BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 8929–8936 8929
DOI: 10.1021/bi101209t

HMP Binding Protein ThiY and HMP-P Synthase THI5 Are Structural Homologues[†,‡]

Shridhar Bale,[§] Kanagalaghatta R. Rajashankar,[§,∥] Kay Perry,[§,∥] Tadhg P. Begley,*[,⊥] and Steven E. Ealick*[,§]

[§]*Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853,* [∥]*Northeastern Collaborative Access Team, Building 436, Argonne National Laboratory, Argonne, Illinois 60439, and* [⊥]*Department of Chemistry, Texas A&M University, College Station, Texas 77843*

Received July 30, 2010; Revised Manuscript Received September 7, 2010

ABSTRACT: The ATP-binding cassette transporter system ThiXYZ transports *N*-formyl-4-amino-5-(aminomethyl)-2-methylpyrimidine (FAMP), a thiamin salvage pathway intermediate, into cells. FAMP is then converted to 4-amino-5-(hydroxymethyl)-2-methylpyrimidine (HMP) and recycled into the thiamin biosynthetic pathway. ThiY is the periplasmic substrate binding protein of the ThiXYZ system and delivers the substrate FAMP to the transmembrane domain. We report the crystal structure of *Bacillus halodurans* ThiY with FAMP bound at 2.4 Å resolution determined by single-wavelength anomalous diffraction phasing. The crystal structure reveals that ThiY belongs to the group II periplasmic binding protein family. The closest structural homologues of ThiY are periplasmic binding proteins involved in alkanesulfonate/nitrate and bicarbonate transport. ThiY is also structurally homologous to thiamin binding protein (TbpA) and to thiaminase-I. THI5 is responsible for the synthesis of 4-amino-5-(hydroxymethyl)-2-methylpyrimidine phosphate in the thiamin biosynthetic pathway of eukaryotes and is approximately 25% identical in sequence with ThiY. A homology model of *Saccharomyces cerevisiae* THI5 was generated on the basis of the structure of ThiY. Many features of the thiamin pyrimidine binding site are shared between ThiY and THI5, suggesting a common ancestor.

Thiamin pyrophosphate (ThDP)[1] is an essential cofactor in all living systems and is involved in branched-chain amino acid synthesis and carbohydrate metabolism. ThDP is implicated in the stabilization of an acyl carbanion intermediate formed in a variety of enzymatic reactions. Bacteria and Archaea, as well as certain lower eukaryotes, synthesize thiamin. Higher eukaryotes cannot produce thiamin and require it as a dietary supplement. Deficiency of thiamin in humans causes beriberi and Wernicke-Korsakoff syndrome (*1*). In all thiamin biosynthetic pathways, the thiazole and pyrimidine moieties of thiamin are synthesized separately and linked to form thiamin monophosphate (ThMP). The thiamin biosynthetic pathway is well studied in prokaryotes, and the details of the eukaryotic pathway are beginning to emerge (*2*−*5*). In prokaryotes, the thiazole ring is synthesized from 1-deoxy-D-xylulose 5-phosphate, cysteine, and glycine (or tyrosine in Gram-negative bacteria) by five gene products. The pyrimidine ring is synthesized by ThiC, a radical SAM enzyme, from 5-aminoimidazole ribonucleotide (*4*). In eukaryotes, the thiazole ring is formed from nicotinamide adenine dinucleotide, glycine, and cysteine by one enzyme, THI4. The mechanistic details of the formation of the thiazole moiety are beginning to emerge, and several steps of the reaction have been biochemically and structurally characterized (*5*, *6*). The pyrimidine moiety is synthesized from histidine and pyridoxal 5′-phosphate by THI5 (*7*−*9*). The THI5 family of enzymes has a conserved CCCXC motif; however, no mechanistic or structural details are available for this complex reaction.

In addition to biosynthesis, most organisms have a thiamin uptake system (*10*−*12*). In *Escherichia coli*, *thiPBQ* encodes an ABC transporter that imports thiamin, thiamin monophosphate (ThMP), and ThDP into the cell (*13*, *14*). The crystal structure of the periplasmic thiamin binding protein (TbpA), with ThMP bound, reveals that the protein belongs to the group II periplasmic binding protein (PBP) family. In addition, the structural similarity of TbpA with thiaminase-I, a thiamin-degrading enzyme, suggests that these proteins evolved from a common ancestor (*15*).

Binding proteins in Gram-negative bacteria are found in the periplasm; however, in Gram-positive bacteria, which lack an outer membrane, binding proteins are normally tethered to the cell membrane through acylation of a cysteine residue in a lipoprotein signal sequence (*16*, *17*). ThiY contains a cysteine residue at position 19, which is preceded by a hydrophobic N-terminus. Cys19 is conserved in ThiY orthologs from other *Bacillus* species and is believed to provide the membrane anchor for ThiY. The hydrophobic lipoprotein signal sequence was removed for structural studies.

Thiamin is degraded in basic soil to *N*-formyl-4-amino-5-(aminomethyl)-2-methylpyrimidine (FAMP) (*18*). In *Bacillus halodurans*, FAMP is transported into the cell by the ABC transporter encoded by *thiXYZ*, where it is deformylated and hydrolyzed to HMP, which is then incorporated into the thiamin biosynthetic pathway (Scheme 1). ThiY is the periplasmic FAMP

[†]This work was supported by National Institutes of Health Grants DK 44083 (T.P.B.) and DK 67081 (S.E.E.). S.E.E. is indebted to the W. M. Keck Foundation and the Lucille P. Markey Charitable Trust.

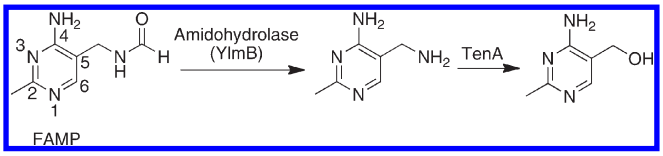
[‡]Coordinates have been deposited as Protein Data Bank entry 3IX1.

*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.

[1]Abbreviations: FAMP, *N*-formyl-4-amino-5-(aminomethyl)-2-methylpyrimidine; HMP, 4-amino-5-(hydroxymethyl)-2-methylpyrimidine; ABC, ATP-binding cassette; PBP, periplasmic binding protein; TbpA, thiamin binding protein; SAD, single-wavelength anomalous diffraction; NiBP, nitrate binding protein; ThMP, thiamin monophosphate; ThDP, thiamin pyrophosphate; Tris, tris(hydroxymethyl)aminomethane; HMP-P, HMP phosphate; PLP, pyridoxal 5′-phosphate.

Published on Web 09/28/2010

Scheme 1



binding component of the ABC transporter. Here we report the 2.4 Å resolution crystal structure of the *B. halodurans* ThiY−FAMP complex and compare the structure with those of the thiamin-related enzymes TbpA and thiaminase-I. We also report a homology model for THI5, the HMP phosphate (HMP-P) biosynthetic protein from yeast, based on the structure of ThiY. The modeling studies predict that the pyrimidine binding site is conserved in ThiY and THI5, suggesting that these proteins evolved from a common ancestor. The model also provides some preliminary insights into HMP-P biosynthesis in eukaryotes.

MATERIALS AND METHODS

Protein Expression and Purification. The construction of the *B. halodurans* ThiY overexpression strain was described previously (*18*). In this construct, the 20 N-terminal amino acids were deleted to remove a hydrophobic membrane anchor and increase protein solubility. The gene was cloned into the pET16b plasmid and transformed into methionine auxotrophic strain B834(DE3) of *E. coli*. A starter culture (10 mL) was grown overnight at 37 °C in LB medium containing 100 mg/mL ampicillin. The starter culture was harvested by centrifugation at 2000 rpm for 15 min, resuspended in 50 mL of minimal medium, and then transferred to 1 L of minimal medium. The minimal medium contained M9 salts supplemented with each amino acid (except for methionine, which was replaced with L-selenomethionine) at 40 mg/mL, 100 mg/mL ampicillin, 0.4% glucose, 2 mM $MgSO_4$, 25 mg/mL $FeSO_4 \cdot 7H_2O$, 0.1 mM $CaCl_2$, and a 1% MEM vitamin solution. The cells were grown to an OD_{600} of 0.6 at 37 °C. Protein expression was induced by addition of 1 mM isopropyl 1-thio-β-D-galactopyranoside, and the temperature was reduced to 15 °C. Cells were harvested after 15 h and stored at −80 °C.

The cell pellet was thawed and resuspended in wash buffer [20 mM tris(hydroxymethyl)aminomethane (Tris) (pH 7.5), 150 mM NaCl, 10 mM imidazole, and 2 mM β-mercaptoethanol]. The cells were lysed by sonication and centrifuged at 24000*g* for 30 min to separate the lysate from the cell debris. The lysate was passed through a Ni-NTA column pre-equilibrated with wash buffer. The column was washed with 150 mL of wash buffer followed by 50 mL of wash buffer containing 35 mM imidazole. The protein was eluted with wash buffer containing 200 mM imidazole and passed through a Sephadex G-75 column pre-equilibrated with 20 mM Tris (pH 7.5), 150 mM NaCl, and 3 mM dithiothreitol. Fractions containing protein were identified by UV absorbance at 280 nm and were collected, pooled, concentrated to 10 mg/mL, and stored in aliquots at −80 °C.

Crystallization Conditions. The protein was thawed and incubated with a 3−4-fold molar excess of FAMP for 2 h prior to crystallization. The crystals were grown using the hanging drop vapor diffusion method at 18 °C in 3 M NaCl, 0.1 M sodium acetate (pH 4.5), and 0.1 M lithium sulfate. Clusters of very thin plates with dimensions of 0.15 mm × 0.1 mm × 0.005 mm grew in a week.

Data Collection and Processing. All data were collected at NE-CAT beamline 24-ID-E at the Advanced Photon Source.

Table 1: Data Collection Statistics for ThiY

	set 1	set 2
wavelength (Å)	0.9792	0.9792
no. of crystals	2	1
space group	$P2_12_12_1$	$P2_12_12_1$
a (Å)	60.0	59.9
b (Å)	97.8	97.0
c (Å)	105.1	104.9
resolution (Å)	2.80	2.40
no. of reflections (total/unique)	135105/29248	94804/23837
redundancy	4.6 (4.0)[a]	4.0 (2.9)[a]
completeness (%)	99.8 (98.6)[a]	96.5 (87.6)[a]
I/σ	10.3 (3.6)[a]	7.7 (2.2)[a]
R_{sym}[b] (%)	16.8 (39.8)[a]	19.2 (42.2)[a]
Matthews no. (Å^3/Da)	2.24	2.21
solvent content (%)	43.1	43.4

[a]Values in parentheses are for 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 the *N* reflections with intensities I_i and common indices *h*, *k*, and *l*.

Crystals were transferred to a well solution containing 25% glycerol and flash-frozen under liquid nitrogen prior to data collection. The crystals were very thin and were sensitive to radiation damage upon being exposed to the X-ray beam. Partial data sets were obtained by exposing each of two crystals multiple times at various spots on the crystal (11 total positions). The partial data sets were merged to obtain a complete, 2.8 Å resolution data set, which was used for initial structure determination. A second higher-resolution (2.4 Å) data set was obtained later by exposing a larger plate at six different positions. A total of 110° was covered with an exposure time of 1 s per frame and a crystal−detector distance of 300 mm. This data set was used for final refinement. All data were indexed, integrated, and scaled using the HKL2000 program suite (*19*). Data collection statistics are summarized in Table 1.

Structure Determination and Refinement. Analysis of the unit cell contents suggested two monomers of ThiY in the asymmetric unit, with six methionine residues per monomer excluding the N-terminal methionine residues. Ten initial Se sites were obtained using SHELXC (*20*, *21*). The sites were input into the AUTO_SOL program in PHENIX (*22*); however, the resulting map was of poor quality because of the very thin crystals and because of pseudotranslational symmetry resulting from the 2-fold noncrystallographic symmetry. Therefore, the Fold and Function Assignment system server (*23*) was used to identify SsuA, an alkanesulfonate/nitrate binding protein, from *Xanthomonas axonopodis* [Protein Data Bank (PDB) entry 3E4R] as the closest structural homologue of ThiY, and a truncated model of SsuA was used for molecular replacement. The SAD phases and molecular replacement phases were combined using PHASER_EP (*24*), and RESOLVE (*25*) was used for density modification. These phases were used to locate the two remaining Se sites. The 12 Se site SAD phases and molecular replacement phases were combined, improved by density modification and noncrystallographic symmetry averaging, and used to calculate an interpretable electron density map. Model building was performed using COOT (*26*), and the model was refined using the CNS program suite (*27*). The model obtained from 2.8 Å SAD phasing was refined against the 2.4 Å data set, which became available after the initial model building. The refinement process included successive rounds of simulated annealing, minimization, *B* factor refinement, calculation of composite omit maps,

Table 2: Refinement Statistics for ThiY

resolution (Å)	2.4
R^a	20.3
$R_{free}{}^b$	26.4
no. of non-H atoms	
protein	4648
ligand	24
water	259
B factor (Å^2)	
protein	22.9
ligand	13.0/15.2
root-mean-square deviation	
bonds (Å)	0.007
angles (deg)	1.3
dihedrals (deg)	22.5
Ramachandran plot (%)	
most favored region	91.9
additional favored region	8.1
generously allowed region	0
disallowed region	0

$^a R = \sum_{hkl}||F_o| - k|F_c||/\sum_{hkl}|F_o|$, where F_o and F_c are observed and calculated structure factors, respectively. ^{b}For R_{free}, the sum is extended over a subset of reflections (8%) excluded from all stages of refinement.

difference Fourier maps, and model building. An $F_o - F_c$ difference Fourier map was used to identify FAMP bound to the enzyme. The ligand was added to the model followed by another round of refinement and identification of water molecules. The parameter and topology files for FAMP were generated using the HIC-UP server (*28*). The final refinement statistics are listed in Table 2.

Homology Modeling of THI5. The sequence alignment of *Saccharomyces cerevisiae* THI5 and *B. halodurans* ThiY was obtained using ClustalW (*29*). The sequence alignment showed three main areas of gaps and/or insertions. (1) THI5 contains an N-terminal deletion of approximately two dozen residues corresponding to the lipoprotein signal sequence of ThiY. (2) THI5 contains an insertion of nine residues near the conserved CCCXC motif, which is absent in ThiY. (3) ThiY contains a deletion of approximately two dozen residues near the C-terminus; however, this deletion is difficult to position because of a low degree of sequence identity. MODELER 9v3 was used to generate five homology models followed by energy minimization of each model based on spatial restraints (*30*). The model with the lowest energy was considered for further analysis. In this model, the conserved CCCXC motif was located in a cavity near the predicted THI5 active site, suggesting a [4Fe-4S] cluster or some other metal center. A [4Fe-4S] cluster was manually placed in the cavity, and residues Cys195, Cys197, and Cys199 were reoriented to make covalent bonds between the sulfur and Fe atoms. The resulting structure was energy minimized to a gradient of 0.01 kJ/mol in vacuo using the AMBER* force field in Macromodel (*31*−*33*). The minimization used a TNCG minimization technique and a distance-dependent 4*r* dielectric treatment (*34*). All the residues located within 5 Å of the [4Fe-4S] cluster were allowed to move during minimization, and the rest of the protein remained frozen.

RESULTS

Overall Structure. ThiY crystallized in space group $P2_12_12_1$ with one dimer per asymmetric unit and the following unit cell dimensions: $a = 59.9$ Å, $b = 97.0$ Å, and $c = 104.9$ Å. The ThiY protein used for crystallization had an N-terminal 20-amino acid

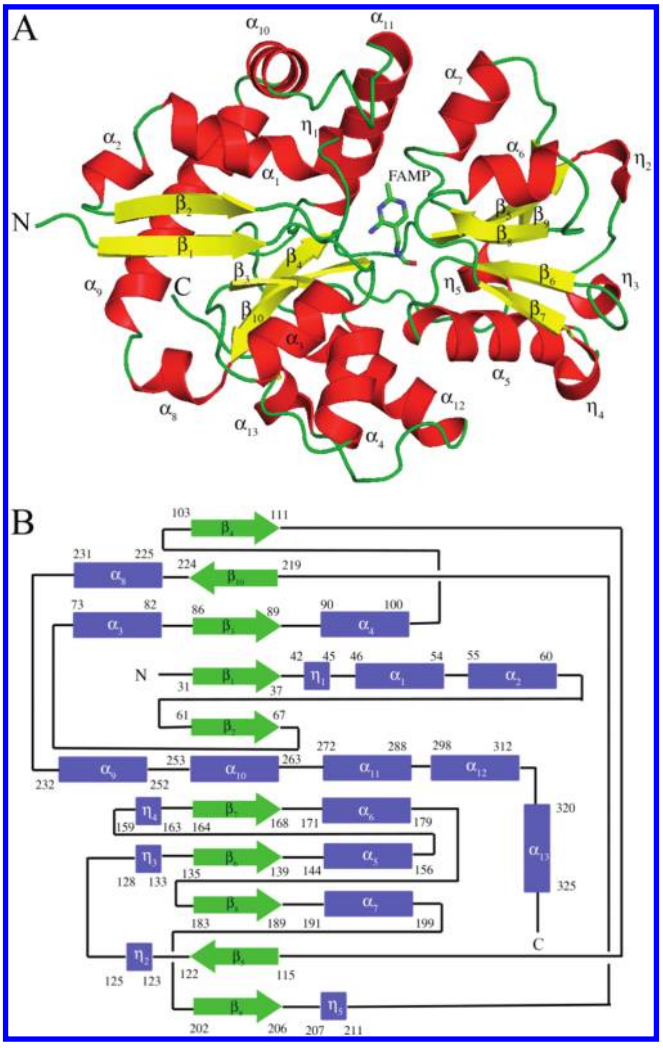


FIGURE 1: Monomeric structure of ThiY. (A) Cartoon diagram of ThiY with the secondary structure elements labeled. The 3_{10}-helices are labeled as η. FAMP is shown as sticks. Carbon atoms in FAMP are colored green, nitrogen atoms blue, and oxygen atoms red. (B) Topology diagram of ThiY showing the domain organization.

deletion. In addition, residues 21−29 were disordered in the crystal structure. Electron density was present for all other amino acid residues, which were included in the final model. The crystal structure reveals that ThiY belongs to the group II PBP family. The monomer is comprised of two domains. Domain 1 is a three-layer αβα sandwich comprised of a mixed β-sheet of five β-strands with $\uparrow\beta_4\downarrow\beta_{10}\uparrow\beta_3\uparrow\beta_1\uparrow\beta_2$ topology. The β-sheet is flanked by 10 α-helices and one 3_{10}-helix. Domain 2 is a three-layer αβα sandwich comprised of a mixed β-sheet containing five β-strands with $\uparrow\beta_9\downarrow\beta_5\uparrow\beta_8\uparrow\beta_6\uparrow\beta_7$ topology. This β-sheet is flanked by three α-helices and four 3_{10}-helices. The two domains are connected by crossover loops between strands β4 and β5 and between helix η5 and strand β10. A ribbon drawing of the ThiY monomer and the topology diagram showing connections between the secondary structural elements are shown in Figure 1.

Dimeric Structure. The enzyme crystallizes as a tightly packed dimer with monomers related by noncrystallographic 2-fold symmetry (Figure 2). The dimer interface is primarily hydrophilic and stabilized by extensive hydrogen bonding between residues Thr70 and Asp167′, Asn71 and Tyr139′, Thr136 and Asn268′, Asp167 and Phe170′, Arg180 and Asn268′, and Arg180 and Phe170′ (where the prime denotes a residue from the other monomer). The dimer buries 1810 Å^2 of surface area, and

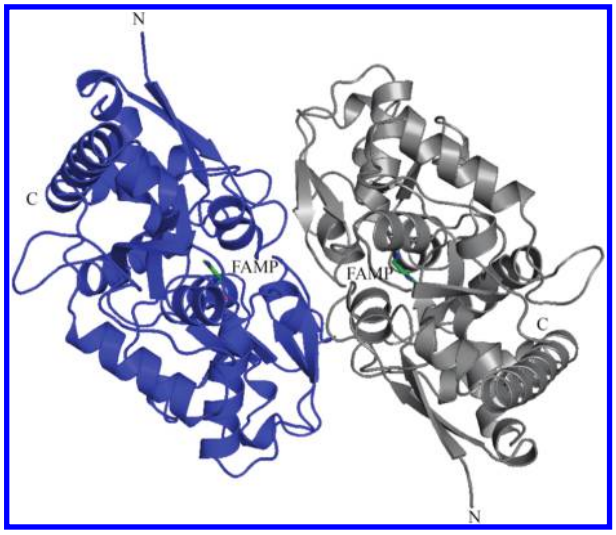


FIGURE 2: Dimeric form of ThiY looking down the 2-fold axis. The monomers are colored blue and gray. FAMP bound in the active site is shown as sticks.

the formation of the dimer seals off the entry−exit path for the ligand.

Ligand Binding Site. The ligand binding site of ThiY is located in a cleft between domain 1 and domain 2. FAMP was added during crystallization and was bound in the crystal structure. A stereoview of FAMP in the binding site is shown in Figure 3A, and a schematic of interactions of FAMP in the binding site is shown in Figure 3B. The pyrimidine ring of FAMP is sandwiched between the side chains of Trp39 and Tyr188. Residues Phe170 and Phe269 interact edge to face with Trp39. Residue Tyr40 interacts edge to face with Phe170. The N2 atom of FAMP is hydrogen bonded to Glu192, and the N1 atom is hydrogen bonded to a water molecule. The water molecule is hydrogen bonded to the side chains of Asp38 and Asn42. The side chain of Glu192 is stabilized by hydrogen bonds to the amide group of Tyr188 and a water molecule. The amino group (N3) of FAMP hydrogen bonds to side-chain atoms of Asp38 and Tyr90. The carbonyl terminus of FAMP hydrogen bonds to the amide group of Asn145. The hydroxyl group of Tyr188 hydrogen bonds to the N4 atom of FAMP and to the carboxylate group of Glu218. The hydroxyl group of Tyr188 is also 2.85 Å from the CE1 atom of Tyr90 in chain B. Residues Trp39 and Phe170 appear to act as a gate covering the entry−exit site for FAMP. Residues Trp39, Phe170, Tyr40, and Phe269 form a hydrophobic pocket shielding the binding site from external solvent.

Homology Modeling of THI5. A homology model of yeast THI5 was generated on the basis of its sequence similarity to ThiY. The sequences are 25% identical and 47% similar. The sequence alignment of ThiY and THI5 shows that the residues involved in binding the pyrimidine ring of FAMP in ThiY are partially conserved in THI5. A superposition of THI5 and the ThiY−FAMP complex suggests that the HMP-P binding residues in THI5 are Trp12, Asn11, Gln165, His18, and Thr15. Residues Asn11 and Trp12 of THI5 align with Asp38 and Trp39 of ThiY, respectively. Residue Gln165 in THI5 aligns with Glu192 of ThiY, which is involved in hydrogen bonding to the N2 atom of FAMP. Residue His45 is hydrogen bonded to Asp38 in ThiY and is also conserved in THI5. These residues are critical to binding FAMP in ThiY and are predicted to correspond to the HMP-P binding site in THI5.

The homology model of THI5 is similar to the group II PBPs with two distinct domains connected by two crossover loops. The conserved CCCXC motif is located in one of the crossover loops between domain 1 and domain 2 and is 15−20 Å from Trp12. A [4Fe-4S] cluster was built in the cavity between the domains facing the active site. The cavity between Trp12 and the [4Fe-4S] cluster is lined with residues Asn11, Gln164, Gln165, Phe117, His66, Glu161, and Gln121.

DISCUSSION

PBPs and ThiY. ATP-binding cassette (ABC) transporters are responsible for the transportation of a variety of ligands across the cell membrane. The importers and the exporters are believed to have diverged very early from a common ancestor. The importers are generally found in prokaryotes and are comprised of a periplasmic substrate binding domain and two cytosolic ATP binding domains. The exporters consist of a transmembrane domain and an ATP binding domain and recruit their ligands directly from the cytoplasm (*35*). The role of PBPs is to deliver small molecules and ions to the membrane-bound transporter for entry into the cell.

The crystal structures of many of the PBPs have been determined, and strong structural homology suggests a common ancestor for this family of proteins. The PBPs are comprised of two domains connected by variable linker regions. This two-domain architecture, which is believed to arise from gene duplication, is present in all PBPs. The PBPs have evolved into three different classes characterized by the number of crossovers between the two domains and the topological arrangement of the central β-sheet in the domains. Group I PBPs have two α/β-type domains with a $\beta_2\beta_1\beta_3\beta_4\beta_5$ β-sheet topology and three crossovers between the domains. Group II PBPs have two domains with a $\beta_2\beta_1\beta_3\beta_n\beta_4$ β-sheet topology with two crossovers between the domains. Strand β_n represents a strand immediately after the crossover from the other domain. Group III PBPs have one crossover between the domains, and little is known about the topological arrangement of the β-sheets in the domains. The substrate binding site in all PBPs is located between the domains and involves residues from both domains. These domains undergo a hingelike bending motion ("Venus flytrap mechanism") during ligand binding (*36*, *37*).

ThiY has two crossover connections between domain 1 and domain 2, revealing that it belongs to the group II PBP family. The topology of the central β-sheet in both domains is in accordance with the group II PBP family. The ThiY construct lacks the first 20 residues (hydrophobic membrane anchor), and nine additional N-terminal residues were not visible in the electron density. The enzyme binds FAMP in the binding site located between domain 1 and domain 2. FAMP makes extensive stacking and hydrogen bonding interactions with the protein and is shielded from the environment by residues Trp39, Phe170, Tyr40, and Phe269. ThiY crystallizes as a dimer in the asymmetric unit, and this dimerization closes the entry−exit path for the ligand. A similar dimeric arrangement was observed in the thiamin binding protein TbpA. In this case, the dimerization also occurs in solution and is dependent on the concentration of the ligand (*15*). The physiological relevance of the TbpA dimerization is unclear. The dimerization of ThiY may be related to the mechanism of delivery of the substrate to the ABC transporter or may be an artifact of crystallization.

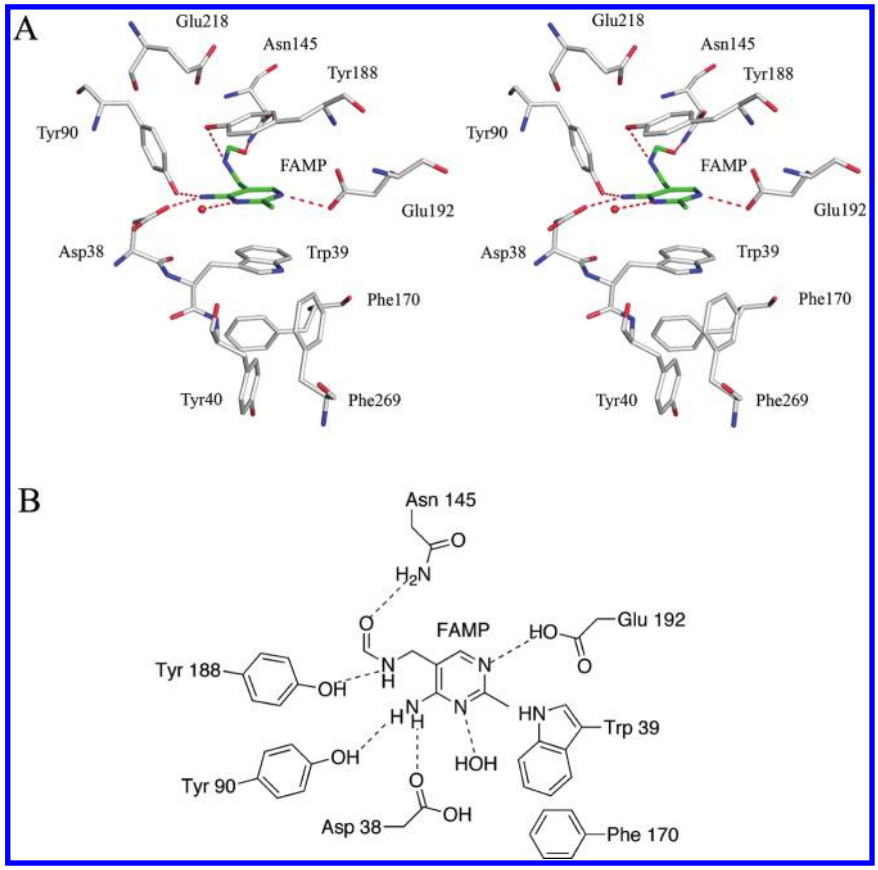


FIGURE 3: FAMP binding site in ThiY. (A) Stereoview of the FAMP binding site. FAMP has carbon atoms colored green. Water molecules are shown as red spheres, and hydrogen bonds are shown as dashed lines. (B) Schematic representation of the binding site showing the key hydrogen bonding and stacking interactions.

Substrate Specificity of ThiY. The ThiXYZ transporter is found in a gene cluster involved in the salvage of FAMP, which is hydrolyzed to HMP for thiamin biosynthesis. The role of ThiY is delivery of FAMP to the ABC transporter, and the equilibrium binding constant for FAMP is 200 nM (*18*). Our structure provides the atomic details of this binding. Tyr188 and Asn145 are involved in binding the formylamino group of FAMP through hydrogen bonding. Tyr188 is also involved in a stacking interaction with the pyrimidine ring. A sequence alignment of diverse ThiY homologues reveals that Tyr188 and Asn145 are not strictly conserved. The tyrosine residue is sometimes replaced with a phenylalanine residue, and the asparagine residue is replaced with a variety of residues such as serine, aspartate, and glutamate. Residues Glu192, Trp39, Asp38, and Tyr90 that interact with the pyrimidine ring of FAMP are strictly conserved. The sequence variability in the ligand binding site suggests that the ThiXYZ system may have evolved to transport a variety of different thiamin degradation products with intact pyrimidines into the cell and therefore suggests the existence of other degradation and salvage pathways for thiamin. In addition to ThiXYZ, other transporters such as YkoEDC and CytX are also likely to be involved in the uptake of HMP or HMP precursors (*10*).

Comparison of ThiY to Other PBPs. A DALI (*38*) search identified several structural homologues of ThiY. The closest homologues are SsuAs from *X. axonopodis* (XaSsuA; PDB entry 3E4R) and *E. coli* (EcSsuA; PDB entry 2X26) (*39*), which are alkanesulfonate/nitrate binding proteins, with Z scores of 25.1 and 24.2, respectively. SsuA belongs to the type II PBP superfamily, and XaSsuA has a molecule of 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) bound in the active site. A superposition of ThiY and XaSsuA is shown in Figure 4A. Despite a low level of sequence idenity, XaSsuA was successfully used as a molecular replacement model in combination with SAD during initial phasing. The Cα trace of XaSsuA also served as a helpful reference during ThiY model building. Other structural homologues of ThiY are hypothetical phosphate binding protein (PDB entry 1ZBM; Z = 21.4), nitrate binding protein (PDB entry 2G29; Z = 22.6) (*40*), and bicarbonate binding protein (PDB entry 2I4B; Z = 22.2) (*41*). Most of the top ThiY structural homologues are group II PBPs involved in the transport of small ligands; however, MqnD, which is an enzyme involved in menaquinone biosynthesis, shows up in the top 10 (PDB entry 3A3U; Z = 22.7) (*42*).

Comparison of ThiY with TbpA. Both ThiY and TbpA are transport proteins involved in thiamin salvage. FAMP is the preferred ligand for ThiY, while the preferred ligands for TbpA are thiamin, ThMP, and ThDP. A DALI search showed that ThiY and TbpA are structural homologues (PDB entry 2QRY; Z = 10.7) (*15*). A superposition of ThiY and TbpA (Figure 4C) shows that the ligand binding sites of both proteins are in a similar position. The structure of TbpA was determined with bound ThMP (*15*), and the binding site residues interacting with ThMP in TbpA are shown in Figure 4D. The thiazole ring stacks between residues Tyr27 and Tyr215 and is positioned edge to face from Trp197. The pyrimidine ring is buried in a pocket formed by Trp197, Trp280, Tyr221, and Tyr201, and the phosphate group is hydrogen bonded to Ser161, Trp197, Gly60, and Asp59. A comparison between thiamin and FAMP binding shows that the pyrimidine ring of FAMP superimposes on the thiazole moiety of thiamin bound to TbpA, and π-stacking is a common

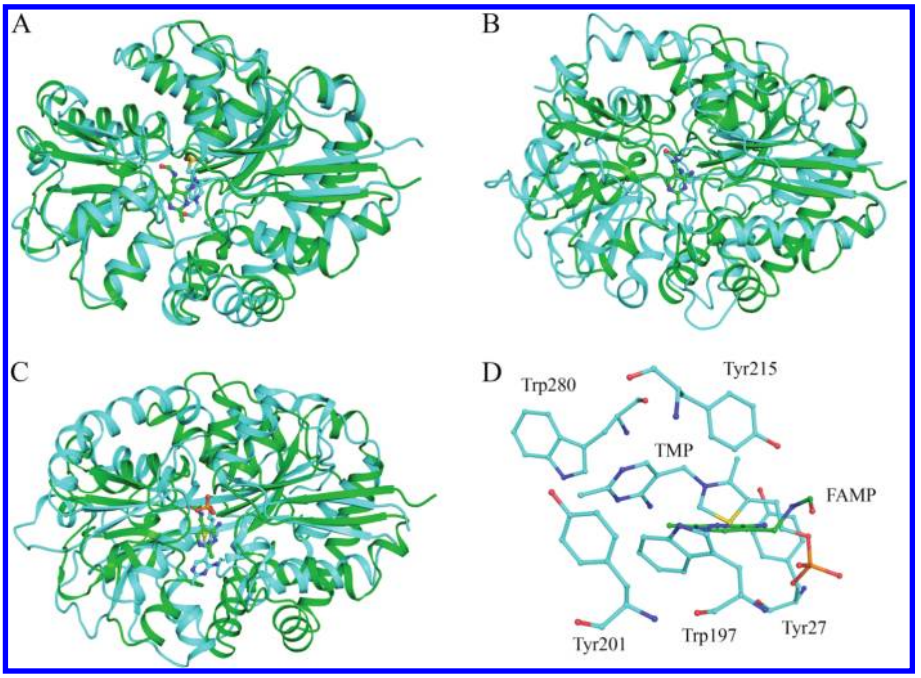
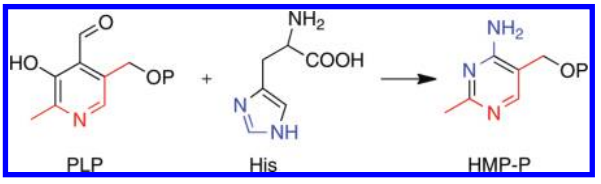


FIGURE 4: Structural homologues for ThiY. Comparisons of ThiY with (A) SsuA, (B) thiaminase-I, and (C) TbpA. ThiY carbon atoms are colored green, and carbon atoms of the structure being compared are colored cyan. The ligands are shown as balls and sticks. (D) Superposition of the binding sites of ThiY and TbpA. Hydrophobic residues involved in stacking of TMP in TbpA are shown.

Scheme 2



feature in the binding of both ligands. The formyl terminus of FAMP is in a position similar to that of the phosphate of thiamin.

Similarity between ThiY and THI5. In eukaryotes, HMP-P is synthesized from PLP and histidine in a reaction catalyzed by THI5, the yeast pyrimidine synthase (Scheme 2). The activity of THI5 has not been reconstituted in solution, and its mechanism has not been established. Iterated BLAST searches (*43*) starting with ThiY identified various proteins annotated as possible periplasmic binding proteins, as well as a few proteins identified as HMP-P synthases. However, all of the proteins identified by this search as HMP-P synthases lacked the conserved CCCXC motif of THI5 and are likely misannotated. Iterated BLAST searches starting with *S. cerevisiae* THI5 identifed orthologs with the conserved CCCXC motif but also periplasmic binding proteins. The sequence of *S. cerevisiae* THI5 is 25% identical and 47% similar to that of *B. halodurans* ThiY, suggesting that ThiY and THI5 are structural homologues. In addition, ThiY binds FAMP, which is structurally similar to HMP-P, the product of THI5. Therefore, we undertook homology modeling in hopes that the THI5 model might provide useful clues for the successful reconstitution of the THI5-catalyzed reaction.

The sequence alignment of ThiY and THI5 shows that the residues involved in binding the pyrimidine ring in ThiY (Asp38, Trp39, Glu192, and His45) are conserved in THI5. In addition, Ser79, Ser82, Tyr113, Asp155, Asn271, Trp304, and Phe209, conserved in diverse sequences of THI5, are also conserved in ThiY. These residues are outside the active site of ThiY and may be involved in maintaining the structural integrity of the protein.

THI5 has a conserved CCCXC motif. Related cysteine rich sequences (e.g., CX_3CX_2C) bind [4Fe-4S] clusters involved in electron transfer or radical reactions (*44*) and are also known to serve as Lewis acids (*45*). The CCCXC motif was included in the homology model to evaluate if THI5 might contain a [4Fe-4S] cluster (or other metal center). The location of the HMP binding site was also predicted to identify residues potentially important for binding or catalysis. The homology model shows that the CCCXC motif is located in a deeply buried cleft between domain 1 and domain 2 (Figure 5). In this model, Cys195, Cys197, and Cys199 serve as ligands for the [4Fe-4S] cluster. The active site (identified from the location of FAMP in the superimposed ThiY) is ~15 Å from the cleft. The space between the putative [4Fe-4S] cluster and the HMP-P binding site could be occupied by the histidine and PLP substrates. The [4Fe-4S] cluster is ~13 Å from Trp12, which is proposed to stack against the pyrimidine ring of the product. While the [4Fe-4S] cluster has reasonable geometry, its occurrence has not been experimentally verified. It is also possible that the cysteine motif could be involved in binding smaller iron−sulfur clusters. The residues lining the proposed active site cavity (Asn11, Gln164, Phe117, His66, Glu161, and Gln121) are strictly conserved in THI5 and may play a role in the reaction mechanism. These predictions will be used in developing a strategy for reconstituting the THI5-catalyzed reaction in vitro.

Comparison of ThiY and Thiaminase-I. DALI searches show that ThiY is structurally homologous to thiaminase-I, an enzyme involved in thiamin degradation (PDB entry 4THI) (*46*), but with a much lower *Z* score of 5.6 for the full-length protein. The homology is comparatively better when only domain I is used with a *Z* score of 7.1. The superposition of ThiY and thiaminase-I is shown in Figure 4B. The crystal structure of thiaminase-I determined with the mechanism-based inhibitor 4-amino-2,5-dimethylpyrimidine shows the pyrimidine moiety covalently attached to Cys113 in the active site (*46*). The comparison of the structures shows that the binding site for the

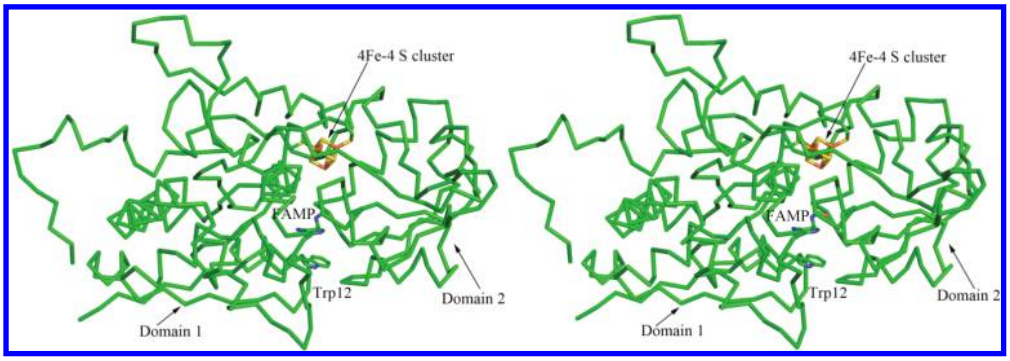


FIGURE 5: Homology model of yeast THI5. Stereo ribbon diagram showing the Cα trace of both domains. The [4Fe-4S] cluster, Trp12, and FAMP (superimposed from ThiY) are shown as sticks. The CCCXC motif binding the cluster is located in the cleft on the other side of the active site.

pyrimidine is in a similar location for both ThiY and thiaminase-I. A closer look at the binding conformations of the substrates reveals that the pyrimidine ring of FAMP is orthogonal to the pyrimidine ring of the inhibitor in thiaminase-I; however, this may result from the covalent attachment to Cys113. The centers of the pyrimidine rings are separated by ~4 Å in the superposition. TbpA and thiaminase-I are structural homologues and are believed to have a common ancestor (*15*). The structure of ThiY described here demonstrates that the thiamin metabolic proteins ThiY, TbpA, thiaminase-I, and THI5 (1) are members of the group II PBP family and (2) have binding sites for a pyrimidine moiety derived from thiamin. These similarities suggest that all four proteins may have evolved from a common ancestor.

ACKNOWLEDGMENT

We thank Leslie Kinsland for assistance in the preparation of the manuscript. We thank the NE-CAT staff of the Advanced Photon Source for assistance in data collection and providing beam time. We thank Dr. Jeremiah Hanes for the synthesis of FAMP. We thank Dr. Cynthia Kinsland of the Cornell Protein Production Facility for the clones for ThiY. This work is based upon research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines, which are supported by Grant RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357.

REFERENCES

1. Harper, C. (2006) Thiamine (vitamin B1) deficiency and associated brain damage is still common throughout the world and prevention is simple and safe!. *Eur. J. Neurol. 13*, 1078–1082.
2. Begley, T. (1996) The biosynthesis and degradation of thiamin (vitamin B1). *Nat. Prod. Rep. 13*, 177–185.
3. 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.
4. Chatterjee, A., Li, Y., Zhang, Y., Grove, T. L., Lee, M., Krebs, C., Booker, S. J., Begley, T. P., and Ealick, S. E. (2008) Reconstitution of ThiC in thiamine pyrimidine biosynthesis expands the radical SAM superfamily. *Nat. Chem. Biol. 4*, 758–765.
5. Jurgenson, C. T., Chatterjee, A., Begley, T. P., and Ealick, S. E. (2006) Structural Insights into the Function of the Thiamin Biosynthetic Enzyme Thi4 from *Saccharomyces cerevisiae*. *Biochemistry 45*, 11061–11070.
6. Chatterjee, A., Schroeder, F. C., Jurgenson, C. T., Ealick, S. E., and Begley, T. P. (2008) Biosynthesis of the thiamin-thiazole in eukaryotes: Identification of a thiazole tautomer intermediate. *J. Am. Chem. Soc. 130*, 11394–11398.
7. Wightman, R., and Meacock, P. A. (2003) The THI5 gene family of *Saccharomyces cerevisiae*: Distribution of homologues among the hemiascomycetes and functional redundancy in the aerobic biosynthesis of thiamin from pyridoxine. *Microbiology 149*, 1447–1460.
8. Zeidler, J., Sayer, B. G., and Spenser, I. D. (2003) Biosynthesis of vitamin B1 in yeast. Derivation of the pyrimidine unit from pyridoxine and histidine. Intermediacy of urocanic acid. *J. Am. Chem. Soc. 125*, 13094–13105.
9. Nosaka, K. (2006) Recent progress in understanding thiamin biosynthesis and its genetic regulation in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol. 72*, 30–40.
10. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2002) Comparative Genomics of Thiamin Biosynthesis in Procaryotes. New genes and regulatory mechanisms. *J. Biol. Chem. 277*, 48949–48959.
11. Hollenbach, A. D., Dickson, K. A., and Washabaugh, M. W. (2002) Overexpression, purification, and characterization of the periplasmic space thiamin-binding protein of the thiamin traffic ATPase in *Escherichia coli*. *Protein Expression Purif. 25*, 508–518.
12. Webb, E., Claas, K., and Downs, D. (1998) thiBPQ encodes an ABC transporter required for transport of thiamine and thiamine pyrophosphae in *Salmonella typhimurium*. *J. Biol. Chem. 273*, 8946–8950.
13. Jones, P. M., and George, A. M. (2004) The ABC transporter structure and mechanism: Perspectives on recent research. *Cell. Mol. Life Sci. 61*, 682–699.
14. Davidson, A. L., and Chen, J. (2004) ATP-binding cassette transporters in bacteria. *Annu. Rev. Biochem. 73*, 241–268.
15. Soriano, E. V., Rajashankar, K. R., Hanes, J. W., Bale, S., Begley, T. P., and Ealick, S. E. (2008) Structural similarities between thiamin-binding protein and thiaminase-I suggest a common ancestor. *Biochemistry 47*, 1346–1357.
16. Hantke, K., and Braun, V. (1973) Covalent binding of lipid to protein. Diglyceride and amide-linked fatty acid at the N-terminal end of the murein-lipoprotein of the *Escherichia coli* outer membrane. *Eur. J. Biochem. 34*, 284–296.
17. Sankaran, K., and Wu, H. C. (1994) Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *J. Biol. Chem. 269*, 19701–19706.
18. Jenkins, A. H., Schyns, G., Potot, S., Sun, G., and Begley, T. P. (2007) A new thiamin salvage pathway. *Nat. Chem. Biol. 3*, 492–497.
19. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307–326.
20. Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr. A64*, 112–122.
21. Pape, T., and Schneider, T. R. (2004) HKL2MAP: A graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr. 37*, 843–844.
22. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr. D58*, 1948–1954.
23. Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. (2005) FFAS03: A server for profile−profile sequence alignments. *Nucleic Acids Res. 33*, W284–W288.
24. Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D50*, 760–763.
25. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D55*, 849–861.

26. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.
27. Brünger, 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.
28. Kleywegt, G. J., and Jones, T. A. (1998) Databases in protein crystallography. *Acta Crystallogr. D54*, 1119–1131.
29. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res. 22*, 4673–4680.
30. Eswar, N., Marti-Renom, M. A., Webb, B., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., and Sali, A. (2007) Comparative Protein StructureModeling with MODELLER. In Current Protocols in Protein Science, pp 2.9.1−2.9.31, Wiley, John & Sons, Inc., New York.
31. Mohamadi, F., Richards, N. G. J., Guida, W. C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., and Still, W. C. (1990) MacroModel: An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem. 11*, 460–467.
32. Weiner, S. J., and Kollman, P. A. (1986) An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem. 7*, 230–252.
33. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., and Weiner, P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc. 106*, 765–784.
34. Ponder, J. W., and Richards, F. M. (1987) An Efficient Newton-like Method for Molecular Mechanics Energy Minimization of Large Molecules. *J. Comput. Chem. 8*, 1016–1024.
35. Oldham, M. L., Davidson, A. L., and Chen, J. (2008) Structural insights into ABC transporter mechanism. *Curr. Opin. Struct. Biol. 18*, 726–733.
36. Shi, R., Proteau, A., Wagner, J., Cui, Q., Purisima, E. O., Matte, A., and Cygler, M. (2009) Trapping open and closed forms of FitE-A group III periplasmic binding protein. *Proteins 75*, 598–609.
37. Fukami-Kobayashi, K., Tateno, Y., and Nishikawa, K. (1999) Domain dislocation: A change of core structure in periplasmic binding proteins in their evolutionary history. *J. Mol. Biol. 286*, 279–290.
38. Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics 24*, 2780–2781.
39. Beale, J., Lee, S. Y., Iwata, S., and Beis, K. (2010) Structure of the aliphatic sulfonate-binding protein SsuA from *Escherichia coli*. *Acta Crystallogr. F66*, 391–396.
40. Koropatkin, N. M., Pakrasi, H. B., and Smith, T. J. (2006) Atomic structure of a nitrate-binding protein crucial for photosynthetic productivity. *Proc. Natl. Acad. Sci. U.S.A. 103*, 9820–9825.
41. Koropatkin, N. M., Koppenaal, D. W., Pakrasi, H. B., and Smith, T. J. (2007) The structure of a cyanobacterial bicarbonate transport protein, CmpA. *J. Biol. Chem. 282*, 2606–2614.
42. Arai, R., Murayama, K., Uchikubo-Kamo, T., Nishimoto, M., Toyama, M., Kuramitsu, S., Terada, T., Shirouzu, M., and Yokoyama, S. (2009) Crystal structure of MqnD (TTHA1568), a menaquinone biosynthetic enzyme from *Thermus thermophilus* HB8. *J. Struct. Biol. 168*, 575–581.
43. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389–3402.
44. Fontecave, M. (2006) Iron-sulfur clusters: Ever-expanding roles. *Nat. Chem. Biol. 2*, 171–174.
45. Saunders, A. H., Griffiths, A. E., Lee, K. H., Cicchillo, R. M., Tu, L., Stromberg, J. A., Krebs, C., and Booker, S. J. (2008) Characterization of quinolinate synthases from *Escherichia coli*, *Mycobacterium tuberculosis,* and *Pyrococcus horikoshii* indicates that [4Fe-4S] clusters are common cofactors throughout this class of enzymes. *Biochemistry 47*, 10999–11012.
46. Campobasso, N., Costello, C. A., Kinsland, C., Begley, T. P., and Ealick, S. E. (1998) Crystal structure of thiaminase-I from *Bacillus thiaminolyticus* at 2.0 Å resolution. *Biochemistry 37*, 15981–15989.