Structural Insights into the Function of the Thiamin Biosynthetic Enzyme Thi4 from *Saccharomyces cerevisiae*[†,‡]

Christopher T. Jurgenson, Abhishek Chatterjee, Tadhg P. Begley,* and Steven E. Ealick*

Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301

Received May 23, 2006; Revised Manuscript Received June 26, 2006

ABSTRACT: The structure of thiazole synthase (Thi4) from *Saccharomyces cerevisiae* was determined to 1.8 Å resolution. Thi4 exists as an octamer with two monomers in the asymmetric unit. The structure reveals the presence of a tightly bound adenosine diphospho-5-(β-ethyl)-4-methylthiazole-2-carboxylic acid at the active site. The isolation of this reaction product identifies NAD as the most likely precursor and provides the first mechanistic insights into the biosynthesis of the thiamin thiazole in eukaryotes. Additionally, the Thi4 structure reveals the first protein structure with a GR_2 domain that binds NAD instead of FAD, raising interesting questions about how this protein evolved from a flavoenzyme to a NAD binding enzyme.

Thiamin pyrophosphate **6** plays a key role in amino acid and carbohydrate metabolism and is an essential cofactor in all living systems. Its biosynthesis requires the separate formation of the pyrimidine **5** and thiazole **4** heterocycles, which are then coupled to form the cofactor (Scheme 1). The mechanism of thiazole formation in bacteria has been elucidated and involves a complex oxidative condensation in which five different enzymes catalyze reactions involving three different substrates (1-deoxy-D-xylulose-5-phosphate **3**, cysteine **2**, and glycine **1** or tyrosine) (*1*). The bacterial biosynthesis of the pyrimidine moiety is still poorly understood.

In contrast to the bacterial system, thiamin biosynthetic studies in eukaryotes are still at an early stage. Labeling studies in *Saccharomyces cerevisiae* have demonstrated that the thiamin thiazole is biosynthesized from an unidentified five-carbon carbohydrate **7**, glycine **1**, and cysteine **2** (*2*−*4*) (Scheme 1). Only one eukaryotic thiazole biosynthetic enzyme has been identified (Thi4) (*5*). This enzyme is a 326-amino acid (35 kDa) cytosolic protein. Sequence analysis suggests a conserved dinucleotide binding motif but does not reveal a biochemical function.

Our recently published results suggest an enzymatic mechanism whereby the cofactor nicotinamide adenine dinucleotide (NAD)[1] is converted into adenosine diphospho-5-(β-ethyl)-4-methylthiazole-2-carboxylic acid (ADT **20**) (*6*). The ADT ligand is tightly bound and can be released in vitro only upon protein denaturation. In this mechanism (Scheme 2), cleavage of the *N*-glycosyl bond of NAD **8** would give ADP ribose **9**. Ring opening of this hemiacetal followed by tautomerization would give **10**. A second tautomerization followed by loss of water would give **12**. Tautomerization of **12** to diketone **13** followed by imine formation with glycine **1** would give **14**. Imine tautomerization to **15** sets up the system to react with the sulfur donor, depicted here as sulfide, to give **16**. Cyclization gives **17**. Two dehydration reactions followed by a final tautomerization complete the thiazole formation in ADT **20**. This proposal suggests that the first step in thiazole formation is similar to the chemistry used in ADP ribosylation.

Here, we report the structure of the *S. cerevisiae* thiazole synthase which reveals the nature of binding to ADT found in the enzyme active site. The identification of this product suggests that NAD is the source of the carbohydrate precursor and provides the first insight into the mechanism of thiamin thiazole biosynthesis in eukaryotes.

MATERIALS AND METHODS

Protein Production. The overexpression vector for the *THI4* gene was constructed at the Cornell Protein Facility by standard methods from genomic DNA harvested from cells obtained from the American Type Culture Collection. *THI4* was cloned into a pET28 overexpression vector encoding a six-polyhistidine tag with a thrombin cleavage site located between the tag and the protein. The vectors were transformed into B834(DE3) (Novagen) methionine auxotrophic *Escherichia coli* cells (Novagen). Cells were grown in LB medium at 37 °C until an OD_{600} of 1.0 was reached. The cells were induced using 1 mM isopropyl-β-D-thiogalactopyranoside and incubated for 16−18 h at 18 °C. Cell pellets that were collected after induction were placed into a beaker containing 50 mL of buffer with 5 mM imidazole, 0.5 M NaCl, and 20 mM Tris (pH 7.9) (lysis buffer) and lysed using sonication. The cell lysate was separated from the insoluble cell particles by centrifugation at 58000*g* at 4

[†] This work was supported by National Institutes of Health Grants DK44083 (to T.P.B.) and DK67081 (to S.E.E.).

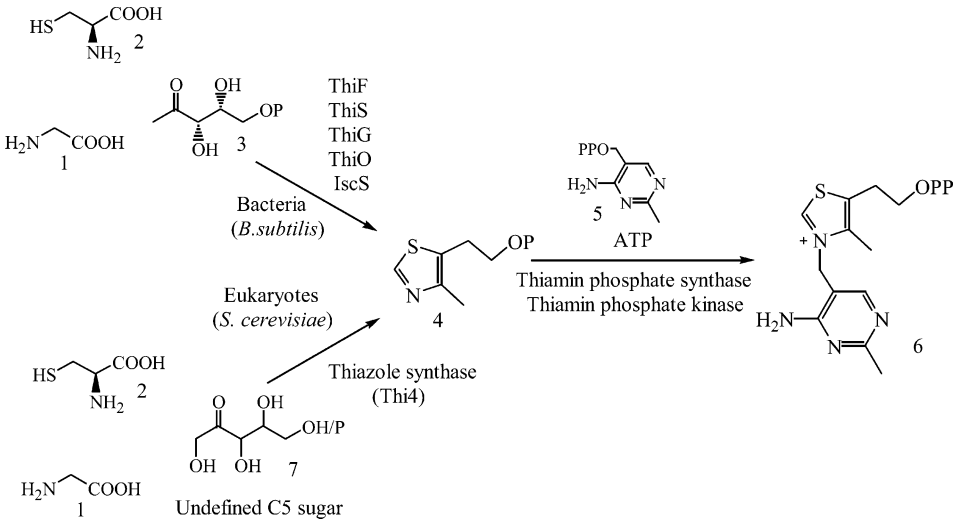
[‡] The Brookhaven Protein Data Bank code for Thi4 is 2GJC.

* To whom correspondence should be addressed: Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853. Telephone: (607) 255-7961. Fax: (607) 255-1227. E-mail: see3@cornell.edu or tpb2@cornell.edu.

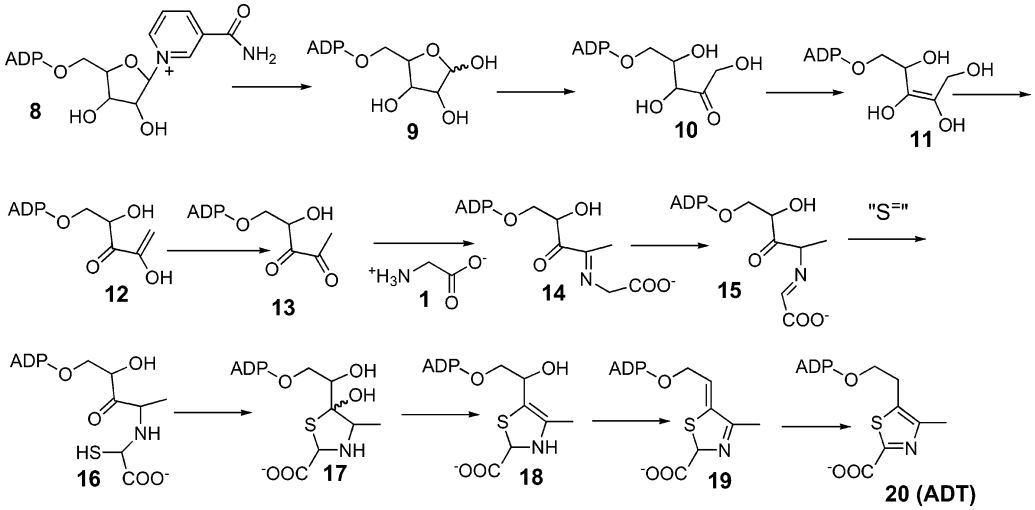
[1] Abbreviations: HET, hydroxyethylthiazole; HMP, hydroxymethylpyrimidine; HMP-P, hydroxymethylpyrimidine phosphate; TPP, thiamin pyrophosphate; ADT, adenosine diphospho-5-(β-ethyl)-4-methylthiazole-2-carboxylic acid; NAD, nicotinamide adenine dinucleotide; FAD, flavin adenine dinucleotide.

10.1021/bi061025z CCC: $33.50
Published on Web 08/25/2006

Scheme 1



Scheme 2



°C for 20 min. The supernatant was decanted and loaded onto a Ni−NTA affinity column equilibrated with the same buffer used in the lysis step. The column with bound Thi4 was washed with lysis buffer containing 60 mM imidazole to remove any remaining proteins that may have affinity for the Ni column. The protein was eluted by washing the column with lysis buffer containing 0.5 M imidazole. The eluate was buffer exchanged into 20 mM Tris (pH 7.9) using Bio-Rad DG size exclusion chromatography columns with a molecular mass cutoff of 6 kDa as per the manufacturer's instructions. The resulting solution was then concentrated in an Amicon Ultra centrifugal filter with a molecular mass cutoff of 10 kDa by centrifugation at 5000*g* for 30−45 min at 4 °C until the concentration reached 10−20 mg/mL as measured using the Bradford assay. The polyhistidine tag was removed using biotinylated thrombin from the Novagen Thrombin Cleavage Capture kit as per the manufacturer's instructions. Biotinylated thrombin was removed by adding streptavidin−agarose beads to the reaction mixture, allowing it to equilibrate for 30 min, and then collecting the beads through a spin filter. The cleaved histidine tags were separated by running the filtrate through a Ni−NTA column equilibrated with 60 mM imidazole, 0.5 M NaCl, and 20 mM Tris (pH 7.9). The eluate was then buffer exchanged into 20 mM Tris (pH 7.9) and concentrated to 10−20 mg/mL as described in the buffer exchange step with histidine-tagged Thi4. The final Thi4 protein product had the endogenously bound ligand ADT, which became apparent during structure determination and refinement.

Crystallization and X-ray Data Collection. The hanging drop vapor diffusion method was used to grow crystals of Thi4. The well solutions consisted of 1.65−2.00 M Li_2SO_4 and 0.1 M HEPES (pH 7.5). A hanging drop containing 1 μL of 10 mg/mL polyhistidine tag-cleaved Thi4 and 2 μL of well solution were mixed and incubated at 22 °C. Cubic crystals appeared in 2−4 weeks and grew 100−200 μm along each edge. Crystals were mounted onto a cryoloop and cryoprotected by being dipped into a saturated solution of Li_2SO_4 as prescribed by Rubinson et al. (*7*) before being plunged into liquid N_2. Data were collected on beamline A1 at the Cornell High Energy Synchrotron Source using a Quantum 210 CCD detector (Area Detector Systems Corp.). Individual frames were collected in 0.5° oscillations for 220 frames with an exposure time of 60 s per frame under a cryostream of liquid N_2 with the crystal mounted 190 mm from the detector. Integration and scaling of the data were carried out using the HKL2000 suite of programs (*8*). Data collection statistics are given in Table 1.

Structure Determination. The structure of Thi4 was determined using the method of molecular replacement with

Table 1: Data Collection Statistics

resolution (Å)	1.82
space group	*I*4
unit cell parameters (Å)	
a	140.70
c	73.33
no. of reflections	280639
no. of unique reflections	63286
redundancy	4.4 (3.1)[a]
completeness	98.5 (94.0)[a]
R_{sym}[b] (%)	25.6 (4.5)[a]
$I/\sigma(I)$	22.7 (3.3)[a]

[a] Numbers in parentheses are from the highest-resolution shell. [b] $R_{sym} = \sum\sum_i|I_i - \langle I\rangle|/\sum\langle I\rangle$, where $\langle I\rangle$ is the mean intensity of N reflections with intensities I_i and common indices *hkl*.

Table 2: Refinement Statistics

resolution (Å)	1.82
no. of non-H atoms	5192
no. of protein atoms	4516
no. of water atoms	598
no. of ligand atoms	76
R_{factor} (%)[a]	19.6
R_{free} (%)[b]	23.1
Ramachandran plot	
most favored region (%)	90.7
additionally allowed region (%)	7.4
generously allowed region (%)	1.0
disallowed region (%)	1.0
rms deviations from ideal	
bonds (Å)	0.027
angles (deg)	2.2

[a] $R_{factor} = \sum||F_{obs}| - K|F_{calc}||/\sum_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. [b] R_{free} was calculated using 5% of all reflections that were excluded at all stages of refinement.

the structure of Thi1 from *Arabidopsis thaliana* (PDB entry 1RP0) as the molecular replacement search model. Refinement was carried out using rigid-body refinement followed by simulated annealing using CNS (*9*) as well as REFMAC from CCP4 (*10*) and model rebuilding using O (*11*) and COOT (*12*). The final refinement statistics are listed in Table 2. All graphics were prepared using PyMOL (*13*) and ChemDraw (CambridgeSoft).

RESULTS

Monomer Structure. Thi4 is a homooctamer with two monomers, designated A and B, per asymmetric unit. The N-terminal regions of chains A and B are disordered from residue 1 to 16 and from residue 1 to 15, respectively. Other areas that have disordered loops for chain A include residues 180−184, 200−202, 239, and 240 and for chain B include residues 179−183, 199−202, and 238−241. The Thi4 monomer consists of 10 β-strands and 9 α-helices. The labeled monomer and a topology diagram representation are shown in Figure 1. The monomer consists of a central five-stranded parallel β-sheet flanked on one side by three α-helices and on the other by an antiparallel three-stranded β-sheet that lies between 8 and 9 Å above the plane of the central β-sheet and is rotated by an angle of approximately 30°. Long α-helices are found on both the N-terminal and C-terminal ends of the monomer. The N-terminal amphipathic helix spans residues 45−62. The C-terminal helix spans residues 307−324 and is also amphipathic. It is located in a

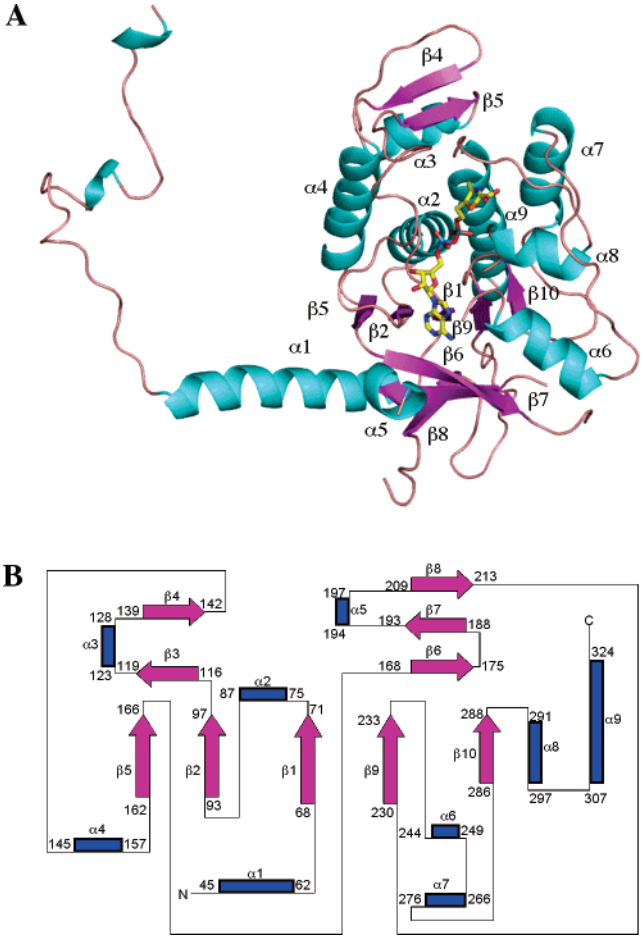


FIGURE 1: Structure of the Thi4 monomer. (A) Ribbon diagram showing 10 β-strands (magenta) and 9 α-helices (blue). The core of the monomer consists of a parallel β-sheet made of strands 1, 2, 5, 9, and 10 which are sandwiched between an antiparallel β-sheet below made of strands 6−8 and an α-helical bundle made of helices 2−4 and 7−9 from above. ADT is represented in stick form in the active site. (B) Topology diagram of Thi4 highlighting the core five-stranded β-sheet.

cleft of the protein in the proximity of helices α7 and α2 and is involved in numerous interactions with the N-terminal region of the adjacent monomer. The N-terminal region is largely disordered with the first 16 residues of chain A and 15 residues of chain B missing from the electron density. The monomer chain contains two 3_{10}-helices with the first α-helix beginning at residue 45. Thirty-one of the first 44 residues are either hydrophilic or alanine, with the few large hydrophobic residues, such as Phe38 and Phe40, pointing toward the hydrophobic core of the protein. The largely hydrophilic nature of the N-terminus may be due to its location on the surface of the octameric structure, thus contributing to the solubility of the protein.

Pro121 is found in the cis conformation and follows a loop consisting of residues 98−115. Within this region is a triglycyl group which gives increased flexibility to the loop. Isomerization of this proline could result in the breakage of some of the hydrogen bond contacts which hold ADT in place such as those contributed by Ser98 and Gly105, thereby reducing the binding energy of ADT and allowing for its release and further catalysis.

Octamer Structure. Thi4 exists as a homooctamer with the disordered and largely hydrophilic N-terminal regions

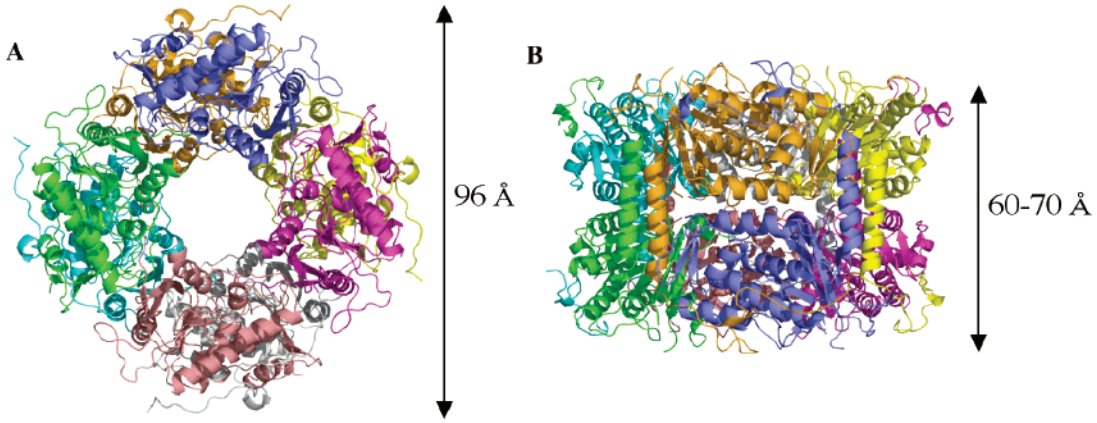


FIGURE 2: Structure of the Thi4 octamer. (A) Ribbon diagram viewed through the central cavity. (B) Ribbon diagram rotated 90° about the horizontal axis relative to panel A. Each monomer is labeled in a different color.

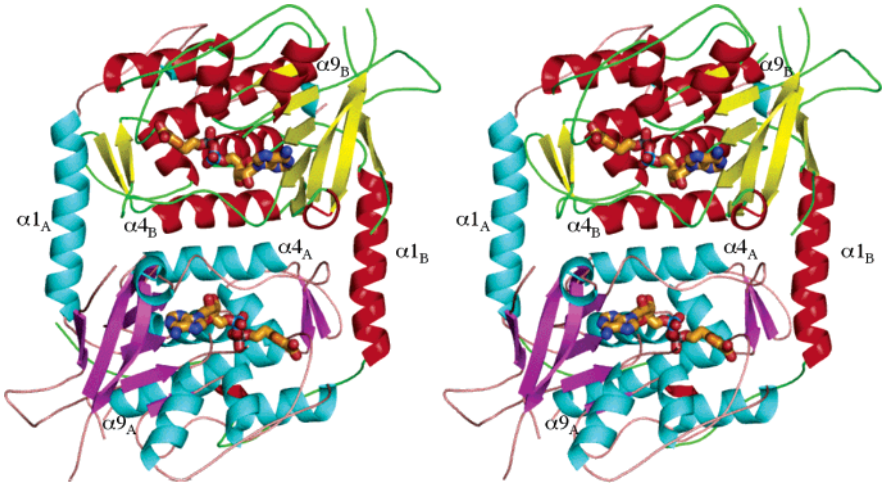


FIGURE 3: Stereoview of the tightly packed dimer. Helices and strands are colored blue and magenta, respectively, in chain A; helices and strands are colored red and yellow, respectively, in chain B. The bound ligand is represented in stick form and can be seen in the active site of each monomer. The first and last secondary structural elements (α1 and α9) are labeled for each monomer. The α4 helices in each monomer are labeled to illustrate the main interaction in the dimer.

located on the exterior of the molecule (Figure 2). The octamer has the shape of a ring with flattened sides with the inner and outer diameters measuring approximately 30 and 96 Å, respectively, and a thickness of approximately 60−70 Å. The active site of Thi4 is found in the 30 Å diameter inner ring, which allows sufficient room for the entry of substrate molecules. The outer and inner surfaces are lined with hydrophilic residues with the overall magnitude of the electrostatic potential being positive both inside and outside the ring. The octameric complex may be viewed as a tetramer of dimers with many interdimeric contacts at the interface between two monomers. The other major interface forms between adjacent dimers. A stereoview of the dimeric complex is shown in Figure 3. Two long α-helices, one from each chain, are seen flanking either side of a 2-fold axis and are arranged in an antiparallel fashion. These helices correspond to α4 (Figures 1B and 3) and facilitate the formation of the dimer largely through hydrophobic and van der Waals interactions, with the exception of residues Thr151 and Ser154 which form short hydrogen bonds to Ser154 (2.71 Å) and Thr151 (2.76 Å) of the adjacent monomer, respectively. Helix α1 passes orthogonally to the 2-fold axis at residue 58 and interacts with the two short antiparallel *β*-strands, *β*3 and *β*4, from chain B. The loop regions between *β*2 and *β*3 and between α8 and α9 from chain B

Table 3: Monomeric Interactions

A chain	B chain	interaction
Tyr56, Phe57	Phe113	*π*-stacking
Ala41	Arg119	N−NH1 hydrogen bond
His24, Trp35	His315	*π*-stacking
Glu33	Lys322	Oϵ−Nζ hydrogen bond
His24	Glu318	Oϵ−Nζ hydrogen bond
Thr23	Asn87	N−Oδ hydrogen bond

are also in the proximity of the N-terminal region. Table 3 lists hydrogen bonds and *π*-stacking interactions seen between chains A and B. The interface between individual dimers occurs mostly along the edge containing helix α1. The helix from one dimer (α1A) interacts with helix α1 from the adjacent dimer (α1B) through three hydrogen bonding interactions between Arg50 and Asp61 of adjacent monomers. Additionally, helix α1A hydrogen bonds to *β*8, α7, and the loop region between *β*6 and *β*7 in the second dimer. A summary of these hydrogen bonding interactions is given in Table 4. Most other interactions between dimers are the result of hydrophobic interactions.

Active Site. The small molecule adenosine diphospho-5-(*β*-ethyl)-4-methylthiazole-2-carboxylic acid (ADT) (*6*) was bound to the Thi4 active site, which is located near the inner ring of the octameric complex (Figure 4A). The adenosine moiety runs along the first *β*-strand in the fold characteristic

Table 4: Dimeric Interactions

A dimer	B dimer	interaction
Arg44	Glu212	Oϵ−Nϵ hydrogen bond
Glu45	Arg176	Oϵ−NH_2 hydrogen bond
Glu45	Met266	Oϵ−NH hydrogen bond
Ser46	Glu212	Oδ−Oϵ hydrogen bond
Ser46	Arg50	O−N hydrogen bond
Ser49	Asn209	Oγ−Oδ hydrogen bond
Arg50	Asp61	Nϵ−Oδ hydrogen bond
Arg50	Asp61	Nϵ−O hydrogen bond
Arg50	Asp61	NH_2−O hydrogen bond

of a dinucleotide binding site. The purine base is stabilized through π-stacking interactions between Phe241 and Phe244. Hydrogen bonding requirements for the N1 and N6 atoms of adenine are satisfied through the amide nitrogen atom and carbonyl oxygen atom of Val170, respectively. The N3 atom forms a hydrogen bond with the amide nitrogen atom of Ser98 with a length of 3.22 Å. The 2′- and 3′-hydroxyl groups on the ribose ring are hydrogen bonded to Glu97, with the 2′-oxygen atom 2.54 Å and the 3′-oxygen atom 2.82 Å from the nearest oxygen atom of the Glu97 side chain. The 2′-hydroxyl group also forms a hydrogen bond with Ser98 and a water molecule with a length of 2.93 Å. The 3′-hydroxyl group forms additional hydrogen bonds with the amide nitrogen atom of Gly104 and another water molecule.

There are two hydrogen bonds to the O1 atom of the α-phosphate. One bond is to the amide nitrogen atom of Gly105, which is one of the three glycine residues seen in the triglycyl segment of the loop region connecting strands β2 and β3 and the other to a water molecule. Two hydrogen bonds with the α-phosphate O2 atom are formed through water molecules, but no bonds are made to side chains or the main chain. O1 from the β-phosphate forms three hydrogen bonds: one to a water molecule and one each to the hydroxyl group and an amide nitrogen atom of Ser76. Two hydrogen bonds are formed by O2: one to a water molecule and the other to the amide nitrogen atom of Met291. The O3 atom which bridges the pyrophosphate group and the thiazole moiety is near the sulfur atom of Met309. This methionine together with Met291 and Met302 forms an interesting 3-fold symmetric triad just above the thiazole ring (Figure 4B).

The thiazole moiety forms several hydrogen bonds with the protein, most of which are using the carboxylate group (Figure 4A). A single hydrogen bond exists between the nitrogen atom of the ring and a carbonyl group of Gly303. The carboxylate group forms two hydrogen bonds to the guanidinium moiety of Arg301, with each oxygen atom binding to a different nitrogen atom. The O1 atom from the carboxylate group also hydrogen bonds to the amide nitrogen atom of Gly303, while the O2 atom of the carboxylate group binds to the Nϵ atom of His237. The thiazole ring is held in place through interactions with Asp207 from an adjacent dimer, which is positioned just above the ring facing the inner surface of the octameric ring. One oxygen atom from Asp207 lies directly above the thiazole nitrogen at a distance of 3.29 Å, while the other oxygen atom lies above the thiazole sulfur at a distance of 3.62 Å. The carboxylic acid of Asp207 forms a hydrogen bond with the carbonyl oxygen atom of Gly303 from another monomer in the octameric complex. The other face of the thiazole ring is held in place by the methionine triad: Met291, Met302, and Met309 (Figure 4B).

The electron density maps for ADT show clear density for the ADP moiety but are less clear in the region of the thiazole ring (Figure 5). The largest peak height corresponds to the sulfur atom as expected, but the density is weaker than expected in the region of the N−C2 bond. The nonideality of the final electron density map probably arises from the presence of ADT as well as minor contamination from one or two protein-bound adenylated metabolites (*6*).

DISCUSSION

Comparison of Thi4 and A. thaliana Thi1. The molecular search model used to determine the structure of the *S. cerevisiae* Thi4 was the ortholog of Thi4 in *A. thaliana*, Thi1 (PDB entry 1RP0) (*14*). The level of sequence identity between Thi1 and Thi4 is 67% with only 24 of 289 Thi1 residues in gap regions. Though the topologies of Thi4 and Thi1 are the same, Thi4 has 326 residues per chain while Thi1 has only 284. Two structural features account for most of this 42-amino acid difference. First, Thi4 contains an N-terminal extension of 32 residues compared to Thi1, and second, Thi4 contains a 15-amino acid insertion relative to Thi1 between β8 and β9. The middle of the loop contains a mostly hydrophilic segment (KNETRMK), faces a solvent channel in the crystal structure, and is located away from the active site. The loop is absent in most Thi4 othologs and probably does not directly affect the catalytic function.

Comparison of the Thi4 and Thi1 active sites reveals a tightly conserved arrangement around the thiazole moiety, despite a rotation of the thiazole ring of approximately 170° about the C5−C6 bond between the two structures. Figure 6 shows a superposition of ADT in both Thi1 and Thi4, highlighting residues that interact with the thiazole moiety. Neither of the sulfur atoms in either structure forms hydrogen bonds or close contacts with the surrounding residues, but the hydrogen bond between the thiazole nitrogen atom and Gly303 in the Thi4 structure is absent in Thi1. Because the thiazole ring in Thi1 is rotated, the nitrogen atom in the thiazole of Thi1 binds to His189, Asp190, and a Zn^{2+} ion. Asp238, which corresponds to Asp190 in Thi1, has poor side chain density in the A chain and therefore has been modeled as alanine. Also, there is no density at all for Asp238 in the B chain of Thi4, so a comparison of binding between proteins for this residue is not possible. However, the electron density peak for the ADT sulfur atom in the Thi4 complex is very high, so the conformation about the C5−C6 bond is well-defined. The electron density for Thi1 shows no density for the C5 atom, which may indicate that different metabolites are present in the two structures.

Conserved Residues. A sequence alignment carried out with 14 different orthologs of Thi4 using ClustalW (*15*) along with secondary structural features is shown in Figure 7. Of those conserved residues, most are involved in interactions either within the protein or with its bound ligand. A conserved region in Thi4 corresponding to residues 68−78 aligns well with a dinucleotide binding motif (*16*). The characteristic motif of $\phi\phi$GXGXXG, where ϕ is any hydrophobic residue and X is any residue, is retained among all the orthologs, with the exception of that from *Methanococcoides burtonii* in which the third glycine is replaced with an asparagine. Many entirely conserved residues are involved

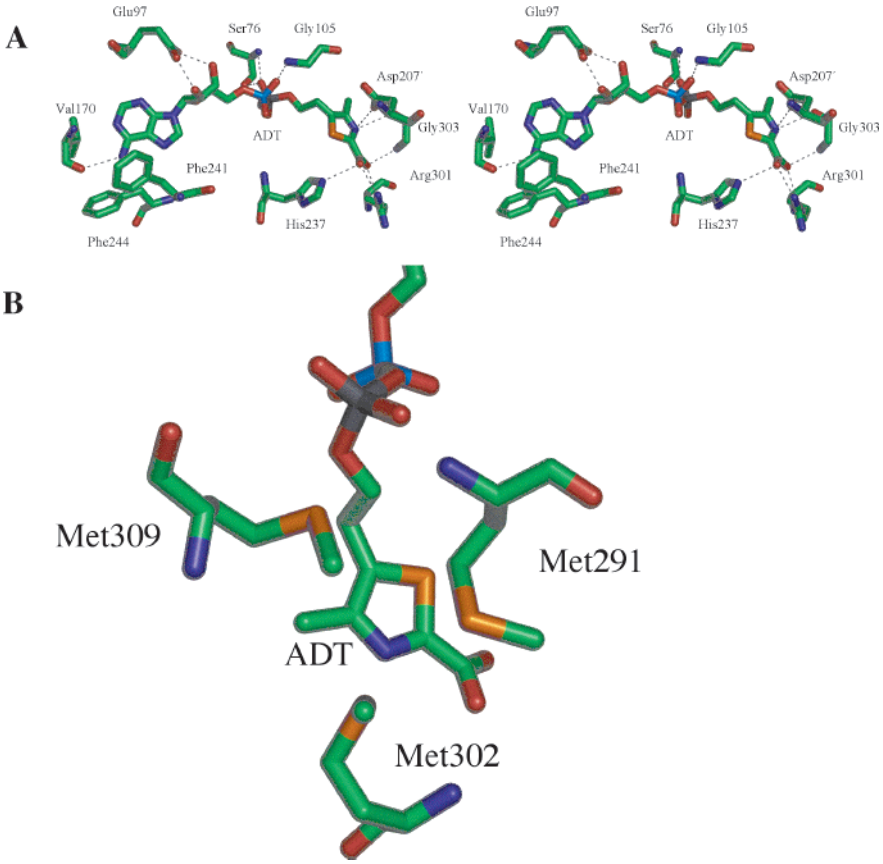


FIGURE 4: Thi4 active site. (A) Stereoview of the Thi4 ADT binding site. All amino acid side chains responsible for binding of ADT are shown. All residues are from a single chain, with the exception of Asp207′, which comes from another Thi4 monomer. Hydrogen bonds between the residues and the ligand are represented as gray dashed lines. (B) View of the thiazole moiety of ADT and the methionine triad which lies on one face of the heterocyclic ring.

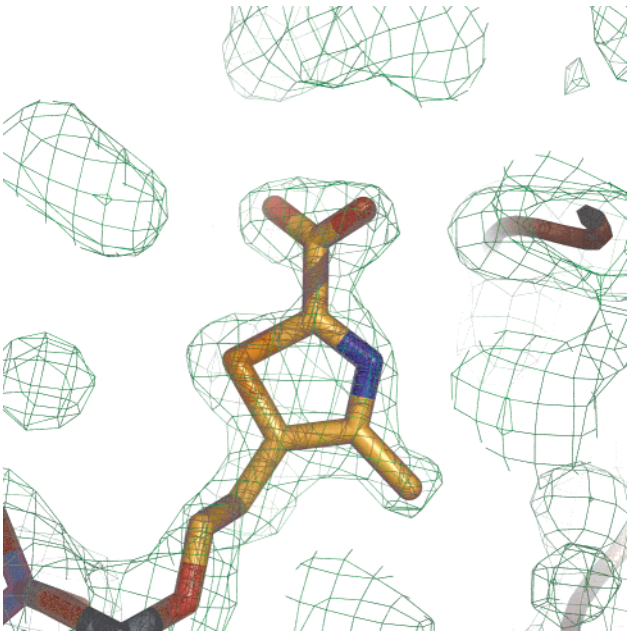


FIGURE 5: Electron density for ADT. The $2F_o - F_c$ map is contoured at 1.0σ. This figure was reproduced with permission from ref *6*.

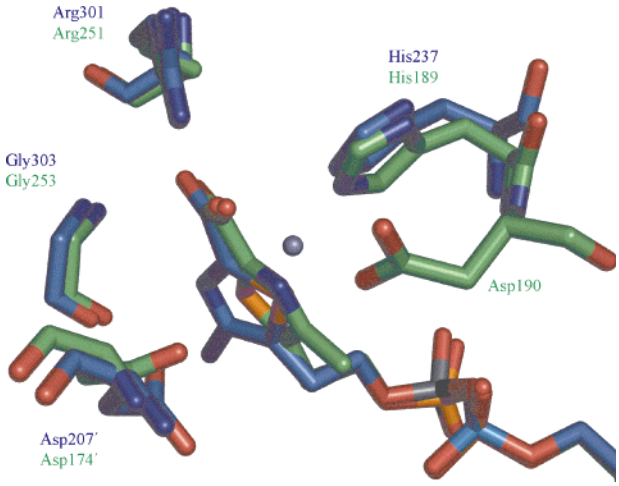


FIGURE 6: Overlay of Thi1 (green) and Thi4 (blue) active site residues in contact with the thiazole moiety. Asp207′ or Asp174′ contributes from an adjacent Thi4 or Thi1 monomer, respectively. The Zn^{2+} in Thi1 is shown as a blue sphere.

in ligand binding, including Glu97, Gly105, Asp207, His237, and Arg301 in Thi4. Arg301 forms a hydrogen bond to both oxygen atoms on the carboxylate group attached to C2 of the thiazole ring. Gly105 is part of a triglycyl region within a flexible loop, with the amide nitrogen of Gly105 satisfying a hydrogen bond with an oxygen atom on the α-phosphate of ADT. Asp207 is found positioned above the thiazole ring with one oxygen atom at a distance of 3.29 Å, while the other oxygen atom lies above the thiazole sulfur atom at a distance of 3.62 Å. The carboxylate of Asp207 is positioned by a water molecule, which bridges the carboxylate moiety and the carbonyl oxygen atom of Gly105 from an adjacent dimer of the octameric complex. Gly303 is entirely conserved and hydrogen bonds with the thiazole ring through its

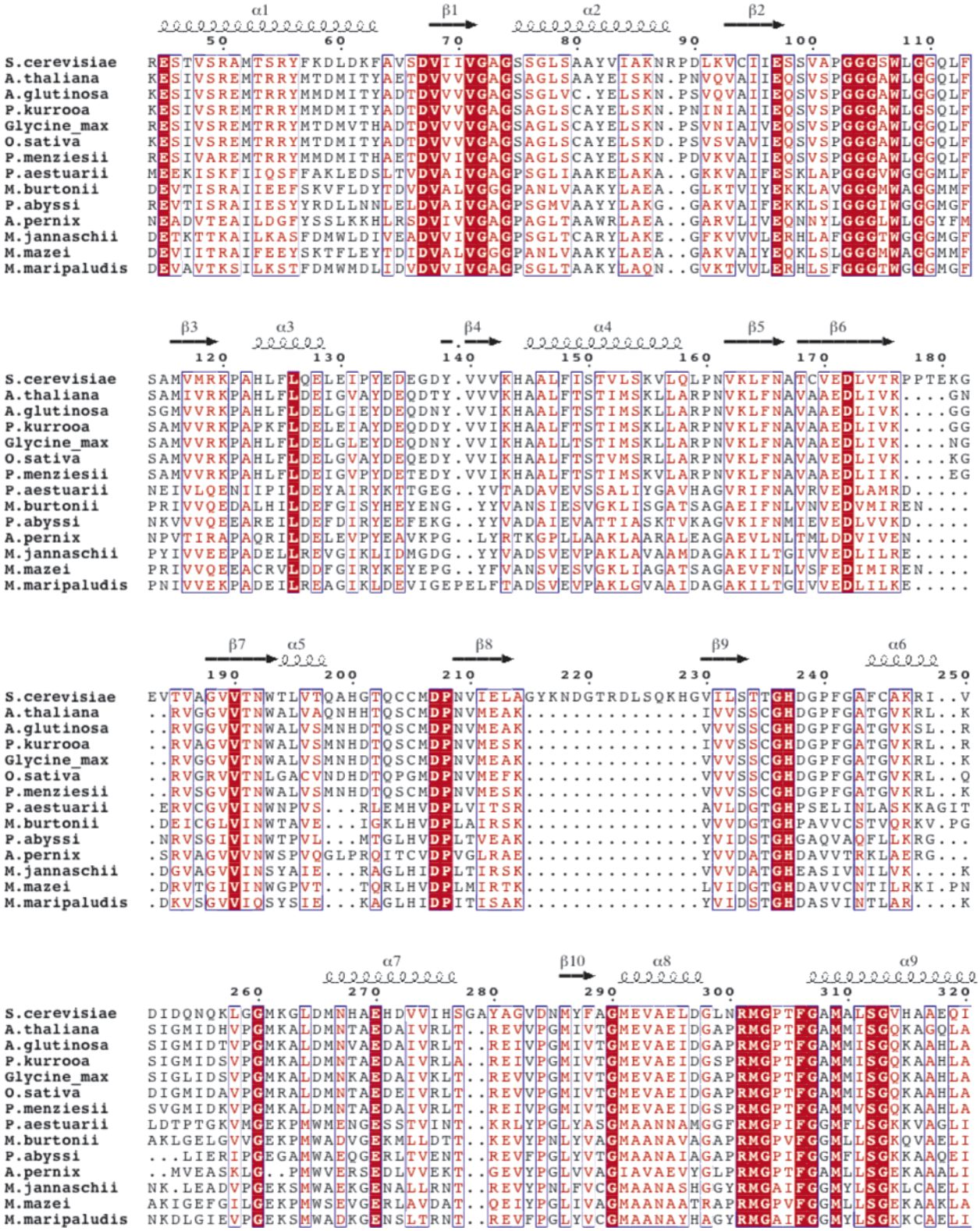


FIGURE 7: Sequence alignment of Thi4 from *S. cerevisiae* with 14 Thi4 orthologs. The Thi4 secondary structures are listed above the corresponding sequence. Mostly conserved residues are colored pink, while those that are exclusively conserved are colored red.

backbone nitrogen atom. The structure reveals that Asp207 hydrogen bonds to Gly303 and positions its backbone nitrogen atom to hydrogen bond to the carboxylate group from the thiazole ring. His237 forms a hydrogen bond with the carboxylate group of the thiazole moiety of ADT via the Nϵ atom and may be involved in catalysis by acting as an active site base through activation by a water molecule which is hydrogen bonded to the Nδ atom of the imidazole ring (see below). Glu270 forms hydrogen bonds with the hydroxyl group of Thr305 and the amide nitrogen of Ala308. Both of these residues are positioned after the loop containing Arg301, which is key in positioning the carboxylate group in ADT. Glu270 forms two hydrogen bonds to the guanidinium moiety of Arg119 which is the last residue in the last secondary structural element, β3, that precedes the *cis*-proline Pro121.

Other exclusively conserved residues within the structure do not directly bind to the ADT molecule. Glu45 forms a salt bridge with Arg176 and creates a key dimer−dimer interaction to stabilize the octameric complex. Though the arginine residue is not entirely conserved, it is usually replaced by lysine. Asp67 forms a hydrogen bond with the amide nitrogen of Lys92 and binds three water molecules which are on the outer surface of the octamer. This interaction serves to orient β2, which contains the Glu97 residue essential for positioning of the ribose ring of ADT. Gly109 forms a hydrogen bond with Phe113 but is also part of the flexible loop region, which also contains the triglycyl region. Asp172 plays a key role in the positioning of Arg301 to bind the carboxylate group of ADT. First, Asp172 is hydrogen bonded to Arg248 which positions the side chain carboxylate in plane with the guanidinium moiety of Arg301. This in turn positions the guanidinium group to lie above the carboxylate of ADT with a geometry that favors strong hydrogen bonds. Though Arg248 is not entirely conserved, it is either arginine or lysine, either of which could orient Asp172 through a salt bridge. Trp193 from one monomer forms π-stacking interactions with a Trp193 in another monomer and is buried among several aromatic residues, including Tyr56, Phe57, Phe113, and Phe165 from both chains. Conservation of Trp193 is probably key in octamer stabilization.

Mostly conserved Val170, Phe241, Phe244, Met291, and Gly303 interact with the ligand. Both phenylalanine residues form π-stacking interactions with the adenine moiety of ADT. Val170 contributes to hydrogen bonding interactions through its amide nitrogen and carbonyl oxygen atoms to N1 and N6 of the adenine ring, respectively. Though this residue is not entirely conserved, all Thi4 orthologs contain a small hydrophobic residue in this position, suggesting that a hydrophobic environment near the side chains is important for orienting the hydrogen bonding main chain atoms. The methionine triad (Figure 4B) found in the helix−turn−helix motif encompassed by helices α8 and α9 is entirely conserved, with the exception of only one organism (*Aeropyrum pernix*) for which the equivalent of Met291 is isoleucine. None of the three methionine residues has significant side chain interactions with the exception of Met309, which forms a short 3.3 Å contact between its sulfur atom and the bridging oxygen atom between the β-phosphate and the thiazole ring. The reason for the strong conservation of this methionine triad is currently unknown.

Comparison with Other Protein Structures. A structural similarity search using DALI (*17*) revealed that all of the top hits belong to the glutathione reductase type II family fold (GR_2) (*18*). A summary of *Z* scores and structural overlaps for the top 10 structures from the DALI output are listed in Table 5. The listed enzymes include phenol hydroxylase (PDB entry 1FOH), fumerate reductase (PDB entry 1QJD), and the prokaryotic thiamin biosynthesis protein glycine oxidase (ThiO) (PDB entry 1NG3), which have the same topology in the ADT binding region. This GR_2 family is characterized by a $GxGxxG(x)_nE/D$ sequence, where *n* can range from 16 to 19. This sequence in Thi4 corresponds to residues 72−97, where $n = 20$, and the 97th residue is an exclusively conserved glutamic acid. The glutamic acid is responsible for hydrogen bonding to the 2′- and 3′-hydroxyl groups of the ribose sugar, a common feature in nucleotide binding proteins. The most unexpected result of Thi4 being a member of the GR_2 family is that every member to which Thi4 is significantly topologically similar via DALI is a flavoenzyme. However, our recently published studies suggest that Thi4 utilizes NAD as its substrate (*6*). Classic NAD-dependent oxidoreductases such as alcohol dehydrogenase, lactate dehydrogenase, and 3-glyceraldehyde-3-phosphate dehydrogenase exhibit a Rossmann fold characteristic of NAD binding proteins (*19*), but the fold in Thi4 has a clearly different topology (Figure 1B).

Table 5: DALI Search Results

PDB entry	*Z* score	LALI[a]	rmsd[b]	% ID[c]
1RP0	36.8	255	1.0	57
1QJD	17.5	212	3.5	18
1TRB	16.1	161	2.4	13
1HYU	16.1	181	8.4	14
1PBE	15.5	181	2.7	13
1KDG	15.4	193	2.6	21
1NG3	15.2	180	2.4	15
1QLA	15.1	198	2.8	20
1FOH	14.6	190	2.9	17
1B8S	14.3	196	2.9	13

[a] Total number of equivalent residues. [b] Positional root-mean-square deviation of superimposed Cα atoms in angstroms. [c] Percentage of sequence identity over equivalent positions.

A structural overlay between Thi4 and ThiO is shown in Figure 8. The FAD binding domain between Thi4 and ThiO overlaps well, with the rest of ThiO having no similarity to the dimeric interface between Thi4 monomers or another Thi4 monomer within the octameric complex. The ADT in Thi4 and FAD in ThiO overlap well in the ADP moiety but diverge after the β-phosphate. Thus, the GR_2 fold serves both as a cofactor binding topology for flavoenzymes and as a fold that can utilize NAD as a substrate to generate a critical component in yeast thiamin biosynthesis. When considering Thi4 from an evolutionary standpoint, it represents an unprecedented scenario in which a protein fold that once served to bind FAD as a cofactor evolved to bind and catalyze NAD as a substrate without any topological rearrangement.

Mechanistic Implications. Previous studies, using labeled precursors, have demonstrated that the thiazole heterocycle of thiamin is biosynthesized in yeast from glycine, cysteine, and an unidentified C5 sugar (*2*−*4*). The Thi4 structure with ADT bound at the active site shows that the unidentified C5 sugar is linked to ADP. This observation suggests that the C5 sugar precursor to the thiazole is derived from an adenylated metabolite synthesized in both *S. cerevisiae* and *E. coli*, such as NAD, NADH, FAD, or FADH (Figure 9). NAD or NADH seems more likely than FAD or FADH because pyridine can be more readily cleaved from ribose than isoalloxazine from ribulose using a NAD-glycohydrolase-like activity. In addition, the pentose sugar of NAD is in the correct oxidation state to form the thiazole moiety, whereas in FAD and FADH, it needs to be oxidized.

While the structure of the yeast thiazole synthase identifies NAD as the most likely thiazole precursor and provides key insights into the mechanism of thiazole formation in eukaryotes, much still remains to be discovered about this interesting protein. The mechanistic proposal outlined in Scheme 2 needs to be tested and refined, and the sulfur source and the mechanism of sulfur transfer remain to be

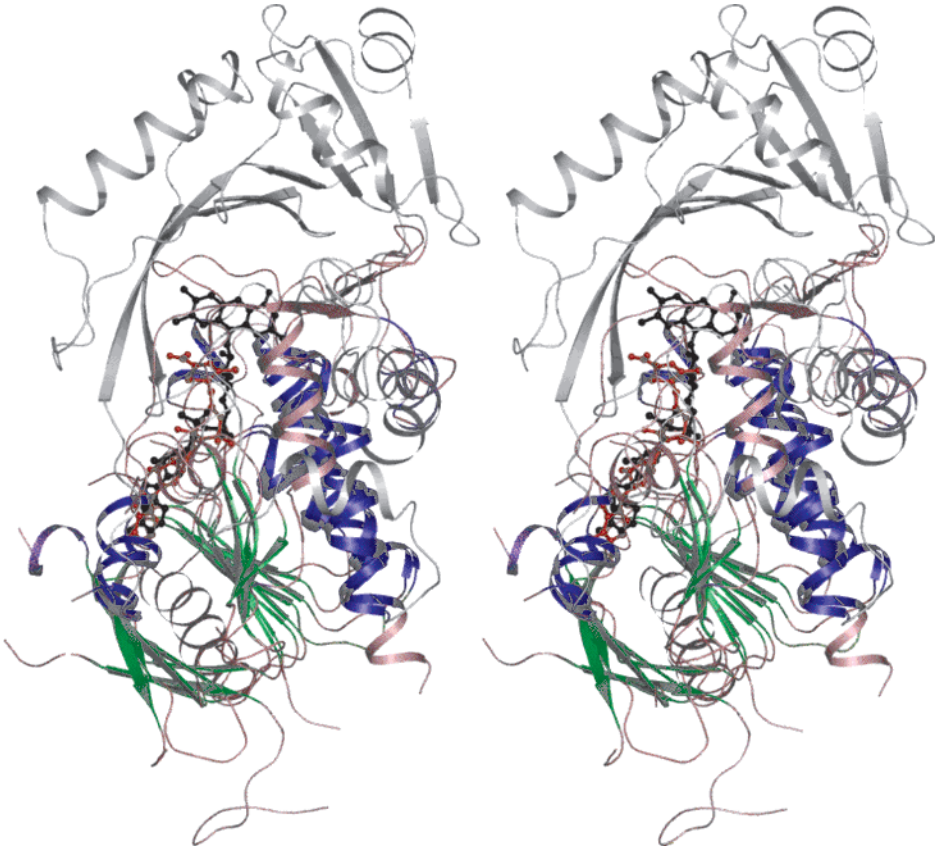


FIGURE 8: Stereoview of a superposition of Thi4 and ThiO showing the overlapping dinucleotide binding domains. Overlapping secondary structural elements are colored blue (α-helices) or green (*β*-strands). Nonoverlapping secondary structural elements are colored pink (Thi4) or gray (ThiO). FAD bound to ThiO is shown as a black ball-and-stick model. ADT bound to Thi4 is shown as a red ball-and-stick model.

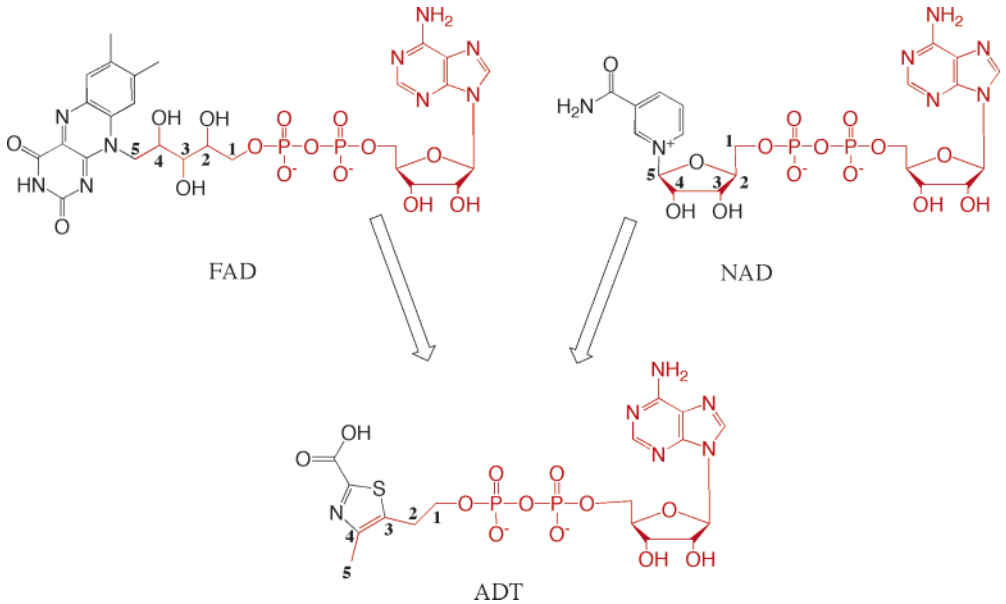


FIGURE 9: Chemical structures of FAD, NAD, and ADT. Atoms common to all three structures are colored red, and the five carbon atoms corresponding to the thiazole are numbered. The remaining atoms of the thiazole are derived from glycine and an unidentified sulfur donor.

determined. Also, the biological significance of the stable enzyme−product complex is unknown. In addition, four other properties of Thi4 and its orthologs are worth mentioning. In the exponential growth phase, Thi4 is an abundant protein (1.5% of the entire proteome) (*20*) in *Neurospora crassa*, grown in the absence of thiamin. Genetic studies indicate that the *Arabidopsis* Thi4 (Thi1) protects mitochondrial DNA from damage (*21*). The *N. crassa* Thi4 (CyPBP37) is posttranslationally modified and forms a stable complex with a cyclophilin (NcCyP41) (*22*). *Fusarium oxysporum* Thi4 (sti35) is induced by various stress conditions (*23*). These observations suggest that in addition to thiamin biosynthesis, Thi4 is likely to have additional functions in the cell.

CONCLUSIONS

Thiamin was the first vitamin discovered but the last to have its biosynthesis elucidated, due to the novelty and

complexity of its biosynthetic chemistry. While the major features of the bacterial pathway have now been elucidated, little is known about the biosynthesis of this important vitamin in eukaryotes. This paper describes the structure of the *S. cerevisiae* thiazole synthase (Thi4) with adenosine diphospho-5-(β-ethyl)-4-methylthiazole-2-carboxylic acid (ADT) bound at its active site. The isolation of this product identifies NAD as the most likely precursor to the thiamin thiazole, suggests that its biosynthesis is mechanistically related to protein ADP ribosylation, which is an important cellular regulatory process, and provides the first mechanistic insights into the biosynthesis of the thiamin thiazole in eukaryotes. Additionally, the Thi4 structure reveals the first protein structure with a GR_2 domain that binds NAD instead of FAD, raising interesting questions about how this protein evolved from a flavoenzyme to an NAD binding enzyme.

ACKNOWLEDGMENT

We thank Leslie Kinsland for help in the preparation of the manuscript, Cynthia Kinsland for the *THI4* overexpression construct, the staff at CHESS beamline A1 [supported by grants from the National Science Foundation (DMR-0225180) and the National Institutes of Health (RR-01646)] and Professor Glaucius Oliva for providing the Thi1 coordinates prior to their release in the Protein Data Bank.

REFERENCES

1. Settembre, E., Begley, T. P., and Ealick, S. E. (2003) Structural biology of enzymes of the thiamin biosynthesis pathway, *Curr. Opin. Struct. Biol. 13*, 739−747.
2. White, R. L., and Spenser, I. D. (1979) Thiamin biosynthesis in *Saccharomyces cerevisiae*. Origin of carbon-2 of the thiazole moiety, *Biochem. J. 179*, 315−325.
3. White, R. L., and Spenser, I. D. (1979) Biosynthesis of vitamin B1 in yeast. Origin of the thiazole unit, *J. Am. Chem. Soc. 101*, 5102−5104.
4. White, R. L., and Spenser, I. D. (1982) Thiamin biosynthesis in yeast. Origin of the five-carbon unit of the thiazole moiety, *J. Am. Chem. Soc. 104*, 4934−4943.
5. Praekelt, U. M., Byrne, K. L., and Meacock, P. A. (1994) Regulation of THI4 (MOL1), a thiamine-biosynthetic gene of *Saccharomyces cerevisiae*, *Yeast 10*, 481−490.
6. Chatterjee, A., Jurgenson, C. T., Schroeder, F. C., Ealick, S. E., and Begley, T. P. (2006) Thiamin Biosynthesis in Eukaryotes: Characterization of the Enzyme-Bound Product of Thiazole Synthase from *Saccharomyces cerevisiae* and Its Implications in Thiazole Biosynthesis, *J. Am. Chem. Soc. 128*, 7158−7159.
7. Rubinson, K. A., Ladner, J. E., Tordova, M., and Gilliland, G. L. (2000) Cryosalts: Suppression of ice formation in macromolecular crystallography, *Acta Crystallogr. D56*, 996−1001.
8. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307−326.
9. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D54*, 905−921.
10. Collaborative Computational Project Number 4 (1994) The CCP-4 suite: Programs for protein crystallography, *Acta Crystallogr. D50*, 760−763.
11. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr. A47* (Part 2), 110−119.
12. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics, *Acta Crystallogr. D60*, 2126−2132.
13. DeLano, W. L. (2002) *PyMol*, DeLano Scientific, San Carlos, CA.
14. Ribeiro, A., Praekelt, U., Akkermans, A. D., Meacock, P. A., van Kammen, A., Bisseling, T., and Pawlowski, K. (1996) Identification of agthi1, whose product is involved in biosynthesis of the thiamine precursor thiazole, in actinorhizal nodules of *Alnus glutinosa*, *Plant J. 10*, 361−368.
15. Pearson, W. R., and Lipman, D. J. (1988) Improved tools for biological sequence comparison, *Proc. Natl. Acad. Sci. U.S.A. 85*, 2444−2448.
16. Moller, W., and Amons, R. (1985) Phosphate-binding sequences in nucleotide-binding proteins, *FEBS Lett. 186*, 1−7.
17. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol. 233*, 123−138.
18. Dym, O., and Eisenberg, D. (2001) Sequence-structure analysis of FAD-containing proteins, *Protein Sci. 10*, 1712−1728.
19. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) Chemical and biological evolution of nucleotide-binding protein, *Nature 250*, 194−199.
20. Faou, P., and Tropschug, M. (2004) *Neurospora crassa* CyPBP37: A cytosolic stress protein that is able to replace yeast Thi4p function in the synthesis of vitamin B1, *J. Mol. Biol. 344*, 1147−1157.
21. Machado, C. R., Praekelt, U. M., de Oliveira, R. C., Barbosa, A. C., Byrne, K. L., Meacock, P. A., and Menck, C. F. (1997) Dual role for the yeast THI4 gene in thiamine biosynthesis and DNA damage tolerance, *J. Mol. Biol. 273*, 114−121.
22. Faou, P., and Tropschug, M. (2003) A novel binding protein for a member of CyP40-type Cyclophilins: *N. crassa* CyPBP37, a growth and thiamine regulated protein homolog to yeast Thi4p, *J. Mol. Biol. 333*, 831−844.
23. Thanonkeo, P., Akiyama, K., Jain, S., and Takata, R. (2000) Targeted disruption of sti35, a stress-responsive gene in phytopathogenic fungus *Fusarium oxysporum*, *Curr. Microbiol. 41*, 284−289.

BI061025Z