The Structural and Biochemical Foundations of Thiamin Biosynthesis

Christopher T. Jurgenson,¹ Tadhg P. Begley,² and Steven E. Ealick²

¹Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06520; email: christopher.jurgenson@yale.edu
²Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853; email: tph2@cornell.edu, see3@cornell.edu

Key Words
degradation, salvage, transport, vitamin B₁

Abstract

Thiamin is synthesized by most prokaryotes and by eukaryotes such as yeast and plants. In all cases, the thiazole and pyrimidine moieties are synthesized in separate branches of the pathway and coupled to form thiamin phosphate. A final phosphorylation gives thiamin pyrophosphate, the active form of the cofactor. Over the past decade or so, biochemical and structural studies have elucidated most of the details of the thiamin biosynthetic pathway in bacteria. Formation of the thiazole requires six gene products, and formation of the pyrimidine requires two. In contrast, details of the thiamin biosynthetic pathway in yeast are only just beginning to emerge. Only one gene product is required for the biosynthesis of the thiazole and one for the biosynthesis of the pyrimidine. Thiamin can also be transported into the cell and can be salvaged through several routes. In addition, two thiamin degrading enzymes have been characterized, one of which is linked to a novel salvage pathway.
INTRODUCTION

When Christiaan Eijkman received the 1929 Nobel Prize in Medicine (see “The Eijkman Nobel Prize” sidebar) “for his discovery of the antineuritic vitamin” (1) dubbed thiamin through the work of Casimir Funk (2), it seemed unlikely that nearly 80 years later thiamin would still be the subject of active research, producing exciting advances in chemistry and biology. We now know that the elaborate process of thiamin biosynthesis utilizes many previously unprecedented biochemical mechanisms. Even the regulation of thiamin was shown to use a sophisticated method of transcriptional control through the use of the thiamin riboswitch (3, 4).

In all organisms, the thiazole and pyrimidine moieties of thiamin monophosphate (ThMP) are generated in separate branches of the pathway and then joined by a coupling enzyme. ThMP is converted to the active form of the cofactor thiamin diphosphate (ThDP) by a specific kinase. The best-studied thiamin biosynthetic pathways are those of *Escherichia coli* and *Bacillus subtilis*, which utilize very similar pathways, yet differ in some notable ways. The enzymes involved in the thiamin biosynthesis pathways for prokaryotes are illustrated in Figure 1. Table 1 gives the gene and enzyme names. Although at one time the gene names varied between *E. coli* and *B. subtilis*, the currently accepted standards, which are the same for both species, are used here.

Collaboration between chemists, enzymologists, and structural biologists has produced a clear understanding of how prokaryotes produce this essential cofactor. The X-ray crystal structures and biochemical functions of nearly every enzyme in this pathway have been determined. Studies on the formation of the thiazole, the pyrimidine, and thiamin itself, along with a variety of required kinases, are an important contribution to the vast array of biochemical knowledge in the field of vitamin biosynthesis.

In this review, we highlight the enzymological and structural studies that have elucidated thiamin biosynthesis in prokaryotes. Mechanisms
and protein structures, where possible, are described for each step. Thiamin salvage, transport, and regulation of thiamin biosynthetic genes are also discussed. The final section highlights a divergent thiazole biosynthetic pathway in *Saccharomyces cerevisiae*. The conclusion presents areas of further thiamin biochemical research in both bacteria and higher organisms.

**THIAZOLE BIOSYNTHESIS IN PROKARYOTES**

The thiazole moiety (4-methyl-5-β-hydroxyethylthiazole or THZ) is made through three distinct steps (Figure 1). First, glyceraldehyde 3-phosphate and pyruvate are coupled together by 1-deoxy-d-xylulose 5-phosphate synthase (Dxs) to give 1-deoxy-d-xylulose 5-phosphate (DXP). Next, the sulfur carrier protein ThiS undergoes an adenylylation by ThiF, followed by a sulfur transfer step using ThiI (*E. coli*) and IscS (NifS) to yield a thiocarboxy at its C terminus. It is this sulfur atom that is incorporated into the THZ ring of thiamin. Finally, glycine (by ThiO in *B. subtilis*) or tyrosine (by ThiH in *E. coli*) is converted to dehydroglycine. The thiocarboxy C terminus of ThiS, along with DXP and dehydroglycine, are all coupled together by thiazole synthase, ThiG, to give thiazole phosphate carboxylate tautomer. The enzyme TenI (*B. subtilis*) then aromatizes the thiazole tautomer to the thiazole phosphate carboxylate. The key enzymes of THZ formation are discussed in the following sections.

### Deoxy-d-Xylulose 5-Phosphate Synthase

The first step in thiazole biosynthesis utilizes Dxs to produce DXP from glyceraldehyde 3-phosphate and pyruvate. Paradoxically, despite being required for the biosynthesis of thiamin, Dxs itself requires ThDP for activity. How bacteria evolved to use a biosynthetic enzyme that requires the very metabolite that it makes is unknown. However, Dxs is also used in the production of pyridoxol (7) and isopentenyl pyrophosphate (PP) (8–11), so it is possible that the function of DXP synthase may have evolved during a time when bacteria produced thiamin through a more ancient pathway.

Dxs exists as a dimer, with the majority of the 3900-Å2 protomer interface consisting of hydrophobic residues. The crystal structures of Dxs from *E. coli* and *Deinococcus radiodurans* show that the monomer contains three distinct domains (Figure 2a) (12). Domains I (residues 1–319), II (residues 320–495), and III (residues 496–629) contain five-, six- and five-stranded β-sheets, respectively. All β-sheets are parallel with the exception of domain III, which has the first β-strand antiparallel to the other four. Dxs is most structurally similar to transketolase (13), pyruvate dehydrogenase E1 subunit (14), and 2-oxoisovalerate dehydrogenase (15). Though these enzymes catalyze similar reactions and use ThDP as a cofactor, the arrangement of domains is different from that seen in Dxs. Dxs is the only enzyme for which the active site is

---

**THE EIJKMAN NOBEL PRIZE**

Christiaan Eijkman’s pioneering work from 1890 to 1900 showed that a component of rice hulls could reverse the effects of beriberi in animals; however, it was not until 1933 that R.R. Williams purified thiamin. Prior to this discovery, Umetaro Suzuki identified rice hull isolates that contained the active ingredient. This work was first published in Japanese and then republished in German in 1912 (5).

The active ingredient called oryzanine, and later patented under the names aberic acid and orizanin, was shown to be an essential dietary component. Dogs fed on a thiamin-deficient diet of polished rice and boiled meat would succumb to beriberi in weeks but would recover quickly when given small amounts of oryzanine. This discovery was published about the time Casimir Funk was able to obtain crystals of a substance preventing polyneuritis, dubbing it “vitamine”—as it was a vital amine—although it was later shown that Funk most likely crystallized nicotinic acid (6). Eijkman further advanced his work after initially attributing the cause of beriberi to poisoning or microbial effects. He won the Nobel Prize in Physiology or Medicine in 1929, “for his discovery of the antineuritic vitamin” (1), also dubbed thiamin through the work of Funk (2).
Figure 1
Complete de novo thiamin biosynthetic pathway in bacteria.

contained within a single monomer—between domains I and II—and not between two monomers as with the other examples.

The active site of Dxs is shown in Figure 2b. The key interactions involved in ThDP binding include hydrogen bonds to the side chain oxygen atom of Glu370 and the amide nitrogen atom of Ser123 of N1 and N3 of the pyrimidine ring, respectively. The thiazole ring does not interact directly with the protein. The PP moiety forms several interactions with the protein. His80 and Tyr288 donate hydrogen bonds to the β-phosphate, and the amide nitrogen of Ala154 donates a hydrogen bond to the α-phosphate. A magnesium ion binds to both phosphate groups as well as to Asn181 and Asp152. The final ligands include the carbonyl oxygen atom from Met185 and, most likely, a water molecule.

The mechanism is depicted in Figure 2c and is typical of ThDP-utilizing enzymes. ThDP exists as an ylide, with the C2 carbon atom from the thiazole moiety acting as the nucleophile that attacks the C2 carbonyl carbon of pyruvate. Loss of CO2 gives the eneamine intermediate, which exists in resonance with the corresponding zwitterion. This acts as the nucleophile attacking the aldehyde group of glyceraldehyde 3-phosphate. Release of the product restores the thiamin ylide.
### Table 1  The bacterial thiamin biosynthetic enzymes

<table>
<thead>
<tr>
<th>Gene product</th>
<th>Enzyme function</th>
<th>Protein Data Bank identifier and organism for representative structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThiC</td>
<td>Hydroxymethyl pyrimidine synthase</td>
<td>3EPM, Caulobacter crescentus</td>
</tr>
<tr>
<td>ThiE</td>
<td>Thiamin phosphate synthase</td>
<td>1G69, Bacillus subtilis</td>
</tr>
<tr>
<td>ThiF</td>
<td>Adenyltransferase</td>
<td>1ZUD, Escherichia coli ThiS-ThiF complex</td>
</tr>
<tr>
<td>ThiS</td>
<td>Sulfur carrier protein</td>
<td>1TYG, B. subtilis ThiS-ThiF complex</td>
</tr>
<tr>
<td>ThiG</td>
<td>Thiazole synthase</td>
<td>1TYG, B. subtilis ThiS-ThiG complex</td>
</tr>
<tr>
<td>ThiO</td>
<td>Glycine oxidase</td>
<td>1NG3, B. subtilis</td>
</tr>
<tr>
<td>ThiH</td>
<td>Thiazole synthase</td>
<td>None</td>
</tr>
<tr>
<td>ThiI</td>
<td>Sulfur transferase</td>
<td>2C5S, Bacillus anthracis</td>
</tr>
<tr>
<td>NifS</td>
<td>Sulfur donor</td>
<td>1EG5, Thermatoga maritima</td>
</tr>
<tr>
<td>ThiM</td>
<td>Thiazole kinase</td>
<td>1ESQ, B. subtilis</td>
</tr>
<tr>
<td>ThiN</td>
<td>Thiamin pyrophosphokinase</td>
<td>None</td>
</tr>
<tr>
<td>ThiD</td>
<td>Hydroxymethyl pyrimidine (phosphate) kinase</td>
<td>1JXI, Salmonella typhimurium</td>
</tr>
<tr>
<td>ThiL</td>
<td>Thiamin phosphate kinase</td>
<td>3C9T, Aquifex aeolicus</td>
</tr>
<tr>
<td>ThiK</td>
<td>Thiamin kinase</td>
<td>None</td>
</tr>
<tr>
<td>Dxs</td>
<td>Deoxy-d-xylulose 5-phosphate synthase</td>
<td>2O1S, E. coli</td>
</tr>
<tr>
<td>TbpA</td>
<td>Thiamin binding protein</td>
<td>2QRY, E. coli</td>
</tr>
</tbody>
</table>

---

**Sulfur Carrier Protein**

The sulfur atom in the THZ ring originates from the thiocarboxylated sulfur carrier protein ThiS. The idea that a protein could be used as a sulfur carrier in the biosynthesis of a cofactor was first postulated in 1993 by Pitterle & Rajagopalan (16) while studying genes involved in the production of molybdopterin. Since then, the identification, function, and, in some instances, the structure of other sulfur carrier proteins have been solved. We now know that sulfur carrier proteins have a ubiquitin-like β-grasp fold and a diglycyl C terminus and are posttranslationally modified to have a thiocarboxy C terminus. They are used not only in the biosynthesis of thiamin, but also in molybdopterin (17), cysteine (18), and thioquinolobactin (19). The structure of ThiS has been solved by NMR (20), and X-ray crystal structures of ThiS in complex with the thiamin biosynthetic enzymes ThiF (21) and ThiG (22) have also been solved. ThiS consists of a five-stranded mixed β-sheet with an α-helix, which crosses over between strands β2 and β3, and a 3_10-helix between strands β4 and β5 (Figure 3a). Despite being structurally similar, there is little sequence similarity between ThiS and ubiquitin. In light of the structure of ThiS, it has been suggested that ubiquitin derives its origin from a prokaryotic ancestor because it also undergoes an adenylylation step followed by AMP displacement by a sulfur nucleophile (20).

ThiS is part of the *E. coli* *thiCEFSGH* operon, which contains most of the genes necessary to make thiamin (23). ThiS was found to be posttranslationally modified with a thiocarboxy C terminus in *thi*+ *E. coli* strains, but it was not modified in *thi*− strains (24). The protein ThiF was bound to ThiS tightly enough to be copurified. The purified ThiF-ThiS complex produced PP when incubated with ATP, and further analysis by electrospray ionization Fourier transform mass spectrometry (ESI/FTMS) showed a mass increase of 329 Da, consistent with the addition of AMP and loss of PP from ATP (24). ESI/FTMS analysis
of ThiS-ThiF and thiI+ *E. coli* strains showed that a mass increase of 16 Da was localized to the final residue on the C terminus of ThiS (24). The mass increase corresponds to a sulfur atom delivered by ThiI (NifS in *B. subtilis*), displacing the C-terminal oxygen atom. This generates ThiS-COS−, which is the source of the sulfur atom in the thiazole ring.

**Adenylyltransferase**

The sequence and structure of ThiF are similar to those of the molybdopterin biosynthetic
enzyme MoeB, the ubiquitin-conjugating enzyme E1 and the ubiquitin-related modifier 1–conjugating enzyme Uba4; each is an adenyllytransferase. This observation led to the conclusion that the ThiF mechanism may be similar to that of the ubiquitin-conjugating system that cross-links ubiquitin through a conserved cysteine residue. Crystal structures of ThiF alone (25) and the ThiF-ThiS complex (21) show how the structure of ThiF changes upon ThiS binding. A crossover loop, encompassing residues 181–185, is not visible in the complex structure but is ordered in the uncomplexed ThiF structure. This loop contains a conserved cysteine residue (Cys184) required for catalysis. It was found that the α-phosphate of a modeled AMP molecule is about 20 Å away from the thiol moiety of Cys184. However, this residue lies on a flexible loop and may still reach the substrate from that distance.

The ThiF-ThiS complex exists as a dimer of ThiF with one ThiS molecule bound to each ThiF protomer. The ThiS molecules do not interact with each other, which results in a cleft, approximately 20 Å wide, in the overall structure (Figure 3a). The protein–protein interface between ThiF and ThiS is 60% hydrophobic, with strands β3 and β4, Leu58, and the last seven residues of the C terminus from ThiS contributing the most interactions. The hydrophilic interactions include 14 hydrogen bonds, seven bridging waters, and a salt bridge. ThiF binds to ThiS using strands β5-β8, residues preceding helices α4 and α9, and helix α10.

The structure of the ThiF-ATP complex was used to generate a model of ATP bound to the active site of the ThiS-ThiF complex (Figure 3b). Asp127 is positioned near the α-phosphate and is predicted to bind a magnesium ion that would activate the α-phosphate for a nucleophilic attack required for adenyllytransferase activity. The C terminus of ThiS is also near the α-phosphate, facilitating transfer of AMP. The 2′ and 3′ oxygen atoms of the ribose ring donate hydrogen bonds to the carboxylate moiety of Asp59. The adenine ring is held by Arg106 through a cation-π interaction. The γ-phosphate accepts hydrogen bonds from the γ-oxygen atom of Thr186 and the guanidinium moiety of Arg70.
Sulfur Transfer

The structures of ThiI from *Bacillus anthracis* (26) and *Pyrococcus horikoshii* (27) have been determined and show that ThiI contains three separate domains (Figure 4a). The N-terminal domain is ferridoxin like, the connecting domain is an RNA-binding THUMP domain and the C-terminal domain is a PP-binding domain. The N-terminal domain consists of a four-stranded antiparallel β-sheet and two α-helices. This region is structurally similar to a wide range of proteins that adopt a ferridoxin-like fold. The final β-strand of the ferridoxin-like domain becomes the first β-strand of the following THUMP domain. The THUMP domain consists of five β-strands and two α-helices. The final domain contains a PP-binding loop and is structurally similar to the "N-type" ATP pyrophosphatases that catalyze an adenylylation step. ThiI is required both for sulfur transfer to ThiS in *E. coli* and for modifying uridine to 4-thiouridine in some prokaryotic tRNAs (28). IscS is also required for the biosynthesis of iron-sulfur clusters and may be responsible for sulfur incorporation in molybdopterin (29).

Figure 4b shows the protein-AMP-binding interactions that are seen in the PP-loop domain. The carbonyl oxygen atom and amide nitrogen atom of Phe209 form hydrogen bonds with N7 and N5 of the adenine ring, respectively. The carbonyl oxygen atom of Leu183 and the amide nitrogen atom of Gly287 form hydrogen bonds with the 2' and 3' oxygen atoms of the ribose ring, respectively. The phosphate moiety of AMP accepts hydrogen bonds from the side chains of Arg265 and Gln296.

Glycine Oxidase

The main difference between the *B. subtilis* pathway and the *E. coli* pathway occurs at the generation of dehydroglycine, which provides the final atoms for the formation of the thiazole. In *B. subtilis*, the reaction is catalyzed by ThiO, a flavoenzyme that uses glycine as a substrate to generate the glycine imine (30). The ThiO monomer has two separate domains (Figure 5a). One domain belongs to the glutathione reductase type 2 family and is responsible for binding flavin adenine dinucleotide (FAD), and the other domain binds substrate. ThiO is a tetramer with 222-point symmetry. Each ThiO protomer interacts with each of the other protomers in the tetramer. Openings in the quaternary structure allow the substrate to enter the active site of the substrate-binding domain.

The active site of ThiO contains an FAD cofactor. Figure 5b depicts the active site of ThiO with the glycine analog N-acetylglycine (NAG) bound. The carboxylate moiety of NAG binds to the guanidinium moiety of Arg302, which positions the substrate for oxidization by the flavin ring. The O7 oxygen atom of the flavin ring accepts hydrogen bonds from the amide nitrogen atoms of Ile332 and Leu333. The ring also forms π-stacking interactions with Tyr246.

ThiH is the *E. coli* enzyme that generates dehydroglycine. Its substrate is tyrosine, and there is no sequence similarity to ThiO. The structure of ThiH has not yet been determined. In *E. coli*, ThiH copurifies with ThiG and has been shown to contain an iron-sulfur cluster (31). Sequence alignments show strong similarity to
biotin synthase, a radical S-adenosylmethionine (SAM) enzyme (32). In addition, in vitro reconstitution of the *E. coli* thiazole biosynthetic pathway requires the ThiH-ThiG complex, ThiS-thiocarboxylate, tyrosine, SAM, and NADPH and occurs only under anaerobic conditions (33–36). In contrast, ThiO from the *B. subtilis* pathway is capable of making THZ in vitro using only ThiG along with sulfide, oxygen, and glycine (37).

The difference in the pathways between ThiO and ThiH lies in the fact that *B. subtilis* is an obligate aerobe, whereas *E. coli* can grow under aerobic and anaerobic conditions. Direct oxidation of glycine in an anaerobic environment is not likely to take place and requires a different catalytic strategy involving the use of an adenosyl radical intermediate. Kriek et al. (36) have shown that the conversion of tyrosine to dehydroalanine is initiated through the generation of a tyrosine radical that ultimately leads to dehydroglycine and p-cresol.

**Thiazole Synthase**

All the components required for THZ formation—ThiS-thiocarboxylate, DXP, and dehydroglycine—are assembled by thiazole synthase ThiG. ThiG shows some substrate tolerance as it is also capable of using 1,4-dideoxy-d-xylulose-5-phosphate as a substrate (38).

The crystal structure of the ThiG-ThiS complex reveals an octamer with 222-point symmetry (Figure 6a). ThiG has a (βα)8 fold, and each protomer binds to two separate ThiG protomers in addition to one ThiS molecule. One ThiG protomer interacts with helices α7 and α8 of the twofold related helices in the other. The other ThiG-ThiG interaction involves the β6/α6 loop of one ThiG protomer that interacts with the C-terminal helix α8. Two areas on ThiG make up the majority of the protein-protein interactions with ThiS. The C-terminal tail of ThiS passes through a clamp loop consisting of 14 residues. Hydrophobic residues on this loop bind to the hydrophobic region between the β-sheet and α-helix of ThiS after the C-terminal tail of ThiS has passed through it. These residues are structurally similar to hydrophobic residues in the enzymes MoaD (17) and NEDD8 (39), which have been responsible for interacting with their respective binding partners. The second area is mostly
Figure 6

(a) X-ray structure of the ThiG-ThiS complex. (b) The reaction mechanism of ThiG.

hydrophobic and involves strands $\beta_3$, $\beta_4$, and $\beta_5$ of ThiS and $\beta_1$, $\beta_2$, $\alpha_1$, $\alpha_2$, $\alpha_3$, and a loop region connecting $\alpha_3$ to $\beta_4$. Although hydrophobic contacts make up the majority of protein-protein interactions in the ThiG-ThiS structure (62%), three salt bridges and 11 hydrogen bonds contribute as well.

As described above, ThiF also binds ThiS but exhibits a different (Rossmann-like) fold. The residues in ThiS that bind both ThiF and ThiG are similar and are mostly located in the C terminus. In both cases, approximately 70% of all the protein-protein interactions with ThiS are hydrophobic. The fact that ThiS is capable of binding to proteins with such dissimilar folds, and differing electrostatic surfaces, supports the proposal that ThiS could be a predecessor to ubiquitin—a protein able to bind to many different targets in the degradation signaling pathway via the proteasome.
The reaction catalyzed by ThiG is outlined in Figure 6b. The final product is not thiazole phosphate (THZ-P), as has long been believed, but rather a carboxy thiazole phosphate tautomer, which must aromatize to form the final product. This step occurs very slowly and is catalyzed by the aromatase TenI in *B. subtilis* (A. Hazra, A. Chatterjee, & T. Begley, unpublished data).

**Aromatization of the Thiazole Ring**

The final protein involved in thiamin biosynthesis, TenI, has a known structure, but initially its catalytic activity was unclear (40). It is now known that TenI facilitates the aromatization of the thiazole tautomer product of ThiG (A. Hazra, A. Chatterjee, & T. Begley, unpublished data). Because the structure of TenI (Figure 7a) is similar to thiamin phosphate synthase (ThiE) (41), it was believed that TenI might catalyze the same reaction. However, the active site of TenI contains a leucine residue in the place of a glycine residue in ThiE, which prohibits proper binding of the substrates necessary to generate ThMP.

The structure of TenI complexed with carboxy thiazole phosphate (TCP) has been

---

**Figure 7**

(a) X-ray structure of TenI. (b) Active-site residues interacting with thiazole carboxylate phosphate (TCP). (c) The proposed reaction mechanism of TenI.
The phosphate moiety accepts hydrogen bonds from the amide nitrogen atoms of Gly156, Met176, and Ser177 (Figure 7b). The carboxylate binds to the ε-nitrogen atom of His102. The proposed aromatization mechanism involves His122 acting as a base to remove a hydrogen from C2 and replacing it on C6 of TCP (Figure 7c).

HYDROXYMETHYL PYRIMIDINE BIOSYNTHESIS

The 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMP-P) ring is generated through a complicated rearrangement reaction catalyzed by ThiC, using 5-aminoimidazole ribotide (AIR) as the substrate. ThiD then phosphorylates HMP-P to give HMP-PP.

Hydroxymethyl Pyrimidine Phosphate Synthase

In contrast to THZ-P, biosynthesis of the HMP-P moiety of thiamin in prokaryotes requires only one enzyme; however, it involves one of the most complicated rearrangement reactions in primary metabolism. Figure 1 shows that AIR is converted to HMP in a single enzyme-catalyzed reaction. Labeling studies have identified the origin of all of the HMP-P atoms (Figure 8c) (42–46). AIR is also an intermediate on the purine biosynthetic pathway. The crystal structure of ThiC from Caulobacter crescentus has been determined (47) and consists of an N-terminal domain, a core (βα)8 barrel domain, and a C-terminal domain, containing three absolutely conserved cysteine residues (Figure 8a). The N-terminal domain has a three-stranded antiparallel β-sheet and four α-helices that fold over the core domain. Sequence alignments show that the N-terminal domain is much smaller in anaerobes and cyanobacteria compared to aerobics. Part of the C-terminal domain is disordered in the crystal structure. Three helices in this domain occur after the final helix of the (βα)8 domain and make up a significant portion of the protein-protein interface in the ThiC dimer. Though the characteristic iron sulfur cluster motif CX2CX,C is present in the ThiC sequence, the cluster is not present when ThiC is overexpressed in E. coli and must be reconstituted in vitro.

The active site of ThiC, modeled with the substrate analog desamino AIR (IMR) along with the cofactor SAM and a [4Fe-4S] cluster, is shown in Figure 8b. SAM and the [4Fe-4S] cluster were modeled into the active site using other radical SAM enzyme structures as a guide, and the IMR is derived from X-ray crystal data. Hydrogen bonds to the phosphate group of IMR come from the side chains of Tyr277, His313, and Arg377 as well as the amide nitrogen atom of Gly335. The iron atoms of the iron sulfur cluster bind to the thiol side chains of Cys561, Cys564, Cys569, and the carboxylate moiety of SAM.

ThiC is required for HMP biosynthesis in plants as well. Sequence comparisons between Arabidopsis thaliana and the prokaryotic HMP synthase enzymes from E. coli, B. subtilis, Salmonella typhimurium, and Azotobacter vinelandii show significant similarity (48). It has been shown that ThiC in plants and algae also requires an iron sulfur cluster, as do their prokaryotic counterparts (49).

4-Amino-5-hydroxymethyl-2-methylpyrimidine Phosphate Kinase

The crystal structure of ThiD shows that it is a dimer with each monomer adopting an αβα sandwich fold consisting of eight α-helices, two 310-helices, and ten β-strands (Figure 9a). The dimeric interaction between ThiD protomers is relatively flat and consists of the first 70 C-terminal residues and a C-terminal loop, spanning residues 249–266. ThiD has a fold similar to that of a family of ribokinases, as does ThiK (below). ThiD is able to utilize both HMP and HMP-P as substrates in the same active site (50) and catalyzes both phosphorylation reactions. The pyrimidine
Figure 8

(a) X-ray structure of 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase, ThiC. (b) Active-site model showing residues interacting with desamino AIR (IMR), S-adenosylmethionine (SAM), and the Fe₄S₄ cluster. (c) The origin of the atoms of HMP-P derived from isotopic labeling studies.

The pyrimidine ring binds similarly whether the substrate is HMP or HMP-P, but the phosphate group of the latter is able to move into different phosphate-binding pockets through rotation about the C5-C7 bond, thus allowing it to be both a product and a substrate. On the basis of the HMP-P three-dimensional structure, it appears that HMP binds first, followed by ATP. After the first phosphorylation of HMP, ADP is released, and another molecule of ATP is bound in order to generate HMP-PP.

The active site of ThiD is depicted in Figure 9b. The N3 and N5 nitrogen atoms of the pyrimidine ring donate hydrogen bonds to the carboxylate side chain of Glu44. A sulfate ion, which mimics phosphate, is situated close to the hydroxyl moiety of the HMP for the phosphate addition reaction in the α position. The second sulfate-binding site (also mimicking phosphate) is positioned through a hydrogen bond to the amino group of Lys176; this marks the position of the second phosphorylation reaction in the β position. Despite the prevalence of pyrophosphorylated metabolites, the only other known example of a dual kinase activity is thymidine kinase from the human herpes virus 8 (51).
THIAMIN PYROPHOSPHATE BIOSYNTHESIS

ThiMP is formed through the coupling reaction of THZ-P and HMP-PP using ThiE. Thiamin phosphate kinase (ThiL) adds the final phosphate group to ThiMP to give ThDP, the active form of the cofactor. ThDP can also be formed in one step from thiamin using thiamin pyrophosphokinase. In bacteria, this enzyme is called ThiN. In higher organisms, the thiamin pyrophosphokinase is THI80.

Thiamin Phosphate Synthase

The high-resolution crystal structure of ThiE revealed a (βα)8-barrel fold, where helices α2–8 and α10 surround the central β-strand-lined barrel (Figure 10a). Helix α1 spans the N-terminal barrel entrance, and α9 is inserted between β8 and α10. ThiE exists as a dimer where helix α3 from each protomer aligns parallel to each other along the dimer interface, which also contains four salt bridges and three bridging water molecules.

Figure 10b depicts the active site of the S130A mutant of ThiE. This structure shows separate pyrimidine, THZ-P, and pyrophosphate moieties, suggesting that a carbocation has formed. The pyrimidine forms a hydrogen bond with an oxygen atom of the pyrophosphate group through the N4 nitrogen atom. The oxygen atom and nitrogen atom of the amide moiety of Gln57 form hydrogen bonds with the N4' and N3 nitrogen atoms of the pyrimidine ring, respectively. The pyrophosphate group also accepts hydrogen bonds from the side chains of Arg59, Lys61, and Lys159. The phosphate moiety of THZ-P forms hydrogen bonds with the amide nitrogen atoms of Gly168, Ser209, and Ile208, as well as with the hydroxyl moieties of Thr156 and Thr158.

The proposed mechanism of ThiMP formation begins with the loss of the PP group from C5 of HMP-PP to generate a carbocation that is stabilized through the delocalized π-system of the pyrimidine ring. Following a proton transfer from the N4' nitrogen atom of the pyrimidine ring to the PP leaving group, a pyrimidine imine methide intermediate is formed, which then undergoes nucleophilic attack from the nitrogen atom of the THZ-P to form ThiMP (52). The importance of a serine residue at position 130 (41) is demonstrated by 8000-fold reduction of activity in the alanine mutant (52).
Investigation of cryptic enzymes in *E. coli* led to the discovery that the previously unannotated gene *yjbQ* is a thiamin phosphate synthase homolog (53). Cryptic genes capable of duplicating contemporary thiamin biosynthetic enzymes are likely to exist because little intracellular thiamin is required for growth. This increases the likelihood that another enzyme within the proteome may be capable of catalyzing the same reaction. Poor catalytic efficiency of the cryptic enzyme may still be sufficient for cellular growth. The *yjbQ* gene was transformed into the MC1061Δ*thiE* strain of *E. coli* and showed thiamin auxotrophy complementation. Additionally, transformation of the *thiE* disruption strain with *yjbQ* genes from *Pyrococcus*, *Sulfolobus*, and *Thermotoga* also exhibited complementation, thereby showing that the thiamin phosphate synthase catalytic activity is a property of the YjbQ protein family.

Thiamin Phosphate Kinase

The final step in cofactor formation is the phosphorylation of ThMP to form ThDP. This reaction is carried out by ThiL in an ATP-dependent manner (54). ThiL contains two domains (Figure 11a). Domain 1 is a half-barrel with four very long β-strands. A ThiL dimer is formed when these β-strands from adjacent protomers join to form an eight-stranded β-barrel at the dimer interface. The second domain is an α/β domain composed of a six-stranded antiparallel β-sheet and a bundle of helices. The dimer interface is mostly hydrophobic with 10 hydrogen bonds and a single disulfide bridge that connects two Cys34 residues: one from each protomer. The structure of ThiL complexed with substrates, products, and analogs suggests that the enzyme utilizes a direct in line phosphate transfer to generate ThDP.

The active site of ThiL is depicted in Figure 11b which shows four of the five Mg$^{2+}$ ions observed to coordinate the phosphate moieties of ThMP and the AMP analog AMP-PCP. Two of the Mg$^{2+}$ ions coordinate both the phosphate moiety of ThMP and the γ-phosphate of AMP-PCP, and Ser209 donates a hydrogen bond from the side chain hydroxyl oxygen atom. One Mg$^{2+}$ ion bridges the two molecules through their phosphate moieties and is coordinated by Asp210. The other two Mg$^{2+}$ ions interact with the γ-phosphate of AMP-PCP as well as Asp27, Asp71, and Asp207. The
α-phosphate of AMP-PCP also accepts a hydrogen bond from Arg142.

A phosphorylated enzyme intermediate was ruled out because no amino acid residue is suitably positioned for phosphorylation. Comparisons with ThDP-binding proteins showed that the pyrimidine moiety is rotated approximately 100° when compared to ThDP in enzymes that require it as a cofactor. This prevents activation of the C2 carbon on the thiazole ring by N4 of the pyrimidine ring to generate the thiamin ylide in the active site of ThiL. ThDP can also be produced by thiamin pyrophosphokinase from *B. subtilis* (55), which is able to carry out the pyrophosphorylation of thiamin. Most higher organisms also contain this enzyme (56).

### THIAMIN DEGRADATION

Thiaminases degrade thiamin into separate thiazole and pyrimidine moieties. Two classes of thiaminases have been identified: thiaminase I (57–61) and thiamine II (62). TenA was shown to be the thiaminase II in *B. subtilis* (40). Thiaminase I has been associated with early mortality syndrome in predatory fish of the Laurentian Great Lakes and the New York Finger Lakes (63) as well as in Atlantic salmon in the Baltic Sea (64). Fish produce eggs with low levels of thiamin, which leads to a variety of illnesses and finally death between the time the fish are hatched and their first feeding. It is believed that the thiamin deficiency results from the egg-carrying female feeding on the non-native forage fish alewife (*Alosa pseudoharengus*) that produces high levels of thiaminase I. The importance of thiaminase I is also indicated in the sidebar “Thiaminase I—The Culprit Behind a Disastrous Nineteenth Century Australian Expedition.”

Catalytically, the difference between thiaminase I and II is seen in the substrates used to cleave the C-N bond connecting the two heterocyclic rings of thiamin. Thiaminase I can use aniline, cysteine, dithiothreitol, pyridine, quinoline, and veratrylamine as substrates (57, 67), and thiaminase II can only use water (Figure 12c) (40). Thiaminase I is also able to accept thiamin analogs, which vary extensively in the thiazole ring moiety but not in the pyrimidine moiety (68). The two classes of thiaminases have been shown to have dissimilar sequences and structures. It is also known that thiaminase II (TenA) promotes the production of degradative enzymes—dubbed Deg proteins in *B. subtilis* (69). These proteins are expressed in bacteria during the transition between a growth phase and a stationary phase. The protein TenI has been shown to have the opposite effect on the production of Deg proteins (70).
Thiaminase I

The structure of thiaminase I from *Bacillus thiaminolyticus* (Figure 12a) shows an overall fold similar to that of group II periplasmic-binding proteins, such as the maltose-binding protein (71) and the spermidine putrescine-binding protein potD (72, 73). The periplasmic binding proteins consist of two domains, each with an α/β fold. The two domains form a deep cleft and are connected by three crossover segments. This led to the proposal that thiaminase I may have evolved from an ancient periplasmic binding protein responsible for the uptake of thiamin (62). The physiological function of thiaminase I is not yet understood.

The active site of thiaminase I is located in the cleft between the two domains (Figure 12b). Six tyrosine residues and four acidic residues line the cleft. The active site was identified using the mechanism-based inhibitor 4-amino-6-chloro-2,5-dimethylpyrimidine, which bonds irreversibly to the active-site cysteine residue (Cys113). The structural arrangement suggests the mechanism outlined in Figure 12c. Thiamin is positioned in the active site by two hydrogen bonds between the pyrimidine moiety and Asp272, one of which occurs through an intervening water molecule. Glu241 then activates Cys113 for attack at C6 of the pyrimidine to form a zwitterionic intermediate. Nucleophilic attack and protonation by Glu241 result in cleavage of the bond between the thiazole and pyrimidine and release of products.

Thiaminase II

TenA was shown to be thiaminase II, which uses water exclusively as the nucleophile to cleave the C-N bond connecting the pyrimidine and thiazole moieties (40). The crystal structure of TenA shows a bundle of 11 helices that surround a deep acidic pocket (Figure 13a). The walls of the pocket are made of helices α4, α5, α9, and α10 with a total volume of approximately 700 Å³. The biological unit is a tetramer, with 222-point symmetry. Each monomer in the quaternary structure interacts with two other monomers. One interaction involves helices α7 and α11, and the other involves helices α4 and α5.

The active site of TenA is shown in Figure 13b. Several residues, Tyr47, Tyr112, Tyr163, and Phe208, contribute to the π-stacking environment around the HMP ligand. The carboxylate side chain of Glu205 forms a hydrogen bond with the N1 nitrogen atom of the pyrimidine ring. The carboxylate side chain of Asp44 and hydroxyl group of Tyr163 form hydrogen bonds with N3 and N4', respectively. The catalytic residue Cys135 is positioned near the C2 atom of the pyridine ring and is required for the cleavage of thiamin into its respective heterocycles (Figure 12c).

The biological function of TenA appears to be the salvage of base-degraded thiamin (Figure 14) (74). Some of the inconsistencies, suggesting that TenA’s physiological role is not thiamin degradation, included the observation that it is found in the thiamin biosynthetic
Figure 12

(a) Thiaminase I crystal structure. (b) Active-site residues with the inhibitor 4-amino-2,5-dimethylpyrimidine (ADP). (c) The reaction mechanism for thiaminase I.
oporon. Also, TenA has activity specifically with thiamin and not THMP or ThDP. Because the products of de novo thiamin biosynthesis are the singly and doubly phosphorylated forms of thiamin, TenA would not degrade thiamin produced in the cell at all. TenA is found in Bacillus halodurans, a bacterium that lives in soil pH > 10, which led to the suggestion that TenA might be involved in the salvage of degraded products of thiamin. Additionally, TenA in B. halodurans is clustered near genes associated with a deformylase YlmB and an ABC transporter complex ThiXYZ, also involved in salvage (75).

The mechanism of thiamin salvage involves uptake of N-formyl-4-amino-5-aminomethyl-2-methylpyrimidine, generated by base degradation of thiamin, followed by deformylation by YlmB, to give aminopyrimidine and TenA-catalyzed hydrolysis to generate HMP (Figure 14). TenA was shown to have 100 times greater activity against aminopyrimidine than with thiamin, further demonstrating that thiamin is not the natural substrate for this enzyme (76). The putative transporter ThiY was also shown to bind aminopyrimidine, which is then delivered to the ThiXZ transport channel.

**THIAMIN SALVAGE**

As an alternative to the de novo pathway described above, thiamin, or components of thiamin, can be salvaged (Figure 14). In bacteria, the thiazole alcohol can be converted to THZ-P by thiazole kinase (ThiM). In addition, HMP can be converted to HMP-P by HMP-P kinase (ThiD). ThiD is also the enzyme that converts HMP-P to HMP-PP in the de novo pathway. In bacteria, thiamin can be converted to THMP by thiamin kinase (ThiK) or to ThDP by thiamin pyrophosphokinase (ThiN). The corresponding enzyme in higher organisms is THI80.

**Thiazole Kinase**

ThiM from B. subtilis (77) and S. typhimurium (78) has been biochemically characterized, and the structure of the B. subtilis ThiM has been determined (Figure 15a) (79). ThiM is a trimer with each protomer containing nine β-strands and 12 α-helices. ThiM from B. subtilis is structurally homologous to ribokinase (80), adenosine kinase (81), and other related kinases. Ribokinase and adenosine kinase each have additional β-sheets that act as a flap over the
Salvage pathways for thiazole, thiamin, and pyrimidine.

active site. In the trimeric structure of ThiM, this flap is replaced by residues from an adjacent protomer.

The active site of ThiM is shown in Figure 15b. The active site contains the product THZ-P and the substrate ATP and is biochemically inert. The γ-phosphate of ATP is seen in close proximity to the phosphate of THZ-P, although it is in a bent conformation to accommodate the presence of the THZ-P. In the absence of the phosphate moiety on thiazole, this phosphate is able to adopt an extended conformation that is suitable for phosphate transfer to the hydroxyl group of the thiazole substrate. The thiazole ring is held in place through a hydrogen bond to the amide nitrogen atom of Met45. The phosphate moiety of THZ-P binds to the side chain of Ser198. Arg121 binds to the β-phosphate of ATP, and Thr168 donates a hydrogen bond to the α-phosphate through the hydroxyl moiety of its side chain. The ribose ring is held in place through a hydrogen bond between the 2′ position of the ribose ring and the carboxylate moiety of Asp172.

**Thiamin Pyrophosphokinase**

Structures have been published for thiamin pyrophosphokinase (THI80) from yeast and mouse (82, 83). The two enzymes share 26% sequence identity and have similar three-dimensional structures. The main difference is that the yeast enzyme has 20 additional N-terminal residues, seven additional C-terminal residues, a 38-residue insertion, and several
small insertions. The overall structure of thiamin pyrophosphokinase is a homodimer (Figure 16a). Each monomer consists of an α/β domain and a β-sandwich domain. Both domains contribute to dimer formation. Each of the two active sites is located in a cleft between the N-terminal domain of one protomer and the C-terminal domain of the other. Iterative BLAST searches suggest that the bacterial thiamin pyrophosphate kinase (ThiN) and THI80 share at least some structural homology.

Figure 15
(a) X-ray structure of thiazole kinase (ThiM). (b) Active site of ThiM with thiazole phosphate (THZ-P) and ATP bound.

Figure 16
(a) Thiamin pyrophosphokinase (THI80) crystal structure. (b) Active-site residues shown with their chain identifications.
Both the yeast and mouse enzymes contain a bound thiamin at the active site (Figure 16b). The binding site (yeast numbering) is formed mainly by residues 123–125 of one protomer and residues 7, 270, and 286–288 of the other. Primary interactions are formed by Glu7 of one protomer and Tyr123 of the other protomer. Hydrogen bonds are formed with Gln122 and Ser286, and Trp270 forms π-stacking interactions with the pyrimidine. Thiamin is observed in the unusual F conformation, and ThDP is usually in the V conformation when utilized as a cofactor. The ATP-binding site has not yet been characterized.

THIAMIN TRANSPORT

Bacteria are able to use exogenous thiamin and components of thiamin to supplement their own de novo biosynthesis through transmembrane transporters. Thiamin, ThMP, ThDP (84), HMP (85), and THZ (86) can all be taken up into the cell from nutrient media. Uptake of these compounds is poorly understood, but the thiamin-regulated operon tbpA thiPQ in E. coli and S. typhimurium (84) encodes an ABC transporter and includes a periplasmic thiamin-binding protein (TbpA), a transmembrane thiamin channel (ThiP), and an ATPase responsible for active transport into the cell (ThiQ) (84).

Thiamin-Binding Protein

TbpA from E. coli is a dimer with one ThMP-binding site per protomer (Figure 17a) (87). The TbpA protomer has two domains connected by a flexible linker. The TbpA-binding site is in a cleft created between these two domains and is typical of periplasmic binding proteins. TbpA is a monomer in solution, suggesting that the dimer observed in the crystal structure may be an artifact. It is also possible that the dimer forms after ThMP binds, resulting in higher binding affinities in the periplasm, which might increase transport efficiency.

The ThMP-binding site of TbpA is shown in Figure 17b. The phosphate moiety accepts

![Figure 17](attachment:figure17.png)

(a) X-ray structure of the thiamin-binding protein TbpA. (b) Thiamin-binding residues are shown with bound thiamin phosphate (TP).
Hydrogen bonds from the amide nitrogen atoms of Gly60 and Trp197. The hydroxyl side chain of Ser161 contributes a third hydrogen bond to the phosphate. The pyrimidine moiety accepts a hydrogen bond from the side chain of Ser218 at the N1 position and forms π-stacking interactions with Trp197. The thiazole moiety is sandwiched between the aromatic rings of the Trp197 and Tyr215.

TpbA is similar in structure to thiaminase I and possesses limited sequence similarity, suggesting that both proteins share a common ancestor. Organisms contain thiaminase I because it degrades thiamin. The sequence identity between TpbA and thiaminase I is only 15%, and the residues required to carry out the degradation reaction are not conserved.

**YkoF Protein**

*B. subtilis* uses an additional ABC transporter found in the *ykoCDEF* operon, which encodes for two transmembrane components (YkoC and YkoE), an ATPase (YkoD), and a thiamin/HMP-binding protein (YkoF) for which a crystal structure is available (88). YkoF has a ferredoxin-like fold and two thiamin (N- and C-terminal) binding domains per protomer that have vastly unequal binding affinities. The high-affinity binding site is 10 μM, whereas the low-affinity binding site is 250 μM (88). Also, there are no direct interactions with the thiazole moiety of ThDP. Because the binding affinity of TpbA is significantly higher at 3.8 nM, it is possible that YkoF may have a biological function other than ThDP uptake.

**Figure 18a** shows the structure of the YkoF dimer. Each monomer contains an eight-stranded antiparallel β-sheet with four α-helices stacked against one face. The monomeric structure has approximate twofold symmetry owing to the internal tandem repeat of the ferredoxin-like fold. Superimposing the 74 structurally similar Cα carbon atoms from each resulting domain shows a root mean square deviation of 1.7 Å. The dimer is arranged so that each monomer is positioned head to tail, with the hydrophobic core of each β-sheet making the majority of the protein-protein interface. The resulting dimer is spherical in shape with each monomer as one hemisphere of the globular structure.
Figure 18b,c show the high-affinity and low-affinity binding sites for thiamin, respectively. Both sites are similar, with the same number of hydrogen bonds to the pyrimidine ring. The high-affinity binding site forms hydrogen bonds between the amide nitrogen atom and the carbonyl oxygen atom with N3 and N4′ of the pyrimidine ring, respectively. Another hydrogen bond is accepted by N1 of the ring by the hydroxyl side chain of Ser154. Ile133 forms a hydrophobic region near the thiazole ring. In the low-affinity binding site, the same hydrogen bonds are seen but in this case are contributed by Leu17 and the hydroxyl side chain of Thr49. Phe15 may be involved in π-stacking, and Ile28 provides a similar hydrophobic environment near the thiazole ring as does Ile133 in the high-affinity binding site.

Eukaryotic thiamin uptake primarily utilizes proteins within the ACT transporter family. The recently discovered thiamin-binding protein Thi9 (89) was identified in Schizosaccharomyces pombe and is found within the APC superfamily of proteins; these proteins bind substrates such as choline, basic amino acids, polyamines, and γ-aminobutyric acid (90). Another protein, Bsu1, was also shown to be a pyridoxine-proton symporter, and Bsu1 is similar to exporting proteins from the multidrug resistance family (91). It is likely that Bsu1 is also able to bind HMP. Both Thi9 and Bsu1 are regulated by intracellular concentrations of thiamin.

THIAMIN REGULATION

The regulation of gene expression of most thiamin biosynthetic proteins is controlled at the transcriptional level. The E. coli operons thiCEFSGH, thiMD, and thiP/thiPQ are all regulated by the presence of ThDP through interactions with the THI-box riboswitch on mRNA encoding for each operon. Also, B. subtilis has a thiamin-regulated tenA-tenI-thiOSGFD operon (92) with a Rho-independent transcriptional terminator site upstream from the thiamin-regulated genes (75). In B. subtilis, the genes thiC and ywlD-ThiME are not downregulated by thiamin but rather are partially repressed by thiazole (77, 93). The genes thiD, thiE, dxs, and iscS are the only thiamin biosynthetic enzymes not regulated by a riboswitch and are not clustered with other thiamin biosynthetic genes. In the case of thiD, dxs, and iscS, this could be because these genes are required for other pathways.

The riboswitch is now known as a common regulatory element in gene expression and is present in biosynthetic genes for amino acids, nucleotides, and other vitamins (94) as well as in many different classes of bacteria (75). Some organisms use riboswitches to regulate over 2% of their genome (95–101), thus making the riboswitch a good drug target for pathogenic bacteria (102). The THI-box remains the only example of a riboswitch seen in eukaryotes, and it inhibits translation of thiamin biosynthetic genes by interfering with intron splicing (4, 48, 103).

The THI-box consists of a 5′ untranslated region that forms a thiamin-binding site (104). In gram-positive bacteria, binding of thiamin induces the formation of a Rho-independent transcriptional terminator. In gram-negative bacteria, binding of thiamin masks the Shine-Dalgarno sequence, which is required for the initiation of translation (105). The thiamin-binding domain of the riboswitch is 1000-fold more specific for ThDP than for ThMP, with KD values of 0.1 μM and 100 μM, respectively (96), thereby ensuring that only the active form of the cofactor inhibits translation.

THI-Box Riboswitch

The fact that ThDP binds to an mRNA molecule presents an interesting problem in that both molecules carry negative charges on their phosphate groups. The recent crystal structure of a THI-box RNA complexed with ThDP sheds light on how the binding occurs (Figure 19a) (3, 4, 106). ThDP binds in a linear conformation as seen in solution. In contrast, ThDP adopts a V conformation in enzymes that utilize it as a cofactor (107). The pyrimidine moiety of ThDP stacks between the
Figure 19
(a) X-ray structure of the THI-box. (b) Stereoview of the ThDP-binding site of the THI-box.

conserved purine bases G42 and A43 of the THI-box (Figure 19b). Electrostatic interactions between two magnesium-binding domains, which are found within the binding pocket of the riboswitch, coordinate the phosphate groups of ThDP. Residue G60 interacts with one Mg2+ directly, and U59, A61, and C77 interact with the same Mg2+ ion through a water molecule. The second Mg2+ is coordinated to A75, C77, and G78 via a water molecule as well. Mn2+, Ca2+, and Ba2+ have also been shown to bind the pyrophosphate group (106). The presence of the pyrophosphate group is required for divalent cation coordination as shown by the THI-box complex structure with pyrithiamine, a thiamin analog lacking a pyrophosphate group, which showed the absence of divalent cations (106). Two additional metal-binding sites contribute to the overall structure of the THI-box/ThDP complex. One interacts with U39 and G40 to help position the J2-3 bulge for pyrimidine binding. This metal-binding site is able to accommodate Mg2+, Mn2+, and Ba2+. The final metal-binding site connects helices P3 and P5 upon ThDP binding and is only able to bind Mg2+ (106). Figure 19b illustrates the THI-box, with two Mg2+ ions coordinating the pyrophosphate group (3).

EUKARYOTIC THIAMIN BIOSYNTHESIS

Though considerable progress has been made in discovering the biosynthesis, uptake, salvage, and regulation of thiamin in prokaryotes, comparatively little is still known about how eukaryotes make thiamin. As is the case for bacteria, the thiazole and pyrimidine moieties are generated through separate pathways, coupled by a single enzyme and phosphorylated by a variety of kinases (Figure 20). The thiazole ring is made from glycine, cysteine, and a five-carbon sugar, using a single enzyme, Thi4 (108–111).
Thiamin biosynthetic pathway in yeast.

Thi5 generates the HMP moiety using pyridoxine and histidine (112, 113). HMP is subsequently pyrophosphorylated by Thi20 (114). The thiazole and pyrimidine are then coupled by the bifunctional enzyme Thi6 (115), which also contains a hydroxyethylthiazole kinase activity (116). Thi20 is also a thiaminase II and has N- and C-terminal sequence homology with B. subtilis ThiD and TenA, respectively.

**Hydroxymethylpyrimidine Kinase/Thiamin Phosphate Synthase**

The thiamin biosynthetic enzyme Thi3 in plants was implicated in the biosynthesis of the hydroxymethylpyrimidine moiety over 15 years ago (117, 118), but no functional annotation was ascribed. The purification of THI3 from Zea mays and biochemical characterization have revealed that it is a dual-function enzyme (119). The N-terminal domain catalyzes the phosphorylation of HMP-P, and the C-terminal domain generates thiamin phosphate from HMP-PP and THZ-P. Structural predictions, using the Swiss Institute of Bioinformatics modeling server SWISS-MODEL, revealed that the N-terminal domain could be similar to the prokaryotic HMP-P kinase enzyme ThiD (50). The predicted residues responsible for HMP-P kinase activity included Gln98 and Met134, which correspond in S. typhimurium to the active-site residues Glu44 and Met80, respectively. Q98L and M134K mutations resulted in loss of HMP-P kinase activity, whereas ThiE activity remained. The C-terminal ThiE domain was predicted to be similar to the ThiE/TenI class of proteins that have a (βα)6-barrel fold. The active-site serine required for catalysis in ThiE from B. subtilis (Ser130) (41, 52) was predicted to be Ser444 in THI3. Mutation of Ser444 to alanine resulted in complete loss of ThiE activity and retention of HMP-P kinase activity. Kinetic analysis also revealed that the ThiE activity is uncompetitively inhibited by the presence of HMP-PP and ATP.

**Thiazole Synthase**

Recent discoveries in yeast thiamin biosynthesis have focused on the formation of the thiazole ring. The enzyme Thi4 in yeast—Thi1 in A. thaliana—is a homooctamer with 35 kDa protomers (30 kDa for THI1) (Figure 21a). Orthologs of THI4 in Fusarium oxysporum, F. solani (STI35), and Aspergillus oryzae (THIA) are transcriptionally regulated by ThDP binding to a THI-box riboswitch (103). Thi4 is a dual function enzyme involved with mitochondrial DNA repair (120) in addition to thiazole biosynthesis. The crystal structure of Thi4 (121) shows a proline in the cis conformation (Pro121) that follows a loop region responsible for donating hydrogen bonds to the α- and β-phosphate moieties of adenylated thiazole carboxylate.
(ADT), an unexpected bound product. It is possible that isomerization of this proline may lead to product release. Interestingly, in N. crassa, a cis-trans prolyl isomerase interacts with Thi4 (122), further supporting this speculation.

**Figure 21a** shows the biological unit of Thi4. The homo octameric structure is arranged as a tetramer of dimers. The dimeric interactions are primarily through helix α4 using both hydrogen bonding and hydrophobic interactions. Four such dimers are arranged into an octamer through interactions with helix α1. The doughnut-shaped octamer is approximately 96 Å across and 65 Å high. The inner ring contains solvent and is approximately 30 Å across. The substrate-binding sites are found on the inner side of the ring.

**Figure 21b** shows key residues that bind to the ADT molecule. The adenine ring is held through π-stacking interactions with Phe241 and Phe244. The carbonyl oxygen atom of Val170 accepts a hydrogen bond from N7' of the adenine ring. The 2' and 3' hydroxyl groups of the ribose sugar each donate one hydrogen bond to the carboxylate moiety of Glu97. The α-phosphate of ADT accepts a hydrogen bond from the backbone nitrogen atom of Gly105. Ser76 donates two hydrogen bonds to the...
COFACTOR CANNIBALISM

The catabolic pathways of many primary metabolites are well-understood and are often utilized as a source of carbon or nitrogen for a growing cell. In the case of some newly discovered enzymes, cofactors are catabolized in order to generate other cofactors. The substrate for thiazole synthase (Thi4) in *S. cerevisiae* is a cofactor NAD (Figure 20) (125). In the case of the cofactor adenosylcobalamin, the protein BluB catalyzes an oxidative ring opening of the isoalloxazine moiety of FAD to generate the dimethylbenzamidazole ligand, which coordinates the cobalt ion of vitamin B12 (126, 127). Another protein in the yeast thiamin biosynthetic pathway has also been found to use a cofactor as a substrate. The enzyme Thi5, which is responsible for generating the HMP moiety of thiamin, uses as substrates pyridoxine and histidine in a complicated rearrangement. Although radical SAM enzymes sometimes require the sacrifice of SAM, in the examples of Thi4, Thi5, and BluB, the cofactor substrates are needed to contribute to the molecular structure of their respective products (128).

β-phosphate moiety through its backbone nitrogen atom and the hydroxyl side chain. The carboxylate of the thiazole ring accepts hydrogen bonds from His237, Arg301, and Gly303. Asp207 from the adjacent monomer is shown just above the thiazole ring and is required for ADT formation.

The structures of Thi4 and Thi1 were analyzed for structural homologs using the DALI server (123) and shown to be most similar to several flavoenzymes from the glutathione reductase type II family (31). This was unexpected because typical NAD-binding proteins have a Rossmann fold (124). Surprisingly, one of the top structural homologs is the prokaryotic thiazole biosynthetic enzyme glycine oxidase (ThiO) from *B. subtilis*. The FAD-binding domain of ThiO overlays with the NAD-binding domain of THI4 with a root mean square difference of 2.4 Å. The second domain showed no closely related structural homologs. In the superposition, the ADP moiety of ADT overlays well with the FAD in ThiO; however, the thiazole carboxylate and isoalloxazine rings, respectively, deviate past the β-carbon. The similarity between Thi4 and flavoenzymes suggests that an ancient predecessor to Thi4 was capable of binding FAD and evolved the ability to bind and catalyze NAD without a rearrangement of topology.

The crystal structures of Thi1 and Thi4 contributed to the identification of the substrates, which for many years remained elusive. Crystal structures of Thi4 (121) and Thi1 (129) showed clear density for ADT in the active site. The unusual metabolite, bound to Thi4 overexpressed in *E. coli*, was confirmed to be ADT by mass spectrometry and NMR experiments (130). The adenylated ligand suggested that the substrate for Thi4 was a dinucleotide, but ADT was bound so tightly that it could not be removed from the native protein without denaturation; however, C204S and H200N mutants lacked a bound ligand but were still partially active. Thi4 mutants incubated in the presence of NAD and glycine showed four intermediates. The four intermediates were identified as ADT, ADP-ribose, ADP-ribulose, and an ADP-ribulose-glycine adduct (125). This represented a biochemical characterization of the consumption of one cofactor to generate a precursor for another cofactor, a type of “Cofactor Cannibalism” (see sidebar and Figure 21c) (128). Figure 21d shows a proposed Thi4 mechanism consistent with biochemical assays and the X-ray crystal structure.

CONCLUSIONS AND FUTURE ISSUES

Thiamin is synthesized by most prokaryotes and by eukaryotes such as yeast and plants. In all cases, the thiazole and pyrimidine moieties are synthesized in separate branches of the pathway and coupled to form thiamin phosphate. A final phosphorylation gives thiamin pyrophosphate, the active form of the cofactor. Over the past decade or so, biochemical and structural studies have elucidated most of the details of the thiamin biosynthetic pathway in bacteria (Figure 1). Formation of the thiazole
requires six gene products (Figure 6), and formation of the pyrimidine requires two. In contrast, details of the thiamin biosynthetic pathway in yeast are only just beginning to emerge (Figure 20). Only one gene product is required for the biosynthesis of the thiazole (Figure 21) and one for the biosynthesis of the pyrimidine.

Considerable progress has been made in understanding the biosynthesis of thiamin in bacteria, and significant inroads have been made for eukaryotes. However, many important questions still remain. The structure of ThiH is not known, although it is likely to be similar to that of biotin synthase. It is also not known how ThiH and ThiG interact. Although structural examples are available for almost all of the prokaryotic enzymes, in many cases, the hard work of preparing and crystallizing complexes will be necessary to fully elucidate the enzyme mechanisms. The recent discovery of a thiamin salvage pathway that utilizes TenA, and the vast number of TenA orthologs, highlights the possibility that other salvage pathways remain to be discovered. Some organisms lack enzymes associated with the standard pathway, and it is likely that completely new pathways or significant variations of the known pathways remain to be discovered.

The field of eukaryotic thiamin biosynthesis still has many questions to be answered, but the remarkable discovery of the structure, function, substrate, and product of thiazole synthase (Thi4/Thi1) shows that new biochemical methods of primary metabolite biosynthesis can still be unveiled. The role of Thi4 in mitochondrial DNA repair is not understood. Structures are not known for Thi3, Thi5, Thi6, Thi20, and others. The possibility of drug discovery using thiamin biosynthetic enzymes or the THI-box as a target may provide new leads in the field of antimicrobial drug development. The structure of the THI-box shows that it binds ThDP differently than typical thiamin-binding proteins, leading to the possibility that thiamin analogs could be developed that bind to the thiamin riboswitch while not interfering with human thiamin-binding proteins. Little is known about the structure and mechanism behind exogenous thiamin, HMP and THZ uptake, so biochemical and structural studies into thiamin transport will be needed.

Other forms of thiamin are also being discovered. Thiamin triphosphate (ThTP) activates chloride channels (131, 132) and is in high concentration in neuronal cells (133). ThTP is also the phosphate donor for the phosphorylation of a histidine residue in 43K rapsyn, which is found in postsynaptic membranes and is required for proper synapse function (134). A recently discovered adenyltransferase (135) generates an adenyalted thiamin species, adenosine thiamin triphosphate (AThTP), which appears in response to carbon starvation in E. coli (136). AThTP is also present in yeast, animal tissue, and the roots of higher plants, illustrating another metabolite found across a diverse range of organisms. Further research on the function of thiamin, its biosynthetic and transport proteins, and its regulation will continue to reveal the biochemical foundations of this essential cofactor.

DISCLOSURE STATEMENT
The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS
This work was supported by National Institutes of Health grants DK44083 to T.P.B. and DK067081 to S.E.E. We thank Leslie Kinsland for help in preparing this manuscript. Dr. Jurgenson was affiliated with the Department of Chemistry and Chemical Biology, Cornell University when this article was written.
LITERATURE CITED

2. Funk C. 1912. The preparation from yeast and certain foodstuffs of the substance the deficiency of which in diet occasions polyneuritis in birds. J. Physiol. 45:75–81


51. Gustafson EA, Schinazi RF, Fingeroth JD. 2000. Human herpesvirus 8 open reading frame 21 is a thymidine and thymidylylate kinase of narrow substrate specificity that efficiently phosphorylates zidovudine but not ganciclovir. J. Viral. 74:684–92
71. Spurlino JC, Lu GY, Quiocco FA. 1991. The 2.3 Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J. Biol. Chem.* 266:5202–19


121. Jurgenson CT, Chatterjee A, Begley TP, Ealick SE. 2006. Structural insights into the function of the thiamin biosynthetic enzyme Thi4 from Saccharomyces cerevisiae. Biochemistry 45:11061–70


134. Nghiem HO, Bettendorff L, Changeux JP. 2000. Specific phosphorylation of Torpedo 43K rapsyn by endogenous kinase(s) with thiamine triphosphate as the phosphate donor. EASEB J. 14:543–54


Contents

Preface ................................................................. v

Prefatory Articles

Frontispiece

E. Peter Geiduschek ......................................................... xii

Without a License, or Accidents Waiting to Happen
E. Peter Geiduschek .......................................................... 1

Frontispiece

James C. Wang ........................................................... 30

A Journey in the World of DNA Rings and Beyond
James C. Wang ............................................................ 31

Biochemistry and Disease Theme

The Biochemistry of Disease: Desperately Seeking Syzygy
John W. Kozarich .......................................................... 55

Biosynthesis of Phosphonic and Phosphinic Acid Natural Products
William W. Metcalf and Wilfred A. van der Donk ...................... 65

New Antivirals and Drug Resistance
Peter M. Colman ........................................................... 95

Multidrug Resistance in Bacteria
Hiroshi Nikaido ............................................................... 119

Conformational Pathology of the Serpins: Themes, Variations,
and Therapeutic Strategies
Bibek Gooptu and David A. Lomas ...................................... 147

Getting a Grip on Prions: Oligomers, Amyloids, and Pathological
Membrane Interactions
Byron Caughey, Gerald S. Baron, Bruce Chesebro, and Martin Jeffrey .... 177

Ubiquitin-Mediated Protein Regulation

RING Domain E3 Ubiquitin Ligases
Raymond J. Deshaies and Claudio A.P. Joazeiro .......................... 399

Regulation and Cellular Roles of Ubiquitin-Specific
Deubiquitinating Enzymes
Francisca E. Reyes-Turcu, Karen H. Ventii, and Keith D. Wilkinson .... 363
Recognition and Processing of Ubiquitin-Protein Conjugates
by the Proteasome
Daniel Finley ................................................................. 477

Degradation of Activated Protein Kinases by Ubiquitination
*Zhimin Lu and Tony Hunter* .............................................. 435

The Role of Ubiquitin in NF-κB Regulatory Pathways
*Brian Skaug, Xiaomo Jiang, and Zhijian J. Chen* ............... 769

Biological and Chemical Approaches to Diseases of Proteostasis
Deficiency
*Evan T. Powers, Richard I. Morimoto, Andrew Dillin, Jeffery W. Kelly,* and *William E. Balch* ........................................... 959

**Gene Expression**

RNA Polymerase Active Center: The Molecular Engine
of Transcription
*Evgeny Nudler* .................................................................. 335

Genome-Wide Views of Chromatin Structure
*Oliver J. Rando and Howard Y. Chang* ................................. 245

The Biology of Chromatin Remodeling Complexes
*Cedric R. Clapier and Bradley R. Cairns* .............................. 273

The Structural and Functional Diversity of Metabolite-Binding
Riboswitches
*Adam Roth and Ronald R. Breaker* ....................................... 305

**Lipid and Membrane Biogenesis**

Genetic and Biochemical Analysis of Non-Vesicular Lipid Traffic
*Dennis R. Voelker* .......................................................... 827

Cholesterol 24-Hydroxylase: An Enzyme of Cholesterol Turnover
in the Brain
*David W. Russell, Rebekkah W. Halford, Denise M.O. Ramirez,* Rahul Shab,* and *Tiina Kotti* .................................................... 1017

Lipid-Dependent Membrane Protein Topogenesis
*William Dowhan and Mikhail Bogdanov* ............................... 515

Single-Molecule Studies of the Neuronal SNARE Fusion Machinery
*Axel T. Brunger, Keith Weninger, Mark Bowen,* and *Steven Chu* ................................................................. 903

Mechanisms of Endocytosis
*Gary J. Doherty and Harvey T. McMahon* ................................ 857
Recent Advances in Biochemistry

Motors, Switches, and Contacts in the Replisome

_Samir M. Hamdan and Charles C. Richardson_ ........................................... 205

Large-Scale Structural Biology of the Human Proteome

_Aled Edwards_ .................................................................................................................. 541

Collagen Structure and Stability

_Matthew D. Shoulders and Ronald T. Raines_ ......................................................... 929

The Structural and Biochemical Foundations of Thiamin Biosynthesis

_Christopher T. Jurgenson, Tadgh P. Begley, and Steven E. Ealick_ ..................... 569

Proton-Coupled Electron Transfer in Biology: Results from Synergistic Studies in Natural and Model Systems

_Steven Y. Reece and Daniel G. Nocera_ .................................................................... 673

Mechanism of Mo-Dependent Nitrogenase

_Lance C. Seefeldt, Brian M. Hoffman, and Dennis R. Dean_ ..................... 701

Inorganic Polyphosphate: Essential for Growth and Survival

_Narayana N. Rao, María R. Gómez-García, and Arthur Kornberg_ ............................ 605

Essentials for ATP Synthesis by F₁F₀ ATP Synthases

_Christoph von Ballmoos, Alexander Wiedenmann, and Peter Dimroth_ ............ 649

The Chemical Biology of Protein Phosphorylation

_Mary Katherine Tarrant and Philip A. Cole_ ................................................................. 797

Sphingosine 1-Phosphate Receptor Signaling

_Hugh Rosen, Pedro J. Gonzalez-Cabrera, M. Germana Sanna, and Steven Brown_ .... 743

The Advent of Near-Atomic Resolution in Single-Particle Electron Microscopy

_Yifan Cheng and Thomas Walz_ .................................................................................... 723

Super-Resolution Fluorescence Microscopy

_Bo Huang, Mark Bates, and Xiaowei Zhuang_ .............................................................. 993

Indexes

Cumulative Index of Contributing Authors, Volumes 74–78 ................................... 1041
Cumulative Index of Chapter Titles, Volumes 74–78 .............................................. 1045

Errata

An online log of corrections to _Annual Review of Biochemistry_ articles may be found at http://biochem.annualreviews.org/errata.shtml