, specifically on the ways that the P–O–P phosphoric anhydride bonds in the triphosphate side chains get cleaved enzymatically to power otherwise unfavorable equilibria in biology. This is the central logic of all macromolecular biosynthesis and essentially every facet of energy metabolism. The single factoid that humans make and utilize 75–80 kg of ATP every day (~2 million kg in a 75 year life span) highlights the central place that the P–O–P anhydride bond plays in enabling the reactions of life.

Section III turns to the different categories of phosphorylated metabolites that are the actual products of phosphoryl, pyrophosphoryl, and nucleotidyl transfers from ATP and other NTPs. The products of enzymatic phosphoryl transfers to substrate alcoholic groups are phosphomonoesters. They and the enzymes that process them hydrolytically are the subject of Chapter 7.

When phosphate oxyanions are the attacking nucleophiles, phosphodiesters are the coproducts along with ADP. Phosphodiester linkages dominate both nucleic acid chemical biology and the reactions of the second messenger cyclic nucleotides cAMP and cGMP (Chapter 8).

Chapter 9 compares the role of water *vs.* inorganic phosphate as nucleophiles in biology. Water is the cosubstrate for phosphomonoesterases (phosphatases), phosphodiesterases, and the few phosphotriesterases. In contrast, there are sets of enzymes that use inorganic phosphate in place of water and transfer electrophilic substrate groups to produce phosphomonoesters. These are critical enzymes for mobilizing carbohydrate monomers

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

from such polymers as starch, glycogen, and cellulose. They showcase the versatility of inorganic phosphate to react either as a phosphorus electrophile or, in this case, as an oxygen nucleophile.

Three categories of phosphoryl groups are found in metabolites where one of the phosphate oxygens is substituted by a nitrogen atom, a carbon atom, or a sulfur atom. Chapter 10 takes up the N–PO_3^- linkages in phosphoramidates while Chapter 11 examines the logic and mechanism of formation of direct C–PO_3^- phosphonate linkages. The last chapter in this section, Chapter 12, examines the relatively small number of S–PO_3^{2-} linkages (phosphorothioates) in biological systems.

CHAPTER 8

Phosphodiesters and Phosphotriesters

Phosphodiester linkages are crucial covalent linkages in several diverse classes of biological scaffolds. The linkage can be deconstructed formally into a $-PO_2^-$ group formally inserted between two alcoholic oxygens. Usually the two ROH partners are distinct, reflecting distinct metabolic origins. In principle, phosphodiesters might arise by some kind of paired sequential phosphoryl transfer processes. In practice they almost always arise by *nucleotidyl transfers* to alcohol nucleophiles as discussed in Chapter 4. In Chapter 4 we focused on Mg–ATP and other Mg–NTPs as nucleotidyl donors. In this chapter we shift focus to examination of the resultant phosphodiester products and their subsequent metabolism. Often those fates are enzymatic hydrolysis by action of phosphodiesterases.

Three disparate molecular cases of biomolecules where phosphodiesters are key backbone linkages are: (1) nucleic acids – both RNA and DNA, (2) cyclic nucleotide second messengers, and (3) a variety of phospholipids that constitute both internal and external cellular membranes. In RNA, DNA and cyclic nucleotides such as cAMP, and di-cyclic GMP, the origin of the core phosphodiester linkages is clear since the XMP moieties comprise one of the alcohol groups in the phosphodiester linkages. In membrane phospholipids the biosynthetic origins from nucleotidyl transfers are cryptic in the final structures.

One of the abiding functional features of biological phosphodiesters is that at physiological pH values they still carry one negative charge on a phosphate oxygen.[1] This is consequential for RNA and DNA as polyanions, probably stabilizing them against adventitious nonenzymatic hydrolysis. It. also explains their association with inorganic and organic cations and positively charged proteins (*e.g.* histones). We noted earlier (in Chapter 7) the estimate that DNA internucleotide phosphodiester bonds might have

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

half-lives of 31 million years.[2] Likewise, cAMP and cGMP are monoanionic and generally interact with partner proteins by electrostatics. The phosphodiester linkages in glycerol-based phospholipids provide the anionic water-soluble head groups to balance the hydrophobic tails of such amphipathic membrane lipids.

We take up enzymatic cleavages of the internucleotide phosphodiester bonds in RNA and DNA first, because so many of the cell-based manipulations of informational biopolymer structures involve formation or cleavage of one or more 3′,5′-phosphodiester bonds. A set of such of such nucleic acid phosphodiester formations were noted in Chapter 4, including RNA splicing, maturation of both 5′ and 3′ ends of mRNAs, and in the actions of DNA ligases on single and double strands breaks in DNAs.

8.1 Phosphodiesterases as Nucleases

Biological phosphodiesters are less diverse scaffolds than the >100 variants of low molecular weight phosphomonoester metabolites. Nor are phosphodiesters common components resulting from posttranslational modification of proteins outside of those containing phosphopantetheine prosthetic groups, such as fatty acid, and polyketide synthases along with nonribosomal peptide synthetases.

However, every nucleotide and 2′-deoxynucleotide unit in both RNA and DNA, respectively, is held together by covalent 3,5′-phosphodiester bonds (Figure 8.1). Those internucleotide phosphodiesters are the *only* covalent linkages between monomers in any RNA or DNA molecule. We have noted in Chapter 4 that they are formed by polymerases carrying out nucleotidyl transfers. In turn, they are cleaved by hydrolytic enzymes termed phosphodiesterases. The products are an alcohol and a phosphomonoester. Enzymes that cleave these phosphodiester bonds in RNA or DNA are also usually called *nucleases*.[3]

Given the nonenzymatic stability of DNA internucleotide phosphodiester bonds of some 31 million years,[2,4] spontaneous hydrolysis of DNA under physiological conditions does not happen. DNA is thus a faithful and stable reservoir of information, absent some external damage. Manipulation of connectivity therefore requires enzymatic cleavage of the internucleotide bonds. We noted phosphodiester exchange enzymes in Chapter 4, notably in RNA splicing, where the sum total of phosphodiester bonds remains constant and the processes are energetically neutral.

Nucleases can cleave phosphodiester bonds hydrolytically from either end of an RNA or DNA chain with distinct directionality of progression (a 5′ nuclease *vs.* a 3′ nuclease) (Figure 8.2).[3] Those are termed exonucleases.[5,6] There are also nucleases that can cleave internal phosphodiester bonds, endonuclease activities[7] (Figure 8.2). The more than 900 bacterial restriction enzymes, so useful in the early years of nucleic acid manipulation for cutting and pasting double stranded DNA at specific oligonucleotide sequences, are sequence-specific endonucleases, a specific subgroup of phosphodiesterases.

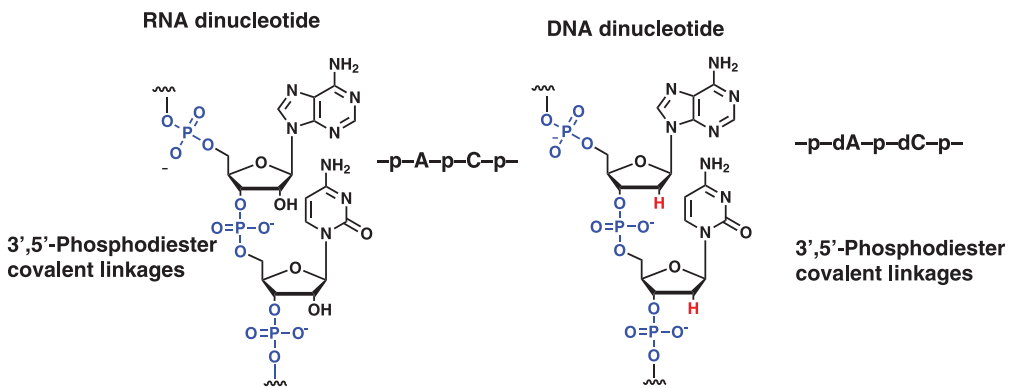


Figure 8.1 The covalent linkages that connect XMP monomer units in RNAs and 2′-dXMP monomer units in DNA are 3′,5′-phosphodiester linkages. They are the only type of covalent bonds in these two classes of information-carrying biomacromolecules.

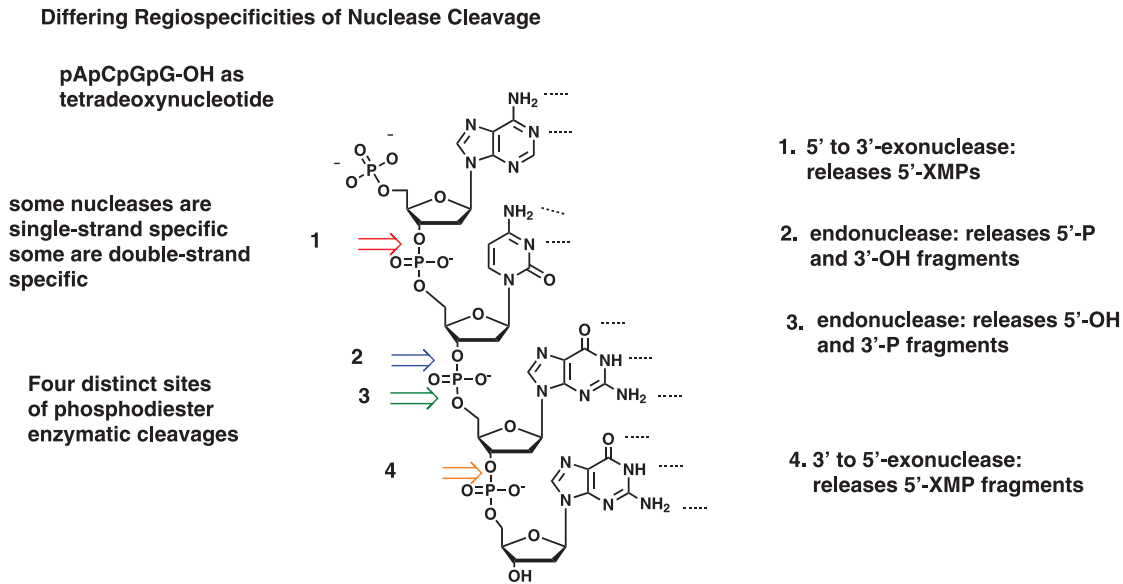


Figure 8.2 Phosphodiesterases acting on nucleic acids (both RNA and DNA) are termed *nucleases.* Four subclasses in two major categories can be distinguished based on location of action and regiospecificity of phosphodiester bond cleavage. The two categories are exonucleases and endonucleases. Exonucleases can act from either the 5′ end or the 3′-end of the nucleic acid substrate. Endonucleases can have different sequence specificities but are also categorized by whether they release 5′-P and 3′-OH fragments or 5′-OH and 3′-P fragments.

8.2 Phosphodiesterase/Nuclease Cleavage Regiospecificity

Nucleases tend to be specific for DNA or RNA.[3] A linear DNA strand or RNA strand has two distinct ends. Most cellular RNAs are single-stranded from transcription of a DNA coding strand but there are short double-stranded (ds) RNAs. Most DNA molecules in a cell are double stranded but there are single stranded regions during replication, recombination, and repair. Figure 8.2 shows four patterns of nuclease action to hydrolyze the phosphodiester bonds in a model DNA tetranucleotide p–A–p–C–p–G–p–G-OH. Exonucleases exist that are specific for recognition of either end before they hydrolyze the terminal phosphodiester bonds. The family of 3′- to 5′-exonucleases hydrolyze phosphodiester bonds sequentially from the 3′-OH terminus, recognizing the unique 3′-OH. This is arrow 4 in Figure 8.2.

Conversely, nucleases that require a free dianionic 5′-phosphate group for recognition are 5′- to 3′-exonucleases, chewing in hydrolytically one phosphodiester bond at a time from the 5′ end, arrow 1. Note that in every phosphodiester bond hydrolyzed, there are two choices. The P–O-3′ or P–O-5′ bond could be broken. In the first case a 3′-OH end is generated. In the latter case a 3′-OPO_3^{2-} end would be generated (5′-OH end). The great preponderance of nuclease action at both internal and external phosphodiester bonds is to liberate a 3′-OH end.[3] This is shown by arrow 2 of Figure 8.2. The two exonucleases (arrows 1 and 4) thereby liberate 5′-XMP monomers, returning immediately to the pool of XMP nucleotides. The endonuclease acting at arrow 2 liberates a 3′-OH and a 5′-phospho-oligonucleotide fragment. The endonuclease acting at arrow 3, would instead yield a 5′-OH fragment and a 3′-OPO_3^{-} terminating fragment.

There are 17 DNA exonucleases[6] identified in the bacterium *E. coli*, reflecting multiple and diverse roles in DNA proofreading during replication, in DNA recombination, and in DNA repair. Three of the seventeen are exonucleases associated with the three DNA polymerases. Exonuclease II is the 3′ to 5′-proofreading catalyst for DNA polymerase I while exonuclease IV is the 5′ to 3′-nuclease associated with that same enzyme. Proofreading is required to remove an incorrectly inserted dXMP unit, and thus remove a mispairing before the polymerase moves on.

DNAse I (the first endonuclease purified historically) acts as a random endonuclease on both DNA strands to liberate 5′-NMPs with free 3′-OH.[3] By contrast, endonuclease IV cleaves 5′-phosphodiester at abasic sites in DNA repair to remove the lesion. Lovett describes the proposed biological roles ascribed to the 17 *E. coli* exonucleases in an extensive 2011 review.[6]

The ubiquitous 3′- to 5′-exonuclease catalysts have been studied for structure and mechanism. While there are one-metal and three-metal versions, the largest category is the two-metal cation-requiring nucleases. This family is termed the DEDD family for the ligating role those four (D and E) carboxylate side chains in nuclease active sites play in ligating and orienting

the two active site divalent cations (most often Zn^{2+} or Mg^{2+}). As depicted in Figure 8.3, the first metal (M^1) has two carboxylate side chain ligands to control its position and it coordinates both the attacking water molecule and one of the internucleotide phosphodiester oxygens. Metal ion M^2 likewise coordinates the phosphodiester oxyanion and also binds to the departing 3′-oxygen, while being oriented by yet another side chain carboxylate of the enzyme. This array allows the two metal cations to organize and bring together the positions of both the attacking water molecule and the phosphodiester electrophile. This active site two metal geometry is a variant of the two-metal cation catalytic array noted for alkaline phosphatase in Chapter 7.

8.3 Ribonucleases

Nucleases acting on RNA most often encounter single stranded RNA,[3] albeit those that see rare double strand RNAs are physiologically important. The classic ribonuclease A acts as an endo(ribo)nuclease, hydrolyzing the phosphodiester bond downstream of pyrimidines (cytidine and uridine). As noted in Figures 8.4 and 8.5 catalytic cycles are initiated by intramolecular action of the 2′-OH as nucleophile (not available in the 2′-deoxyribose units of DNA). The cleavage step generates a 5′-OH fragment and a nascent 2′,3′-cyclic phosphodiester. Thus, the internucleotide phosphodiester cleavage in this case is a phosphodiester exchange reaction. Water only participates in the second step: hydrolysis of the cyclic 2′,3′-phosphodiester. As anticipated for enzymatic catalysis, the ring-opening is regiospecific, yielding the 3′-OPO_3^{2-} upstream RNA fragment. Note this product fragment regiospecificity is the opposite of DNA fragments generated by most DNA nucleases.

Other RNAse classes are worth specific mention.[3] One is the RNAse III family that cleaves double-stranded RNA. These include human DROSHA enzymes.[8] DROSHA endonucleases initiate nuclear processing of microRNAs in RNA silencing situations and also ribosomal RNA precursors. Two other dsRNAses involved in micro RNA processing are DICER and argonaute. DICER cleaves dsRNA into 25–27 base single-strand fragments[9] while argonaute cleaves mRNA targets identified by these single-strand RNAs.[10] The second type of RNAse of note is RNAse H whose members cleave the RNA strand in RNA–DNA hybrid duplexes without cleaving the DNA strand.

There are dozens of other nucleases characterized by type of substrate (RNA or DNA), preference for single or double stranded DNA or RNA, exonucleolytic *vs.* endonucleolytic activity, metal dependence or metal independence, and product regiospecificity (3′-OH and 5′-phosphoester or 3′-phosphoester and 5′-OH) (Figure 8.2). Yang[3] has compiled the summary characteristics of more than a hundred nucleases and observed that all the metal ion-dependent nucleases, both RNAses and DNAases, generated 3′-OH and 5′-phosphate products. On the other hand, some of the metal-independent nucleases gave the opposite 3′-phosphate and 5′-OH products[3] (Figure 8.2).

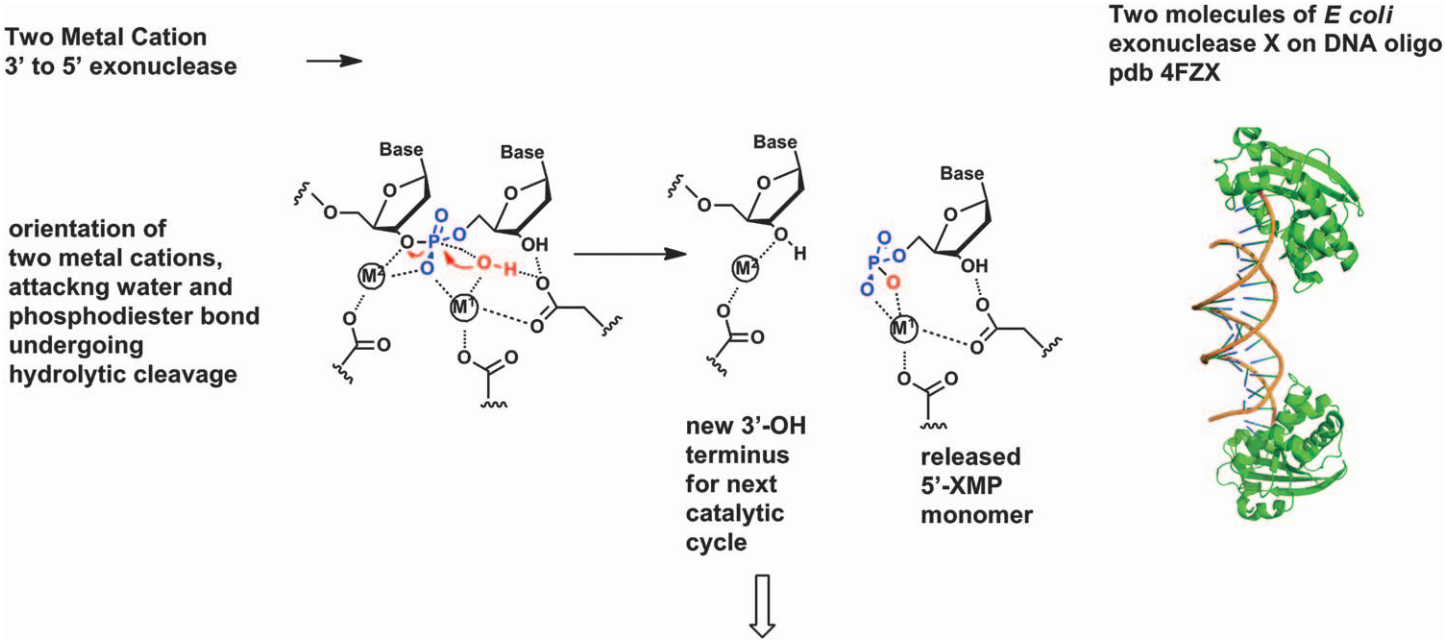


Figure 8.3 Schematic of mechanism for a 3′ to 5′ exonuclease, working in from the free 3′-OH end, one dXMP nucleotide removed per catalytic cycle. The position and role of the two metal cations at the active site are shown. The next catalytic cycle will act on the new 3′-OH end of the nascent oligonucleotide product. PDB file 4FZX shows two molecules of *E. coli* exonuclease Q on either end of a DNA oligonucleotide.

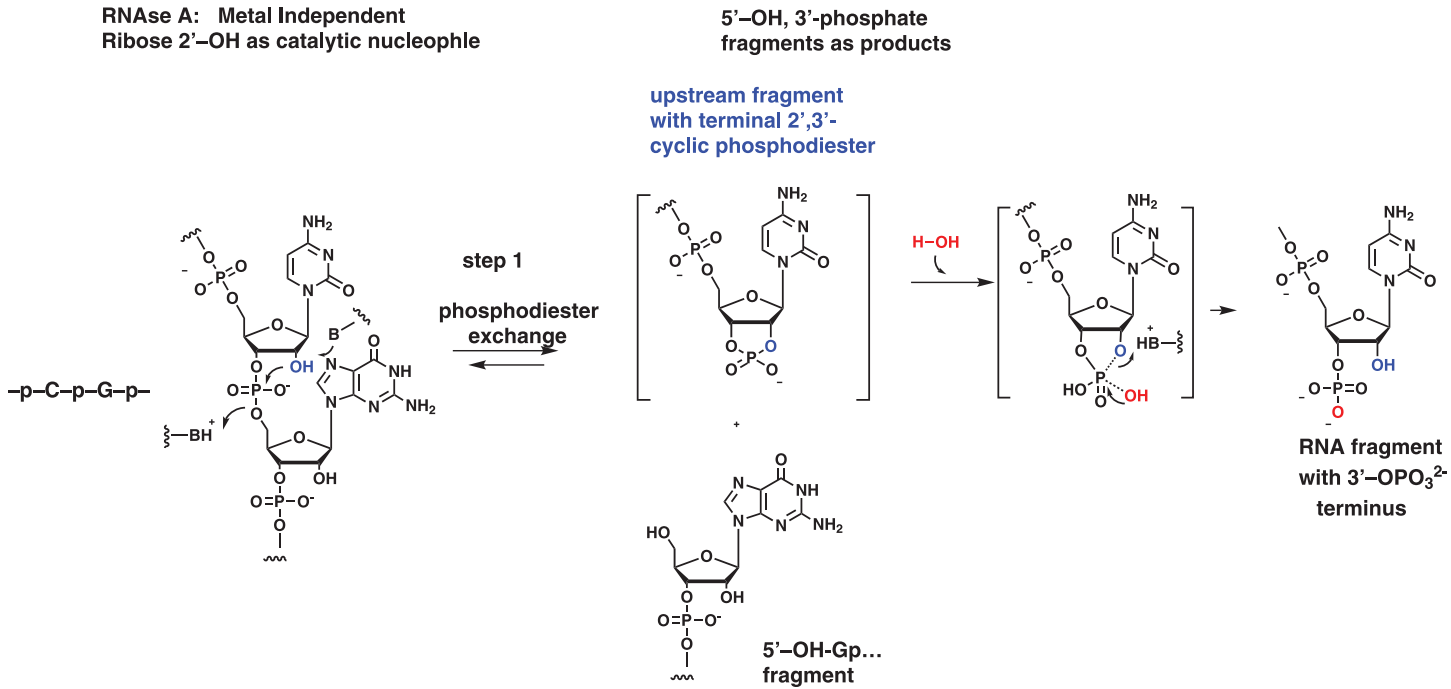


Figure 8.4 Ribonuclease A is a classic single-strand RNAse, acting endonucleolytically downstream of pyrimidine bases, to cut RNAs into limit fragments. The 2′-OH, available in RNA but not DNA bases, initiates intramolecular attack on the adjacent 3′-phosphodiester. This results in expulsion of the downstream oligonucleotide with a free 5′-OH, as a cyclic 2′,3′-phosphodiester is formed (phosphodiester interchange). The 2′,3′-cyclic phosphodiester intermediate is then hydrolyzed regiospecifically to release the upstream oligonucleotide as the 3′-phosphate.

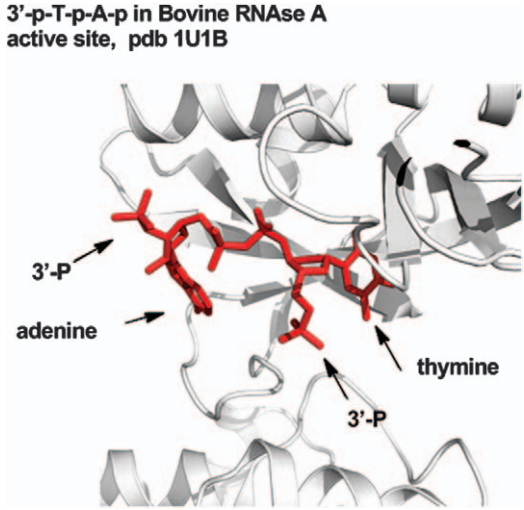


Figure 8.5 Dinucleotide orientation in bovine RNAse A, from pdb 1U1B.

8.3.1 Catalytic RNAs: Ribonucleic Acids Can Catalyze Phosphodiester Transesterifications and Hydrolyses

A variety of analyses and conjectures have been put forth on the possible origins of life – among them the concept of an RNA world, both for self-assembly of ribonucleotide polymers and for information storage and replication. As with many topics where phosphorus impinges on biological processes, RNA biology is far too extensive to be examined in this monograph and readers should delve into the primary and review literature on specific RNA topics.

On the other hand, we have noted that the only covalent bond between monomer units in RNA (and DNA) are the internucleotide phosphodiester bonds, anionic at physiological pH. Any formulation of an RNA world argues for catalytic RNAs.[11] The requisite chemistry could have been many-faceted but should have included manipulation of phosphodiester chemistry and also peptide bond-forming chemistry to transfer information from an RNA sequence to an encoded protein (Figures 8.6–8.10).

Indeed, both those capacities are still extant in contemporary organisms.[12,13] RNA splicing (briefly addressed in Chapter 4) of the great majority of eukaryotic RNAs by spliceosomal machinery based on RNA components is *prima facie* evidence for ongoing RNA catalysis: in those cases, reversible phosphodiester bond cleavage and religation in sequence-specific contexts. For protein biosynthesis, transcribing mRNAs into proteins is the province of RNA structural and catalytic machinery, the ribosome. The peptidyl transferase center of ribosomes in their millions, responsible for hundreds of millions of peptide bonds assembled in every cell cycle is an RNA structural and catalytic machine. Herschlag[13] has noted that the tendency of

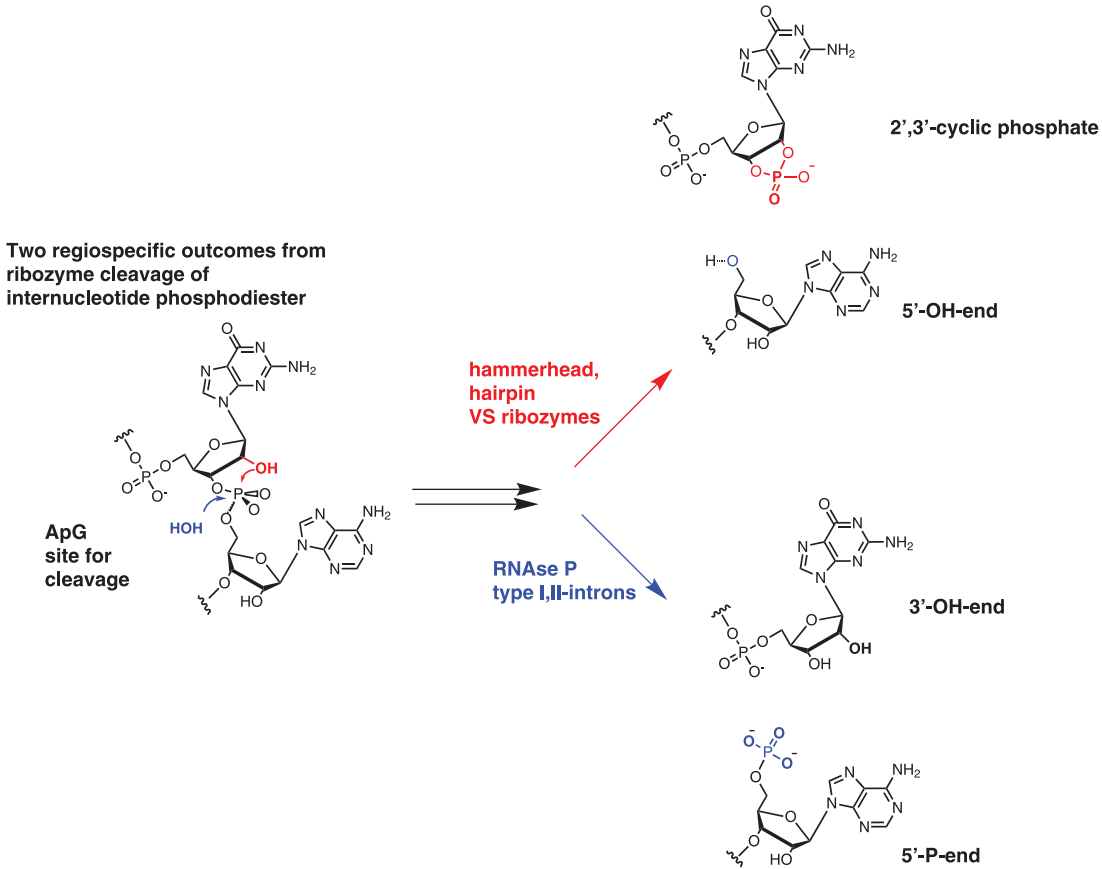


Figure 8.6 Ribozymes and catalytic RNA show two patterns of cleavage of internucleotide phosphodiester bonds. The group I and II introns and RNAse P release RNA fragments with 3′-OH ends and 5′-P-ends. The small hammerhead, hairpin, and VS ribozymes instead release 5′-OH ends and 2′,3′-cyclic phosphate terminal RNA fragments (reminiscent of the intramolecular 2′-OH nucleophilic mechanism of RNAse A).

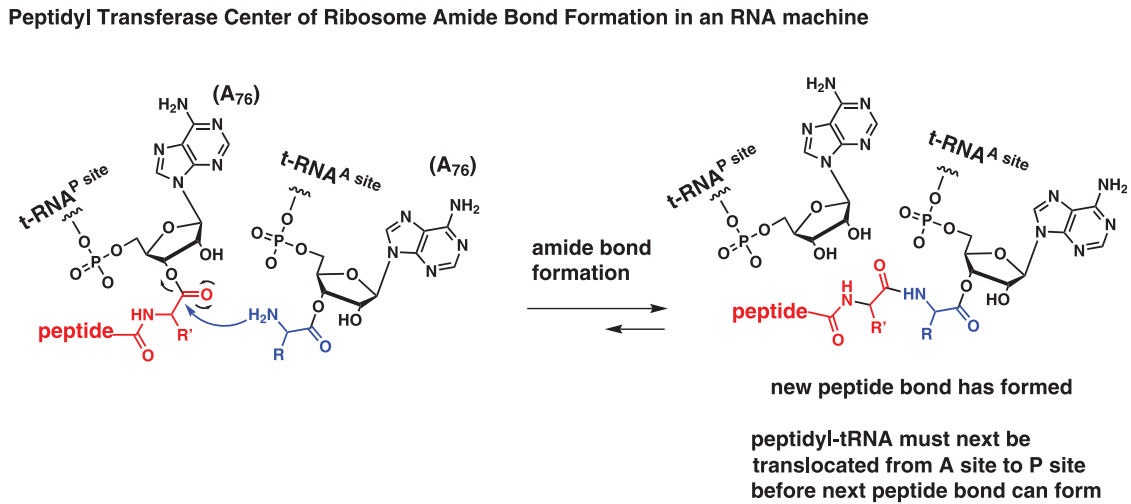


Figure 8.7 Peptide bond formations in the ribosomes represents catalytic ribozyme actions billions of times pre cell cycle.

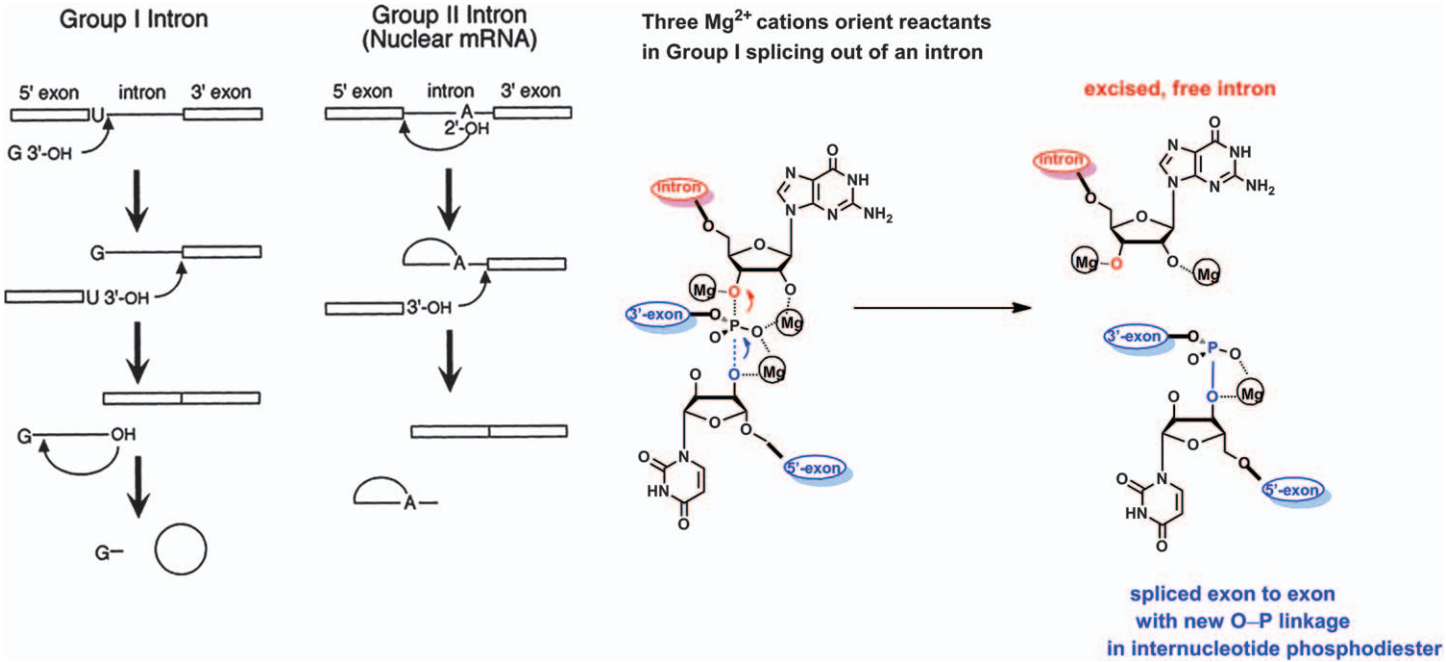


Figure 8.8 Group I and group II introns are both-self splicing ribozymes that ligate upstream and downstream coding sequences (exons) as the introns splice themselves out. Group I ribozymes use a free guanosine in an intermolecular transesterification initiation event, while group II introns utilize the intramolecular 2′-OH (akin to mammalian mRNA splicing routes). Reproduced from ref. 15 with permission from Oxford University Press, Copyright 1999.

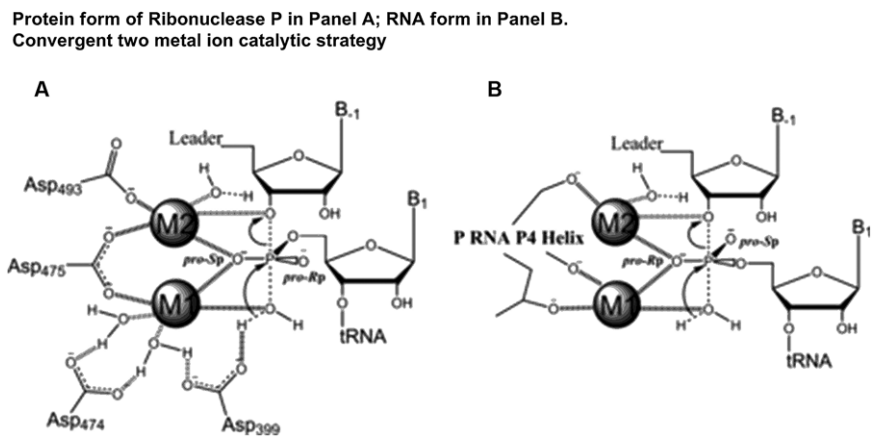


Figure 8.9 Similarities and minor distinctions in the hydrolysis logic and mechanism for the RNAse P ribozymes *vs.* the RNAse P protein enzymes. Both strategies employ the two-metal cation logic seen in other phosphodiesterase enzymes. The nature of the ligands, amino acid carboxylate side chains *vs.* internucleotide phosphodiester anions to orient the two magnesium divalent cations vary.
Reproduced from ref. 20 with permission from American Society for Biochemistry & Molecular Biology, © 2015 by The American Society for Biochemistry and Molecular Biology, Inc.

RNAs to visit many misfolded states has led to understanding of the role of proteins as chaperones and the utility of ribonucleoproteins where the proteins allow the RNAs to populate active conformations.

The past half century has been a golden era of RNA biology, including discovery and characterization of self-splicing RNAs, ribozymes. Most of the half dozen variants of catalytic RNA discovered to date act *in cis*, catalyzing nucleophilic attack on the electrophilic phosphorus centers of constituent phosphodiester bonds, specified by productive conformers of RNAs complexed with metal cations that assist in two and then three dimensional folding ensembles.[14]

The seven or so RNAs that have been confirmed to have catalytic activity are classified according to size and to reaction mechanism.[15,16] RNAse P, described below, and the group I and group II self-excising introns are in the size range from 300 to 3000 nucleotides, and create product fragments with 5′-phosphates and 3′-hydroxyl ends.[16] Small catalytic RNAs, 35–155 nucleotides, include hammerhead, hairpin, hepatitis delta and Varkud satellite (VS) RNA. They generate 5′-OH ends and 2′,3′ cyclic phosphate-terminating fragment pairs.

The distinct product regiochemistries are reminiscent of patterns seen with different types of protein-based nucleases. The release of the 2′,3′-cyclic phosphate is a mimic of, perhaps a progenitor evolutionarily, of the pattern seen with the protein enzyme RNAse A (Figures 8.4–8.10). These data

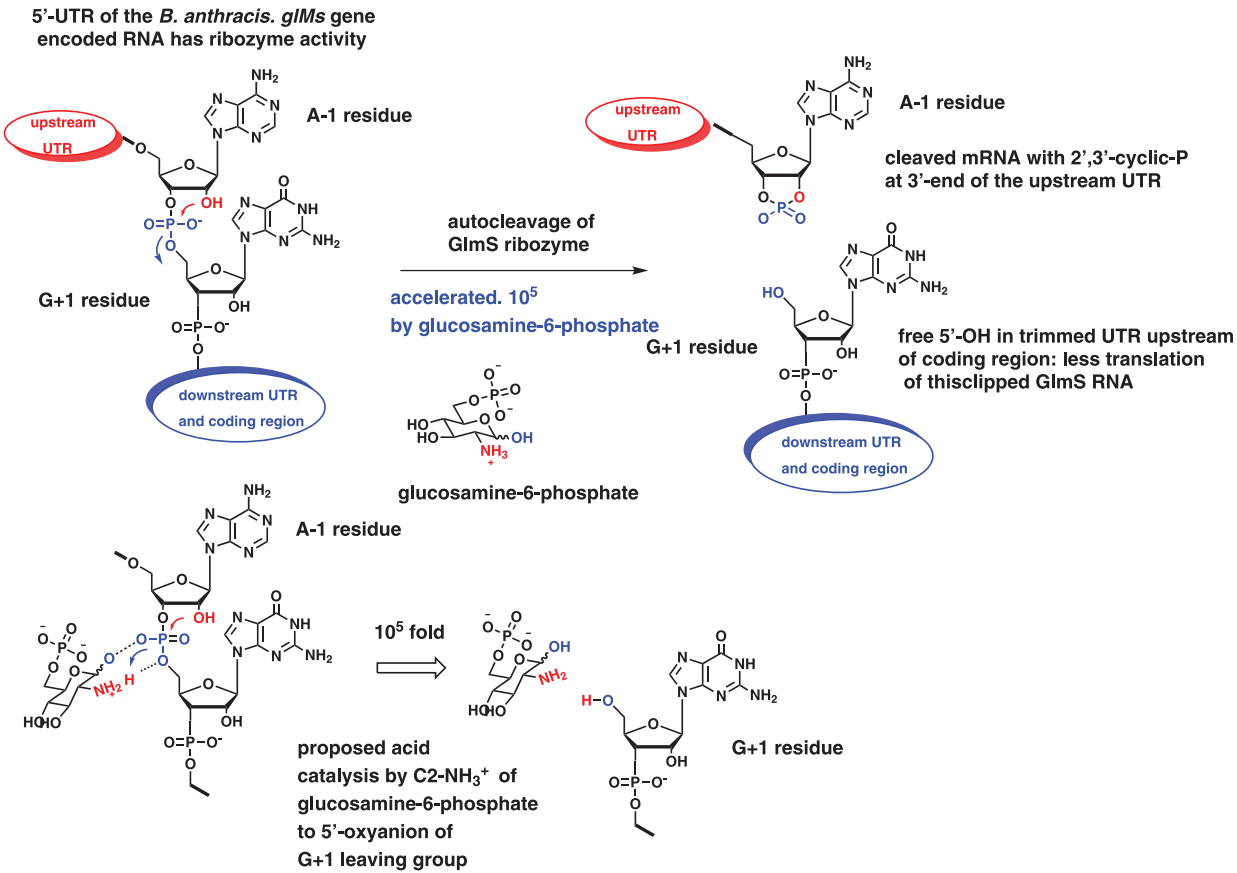


Figure 8.10 Proposed proton transfer role for the 2-ammonium group of glucosamine-6-phosphate in autocleavage of the GlmS 5′-untranslated region of GlmS RNA encoding glucosamine synthase.

indicate that this class of ribozymes can activate the 2′-OH of a ribose moiety as a nucleophile to attack the adjacent 3′-phosphodiester phosphorus. Ribozymes have the heteroatoms of the purine and pyrimidine bases and the ready availability of divalent cations such as Mg^{2+} as potential catalytic bases and acids. For the ribozymes and catalytic RNAs that cleave to 5′-P and 3′-OH ends, two and even three divalent cations are proposed in transition states for these relatively slow catalysts (1–5 min^{-1}). Figure 8.8 depicts a three metal cation active site for a group I intron cleavage (see review by Fedor & Williamson[17]).

8.3.2 Group I and Group II Introns

The introns are self-splicing ribozymes that can cut themselves out of precursor RNAs.[18] Their discovery predated the findings of mRNA splicing but they represent similar logic and chemistry of internucleotide phosphodiester transesterification catalysis. In that spirit, Figures 8.6–8.10 schematize group I and group II introns and ribozyme catalytic action in terms of intron and exon terminology.[19]

In the group I cases, exemplified by the original Tetrahymena pre-rRNA intron discovered by Cech and colleagues in 1982, a low molecular weight, free guanosine is the initiator. Its 3′-OH is the attacking nucleophile at an intron–exon boundary downstream of a uridine residue. The released upstream intron, terminating in a 3′-OH, uses that 3′-OH as nucleophile on the internucleotide bond at the intron–3′ exon boundary.[12,14] This restores the phosphodiester bond, in frame, and liberates the intron with the initial free guanosine now tethered at the 5′-end (Figure 8.4).

Type II introns do not use an external low molecular weight nucleoside as an external nucleophile. Instead, they use one of the 2′-OHs as intramolecular nucleophile to create a lariat intermediate (Figures 8.6–8.10). The exon joining step is similar to the group I intron logic in using the 3′-OH end of the upstream exon as a nucleophile at the intron–3′exon boundary. While the introns excised in type II introns are relatively small stretches of RNA oligonucleotides, they prefigure the mRNA splicing seen in eukaryotes where hundreds of thousands of bases are excised during splicing and several RNAs are called into play to act *in trans*.

8.3.3 Ribonucleases: Evolution from RNA Catalysts to RNA–Protein Catalysts to Proteins: Ribonuclease P

Among the RNA precursor species that get processed hydrolytically and specifically to mature RNA functional forms are pre-tRNAs where oligonucleotides at both the 5′-end and the 3′-end are trimmed by ribonuclease P family members. Typically, these are RNA–protein ribozyme complexes with one to ten protein subunits. Some forms of RNAse P in *Arabidopsis thaliana* are protein only catalysts[20] while humans use both the protein only

(in mitochondria) *and* the protein-associated ribozyme (in the nucleus) versions.

The active site of the protein-only form of RNAse P shows the classic two metal cation active site-oriented by interaction with four aspartate side chains (figure 1B of[20]) found in large families of phosphodiesterases. Howard *et al.*[20] show similar two metal cation mechanisms (their figure 9) (Figure 8.9) for both the protein-only version of RNAse P and the RNA-only ribozyme form of the catalyst. Strategies for orientation of the catalytic metal cations do differ. In the protein catalyst there are four aspartate beta carboxylate anions. In the RNAse P ribozyme the P4 RNA helix is proposed to provide anionic phosphodiester oxygen anions as the oxygen ligands to the two Mg ions. The ribozyme is known to use both metal cations to coordinate the prochiral proR oxygen of the phosphodiester bond to be cleaved. Howard *et al.*[20] suggest that the protein catalyst instead coordinates the proS oxygen of the precursor tRNA substrate internucleotide to be cleaved. The catalytic nucleophile in each case is proposed to be a metal cation-coordinated hydroxide equivalent. RNA and protein may have converged on a strategy to use metal-bound hydroxide equivalents for cleavage (and shuffling) of internucleotide phosphodiester bonds.

8.3.4 GlmS Ribozyme

In the absence of the inventory of the side chains of the twenty amino acids to act or assist in catalysis by protein enzymes, catalytic RNAs might have used low molecular weight heterocyclic metabolites as cofactors or coenzymes. There are hundreds of riboswitches[21] known for two such coenzymes, thiamine-diphosphate and FMN. They could have supplied carbanion chemistry and redox chemistry respectively to ribozymes. The bacterial mRNA encoding the glucosamine synthase (GlmS) enzyme converting fructose-6-phosphate plus glutamine to glucosamine-6-phosphate and glutamate has a glmS riboswitch in the 5′-untranslated region of the mRNA.[22] This riboswitch binds glucosamine-6-phosphate and activates the riboswitch to act as a ribozyme, cutting the mRNA and destroying the message. This constitutes a cofactor-dependent ribozyme or catalytic RNA for regulating metabolic flux. The suggested role for the glucosamine-6-phosphate is to function as a general acid catalyst, supporting protonation of the 5′-oxyanion of the leaving group fragment as the 2′,3′-cyclic-phosphodiester forms in the new 3′-fragment[23,24] (Figure 8.10).

8.4 Polynucleotide Phosphorylase: A Two-way Enzyme

Inorganic phosphate acts as a substrate for the enzyme polynucleotide phosphorylase, an exonucleolytic RNA cleaving enzyme working from the 3′ end. Because this enzyme is a phosphorylase, not a hydrolase and it is a

phosphodiesterase bond undergoing cleavage, the NDP product monomers released contain a pyrophosphate (phosphoric anhydride) linkage. Regiospecificity of action is such that 5′-nucleoside diphosphates are released from the 3′ end (Figure 8.11). In the case illustrated the terminal XMP residue in the mRNA is an AMP such that the released product is ADP.

In the presence of high concentrations of ADP, polynucleotide phosphorylase will run in the opposite direction, adding AMP residues to the end of the RNA chain.[25] As shown, this would be an alternate route to adding oligo- or poly-AMP moieties to mRNA tails; polynucleotide phosphorylase is an ancillary enzyme for forming mRNA poly A 3′-tails (Figure 8.12).

8.5 Sequence-specific DNA Restriction Endonucleases Make Double Strand Cuts

Two examples of nuclease families that recognize specific sequences in bacterial DNA and cut both strands are involved in defense roles, mediating distinction of ''nonself'' foreign DNA that has made its way into a bacterial cell from ''self'' DNA. The foreign DNA is usually bacterial DNA in the case of restriction endonucleases[7,26] and often viral DNA in the case of CRISPR endonucleases.[27]

More than 3000 restriction endonucleases have been detected in bacterial genomes.[28] The restriction terminology stems from the fact that the endonucleases are usually one of two components in a restriction–modification (RM) system that bacteria use to mark their own DNA. They then detect unmodified DNA as foreign and remove it or destroy its coding integrity by double strand endonuclease digestion.

The restriction modification is invariably enzymatic methylation by an *S*-adenosylmethionine-dependent methyltransferase, typically methylating one of the exocyclic substituents on a given DNA base within a given recognition sequence. There are several variations of the restriction methylase–endonuclease paired systems in terms of length of double strand DNA sequence recognized and sites of phosphodiester cleavage within or outside the recognition sites. The Type II restriction–modification systems number more than 900 members and are the simplest to understand for specific cuts at given oligonucleotide sequences.[28]

The two most common self-identifying methylations occur at the exocyclic N6–NH_2 of adenine bases and at C5 of cytidine residues. (Figure 8.13). In the *E. coli* RI system adenine methylation occurs in the hexanucleotide sequence G^{m6}AATTC. The corresponding *E. coli* RII methylase acts to create C^{m5}CNGG pentanucleotide sequences. Methylations usually occur protectively just after DNA replication.

Methylation at these sites blocks the subsequent action of the companion RI and RII endonucleases (Figures 8.13 and 8.14). The methylation mark indicates self-DNA and protects it from double strand hydrolysis. In the

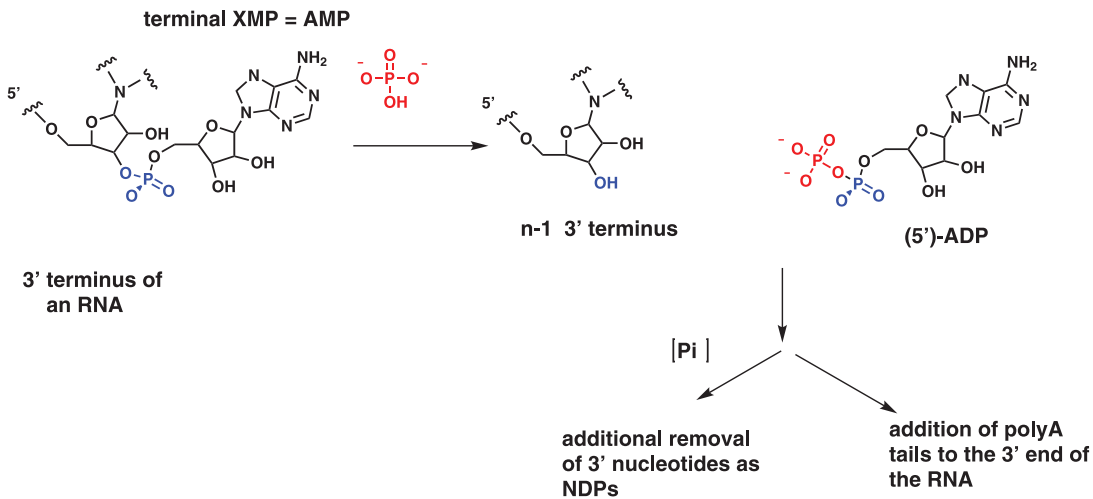


Figure 8.11 Polynucleotide phosphorylase can function as a phosphorylase or a sequence independent adenylyl transferase. In the phosphorolytic cleavage direction polynucleotide phosphorylase transfers a terminal XMP residue from the $3'$ end of an RNA to inorganic phosphate as an attacking nucleophile. The result is an $n-1$ shortened RNA and a molecule of product ADP (a net phosphodiester interchange with AMP).

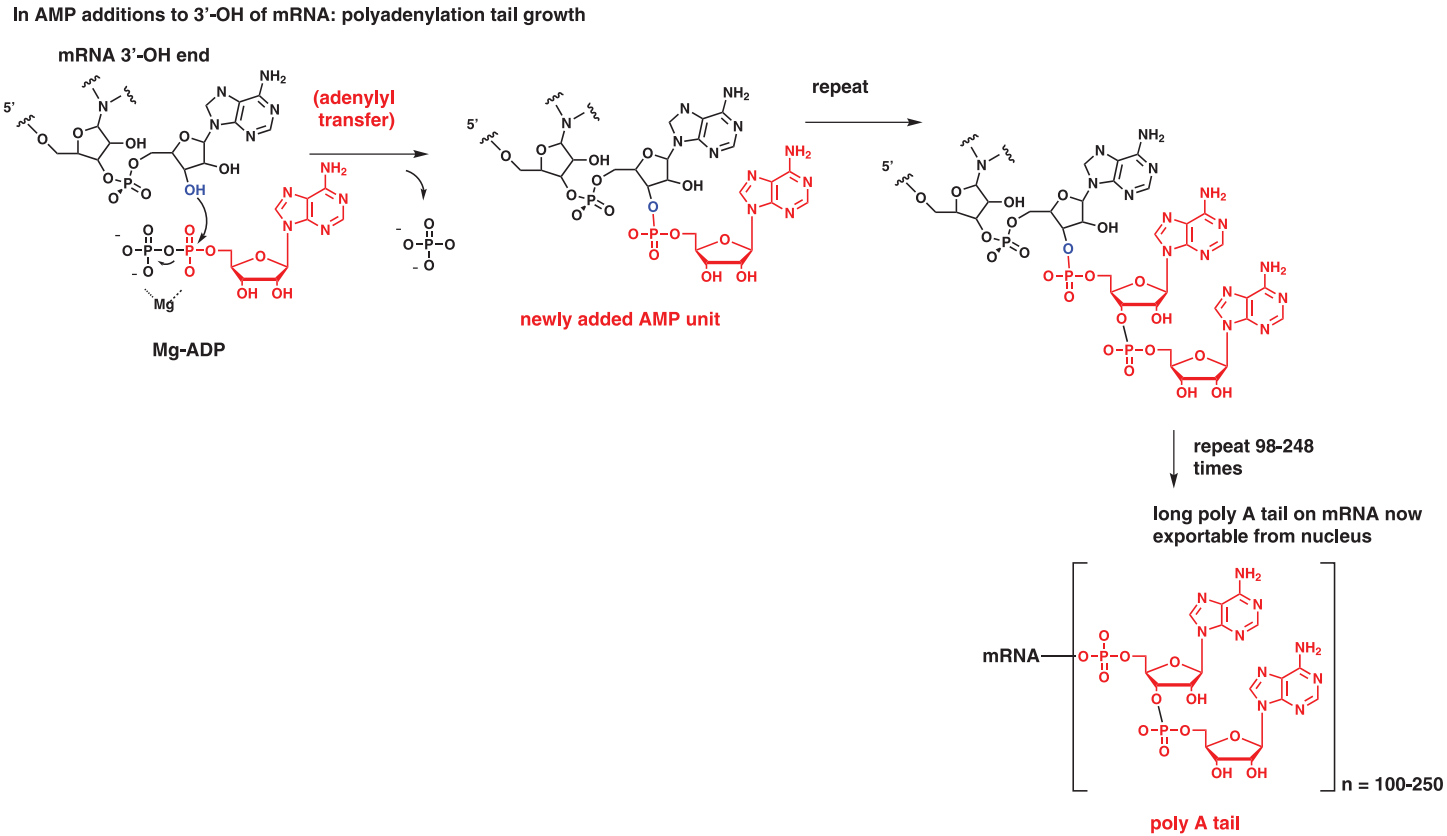


Figure 8.12 In the polymerase direction ADP, not ATP, is the donor of a nucleotidyl group to the 3′-OH end of an RNA. These are adenylyl transfers from a rare adenylyl donor and give Pi as a coproduct.

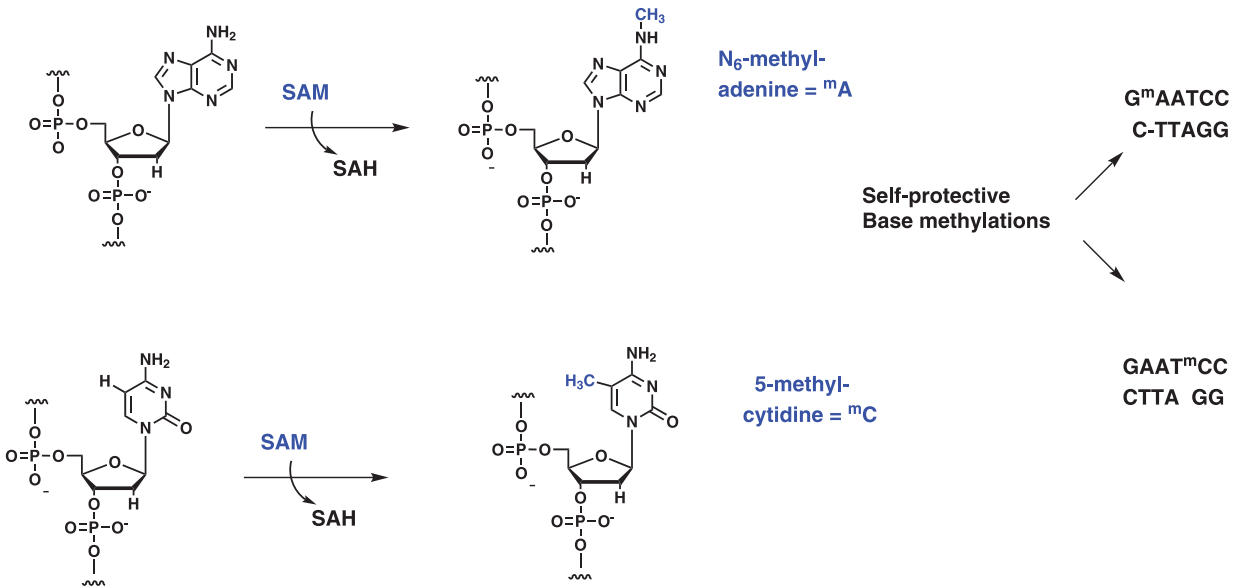


Figure 8.13 Logic and machinery of restriction modification system to distinguish ''self'' DNA from ''nonself''. Bacteria ''restrict'' their own DNA at four-base to six-base sequences by methylation, typically at an exocyclic amine of an A or C residue. SAM is the electrophilic methyl group donor.

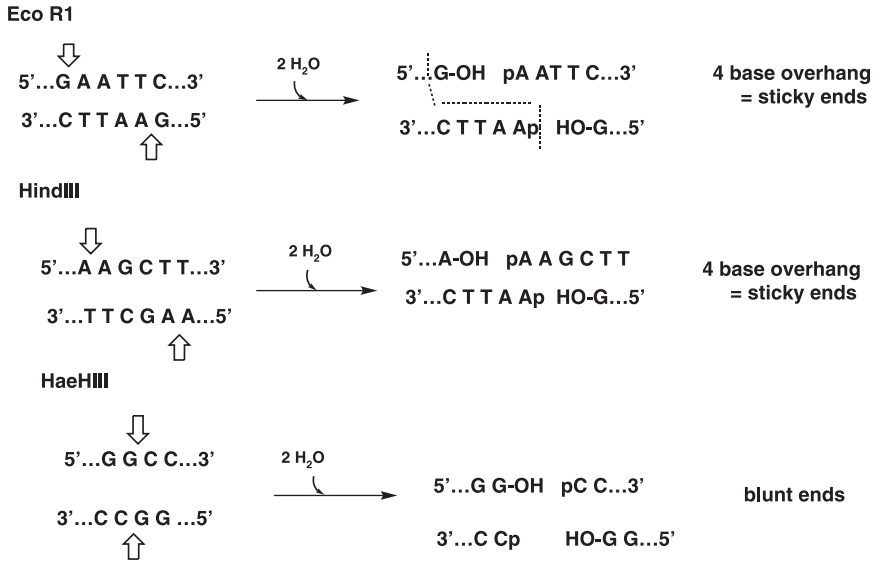


Figure 8.14 The paired restriction endonucleases hydrolyze unmethylated DNA sequences on both DNA strands. Shown are EcoR1 and HindIII sites of hydrolytic cleavage, leaving sticky end overhangs.

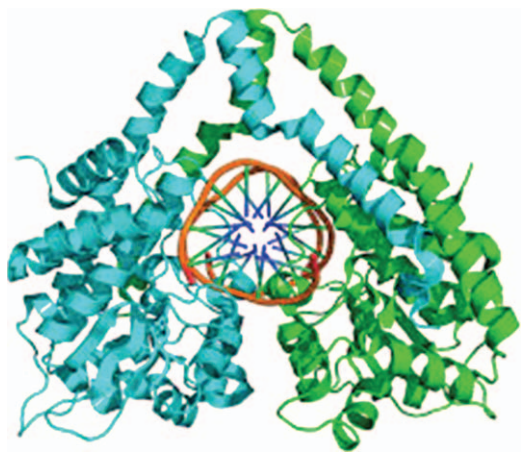


Figure 8.15 Picture of an endonuclease dimer wrapped around its double helical DNA target site (from pdb 1ESG, 2VLA).

absence of the covalent methyl marks the paired endonucleases recognize the unmethylated palindromic sequences in foreign DNA, often binding very tightly as dimers, one subunit on each DNA strand (Figure 8.14). The endonuclease dimers then proceed to execute hydrolysis of a specific phosphodiester bond within the recognition sequence, one on each DNA strand. Figure 8.15 depicts the two EcoRI subunits wrapped around a target DNA double strand helix.[29]

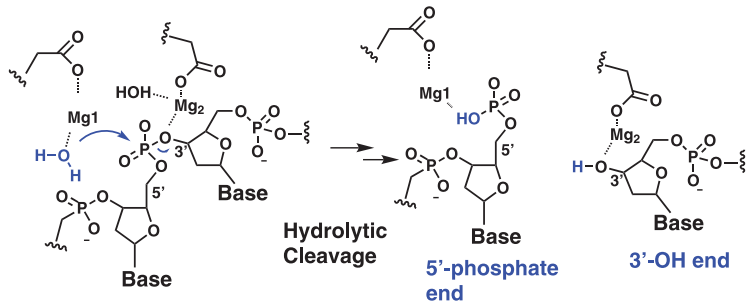


Figure 8.16 A two-metal cation-based hydrolytic cleavage of one of the two DNA strands by a restriction endonuclease, Regiospecificity is to create a 5′-P-end and a 3′-OH end in the cut strands. The second strand is also hydrolyzed at the cognate phosphodiester before release of the DNA product with a double strand break.

The dimeric nature of restriction endonucleases allows then to bind to both strands of palindromic DNA sequences and execute coordinated cuts at a specific internucleotide phosphodiester bond in each strand. The classic case in *E. coli* is the EcoRI restriction endonuclease dimer. Coordinated cleavage occurs at the internucleotide bond between the two thymidine (T) residues in each DNA strand (Figure 8.14). The active site in each enzyme subunit is a classic two metal phosphodiesterase, activating the bound water molecule for attack on the specific electrophilic phosphorus center (Figure 8.16). Both strands are cut before product release of two fragments that are double stranded except for the four base single strand overhangs, so called sticky ends (useful for reannealing the two fragments for *in vitro* pasting together of DNA strands[28]).

In contrast, the first restriction endonuclease studied, from *Hemophilus influenzae,* HindIII (the third RM system found in Hind strains), recognizes the palindromic double strand DNA sequence AAGCTT/TTCGAA. This dimeric nuclease hydrolyzes the GpC internucleotide phosphodiester on each strand before fragment release and also generates sticky ends. HaeHIII on the other hand makes blunt end double strand hydrolytic cuts, but at more promiscuous GGCC/CCGG palindromic tetranucleotide sequences (Figure 8.14). Hundreds of characterized RM bacterial pairs confer different recognition sites and different sticky or blunt end fragments for thousands of DNA constructs.

8.6 CRISPR–CAS9: Another Bacterial Defense Strategy Involving Specific Double Strand Cuts of Foreign DNA

A second bacterial defense strategy for distinguishing ''self *vs.* nonself'' DNA sequences and cleaving the foreign ones by double strand endonuclease

action is exemplified by CRISPR. CRISPR stands for ''Clusters of Regularly Interspersed Short Palindromic Repeats''.[30] The phenomenon and associated genetic editing technology have mushroomed in the last decade.[31–33] Thousands of papers have been written on different types of CRISPR systems, their biological roles and most abundantly the biotechnology roles of CRISPR and its associated DNA endonuclease Cas9 from *Streptococcus pyogenes*.

Our comments on CRISPR–Cas9 are only summary in form and content to compare the biological function akin to restriction endonucleases in cleaving a specific pair of phosphodiester linkages in a foreign DNA substrate. There are several mechanistic differences that distinguish execution of the CRISPR strategy from the restriction–modification strategy although both use specific endonuclease catalysts to make double strand hydrolytic cuts in target foreign DNAs.

One distinction is that the Cas9 endonuclease is guided to a DNA target by a pair of RNA molecules. One RNA [termed trans-activating CRISPR (tracRNA)] activates the nuclease from an inactive basal state to an activated nuclease conformer. The second RNA [crispr RNA (crRNA)] is the guide RNA. It has a sequence of 17–20 bases complementary typically to a phage DNA sequence that has previously been packaged into the bacterial genome as a memory unit for transcription (Figure 8.17). The guide RNA unwinds the target DNA double strand, forming an RNA–DNA double strand hybrid (Figure 8.18). That is the target for the Cas9 endonuclease,[34] acting as a bimetallic active site hydrolytic catalyst on each of the two unwound DNA strands. The target DNAs are the CRISPR repeat loci that reflect prior invasions and packaging of invader DNA (Figure 8.17).

The transformation of the native bacterial system to a biotechnological user- friendly system in 2012 was the synthetic connection of the two RNA pieces into a single guide RNA. Now the Cas9 endonuclease was efficiently recruitable to any desired stretch of DNA *via* formation of the RNA–DNA duplex[36] (Figure 8.18). The endonuclease action occurs three base pairs upstream of a N–G–G trinucleotide anchoring sequence (designated the PAM sequence Figure 8.18) and cuts a single phosphodiester bond in each DNA strand.

Intriguingly, Cas9 acts as a monomer. It has some six domains, two of which are endonucleases (Figure 8.19A). One endonucleolytic domain of the HNH family of nucleases cleaves the coding strand. The second endonuclease domain is a member of the resistance to ultraviolet C (RuvC) nuclease superfamily. It cleaves the noncoding strand before release of the blunt end DNA fragments (Figures 8.17 and 8.19B). The double strand cleavage, as in the restriction endonuclease cases, makes endogenous repair of the foreign target DNA fragments difficult to impossible and nullifies the gene-coding information in the foreign DNA. An enormous range of modification to the CRISPR–Cas reagents have been reported with many more to come to improve site specific editing in almost any genome.

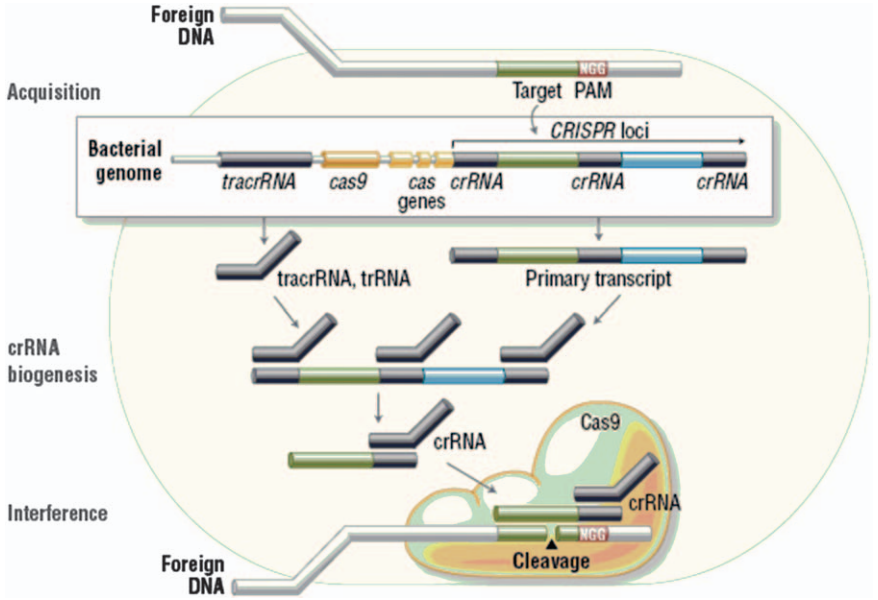


Figure 8.17 Schematic for action of CRISRP–Cas9 system as a site-specific double stranded DNA endonuclease.[35] Prior exposure of bacteria to foreign DNA results in prior packing of small bits of foreign DNA sequence in the CRISPR loci. In this particular type of CRISPR response bacteria make a tracRNA (to activate Cas9) and express many crRNAs from the CRISPR loci. The crRNA hybridizes to the complementary sequence on the invading foreign DNA. The associated CAS9 endonuclease activated by tracRNA cuts just upstream of the PAM sequence in the foreign DNA. Cas9 has two nuclease domains, one for each foreign DNA strand. Reproduced from ref. 35, www.neb.com (2020), with permission from New England Biolabs, Inc.

Drawing back from the details of the restriction–modification enzyme pairs and the CRISPR–Cas with its RNA-guided endonuclease, both are strategies for finding foreign DNA amidst a sea of host bacterial DNA. Bacteria are under constant assault from foreign competitors, bacteria and viruses. There are powerful selections to maintain and express the ''self *vs.* nonself'' DNA detectors. The destruction of foreign coding information is to enact multiple, specific double strand cuts of target DNAs before the fragments are released. Double strand breaks are much more difficult than single strand breaks to repair. Phosphodiesterases are powerful weapons for manipulation of DNA structure and content.

The many scholars and practitioners of CRISPR–Cas systems will find many alternate passages into the literature, including the less widely used Type III CRISPR–CAS systems. We noted in Chapter 4 that bacteria harboring the type III machinery convert ATP to cyclo-oligoadenylates such as $c(AMP)_4$ and $c(AMP)_6$ that then function as allosteric activators for certain RNAses that act to hydrolyze the foreign nucleic acid.[38,39]

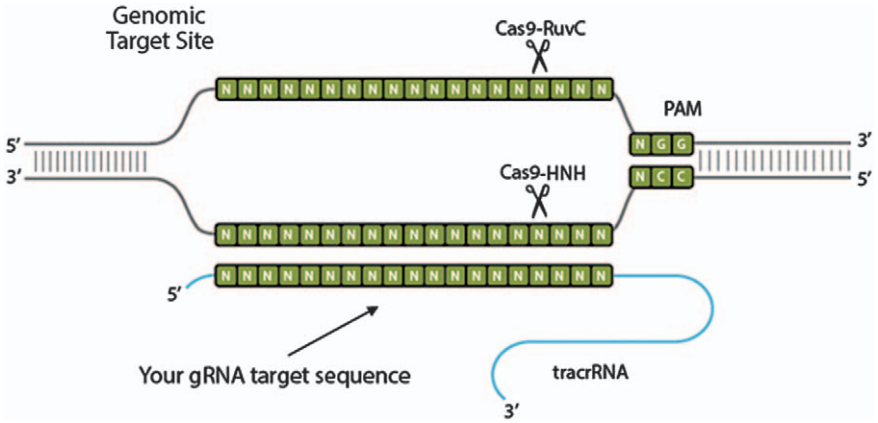


Figure 8.18 Schematic for action of CRISRP–Cas9 system as a site-specific double stranded DNA endonuclease.[35] Biotechnology adaptation of fusing tracRNA to any desired RNA guide sequence allows recruitment of Cas9 to any desired DNA sequence where Cas9 will do double strand cuts three nucleotides upstream of the PAM sequence.
Reproduced from https://www.sigmaaldrich.com/technical-documents/articles/biology/genome-editing-in-plants-with-crispr-cas9.html with permission.

8.7 Type II Topoisomerases

A variant of enzymatic phosphodiester bond cleavage in DNA is practiced by topoisomerase catalysts, enzymes that cut and then religate either DNA single strands or DNA double strands to reduce the twist in newly synthesized DNA or to decatenate circular DNAs produced by DNA polymerases.[40] Type I topoisomerases effect single strand cleavages and religations. The type II topoisomerases cleave both strands of a target DNA four bases apart on each strand (Figure 8.20) by engaging in covalent catalysis. The dimeric enzyme uses a phenolate side chain of a tyrosyl residue in each subunit to attack the particular internucleotide phosphodiester on each DNA strand (Figure 8.21).

The enzyme uses a two-metal cation strategy to align the cleavable but otherwise stable DNA phosphodiester bond in each DNA strand and to activate the phenol of Tyr_{805} and Tyr_{821} as the catalytic nucleophiles that act on the two DNA strands. The strategy for each strand's phosphodiester bond cleavage parallels that of alkaline phosphatase In the alkaline phosphatase case[41] the substrates are typically phosphate monoesters (Figure 7.18) but can also be phosphodiesters (where that enzyme was activating a serine side chain alkoxide rather than a tyrosyl side chain phenoxide.)

The result is creation of a nick on each strand, four bases displaced from each other (Figure 8.20). Each DNA strand has become covalently attached to the active site tyrosyl residue through a new phosphodiester link. The gap in duplex DNA thereby created is wide enough for another DNA double helix to

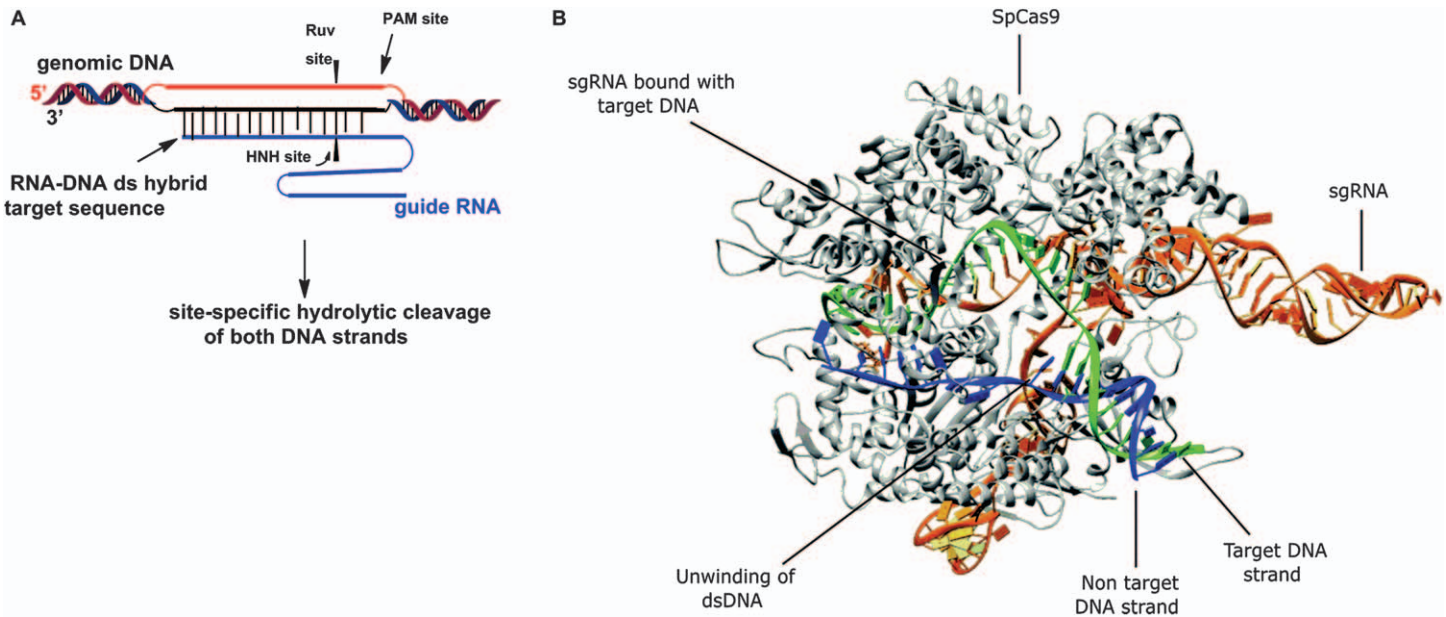


Figure 8.19 Cas9 a double strand sequence specific DNA endonuclease.[37] (A) The primary sequence of Cas9 indicates two catalytic nuclease domains (from two distinct families) and regulatory and recognition domains. (B) The X-ray structure of Cas bound with its single guide RNA (sgRNA) and the two strands (target strand and complementary strand) of substrate DNA. Note that the endonuclease unwinds the DNA to allow formation of the RNA–DNA hybrid. (PDB 6AI6).

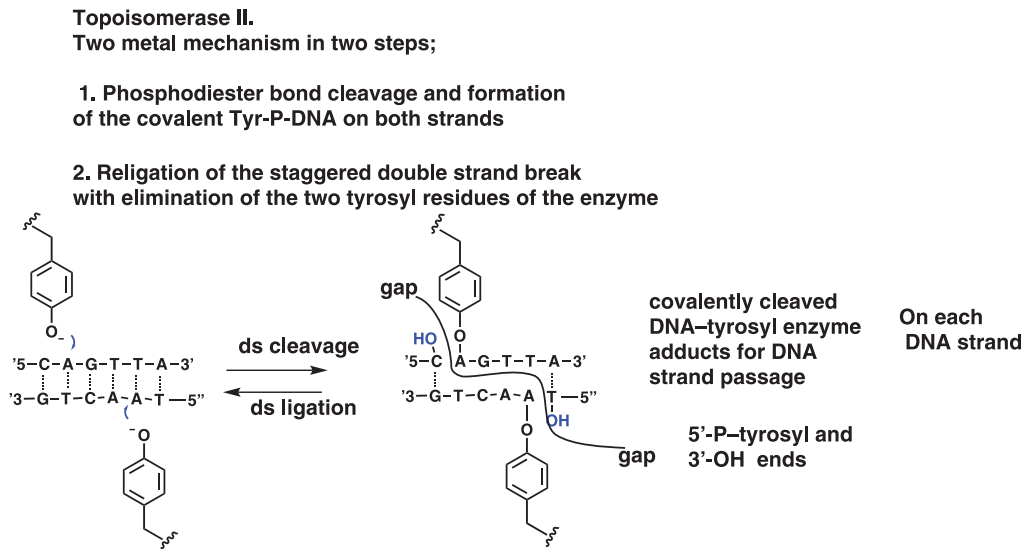


Figure 8.20 Topoisomerases reversibly cleave DNA, pass DNA strands through the cut, then religate the strand breaks, in the service of altering the topological winding numbers on overwound or underwound DNA. Topoisomerases I pass one strand at a time. Topoisomerases II pass both double strands of DNA through a transient break. The gapped ends on the 5′-side of the break are held covalently in phosphodiester linkage to a tyrosyl residue. Thus, the strand break is an energy neutral phosphodiester interchange.

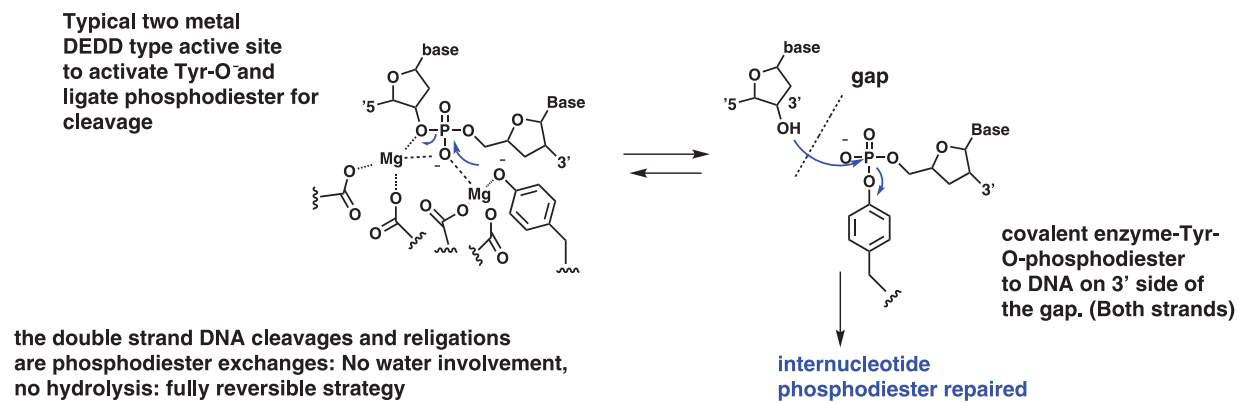


Figure 8.21 Topoisomerases I pass one strand at a time. Topoisomerases II pass both double strands of DNA through a transient break. The relegation: active site of topoisomerase is a typical two metal cation environment for holding the phosphodiester and activating the tyrosyl residue for attack in the first step (strand cleavage) and elimination in the second step (religation).

pass through the (four bp) gap, deconcatenating circles or reducing twist in linear DNA molecules. Once strand passage has occurred, the 3′-OH terminal residue on one side of the gap, attacks the protein-5′ phosphotyrosine ester on the other side of the gap. As the tyrosyl residue is expelled, the internucleotide 3′,5′-phosphodiester bond is religated on each DNA strand, repairing the two single-strand gaps.

The double strand cleavage and reverse step of double strand relegation are each phosphodiester exchanges. No water and no hydrolysis is involved. The forward and back direction reactions are fully reversible, reflecting two chemically difficult maneuvers on otherwise stable DNA. If water had been the kinetically competent nucleophile for double strand breaks, the topoisomerase would instead be a restriction endonuclease variant (Figures 8.17 and 8.18).

The covalent tyrosyl-DNA intermediates hold onto the gapped strands so they do not dissociate away from the topoisomerase active site (Figure 8.22). Furthermore, the phosphodiester bond between the 5′ phosphoryl group and the tyrosyl-phenol-oxygen atom sets up the gap for religation, repairs the double strand breaks, and allows for error-free DNA strand passage. One might compare this logic to that used by DNA ligases (Chapter 4).

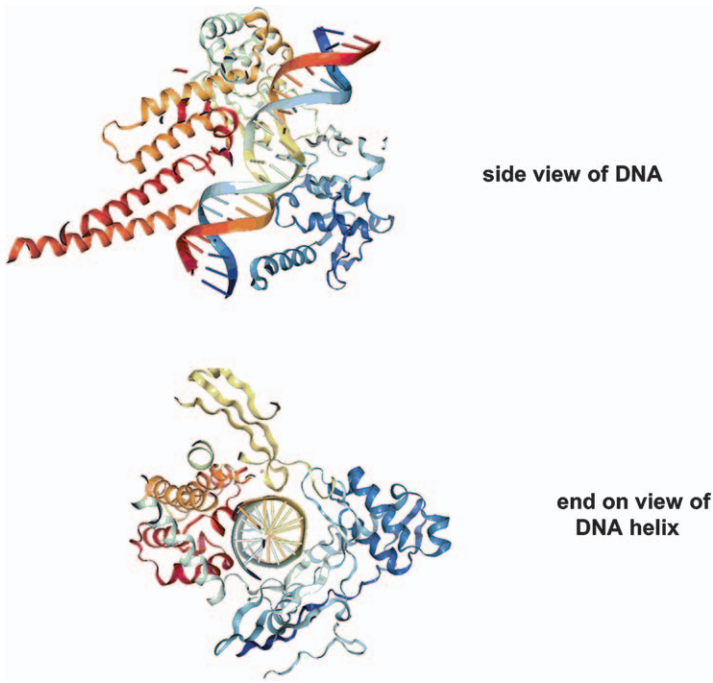


Figure 8.22 Topoisomerases I pass one strand at a time. Topoisomerases II pass both double strands of DNA through a transient break. Two views of a DNA oligonucleotide bound to human topoisomerase I (drawn from PDB 1A36).

8.8 Cyclic Nucleotide Phosphodiesterases

A second separate metabolic arena for control of spatial and temporal phosphodiester molecular scaffolds involves the cyclic nucleotides. These include the first generation, now classical, cyclic AMP and cyclic GMP second messengers. They also encompass the more recently discovered cyclic-dinucleotides such as the cyclic di-GMP and also the 2′,3′-cGAMP ligand active in the human immune system. Together with the nucleotidyl transferases that form cyclic nucleotides (discussed in Chapter 4), the corresponding phosphodiesterases (PDEs) control their lifetime and intracellular concentrations of these cyclic nucleotide second messengers.[42]

Figures 8.23–8.27 schematize features of the cellular activities of cAMP and cGMP and the intracellular sites of action and selectivities of some of the multiple PDE isoforms for controlling the lifetimes of these second messenger molecules in distinct cellular compartments. In addition to the

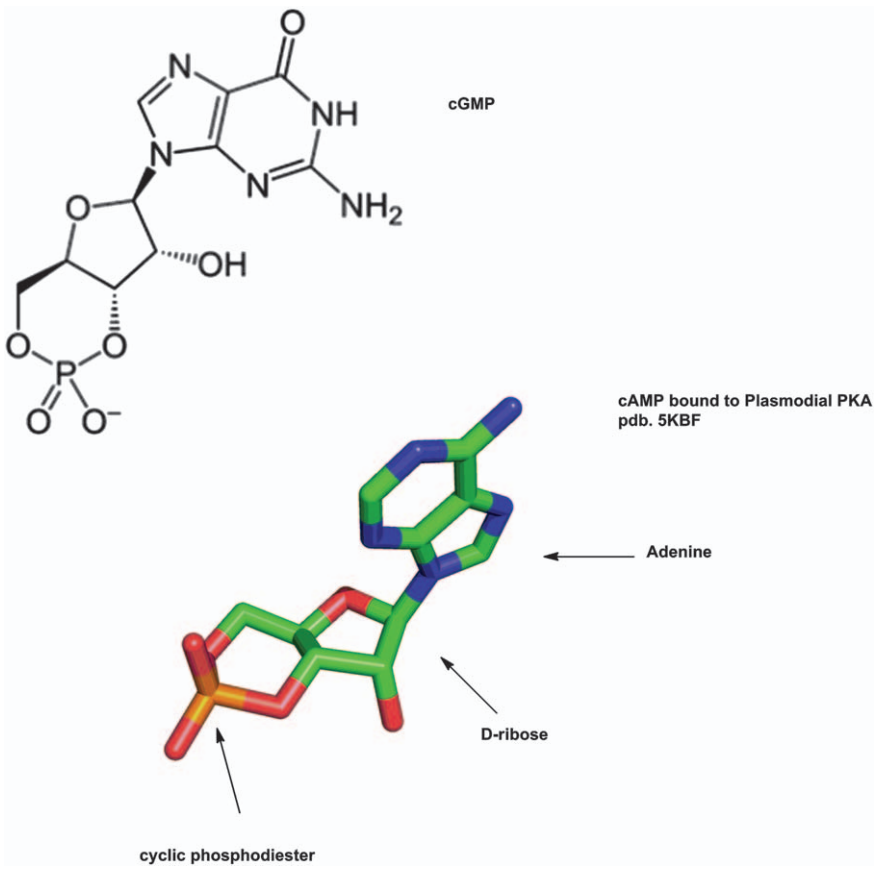


Figure 8.23 Cartoons reflecting some of the multiple signaling roles of cAMP and cGMP and the controls imposed by different cyclic nucleotide phosphodiesterases. cAMP and cGMP.

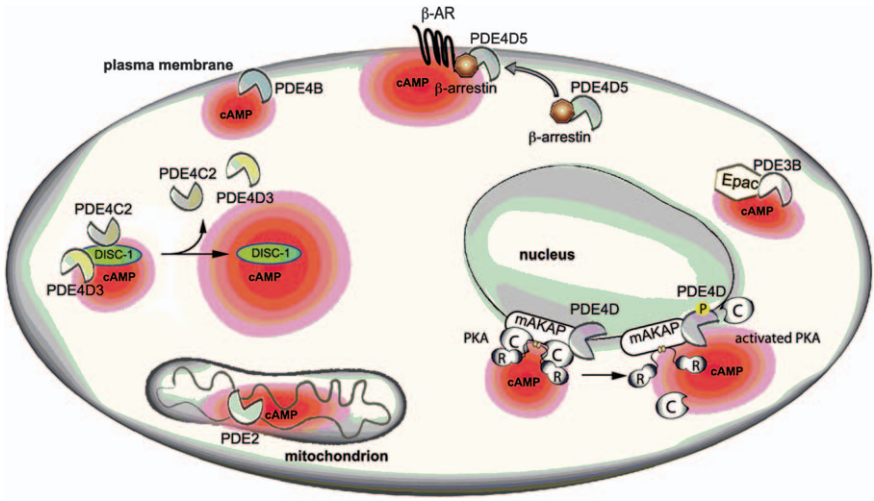


Figure 8.24 Seven phosphodiesterases (PDEs) are localized at different subcellular locations to control cAMP or cGMP lifetimes in those regions.

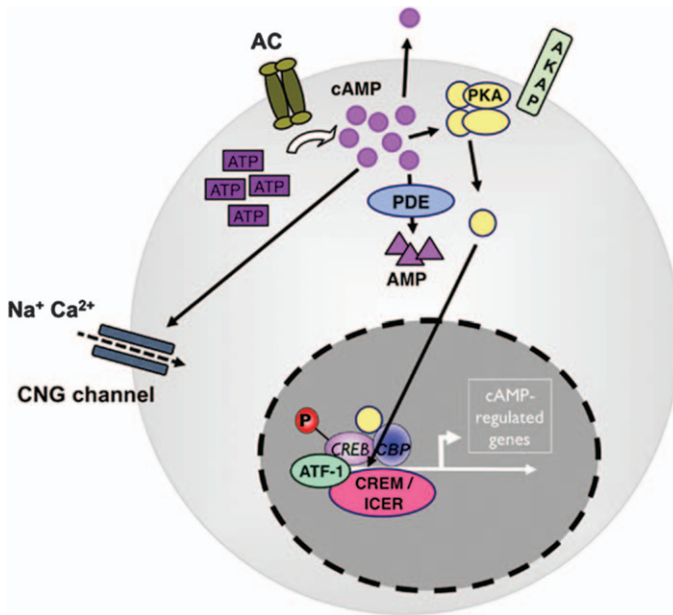


Figure 8.25 Roles for cAMP in activation of PKA and sending its subunits to the nucleus and role of cAMP in gating cell membrane Na/Ca channels.[44] Reproduced from ref. 44, https://doi.org/10.3389/fimmu.2016.00123, under the terms of a CC BY 4.0 license, https://creativecommons.org/licenses/by/4.0/.

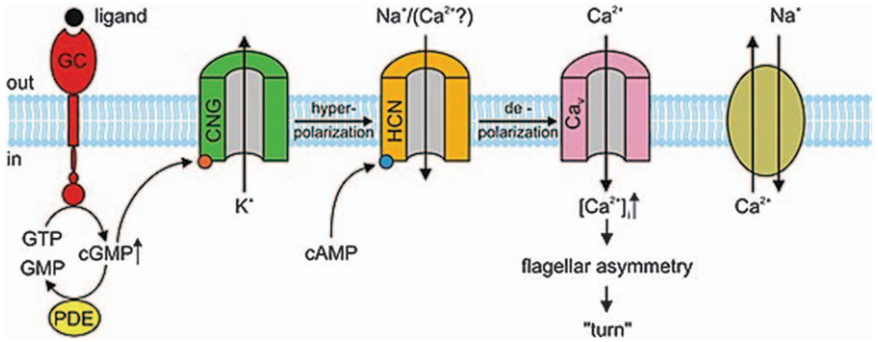


Figure 8.26 cGMP regulates the K channel CNG, while cAMP regulates the HCN Na/Ca ion channel.[43]
Reproduced from https://commons.wikimedia.org/wiki/File:Signal_transduction_in_sea-urchin_sperm_chemotaxis.jpg, courtesy of Prof. Dr. U.B. Kaupp, under the terms of a CC BY-SA 3.0 license, https://creativecommons.org/licenses/by-sa/3.0/deed.en.

roles of cAMP and cGMP as activating ligands for protein kinase A (PKA) and protein kinase G (PKG), respectively, the cyclic nucleotides are allosteric activators of different ion channels (Figure 8.26). Figure 8.25 is a reminder that cAMP is an activator of the CREB transcription factor for selective gene expression in nuclei. Figures 8.24 and 8.27 indicate subcellular localized actions of distinct PDEs and indicate the activity of PDE2 and PDE3 for hydrolysis of both cAMP and cGMP while PDE5 is selective for cGMP hydrolysis.

There are some eleven separate families classified as cyclic nucleotide PDEs, numbered PDE1–PDE11, several of which have been targets for pharmacological agents[42] (see Figure 8.24). Typically, the organization encompasses a catalytic phosphodiesterase domain attached to a regulatory domain that may confer selectivity for cellular location and/or partner protein recognition to attain effects on subsets of signaling pathways (Figure 8.24). Inhibitors of specific PDE isoforms, for example PDE5, have led serendipitously to Viagra (sildenafil), Cialis (tadalafil), Levitra (vardenafil) and other drugs for treating dysfunctional male penile erection[46] (Figure 8.28). These are competitive inhibitors of cGMP and consequent inhibition raises the cellular levels of cGMP. One can detect the structural analogies to the bicyclic guanine nucleus embedded within the three inhibitors of Figure 8.28.

The 11 subfamilies vary in cytoplasmic *versus* membrane locations and the nature and degree of their posttranslational modifications (affecting subcellular locations, including membrane targeting isoprenylations), some of which are schematized in Table 8.1.[42]

Figure 8.29, taken from a review by Bender and Beavo,[42] indicates some partially overlapping functions and tissue expressions of the eleven PDE subgroups as of the year 2006. All that have been examined structurally and

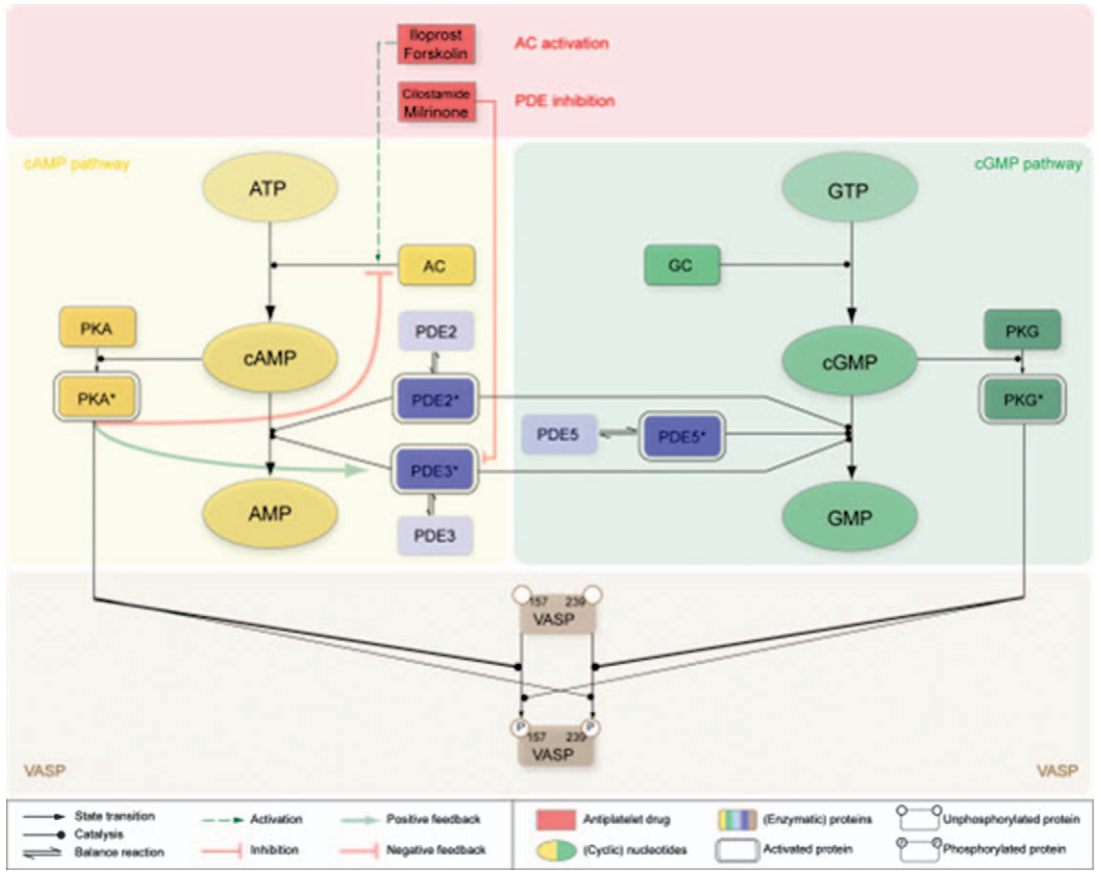


Figure 8.27 Orientation of cyclic GMP in a phosphodiesterase active site, with placement of the two metal cations required for activity.[45] Reproduced from ref. 72 with permission from The American Physiological Society, Copyright 2012.

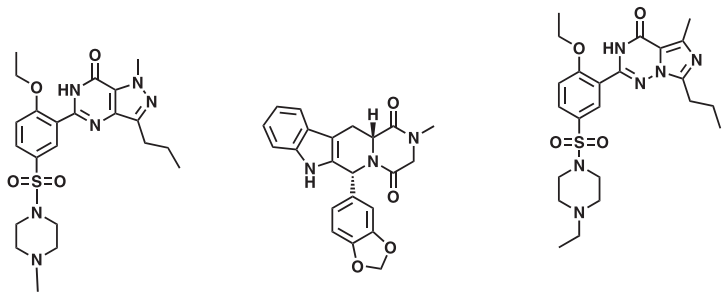


Figure 8.28 Three inhibitors of human PDE5 that have been approved for erectile dysfunction.

Table 8.1 Proposed functions for eleven PDE family members.[42] Functions for PDE1–6 and specific isoforms thereof have been characterized physiologically and pharmacologically. Predicted isoforms PDE7–11 have been less well differentiated.

Phosphodiesterase family member	function
PDE1	Vascular smooth muscle contraction; PDE1B dopaminergic signaling and immune cell activation; PDE1C vascular smooth muscle, sperm function, neuronal signaling
PDE2	Mediates cross talk between cAMP and cGMP pathways, aldosterone secretion from adrenals; cAMP regulation of cardiac muscle calcium channels
PDE3	3A regulates cardiac contractility, platelet aggregation, vascular smooth muscle, contraction; 3B mediates leptin signaling, insulin signaling, and renin release
PDE4	Monocyte and macrophage activation, B neutrophil infiltration, fertilization, brain function
PDE5	Smooth muscle contraction regulation (penis and lung); platelet aggregation
PDE6	Signal transduction of the photoresponse
PDE7–11	Less well characterized functions

mechanistically are members of the superfamily of two-metal ion phosphodiesterases, with regioselectivity to cleave the P–O bond to the 3′ carbon to produce 5′-GMP or 5′-AMP (Figure 8.29). Consistent with hydrolytic cleavage of a phosphodiester the fragments are an alcohol, and a phosphate monoester. Because the substrates are cyclic phosphodiesters, both the alcohol and the phosphate monoester groups are contained in the same product scaffold of AMP and GMP. The linear AMP and GMP products instantly rejoin the major nucleotide monophosphate cellular pools. Presumably, the different PDEs act on distinct subpools of cAMP and cGMP in distinct subcellular regions and in distinct cell and tissue types.

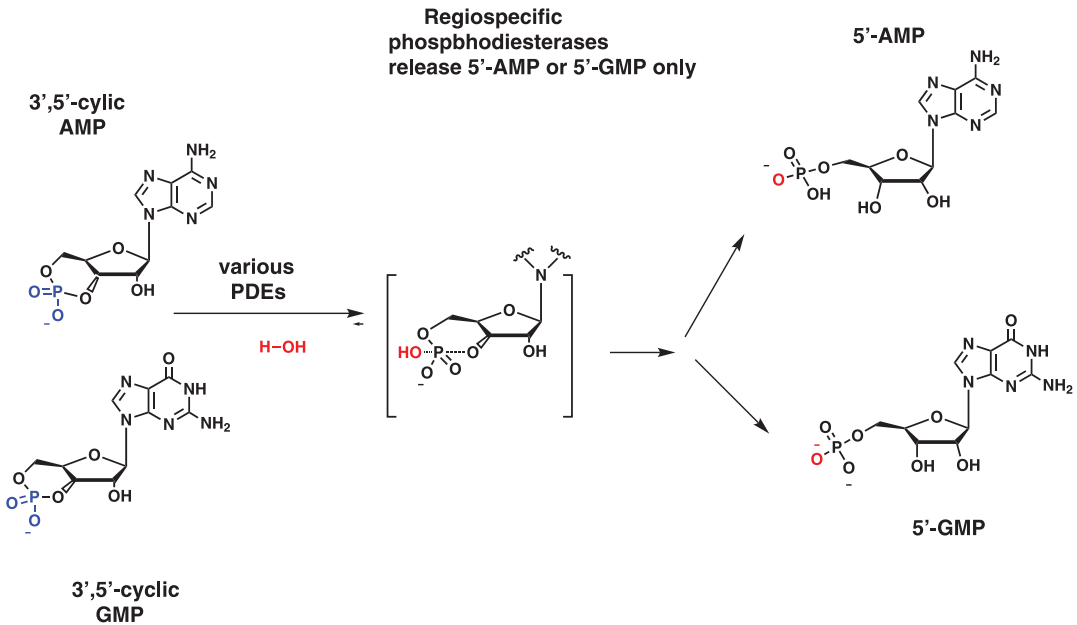


Figure 8.29 Schematic for cAMP and cGMP hydrolysis to the linear 5′-AMP and 5′-GMP.

For example, PDE1 members are activated by calmodulin and calcium ions. PDE2s hydrolyze both cyclic AMP and cyclic GMP, and are allosterically activated by both cyclic nucleotides. PDE3s hydrolyze cAMP but are inhibited by cGMP. PDE4s are selective for cAMP hydrolysis and up to 20 variants are known. PDE5s are regulators of smooth muscle contraction and the selective inhibitors sildenafil (Viagra) and atadalafil (Cialis) have been widely used for erectile dysfunction as noted in Figures 8.23–8.27.[46] PDE6s encompass photoreceptor PDEs for cGMP hydrolysis, when activated by the small GTPase protein transducing in visual cycles. Comments that lead to identification of the functions of the remaining five PDEs are described in depth in the 2006 review article by Bender and Beavo.[42]

8.9 Cyclic Dinucleotide Phosphodiesterases

8.9.1 Bacterial Phosphodiesterase Activity on 3′,3-di-cyclic-GMP and di-cyclic-AMP

Given the enormous range of bacterial responses to di-cyclic-GMP and di-cyclic-AMP – initiation and anti-termination effects on gene transcription, bacterial cell cycle progression, flagellar assembly, and specific target proteolysis (Figure 8.30) – it is not surprising that bacteria tightly control both synthesis and degradation of 3′,3′-di-cyclic-GMP and the corresponding di-cyclic-AMP.[47,48]

As one measure of the importance of temporal and spatial balance of subpools of these di-cyclic nucleotides as regulators, the results of bioinformatics analysis indicate that the genome of the *E. coli K12* strain encodes 12 di-cyclic nucleotidyl transferases. These di-c-GMP or di-cAMP synthases are identified by a GGDEF pentapeptide signature. Furthermore, the *E. coli K12* genome encodes 13 proteins with the EAL or HD-GYP residue signatures of di-cyclic nucleotide hydrolases: 25 proteins presumptively dedicated to balancing synthesis and hydrolytic degradation of di-cGMP and/or di-cAMP.[49]

The EAL-signature di-cyclic-GMP phosphodiesterases, and corresponding hydrolases for di-cyclic-AMP are, as might be expected, members of the large family of two-metal catalysts, one divalent cation bridging the phosphodiester, the second activating the nucleophilic water molecule.[50] Both the di-cyclic AMP and di-cyclic GMP phosphodiesterases carry out two catalytic cycles on the starting substrates. Enzymatic hydrolysis of the first 3′,5′-phosphodiester bond creates a linear dinucleotide, either pGpG or pApA. These are canonical RNA dinucleotides: a 5′-phosphate-then AMP or GMP in 3′,5′-phosphodiester linkage to AMP or GMP, with free 3′-OH (Figure 8.30). These dinucleotides can also then be acted on by the same PDEs to release two 5′-GMP or two 5′-AMP product molecules. If the regiospecificity had been the other way, ultimately releasing 3′-GMP or 3′-AMP, these unusual nucleotides would represent stranded material; they would require

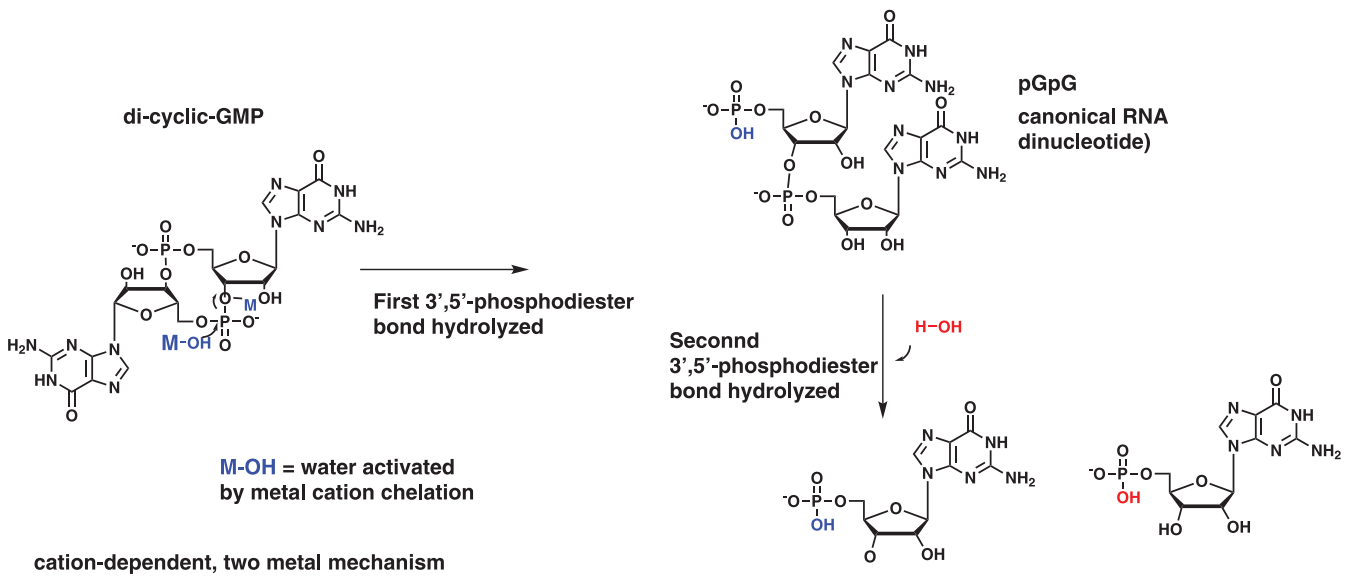


Figure 8.30 Enzymatic hydrolysis of 3′,3′-dicyclic GMP proceeds in two enzymatic steps. Cleavage of the first 3′,5′-phosphodiester bond gives the linear dinucleotide pGpG. Hydrolysis of the internucleotide 3′,5′-phosphodiester bond then releases two molecules of GMP.

phosphatase action and then expenditure of an ATP in a kinase reaction to get back to canonical 5′-nucleoside monophosphates.

8.9.2 Phosphodiesterase Activity on Both 2′,5′ and 3′,5′ Phosphodiester Bonds of cGAMP

The activating role of cGAMP on STING and the expression of interferon genes in innate immune responses to double stranded DNA detection in cell cytoplasm (detailed in Chapter 4 above)[51,52] indicate that it is important for cells to balance synthesis and degradation of cGAMP.[53,54] An intrinsic transmembrane glycoprotein in plasma membranes and in the Golgi complex organelle of secretory pathways has been found to hydrolyze Mg–ATP to AMP and PPi, but also has high activity in hydrolytic opening of 2′,3′-cGAMP (Figure 8.31). The protein is named ecto-nucleotide pyrophosphate phosphodiesterase 1 (ENPP1) for the ATP cleavage activity.[51,55]

The PPi released from ATP hydrolysis (a highly unusual nucleotidyl transfer to water) is thought to be an inhibitor of bone mineralization in osteoclasts and osteoblasts. ENPP1 does not hydrolyze a synthetic 3′,3′-cGAMP regioisomer, consistent with physiological specificity as an enzyme that opens and thereby deactivates 2′,3′-cGAMP, removing this dicyclic nucleotide as stimulator for detection of foreign DNA in cell cytoplasm by the innate immune system.

As with the bacterial 3′,3′-cyclic dinucleotide phosphodiesterases that linearize their substrates to pApA (or pGpG), the 2′,3′-cGAMP is opened enzymatically first at the 2′,5′-phosphodiester bond to produce the linear dinucleotide pApG (5′-phospho AMP-3′,5′-GMP) (Figure 8.31). This linear dinucleotide can be released or can also undergo hydrolysis of the remaining 3′,5′-phosphodiester bond to give 5′-AMP and 5′-GMP, standard mononucleotide metabolites. In this second step the nascent pApG is proposed to flip over in the active site as depicted in Figure 8.31, giving the catalytic di-zinc ions a similar architecture to act on the remaining 3′,5′-phosphodiester in the bound pApG conformer. As shown, the water from the first phosphodiester cleavage–linearization ends up in the 5′-phophoryl group of AMP. The water from the second hydrolysis, of the flipped 3′,5′-phosphodiester in pApG is incorporated into the 5′-phosphate of GMP (Figure 8.31).

8.10 Phospholipases C and D

Four classes of phospholipases act hydrolytically on the scaffolds of glycerol-based membrane phospholipids.[56,57] Two of them are actually carboxyesterases hydrolyzing the long chain acyl-*O*-ester linkages at carbons 1and 2. In this family are phospholipase A_1 and A_2 family members: acyl chain esterases regiospecifically hydrolyzing the acyl chain oxoesters at C_1 or C_2 of the glycerol backbone of phospholipid substrates (Figure 8.32). The other

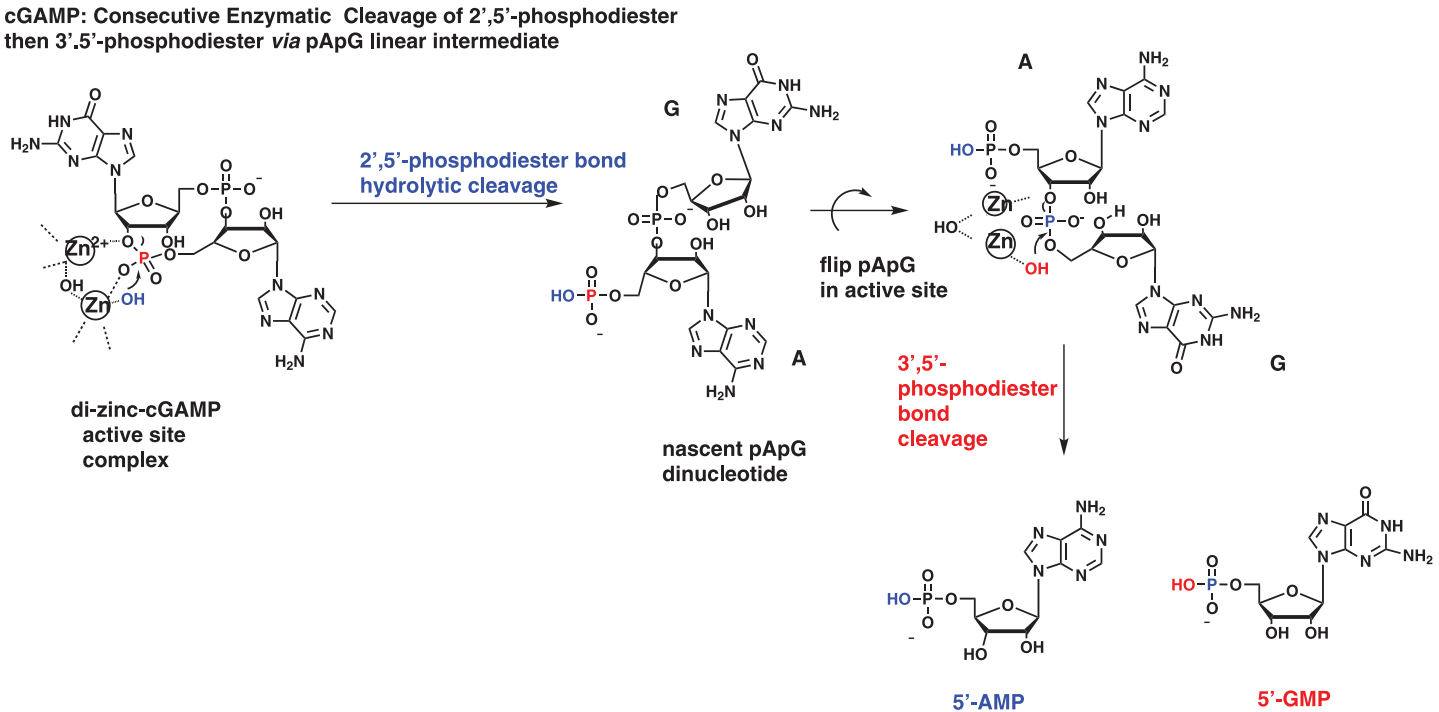


Figure 8.31 Enzymatic hydrolytic removal of the signaling molecule cGAMP. As with other phosphodiesterase family members a di-zinc enzyme carries out hydrolysis. First, the unusual 2′,5′ phosphodiester bond is hydrolyzed to give nascent pApG as bound dinucleotide intermediate. It is proposed that this flips over in the active site to allow activated water to attack the 3′,5′-phosphodiester internucleotide bond to release GMP and AMP.

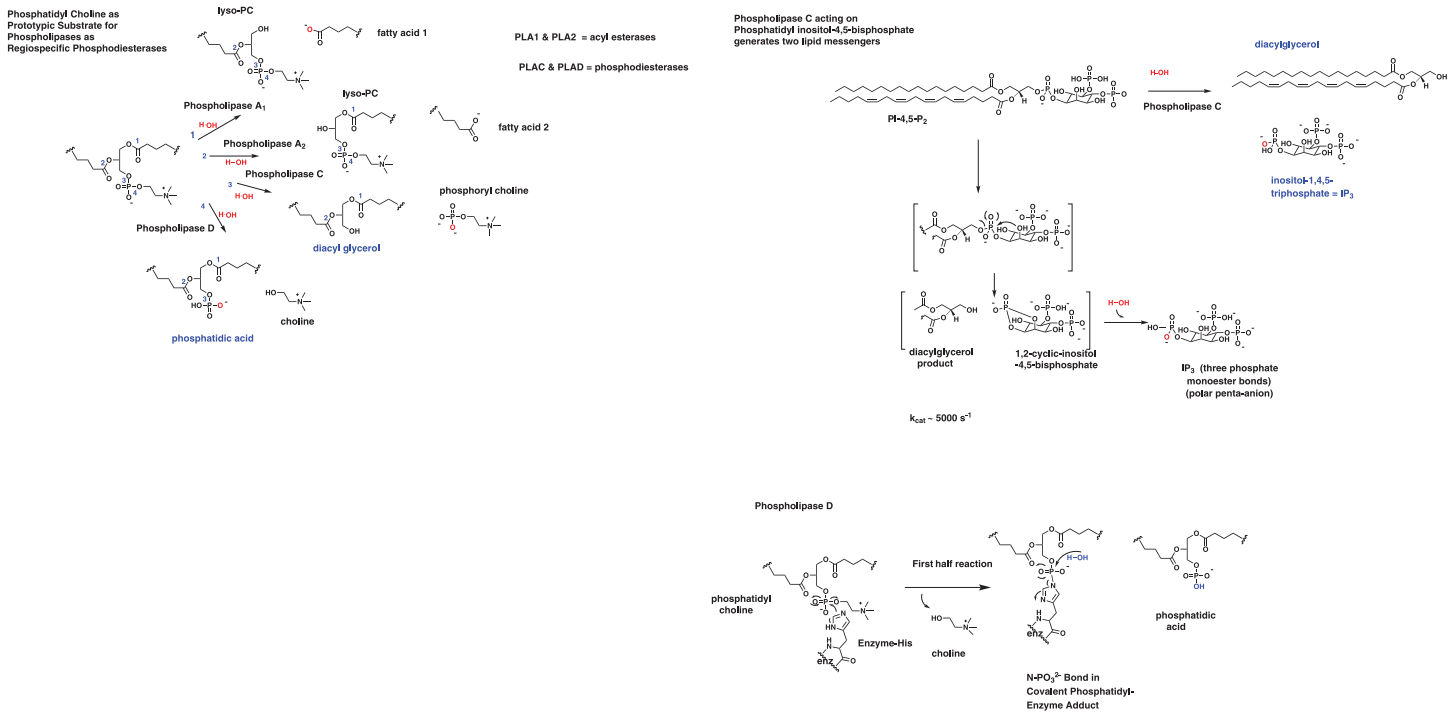


Figure 8.32 Phosphatidylcholine an abundant membrane phospholipid as prototypic substrate for four categories of phospholipases. Phospholipase A1 and A2 are carboxylesterases acting at the indicated acyl chains. Phospholipase C and Phospholipase D cleave on either side of the phosphodiester linkage in the polar head group. Phospholipase C would release phosphorylcholine and the diacylglycerol. Phospholipase D would release choline and the phosphatidic acid.

two phospholipase categories, phospholipase C and D family members, are actual phosphodiesterases that show distinct regiospecificities for cleavage on either side of the phosphodiester bond.

8.10.1 Phospholipase C

Phospholipase C isoforms are sluggish catalysts for the major membrane phospholipid phosphatidylcholine, but highly active against the phosphatidylinositol-4,5-bisphosphate (PIP_2), the minor phospholipid[58] discussed in Chapter 7. Phospholipase C acts on that substrate to cleave the O–P bond between the glycerol backbone and the phosphorus atom as shown (Figures 8.33 and 8.34). This yields a diacyl glycerol product as one fragment and the inositol 1,4,5-triphosphate IP_3 as the other product. Both the diacylglycerol and IP_3 are active regulatory metabolites, as is the substrate PIP_2. Thus, phospholipase C controls the availability of three regulatory molecules, the anionic lipid substrate PIP_2, the neutral lipid product diacyl glycerol and the water soluble, penta-anionic sugar triphosphate IP_3 (Figures 8.33 and 8.34).

Not surprisingly, phospholipase C activity is highly regulated.[56,58,59] There are six mammalian isoforms (β, γ, δ, ε, ζ, and η), encompassing 13 multidomain variants. The catalytic domains have essential similarities, but the regulatory domains are distinctly modulated in subtypes by different activators including GTPase subunits, calcium ions, and protein tyrosine kinases.[60] The diacylglycerol products, varying in acyl chain composition at either carbon 1 or carbon 2 of the glycerol backbone, stimulate a variety of enzyme activities. They also undergo kinase-mediated phosphoryl transfers to phosphatidic acids. The phosphatidic acids are also direct products from phospholipase D action and also have cellular regulatory effects.[56]

Concordantly, the IP_3 product, released by PLC turnover, gates a Ca^{++} channel protein in endoplasmic reticulum membranes.[61] The controlled, transient, and reversible flow of calcium ions from ER to cytoplasm also controls a variety of calcium-activated enzymes (such as kinases and the protein phosphatase calcineurin). Clearly, the multiple isoforms of mammalian phospholipase C isozymes sit at the intersection of many metabolic signaling pathways. They act at membrane–aqueous interfaces, can be activated up to 100-fold by specific modulatory ligands and control directly or indirectly at least four regulatory ligands (PIP_2, IP_3, diacylglycerols, and phosphatidic acids[62]).

Phospholipase C family members use the familiar two-cation motifs in their active sites to collect substrates and activate attacking nucleophiles.[63] In the hydrolysis of the preferred PIP_2-4,5-bisphosphate, the catalytic nucleophile in the first step is not water but the adjacent 2-OH of the inositol ring. As the diacylglycerol product departs the active site (Figures 8.33 and 8.34), a 1,2 cyclic phosphodiester of IP_3 is generated. This adduct is attacked by a water molecule (activated by one of the two zinc ions in the active site) before release such that the dissociating product is acyclic IP_3.

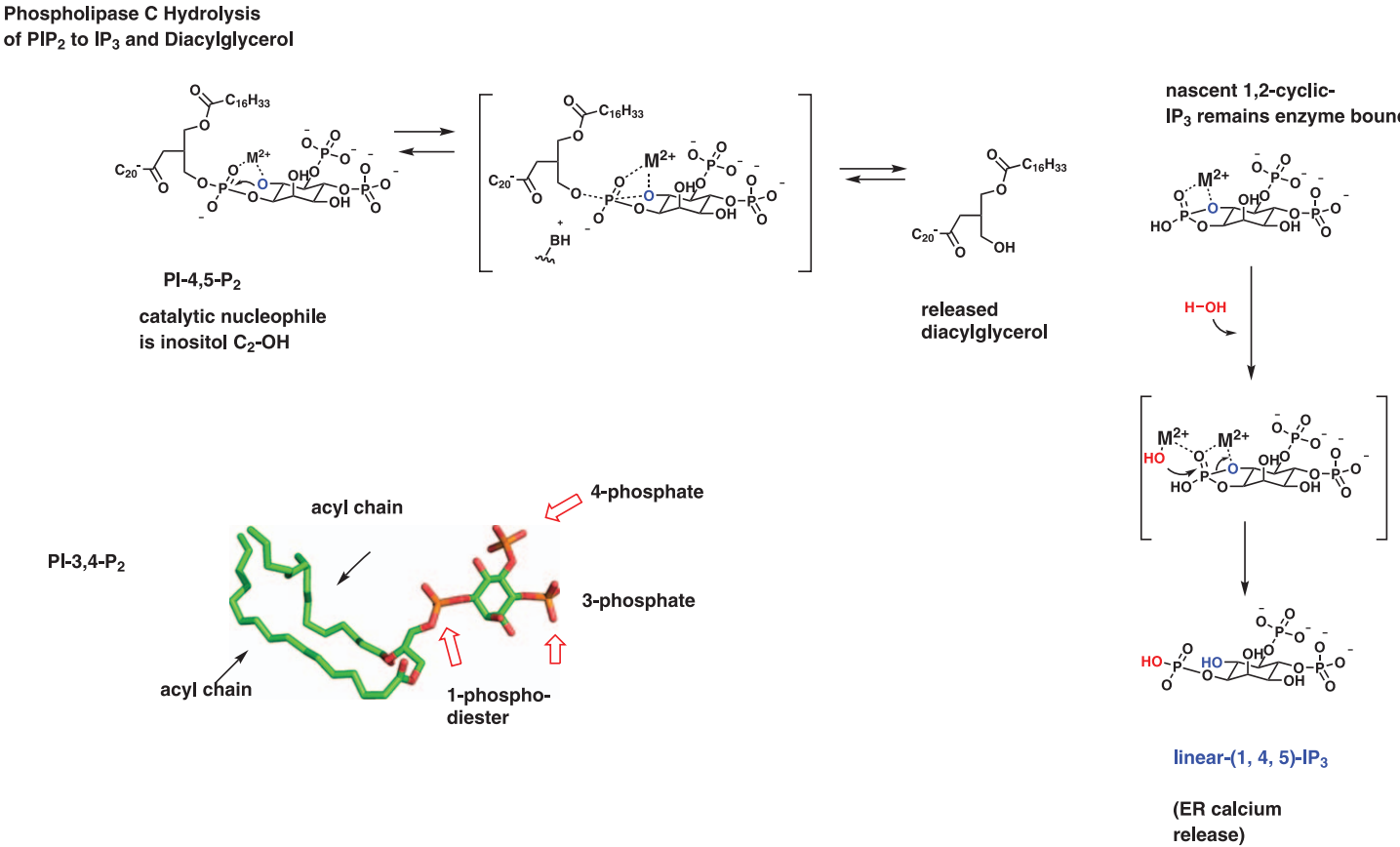


Figure 8.33 Phosphatidylinositols are preferred substrates for phospholipase C as are their phosphorylated derivatives phosphatidylinositol-4-phosphate, phosphatidylinositol-5-phosphate, and phosphatidylinositol-4,5-bisphosphate. PI cleavage by phospholipase C yields inositol-1-phosphate whereas PIP_2 cleavage yields IP_3 (inositol 1,4,5-triphosphate), a calcium signaling molecule.

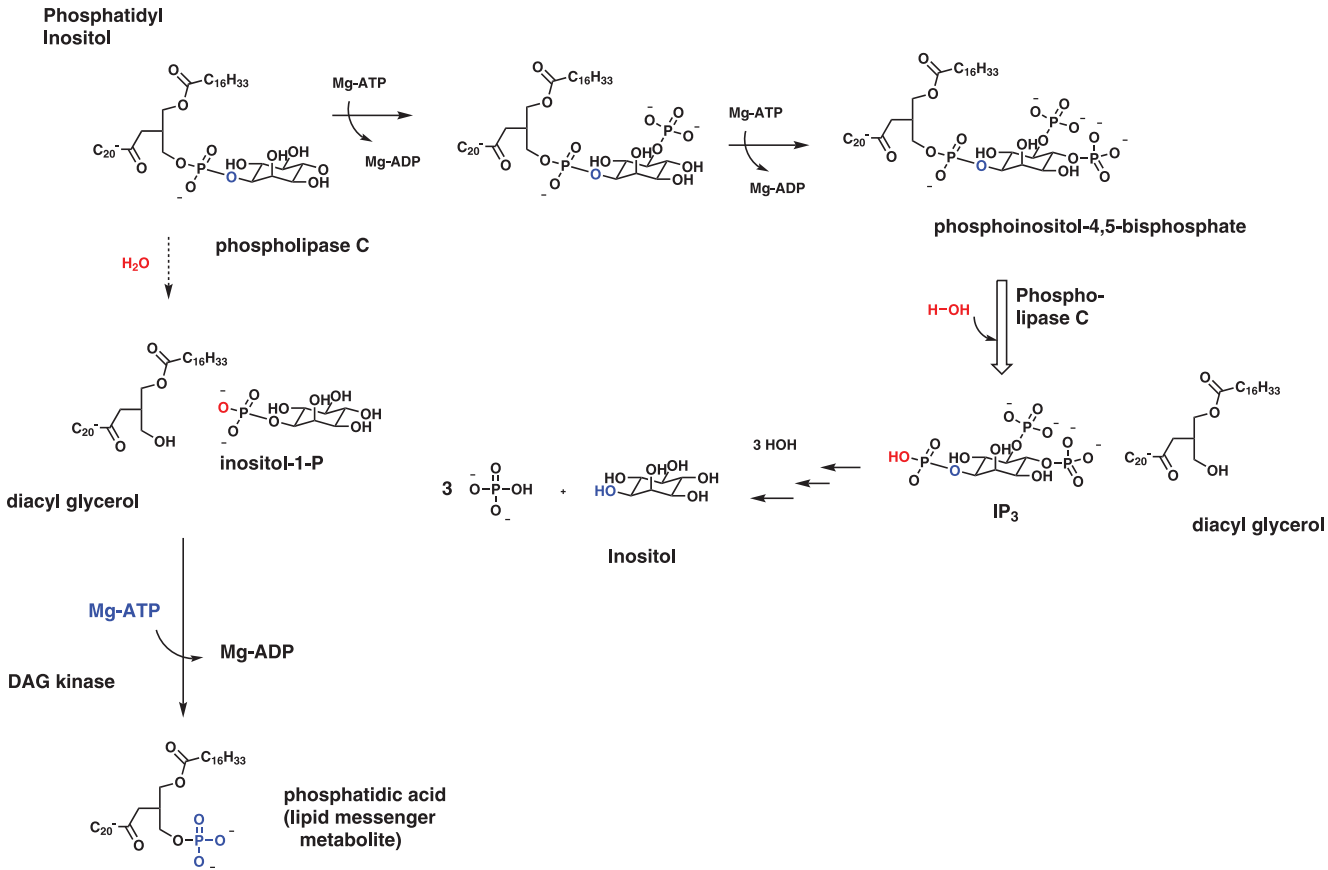


Figure 8.34 Reaction mechanism for PIP_2 cleavage by phospholipase C involves initial attack by the C2–OH of the inositol ring on the 1-phosphodiester phosphorus atom. As the diacylglycerol product is released, a 1,2-cyclic phospho-inositol-4,5-diphosphate intermediate is formed. This is the species hydrolyzed by water in the next step to release IP3.

Phospholipase C enzymes are robust catalysts. For the above transformation, k_{cat} values can approach 5000 s^{-1}. A catalytic cycle is complete in 200 microseconds.

8.10.2 Phospholipase D

The fourth class of phospholipases acting on glycerol-based phospholipid scaffolds (Figure 8.32) includes phospholipase D family members.[56] Their preferred phospholipid substrates are phosphatidylcholine molecules. As depicted in Figure 8.32 (route 4), they cleave the other P–O ester bond from that attacked by phospholipase C members. For example, a phosphatidylcholine substrate is regiospecifically hydrolyzed by phospholipase D action to choline and the partner phosphatidic aid. The choline metabolite is not active at the membrane but the dianionic phosphatidic acids can induce membrane curvature that may be preamble to vesicularization and membrane budding.

Mammalian cells express two isoforms, PLAD1 (mostly in internal compartment membranes) and PLAD2 (mostly in plasma membranes), that enable aspects of membrane trafficking, and cell migration, as well as endocytic and exocytic vesicle formation.[64,65]

In contrast to the 1,2-cyclic phosphodiester intermediate from PIP_2 during phospholipase C catalytic cycles, phospholipase D enzymes use a side chain imidazole ring of an active site histidine as primary nucleophile[66] (Figure 8.35). This covalent substrate–enzyme adduct strategy is analogous to the covalent histidine mechanism of the phosphomonoesterase glucose-6-phosphatase noted in Chapter 7. The imidazole nitrogen attacks the electrophilic phosphorus atom of the phosphodiester substrate to form the pentacovalent adduct in the transition state. Expulsion of the ROH alcohol product (in this instance free choline) leads to a covalent phosphatidic-histidine–enzyme adduct. The second half reaction involves activation of the water molecule that acts as a nucleophile to decompose the enzyme–phosphatidic acid adduct to give free phosphatidic acid and regenerate the enzyme active site histidine for subsequent catalytic cycles (Figure 8.35).

There is a distinction between the two covalent phosphoryl-histidinyl enzyme adducts formed in mid catalytic cycle of glucose-6-phosphatase *vs.* phospholipase D. From the phosphomonoester glucose-6-phosphate, the covalent enzyme has an inorganic phosphoryl group on the imidazole nitrogen (simple phosphoramidate). In contrast, the starting phosphodiester substrate (a phosphatidylcholine) in a phospholipase D cycle yields a lipid soluble phosphatidyl phosphoramidate as covalent adduct. In turn, the hydrolytic second half reactions yield inorganic phosphate from glucose-6-phosphatase, but the phosphomonoester grouping in the phosphatidic acid coproduct. Phospholipase D isozymes are much more hydrophobic to handle the complex long chain membrane-embedded phosphatidyl substrates.

This chapter so far has been a brief and selective look at only three groups of phosphodiesterases in three distinct areas of metabolism: nucleic acids,

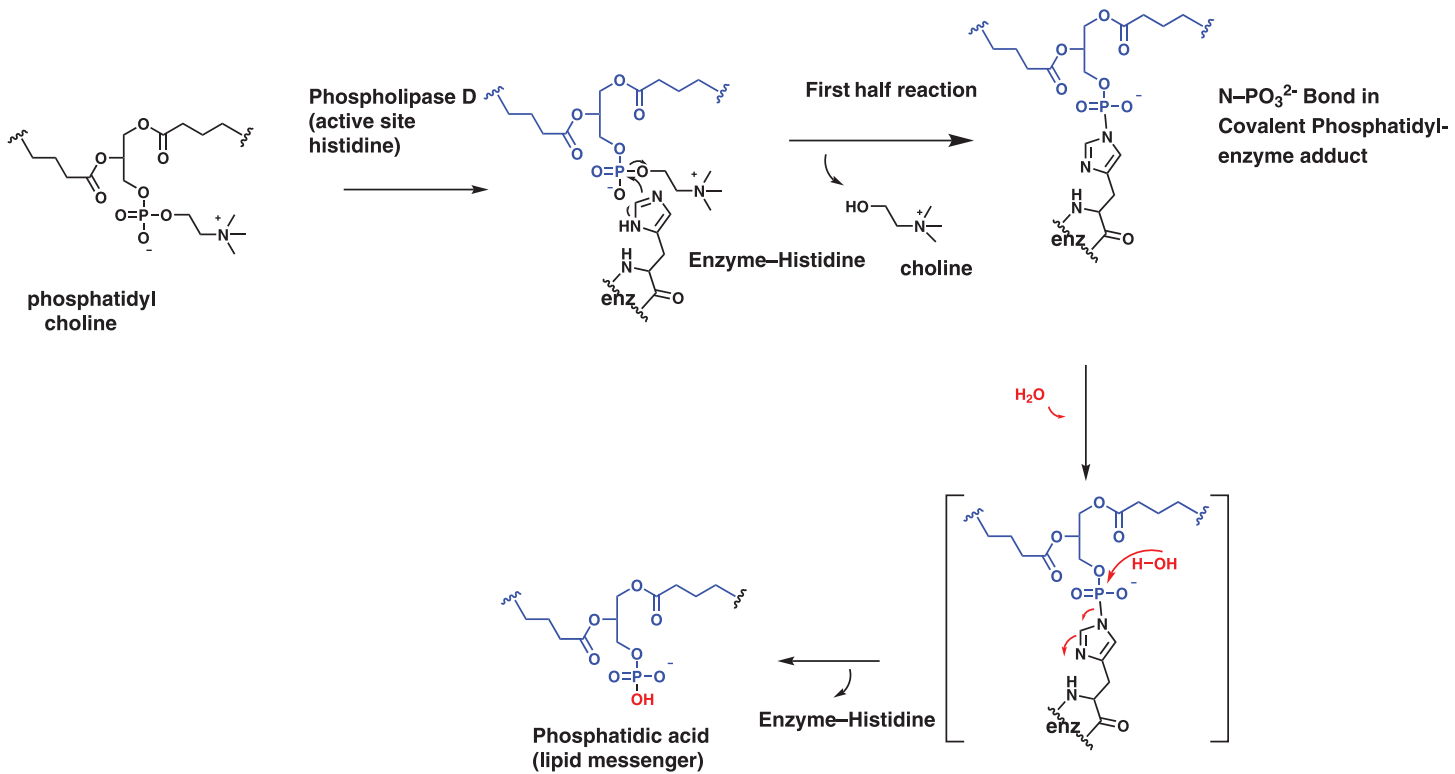


Figure 8.35 Phospholipase D catalysis of phospholipid hydrolysis proceeds by covalent catalysis. A phosphatidylcholine substrate is attacked on phosphorus by the imidazole side chain of an active site histidine residue. This yields an *N*-phosphatidyl–enzyme intermediate and released choline. In the second step a water molecule becomes a kinetically competent nucleophile to release the phosphatidic acid form of the phospholipid.

cyclic nucleotide second messengers, and phospholipids. Yet, they suffice to emphasize the broad reach of nucleotidyl transferases in formation and their opposing or balancing hydrolases in enzymatic cleavage of distinct phosphodiester scaffolds. The chemical biology of phosphodiesters permeates many of the key events in biological information transfer, the strategy of low molecular weight second messengers, the dynamics of membrane phospholipid composition, and control of lipid signaling and regulatory molecules.

8.11 Phosphotriesters and Phosphotriesterases

There is a sharp contrast in the abundance of phosphate monoesters and phosphodiesters, and the enzymatic machinery for their construction and hydrolyses in phosphorus chemical biology *versus* the *absence of phosphotriester* functions in normal metabolites.

Harking back to Chapter 1 and the pK_as of phosphoric acid, the first two pK_as (p$K_{a1} \sim 2$, p$K_{a2} \sim 6$) ensure dissolved inorganic phosphate at physiological pH of ~7.2 is a mixture of monoanions and dianions. The third pK_a of 12 (50% ionized at pH 12) is so far removed from the biological pH range that the phosphate trianion is not available. Thus, while phosphate oxyanions can function as nucleophiles for monoesterification and diesterification with alcohol partners, the third oxygen of inorganic phosphate does not normally participate in enzymatic esterifications.

There are two exceptions where organisms are presented with triesters and have adapted machinery to their removal.

8.11.1 DNA Alkylation at the Internucleotide Phosphodiester Bond

The first enumeration of phosphotriesters in biologic systems centers is on DNA. *O*-Alkylation of DNA internucleotide phosphodiesters can occur, typically on exposure to chemical mutagens such as diethyl sulfate, ethyl diazonium ions, or *N*-alkylnitrosoureas (Figure 8.36). While the exocyclic amino groups of adenine and guanine bases are kinetically most reactive, alkylation of the oxyanion of the internucleotide phosphodiester bond occurs and both R_P and S_P adduct diastereomers are known.[4]

In vitro treatment of double stranded DNA with methylmethane sulfonate gives about 5% triester adducts while ethylmethane sulfonate gives 16% adducts presumed to be phosphotriesters (Figure 8.36). The triester linkages are some 10^9-fold more prone to hydrolysis than the unalkylated, native phosphodiester internucleotide link,[4] but are typically still stable under normal *in vivo* conditions for DNA to be detectable in cells. While bacteria can repair phosphotriester adducts with the O^6-methyl-G-DNA methyltransferase, there are no identified enzymatic routes to repair the triester adducts in eukaryotes. The biological consequences of internucleotide alkylation to phosphotriesters remains largely unexplored.[4]

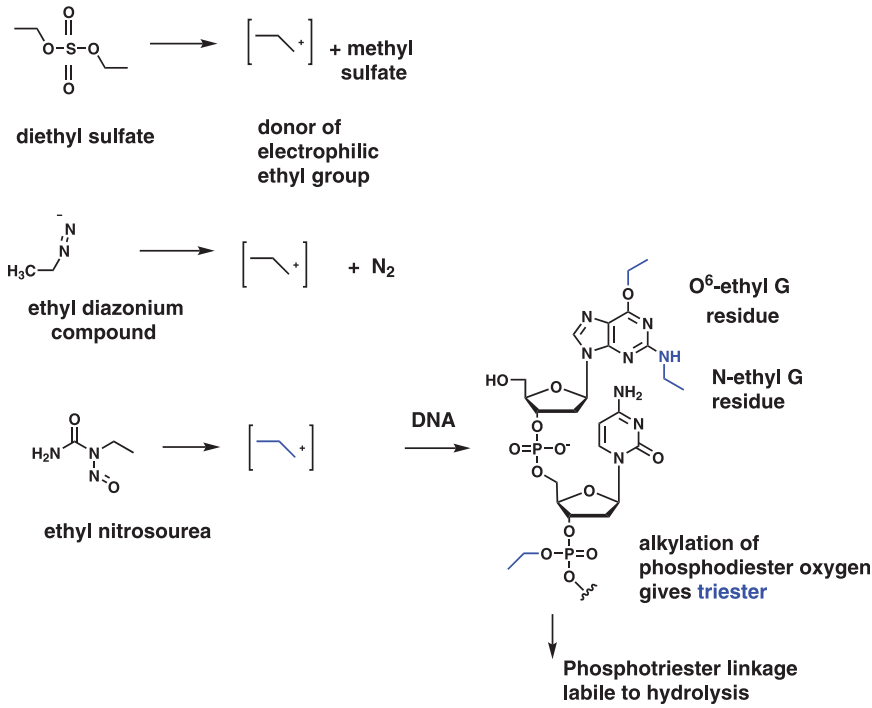


Figure 8.36 Some chemical mutagens are alkylating agents for DNA. While exocyclic amines and oxyanions are major targets, the oxyanion of the internucleotide phosphodiester bonds in DNA are also subject to *O*-alkylation as shown for ethyl nitrosourea.

8.11.2 Xenobiotic Phosphotriesters

Synthetic molecules with phosphotriester linkages are among the categories of both insecticides for agricultural purposes and nerve gases from chemical warfare programs.[67] (Figure 8.37) The convergent biological target – insect or human – is the enzyme acetylcholine esterase. Acetylcholine is one of several neurotransmitters that function at synapses, including neuromuscular junctions. Hydrolysis of the oxoester linkage to choline and acetate by the acetylcholine esterases deactivate the neurotransmitter and prevent excess nerve stimulation (Figure 8.37). Inhibitors of the enzyme lead to prolonged lifetimes and high synaptic concentrations of the neurotransmitter, causing muscular paralysis and, on prolonged exposure, death.

The phosphate-containing insecticides and nerve gases come in two classic chemical forms. They can have one or more carbon–phosphorus bonds. These are phosphonates. We will see natural examples of phosphonate metabolites in Chapter 11. Alternatively, the toxins or nerve gas structures can be triesters of inorganic phosphate. The three alcohols can all be different which results in a chiral phosphorus triester (R_P or S_p as shown in

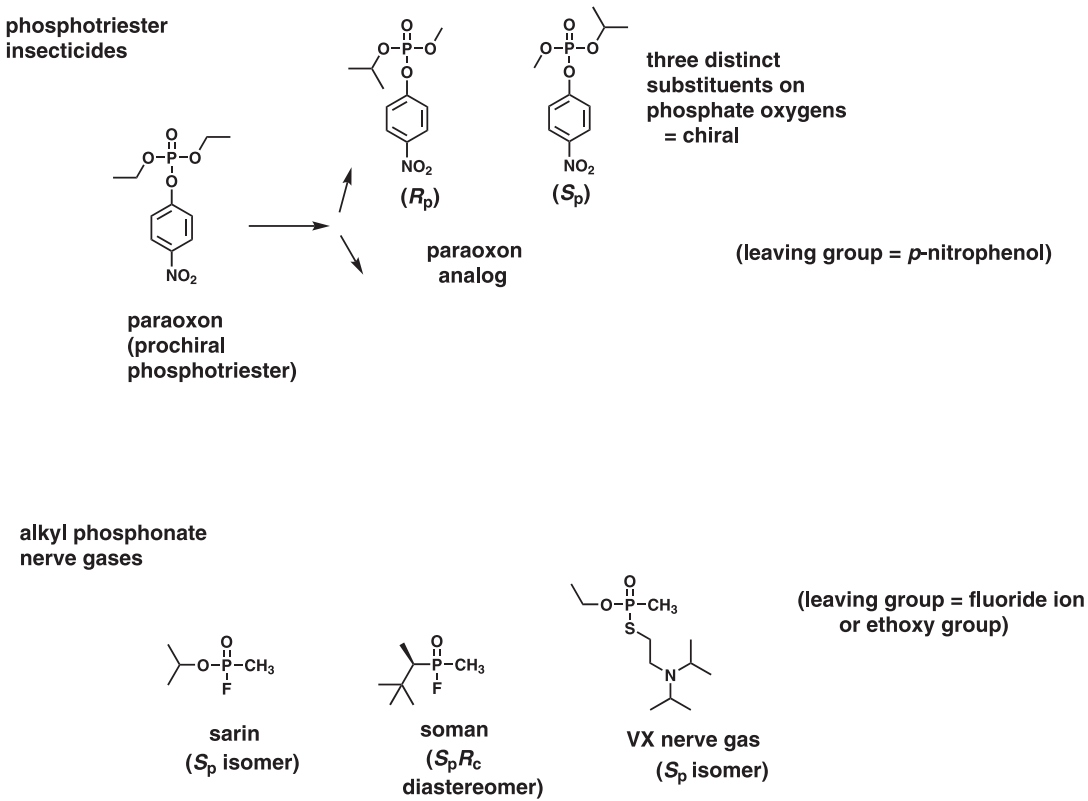


Figure 8.37 Xenobiotic phosphotriesters have been widely used for decades as insecticides. Note that when three distinct alcohol substituents are on phosphorus, two diastereomers, the S_P and R_P isomers, exist and can be processed differently by biological systems and enzyme targets. Many nerve gases are also centered on phosphorus chemistry but tend to differ from the agricultural phosphotriester scaffolds in having one or more direct C–P bonds and also a P–F bond to allow the nerve gas to act as a phosphorylating agent.

Figure 8.37). The nerve gases almost always have one noncleavable C–P phosphonate bond and a fluorine in a P–F bond. That fluoride ion is typically the low energy leaving group as a phosphoryl enzyme adduct forms in the acetylcholine esterase active site (Figure 8.38).

Both categories, insecticides and nerve gases, inhibit acetylcholine esterase by forming covalent phosphoryl enzyme intermediates as analogs of the normal covalent acetyl–enzyme intermediate formed in catalytic hydrolysis cycles[68] (Figure 8.38). The bulky phosphoryl groups are prone to long lifetimes and rates of hydrolyses are many orders of magnitude slower than the normal acyl enzyme. This long lifetime of the phosphorylated forms of acetylcholine esterase are the molecular basis of insecticide and nerve gas action.

The phosphotriester inhibitors typically have two alkyl and one aryl substituent as the triesters (Figure 8.38). The aryl alcohol is the lower energy leaving group (lower pK_a of aryl–OH *vs.* alkyl–OH) and is the alcohol displaced as acetylcholinesterase carries out an addition–elimination reaction that leads to the long-lived covalent phosphoryl enzymes.

The phosphotriester subclass of organophosphate insecticides and nerve gases are substrates for hydrolysis by phosphotriesterases, catalysts that detoxify these problematic xenobiotic scaffolds.[69,70] The leaving group ability carries over to the phosphotriesterase mechanisms, such that hydrolysis of the aryl–O–P bond is most rapid in producing the dialkyl phosphodiester products (Figure 8.39A). The phosphotriesterase-mediated hydrolysis of the insecticide paraoxon is shown in Figure 8.39, yielding the *para*-nitrophenolate anion and diethylphosphate.

Bacterial phosphotriesterases may have evolved from bacterial lactonases to deal with ambient levels of phosphotriesters and organophosphate insecticides in soil.[70] The first phosphotriesterase isolated from *Pseudomonas diminuta* came from an organophosphate-contaminated soil.[71] Phosphotriesterases have different folds and probably distinct evolutionary histories as they converge to acquire this common xenobiotic hydrolase activity.[69] All such enzymes yet characterized are two metal-dependent in a binuclear metal center (most commonly a Zn^{2+} pair, Figure 8.39), a common theme for phosphate coordination and water activation, seen in both phosphomonoesterases and phosphodiesterases in Chapter 7 and earlier in this chapter. With chiral phosphotriester substrates the Pseudomonas enzyme shows a preference for R_P isomers but will also less efficiently hydrolyze the S_P isomers of phosphotriesters.

Some phosphotriester insecticides have one of the phosphate oxygens replaced by a stable bond to sulfur. For example, parathion, the precursor metabolite to paraoxon is in this category. In eukaryotes the sulfur atom gets oxygenated enzymatically by a liver cytochrome P450 and rearranges to the indicted three-membered ring shown in Figure 8.39B before extruding sulfur and creating paraoxon as substrate for phosphotriesterase cleavage.

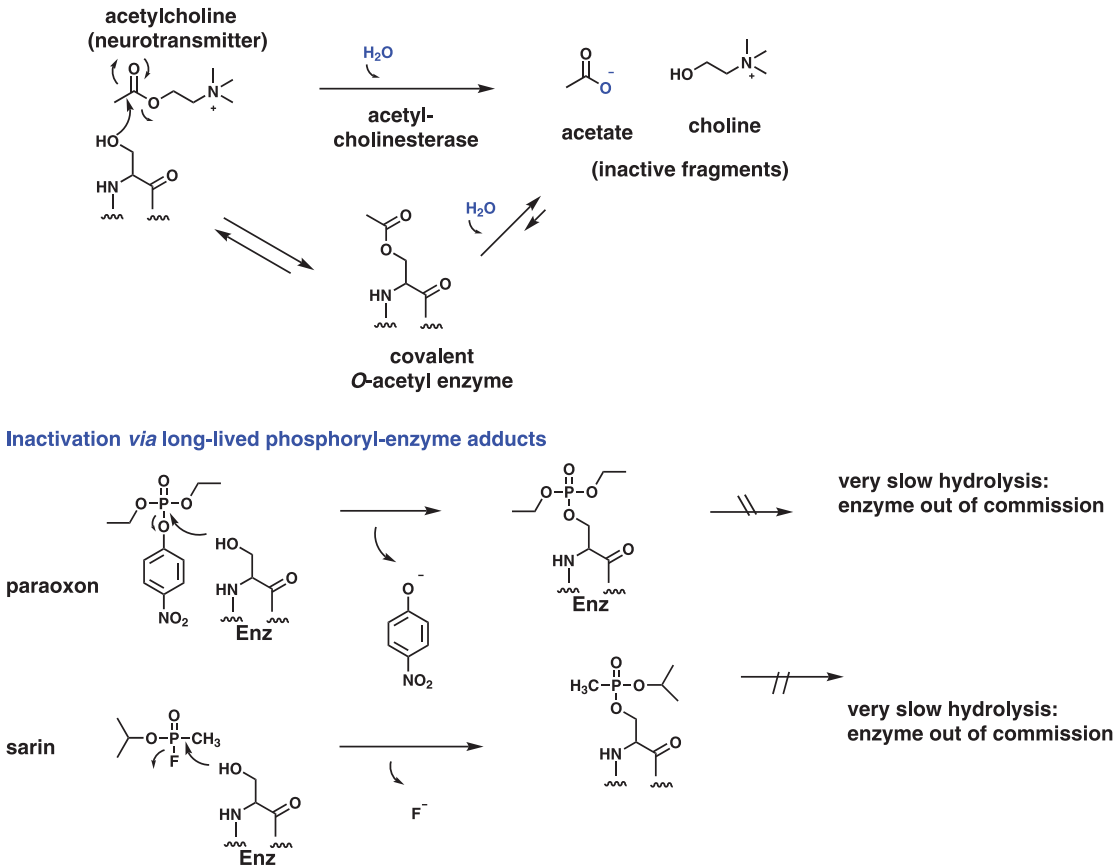


Figure 8.38 Acetylcholine esterase (insect or human) is the target for phosphotriester insecticides or phosphonyl fluoride nerve gases, respectively. Both types of phosphoryl scaffolds have a low energy leaving group (*e.g.* a phenol in the phosphotriesters, a fluoride ion in sarin) to allow facile covalent phosphorylation of the active site serine, the normal catalytic nucleophile in acetylcholinesterase. Unlike the normal covalent acetyl enzyme intermediate, which is hydrolyzed rapidly, the phosphoryl enzymes are long-lived, shut down acetylcholinesterase activity for hours to days, and promote excessively high levels of the neurotransmitter acetylcholine.

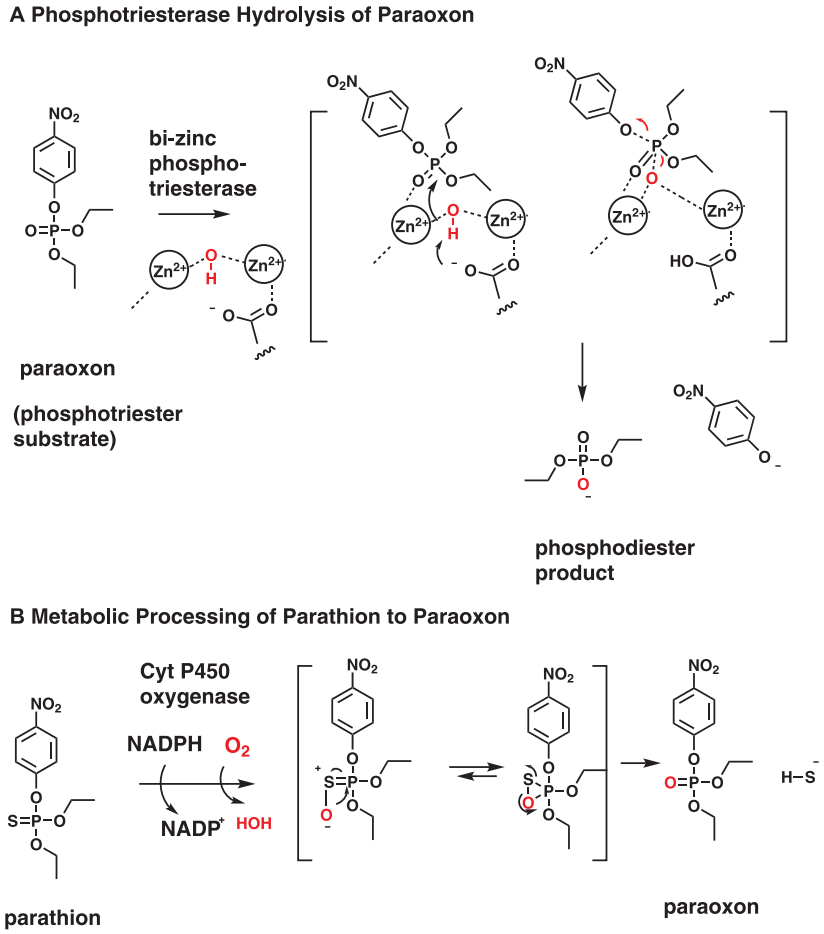


Figure 8.39 (A) Bacterial phosphotriesterases use a variant of the two-metal zinc cation hydrolytic mechanism used by phosphatases and phosphodiesterases. Hydrolysis of the insecticide paraoxon proceeds to release *para*-nitrophenol and the dialkyl phosphodiester product. (B) Parathion is also an insecticide. In mammalian hepatocytes it can be desulfurized to paraoxon, perhaps by the proposed *S*-oxygenation route depicted.

References

1. M. W. Bowler, M. J. Cliff, J. P. Waltho and G. M. Blackburn, Why did nature select phosphate for its dominant roles in biology, *New J. Chem.*, 2010, **34**, 784–794.
2. G. K. Schroeder, C. Lad, P. Wyman, N. H. Williams and R. Wolfenden, The time required for water attack at phosphorus, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 4052–4055.

3. W. Yang, Nucleases: diversity of structure, function and mechanism, *Q. Rev. Biophys.*, 2011, **44**, 1–93.
4. K. S. Gates, An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals, *Chem. Res. Toxicol.*, 2009, **22**, 1747–1760.
5. U. Shevelev IV, Hubscher, The 3′ 5′ exonucleases, *Nat. Rev. Mol. Cell Biol.*, 2002, **5**, 364–376.
6. S. T. Lovett, The DNA exonucleases of Eschericia coli, *EcoSal Plus*, 2011, **4**, DOI: 10.1128/ecosalplus.4.4.7.
7. A. Pingoud and A. Jeltsch, Structure and function of type II restriction endonucleases, *Nucleic Acids Res.*, 2001, **29**, 3705–3727.
8. V. Filippov, V. Solevyev, M. Filippova and S. S. Gill, A novel type of RNAse III family proteins in eukaryotes, *Genes*, 2000, **245**, 213–221.
9. L. Jaskiewicz and W. Fillipowicz, Role of Dicer in posttranscriptional RNA silencing, *Curr. Top. Microbiol. Immunol.*, 2008, **320**, 77–97.
10. J. Hock and G. Meister, The argonaute protein family, *Genome Biol.*, 2008, **9**, 210.
11. T. J. Wilson and D. M. Lilly, RNA catalysis- is that it?, *RNA*, 2015, **21**, 534–537.
12. G. J. Narlikar and D. Herschlag, Mechanistic aspects of enzymatic catalysis: lessons from comparison of RNA and protein enzymes, *Annu. Rev. Biochem.*, 1997, **66**, 19–59.
13. D. Herschlag, Learning from ribozymes, *RNA*, 2015, **21**, 527–528.
14. T. R. Cech, D. Herschlag, J. A. Piccirilli and A. M. Pyle, RNA catalysis by a group 1 ribozyme, *J. Biol. Chem.*, 1992, **267**, 17479–17482.
15. N. K. Tanner, Ribozymes: the characteristics and properties of catalytic RNAs, *FEMS Microbiol. Rev.*, 1999, **23**, 257–275.
16. A. M. Pyle, Group II inron self-splicing, *Annu. Rev. Biophys.*, 2016, **45**, 183–205.
17. M. J. Fedor and J. Williamson, The catalytic diversity of RNAs, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 399–412.
18. M. R. Stahley and S. Strobel, Structural evidence for a two metal ion mechanism of group I intron splicing, *Science*, 2005, **309**, 1587–1590.
19. D. M. Lilley, Mechanisms of RNA catalysis, *Philos. Trans. R. Soc., B*, 2011, **366**, 2910–2917.
20. M. J. Howard, B. P. Klemm and C. A. Fierke, Mechanistic studies reveal similar catalytic strategies for phosphodiester bond hydrolysis by protein-only and RNA-dependent RNAse P, *J. Biol. Chem.*, 2015, **290**, 13454–13464.
21. W. C. Winkler and R. R. Breaker, Regulation of bacterial gene expression by riboswitches, *Annu. Rev. Microbiol.*, 2005, **59**, 487–517.
22. W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins and R. R. Breaker, Control of gene expression by a natural metabolite-responsive ribozyme, *Nature*, 2004, **428**, 281–286.
23. D. J. Klein and A. R. Ferre-D'Amare, Structural basis of glmS ribozyme activation by glucosamine-6-P, *Science*, 2006, **313**, 1745–1747.

24. J. C. Cochran, S. A. Lipchock and S. A. Strobel, Structural investigation of the GlmS ribozyme bound to its catalytic cofactor, *Chem. Biol.*, 2007, **14**, 97–105.
25. S. Yehudai-Resheff, M. Hirsch and G. Schuster, Polynucleotide phosphorylase functions as both an exonuclease and a poly (A) polymerase in spinach chloroplasts, *Mol. Cell. Biol.*, 2001, **21**, 5408–5416.
26. B. K. Mohanty and S. R. Kushner, Polynucleotide phosphorylase functions both as a 3′ to 5′ exonuclease and a poly (A) polymerase in Eschericicai coli, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 11966–11971.
27. J. Wu, J. Xiao, Z. Zhang, X. Wang, S. Hu and J. Yu, Ribogenomics: the science and knowledge of RNA, *Genomics, Proteomics Bioinf.*, 2014, **12**, 57–63.
28. R. J. Roberts, How restriction enymes became the workhorses of molecular biology, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 5905–5908.
29. J. A. McClarin, C. A. Frederick, B. C. Wang, P. Greene, H. W. Boyer, J. Grabl and J. M. Rosenberg, Structure of the DNA-Eco R1 endonuclease recognition complex at 3 A resolution, *Science*, 1986, **234**, 1526–1541.
30. F. J. Mojica, C. Diez-Villasenor, J. Garcia-Martinez and E. Soria, Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements, *J. Mol. Evol.*, 2005, **60**, 174–182.
31. K. S. Makarova, Y. I. Wolf, O. S. Alkhnbashi, F. Costa, S. A. Shah, S. J. Saunders, R. Barrangu, S. J. Brouns, E. Charpentier, D. H. Haft, P. Horvath, S. Moineau, F. J. M. Mojica, R. M. Terns, M. P. Terns, M. F. Whie, A. F. Yakunin, R. A. Garrett, J. van der Oost, R. Backofen and E. V. Koonin, An updated evolutionary classification of CRISPR-CAS systems, *Nat. Rev. Microbiol.*, 2015, **13**, 722–736.
32. A. Pickar-Oliver and C. A. Gersbach, The next generation of CRISPR-Cas technologies and applications, *Nat. Rev. Mol. Cell Biol.*, 2019, **20**, 490–507.
33. M. Adli, The CRISPR tool kit for genome editing and beyond, *Nat. Commun.*, 2018, **9**, 1–13.
34. F. Jiang and J. A. Doudna, CRISPR-Cas9 structure and mechanisms, *Annu. Rev. Biophys.*, 2017, **46**, 509–529.
35. New England BioLabs website (downloaded 11-04-2019) CRISPR/Cas9 & targeted genome editing: new era in molecular biology.
36. M. Jinek, K. Chylinski, I. Fonfara, Hauer, J. A. Doudna and E. Charpentier, A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity, *Science*, 2012, **337**, 816–821.
37. L. F. Ribeiro, L. F. C. Riberio, M. Q. Barreto and R. J. Ward, Protein engineering strategies to expand CRISPR/Cas9strategies, *Int. J. Genomics*, 2018, 1652567.
38. O. Niewoehner, C. Garcia-Doval, J. T. Rostol, C. Berk, F. Schwede, L. Bigler, J. Hall, Marraffini and M. Jinek, Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers, *Nature*, 2017, **548**, 543–548.

39. C. Rouillon, J. S. Athukoralage, S. Graham, S. Gruscgow and M. F. White, Control of Oligoadenylate synthesis in a type III CRISR system, *eLife*, 2018, **7**, e36734.
40. J. J. Champoux, DNA topoisomerases:structure, function, and mechanism, *Annu. Rev. Biochem.*, 2001, **70**, 369–413.
41. J. K. Lassila, J. G. Zalatan and D. Herschlag, Biological phosphoryltransfer reactions: understanding mechanism and catalysis, *Annu. Rev. Biochem.*, 2011, **80**, 669–792.
42. A. T. Bender and J. A. Beavo, Cyclic nucleotide phosphodiesterases: molecular regulation to clinical use, *Pharmacol. Rev.*, 2006, **58**, 488–520.
43. N. Manatage, Revisiting bacterial cyclic nucleotide phosphodiesterases: Cyclic AMP hydrolysis and beyond, *FEMS Microbiol. Lett.*, 2015, **362**, fnv183.
44. V. K. Raker, C. Becker and K. Steinbrink, The cAMP pathway as therapeutic target in autommune and inflammatory diseases, *Front. Immunol.*, 2016, **7**, 123.
45. A. Stragherlin and M. Zaccolo, Phosphodiesterases and subcellular compartmentalized cAMP, *Am. J. Physiol.*, 2012, **302**, H379–390.
46. D. H. Maurice, H. Ke, F. Ahmad, Y. Wang, J. Chung and V. C. Manganiello, Advances in targeting cyclic nucleotide phosphodiesterases, *Nat. Rev. Drug Discovery*, 2014, **13**, 290–314.
47. P. Ross, H. Weinhouse, Y. Aloni, D. Michaeli, P. Weinberger-Ohana, R. Mayer, E. de Vroom, G. A. van der Marel, J. H. van Boom and M. Benziman, Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid, *Nature*, 1987, **325**, 279–281.
48. L. Chao, S. Rakshe, M. Leff and A. M. Spormann, PdeB, a cyclic di-GMP-specific phosphodiesterase that regulate *Sawenella oneidensis MR-1* motility and biofilm formation, *J. Bacteriol.*, 2013, **195**, 3827–3833.
49. O. Sarenko, G. Klauck, F. M. Wilke, V. Pfiffer, A. M. Richter, S. Herbst, V. Kaever and R. Hengge, More than enzymes that make or break cyclic di-GMP-local interactions in the interactome of GGDEF/EAL domain proteins of E. coli, *mBio*, 2017, **8**, e01639–01617.
50. F. Rao, Y. Yang, Y. Qi and Z.-X. Liang, Catalytic mechanism of cyclic di-GMP specific phosphodiesterase: a study of the EAL domain-containing RocR from Pseudomonas aeruginosa, *J. Bacteriol.*, 2008, **190**, 3622–3631.
51. J. Gao, J. Tao, W. Liang, M. Zhao, X. Du, H. Duan, B. Kan, X. Su and Z. Jinag, Identification and characterization of phosphodiesterases that specifically degrade 3′,3′-cyclic GMP-AMP, *Cell Res.*, 2015, **25**, 539–550.
52. X. D. Li, J. Wu, D. Gao, H. Wang, L. Sun and Z. J. Chen, Pivotal roles of cGAS-cGAMP sgnaling in antiviral defense and immune adjuvant effects, *Science*, 2013, **341**, 1390–1394.
53. J. Wu, L. Sub, X. Chen, F. Du, H. Shi, C. Chen and Z. J. Chen, Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA, *Science*, 2012, **339**, 826–830.

54. D. Gao, J. Wu, Y. T. Wu, F. Du, C. Aroh, N. Yan, L. Sun and Z. J. Chen, Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses, *Science*, 2013, **341**, 903–906.
55. L. Li, Q. Yin, P. Kuss, Z. Maliga, J. L. Milan, H. Wu and T. Mitchison, Hydrolysis of 2′,3′-cGMAP by ENPP1 and design of non-hydrolyzable analogs, *Nat. Chem. Biol.*, 2014, **10**, 1043–1048.
56. X. Wang, The role of phospholipase D in signaling cascades, *Plant Physiol.*, 1999, **120**, 645–651.
57. E. A. Dennis, Thematic Review Series on Phospholipases: central role in signaling and disease, *J. Lipid Res.*, 2009, **56**, 1245–1247.
58. S. G. Rhee and K. D. Choi, Regulation of inositol phospholipid-specific phospholipase C isozymes, *J. Biol. Chem.*, 1992, **267**, 12393–12396.
59. J. Mackenzie, L. S. Lynn and A. D. Hughes, Linking phospholipase C isoforms with differentiation function in human vascular smooth muscle cells, *Biochim. Biophys. Acta*, 2013, **1833**, 3006–3012.
60. S. B. Hansen, Lipid agonism: the PIP2 paradigm on ligand-gated ion channels, *Biochim. Biophys. Acta*, 2015, **1851**, 620–628.
61. M. J. Berridge, The inositol triphosphate/calcium signaling pathway in health and disease, *Physiol. Rev.*, 2016, **96**, 1261–1296.
62. M. H. Rebeecchi and S. N. Pentyala, Structure, function, and control of phosphoinositide-specific phospholipase C, *Physiol. Rev.*, 2000, **80**, 1291–1335.
63. M. V. Ellis, S. R. James, O. Perisic, C. P. Downes, R. L. Williams and M. Katan, Ctalaytic Domains of Phosphoinositide-specific phospholipase C; mutaional analysis of residues within the active site and hydrophobic ridge of PLC delta1, *J. Biol. Chem.*, 1998, **273**, 11650–11659.
64. K. E. Meier, T. C. Gibbs, S. M. Knoepp and K. M. Ella, Expression of phospholipase D isoforms in mammalian cellls, *Biochim. Biophys. Acta*, 1999, **1439**, 199–213.
65. R. C. Bruntz, C. W. Lindsley and H. A. Brown, Phospholipase D signaling pathways and potential therapeutic targets in cancer, *Pharmacol. Rev.*, 2014, **66**, 1033–1079.
66. E. B. Gottlin, A. E. Rudolph, Y. Zhao, H. R. Mattjews and J. E. Dixon, Catalytic mechanism of the phospholipase D superfamily proceeds via a covalent phosphohistidine intermediate, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 9202–9207.
67. F. M. Raushel, Bacterial detoxification of organophosphate nerve agents, *Curr. Opin. Microbiol.*, 2002, **5**, 288–295.
68. L. G. Costa, Organophosphorus compounds at 80: some old and new issues, *Toxicol. Sci*, 2018, **162**, 24–35.
69. A. N. Bigley and F. M. Raushel, Catalytic mechanisms of phosphotriesterases, *Biochim. Biophys. Acta*, 2013, **1834**, 443–453.
70. R. J. Russell, C. Scoot, C. J. Jackson, R. Pandey, G. Pandey, M. C. Taylor, C. W. Coppin, J.-W. Liu and J. G. Oakeshott, The evolution of new enzyme function: lessons from xenobiotic metabolizing bacteria versus insecticide-resistant insects, *Evol. Appl.*, 2011, **4**, 225–248.

71. C. Zhongli, L. Shumpeng and F. Guoping, Isolation of methyl-parathion-degrading strain M6 and cloning of the methyl parathion hydrolase gene, *Appl. Environ. Microbiol.*, 2001, **67**, 4922–4925.
72. A. Stangherlin and M. Zaccolo, *Am. J. Physiol.: Heart Circ. Physiol.*, 2012, **302**, H379–H390.

CHAPTER 9

Phosphorylases: Inorganic Phosphate as Oxygen Nucleophile

The enzyme class known as phosphorylases are contrasts and comparators to the phosphoryl transfer enzymes: both the kinases described in Chapter 3 and the phosphatases examined in Chapter 8. In kinase action the terminal phosphoryl group [$-PO_3^{2-}$] is transferred from ATP as a phosphorus electrophile to cosubstrate nucleophiles. In phosphatase action the phosphoryl group [$-PO_3^{2-}$] of a phosphomonoester substrate is likewise transferred as an electrophile at phosphorus to cosubstrate water activated to function as a nucleophile (Figure 9.1).

9.1 Phosphate Anion as Oxygen Nucleophile

Phosphorylases also generate phosphomonoesters as products, in analogy to the products from kinase action. However, Mg–ATP is *not* a substrate. Instead it is inorganic phosphate acting as oxyanion nucleophile. Compared with phosphatases where water attacks an electrophilic phosphoryl group, inorganic phosphate is the nucleophile, typically attacking some electrophilic carbon atom in a cosubstrate (Figure 9.1). The pK_a values of phosphoric acid mean that inorganic phosphate at physiological pH is a mix of monoanions and dianions of the peripheral oxygens. Those oxyanions are the nucleophiles in action of phosphorylases.

In the example of line 3 of Figure 9.1 phosphorolytic cleavage of cellulose at the terminal glucosyl unit converts the itinerant glucosyl C_1 oxocarbenium ion into glucose-α-1-phosphate. Suites of both hydrolases and phosphorylases

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

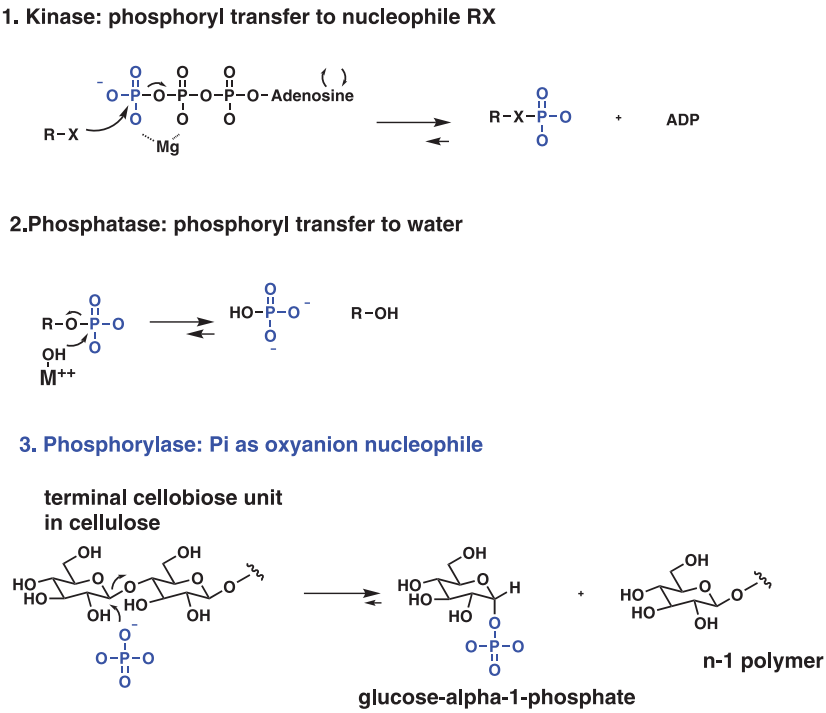


Figure 9.1 In the action of both kinases and phosphatases, the $-PO_3^{2-}$ group is transferred as an electrophile at phosphorus to cosubstrate nucleophiles. In phosphorylase reactions, all four oxygens of inorganic phosphate are incorporated into the product as an oxyanion of inorganic phosphate acts as a nucleophile on a cosubstrate electrophile.

are expressed in organisms that can degrade oligosaccharide fragments of cellulose.[1] All four oxygens in this phosphate ester originate in inorganic phosphate, reflecting its function as a nucleophile. The phosphorylase class of enzymes demonstrate and utilize this nucleophilic feature of inorganic phosphate in their chemical biology.

Phosphorylases need electrophilic carbon atoms in their cosubstrates to be attacked by inorganic phosphate. The C_1 acetal carbons of oligosaccharides and polysaccharides give rise to C1-oxocarbenium ion transition states, so in principle and in practice they are candidates. Furthermore, phosphorolysis instead of hydrolysis gives the sugar-1-phospates that are the activated species on the way to NDP–sugars, such as UDP–glucose. By using inorganic phosphate as nucleophile, organisms avoid having to spend an ATP to convert free glucose to glucose-1-phosphate or free mannose to mannose-1-phosphate. In this sense polysaccharide phosphorylases function as energy-saving nanomachinery.

As a sugar monomer unit is mobilized from a homopolymeric glucan such as starch or cellulose by enzymatic phosphorolysis, the resultant monomer

typically has the phosphate ester grouping at C1 of the sugar as the α-anomer, *e.g.* glucose-α-1-phosphate (Figure 9.1). This may be due to the immediate utility of the hexose-α-1-phosphate isomer as a usable metabolite. In contrast, hexose-β-1-phosphates, if formed, would be dead-end metabolites.

Among well studied phosphorylases in this class are the disaccharide phosphorylases trehalose phosphorylase and sucrose phosphorylase[2] (Figure 9.2). The classic polysaccharide phosphorylase is glycogen phosphorylase[3] (Figure 9.2). Together, these enzymes represent cleavage of distinct 1,1-, 1,2-, and 1,4 linked sugar residues (Figure 9.2). α-Glycosidases that act on the amylopectin component of starch must also be able to hydrolyze glucosyl-α-1,6-glucose linkages (not shown).

A number of glycoside phosphorylases have been identified[4] and include enzymes that retain the configuration at C_1 of the mobilizing sugar residue as well as those that invert the configuration. Maltose phosphorylase and glycogen phosphorylase retain the alpha configuration in the glucose-1-phosphate product found in the α-1,4-linked maltose disaccharide or glycogen starting polysaccharide. In contrast, the β-linkage in sophorose (β-1,2) proceeds with inversion at C1 to glucose-α-1-phosphate (Figures 9.3 and 9.4). The number of glycoside phosphorylases identified to date is around 30.[5] Many have equilibrium constants not far from unity and so represent enzymes that could be run backwards and used to make oligosaccharides of defined constitution and stereochemistry.[6]

The consensus mechanism for glycoside phosphorylases is a transition state with early C_1–O bond cleavage, generating a C1-carbocation.[5,7,8] This is resonance stabilized by delocalization into the adjacent oxygen such that the positive charge is spread over C1–O as an oxocarbenium ion (Figures 9.3 and 9.4). The retention *vs.* inversion outcome in configuration at C1 of the sugar 1-phosphates, is thought to reflect restricted access of Pi to the sugar oxocarbenium ion. Approach from the back as in sophorose phosphorylase leads to inversion from beta to alpha. In glycogen- and maltose phosphorylases, approach from the front leads to retention (alpha to alpha).

On the other hand, the sucrose phosphorylase from *Pseudomonas saccharophila* carries out hydrolysis by a two-step mechanism where the nucleophile in the first step is the beta carboxylate oxygen of an aspartate residue in the active site. This releases fructose and produces a glucosyl-aspartate oxoester covalent enzyme adduct.[9] In the second half reaction the enzyme uses inorganic phosphate to discharge the glucosyl enzyme intermediate. The results of stereochemical analysis indicated that the original glucosyl-α-1,2-link to fructose in the sucrose substrate was converted to a β-glucosyl enzyme intermediate[10] (Figure 9.2). Phosphorolysis by inorganic phosphate again proceeded with conversion to give the observed glucose-α-1-phosphate.

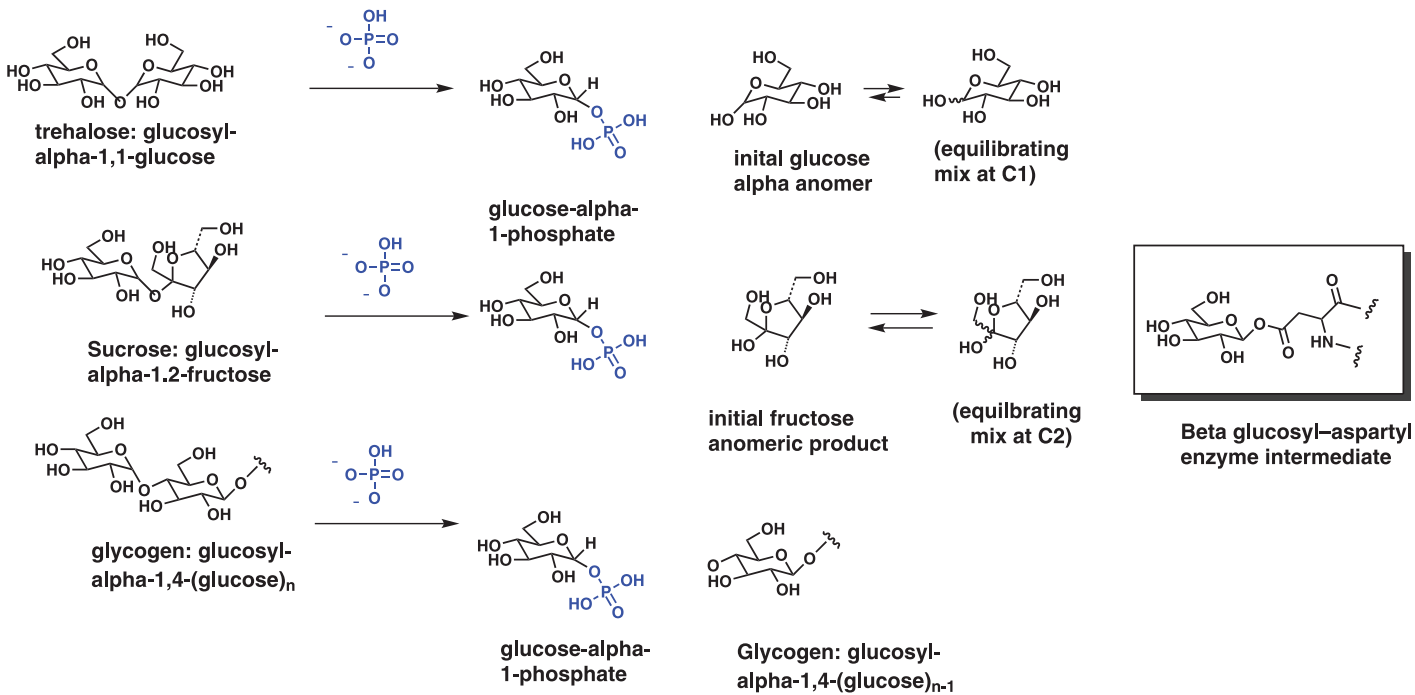


Figure 9.2 Three carbohydrate phosphorylases: trehalose phosphorylase, sucrose phosphorylase, and glycogen phosphorylase act on saccharide substrates of different regiochemistry.

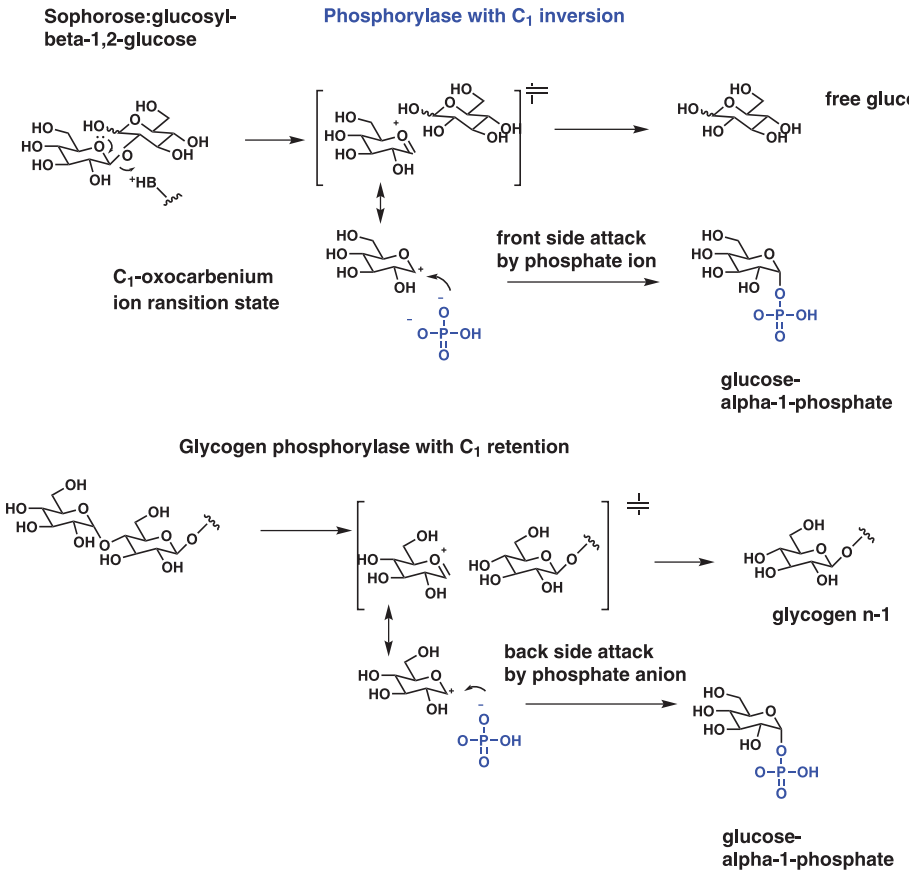


Figure 9.3 Distinct stereochemical outcomes for enzymatic glycosyl transfers to inorganic phosphate. Sophorose phosphorylase catalyzes inversion, glycogen phosphorylase catalyzes retention of stereochemistry at C_1 of the migrating glycosyl unit.

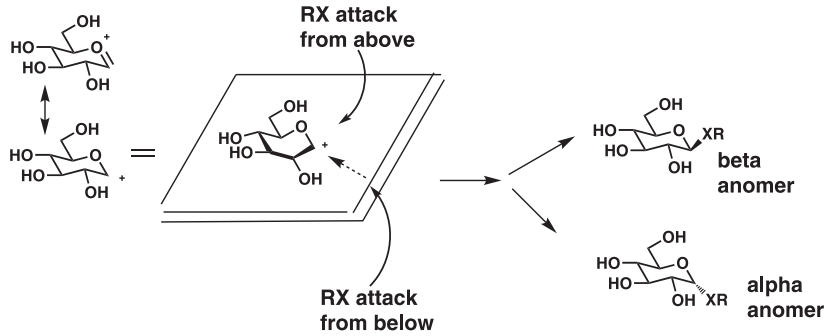


Figure 9.4 The oxocarbenium ion transition state can be attacked from above or below by cosubstrate nucleophiles.

9.2 Equilibrating Acyl Thioesters and Acyl Phosphates

A second class of electrophilic carbon centers in metabolites that can be captured by inorganic phosphate are acyl thioesters. The thermodynamics of acyl phosphates (mixed acyl phosphoric anhydrides) and acyl thioesters are close enough that these can be reversible transformations. One such enzyme is termed phosphotransacetylase,[11] a bacterial enzyme interconverting acetyl-CoA and acetyl-phosphate with $K_{eq.} \sim 1$. (Figure 9.5). The physiological role of this enzyme in bacteria such as *E. coli* is to function in the tandem enzyme sequence acetate kinase, phosphotransacetylase (Figure 9.5) to generate acetyl CoA. The kinase spends an ATP to activate the resonance stabilized carboxylate anion of acetate to the mixed acyl–phosphoric anhydride. The acetyl-phosphate is then harvested by the thiolate anion of coenzyme A to produce acetyl-*S*-CoA as the major circulating activated acetyl metabolite. The thermodynamic acetyl group transfer potentials in acetyl-phosphate and acetyl CoA are about equivalent, but the thioester linkage in acetyl CoA is kinetically much more stable in cells.

Analogously, in the middle of the citric acid cycle, succinyl phosphate is generated as a nascent product from succinyl CoA.[12] The reaction is phosphorolysis of a succinyl thioester (Figure 9.5). The phosphoryl group of succinyl phosphate is then attacked by one of the terminal phosphate oxygens of GDP to create GTP and succinate.

A central enzyme in glycolysis that also operates in the phosphorolysis direction is glyceraldehyde-3-phosphate dehydrogenase.[13] In the glyceraldehyde-3-phosphate dehydrogenase catalytic site, the covalent thiohemiacetal adduct between the active site cysteine thiolate side chain and the C1-aldehyde of the substrate is oxidized. The vehicle for the oxidation event is hydride ion transfer to NAD^+, generating NADH (energy capture) and the covalent 3-phosphoglyceryl group in thioester linkage to the sulfur of the active site cysteine.

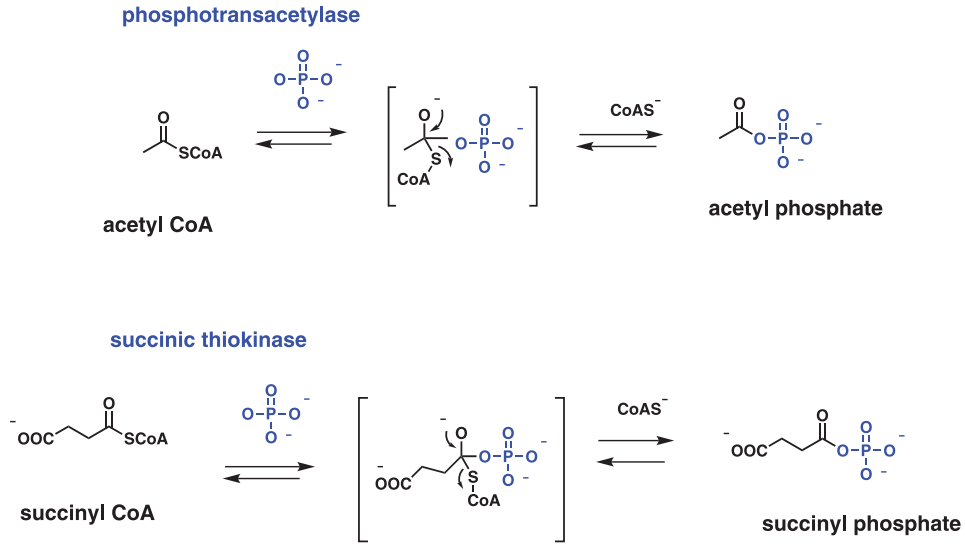


Figure 9.5 Phosphate capture of acyl thioesters can be reversible, equilibrating acyl phosphates with acyl thioesters.

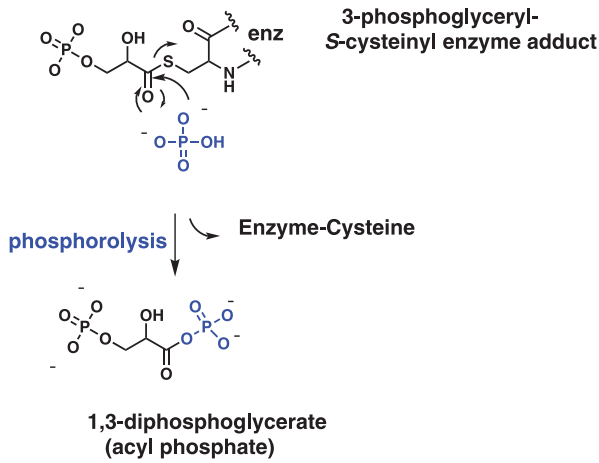


Figure 9.6 Although glyceraldehyde-3-phosphate dehydrogenase is named for its two-electron oxidative conversion of the aldehyde to acyl oxidation state, in the second step of the reaction the enzyme acts as a *phosphorylase*, releasing the nascent enzyme-bound 3-phosphoglycerly thioester as the 1,3-diphospholglycerate product.

As shown in Figure 9.6 phosphorolysis regenerates the free enzyme as the 1,3-diphosphoglycerate is released. The phosphoryl group at C_3 is a typical phosphate monoester. The phosphoryl group at C_1 is part of the acyl-phosphoric anhydride linkage. The phosphorolysis is an energy harvesting step, moving the activated 3-phospholgyceryl group from its tethered locus in the dehydrogenase active site to a thermodynamically activated product species free to diffuse in solution. In the next step of glycolysis the 1,3-diPGA diffuses to the active site of the 3-PGA kinase as a strategy for moving the phosphoric anhydride part of the acyl phosphate to ADP to form ATP. During gluconeogenesis glyceraldehyde 3-phosphate dehydrogenase does run in the opposite direction, using 1,3-diposphoglycerate for acyl group transfer to the active site cysteine side chain: the dual activation of acyl phosphoric anhydrides utilized.

9.3 Polynucleotide Phosphorylase

A third class of phosphorylases act on RNA. These are the polynucleotide phosphorylases noted previously in Chapter 5 as nucleotidyl transferases[14,15] (Figure 9.7). The enzyme is named for its activity as a 3′-exonucleolytic phosphorylase, but under some conditions the enzyme can function to add AMP residues to the 3′-ends of RNAs. This enzyme is a minor contributor to poly A tail elongation. A separate poly-A polymerase is the major contributor, adding the >100 AMP residues that constitute polyadenylation tails characteristic of mature messenger RNAs that exit nuclei.

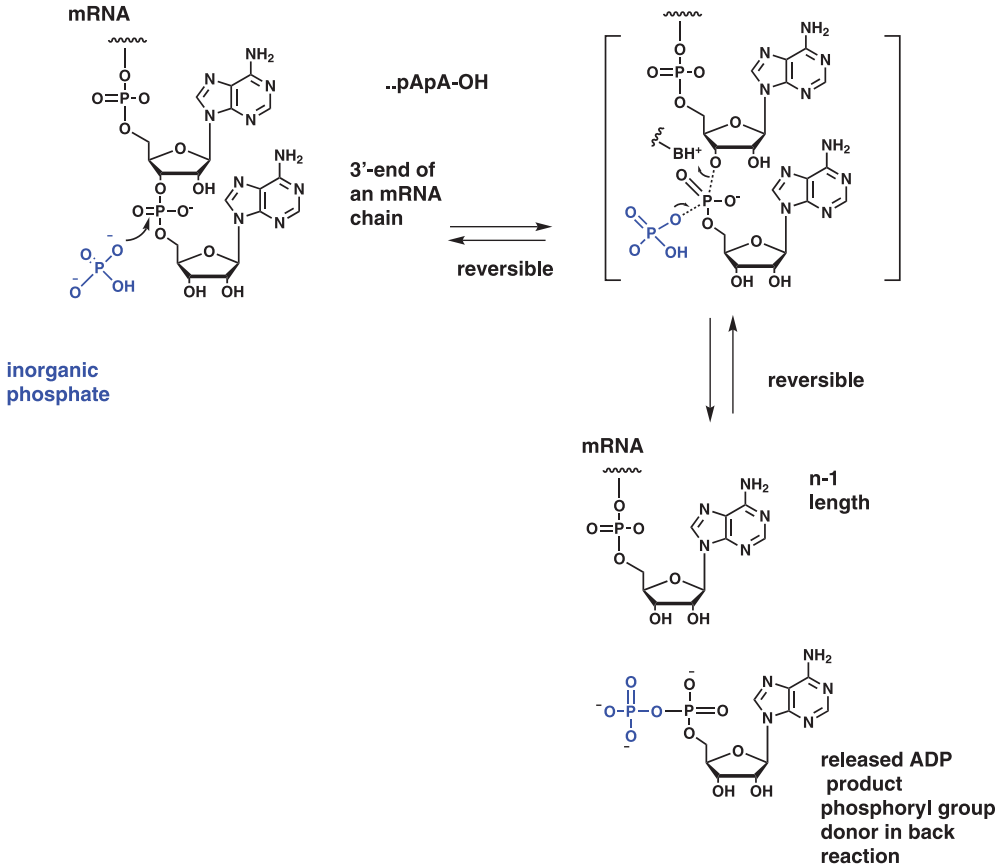


Figure 9.7 Polynucleotide phosphorylase is an unusual phosphorylase catalyst. It mediates sequential adenylyl transfers from the polyadenylated 3′–OH terminus of mRNAs, releasing adenylated phosphate = ADP as product in each phosphorolytic catalytic cycle. In the reversible back direction ADP, rather than ATP, is the unusual adenylyl donor.

In the phosphorylase direction, the 3′ end of an RNA chain gives up its terminal NMP moiety to an incoming Pi substrate. Each cycle of phosphorolysis shortens the RNA chain by one nucleotide from the 3′-end. The coproduct in each such catalytic cycle is ADP.

One unusual feature of polynucleotide phosphorylase is that the electrophilic fragment undergoing transfer to Pi is *not a carbon center but instead the electrophilic phosphorus of a phosphodiester*. The cleavage breaks that internucleotide phosphodiester linkage as the phosphorus atom embedded in the NMP group is attacked by one of the oxygens of Pi. In turn the released product is the nucleotide diphosphate ADP. There is net preservation of phosphodiester linkages between substrates and products. From this perspective polynucleotide phosphorylase can be viewed three ways: as a phosphodiester interchange catalyst, a nucleotidyl transferase in one direction, and a phosphorylase in the opposite direction.

Finally, viewed in the back direction as a poly-A adding enzyme, polynucleotide phosphorylase is almost unique in using ADP rather than ATP as the nucleotidyl (AMP) unit donor. No other extant nucleotidyl transferase discussed in Chapter 5 can use ADP in place of ATP, yet that might have been an early evolutionary step in nucleotidyl transfer catalysis in biology.

9.4 Phosphorylase Summary

This brief summary of phosphorylases illuminates the orthogonal role of inorganic phosphate as a nucleophile in contrast to the hundreds of electrophilic phosphoryl group transfers that populate all aspects of metabolism. In principle any electrophilic center would qualify as a substrate for a phosphorylase. Two types of (potential) carbon electrophiles subject to enzymatic phosphorolysis are glycoside linkages and acyl thioesters. The use of internucleotide phosphodiesters as electrophilic partners with Pi reminds us about the interconnected scope of inorganic phosphate and phosphoryl groups as two sides of the same coin of phosphate chemical biology. It may reprise earlier metabolic roles as a nucleophile and electrophile for phosphate during evolution.

References

1. S. Devendran, A. Abdel-Hamid, A. F. Evans, M. Iakiviak, I. H. Kwon, R. I. Mackie and I. Cann, Multiple cellobiohydrolases and cellobiose phosphorylases cooperate in the ruminal bacterium *Ruminococcus albus* 8 to degrade cellulo-oligosaccharides, *Sci. Rep.*, 2016, **6**, 35342.
2. C. Goedl, A. Schwarz, M. Mueller, L. Brecker and B. Nistzky, Mechanistic differences among retaining disaccharide phosphorylases: insights from kinetic analysis of active site mutants of sucrose phosphorylase and alpha, alpha-trehalose phosphorylase, *Carbohydr. Res.*, 2008, **343**, 2032–2040.

3. M. M. Adeva-Andany, M. Gonzalez-Lucan, C. Donaperty-Garcia, C. Fernandez-Fernandez and E. Ameneiros-Rodriguez, Glycogen metabolism in humans, *BBA Clin*, 2016, **5**, 85–100.
4. G. Pergolizzi, S. Kuhaudimiarp, E. Kalita and R. A. Field, Glycan phosphorylases in multi-enzyme synthetic processes, *Protein Pept. Lett.*, 2017, **24**, 696–709.
5. S. S. Macdonald, A. Patel, V. L. C. Larmour, C. Morton-Lang, S. J. Hallam, B. L. Mark and S. G. Withers, Structural and mechanistic analysis of a beta glycoside phosphorylase identified by screening a metagnomic library, *J. Biol. Chem.*, 2018, **293**, 3451–3467.
6. M. Kitaoka, Diversity of glycoside phosphorylases in glycoside hydrolase families, *Appl. Microbiol. Biotechnol.*, 2015, 8377–8390.
7. D. L. Zechel and S. Withers, Glycosidase mechanisms: anatomy of a finely tuned catalyst, *Acc. Chem. Res.*, 2000, **45**, 308–316.
8. H. Su, X. Sheng and Y. Liu, Insights into the catalytic mechanism of N-acetylglucosamidase glycoside hydrolase from *Bacillus subti*lis: a QM/MM study, *Org. Biomol. Chem.*, 2016, **14**, 3432–3442.
9. R. Silverstein, J. Voet, D. Reed and R. H. Abeles, Purification and mechanism of action of sucrose phosphorylase, *J. Biol. Chem.*, 1967, **242**, 1338–1346.
10. J. G. Voet and R. H. Abeles, The mechanism of action of sucrose phosphorylase. Isolation and properties of a beta-linked covalent glucose-enzyme complex, *J Biol. Chem.*, 1970, **245**, 1020–1031.
11. C. Castano-Cerezo, J. Pastor, S. Renilla, V. Bernal, J. L. Iborra and M. Carbovas, *Microb. Cell Fact.*, 2009, **8**, 54.
12. D. Phillips, A. M. Aponte and S. A. French, Succinyl-CoA synthetase is a phosphate target for the activation of mitochondrial metabolism, *Biochemistry*, 2009, **48**, 7140–7149.
13. Nicholls, GAPDH: a common enzyme with uncommon functions, *Clin. Exp. Pharmacol. Physiol.*, 2012, **39**, 674–679.
14. B. K. Mohanty and S. R. Kushner, Polynucleotide phosphorylase functions both as a 3′ to 5′ exonuclease and a poly (A) polymeraase in Eschericicai coli, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 11966–11971.
15. S. Yehudai-Resheff, M. Hirsch and G. Schuster, Polynucleotide phosphorylase functions as both an exonuclease and a poly (A) polymerase in spinach chloroplasts, *Mol. Cell. Biol.*, 2001, **21**, 5408–5416.

CHAPTER 10

N–P Bond Chemical Biology

10.1 *N*-Phosphoryl Proteinogenic Amino Acids

In comparison to the ubiquity of phosphate oxygen ester chemical biology, N–P bonds in metabolites are substantially more limited. One major class of N–P scaffolds involves enzymatic phosphoryl transfers to nitrogen atom-containing nucleophiles. Figure 10.1 notes that five of the twenty proteinogenic amino acids can become phosphorylated on a side chain nitrogen atom. The nitrogen groups can be primary amines, presumably as the free base forms, and two less likely nitrogen nucleophiles: amides and amidines.

A sixth example in that same figure shows phosphorylation of the α-amino group of leucine in a L-Leu–L-Trp dipeptide backbone[1] (in a pair of natural antimetabolites phosphoramidon and talopeptin, noted in detail in Section 10.7 of this chapter). Given this Leucyl α-N–PO_3^{2-} precedent, any free primary amine of an amino acid moiety could thereby undergo enzymatic *N*-phosphorylation in nature. In addition, a recent review in 2018 catalogued some 55 secondary (conditional) metabolites that contain one or more N–P bonds.[1]

The other primary amine in Figure 10.1 is the C_6-side chain amine of lysine but the context is not the free amino acid but rather a lysine residue side chain found in the chromosomal protein histone H1. The imidazole side chain of histidine residues has a much higher proportion of the side chain as a free base than the C_6 ammonium form of lysine (imidazolium ion pK_a ~6.5 *vs.* lysine-C6-ammonium pK_a ~10.5) and histidine-*N*-phosphorylation is substantially more frequent than lysine residue *N*-phosphorylation. The two secondary amine nitrogens of the imidazole ring are numbered N_1 and N_3 where the alanyl side chain is at C_2. Figure 10.1 shows the N_3-imidazolyl-phosphate regioisomer of N_3-phosphoryl-His but the N_1 regioisomer is also known.[2] The known biological contexts for histidine phosphorylation are in phosphoproteomics. We will note that *N*-phospho-His residues retain the

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

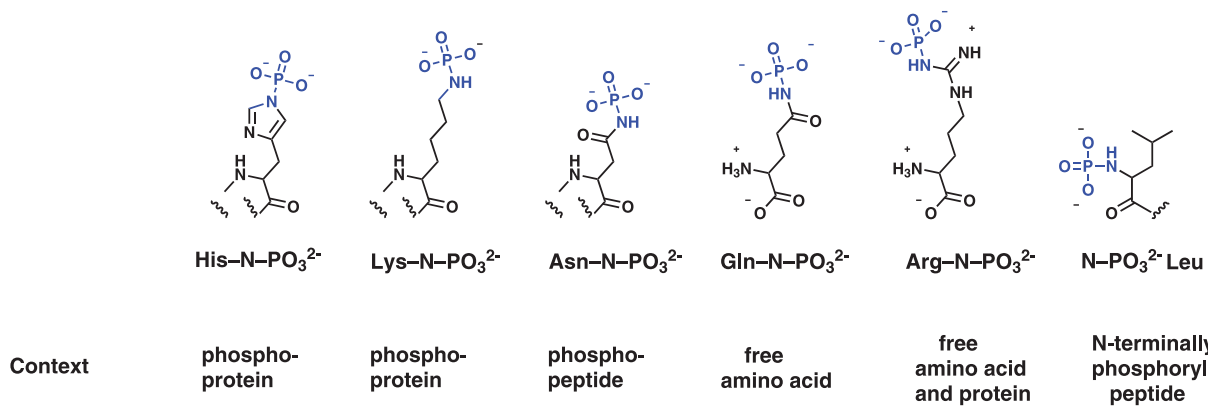


Figure 10.1 Five of the 20 proteinogenic free amino acids can be *N*-phosphorylated during metabolism. The sixth, Leu, is *N*-phosphorylated on the free amino group in the dipeptides phosphoramidon and talopeptin.

high group transfer potential of the phosphoryl group: most of the phospho-His residues are catalytic intermediates in enzymatic phosphoryl transfer reactions. These examples including the category of bacterial histidine sensor kinases taken up in Section IV.

Two examples of *N*-phosphorylated amides, *N*-phosphoryl glutamine and *N*-phosphoryl-asparagine are included in Figure 10.1. Amides are at best weak nitrogen nucleophiles and, *ab initio*, are surprising participants. We delve into these two examples below.

10.2 Phosphagens and Phosphoramidates

The third category of proteinogenic amino acids that become *N*-phosphorylated are modified at the terminal amine of an amidino group. Arginine phosphate is the representative here as an *N*-phospho-guanidinium metabolite. Again, the guanidinium group is an unlikely nitrogen nucleophile, with a typical pK_a for loss of a proton from the cationic guanidinium to the free base form in the range of pH 12–14.

The *N*-phosphoguanidinium-containing metabolites are sometimes referred to as *phosphagens.* The term phosphagen implies that the molecule is a source of phosphoryl groups and refers to the biological function of this small class of eight such *N*-phosphoryl metabolites. As shown in Figure 10.2, *N*-phosphoarginine is one of eight related phosphagens found in invertebrates, including earthworms. Lommbricine for example was isolated from *Lombricus terrestris*, the common earth worm. In *Trypanomsa brucei*, there are three isoforms of arginine kinase: one in the flagellum, one in glycosomes, and one in the cytoplasm, presumably serving local energy buffer functions.[3]

The phosphagens fall into distinct subclasses. Lombricine and thalassamine are *O*-phosphoserine derivatives to which an ethanolamine and then an amidino extension have been added. (These are unusual scaffolds with both a phosphodiester and a phosphoramidate linkage). The hypotaurocyamine and taurocyamine derive biosynthetically from taurine and hypotaurine. Glycocyamine and creatine form a pair with carboxylate anion replacing sulfonate or sulfinate anions. All bear the guanidine group as the site of *N*-phosphorylation as phosphagens.

The eighth representative, creatine phosphate, is the sole phosphagen in vertebrates. It is well-studied as a buffer molecule for ATP production in high metabolic demand tissues such as skeletal muscle. Creatine kinase catalyzes the fully reversible transfer of the terminal γ-PO_3^{2-} of ATP to the deprotonated guanidine terminal nitrogen in the active site of creatine kinase[4,5] (Figure 10.3). Creatine phosphate can be stored in skeletal muscle in its phosphagen role. During active muscle contraction as ATP is rapidly consumed by the ATPase activity of actomyosin, creatine kinase can run backwards, converting ADP into ATP for continued use by the sarcoplasmic contractile machinery.

The phosphoryl group stored in creatine phosphate also is an alternative to Pi generation from ATP hydrolysis. There is thought that phosphagens

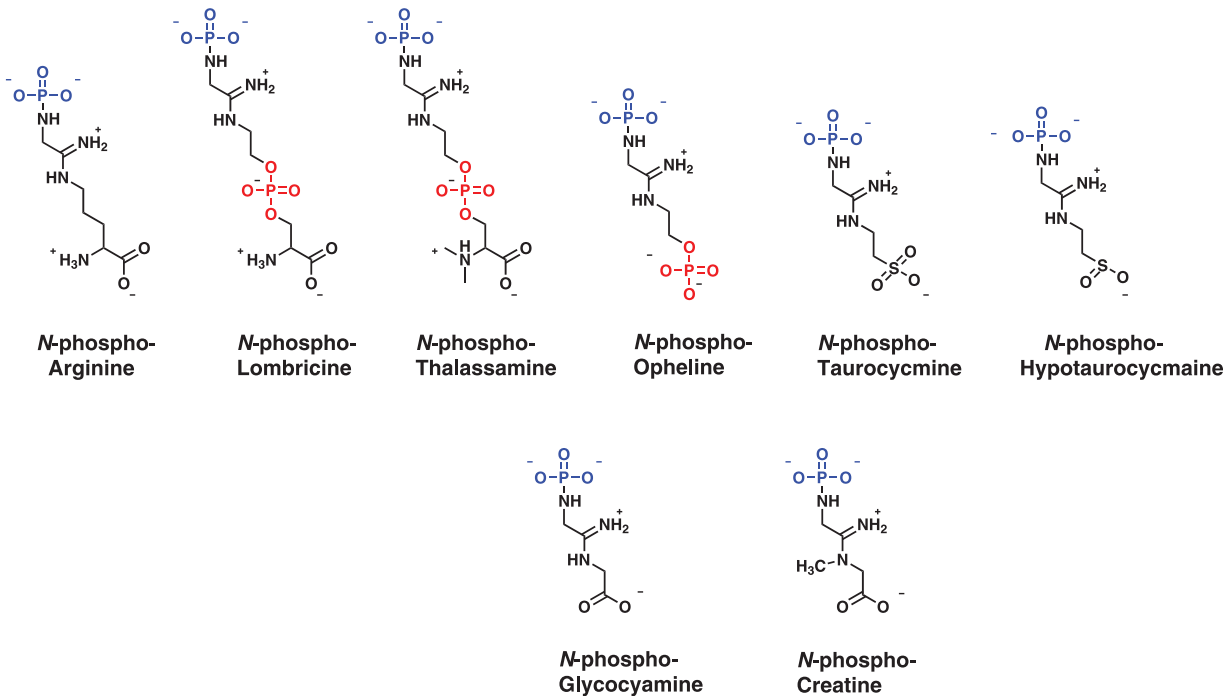


Figure 10.2 *N*-phosphorylated metabolites (*phosphoramidates*) serve as reservoirs of high phosphoryl group transfer potential to react under enzymatic catalysis with ADP to form ATP. They are collectively termed phosphagens.

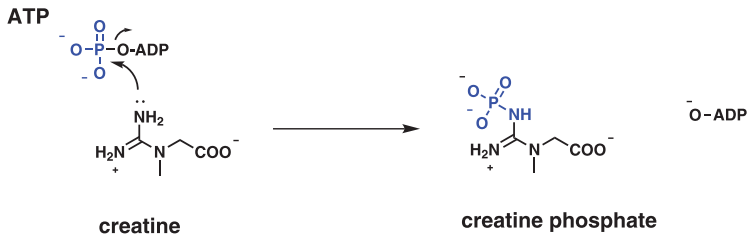


Figure 10.3 Creatine kinase is an *N*-amidino kinase, reversibly transferring the terminal $-PO_3^{2-}$ group between ATP and the deprotonated amidine (guanidine) group of creatine.

may have evolved to control levels of inorganic phosphate in high demand tissues with episodic high bursts of ATP usage. Of course, as creatine phosphate runs backwards through creatine kinase to form ATP and the ATP is then hydrolyzed, a stoichiometric amount of inorganic phosphate is formed, but it may be the instantaneous concentrations of phosphagen *vs.* [Pi] that matter. Both ADP and Pi have to make their way from sites of ATPase action throughout a cell to and through dedicated ADP- and Pi transport proteins in the mitochondrial membranes to be available (~70 kg per day in an adult) for the ATP synthase machinery. The seven invertebrate phosphagens play similar roles as ancillary stored supplies of ''high energy'' (thermodynamically activated) phosphagens.

Creatine biosynthesis originates with the two-carbon achiral proteinogenic amino acid glycine[6] (Figure 10.4). An amidinotransferase enzyme transfers the amidino group from arginine to the amine of glycine, creating guanidinoacetate in the rate limiting step in the pathway. This occurs principally in kidneys: guanidinoacetate is secreted into the blood and subsequently methylated by the *S*-adenosylmethionine-consuming guanidinoacetate *N*-methyltransferase (Figure 10.4B) to produce creatine. Creatine is then transported across the plasma membrane into tissues such as cardiac and skeletal muscle and brain by a sodium ion-coupled creatine transport protein. Biosynthesis accounts for about 1 g of new creatine per day and another 1 g would derive from an omnivore diet.

The enzymatic amidino transfer involves a two step mechanism and a covalent amidino enzyme intermediate[7] (Figure 10.4A). The thiolate anion form of the side chain of an active site cysteine adds into the imine carbon of L-arginine. The resultant tetrahedral adduct can expel coproduct ornithine and yield a covalent thiohemiamidine adduct. This is the species captured in the second half reaction by the cosubstrate glycine's deprotonated amino group. Similar amidinotransferases are involved in generation of the other six amidino scaffolds of the invertebrate phosphagens.

In the subsequent action of creatine kinase (and the other seven invertebrate guanidinium kinases that create phosphagen energy reserves), the most difficult step in catalysis may be to effect deprotonation of the substrate guanidinium group. The guanidinium pK_a of creatine is 12.7 while the

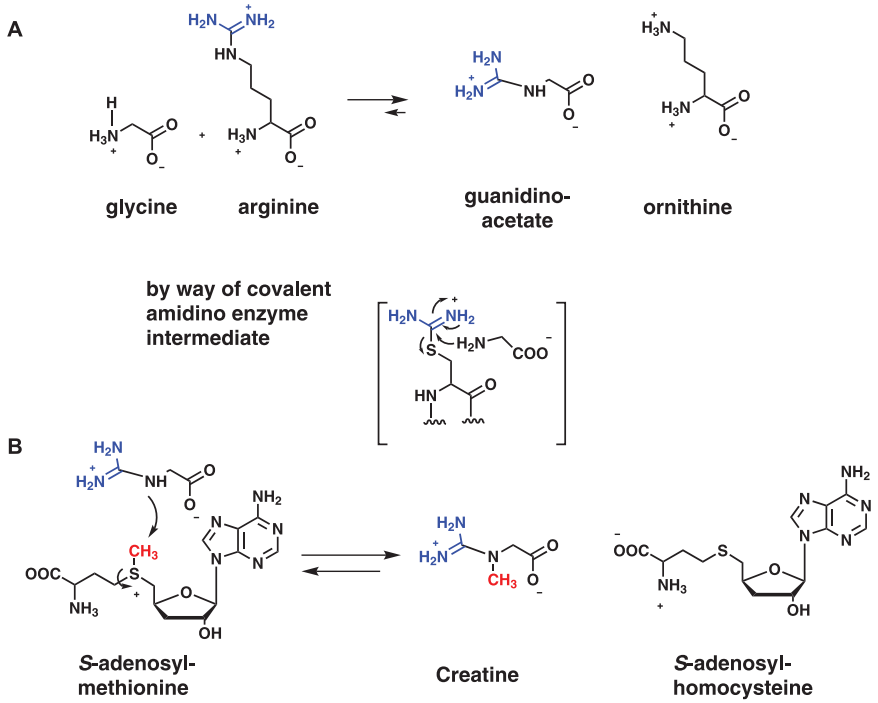


Figure 10.4 Two enzymatic steps convert glycine to creatine. (A) The first enzyme is an amidino transferase using L-arginine as amidino donor and glycine as acceptor *via* an intermediate amidino-*S*-enzyme intermediate. (B) The second enzyme is a conventional SAM-dependent *N*-methyltransferase acting on guanidinoacetate as nitrogen nucleophile.

guanidinium group in arginine has been remeasured to be 13.7. Thus, at near neutral pH there is less than one part in 10^5 to one part in 10^6 of the deprotonated guanidine form of creatine or arginine required for the nitrogen to function as a nucleophile. An ion pair with a conserved aspartate residue β-carboxylate side chain in the kinase active sites is likely to be the preamble to proton transfer in the phosphoryl transfer transition state.

The guanidinium nitrogens are, to put it mildly, unlikely groups to have been conscripted by biology as nucleophiles. On the other hand, once formed the positive nitrogen charge on the *N*-phospho-guanidinium product ensures it is thermodynamically activated. The phosphoryl transfer to ADP to refurnish ATP levels in tissues actively consuming large amounts of energy (ATP) makes them a useful ready energy store – the function of a phosphagen.

10.3 Phosphoramidate Nomenclature

Another piece of terminology for describing N–P compounds is the word phosphoramidate. When all three –OH groups of phosphoric acid are

substituted by NH_2 this triamino parent compound is known as phosphoramide. There is no known biology for the parent phosphoramide. When there is one substitution of NH_2 for one –OH, the parent compound is phsphoramidic acid (Figure 10.5). At physiological pH values, the anion is termed a phosphoramidate. All phosphorus-substituted amines are likewise termed phosphoramidates. Thus, all the *N*-phosphoryl amino acids of Figure 10.1 and all the phosphagens of Figure 10.2 are formally phosphoramidates.

There are some naturally occurring phosphorotriamidates. Two such metabolites are octicidin and sulphostin[1] (Figure 10.5). Two nucleotides with replacement of one of the terminal phosphate side chain oxygens by NH_2 are known: AMP phosphoramidate and 3′5′-CDP phosphoramidate (Figure 10.5). We will note later in this chapter that the latter compound is a phosphoramidyl donor in the maturation of the capsular polysaccharide of the intestinal pathogen *Campylobacter jejuni*. A few compounds with S–P–N bonds are also known; those are phosphorothioamidates.

10.4 *N*-Phospho-asparaginyl Residue in Microcin C7 Maturation

Microcins are low molecular weight, ribosomally-generated peptides synthesized as antimetabolites by bacterial strains to kill neighboring strains in order to create nutrients and space for the microcin producers to grow.[8,9] Among the intriguing structures and mechanisms of bacteriocidal actions is the heptapeptide microcin C7[10] (Figure 10.6). Strains of *E. coli* encode the gene cluster that contains both the small 21 base pair *mccA* structural gene and ancillary genes including *mccB* necessary for creation of the unusual AMP–phosphoramidate linked to the C-terminal aspartate residue of the active, mature microcin C.

Mature microcin C is secreted by producer *E. coli* strains and taken up by neighboring strains that lack the plasmid and are vulnerable. Peptidases in the recipient strain hydrolyze the heptapeptidyl–AMP phosphoramidate down to the free aspartyl–N–AMP that is a potent nonhydrolyzable analog of the normal asparaginyl–AMP intermediate for asparaginyl tRNA synthetase (Figure 10.7). Competitive blockade of asparagine incorporation into proteins by the tight binding nonhydrolyzable phosphoramidate brings protein synthesis to a halt and that sets off responses that lead to bacterial death.

The mechanism of formation of the P–N phosphoramidate bond to AMP by the MccB enzyme has been investigated and found to consume two ATP molecules for each N–P linkage generated.[11] The first ATP is cleaved to PPi and AMP, indicating a nucleotidyl transfer. The second ATP releases PPi while the AMP portion becomes part of the phosphoramidate linkage. A third factoid of import is that the *mccA* gene encodes an asparaginlyl residue at the C terminus of the nascent heptapeptide, not the aspartate observed in the mature microcin C.

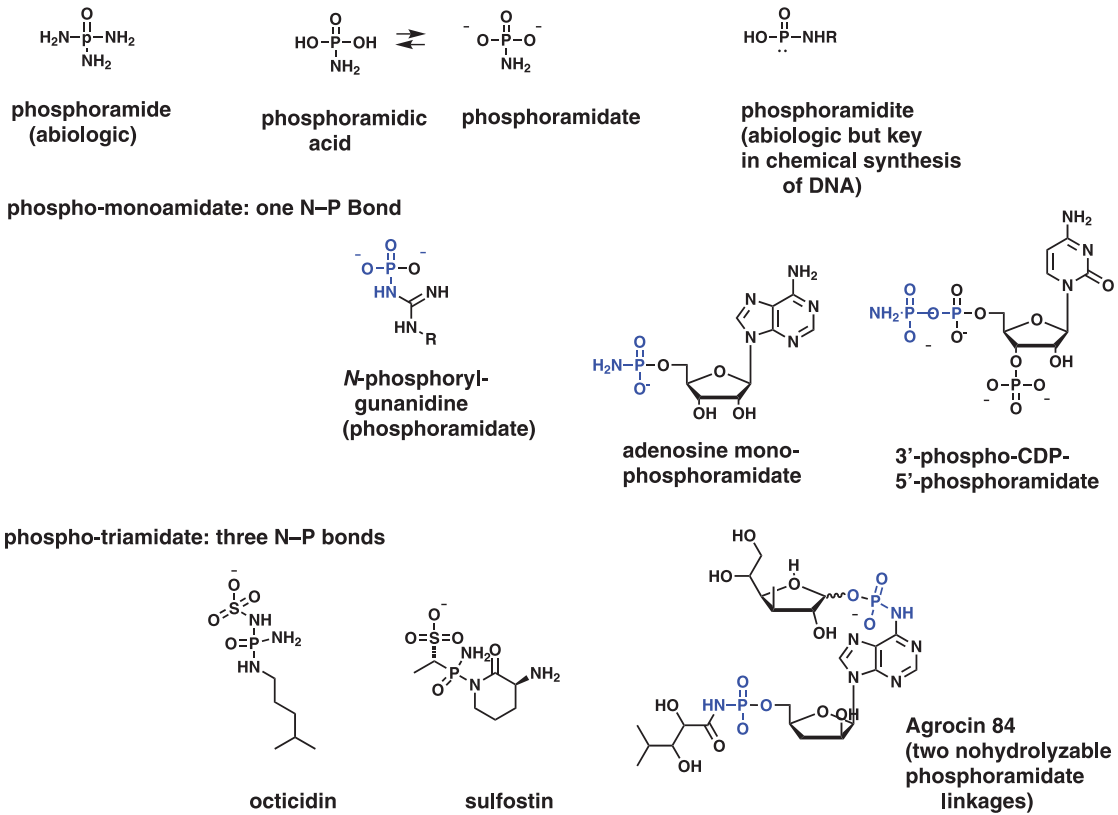


Figure 10.5 Phosphoramide and phosphoramidate nomenclature. Monoamidates and triamidates. Phosphoroamidites (P^{III} oxidation state) are not biological metabolites but are key intermediates in chemical synthesis of oligonucleotides.

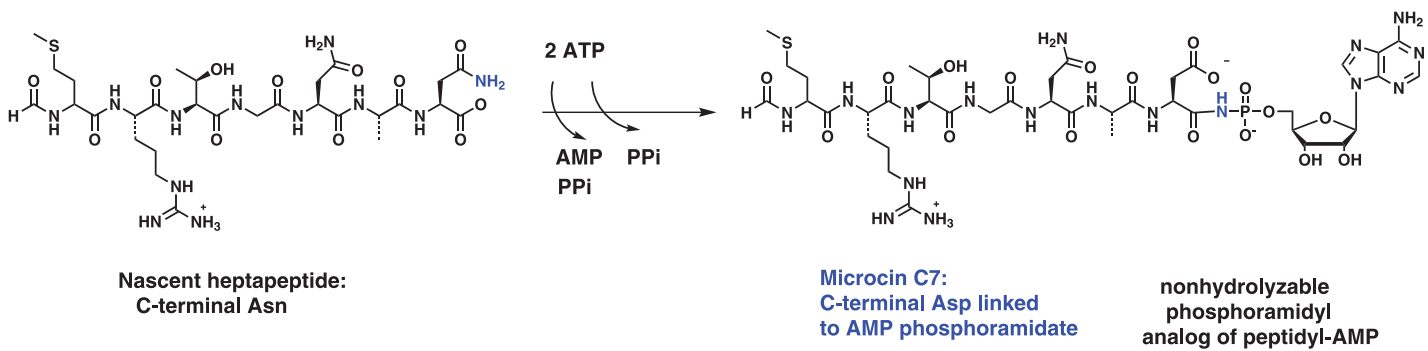


Figure 10.6 Microcin C7 is a posttranslationally modified bacterial heptapeptide with an unusual *N*-phosphoramidyl-AMP group as nonhydrolyzable analog of the corresponding acyl-AMP.

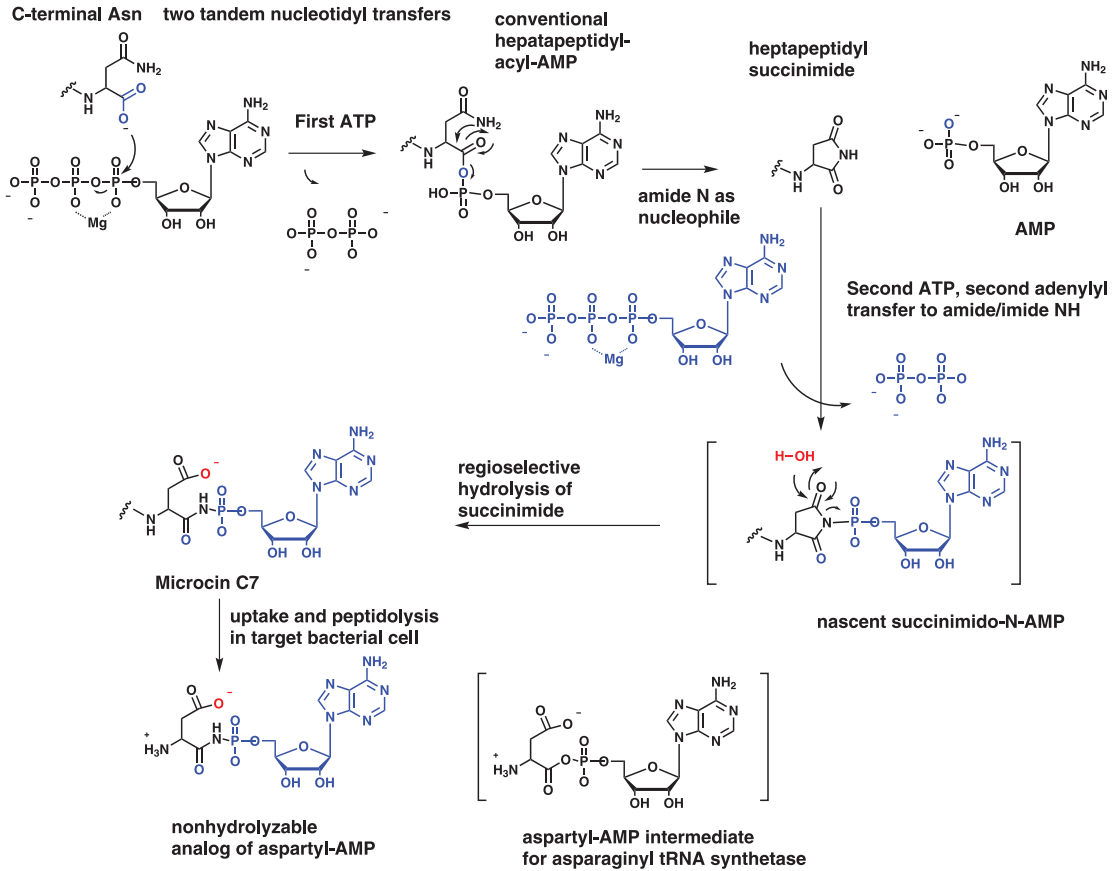


Figure 10.7 Two adenylyl transfers in the posttranslational maturation of microcin C7. The first is a conventional adenylyl transfer to the C-terminal carboxylate of the heptapeptide. The second is AMP group transfer to the N–H of a heptapetidyl–succinimide intermediate.

MccB starts a catalytic cycle with a nucleotidyl transfer of the AMP moiety of the first ATP to the carboxylate of the nascent heptapeptide C-terminal asparagine (Figure 10.7). This forms the canonical hexapeptidyl–AMP (acyl-phosphoric anhydride). The AMP moiety then undergoes displacement with intramolecular assistance from the side chain of asparaginyl residue$_{7.}$ The result is free AMP and conversion of the linear side chain amide of Asn_7 into a cyclic succinamide. Now, that C-terminal succinamide acts as nucleophile for nucleotidyl transfer of the AMP moiety of the second ATP. This step releases the second overall equivalent of PPi and tethers this AMP moiety covalently to the heptapeptidyl backbone with a new N–P bond. This is the phosphoramidate formation step.

The MccB enzyme has one last step to perform. It mediates a regiospecific hydrolysis of the cyclic imide to the linear microcin C product. The N–P bond is undisturbed in this imide hydrolysis step. The C terminal amino acid is now an aspartate rather than the starting asparagine, creating part of the recognition for binding to and inhibiting aspartyl tRNA synthetase. To get to the N–P bond, though, the system must start with a C-terminal Asn rather than Asp residue.

At two stages during this consumption of two ATP molecules the MccB enzyme coaxes first a weak amide nitrogen, then a more active imide nitrogen to function as a nucleophile. The value of the N–P bond is that it creates a kinetically stable isostere to the O–P bond in the natural aspartyl–tRNA substrate.

10.5 *N*-Phospho-glutamine and Tailoring of *C. jejuni* Capsular Polysaccharide

Campylobacter strains can be pathogens causing gastroenteritis in the human gastrointestinal tract by way of ingesting contaminated poultry. Among the pathogenic mechanisms, some invasive C. *jejuni* strains produce a capsular polysaccharide that is decorated with an unusual *O*-methyl-phosphoramidate group on two of the sugars in a repeat saccharide unit.[12] One sugar is 2-acetamido-2-deoxy-D-galatofuranose. The other is D-glycero-α-L-glucoheptapyranose (shown in Figure 10.8).

Four enzymes are required for formation and transfer of the *O*-methyl-phosphoramidate group. The first enzyme is an unusual ATP-dependent *amide "kinase"* generating the N–P bond as an $N–PO_3^{2-}$ group on the

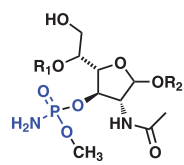


Figure 10.8 Structure of the *O*-methylphosphoramidate group in the capsular polysaccharide of some pathogenic strains of *Camphylobacter jejuni*.

glutamine amide NH group.[13] The coproducts are AMP and Pi. The side chain amide of glutamine, like the side chain of asparagine above, is a weak nucleophile. Nonetheless, hundreds to thousands of *N*-glycoproteins are biosynthesized by that carboxamido nitrogen of asparaginyl side chains in proteins acting as a nucleophile for glycosylation.

The cleavage to AMP and Pi by the purified enzyme, in the absence of contaminating PPiase, indicates an as yet unexamined unusual mechanism. The formation of AMP as one product is consistent with a nucleotidyl (adenylyl) transfer at some stage of the catalytic cycle. It is also consistent with a rare pyrophosphoryl transfer and then a P–O–P anhydride bond cleavage to further drive the biosynthetic equilibrium to $N–PO_3$ formation (Figure 10.9). One would need to determine if Pα (adenylyl transfer) or Pβ (pyrophosphoryl transfer) becomes attached to the glutamine NH group as preamble to mechanism. There is also the question of a possible isoamide form of Gln as nucleophile followed by a PO_3^{2-} group O to N transfer (*e.g.* followed by cleavage of a pyrophosphoryl to phosphoryl *N*-linked product).

10.6 Nucleotide Phosphoramidates: Transfer of a Phosphoramidinyl Group

The second and third enzymes in the *C. jejuni* pathway bring nucleotides into play.[14] Specifically, CTP is the donor of a CMP moiety in a variant of a classic nucleotidyl transfer. Traditionally an oxyanion of a phosphate ester would attack CMP to give a cytidine diphospho diester product. In this unusual case the phosphoramidate oxyanion of the glutaminyl–$N–PO_3^{2-}$ is the nucleophile and the product is the CDP-L-glutamine with an internal phosphoramidate-ester linkage instead of the canonical phosphodiester[15] (Figure 10.9).

The third enzyme is a version of a classic glutaminase, hydrolyzing the amide linkage by C–O bond cleavage. In this case the product glutamate is accompanied by a molecule of CDP-phosphoramidate. The $N–PO_3^{2-}$ group has been transferred from its original home on the glutamine scaffold to the CMP moiety in a short three enzyme pathway.

The fourth enzyme takes a page out of sulfate activation strategy. In sulfate activation for $–SO_3^-$ transfers, two ATPs are spent. The first enzyme uses a sulfate oxygen as nucleophile to attack Pα of ATP to give product adenylsulfate (APS) (Figure 6.7 in Chapter 6). The second ATP is then utilized as phosphoryl donor to the 3′-OH to yield 3′-phospho-5′-adenylsulfate (PAPS). This is the preferred donor for sulfuryl transferases to transfer the SO_3^- group to cosubstrate nucleophiles.

In the *Campylobacter jejuni* pathway the fourth enzyme is a similar phosphoryl transferase, transferring a PO_3^{2-} group to the 3′-OH to yield 3′,5′-CDP-phosphoramidate (Figure 10.9). The latter steps in the capsular polysaccharide maturation pathway are as yet undetermined but the PAPS precedent would predict transfer of the phosphoramidate moiety to the

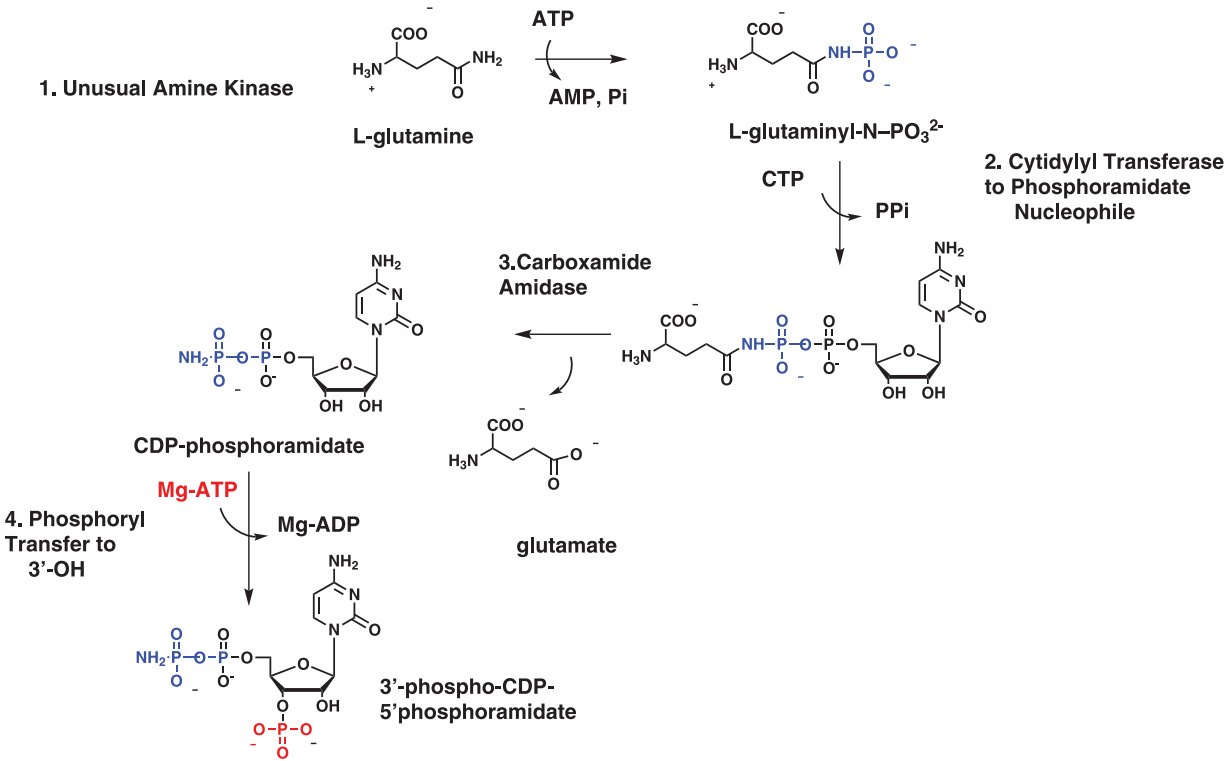


Figure 10.9 Four enzyme pathway from glutamine to 3′-phospho-CDP-5′-phosphoramidate involve a novel glutamine carboxamide *N*-kinase as first step and subsequent donor of the phosphoramidyl group.

galactofuranose and glucoheptose –OH groups. Then, it would only remain for an *S*-adenosylmethionine dependent *O*-methyltransferase to finish the *O*-methylphosphoramidate tailoring (Figure 10.10).

As an aside, there is indeed an enzyme that uses NH_3 and adenylsulfate (APS) to carry out adenylyl transfer to NH_3, forming adenyl phosphoramidate[1] (Figure 10.11). APS is an alternate sulfuryl group donor to PAPS. Perhaps adenyl phosphoramidate may yet function for phosphoramidyl transfer (with adenosine as coproduct) to cosubstrate nucleophiles. It seems unlikely that adenyl phosphoramidate would transfer ammonia as a nucleophile. Such chemical biology has not yet been reported.

10.7 Talopeptin/Phosphoramidon: α-Aminokinase Action

Phosphoramidon is a streptomycete metabolite that functions as an inhibitor of metallopeptidases by binding to their active sites and chelating the active site divalent cations (Figure 10.12). Phosphoramidon has a Leu–Trp dipeptide scaffold with the amino group of the leucyl residue in phosphoramidate linkage. One of the phosphoramidate oxygens is also in ester linkage to L-rhamnose, such that the phosphoramidyl moiety is also part of a phosphoramidate diester linkage. A congener known as talopeptin differs only on the stereochemistry at one of the deoxyhexose–OH groups (Figure 10.12). The biosynthetic path to talopeptin has recently been deciphered, demonstrating the involvement of TalE as a novel amino kinase.[16] The primary amine of the leucine moiety of the Leu–Trp dipeptide is the nucleophile attacking Pγ of ATP for phosphoryl transfer to yield the N-terminal phosphoramidate (Figure 10.12). The Leu primary amine is a much more robust nucleophile than the carboxamide nitrogens of Asn or Gln in the two previous phosphoramidate examples.

The short pathway concludes with L-rhamnosyl transfer from the prototypic nucleoside diphosphate glycosyl donor, 2′deoxythymidine-L-rhamnose (Figure 10.12). This is a variant of the standard transfer of glycosyl oxocarbenium units to phosphate monoester anions discussed in Chapter 5. In this case the nucleophile is a phosphoramidate monoanion to yield the phosphoramidate ester final metabolite talopeptin.

Three other examples of phospohramidate natural products are worth passing mention (Figure 10.5). The first is agrocin 84 an agrobacterium-generated antimetabolite that has elements of a complex AMP scaffold.[1] It harbors *two* phosphoramidate linkages, one at N_6 and the other in the side chain phosporamidate moiety. The biosynthetic logic has not been deciphered but the N6-phosphormaidate could arise by attack of the N_6-amine of ATP as in the talopeptin example. The linkage at the side chain phosphoramidate is an amide bond to a dihydroxy acid. That might arise by prior formation of the 5′-phosphoramidate that then attacks an acyl CoA derivative to form the 5′-phosphoramidate amide scaffold.

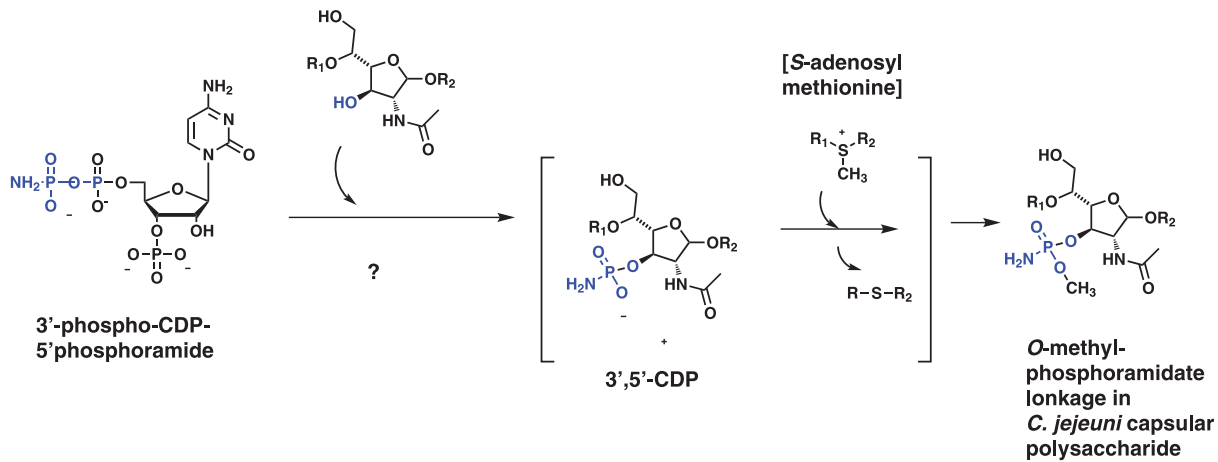


Figure 10.10 Proposed transfer of a [CH_3^+] equivalent from *S*-adenosylmethionine to one of the oxyanions of the phosphoramidate group in the *C. jejuni* capsular polysaccharide ensures the product is a monoanion not a dianion.

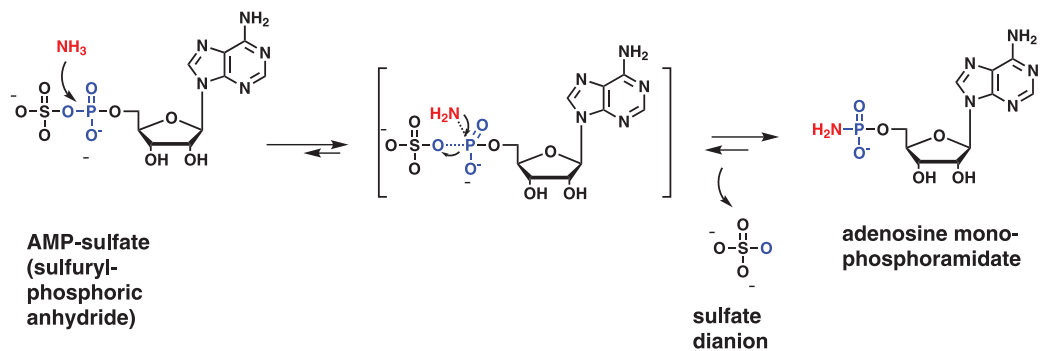


Figure 10.11 Enzymatic formation of adenosine monophosphoramidate by enzymatic attack of free NH_3 on adenyl sulfate: net adenylyl transfer to NH_3 with release of SO_4^{2-}.

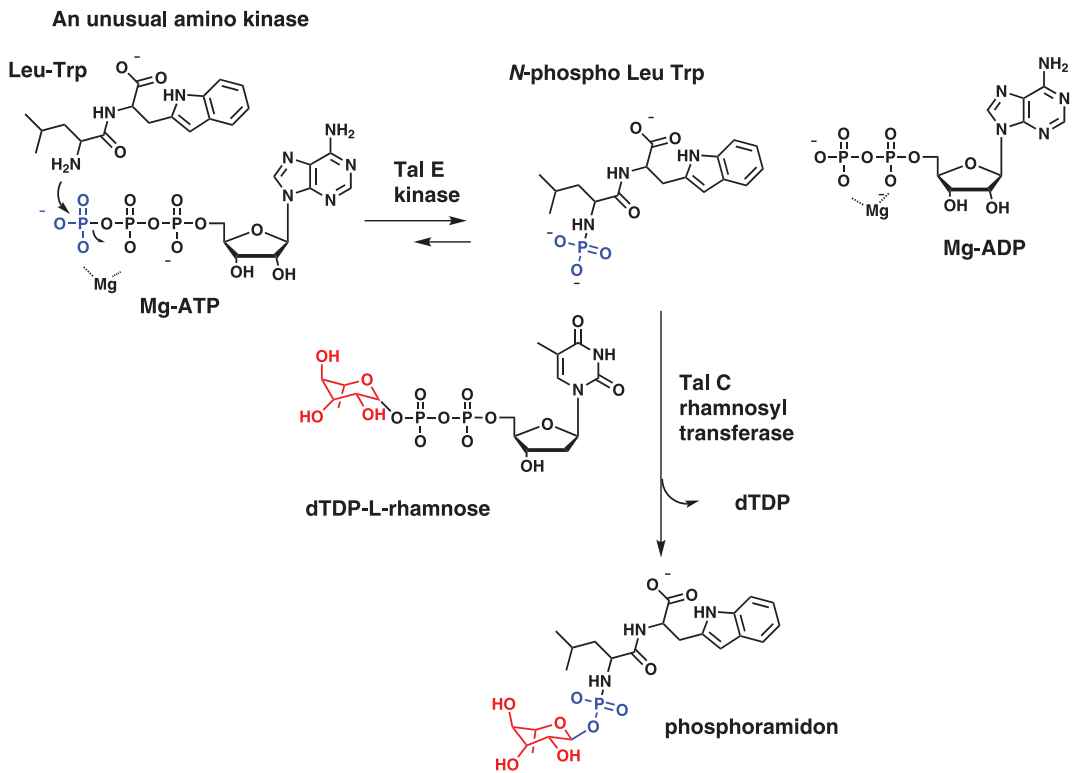


Figure 10.12 The formation of phoshoramidon and talopeptin happens in two enzymatic steps from the dipeptide Leu–Trp. First is an aminokinase (TalE in the talopeptin pathway). Then, one of the two oxyanions of that phosphoramidate group attacks C1 of the rhamnosyl moiety of dTDP–L-rhamnose to transfer the electrophilic rhamnosyl moiety to complete the natural product scaffold.

The other two structures in Figure 10.5, octicidin and sulfostin, are rare examples of a phosphotriamidate group – phosphorus with three N–P bonds and one O–P bond (present as O=P).[1] The biosynthetic assembly strategy for phosphotriamidate construction is not deciphered. One can speculate that the sulfonamide bond forms late from a sulfuryl transfer from adenylsulfate. The branched hexylamine might be the nucleophile on a phosphodiamidate intermediate. Whether such a phosphodiamidate could arise from iterative transfers from ATP is unknown. No examples of stable phosphodiamidates are yet known in metabolism.

10.8 Phosphoramidites in Synthetic Oligonucleotide Chemistry

In Chapter 1 we noted that there are rare reports of lower valent states of phosphorus other than P^V in biology including detection of PH_3 as a product of anaerobic bacterial metabolism. In addition, other than rare NAD^+-dependent microbial oxidation of phosphite ($P = +3$) to phosphate by phosphite dehydrogenase,[17] essentially all of phosphorus chemical biology occurs in the (+5) oxidation state. The equivalent of phosphite for a (P^{III}) N–P containing species is a *phosphoramidite* (Figures 10.5 and 10.13). To date there are no indications that phoshoramidites, as compared with the *phosphoramidates*, feature in phosphorus metabolic biology.

However, we note in passing that *phosphoramidite chemistry* has been the mainstay for oligonucleotide chemical synthesis over the past forty years.[18] Among the enabling discoveries that have transported biology into the postgenomic era, one can place DNA sequencing and the orthogonal ability to synthesize oligonucleotides and then whole genes as dramatic technologies. DNA synthesis has been automated for decades and each internucleotide bond-forming step is estimated to proceed with accuracy and yield in the region of 99.5%. A quick calculation indicates that 200 cycles to get a 201 oligonucleotide with this coupling efficiency yields ~37% of the desired product. Ligation of five 200mers will produce a 1 kilobase synthetic DNA that can encode a protein of ~30 000 Daltons.

The core strategy for synthetic DNA chemistry is (1) to protect exocyclic amino groups in the adenine, guanine and cytosine bases with acid-labile groups and (2) use phosphorus in the [+3] oxidation state as a cyanoethyl *phosphoramidite* as a nucleophile for the coupling step. The ligated phosphite diester product has created the internucleotide bond as the P–N bond cleaves with diethylamine departure (Figure 10.13). (3) The phosphite triester is then oxidized quantitatively to the phosphate [+5] triester and (4) the base-labile cyanoethyl protecting group removed by ammonia only after release of the full length deoxy-oligonucleotide from the resin.

The phosphite to phosphate oxidation step seems an unlikely route to DNA at first glance,[19] but the dialkyl-phosphoramidite group has proven to be among the most reactive in ligating the two nucleotides together. Each

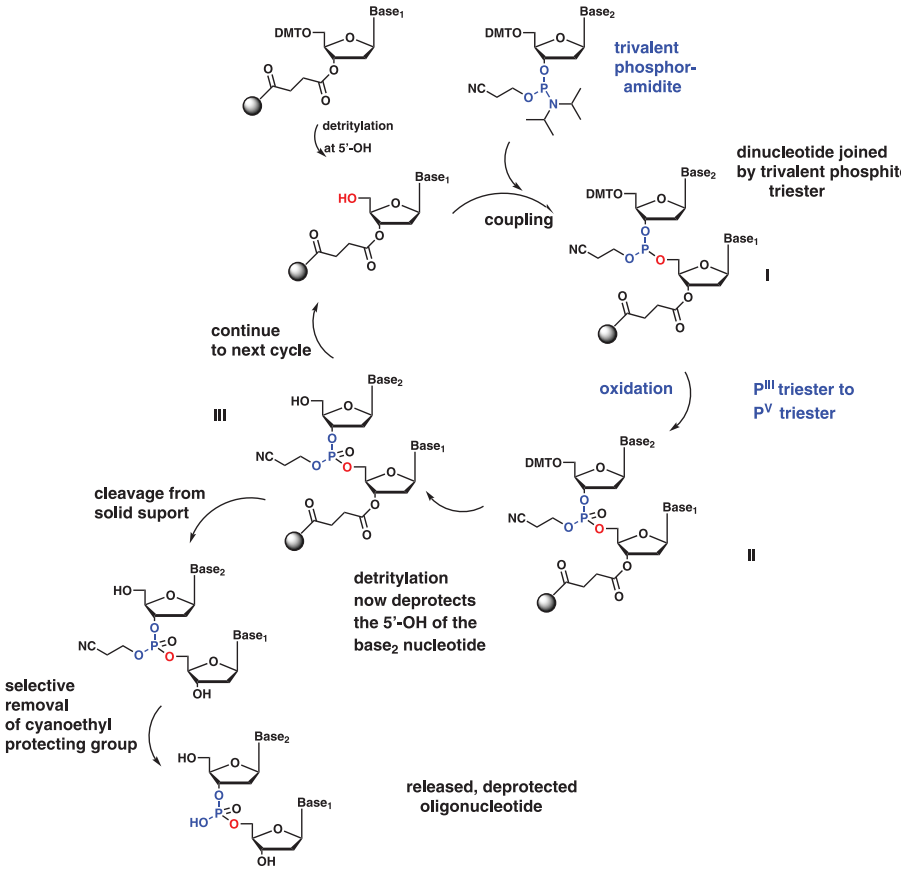


Figure 10.13 Schematic for DNA oligomer synthesis chain elongation by addition of 5′ nucleotides in each cycle. The incoming nucleotide is a 3′-phosphoramidite with phosphorus in the nucleophilic P^{III} oxidation state. After coupling to form the internucleotide *phosphite diester*, the phosphorus is oxidized to the P^{V} electrophilic oxidation state in the internucleotide *phosphodiester.* Detritylation frees up the 5′-OH of the just added nucleotide for the next coupling cycle.

incoming mononucleotide is protected at the 5′-OH as a -4,4′dimethoxytrityl (DMT) group to block reactivity at that position. Each is meanwhile activated as the 3′-phosphoramidite. The coupling direction is from the 5′-OH of the growing deoxyoligonucleotide tethered to the solid support onto the 3′-phosphoramidite (the opposite of the directionality of DNA polymerases). The O–P bond formation step creates an internucleotide *phosphite* triester as the diisopropyl amine departs (structure I in Figure 10.13). Conversion of I to II is the oxidation step from trivalent *phosphite triester* to pentavalent *phosphate trimester.* This proceeds with high chemical efficiency but no stereo control such that both R_p and S_p phosphate triesters are formed (More on

this conundrum in Chapter 12). To prepare for the next elongation cycle, species II is deprotected at its remaining 5′-trityl group to free up that 5′-OH as the nucleophile in the next elongation cycle.

After any elongation cycle one could remove the growing deoxynucleotide chain from the solid support and only then remove the cyanoethyl protectant group from each internucleotide phosphotriester, producing the final deoxyribonucleotide with internucleotide phospho*diester* linkage. Each ligation or chain elongation step is iterated during solid phase machine synthesis of oligonucleotides.

References

1. J. Petkowski, W. Bains and S. Seager, Natural products containing ''rare'' organophosphorus functional groups, *Molecules*, 2019, **24**, 866.
2. S. R. Fuhs, J. Meisenhelder, A. Asianian, L. Ma, A. Zagorska, M. Stankova, A. Binnie, F. Al-Obeidi, J. Mauger, G. Kemke, J. R. Yates 3rd and T. Hunter, Monoclonal 1- and 3-phosphohistidine antibodies: new tools to study histidine phosphorylation, *Cell*, 2015, **162**, 198–210.
3. F. Voncken, F. Gao, C. Wadforth, M. Harley and C. Colasante, The phosphoarginine energy-buffering system of Trypanosoma brucei involves multiple arginine kinase isoforms with different subcellular locatons, *PLoS One*, 2013, **8**, e65908.
4. M. J. McLeish and G. L. Kenyon, Relating structure to mechanism in creatine kinase, *Crit. Rev. Biochem. Mol. Biol.*, 2005, **40**, 1–20.
5. Q. Li, S. Fan, X. Li, Y. Jun, W. He, J. Zhou, S. Cen and Z. Yang, Inisghts into the phosphoryl transfer mechanism of the human ubiquitous mitochondrial creatine kinase, *Sci. Rep.*, 2016, **6**, 1–10.
6. A. Bera, T. Wallimann, S. Ray and M. Ray, Enzymes of creatine biosynthesis, arginine annd methionine metabolism in normal and malignant cells, *FEBS J.*, 2008, **275**, 5899–5909.
7. A. Humm, E. Fritsche, S. Steunbacher and R. Huber, Crystal structure and mechanism of human L-arginine:glycine amidinotransferase: a mitochondrial enzyme involved in creatine biosynthesis, *EMBO J.*, 1997, **16**, 3373–3385.
8. K. Severinov, E. Semenova, A. Kazakov, T. Kazakov and M. S. Gelfand, The post-translationally moidified microcins, *Mol. Microbiol.*, 2007, **6**, 1380–1394.
9. S. Rebuffat, Microcins in action: amazing defense strategies in Enterobacteria, *Biochem. Soc. Trans.*, 2012, **40**, 1456–1462.
10. K. Severinov and J. K. Nair, Microcin C: biosynthesis and mechanisms of bacterial resistance, *Future Microbiol.*, 2012, **7**, 281–289.
11. R. F. Roush, E. Nolan, F. Lohr and C. T. Walsh, Maturation of an *E.coli* ribosomal peptide antibiotic by ATP-consuming N–P bond formation in Microcin C7, *J. Am. Chem. Soc.*, 2008, **130**, 3603–3609.

12. L. B. van Alen, C. Q. Wenzel, M. R. Richards, C. Fodor, R. A. Ashmus, M. Stahl, A. V. Karlyshev, B. W. Wren, A. Stintzi, W. G. Miller, T. L. Lowary and C. Szymanski, Biological roles of the O-methyl phosphoramidate capsule modification in *Camphylobacter jejuni*, *PLoS One*, 2014, **9**, e87051.
13. Z. W. Taylor, H. A. Brown, T. Narindoshvilli, C. Q. Wenzel, C. M. Syzmanski, H. M. Holden and F. M. Raushel, Discovery of a glutamate kinse required for the biosynthesis of the O-methyl phosphoramidate modifications found in the capsular polysaccharides of *Camphylobacter jejuni*, *J. Am. Chem. Soc.*, 2017, **139**, 9463–9466.
14. Z. W. Taylor, H. A. Brown, H. M. Holden and F. M. Raushel, Biosynthesis of nucleoside diphosphoramidates in *Camphylobacter jejuni*, *Biochemistry*, 2017, **56**, 6079–6082.
15. Z. W. Taylor and F. M. Raushel, Cytidine diphosphoramidate kinase: an enzyme required for the biosynthesis of O-methyl phosphoramidate modification in the capsular polysaccharides of *Camplylobacter jejuni*, *Biochemistry*, 2018, **57**, 2238–2244.
16. A. Baulig, I. Helmle, M. Bader, F. Wolf, A. Kulik, A. Al-Dilami, D. Wibberg, J. Kalinowski, H. Gross and L. Kaysser, Biosynthetic reconstitution of deoxysugar phosphoramidate metalloprotease inhibitors using an N–P bond-forming kinase, *Chem. Sci.*, 2019, **10**, 4486–4490.
17. H. A. Relyea and W. van der Donk, Mechanism and applications of phosphite dehydrogenase, *Bioorg. Chem.*, 2005, **33**, 171–189.
18. C. Reese, Oligo-and poly-nucleotides: 50 years of chemical synthesis, *Org. Biomol. Chem.*, 2005, **3**, 3851–3868.
19. S. L. Beaucage and R. P. Iyer, Advances in the synthesis of oligonucleotides by the phosphoramidite approach, *Tetrahedron*, 1992, **48**, 2223–2231.

CHAPTER 11

C–P Bonds in Biology: Phosphonates and Phosphinates

11.1 C–P Bonds: Naturally Occurring Phosphonates

The vast majority of phosphorus containing metabolites in biological systems are inorganic phosphate or ester and anhydride derivatives of inorganic phosphate as detailed in Chapters 2–9. In the preceding Chapter 10, we have noted the existence of some 50–60 phosphorus-containing metabolite scaffolds with N–P bonds and detailed some of the common mechanisms for enzymatic formation and phosphoramidate transfer.

There are also a set of phosphorus metabolites, largely but not exclusively in prokaryotic metabolism, with one or two *direct carbon–phosphorus bonds* (C–P bonds). At present, they number ~200 metabolites, with >98% containing a single C–P bond.[1–3] The remaining ~2% are much rarer scaffolds with two C–P bonds. Metabolites with one C–P bond are termed phosphonates. As noted in Figure 11.1 they are derivatives of phosphorous acid (thee oxygens around the central P atom rather than the four oxygens in phosphoric acid). We will note enzymatic formation and breakdown of these C–P containing phosphonates. Analogously, the dialkyl phosphonates with two C–P bonds are formally in the hypophosphorous acid family. Hypophosphorous acid has two hydrogens and two oxygens attached to the central phosphorus atom. The species with one C–P and one H–P is also termed a phosphinate and we will observe that such metabolite scaffolds are found in bacteria that make the antimetabolite phosphinothricin[4] (Figure 11.1).

While direct C–P bonds in metabolites were once thought to be quite rare, the finding that the phosphonolipid phosphatidyl aminoethylphosphonate

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

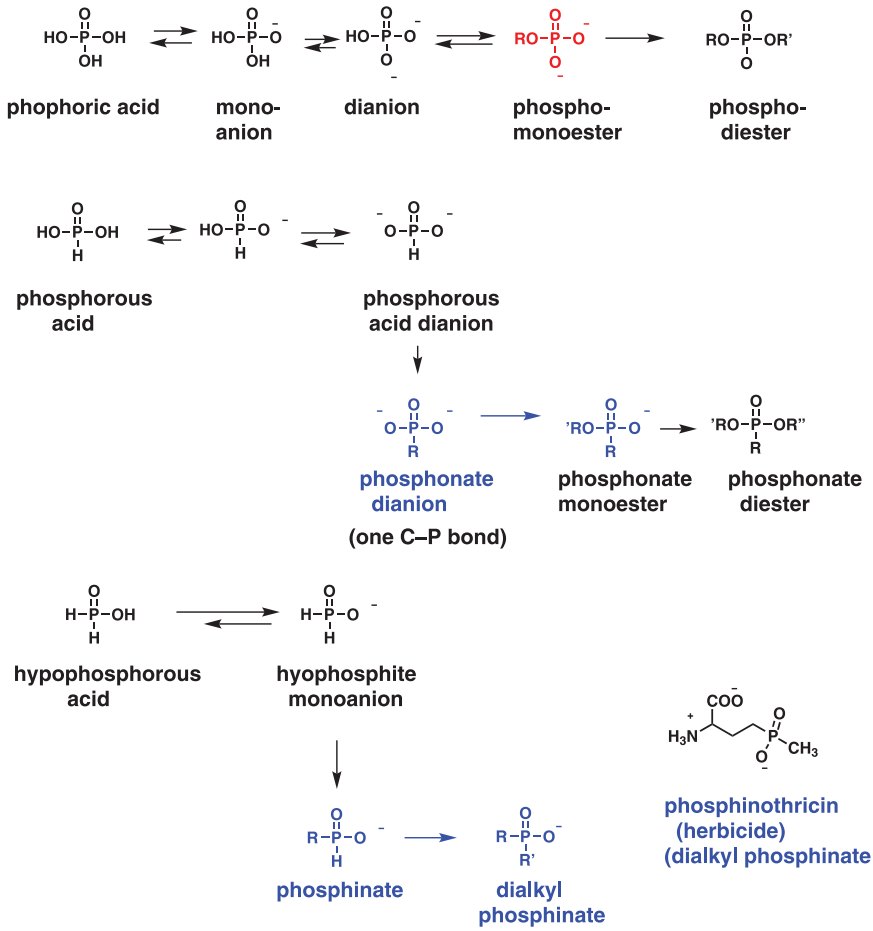


Figure 11.1 Phosphorous acid, hypophosphorous acid, phosphonates and phosphinates.

was a major component of Tetrahymena membranes altered that view.[5] This phosphonate is part of a major membrane phospho(no)-lipid with aminoethylphosphonate replacing the normal aminoethyl phosphate of the common membrane lipid phosphodiester phosphatidylethanolamine (Figure 11.2).

The C–P bond is much more stable chemically than an O–P bond and only a few specialized phosphonate hydrolases exist so this may be a strategy to conserve phosphorus from competitors. Indeed, in many microenvironments, including the open ocean, phosphorus limitation constrains growth.[1] Organisms have gone to great lengths to accumulate phosphorus and keep it from competitors. The C–P bond phosphonate strategy is an effective effort. Of course, some microbes have learned to cleave some C–P bonds to convert the phosphonate to readily usable inorganic phosphate as noted in section 11.6 below.

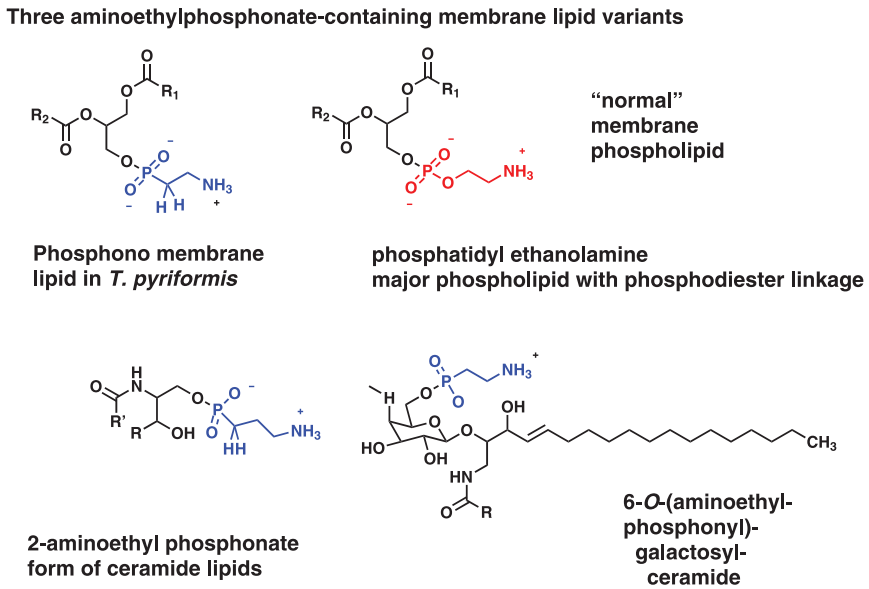


Figure 11.2 Aminoethyl phosphonate diester linkages in lipids with direct C–P bonds.

11.2 PEP Mutase: The Singular Route to C–P Bonds

Some fifty years after phosphonates were detected as biological metabolites, it appears that microbes have evolved only one pathway to construct C–P bonds, all in the form of the single metabolite phosphonopyruvate. The direct C_3–P bond arises from phosphoenolpyruvate (PEP), the thermodynamically activated phosphorylated enol of central metabolism, by the catalytic action of a single enzyme encoded by the *pepM* gene. The enzyme is termed PEP mutase[6,7] (Figure 11.3).

William Metcalfe as microbiology lead and Wilfred van der Donk as chemistry lead of a team at the University of Illinois have led the way in C–P chemical biology – from pathway genetics to biochemistry of the enzymes and determination of both enzyme and metabolite structures that reveal novel chemistry.[8] Their survey of 10 000 actinomycete isolates over the years of collection indicated that 278 strains have the *pepM* gene and associated genes likely to transform phosphonopyruvate into downstream metabolites.[9,10] That is about a 3% abundance rate from those spore-forming bacterial strains, responsible for a wide range of natural product frameworks, including the phosphonate and phosphinate examples noted below.

Much mechanistic and structural analysis has gone into analysis of this single pathway from phosphate ester metabolism to phosphonate metabolism though the narrow catalytic valley of PEP mutase. The consensus mechanism favors (1) retention of stereochemical configuration at the migrating PO_3^{2-} but (2) not a covalent phosphoenzyme intermediate. Instead,

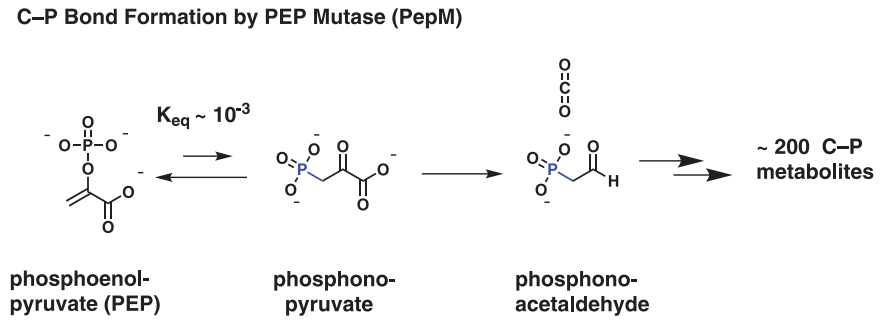


Figure 11.3 The sole enzymatic route to C–P bonds in biology is mediated by members of the PEP mutase family. Some 200 C–P bond-containing metabolites are downstream of PEP mutase.

as shown in Figure 11.4, a dissociative mechanistic route involving early O–P bond cleavage to yield the enolate form of pyruvate and the metaphosphate anion has been proposed.[11] Such a transient metaphosphate anion can either diffuse locally from O_2 to C_3 of the bound enolpyruvate anion or the pyruvate could rotate around the C_1–C_2 bond. The rotation in particular is minor in terms of conformational requirements and would place the C_3 carbon nucleophile (the carbanion resonance contributor to the enolate structure) in place to capture the metaphosphate ion as the C–P bond forms. The product is then phosphonopyruvate.

The equilibrium actually favors the back reaction (phosphonopyruvate to PEP) and has to be pulled in the forward direction by the next enzyme that typically works in tandem with PEP mutase[7,12] (Figure 11.3). That second enzyme is a classic thiamin-pyrophosphate-dependent α-keto acid decarboxylase, yielding CO_2 and phosphonoacetaldehyde (Figure 11.5). As CO_2 volatilizes, the reaction is largely irreversible and draws PEP mass into the phosphonate metabolite universe.

Before turning to some of the noteworthy enzymatic chemistry practiced on the C–P bond of phosphonate metabolites,[13] we make three summary comments on PEP mutase. (1) It is the sole catalyst allowing phosphate flux to phosphonate pathways. (2) PEP mutase appears to catalyze a dissociative phosphoryl transfer, *via* a transient, bound metaphosphate anion: the most likely case in many hundreds to thousands of phosphoryl transfers metabolism. (3) The nucleophile to which the itinerant $-PO_3^{2-}$ is transferred is a carbon nucleophile, the C_3 carbanion contributor to the enolate anion structure. That is the transfer that creates the direct C–P bond.

11.3 Phosphonate Metabolism I

Given the point of entry to phosphonate metabolism is at the three-carbon level, phosphonopyruvate from PEP, and given that the phosphonopyruvate decarboxylase creates an effectively irreversible step that pulls on the equilibrium, three-carbon, two-carbon, and one-carbon scaffolds with C–P bonds

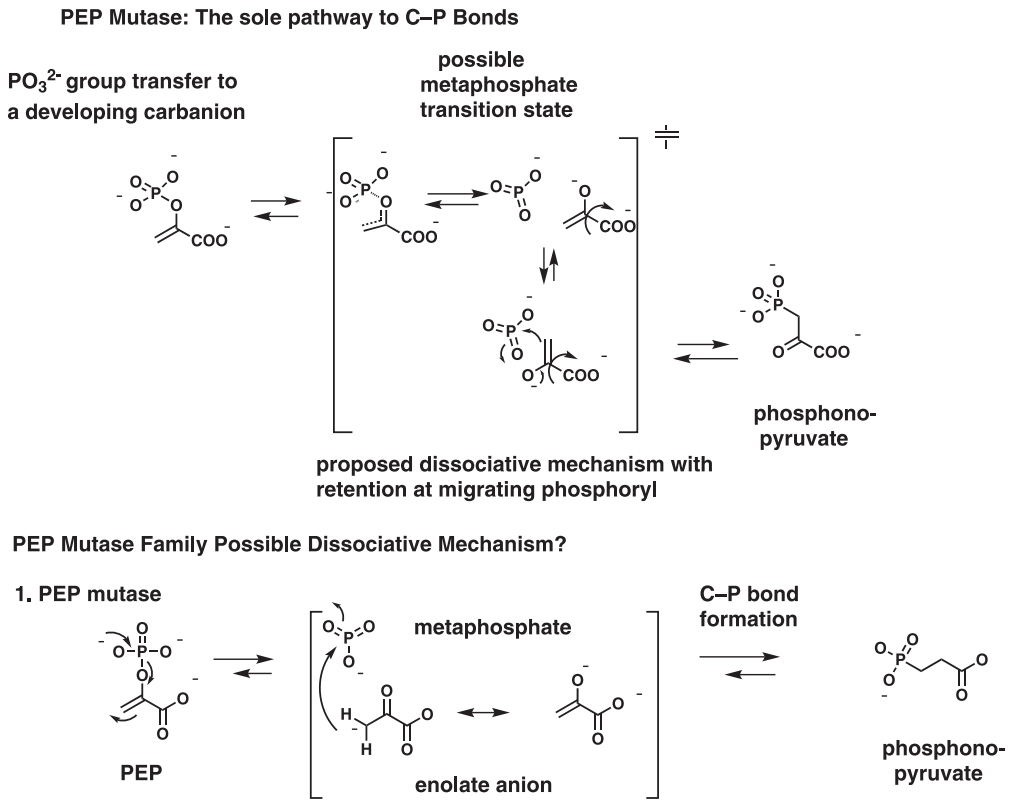


Figure 11.4 The sole enzymatic route to C–P bonds in biology is mediated by members of the PEP mutase family. The reaction mechanism for PEP mutase is formulated as $-PO_3^{2-}$ group transfer to the developing pyruvate carbanion. A dissociative transition state has been proposed.

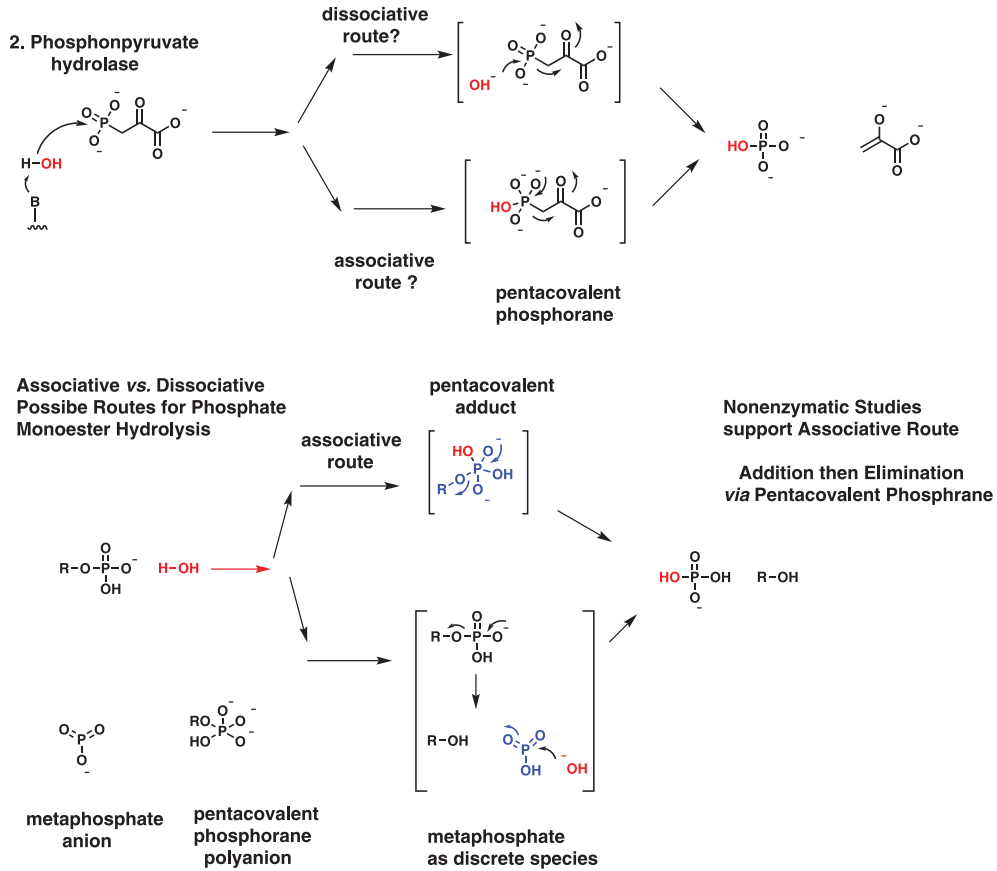


Figure 11.4 (Continued).

are within easy reach of variants of normal metabolic enzymes[5,8,13] (Figure 11.5). For example, transamination of phosphonopyruvate would yield phosphono-alanine (L- or D-). To the extent that the phosphonic acid is a carboxylate mimic, phosphonoalanine may be mistaken for an aspartate analog. To the extent the phosphonate group is misread by enzymes as a phosphate analog, it will be a mimic of phosphoserine. In either case it could be an antimetabolite. We will see in the next section that enzymes process phosphinates as though they were carboxylate groups.

Two-carbon phosphonate scaffolds are available from phosphonoacetaldehyde by two-electron oxidation to phosphonoacetate or two-electron reduction to hydroxyethylphosphonate; both are observed metabolites in PEP mutase-expressing actinomycete bacteria (Figure 11.5).

Conversion of the two-carbon scaffold of hydroxyethyl phosphonate to two one-carbon phosphonate frameworks is carried out by a pair of nonheme iron-based monooxygenases. One of them effects C–C radical cleavage and $OH^{\bullet}$ transfer to the carbon centered methylene phosphonate radical to yield the one carbon hydroxymethyl phosphonate (Figure 11.5). The hydroxymethyl C_2 carbon of starting substrate has been extruded as formate.

The other enzyme with comparable ligands to active site Fe^{2+} carries out the same initial carbon radical chemistry but the intermediate $^{\bullet}CH_2PO_3^{2-}$ radical abstracts an $H^{\bullet}$ from the nascent formate. Oxygen $OH^{\bullet}$ transfer is outcompeted by this hydrogen atom transfer. This enzyme is a ''thwarted oxygenase''. The released products are methyl phosphonate and CO_2 (Figure 11.5). The C_1 of starting hydroxyethyl phosphonate has been oxidized while the C2-CH_2 carbon of substrate has been reduced. Methyl phosphonate is a source of global methane and inorganic phosphate as bacteria extract inorganic phosphate from this unactivated phosphonate scaffold by C–P lyase action (Section 11.5.2).

11.4 Phosphonate Metabolism II: A Route to Phosphinates in Phosphinothricin Biosynthesis: Phosphinates as Carboxylate Isosteres

Some strains of bacteria make the antibiotic phosphinothricin as an inhibitor of glutamine synthetase in susceptible neighboring cells[4,13] (Figures 11.2 and 11.6). The producers make the phosphinate scaffold of phosphinothricin as the *N*-acetyl form to protect themselves from its actions. They also add two alanine residues to the phosphinothricin (Figure 11.6), before removing the *N*-acetyl group and secreting the tripeptide. The phosphinothricin–Ala–Ala tripeptide is taken up by oligopeptide membrane transport proteins in neighboring bacteria. Inside those cells, peptidases hydrolytically convert the tripeptide to two alanine molecules and free phosphinothricin.[8]

This free internalized phosphinothricin, a structural analog of glutamate, inhibits the essential enzyme glutamine synthase, eventually leading to cell

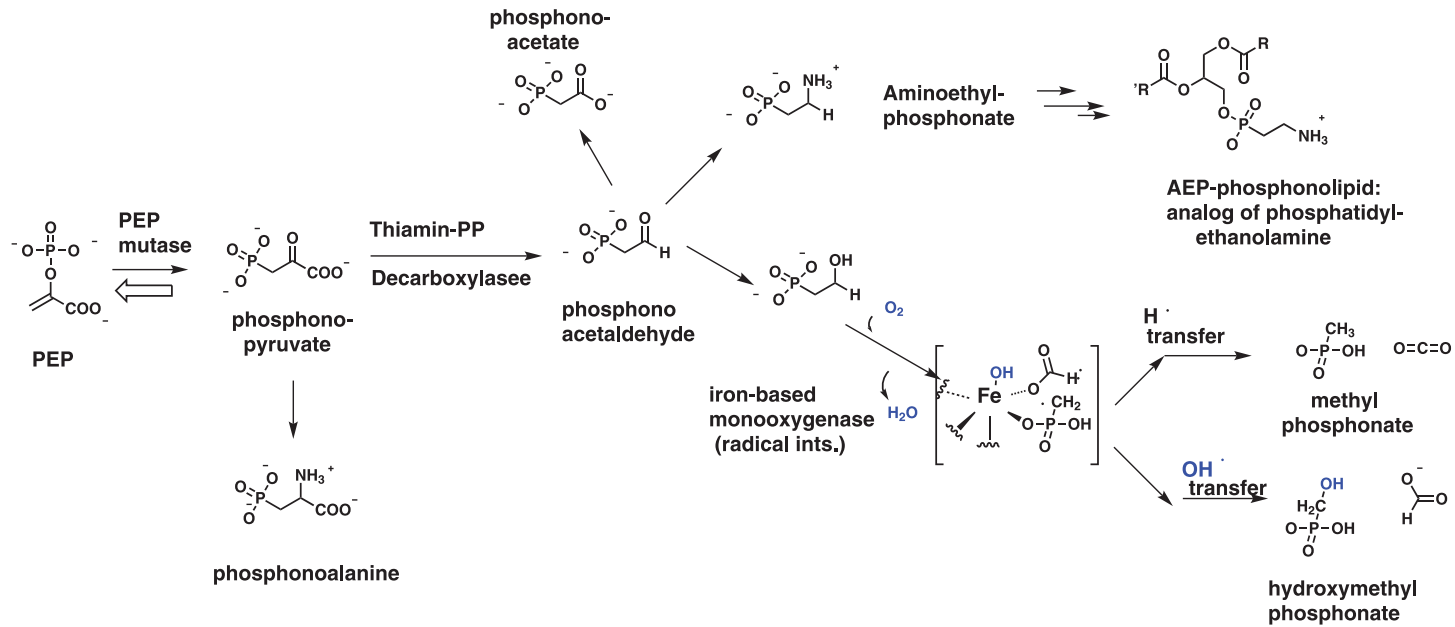


Figure 11.5 The PEP mutase equilibrium is pulled in the direction of phosphonopyruvate by its subsequent enzymatic decarboxylation to phosphonoacetaldehyde. In turn that aldehyde is progenitor to aminoethylphosphonate (AEP), hydroxymethyl phosphonate, and methyl phosphonate as depicted. (Radical ints is an abbreviation for radical intermediates.)

death and the harvesting of the released metabolites by the neighboring phosphinothricin producers. The *phosphinate* $-PO_2^-$ group is a close isostere to the pervasive CO_2^- carboxylate group which accounts for its misrecognition by glutamine synthase as a homolog of the normal glutamate substrate.

Inspection of the phosphinothricin scaffold reveals that the methylphosphinate anionic group is covalently attached to a four-carbon amino acid scaffold. Two challenges (at least) need to be surmounted to convert the three-carbon point of entry phosphonate metabolite phosphonopyruvate to phosphinothricin. One is to elongate the C_3 to a C_4 framework. The second is to construct a second C–P bond to go from phosphonopyruvate to two C–P bonds in the phosphinate framework. Deconvolution of the biosynthetic pathway to phosphinates has unearthed several novel facets of C–P chemical biology.

The biological strategy is to start with phosphonoformate, available by enzymatic double oxidation of hydroxyethyl phosphonate (Figure 11.6). The first remarkable pathway intermediate is a four-carbon carboxy-phosphono-

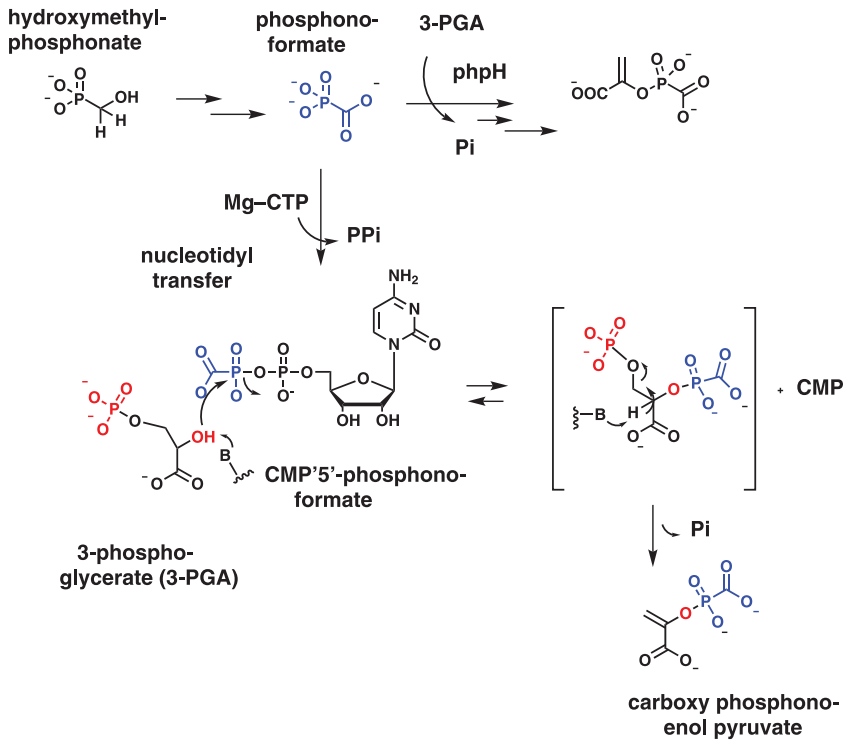


Figure 11.6 Hydroxymethyl phosphonate to carboxy-phosphono-enolpyruvate precedes a second phosphonomutase on the way to phosphinopyruvate. That is followed by three steps of a pseudo-citrate cycle ensemble to generate the C–P–C scaffold of *N*-acetylphosphinothricin.

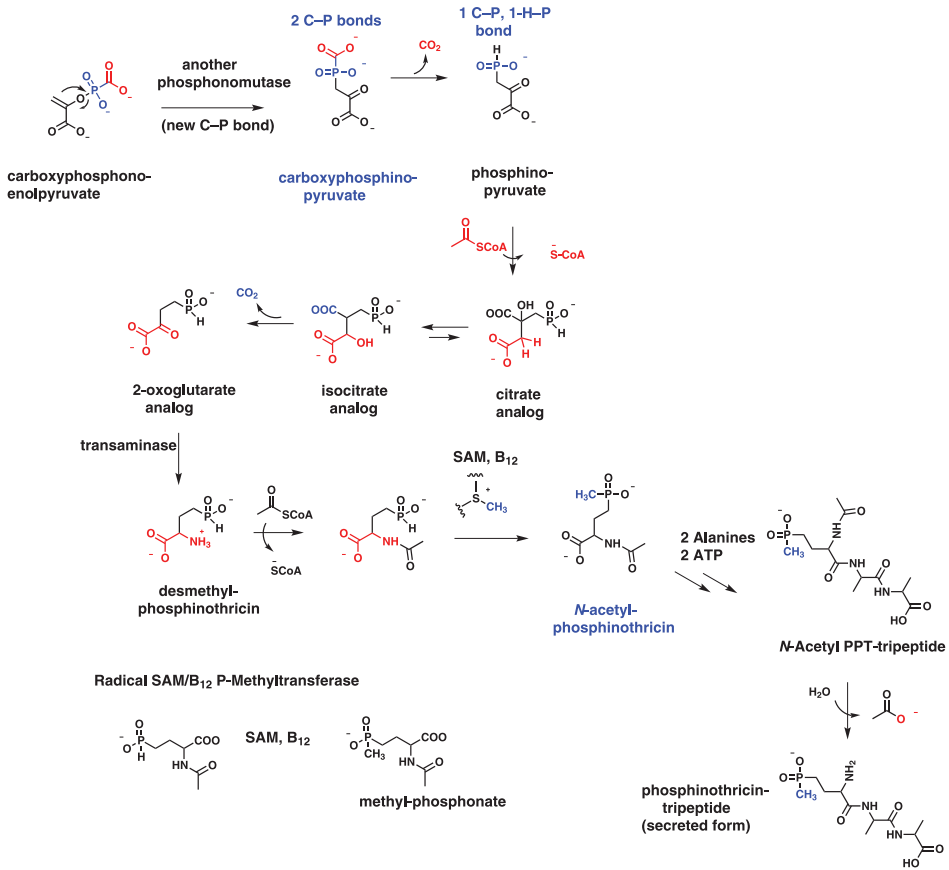


Figure 11.6 (Continued).

enolpyruvate framework.[8,13] The four carbons are not all connected. They are separated by a phosphono–phosphate ester bridge (Figure 11.6).

The logic to get to this metabolite involves cytidine triphosphate and nucleotidyl transfer of the CMP moiety to one of the phosphonate oxyanions of phosphonoformate. This creates the CMP–5′-phophonoformate metabolite. As always, nucleotidyl transfer to phosphate or phosphonate anions creates phosphodiester or phosphono-ester linkages. This nucleotidyl transfer also activates the phosphonyl group of CMP-phosphonoformate as an electrophile in a transfer of the $-PO_2-COO$ fragment to an incoming nucleophile.

The cosubstrate in the next enzymatic step is the 2-OH of the glycolytic central metabolite 3-phosphoglycerate. As depicted in Figure 11.6, the potential transition state can deconvolute in the forward direction to release CMP and generate the 2-phosphonoformate ester of 3-PGA. That undergoes elimination of inorganic phosphate (that had been the 3-phosphate from 3-PGA), creating the double bond of a trapped enolpyruvate. In this case the trapping group is not just $-PO_3^{2-}$ as is the case in phosphoenolpyruvate (PEP), but instead a phosphonoformate moiety.

The novel phosphonate enzymatic chemistry is just beginning. The next enzyme is the second example of an O–P to C–P bond-forming mutase[14] (PEP mutase that is the gateway to all C–P metabolites is the first) (Figure 11.6). Although the mechanism of this phosphonate mutase has not been fully explored, the early dissociation of the O–P bond would yield the pyruvyl enolate with the requisite C_3-carbanionic character. The other dissociated fragment would be a formate derivative of metaphosphate (not shown), that can be captured on phosphorus by the C_3-pyruvyl carbanion. This C–P bond formation, on a pre-existing phosphono phosphorus, creates the phosphinate group in carboxyphosphinopyruvate.

There are four carbon atoms in this metabolite, two of them attached to carbon but the phosphinate group intervenes.[8] One could argue that the function of the next four enzymes is to perform a formal transposition to put the phosphinate at the end of a linear four-carbon chain (Figure 11.6). The strategy is an intriguing reprise of the logic of the citric acid cycle and relies on the phosphinate anion as an isosteric mimic of carboxylate anions to be processed analogously to the way that oxaloacetate is processed in the citric acid cycle.

To get to the phosphinate analog of oxaloacetate, the carboxy phosphinopyruvate is enzymatically decarboxylated with regiospecific loss of the carboxy group attached to phosphorus. The resultant product is phosphinopyruvate with the departing COO^- replaced by a proton. Even though the phosphorus atom now has only one bond to the carbon atom, the new H–P bond keeps the phosphorus in a mono-phosphinate state (see Figure 11.1). This is the first hydrogen–phosphinate grouping in a pathway metabolite, cryptic in the overall process.

The phosphinate anion in phosphinopyruvate is such a good mimic of the carboxylate anion in oxaloacetate, that phosphinopyruvate is taken through

three enzymatic steps paralleling those of the citrate cycle (Figure 11.6). The first enzyme of the cycle, citrate synthase, is mimicked by enzymatic addition of the acetyl group of acetyl CoA to the ketone carbonyl. Addition of these two carbons to the three-carbon backbone of phosphinopyruvate yields the phosphinate analog of citrate. Aconitase-type dehydration and opposite regiochemical rehydration produces the phosphino equivalent of isocitrate. The third enzyme follows the mechanism of isocitrate dehydrogenase, using NAD to accept electrons as the –OH group is oxidized to the ketone, providing an electron sink beta to the carboxylate that will be lost as CO_2. Ketonization of the resulting enolate yields the hydrogen phosphino equivalent of 2-ketoglutarate.

Before considering the final steps in the phosphinothricin pathway let us consider what has been achieved by these four enzymatic steps. There are two decarboxylations. The first generates the hydrogen phosphinate and that is carried all the way through the three steps of the pseudo citrate cycle because that hydrogen phosphinate looks like an ordinary carboxylate. The second decarboxylation step builds the linear four-carbon-2-keto array with a terminal hydrogen phosphinate.

From one perspective, the bacterium starts with a four carbon carboxyphosphinopyruvate and then the pathway machinery throws away two carbons as CO_2. A starting C4-phosphinate goes to a C3-hydrogen phosphinate, then a C5-framework and then the final C4 framework. However, the final C4 framework has all four carbons connected to each other, a remarkable piece of scaffold refashioning in only four enzymatic steps, using the enzyme logic of primary metabolism.

Transamination of the phosphinate analog of 2-oxogutarate yields the C4 amino acid framework of desmethyl phosphinothricin (Figure 11.6). Now, the producing streptomycete bacteria protects itself against the actions of the end point antimetabolite by *N*-acetylation, preventing recognition by its own glutamine synthetase.

The reconstruction of the second C–P bond, from P–H to P–CH_3, is difficult chemistry[13] (Figure 11.6). The catalyst that acts as a phosphorus methyltransferase is a member of the radical *S*-adenosylmethionine (SAM) methyl transferase superfamily. These all have a 4Fe–4S cluster as a one electron redox agent to homolyze the S–C5′-adenosyl bond in SAM bound at one of the corners of the 4Fe–4S redox cube. The 5′-deoxyadenosyl radical ($dA^{\bullet}$) is the radical initiator of specific chemistry in the $>100\,000$ member enzyme family.[15] The exact sequence of steps of transfer of the methyl group from the sulfonium cationic center of SAM to the cobalt atom in the required cofactor vitamin B_{12} is unclear. If a methyl B_{12} species is the proximal donor to the hydrogen-phosphinate phosphorus of desmethyl phosohinothricin it could be a $[CH_3]^+$ equivalent to a nucleophilic phosphorus, or, perhaps less likely, a $[CH_3^{\bullet}]$ transfer to a phosphinate radical. The product, by whatever route, is *N*-acetyl phosphinothricin. There are, again, two C–P bonds in this phosphino-amino acid framework. Tandem addition of two alanines by nonribosomal peptide synthetase machinery is

then followed by *N*-deacetylation since the tripeptide is not the active antimetabolite (Figure 11.6).

In sum, the conversion of phosphonoformate to phosphinothricin reveals novel features of phosphorus chemical biology, including both a C–P bond-forming mutase and a novel C–P methyl transferase. One of the C–P bonds is lost by enzymatic decarboxylation, generating hydrogen phosphinate forms, *e.g.* phosphinopyruvate, that are cryptic in the overall pathway. The phosphinates are such good isosteres of carboxylates that phosphinopyruvate is a mimic of oxaloacetate for processing by citrate cycle logic for C–C bond formations and cleavages that resculpt the carbon scaffold of the maturing phosphinate antimetabolite.

11.5 Cleaving C–P Bonds

Although only a minority of soil and marine actinomycetes (actinomycetales comprise an anaerobic bacterial Order) (~3% of isolates) contain the biosynthetic genes for C–P bond formation and subsequent scaffold modulations, a larger number of bacteria have evolved catalysts to crack the chemically stable C–P bond. Their primary goal is to liberate the inorganic phosphate ion, often a limiting metabolite in many environmental niches.

Thus, there is a great survival premium on microbes that can cleave phosphonate metabolites described above, such as phosphonoacetaldehyde, phosphonoacetate, and phosphonoformate (Figure 11.7). Given the abundance of microbial phosphonolipipds containing the aminoethylphosphonate head group, enzymes that cleave the C–P bond in aminoethylphosphonate and its metabolites are relatively widespread. Additionally, the methylphosphonate, which we noted above as a metabolite from the two-carbon hydroxyethylphosphonate, is acted on by many bacteria to yield the usable inorganic phosphate for macromolecular and primary metabolite constructions. They discard the coproduct methane as a volatilizing gas (Figure 11.7).

Three different enzymatic routes, employing distinct strategies to cleave the otherwise chemically and hydrolytically stable C–P bond have been recognized. As set forth in Figure 11.7, the first group comprise enzymes known as phosphonatases. The second C–P cleaving catalyst is a multiprotein complex known as C–P lyase. The most recent, third entry is a pair of nonheme iron, O_2-consuming enzymes that act in tandem, specifically on the aminoethylphosphonate product available from phosphonolipid hydrolysis by phospholipase D family members (Chapter 8).

11.5.1 Phosphonatase Acting on Phosphonoacetaldehyde

The best studied of the phosphonatase subclass of C–P cleaving enzymes acts on phosphonoacetaldehyde[16] (the decarboxylation product of the point of entry metabolite phosphonopyruvate). This enzyme uses an active site lysine side chain amine to engage in iminium catalysis with the aldehyde group of the phsphonoacetaldehyde. This is preamble to C–P cleavage. The

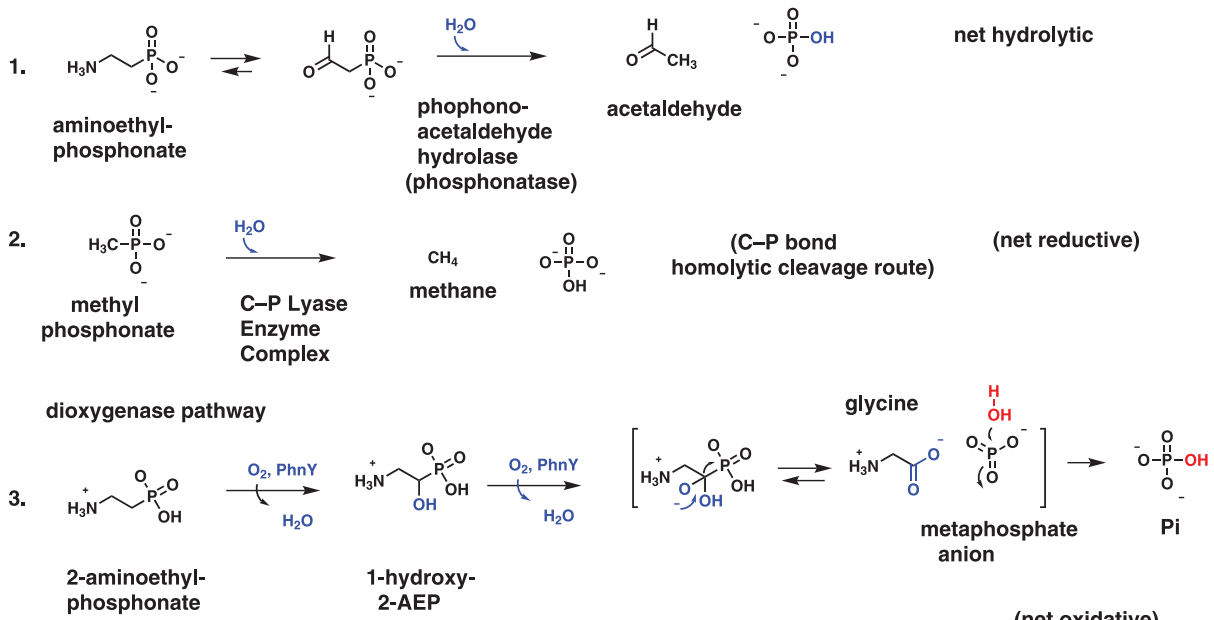


Figure 11.7 Three distinct variants of chemical logic in enzymes that carry out net hydrolysis of C–P bonds.

iminium group serves as an electron sink to take up the electrons released as a cosubstrate nucleophile next attacks the phosphorus atom and the C–P bond cleaves[17] (Figure 11.8).

The kinetically competent nucleophile attacking the phosphonoacetaldehyde imine is not water, but instead the side chain β-carboxylate of an active site aspartyl residue. This O–P bond forming step is also the C–P bond cleaving step, possibly *via* a transition state with some degree (100% of equivalent O–P formation with C–P breaking would be the concerted end of the reaction spectrum) of pentacovalent phosphorane formation and then breakdown.

The immediate result is creation of *two covalently bound enzyme-product fragments.* The two-carbon acetaldehyde moiety is the enamine adduct. Its ketonization to the lower energy ketimine form is the step that converts the CH_2 of starting material to the CH_3 in the acetaldehyde product. Hydrolysis of the ketimine isomeric adduct releases free acetaldehyde. Meanwhile the $-PO_3^{2-}$ group attached to the enzyme carboxylate now has four oxygen bonds to phosphorus. It is a covalent phosphoryl enzyme: more specifically, an acyl phosphate, a mixed anhydride between phosphoric acid and the enzyme aspartate carboxylate residue (see Chapter 14). Its subsequent hydrolysis is downhill energetically and is the step that introduces a solvent oxygen into the product inorganic phosphate. A remarkable solution to cracking a stable C–P bond: covalent iminium ion and acyl phosphate covalent product fragment strategies in a single active site.

11.5.2 C–P Lyase: Methyl Phosphonate Fragmentation to Methane and Inorganic Phosphate

Microbes that express the phosphonoacetaldehyde phosphonatase also usually express a second enzyme with broader specificity for alkyl phosphonate fragmentation to the parent alkane and inorganic phosphate. The prototypic substrate is methyl phosphonate, whose metabolic origin from hydroxyethyl phosphonate we noted above in Figure 11.6. Early studies established that C–P lyase activity required a set of adjacent genes to be turned on, indicating multiple protein players. Furthermore, early stereochemical studies on alkyl phosphonates gave racemic products consistent with a radical mechanism for C–P bond cleavage by the C–P lyase complex.

Eventually, Frank Raushel's research group at Texas A&M deciphered the logic and novel chemistry utilized by C–P lyase protein components to set up methyl phosphonate for C–P bond cleavage.[18,19] The pathway starts, as so many involving phosphorus chemical biology, with attack on ATP by a phosphate oxyanion (Figures 11.9 and 11.10). In this case the oxyanion is from the methylphosphonate. Most remarkable is that the site of attack in ATP is not any of the electrophilic Pα, Pβ, or Pγ anhydride linkages in the ATP side chain, nor even the C5′ carbon for adenosyl transfer.

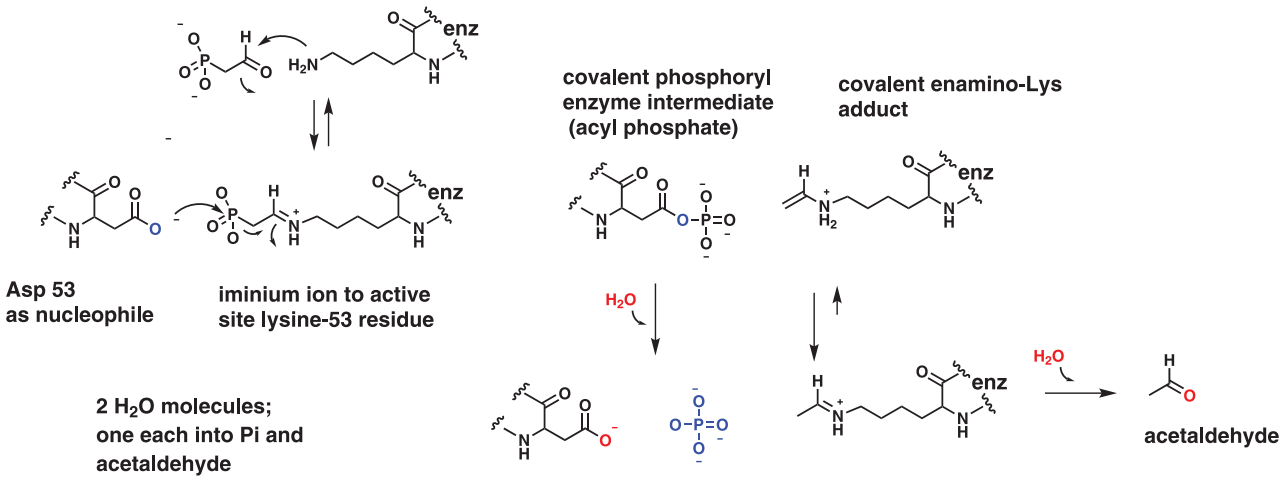


Figure 11.8 The phosphonatase that cleaves the C–P bond in the metabolite phosphonoacetaldehyde uses the aldehyde of that substrate to set up iminium ion catalysis to lower the energy barrier to C–P hydrolytic cleavage, while generating a covalent acyl phosphate–enzyme adduct.

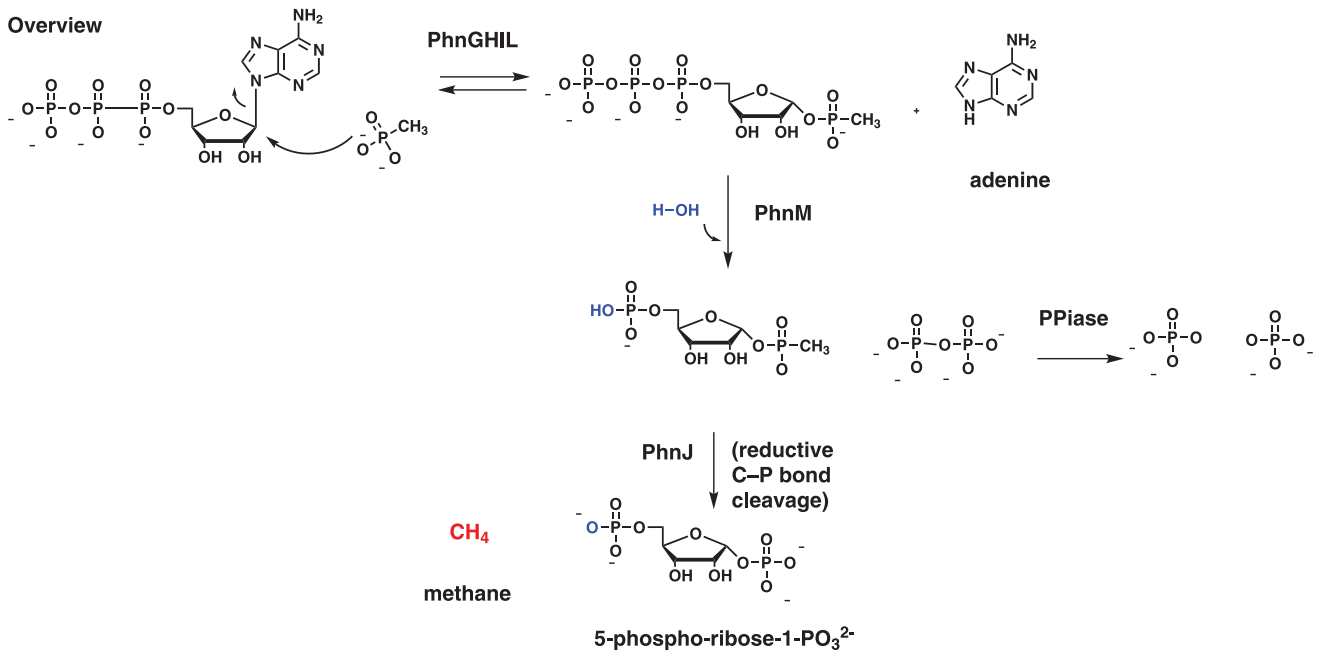


Figure 11.9 C–P lyases cleaving unactivated alkyl phosphonates such as methyl phosphonate to inorganic phosphate and methane are multi-enzyme complexes that use radical chemistry for the C–P cleavage step. The oxyanion of methylphosphonate attacks cosubstrate ATP at the C1′-ribose carbon in a remarkable ribosyl triphosphate transfer to yield PPi and 5′phospho-ribose-1-methylphosphate ester.

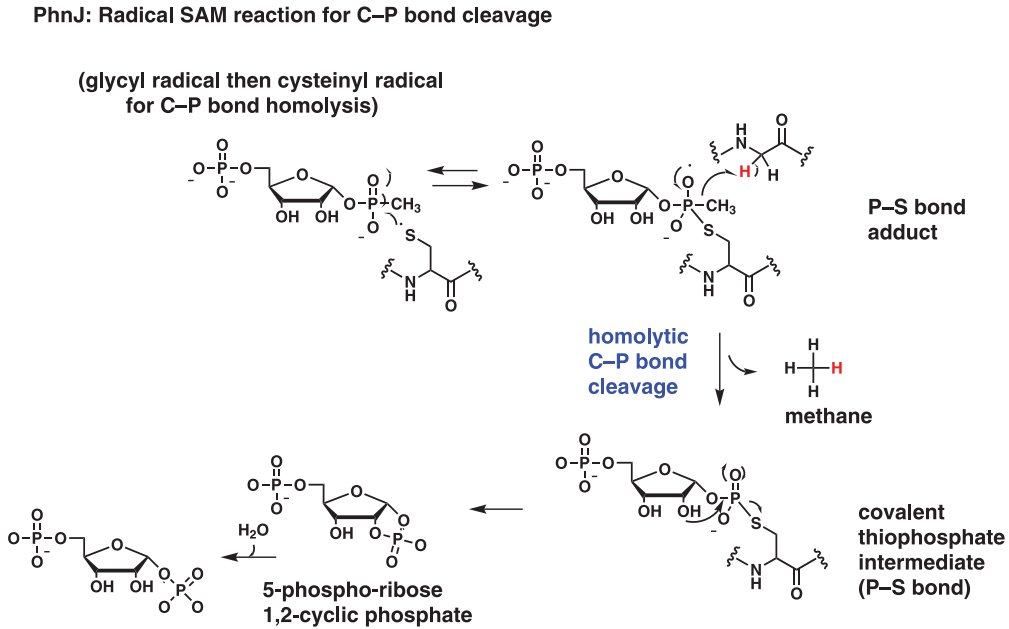


Figure 11.10 C–P lyases cleaving unactivated alkyl phosphonates such as methyl phosphonate to inorganic phosphate and methane are multi-enzyme complexes that use radical chemistry for the C–P cleavage step. This metabolite undergoes radical SAM-mediated reductive cleavage to methane and a covalent 5′phospho-ribosyl-1-thiophosphoryl enzyme adduct that unravels to release 5′-phospho-ribose-1,2-cyclic phosphate.

Instead, the methyl phosphonate attack is on C1′ of the ribose moiety: a net 5′-triphosphoribosyl transfer, ejecting the free adenine base. To date this is a *unique* fragmentation mode for ATP in all the hundreds of transformations that have been detected in cellular metabolism. A summarized in Figure 11.9, the mechanism of this unusual transfer is not yet deciphered and requires the four enzymes PhnG, H, I and L, probably acting as a complex.[20] The next enzyme PhnM is a hydrolase, also categorizable as a variant of a pyrophosphatase, as it cleaves between Pα and Pβ to release the Pβ–Pγ fragment as pyrophosphate (cleavable in tandem by inorganic pyrophosphatases in the cell). The structure of a four protein *E. coli* C–P lyase core unit (PhnGHIJ) has been determined by X-ray analysis.[21]

The coproduct from PhnM action is the 5′-phospho-ribose-1′ methylphosphonate, a classical phosphate monoester at the 5′-carbon and an unusual phosphonate ester at the 1′-carbon. This is the substrate for PhnJ that acts as the C–P lyase core catalyst. As shown in Figure 11.10, an active site cysteine radical is proposed to add into the 1′-phosphorus atom to create a transient pentacovalent thiophosphorane radical adduct.

The source of the cysteine radical is ultimately from homolytic cleavage of the sulfur-5′-adenosyl *S–C* bond in cosubstrate *S*-adenosylmethionine (SAM). PhnJ is a member of a large family of radical SAM enzymes,[15] so called because a one electron input to an enzyme bound 4Fe–4S cluster leads to homolytic scission of the SAM coordinated to one of the edges of the 4Fe–4S cube. The 5′-deoxyadenosyl radical (dA$^{\bullet}$) fragment initiates one electron reaction flux in the active site. In PhnJ, the first enzyme radical formed is a glycyl backbone radical *via* H$^{\bullet}$ abstraction by dA$^{\bullet}$. Subsequently, the glycyl radical abstracts an H$^{\bullet}$ from an active site cysteine and that sulfur radical is the one electron nucleophile at phosphorus.

The P=O double bond formally gives up one electron to the incoming sulfur radical to create the new S–P bond, while leaving the unpaired electron on the phosphoryl oxygen as depicted (Figure 11.10). Collapse of that proposed phosphono-thiyl-oxy radical is hypothesized to be the actual P–CH_3 cleavage step *via* a one-electron reaction manifold. The abstraction of the H$^{\bullet}$ from the glycyl backbone may be concerted with C–P bond cleavage. The CH_3 moiety has become CH_4, methane, as it leaves, [picking up one of the glycine active site residue methylene hydrogens (as H$^{\bullet}$)]. The enzyme-glycyl radical has been regenerated and can run the next catalytic cycle. Or it can get an H$^{\bullet}$ back from dA–H and regenerate the initial 5′-deoxy radical (dA$^{\bullet}$). In turn, that could reform the SAM bond such that SAM acts only catalytically. Or the dA$^{\bullet}$ could pick up another H$^{\bullet}$ from some external electron donor. Then, SAM would be cleaved to methionine and dAH in every catalytic cycle.

In passing, we note that radical SAM enzymes such as PhnJ are exquisitely sensitive to inactivation by air. The O_2 scavenges radicals (by undergoing facile one-electron reduction to the superoxide radical anion) and destroys many Fe–S centers. Thus, C–P lyase is an anaerobic nanomachine for C–P disconnection.

The enzyme at this point is still in mid catalytic cycle. It has achieved C–P bond cleavage and CH_4 is released. However, the phosphorus moiety at the 1′-carbon of the 5′-phospho-ribose scaffold is now both a *thiophosphoryl covalent enzyme adduc*t and a hybrid oxo/thio-phosphodiester. This thiophosphoryl bond is unusual but has analogies in other phosphoproteomics cases, as discussed in Section IV.

Decomposition of the thiophosphoryl enzyme is proposed to result from intramolecular attack of the ribosyl-2′-OH on the 1′ thiophosphoryl group. Breakdown in the forward direction then ejects the enzyme-thiolate moiety (two-electron chemistry) and produces the 5-phospho-ribose-1,2-cyclic phosphate. In a subsequent enzymatic step regioselective action of a phosphodiesterase converts that cyclic phosphodiester to the 5-phospho-ribose-1-phosphate. Bacteria also have an ATP-dependent kinase that converts the 1-phosphate into the 1-pyrophosphate group. This is PRPP, the key activated form of ribosyl groups in nucleotide metabolism.

In sum, the cleavage of the chemically stable C–P bond in the unactivated methyl phosphonate has required some inventive chemistry. ATP is cleaved to give *adenine*, with the 5′-triphospho-ribosyl moiety transferred onto the methyl phosphonate oxyanion. The C–P bond is ultimately cleaved homolytically, consistent with the stereo-random outcome at the itinerant alkyl group. The radical reaction manifold is set in motion by homolysis of the $C–S^+$ bond in *S*-adenosylmethionine on the way to a kinetically and thermodynamically accessible active site thiolate radical. Three notable features are: (1) the one-electron nucleophilicity of the sulfur side chain, (2) the cleavage of the C–P bond in the one-electron reaction manifold, and (3) the formation of the covalent thiophosphoryl adduct between enzyme and phosphoester nascent product. The participation of the 2-OH for intramolecular chemistry on ribose scaffolds is mirrored elsewhere in metabolism, including NAD^+-cleaving sirtuins, and also ribonucleases.

11.5.3 Oxygenative C–P Bond Cleavage

The third route that bacteria have evolved for C–P cleavage is oxygenative and to date only applies to the common aminoethylphosphonate (AEP) substrate. Tandem oxygenation is the strategy at the methylene carbon adjacent to phosphorus[22] (Figure 11.7, line 3). The first monooxygenation yields the 1-amino 2-hydroxyphosphonate framework. Tandem action of a second oxygenase is proposed to get to a 1-amino-2-dihydroxyphosphonate intermediate as shown. This can collapse to the two-carbon product glycine (one of the observed products), with proposed extrusion of a metaphosphate anion. Capture of the nascent metaphosphate by water gives the other observed product, inorganic phosphate. The driving force for C–P bond cleavage may be the gain in resonance energy as the carboxylate anion of glycine forms. This enzyme would join the PEP and phosphonopyruvate mutases as catalysts with proposed metaphosphate type mechanisms (Figure 11.11).

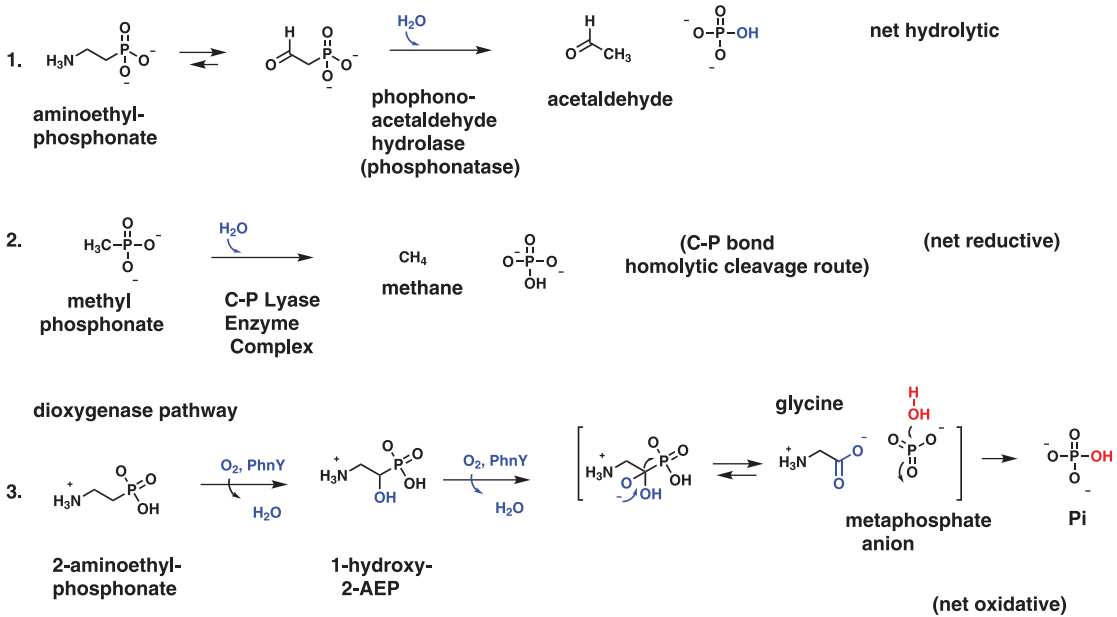


Figure 11.11 Three enzymes involving phosphonate substrates where dissociative transition states and metaphosphate ions could be involved.

The three enzymatic strategies found in contemporary microbes that can mineralize organophosphate (phosphonate) frameworks represent three distinct catalytic approaches to cleavage of the strong, stable C–P bond. In the phosphonoacetaldehyde phosphonatase case, the enzyme uses iminium ion formation to create the electron sink that lowers the energy barrier for C–P cleavage. In the C–P lyase case with unactivated alkyl phosphonates there is no chemical handle on the carbon fragment. The big gun of cellular anaerobic radical chemistry at unactivated carbon chains, *S*-adenosylmethionine coordinated to a 4Fe–4S center, is wheeled into play. A homolytic reaction path occurs with cysteine thiol radical participation, and covalent thiophosphoryl enzyme adduct formation as the immediate result of C–P homolysis. The third approach is to use molecular oxygen as cosubstrate, twice, to activate the carbon adjacent to the phosphorus atom (the C–P) bond by successive oxidation of that carbon. The CH_2 of substrate AEP goes to the CHOH and then C–$(OH)_2$ state. At that four-electron oxidized state, there is enough driving force in rearrangement of C–$(OH)_2$ to COO^- to enable C–P bond cleavage. This proposal involves transient metaphosphate anion.

The three enzymatic routes to C–P bond cleavage reveal a whole array of chemical strategies. Electrophilic catalysis lowers energy barriers when the substrate permits. Anaerobic radical SAM generators create radical paths for homolytic C–P cleavage at unactivated phosphonates. Finally, the aerobic route consumes two molecules of O_2 as the carbon attached to phosphorus gets oxidized by four electrons to set up heterolytic cleavage to a metaphosphate species. The aggregated chemical creativity shows how valuable or essential supplies of inorganic phosphate are to microbes.

References

1. J. F. Villareal-Chiu, J. P. Quinn and J. W. McGrath, The genes and enzymes of phosphonate metabolism by bacteria and their distribution in the marine envronment, *Front. Microbiol.*, 2012, **3**, 19.
2. X. Yu, J. Doroghazi, S. C. Janga, J. K. Zhang, B. Circello, B. M. Griffin, D. P. Labeda and W. W. Metcalfe, Diversity and abundance of phosphonate genes in nature, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 20759–20764.
3. J. Petkowski, W. Bains and S. Seager, Natural products containing ''rare'' organophosphorus functional groups, *Molecules*, 2019, **24**, 866.
4. J. A. Blodgett, P. M. Thomas, G. Li, J. E. Velasquez, W. A. van der Donk, N. Kelleher and W. W. Metcalf, Unusual transformataion in the biosynthesis of the antibiotic phosphinothricin tripeptide, *Nat. Chem. Biol.*, 2007, **3**, 480–485.
5. G. P. Horsman and D. L. Zechel, Phosphonate biochemistry, *Chem. Rev.*, 2016, **117**, 5704–5783.
6. E. Bowman, M. McQueeny, J. D. Scholten and D. Dunaway-Mariano, Catalysis and thermodynamics of the phosphoenolpyruvate/phosphonopyruvate rearrangement. Entry into the phosphonate class of naturally

occuring organophosphorus compounds, *J. Am. Chem. Soc.*, 1988, **110**, 5575–5576.
7. J. Kim and D. Dunaway-Mariano, Phosphoenolpyruvate mutase catalysis of phosphoryl transfer in phosphoenolpyruvate: kinetics and mechanism of phosphorus-carbon bond formation, *Biochemistry*, 1996, **35**, 4628–4635.
8. W. W. Metcalf and W. van der Donk, Biosynthesis of phosphonic and phosphinic acid natural products, *Annu. Rev. Biochem.*, 2009, **78**, 65–94.
9. X. Yu, J. Doroghazi, S. C. Janga, J. K. Zhang, B. Circello, B. M. Griffin, D. P. Labeda and W. W. Metcalf, Diversity and abundance of phosphonate biosynthetic genes in nature, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 20759–20764.
10. K.-S. Ju, J. Gao, J. R. Doroghazi, K.-K. A. Wang, C. J. Thibodeaux, S. Li, E. Metzger, J. Fudala, J. Su, J. K. Zhang, J. Lee, J. P. Cioni, B. S. Evans, R. Hirota, D. P. Labeda, W. A. van der Donk and W. W. Metcalf, Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 12175–12180.
11. S. Liu, Z. Lu, Y. Jia, D. Dunaway-Mariano and O. Herzberg, Dissociative phosphoryl transfer in PEP mutase catalysis: structure of the enzyme/sulfopyruvate complex and kinetic properties of mutants, *Biochemistry*, 2002, **41**, 10270–10276.
12. S. C. Peck, J. Gao and W. A. van der Donk, Discovery and biosynthesis of phosphonate and phosphinate natural product, *Methods Enzymol.*, 2012, **516**, 101–123.
13. S. C. Peck and W. van der Donk, Phosphonate biosynthesis and catabolism: a treasure trove of unusual enzymology, *Curr. Opin. Chem. Biol.*, 2013, **17**, 580–588.
14. T. Hidaka, S. Imai, O. Hara, H. Anzai, T. Murakami, K. Nagaoka and H. Seto, Carboxyphosphoenolpyruvate phosphonomutase, a novel enzyme catalyzing C-P bond formation, *J. Bacteriol.*, 1990, **172**, 3066–3072.
15. H. J. Sofia, G. Chen, B. G. Hetzler, J. F. Reyes-Spindla and N. E. Miller, Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods, *Nucleic Acids Res.*, 2001, **29**, 1097–1106.
16. M. C. Morais, G. Zhang, W. Zhang, D. B. Olsen, D. Dunaway-Mariano and K. N. Allen, X-ray crystallographic and site-directed mutagenesis analysis of the mechanism of schiff-base formation in phosphonoacetaldehyde hydrolase catalysis, *J. Biol. Chem.*, 2004, **279**, 9353–9361.
17. A. S. Baker, M. J. Cioci, W. W. Metcalf, J. Kim, P. C. Babbit, B. L. Wanner, B. M. Martin and D. Dunaway-Mariano, Insights into the mechanism of catalysis by the P-C bond-cleaving phosphonoacetaldehyde hydrolase derived from gene sequence analysis and mutagensis, *Biochemistry*, 1998, **37**, 9305–9315.

18. S. S. Kamat and F. M. Raushel, The enzymatic conversion of phosphonates to phosphate by bacteria, *Curr. Opin. Chem. Biol.*, 2013, **17**, 589–596.
19. S. S. Kamat and F. M. Raushel, PhnJ-a novel radical SAM enzyme from the C-P lyase complex, *Perspect. Sci.*, 2015, **4**, 32–37.
20. P. Seweryn, L. B. Van, M. Kjeldgaard, C. J. Russo, L. A. Passmore, B. HOve-Jensen, B. Jochimsen and D. E. Brodersen, Structural insights into the bacterial carbon-phosphorus lyase machinery, *Nature*, 2015, 68–72.
21. P. Seweryn, L. B. Van, M. Kjelgaard, C. J. Russo, L. A. Passmore, B. Hove-Jensen, B. Jochimsen and D. E. Brodersen, Structural insights into the bacterial carbon-phosphorus lyase machinery, *Nature*, 2015, **525**, 68–72.
22. L. J. Rajakovich, M.-E. Pandelia, A. J. Mitchell, W.-c. Chang, B. Zhang, A. K. Boal, C. Krebs and J. M. Bollinger Jr., A new microbial pathway for organophosphonate degradation catalyzed by two previosuly misannotated non-heme iron oxygenases, *Biochemistry*, 2019, **58**, 1627–1647.

CHAPTER 12

P–S Bonds: Phosphorothioates

12.1 Phosphorothioates/Thiophosphates

As one proceeds from phosphorus compounds with four oxygens bound to phosphorus (phosphates) to one C–P bond (phosphonates [–C–PO_3^{2-}]), to one N–P bond (phosphoramidates [–N–PO_3^{2-}]), the number of natural metabolites in those classes decreases. For the S–P bond-containing phosphorothioates (also known as thiophosphates), the number of low molecular weight metabolites drops towards unity. The icthyotoxin from the Florida-dwelling red tide dinoflagellate *Karenia brevis* produces a phosphorothione that also has a hydrazone moiety[1] (also termed a phosphorohydrazidothioate) as shown in Figure 12.1. In this molecule the phosphorus atom has both a nitrogen and a sulfur attached, so it is a phosphorothioamidate diester. The biosynthetic route to this rare P–S metabolite is totally undetermined at present.

A brief comment about nomenclature. One descriptor of the monosulfur analog of phosphoric acid (H_3PO_4) is thiophosphoric acid (H_3PO_3S) (Figure 12.1). In turn, the mono and dianions at physiological pH are inorganic thiophosphate monoanions and dianions. The alternative nomenclature is phosphorothioate. Some nerve gases and insecticides have two P–S bonds. They are phosphorodithioates. (Figure 12.1) We will use the phosphorothioate terminology preferentially as much of the original literature is couched that way. The rare naturally occurring stretches of oligonucleotides in some microbial DNAs are 3′,5′-phosphorothioate diesters in place of the normal 3′,5′-phosphodiester internucleotide linkages.

12.2 Inorganic Thiophosphate: Not a Known Metabolite

Inorganic phosphate with one oxygen replaced by a sulfur atom substituent is inorganic thiophosphate (Figure 12.1). Despite its utility as a synthetic

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

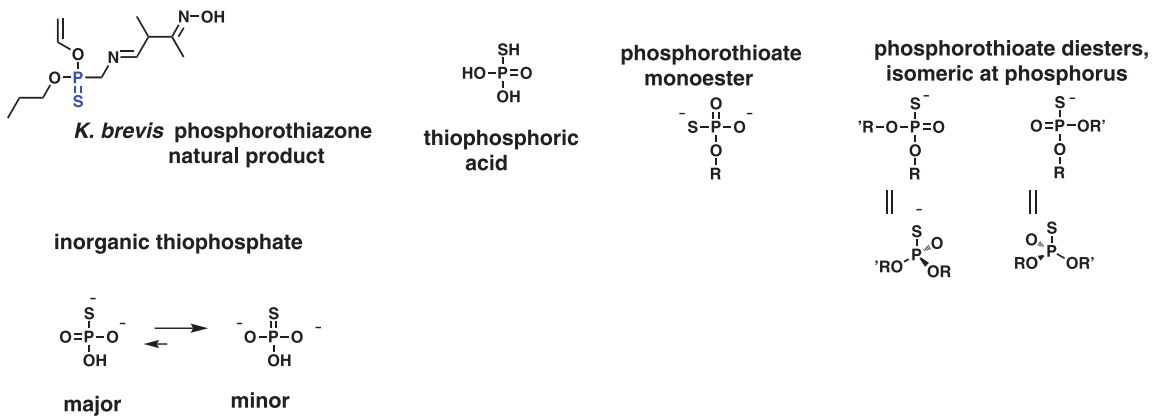


Figure 12.1 P–S compounds are phosphorothioates. The ichthyotoxin from *K. brevis* is the sole known P–S low molecular weight natural metabolite.

analog of inorganic phosphate, neither the thiophosphoric acid nor the thiophosphate anionic forms are known biologic metabolites. Thiophosphate chemical biology that is discussed in this chapter has been most useful in two realms. One realm has been in the arena of synthetic probes of phosphoryl and nucleotidyl transferase mechanisms. The second has been in the chemical syntheses of antisense synthetic oligonucleotide (mostly DNA oligonucleotides but RNA oligonucleotides as well.)

When a P–S bond is present, it has the opportunity to be represented either as a P–S^- single bond or as a P=S resonance form. Although in many early publications, the P=S contributor in phosphorothioates was shown as the major species, different types of experimental observation indicate that the P–S^- single bond contributor dominates the population[2] (Figure 12.1). We therefore show inorganic thiophosphates at neutral pH as the P–S^- dianion form. Analogously, the single P–S bond is shown in thiophosphoryl monoesters and diesters.

12.3 Phosphorothioate Monoesters and Diesters

Phosphorothioate monoesters are likewise not normal metabolic species (Figure 12.2). However, the ability to chemically synthesize nucleoside triphosphates and the corresponding 2′-deoxynucleoside triphosphates, with one sulfur atom at either Pα, Pβ, or Pγ (Figure 12.2) opened up the study of reaction mechanism of kinases, nucleotidyl transferases, and even pyrophosphoryl transferases with phosphorothioate analogs as stereochemical probes.

One of the consequences of replacing one of the two chemically equivalent prochiral oxygens in a phospho-monoester, diester, or phosphoric anhydride linkage is the conversion of prochiral phosphoryl groups to chiral phosphorus centers. The chirality convention of *R* (*rectus* = right-handed = clockwise) and *S* (*sinister* = left-handed = counterclockwise) for assignment of stereochemistry at other chiral centers, such as tetrahedral carbons, obtains also for phosphates. An R_P chiral phosphorus has the four substituents arranged, in decreasing priority of atomic number or weight, in a clockwise direction. The S_P designation has those substituents describing a counterclockwise direction.

12.4 Stereochemical Inversion at Phosphorus in Enzymatic Nucleotidyl Transfers

Figure 12.3 depicts two phosphorothioate diastereomers, the S_P and R_P forms respectively of a dG–dA–(phosphorothioate)-dinucleotide. The ability to conduct unambiguous chemoenzymatic syntheses of ATP-α–S, ATP-β–S and ATP-γ–S as well as regioselective phosphorothioate linkages in congener nucleotides opened up the ability to analyze phosphoryl and nucleotidyl transfers for stereochemistry in the 1970s and 1980s.[3–5] The designations

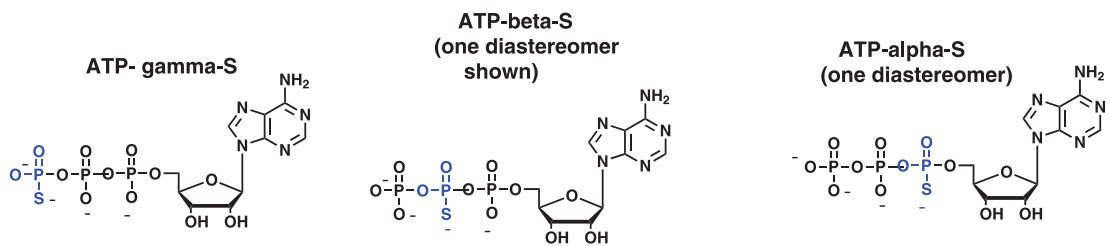


Figure 12.2 Synthetic analogs of ATP with P–S bonds at Pγ, Pβ, or Pα are ATPγ–S, ATPβ–S, and ATPα–S respectively. ATPβ–S and ATPα–S exist as S_P and R_P diastereomers while ATPγ-S is prochiral.

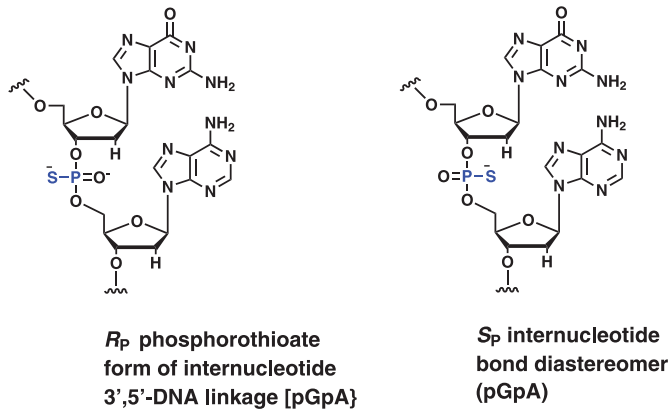


Figure 12.3 R_P and S_P diastereomers of the dinucleotide pG–(S_P)A and pG–(R_P)A.

ATP-α–S, ATP-β–S and ATP-γ–S (Figure 12.2) are shorthand for the α-thiophosphoryl group in the triphosphate side chain, the β-thiophosphoryl group in the triphosphate side chain, or the γ-thiophosphoryl in the triphosphate side chain of sulfur-substituted ATPs, respectively. Sulfur at Pα or Pβ in place of one of the P–O bonds creates the two diastereomers (R_P or S_P) at that phosphorus.

In contrast, a thiophosphoryl group at the terminal Pγ of ATP-γ–S still leaves two prochiral oxygens in the $-PSO_2^{2-}$ moiety. That γ-thiophosphoryl in ATP-γ–S is itself now prochiral (Figure 12.2). It is necessary to distinguish the remaining two oxygens before that phosphorus center can report on stereochemistry of any (thio)phosphoryl transfers. That could be achieved by replacing O16 with a heavy O^{18} atom. The (–P–S, O^{16}, O^{18}) thiophosphoryl group is chiral, has both R_P and S_P isomers, and needs to be prepared to give only one diastereomer at a time to be useful for stereochemical determination of the fate of such a thiophosphoryl group in flight. The determination of whether a migrating (thio)phosphorus center moved with inversion or retention (or racemization) of stereochemistry at phosphorus allowed investigators to rule out mechanistic ambiguities. The R_P and S_P forms of ATP-γ-P–[S, O^{16}, O^{18}] allowed stereochemical studies of both phosphoryl *and then nucleotidyl* transfers that complemented the use of chiral *S*- and *R*- samples of ATP-γ-P-[O^{16}, O^{17}, O^{18}] described in Chapter 7 for determination of enzymatic phosphoryl transfer stereochemical outcomes.

An early study carried out by Eckstein and colleagues on DNA polymerase *in vitro* used the S_P diastereomer of 2′-deoxy-ATP-α–S as substrate for incorporation into the growing DNA chain.[3] As depicted in Figure 12.4, the S_P diastereomer (but not the R_P diastereomer) was smoothly added to the 3′-end of an oligonucleotide by enzymatic nucleotidyl transfer mediated by a DNA polymerase. The product, after stereochemical analysis, yielded only the R_P configuration at the newly formed 3′,5′ thiophosphoryl diester internucleotide link. Since DNA polymerization steps are nucleotidyl transfers

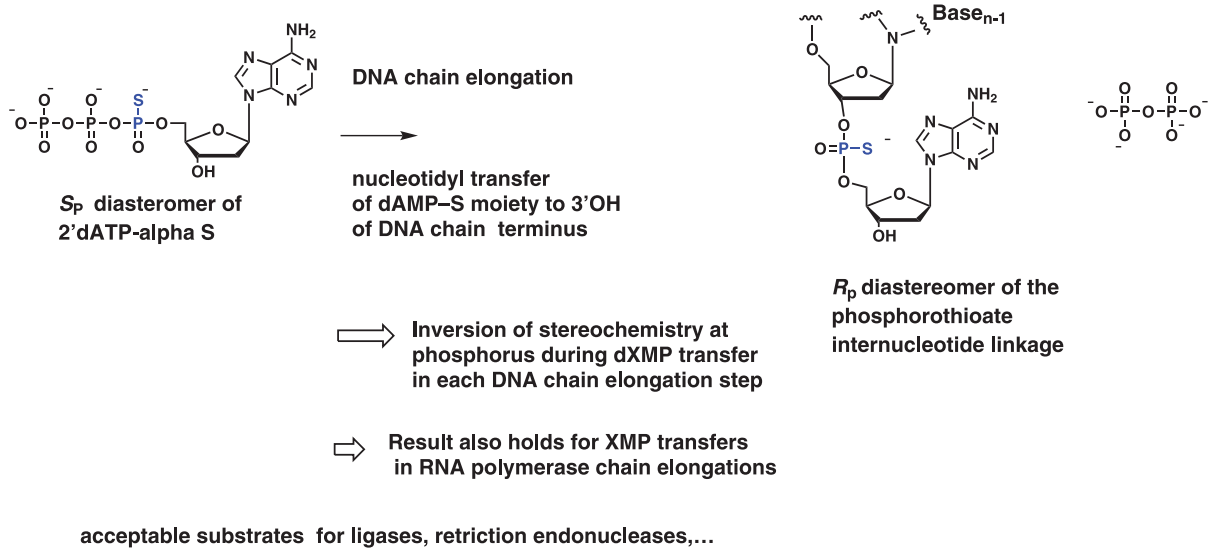


Figure 12.4 Inversion of stereochemistry at phosphorus revealed during nucleotidyl transfer in DNA chain elongation step. The 2′dATP (S_P) diastereomer of ATPα–S gives only the R_P diastereomer of the internucleotide thiophosphoryl product.

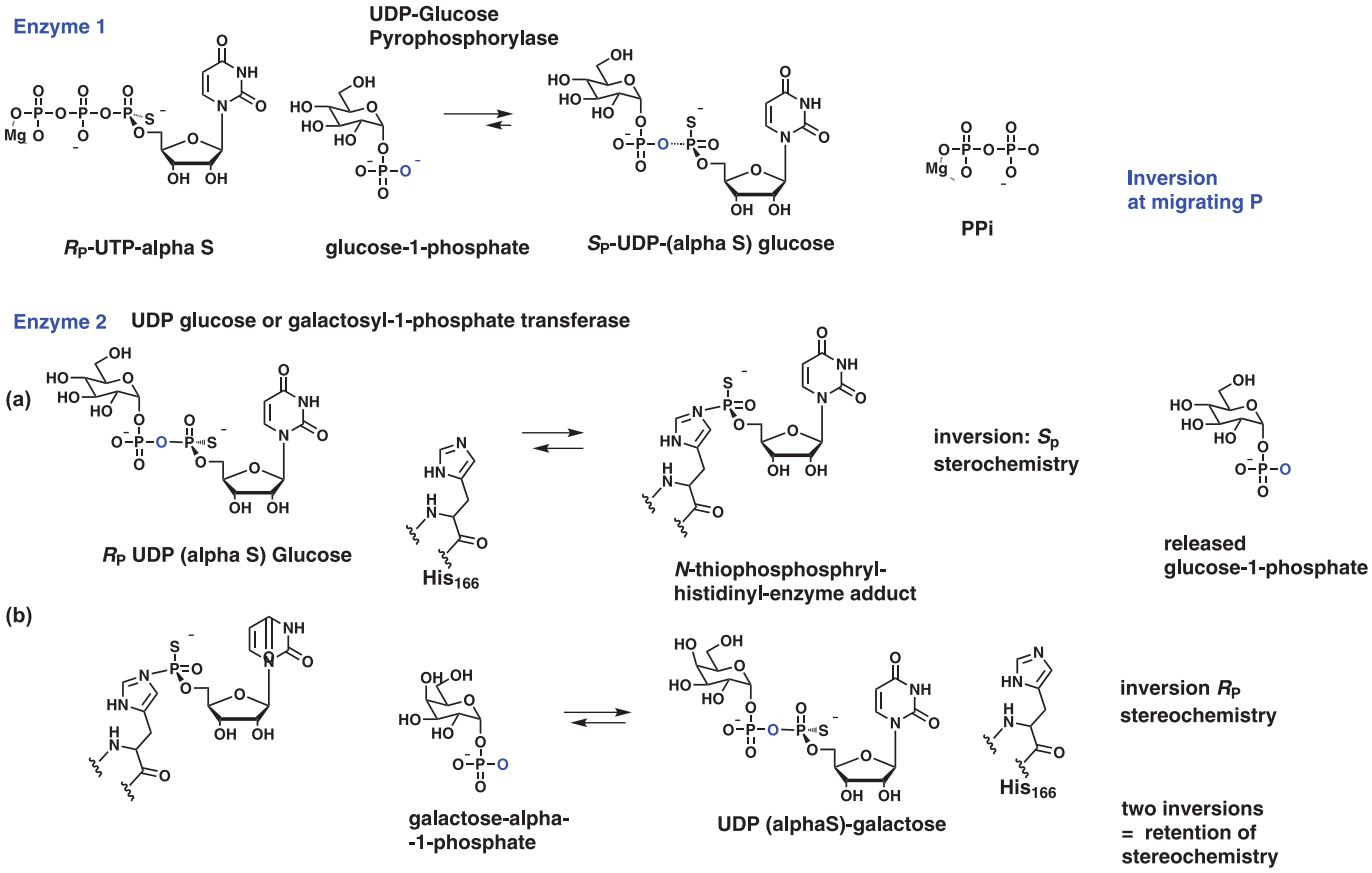
as discussed in Chapter 4, the AMP–S transfer in this DNA polymerase (and by extension all such XMP chain elongation steps in DNA polymerization) proceeded with inversion at phosphorus, consistent with a direct attack of the 3′-OH on the incoming dNTP. A similar stereochemical outcome of inversion at the itinerant phosphorus center of an XMP moiety was observed for RNA polymerase.

One of the key research groups in phosphorus enzymatic stereochemical studies was that of Perry Frey and colleagues in Wisconsin.[6] They elucidated mechanistic aspects of the intersection of glucose and galactose metabolism, arising from digestion of the lactose in mother's milk, to glucose and galactose-1-phosphate. In particular, Frey and colleagues studied the mechanism of two enzymes: UDP–glucose pyrophosphorylase and the companion enzyme UDP–galactosyl transferase (Figure 12.5)

To probe stereochemistry and mechanism of UDP-glucose pyrophosphorylase, they prepared as substrate only the R_P diasteromer of UTP-α–S. After exposure to the enzyme in the presence of cosubstrate glucose-1-phosphate, the UDP(αS)–glucose product was isolated and analyzed for stereochemistry by diastereomer-specific ^{31}P-NMR chemical shift differences. Only the S_P diastereomer of UDP-α–S–glucose was detected (Figure 12.5). This clean inversion of stereochemistry at the migrating phosphorus was consistent with an in line associative transition state and against both a covalent phosphoryl enzyme intermediate (two inversion steps anticipated for net retention at phosphorus) or a dissociative transition state involving free metaphosphate anion, as that could lead to racemization to a mixture of R_P and S_P geometries (not observed).

The other, R_P diastereomer of UDP-α–*S*–glucose served as substrate for the companion enzyme, a UDP–glucose galactosyl-1-phosphate uridylyltransferase, This enzyme, GalT, catalyzes the net transfer of the UMP moiety between galactose-1-phosphate in the forward direction and glucose-1-phosphate in the reverse direction[7] by way of a covalent UMP–enzyme intermediate. The observed product UDP-α–S–galactose was solely the R_P isomer, a net retention of configuration at the itinerant phosphorus atom (Figure 12.5). Because it was known from other evidence that the transferase proceeded through a covalent phosphoryl enzyme intermediate, this stereochemical outcome was anticipated and explainable. Formation of the phosphoryl enzyme and then its subsequent breakdown would each proceed with inversion at phosphorus. Two tandem inversions present as net retention of stereochemistry. In this study, the intermediate would be an UMP-α–S–enzyme intermediate.[8]

As diagramed in Figure 12.5, N3 of the imidazolyl side chain of histidine residue166 in the active site acts as a nitrogen nucleophile to attack the R_P thiophosphoryl group of bound UDP-αS–glucose. As the transition state decomposes in the forward direction glucose–α-1-phosphate is released and the covalent *N*-thiophosphoryl–histidinyl enzyme adduct is formed. This is the first inversion at the migrating thiophosphoryl group and gives an S_P configuration in the covalent enzyme adduct. Not only is this an example


Enzyme 1
UDP-Glucose
Pyrophosphorylase
R_P-UTP-alpha S
glucose-1-phosphate
S_P-UDP-(alpha S) glucose
PPi
Inversion
at migrating P
Enzyme 2 UDP glucose or galactosyl-1-phosphate transferase
(a)
R_P UDP (alpha S) Glucose
His166
N-thiophosphosphryl-
histidinyl-enzyme adduct
inversion: S_p
sterochemistry
released
glucose-1-phosphate
(b)
galactose-alpha-
-1-phosphate
UDP (alphaS)-galactose
His166
inversion R_P
stereochemistry
two inversions
= retention of
stereochemistry

of thiophosphoryl enzyme adduct to histidine, in this stereochemical experiment the adduct has both an N–P and an S–P bond so it is actually a phosphorothioamidate intermediate.

In the second half reaction the released gucose-1-phosphate is replaced by galactose-α-1-phosphate. The nucleophile is one of the oxyanions of the phosphate of galactose-1-phosphate, attacking the electrophilic thiophosphoramidate phosphorus. Net inversion at P converts the S_P configuration of that covalent enzyme intermediate to the observed R_P as sole diastereomer of the UDP-α–*S*–galactose. The presumption in this and other studies with thiophosphoryl analogues of normal NTP metabolites was that the thiophosphoryl group reactions ($-PO_2S^{2-}$ group transfers) were faithful and predictive mimics of the natural $-PO_3^{2-}$ group transfers.

12.5 Stereochemical Inversion at Phosphorus in Phosphoryl Transfers

As noted earlier in this chapter and previously in Chapter 7, determination of the stereochemical outcome of phosphoryl transfer enzymes reactions (kinases) compared with nucleotidyl transferases just examined, was a more formidable task. Both the γ-PO_3 of ATP and NTP congeners, and the $-PO_3$ group in product monophosphate esters ($R–O–PO_3^{2-}$) have three equivalent oxygens around phosphorus. Replacement of one oxygen by sulfur still leaves two chemically identical prochiral oxygens (Figure 12.6). Thus, one needs a way to distinguish among the two remaining oxygens. Heavy isotopes of the normal O^{16} such as O^{18} and even the less familiar O^{17} were available and would serve to generate chiral thiophosphoryl groups. Synthetic methodologies were developed both to produce R_P and S_P diastereomers of ATP-γ–S, 0^{16}, O^{18} and the chemoenzymatic means to deconvolute the stereochemistry (R_P or S_p) of kinase products RO–(P–S, O^{16}, O^{18}).

In parallel as noted earlier in Chapter 7, Jeremy Knowles and colleagues at Harvard took the even more elegant, less perturbative approach to generate chiral-γ-ATP diastereomers using only the three isotopes of oxygen (ATP-γ–P (O^{16}, O^{17}, O^{18}), as separate unambiguously assigned diastereomer.[9,10] Then they had to work out methods to solve the absolute stereochemistry of the enzymatic RO–P–(O^{16}, O^{17}, O^{18}) products.

Figure 12.5 Stereochemical analysis of a tandem pair of enzymes in phosphorylated galactose metabolism. UDP glucose phosphorylase takes the R_P diastereomer of UTPα-S and generates only the S_P diastereomer of UDPα-*S*-glucose. The next enzyme, the UDP-galactose/UDP-glucose transferase, takes the product of the first enzyme and produces UDPαS-galactose, only the S_P diastereomer. The first enzyme gives inversion at the migrating phosphorus center. The second enzyme acts with net retention of stereochemistry *via* two inversion steps, involving a covalent UMPαS–enzyme adduct.

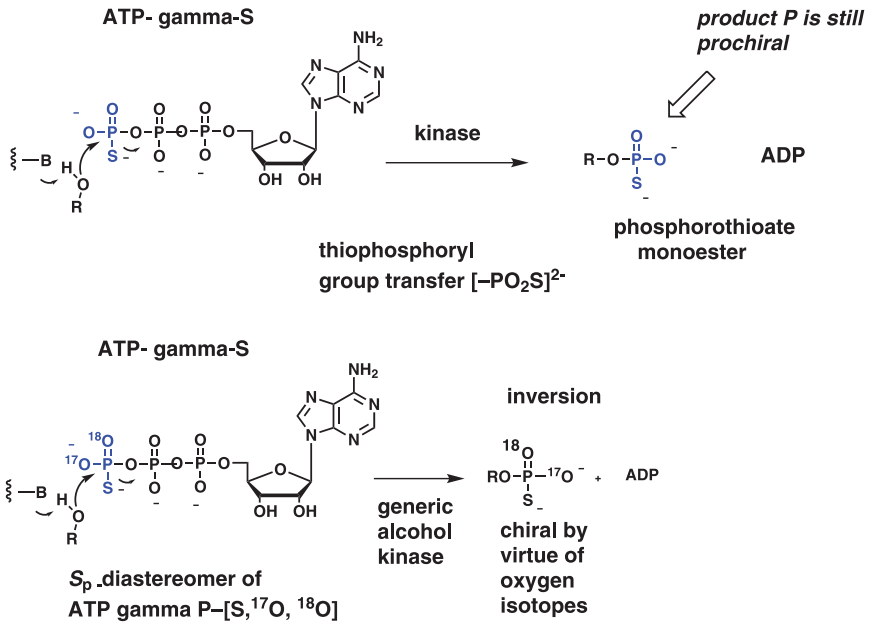


Figure 12.6 To use the prochiral ATPγS as a stereochemical probe, one needs to convert one of the two prochiral oxygens at the phosphorothioate center to a distinguishable substituent. This could be accomplished with a distinct oxygen isotope (*e.g.* ^{17}O or ^{18}O in place of the normal ^{16}O isotope). For example, the S_P diastereomer of ATP–[γS, ^{18}O, ^{17}O] could be used to show that alcohol kinases gave the R_P phosphorothioate monoester (net inversion).

After much ingenious syntheses, phosphoryl transferase enzyme studies, and chemoenzymatic methods for product chirality determinations, the conclusions for the half dozen or more kinases examined were unambiguous. *Enzymatic phosphoryl transfers proceed with inversion at the migrating phosphoryl atom* (Figure 12.6) Thus, both nucleotidyl and phosphoryl transfers of Pα and Pγ of nucleoside triphosphates and derivatives thereof (*e.g.* nucleoside diphosphosugars) seem to follow in line displacement; associative mechanisms with likely participation of transition states with at least partial bond formation and bond cleavage to phosphorus).

Thus, sulfur substitution for oxygen at phosphorus, albeit an unnatural change, has proven of substantial mechanistic utility.[4] Both NTP-γ–S for phosphoryl transfer enzymes (kinases) and NTP-α–S analogs of natural nucleoside triphosphates have been of great value in probing the two main types of ATP (NTP) fragmentation patterns that account for > 99% of all ATP consumption in biology. [There is always a caveat of the possibility of slower rates and of altered recognition or interactions of the phosphorothioate moiety (including chirality in phosphorothioate diesters) with partner proteins leading to changes in mechanism from the physiological all oxygen phosphorus-containing substrates.]

12.6 Naturally Occurring DNA-based Phosphorothioates

Bacteria in natural populations are typically under constant assault from foreign DNA, largely but not exclusively *via* myriad bacterial viruses and phages. Recognition of ''self'' DNA from foreign DNA has been an ongoing challenge for bacteria through the eons of evolution. Two of the more famous bacterial self-defense systems to recognize their own genome from foreign DNA are (1) restriction–modification systems and (2) CRISPR. Each of these systems uses a different principle for guarding the host genome and each have proven remarkably useful in gene editing applications, as has been summarized in Chapter 8.

The restriction–modification systems involve initial recognition of a local DNA sequence, often a four base pair stretch, by a surveillance self-defense enzyme. Recognition and binding are then followed by enzymatic self-methylation of one of the bases in that tetranucleotide sequence, *e.g.* of the N_6–NH_2 of adenine (Figure 8.6). A second enzyme partner to the methyltransferase is a restriction endonuclease that carries out surveillance of DNA. If a sequence is found that is not methylated, reflecting a foreign DNA double strand molecule, that sequence undergoes double strand cleavage by the endonuclease (phosphodiesterase chemistry).

The restriction–modification self-protection logic has a counterpart in recently discovered self-modification of bacterial DNA by an enzyme system that installs a new S–P bond at an internucleotide phosphodiester bridge within a given four base recognition sequence.[11,12] Presumably, this covalent modification, always generating the R_P configuration of the phosphorothioate diester, allows the host DNA to be recognized as self and protected from nuclease digestion of unmodified ''foreign'' DNA[13] (Figures 12.7–12.9). It has been known for decades that phosphorothioate internucleotide bridges in place of the normal phosphodiester bridge stabilizes those bonds against nuclease digestion.

This is the first example of naturally occurring P–S bonds in DNA. Originally detected at low levels in DNA of some soil actinomycetes, the DNA modification system has since been detected in a variety of other bacteria including *E. coli.* Internucleotide phosphodiester conversion to R_P phosphorothioate diester requires four proteins encoded by the DNA degradation *dndACDE* genes. Although homologies to known enzyme classes indicate a logic for sulfur introduction, detailed mechanistic studies remain to be conducted.

Sulfur insertion most probably starts with recognition of a four nucleotide GAAC stretch on one DNA strand and then a presumed single-strand nick by DndD, acting as a classic sequence-specific DNA endonuclease (Chapter 6), to give a 3′-OH and a 5′-phosphate on either side of the nick (Figure 12.7). Typically, such nicks undergo nucleotidyl transfer from ATP to the 5′-phosphate oxyanion to create an ADP derivative on the 5′-side of the nick or gap (*e.g.* the first half of a DNA ligase catalytic cycle (Chapter 4).

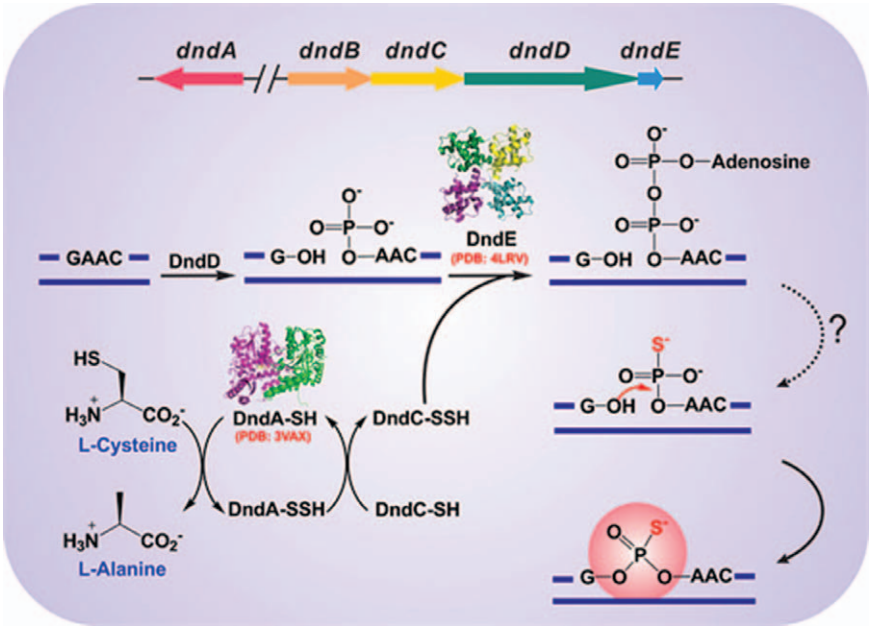


Figure 12.7 Naturally occurring internucleotide phosphorothioate bonds in bacterial DNA. A five gene cluster, *dndA–E*, encode enzymes required to modify specific 3′,5′-internucleotide phosphodiester bonds in a given DNA sequence into the R_P phosphorothioate diastereomers, recognizable as ''self'' DNA sequence. The sulfur donor is proposed to be a persulfide form of an active site cysteine side chain in DndC.
Reproduced from rref. 12 with permission from Oxford University Press, Copyright 2018.

Meanwhile, DndA is proposed to act as a cysteine desulfhydrylase, using pyridoxal-phosphate as a coenzyme to cleave the C3–SH carbon sulfur bond of cysteine.[12] DndC is a partner subunit of DndA and thought to provide an internal cysteine side chain to accept the sulfur released from cysteine as a six-electron sulfane.[14] DndC would then be a chaperone carrying its own cysteine residue as a persulfide–Cys–SSH) (Figure 12.7).

At this juncture, the process would have the persulfide thiolate anion of modified DndC enzyme attack the ADP moiety on the 5′-side of the gap. (Figure 12.8) Addition of the sulfur anion could lead to expulsion of AMP. Cleavage of the DndC–S–S–PO_2–O–DNA adduct or intermediate could happen by persulfide reduction to give a thiophosphoryl group on the 5′ side of the gap. Ligation by an as yet unclear mechanistic path would create the R_P-phosphorothioate internucleotide ligand (Figure 12.7). This modification would be the signal that the DNA is ''self.''

Consistent with this interpretation of a four base pair recognition sequence in a three megabase genome, there would be ~5000 GAAC sites in

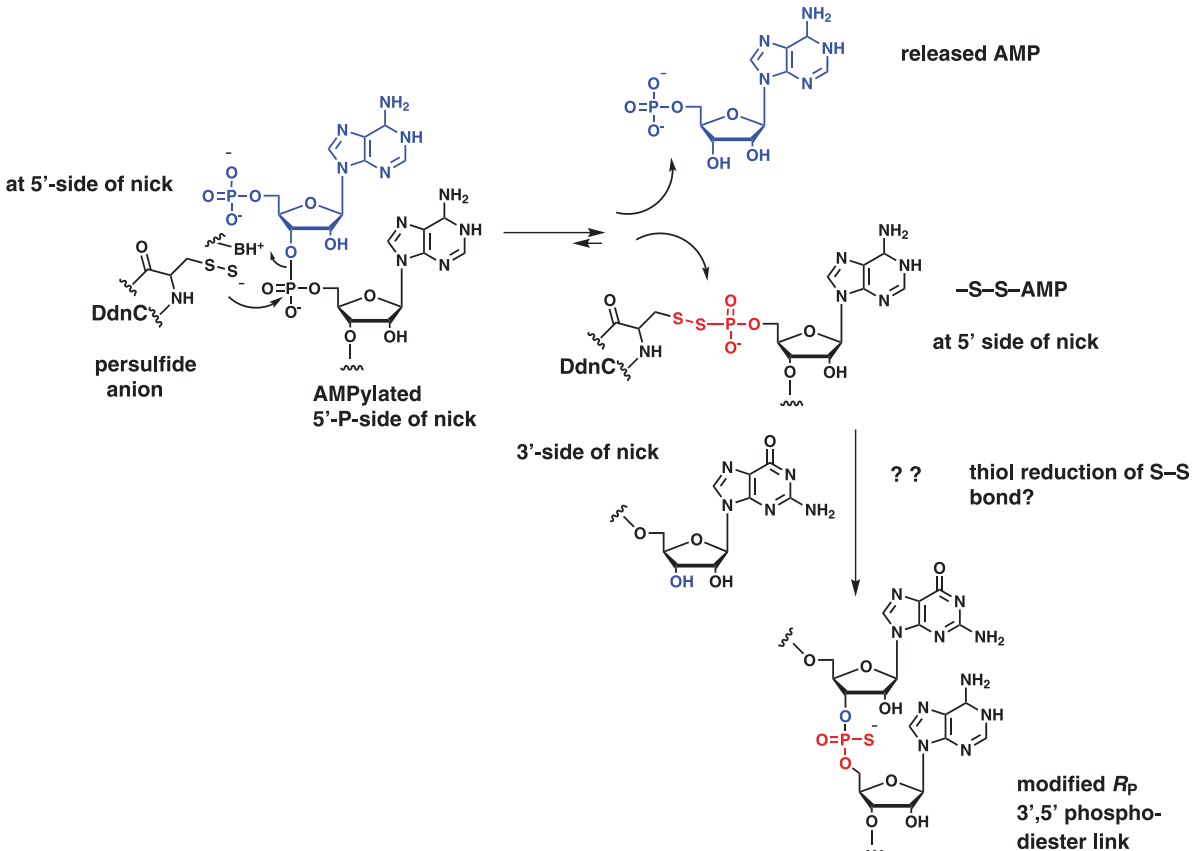


Figure 12.8 Naturally occurring internucleotide phosphorothioate bonds in bacterial DNA. The reaction mechanism for DndC could involve a single-strand nick AMPylated on the 5′ side of the gap as the electrophile for attack by the DndC–SS$^-$ to create a persulfido AMP adduct that could undergo reductive closure of the gap.

E. coli DNA. The observed modification appears to be sub-stoichiometric at those 5000 sites. This may mean the thiolation system is dynamic and/or the phosphorothioate link may serve another or an additional purpose, such as being a sentinel for oxidative conditions (Figure 12.9). Oxidative removal of sulfur from S–P bonds is precedented in metabolism as noted in Section 12.7 below.

Most remarkable is the detection of these first S–P bonds in DNA[11] and their likely function in analogy to known restriction–modification systems to protect integrity of bacterial host genomes.

12.7 Oxygenative Conversion of Parathion to Paraoxon

The xenobiotic organophosphate insecticides (see Chapter 8 on phosphotriesters and related phosphorothioates) have been widely used for agricultural applications. Among the best studied commercial thiophosphosphoryl agricide is parathion (Figure 12.9). It is a thiophosphoryl triester. Since all three oxygens are esterified, two as alkyl esters and the third as the aryl nitrophenyl ester that is the leaving group for acetylcholinesterase inhibition, the P–S bond must be a formal P═S double bond as shown.

Among the metabolic routes for parathion in mammalian liver is oxygenative desulfuration by a set of hepatic cytochrome P450 monooxygenases. The desulfuration produces H_2S as a byproduct as the P═S in parathion is converted to the P═O in paraoxon, still a potent acetylcholinesterase inhibitor. The mechanism is thought to involve initial *S*-oxygenation to an $-S^{+}-O^{-}$ structure. That could readily rearrange to the indicated oxaphosphothiirane three-membered ring where all three atoms are heteroatoms: sulfur, oxygen, and phosphorus. One expects that oxaphosphothiirane to be unstable and it could further rearrange to the P═O product with extrusion of H_2S, probably as the hydrosulfide monoanion. It is possible the oxygenative conversion of parathion to paraoxon is a precedent for nonenzymatic oxygenative desulfurization of the phosphorothioate internucleotide links in the previous subsection back to the unmodified phosphodiester. If so, that could be an element in dynamic turnover of the sulfur mark in those stretches of bacterial DNA.

12.8 Synthetic Phosphorothioate Oligonucleotides and the Problem of Chirality Control

Decades before this naturally occurring internucleotide R_P phosphorothioate DNA self-protection strategy noted above was discovered, nucleotide chemists had been synthesizing sequence-specific oligonucleotides with various alterations in the covalent 3′,5′-internucleotide link. Sulfur in the

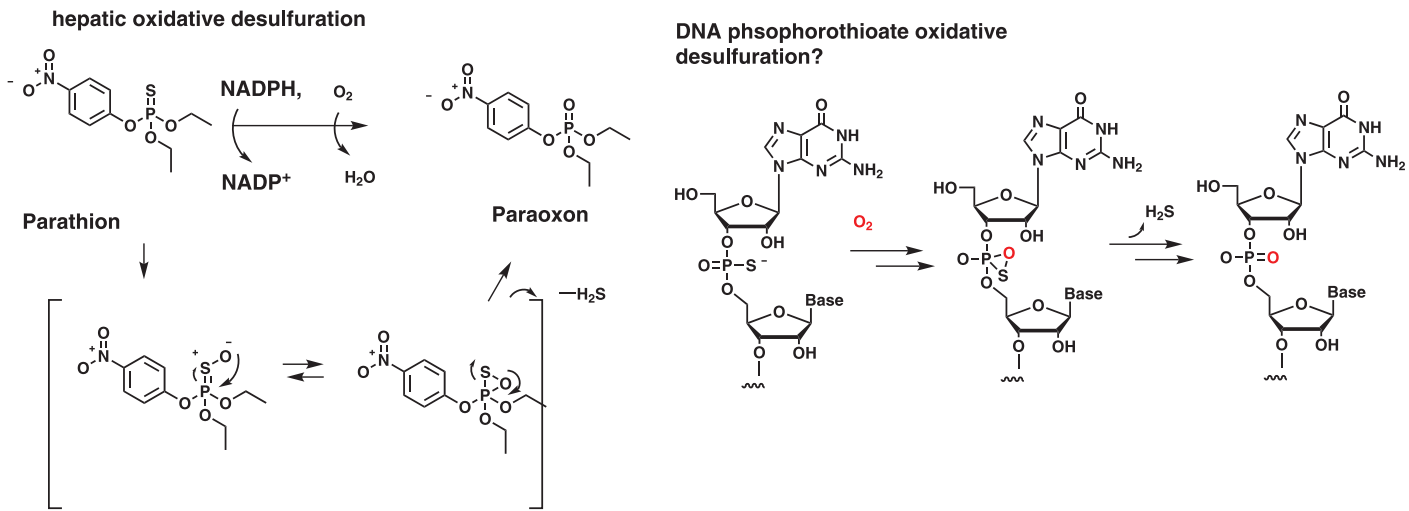


Figure 12.9 Naturally occurring internucleotide phosphorothioate bonds in bacterial DNA. Proposed dynamic oxidative desulfurization of such R_P phosphorothioate internucleotide bonds (dynamic reversal of the modification) might have precedence in the hepatic desulfurization of parathion to paraoxon.

form of phosphorothioate internucleotide links in oligodeoxynucleotides became a strategic and synthetic mainstay for antisense applications to form hybrid duplexes with target DNA strands.[15] It was observed that the sulfur substitution for one oxygen atom in the internucleotide bond helped slow hydrolytic decomposition by cellular nucleases and also somewhat reduced polarity to facilitate cellular uptake of these anionic oligomers.

Most synthetic routes to phosphorothioate oligonucleotides use a variant of DNA synthesis protocols that feature the coupling steps for each nucleoside addition *via* a phosphorus^{III} nucleophile (see Chapter 10 Section 10.8). For the thiophosphoryl links, the P^{III} oxidation state is oxidatively sulfurized to form the biologically relevant P^{V} oxidation state. Decades of process chemistry and automation have increased the yields of phosphorothioate oligonucleotides, allowing clinical studies for human therapeutics targeted at gene variants. Thus, spinraza, an 18-residue oligonucleotide where each covalent 3′,5′-internucleotide linkage is a phosphorothiodiester, was recently approved as a therapeutic agent for treatment of spinal muscular atrophy in the USA (Figure 12.10).

One of the consequences of the P^{III} synthetic route to spinraza noted by Knouse *et al.*[17] is that there is no stereocontrol over R_P *vs.* S_P phospho diastereomers at each of the 18 coupling steps. That allows for 2^{18} possible configurational isomers, or up to ~132 000 possible isomers in the approved human therapeutic product!

Recently, a joint team of medicinal chemists between Scripps Research Institute and Bristol Meyers Squibb published an approach to bypass the stereorandom defects of using P^{III} coupling steps.[17] They designed and tested a P^{V} coupling reagent for oligodeoxynucleotide synthesis with complete stereochemical control in the formation of each phosphorus–heteroatom bond.

To that end they prepared in crystalline, scalable amounts both the separate R_P and S_P forms of the phosphorus^{V} oxidation state reagent shown in Figure 12.11. They demonstrated the chemical and stereochemical utility for making all relevant dinucleotides with internucleotide 3′,5′-phosphorothioate bonds under absolute stereocontrol to make either the R_P or S_P diastereomers. In initial tests the method was also used to make a single pentamer diastereomer. Also, a 16-*mer* could be prepared by these reagents. This strategy should revolutionize the synthetic approaches to phosphorothioate oligonucleotides as diagnostic and perhaps therapeutic agents.

12.9 Thiophosphoryl Enzyme Intermediates

In the realm of proteins and enzymes, we have already noted a couple of cases of *S*-phosphoryl–cysteinyl adducts as kinetically and chemically competent enzymatic intermediates. We defer further examination of

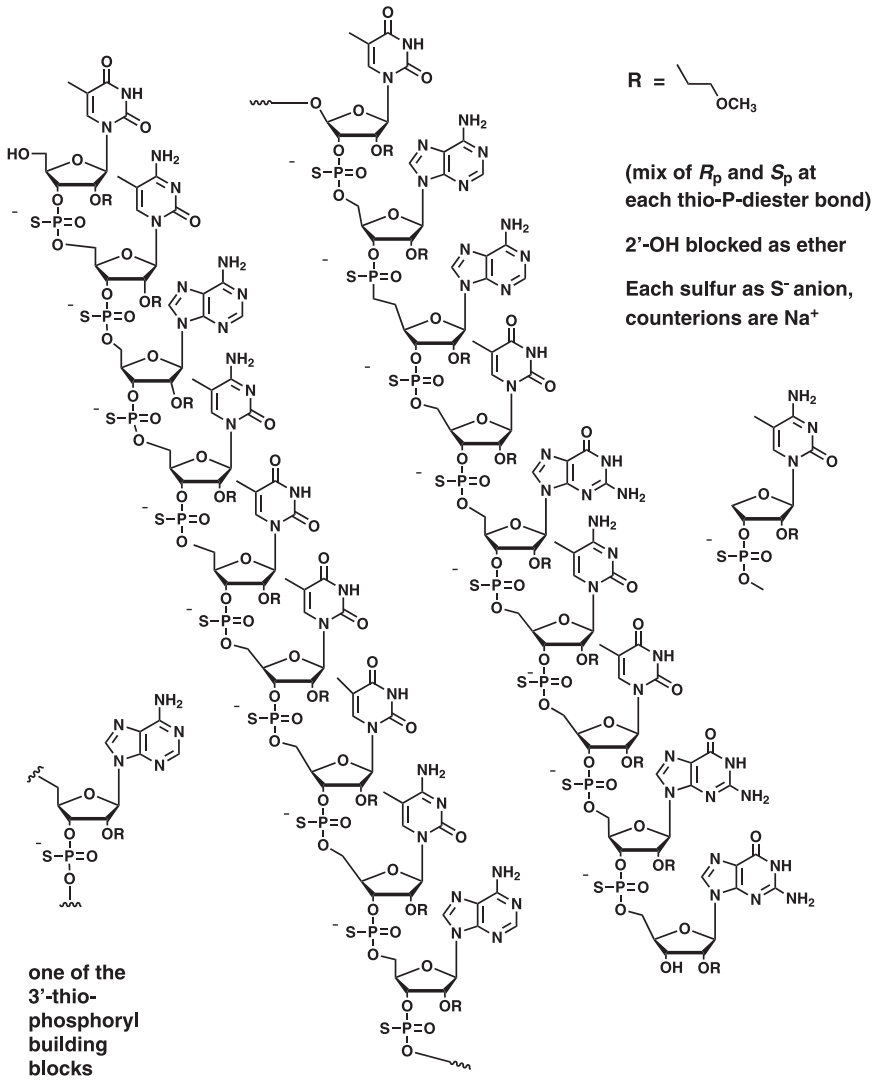


Figure 12.10 FDA approved drug Spinraza:[16] an 18-*mer* oligonucleotide with 17 thiophosphoryl internucleotide bonds. In principle, given its mode of synthesis *via* P^{III} *phosphoramidite* chemistry, there could be up to 2^{18} possible diastereomers in a preparation.

formation, function, and breakdown to the phosphoproteomics Chapter 14 in Section IV. In the documented cases of *S*-phosphoryl cysteine proteins the sulfur is the bridging atom from the amino acid side chain to the phosphorus atom (in contradistinction to the nonbridging substitutions in thionucleotides discussed earlier in this chapter).

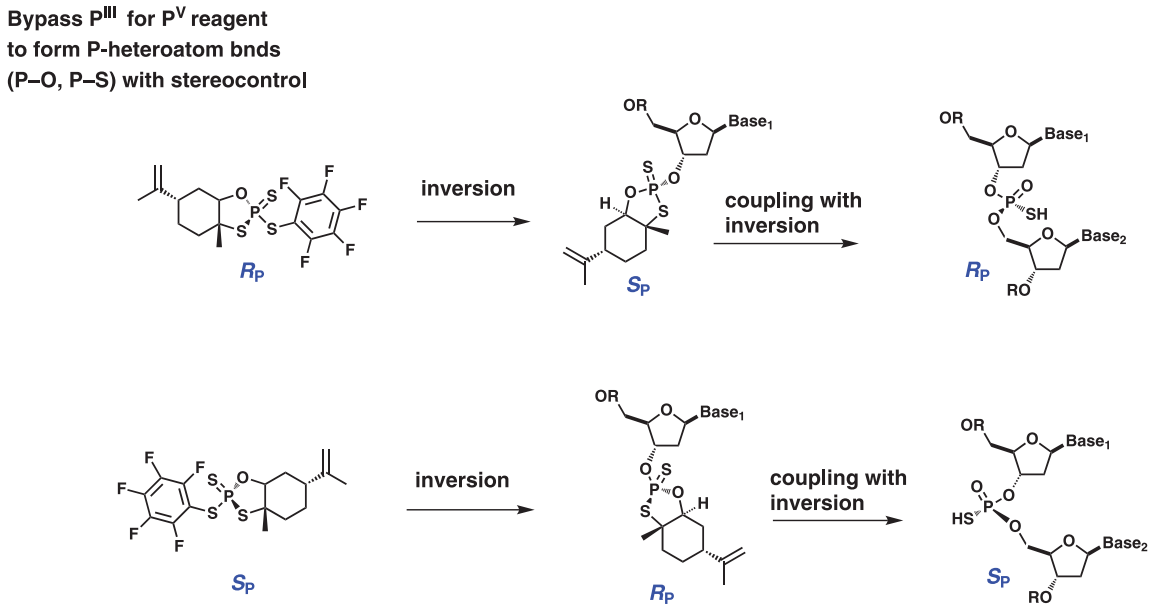


Figure 12.11 Methodology for oligonucleotide synthesis that bypasses P^{III} chemistry by P^{V} reagents that enable diastereomeric control at each phosphorothioate diester internucleotide bond.

References

1. M. Alam, R. Sandula, M. B. Hossain and D. van der Helm, Gymnodinium brevetoxins 1. Isolation and x-ray structure of O,O-diisopropyl (E)-2-(1-methyl-2-oxopropylidene)phosphorohydrazothioate (E)-oxime from the red tide dinoflagellate *Gymnodinium breve,*, *J. Am. Chem. Soc.*, 1982, **104**, 5232–5234.
2. P. A. Frey and R. D. Sammons, Bond order and charge localization in nucleoside phosphorothioates, *Science*, 1985, **228**, 541–545.
3. F. Eckstein, Phosphorothioate analogs of nucleotides: tools for the nvestigation of biochemical processes, *Angew. Chem., Int. Ed. Engl.*, 1983, **22**, 423–439.
4. F. Eckstein, Nucleoside phosphorothioates, *Annu. Rev Bicohem.*, 1985, **54**, 367–402.
5. F. Eckstein and G. Gish, Phosphorothioates in molecular biolgy, *Trends Biochem. Sci.*, 1999, **14**, 97–100.
6. P. A. Frey, Sulfur as a mechanistic probe in enzymatic and nonenzymatic substitution at phosphorus, *New J. Chem.*, 2010, **34**, 820–828.
7. J. E. Wedekind, P. A. Frey and I. Rayment, The structure of nucleotidylated histidine-166 of galactose-1-phosphate uridylyl transferase provides insight into phosphoryl group transfer, *Biochemistry*, 1996, **36**, 11560–11569.
8. P. A. Frey, Transient intermediates in enzymology 1964-2008, *J. Biol. Chem.*, 2015, **290**, 10610–10626.
9. S. R. Jones, L. R. Kindman and J. R. Knowles, Stereochemistry of phosphoryl group transfer using a chiral [^{16}O,^{17}O,^{18}O)] stereochemical course of alkaline phosphatase,, *Nature*, 1978, **275**, 564–565.
10. J. R. Knowles, Enzyme-catayzed phosphoryl transfer reactions, *Annu. Rev. Biochem.*, 1980, **49**, 877–919.
11. L. Wang, S. Chen, T. Xu and K. Taghizadeh, Phosphorothioation of DNA in bacteria by *dnd* genes, *Nat. Chem. Biol.*, 2007, **3**, 709–710.
12. L. Wang, S. Jiang, Z. Deng, P. C. Dedon and S. Chen, DNA phosphorothioate modification- a new multi-functional epigenetic system in bacteria, *FEMS Microbiol. Rev.*, 2018, **43**, 109–122.
13. R. Gan, X. Wu, W. He, Z. Liu, S. Wu, C. Chen, S. Chen, Q. Xiang, Z. Deng, D. Linag, S. Chen and L. Wang, D. N. A. phosphorothioate modifications influence the global transcriptional response and protect DNA from double strand breaks, *Sci. Rep.*, 2014, **4**, 6642.
14. C. T. Walsh, *The Chemnical Biology of Sulfur*, Royal Society of Chemistry Press, London, 2020.
15. F. Eckstein and G. Gish, Phosphorothioates, essential components of therapeutic nucleotides, *Nucleic Acid Ther.*, 2014, **24**, 374–382.
16. (downloaded 12-01-2019) SPINRAZA (nusinersen), www.spinraza.com.
17. K. W. Knouse, J. N. deGruyter, M. A. Schmidt, B. Zheng, J. C. Vantouront, C. Kingston, S. E., Mercer, I. M. McDonald, R. E. Olson, Y. Zhu, C. Hang, J. Zhu, C. Yuan, Q. Wang, P. Park, M. D. Eastgate and P. S. Baran, Unlocking $P^{(v)}$:reagents for chiral phosphorothioate synthesis, *Science*, 2018, **361**, 1234–1238.

SECTION IV

Phosphoproteomics

Over the past few decades phosphoproteomics has dominated the roles of phosphates in biology. From the initial discovery of protein kinases and opposing phosphoprotein phosphatases a half century ago, tens of thousands of publications have defined and filled out the pictures of protein posttranslational phosphorylation as a dominant mode of signal transduction in eukaryotic organisms. With mass spectrometry of digested phosphopeptides as the main analytical tool, the number of stable protein phosphorylation sites in human proteomes have surpassed the staggering number of 250 000 distributed among the ~26 000 human proteins.

Indeed, phosphoproteomics and the behavior of certain sets of protein kinases have dominated much of the literature on cancer metabolism for the past three decades. There is no way four chapters in this volume on phosphorus chemical biology can reach meaningfully into the many levels of phosphoprotein signaling biology and provide detailed new insights into mechanism, logic, or strategy not already covered in more detail by primary and review literature.

Instead, we stay focused in this section, as we have in the earlier three sections, on how the chemistry of the differing phosphorylated amino acids in phosphoproteomes condition both initial detection of the phosphorylated side chains and constrain or enable function. Among these chemical matching issues are whether dedicated phosphoprotein phosphatase families are required to mediate dehosphorylation and thereby reversible signaling.

As noted in Chapter 13, nine of the twenty proteinogenic amino acids have side chain nucleophiles that could qualify as targets for subclasses of protein kinases. Five are potential oxygen nucleophiles, three potential nitrogen nucleophiles, and one is the thiolate side chain of cysteine residues. Yet, the mass spectrometry-based workflows over the past two decades have included

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

steps requiring the phosphoryl linkage to the protein side chain to be acid-stable. This tends to miss six of the nine side chains and focus on the three acid-stable ones: P-Ser, P-Thr, and P-Tyr.

Therefore, Chapter 14, takes up the six residues of the noncanonical phosphoproteomes that have been missing or severely under-reported in classic mass spectrometry workflows. These include three side chain phosphoramidate P-His, P-Arg, and P-Lys residues. They are all acid-labile, base-stable and missing from canonical phsphoproteomics analyses. There are multiple P-His- and P-Arg- protein examples but few P-Lys- proteins yet identified. The P-His have high group transfer potential and do not require dedicated phosphoprotein phosphatases for reversal as they hydrolyze spontaneously.

The protein aspartyl-phosphate and glutamyl-phosphate intermediates would be similarly thermodynamically activated and may be kinetically labile. P-Type ion-pumping ATPases and the response regulator proteins of two-component regulatory systems of bacteria are replete with transient side chain aspartyl phosphate intermediates.

Finally, the sixth side chain nucleophile that gives activated phosphoprotein intermediates is cysteine. Some 95 of the 107 predicted human phosphotyrosyl protein phosphates use an active site cysteine as nucleophile, generating covalent *S*-phosphocysteinyl enzyme intermediates which are rapidly hydrolyzed to Pi and the regenerated, free cysteinyl enzyme.

All six of these phosphoprotein classes are typically missed in standard counting of phosphoproteins by acid-stable phosphopeptide detection. Recent altered workflow protocols indicate that these missing phosphoprotein classes could reach one third of the total phsphoproteomes in HELA cells.

Chapter 15 then turns in summary fashion to the vast arena of canonical phosphoproteomics. With ~500 human protein kinases known to modify up to 250 000 Ser, Thr or Tyr side chains, sorting out dynamics, regulation, and signal integration continues to be a beehive of research activity. About 100 of these are Tyrosine kinases and ~420 are Ser–Thr kinases. Abundance counts indicate P-Ser residues at close to 90% of the total, P-Thr at 6–8% and P-Tyr in the range of 1–2% of reported protein phosphorylation sites. Despite their low abundance, tyrosine kinases, especially the transmembrane receptor subset, have a large presence in signal transduction biology.

Finally, the rock stable chemical stability of the two aliphatic alcohol phosphate monoesters P-Ser and P-Thr and the less stable but still long-lived stability of the phenolic phosphate ester in P-Tyr residues necessitates the evolution and deployment of opposing phosphatases to control the dynamics of protein phosphorylation states. We have noted that most P-Tyr-protein phosphatases generate covalent enzyme–Cys–S–PO_3^{2-} intermediates. Contrarily, most Ser and P-Thr protein phosphatases use two metal cations (*e.g.* Zn^{2+}, Fe^{2+}) to coordinate substrates and accelerate phosphoester hydrolysis.

Chapter 16 takes up additional noncanonical posttranslational modifications that start from ATP, CDP-choline or NAD^+ that yield diverse types of posttranslational modifications. These are typically not scored in canonical phosphoproteomics, but reveal the scope of phosphoric anhydride and phosphoric ester chemical biology in proteomics.

CHAPTER 13

Scope and Roles of Posttranslational Protein Phosphorylations

13.1 Proteomics

Starting with the first publication of organismal genomes in the 1990s and culminating in the back to back pair of papers on the first human genomes in 2001, it became possible to predict the primary sequences of all encoded proteins.[1,2] This genomics knowledge base changed the downstream field of protein analyses from one at a time to the goal of ''all at a time''. That is, it seemed possible to analyze how systemic changes in protein composition within a cell or tissue or organism changed over time in response to a specific set of controlled perturbations. This is the essence of *proteomics*, a systems-wide analysis of correlated changes of proteins in a cell, under different conditions.[3] It could lead to deconvoluting the dynamic wiring diagrams for protein responses within cells, tissues, and organisms.

Proteomics are just one level of systems approaches to information in biology.[4,5] They fit into the hierarchical rubric of genomics to transcriptomics to proteomics to metabolomics, following changes in gene expression, mRNA transcriptions, protein expression, and function to the changes in low molecular weight metabolites from moment to moment and from place to place (Figure 13.1).

The ability to monitor system-wide changes in proteins was then coupled to examination of the *posttranslationa*l modifications of proteins.[7,8] Some 200 types of covalent modifications have been described that are largely monitorable by the changes in mass of the side chains (or backbones) of the amino acids that have undergone a given covalent modification. The top five

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

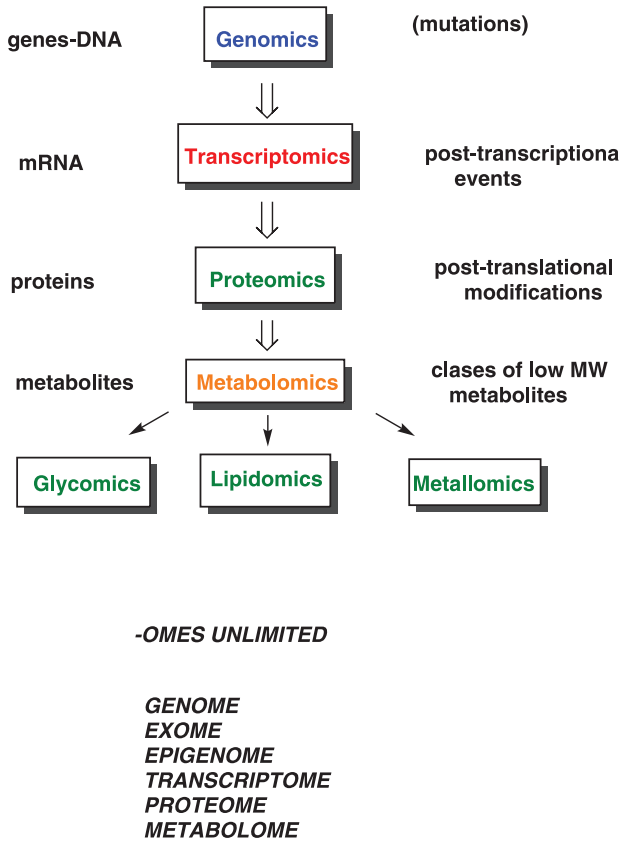
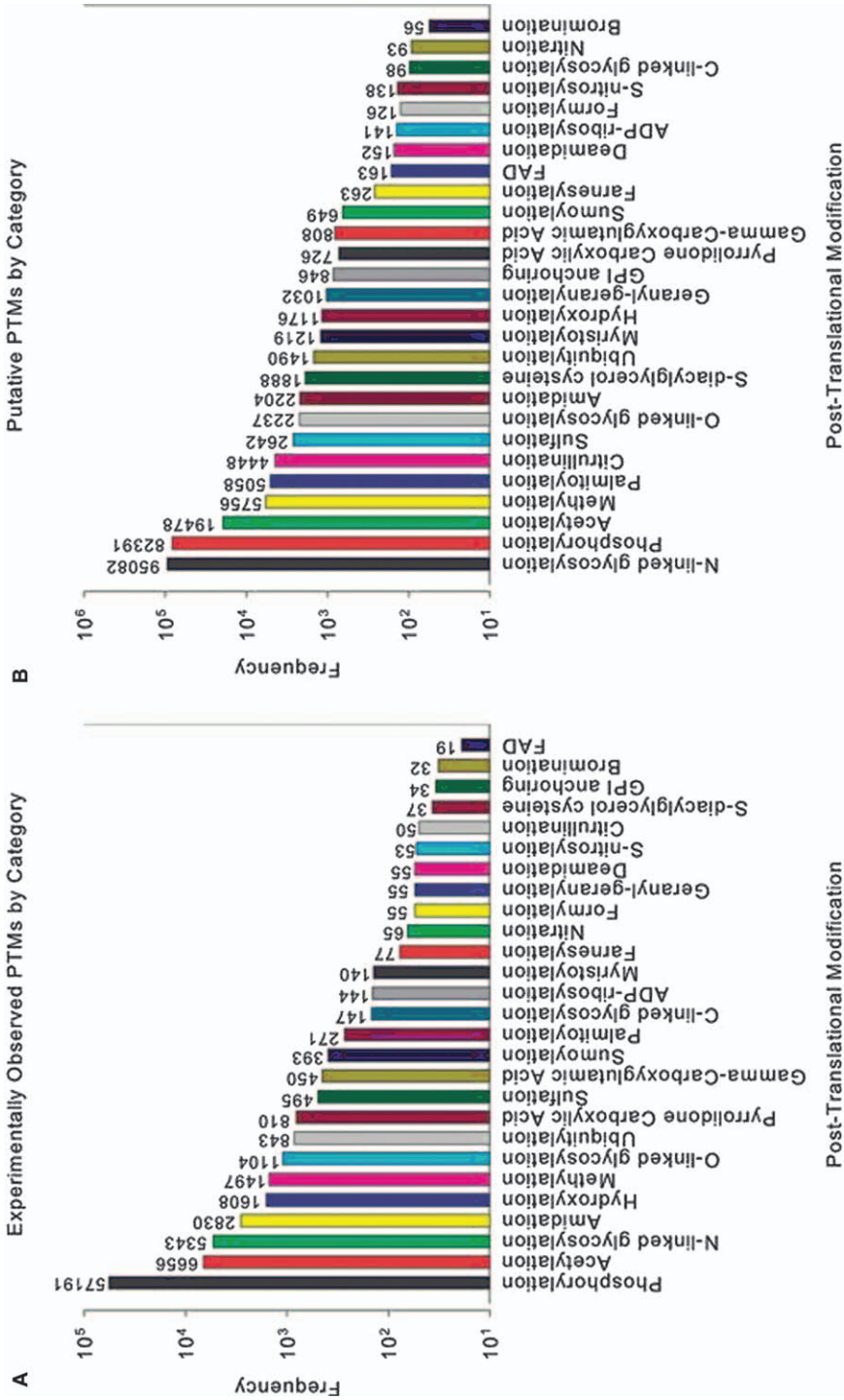


Figure 13.1 Omics Depictions.[6] A typical chart of omics might reflect information flow from genes (genomics) to mRNA species, spliced with splice variants (transcriptomics) to a collection of proteins including posttranslationally-modified forms (proteomics), to metabolites of molecular weight <1000 Daltons (metabolomics). The lower section indicates that the term "ome" has become almost a universal suffix in biology from a systems perspective. Thus, one can define lipidpomics, cellulomics, epigenomics, and other omics *ad libidum*.

modifications in abundance have been thought to be protein side chain phosphorylations, acylations (both acetyl and longer chain palmitoyl and myristoyl), alkylations (both methyl and prenyl groups), glycosylations (*N*- and *O*-glycosylations), and information-rich ubiquitylations (mono- to oligo-Ubs) (Figure 13.2).

13.2 Phosphoproteomics: Version 1.0

The most abundant of the posttranslational modifications catalogued over the past twenty years with mass spectrometry-based analyses of peptide fragments that can be traced back to unique intact proteins are


A
Experimentally Observed PTMs by Category
Frequency
Post-Translational Modification
Phosphorylation 57191
Acetylation 6656
N-linked glycosylation 5343
Amidation 2830
Hydroxylation 1608
Methylation 1497
O-linked glycosylation 1104
Ubiquitylation 843
Pyrrolidone Carboxylic Acid 810
Sulfation 495
Gamma-Carboxyglutamic Acid 450
Sumoylation 393
Palmitoylation 271
C-linked glycosylation 147
ADP-ribosylation 144
Myristoylation 140
Farnesylation 77
Nitration 65
Formylation 55
Geranyl-geranylation 55
Deamidation 55
S-nitrosylation 53
Citrullination 50
S-diacylglycerol cysteine 37
GPI anchoring 34
Bromination 32
FAD 19
B
Putative PTMs by Category
Frequency
Post-Translational Modification
N-linked glycosylation 95082
Phosphorylation 82391
Acetylation 19478
Methylation 5756
Palmitoylation 5058
Citrullination 4448
Sulfation 2642
O-linked glycosylation 2237
Amidation 2204
S-diacylglycerol cysteine 1888
Ubiquitylation 1490
Myristoylation 1219
Hydroxylation 1176
Geranyl-geranylation 1032
GPI anchoring 846
Pyrrolidone Carboxylic Acid 726
Gamma-Carboxyglutamic Acid 808
Sumoylation 649
Farnesylation 263
FAD 163
Deamidation 152
ADP-ribosylation 141
Formylation 126
S-nitrosylation 138
C-linked glycosylation 98
Nitration 93
Bromination 56

O-phosphorylations. The work flow for both base line identification, and changes on some specific cellular perturbations, of phosphate monoesters in side chains of Ser, Thr, and Tyr residues takes advantage of their acid stability during peptide cleavage and isolation procedures[10–12] (Figure 13.3).

Some astounding conclusions have emerged. One generalization is that each human protein has a 30% chance of undergoing at least one phosphorylation event during its cellular lifetime. The inventory of Ser, Thr, and Tyr protein kinases encoded in human genomes is around 520. Those 520 encoded enzymes, protein kinases, comprise the human kinome. Some are specific for a few Ser, Thr, or Tyr sites in protein targets; others are dramatically more promiscuous. Basal levels of activity can vary greatly but many or most of the kinases are in low activity, idle mode until a specific upstream signal is generated. In this strategic mode, they can affect essentially every cell parameter: cell cycle stage, DNA replication and repair, selective activation of certain metabolic pathways, decisions on growth, differentiation, and apoptotic cell death programs, to name just a subset of the myriad consequences.

The primary and review literature on phosphoproteins, singly, on pathways from tandem phosphorylations, on multiple phosphorylations of target and partner proteins in complexes, and on cellular effects, on phosphoproteomics is consequently immense. The many facets of the subject are beyond the scope of this chapter, and indeed any single contemporary book volume, or web site, especially the biological consequences. Therefore, we stay focused on some of the general properties that the PO_3^{2-} group(s), once introduced into a target protein side chain by kinase action, are likely to have on protein conformation and function as starting points for any consideration of biological consequences.

In one of many cogent reviews on phosphoproteomics (a dramatically large literature given that phosphoproteomics have dominated the attention to phosphate metabolism over the past three decades), Adam and Hunter[13] recently noted that the full human proteome contains ~2 million serine residues, 1.3 million threonine residues, and 637 000 tyrosine residues. Examination of many data sets in the range of a few thousand to 30 000 phosphopeptide detections indicates that 8% of serine residues are currently known to be phosphorylated, 5% of threonine residues, and up to 7% of tyrosine residues. If all of these were phosphorylated at the same time (unlikely given the dynamic balance of protein kinases and phosphoprotein

Figure 13.2 Two snapshots of posttranslational frequencies in a tabulation of data sets on abundance of experimentally observed posttranslational modifications of a proteome and comparative abundance by chemotype.[9] (A) The top three categories are protein phosphorylations, acetylations, and *N*-linked glycosylations. The same three categories are most abundant in (B).

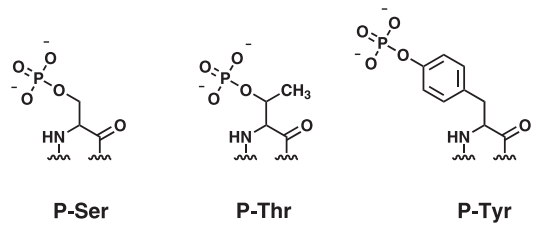


Figure 13.3 The three acid-stable protein-based phosphomonoesters are P-Ser, P-Thr, and P-Tyr, amenable to detection by conventional proteomic mass spectrometry workflows.

phosphatases), that would sum to ~160 000 P-Ser residues in proteins (some with multiple sites phosphorylated in the same protein), ~65 000 P-Thr residues, and ~45 000 P-Tyr residues, for a possible total of ~270 000 phosphorylation sites for just these three residues. The diversity of protein forms that could be generated is enormous, and the functional consequences of even a small defined set of residue phosphorylations are known to be profound.

We will turn to distinct and specific subsets of P-Ser and P-Thr kinases *vs.* P-Tyr kinases in Chapter 15, but just note here that they tend to be mutually specific for the alkyl alcohol *vs.* the phenol side chains, although there are a small number of dual and triple specificity protein kinases for particular protein substrates.

13.3 How do Phosphate Groups Have Their Effects in Proteins

The comprehensive review by Westheimer[14] *"Why Nature Chose Phosphates"* focused on the chemical properties of phosphoric acid, its ionization states, and solubilizing effects, its ability to function both as a nucleophilic oxyanion at the periphery and an electrophilic phosphorus (P = [+5]) in the center. The electrostatic barriers provided by the negative charge allow phosphate monoesters to have half-times estimated to be 10^{12} years at 25 °C but to react rapidly in phosphatase active sites where charges can be neutralized or opposing charges used to raise bound E–S complex energies. The special utility of the 3′,5′-phosphodiester bond as the covalent linkage between monomer units in both RNA and DNA means the phosphodiester bonds still have a negative charge to repel water and ensure hydrolytic stability.

Twenty five years after the Westheimer review, Hunter[15] provided a philosophic update on "*Why Nature Chose Phosphate to Modify Proteins*" noting how the chemical properties of phosphate as a low molecular weight metabolite were also admirably adapted to function as a novel entity when covalently attached to an amino acid side chain in a protein microenvironment. He emphasized the introduction of ($-PO_3^{2-}$) as a divalent anion, its tetrahedral nature, and its hydration sphere as a strong perturbant of preexisting protein structure. None of the twenty proteinogenic amino acid

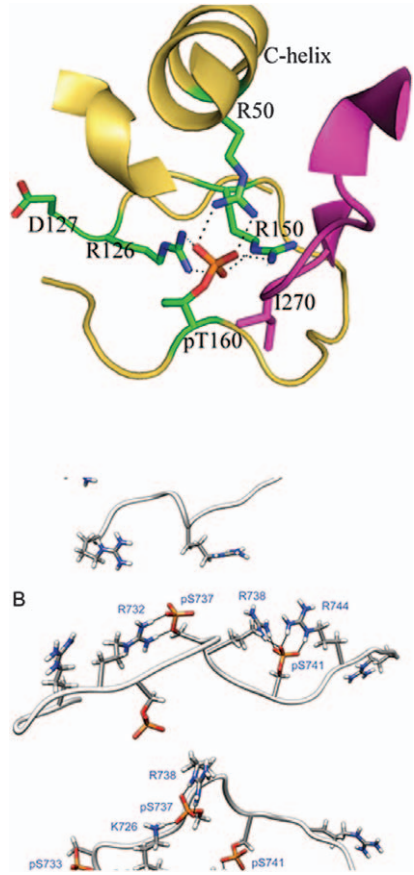
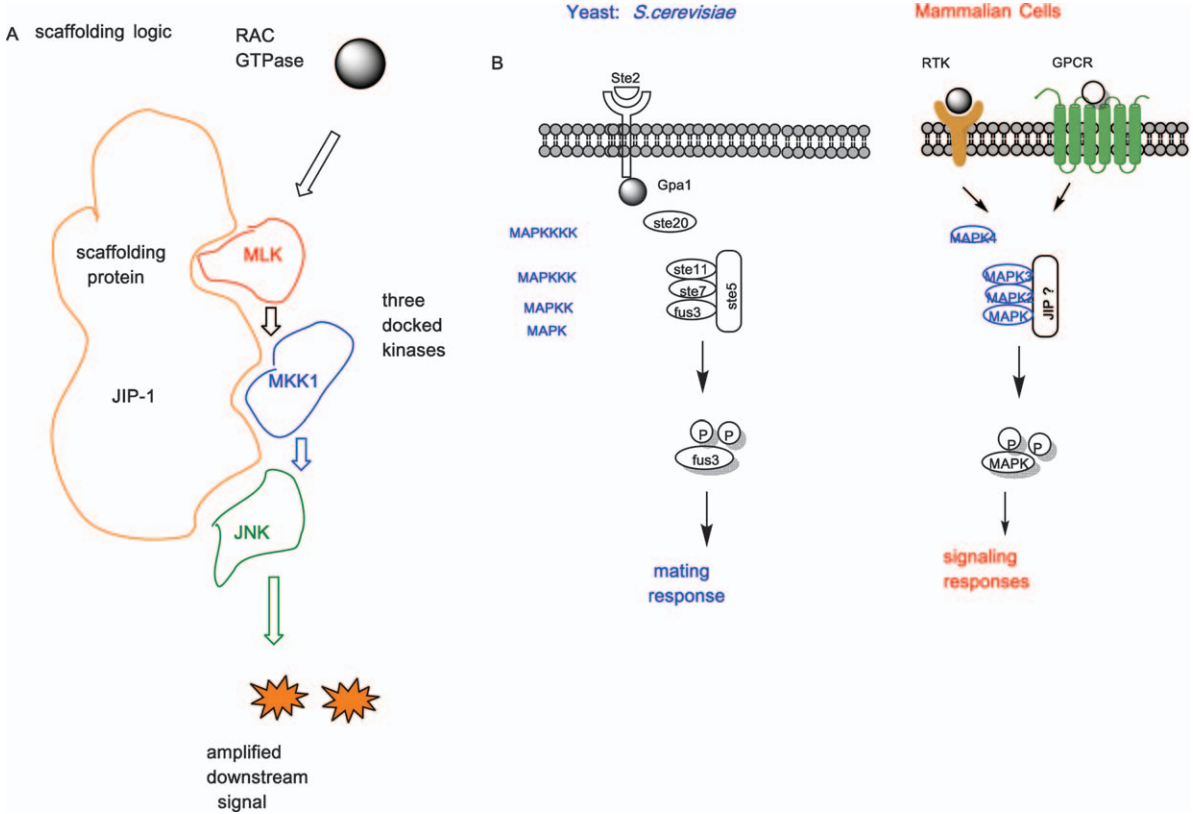


Figure 13.4 The dianionic tetrahedral side chains of phosphate monoesters on Ser, Thr, and Tyr residues can perturb protein structure and form new intramolecular and intermolecular interactions. (A) The interaction of phospho-Thr_{160} of the cyclin-dependent cdk2 protein kinase with three intramolecular arginine residues (Arg_{50}, Arg_{126}, and Arg_{150}) drives a distinct conformation in this region of the kinase.[16] Reproduced from ref. 16, https://doi.org/10.1042/BST0370627, under the terms of a CC BY 4.0 license, https://creativecommons.org/licenses/by/4.0/legalcode. (B) In the CLASP2 protein, P-Ser_{737} and P-Ser_{741} likewise pair electrostatically with arginine guanidinium cationic side chains, in this case nearby either in space or primary sequence. Reproduced from ref. 17 with permission from American Society for Biochemistry & Molecular Biology, Copyright 2012.

building blocks in proteins provide a dianion at physiological pH ranges. Two, Glu and Asp, are mostly monoanionic at physiological pH values.

The newly introduced dianionic P-Ser, P-Thr, and P-Tyr side chains are strong electrostatic attractors for the guanidinium cationic ends of arginine side chains. These bidentate charged interactions are recurring themes in phosphoprotein structures as exemplified in Figure 13.4A. In this


A scaffolding logic
RAC GTPase
scaffolding protein
MLK
three docked kinases
JIP-1
MKK1
JNK
amplified downstream signal
Yeast: S.cerevisiae
B
Ste2
Gpa1
ste20
MAPKKKK
MAPKKK
MAPKK
MAPK
ste11
ste7
fus3
ste5
P P
fus3
mating response
Mammalian Cells
RTK
GPCR
MAPK4
MAPK3
MAPK2
MAPK
JIP ?
P P
MAPK
signaling responses

example the side chain-OH of Thr160 of cyclin dependent kinase2 (CDK2), in complex with its cyclin activating subunit, has become phosphorylated by a kinase.[16] The new P-Thr160 side chain interacts with and collects three arginine guanidinium side chains. These are Arg50 (PDB code 1QMZ), Arg127 (next to the catalytic aspartate) and Arg150 in the activation segment.

Analogously, phosphorylation of the cytoplasmic linker associated protein 2 (CLASP2) protein (tracking the ends of microtubules) on two nearby serine side chains, Ser737and Ser741, induces reorganization (Figure 13.4B) by charge pairing with a total of three arginine residue guanidinium groups,[17] in the process disrupting pre-existing Arg–Glu salt bridges. The protein kinase isoform glycogen synthase kinase3β (GSK3β) inactivates CLASP2 function *via* these two-serine side chain phosphorylations, by disrupting its interactions with kinetochore microtubules during the metaphase portion of the cell cycle.

These charge pairing interactions can be intramolecular within subunits or intermolecular between protein subunits and can immobilize specific conformations. In that context phosphorylations can drive a population of specific phosphoprotein conformers to act as high affinity sites for recognition by and binding to partner proteins. This is emphasized by scaffolding protein domains that can be multiply phosphorylated and also serve as nodes to collect sets of partners that then interact for phosphorylation cascades[18] (Figure 13.5A,B)(*MBC web site, 2017*). The cartoon in Figure 13.5A indicates the three sequentially acting Ser or Thr kinases in the Janus kinase (JNK)–mitogen-activated protein (MAP) kinase cascade are gathered in proximity by the scaffolding protein JNK-interacting protein 1 (jip1).[19] Figure 13.5B is a more elaborate cartoon for a signaling pathway for a membrane-based receptor tyrosine kinase (RTK) that proceeds *via* sequential activation of three comparable Ser or Thr kinases, rapidly accelerated fibrosarcoma (Raf), mitogen activated protein kinase/extracellular signal–regulated kinase kinase 1 or 2 (MEK1/2), and extracellular signal–regulated kinase 1 or 2 (ERK1/2). Continuing the theme of protein-collecting scaffolds, four such kinase way stations are depicted in Figure 13.5C. All of these features sum to protein phosphorylations serving as molecular switches for protein conformation changes and protein–protein interaction effects.

Figure 13.5 Three cartoons of scaffolding proteins that collect individual kinases at docking stations in protein kinase signal cascades. (A) JIP-1 as a docking partner.[20] (B) One of the JIP family members is a kinase docking station for several ERK isoforms in signal transduction from outside the cell to nucleus.[21] (C) The classic example of sterility 5 (Ste5) in yeast for Ste11, Ste7, and fusion 3 (Fus3), the three yeast homologs of mammalian mitogen activated protein kinase kinase kinase (MAPKKK), mitogen activated protein kinase kinase (MAPKK), and mitogen activated protein kinase (MAPK).[22]

13.4 Nine of Twenty Protein Residues Can Be Phosphorylated

The potential 270 000 phosphoprotein isoforms from modification of any or all of the serine, threonine, or tyrosine residues in the human proteome is an amazing level of structural and functional diversity, even as it requires the biosynthesis and expression of the >520 enzymes of the human kinome. The Ser and Thr side chains are related alcohols with nominal pK_a values of between 13 and 14. The phenolic OH of tyrosine has a pK_a some four orders of magnitude lower at ~10. The relevant kinases must activate those three alcohols to alkoxide or phenoxide anions in the transition states for phosphoryl transfers in the respective protein kinase active sites. The kinases typically work on folded protein structures as substrates although there are many loops and conformationally mobile regions of protein targets that offer side chain accessibility to ATP.

Astounding as this expansion of the proteome is by enzymatic phosphorylation of these alcoholic side chain groups, these are only three out of nine proteinogenic side chains that can be enzymatically phosphorylated.[13] They are the three that have been detected by far most commonly in eukaryotic proteomes. However, we have briefly noted the bias in the strategies of MS-based phosphoproteome workflow protocols for acid-stable phosphate monoester linkages.[23]

The other six amino acid residues that can be enzymatically phosphorylated are subcategorized by the identity of the nucleophilic side chains attacking the cosubstrate Mg–ATP at Pγ (Figure 13.6). Two of them provide the fourth and fifth example of oxygen nucleophiles. However, those Asp and Glu side chains are carboxylates, not alcohols. The corresponding phosphorylated species are acyl phosphates not phosphomonoesters and we have dissected their reactivity differences in earlier chapters (Chapters 6 and 9). Of the two sulfur-containing amino acids cysteine and methionine, the thioether sulfur seems inert to *S*-phosphorylation, but the thiolate side chain can be *S*-phosphorylated in about 100 members of the phosphotyrosine phosphatase class of eukaryotic enzymes (Figure 13.6).

The remaining three residues offer nitrogen atoms as the nucleophiles. The *N*-phosphorylated product residues would be phosphoramidates. We have noted the guanidinium group of free arginine is a phosphagen in invertebrates (Chapter 10), and protein *N*-phosphoarginine is also a known, albeit little studied phosphoprotein class. A more likely nitrogen nucleophile is the ε-NH_2 of lysine side chains to give N–P-Lys residues. The third relevant nitrogen side chain is in the heterocyclic imidazole ring of histidine. There are two nitrogens there, N_1 and N_3, and both are known sites of enzymatic phosphorylation. Our discussion of free amino acid phosphoramidates in Chapter 10 noted the unlikely utilization of the caboxamido nitrogens of both glutamine and asparagine in *N*-phosphorylations (in principle expanding the proteinogenic amino acids that can be phosphorylated to

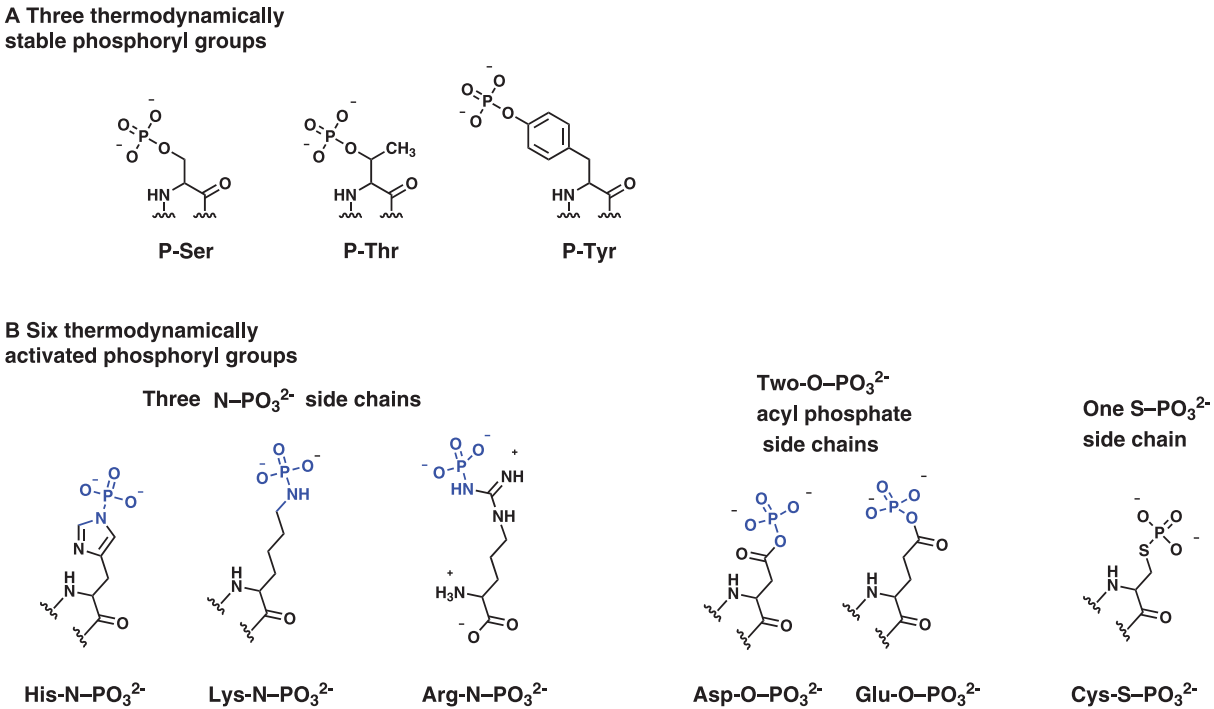


Figure 13.6 Nine of the twenty proteinogenic (building blocks for protein biosynthesis) amino acids can be phosphorylated by ATP on different side chain functional groups. (A) Of those nine, three (P-Ser, P-Thr, and P-Tyr) are acid-stable and the ones detected in contemporary mass spectrometry-based phosphopeptide proteomics workflows. (B) The other six are differentially labile and thermodynamically activated for onward phosphoryl group transfers. They have been under-detected in traditional phosphoproteomic studies but have been seen in recent studies designed specifically to detect these six phosphopeptide types.[23]

11 out of 20). To date, there are no examples of the comparable Gln or Asn *N*-phosphoramides in protein posttranslational modifications. The remaining proteinogenic amino acids do not have functional groups in their side chains.

If we rank order the nine phosphorylated amino acid side chains of phosphoproteins by anticipated thermodynamic activation, we might get the rank order of Figure 13.6. The phosphoramidates would be at the top [free arginine-N–PO_3^{2-} is a widely dispersed metabolic phosphagen (Chapter 10)] as would the mixed acyl phosphoric anhydrides of Asp-β-phosphate and Gln-γ-phosphate. It is also likely that the Cys-S–PO_3^{2-} would have noticeably high phosphoryl group transfer potential.

The only three phosphoprotein groups likely to be stable chemically are the phosphomonoesters of P-Ser, P-Thr, and P-Tyr. For that last set of three, one would expect to have to evolve protein phosphatases to remove them. That is the case in contemporary organisms. The other six might be reactive enough as phosphorylated proteins that they would qualify as transient intermediates in enzymatic catalyses and not need dedicated chemospecific phosphoprotein phosphatases.

Two points emerge from this type of consideration of the chemical reactivity of the nine phosphorylated amino acid residues of Figure 13.6. The first is the well-known realization that the current phosphoproteomes, extensive as they are, are likely to be incomplete, based on the bias of MS-based workflow for detecting only P-Ser, P-Thr, and P-Tyr in peptide fragments.[13,23] Indeed, isolation methods restructured to detect base-stable, acid-labile phosphoramidate linkages in proteins reveal in initial range-finding studies hundreds of sites of P-Arg, P-His, and even P-Lys modifications. These may gain prominence as their biology gets uncovered in phosphoproteomics 2.0.

The second point guides the organization of the next two chapters in this section. We turn first to those understudied six phospho-amino acid residues (phosphoramidates, acyl phosphates, and phosphorothioates) that should have enough thermodynamic activation to function as transient phosphoryl intermediates in catalytic cycles, and not require dedicated protein phosphatases for their removal. Then, we turn back to the known universe of P-Ser, P-Thr, and P-Tyr residues for a summary of some of the central features of the many thousands of papers on their biological import.

13.5 Criteria for Phosphoproteomics 2.0

One starting point for a phosphoproteomics 2.0 is the report by Hardman *et al.*[23] analyzing phosphoprotein content of HeLa cells by treating the protein extract with the protease trypsin to yield a mix of tryptic peptides. Then the phosphopeptides were separated and purified by strong anion exchange chromatography at pH 6.8. Mass spectrometric analysis captured

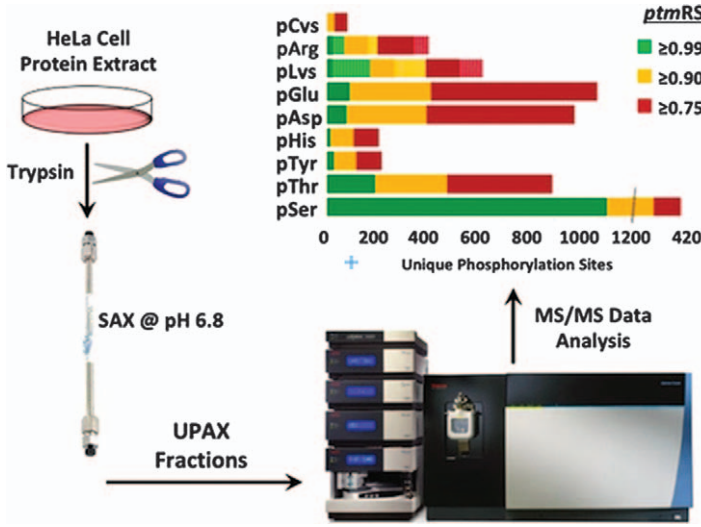


Figure 13.7 Scheme of a phosphopeptide workflow of proteins from HeLa cells and shows a bar chart of the nine phosphopeptide chemotypes detected by MS–MS data analysis. Most notable are the high abundance of P-Arg, P-Lys, P-Glu, and P-Asp phosphopeptide categories compared with rare occurrence in other data sets with different workflows. The headings pCvs and pLvs should be read as pCys and pLys. Reproduced from ref. 23, https://doi.org/10.15252/embj.2018100847, under the terms of a CC BY 4.0 license, https://creativecommons.org/licenses/by/4.0/.

all nine anticipated chemical classes of phosphopeptides, as summarized in Figure 13.7.

While *S*-phospho-Cys peptides were in the rarest category, consistent with prior experience, P-Arg, P-Lys, and P-His phosphoramidate linkages survived this workflow to yield ~500, 800, and 200 unique phosphopeptides each. The total of ~1500 N–PO_3^{2-} protein forms gives credence to the concept of a substantial missing fraction of phosphoproteomes.

Equally surprising was the large number of peptide acyl phosphate species, P-Glu and P-Asp, each numbering in the ~1000 phosphopeptide levels, auguring for much discovery in this activated class of phosphopeptides.

The three stable, canonical alcoholic phosphopeptides P-Tyr, P-Thr, and P-Ser were also obtainable by this workflow. The P-Tyr peptide level at ~200 unique phosphopeptides equaled the P-His peptide yield but was unexpectedly dwarfed by P-Arg, P-Lys, P-Asp, and P-Glu peptides. In accordance with expectations the P-Ser phosphopeptides were most abundant. All told, the set of six noncanonical phosphopeptides from the HeLa cell proteome was about one third the level of the canonical phosphoproteome version 1.0. These findings set the stage for the next chapter on noncanonical phosphoproteomics.

References

1. J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. G. Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. D. Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R.-R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Y. Wang, A. Wang, X. Wang, J. Wang, M.-H. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. C. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.-H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigó, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stin, P. Caulk, Y.-H. Chiang, M. Coyne, C. Dahlke, A. D. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodel, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter,

M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu, The sequence of the human genome, *Science*, 2001, **291**, 1304–1351.
2. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody and J. Baldwin, *et al.*, Initial sequencing and analysis of the human genomes: International Human Genome Cosortium, *Nature*, 2001, **409**, 860–921.
3. N. L. Anderson and N. G. Anderson, Proteome and proteomics: new technologies, new concepts, new words, *Electrophoresis*, 1998, **19**, 1853–1861.
4. W. P. Blackstock and M. P. Weir, Proteomics: quntitative and physical mapping of cellular proteins, *Trends Biotechnol.*, 1999, **17**, 121–127.
5. R. Breitling, What is systems biology?, *Front. Physiol.*, 2010, **1**, 159.
6. Chemometrics for enviromental omics (downloaded 10-31-2019).
7. C. T. Walsh, *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, Roberts & Co, Englewood, Colorado, 2006.
8. G. Duan and D. Walther, The roles of post-translational modifications in the context of protein interaction networks, *PLoS Comput. Biol.*, 2015, **11**, e1004049.
9. G. A. Khoury, R. C. Baliban and C. A. Floudas, Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database, *Sci. Rep.*, 2011, **1**, 90.
10. S. J. Humphrey, O. Karayel, D. E. James and M. Mann, High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform, *Nat. Protoc.*, 2018, **13**, 1897–1916.
11. V. Mayya and D. K. Han, Phosphoproteomics by mass spectrometry: insights, impicatons, applications, and limitations, *Expert Rev. Proteomics*, 2010, **6**, 605–618.
12. E. J. Needham, B. L. Parker, T. Burykin, D. E. James and S. J. Humphrey, Illuminating the dark phosphoproteome, *Sci. Signaling*, 2019, **12**, eaau8645.
13. K. Adam and T. Hunter, Histidine kinases and the missing phosphoproteome from prokaryotes to eukaryotes, *Lab. Invest.*, 2018, **98**, 233–247.
14. F. H. Westheimer, Why nature chose phosphates, *Science*, 1987, **235**, 1173–1178.
15. T. Hunter, Why nature chose phosphate to modify proteins, *Phil*, *Trans R. Soc. B*, 2012, **367**, 2513–2516.
16. L. N. Johnson, The regulation of protein phosphorylation, *Biochem. Soc. Trans.*, 2009, **37**, 627–641.
17. P. Kumar, M. S. Chimenti, H. Pemble, A. Schonichen, O. Thompson, M. P. Jacobson and T. Wittman, Multiple phosphorylation disrupts arginine-glutamate salt bridge networks required for binding cytoplasmic linker-associated protein2 (CLASP2) to end-binding protein 1 (EBP1), *J. Biol. Chem.*, 2012, **287**, 17050–17064.
18. J. A. Printen and G. F. Sprague, Protein-protein interactions in the yeast pheromone response pathway: Ste5p interacts with all members of the MAP kinase cascade, *Genetics*, 1994, **138**, 609–619.
19. T. Q. Nguyen, M. Aumont-Nicaise, J. Andreani, C. Vellours, M. Chenon, F. Vilela, C. Geneste, P. F. Varela, P. Llinase and J. Menetrey,

Characterization of the binding mode of JNK-interacting protein (JIP1) to kinesin light chain 1 (KLC1), *J. Biol. Chem.*, 2018, **293**, 13946–13960.
20. Webpage B for Biology (downloaded 11-04-2019) signal transduction pathway map, post for Tuesday March 12, 2013.
21. M. Meister, A. Tomasovic, A. Banning and R. Tikkanen, Mitogen-activated protein (MAP) kinase scaffolding proteins: a recount, *Int. J. Mol. Sci.*, 2013, **14**, 4854–4884.
22. D. N. Dhanasekaran, K. Kashef, C. M. Lee, H. Xu and E. P. Reddy, Scaffold proteins of MAP-kinase modules, *Oncogene*, 2007, **26**, 3185–3202.
23. G. Hardman, S. Perkins, P. J. Brwnridge, C. J. Clarke, D. P. Byrne, A. E. Campbell, A. Kalyuzhnyy, A. Myall, P. A. Eyers, A. R. Jones and C. E. Eyers, Strong-anion exchange-mediated phosphoproteomics reveals extensive human noncanonical phosphorylation, *EMBO J.*, 2019, **38**(21), e100847.

CHAPTER 14

Noncanonical Phosphoproteomes

In considering the universe of phosphorylated proteins, one can categorize them by the identity of the amino acid side chain phosphorylated. One can also categorize them by the lifetime in cells, a mix of chemical and biological stability. The two modes of categorization turn out to be correlated and lead to distinct catalytic *vs.* regulatory functions. Short-lived phosphoproteins are unlikely to be counted in phosphoproteomic analyses but may be crucial intermediates in catalytic logic throughout biology.

14.1 Types of Phosphate Scaffolds in Biology

In our prior discussions of the universe of low molecular weight phosphorus-containing metabolites, we noted that organisms can combine the readily available inorganic phosphate monoanion–dianion mixture to phosphoric anhydrides, phosphomonoesters, and phosphodiesters (*e.g.* RNA, and DNA). We also noted enol phosphates, N–P bonds (phosphoramidates), C–P bonds (phosphonates), and S–P bonds (phosphorothioates). One can step back and inquire which of these bonding strategies are found in phosphoproteins and what are the chemical consequences.

To date no C–P phosphonate bonds have been found in phosphoproteins, consistent with the lack of candidates for carbon nucleophiles in proteins (one might have anticipated C-phosphorylations of tryptophan, given the existence of indole natural products prenylated at every carbon and the nitrogen of indole, but not so). Nor have enol phosphates been detected to date in phosphoproteomes. This accords with the lack of enolizable ketones and aldehydes in the side chains of normal proteomes. Surprisingly, there has been evidence for a lone *pyrophosphoryl*-histidine residue in the enzyme literature since 1968.[1] We will examine the formation and breakdown of this

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

singular phosphoric anhydride-containing adduct later in discussion of the bacterial enzyme pyruvate phosphate dikinase.

We turn in this chapter first to the phosphoramidates, then the protein bound acyl phosphates, and finally the Cys–S–PO_3^{2-} phosphoproteomes. Relative to the conventional P-Ser, P-Thr, and P-Tyr residues of eukaryotic phosphoproteomes these phosphoproteomes are understudied, perhaps underappreciated. The chemical biology of low molecular weight acyl phosphates noted in Chapter 6, of phosphoramidates in Chapter 10 and of phosphorothioates in Chapter 12 provide precedents on thermodynamics and kinetic reactivities of these three classes of phosphoproteins.

The results of the recent study by Hardman *et al.*,[2] discussed at the end of the preceding chapter, indicate that they are much more widespread than conventionally assumed. The ~1500 N–PO_3^{2-} unique phosphopeptides in HeLa cell proteins indicates a broad biological involvement that prompts examination of their likely chemical properties.

The thermodynamic activation of covalent protein–phosphoramidates, protein–acyl phosphates and protein–thiophosphates make them more suited to transient adducts and thereby more likely to qualify as kinetically and chemically competent reaction intermediates.[3] In many cases they may not need protein phosphatases to effect dephosphorylations. In contrast, the alcoholic phosphomonoesters in P-Ser, P-Thr, and P-Tyr are deactivated covalent adducts, stable, need dedicated enzymes for their reversal, and are adapted almost exclusively to regulatory, noncatalytic functions of stable, expanded (largely) eukaryotic proteomes.

14.2 The *N*-phosphoproteomes: Phosphohistidine Residues

14.2.1 Bacterial Two Component Sensor–Response Regulator Systems

The best studied context for phosphorylated histidine residues in proteins has been the two-component sensor–response regulator systems in bacteria (Figure 14.1). First detected in 1980, they are termed two-component systems[4] because a transmembrane sensor protein is a first component receptor for some environmental chemical or condition. It could be pH, osmotic pressure, heat, or a defined molecular framework. Often, it is the sensing of inorganic ions, from Mg^{2+} cations to PO_4^{2-} dianions. The second component is typically a cytoplasmic partner protein that is a conditional transcriptional activator or repressor. Signals are transmitted from outside the cell by the transmembrane sensor kinase to the response regulator protein for selective modulation of a set of bacterial genes.

Protein kinases come in two major architectures based on their cellular functions. Most are intracellular, have a relatively conserved catalytic domain and varying amounts of other domains that may regulate activity,

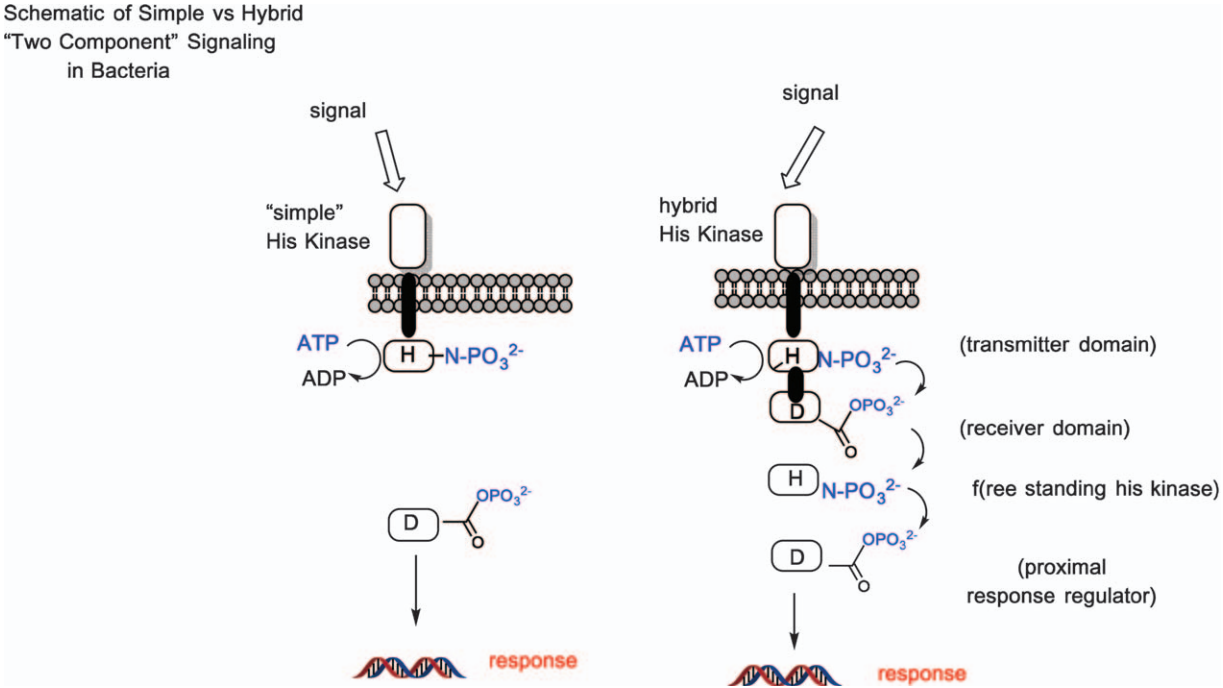


Figure 14.1 Bacterial two-component sensor protein schematic architecture. Left, a transmembrane histidine autokinase receives a signal on the extramembrane domains and autophosphorylates on a histidine residue. An aspartate side chain in a cognate response regulator protein attacks and transfers the $-PO_3^{2-}$ from the P-His residue. The resultant P-Asp protein with altered conformation transmits the response. Right, iteration of domains in the cytoplasmic domain of the sensor kinase allows tandem transfer of the $-PO_3^{2-}$ group between domains in the route to transmittal of response.[5]

control subcellular location, and provide recognition surfaces, *e.g.* for homodimerizations or heterodimerizations or higher order complexes. The second major kinase category contains transmembrane kinase architectures in both prokaryotic and eukaryotic membranes.[5–7] Typically, they have an extracellular sensor domain for binding an incoming ligand, a transmembrane domain, often largely helical, that communicates ligand occupancy from outside to inside, then intracellular domains that include the catalytic kinase domain and perhaps additional regulatory domains. Often the transmembrane kinases are dimers in their active form and catalyze cytoplasmic autophosphorylations to recruit partner proteins in signal cascades.

The bacterial sensor histidine kinases are transmembrane in function and architecture. (Figure 14.2). However, the predominant class of eukaryotic transmembrane kinases are instead protein tyrosine kinases[8–10] (Chapter 15). The schematic of the prototypical bacterial transmembrane histidine autokinase as environmental sensor in Figure 14.2 is of PhoQ, where Pho is an abbreviation for phosphate.[11] The periplasmic domain is the sensor for inorganic phosphate. The kinase domain is in the wing of the cytoplasmic domain. The histidine residue that becomes phosphorylated is His277 and it is the portal out to the partner protein PhoP.

Gram-negative bacteria have tens to many dozens of distinct transmembrane sensor proteins per cell.[4] They transmit occupancy of distinct ligands in a periplasmic domain accessible to the outside medium. Figure 14.3 shows a representative transmembrane histidine kinase in the cytoplasmic membrane of the Gram-negative bacterium *E. coli*. Envelope Z (EnvZ) is a transmembrane osmotic strength reader. It is paired with a cognate cytoplasmic response regulator outer membrane porin R (OmpR). OmpR, when activated by phosphorylation, is a direct transcriptional activator and controls whether the genes for outer membrane proteins OmpC or instead OmpF are expressed to remodel the size of the passageway through the outer membrane pores.

Occupancy of the periplasmic domain of a sensor kinase is transmitted down paired transmembrane helices to a juxtamembrane domain on the inside of the membrane[7] (Figures 14.4 and 14.5). In turn that domain interacts with a kinase domain that uses ATP to *autophosphorylate* itself. The residue that undergoes autophosphorylation on the adjacent dimerization and histidine phosphorylation (DHP) domain can be *in cis* [on the same subunit – see VicK in Figure 14.5] or *in trans* (on the adjacent dimer – as in EnvZ[15,16]), depending on the twist of the helical segment in the DHP domain that contains the His residue to be phosphorylated. Autophosphorylation rates can be slow, from 0.01 to 5 min^{-1}.

Each transmembrane bacterial sensor histidine kinase has a paired response regulator partner protein (Figures 14.3 and 14.6). Specificity is imposed structurally at the tip of the DHP domain and is mediated both by dissociation constant (K_D) values for the sensor–response regulator pair and by mass action, with response regulator often in molar excess.[4,7] Although there are variant mechanisms and functions, most response regulators are

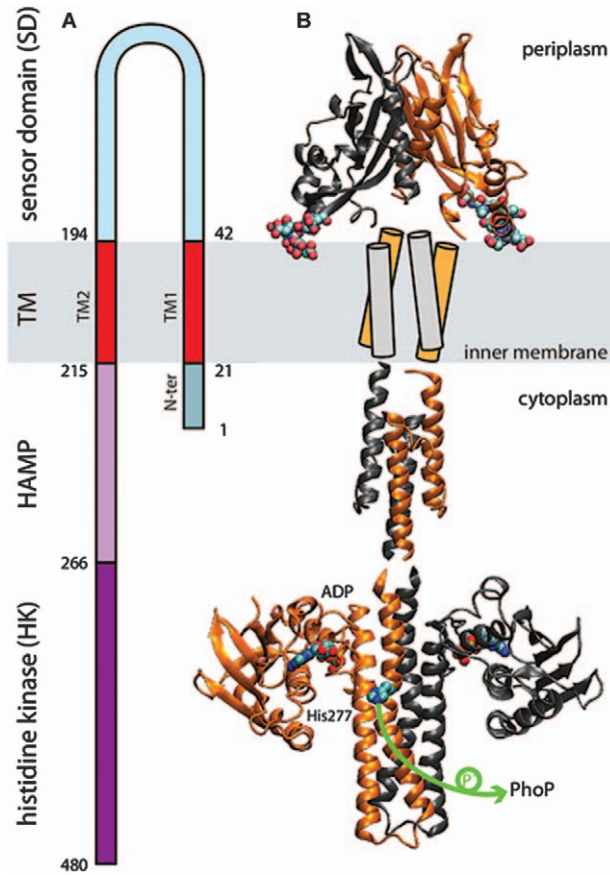


Figure 14.2 Schematic of organization of the bacterial PhoQ transmembrane kinase for inorganic phosphate sensing. (A) The hairpin represents the distinct domains in PhoQ with two transmembrane domains (TM1, and TM2). (B) A dimeric active form of PhoQ showing the ATP site, the location of His277 that undergoes phosphorylation from bound ATP. The PHis277 form is then the phosphoryl donor transfer to a specific aspartate side chain in PhoP. Reproduced with permission from.[12]

DNA binding proteins, where binding to a particular promoter region of DNA controls expression of sets of genes that mount the particular physiological responses of the bacterium.

Transmission of the signal from the phospho-His form of a sensor kinase to its response regulator involves (electrophilic) phosphoryl group transfer to a conserved aspartyl residue carboxylate side chain in the receiver domain (Figure 14.6). The introduction of the phosphate (in acyl phosphate linkage) drives a conformational change in the DNA binding region to alter

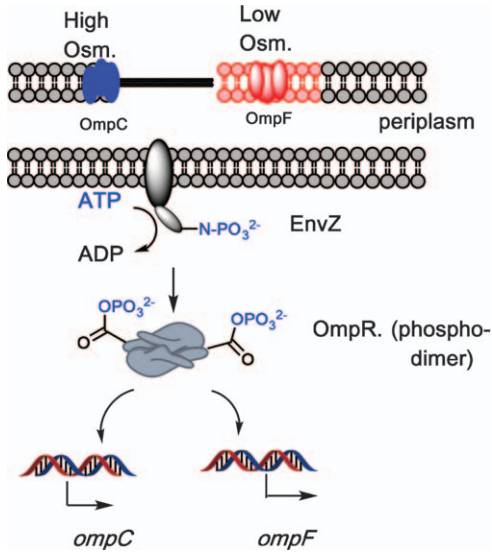


Figure 14.3 Schematic of the EnvZ–OmpR two-component system for sensing and responding to changes in environmental osmotic pressure by altered expression of outer membrane porins. OmpR is a transcriptional regulator, active in the phosphorylated state as a dimer for transcription of porin genes.[13,14]

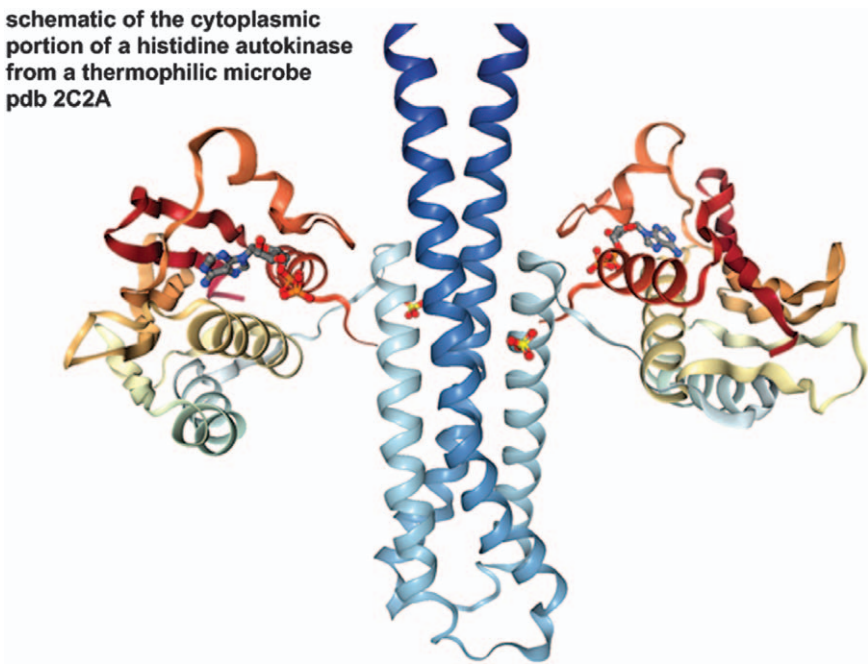


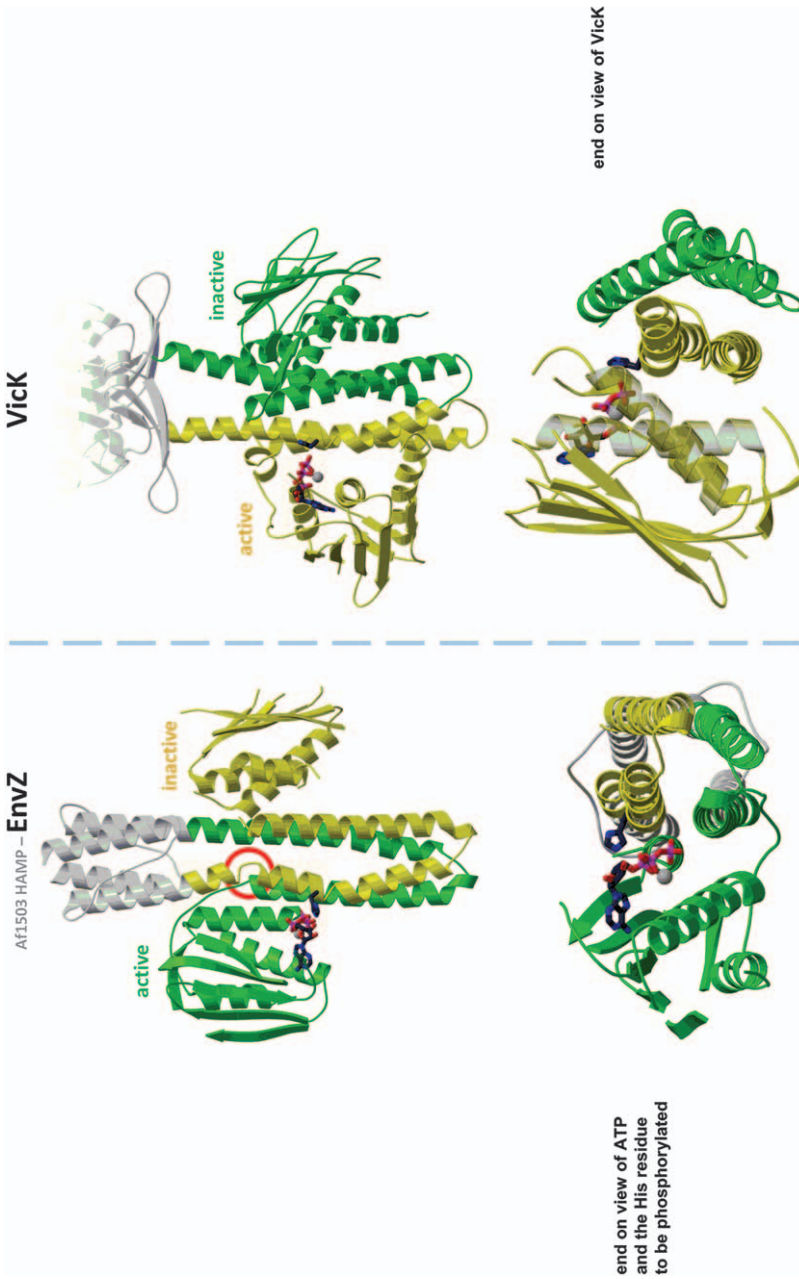
Figure 14.4 The cytoplasmic domains of dimers of a thermophilic microbial histidine kinase. The active kinase domain brings ATP in proximity to the His residue in the central helix.

interaction with target DNA. Often, the phosphorylated response regulator undergoes dimerization and thereby conversion to the DNA-binding conformer (Figure 14.6). Depending on the specific sensor kinase–response regulator system, specific gene expression can be turned on or turned off. With dozens of sensor kinase–response regulators per cell bacteria can orchestrate real time complex behavior and physiological responses to many environmental signals.[4]

Figure 14.7 shows in cartoon form three such two-component systems in Salmonella bacteria. The PhoQ–PhoR two component pair is involved in sensing of extracellular inorganic phosphate. The tandem phosphoryl transfers from ATP to P-His-PhoQ and on to P-Asp-PhoR, turning on target genes for external phosphoester hydrolysis and Pi uptake. The EnvZ–OmpR pair constitute the sensing of osmotic pressure changes. The P-Asp form of OmpR is a transcriptional regulator for outer membrane porin production and porin class change. The third pair, small stable RNA A and B (SsrA–SsrB), turn on hybrid transfer messenger RNA (tmRNA) production to shut down protein synthesis under starvation conditions.

The chemistry of the interprotein signal transmission is attack by the aspartyl carboxylate on the phosphoryl group attached to the imidazole group nitrogen of the histidine in the DHP region of the sensor kinase (Figure 14.8). These have been measured at up to 20 phosphoryl transfers per minute. The $-PO_3^{2-}$ group transfer from imidazole to carboxylate in the response regulator partner validates that the N–P phosphoramidate linkage in the phosphohistidine residues is thermodynamically activated. Albeit relatively slow on an absolute time scale compared with many enzyme rates, the P-His in the sensor kinases qualifies as a competent reaction intermediate. The P-Asp residue in response regulators is more akin to a classic posttranslational phosphoprotein, albeit with variable stability. The dramatic conformation change on phosphorylation (*e.g.* protein dimerization) communicated to the DNA binding domain emphasizes the power of introduction of a tetrahedral dianionic PO_3^{2-} group to alter conformation and biological function-in these cases transcriptional activation or repression (Figure 14.8).

Figure 14.1 on the left side schematizes the architecture of a simple two component circuit with one sensor domain and one response domain. More complex domain architecture has evolved, *e.g.* by domain duplications and insertions.[4] The right-hand side sketches the sequential movement of one phosphoryl group from the P-His residue to an aspartyl side chain within the same protein. Then, it is transferred out to a separate protein on a histidine residue and then interfaces with a response regulator. Counting the initial $-PO_3^{2-}$ transfer from ATP to the first histidine (not shown), the *same* $-PO_3^{2-}$ *group* is transferred *four times* (His to Asp to His to Asp) in a cascade before reaching the cognate response regulator or DNA transcription factor. This cascade, and related ones on other hybrid histidine kinases with fused receiver domains, emphasizes the reversible transfer of thermodynamically activated $-PO_3^{2-}$ groups. At each of the interdomain transfers there is


VicK
inactive
active
end on view of VicK
Af1503 HAMP – EnvZ
inactive
active
end on view of ATP
and the His residue
to be phosphorylated

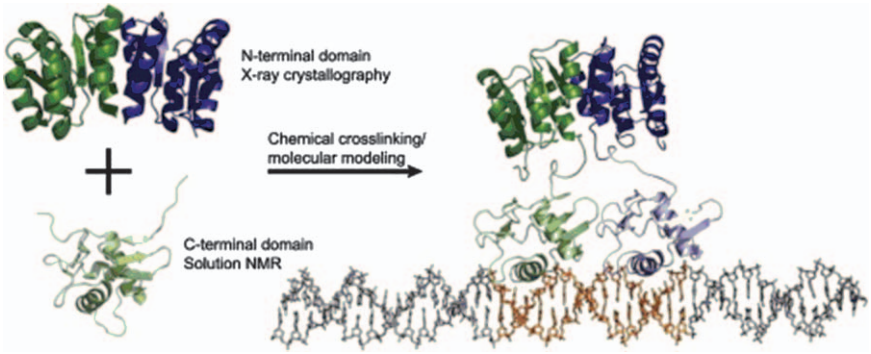


Figure 14.6 Diagrammatic conversion of an inactive monomeric response regulator to an active aspartyl-phosphate dimeric form that can bind its target DNA as a transcriptional activator.[17]
Reproduced from ref. 17 with permission from Elsevier, Copyright 2018.

opportunity for $-PO_3^{2-}$ transfer in from another signaling system to allow for integration and/or cross talk (including competing phosphatase catalytic removal of the phosphoryl group signal).

The logic of these two-component phosphoryl transfer kinase–gene regulators is explicit tandem transfer of a single $-PO_3^{2-}$ group from ATP to imidazole nitrogen and then on to the acyl phosphate group in the second protein. This is not widely replicated in *eukaryotic phosphoproteomics, where tandem phosphoryl transfers are not at all common*. This may be as much a consequence of looking (under the lamp post?) only at the stable phosphate esters in P-Ser, P-Thr, and P-Tyr, as they are not thermodynamically activated for further phosphoryl group transfers.

Indeed, the use of an aspartyl side chain carboxylate in response regulator logic provides a built-in timer for the signal to disappear. Acyl phosphate lifetimes in these proteins can vary from seconds to hours, based on local microenvironment around the aspartyl-phosphate side chain and the accessibility of water in its active site.[4] In addition, some of the sensor kinases show acyl phosphatase hydrolytic activity to accelerate the transfer of the acyl phosphate group to water to release inorganic phosphate and by default turn off downstream gene regulations.

Figure 14.5 The cytoplasmic domains of dimers of an EnvZ–Af1503 hybrid and dimers of VicK, an essential histidine autokinase in several types of streptococci. For EnvZ the His residue is in the opposite subunit of the dimer. For VicK, the histidine nucleophile is in the same chain as the ATP domain. Thus, EnvZ acts *in trans* while VicK acts *in cis to* generate the P-His residues. The bottom views are orthogonal and looking at the two pairs of helices in the dimers end on.
Reproduced from ref. 16 with permission from Elsevier, Copyright 2014.

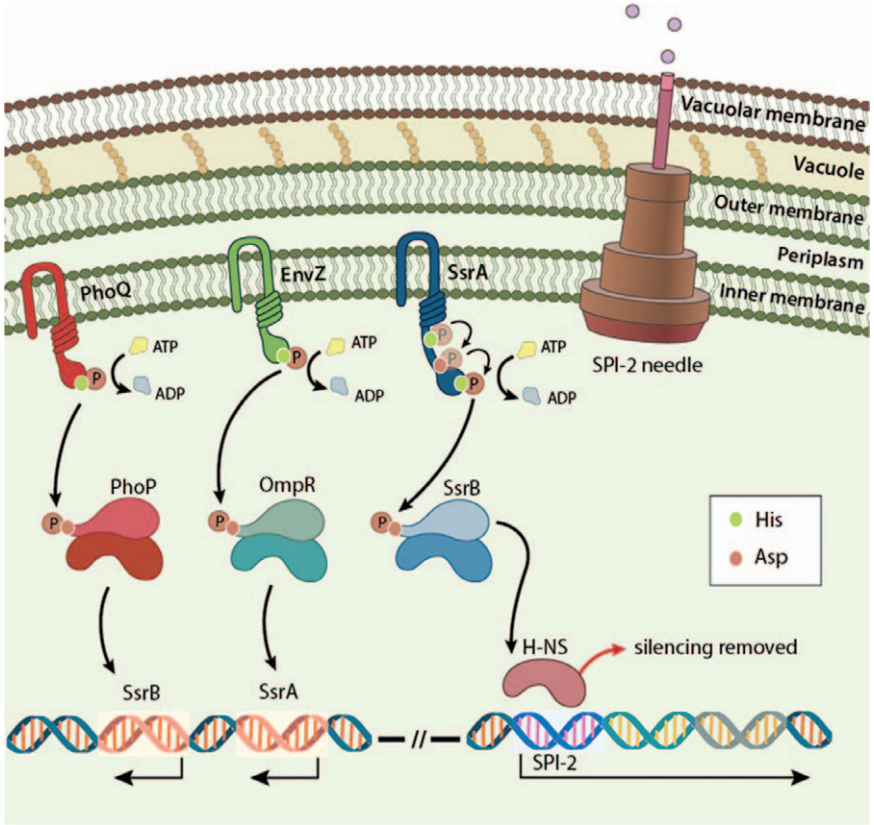


Figure 14.7 Schematic of three parallel two component systems in Salmonella. PhoQ–PhoP for inorganic phosphate, EnvZ–OmpR for osmotic pressure, and SSrA–SSrB for nutrient limitation stress. Phosphorylated forms of PhoP and OmpR are transcriptional activators, SsrB displaces a repressor.
Reproduced from ref. 13, image provided courtesy of the Mechanobiology Institute, National University of Singapore.

Three further conclusions emerge. First, the tandem phosphoryl group transfers involve two thermodynamically activated phosphoryl species: phosphohoramidates and acyl phosphates. This allows for their chemical competence to act as reaction intermediates in phosphoryl transfers (most notably in the hybrid sensor kinase–response regulator cascades in the right-hand side of Figure 14.1).

Second, the thermodynamic lability ultimately imparts some kinetic lability to hydrolysis. Therefore, one would *not* expect the need for large families of protein phosphatases to reverse the phosphorylation state of these two classes of phosphoproteins. (There are a few acyl phosphatases and protein histidine phosphatases but nothing like the cadre of protein phosphatases for P-Ser, P-Thr, and P-Tyr residues noted in Chapter 15).

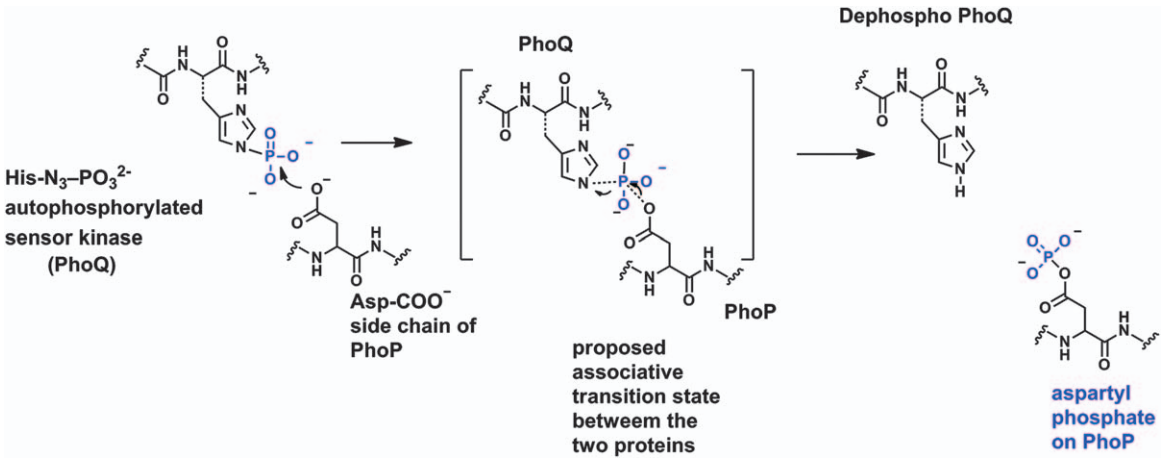


Figure 14.8 Interprotein phosphoryl transfer of a phosphoryl group from the phosphoramidate *N*-phosphoryl-His of a sensor kinase to the β-aspartyl acyl phosphate residue of a response regulator.

Third is the question of why this widespread tandem phosphohistidine–phosphoaspartate signaling logic so pervasive in microbes, has not been found or widely utilized in multicellular eukaryotes. One suggestion has been that the short time frame and duration of the responses (a consequence, at least in part, of the intrinsic lability of aspartyl acyl phosphoric anhydrides to chemical hydrolysis) may not be compatible with longer time constants for signaling in animal and human cells and tissues. The chemically more stable P-Ser, P-Thr, P-Tyr phosphoesters may be better suited for responses slow to develop and of longer duration.

In such a scenario the transmembrane sensor histidine kinase architecture and logic may have been forerunners of similar transmembrane receptor tyrosine kinases. (Bacteria *do* express transmembrane tyrosine kinases[18] as well as His, Arg, and Ser kinases). Both the bacterial transmembrane histidine kinases and the eukaryotic transmembrane tyrosine kinases are almost exclusively *autokinases*, phosphorylating an internal residue (His or Tyr) and typically stopping at that point. It is also possible that some P-His–P-Asp tandem $-PO_3^{2-}$ transfers are at work in eukaryotic phosphoproteomes but have been going on in the dark regions of the phsosphoproteomes. There are those recently reported, as yet unattributed, ~2000 P-Asp and P-Glu peptides detected by Hardman *et al.*[2] in their modified mass spectrometry workflows.

Finally, recent bioinformatic compilations on some 2000 eukaryotic protein kinases have indicated that the explosion in P-Ser, P-Thr, and P-Tyr kinases happened early after the prokaryote to eukaryote shift.[19,20] Eight major protein kinase groups emerged or evolved rapidly and then gave rise to the subgroups that populate the current P-Ser, P-Thr, and P-Tyr kinome tree[21] (Chapter 15).

14.2.2 Tandem Phosphoryl Transfers in Bacterial Phosphotransferse Systems for Coupled Import and Phosphorylation of Sugars

The first report of phosphoenolpyruvate serving as phosphoryl donor, instead of ATP, during phosphorylation of hexose imported into *E. coli* was in 1964.[22] Fifty years later, knowledge of the (hexose) phosphotransferase system (PTS) proteins has expanded to several auxiliary protein components. In aggregate, they control not only sugar transport and phosphorylation in bacteria, but also a variety of transcription factors, potassium transport, expression of catabolic enzymes, virulence, biofilm formation, phosphate and nitrogen utilization, and adenylate cyclase activity.[23]

The central components of PTS permease systems include two peripheral membrane proteins or domains (Enz IIA, IIB) and an integral membrane protein or domain (Enz IIC), that can be linked together in a single polypeptide or be fully separate. Additionally, there is a phosphoryl transferase (Enz I) mobilizing the $-PO_3^{2-}$ group from cytoplasmic PEP[23,24] (Figure 14.9).

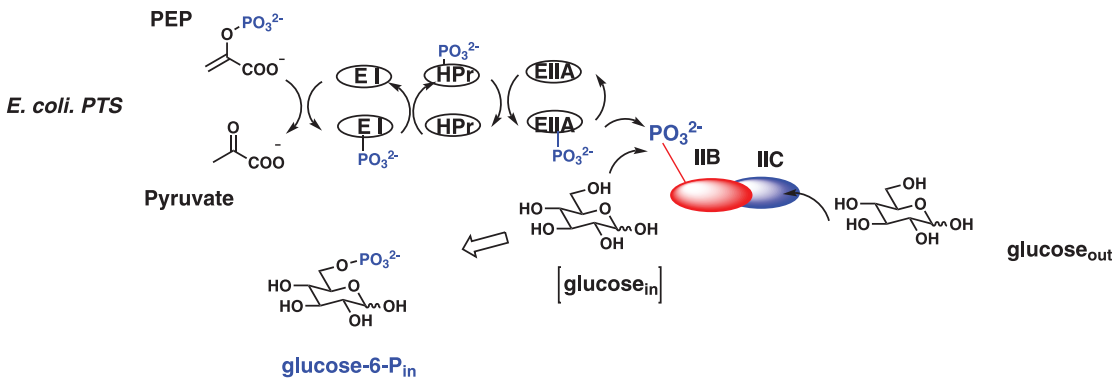


Figure 14.9 Schematic of tandem transfer of the same $-PO_3^{2-}$ group for PEP to four proteins in the *E. coli* and *B. subtilis* phosphotransferase complexes as glucose is captured as glucose-6-phosphate (the fifth $-PO_3^{2-}$ transfer) inside bacterial cells.

The itinerant $-PO_3$ can undergo *four tandem transfers*, the latter three between imidazole side chains of histidine residues on distinct enzyme II forms. The fifth and final transfer of the same $-PO_3^{2-}$ is to the alcohol of the sugar being transported into the cell and irreversibly trapped there by covalent attachment of the dianionic phosphate ester group (Figure 14.10).

The order of covalent phosphoenzyme formation and decay is typically Enz I, histidine protein (HPr), Enz IIA, Enz IIB, to sugar. The four phosphoryl enzymes in some bacteria are all histidine phosphoramidates (Figure 14.11). For example, the active site His in Enz I is His189 and it attacks PEP. The next transfer is between imidazole rings on two proteins. His15 of the small (9.5 kDa) HPr[22] attacks the $N-PO_3^{2-}$ on enzyme I (Figure 14.10). Two more imidazolyl N–P transfers *via* IIA and IIB proteins finally put the still

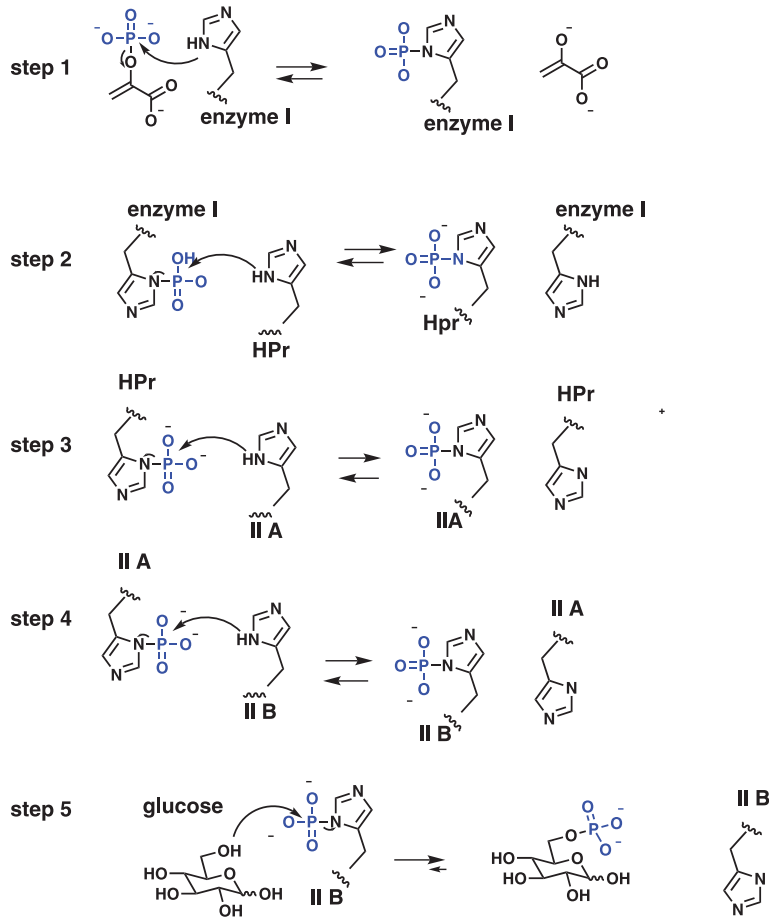


Figure 14.10 The five steps involving $-PO_3^{2-}$ tandem transfers: first from the trapped enol in PEP, then 4-distinct *N*-phosphoryl histidine residues in four different proteins before the ''downhill'' transfer to C6–OH of glucose.

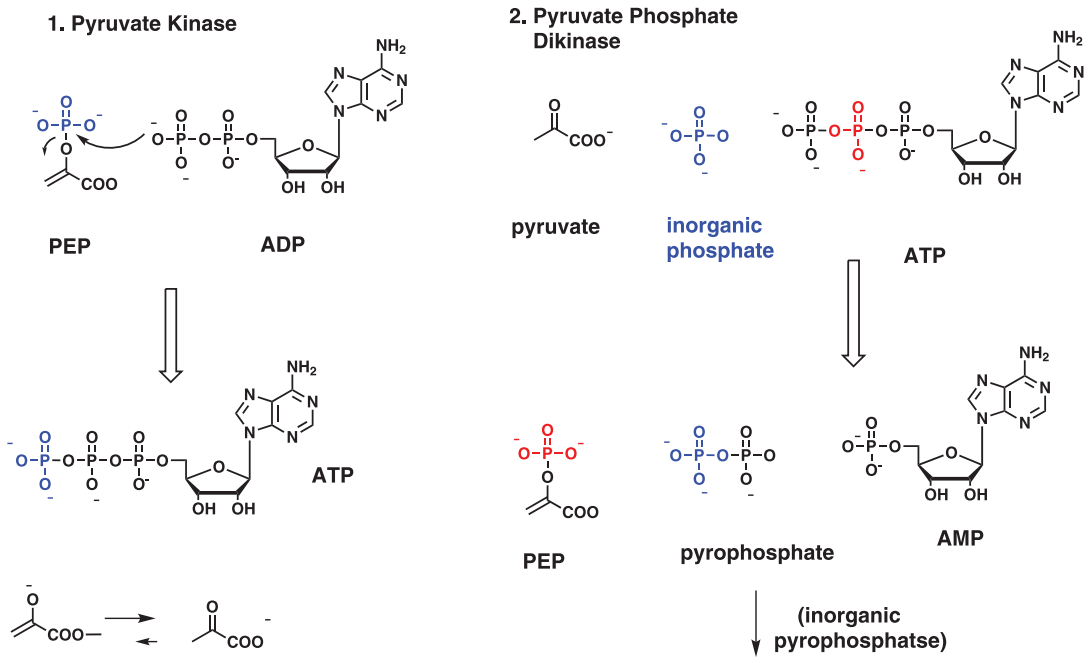


Figure 14.11 Comparison of the reaction stoichiometries for pyruvate kinase and pyruvate phosphate dikinase. Pyruvate kinase runs in the direction PEP + ADP to pyruvate and ATP. Pyruvate phosphate dikinase runs in the direction of pyruvate to PEP and carries out both phosphoryl and pyrophosphoryl transfers, cleaving ATP to AMP and PPi. The sources of the phosphoryl moieties in PEP and PPi indicate a complex reaction mechanism.

activated –N–PO_3^{2-} group on the protein that has binding affinity for the sugar that has come through the transmembrane IIC protein transporter and is finally phosphorylated and trapped as it emerges from transmembrane passage, as shown for glucose on the C6–OH group.

The sugars undergoing transport fall into four families, of which the largest family acts on glucose, fructose, and lactose. There is also an ascorbate and galactitol, a mannose, and then a dihydroxyacetone family. Each of the families has their own variant type II protein components.[25]

This phosphoryl transfer *cascade* (five transfers of the same phosphoryl group) is remarkable both in its use of PEP instead of ATP as an electrophilic phosphoryl donor and as a testament to the thermodynamic activation of the four intermediate protein histidinyl N–P phosphoramidate bonds. The final transfer from an Enz IIB–N–PO_3^{2-} generates a ''low'' energy phosphate monoester bond on the hexose. As denoted in the last line of Figure 14.10, while PEP is the ultimate donor of the electrophilic phosphoryl group, the actual, proximal transfer step to the C6–OH of glucose bound to protein IIB is of the N–PO_3^{2-} of the histidine moiety in enzyme IIB.

Both this cascade and the cascades in the two-component sensor kinase–response regulator systems above illustrate the chemical competence of histidinyl phosphoramidates to carry out both intramolecular and intermolecular catalytic phosphoryl transfers between protein components. They stand as existence proofs that such, as yet unexamined, tandem phosphoryl group transfers could be occurring in eukaryotes.

14.2.3 Bacterial Pyruvate Phosphate Dikinase: A Relative of PTS Logic?

A third example of bacterial metabolism involving a protein phosphohistidine with catalytic importance is the enzyme pyruvate phosphate dikinase. It is at a metabolic branchpoint in both plant and bacterial metabolism to get around the energetically irreversible block in pyruvate kinase and send pyruvate uphill energetically to PEP (Figure 14.11). Thus, the pyruvate phosphate dikinase couples ATP and PEP as intersecting activated –PO_3^{2-} donors, involves histidine phosphorylation, but additionally involves histidine-*N*-pyrophosphorylation. The dikinase falls into the rare category of pyrophosphoryl transferase (see Chapter 5). It is an even more rare example of pyrophosphoryl enzyme formation, in particular of a *pyrophosphoramidate* linkage. From a stoichiometry perspective, pyruvate and the cosubstrate inorganic phosphate are converted into the activated enol phosphate linkage of PEP as ATP is split to AMP and PPi. The source of the PO_3^{2-} incorporated into PEP, however, originates as the *beta* phosphoryl group of ATP not inorganic phosphate.[1]

From its discovery in 1968 it was clear that an enzyme residue, later shown to be the active site histidine's imidazole nitrogen, attacked Pβ of ATP. This step yields free AMP (an observed product) and the pyrophosphoramidinyl-enzyme

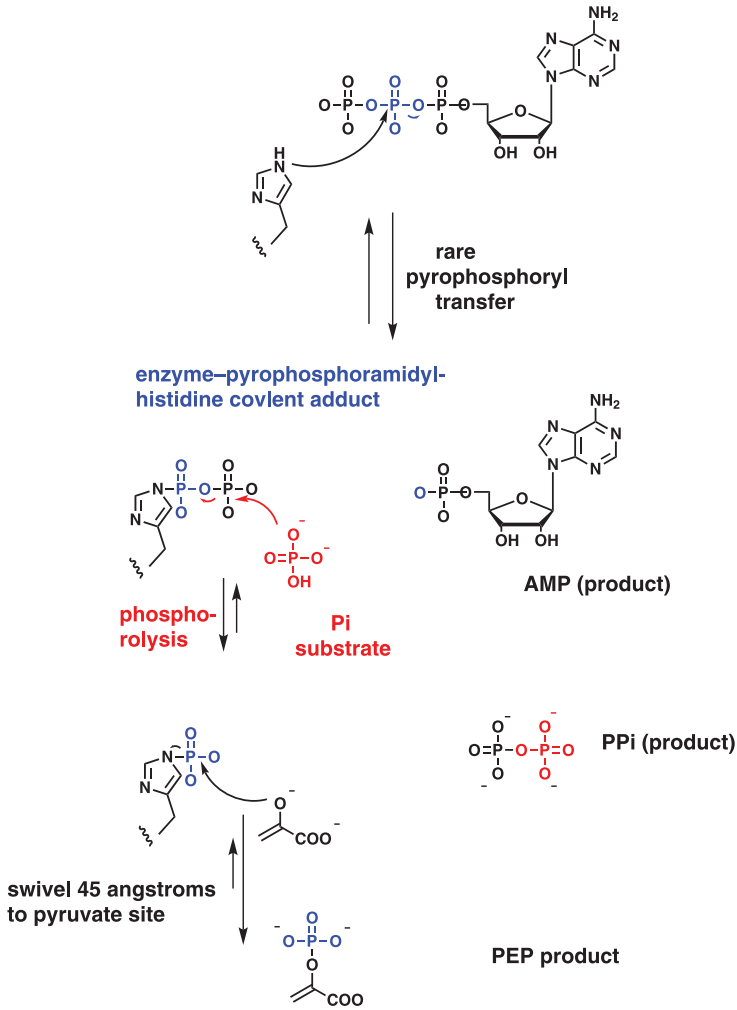


Figure 14.12 The catalytic cycle of pyruvate phosphate dikinase starts with an unusual attack by a histidine residue of the enzyme on Pβ of ATP to give a pyrophosphoryl histidine intermediate. Also surprising is the subsequent phosphorolysis by Pi to release PPi and leave the *N*-phosphoryl-His–enzyme intermediate. That is the $-PO_3^{2-}$ donor to the enolate anion of bound pyruvate.

(Figure 14.12). The second substrate is inorganic phosphate and it attacks the pyrophosphoryl-*N*–His-enzyme on the terminal phosphate (what had been Pγ of ATP). This is a *phosphorolysis step*, inorganic phosphate acting as an oxygen nucleophile as described in Chapter 9. The result is a molecule of PPi to go along with the AMP. ATP has undergone a net cleavage to AMP and PPi, but by a particularly circuitous tandem pyrophosphoryl transfer and a phosphorolysis.

These two steps have left the original Pβ phosphoryl group of ATP as the covalent phosphohistidinyl enzyme intermediate. We know from the PTS system above that phosphoryl transfer between histidine and enolpyruvate is reversible. It remains for the enzyme–N–PO_3^{2-} form of the dikinase to be able to bind and enolize pyruvate to create the enolate anion as attacking oxygen nucleophile and yield PEP as the third product.

Structural studies on pyruvate phosphate dikinase show separate binding domains for ATP and PEP and a core domain for phosphoryl transfer from these two cellular activated phosphoryl transfer metabolites.[26] The catalytic histidine is in a middle domain that undergoes dramatic swiveling movement to cycle from the ATP domain to the PEP domain (Figure 14.13). The PTS enzyme I above is also homologous in its P-His swiveling movement to go from a PEP domain to the HPr recognition site as it transfers the attached –PO_3^{2-} group across long distances.

We sum up the chemical biology lessons from pyruvate phosphate dikinase before moving to eukaryotic histidine phosphoproteomics. (1) The enzyme is at a mechanistic and structural intersection of ATP and PEP as cellular phosphoryl donors. (2) Pyrophosphoryl transfers are in play: in the forward direction as pyrophosphoryl transfer to the imidazole nitrogen of the active site histidine. (3) In the reverse direction PPi acts as a phosphoryl donor to the enzyme–PO_3^{2-} species. This may be a strategic remnant of PPi as a common phosphoryl donor in early cellular evolution. (4) Finally, while the –PO_3^{2-} group acts as conventional electrophile in transfer to the enolate anion of bound pyruvate, free inorganic phosphate shows orthogonal reactivity as an oxygen nucleophile in the *phosphorolysis* step. The pyruvate phosphate dikinase logic embodies several of the many reactivities of phosphorus in biology.

14.2.4 Eukaryotic Phosphohistidinyl Protein Residues

Relatively few eukaryotic proteins have been documented to undergo phosphorylation on histidine residues. Those that do so use the N–PO_3^{2-} group as a phosphoryl transfer agent rather than accumulating as a stable group for some structural or regulatory purpose. Three metabolic enzymes known to proceed *via* intermediate phosphoryl-histidine enzymes include succinyl CoA synthetase in the Krebs cycle,[28] ATP citrate lyase as the first committed step in cytoplasmic fatty acid biosynthesis,[29] and nucleoside diphosphokinase.[30]

The succinyl CoA synthetase, running in the physiological direction, succinyl CoA to succinate, converts GDP and Pi into GTP, a net GTP synthetase action. The substrate inorganic phosphate acts as nucleophilic anion (phosphorylase action) on the thioester carbonyl of succinyl CoA to make bound succinyl-phosphate (the acyl phosphate). This mixed acyl–phosphoric anhydride is then captured by N3 of the imidazole side chain of the histidine residue to give *N*-phospho-His–enzyme. It is this phosphoramidate group

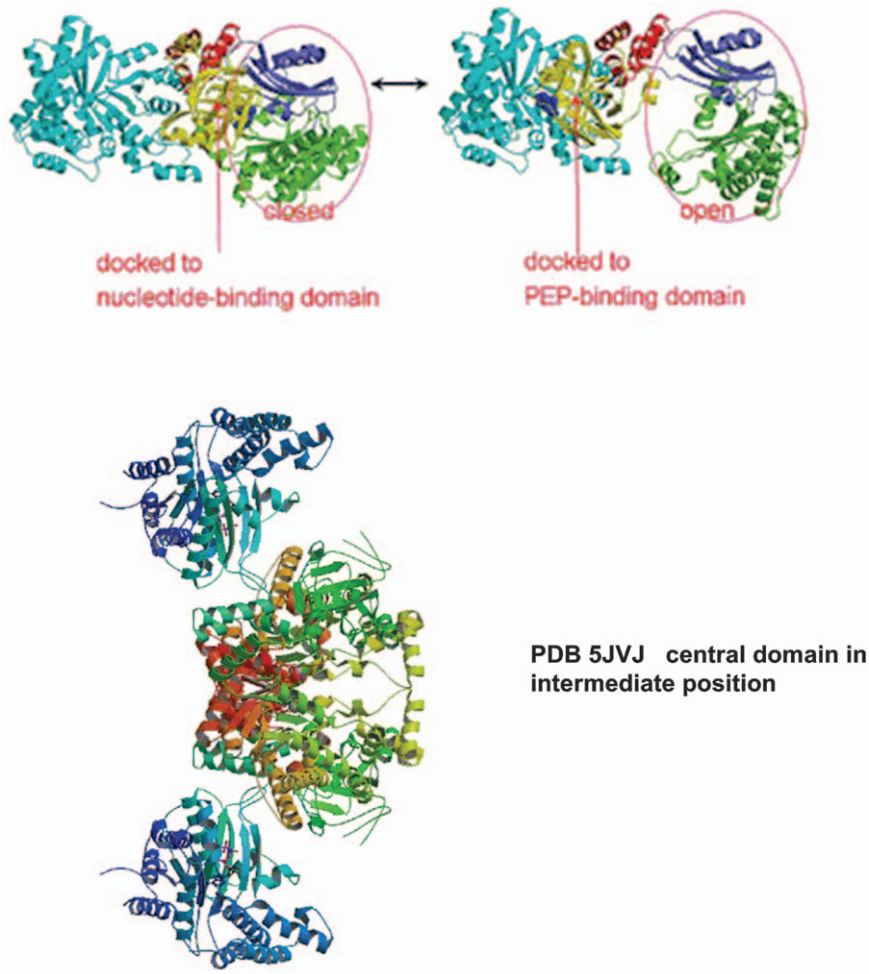
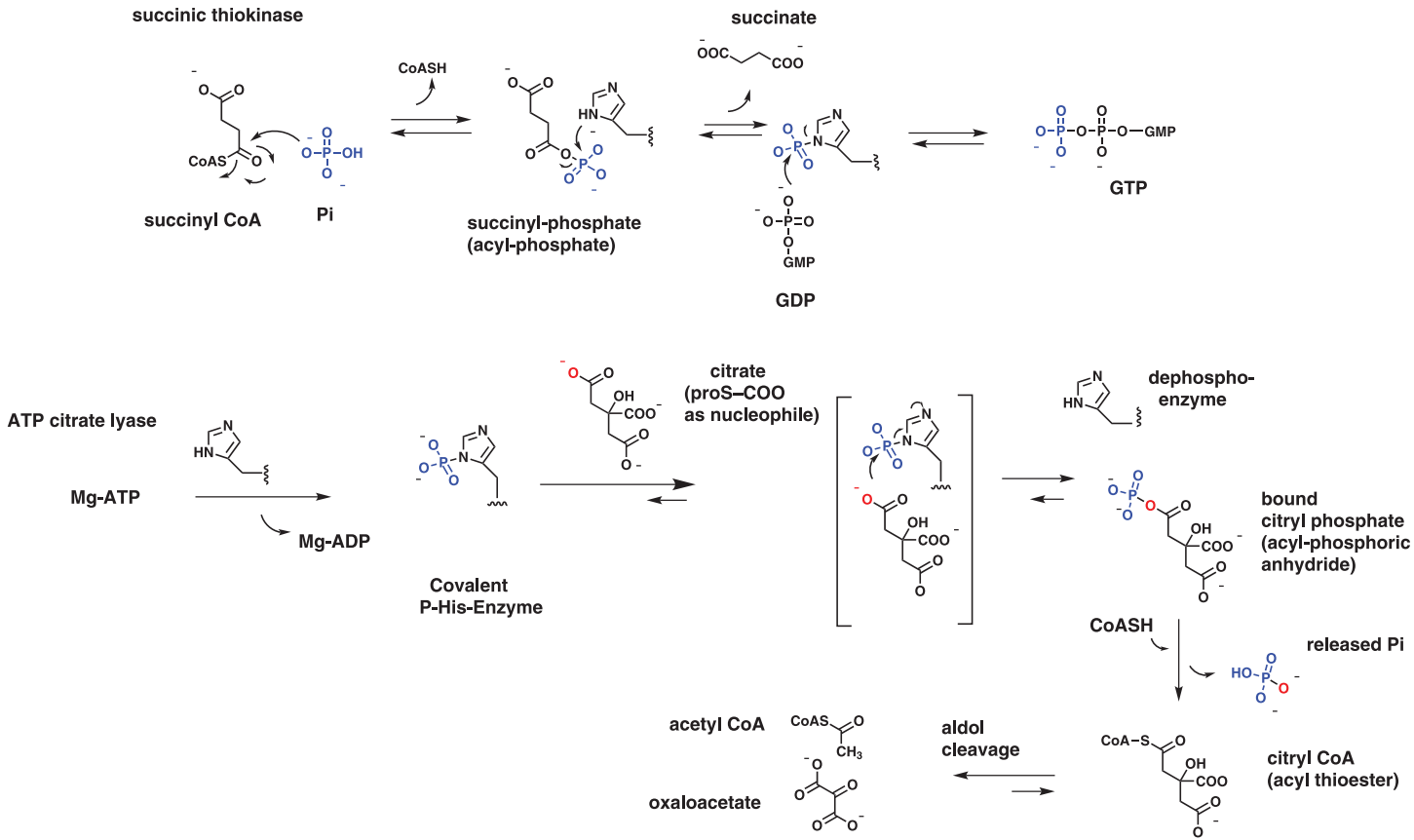


Figure 14.13 Pyruvate phosphate dikinase has two distinct domains, one for ATP, one for PEP binding and catalysis with a connecting middle domain that acts as a swivel. X-ray analyses find the central domain docked either to the nucleotide-binding (ATP) domain or docked to the PEP domain or positioned equidistant, giving a mechanical view of how the phosphoryl groups move from substrate to enzyme to cosubstrate. (Top)
Reproduced from ref. 27, image courtesy of Osnat Herzberg. Bottom panel from RSCB PDB structure 5JVN.

phosphorus atom that gets attacked by one of the oxygen anions of substrate GDP to form GTP (Figure 14.14).

This second half reaction of succinyl CoA synthetase (it actually runs physiologically in the reverse direction) is the basis for the catalytic logic of nucleoside diphosphokinases (Chapter 3). These enzymes feature prominently in cellular expenditure of ATP to bring the other canonical nucleoside


succinic thiokinase
succinate
CoASH
succinyl CoA
Pi
succinyl-phosphate
(acyl-phosphate)
GDP
GTP
ATP citrate lyase
Mg-ATP
Mg-ADP
Covalent
P-His-Enzyme
citrate
(proS–COO
as nucleophile)
dephospho-
enzyme
bound
citryl phosphate
(acyl-phosphoric
anhydride)
CoASH
released Pi
acetyl CoA
oxaloacetate
aldol
cleavage
citryl CoA
(acyl thioester)

diphosphates (GDP, CDP, UDP) up to the triphosphate level for RNA polymerase to act. In a given catalytic cycle the active site imidazole side chain of histidine attacks Pγ of ATP, releasing ADP and generating the covalent His–N–PO_3^{2-} intermediate. Then ADP is replaced by any of UDP, CDP, or GDP for –PO_3^{2-} transfer to the bound NDP-β-oxyanion as a nucleophile.

The catalytic logic of ATP citrate lyase is to fragment an ATP molecule while cleaving six-carbon citrate to get to the four-carbon oxaloacetate and an activated form of acetate that can be fed into the fatty acid synthase machinery. That two-carbon fragment is acetyl-*S*-CoA.

The reaction path involves initial cleavage of ATP with formation of the His_{760}–N–PO_3^{2-} covalent intermediate. Attack by a citrate carboxylate yields the acyl phosphate, citryl-phosphate. Thiolysis by CoASH gives citryl-CoA and retro Claisen fragmentation gives oxaloacetate and the desired acetyl CoA.

The findings of Hardman *et al.*[2] of more than 500 phosphohistidinyl peptide fragments from trypsin cleavage of the crude proteins from HELA cells indicates a vast landscape of as yet unidentified phosphohistidine proteomics (Figure 13.7). Many such N–PO_3^{2-} species may be transient, but if even only 10% were stable under cellular conditions, that would be ~50 as yet uncharacterized phosphoproteins. The possibility or probability that initial His–N–PO_3^{2-} adducts could transfer –PO_3^{2-} groups to other side chains of residues is also likely. One documented case occurs from nucleoside diphosphokinase to aldolase (as an acyl phosphate on the side chain of Asp319[31]) but that may be just the start of understanding phosphoprotein dynamic transfers in eukaryotes. There is one characterized bacterial phosphohistidine protein phosphatase,[32,33] compatible with the idea that as yet unknown protein substrates require hydrolytic cleavage on otherwise stable (or perhaps stranded) N–PO_3^{2-} bonds.

14.3 Arginine Phosphoramidates in Proteins

In Chapter 10 we noted that the free amino acid *N*-phosphoarginine is widely used in invertebrates as a reservoir of high phosphoryl group transfer potential.[34] Arginine kinase catalyzes the reversible reaction of ADP and *N*-phosphoarginine in a reversible equilibrium with ATP and L-arginine (Figure 14.15). This reversibility is the classic function of a phosphagen. Given this precedent it would not be surprising that organisms have also evolved ATP dependent phosphoryl transferases that also transfer –PO_3^{2-} groups to the side chains of arginine residues in proteins. Indeed, phosphorylation of a specific arginine side chain in eukaryotic histone 4 has been

Figure 14.14 Two examples of phosphoramidate intermediates (*N*-phosphoryl-His residues) in eukaryotic central metabolism: succinic thiokinase in the citrate cycle and ATP citrate lyase the first enzyme in cytoplasmic fatty acid biosynthesis.

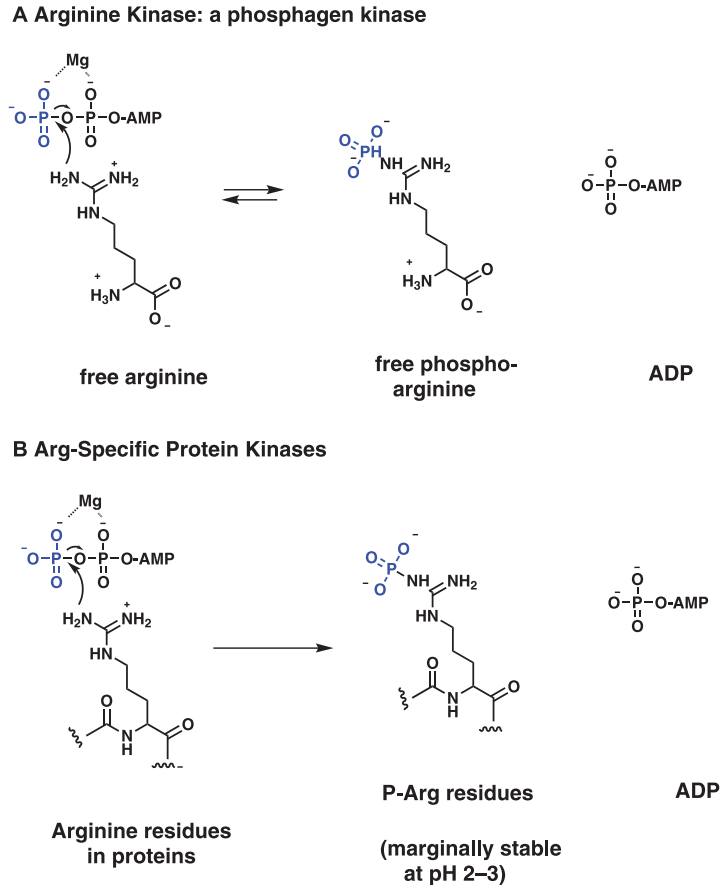


Figure 14.15 Analogy between kinases that phosphorylate free arginine (A) and those that phosphorylate protein-bound arginine guanidino side chains (B).

known for decades as an existence proof[35] but was not of obvious functional consequence and had garnered almost no follow up or generalization.

The knowledge base changed about a decade ago from investigations of heat shock responses and arginine phosphoproteomics in two types of Gram-positive bacteria – bacilli and streptococci.[36] It had been known that a repressor protein class three stress repressor (CtsR) controlled the expression of both protein chaperones and the chambered caseinolytic protease (ClpP) in the thermophilic bacterium *Bacillus stearothermophilus.* Expression of the *mcsb* gene (encoding a protein arginine kinase) relieved repression, with the anticipated enhanced expression of both protein chaperones and the ClpP protease to help with hydrolytic chopping up of proteins having difficulty and lowered efficiency in reaching their native folded states.[36] Both sets of proteins would aid in dealing with temperature-sensitive proteins that unfold at the

heat shock temperatures and get marked for proteolytic degradation by being fed into the ClpP protease chambers.

The encoded enzyme McsB turned out to be an arginine protein kinase, *N*-phosphorylating up to three arginines in the DNA binding domain of the repressor CtsR[36] (Figure 14.16A). The phosphoarginine forms of CtsR, with negative charges replacing the former cationic guanidinium groups, no longer bound the promoter regions of target genes, relieving repression and enabling the heat shock response. Subsequent X-ray analysis of both CtsR and McsB revealed high affinity binding sites for phosphoarginine residue docking[37](Figure 14.16B).

X-ray structural analysis of the McsB arginine binding site revealed two anionic side chains that position the guanidinium side chain of arginine residues to set up the chemospecific Arg *N*-phosphorylations. As sketched in Figure 14.16A this is distinct from the binding site in canonical serine-specific protein kinases and explains the specificity. The catalytic domain of

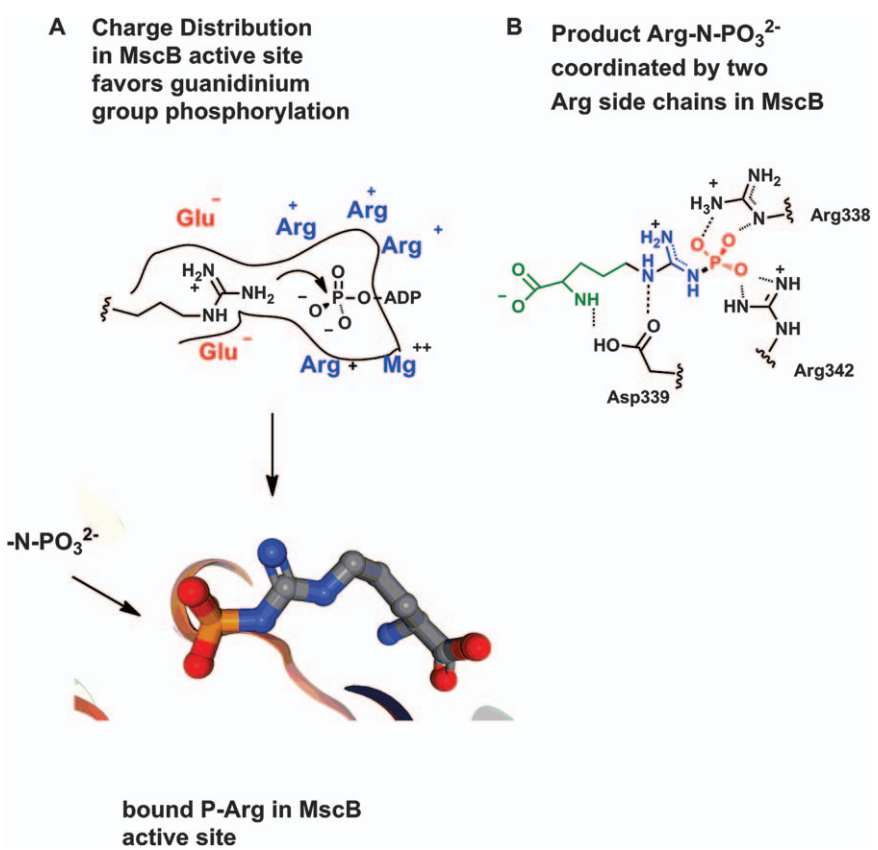


Figure 14.16 Protein arginine kinase McsB.[37] (A) An electrostatically optimized binding site for recognition of guanidine side chains in arginine residues: (B) Phospho-arginine residues bound in McsB.

McsB is related to the catalytic architecture of the free arginine phosphagen kinases.

Most intriguingly, it turned out that McsB could phosphorylate many other bacterial cellular proteins under the heat shock conditions.[38,39] Those *N*-phosphorylated proteins were selectively fed into the chambers of ClpP for proteolytic digestion (Figure 14.17). Presumably, these were unfolded or otherwise damaged proteins that were being marked for proteolytic removal. These observations have been generalized to a proposed role for arginine residue *N*-phosphorylations as signals for accelerated protein degradation in bacilli and in staphylococci, that parallel the eukaryotic ubiquitin-tagging system for accelerated degradations (Section 14.5). Consistent with an editing role for damaged proteins, there is a phosphoarginine protein phosphatase YwiE (Figure 14.18) that can specifically remove the N–PO_3^{2-} groups from phosphorylated proteins before they are proteolyzed. This could be a quality check function, akin to the role of deubiquitinases in eukaryotic protein quality control processing. There is an analogy between the canonical Ser, Thr, and Tyr kinase-phosphatase cycles and the corresponding P-Arg kinase–phosphatase balance. The disposal of trash proteins by intracellular bacterial proteases has been dubbed the ''phospho kiss of death''.[40]

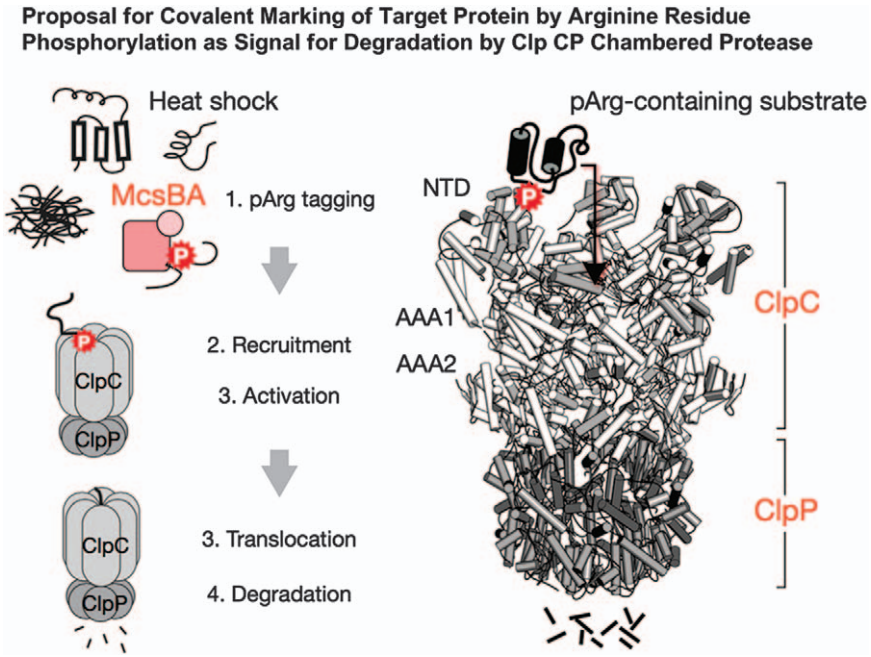


Figure 14.17 Phosphorylation of Arg residues by bacterial arginine protein kinases as signals for accelerated destruction by the chambered protease ClpP.[37]
Reproduced from ref. 39 with permission from Springer Nature, Copyright 2016.

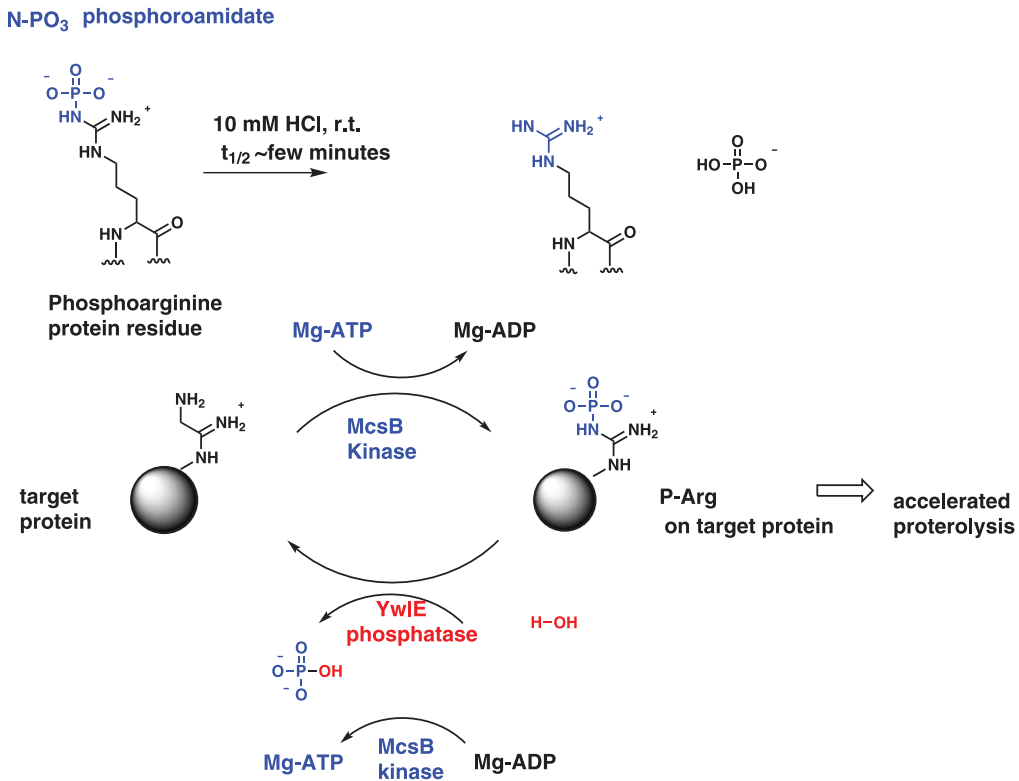


Figure 14.18 Balance between the protein arginine kinases and protein P-Arg phoshatases.[41]

This logic would be akin to the covalent posttranslational marking of eukaryotic proteins by polyubiquitin chains for delivery to the chambered proteasomes for accelerated degradation. There are also eukaryotic examples of degrons – short stretches of proteins that become targets for the ubiquitinylation machinery that become tagged by serine or threonine phosphorylation.[42,43]

Future investigations will test the generality of the proposal that post-translational N–PO_3^{2-} marks on arginine residues in proteins are bacterial signals for protein transport and breakdown by the ClpP class of proteases. If validated, another broad chapter in phosphoproteomics will have opened up.

Turning back to the recent findings of ~500 *N*-phosphoarginyl peptides derived from HELA cell proteins[2] (Figure 13.7), it is clear that extensive arginyl phosphoproteomics represent a significant fraction of the yet to be uncovered eukaryotic phosphoproteomics 2.0 universe.

An independent approach using an immobilized phosphoarginine-specific antibody has recently led to initial identification of 830 P-Arg sites in 500 Jurkat cell proteins: further analysis confirmed 152 P-Arg sites in 118 identifiable proteins from a human Jurkat cell line.[44] Of that 118, 37% were nucleic acid binding proteins and another 16% were transcription factors, indicating an enrichment of protein arginine phosphorylation in nucleic acid-based information transfer processes. One specific functional example was the observation that phosphorylation of the prohormone proglucagon at Arg_{91} blocked the subsequent proteolytic processing by the RR-selective proteases (that remove the pro sequence) in the secretory pathway. Thus, Arg91 *N*-phosphorylation thereby blocked formation of the active glucagon peptide hormone forms.

14.4 *N*-phospho-lysine Protein Residues

The nature and function of *N*-phosphorylated lysine chains in eukaryotic phosphoproteomics remains largely dark matter at this juncture. There are no specific examples to highlight. This is the case despite the heavy involvement of lysine side chains in *N*-methylations, *N*-acetylations and even longer chain *N*-acylations in eukaryotic biology. The many sets of acetylations and mono- to tri-*N*-methylations of lysine residues in histone tails of nucleosomes are the most celebrated of those posttranslational protein modifications.

Two points to note. The results of recent mass spectrometry analyses of Hardman *et al.*[2] indicate that there are up to 800 mammalian *N*-phsopho-Lys peptides from HELA cell proteins (Figure 13.7). Even if only a fraction of those initial detections are borne out for identification of specific mammalian proteins, the lysine-phosphoramidate proteome may prove to be extensive. Finally, as a brief harbinger of posttranslational protein nucleotidylation in Chapter 16, contemporary terminology for adenylylation is protein AMPylation, we have noted in Chapter 5 that DNA ligases use

covalent *N*-Lysyl-AMP active site adducts in every catalytic cycle of DNA nick, gap or blunt end repair. Where there are N–AMP adducts to lysine, one expects N–PO_3^{2-} adducts to follow.

14.5 *S*-phosphocysteinyl Proteomes

The thiol side chain of cysteinyl residues in proteins is the most reactive of all twenty amino acid building blocks, both in two-electron and one-electron reactions. Figures 14.19–14.21 show eight types of covalent adducts between cysteinyl side chain thiolate anions in proteins.[45] They do not form all at once or in those specific sequences but are notional indications of the types of chemistry in which the nucleophilic thiolate anion of various protein substrates has engaged. Acylation and alkylation posttranslational modifications are exemplified by *S*-palmityl thioesters and *S*-geranylgeranyl thioether linkages, respectively. Oxidative conversion of thiols to the monoxygenated sulfenate, *S*-nitrosyl cysteine and formylglycine residues represent three distinct oxidative posttranslational modifications. The formation of disulfides dwarves all the others for proteins passing through the secretory pathways of cells. The persulfide forms of cysteine are likewise formal oxidations and ultimately the sources of inorganic sulfide in both 2Fe–2S and 4Fe–4S redox clusters in hundreds of proteins.

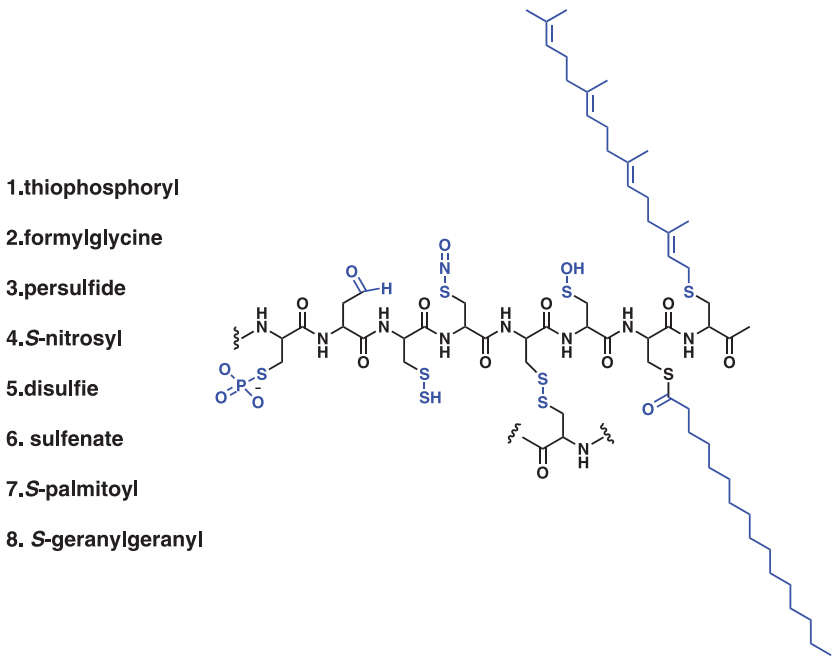


Figure 14.19 Eight different types of posttranslational modifications on cysteinyl residue side chains. One of these eight chemotypes is phosphorylation of cysteinyl thiolates acting as nucleophiles.

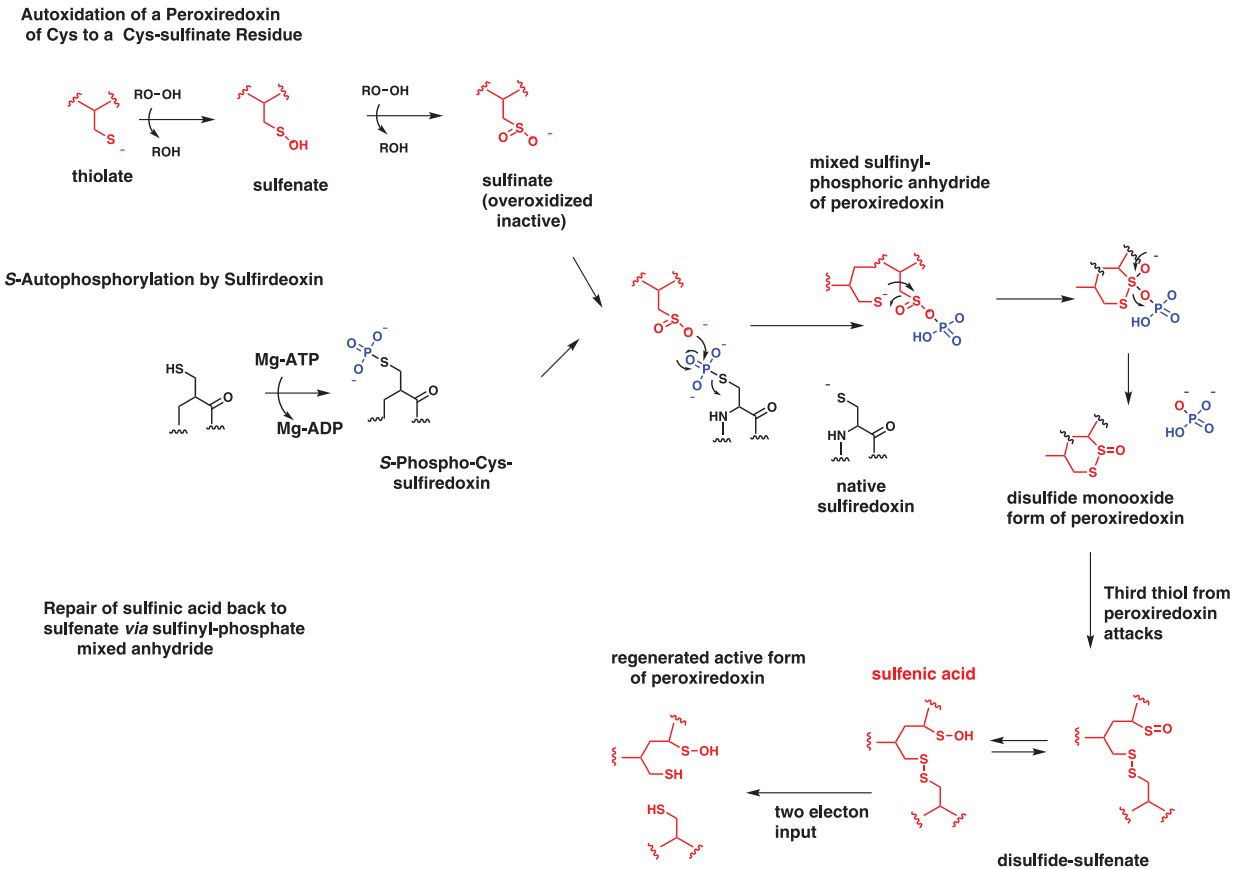


Figure 14.20 Autoxidation of a peroxiredoxin of Cys to a Cys-sulfinate residue, *S*-autophosphorylation by sulfirdeoxin and repair of sulfinic acid back to sulfenate *via* sulfinyl-phosphate mixed anhydride.

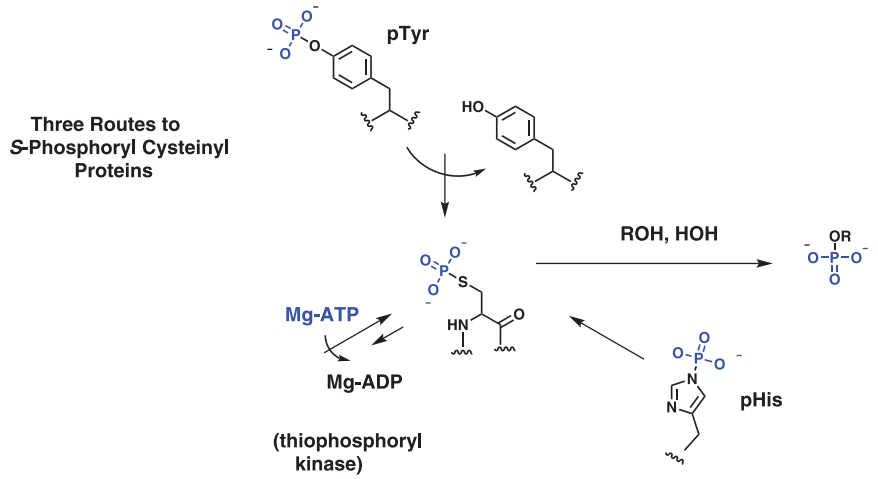


Figure 14.21 Three routes to *S*-phosphoryl cysteinyl proteins.

The eighth covalent modification depicted is *S*-phosphorylation of cysteinyl side chains in proteins. The S–PO_3^{2-} linkage is a phosphorothioate bond. We have discussed some of the low molecular weight phosphorothioate metabolite chemical biology in Chapter 12. We also noted the remarkable findings of *S*-phosphorothioate diester replacement of some of the internucleotide phosphodiester bonds of DNA in bacteria as restriction–modification machinery to detect foreign DNA in the cytoplasm.

Given the broad swath of nucleophilic chemistry of cysteine side chains in other aspects of cysteinyl proteomes, the phosphocysteinyl proteome is surprisingly modest and limited. We note two bacterial examples and then two higher eukaryotic phosphoproteomic examples to highlight the underlying logic for forming and utilizing *S*-phosphocysteine residues. As with the other members of the noncanonical phosphoproteomes described in this chapter, the few examples known for covalent phosphorothioate-protein adducts are for the most part transient rather than stable derivatives that accumulate with long lifetimes.

14.5.1 Bacterial *S*-phosphoryl Enzymes

The first bacterial evidence of posttranslational modification of protein cysteinyl sidechains came from investigation of virulence mechanisms in staphylococci that involve a network of general transcription factors staphylococcal accessory regulator A (SarA) and multiple gene regulator A (MgrA).[46] It had been known that specific cysteinyl residues were required for these two proteins to bind to target DNA promoter regions and turn on genes such as that for hemolysins.

Sun *et al.*[46] investigated the function of a conserved pair of homologs to eukaryotic serine–threonine kinases (bacterial Stk1) and its presumed

companion phosphoprotein phosphatase serine–threonine phosphatase (Stp1). As sketched in Figure 14.22 despite the Ser–Thr homology, it turned out that Stk1 was an *S*-phosphocysteine kinase for both SarA and MgrA, attenuating their binding to target DNA and lowering expression levels of the hemolysin virulence factor. The *S*-phosphoryl group could then be removed hydrolytically by Stp1, acting as a specific phosphorothioate hydrolase. Two conclusions emerge. One is that homology to canonical Ser–Thr protein kinases is not a foolproof predictive algorithm and may miss other protein phosphorothioate kinases. The other is that the S–PO_3^{2-} Cys linkage on the staphylococcal transcription factors was sufficiently long-lived to (need to) be removed functionally by the Stp1 phosphatase.

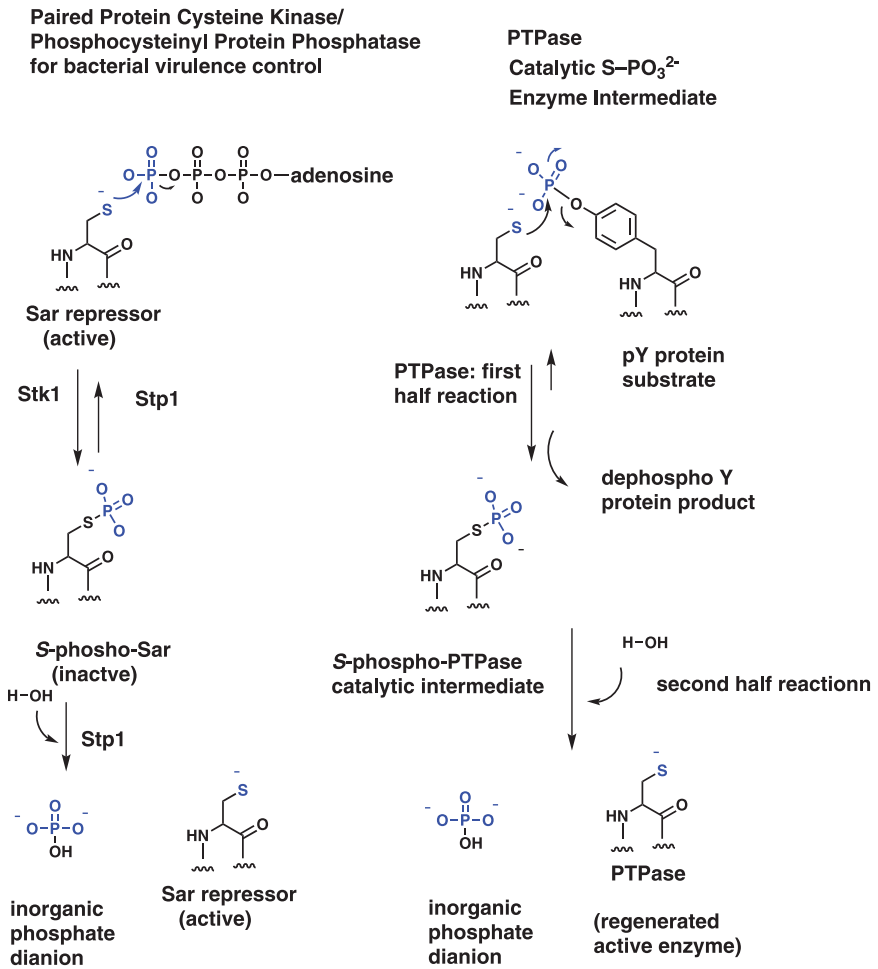


Figure 14.22 Paired protein cysteinyl *S*-kinase (Stk1) and thiophosphoryl protein phosphatase (Stp1) control activity of SarA and MgrA bacterial transcription factors.

The second bacterial venue for *S*-phosphocysteinyl proteins harks back to the hexose phosphate transferase system for bacterial uptake and trapping of up to twenty distinct hexose metabolites by coupled phosphorylation.[23] We noted above that there is a cascade of phosphoryl transfers from PEP, not ATP, to histidine residues on three proteins: enzyme I, HPr, and the soluble component enzyme IIA. Enzyme IIC is typically the transmembrane carrier of a hexose subfamily and enzyme IIB a peripheral enzyme that interacts with the hexoses to be phosphorylated and the P-His form of enzyme IIA (Figure 14.9). For the sugar mannose the enzyme IIB indeed becomes covalently phosphorylated on a histidine imidazole ring, as the fourth of four tandem phosphoryl transfers between histidines on distinct proteins, prior to the final PO_3^{2-} transfer to C6–OH of mannose.

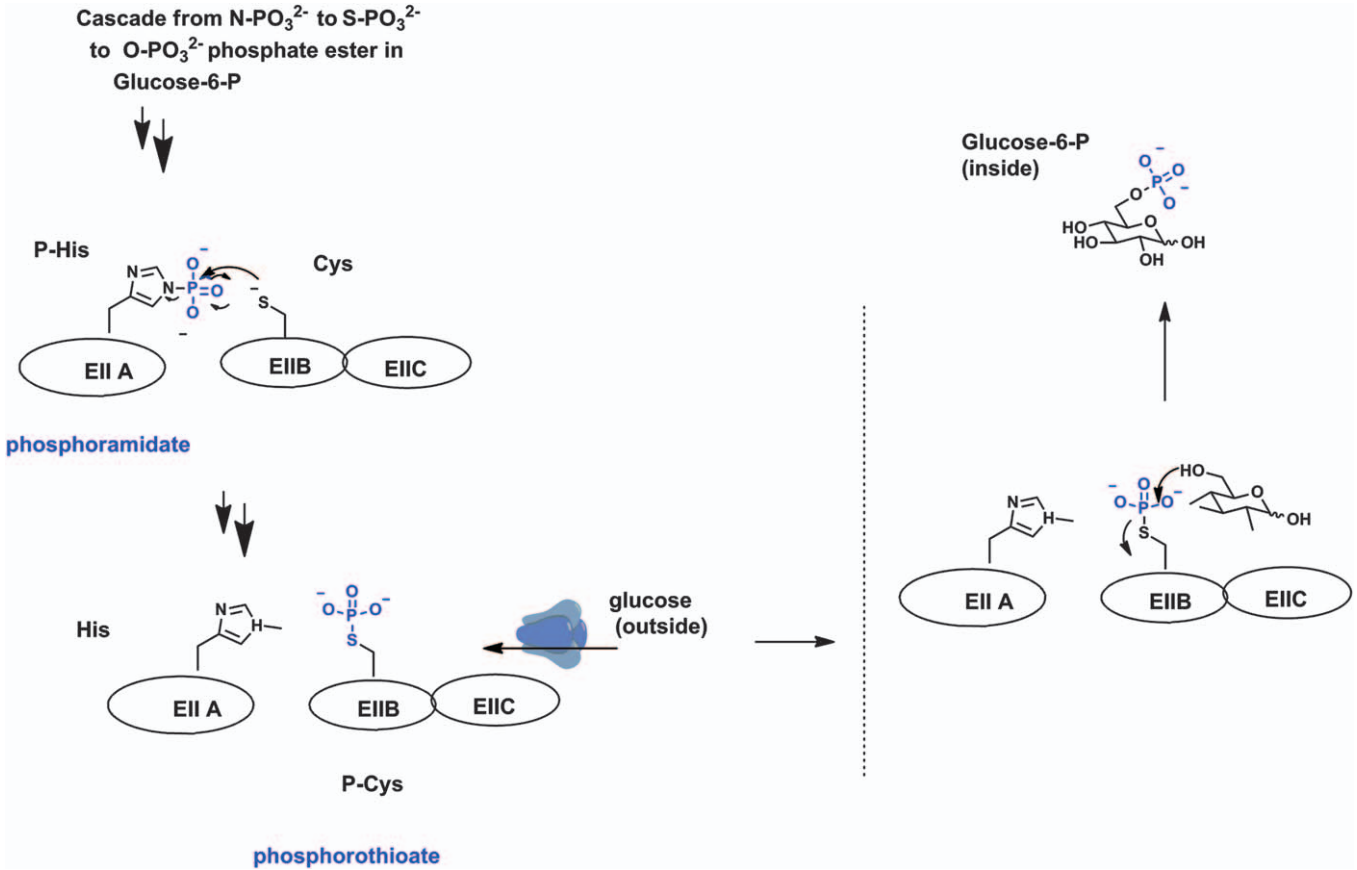
However, this is the exception. For the IIB proteins that bind to and phosphorylate mannitol, sorbitol, glucose, fructose, and lactose, the active site nucleophile in the IIB enzyme is a cysteine thiolate not an imidazole ring of histidine.[47] As depicted in Figure 14.23 for the bifunctional enzyme IIB specific for glucose, the pathway for the last two phosphoryl transfer steps are *phosphoramidate* in enzyme IIA, *phosphorothioate* in enzyme IIB, and finally to the *phosphate oxoester* grouping in glucose-6-phosphate. The cascade started three phosphoryl transfers back, with PEP as the PO_3^{2-} donor, so this tandem phosphoryl transfer PTS protein machinery illustrates four of the common phosphoryl groups in metabolism: *enol phosphate, phosphoramidate, phosphorothioate, and phosphate monoester*. The first three are thermodynamically activated for phosphoryl group transfer, the fourth is generally not. The $-PO_3^{2-}$ group stops there.

14.5.2 Eukaryotic *S*-phosphoenzymes

In eukaryotic phosphoproteomics we note two cases of phosphorothioate enzyme intermediates. Both involve transient *S*-phosphoryl cysteinyl–enzyme intermediates in quite different metabolic contexts. The first set are found in ~90% of phosphotyrosine protein phosphatases,[48,49] the second in the resuscitation of cellular peroxide detoxification by peroxiredoxins.

14.5.2.1 S-Phosphocysteinyl–Enzyme Intermediates in PTPase Action

We defer full discussion of architecture, context and function of protein tyrosine kinases to Chapter 15, except to note that the various subclasses of tyrosine protein kinases all autophosphorylate themselves in certain tyrosine-rich domains, once activated. Many of these kinases, especially the cytoplasmic ones, also phosphorylate tyrosines in other proteins, numbering in the hundreds in mammalian cell proteomes. The phosphotyrosyl bonds in proteins, in the chemical category of phosphomonoesters of phenols, are indefinitely stable chemically, both on the target proteins and on the kinase regions themselves.


Cascade from N-PO3 2- to S-PO3 2-
to O-PO3 2- phosphate ester in
Glucose-6-P
P-His
Cys
EII A
EIIB
EIIC
phosphoramidate
His
glucose
(outside)
EII A
EIIB
EIIC
P-Cys
phosphorothioate
Glucose-6-P
(inside)
EII A
EIIB
EIIC

To counteract the P–Tyr group chemical stability and provide reversibility to the signaling mechanisms entrained by those P–Tyr groups, cells have evolved P–Tyr specific phosphoprotein phosphatases. We noted mechanisms of phosphatases acting on low molecular weight phosphomonester metabolites in Chapter 6, noting that more than 100 metabolites are enzymatically dephosphorylated during metabolic pathway functioning. Some are quite nonspecific, *e.g.* alkaline phosphatase, while others show restricted hydrolytic activity.

The great majority of phosphoprotein tyrosine phosphatases (PTPases) operate *via* transfer of the itinerant $-PO_3^{2-}$ group from a protein tyrosine side chain first to an enzyme nucleophile to form a covalent phosphoryl enzyme intermediate while the dephosphorylated protein is released (Figure 14.24). In PTPases examined for mechanism and structure the enzyme nucleophile is an active site cysteine thiolate.

The first half reaction then is transfer of the $-PO_3^{2-}$ group from the phenolic oxygen of the protein tyrosine substrate to form a phosphorothioate enzyme. In the second step the phosphoenzyme activates a water molecule to act as nucleophile to release inorganic phosphate and the starting form of the PTPase, ready for the next catalytic cycle. The presumption is that the $-Cys-S-PO_3^{2-}$ covalent enzyme intermediate is uphill energetically from the starting P-Tyr protein substrate while the hydrolytic second half reaction is downhill.

This is a classic case, along with some of the phosphohistidinyl enzymes noted earlier in this chapter, of *transient phosphoprotoeomics* due to the thermodynamic activation of the chemical form of the enzyme–$X-PO_3^{2-}$. There are some 95 predicted human PTPases with active site cysteines as catalytic nucleophiles in mammalian proteomes. The large number of PTPases indicate distinctions in subcellular locations, recognition of P-Tyr sequence contexts (*e.g.* runs of P-Tyr residues or mixed P–Ser and P-Tyr sequences), and selective abrogation of some cellular signals and not others. A small number of PTPases have tumor suppressor activities,[50] detected by mutations to reduce or abolish the PTPase activity, allowing inappropriate buildup of the set of P-Tyr proteins that would normally serve as dynamic dephosphorylation substrates.

The chemical surprise may not be the use of covalent cysteine thiolate side chain catalysis in the PTPase superfamily. After all, one finds superfamilies of proteases with cysteines as catalytic nucleophiles, and also thymidylate synthase and DNA methylases where cysteine–substrate alkylation adducts

Figure 14.23 Revisiting the phosphoryl transfer cascade in bacterial hexose phosphotransferase intracellular capture of sugars. The canonical five $-PO_3^{2-}$ transfers of Figure 14.8 are true for mannitol but not for more common hexoses such as glucose. As shown, the first transfer from PEP is to a histidine residue in enzyme 1. Then two more phosphoryl transfers involve histidine residues. The final enzyme adduct on enzyme EIIB is a thiophosphoryl cysteinyl adduct rather than the *N*-phosphoryl-His adduct. In turn the final phosphoryl transfer step is from the thiophosphoryl-EIIB ''downhill'' to the C6–OH of glucose.

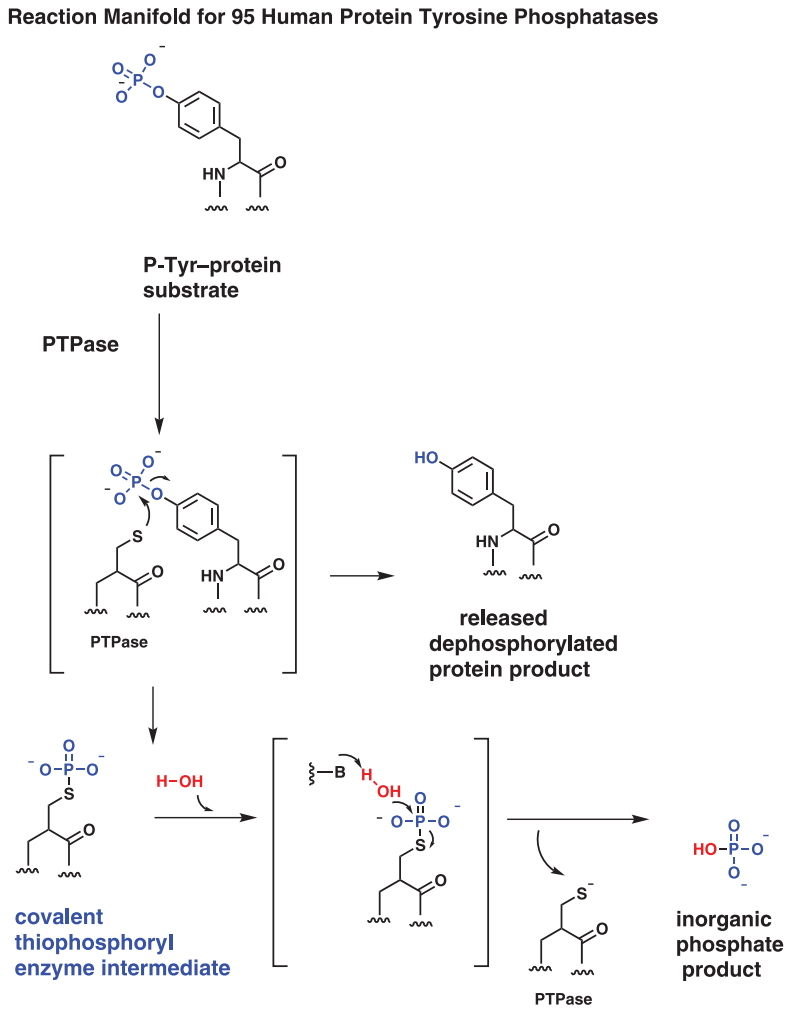


Figure 14.24 The major class of protein tyrosine phosphatases carry out two-step mechanisms for transfer of the $-PO_3^{2-}$ to water in hydrolyses reactions. The first step is formation of a phosphorothioate covalent enzyme adduct from nucleophilic attack of the enzymes' active site cysteine thiolate anion on the phosphotyrosine protein substrates.

are a central part of the catalytic logic.[45] The surprise may be that phosphoprotein phosphatases for P-Ser and P-Thr residue hydrolyses do not follow the covalent phosphoenzyme logic.

14.5.2.2 S-phosphocysteinyl Involvement in Sulfiredoxin and Peroxiredoxin Resuscitation

The second mammalian example of *S*-phosphocysteinyl–enzyme adducts as key reaction intermediates occurs during the action of the protein

sulfiredoxin to resuscitate over-oxidized molecules of the protein forms of peroxiredoxins. The peroxiredoxins, as the name indicates, play surveillance and defense roles in lowering the ambient concentration of both HOOH (hydrogen peroxide) and other ROOH, alkyl peroxides, to protect against oxidative damage[51] (Figure 14.25).

A typical peroxiredoxin catalytic cycle would involve attack of an active site cysteine thiolate on the weak O–O bond of HOOH, generating a molecule of water (two electron-reduced, safe) and the two electron oxidized sulfenic acid (–S–OH) form of the enzyme active site cysteine. Peroxiredoxins have a second active site cysteine thiol, disposed architecturally to attack the now electrophilic sulfur in the S–OH group. This releases the second molecule of water (Figures 14.26–14.28) and yields the peroxiredoxin disulfide. That in turn is reduced by an exogenous encounter with the dithiol form of thioredoxin (generated from NADPH by the FAD enzyme thioredoxin reductase). This can be a high flux detoxification route for peroxide removal within cells.

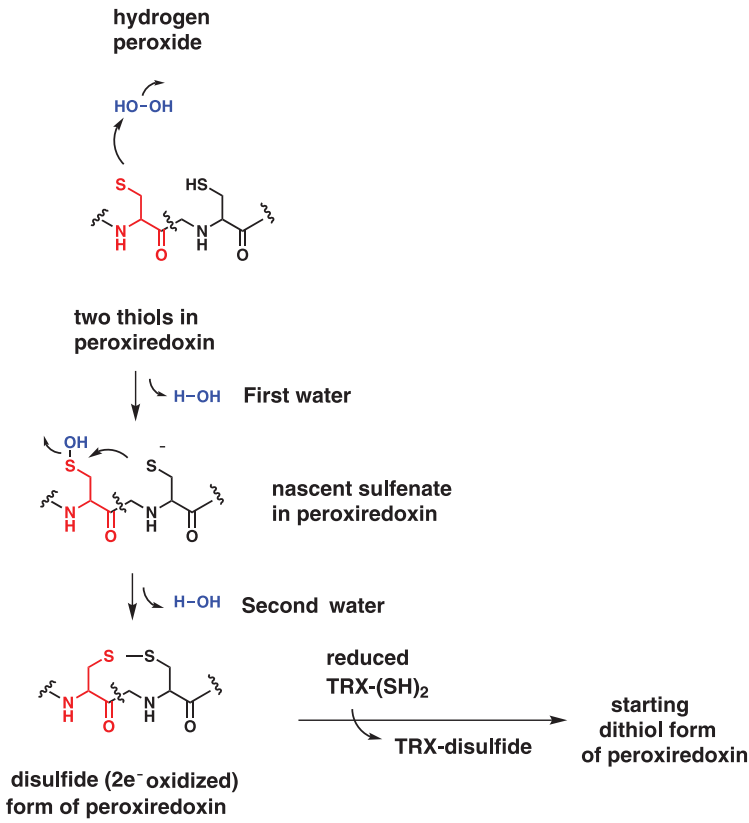


Figure 14.25 Reductive detoxification of hydrogen peroxide (to two water molecules) is catalyzed by peroxidredoxins by way of sulfenate intermediates. These are captured by a second thiolate in the peroxiredoxin, giving a disulfide form of the protein as product. The disulfide needs to be reduced back to the dithiol form of peroxiredoxins, typically by NADPH, thioredoxin, and the thioredoxin reductase system.

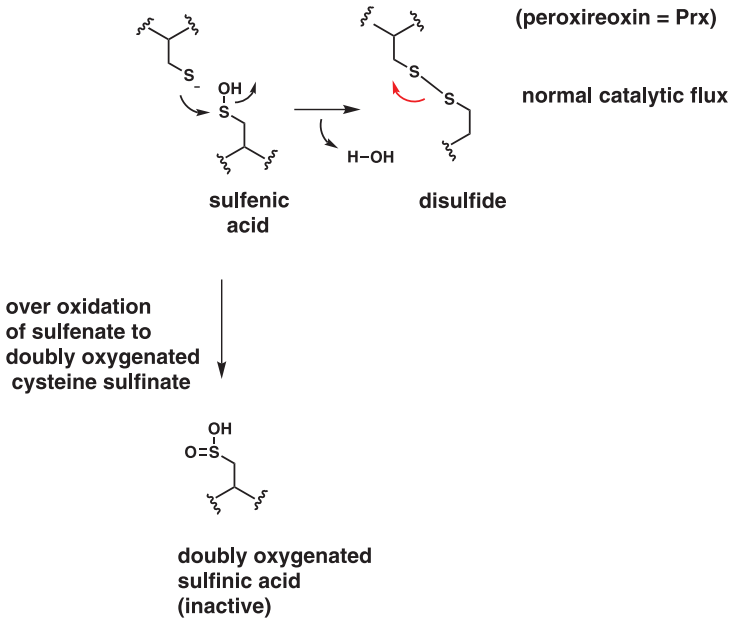


Figure 14.26 In some cycles of detoxification of peroxides by peroxiredoxins, the active site sulfenate (one oxygen) intermediate can be over-oxidized to the sulfinate (two oxygens) and become inactivated.

So far all is well, but in occasional catalytic cycles the sulfenate product form of a peroxiredoxin enzyme can be further oxidized by two electrons up to the dioxo sulfur form, termed a sulfinic acid. That is an *inactivated form* of peroxiredoxin (Figure 14.26). Sulfinic acid forms of cysteinyl side chains in proteins had been thought to be irreversible, inactive dead ends, but that was before the discovery of sulfiredoxin two decades ago. Sulfiredoxin can act to initiate reduction of the sulfinic group back to a sulfenate by some dramatically unusual enzyme chemistry.[52]

Figure 14.27 shows an overoxidized sulfinic acid form of a peroxiredoxin (in red). In its neighborhood a sulfiredoxin protein molecule (in blue) will use ATP to autophosphorylate itself on an active site cysteine: the only known mammalian case of autokinase *phosphorothiate* formation. The S–PO_3^{2-} adduct of sulfiredoxin can now be transferred to one of the sulfinate oxygens of the peroxiredoxin, acting as a nucleophile[52] (Figures 14.26–14.28). Sulfiredoxin has done its job, primed the repair *via* phosphorothioate chemistry –*S*-PO_3^{2-} formation and subsequent PO_3^{2-} transfer to the peroxiredoxin partner – and sulfiredoxin exits. The resultant peroxiredoxin adduct is an unusual variant of a mixed *sulfinic–phosphoric anhydride*, activated for transfer of ether the phosphoryl or sulfinyl moiety.

In the event, it is internal capture of the electrophilic sulfinyl sulfur by a second thiolate of the peroxiredoxin active site (Figure 14.27). The transition state decomposes in the forward direction to release inorganic phosphate,

bearing one of the two oxygens initially present in the sulfinic acid group. The second oxygen from the starting sulfinate is still present on the peroxiredoxin, now as a disulfide monoxide. Attack by a third cysteine side chain serves to fragment the S–S=O bond and give a disulfide and a sulfenate. Reduction of the disulfide by an external thioredoxin gets to the sulfenate form of peroxiredoxin. Its return to a thiol form also goes *via* the normal route of disulfide formation and reduction. Four-electron input, in the form of two pairs of thioredoxin dithiols, has returned a peroxiredoxin sulfinate to the starting thiol form.

Notable intermediates in this resuscitation logic include a t*hiophosphoryl* enzyme in sulfiredoxin, a *sulfinyl-phosphoric* anhydride adduct on peroxiredoxin, and a *disulfide monoxide* in the later stages of peroxiredoxin repair (Figure 14.29). The sulfiredoxin–peroxiredoxin intersection shows the intimate interface between protein thiolate oxidation and thiolate phosphorylations.

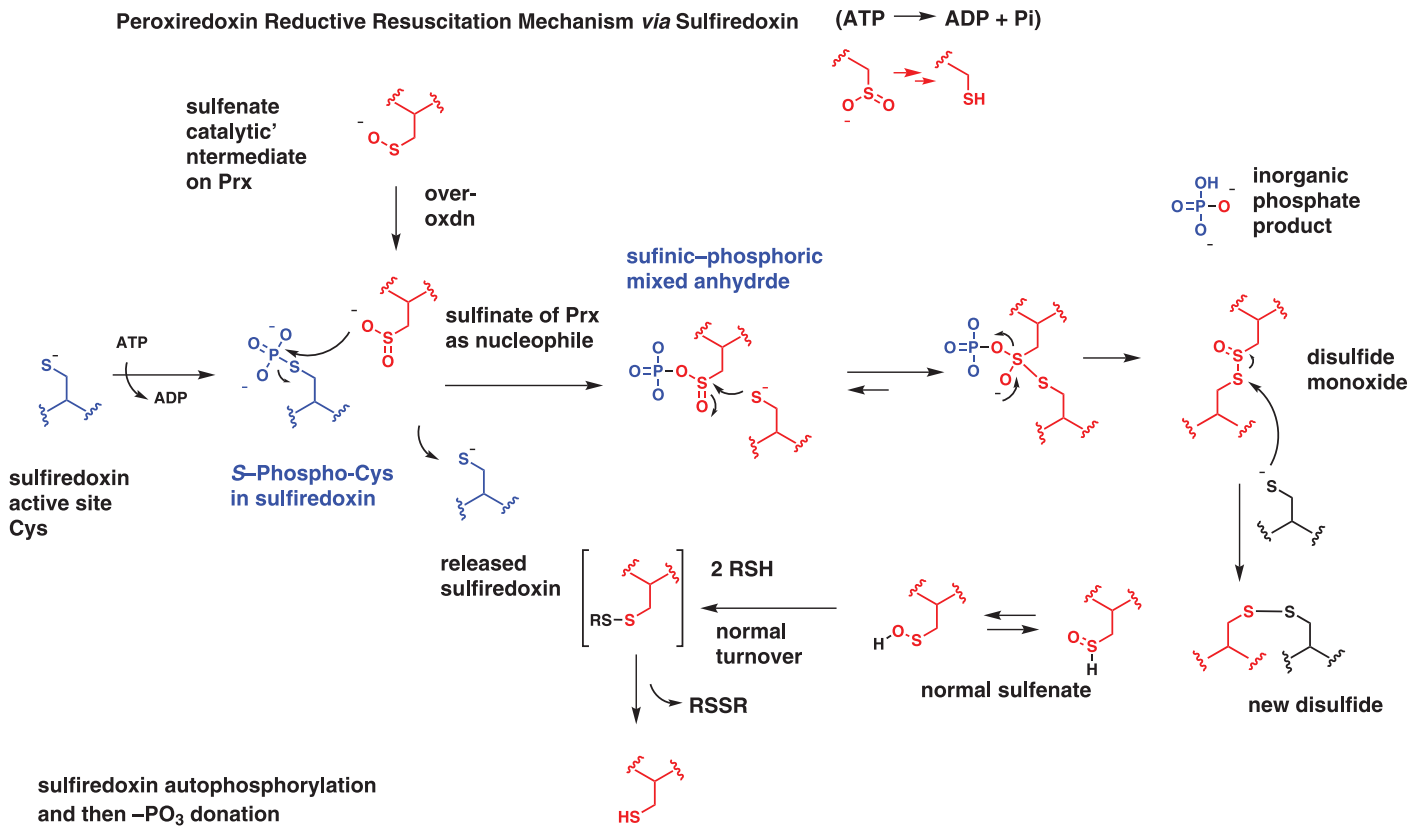
14.6 Acyl Phosphates in Phosphoproteomes

The acyl phosphates, mixed acyl phosphoric anhydrides, have been discussed in Chapters 6 and 9 in the context of activated phosphoryl transfer structures, capable of both acyl transfer and phosphoryl transfer to nucleophiles. In primary metabolic pathways both free aspartate and free glutamate are converted to acyl phosphates as metabolic intermediates in biosynthetic pathways to homoserine and glutamine, respectively as noted previously in Chapter 6. The β-aspartyl phosphate is captured in a relatively unusual reaction with a hydrogen species as nucleophile: a hydride ion from NADH. The γ-glutamyl phosphate is captured in the active site of glutamine synthetase by a cosubstrate ammonia molecule (Figure 14.30A).

The thermodynamic activation of both aspartyl and glutamyl phosphates makes them *a priori* unlikely stable phosphoprotein modifications. The free amino group in glutamyl phosphate, if deprotonated, provides a high local concentration of intramolecular nitrogen atoms to create the more stable five-member lactam, 5-oxoproline (Figure 14.30B). In proteins the amino groups of Asp and Glu residues are tied up in peptide bonds so that intramolecular decomposition route is abrogated. Thus, one might expect protein-based aspartyl and glutamyl phosphate side chains to have finite lifetimes if water can be excluded from hydrolytic reaction modes.

14.6.1 Phosphoaspartyl Response Regulators in Bacteria

Aspartyl phosphate-containing proteins as the response regulator–receiver domains of the two-component sensor kinase–response regulator systems in bacteria are the most prevalent sensing machinery in bacteria (noted in the context of the first step – the histidine autokinases – in Chapter 13). *E. coli* strains have over 50 such histidine sensor kinase–aspartate response regulators.[4] Myxococcal strains can have more than 150 two-component systems. They have relatively large genomes for bacteria.


Peroxiredoxin Reductive Resuscitation Mechanism *via* Sulfiredoxin
(ATP → ADP + Pi)
sulfenate catalytic' ntermediate on Prx
over-oxdn
inorganic phosphate product
sufinic–phosphoric mixed anhydrde
sulfinate of Prx as nucleophile
ATP
ADP
disulfide monoxide
sulfiredoxin active site Cys
S–Phospho-Cys in sulfiredoxin
released sulfiredoxin
2 RSH
normal turnover
normal sulfenate
new disulfide
RSSR
sulfiredoxin autophosphorylation and then –PO3 donation

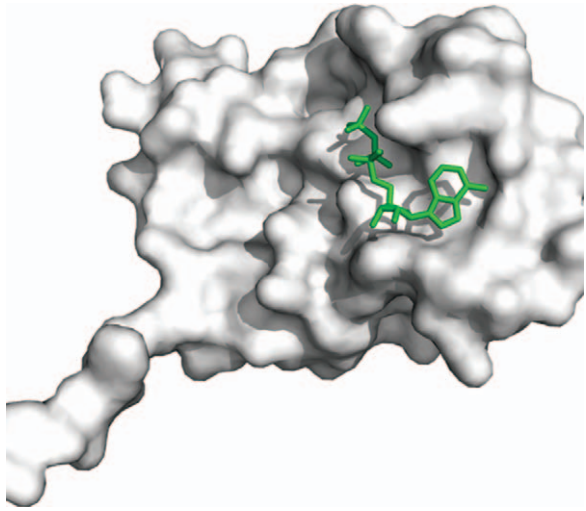


Figure 14.28 ATP bound to active site of sulfiredoxin, from pdb 3CYI.

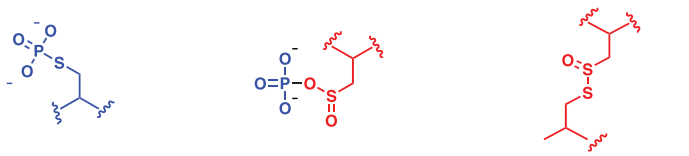


Figure 14.29 Three unusual phosphorus-containing adducts in the sulfiredoxin–peroxiredoxin system.

An average myxococcal genome size of 12 million base pairs is about 1/250 of the size of the human genome. If the number of sensor kinase–response regulator pairs scaled with genome size that would correspond to 3750 two-component systems in a three billion base pair genome, the size of the human genome. The human serine, threonine and tyrosine kinomes actually contains about ~550 members. This back of the envelope comparison shows that bacteria use more of their genome space for histidine kinases and for phosphoaspartyl regulator proteins than humans do for canonical protein kinases.

Figure 14.27 Resuscitation can be achieved by the enzyme sulfiredoxin that undergoes ATP-mediated thiophosphorylation on its active site cysteine thiolate anion. The sulfinate form of peroxiredoxin attacks the thiophosphoryl group, with $-PO_3^{2-}$ transfer to create the mixed sulfinic–phosphoric anhydride of the inactive peroxiredoxin. As shown, this can rearrange to a disulfide monoxide with release of Pi and ultimately lead to the sulfenate form of peroxiredoxin.

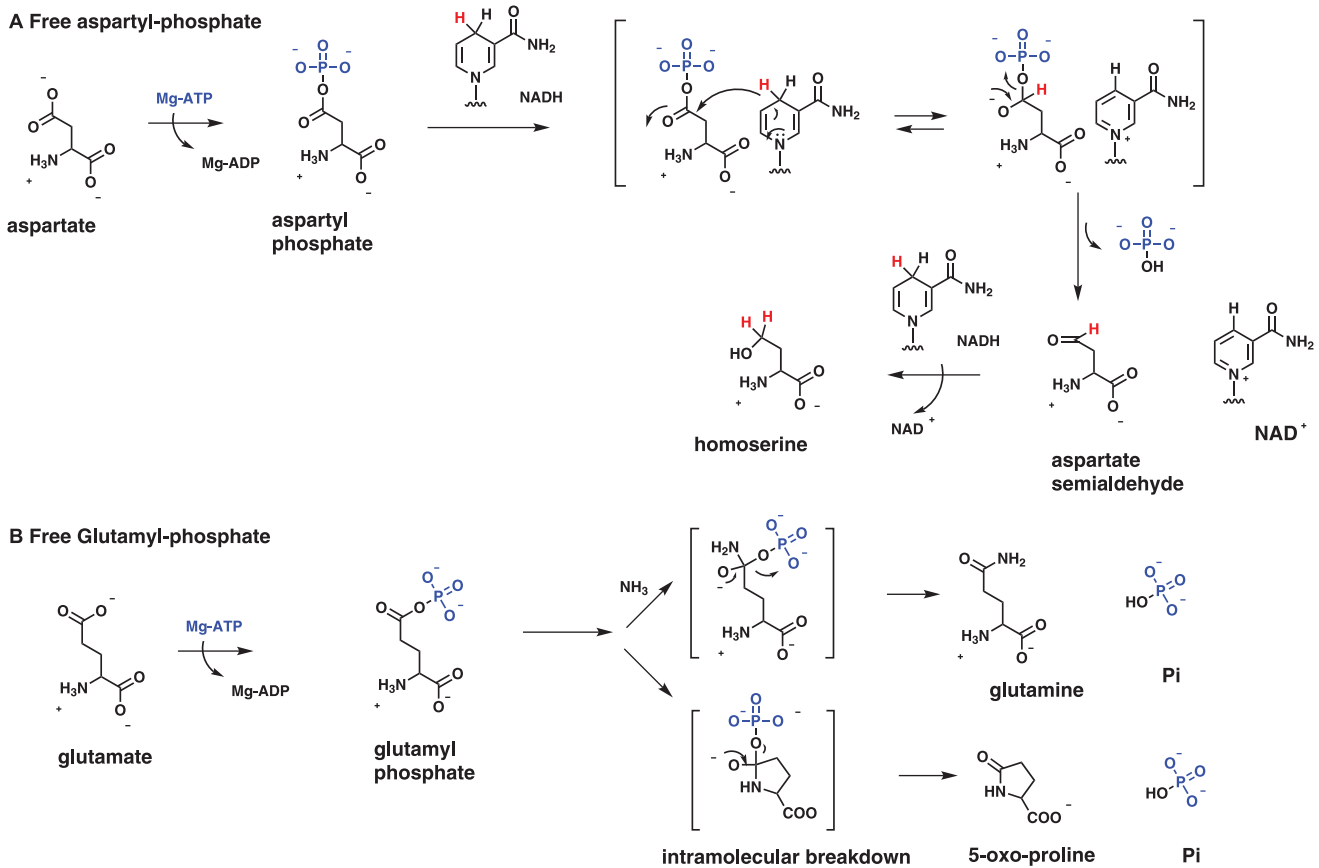


Figure 14.30 Two proteinogenic amino acids form acyl phosphates in kinase reactions in central amino acid biosynthetic metabolism. (A) aspartate to homoserine; (B) glutamate to glutamine.

A compilation of >4000 aspartate-containing response regulator proteins from some 200 bacterial species allows for some generalizations to be made in logic and protein architecture.[53] Most of the variants in a given bacterial cell probably arose by gene duplication and evolution of new specificities. The core response regulator domain that undergoes aspartyl phosphorylation from a cognate sensor phosphohistidine residue could be joined to other kinds of domains. Most common are the DNA binding domains of bacterial helix–turn–helix transcription factors.

A typical consequence of phosphorylation of the conserved aspartyl-β carboxylate side chain in a receiver domain is dimerization, engendering the consequent dimerization of the attached, typically downstream, DNA binding domain[4] (Figure 14.6). These latter domains tend to function as dimers since each subunit recognizes an oligonucleotide sequence on one strand of DNA. The dimer, driven by Asp-phosphorylation in the adjacent receiver domain, can now read both strands of a target DNA promoter region, usually turning on gene expression. This logic may play out in 50–150 parallel, converging, intersecting two-component systems to offer an integrated response to factors in the local environment of a bacterial cell.

The other, less commonly observed but canonical architecture of receiver domains is fusion to some enzymatic activity that is regulated, either up or down in a specific context, by aspartate phosphorylation. One such adjacent catalytic domain is found in the carboxymethylesterase domain of the chemotaxis protein CheB. Methylation of one or more glutamate side chain carboxylates alters the electrostatic and solubility properties of companion chemotactic protein components.

Of more direct relevance to phosphorus chemical biology in bacteria, the receiver domain is also sometimes found attached to both di-cyclic GMP synthases and to the corresponding cyclic di-cGMP phosphodiesterases, two opposing sets of enzymes that control the amount and duration of di-cyclic-3′,5′-GMP discussed back in Chapter 5.[54] Aspartate phosphorylation by the upstream sensor histidine kinase tends to activate the synthase domain to produce more of the signal molecule di-cGMP and also to inhibit the phosphodiesterase.

The off-switch logic in phosphoproteomics, to undo the effects of phosphorylation of a set of protein targets, varies with the chemical nature of the phosphorylated amino acid side chain. We will turn to the well-known triad of P-Ser, P-Thr, and P-Tyr side chains in the next chapter. Their chemical stability as alcoholic phosphate monoesters required the invention and evolution of phosphoprotein phosphatases to remove the $-PO_3^{2-}$ groups in a meaningfully fast cellular time frame.

In this chapter we have noted that that $N-PO_3^{2-}$ groups tend to be thermodynamically activated *and* kinetically labile, functioning largely as transient, rather than stable phosphoproteins. For the most part the P-His residues breakdown as transient intermediates in enzymatic catalytic cycles. There are isolated examples of phosphohistidine protein phosphatases, noted earlier in this chapter Separately, the P-Arg residues in bacterial

phosphoproteins appear to be used as stable markers of proteins for chambered protease hydrolytic degradation. There is the need for and the experimental validation that an editing phosphatase is expressed to occasionally remove the PO_3^{2-} from a phosphoarginine protein to rescue it from an accelerated cellular death. There are few S–PO_3^{2-} Cys proteins, relegated largely to the phosphotyrosyl protein phosphatases category.

The half-lives of aspartyl phosphates in the thousands of known bacterial two-component systems vary from a few minutes (~5 minutes) to hours.[4] Some of the sensor histidine kinases accelerate the hydrolysis of their cognate aspartyl phosphate response regulator pairs. There are no other dedicated categories of protein phosphatases for those acyl phosphate forms of bacterial response regulators.

14.6.2 Aspartyl Phosphate Intermediates in P-type ATPase Catalytic Cycles

In Chapter 3 on ATPases, we noted that the thermodynamically favored hydrolysis of ATP was typically coupled to a set of conformational changes in the catalytic subunit of an ATPase, transmitted to any partner or regulatory subunits and from there used to drive changes in more distant protein partners to achieve disparate forms of cellular work. As discussed in Chapter 3, such membrane-embedded ATPases are often categorized into three separate structural and functional classes. The multi-subunit F-Type ATPases in prokaryotic plasma membranes and in eukaryotic chloroplast and mitochondrial membranes function in the opposite direction to ATP synthases and operate as the major sources of ATP biosynthesis. The V (vacuolar) type ATPases are usually proton pumps, moving H^+ ions against their concentration gradients to acidify the vacuole or compartment lumens. The third class are the transmembrane P-type ATPases that couple ATP hydrolysis to pumping specific cations across either plasma membranes or internal membranes, typically endoplasmic reticular compartments[55] (Figure 14.31).

The P in P-type ATPases reflects the designation that they all form covalent phosphoenzyme intermediates, specifically aspartyl phosphate residues, as drivers of conformational changes that allow pumping by specific ATPase members of H^+, Na^+, K^+, Ca^{2+}, Zn^{2+}, Mg^{2+}, Ca^{2+}, and Cu^{++} against their concentration gradients.[58] P-type ATPases have probably evolved multiple times to carry out the essential pumping of ions in and out of cells and cell compartments to establish and maintain the electrochemical gradients that govern so much of cell biology.

There are some 30 P-type ATPase variants in animal cells with alternative splicing in specific tissues to provide distinct functional capacities, both for ion specificities and for different flux capacities.[59,60] P-Type ATPases labor unceasingly to pump their specific cations uphill against thermodynamic concentration gradients, as illustrated by the Na pump, extruding three Na^+ cations for each ATP hydrolyzed and allowing two external K^+ ions to run

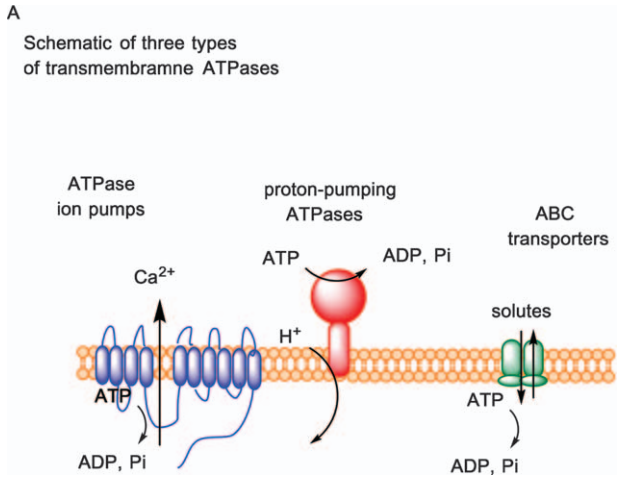


Figure 14.31 Schematic orientation of three types of transmembrane ATPases. Note the ATP binding and active sites are all in the cytoplasmic domains.[56]

down their concentration into cell cytoplasm. The Na pump is thus electrogenic. Estimates are that P-type ATPases in aggregate may account for between 30 and 40% of all ATP consumed in higher organisms. In humans we have noted earlier that the daily total of ATP synthesis can reach 75 kg. The ATPase ion pumps would then be spending 20–30 kg of ATP per day!

Typically, P-type ATPases have a single catalytic subunit with three cytoplasmic domains and a transmembrane domain with 10 or more transmembrane helices. Figure 14.32 shows a schematic of the Na/K-ATPase catalytic subunit, often termed the sodium pump. Two additional noncatalytic subunits comprise the functional pump.

Figures 14.33–14.35 compare the overall architecture and membrane placement of the Na/K ATPase catalytic subunits as an ion pump *vs.* the simpler K^+ ion channel, Ksca from bacteria. The Ksc bacterial K^+ channel is shown in part as a reminder of the definitional distinction between ion channels and ion pumps. Ion channels allow specific ions to flow *down* their electrochemical gradients. These do not require energy inputs. The flux can exceed 10^7 ions per second per channel protein. In contrast the P type ATPase pumps have fixed stoichiometries of only 1–3 cations transported per ATP hydrolyzed and have turnover rates in the range of 5–10 catalytic events per second. This is a dramatic difference of about a million-fold in throughput rates of channels over energy-dependent pumps. Structurally, channels may have one gate that closes and opens. It can be as small as a single residue movement to open or close a channel or be as large as movement of a four-helix bundle gate. Ion pumps on the other hand are thought to have two gates that open and close reciprocally (Figure 14.34) to control unwanted flux in the thermodynamically favored direction that would undo all the uphill movement of ions.

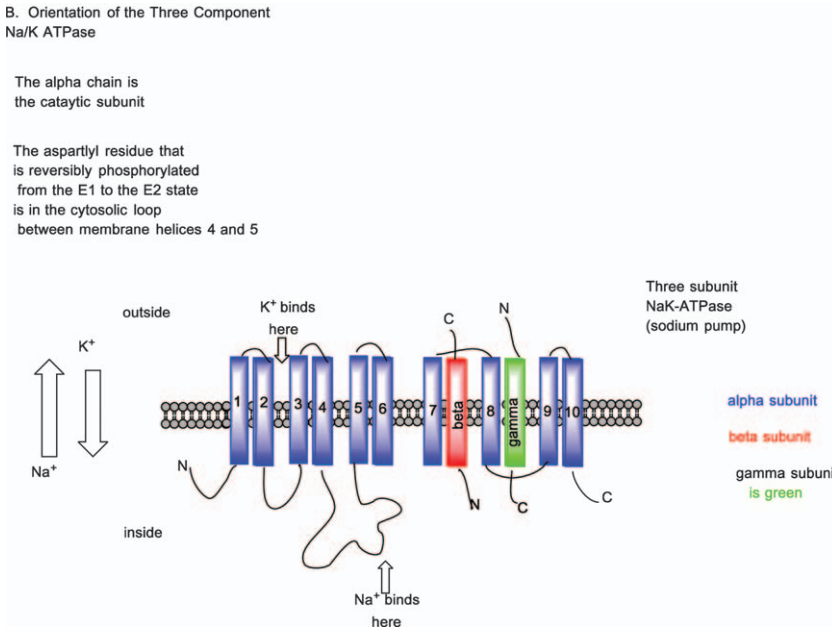


Figure 14.32 The Na/K P-type ATPase has a catalytic α-subunit and two small, noncatalytic β- and γ-subunits with single membrane spanning domains. The aspartate residue that forms the transient catalytic aspartyl-phosphate intermediates is in the cytoplasmic loop of the α-subunit between transmembrane helices 4 and 5.[57]
Adapted from ref. 57, https://doi.org/10.3390/molecules22040578, under the terms of a CC BY 4.0 license, https://creativecommons.org/licenses/by/4.0/.

The ion pumps of the P-type ATPase by contrast work to pump specific ions uphill against their concentration gradients. Three examples are the H^+/K^+ ATPases in the plasma membrane of gastric parietal cells that pump protons into the acidifying gastric lumen, the Na^+/K^+ pump noted above (Figure 14.36), and the sarcoplasmic endoplasmic reticulum calcium ATPase (SERCA) pump for calcium ions in skeletal and cardiac muscle cells.[60] After muscle contraction has released Ca^{2+} ions from sarcoplasmic reticulum stores into the cytoplasm, the SERCA pump spends ATP to pump the Ca^{2+} ions back into the sarcoplasmic (endoplasmic) reticular lumen against the concentration gradient to lower cytoplasmic Ca^{2+} below 10^{-7} M and thus shut down the several enzymes that show calcium-dependent activities. The stoichiometry is two Ca^{2+} ions pumped for each ATP hydrolyzed; in turn two or three H^+ flow in the opposite direction, from the sarcoplasmic reticular lumen into the cytoplasm in the second half reaction of the SERCA pump.

The first P-type ATPase characterized was reported in 1957[65] and 70 plus years of study of this class of energy-requiring ion pumps have led to a

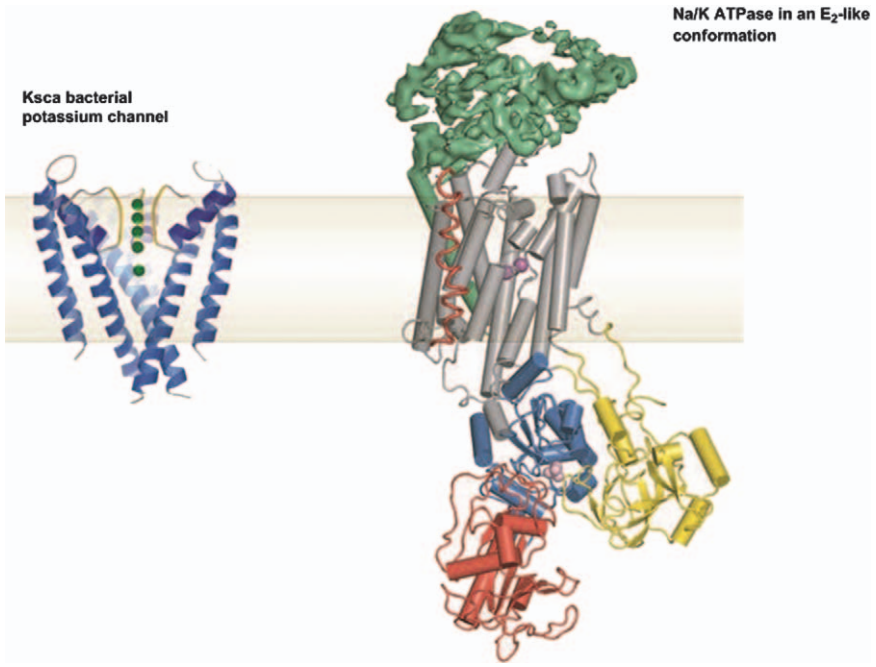


Figure 14.33 Comparison of a bacterial ion channel for potassium (Ksca) with the eukaryotic Na/K ATPase; there is distinctly greater complexity in the ion pump compared with the ion channel.[61]
Reproduced from ref. 61 with permission from Springer Nature, Copyright 2009.

mountain of literature and deep molecular insights into the structures of these protein machines.[60] As in many other aspects of molecular and cell biology where phosphorus-containing molecular scaffolds enable complex biology – from ATP synthases, to nucleic acid replication and repair, RNA splicing, and the two-component sensing and response regulating systems of bacteria – there are many thousands of primary articles and dozens to hundreds of reviews that allow deep dives into the details of the biology. Such is the case with the ATP-dependent ion pumps. We endeavor to illustrate only the principles of phosphorus chemical biology in the specific context. Here, the key question is how ATP hydrolysis 50 angstroms away from the ion binding sites leads to anisotropic and thermodynamically uphill pumping of ions from one side of a cellular membrane to the other.

The three cytoplasmic subdomains of the Ca^{2+} and the Na^{+} ATPase pumps have been examined intensely for mechanistic and structural features. The ATP binding site coordinates two Mg^{2+} ions that help align the Pγ of bond ATP in proximity to the aspartyl side chain in the next domain and to shield the charges of the triphosphate side chain that accelerates $-PO_3^{2-}$ transfer to generate the covalent aspartyl-P enzyme intermediate (Figures 14.34 and 14.36).

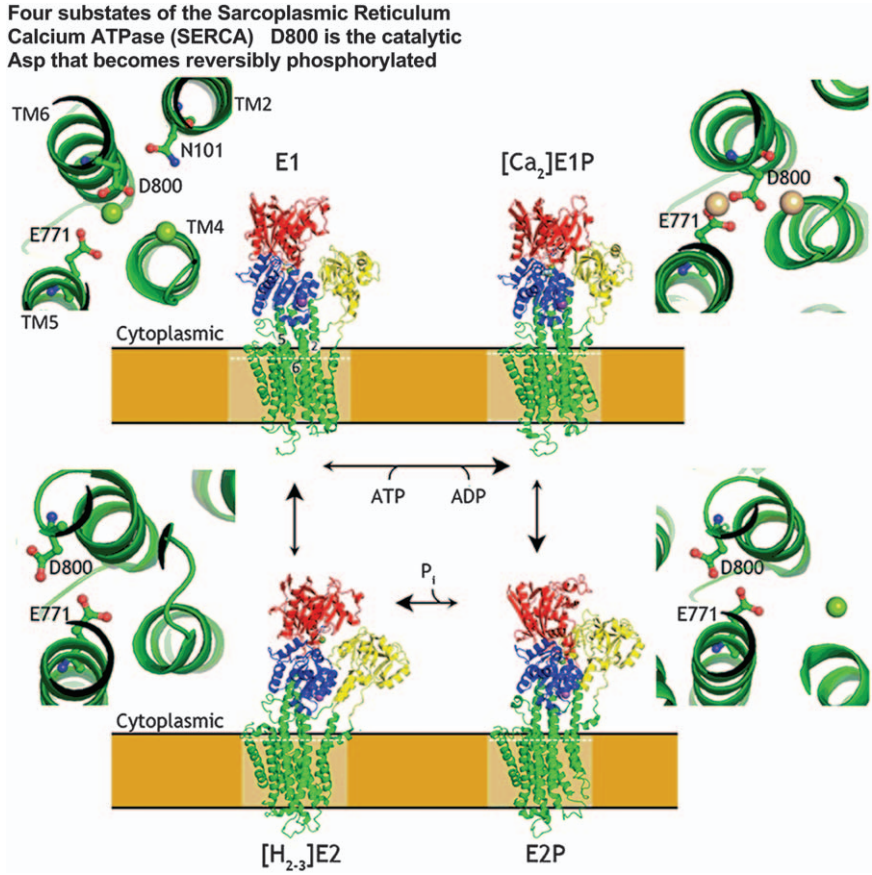


Figure 14.34 Snapshots of four states of the Ca^{++}-pumping ATPase of sarcoplasmic/endoplasmic reticulum (SERCA) in the E1, and E2 (inward and outward facing conformations).[62]
Reproduced from ref. 62, https://doi.org/10.3389/fphys.2017.00202, under the terms of a CC BY 4.0 license, https://creativecommons.org/licenses/by/4.0/.

The phosphoryl group transfer from ATP to the enzyme aspartate side chain in SERCA occurs after two Ca^{2+} ions have been bound to the cytoplasmic side of the ten transmembrane helical domain some 50 angstroms away. The introduction of the tetrahedral dianionic phosphate group on the aspartate side chain propagates conformational change all the way to the transmembrane domain. Half of the helixes separate enough from the other set to allow passage of the two calcium ions through an aqueous microenvironment. In that conformation the Ca^{2+} ions are less tightly bound, perhaps due to distortion of their ligand sets on the protein. They dissociate and are replaced in SERCA by two to three protons. That occupancy helps drive dephosphorylation and return of the catalytic subunit back to the

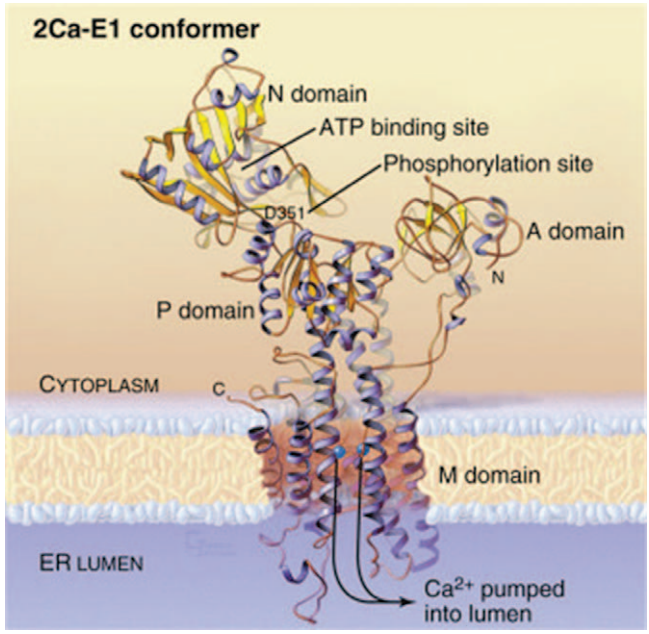


Figure 14.35 Proposed architecture of the 2Ca-E1 state of SERCA in the ER membrane of muscle cells where calcium ions are pumped back out of the cytoplasm into the ER lumen.[63]
Reproduced from ref. 63 with permission from Elsevier, Copyright 2017.

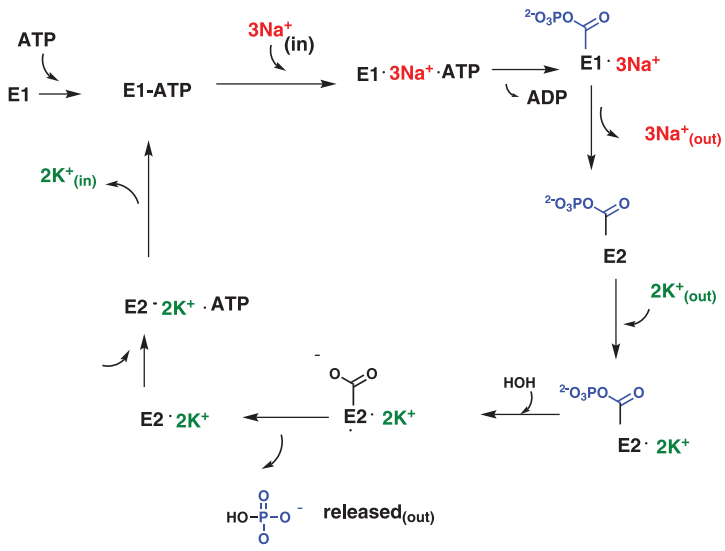
original state, with the protons coming through from sarcoplasmic reticulum to cytoplasm.

In a sense one can generalize that the 20–30 kg of ATP consumed hydrolytically by the P-type ATPases every day are classic examples of coupling of the energy released across large protein conformational changes to carry out otherwise unfavorable tasks. Those unfavorable equilibria are the pumping of protons, sodium ions, calcium ions, and other ions against their concentration gradients to establish and maintain the transmembrane electrochemical potential differences across plasma and intracellular membranes of cells.

The evolution of aspartyl–PO_3^{2-} as the covalent adduct distinguishing two conformational extremes of these coupling ATPases maintains the high phosphoryl group transfer potential and offers a rapidly hydrolysable acyl phosphate, functional reversal, without intervention of separate phosphoprotein phosphatases. This allow the ion pumps to run at the 5–10 s^{-1} turnover rates because of that kinetic lability. The role of the covalent aspartyl phosphate as driver of conformational change across substantial protein landscapes in the P-type ATPases[66] is notionally analogous to the role that the aspartyl phosphate side chains play in driving conformational changes in adjacent DNA binding domains in bacterial response regulators.

Catalytic Cycle for Na^+/K^+ ATPase

E1. Conformation: Active site accesible in cytoplasm



E1 to E2 conversion extrusion of Na^+ as enzyme changes its orientation from "in" to "out"

E2 conformation: enzyme open to extracellular milieu

Net: ATP hydrolysis to ADP and Pi drives extrusion of $3Na^+$ and entry of $2K^+$ ions:

electrogenic transmembrane pump

14.6.3 Phosphoserine Phosphatase

One final metabolic example of an aspartyl phosphate covalent intermediate occurs in each catalytic cycle of the last enzyme in the serine biosynthetic pathway. The three-carbon framework of serine derives ultimately from glucose-6-phosphate *via* glycolysis. Some of the glycolytic intermediate 3-phosphoglycerate (Figures 14.37 and 14.38) is siphoned off for the serine biosynthetic pathway. Enzymatic oxidation of the 2-OH to the 2-ketone in an NAD-dependent enzymatic process yields 3-phosphopyruvate. Transaminase action reductively aminates the ketone to the amino group of 3-phospho-L-serine.

The final enzyme, phosphoserine phosphatase, engages in covalent catalysis. The kinetically and chemically competent enzyme nucleophile is the side chain of an active site aspartate, creating the aspartyl phosphate intermediate as serine is released.[67] The phosphoenzyme is then hydrolyzed in the second half reaction to regenerate the starting form of the enzyme and yield coproduct inorganic phosphate. In preceding chapters we have noted covalent phosphoenzyme intermediates involving a serine side chain (phosphoglucomutase) and a histidine side chain (glucose-6-phosphatase). The *S*-phospho cysteinyl intermediate in PTPases is the third class. The phosphoserine phosphatase adds a fourth enzyme nucleophile to the list (Figure 14.38).

From the perspective of this chapter, the rapid turnover of the phosphoenzyme intermediate in phosphoglucomutase (Chapter 7) of the stable phosphate monoester covalent adducted form is the surprise We will come back to transient serine residue phosphorylation in a discussion of lanthipeptide synthetases in the next chapter.

14.6.4 The Search for Protein Glutamyl Phosphates

Given the primacy of acyl phosphate derivatives of aspartate side chains in the widespread two component systems of bacteria and the P-type ATPases noted above, one might expect similar roles for glutamyl side chain carboxylates in phosphoproteins. There are examples from maturation of the coenzyme forms of vitamin B9, tetrahydrofolic acid, and also from post-translational side chain glutamylations of proteins that implicate peptidyl- and protein-based glutamyl-γ-phosphates as reaction intermediates.

Figure 14.36 Schematic cycle for the two state model for Na/K ATPase open to cytoplasm (E1 form) or open to external medium(E2 form) with proposed timing of binding of three Na^+, enzyme phosphorylation on the aspartyl side chain, pumping the sodium ions out of the cell, binding of two K^+ on the outside, dephosphorylation, and release of potassium ions before the next cycle.[64]

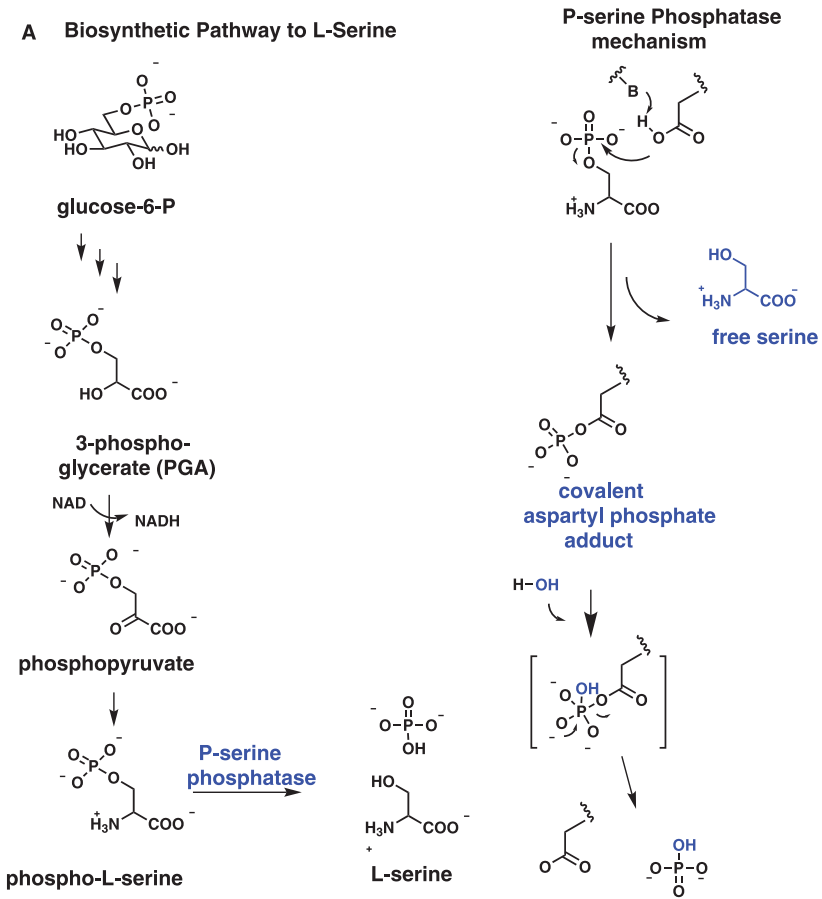


Figure 14.37 Phosphoserine is on pathway from glycolysis to serine biosynthesis with phosphatase action as the last enzymatic step. Phosphoserine phosphatase engages in covalent catalysis *via* an acyl-phosphate intermediate.

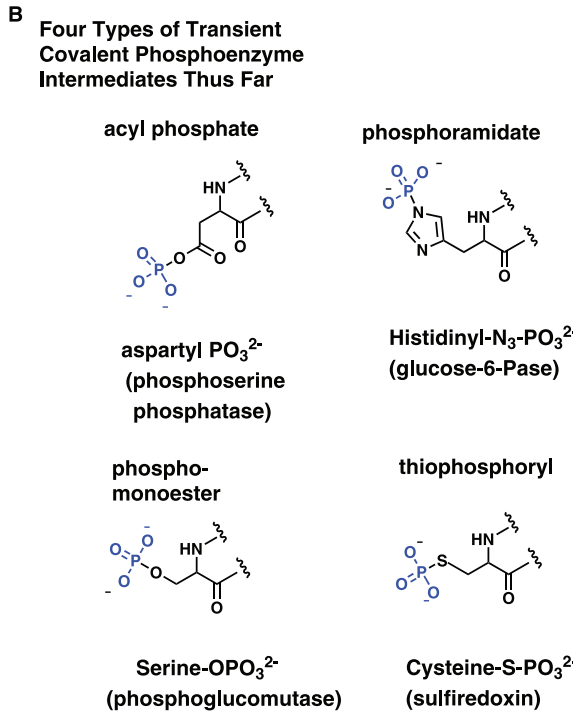


Figure 14.38 Types of covalent phosphoenzyme catalytic intermediates discussed thus far.

14.6.4.1 Folyl Polyglutamate Synthetase

Folic acid is the human vitamin B9.[68] It is taken up from plant dietary sources as the fully oxidized bicyclic pterin heterocycle with a single glutamate side chain (Figure 14.39). Two metabolic transformations are required to convert the vitamin to active coenzyme forms. One is double two-electron reduction of the right-hand pyrazine ring to the tetrahydro form (H_4-folyl monoglutamate) for N5 to become basic to function as a nucleophile in one-carbon metabolic transfers to and from that nitrogen atom.

The second required transformation is the addition of from one to six glutamate residues to the monoglutamate tail[69] (Figure 14.39). The suite of human enzymes using folate coenzymes are inactive with the monoglutamate and require the oligoglutamyl polyanionic tails for high-affinity recognition.[68] The folylpolyglutamate synthetase cleaves ATP to ADP and ultimately inorganic phosphate as it activates the terminal γ-carboxylate of the glutamate tail as the indicated acyl phosphate. This sets up formation of the isopeptide (Glu-γ-Glu) linkage in the product H_4-folyl-diglutamate. Iteration of this acyl phosphate activation logic yields the triglutamate as shown. The tail can eventually grow to a heptaglutamate, *via* another four acyl-phosphate

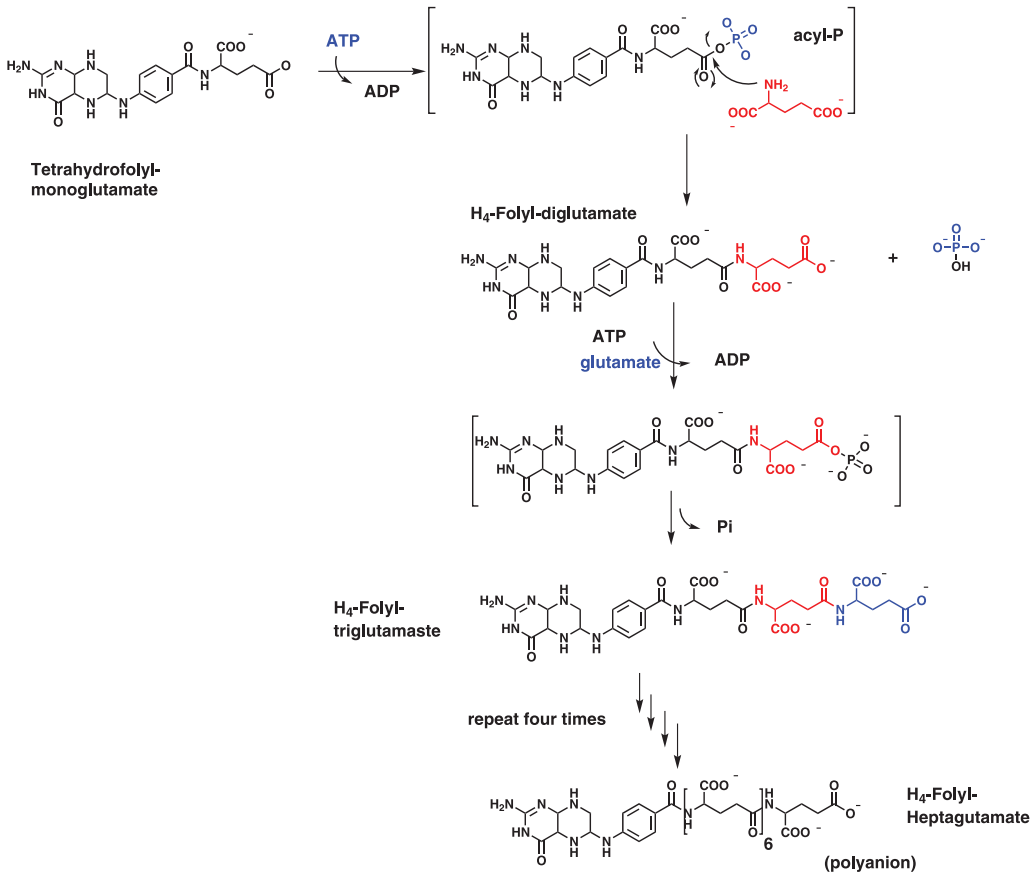


Figure 14.39 A cascade of peptidyl γ-glutamyl phosphates are intermediates in the chain elongation reactions of ATP-dependent folyl oligoglutamate synthetase.

intermediates. To get to the predominant hexaglutamyl form of folate in human tissues, six γ-glutamyl phosphates are intermediates.

14.6.4.2 Protein-γ-Glutamyl Phosphates as Intermediates in Posttranslational Oligoglutamylations

While γ-phosphoglutamyl acyl phosphates have not been detected as *accumulating* covalent adducts in proteomes, they are implicated as intermediates in the oligoglutamylations and polyglutamylations of both α- and β-chains of mammalian tubulins and recently in the oligoglutamylation of a protein SidE that *Legionella pneumophila* bacterial pathogens inject into infected host cells.[70] Each glutamyl residue added to a target protein requires Mg–ATP as a cosubstrate. The ATP is cleaved to ADP and ultimately Pi is also released. The γ-carboxylate of the glutamyl moieties are activated as the acyl phosphates that then serve as the electrophiles in peptide bond formations (as in the folyl oligoglutamate synthase case above).

The recently identified *Legionella pneumophila* SidJ protein adds up to three glutamyl residues, one at a time, to a specific glutamate side chain, Glu_{880}, on its target protein SidE.[70] SidE in turn, as will be detailed in Chapter 16, is itself *a L. pneumophila* protein that, like SidJ, is one of some 300 proteins the bacterial pathogens release into host human host cells. The modification of SidE by SidJ is thought to proceed by initial formation of the acyl phosphate (mixed anhydride) on the side chain of Glu_{880} of SidE (Figure 14.40). This mixed anhydride is then attacked by the amine of the cosubstrate free glutamate, forming an isopeptide bond as the monoglutamyl–SidE is generated. This process can be repeated a second time. Now, the γ-carboxylate of the newly added glutamyl group acts as nucleophile to form the next iteration of a protein-γ-glutamyl-phosphate, to be attacked by the amino group of the free glutamate cosubstrate. This builds a bis-glutamylated form of the Glu_{880} residue of SidE, inactivating it.

Tubulin chain glutamylations follow the same logic and glutamylations are the main rheostat for control of severing action on microtubules.[71] As few as two to as many as 7 to 11 oligoglutamyl residues can decorate the C-termini of tubulin subunits in neurons; axonemal tubulins are decorated with up to 21 isopeptide glutamyl chains. In each case one presumes protein-based γ-glutamyl phosphates are the activated side chains enabling isopeptide bond formation and oligoglutamyl chain elongation Glutamyl tails recruit the severing enzyme spastin, itself an AAA-type ATPase (see Chapter 3) to microtubules and to initiate microtubule cleavages. Although the glutamyl-γ-PO_3^{2-} species are cryptic and phosphate does not end up as part of the covalent oligoglutamyl chains, the protein-glutamyl phosphates are obligate intermediates.

Against this backdrop, the literature on phosphoproteins does not yet include a report of any authenticated identification of phosphoglutamyl residues as stable acyl phosphate derivatives. Yet, the recent report of Hardman *et al.*[2] revealed that their mass spectrometry-based analysis of

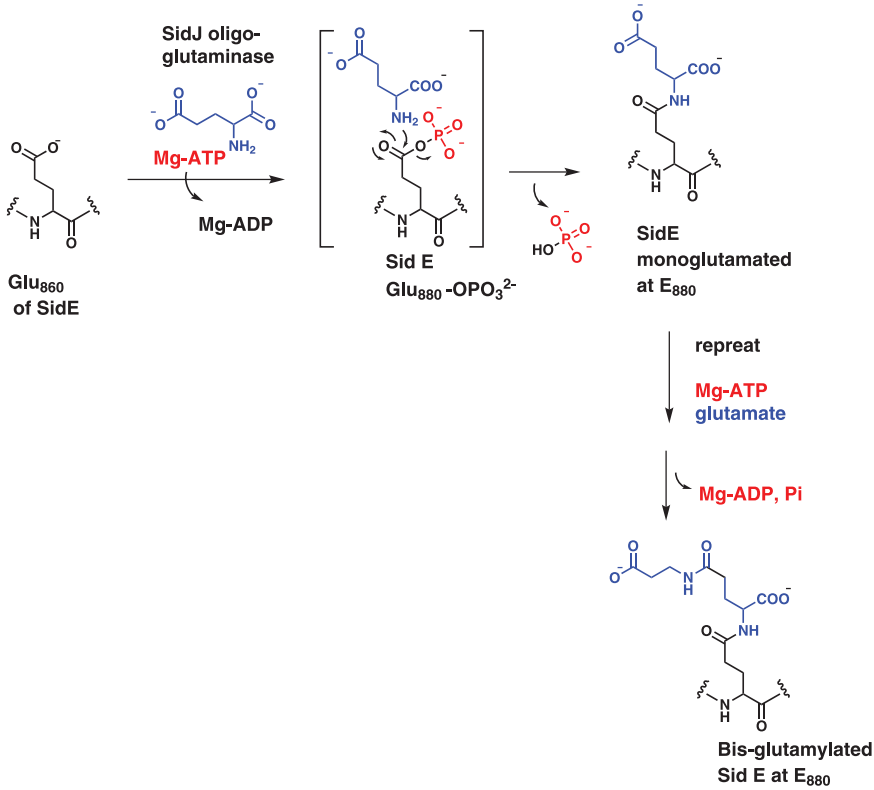


Figure 14.40 Oligoglutamylation of the *Legionella pneumophila* enzyme SidE by SidJ involves Mg–ATP and glutamate in each elongation step. The side chain of glutamate$_{880}$ in SidE is activated as the acyl phosphate and then captured by the cosubstrate free glutamate to make an isopeptide bond. The process is repeated in each chain elongation step. The protein glutamate residue acts as nucleophile towards Pγ of ATP. Then in the next step, that acyl phosphate carbonyl is the electrophile for capture by the amino group of cosubstrate free glutamate.

phosphopeptides from HeLa cell protein extracts identified more than 1000 phosphoglutamyl peptides as candidates for further characterization. It may yet turn out that kinetically stable phosphoglutamyl proteins join the ranks of noncanonical phosphoproteomes, unless all those mass-based phosphoglutamyl peptide data are somehow artifactual.

References

1. H. J. Evans and H. G. Wood, The mechanism of the pyruvate phosphate dikinase reaction, *Proc. Natl. Acad. Sci. U. S. A.*, 1968, 1448–1453.

2. G. Hardman, S. Perkins, P. J. Brwnridge, C. J. Clarke, D. P. Byrne, Campbell Ae, A. Kalyuzhnyy, A. Myall, P. A. Eyers, A. R. Jones and C. E. Eyers, Strong-anion exchnange-mediated phosphoproteomics reveals extensive human noncanonical phosphorylation, *EMBO J.*, 2019, **38**, e100847.
3. K. Adam and T. Hunter, Histidine kinases and the missing phosphoproteome from prokaryotes to eukaryotes, *Lab. Invest.*, 2018, **98**, 233–247.
4. E. J. Capra and M. T. Laub, The evolution of two-component signal transduction systems, *Annu. Rev. Microbiol.*, 2012, **66**, 325–347.
5. E. Li and K. Hristova, The role of tyrsoine kinase transmembrane domains in cell signaling and human pathologies, *Biochemistry*, 2006, **45**, 6241–6251.
6. G. Jones and P. Dyson, Evolution of transmembrane protein kinases implicated in coordinating remodeling of gram-positive peptidoglycan: inside vs outside, *J. Bacteriol.*, 2006, **188**, 7440–7476.
7. M. P. Bhate, K. S. Molnard, M. Goulian and W. F. DeGrado, Signal transduction in histidine kinases: insights from new structures, *Structure*, 2015, **23**, 981–992.
8. T. Hunter, Tyrosine phosphorylation: 30 years and counting, *Curr. Opin. Cell Biol.*, 2009, **21**, 140–146.
9. S. R. Hubbard and J. H. Till, Protein tyrosine kinase structure and function, *Annu. Rev. Biochem.*, 2000, **69**, 373–398.
10. D. R. Robinson, Y.-M. Wu and S.-F. Lin, The protein tyrosine kinase family of the human genome, *Oncogene*, 2000, **19**, 5548–5557.
11. E. Groisman, The pleiotropic two component regulatory system PhoP-PhoQ, *J. Bacteriol.*, 2001, **183**, 1835–1842.
12. T. Lemmin, C. Soto, Clinthorne, W. F. DeGrado and M. Dal Pararo, Assembly of the transmembrane domain of E.coli PhoQ Histidine Kinase: implications for signal transduction from molecular simulations, *PLoS Comput. Biol.*, 2013, **9**, e1002878.
13. MBINFO website (downloaded 11-09-2019) What are two-component regulatory systems, https://www.mechanobio.info/pathogenesis/what-are-two-component-regulatory-systems/.
14. S. Francis, K. Wilke, D. E. Brown and E. E. Carlson, Mechanistic insight into inhibition of two component signaling, *Med. Chem. Commun.*, 2013, **4**, 269–277.
15. P. Casino, L. Miguel-Romero and A. Marina, Visualizing autophosphorylation in histidine kinases, *Nat. Commun.*, 2014, **5**, 1–12.
16. H. U. Ferris, M. Coles, A. N. Lupas and M. D. Hartmann, Crystallographic snapshot of the Eschericia coli EnvZ histidine kinase in an active conformation, *J. Struct. Biol.*, 2014, **186**, 376–379.
17. G. Draughton, M. E. Milton, E. A. Feldmann, B. G. Bobay, B. M. Roth, A. L. Olson, R. J. Thompson, L. A. Actis, C. Davies and J. Cavanagh, The structure of the biofilm-controlling response regulator BfmR from Acinetobacter baumanii reveals details of its DNA-binding mechanism, *J. Mol. Biol.*, 2018, 430.

18. I. Mijakovic, C. Grangeasse and K. Turgay, Exploring the diversity of protein modifications: special bacterial phosphorylation systems, *FEMS Microbiol. Rev.*, 2017, **40**, 398–417.
19. C. J. Leonard, L. Aravind and E. V. Koonin, Novel families of putative protein kinases in bacteria and archaea: evolution of the ''eukaryotic'' protein kinase superfamily, *Genome Res.*, 1998, **8**, 1038–1047.
20. S. S. Taylor, M. M. Keshwani, J. M. Steichen and A. P. Kornev, Evolution of the eukaryotic protein kinases as dynamic molecular switches, *Philos. Trans. R. Soc., B*, 2012, **367**, 2517–2528.
21. G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsnam, The protein kinase complement of the human genome, *Science*, 2002, **298**, 1912–1934.
22. W. Kundig, S. Ghosh and S. Roseman, Phosphate bound to histidine in a protein as an intermediate in a novel phosphotransferase system, *Proc. Natl. Acad. Sci. U. S. A.*, 1964, **52**, 1067–1074.
23. M. Saier Jr, The bacterial phosphotransferase system: new frontiers 50 years after its discovery, *J. Mol. Microbiol. Biotechnol.*, 2016, **25**, 73–78.
24. P. Kotbra, M. Inui and H. Yukawa, Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism, *J. Biosci. Bioeng.*, 2001, **92**, 502–517.
25. J. Deutscher, C. Francke and P. W. Postma, How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria, *Micorbiol. Mol. Biol. Rev.*, 2006, **70**, 939–1031.
26. O. Herzberg, C. H. C. Chen, S. Liu, A. Tempczyk, A. Howrd, M. Wei, D. Ye and D. Dunaway-Mariano, Pyruvate site of pyruvate phosphate dikinase: crystallization of the enzyme phosphonpyruvate complex and mutant analysis, *Biochemistry*, 2002, **41**, 780–787.
27. Protopedia. (donloaded 12-01-2019) Pyruvate phosphate dikinase.
28. D. Phillips, A. Aponte and S. A. French, Succinyl-CoA synthetase is a phosphate target for the activation of mitochondrial metabolism, *Biochemistry*, 2009, **48**, 7140–7149.
29. K. H. G. Verschueren, C. Blanchet, J. Felix, A. Dansercoer, D. de Vos, Y. Bloch, J. van Beeumen, D. Svergun, I. Gutsche, S. N. Savvdes and K. Verstratete, Structure of ATP citrate lyase and origin of citrate synthase in the Krebs cycle, *Nature*, 2019, **568**, 571–575.
30. S. Morera, M. Chiadmi, G. LeBras, I. Lacu and J. Janin, Mechanism of phosphate transfer by nucleoside diphosphate kinase: x-ray structure of the phosphohistidine intermediate of the enzymes from Drosophila and Dictyostelium, *Biochemistry*, 1995, **34**, 11062–11070.
31. P. D. Wagner and N.-D. Vu, Histidine to aspartate phosphotransferase activity of nm23 proteins: phosphorylation of aldolase C on Asp319, *Biochem. J.*, 2000, **346**, 623–630.
32. J. E. Schulte and M. Goulian, The phosphohistidine phosphatase SixA targets the phosphotransferase system, *mBio*, 2018, 9.

33. P. Ek, G. Petterson, B. Ek, F. Gong, J. P. Li and O. Zetterqvist, Identification and characterzation of a mammalian phosphohistidine phosphatase, *Eur. J. Biochem.*, 2002, **269**, 5016–5023.
34. F. Voncken, F. Gao, C. Wadforth, M. Harley and C. Colasante, The phosphoarginine energy-buffering system of *Trypanosoma brucei* involves multiple arginine kinase isoforms with different subcellular locatons, *PLoS One*, 2013, 8.
35. E. Tan, P. G. Besant, X. L. Zu, C. W. Turck, M. A. Bogoyevitch, S. G. Lim, P. V. Attwood and G. C. Yeoh, Histone 4 histidine kinase dsiplays the expression pattern of a liver oncodevelopmental marker, *Carcinogenesis*, 2004, **25**, 2083–2088.
36. J. Furhmann, B. Mierzwa, D. B. Trentini, S. Spiess, A. Lehner, E. Charpentier and T. Clausen, Structural basis for recognizing phosphoarginine and evolving residue-specific protein phosphatase in grampositve bacteria, *Cell Rep.*, 2013, **3**, 1832–1839.
37. M. J. Suskiewicz, B. Hajdusits, R. Beveridge, A. Heuck, L. D. Vu, R. Kurzbauer, K. Hauer, V. Thoeny, K. Rumpel, K. Mechtler, A. Meinhart and T. Clausen, Structure of McsB, a protein kinase for regulated arginine phosphorylation, *Nat. Chem. Biol.*, 2019, 510–518.
38. D. B. Trentini, J. Furhmann, K. Mechtler and T. Clausen, Chasing phosphoarginine proteins: development of a selective enrichment using a phosphatase trap, *Mol. Cell. Proteomics*, 2014, **13**, 1958–1964.
39. D. B. Trentini, M. J. Suskiewiz, A. Heuck, R. Kurzbauer, L. Deszcz, K. Mechtler and T. Clausen, Arginine phosphorylation marks proteins for degradation by Clp protease, *Nature*, 2016, **539**, 48–53.
40. A. Tripathi and S. Gottesman, Phosphate on, rubbish out, *Nature*, 2016, **539**, 38–39.
41. H. Jung, Y. Choi, D. Lee, J. K. Seo and Kee J-m, Distinct phosphorylation and dephosphorylation dynamics of protein arginine kinases reveled by fluorescent activity probes, *Chem. Commun.*, 2019, 55.
42. X. Ye, G. Nalepa, M. Wiecker, B. M. Kessler, E. Spooner, J. Qin, S. J. Elledge, B. E. Clurman and J. W. Harper, Recognition of phosphodegron motifs in human cyclin E by the SCF (Fbw7) ubiquitin ligase, *J. Biol. Chem.*, 2004, **279**, 50110–50119.
43. C. Li, Y.-Y. Liang, X.-H. Feng, S. Y. Tsai, M. T. Tsai and B. W. O'Malley, Essential phosphatases and a phospho-degron are essential for regulation of SRC-3/A1B1 coactivator function and turnover, *Mol. Cell. Biol.*, 2008, **31**, 835–849.
44. S. Fu, C. Fu, Q. Zhou, R. Lin, H. Ouyang, M. Wang, Y. Sun, Y. Liu and Y. Zhao, Widespread arginine phosphorylation in human cells-a novel proteim PTM revealed by mass spectrometry, *bioRXiv*, 2019.
45. C. T. Walsh, *The Chemical Biology of Sulfur*, Royal Society of Chemistry, London, 2020.
46. F. Sun, Y. Ding, Q. Ji, Z. Liang, X. Deng, C. C. Wong, C. Yi, L. Zhang, S. Xie, S. Alvarez, L. M. Hicks, C. Luo, H. Jiang, L. Lan and C. He, Protein cysteine phosphorylation of SarA/MgrA family of transcriptional

regulators mediates bacterial virulence and antibiotic resstance, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 15461–15466.
47. H. H. Pas and G. T. Robillard, *S*-Phosphocysteine and phosphohistidine are intermediates in the phosphoenolpyruvate-dependent mannitol transport catalyzed by *E.col*i EIImtl, *Bicohemistry*, 1988, **27**, 5835–5839.
48. K. L. Guan and J. E. Dixon, Evidnce for protein tyrosine phosphatase catalysis proceeding via a cysteine-phosphate intermediate, *J. Biol. Chem.*, 1991, **266**, 17026–17030.
49. J. M. Denu, D. Lohse, J. Vijayalakshmi, M. Saper and J. E. Dixon, Visualization of intermediate and transition state structure in protein tyrosine phosphatase catalysis, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 2493–2498.
50. A. Scott and Z. Wang, Tumor suppressor function of protein tyrosine phosphatase receptor T, *Biosci. Rep.*, 2011, **31**, 303–307.
51. S. G. Rhee and H. Woo, Multiple functions of preoxiredoxins: peroxidases, sensors, and regulators of the intracellular messenger H2O2 and protein chaperones, *Antioidants Redox Signaling*, 2010, **15**, 781–794.
52. T. J. Jonsson, M. Murray, L. C. Johnson and W. T. Lowther, Reduction of cysteine sulfinic acid in peroxiredoxin by sulfiredoxin proceeds through a sulfinic phosphoric ester intermediate, *J. Biol. Chem.*, 2008, **283**, 23846–23851.
53. M. Y. Galperin and E. V. Koonin, A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity, *Protein Sci.*, 1997, **6**, 2639–2643.
54. M. Christen, B. Christen, M. Folcher, A. Schauerte and U. Jenal, Identification and characterization of a cyclic di-GMP phosphodiesterase and its allosteric control by GTP, *J. Biol. Chem.*, 2005, **280**, 30829–30837.
55. M. Bublitz, J. P. Morth and P. Nissen, P-type ATPases at a glance, *J. Cell Sci.*, 2011, **124**, 2515–2519.
56. Gardner. (downloaded 11-09-2019) ATPase pumps.
57. C. F. Goncalves-de-Alberquerque, A. R. Silva, C. I. de Silva, H. C. Castro-Faria-Neto and P. Burth, Na/K pumps and beyond: Na/K ATPase as a modulator of apoptosis and autophagy, *Molecules*, 2017, **22**, 578.
58. T. L.-M. Sorensen, J. V. Moller and P. Nissen, Phosphoryl transfer and calcium ion occlusion in the calcium pump, *Science*, 2004, **304**, 1672–1675.
59. M. G. Palmgren and K. B. Axeson, Evoution of P-type ATPases, *Biochim. Biophys. Acta*, 1998, **1365**, 37–45.
60. M. G. Palmgren and P. Nissen, P-type ATPases, *Annu. Rev. Biophys.*, 2011, **40**, 243–266.
61. D. C. Gadsby, Ion channels vs ion pumps: the principal difference in principle, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**, 344–352.
62. D. Focht, T. I. Croll, B. R. Pedersen and P. Nissen, Imporved model of proton pump crystal structure obtained by interactive molecular

dynamics fitting expands the mechanistic model for proton translocation in P-type ATPases, *Front. Physiol.*, 2017, 11.
63. *Cell Biology*, ed. T. D. Pollard, W. C. Earnshaw, J. Lippincott-Schwartz, G. T. and Johnson, Structure of a P-type ATPase pump, Membrane Pumps, 3rd edn, ch. 14, 2017, pp. 241–252.
64. R. H. Morris and A. Hadzovic, A proposed cycle for the Na/K ATPase pump (U of Toronto chemistry course notes), 2012.
65. J. C. Skou, The influence of some cations in an adenosine triphosphatase from peripheral nerves, *Biochim. Biophys. Acta*, 1957, **23**, 394–401.
66. X. C. Zhang and H. Zhang, P type ATPases use a domain association mechanism to coule ATP hydrolysis to conformational change, *Biophys. Rep.*, 2019, **5**, 167–175.
67. J.-F. Collet, V. Stroobant and E. Van Schaftingen, Mechanistic studies of phosphoserine phosphatase, an enzyme related to P-type ATPases, *J. Biol. Chem.*, 1999, **274**, 33985–33990.
68. C. T. Walsh and Y. Tang, *The Chemical Bilogy of Human Vitamins*, Royal Society of Chemistry, Cambridge, 2018.
69. K. C. Scrimgeour, biosynthesis of polyglutamates of folates, *Biochem. Cell Biol.*, 1986, **64**, 667–674.
70. A. Sulpizio, Minelli, M. Wan, P. D. Burrowes, X. Wu, E. J. Sanford, J.-H. Shin, B. C. Williams, M. L. Goldbverg, M. B. Smolka and Y. Mao, Protein polyglutamylation catalyzed by the bacterial calmodulin-dependent pseudokinase SidJ, *eLife*, 2019, **8**, e51162.
71. M. L. Valenstein and A. Roll-Mecak, Graded control of microtubule severing by tubulin glutamylation, *Cell*, 2016, **164**, 911–921.

CHAPTER 15

Canonical Phosphoproteomics: Phosphoserine, Phosphothreonine, and Phosphotyrosine

15.1 Scope of the Canonical Phosphoproteome

The noncanonical phosphoproteome of the preceding chapter examined three types of covalent phosphoproteins: N–P bonds in phosphoramidates, acyl phosphoric anhydrides, and S–P bonds in *S*-phosphocysteinyl–enzyme adducts. In this chapter the canonical phosphoproteome encompasses only phosphate monoester linkages to proteins. Two of them (Ser and Thr) are alkyl alcohols, the third (Tyr) is a phenol.

Canonical phosphoproteomics, comprising tens of thousands of research publications and reviews on the subject, deals with the compilation of phosphorylated serine, threonine, and tyrosine residues detected in various proteomes under different cellular conditions. More than 250 000 such phosphorylation sites have been projected in mammalian proteomes in the past two decades as acid-stable phosphopeptides whose sequences can be matched back to the intact proteins. The sheer numbers attest to the enormous reach of protein phosphorylation on these three residues.

A recent statistical analysis[1] based on 187 published data sets with identified phosphopeptide sequences computed that most of the anticipated phosphoproteins in the human proteome have been detected by this point, allowing projection of the probable full set. This methodology predicts that 13 000 human proteins (out of 21 000 in total) undergo phosphorylation at some point, comprising some 230 000 phosphorylation sites. This ratio gives

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

an average of ~17 residues per protein that can be phosphorylated in the human proteome. The corresponding values for the model organism yeast are 3000 phosphoproteins with ~40 000 phosphorylation sites, for a hypothetical average of 13 phosphorylatable residues for each of those yeast proteins that can be substrates. The 3000 number is about half of the 6000 proteins in the yeast proteome.

Earlier estimates on the frequency of P-Ser, P-Thr, and P-Tyr residues from a set of ~6000 phosphopeptides has indicated that P-Ser residues dominate at around 86% of the total. Phospho-Thr residues follow at around 12% abundance, with P-Tyr lagging at 1.8% abundance. Other studies put the ratio at 1800 : 200 : 1 = 90% : 10% : 0.05%.[2] More recent isolation of phosphopeptides by methods that should detect all nine amino acids that can be phosphorylated in HELA cells[3] found 4200 P-Ser peptides, 900 P-Thr peptides, and 220 P-Tyr peptides, for an abundance ratio of 79% to 17% to 4%. One additional global statistic is that human genome analysis predicts 90 protein tyrosine kinases and 428 serine or threonine protein kinase from the 518 total members of these two enzyme superfamilies.

Changes in phosphorylation patterns in response to both intracellular and extracellular perturbations have sustained an entire arena of protein-based signal transductions in both health and a plethora of disease states. In turn this knowledge base has fueled an onslaught of protein kinase inhibitors by the pharmaceutical and biotechnology industries to target specific proteins that undergo aberrant phosphorylation levels or sites in disease.

Dozens of authoritative reviews over the past two decades and entire volumes on distinct aspects of canonical phosphoproteomics have been written. Therefore, a single chapter is not able to go deeply into the many facets of the diverse biologies of this stable phosphoproteomics class. Expert readers will seek more detailed works of scholarship. Here, we will focus only on some general themes that tie this phosphoproteomics chemical biology back to several of the concepts set forth in the previous chapters.

15.2 Phosphoserine, Phosphothreonine and Phosphotyrosine are Chemically Stable Phosphomonoesters

The most notable chemical constraint on the canonical phosphoproteomics collection is that each of the three residues that undergo enzymatic phosphoryl transfers from ATP by action of protein kinases are alcoholic side chains (aliphatic for Ser and Thr; phenolic for Tyr), forming canonical phosphomonoesters. The phosphoryl transfer mechanisms follow the pattern for phosphoryl transfers in low molecular weight metabolites (*e.g.* glucose). Probably (partially) associative transition states from addition of the alcoholic oxyanion of Ser, Thr, Tyr side chains at Pγ of Mg–ATP decompose in the forward direction by a net in-line elimination of ADP to give the protein-linked phosphate monoesters (Figure 15.1). As with other

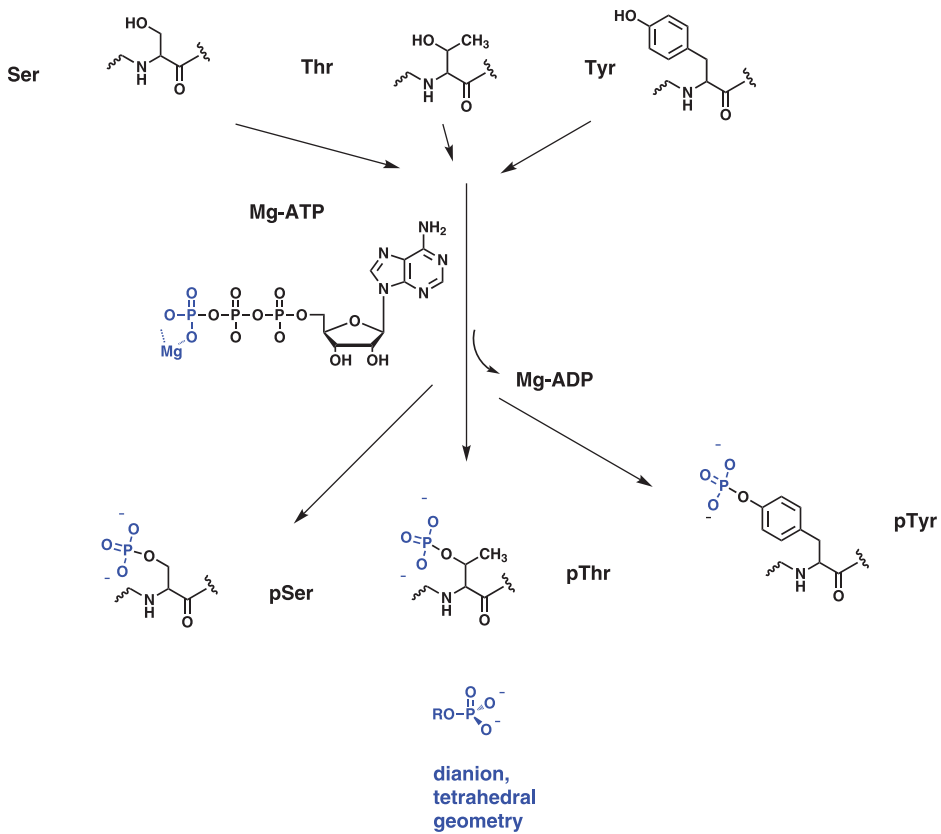


Figure 15.1 Creation of dianionic tetrahedral phosphomonoesters during protein phosphorylation of the canonical three residues Ser, Thr, and Tyr.

phosphomonoesters, these are chemically stable entities (half-lives in model phosphopeptides of years to decades). From the perspective of phosphoryl transfer mechanisms, all 517 human Ser–Thr and Tyr protein kinases would follow the same fundamental mechanism. Overlaid on the chemospecificity can be sequence specificity for neighboring residue side chains (*e.g.* negatively or positively charged residues), and secondary structures as well (*e.g.* a substrate residue at the end of a helix).

15.3 Consequences of Phosphate Monoester Chemical Stability

In contrast to the thermodynamically activated phosphoramidates, phosphorothioates, and acyl phosphate covalent protein adducts of the previous chapter, these phosphomonoesters are in general *not* activated for further $-PO_3^{2-}$ transfer on to other nucleophiles. They do *not* undergo *''cascade''* transfers chemically or enzymatically.

The chemical stability of P-Ser, P-Thr, and P-Tyr residues has several consequences. First, it is the basis for their detection by the tens of thousands by the canonical mass spectrometry workflow for identification of acid-stable phosphopeptides[4,5] (largely to the exclusion of the six phosphoresidues in the preceding chapter). Second, the chemical stability allows them to function as long-lived posttranslational modifications of many proteins within cells: a condition favorable for widespread protein-based signaling regimes. Third, the stable alcohol–O–PO_3^{2-} groups are not subsequently moving from one protein residue to the next. They cannot enable the kind of cascades seen in both bacterial two-component systems and the PEP-dependent hexose phosphotransferase systems of Chapter 14.

Thus, to signal phosphorylation state information about these three side chains from one protein to the next, eukaryotes invented a set of protein domains to recognize and interact with phosphoserine or phosphothreonine and separately phosphotyrosine residues (Figure 15.2). Signal propagation of a phosphorylated state is not by $-PO_3^{2-}$ group mobile transfer downstream but rather by recognition of the presence of one or more dianionic tetrahedral phosphate ester groups by a partner protein or domain. Figure 15.3 emphasizes the unique tetrahedral geometry of a stable phosphoseryl residue side chain in the upper panel. The lower panel shows a dianionic phosphotyrosyl residue compared with a monoanionic planar glutamyl residue at the end of a stretch of protein helix.

A fourth consequence of the ''almost infinite'' chemical stability of P-Ser, P-Thr, and P-Tyr residues under physiological conditions is the need for phosphoprotein phosphatases, if the $-PO_3^{2-}$ groups are to be removed. In principle, the cellular logic of canonical protein phosphorylation on Ser, Thr, and Tyr side chains of proteins could have been irreversible, akin to proteolysis. However, that would limit the capacity of signaling regimes.

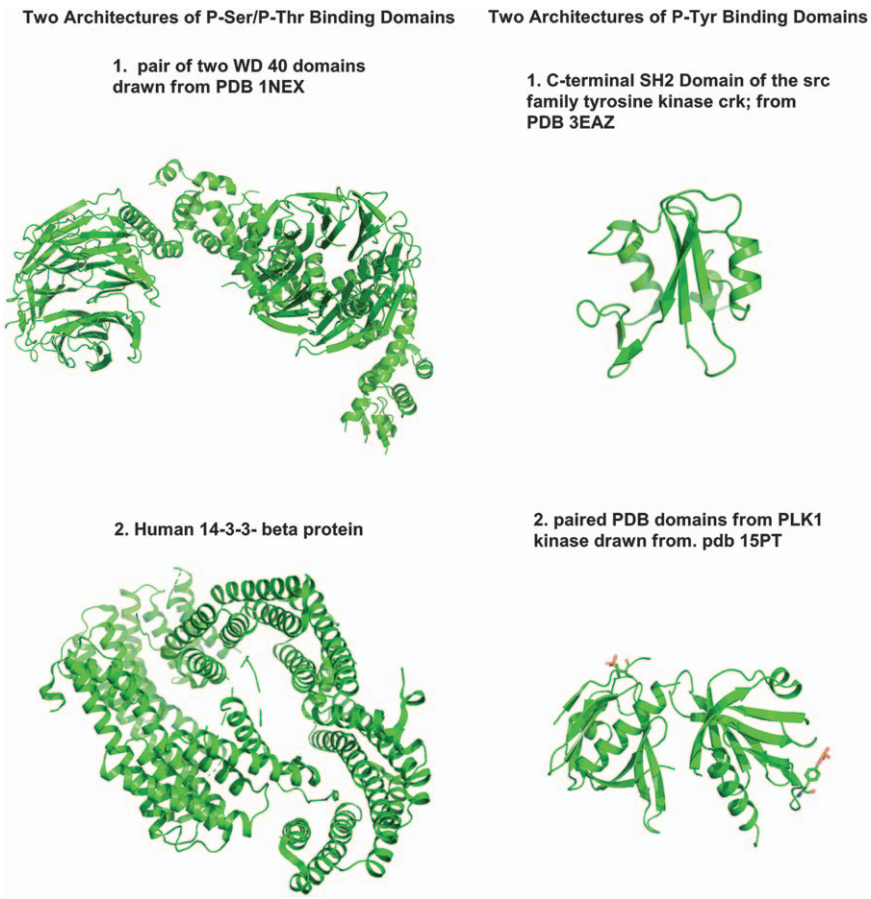


Figure 15.2 A sampling of protein domains that have evolved to bind P-Ser and P-Thr or P-Tyr residues include PBD (polo box domains), WD40 for P-Ser or P-Thr residues and SH2 and PTB domains for recognition of P-Tyr side chains.[6,7]

Biology is parsimonious about irreversible switches in steady state cellular situations. Much more useful in broad signaling regimes is the opportunity to turn on signals that become limited in time and space by subsequently turning those signals off. Execution of that logic amounts to dramatically accelerated hydrolysis of the stable phosphomonoester bonds by dedicated sets of enzymes. This is in contrast to the multiple cases of high kinetic lability of phosphoramidates, phosphorothioates, and acyl phosphate forms of side chains in the preceding chapter.

Then, the essential logic of *reversible* posttranslational protein phosphorylation of these three residues, sketched in Figure 15.4, was to evolve companion P-Ser, P-Thr, and P-Tyr phosphoprotein phosphatases. Regulation, separate and complementary, of the protein kinases and the phosphoprotein phosphatases allows for maximal flexibility in balancing the

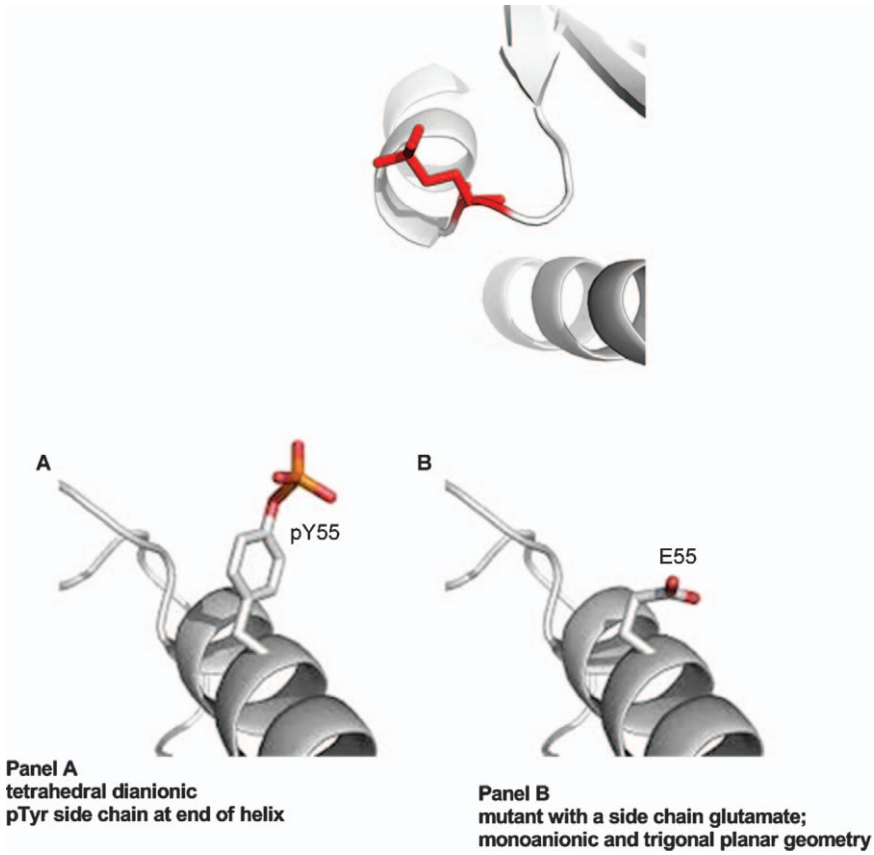


Figure 15.3 Emphasis on the newly introduced tetrahedral geometry on side chain phosphorylation. (Top) phosphoserine residue with the tetrahedral phosphate group in red. (Bottom) Panel A illustrates a tetrahedral P-Tyr side chain at the end of a helix and panel B shows its mutant variant with a planar glutamate anionic side chain.[8]
Reproduced from ref. 8, https://doi.org/10.18632/oncotarget.22829, under the terms of a CC BY 3.0 license, https://creativecommons.org/licenses/by/3.0/.

mole fraction of target proteins bearing the (up to 270 000) P-Ser, P-Thr, or P-Tyr groups. We will note separate catalytic strategies for the P-Ser and P-Thr *vs.* P-Tyr phosphatases below.

15.4 The Human Kinome

Although there are a few cases of bacterial protein kinase–phosphatase pairs that fall in the P-Ser–P-Thr superfamily, they are in minor abundance compared with the dozens of phospho-His sensor kinase components of the two-component sensor–response regulator pairs. In contrast, the P-Ser–P-Thr protein kinase family exploded in eukaryote evolution and

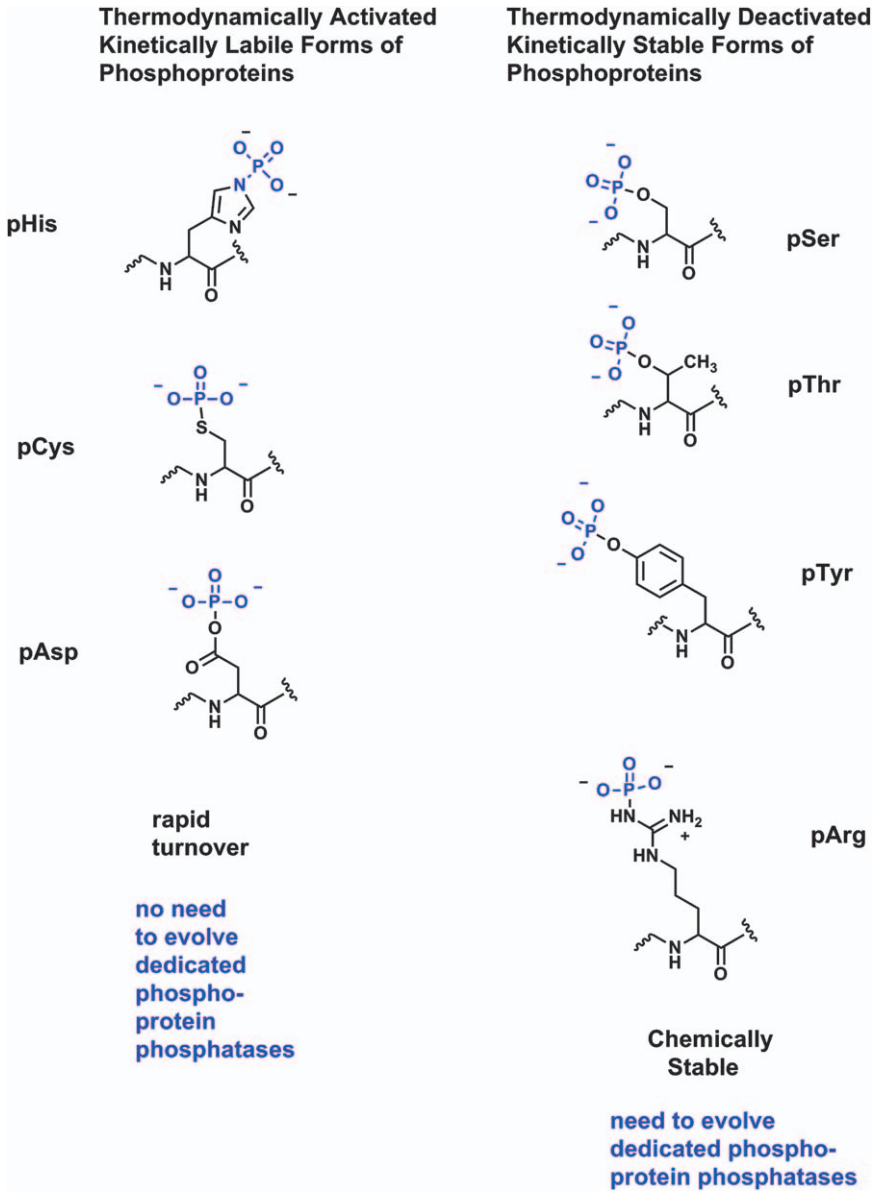


Figure 15.4 Division of phosphoproteins according to kinetic stability and the consequent need (or not) for evolution of paired classes of chemo-specific protein phosphatases to reverse the phosphorylation.

reaches into the hundreds in the human kinome.[9,10] The now classic representation of kinomes as unrooted trees of super families,[11] for both P-Ser–P-Thr and also the P-Tyr families is shown in Figures 15.5–15.7.

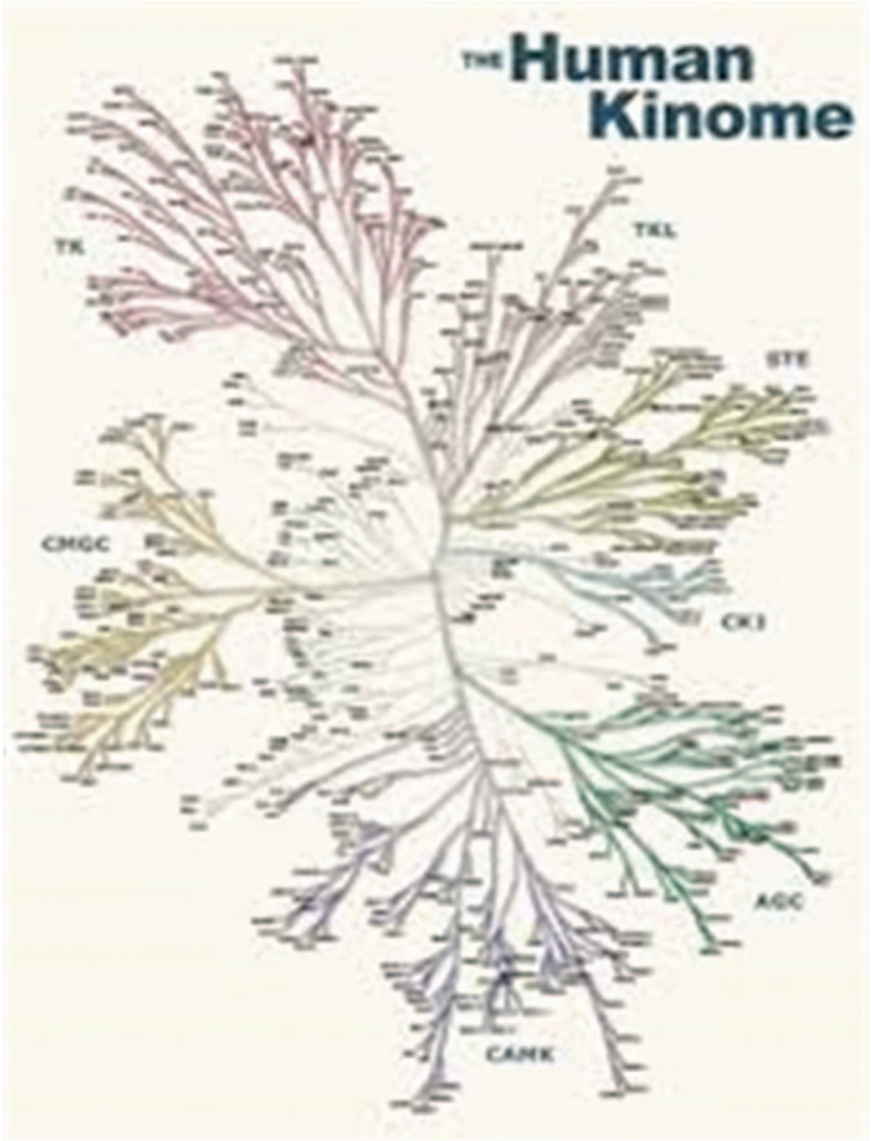


Figure 15.5 Collection of all known or predicted human protein kinases as the human kinome. The iconic rendering of the kinome as an unrooted tree from Manning *et al.*, with eight subfamilies.[11] (This includes all P-Ser and P-Thr and P-Tyr kinases but not the noncanonical members of Chapter 14).
Adapted from ref. 11 with permission from AAAS, Copyright 2002.

The human kinome tree now contains >500 predicted members, the great bulk of which have been assayed for kinase activity and evaluated for selectivity against banks of peptide and protein substrates.

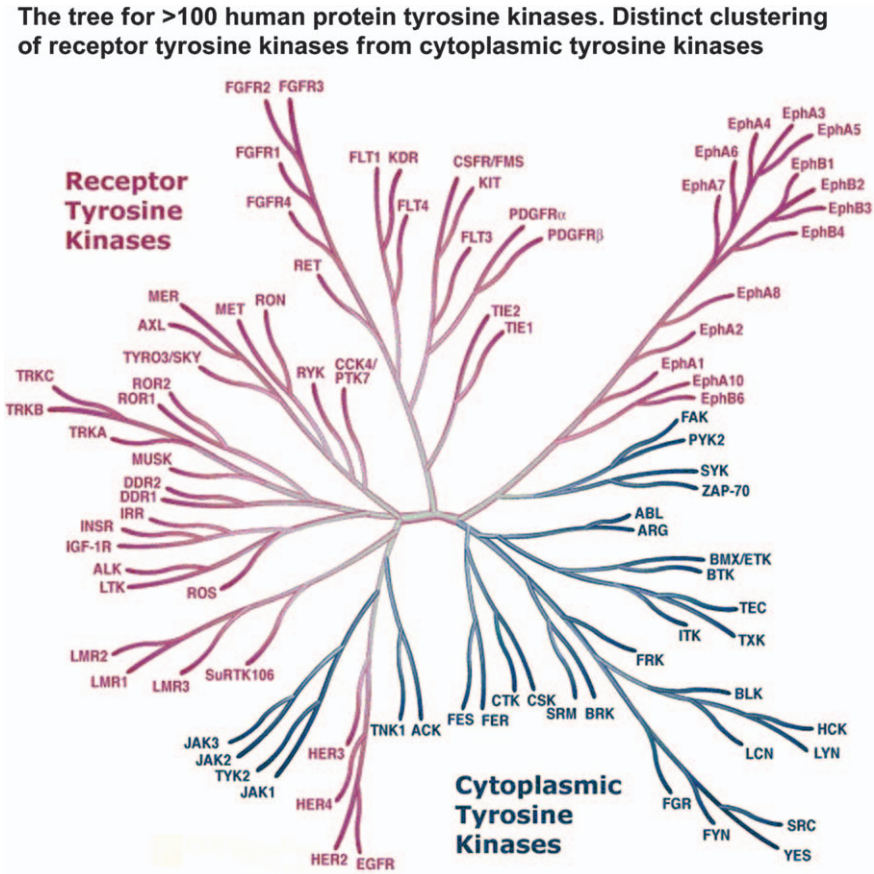


Figure 15.6 Collection of all known or predicted human protein kinases as the human kinome. A sub-tree shows the relationship of the ~100 human protein tyrosine kinases to each other.[12]

One of the sources for the evolution of Ser–Thr protein kinases may have been the metabolic kinases that act on low molecular weight primary metabolites.[14] Contemporary metabolic kinases involved in energy metabolism and biosynthetic and catabolic pathways that can act as protein kinases include hexokinase, the M2 isozyme of pyruvate kinase, an isozyme of phosphoglycerate kinase, and isoforms of nucleoside diphosphate kinases.[14]

As noted above, P-Ser residues are about ten times more abundant than P-Thr residues, reflecting perhaps a mix of accessibility of Ser over Thr side chains and the steric blocking influence of the C3 methyl in threonyl side chains. There are about 100 protein tyrosine kinases in the human kinome (Figure 15.8) with a small subset of those kinases also showing dual specificity for phosphorylation of Tyr and Ser side chains.[15]

Even though tyrosine protein kinases make up 18–20% of the full kinome, the abundance of P-Tyr groups is only 0.05% to ~1% of the abundance of

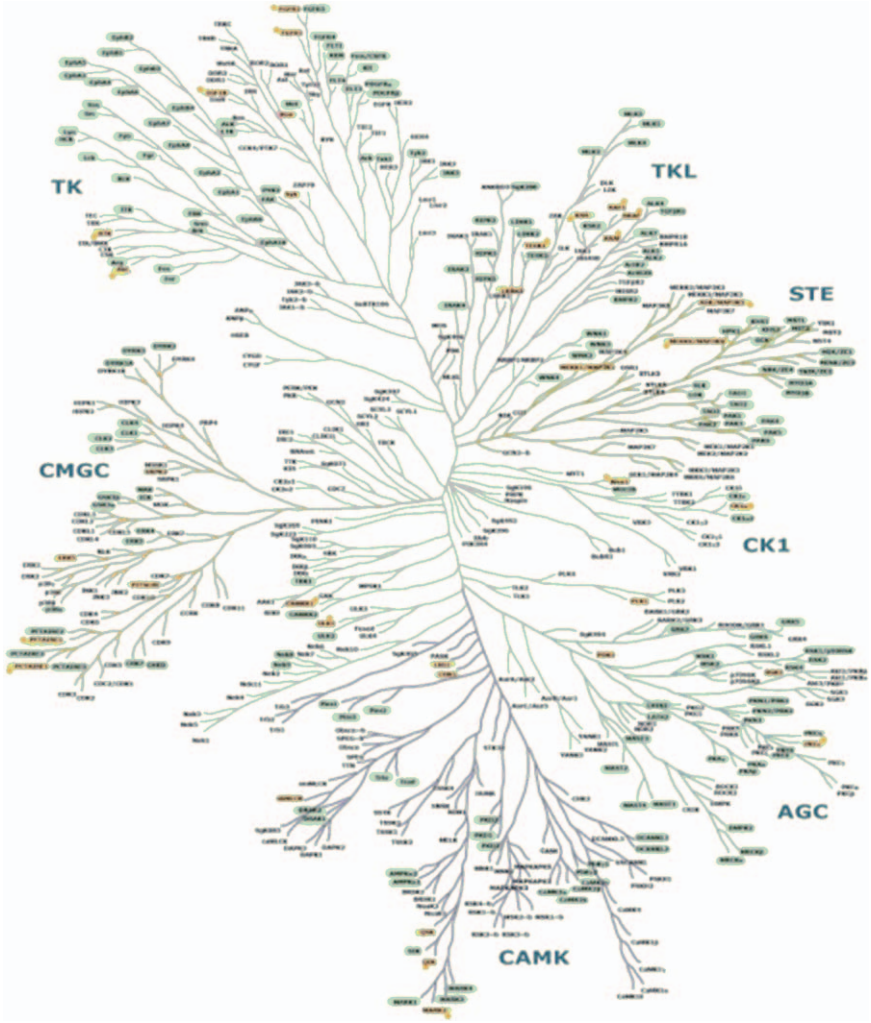


Figure 15.7 Collection of all known or predicted human protein kinases as the human kinome. P-Ser–P-Thr- binding domains include 14-3-3 proteins. The kinases that form the 14-3-3 interactome are shown in green.[13]

the P-Ser and P-Thr combination of phosphoproteins in humans. This lowered ratio may indicate a combination of lessened promiscuity and lower turnover numbers and reduced mole fraction of activated P-Tyr kinases in cells compared with the dominant P-Ser and P-Thr kinase flux.

Figure 15.8 indicates that hundreds of proteins are multiply phosphorylated[16] (at up to 15–20 distinct residues). The protein kinases (mostly) act posttranslationally, on proteins that have achieved their native folds in cells and so the extant phosphoproteomes represent accessible side chain phosphorylation sites in target proteins. These sites of side chain

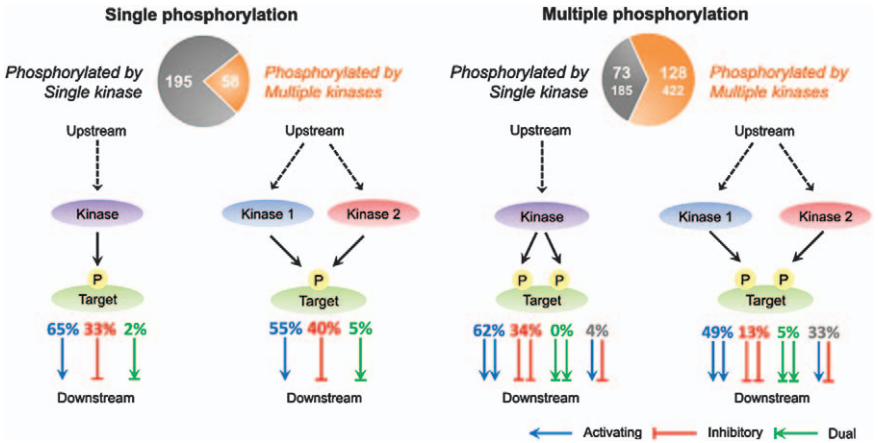


Figure 15.8 Interplay of multiple protein phosphorylations. Proteins can be targeted by a single specific kinase or by multiple kinases acting on the same residue of that target protein.[17] Alternatively, proteins at nodal points of metabolic and signaling control may be phosphorylated on multiple side chains, either by a single kinase (*e.g.* by protein kinase A) or by multiple protein kinases. Downstream fates of those modifications can be inhibitory, activating, or a mixture of both, depending on downstream targets.
Reproduced from ref. 17 with permission from Elsevier, Copyright 2015.

phosphorylations can reflect not only accessible surface residues, but also side chains in flexible loops, or in equilibrating conformational states. A number of kinases act only after a preceding kinase has put in a phosphorylation mark, presumably reflecting one or more conformational dynamic changes arising from the action of the first kinase to make a section of the phosphorylated protein target now available for the second kinase (Figure 15.8).

The Abelson tyrosine-protein kinase 1 (Abl) protein has 11 documented sites of phosphorylation by other kinases, undoubtedly modulating activity across its own catalytic and multiple regulatory domains[18] (Figure 15.9). Nine residues that get phosphorylated in Abl or its activated breakpoint cluster region–Abl (Bcr–Abl) fusion oncogenic protein form are tyrosines, one is a serine, one is a threonine residue, indicating multiple kinase inputs.

The catalytic subunit of protein kinase A has 10 reported autophosphorylation sites. Phosphopeptide mapping of the ~500 residue protein paxillin, implicated in cell–cell junctions, indicated five sites of tyrosine phosphorylation and up to 24 sites of serine phosphorylation:[19] an enormous potential diversification of protein structural and functional inventory.

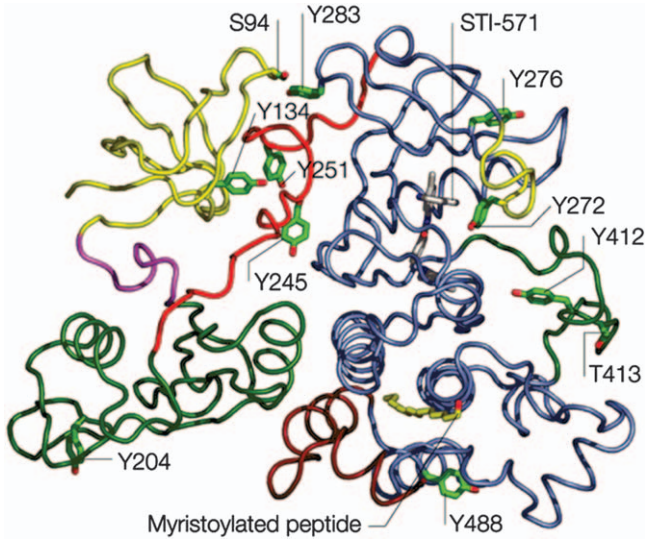


Figure 15.9 Eleven phosphorylation sites in the Abl tyrosine kinase: nine tyrosines, one serine, one threonine.[18]
Reproduced from ref. 18 with permission from Springer Nature, Copyright 2004.

15.5 "Cascades" of Protein Kinase Activation from Outside In: Cell Surface to Cytoplasm and Nucleus

15.5.1 Serine–Threonine Protein Kinases

Several examples of the consecutive action of sets of kinases that become activated by phosphorylation and then turn on a downstream kinase by its phosphorylation are known and support the view of linked temporal and spatial cascades of information transfer by protein phosphorylations. Some are summarized in cartoon fashion in Figure 15.10. At the heart of this strategy is keeping the basal state of a protein kinase at a low level, an almost "idle" state.[20] When a signal is sensed, such as a so called "first messenger" growth factor or hormone reaching the external domain of a membrane receptor (Figures 15.11–15.14), transduction across the membrane results in autophosphorylation and activation of the first-level protein kinase (*e.g.* at or near the cytoplasmic face of the cell membrane). For example, in a three kinase amplification cascade[21] the mitogen activated protein (MAP) kinase kinase kinase (MKKK) represents that first level. Those MKKK enzyme molecules activate an MAP kinase kinase (MKK) partner by phosphorylation. In turn an MKK enzyme activates MAP kinase. Now, sets of protein substrates get phosphorylated on serine or threonine by the active form of MAP kinase.

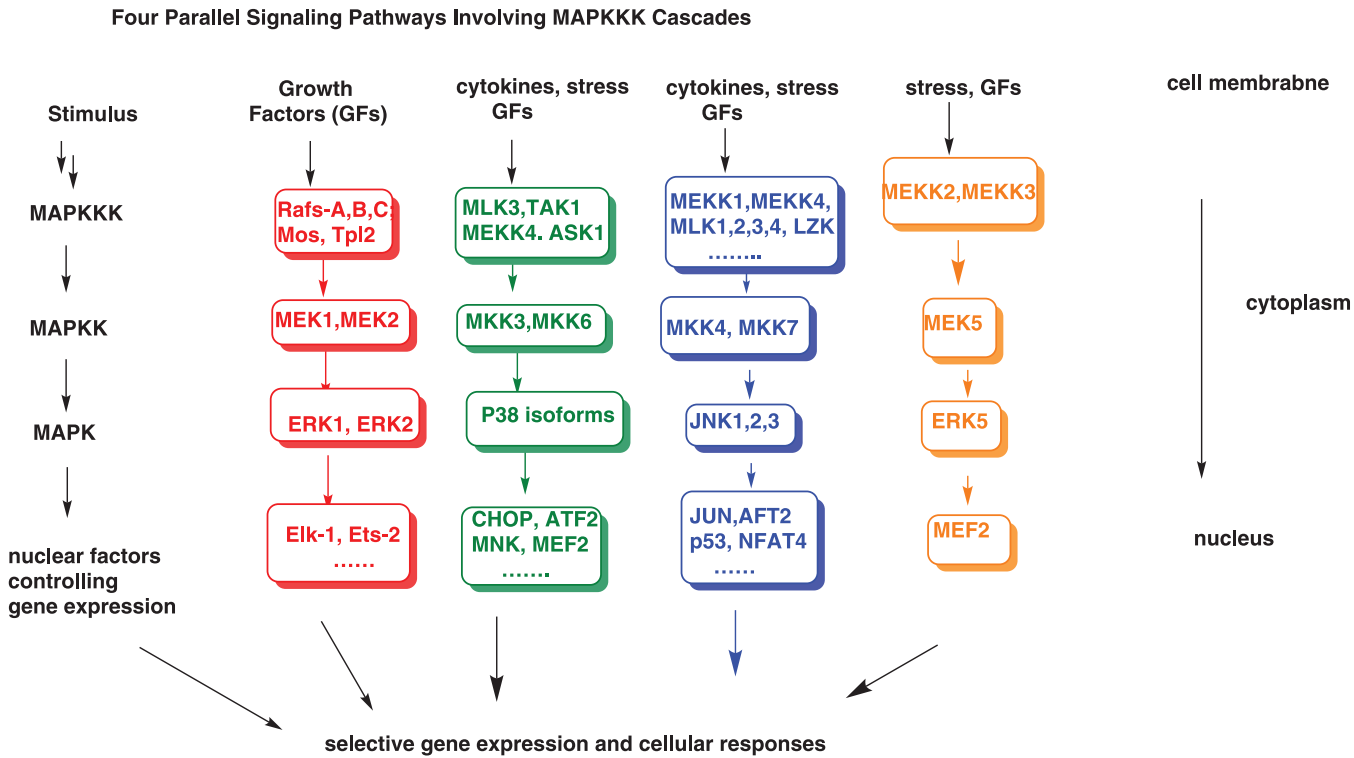


Figure 15.10 Two representations of protein kinase signaling cascades. The left-hand panel shows five parallel cascades of protein kinases carrying information from cell surface to interior and often to nuclear transcription factors for control of for gene expression.[22] The right-hand panel indicates that each level of information reaches further into cytoplasmic pathways and integration of signals.[23]
Adapted from ref. 22 with permission from Springer Nature, Copyright 2007.

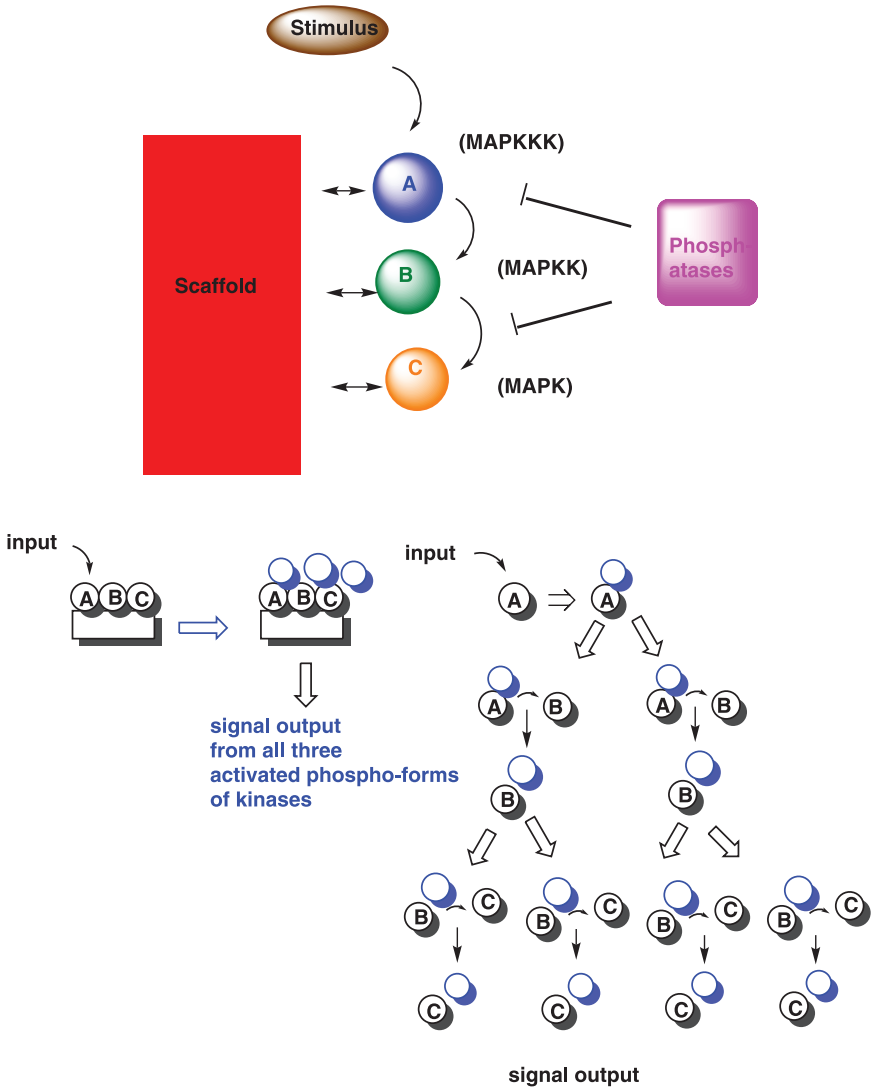


Figure 15.11 Proteins with docking scaffolds collect multiple protein kinases that act tandemly in a pathway and increase local concentrations and efficiency of signal transmission. The troika of MAPKKK, MAPKK, and MAPK on a common scaffold.[29]
Adapted from ref. 29 with permission from PNAS, Copyright (2007) National Academy of Sciences, U.S.A.

In terms of signal amplification, if one MKK activates ten MKKs and ten MKKs activate each of ten MKs, and so on there would rapidly be a thousand activated MAPK molecules. If they each phosphorylate a hundred protein targets, the net amplification of initial signal is 100 000 fold. The gain could easily be higher given that catalysts are involved in all four steps, but there

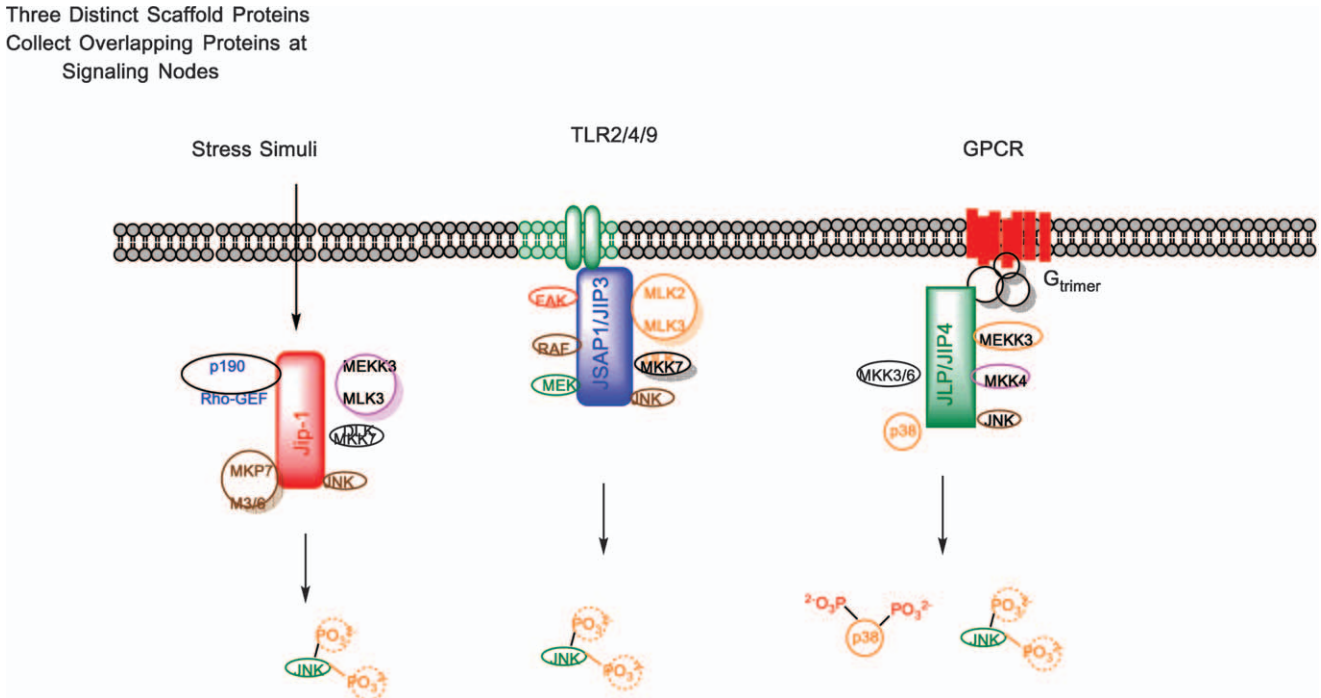


Figure 15.12 Proteins with docking scaffolds collect multiple protein kinases that act tandemly in a pathway and increase local concentrations and efficiency of signal transmission. A more complete and complex view of the role of different JIP scaffold isoforms in directing multiple different phosphorylations of downstream target proteins.[30]
Reproduced from ref. 30 with permission from Springer Nature, Copyright 2007.

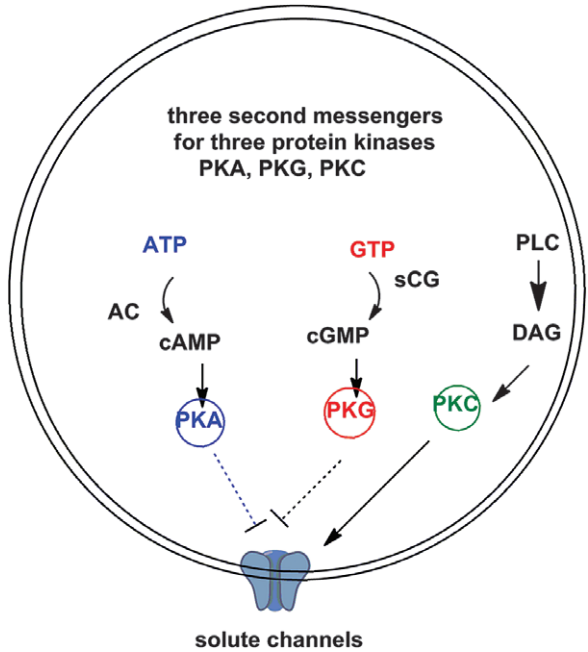


Figure 15.13 Schematic for cellular roles of protein kinases A, G, and C with physiological activators (cAMP, cGMP, and phospholipase C), with specific synthetic inhibitors.[34,35]

may not be enough protein target molecules in a given volume of the cell to get the full burst of amplification.[24] (One nanomolar concentration of a protein component may equate to $\sim 10^3$ molecules in an epithelial cell.[25])

There are parallel and converging activation cascades for sets of protein kinases to integrate multiple signal inputs. One physical node for such integrations are scaffolding proteins with ''docking stations'' (high affinity binding sites for particular protein partners; Figures 15.11 and 15.12) that increase the local concentration of several protein substrates (and thereby the kinetics of successive interprotein phosphorylations) that may be integrative in terms of phosphorylation routes.[26,27] Figure 15.11 schematizes the collection of the three tandem P-Ser–P-Thr kinases MAPKKK, MAPKK, MAPK and the potential advantage in catalytic efficiency from collection on a common scaffold protein. Figure 15.12 depicts a family of c-Jun N-terminal kinase-interacting protein (JIP) scaffolding proteins connected to different inputs at the cell membrane with partially overlapping sets of kinases. For example, the c-Jun NH_2-terminal kinase (JNK)–stress-activated protein kinase-associated protein 1 (JSAP1) aka JIP3 scaffold protein can interact with up to eight distinct protein Ser–Thr kinases to integrate signals at this intermediate signaling node.[28]

The degree of activation of any protein kinase in a cell or tissue could vary from inactive, to a population where some are active and others remain inactive, to a population where all members are active. Thus, there will be a

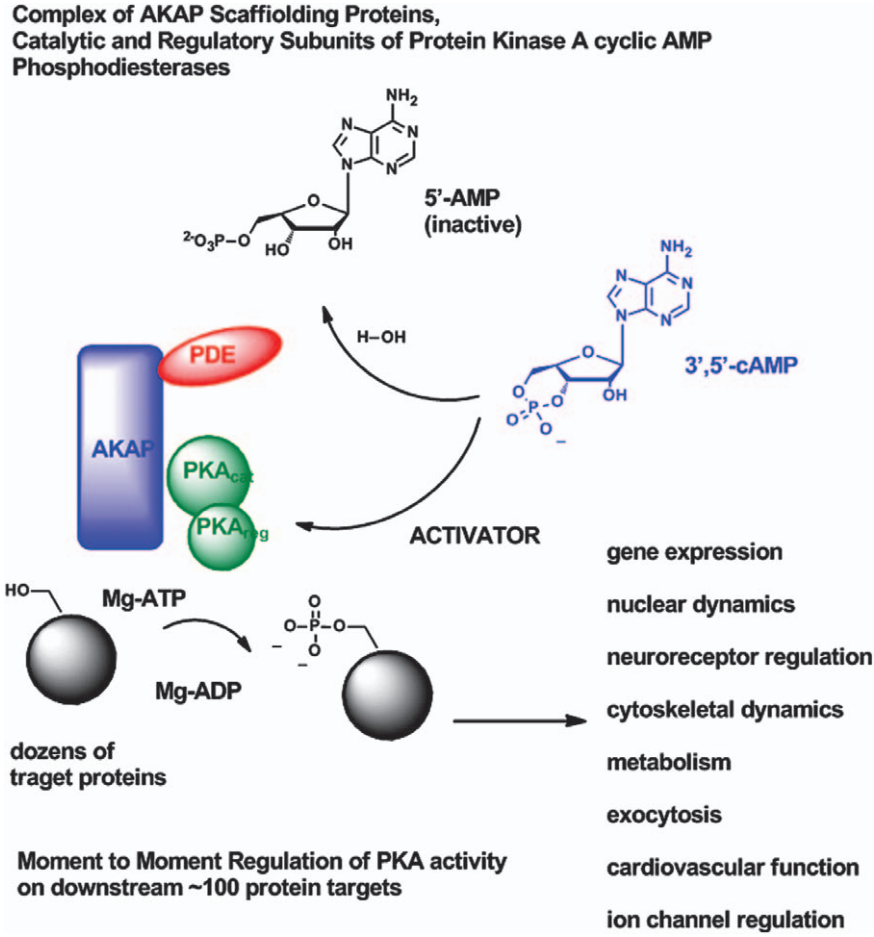


Figure 15.14 The PKA complexes have anchoring scaffold proteins [a kinase anchoring proteins (AKAPs)] and phosphatases to balance the net kinase activity with its plethora of downstream effects.[36]

spectrum of both kinase activities and their target phosphoproteomes in time and space that will show fractional stoichiometries of phosphorylations (*e.g.* 25 molecules out of 100 of a particular target protein may be phosphorylated and 75 unphosphorylated. Also, a target protein or intermediate protein kinase may, for example, have up to four phosphorylations at maximum and at any time present a mixture of mono- to tetra-phosphorylated forms).

As a final point in this short summary of many thousands of different microstates of the canonical phosphoproteomes, different protein kinases show distinct degrees of promiscuity. Some, such as cyclic AMP-activated protein kinase A (PKA), have multiple target proteins: over 100 for PKA. Others, even when fully activated are highly specific. The protein kinase that phosphorylates pyruvate dehydrogenase acts only on that target protein and no other. (Once phosphorylated the

mitochondrial dehydrogenase activity drops off dramatically. As this is a key conduit for acetyl CoA generation for oxidative phosphorylation, this regulatory phosphorylation is highly consequential in switching cells from oxidative phosphorylation for 34–38 ATP or glucose molecule sent down glycolysis to the net two ATP or glucose when pyruvate dehydrogenase is blocked).

Early on in the characterization of the activity and regulatory profiles of the initially discovered protein kinases some were classified according to differing low molecular weight activators. These include the cyclic AMP-activated protein kinases, designated protein kinases A,[31] the corresponding cyclic GMP-activated PKs (PKG),[32] and the lipid-activated multiple isoforms of protein kinase C.[33] These well studied examples also represent nodal points of intersection of second messenger logic and machinery, integrating the two signaling universes: low molecular weight second messengers and phosphoproteomics. They are also among the best studied mechanisms by which low molecular weight second messengers (*e.g.* cAMP, and cGMP) achieve multiplicative to exponential amplification of signals, from stoichiometric to catalytic (Figures 15.13 and 15.14).

The 400 or so serine–threonine protein kinases in humans all have a common catalytic core for binding ATP and an adjacent region for the serine or threonine side chains in a protein substrate to act as catalytic nucleophile towards Pγ of Mg–ATP. An amino terminal (N) and carboxy terminal (C) lobe together make up the ATP binding site of this Ser–Thr kinase family[37] as depicted in Figure 15.15A. A view of the ATP binding site in Figure 15.15B indicates placement of donor ATP substrate.[38] The flexible glycine-rich G loop that anchors ATP in the active site is highlighted.

15.5.2 Tyrosine Kinases: Membrane Receptor *vs.* Soluble Forms

There are more than 90 tyrosine kinases in human proteomes. They all operate by the same chemical mechanism of direct $-PO_3^{2-}$ group transfer from Pγ of Mg–ATP to a nascent phenolate oxygen of one or more tyrosine residues[40] (Figure 15.1). Phosphoryl group transfers occur both in autophosphorylations and *via* external phosphorylations of substrate proteins once the P-Tyr kinases have become autoactivated. However, the 95 members of that tyrosine kinome split into two different subgroups based on subcellular location and consequent domain architecture (Figure 15.8).

There are 58 human examples of receptor tyrosine kinases and 32 members of the non-receptor, cytoplasmic tyrosine kinase subfamily[41] Figure 15.16. schematizes two membrane-initiated signaling pathways: the MAP kinase cascade and the Janus kinase–signal transducer and activator of transcription (JAK–STAT) pathway. Autophosphorylating transmembrane kinases are initiators in each pathway, representing the large class of receptor tyrosine kinases. The MAP kinase cascade entails a tandemly acting set of nonreceptor Ser–Thr kinases. The JAK–STAT cascade involves two variants of the nonreceptor tyrosine kinase class.

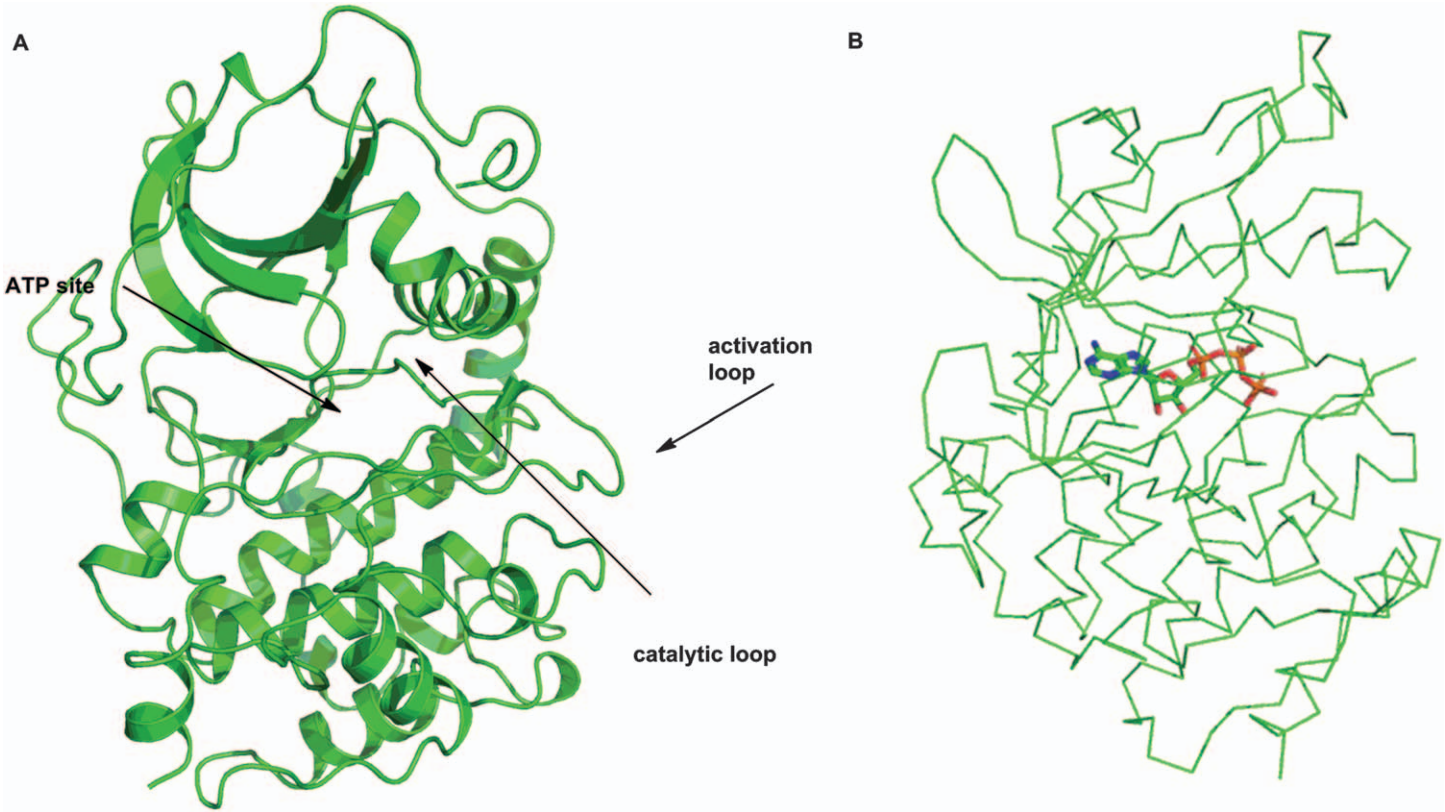


Figure 15.15 Two views of a serine–threonine kinase family member folds. (A) Overall fold of a kinase catalytic subunit or domain, drawn from pdb 2CPK.[39] (B) Orientation of ATP in PKA active site, drawn from pdb file 4WB5.[38]

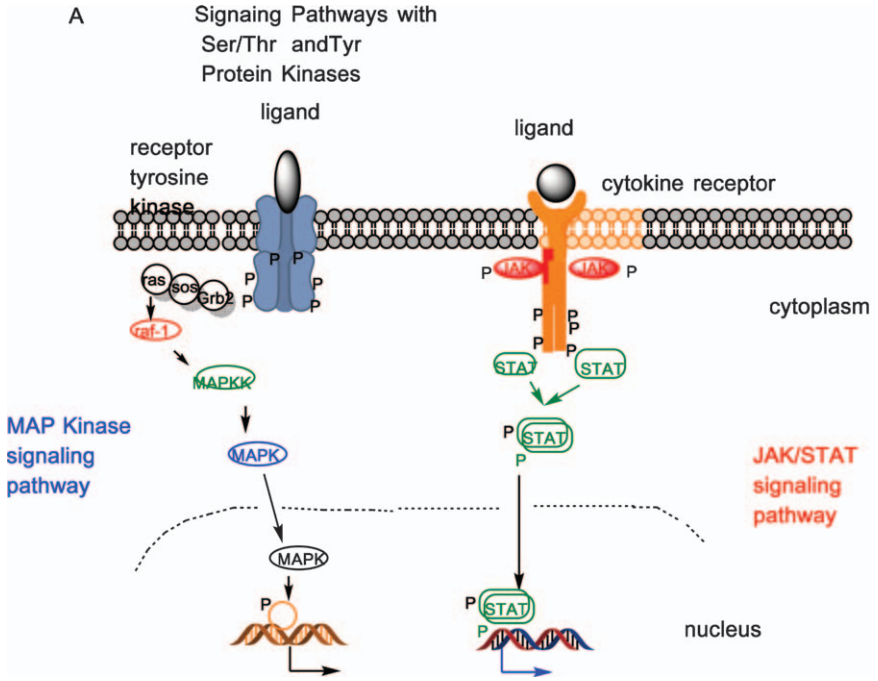


Figure 15.16 Two surface-initiated transmembrane signaling routes involving serine–threonine kinases and tyrosine kinases. Two receptor tyrosine kinases auto-phosphorylate themselves on multiple tyrosine side-chains in or near the kinase domains. For the left-hand receptor the subsequent signaling is *via* the MAPKKK–MAPKK–MAPK serine–threonine protein kinase cascade. For the right-hand pathway, the cytoplasmic onward signal transmission is instead mediated by cytoplasmic JAK and STAT tyrosine kinases.

In aggregate, the tyrosine kinases affect or control aspects of growth, differentiation, cell–cell adhesion, cell motility, and apoptotic cell death. Members of the receptor tyrosine kinases include receptors for epidermal growth factor (EGF), platelet derived growth factor (PDGF), insulin, fibroblast growth factor (FGF), and many other protein growth factors. Figure 15.17 notes seven subfamilies of receptor protein kinases based on additional domains both external and internal to the plasma membrane. Thus, the insulin receptor is a disulfide linked dimer, while the PDGFR, FGFR, and vascular endothelial growth factor receptor (VEGFR) have iterated and/or split catalytic domains. Several of these subfamilies have external immunoglobulin-like domains as part of their receptor or sensor architectures for specific extracellular ligands.

In the cytoplasmic group of tyrosine kinases the Rous sarcoma virus (v-src) oncogenic protein is the founding member but it also includes fyn and

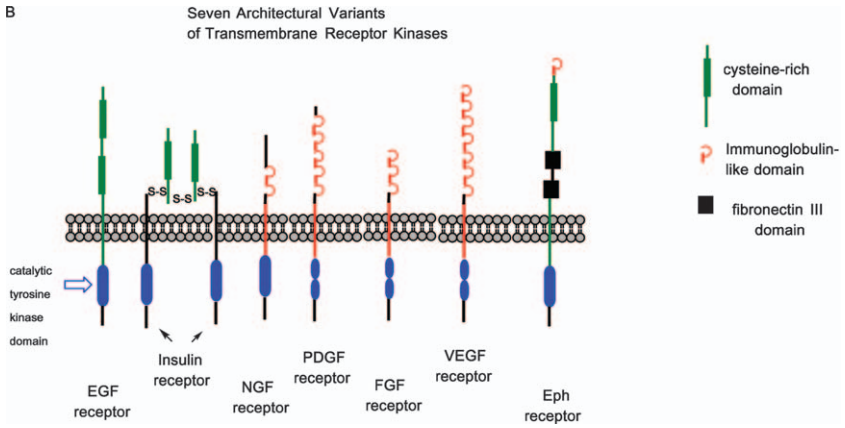


Figure 15.17 Two surface-initiated transmembrane signaling routes involving serine–threonine kinases and tyrosine kinases. Seven subfamilies of receptor tyrosine kinases are listed with variations in types of extramembrane recognition folds and split tyrosine kinase domains.[42]

lymphocyte-specific protein tyrosine kinase (lck) enzymes downstream of T cell receptor signaling, zeta-chain-associated protein kinase 70 (zap70), and isozymes 1–3 of the JAK family of signal transduction cytoplasmic tyrosine kinases.[18]

These joint receptor–enzyme kinases are typically termed receptors, such as the EGF receptor (EGFR), PDGF receptor (PDGFR), fibroblast growth factor receptor (FGFR), VEGFR, and so on.

Receptor tyrosine kinases have at least three domains. The first, external to the plasma membrane, is the receptor domain. Occupancy by an external ligand (*e.g.* insulin for the insulin receptor) is transmitted through the second, transmembrane helical domain to the (third) cytoplasmic catalytic domain, activating it for phosphoryl transfer. Often, the ligand binding to the external receptor domain induces dimerization as sketched for the EGFR[43] in Figure 15.18A. Then, the activated cytoplasmic domains *autophosphorylate* each other, typically *in trans* on the adjacent subunit.

As depicted in the cartoon of Figure 15.18 up to five tyrosyl residues in the cytoplasmic domain adjacent to the kinase domain can be phosphorylated. Five per chain times two is a total of up to ten newly introduced dianionic tetrahedral phosphate groups. These not only perturb the cytoplasmic domain conformations, but also recruit partner proteins with their own domains that recognize the newly introduced P-Tyr side chains.

As shown in Figure 15.18B for one activated subunit of the PDGFR kinase, eight tyrosine side chains in the cytoplasmic domain become autophosphorylated (*in trans*). These can bind adaptor proteins [such as growth factor receptor-bound protein 2 (Grb2)] that in turn recruit additional protein components. Sometimes they are enzymes, including the src

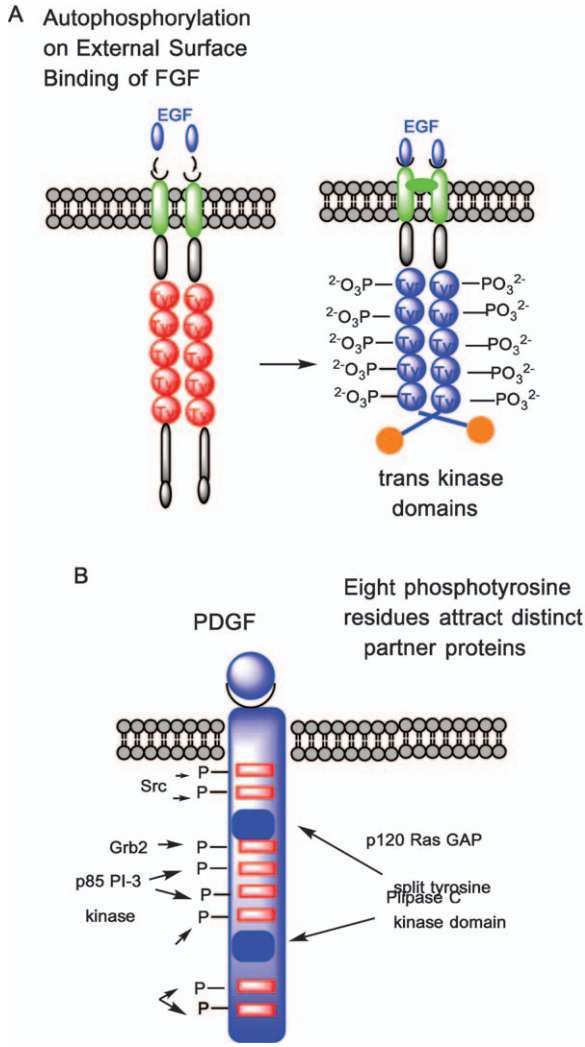


Figure 15.18 Multiple autophosphorylations by the cytoplasmic tyrosine kinase domain is the rule for activated receptor tyrosine kinases. (A) Ligation of the epidermal growth factor receptor by its soluble, circulating EGF protein drives dimerization, transmembrane activation of the cytoplasmic tyrosine kinase domains. Those six tyrosine phosphate side chains then recruit partner proteins with P-Tyr-recognition domains.[44] (B) Analogously, binding of the platelet-derived growth factor (PDGF) to its extracellular receptor domain activates the kinase domain to phosphorylate up to eight tyrosine side chains. Those eight P-Tyr tetrahedral dianions recruit the indicated proteins to the specific P-Tyr sequences, including the src cytoplasmic tyrosine kinase, the adaptor protein Grb2, the GTPase accelerating protein (GAP) for the p120 form of RAS, plus the two indicated phospholipid-metabolizing enzymes phosphoinositol-3 (PI-3)-kinase and phospholipase C.[45]

cytoplasmic tyrosine kinase and the p85 subunit of the lipid phosphatidylinositol-3-kinase that amplify phosphorylation cascades for both proteins and phospholipids. Also recruited to distinct stretches of P-Tyr residues are the p120 subunit of the rat sarcoma (Ras) GTPase accelerating protein (GAP), increasing the turnover and deactivation rate of Ras–GTP to Ras–GDP and the phospholipase C-zeta subform. We have noted in Chapter 8 that PLC isoforms hydrolyze PIP-2 to inositol 1,4,5-triphosphate (IP_3) and diacylgycerols, two distinct kinds of low molecular weight second messengers.

Clearly, the multiple tyrosine side chain autophosphorylations of receptor tyrosine kinases have explosive amplification consequences for signals that reach into distinct corners of cell metabolism.

As shown in the left-hand side of Figure 15.16, the multiply autophosphorylated cytoplasmic domain of a growth factor receptor can switch on the MAPKK cascade, ending in selective nuclear gene transcription. The right-hand portion of that same figure shows the role for the Janus family of kinases (JAKs) that analogously cause multiple phosphorylations of the cytoplasmic tail of a cytokine receptor to turn on a gene activation pathway. In the left-hand pathway, the phosphoproteomic logic switches from tyrosine phosphorylation at the receptor tyrosine kinase first step to the Ser–Thr kinase tandem relay of MAPKK–MAPK further downstream, showing how tyrosine kinase and serine–threonine kinase machinery mesh and integrate.

15.6 Phosphoserine and Phosphotyrosine Binding Domains

Tony Pawson and colleagues over the years have collected information on a variety of protein domains that show affinity for the phosphorylated side chains of serine or threonine in proteins.[46] These include as many as 350 proteins that have bioinfomatically-predicted domains that have the acronyms: 14-3-3, polo box, FHA, FF, BRCT, WW, WD40 and MH2 subfamilies. Only a subset in any one of these eight families have been characterized for heightened affinity to P-Ser and P-Thr side chains so these may be overestimates of *in vivo* function.[47] Nonetheless, they do indicate the likelihood of complexation with target proteins that have been phosphorylated on one or more serine or threonine residue by one of the ~420 human serine–threonine protein kinases. The different domains may have evolved *via* independent and convergent evolution to mediate the interactomes for protein-based phosphoryl group signaling.

The signal transduction logic noted above for receptor tyrosine kinases requires three aspects of protein chemistry: (1) ligand-induced dimerization; (2) *in trans* autophosphorylation of the dimeric subunits by the catalytic domain; (3) recruitment by the phosphotyrosine side chains of partner proteins to mediate the cytoplasmic and then nuclear steps of the information cascade.

Two widely distributed phosphotyrosine residue binding motifs have evolved to mediate point 3 above. They are termed src homology 2 (SH2) and phosphotyrosine binding (PTB) domains and have independently evolved folds. SH2 was first detected in src family proteins.[46,48]

15.7 Visualization of Binding Pockets for P-Thr, P-Ser, and P-Tyr in Protein Partner Domains

Figure 15.19A shows the local structure of a polo box domain of polo like kinase 1 (PLK1) P-threonine binding domain with P-threonine as ligand in both ribbon diagram and surface representation. Analogously Figure 15.19B depicts two views of how P-serine interacts with the 14-3-3 domain of P-Ser binding protein. Finally, Figure 15.19C shows how P-tyrosine fits into a groove of the SH2 domain of the src-family tyrosine kinase lck.

15.8 FDA-approved Protein Kinase Inhibitors

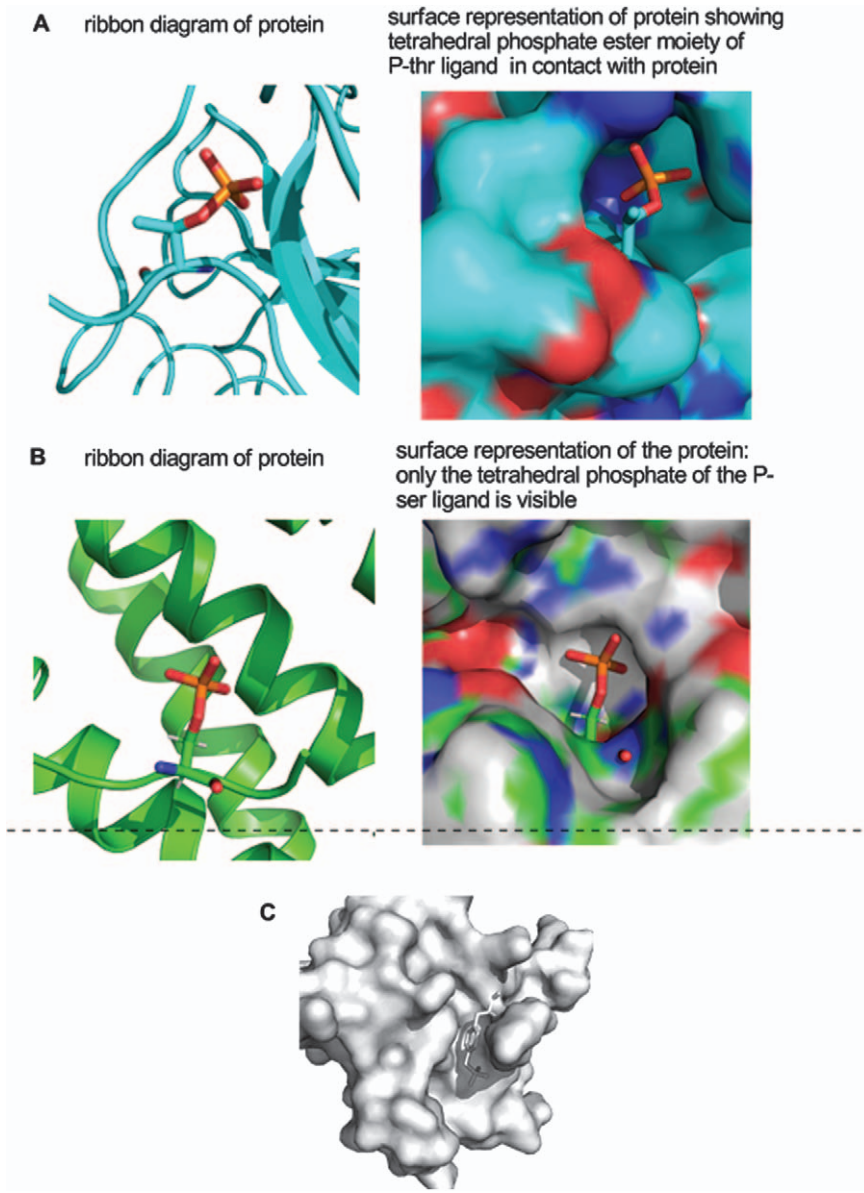
Thus far in this volume we have not taken up the topic of inhibitors of various phosphorus-utilizing enzymes. It could easily be a different book that delves into the molecular pharmacology of many types of phosphorus-recognizing enzymes. These range from DNA ligases, DNA and RNA polymerases, particularly of viral origin, to many structural scaffolds that inhibit one or more of F-type, V-type, and P-type ATPases. We make a brief exception for inhibitors of protein kinases because more than a dozen have been approved over the past three decades by the United States Food and Drug Administration (US FDA) as human medicines.

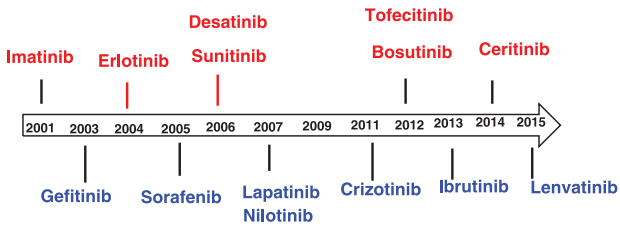
The pharmaceutical and biotechnology industries have focused on members of the human kinome as perhaps the most intensively investigated set of proteins for pharmacological intervention.[49] Excess protein kinase activity has proven to be a nodal point in many human cancers. Figures 15.20 and 15.21 list a dozen or so selective to semiselective inhibitors with trade names, kinase targets and human cancer type for which they are prescribed.

Originally, the medicinal chemistry goals were to achieve absolute monoselectivity for a particular kinase where there was evidence that either excess levels of the native kinase or excess activity due to a mutated version was a driver mutation in a particular set of cancer patients. However, the close homology in ATP binding sites among neighbors on the kinome tree (Figures 15.5–15.7) made that impractical in many cases as noted in Figures 15.20 and15.21. As it happens, sometimes inhibiting closely related kinases may have therapeutic benefit. Figures 15.20 and 15.21 indicate that imatinib and desatinib are relatively selective for the abl tyrosine kinase to treat chronic myeloid leukemia (CML) while gefitanib is selective for the human epidermal growth factor receptor (HER) 1 tyrosine kinase isoform in non-small cell lung cancer. Lapatanib targets both HER 1 and 2 isozymes and is used for breast cancer chemotherapy. On the other hand, serafinib

inhibits three related kinases. Figure 15.21 shows X-ray snapshots of some of these FDA-approved kinase inhibitors on pharmacologically relevant protein kinase targets.

While many of the cancer drug targets are in the receptor tyrosine kinase family, Raf murine sarcoma viral oncogene homolog B (BRAF) is a serine–threonine protein kinase. Most of the first- and even second-generation clinical inhibitors have been targeted to the ATP binding site as competitive





Kinase Inhibitor	Trade Name	Target	Cancer Type
Imatinib	Gleevec	ABL, c-KIT, PDGFR	CML, GIST, ALL
Dasatinib	Sprycel	src family, ABL	CML, ALL
Gefitinib	Iressa	HER2	NSCLC
Erlotinib	Traceva	HER2	Lung
Lapatinib	Tykerb	HER2, EGFR	Breast
Crizotinib	Xalkori	EML4-ALK	NSCLC
Sorafenib	Nexavar	B-RAF, VEGFR, PDGFR	Renal, Liver

Figure 15.20 FDA-approved protein kinase inhibitors for human therapeutic use. Starting with Imatininb in 2001, some 29 kinase inhibitors had been approved by 2015.[50] A partial list is shown in the list below the time line. CML, chronic myeloid leukemia; GIST, gastrointestinal stromal tumor; ALL, acute lymphoblastic leukemia; NSCLC, non small cell lung carcinoma.

inhibitors. Among contemporary research programs are allosteric inhibitors, that should be free of upstream substrate proteins overcoming inhibition by mass action. There are also perhaps 40 of the >500 human protein kinases that have cysteine thiolate side chains in or near the active site that have the propensity to be targeted covalently and irreversibly by inhibitors bearing electrophilic groups such as conjugated enones.[52]

Figure 15.19 A snapshot of some of the protein domains that selectively recognize and bind to P-Ser, P-Thr, or P-Tyr domains of proteins and thereby enable signal propagation (A) P-Thr in the polo box domains (PBD) of Polo-like kinase-1: domain the left side is the ribbon diagram of the protein, on the right side is the surface representation. The surface view shows that only the beta methyl and the tetrahedral phosphate of the P-Thr ligand is visible at the surface; drawn from pdb file 1UMW. (B) P-Ser bound to the human 14-3-3 beta protein is shown on the left with the protein as a ribbon diagram, on the right as a surface depiction (drawn from pdb file 6G8K). (C) P-Tyr bound to the SH2 domain of the 56 kDa lck member of the src family of cytoplasmic tyrosine kinases. The protein is shown in space-filling, surface representation to show the shape of the p-Tyr binding site (drawn from pdb file 1CWD).

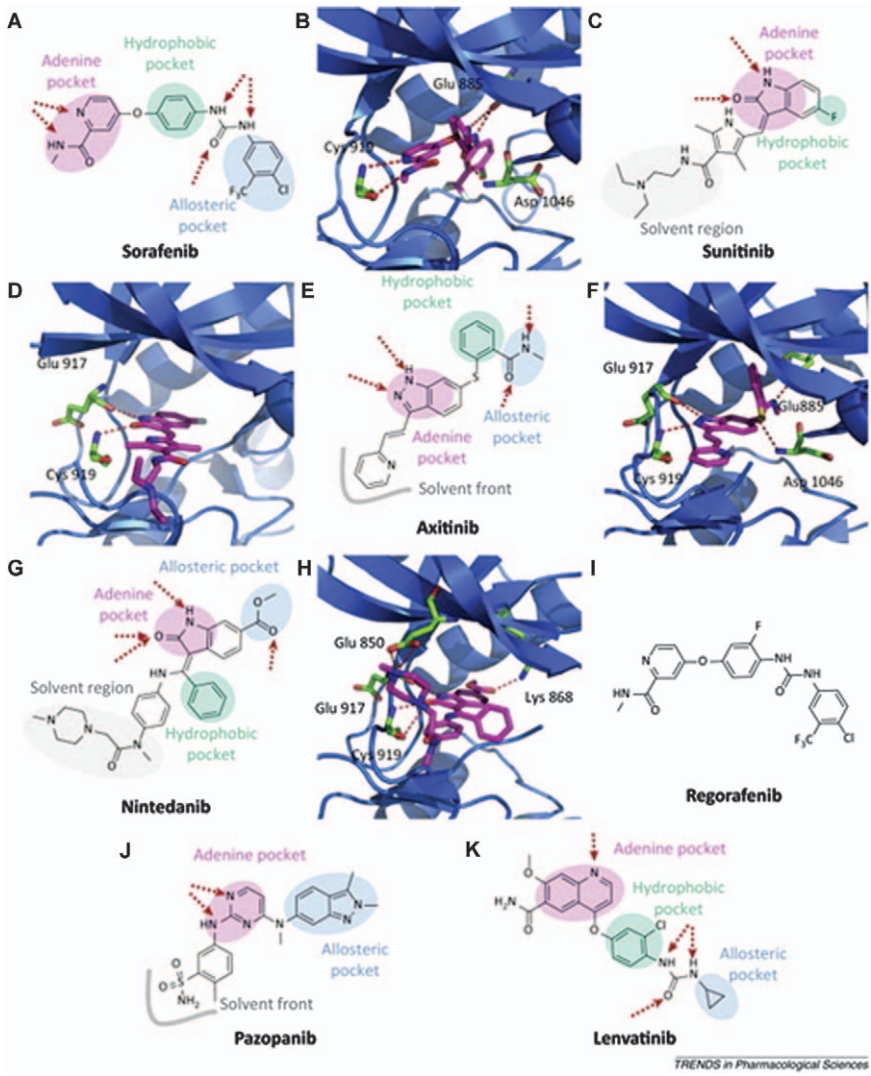


Figure 15.21 FDA-approved protein kinase inhibitors for human therapeutic use. Given the high homology of protein kinase family members, especially in the ATP binding site typically targeted by kinase inhibitors,[51] it is not surprising that the inhibitors may act on more than one kinase *in vivo*. The arrows show the key specificity points of contact of the specific drugs with their target kinases.

15.9 Phosphoprotein Phosphatases: Two Limiting Strategies for Specificity

The prediction of ~230 000 P-Ser, P-Thr, and P-Tyr residues distributed over ~13 000 human phosphoproteins noted earlier gives an upper limit to the

human phosphoproteomics inventory. The actual fraction forming at any given time in any cell depends, among other parameters, on two balancing factors. One is certainly the catalytic flux of the actual protein serine–threonine kinases and tyrosine kinases that are activated in that interval. The second is the corresponding inventory of phosphoprotein phosphatases, capable of acting on P-Ser, P-Thr, or P-Tyr residues in that cell in that time frame.

The need for a dedicated set of phosphoprotein phosphatases to have protein phosphorylation be reversible, and therefore qualify for the full range of signal dynamics (Figure 15.22) is mandated by the chemical

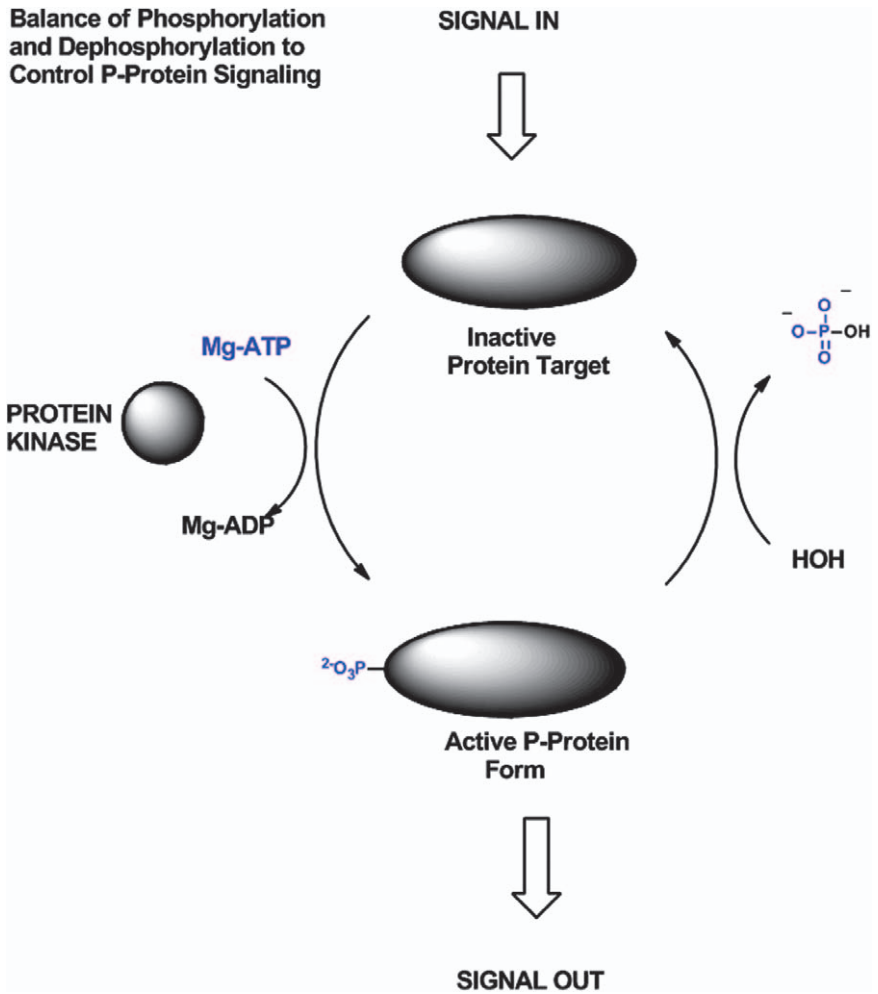


Figure 15.22 The logic for signal ''on'' and signal ''off'' by phosphorylation of Ser, Thr, or Tyr residues in proteins requires catalytic removal of these chemically stable phosphoryl groups. The ''off'' role is served by protein phosphatases.

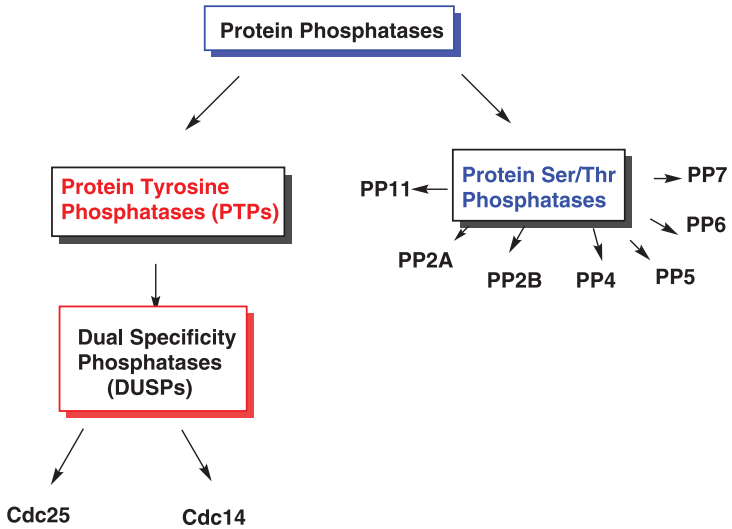


Figure 15.23 Categories of protein phosphatases: The two major chemoselective and mechanistically distinct classes of phosphatases are the protein tyrosine phosphatase (PTPs) and the Protein Ser–Thr phosphatases (PSPases).[53] The dual specificity group is a subclass architecturally and mechanistically of PTPases. Compared with the ~100 human PTPases, there is a much smaller number of PSPases to deal with the ~230 000 potential P-Ser and P-Thr sites in the proteome.

stability of the two aliphatic alcohol phosphomonoesters (P-Ser and P-Thr residues) and the equivalent stability of the phenolic ester (P-Tyr residues) (Figure 15.1). Otherwise the three phosphorylated alcoholic side chains would persist for the life of the cell, much like proteolytic cleavages.

There are dual specificity protein phosphatases (DUSPs; see later in this Chapter) that will transfer all three phosphoryl groups on Ser, Thr, and Tyr residues to water, but the majority of protein phosphatases are selective either for P-Ser and P-Thr side chains or P-Tyr side chains (Figure 15.23)

In principle, one might have expected three possible enzymatic strategies for such protein dephosphorylations based on acceptor precedents from primary metabolism. One route would be $-PO_3^{2-}$ transfer to Mg–ADP to form ATP (akin to an F-type ATP synthase). A second route would be to use inorganic phosphate, Pi, as an external substrate. The protein dephosphorylation would then produce inorganic pyrophosphate, PPi, as coproduct. The problem with each of these routes is that they would generate phosphoric anhydride bonds in the ATP and PPi products and so be uphill thermodynamically and disfavored energetically. Cells instead use the third reagent, water, as cosubstrate for protein dephosphorylations. The hydrolysis to yield inorganic phosphate is at worst energetically neutral and typically has a positive energetic outcome, with a relatively low energetic

barrier. (This route does lead to net ATP consumption when the protein kinase–protein phosphatase pairs are considered as an ensemble.)

The protein phosphatases hydrolyzing P-Ser and P-Thr alcoholic phosphate monoester bonds follow a mechanistic path analogous to phosphatases that act on low molecular weight phosphoester metabolites,[56,57] discussed in Chapter 6. The enzymes are metalloprotein catalysts, using divalent cations to orient both the dianionic phosphate side chains of the Ser and Thr phosphate esters and to orient and activate a water molecule for nucleophilic attack. Intriguingly, protein phosphatase 1 and 2 (PP1 and PP2) complexes tend to be isolated with Zn^{2+} and Fe^{2+} (rather than two Mg^{2+}) as the bimetallic catalytic center. Direct in line phosphoryl transfer *via* an at least partially associative transition state is the proposed mechanism for all those (~230 000) P-Ser and P-Thr phosphoprotein hydrolyses (Figure 15.24).

In contrast, some 95 of the 107 human phosphotyrosyl protein phosphatases[58] employ covalent catalysis as described earlier in Chapter 14 (Figures 14.24 and 15.24). These all form cysteinyl–S–PO_3^{2-} covalent enzyme intermediates, resulting from attack of the active site cysteine thiolate on the P-Tyr–PO_3^{2-} group. The remaining dozen or so PTPases are metalloenzyme hydrolases with the conventional phosphatase logic and metal cation active site machinery. To reverse the autophosphorylation events of the transmembrane receptor tyrosine kinases, some of the protein tyrosine phosphatases are themselves transmembrane proteins with catalytic domains (often tandemly repeated) on the cytoplasmic side of the membrane. Figure 15.25 sketches the orientation of three PTPases CD45, PTPζB, and Drosophila PTP (DPTP). The extracellular domains can be occupied by external ligands to control the PTPase activities in the cytoplasmic domains. Also depicted are the congeneric Src homology 2 domain tyrosine phosphatases SHP-1 and SHP-2 that have tandem P-Tyr-binding SH2 domains to recruit P-Tyr substrate proteins into proximity for dephosphorylation.

To effect specific hydrolysis of subsets of the estimated ~230 000 P-Ser–P-Thr sites and ~10 000 P-Tyr sites in human phosphoproteins to reverse the action of specific subsets of protein kinases, one can imagine two limiting strategies.[59] One strategy would be to pair up each protein kinase with a partner protein phosphatase. This is the approach bacteria use in their two-component sensor kinase–response regulator logic. Given that humans express 90 protein tyrosine kinases and 107 phosphotyrosyl protein phosphatases, that is effectively the strategy for making phosphotyrosine proteomics a usefully reversible posttranslational signaling modality.

The other limiting strategy might be to express only a small number of protein phosphatases that are chemospecific (P-Ser and P-Thr but not P-Tyr) where the catalytic subunit is intrinsically promiscuous with regard to protein sequence or architecture. Specificity could then be imposed with regulatory subunits that impart partner protein recognition and/or serve to anchor the catalytic phosphatase subunit in certain subcellular locations or complexes. This is the limiting strategy for the P-Ser and P-Thr protein

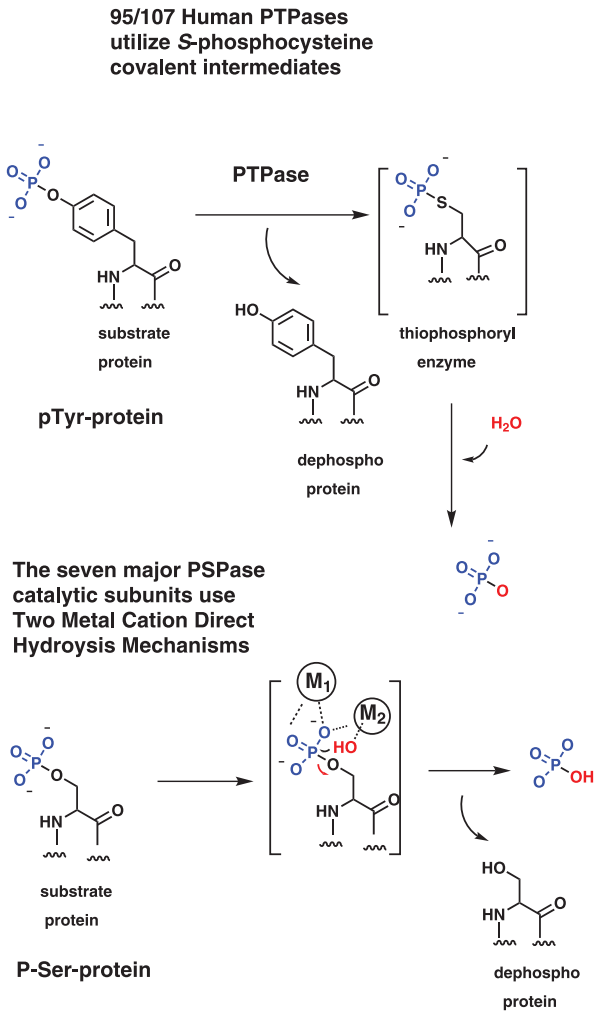


Figure 15.24 Categories of protein phosphatases: The great bulk of PTPases use covalent catalysis *via* phosphorothioate–enzyme adducts while the small class of PSPases use water as a direct cosubstrate in a two-metal cation active site.

phosphatases.[57] Only ~40 genes are used for such phosphatase catalytic subunits, approximately one tenth the number of the 400 Ser–Thr kinases to be balanced. The most important are as few as seven PSPases, numbered PP1–7 (Figure 15.26). Within this family, PPI and PP2 isoforms are thought to carry ~90% of the catalytic flux for those potential ~230 000 P-Ser–P-Thr dephosphorylations.[57,59]

Specificity is achieved by regulatory subunit combinatorics.[60,61] There are over 100 regulatory subunits for PP1. Some are phosphorylated and become substrates, some are scaffolding proteins that bring substrate

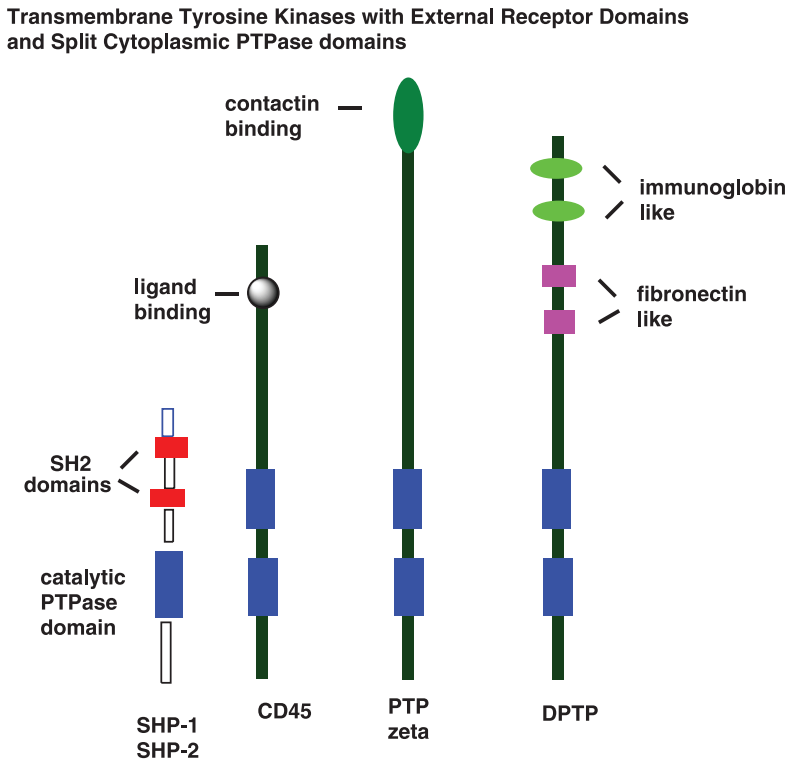


Figure 15.25 Categories of protein phosphatases: Schematic of three receptor tyrosine kinases – CD45, PTPase zeta B, and DPTP – and one cytoplasmic SHP-1/2 member.[55]

phosphoproteins into proximity, some anchor the PP1 and PP2 complexes to membrane surfaces of specific organelle membranes to increase proximity to classes of localized phosphoproteins (Figure 15.26). The combinatorial use of a catalytic phosphatase subunit (of which there are four different isozymes for PP1, derived from three genes) with a regulatory subunit and a scaffolding subunit in PP1 heterotrimers means there are thousands of molecular variants that can act in specific subcellular locations towards specific sets of P-Ser–P-Thr phosphoproteins for hydrolysis (Figure 15.26). There are also protein inhibitors that can tie up catalytic or regulatory subunits of PP1 and PP2A to modulate or control phosphatase activity. One should think of the P-Ser–P-Thr protein phosphatase cellular inventory in terms of dynamic combinatorics to achieve specificity.

Among the family of phosphotyrosine protein phosphatases is the enzyme $SHPTP_2$ that has two SH_2 (phosphotyrosine binding) domains tethered intramolecularly upstream of the catalytic phosphatase domain. In the absence of P-Tyr protein substrates, the intramolecular SH_2 domains, particularly the N-terminal one, inhibit target protein access and the phosphatase activity is negligible[62] (Figure 15.25).

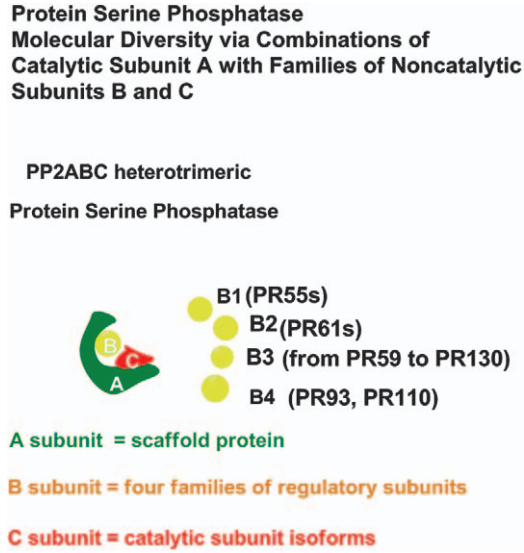


Figure 15.26 Categories of protein phosphatases: Schematic of trimeric PSPases with catalytic subunit A and a large set of regulatory subunits B and C that alter location and substrate selectivity and enable many combinations of the ABC trimeric PSPases.[54]

In the presence of a protein with multiple P-Tyr residues, such as the cytoplasmic domains of receptor tyrosine kinases (EGFR, IRS, PDGFR, FGFR), engagement of the SH_2 domain of $SHPTP_2$ with one of the target protein P-Tyr domains, frees up the phosphatase active site for hydrolytic activity. The catalytic domain is now in the vicinity of those multiple P-Tyr residues and can act to hydrolyze them. If the SH_2 domain of the phosphatase dissociates and then reassociates with a different P-Tyr residue, one can imagine removal of all the P-Tyr residues from the cytoplasmic domain of a multiply phosphorylated P-Tyr receptor kinase to turn off all signaling modalities. The fusion of an SH_2 phosphotyrosine binding domain to a catalytic phosphatase domain thus can increase the efficiency of tandem P-Tyr hydrolyses.

A third subcategory of protein phosphatases, that are structurally related to classical PTPases, have dual specificity (dual specificity protein phosphatases = DUSP) for hydrolytic removal of $-PO_3^{2-}$ groups from both serine and threonine residues within a single target protein (Figure 15.27). Six subgroups of DUSPs have been categorized[64] and several act in the MAP kinase cascades. All three types of protein phosphatases work in signaling cascades to control the timing and extent of protein phosphorylations and strength of various protein-based signals. Figure 15.27 schematizes two P-Ser–P-Thr- and one P-Tyr-based signaling pathway from membrane to cytoplasm to nucleus. The locations of action of PP1, PTPs, and DUSPs in the cytoplasm and nucleus are indicated.

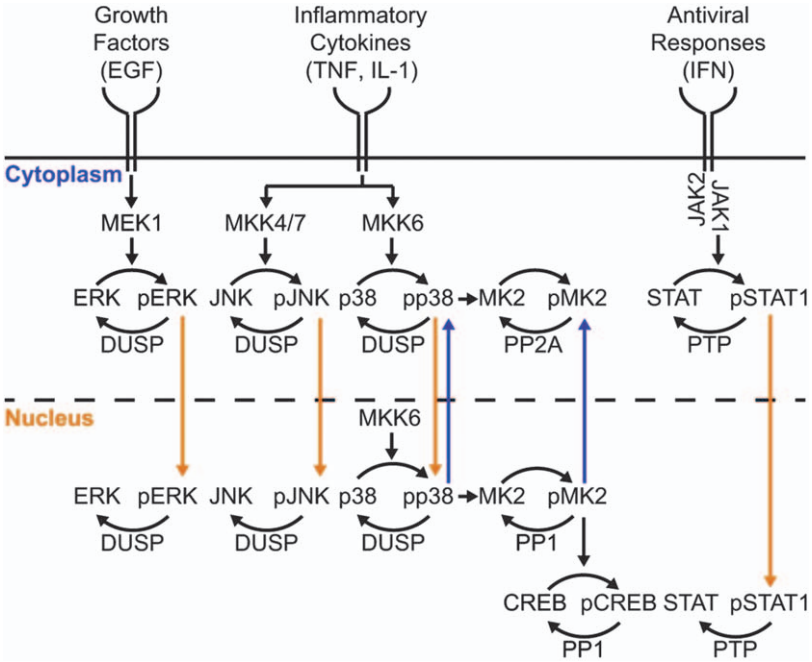


Figure 15.27 Schematic of three signaling pathways initiated on the external face of the cell membrane by growth factors, inflammatory cytokines (tumor necrosis factor (TNF)), interleukin-1 (IL1), or antiviral interferons (IFN).[63] The left and middle routes transmit cytoplasmic signals *via* Ser–Thr protein kinases, while the rightmost route uses tyrosine kinases. The physiological phosphatases dephosphorylating the phosphoserine forms of extracellular signal regulated kinase (pERK), pJNK, and pp38 in both cytoplasm and nucleus (to terminate signaling) are double specificity phosphatases (DUSPs). The phosphorylated mitogen-activated protein kinase-activated protein kinase 2 (pMK2) and the phospho-form of the cyclic AMP response element binding protein (CREB) transcription factor are hydrolyzed by the PSPase PP1. PP2A is active in the cytoplasm, PP1 in the nucleus on pMK2. pSTAT1 in both the cytoplasm and nucleus is phosphorylated on tyrosine and hydrolyzed by PTPases. Thus, all three types of protein phosphatases are at work at distinct signaling nodes.
Reproduced from ref. 63 with permission from American Society for Biochemistry and Molecular Biology, Copyright 2017.

References

1. P. Vlastaridis, P. Kyriakidou, A. Chaloiotis, Y. van de Peer, S. G. Oliver and G. D. Amoutzias, Estimating the total number of phosphoproteins and phosphorylation sites in eukaryotic proteomes, *GigaScience*, 2017, **6**, 1–11.
2. J. Ding, W. Qian and R. D. Smith, Quantittative proteomic approaches for studying phosphotyrosine signaling, *Expert Rev. Proteomics*, 2007, **4**, 13–23.

3. G. Hardman, S. Perkins, P. J. Brwnridge, C. J. Clarke, D. P. Byrne, A. E. Campbell, A. Kalyuzhnyy, A. Myall, P. A. Eyers, A. R. Jones and C. E. Eyers, Strong-anion exchange-mediated phosphoproteomics reveals extnsive human noncanonical phosphorylation, *EMBO J.*, 2019, **38**(21), e100847.
4. R. Aebersold and M. Mann, Mass spectrometry-based proteomics, *Nature*, 2003, **422**, 198–207.
5. R. Aebersold and M. Mann, Mass-spectrometric exploration of proteome structure and function, *Nature*, 2016, **537**, 347–355.
6. H. C. Reinhardt and M. B. Yaffe, Phospho-Ser/Thr-binding domains: navigating the cell cycle and DNA damage response, *Nat. Rev. Mol. Cell Biol.*, 2013, **14**, 563–580.
7. M. Yaffe, Phosphotyrosine-binding domains in signal transduction, *Nat. Rev. Mol. Cell Biol.*, 2002, **3**, 177–186.
8. J. V. Chan, D. X. P. Koh, Y. Liu, T. L. Joseph, D. P. Lane, C. S. Verma and Y. S. Tan, Role of the N-terminal lid in regulating the interaction of phosphorylated MDMX with p53, *Oncotarget*, 2017, **8**, 112825.
9. C. J. Leonard, L. Aravind and E. V. Koonin, Novel families of putative protein kinases in bacteria and archaea: evolution of the ''eukaryotic'' protein kinase superfamily, *Genome Res.*, 1998, **8**, 1038–1047.
10. S. S. Taylor, M. Keshwani, J. M. Steichen and A. P. Kornev, Evolution of the eukaryotic protein kinases as dynamic molecular switches, *Philos. Trans. R. Soc., B*, 2012, **367**, 2517–2528.
11. G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsnam, The protein kinase complement of the human genome, *Science*, 2002, **298**, 1912–1934.
12. Slide 5 of 105 on tyrosine kinase network signaling in normal physiology and cancer in Avraham readings on tyrosine kinases (downloaded 11-09-2019).
13. M. Tinti, F. Madeira, G. Mururgesan, G. Hoxhai, R. Toth and C. Mackintosh, Annotation and Integrated analysis of the 14-3-3 interactome, *Database*, 2014, bat085.
14. A. Liu and T. Hunter, Metabolic kinases moonlighting as protein kinases, *Trends Biochem. Sci.*, 2018, **43**, 301–310.
15. D. R. Robinson, Y.-M. Wu and S.-F. Lin, The protein tyrosine kinase family of the human genome, *Oncogene*, 2000, **19**, 5548–5557.
16. C. Stark, T.-C. Su, A. Breitkreutz, P. Lourenco, M., Dehabieh, B.-J. Breitkreutz, M. Tyers and I. Sadowski, PhosphoGRID a database of experimentally verified *in vivo* protein phosphorylation sites from the budding yeast *Saccharomyces cerevisiae*, *Database*, 2010, **2010**, bap026.
17. H. Nishi, E. Demir and A. R. Pancheko, Crosstalk between signaling pathways by single and multiple protein phosphorylation sites, *J. Mol. Biol.*, 2015, **427**, 511–520.
18. O. Hantschel and G. Superti-Furga, Regulation of c-Abl and Bcr-Acl tyrosine kinases, *Nat. Rev. Mol. Cell Biol.*, 2004, **5**, 33–44.

19. D. J. Webb, M. J. Schroeder, C. J. Brame, L. Whitmore, J. Shabanowitz, D. F. Hnt and A. R. Horwitz, Paxillin phosphorylation sites mapped by mass spectrometry, *J. Cell Sci.*, 2005, **118**, 4925–4939.
20. J. A. Endicott, M. E. M. Noble and L. N. Johnson, The structural basis for control of eukaryotic protein kinases, *Annu. Rev. Biochem.*, 2012, **81**, 587–613.
21. A. Plotnikov, E. Zehorai, S. Procaccia and R. Seger, The MAPK cascades: signaling components, nuclear roles and mechanisms of nuclear translocation, *Biochim. Biophys. Acta*, 2011, **1813**, 1619–1633.
22. P. J. W. Roberts and C. J. Der, Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer, *Oncogene*, 2007, **26**, 3297–3310.
23. The Biology Project website (downloaded on 10-31-2009) Tutorial on activation of protein kinases; figure labeled protein kinase cascade.
24. Y. Keshet and Seger, The MAP kinase signaling cascades: a system of hundreds of components regulates a diverse array of physiologic functions, *Methods Mol. Biol.*, 2010, **661**, 3–38.
25. R. Milo and R. Phillips, *Cell Biology by the Numbers*, Taylor & Francis Books, Garland SciencePress, 2015.
26. J. A. Printen and G. F. Sprague, Protein-protein interactions in the yeast pheromone response pathway: Ste5p interacts with all members of the MAP kinase cascade, *Genetics*, 1994, **138**, 609–619.
27. T. Q. Nguyen, M. Aumont-Nicaise, J. Andreani, C. Vellours, M. Chenon, F. Vilela, C. Geneste, P. F. Varela, P. Llinase and J. Menetrey, Characterization of the binding mode of JNK-interacting protein (JIP1) to kinesin light chain 1 (KLC1), *J. Biol. Chem.*, 2018, **293**, 13946–13960.
28. H. Matsuura, H. Nishitoh, K. Takeda, A. Matsuzawa, T. Amagasa, M. Ito, K. Yoshioka and H. Ichijo, Phosphorylation-dependent scaffolding role of jsap1/jip3 in the ASK1-JNK signaling pathway, *J. Biol. Chem.*, 2002, **277**, 40703–40709.
29. J. W. Locasale, S. Shaw and A. K. Chakraborty, Scaffold proteins confer diverse regulatory properties to protein kinase cascades, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 13307–13312.
30. D. N. Dhanasekaran, K. Kashef, C. M. Lee, H. Xu and E. P. Reddy, Scaffold proteins of MAP-kinase modules, *Oncogene*, 2007, **26**, 3185–3202.
31. J. D. Smith and J. D. Scott, Protein kinase A; something new under the sun?, *J. Cell Biol.*, 2018, **217**, 1895–1897.
32. D. E. Casteel, E. Smith-Nguyen, B. Sankaran, S. H. Roh, R. B. Pilz and C. Kim, A crystal structure of the cyclic GMP-dependent protein kinase dimerization/docking domain reveals molecular details of isoform-specific anchoring, *J. Biol. Chem.*, 2010, **285**, 32684–32688.
33. S. F. Steinberg, Structural basis of protein kinase C isoform function, *Physiol. Rev.*, 2008, **88**, 1341–1378.
34. I.-s. Chen, Z.-k. Dai, D. G. Welsh, I.-j. C Chen and B.-n Wu, Protein kinases modulate store-operated channels in pulmonary artery smooth muscle cells, *J. Biomed. Sci.*, 2011, **18**, 2.

35. Medical College of Wisconsin website (downloaded 10-30-2109) Pathway diagram.
36. Cell Migration Lab, U of Reading website (downloaded 10-21-2019) Cell migration.
37. S. K. Hanks and T. Hunter, The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification, *FASEB J.*, 1995, **9**, 576–596.
38. S. F. Steinberg, Post-transational modifications at the ATP-positioning G-loop that regulate protein kinase activity, *Pharmacol. Res.*, 2018, **135**, 181–187.
39. V. Modi and R. L. Dunbrack, Defining a new nomenclature for the structure of active and inactive kinases, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 6818–6827.
40. Z. Wang and P. A. Cole, Catalytic mechanisms and regulation of protein kinases, *Methods Enzymol.*, 2015, **548**, 1–21.
41. S. R. Hubbard and W. T. Miller, Recptor tyrosine kinases: mechanisms of activation and signaling, *Curr. Opin. Cell Biol.*, 2008, **19**, 117–123.
42. MBL website (downloaded 10-29-2019) tyrosine kinase assay.
43. M. Z. Cabail, S. Li, E. Lemmon, M. E. Bowen, S. R. Hubbard and W. T. Miller, The insulin and IGF receptor kinase domains are functional dimers in the activated state, *Nat. Commun.*, 2015, **6**, 6406.
44. Memorial Univesrsity of Newfoundland website (downloaded 10-29-2019) Principles of Cell Biology (BIOL2060) copyright 2009 Pearson Education Inc.
45. OpenLearn website, (downloaded 11-01-2019) PDGF receptor in cell signaling.
46. M. J. Wagner, M. M. Stacey, B. A. Liu and T. Pawson, Molecular mechanisms of SH2- and PTB-domain-containing proteins in receptor tyrosine kinase signaling, *Cold Spring Harbor Perspect. Biol.*, 2013, **5**, a008987.
47. M. Shohag, T. Nishioka, R. U. Ahammad, S. Nakamuta, Y. Yura, T. Hamaguchi, K. Kaibuchi and M. Amano, Phosphoproteomic analysis using the WW and FHA domains as biological filters, *Cell Struct. Funct.*, 2015, 95–104.
48. J. Schlessinger and M. A. Lemmon, SH2 and PTB domains in tyrosine kinase signaling, *Sci. STKE*, 2003, **191**, re12.
49. Y. H. Li, P. P. Wang, X. X. Li, C. Y. Yu, H. Yang, J. Zhou, W. W. Xue, J. Tan and F. Zhu, The Human kinome trageted by FDA Aproved Multi Target Drugs and Combination Products: a comparative study from the drug-target interaction network perspective, *PLoS One*, 2016, e0165737.
50. K. Gluza and M. Dobranska (2016) New horizons in next-generation small molecule kinase inhibitors. (Drug Traget Review website-downloaded 2-14-2020).
51. M. J. Miller, K. C. Foy and P. T. Kaumaya, Cancer immunotherapy: present status, future perspective and a new paradigm of peptide immunotherapeutics, *Discovery Med.*, 2013, **15**, 166–176.

52. Z. Zhao and P. A. Bourne, Progress with covalent small-molecule kinase inhibitors, *Drug Discovery Today*, 2018, **23**, 727–735.
53. A. Tomiyama, T. Kobayashi, K. Mori and K. Ichimura, Protein phosphatases- a touchy enemy in the battle against glioblastomas: a review, *Cancers*, 2019, **11**, 241.
54. J. J. Thompson and C. S. Williams, Protein phosphatase 2A in the regulation of wnt signaling, stem cells and cancer, *Genes*, 2018, **9**, 121.
55. OpenLearn website (downloaded11-02-2019) section 3.7 protein phosphatases. Figure 44.
56. D. Barford, A. K. Das and M. P. Egloff, The structure and mechanism of protein phosphatases: insights into catalysis and regulation, *Annu. Rev. Biophys. Biomol. Struct.*, 1998, **27**, 133–164.
57. Y. Shi, Serine/Threonine phosphatase: mechanism through structure, *Cell*, 2009, **139**, 468–484.
58. N. K. Tonks, Protein tyrosine phosphatases:from genes to function to disease, *Nature Revs*, *Mol. Cell. Biol.*, 2006, **7**, 833–846.
59. F. Sacco, L. Perfetto, L. Castagnoli and G. Ceasareni, The human phosphatase interactome. An intricate family portrait, *FEBS Lett.*, 2014, **586**, 2732–2739.
60. P. T. W. Chen, Protein phosphatase 1-targeted in many directions, *J. Cell Sci.*, 2002, **115**, 241–256.
61. F. Wang, S. Zhu, L. A. Fisher, W. Wang, G. G. Oakley, C. Li and A. Peng, Protein interactomes of protein phosphatase 2A B55 regulatory subunits reveal B55-mediated regulation of replication protein A under replication stress, *Sci. Rep.*, 2018, **8**, 1–10.
62. S. Sugimoto, T. Wandless, S. E. Shoelson, B. G. Neel and C. T. Walsh, Activation of the SH2-containing protein tyrosine phosphatase SH-PTP2 by phosphotyrosine-containing peptides derived from insulin receptor substrate-1, *J. Biol. Chem.*, 1994, **269**, 13614–13622.
63. M. Shah, C. M., Smolko, S. Kinicji, Z. D. Chapman, D. L. Brautigan and Jones, Profiling subcellular protein phosphatase responses to coxsackievirus B3 infection of cardiomyocytes, *Mol. Cell. Proteomics*, 2017, **16**, S244–S262.
64. K. L. Patterson, T. O. Brummer, P. M. O'Brien and R. J. Daly, Dual-specificity phosphatases: critical regulators with diverse targets, *Biochem. J.*, 2009, **418**, 475–489.

CHAPTER 16

Noncanonical Phosphoproteomics – II

16.1 Noncanonical Adducts in a Broadened View of Phosphoproteomics

The preceding two chapters, Chapters 14 and 15, have surveyed both: (1) sets of noncanonical phosphoproteins that have escaped conventional mass spectroscopic detection, most probably due to their acid lability, and (2) the conventional P-Ser, P-Thr, and P-Tyr stable troika of canonical phosphorylated side chains thought to carry out the bulk of phosphoproteomic signaling in eukaryotic cells. There are some additional protein posttranslational covalent modifications that incorporate phosphoryl groups into target proteins that fall outside the boundaries of the above two categories. Most often, they are not simple *O*-phosphomonoesters of Ser, Thr, or Tyr residues but nonetheless reveal how phosphorus containing groups fit into the posttranslational proteome universe.

We summarize some of these in this chapter to give further scope to the logic and machinery that cells use for derivatization of protein residue side chains with electrophilic fragments that contain one or more phosphoryl groups. While the first two examples use ATP as donor, the two subsequent enzyme families use alternate donors that contain more complex phosphorus-containing substituents.

One can quarrel with the inclusion of these posttranslational categories as ''simple'' phosphoproteomics, but they all contain anionic phosphoryl groups. They thereby fit in our ''noncanonical'' categories of enzymatic strategies that use phosphorus-containing metabolites as donors of electrophilic fragments to nucleophilic side chains of protein residues (Figure 16.1). Not surprisingly, the biological niches for these posttranslational modifications can be quite

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

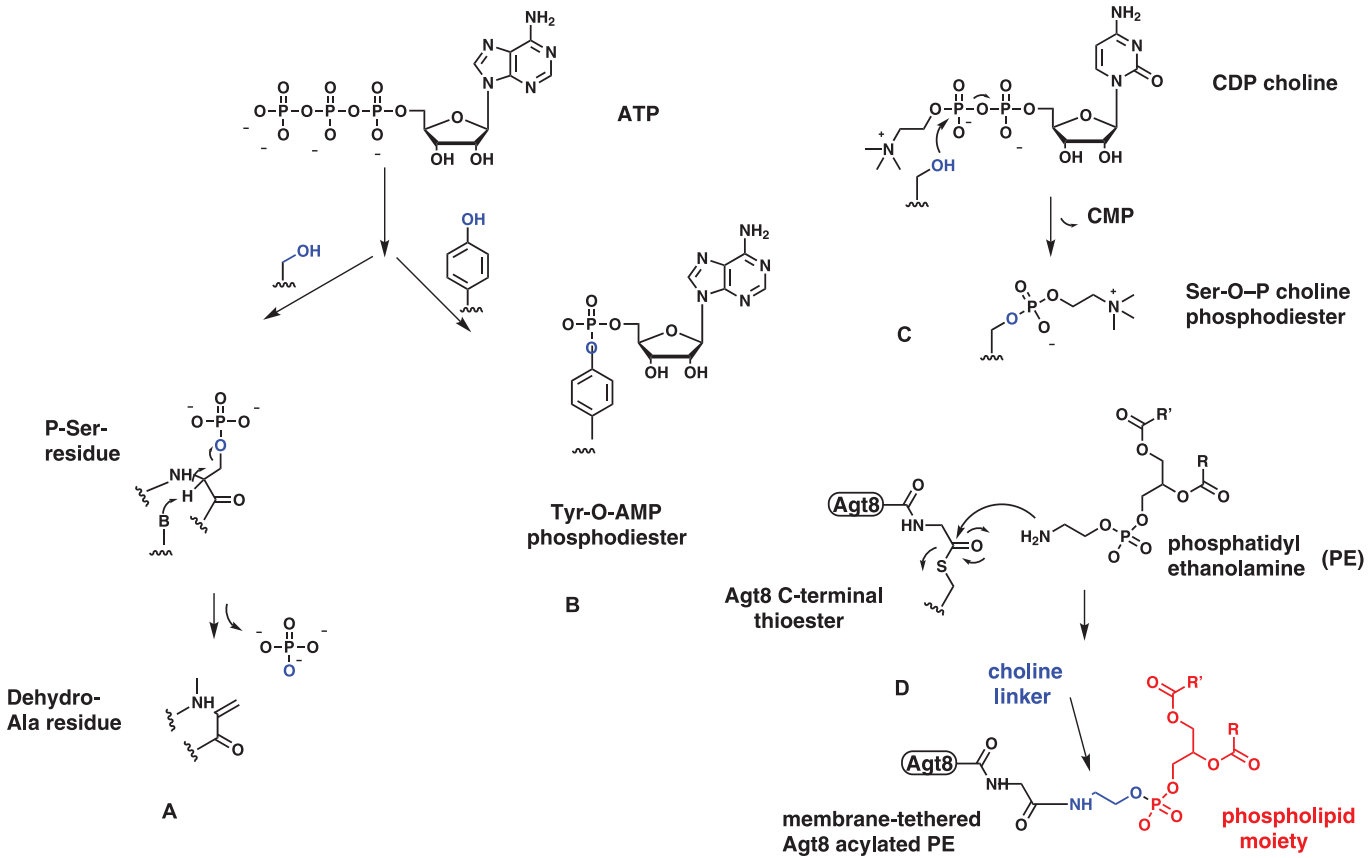


Figure 16.1 Four noncanonical proteome modifications: (A) Dehydroalanyl residues *via* P-Ser eliminations; (B) Tyrosyl–O–AMP phosphodiesters; (C) Ser–O–phosphorylcholine diesters; (D) Autophagy-related protein 8 (Agt8) protein linked to the head group of the membrane phospholipid phosphatidylethanolamine *via* a phosphodiester linkage.

distinct and specialized functionally. There are probably other examples of posttranslational adducts that might make the list for interested readers beyond the four examples in Figure 16.1.

16.2 Cryptic Phosphoserines in Lanthipeptide Formation

There is one well studied protein context in which serine and threonine residues are posttranslationally phosphorylated but then subjected to enzyme-mediated elimination of inorganic phosphate across the C2 and C3 carbons of the serine residue (Figures 16.2 and 16.3). These are the family of

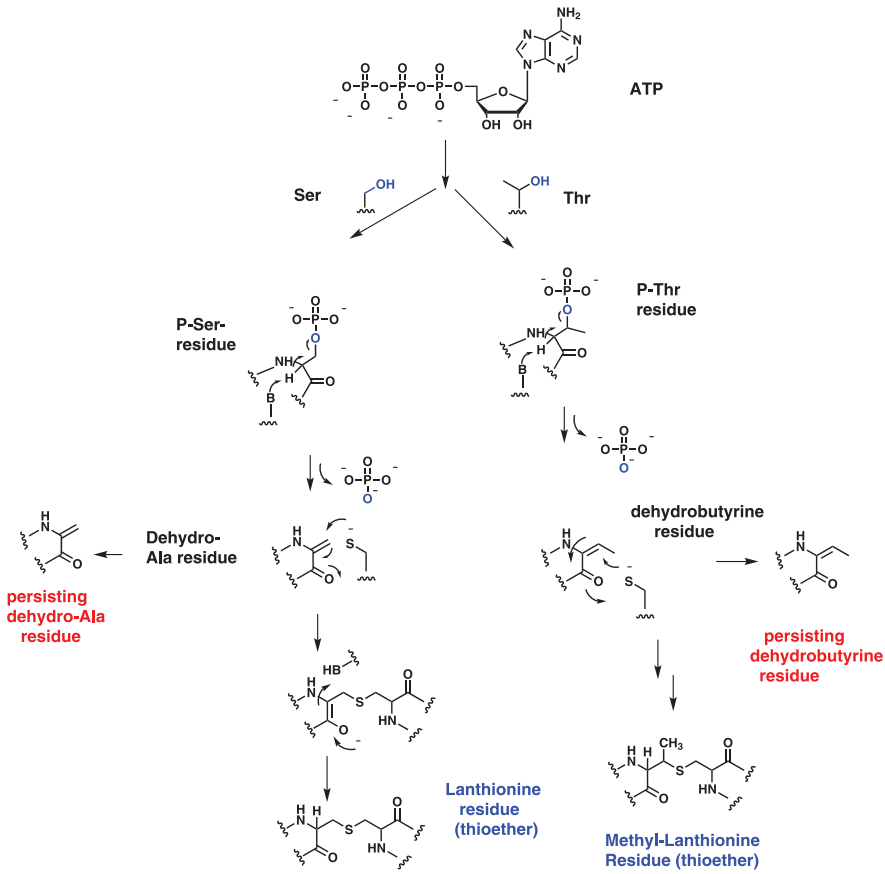


Figure 16.2 Enzymatic processing of Ser or Thr residues in nascent lanthipeptide precursor proteins to P-Ser or P-Thr residues, followed by elimination of inorganic phosphate across C2 and C3. These olefinic side chains may persist in the mature peptide or be captured by neighboring cysteine thiolate anions to create the thioether residues lanthionine or methyllanthionine.

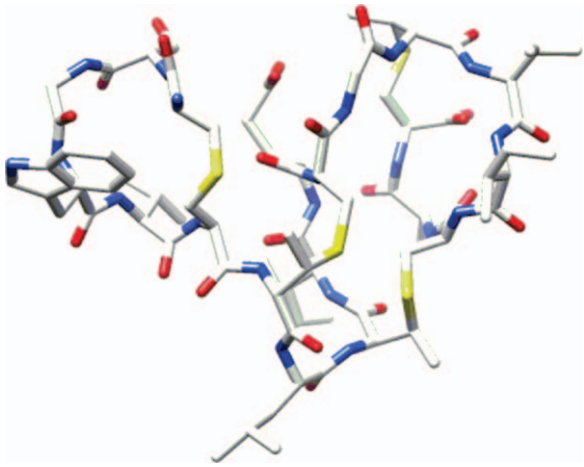


Figure 16.3 NMR structure of lanthipeptide antibiotic actagardine (pdb 1AJ1), with thioether linkages in yellow.

lanthipeptides.[1] The P-Ser or P-Thr residues do not accumulate. They are transient intermediates on the way to olefinic dehydroalanlyl (from P-Ser) and dehydrobutyryl residues (from P-Thr). This is a posttranslational strategy in hundreds to thousands of bacterial proteins. The fate of many but not all of the olefinic side chains is subsequent intramolecular capture by a thiolate anion of a particular cysteine side chain.[2] The resultant nascent isoamide adducts isomerize in the enzyme active sites, with C2 protonation, (often to give the D-isomers) to yield lanthionine or methyllanthionine residues. Lanthionyl and methyl-lanthionyl residues are bifunctional with a stable thioether linkage, representing conformationally restrictive covalent crosslinks.

Subsequent limited proteolysis digests the crosslinked proteins into compact lanthipeptides with a diverse range of biological activities. Among the functions are potent complexation of bacterial lipid II by nisin acting as a food preservative. The saposin B (SapB) lanthipeptide instead functions as a developmental morphogen for streptomycete spore-forming bacteria.[3]

Lanthipeptide bacterial gene clusters typically encode the substrate protein to be modified and a set of modifying enzymes. One protein partner is a Ser–Thr kinase homolog. It is dedicated to forming the P-Ser or P-Thr modifications noted above. This is classic Ser–Thr protein kinase activity on a specific substrate protein. For nisin maturation, eight such $O–PO^{2-}$ side chain monoesters are generated (Figures 16.2 and 16.3). We have noted iteratively that such P-Ser and P-Thr side chains are chemically stable.

The elimination of inorganic phosphate is therefore enzyme-mediated, presumably initiated by abstraction of the C2–H of a newly formed P-Ser–P-Thr residue as a proton, setting up a low energy elimination of $HO–PO_3^{2-}$ from C3. This is the *olefin-forming step* as each of the eight P-Ser and P-Thr phosphate esters undergo elimination of Pi across the C2–C3 bonds in nisin. These

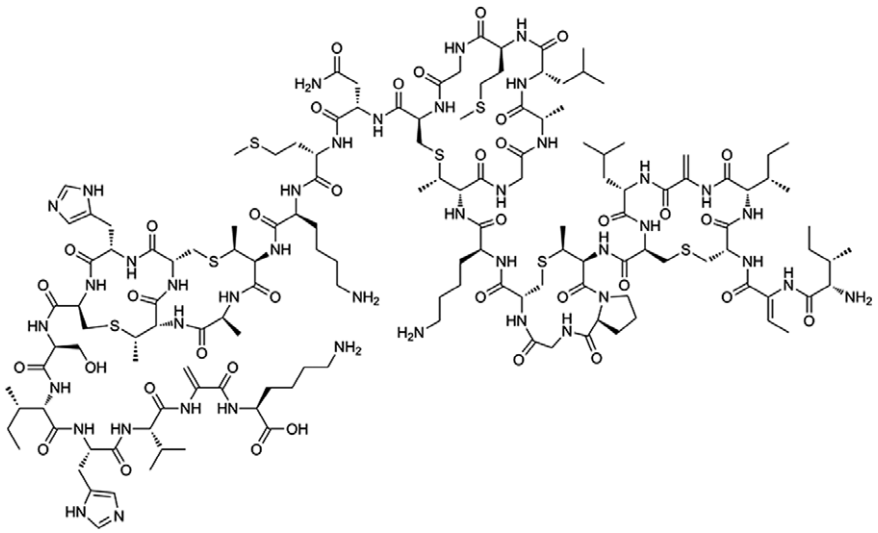


Figure 16.4 The food preservative nisin is generated as a small 34 residue preprotein, in which five Thr and three Ser side chains are phosphorylated enzymatically by ATP and then eliminated to produce five dehydrobutyryl and three dehydroalanyl resides.

eliminations subsequent to phosphorylations do *not* happen in the action of the several hundred traditional kinases of the human kinome.

The olefins in nascent dehydrolalanyl and dehydrobutyryl residues are also chemically stable (in the absence of nearby nucleophiles) and some persist in a range of peptides derived from such posttranslational processing enzyme machinery. However, those olefins are electrophilic and capturable by cysteinyl thiolates in the immediate (usually intramolecular) microenvironment. Mature nisin has five thioethers, a mix of one lanthionyl and four methyllanthionyl residues (Figures 16.4 and 16.5). It is the thioether formations that constrain conformational flexibility and introduce rigidity into the maturing modified peptide scaffolds. Mature nisin is resistant to proteases.

Thousands of lanthipeptide modification genes have been detected in cyanobacterial genomes,[4] indicating that the tandem steps of serine–threonine protein kinases followed by P-Ser–P-Thr eliminase actions to generate the dehydro amino acid residues are widespread in prokaryotes. In turn, the presence of a dehydroalanyl or dehydrobutyryl side chain in a bacterial protein or peptide is most often a signature that a serine or threonine residue had been modified to set up the elimination. Not all go *via O*-phosphorylation.[1] Some lanthipeptides form *via* Ser–O–glutamyl or Thr–O–glutamyl adducts as an alternate route to deacylating eliminations to the dehydroalanyl and dehydrobutyryl residues (not shown).

In contrast, none of the thousands of P-Ser and P-Thr residues in proteins of higher eukaryotes are known to be enzymatically processed for such

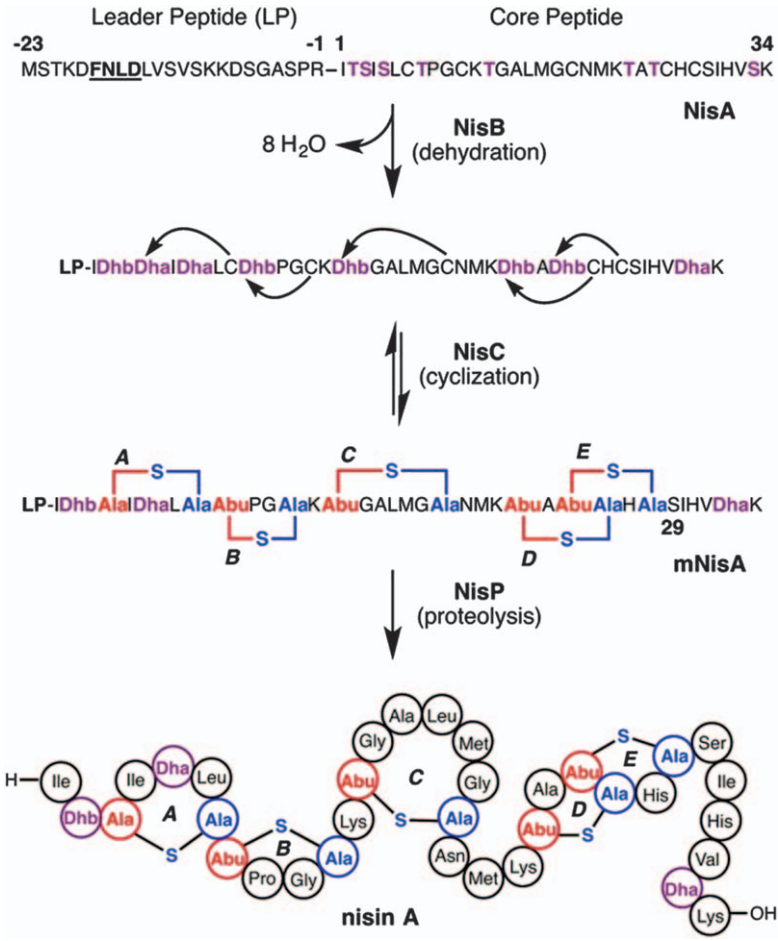


Figure 16.5 There are five cysteine side chains in nascent nisin. All of these capture upstream olefinic side chains, creating five thioethers (four methyllanthionine, one lanthionine), leaving two dehydroalanyl and one dehydrobutyryl side chain in the mature nisin. All five thioethers are formed in the (retrograde) C to N direction.

olefin-forming eliminations. Olefins are not posttranslational products in higher eukaryotic proteomes.

16.3 Nucleotidyl Transfers to Thr and Tyr Side Chains: Protein AMPylations

Back in Chapters 3 and 4 we noted that in low molecular weight metabolites, the two major routes of ATP cleavage were phosphorylation (nucleophilic attack at Pγ) and nucleotidyl transfer (nucleophilic attack at Pα), with dozens to

hundreds of examples of each route to P–O–P anhydride cleavage. *A priori* then, it would not be surprising to expect that there might be similar protein-modifying enzymes that carried out nucleotidylation rather than phosphorylation, particularly given the abundance of modifications of Ser, Thr, and Tyr alcohol side chains (Figure 16.6). When ATP is substrate the nucleotidyl transfer would be classified as adenylylation. In recent years as such protein posttranslational adenylylations have been detected, largely in bacterial metabolism, authors have introduced the colloquial term AMPylation, perhaps to AMP up the visibility of the phenomena in their publications.

16.3.1 Adenylylation and Uridylylation Cascades in Bacterial Glutamine Synthetase Modification

The initial observations on protein side chain nucleotidylation go back almost 50 years to studies by Earl Stadtman's group on regulation of bacterial glutamine synthetase under high and low nitrogen conditions in the culture media.[5] Glutamine synthetase is the primary conduit for fixing NH_3 into amino acids (Figure 16.7) and its activity levels are closely regulated.

Bacterial glutamine synthetase is the key catalyst for fixation of inorganic NH_3 into organic metabolite scaffolds and is essential for life. The strategy is to activate a γ-carboxylate of substrate glutamate *via* phosphoryl transfer from Mg–ATP to produce the intermediate glutamyl phosphate, now activated for capture of ammonia and formation of the amide in glutamine. Glutamine is an amino acid framework to carry around a source of ammonia in a nonnucleophilic form. Close regulation of the enzyme is anticipated to deal with environmental nitrogen limitations on growth. The *E. coli* enzyme is a dodecamer of 50 kDa subunits, a surprisingly large homo-oligomer. This architecture allows cooperative interaction among subunits by both covalent and noncovalent regulators for fine tuning of the flux of NH_3 into metabolite frameworks.

16.3.2 Regulation by Covalent Adenylyl Transfer to a Tyrosyl Residue in Each Subunit

A major mode of regulation turned out to be covalent adenylylation of tryosine residue 397 in each subunit of bacterial glutamine synthase. The Tyr 397–O–AMP phosphodiester form of the enzyme is largely inactive (Figure 16.8). One can find intermediate states of activity, reflecting intermediate levels of covalent adenylation from one to all twelve subunits, but all 12 subunits can be adenylated to turn the enzyme down to very low activity levels.

When exogenous NH_3 levels drop, glutamine synthetase needs to be reactivated to sustain central metabolism. It turns out that the regulatory adenylyl transferase has separate deadenylylation activity. (Figure 16.8).

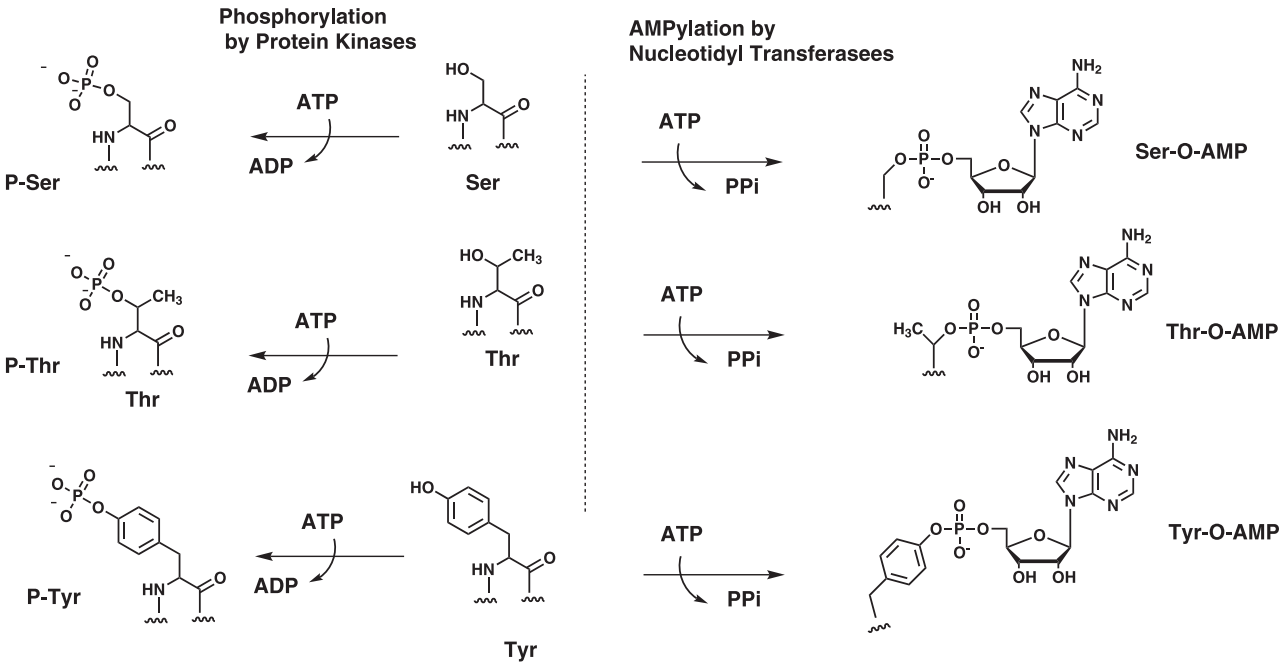


Figure 16.6 Two possible route of Ser, Thr, Tyr side chain modifications by Mg–ATP. Phosphoryl transfers give the familiar troika of P-Ser, P-Thr, and P-Tyr stable phosphoproteins. On the other hand, nucleotidyl transfers would instead give the indicated three adenylylated protein forms.

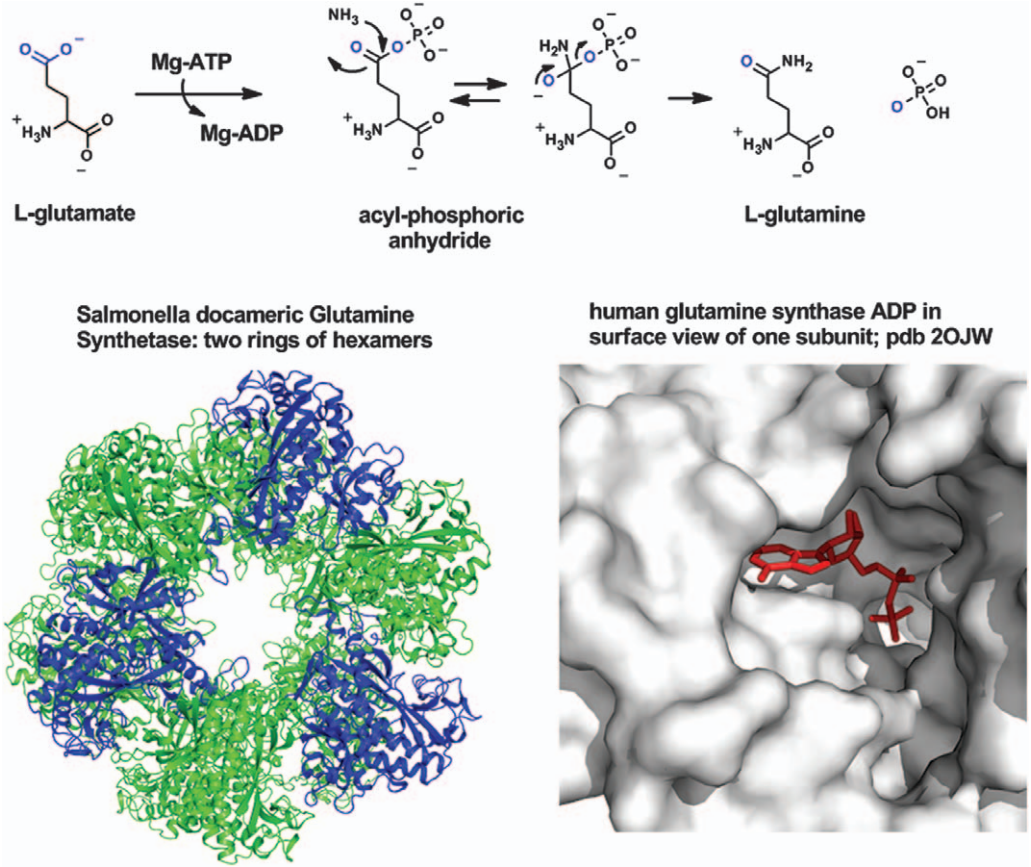


Figure 16.7 Glutamine synthetase is the enzymatic conduit for fixation of inorganic ammonia, NH_3, into amino acid, purine, and pyrimidine scaffolds. It is a highly regulated enzyme. The catalytic strategy is to activate the γ-carboxylate of Glu by phosphoryl transfer from ATP and then aminate that activated carboxyl group. The dodecameric glutamine synthetase is shown from X-ray structure determination (RSCB PDB structure 1F1H from *Salmonella typhimurium*). Also displayed is the binding of ADP to one of the twelve identical subunits.

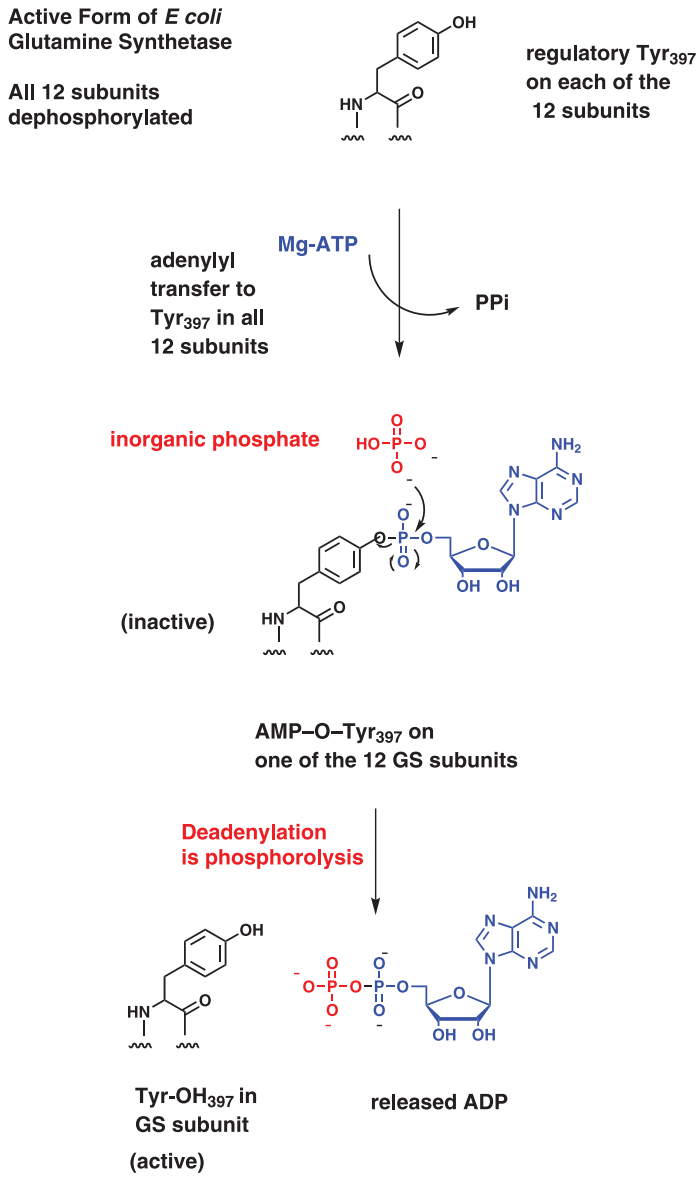


Figure 16.8 The dodecameric bacterial glutamine synthetase can be adenylated on Tyr 397 by a regulatory enzyme using Mg–ATP, acting as an inactivating adenylyl transferase (Tyr-AMPylation in current parlance). This covalent inactivation can be reversed by the same enzyme, acting in a *deadenylation* mode. The deadenylation cosubstrate is not water, but instead inorganic phosphate. Thus, on reactivation of glutamine synthetase ADP is released as coproduct by deadenylylating phosphorolysis.

In turn, another layer of covalent regulation is imposed on the system by a partner enzyme that alters the balance of the adenylyl transferase (ATase) activity between adenylylation and deadenylylation. That balance of adenylylation and deadenylylation can be controlled by the interaction of this ATase modification catalyst with a partner protein known as PII. PII, in turn, can also be in a nucleotidylated form (Figure 16.9). In this case the

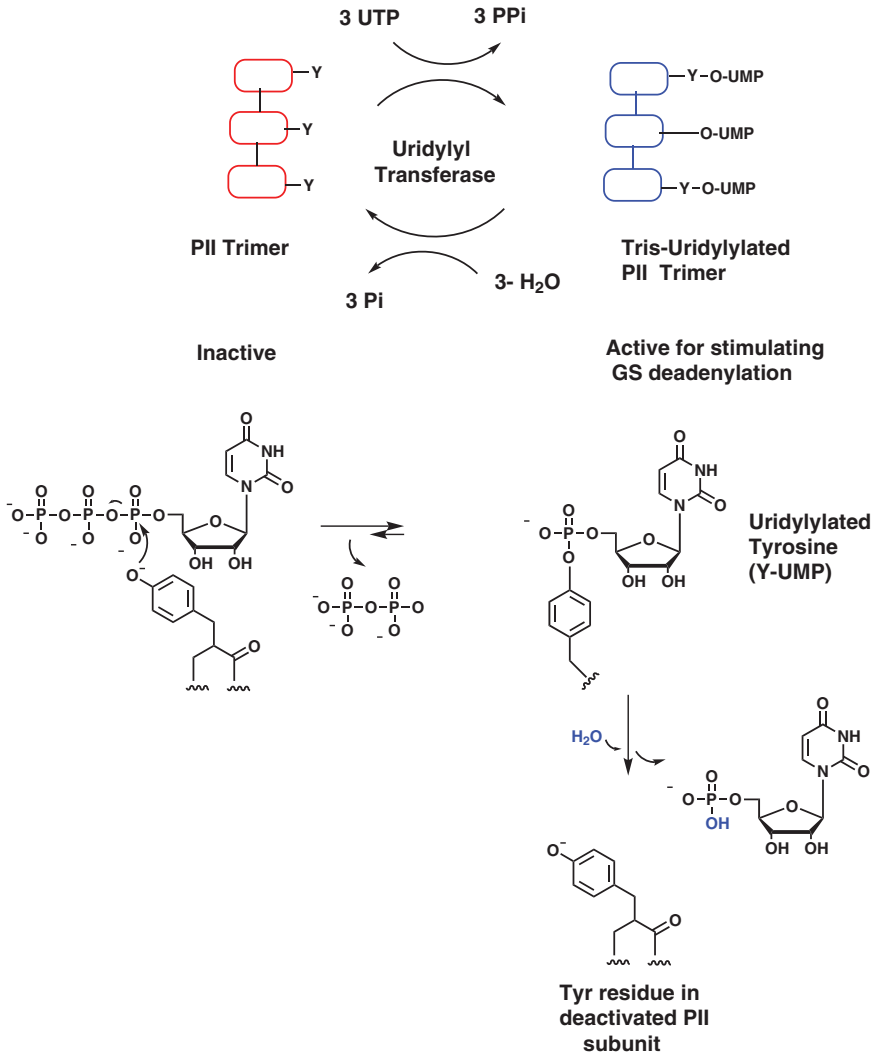


Figure 16.9 A second level of regulatory nucleotidylation in the glutamine synthetase system is provided by a uridylyl transferase (UTase) that adds (or subsequently removes) a uridylyl (UMP) group to a tyrosine residue in each of three subunits of regulatory protein PII. PII interacts with the regulatory adenylyl transferase (ATase).

nucleotidylation is *via* a UMP residue rather than an AMP residue. Up to three tyrosyl residues on PII can be uridylylated by a companion uridylyl transferase (UTase). The UTase itself can also deuridylylate PII under specific conditions. We have noted the role of UTases in RNA processing by 3′-mono- and oligo-uridylylations in Chapter 4.

The unmodified uridylylated PII interacts with the GS-adenylyl transferase and stimulates its deadenylylation activity on its target glutamine synthetase. Two ways of looking at the multilayered two-tiered protein nucleotidylation logic are depicted in Figure 16.10. The left-hand circular diagram shows the cyclical behavior of the four-protein system [UTase, PII, ATPase, and ultimate target glutamine synthetase (GS), with two levels of regulatory nucleotidyl transfers, and one catalytic phosphorylation (GS catalysis)]. The right-hand schematic indicates the costs of such regulatory control expenditures of ATP and UTP to modulate a glutamine synthetase dodecamer and the preferred kinetic directions It also indicates that the activity of glutamine synthetase is also controlled by the levels of glutamate, and 2-ketoglutarate in the bacteria.

The activity of *E. coli* glutamine synthetase is thus controlled by a pair of bifunctional nucleotidyl transferases which allows moment to moment control of the uridylylation and adenylylation levels of the two proteins that integrate metabolite levels to fine tune glutamine synthetase activity.

16.3.3 FIC Adenylation Domains

The bacterial glutamine synthetase paradigm appeared to be a singular use of posttranslational nucleotidylation logic for almost four decades from its original discovery in the 1960s and 1970s. This changed with the discovery that bacterial filamentation phenotype induced by cAMP (FIC) domains contained activatable protein nucleotidyl transferase domains.[6,7] Specifically, these use ATP, transfer an AMP equivalent, and have been termed AMPylation domains. There are thousands of FIC-like open reading frames encoded in bacterial genomes. Most of them are thought to be inactive in the basal state due to an α-helix that charge pairs with an active site residue to occlude access of ATP to the active site.

Activation of FIC domains includes auto-AMPylation of Tyr-OH side chains to relieve the inhibition of the mobile α-helix (Figure 16.11). One of the targets of bacterial FICs is adenylyl transfer to tyrosine residues in the DNA gyrase B (GyrB) or partitioning protein E (ParE) subunits of the two classes of DNA toposiomerase II dimers (Figure 16.11). The consequent inactivation of gyrase activity turns on the SOS DNA damage response genes and leads, among other effects, to the elongation and filamentation of cells that was the original FIC phenotype.

Several bacterial pathogens, including *Vibrio parahemolyticus* and *Histophilius somni* secrete FIC type AMPylation enzymes that modify threonine residues in various members of the small GTPase superfamily (Figure 16.11). The *V. parahemolyticus* Vibrio outer protein S (VopS) targets a threonine in the switch I region of RhoGTPases, slowing down hydrolysis of GTP to tightly

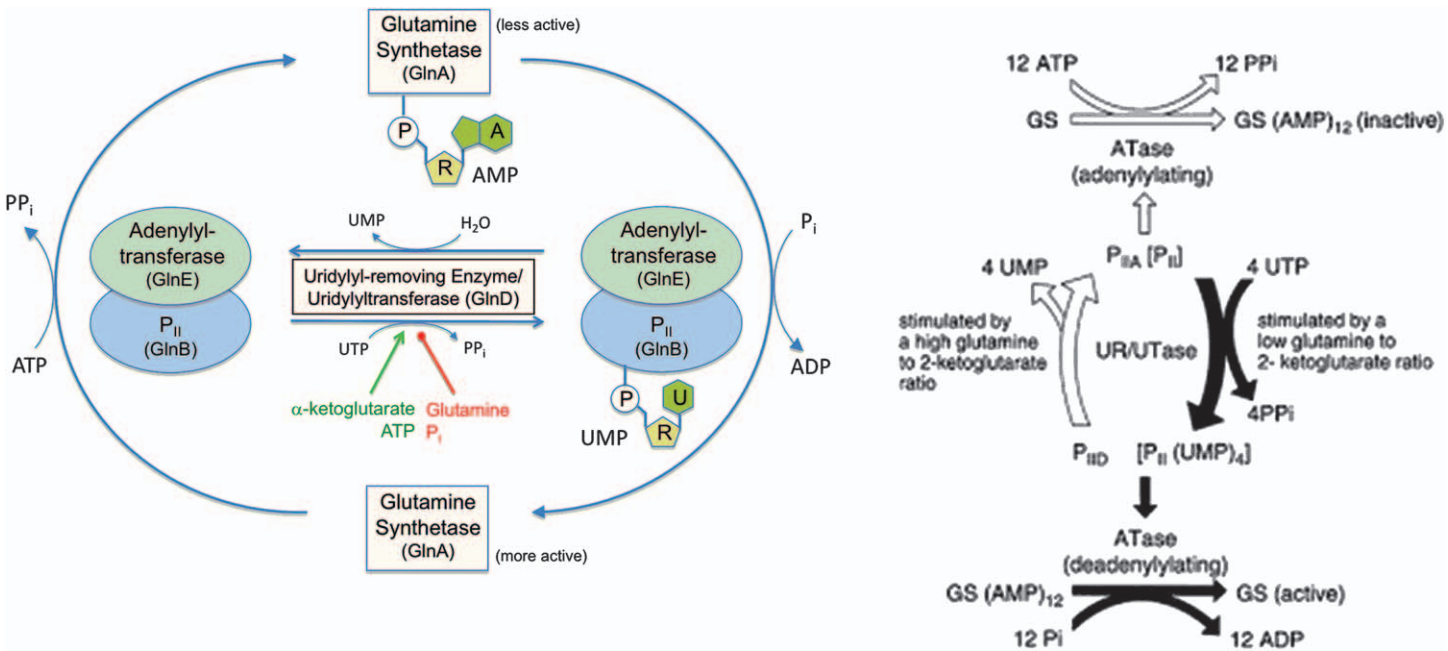


Figure 16.10 Two views of the multi-layered regulation of bacterial glutamine synthetase by two-tiered nucleotidylation. (Left) The adenylation state of GlnA subunits is controlled by the balance of action of the GlnE adenylyltransferase in adenylation mode and the deadenylation mode, in turn controlled by the GlnD uridylyltransferase. (Right). The dodecameric glutamine synthetase (GlnA) can be adenylylated at any or all of its 12 subunits, modulated by the activity of the PII protein that in turn can undergo four uridylylations.

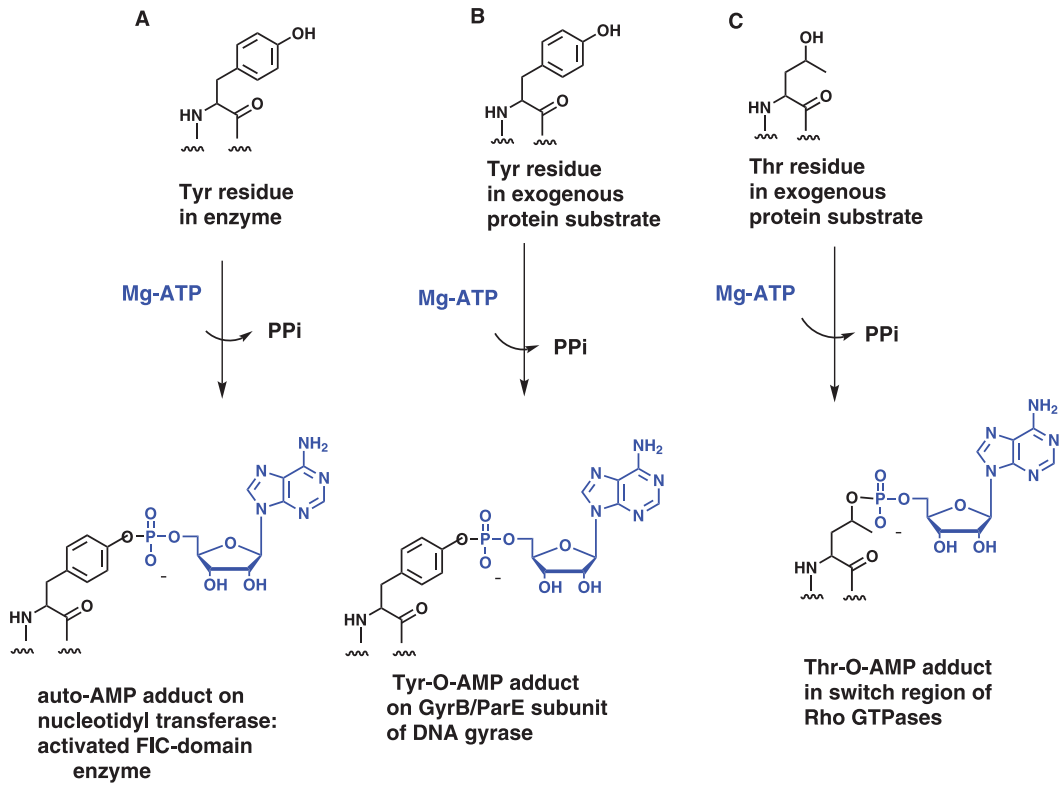


Figure 16.11 (A) Auto-AMPylation of FIC domains in proteins activates them, *e.g.* for bacterial filament formation; (B) among exogenous proteins AMPylated are tryosyl residues in the GyrB or ParE subunits of the type II bacterial DNA gyrases. Those are inactivating modifications and are distinct from the tyrosine residues that get covalently modified to 5′phosphate ends of DNA gaps during catalytic cycles of single or double strand DNA breaks. (C) The aliphatic alcohol side chain of a threonine residue in conditional RhoGTPases can undergo adenylylation (AMPylation) with resultant decrease in catalytic GTPase activity.

bound GDP, thereby blocking downstream signaling and disrupting interactions with actin. Analogously, the immunoglobulin-binding protein A (IbpA) protein of *H. somni* covalently adenylylates tyrosines in the switch I region of RhoA, Rac, and Cdc42 subfamily members, with similar cell rounding phenotypes, resulting from blocking of GTPase activity.[7–9]

Recently a pseudokinase, YidU, conserved from *E. coli* and *Salmonella typhimurium* to humans has been demonstrated to have two nucleotidyl transfer activities. One is AMPylation, both of itself, and proteins on Ser, Thr, or Tyr residues that turns down activity.[10] Variation in nucleotidyl transfer donor specificity for the Salmonella enzyme has indicated that UTP is recognized preferentially to ATP as a nucleotidyl donor. Among some 50–60 substrate proteins UMPylated were the major protein chaperones GroEL, DNAK, HtpG and ClpB in times of stress.[11] Uridylylation of tyrosine or *histidine* residues (UMP-phosphoramidates) are reminiscent of the logic of regulation of the bacterial glutamine synthetase system noted above. The UMPylated chaperones no longer bind client proteins to help shepherd them through folding processes. In turn less ATP is used to help such protein clients fold and[11] there is less stress-induced ATP depletion in those cells where YidU is overexpressed.

FIC homologs are found in eukaryotes but have not been examined deeply yet. It is known that a drosophila FIC adenylylates binding-immunoglobulin protein homolog (BIP), an endoplasmic reticulum-resident protein folding chaperone.[9]

In *E. coli* and *Sacchromyces cerevsiiae* the selenoprotein SelO, thought to be a pseudokinase because it lacks a crucial active site aspartate residue, shows AMPylation activity towards a number of proteins involved in oxidative phosphorylation.[12] Further structural and catalytic examination of both a pseudomonal and a human analog of that pseudokinase SelO demonstrated that indeed it was not active for phosphoryl transfer from Mg–ATP. Instead, the ATP substrate was flipped around in the active site such that Pα- rather than Pγ of bound ATP was attacked by alcoholic or phenolic side chains of threonine or tyrosine residues in proteins. The result is covalent AMPylation.[10] In humans a SelO homolog is mitochondrial in subcellular location and led to AMPylation of mitochondrial proteins involved in redox metabolism, including the E1 enzyme component of α-ketoglutarate dehydrogenase (Thr–O–AMP) and of glutaredoxin (Tyr–O–AMP). The authors suggested a reexamination of pseudokinases for AMPylation rather than phosphorylation catalytic activity towards protein side chains was in order.

To date, the extent of protein posttranslational nucleotidylations compared with protein phosphorylations is in the order of 1 : 10 000. While there may be more protein AMPylations to be discovered, it appears that in contrast to low molecular weight metabolism, posttranslational proteomics has not utilized nucleotidyl transfers as a common route for protein-based signal transductions.

16.4 Transfers of Phosphorylcholine to Proteins

Legionella strains have two types of Fic enzymes that they can inject into the cytoplasm of host cells. The defect in Rab1 (one of the small GTPase family

members) recruitment (Drr) protein transfers an AMP moiety from ATP to the tyrosyl 77 side chain phenolic oxygen in the switch II region of Rab1 GTPase, causing inactivation (Figure 16.12). An intriguing alternate activity of a separate FIC enzyme, ankyrin repeat-containing protein X (AnkX), from the same *Legionella* strain instead modifies the adjacent serine side chain-OH but not by AMP. Instead, a phosphorylcholine is attached in both Rab1 and Rab35 GTPase isoforms[10] (Figure 16.13).

Reasoning that CDP-choline is a common metabolically activated form of phosphorylcholine, the investigators determined that it was indeed the phosphorylcholine donor in this alternative posttranslational modification. One would formulate the enzymatic transfer as a regiospecific attack by the target protein-serine alkoxide anion in the transition state to attack the distal phosphate of CDP-choline. Subsequent study revealed that AnkX carries out *autophosphorylcholination* at Ser 521, Thr 620, and Thr 943, establishing chemospecificity to Ser and Thr side chains[13] (Figure 16.14). These authors propose a covalent phosphorylcholinyl-His 229 intermediate on AnkX, indicating phosphoramidate chemistry (see Chapter 10) in play

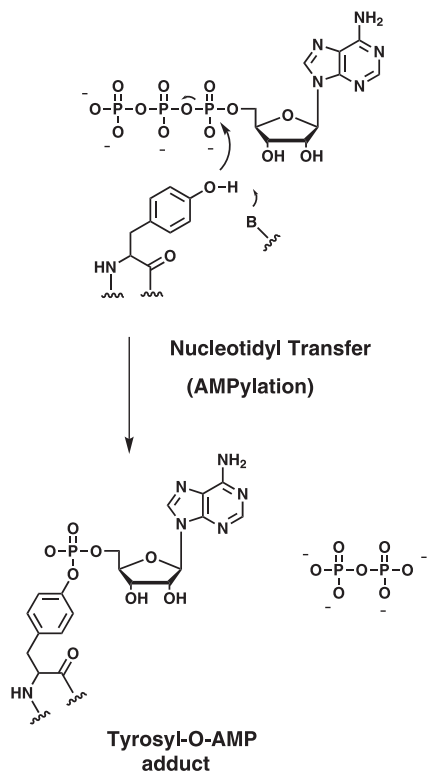


Figure 16.12 Rab1 unusual modifications by bacterial pathogen enzymes. *Legionella* enzyme Drr is an AMPylase for Tyr 177 of the Rab1 GTPase.

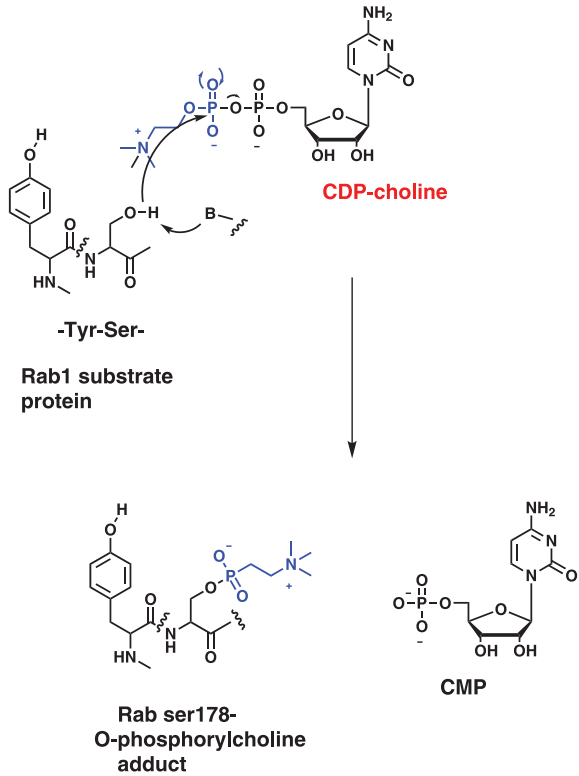


Figure 16.13 Rab1 unusual modifications by bacterial pathogen enzymes. The *Legionella* AnkX instead transfers a choline phosphate moiety to the adjacent Ser 178 side chain.

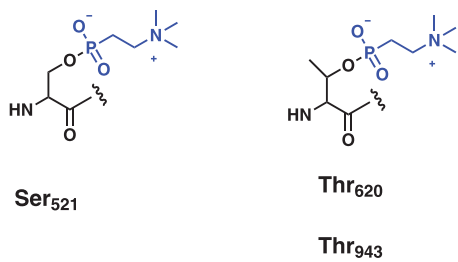


Figure 16.14 Rab1 unusual modifications by bacterial pathogen enzymes. AnkX automodifies up to three of its own side chains with a phosphorylcholine diester linkage.

during phosphoryl transfer to Rab1 (Figure 16.15). Finally, a Legionella protein can remove three phosphocholine groups from Rab, presumably acting as a phosphodiesterase[14] (not shown).

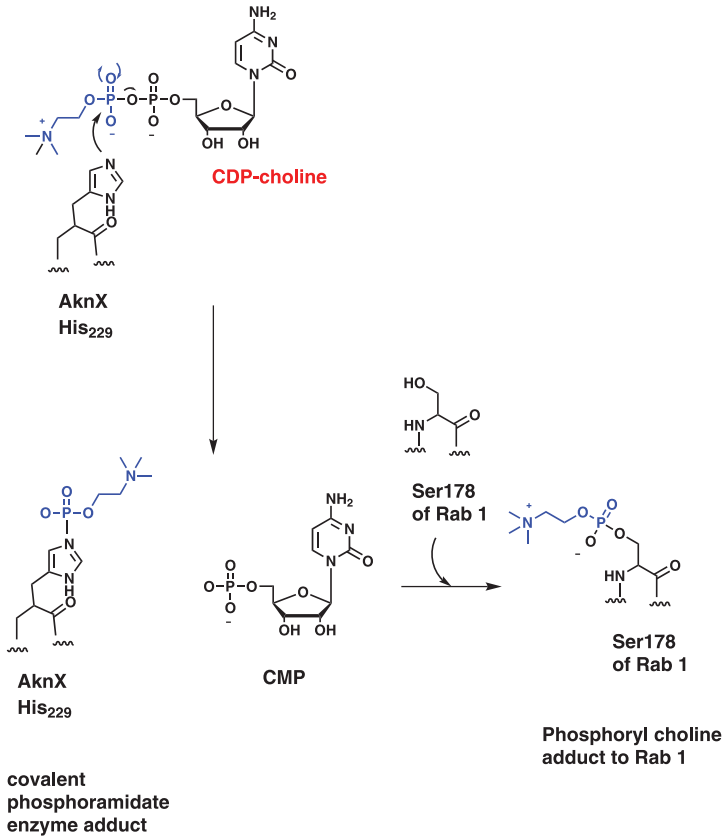


Figure 16.15 Rab1 unusual modifications by bacterial pathogen enzymes. AnkX first transfers the phosphorylcholine moiety from CDP–choline to an active site histidine before transfer onto Ser 178 of Rab1.

Typical nucleotidyl (cytidylyl) transfers would have seen attack on the proximal phosphorus of CDP-choline *not the distal phosphorus observed for AnkX*. It remains to be seen if there are other examples of such phosphorylcholine phosphodiester formations by other FIC enzymes on the small GTPase superfamily members and what the balance between adenylyl transfer from ATP *vs.* phosphorylcholine transfer from CDP-choline turns out to be.

16.5 Covalent Tethering of Atg8 to Autophagosomal Membranes

Separate from phosphoproteomics is an extensive category of lipidated proteomics in eukaryotic cells. The two major types of lipid protein covalent

posttranslational linkages are (1) long chain fatty acids in either amide (N-terminal glycine or side chain lysine amine) or thioester linkages (cysteine thiolate), or (2) prenylated thioethers (cysteine thiolate as nucleophile).[15]

An alternate strategy for lipidation and directed location to an internal cell membrane occurs for the protein autophagy 8 (Atg8), required for autophagy.[16] The core of the ~14 kDa Atg8 resembles ubiquitin and that core is processed by Atg enzymes that mimic the ubiquitin cascade. The C-terminus of Atg8 undergoes precise limited proteolysis by Atg4 to uncover a C-terminal glycine. The new glycine C-terminal carboxylate acts as nucleophile in a classic nucleotidyl transfer to yield the Atg8–Gly–AMP mixed acyl phosphoric anhydride. This is converted to a covalent Atg8 thioester on a Cys residue of another Atg protein before undergoing thermodynamically favored capture by the amino group of a phosphatidylethanolamine molecule under the aegis of an Atg5–12–16 enzyme complex (Figure 16.16).

This lipidation in the form of a C-terminal amide between Atg8 and the phosphatidyl ethanolamine tethers Atg8 to newly forming autophagosomal membranes and enables assembly of other Atg protein components that result in maturation of the autophagosome. The logic and mechanism of attachment of the phosphatidylethanolamine head group to Atg8 is *mechanistically orthogonal* to the phosphocholination of Rab1 and Rab35 noted above. In the Atg8 case, the C-terminus in the thioester linkage is the electrophile. The amine of the membrane-localized phosphatidylethanolamine is the nucleophile. In the Rab1–Rab35 case above, the Ser–OH side chain is the nucleophile and CDP–choline is the electrophilic partner.

From a strict definition, neither of these two modifications fit into conventional phosphoproteomics by mechanism or constitution of the protein adducts. Yet, each has an anionic phosphodiester grouping covalently tethered to a protein side chain or C-terminus and illustrates the diversity of phosphorus posttranslational chemical biology. In such an expanded definition one could also fit any enzyme that cleaves NAD^+ in protein ADP ribosylations, both as mono-ADP ribose or poly-ADP ribosylations.[15] Although the diphosphate moiety of NAD^+ is a bystander in such posttranslational modifications, the newly introduced ADP ribosyl moieties have a dianionic pyrophosphate group (Figure 16.17). Furthermore, NAD^+ turns out to be the source of a novel phosphodiester crosslink in more than 180 mammalian proteins as detailed in the next section.

16.6 NAD^+ as a Donor of Electrophilic Fragments to Protein Side Chains: Proteomic Roles

Although the reduced forms of the nicotinamide coenzymes (NADH, NADPH) are rightly considered the mobile electron transfer currency for hundreds of two-electron redox reactions cellular metabolism, the oxidized form NAD^+ has an extensive set of non-redox roles in posttranslational modification of proteins.[17] Inspection of the structure of oxidized NAD^+ in

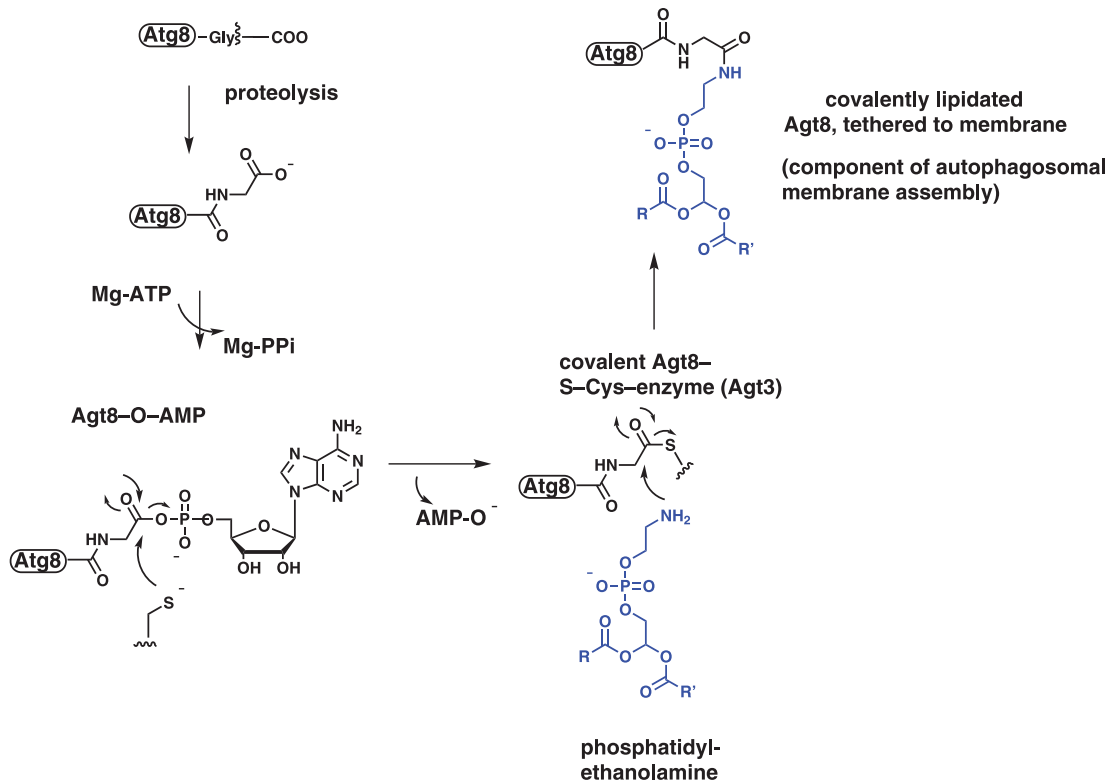


Figure 16.16 Atg8 is activated by ubiquitin type logic to form a C-terminal-AMP and then a C-terminal S–enzyme intermediate before transfer of the activated Agt8 protein moiety to the head group of a phosphatidylethanolamine lipid molecule. This is a key step in organellar apoptosis in cells.

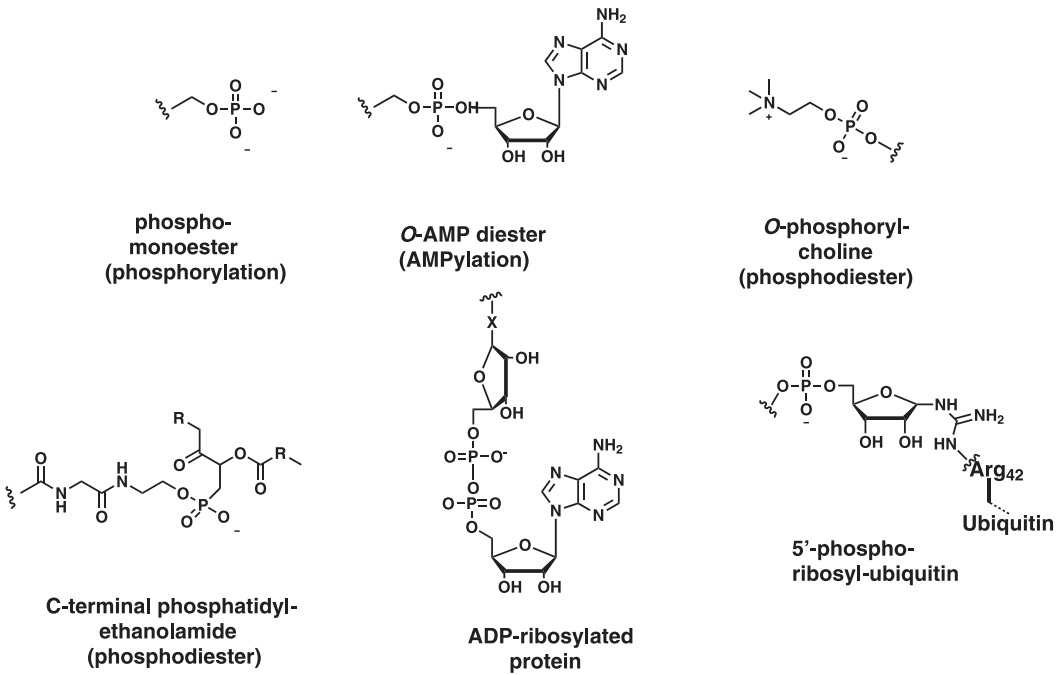


Figure 16.17 Six distinct posttranslational modifications containing one or more phosphate-derived group.

Figure 16.18 reveals three sites of enzyme-mediated fragmentation of NAD^+ with transfer of electrophilic fragments to nucleophilic side chains in proteins. The corresponding biosynthetic logic to NAD^+ assembly, utilizing three distinct types of phosphoryl group chemistry has been described earlier in Figure 4.24.

The first reaction arrow in Figure 16.18 points to C1 of the ribose moiety in the NMN^+ portion of the NAD^+ scaffold, reflecting the permanent charge on the adjacent quaternized nicotinamide ring in the oxidized state. That permanent charge makes C1 of that ribose moiety electrophilic, in analogy to the CH_3 group adjacent to the cationic S atom in *S*-adenosylmethionine. This C1–N^+ bond turns out to be the most labile for enzymatic reactions that generate fragmentation of NAD^+, transferring the ADP ribosyl moiety to cellular nucleophiles. As shown in Figure 16.19, typical protein side chains undergoing ADP ribosylation in target proteins include the thiolate anion of cysteine, the γ-carboxylate side chain of glutamyl residues, and, perhaps unexpectedly, the amidino side chain of arginine residues.

The second arrow in Figure 16.18 indicates the cleavage point of NAD^+ when it is a substrate for the subclass of DNA ligases that use NAD^+ rather than ATP as a nucleotidyl donor (see Chapter 6, Figure 6.13). That mode of

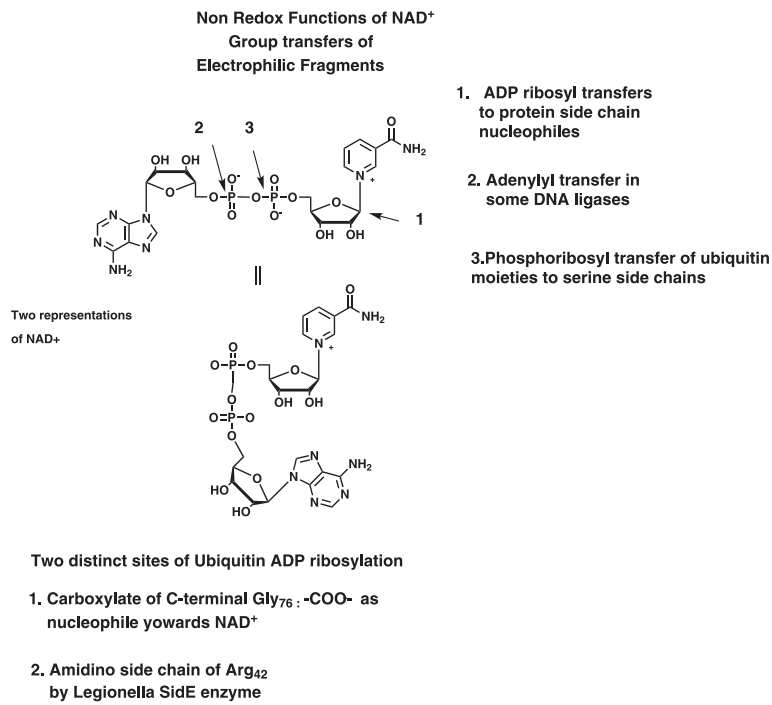
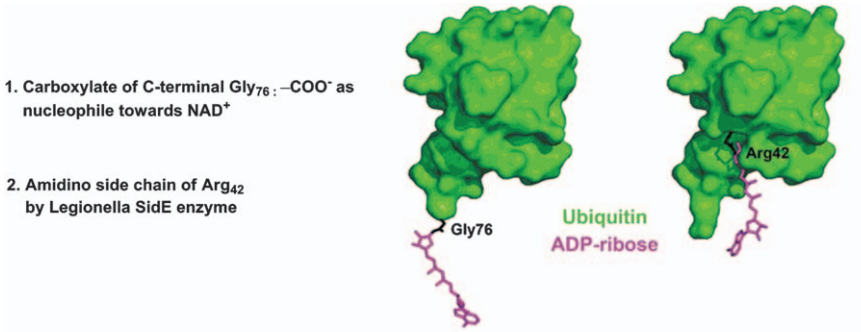


Figure 16.18 Three sites of enzymatic cleavage of NAD^+ in nonredox protein modifications. ADP ribosyl transfers, adenylations, and phosphoribosyl-ubiquitin transfers.



	Dtx3L (E3 ubiquitin ligase)/ Parp9 (ADP-ribosyltransferase)	SidE (ADP-ribosyltransferase)
Cofactor	NAD^+	NAD^+
ADP-ribosylation site	Glycine 76	Arginine 42
Effect on ubiquitin	Blocks conjugation	Alternative activation
Subsequent Processing	Hydrolase	SidE (phosphodiesterase)

Figure 16.19 Two distinct sites of ADP ribosylation on ubiquitin.

fragmentation utilizes the pyrophosphate bridge between the AMP and NMN^+ halves of NAD^+ as a *phosphoric anhydride* donor of the adenylyl moiety.

The third arrow indicates cleavage of that same pyrophosphate bridge at the other electrophilic phosphorus in the P–O–P anhydride functionality. In this case the phosphoribosyl moiety is transferred to a serine residue in a range of mammalian target enzymes. To examine this third mode of NAD^+ reactivity we take up the recent finding that this mode of cleavage is effected by a set of enzymes injected into host cells by the bacterial pathogen *Legionella pneumophila*. This also involves the mammalian posttranslational protein modifier ubiquitin and ADP ribosylation as part of this pathogen's host cell biology.

A large number of mono-ADP ribosylations and poly-ADP ribosylations have been reported and characterized for different biological effects in cellular contexts and are largely beyond the scope of discussion in this section. Notably though, ADP-ribosylation of the 76-residue mammalian protein ubiquitin has recently been discovered.

Ubiquitin itself, a 76-residue small helical protein, is widely used as an information-rich posttranslational modification covalent tag in a wide range of mammalian biology. Typically, ubiquitin is activated on its C-terminal glycine 76 carboxylate, first as an –AMP mixed carboxylic anhydride, then as an acyl thioester on two consecutive enzymes before undergoing transfer as a protein acyl group on to the side chains of lysine residues in hundreds of cellular protein targets. The resultant ubiquitinyl–Lys isopeptide bonds are stable to normal peptidases, although susceptible to hydrolysis by deubiquitinylating hydrolases.

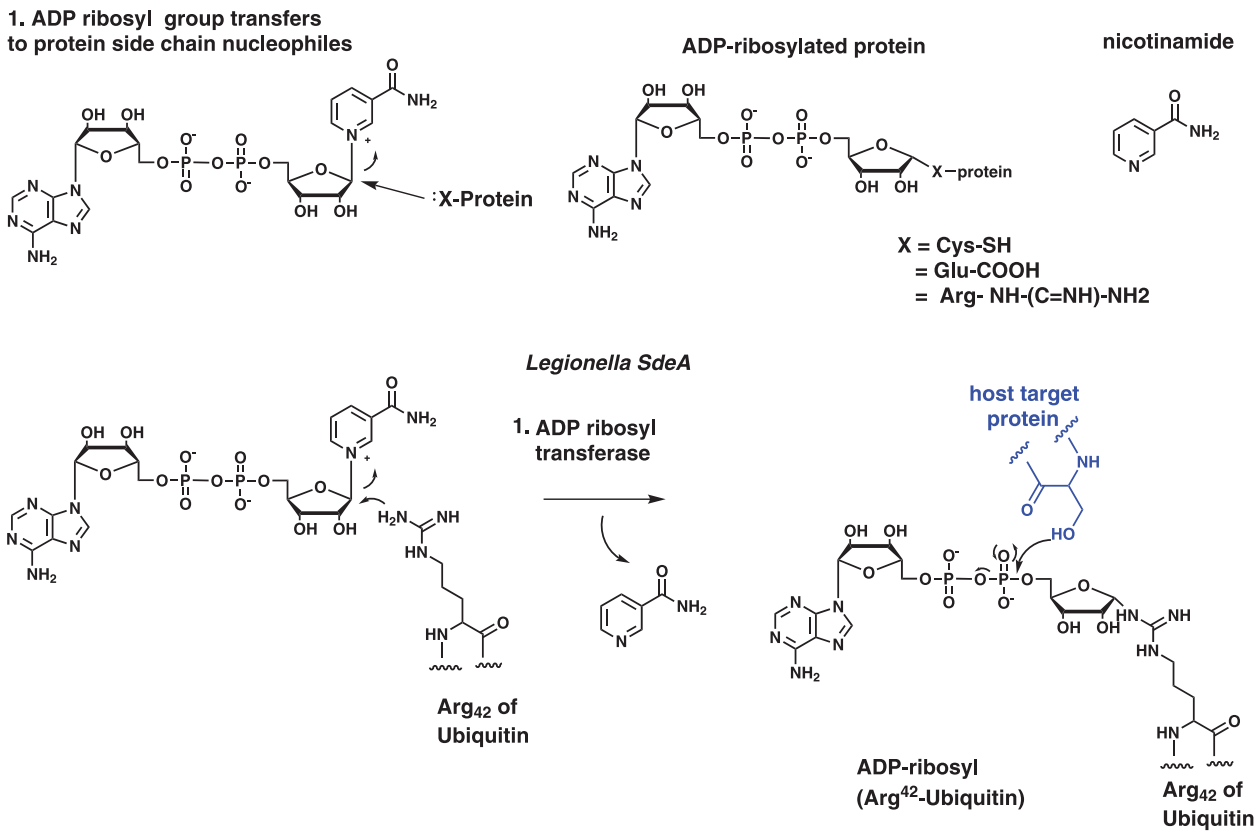


Figure 16.20 ADP ribosyl transfer to Arg 42 of ubiquitin by the *legionella* SdeA enzyme.

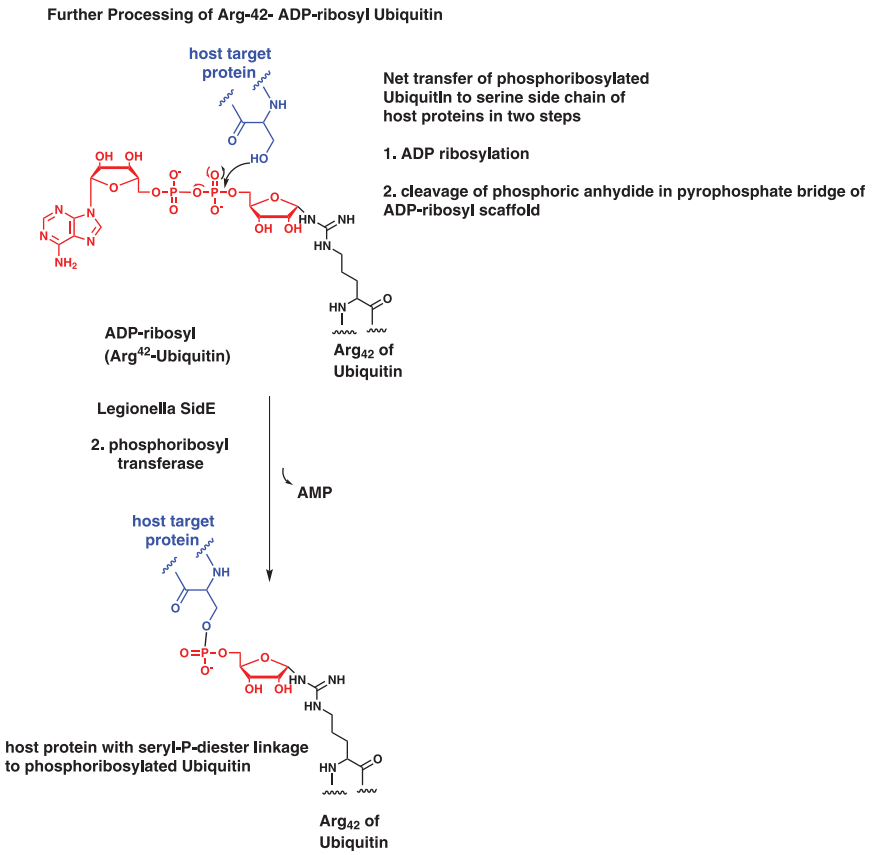


Figure 16.21 The second catalytic activity of SdeA involves release of AMP and transfer of a phosphoribosyl-ubiquitin to serine-OH side chains of host proteins.

Two regio-distinct ADP-ribosylations of ubiquitin have been observed as noted in Figure 16.19. The poly (ADP–ribose) polymerase 9 (PARP9) ADP ribosyltransferase modifies the C-terminal glycine 76 carboxylate, the same kind of regiochemistry seen in canonical ubiquitin activation enzymology. However, when *L. pneumophila* infects mammalian cells it uses NAD^+ to ADP-ribosylate the amidino side chain of Arg 42 in an ADP–ribosyl–Ub *aminal linkage*[18,19] (Figure 16.20). The *Legionella* enzyme is one of about 300 enzymes injected into host cells, some 10 of which disrupt normal ubiquitin biology.[20] The requisite ADP ribosyl transferase is SdeA, but SdeA is not finished when it has ADP-ribosylated ubiquitin.

It turns out that SdeA and its *Legionella* homologs are bifunctional enzymes. The second activity in an adjacent catalytic domain is structurally related to some phosphodieterases.[21] However, this second activity of SdeA is in fact the third arrow for NAD^+ cleavage in Figure 16.18. The

ADP–ribosyl–Arg 42–Ub is one substrate and any of ~180 mammalian host proteins are cosubstrates, each using a serine side chain as a nucleophile to transfer the phosphoribosyl-Ub fragment to the attacking host proteins. As shown in Figure 16.21, this action converts each of those serine side chain to phosphodiesters. The other substituent on phosphorus is the -ribosyl–Arg 42–Ub tag.[22,23] This second activity of SdeA is required for *the L. pneumophila* intracellular disease phenotype in infected human cells: Golgi complex fragmentation as the bacteria encase themselves in a newly created intracellular vacuole and multiply therein. Finally, *L. pneumophila* can also inject enzymes into host cells that act as specific phosphodiesterases, cleaving the Protein–Ser–P–ribose–Ub crosslink giving back native host protein and releasing the Ub–Arg 42_ribose-5-phosphate fragment.[24]

This is a new nonredox role for NAD^+ with a distinct regiospecificity of enzymatic cleavage of its pyrophosphate linker. An engulfed *L. pneumophila* bacterium injects hundreds of proteins into its mammalian host cell, among them the bifunctional SdeA that does a pair of tandem reactions on the NAD^+ scaffold: first an ADP ribosylation to ubiquitin on an unusual side chain and then the novel phosphoribosyl-Ub transfer to host proteins. This affects not just one protein but ~180 of them, opening up a new chapter in posttranslational protein modifications by phospho-ribosyl-ubiquitin frameworks. The second step is a special form of protein ubiquitination, likely to disrupt normal ubiquitin trafficking and signaling (*e.g.* the fragmentation of the Golgi complex in Legionella-infected cells).

References

1. L. M. Repka, J. R. Chejan, S. K. Nair and W. A. van der Donk, Mechanistic understanding of lanthipeptide biosynthetic enzymes, *Chem. Rev.*, 2017, **117**, 5457–5520.
2. X. Yang and W. van der Donk, Michael-type cyclizations in lantibiotic biosynthesis are reversible, *ACS Chem. Biol.*, 2015, **10**, 1234–1238.
3. S. Kodani, E. Hudson, M. C. Durrant, M. J. Buttner, J. R. Nodwell and J. M. Willey, The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene ramS in *Streptomyces coelicolo*r, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11448–11453.
4. A. Cubillos-Ruiz, J. Berta-Thomson, J. W. Becker, W. A. van der Donk and S. W. Chisholm, Evoutionary radiation of lanthipeptides in marine cyanobacteria, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E5424–E5433.
5. E. R. Stadtman, The story of glutamine synthetase regulation, *J. Biol. Chem.*, 2001, **276**, 44357–44364.
6. M. Y. Yarbrough and K. Orth, AMPylation as a new post-translational modifFICation, *Nat. Chem. Biol.*, 2009, **5**, 379–380.
7. A. K. Casey and K. Orth, Enzymes involved in AMPylation and deAMPylation, *Chem. Rev.*, 2017, **118**, 1199–1213.
8. M. C. Truttman and H. Ploegh, rAMPing up stress signaling: protein AMPylation in metazoans, *Trends Cell Biol.*, 2017, **27**, 608–620.

9. M. C. Truttmann, D. Pincus and H. Ploegh, Chaperone AMPylation modulates aggregation and toxicity of neurodegenerative disease-associated polypeptides, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E5008–E5017.
10. A. Sreelatha, S. S. Yee, V. A. Lopez, B. C. Park, L. N. KInch, S. Pilch, K. Servage, J. Zhang, J. Jiou, M. Karasiewicz-Urbanskka, M. Lobocka, N. V. Grishin, K. Orth, R. Kucharczyk, K. Pawlowski, D. R. Tomchick and V. S. Tagliabracci, Protein AMPylation by an evolutionarily conserved pseudokinase, *Cell*, 2018, 809–821.
11. Y. Yang, Y. Yue, C. Li, N. Song, Z. Yuan, Y. Wang, Y. Ma, H. Li, F. Zhang, W. Wang, H. Jia, P. Li, X. Li, H. Dong, L. Gu and B. Li, The YidU domain modulates bacterial stress through UMPylation, *BioRxiv*, 2019.
12. W. Wong, Flipping in a kinase fold for AMPylation, *Sci. Signaling*, 2018, **11**, eaav7869.
13. K. Gavriljuk, J. Schartner, H. Seidel, C. Dickhut, R. P. Zahedi, C. Hedberg, C. Kötting and K. Gerwert, Uraveling the phosphocholination mechanism of the *Legionella pneumophila* enzyme AnkX, *Biochemistry*, 2016, **55**, 4375–4385.
14. P. R. Goody, K. Heller, L. K. Oesterlin, M. P. Muller, A. Itzen and R. S. Goody, Reversible phosphocholination of Rab proteins by *Legionella pneumophila* effector proteins, *EMBO J.*, 2012, **31**, 1774–1784.
15. C. T. Walsh, *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, Roberts & Co, Englewood, Colorado, 2006.
16. T. Shpilka, H. Weidberg, S. Pietrokovski and Z. Elazar, Atg8: an autophagyprelated ubiquitn-like protein family, *Genome Biol.*, 2011, **12**, 226–237.
17. C. T. Walsh and Y. Tang, *The Chemical Biology of Human Vitamins*, Royal Society of Chemistry, Cambridge, 2018.
18. J. Qiu, M. J. Sheedlo, K. Yu, Y. Tan, E. S. Nakayasu, C. Das, X. Liu and Z. Q. Luo, Ubiquitination independent of E1 and E2 enzymes by bacterial effectors, *Nature*, 2016, **533**, 120–124.
19. D. Komander and F. Randow, Strange new world: bacteria catalyze ubiquitylation via ADP ribosylatioin, *Cell Host Microbe*, 2017, **21**, 127–129.
20. R. K. Sherwood and C. R. Roy, Autophagy evasion and endoplasmic reticulum subversion: the yin and yang of legionella invasion, *Annu. Rev. Microbiol.*, 2016, **70**, 413–433.
21. K. M. Kotewicz, V. Ramabadhran, N. Sjoblom, J. P. Vogel, E. Haenssler, M. Zhang, J. Behriner, R. A. Scheck and R. R. Isberg, A single legionella effector catalyzes a multistep ubiquitin pathway to rearrange tubular endoplasmic reticulum for replication, *Cell Host Microbe*, 2017, **21**, 160–181.
22. S. Bhogaraju, S. Kalayil, Y. Liu, F. Bonn, T. Colby, I. Matic and I. Dikic, Phosphoribosylation of ubiquitin promotes serine ubiquitination and impairs conventional ubiquitination, *Cell*, 2016, 1636–1649.

23. A. Akturk, D. J. Wasilko, X. Wu, Y. Liu, Y. Zhang, J. Qiu, Z.-Q. Lio, K. H. Reiter, P. S. Brzovic, R. E. Levit and Y. Mao, The mechanism of phosphoribosyl-ubiquitination mediated by a single legionella effector, *Nature*, 2018, **557**, 729–733.
24. M. Wan, A. G. Sulpizio, A. Akturk, W. H. J. Beck, M. Lanz, V. M. Faca, M. B. Smolka, J. P. Vogel and X. Mao, Deubiquitination of phosphoribosyl-ubiquitin conjugates by phosphodiesterase-domain-containing Legionella effectors, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 23518–23526.

CHAPTER 17

Broad Biological Arcs from Only Four Types of Phosphate Molecules

The vast sweep of phosphate chemistry in biological systems can be categorized by multiple criteria. One is by metabolite category as we shall summarize below. A second is by the economically small number of chemical forms in which phosphorus exerts its biologic effects.

17.1 The Simplicity of Phosphate Chemistry in Biological Systems: Inorganic Phosphate, Phosphoric Anhydrides, Phosphate Monoesters, Phosphate Diesters

This second categorization has been the backbone of this volume. The default compound is inorganic phosphoric acid, as its monoanionic and dianionic salts at physiological pH ranges. Phosphate ions are essentially redox inert in >99% of the metabolic niches of life, simplifying (and perhaps also restricting) the scope of phosphometabolite reactivities to P^{+5} oxidation state chemistry. Three derivatives of inorganic phosphate then make up >90% of the phosphate forms across the metabolic landscape (Figure 17.1). Two other notable transient phosphate metabolites derived from these core categories are acyl phosphates and enol-phosphates noted later as metabolic intermediates that donate $-PO_3^{2-}$ groups to ADP to form ATP in the glycolytic pathway.

All forms of biological phosphate compounds (except the xenobiotic phosophotriester insecticides and nerve toxins) are negatively charged in

Chemical Biology No. 13
The Chemical Biology of Phosphorus
By Christopher T. Walsh

Published by the Royal Society of Chemistry, www.rsc.org

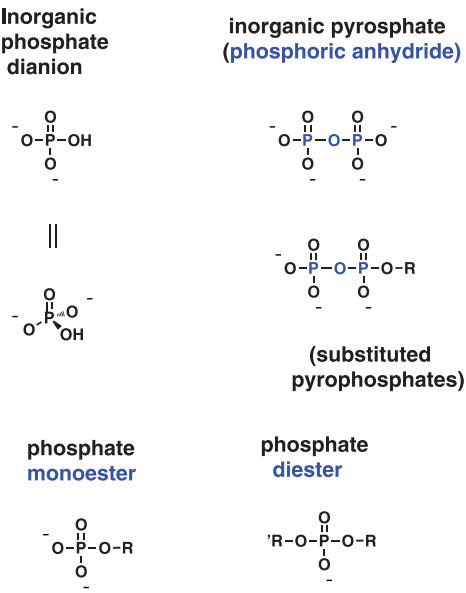


Figure 17.1 Inorganic phosphate and three other derivatives are the small set of phosphate forms that power metabolism.

biologic milieus. Those charges affect kinetic stability, and enable phosphates to function as double-headed, orthogonal reactants: nucleophilic at the peripheral oxyanions, electrophilic at the central phosphorus atoms.

Another essentially simplifying feature of phosphate chemical biology is that there are only a handful of phosphate-containing molecular frameworks, used over and over. There is inorganic phosphate itself, often limiting for prokaryotic growth and yet maintained at a tight concentration of around 1 mM in eukaryotic sera. The two biologically most important derivatives of phosphoric acid (and its anions) are phosphoric anhydrides and phosphate esters to alcohol(s), either monophosphate esters (one ROH) or phosphodiesters (ROH and R′OH).

Depending on one's biological interest, one might advance phosphoric anhydrides (ATP and other NTPS), phosphate monoesters (protein kinases and phosphoproteomics), or phosphodiesters (RNA and DNA) ahead of the other two phosphate derivatives. They represent as few as three central facets of chemical logic to power most of metabolism, growth, and biological information transfer.

Phosphoric anhydrides (from the simple inorganic pyrophosphate to the nucleoside triphosphate building blocks of RNA and DNA) ultimately power every category of biomacromolecular synthesis. In the specific guise of ATP, they serve as ion pumps in every cell type. The astounding calculation that organisms make and fully utilize their body weight in ATP every day is a triumphant paean to the combined thermodynamic activation and

conjoined kinetic stability of phosphoric anhydrides as chemically central sources of biologic energy.

Phosphate monoesters are generated from the *ur*-energy source ATP by attack of cosubstrate alcohols on the terminal Pγ phosphoric anhydride group in the side chain of ATP (Figure 17.2). This is the default reaction in trapping glucose in every cell type. It is also the most common chemical reaction in the more than 200 types of protein posttranslational modifications. Protein kinases in their hundreds provide the protein-based signaling network that enable moment to moment coordination and integration of the thousands of tasks cells and tissues have to choreograph incessantly.

Phosphodiesters can arise by two tandem phosphorylation events from two ATP molecules. However, this logic is dwarfed in the biosynthetic pathways for DNA, RNA, proteins, polysaccharides, and membrane phospholipids by phosphodioester formation by way of nucleotidyl transfers (Figure 17.2). These, like phosphoryl transfers above, rely on energetically favored ATP side chain phosphoric anhydride cleavage. The pattern of fragmentation is distinct: Cosubstrate nucleophiles, most typically R-OH groups, are directed to attack Pα, not Pγ, in the ATP side chain. The nucleophile gets adenylated when ATP is cosubstrate or guanylated (GMP transfer), uridylylated (UMP), or cytidylylated (CMP). There are multiple and chemically rich consequences of nucleotidyl transfer logic and machinery for building phosphodiesters, none more central than making RNA and DNA. It still seems astoundingly simple that the sole covalent

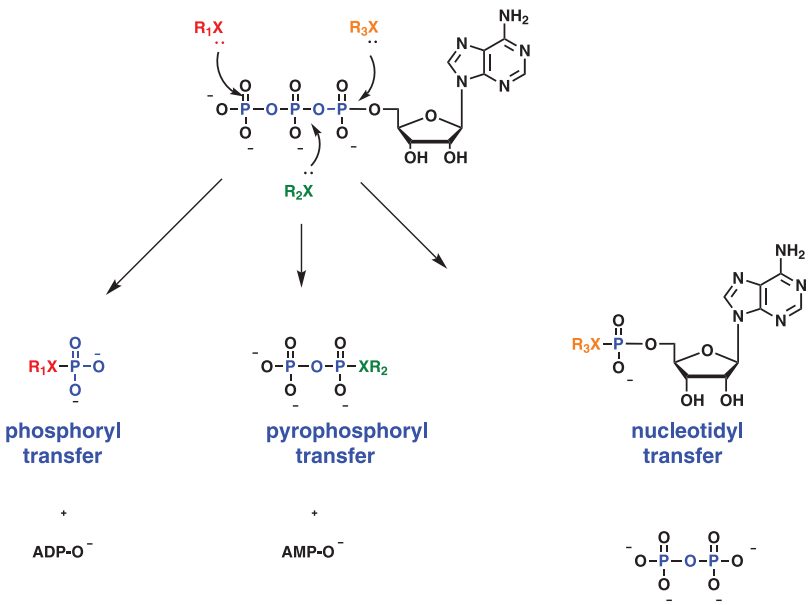


Figure 17.2 Two phosphoric anhydride linkages in ATP enable three modes of triphosphate side chain fragmentation: phosphoryl transfers, pyrophosphoryl transfers, and nucleotidyl transfers.

bonds connecting the millions of monomer (NMP or 2′dNMP) units in long chain information-rich RNA and DNA are the resultant internucleotide phosphodiesters. The use of these (*always monoanionic*) internucleotide phosophoidiesters has many ramifications, stretching from evolution, differential chemical stability of DNA *vs.* RNA, replication, repair, editing, and splicing. Nucleic acid biology in all its glory and complexity is enabled, directed, and constrained by the chemistry available to those monoanionic internucleotide phosphodiester bonds.

17.2 Phosphate Involvement by Metabolite Category

17.2.1 Phosphate Involvement in Low Molecular Weight Metabolic Pathways

17.2.1.1 Carbohydrates

Carbohydrates are perhaps the highest visibility metabolic class where monomers, largely neutral sugars, are routinely phosphorylated, converting neutral polyols, passively diffusible through membranes, into membrane impermeant dianions. These are largely sugar monophosphate esters formed in classical phosphoryl transfers from ATP, but can be transformed to partition glucose-phosphates to distinct metabolic fates.

The most celebrated processing route of glycolysis reveals the logic and mechanisms for converting the ‘‘low energy’’ glucose-6-phosphate monoester into two notable activated phosphoryl groups (Figure 17.3). The first is an acyl phosphate – mixed acyl phosphoric anhydride (1,3-diphosphglycerate) – and the second an enol phosphate (PEP) as thermodynamically activated, kinetic stable phosphoryl agents to drive ATP synthesis.

The phosphoryl group on glucose acts as a chemical traffic cop for routing glucose between glycolysis, the pentose phosphate pathway, and oligosaccharide and polysaccharide biogenesis.

The interconversion of glucose-6-phosphate to glucose-1-phosphate *via* transient 1,6-diphospho intermediate is an inventive enzymatic turnstile mechanism.

The glucose-1-phosphate is used then as a *nucleophilic phosphate* equivalent, a reversal of reactivity from the electrophilic phosphoryl group transfers. This reactivity sets in motion nucleotidyl transfers to create nucleoside diphosphosugars as sources of C1–glucosyl oxocarbenium ion equivalents for all biological glycosyl transfers (such that *all* glycosyl units are connected to what had been the cosubstrate nucleophile at C1). Analysis of the underlying chemical logic is that the net conversion of, for example, glucose-6-phosphate to glucose-1-UDP, has set up low energy cleavage of the C1–O bond, lowering the barrier for transfer of glucosyl C1 oxocarbenium equivalents into the biological universe.

Finally, a quintessential strategy for mobilizing glucosyl units, from starch, glycogen, cellulose and other glycans, is enzyme-mediated phosphorolysis.

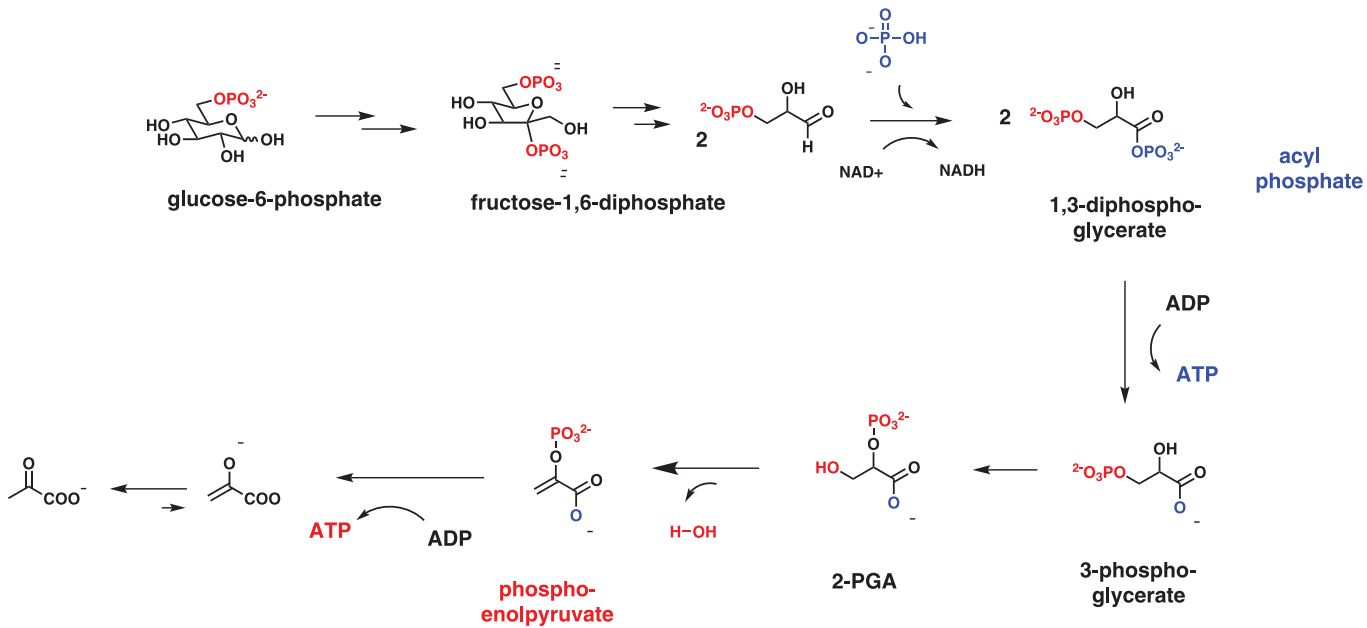


Figure 17.3 The phosphate monoester in glucose-6-phosphate is converted to two activated phosphoryl group during glycolysis: an acyl phosphate and a phosphorylated, trapped enol.

Inorganic phosphate acts as a nucleophile, *via* one of its anionic oxygen substituents, towards the glucosyl–C1–oxocarbenium ions in transition states to give the sugar-1-1-phosphates. Phosphohexose mutases in their reverse turnstile catalytic modes generate the 6-phosphoglucose regioisomers to feed into glycolysis for energy.

17.2.1.2 Nucleotides

Nucleotides are clearly at the intersection of ribose metabolism as building block on the one hand and DNA and RNA assembly as the downstream linear condensed products on the other hand. Several attributes of phosphate chemical logic are exercised as glucose scaffolds are funneled down its third metabolic pathway to ribose and then nucleic acids. On the order of tens of billions of 6-phosphoglucose molecules per cell cycle get oxidized at C1, then at C3 to set up decarboxylation to D-ribose-5-phosphate. This is a major flux during the DNA synthesis part of the cell cycle (S phase) and at all times when RNAs are being made.

The *de novo* biosynthesis of both pyrimidine and purine nucleotides from 5-phosphoribose utilizes the third mode of ATP fragmentation to drive biosynthetic equilibria, pyrophosphoryl transfers (Figure 17.2). The 5-phospho-ribose-1-pyrophosphate (PRPP, reflecting two of the three modes of ATP cleavage) molecules are activated at C1 as electrophilic sites for subsequent displacement by amine nucleophiles so both purine and pyrimidine-5′-phosphates are assembled this way.

To go from nucleoside-5′-phosphates that emerge from the purine and pyrimidine pathways to the triphosphates, two consecutive sets of kinase-mediated phosphoryl transfers occur. These build up pyrophosphate and then triphosphate side chains in the NDP and NTP products. Those are energy neutral *phosphoric anhydride interchange r*eactions, not phosphate ester transformations.

17.2.1.3 Biomacromolecular Condensed Linear Polymer Growth: Driven Thermodynamically and Mechanistically by Nucleotidyl Transfers.

Much of chapter four dealt with the common logic that the assembly of dehydrated, condensed linear biopolymers – RNA, DNA, proteins, and oligosaccharides – are driven by coupled nucleotidyl transfers that ultimately end in net hydrolysis of the NTP to NMP (captured in the growing RNA chain) and released PPi.

The DNA and RNA chains, the proteins, and the polysaccharides are all accumulating against unfavorable equilibria. The combination of nucleotidyl transfer logic that releases PPi as a phosphoric anhydride linkage is utilized, and inorganic pyrophosphatase in cells, means that NTPs are cleaved *de facto* into an NMP fragment and two PPis. Two phosphoric anhydride linkages are

spent in each elongation step to push the otherwise unfavorable equilibria of biopolymer accumulations.

The nucleotidyl transfers in the macromolecular biosynthetic and membrane phospholipid classes are in the context of distinct types of cosubstrate scaffolds. At first glance they look to be quite disparate processes with dramatically different functional group chemical biology in play. However, all four high flux pathways are united by underlying nucleotidyl transfer logic, even though only the nucleic acids have incorporated the NMP or dNMP moiety in flight in each reactant class (Figures 17.4–17.7).

In RNA and DNA polymerase action the NMP or dNMP moiety from the triphosphate substrates are incorporated *permanently* in the growing chains *via* direct formation of each internucleotide phosphodiester bond. In protein biosynthesis, each of the 20 proteinogenic amino acid carboxylates react as oxygen nucleophiles to form the mixed aminoacyl–AMP anhydrides. These are not the final products because of kinetic instability with release of AMP. Indeed, they are one step away from the stable amide bonds in proteins by way of mobile aminoacyl–oxoester linkages in the aminoacyl-tRNAs. Each ATP utilized is split to AMP and PPi.

For oligosaccharide, polysaccharide, *O*- and *N*-glycoprotein arrays, yet a third context is operant. A sugar-1-phosphate, *e.g.* glucose-1-phosphate, *via* one of that phosphate's oxyanions, is the nucleophile towards UTP for the enzyme-mediated uridylyl (nucleotidyl) transfer step. This deploys a third kind of oxygen nucleophile (3′–OH for RNA and DNA, amino acid carboxylate for protein biosynthesis, here the phosphate monoester oxygen). The result of this nucleotidyl transfer is a sugar-1–P–*P–ribose–uracil* = UDP glucose. We have noted that such UDP sugars are then used as chemical reagents for offering up C1-glucosyl oxocarbenium ion equivalents, *e.g.* to C4–OH of oligosaccharide chains in forming the characteristic linear condensed chains of starch, glycogen, and cellulose.

17.2.1.4 Membrane Phospholipids

The fourth major biosynthetic metabolite class involving phosphoryl and/or nucleotidyl group transfer chemical logic and enzymatic machinery are not condensed informational biopolymers. They are the amphiphilic diacylglycerol-based phospholipid diesters that in their hundreds of millions and billions of individual monomer molecules make up the lipid biolayers of cellular membranes. These can still be assembled by nucleotidyl transfer logic, *e.g.* CDP–choline and diacyl glycerols as substrates. The product – phosphatidylcholine – contains one of the phosphate groups from CDP choline but the CMP moiety has been released as free product (a smoking gun for cytidylyl transfer?). The membrane phosphatidyl lipids are phosphodiesters, with hydrophobic acyl tails and hydrophilic monoanionic phosphodiester groups. The amphiphilic two portions, hydrophobic tails and polar head groups, form the physical basis for self-assembly into lipid membrane bilayers.

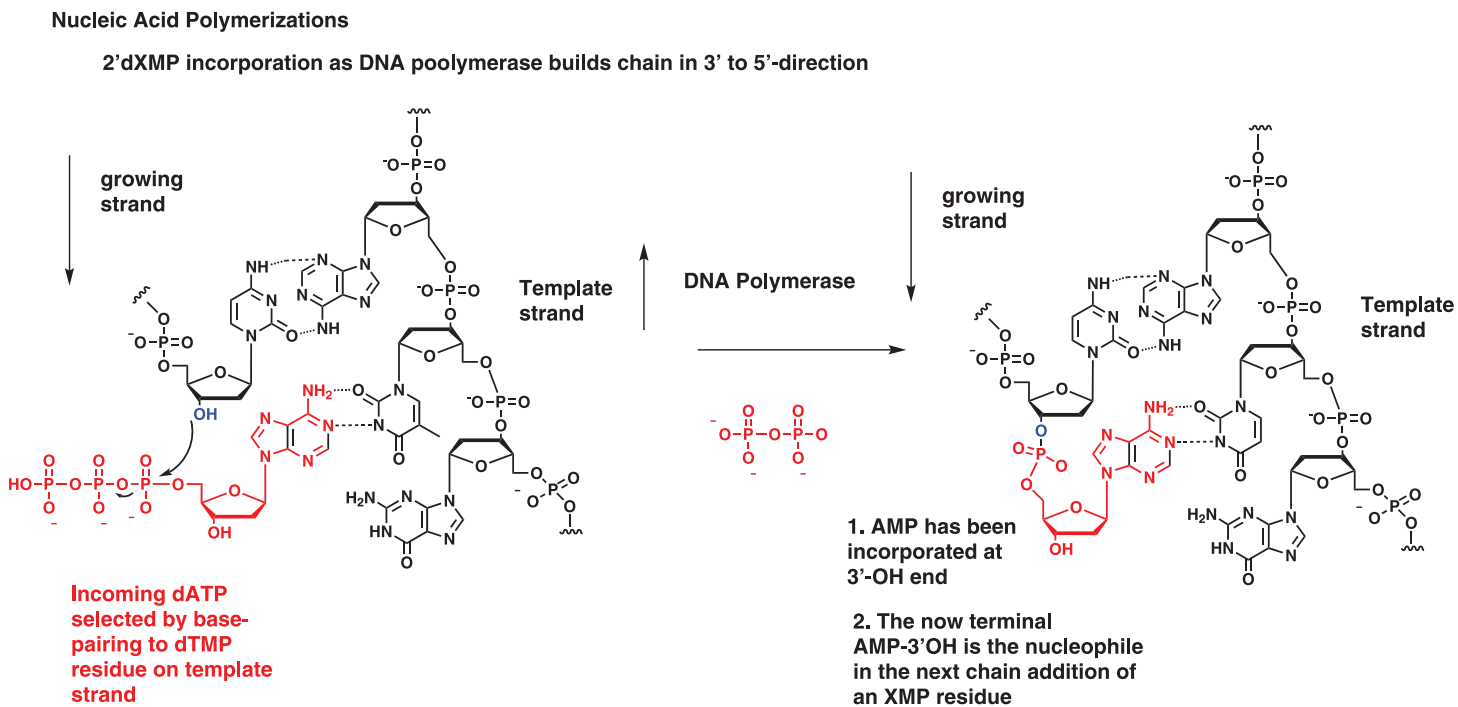


Figure 17.4 Nucleotidyl transfer is the common biosynthetic logic in activation of monomers for RNA and DNA: examination of the fate of the nucleotidyl groups in flight.

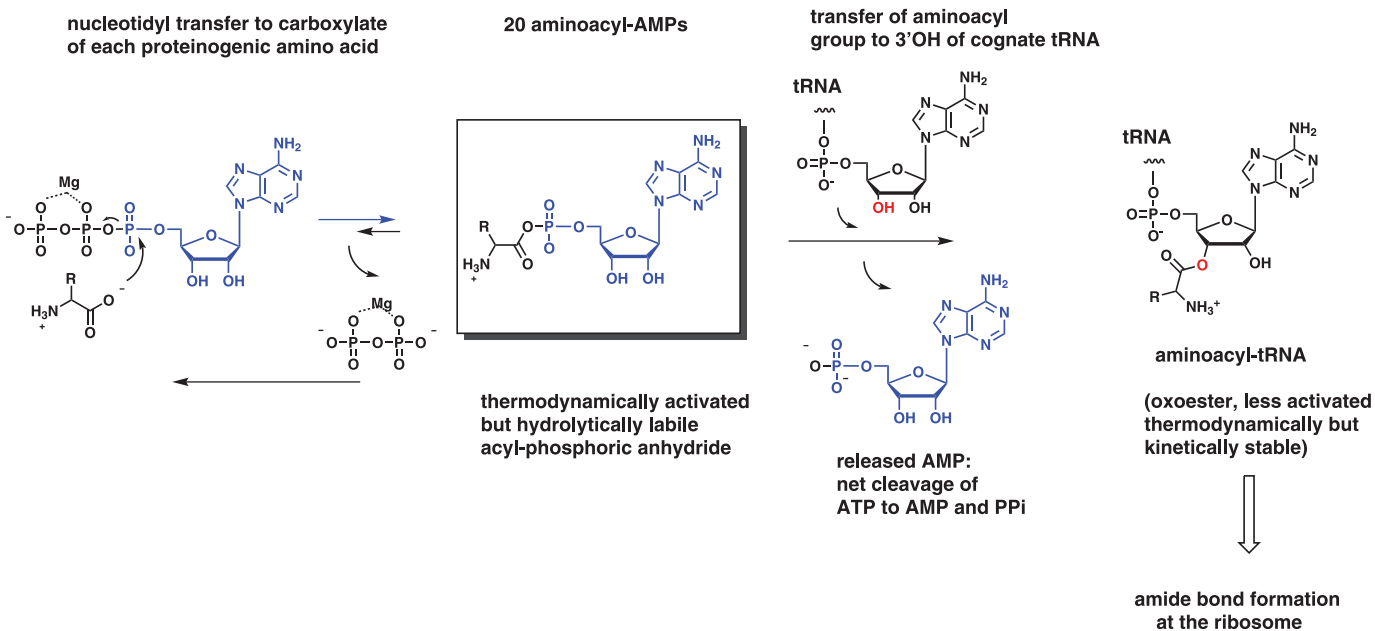


Figure 17.5 Nucleotidyl transfer is the common biosynthetic logic in activation of monomers for proteins: examination of the fate of the nucleotidyl groups in flight.

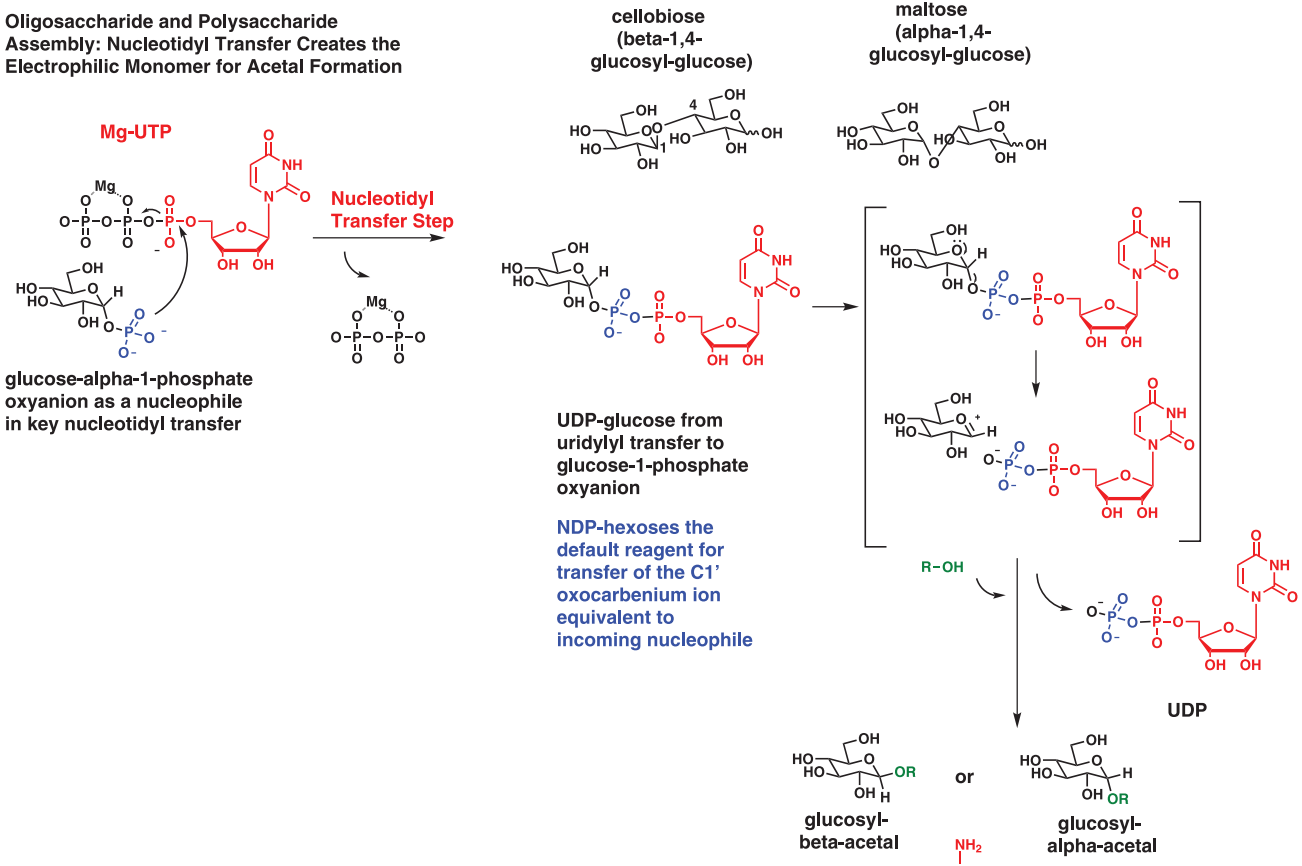


Figure 17.6 Nucleotidyl transfer is the common biosynthetic logic in activation of monomers for polysaccharides: examination of the fate of the nucleotidyl groups in flight.

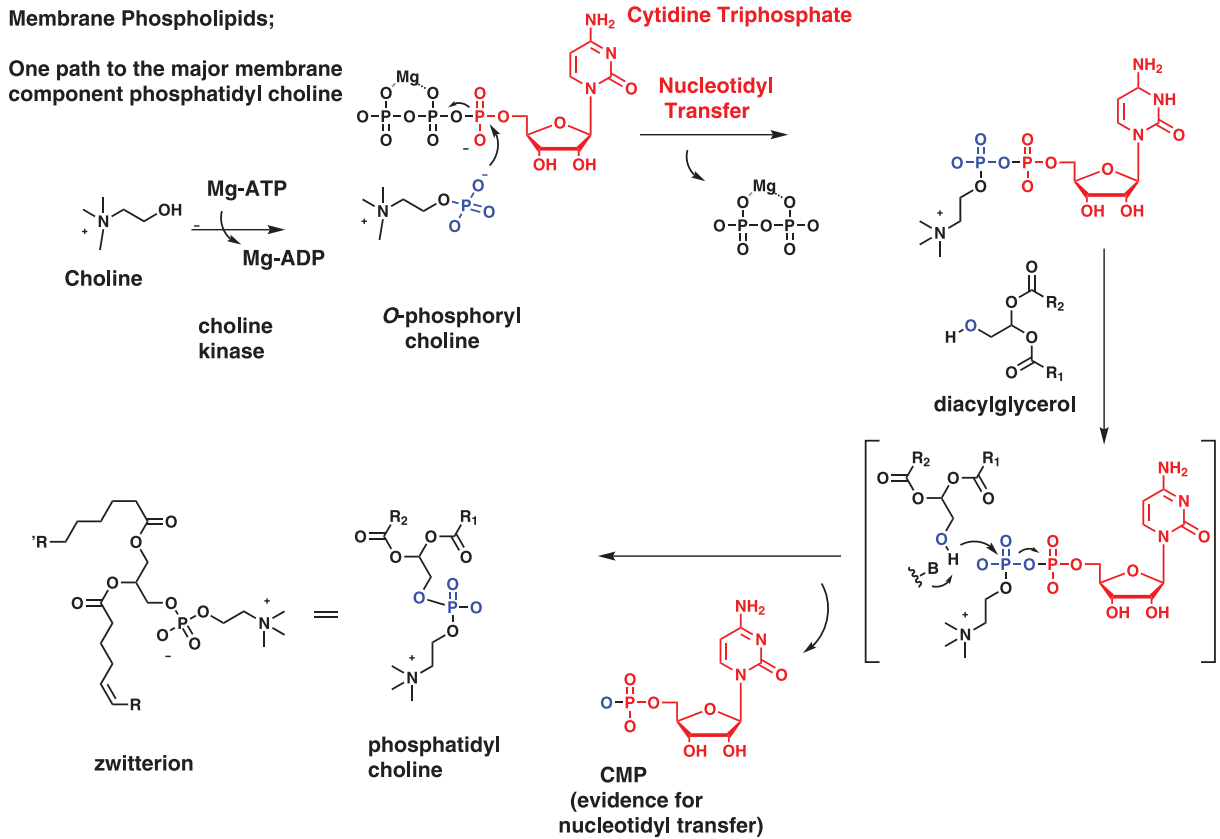


Figure 17.7 Nucleotidyl transfer is the common biosynthetic logic in activation of monomers for membrane phospholipids: examination of the fate of the nucleotidyl groups in flight.

17.2.2 Protein Posttranslational Modifications: Stable *vs.* Transient Phosphoproteomics

If one woke up a biologist from a sound sleep and uttered the word phosphate biochemistry, intelligible replies would surely vary but ''phosphoproteomics'' would most likely dominate. The explosive reach of phosphoproteomics over the past four decades has uncovered far-reaching and intricate networks of protein-based sensing and signal integration in essentially every subcellular compartment. Up to ~250 000 P-Ser, P-Thr and P-Tyr residues are predicted in human phosphoproteomes. This all happens with one type of substrate ATP, acted on by ~520 protein kinases in the human kinome, provides its electrophilic γ–PO_3^{2-} to any of the three -OH side chains of Ser, Thr, or Tyr residues of target proteins. Chemically stable phosphate monoesters are the products and require the action of phosphate monoesterases (phosphatases) for reversal. The enzyme chemistry is almost unitary, albeit protein kinases tend to be highly chemoselective for the aliphatic Ser–OH and Thr–OH *vs.* the phenolic –OH of Tyr residues. This is a high level and oversimplified description of what has become canonical phosphoproteomics (Figure 17.8A)

17.2.3 Protein Posttranslational Modifications: Canonical *vs* Noncanonical Phosphoproteomics

We (and others) have distinguished these canonically stable phosphoproteomes from *transient noncanonical phosphoproteomes*. Of the 20 proteinogenic amino acids, nine have potentially nucleophilic side chains. Above and beyond the three Ser, Thr, and Tyr-alcohol groups, the nitrogens of His, Lys, and Arg side chains, the carboxylate side chain oxygens of Asp and Glu, and the thiolate anion of Cys residues are nucleophilic candidates for enzymatic phosphoryl transfers. They rarely lead to stable phosphoproteins, consistent with the expected thermodynamic activation of –N–PO_3^{2-} phosphoramidates, –$COOPO_3^{2-}$ mixed acyl phosphoric anhydrides, and –S–PO_3^{2-} phosphorothioate linkages (Figure 17.8B). It is likely that an updated phosphoproteomics version 2.0 should encompass *both* the thermodynamically activated transient phosphoproteome (Figure 17.5), many of which are covalent catalytic intermediates in ATP-consuming enzymes, *and* the deactivated stable phosphoproteome of P-Ser, P-Thr, and P-Tyr, chemically suited for stable transduction of protein-based signals.

Histidine autokinases and paired aspartyl phosphate receiver proteins dominate bacterial integration of external signals *via* such transient phosphoproteomics. In eukaryotes, this transient phosphoproteomics pair have been replaced by an explosive radiation of tyrosine autokinases and protein serine–threonine kinases, giving the canonical stable phosphoproteomes. In turn, that stability has led to the evolution of sets

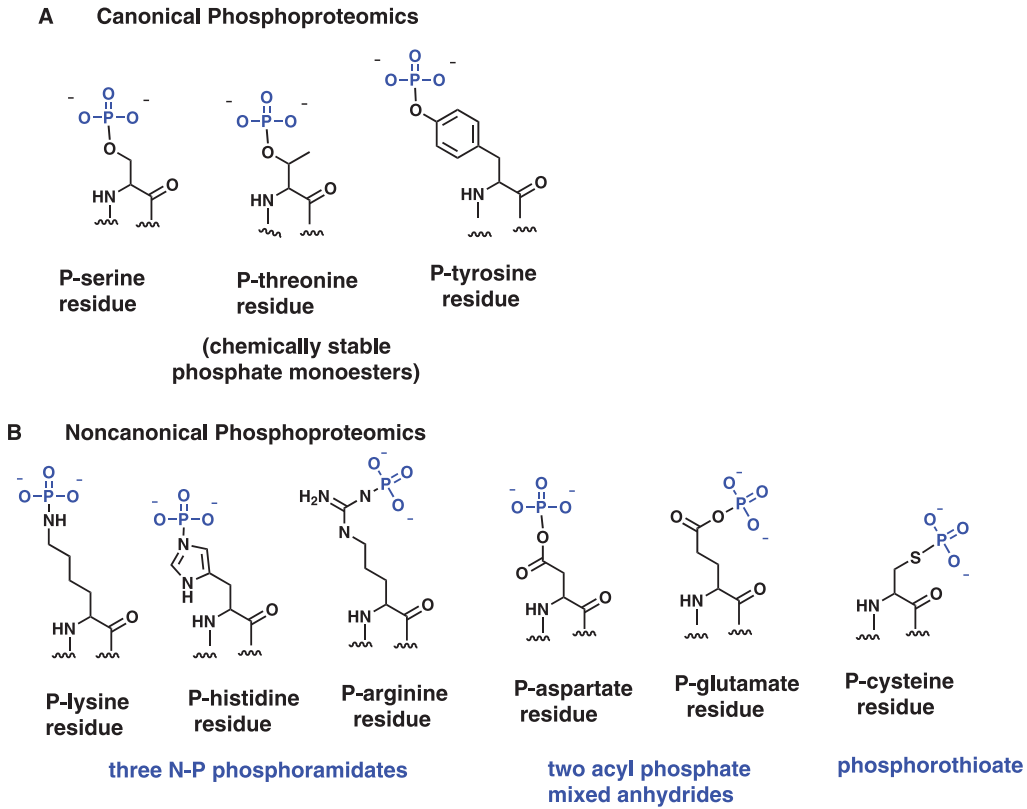


Figure 17.8 (A) Stable canonical *vs.* (B) Transient noncanonical phosphoproteomics: Nine nucleophilic amino acid side chains in proteins can be phosphorylated; Three of them are chemically stable as the canonical P-Ser, P-Thr, and P-Tyr phosphoproteomes.

of phosphoprotein-recognition partner protein domains and also families of both protein serine phosphatases and protein tyrosine phosphatases, features not required or present in transient phosphoproteomics networks.

17.3 Glucokinase Regiochemistry Connected to Directionality of RNA and DNA Polymerase Replication and Transcription

We conclude with a short vignette about the downstream fate of the phosphoryl group transferred from Pγ of ATP to the hydroxyl substituent at C6 of glucose, arguably the first step in capturing a key metabolite intracellularly. This is meant to illustrate the linkage of upstream and downstream cellular phosphoryl transfers and the imperatives the upstream paths lay out for the downstream catalysts; in this case hexokinase regiospecificity and directionality of DNA and RNA polymerase-catalyzed nucleic acid chain growth.

The first chemical event on the scaffold of glucose after it has been transferred across membranes into cells is enzymatic phosphorylation at the sole primary hydroxyl group at carbon six, available in both the major cyclic hemiacetal glucopyranose anomers and the minor open chain aldehyde form. The several isomers of mammalian hexokinases (glucokinases) all exhibit the same regiochemical specificity for the more reactive, less hindered C6–OH as do those bacterial phosphotransferase systems where the phosphoryl donor is PEP not ATP (Chapter 3).

This has been noted for decades as an electrostatic trap to keep the now anionic glucose-6-phosphate from passively diffusing across lipid plasma membranes and back out of cells. We also noted that the low energy phosphate monoester in glucose-6-phosphate is ultimately converted during the ten enzyme steps of glycolysis to the phosphorylated enol in PEP and thereby converted to a phosphoric ester (by enolase) thermodynamically activated for transfer to ADP [by pyruvate kinase (Chapter 6)].

In a very real sense this initial glucokinase regiochemistry for C_6-OH phosphorylation of glucose has ramifications that reach far downstream into the catalytic machineries of primary metabolism. For example, one of the three major metabolic fates of glucose-6-phosphate is transit through the pentose-phosphate pathway with 5-phosphoribose as a major endpoint metabolite. The first four enzymes of the pentose phosphate pathway oxidize 6-phospho-glucose at C1 from the hemiacetal to the lactone oxidation state, open it hydrolytically to the acid, oxidize C3 *via* NADP reduction, and then decarboxylate 6-phospho-3-keto-gluconate to D-ribulose-5-phosphate (Figure 17.9). A keto–aldo isomerase (ribulose-phosphate isomerase) then equilibrates the C-1,2-hydroxycarbonyl moiety between the ribulose ketose and the ribose-aldehyde form.

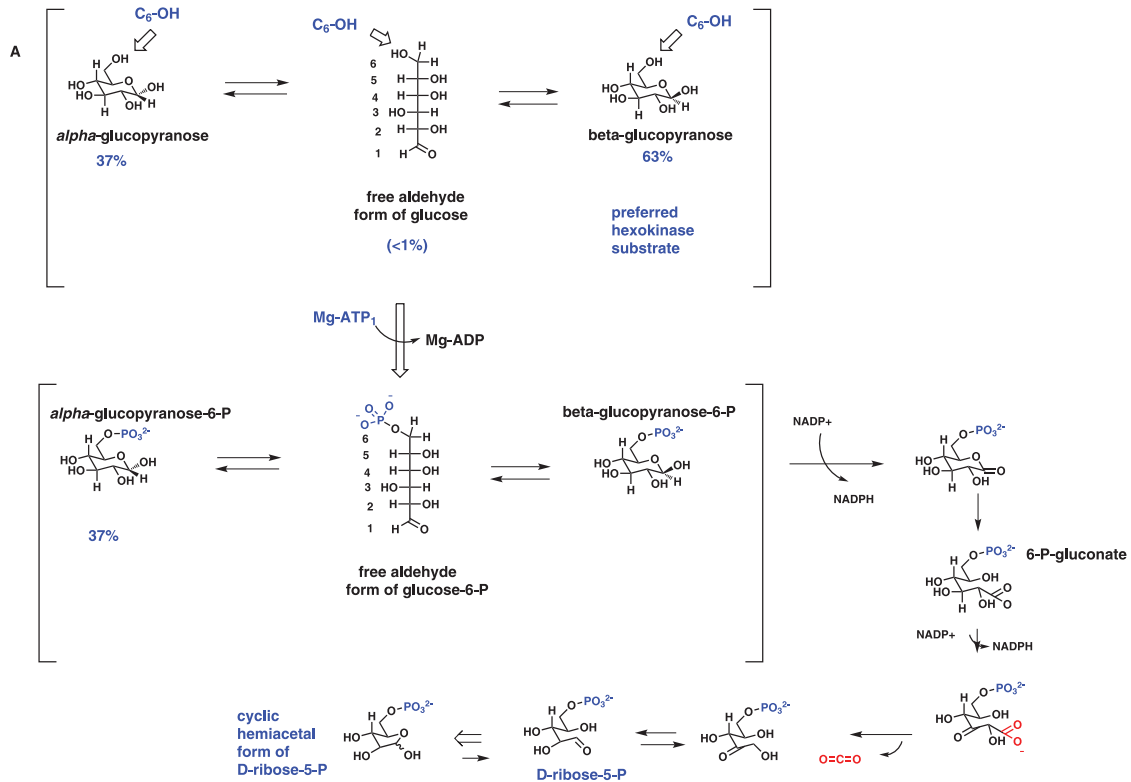


Figure 17.9 The regiochemical preference of hexokinase for C6 of glucose reads out in the 5′ to 3′ directionality of chain growth by RNA and DNA polymerases. Glucose-6-phosphate to ribulose-5-phosphate.

No change in the phosphate monoester group has occurred in this five-enzyme (pentose-phosphate) pathway from 6-phospho-glucose to 5-phospho-ribose. As noted in Chapter 5, the 5-phospho-ribose is then substrate for a unusual pyrophosphoryl transfer, with the α-C1–OH of the hemiacetal (furanose) form of 5-phospho-ribose attacking Pβ of Mg–ATP to give 5-phosphoribose-ribose-1-diphosphate (PRPP) that serves as the key building block for both purines and pyrimidines as well as the nicotinamide coenzymes. Thus, the pyrimidine scaffolds (uracil, and then to cytosine) and the purine scaffolds (inosine and then to adenine and guanine) are all constructed *de novo* on C1 of the 5-phospho-ribose framework (Figure 17.10). This biosynthetic chemistry leads inevitably to 5′-monophosphates (UMP, AMP, GMP) and then as noted in Chapter 3, by tandem action of two kinases up to the 5′-nucleoside triphosphates (Figure 17.11). The corresponding building blocks for DNA replication are processed as 5′-ribonucleoside diphosphates to 2′deoxy-5′-NDPs by ribonucleotide reductases.

Thus, the original regiochemical choice of glucokinases as the first entry step in glucose metabolism leads straightforwardly to 5′-NMPs and 5′-NTPs and 2′-deoxy-5′-NTPS. The original $-PO_3^{2-}$ group on glucose-6-phosphate has become the α-phosphoryl group of NMPs and NTPs or dNTPs. There is no similar widespread enzymatic path to 3′NMPs, NTPs, and 2′deoxy-3′-NTPs, and they are not common metabolites in cells. DNA polymerases and RNA polymerases have presumably evolved and adapted to this existing pool of activated monomers for chain elongation strategies.

To build the elongating chains with the 3′,5′-phospohodiester as the sole covalent linkage that anchors the NMPs and 2′dNMP units together in the long RNA and DNA chains, the polymerases use the 5′-NTP and 5′-dNTP building blocks: monomers as electrophiles at phosphorus and the 3′-OH of the nucleic acid chain termini as nucleophiles. Each such attack releases PPi and incorporates the NMP and dNMP monomer into the growing chain (Figure 17.12). The phosphate group coming in with each NMP is the one put on by glucokinases in the first place. Chain growth of DNA and RNA can only proceed in the 5′ to 3′ direction because the pool of activatable monomers are 5′-NTPs and dNTPs not the 3′NTPs and dNTPS.

This is a tidy and chemically logical connection between phosphoryl group trapping of glucose inside cells and its metabolic morphing into the phosphodiester group that links each monomer in RNA (and DNA) chains. The hexokinase or glucokinase regiochemical preference for the lone primary hydroxyl group of glucose, leading to accumulating 6-phospho-glucose, ensures that 5-phospho-ribose is the pentose-phosphate that is available for *de novo* assembly of purines and pyrimidines. Chain extension to the 5′ NTPs or dNTPS then dictates the chemical logic for directionality of RNA and DNA polymerases. Four ATPs (labeled ATP_{1-4} in Figures 17.9–17.11) are consumed between glucose and any RNA (mRNA, tRNA, rRNA, non-coding RNAs, . . .) chain elongation event. Three of the four ATPs undergo phosphoryl

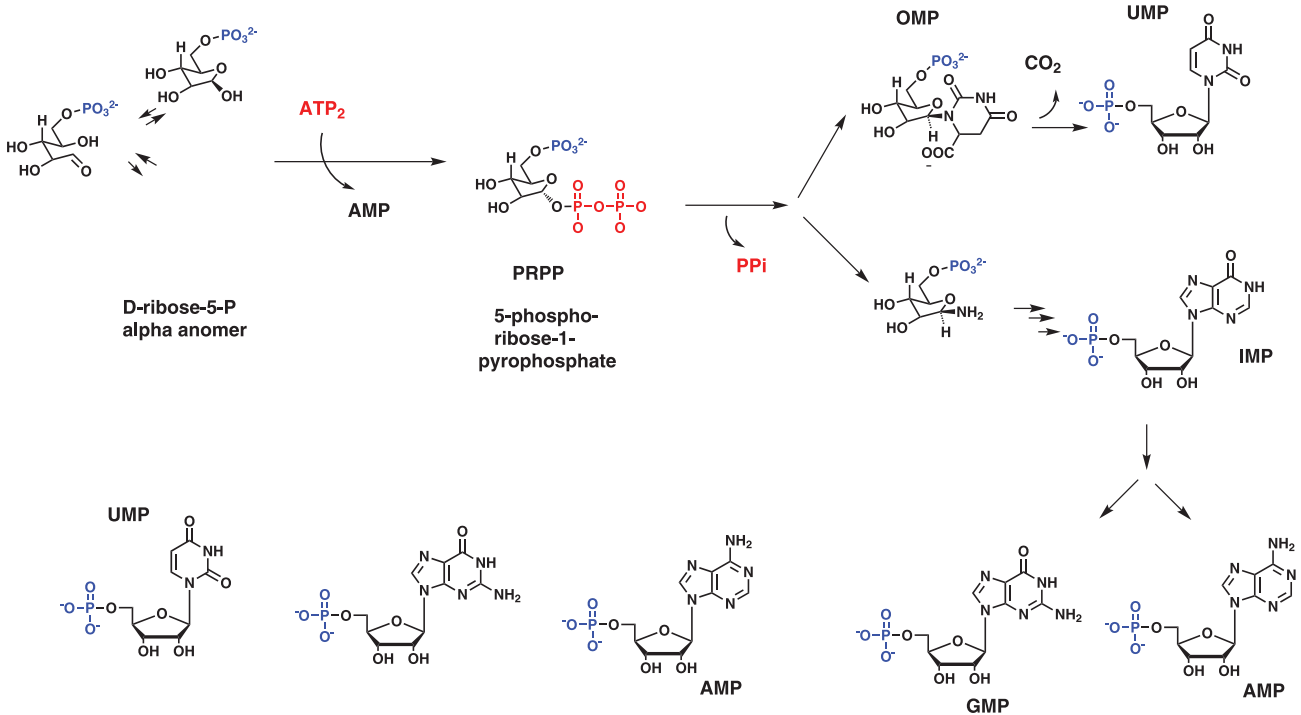


Figure 17.10 The regiochemical preference of hexokinase for C6 of glucose reads out in the 5′ to 3′ directionality of chain growth by RNA and DNA polymerases. Ribose-5-phosphate to PRPP to purine and pyrimidine nucleotide 5′-monophosphates.

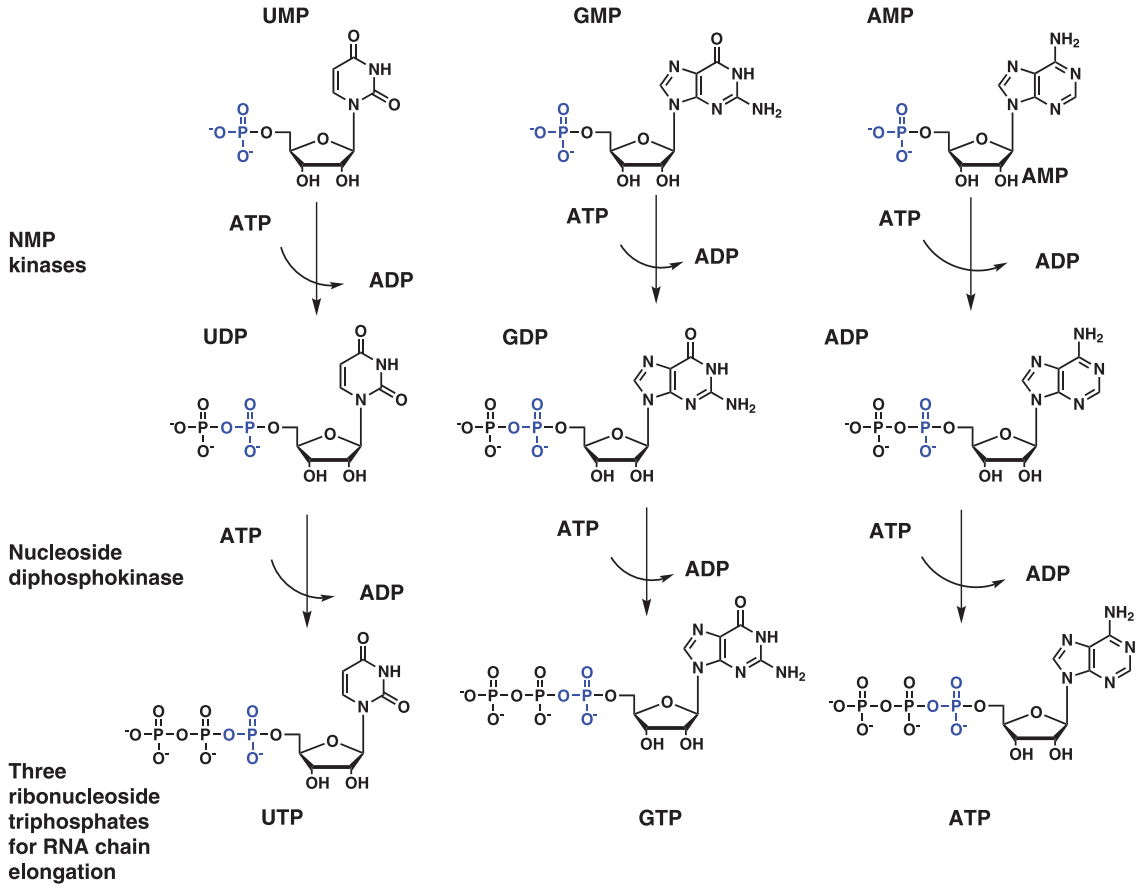


Figure 17.11 The regiochemical preference of hexokinase for C6 of glucose reads out in the 5′ to 3′ directionality of chain growth by RNA and DNA polymerases. 5′-NMPs to 5′-NTPs by successive phosphoryl transfers.

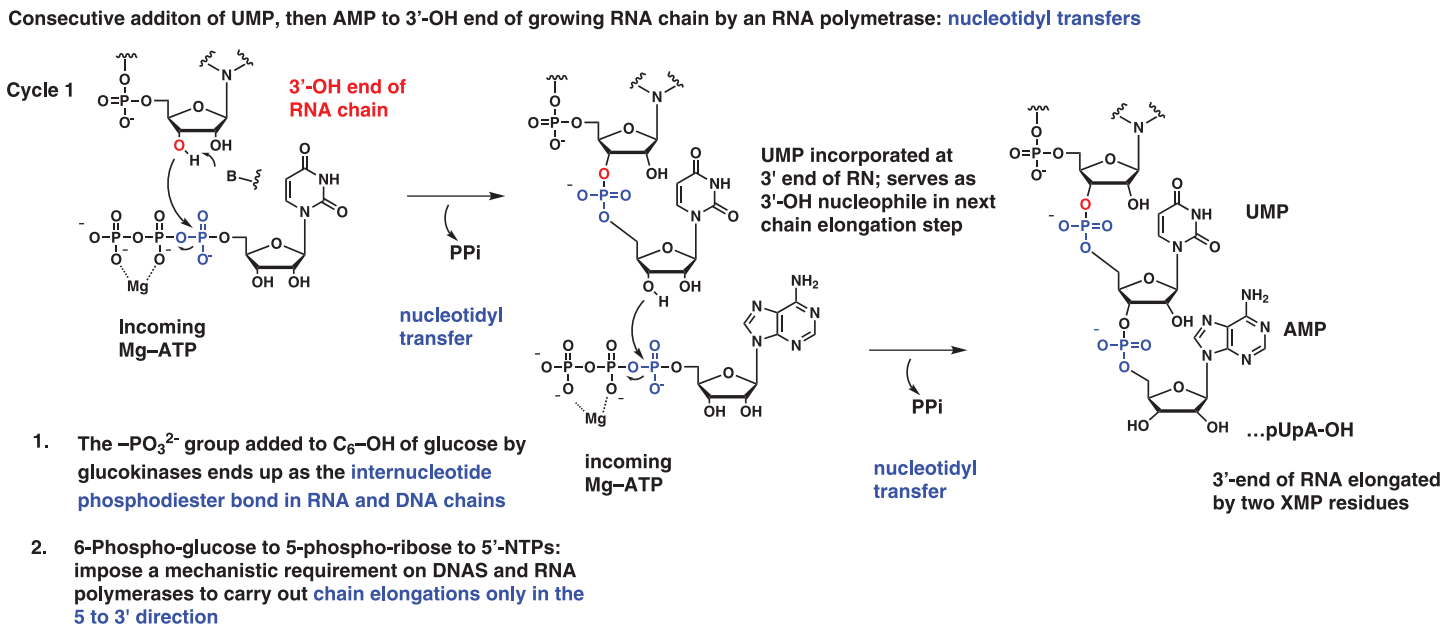


Figure 17.12 The regiochemical preference of hexokinase for C6 of glucose reads out in the 5′ to 3′ directionality of chain growth by RNA and DNA polymerases. Incorporation of 5′-NTPs or dNTPs into nucleic acid growing chains.

transfers, one experiences pyrophosphoryl transfer in the PRPP formation step. Each XMP group incorporation (AMP, GMP, UMP, CMP) in RNA chain extension is a nucleotidyl transfer. Thus, all three types of ATP triphosphate side chain fragmentations are involved in connecting glucose-6-phosphate formation to nucleic acid polymerase directionality.

Subject Index