*Articles*

# Structure of the *Escherichia coli* ThiS−ThiF Complex, a Key Component of the Sulfur Transfer System in Thiamin Biosynthesis†,‡

Christopher Lehmann, Tadhg P. Begley,* and Steven E. Ealick*

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

*Received July 28, 2005; Revised Manuscript Received October 28, 2005*

ABSTRACT: We have determined the crystal structure of the *Escherichia coli* ThiS−ThiF protein complex at 2.0 Å resolution. ThiS and ThiF are bacterial proteins involved in the synthesis of the thiazole moiety of thiamin. ThiF catalyzes the adenylation of the carboxy terminus of ThiS and the subsequent displacement of AMP catalyzed by ThiI-persulfide to give a ThiS−ThiI acyl disulfide. Disulfide interchange, involving Cys184 on ThiF, then generates the ThiS−ThiF acyl disulfide, which functions as the sulfur donor for thiazole formation. ThiS is a small 7.2 kDa protein that structurally resembles ubiquitin and the molybdopterin biosynthetic protein MoaD. ThiF is a 27 kDa protein with distinct sequence and structural similarity to the ubiquitin activating enzyme E1 and the molybdopterin biosynthetic protein MoeB. The ThiF−ThiS structure clarifies the mechanism of the sulfur transfer chemistry involved in thiazole biosynthesis.

Thiamin pyrophosphate (**11**, vitamin $B_1$) is an important cofactor in all living systems. It is involved primarily in carbohydrate and branched chain amino acid metabolism, where it stabilizes acyl carbanion intermediates (*1*). Vitamin $B_1$ is an essential nutrient in humans, and the thiamin deficiency diseases beriberi and Wernike−Korsakoff syndrome have been extensively studied (*2*, *3*). The thiamin biosynthetic pathway in *Escherichia coli* is outlined in Scheme 1 (*4*, *5*).

The biosynthesis of thiazole **8** is complex and involves six proteins (ThiS, ThiF, ThiG, ThiH, ThiI, and IscS). IscS, an enzyme also involved in iron−sulfur cluster biosynthesis, catalyzes the transfer of the cysteine sulfur to ThiI to give ThiI-persulfide **3** (*6*, *7*). ThiF catalyzes the adenylation of ThiS, at its carboxy terminus, to give **2** and the subsequent displacement of AMP by the ThiI-persulfide to give **4**. Disulfide interchange, involving Cys184 on ThiF, generates the mixed acyl disulfide **5**, which functions as the sulfur donor for thiazole formation (*6*, *8*). Thiazole formation in *E. coli* has been reconstituted, at trace levels, but the mechanistic details are not yet well understood (*9*).

The corresponding thiazole biosynthesis in *Bacillus subtilis* (Scheme 2) involves five proteins (ThiS, ThiF, ThiG, ThiO, and IscS) and differs in the detailed mechanism of the redox and sulfur transfer chemistry from the *E. coli* pathway (*10*). In both prokaryotes and eukaryotes, thiazole **8** displaces the pyrophosphate of pyrimidine **9** to form thiamin phosphate, and a final phosphorylation generates thiamin pyrophosphate **11**, the biologically active form of the cofactor (*11*, *12*).

ThiS is a 7.2 kDa protein (66 amino acids) that is structurally similar to ubiquitin and MoaD (*13*). All three proteins are small (<100 amino acids) and have a low degree of sequence similarity except for a carboxy-terminal Gly-Gly pair. ThiF is a 27 kDa protein (251 amino acids) with sequence and structural similarity to Ubl[1] activating enzymes, and to the molybdopterin biosynthetic protein MoeB (*14*−*17*). The molybdopterin biosynthetic proteins MoeB and MoaD have functions similar to those of the ThiF−ThiS complex. MoeB catalyzes the ATP-dependent formation of MoaD-thiocarboxylate, which functions as the sulfur source for molybdenum cofactor biosynthesis. The relationship between the ThiF−ThiS complex and the ubiquitinating system is less direct and involves thioester rather than acyl disulfide formation. In this system, the ubiquitin-activating enzyme E1 catalyzes the formation of the acyl adenylate of ubiquitin, which then reacts with a cysteine on E1 to form a thioester. This thioester functions as the ubiquitin donor, ultimately tagging proteins for degradation in the proteasome (*18*, *19*). While no prokaryotic genome has been found to encode the ubiquitinating enzymes, these similarities suggest that the ubiquitin−E1, ThiS−ThiF, and MoaD−MoeB systems have evolved from a common ancestor.

The structural enzymology of the metabolism of thiazole **8** is now at an advanced stage, and representative examples

† This work was supported by National Institutes of Health Grants DK44083 (to T.P.B.) and DK067081 (to S.E.E.). S.E.E. is indebted to the W. M. Keck Foundation and the Lucille P. Markey Charitable Trust.

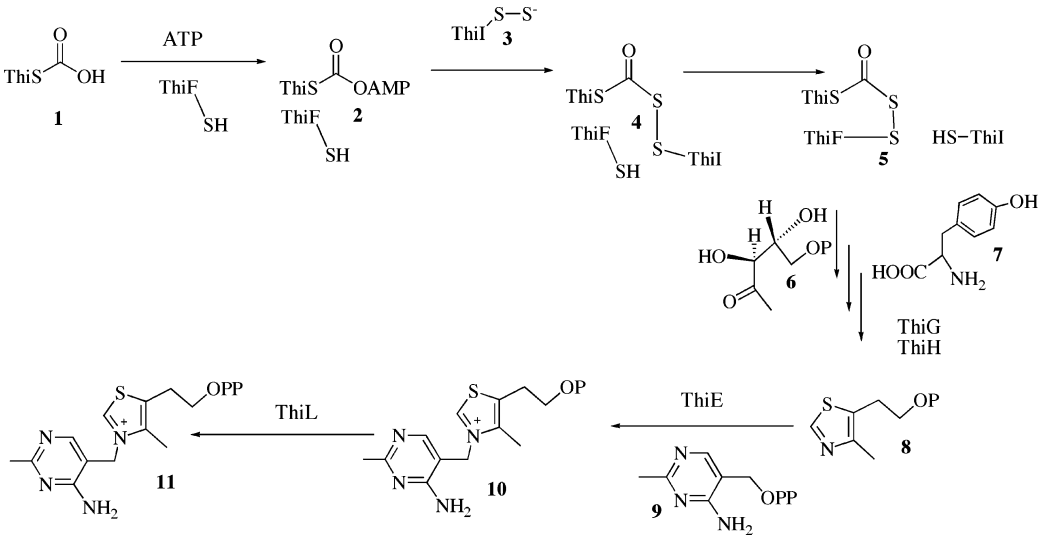
‡ The Brookhaven Protein Data Bank code for the ThiS−ThiF complex is 1ZUD.

* 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 tbp2@cornell.edu.

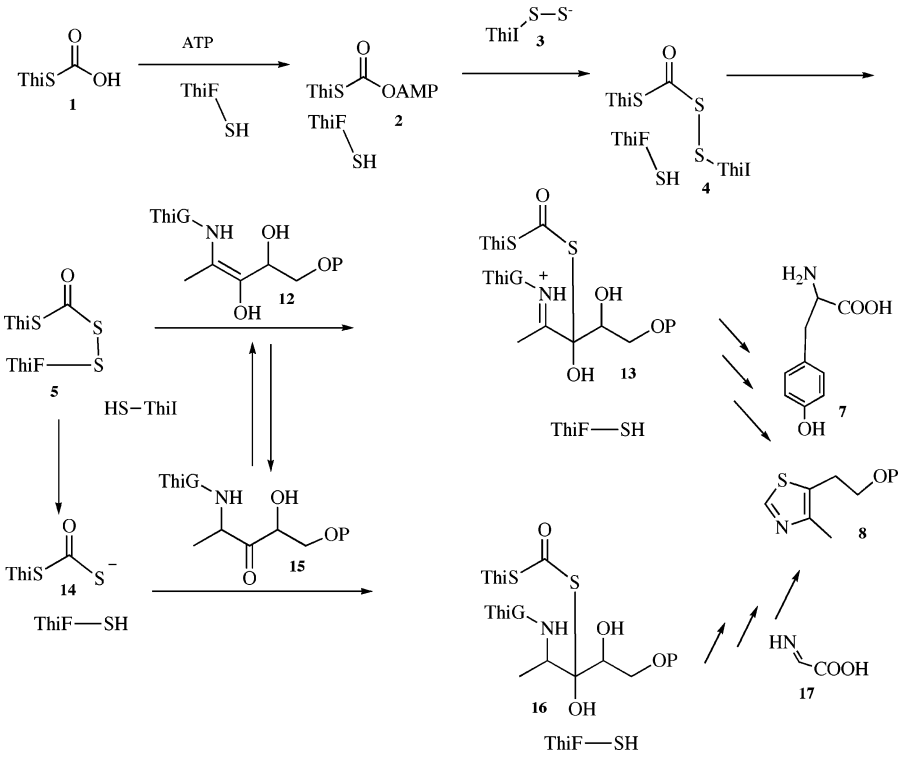
[1] Abbreviations: Ub, ubiquitin; Ubl, ubiquitin-like; CHESS, Cornell High Energy Synchrotron Source; APS, Advanced Photon Source; rmsd, root-mean-square deviation.

10.1021/bi051502y CCC: $33.50 
Published on Web 12/06/2005

Scheme 1




Scheme 2




of all of the enzymes involved have been structurally characterized. These enzymes include thiaminase I (*20*) and thiaminase II (*21*), both of which catalyze the cleavage of the thiazole from thiamin, thiazole kinase (*22*), which catalyzes the phosphorylation of the thiazole alcohol, thiamin phosphate synthase (*23*), which catalyzes the attachment of the thiazole to the pyrimidine and ThiO (*24*), ThiF (*25*), ThiS (*13*), and ThiSG (*26*), which are all involved in thiazole assembly. Here we describe the crystal structure of the protein complex between ThiS and ThiF from *E. coli* at 2.0 Å resolution. This structure clarifies the mechanism of the sulfur transfer chemistry involved in thiazole biosynthesis.

## MATERIALS AND METHODS

*Protein Production.* Construction of the *thiFS*-containing pCLK1405 plasmid was described previously (*6*). For protein production, the pCLK1405 plasmid was transformed into *E. coli* overexpression strain BL834(DE3) (Novagen). A starter culture was prepared by inoculating a single colony into 5 mL of LB medium containing 100 $\mu$g/mL ampicillin and shaking overnight at 200 rpm and 37 °C. The starter culture was inoculated into 1 L of LB medium containing 100 $\mu$g/mL ampicillin. This culture was grown at 37 °C and 200 rpm to a cell density corresponding to an $OD_{600}$ of ~0.6 and induced with 300 $\mu$M isopropyl $\beta$-D-thiogalactoside. After being induced for 3 h, the cells were harvested and centrifuged at 4000*g* for 5 min. All subsequent purification steps were carried out at 4 °C. Cells were resuspended in 50 mL of 25 mM Tris-HCl (pH 8) (buffer A) and lysed using sonication. The lysate was centrifuged at 23000*g* for 30 min, and ammonium sulfate was added to bring its concentration in the supernatant to 50%. The resulting precipitate was

isolated by centrifugation at 23000*g*, resuspended in buffer A, and dialyzed against the same buffer for 12 h. Protein purification was performed on an AKTA explorer FPLC system (Amersham Biosciences). The dialyzate was applied to a 5 mL Hi-Trap QFF column (Amersham Biosciences), washed with buffer A for 30 min, and eluted with buffer A containing 250 mM NaCl. Fractions were analyzed by SDS−PAGE and pooled. Size exclusion chromatography was performed using a Superdex 200 high load 26/60 column (Amersham Biosciences) at a flow rate of 1 mL/min using buffer A. Fractions were analyzed by SDS−PAGE, pooled, and concentrated to 20 mg/mL using an Amicon Ultra 15 concentrator (Millipore) with a 5 kDa molecular mass cutoff. Protein concentrations were determined using the Bradford assay (*27*). The sample was aliquoted, frozen in liquid nitrogen, and stored at −80 °C.

*Crystallization of the ThiS−ThiF Complex.* The ThiS−ThiF complex was crystallized using the hanging drop method with each drop containing 1.2 $\mu$L of protein solution (8 mg/mL ThiF−ThiS complex in buffer A) and 1.2 $\mu$L of well solution [7−8% polyethylene glycol 400, 35 mM calcium chloride, and 100 mM Tris-HCl (pH 7.0−7.3)]. Crystals appeared sporadically within 3 days to 2 weeks and grew to a maximum size of 0.1 mm × 0.1 mm × 0.1 mm. Preliminary X-ray analysis showed that the crystals belong to orthorhombic space group $P2_12_12_1$ with the following unit cell dimensions: $a = 49.51$ Å, $b = 111.17$ Å, and $c = 114.15$ Å. The Matthews coefficient, assuming two monomers of ThiF and two monomers of ThiS per asymmetric unit, is 2.3 Å$^3$/Da. This corresponds to 47% solvent. For cryoprotection, the crystals were transferred into solutions containing an additional 25% glycerol, frozen directly in the cryostream, and stored under liquid nitrogen for later use.

*X-ray Data Collection and Processing.* In-house data were collected using a Rigaku RTP 300 RC copper rotating anode X-ray generator operating at 50 kV and 100 mA, equipped with a Rigaku RAXIS IV$^{++}$ image plate detector. Integration and scaling of diffraction intensities were performed using CRYSTAL CLEAR (Rigaku/Molecular Structure Corp.). Two synchrotron data sets were collected. First, data were collected at Cornell High Energy Synchrotron Source (CHESS) station F2 at a wavelength of 1.49 Å to a resolution of 2.4 Å. This wavelength was chosen to exploit the anomalous signal for heteroatom identification. Station F2 was equipped with a Quantum 210 X-ray detector (Area Detector Systems Corp.). Data were collected over 110° using 50 s for each 0.5° oscillation with a crystal to detector distance of 150 mm. The second data set was collected, using the identical crystal, at Northeastern Collaborative Access Team (NE-CAT) beamline 8-BM at the Advanced Photon Source (APS), at a wavelength of 0.98 Å to a resolution of 2.0 Å. Beamline 8-BM was equipped with an ADSC Quantum 315 X-ray detector. Data were collected over 158.5° using 30 s for each 0.5° oscillation with a crystal to detector distance of 310 mm. The data were integrated and scaled using HKL2000 (*28*). Data collection statistics are summarized in Table 1.

*Structure Determination.* The structure was determined by molecular replacement using MOLREP (*29*). MoeB (chain A) from the MoeB−MoaD complex (*15*) (PDB entry 1JW9) was used as a search model. Residues Ala190−Ser202, which based on sequence alignments are not present in ThiF, were deleted from the model. In-house data were used to a resolution of 3 Å. Two monomers of ThiF were found. The first gave a correlation coefficient of 0.204 and an *R*-factor of 0.572, while the second gave a correlation coefficient of 0.300 and an *R*-factor of 0.542. A $\sigma_A$-weighted $2F_o - F_c$ map clearly showed density for two ThiS molecules. *B. subtilis* ThiS (PDB entry 1TYG) was truncated at the C-terminus after residue Val63. Two of these ThiS models were manually fitted into this electron density using the MoeB−MoaD complex as a guide.

Table 1: Data Collection Statistics

| | in house | CHESS (F2) | APS (8BM) |
|---|---|---|---|
| space group | $P2_12_12_1$ | | |
| cell dimensions (Å) | $a = 49.51$, $b = 111.17$, $c = 114.15$ | | |
| wavelength (Å) | 1.54 | 1.49 | 0.98 |
| resolution (Å) | 16−2.5 | 49−2.4 | 27−2.0 |
| no. of reflections | 56485 | 92223 | 224056 |
| no. of unique reflections | 19428 | 24000 | 42395[c] |
| redundancy | 2.5 | 3.8 | 5.3 |
| completeness (%) | 86.8 (87.9)[a] | 94.9 (68.4)[a] | 94.9 (93.6)[a] |
| $R_{sym}$[b] (%) | 7.2 (23.5)[a] | 9.1 (32.5)[a] | 8.9 (26.3)[a] |
| $I/\sigma$ | 9.8 (3.9)[a] | 11.9 (2.9)[a] | 21.1 (4.0)[a] |

[a] Values for the outer resolution shell are given in parentheses. [b] $R_{sym} = \sum_{hkl}[(\sum_j|I_j - \langle I\rangle|)/\sum_j|I_j|]$, for equivalent reflections. [c] Includes 880 reflections with negative *I*, which were omitted from the refinement.

*Model Building and Structure Refinement.* All model building was performed using O (*30*). *E. coli* ThiF side chains were inserted into the MoeB molecular replacement model, and the *E. coli* ThiS side chains were inserted into the ThiS model, after which all side chains were manually adjusted. The structure was refined against data between 25 and 2.0 Å resolution using CNS (*31*). The refinement procedure involved rigid body refinement followed by successive rounds of simulated annealing, individual temperature factor refinement and model rebuilding. Early refinement was aided by the use of noncrystallographic symmetry restraints. Water molecules were added to the model in the final stages of the refinement if the $F_o - F_c$ difference electron density peaks were $\geq 3\sigma$. Anomalous and dispersive difference electron maps were calculated in CNS using the two synchrotron data sets. These maps were used to aid in the identification of metal ions. Structural analysis was carried out using PROCHECK (*32*) for geometric analysis, MOLSCRIPT (*33*) and RASTER3D (*34*) for structure depiction, and DALI (*35*) for identification of structural homologues. The protein−protein interfaces were analyzed using SPOCK (*36*), and the protein interaction server (http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html). The final refinement statistics are given in Table 2.

## RESULTS

*The Final Model.* The final crystallographic *R*-factor is 17.4%, and $R_{free}$ is 22.9%. The asymmetric unit contains one ThiF dimer (chains A and C) and two molecules of ThiS (chains B and D). The model also contains 390 water molecules, two zinc ions, two calcium ions, and two sodium ions. Electron density was missing for Asp176 to Thr186 and the six C-terminal residues of chain A, Met1 of chain B, and Pro181−Arg185 and the six C-terminal residues of chain C. These residues were omitted from the final model. Twenty-seven residues were refined with multiple conformations.

Table 2: Refinement Statistics

| | |
|---|---|
| resolution (Å) | 25−2.0 |
| wavelength (Å) | 0.9795 |
| completeness (%) | 92.9 (87.3)[a] |
| no. of unique reflections | 41515 |
| no. of protein atoms | 4700 |
| no. of waters | 390 |
| no. of heteroatoms | 6 |
| $R_{cryst}$[b] (%) | 17.4 |
| $R_{free}$ (%)[c] | 22.9 |
| rms deviation from ideal geometry | |
| bond lengths (Å) | 0.017 |
| bond angles (deg) | 1.8 |
| Ramachandran plot | |
| most favored region (%) | 90.5 |
| additional allowed region (%) | 9.1 |
| generously allowed region (%) | 0.4 |
| disallowed region (%) | 0.0 |
| average *B*-factors (Å$^2$) | |
| ThiF chain A | 30.1 |
| ThiF chain B | 50.5 |
| ThiS chain C | 29.0 |
| ThiS chain D | 36.9 |
| waters | 43.1 |

[a] The values in parentheses are for the highest-resolution shell. [b] $R_{cryst} = \sum_{hkl}||F_o| - |F_c||/\sum_{hkl}|F_o|$, where $F_o$ and $F_c$ are the observed and calculated structure factors, respectively. [c] $R_{free}$ is computed from 2080 reflections that were randomly selected and omitted from the refinement.

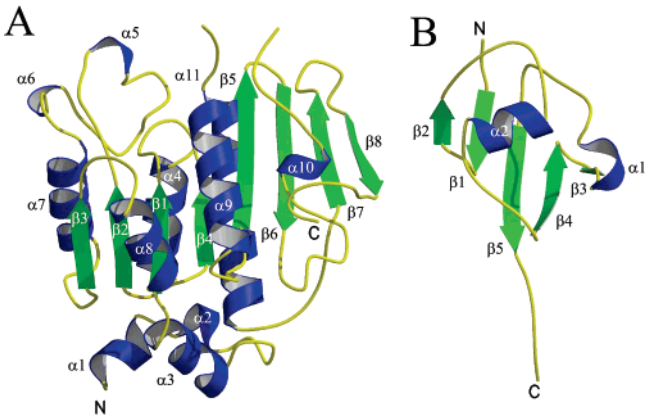



FIGURE 1: Overall folds of ThiF and ThiS. Helices are colored blue, and strands are colored green. The N- and C-termini are labeled N and C, respectively. α-Helices and *β*-strands are labeled with Greek letters. (A) Ribbon diagram of ThiF. (B) Ribbon diagram of ThiS.

*Description of the Structure.* ThiF is an α/*β* protein consisting of an eight-stranded mixed *β*-sheet with the topology *β*3↑*β*2↑*β*1↑*β*4↑*β*5↑*β*6↓*β*7↑*β*8↓, which is flanked by six α-helices and two $3_{10}$-helices (5 and 6) on one side and three α-helices on the other side (Figure 1A). The root-mean-square deviation (rmsd) for the two monomers of ThiF in the asymmetric unit is 0.27 Å. Comparison of ThiF to known structures in the Protein Data Bank using DALI revealed several structural relatives. MoeB (PDB entry 1JW9) gave the best fit based on the DALI Z score, with an rmsd of 1.3 Å for 235 (240 total) aligned Cα positions. This is consistent with the high level of sequence identity (46%) between MoeB and ThiF. The second best fit corresponded to the small ubiquitin-related modifier (SUMO) activating enzyme Sae1 (PDB entry 1Y8Q, chain A) from the Sae1−Sae2 complex structure (*37*), with an rmsd of 2.0 Å for 219 (313 total) aligned Cα positions. The third best fit corresponded to the APPBP1 monomer (PDB entry 1NGV, chain A) from

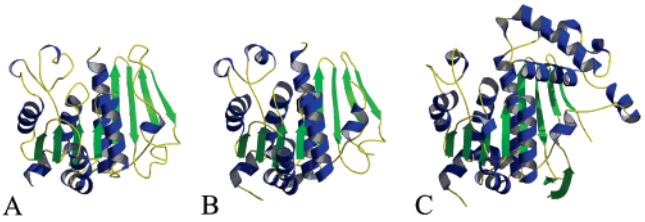



FIGURE 2: Comparison of ThiF, MoeB, and ubiquitin activating enzyme. The molecules are oriented on the basis of the DALI alignments. Ribbon diagrams of (A) ThiF, (B) MoeB (PDB entry 1JW9) (*15*), and (C) a Ubl activating enzyme (PDB entry 1Y8Q) (*37*). The color scheme is the same as in Figure 1.

the NEDD8 activating enzyme APPBP1−UBA3 (*38*), with an rmsd of 1.8 Å for 218 (529 total) aligned Cα positions (Figure 2). Eukaryotic ThiF homologues are generally larger than ThiF and contain additional domains. Furthermore, the level of sequence identity in the region of structural similarity is only ∼20%. In contrast to homodimeric ThiF, the Sae1−Sae2 and APPBP1−UBA3 species are heterodimers that show structural similarity between the two monomers. As a result, the Sae1 partner, Sae2 (PDB entry 1Y8Q, chain B), and the APPBP1 partner, UBA3 (PDB entry 1NGV, chain B), are the next two structures in the DALI search, with rmsd's and sequence identity levels comparable to those of their partners.

ThiF contains a Zn−sulfur center in which zinc is tetrahedrally coordinated by the sulfur atoms of cysteine residues Cys169, Cys172, Cys240, and Cys243. The presence of zinc is consistent with anomalous and dispersive difference maps calculated using the two synchrotron data sets (Table 1). Additional features in the difference electron density maps (calculated between the two synchrotron data sets) indicated radiation damage, especially for the carboxylate groups of Asp124 and Asp59 and the methylthio group of Met159. Interestingly, the radiation damage occurred in both ThiF chains at the same locations and to similar extents as judged by the peak heights in the difference map.

ThiS is an α/*β* protein adopting a ubiquitin-like fold. A five-stranded mixed *β*-sheet with the topology *β*2↑*β*1↓*β*5↓*β*4↑*β*3↓ is flanked on one side by an α-helix and a $3_{10}$-helix (Figure 1B). The rmsd for the Cα positions between the two ThiS molecules in the asymmetric unit is 0.9 Å. Difference dissected distance matrices (*39*) indicated that the C-terminus is the most rigid part of the molecule in this structure. Neglecting residues 1−28 and 41−52 from the calculation gives an rmsd of 0.25 Å. In addition, the temperature factors are lowest in the C-terminal region and highest for the $3_{10}$-helix and the two loops containing residues 10−18 and residues 46−56. These two loops run in parallel, and their structural differences are directionally correlated perpendicular to that direction.

Comparison to the ThiS solution structure (*13*) (PDB entry 1F0Z) gives an rmsd of 2.1 Å. Neglecting residues 59−66, which are poorly defined in the NMR structure, yields an rmsd of 1.2 Å. Comparing *E. coli* ThiS to the ThiS from the *B. subtilis* ThiS−ThiG complex (*26*) (PDB entry 1TYG), which are 22% identical in sequence, gives an rmsd of 1.8 Å for 64 of 66 Cα positions. The structure of MoaD (*17*) (PDB entry 1FMA) superimposes with an rmsd of 2.4 Å for 59 of 81 aligned Cα positions. The sequence of MoaD is 24% identical with that of ThiS. ThiS also is structurally

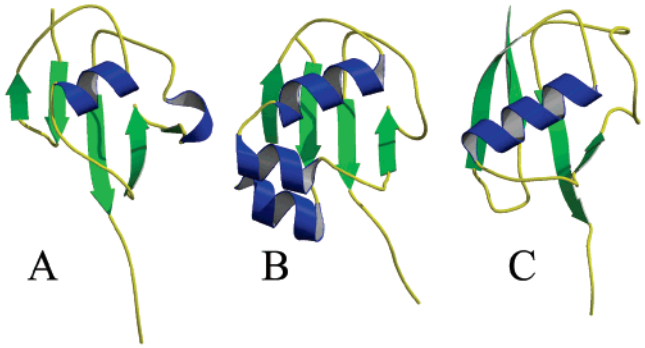



FIGURE 3: Comparison of ThiS, MoaD, and ubiquitin. The molecules are oriented on the basis of the DALI alignments. Ribbon diagram of (A) ThiS, (B) MoaD (PDB entry 1JW9) (*15*), and (C) ubiquitin (PDB entry 1UBI) (*42*). The color scheme is the same as in Figure 1.

similar to ubiquitin (*40*) (PDB entry 1UBI). Superposition gives an rmsd of 2.6 Å for 60 of 79 aligned Cα positions (Figure 3). The level of sequence identity between ThiS and ubiquitin is 13%.

*Quaternary Structure.* The ThiS−ThiF complex is a heterotetramer consisting of a dimer of heterodimers. The heterotetramer displays 2-fold symmetry and has approximate dimensions of 55 Å × 55 Å × 65 Å (Figure 4). ThiF forms a central globular dimer, which is complexed with two ThiS molecules. There are no contacts between the two ThiS molecules, resulting in a large cleft with a minimum ThiS−ThiS distance of 20 Å. Each ThiS binds a deep crevice of one ThiF with its C-terminus; ThiF chain A binds to ThiS chain B, and ThiF chain C binds to ThiS chain D.

The main interactions between the two ThiF molecules are hydrophobic, in which 68% of the atoms are nonpolar. Surface atoms belong mainly to residues located on helices, namely, α2, α4, $3_{10}$-helix 5, α7, and α11. There are 14 hydrogen bonds involved in the ThiF−ThiF interface, and 22 bridging waters can be found between chains A and C. The ThiF dimer interface buries 22% of the total accessible surface area of ThiF.

The interface between ThiF and ThiS is also mostly hydrophobic with 60% of the atoms involved in the ThiS−ThiF interface being nonpolar. The ThiF surface utilizes the turn preceding α4, the stretch before α9, Arg70, the four strands from β5 to β8, and α10 from the antiparallel part of the central β-sheet to form the protein−protein interface, while the ThiS surface utilizes β3, β4, Asp7, Leu58, and the C-terminus from Phe60 to Gly66. There are 14 hydrogen bonds between ThiS and ThiF, seven bridging waters, and a salt bridge between Asp7 of ThiS and Arg230 of ThiF. The ThiS−ThiF interface buries 9% of the total accessible surface area of ThiF and 24% of the surface area for ThiS.

*ThiF Active Site.* The structure of ThiF alone and with bound ATP was recently determined (*25*), and this allowed us to model the ThiF−ThiS−ATP complex (Figure 5). In this model, each ThiSF tetramer contains two active sites and each active site is composed of residues from both ThiF monomers. The ThiS C-terminus can be found at the bottom of the active site, and close to the ATP α-phosphate. The active site is lined with residues from five conserved sequence stretches. The residues beneath the adenosine base and the ribose are composed of a glycine rich nucleotide-binding motif reaching from residue Leu32 to Gly40. The nucleotide-binding loop can be found between β1 and α3, and it interacts with the ATP in the active site. Asp59 and Asp61 make up a conserved DXD motif located on the loop between β2 and $3_{10}$-helix 5, and are within hydrogen bonding distance of the ribose hydroxyl groups. The $3_{10}$-helix 5 is highly conserved in all E1's and maintains several interactions with the ATP β- and γ-phosphate groups. This helix is believed to stabilize the outgoing pyrophosphate group after nucleophilic attack of the ThiS C-terminal carboxylate on ATP's α-phosphate. Asp127, located in the loop between β4 and α9, has been predicted to bind a $Mg^{2+}$ ion (*15*), which would activate the α-phosphate for nucleophilic attack. An equivalent $Mg^{2+}$ ion was observed in the structure of the Sae1−Sae2 complex (*37*). Arg132 on α9 is responsible for the positioning of the ThiS C-terminal Gly-Gly motif. Arg132 maintains a hydrogen bond to Ala64 of ThiS, thus leaving some flexibility in the C-terminus for attacking ATP. Arg11 from the other ThiF monomer in the dimer makes a salt bridge to the α-phosphate of ATP. The disordered loop region between Pro181 and Arg185, called the "crossover" loop, contains the highly conserved Cys184 involved in the formation of ThiF−ThiS acyl disulfide **5**. In the ThiF−ATP structure, the thiol group of this cysteine is nearly 20 Å from the α-phosphate of ATP, suggesting that a large conformational change in this region of the protein is required for the formation of acyl disulfide **5**.

ThiF is a zinc-containing protein. The structure demonstrates that the zinc ion is more then 20 Å from the active site and most likely has a structural rather than a catalytic role (*25*).

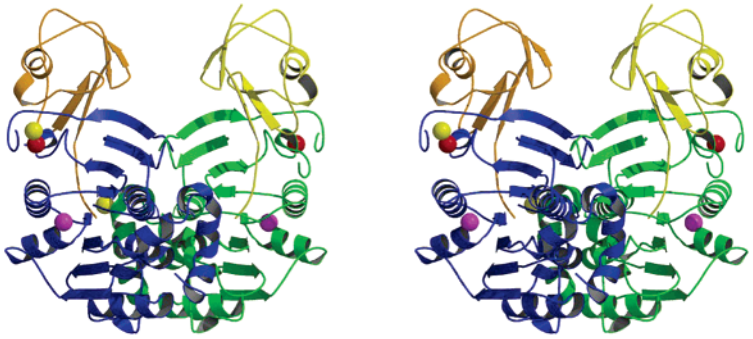


FIGURE 4: Stereoscopic ribbon diagram of the ThiS−ThiF complex. The view is perpendicular to the molecular 2-fold axis. The complex is labeled according to chain names. ThiS molecules are colored yellow and orange, and ThiF monomers are colored green and blue. Zinc, calcium, and sodium ions are depicted as spheres colored red, yellow, and magenta, respectively.

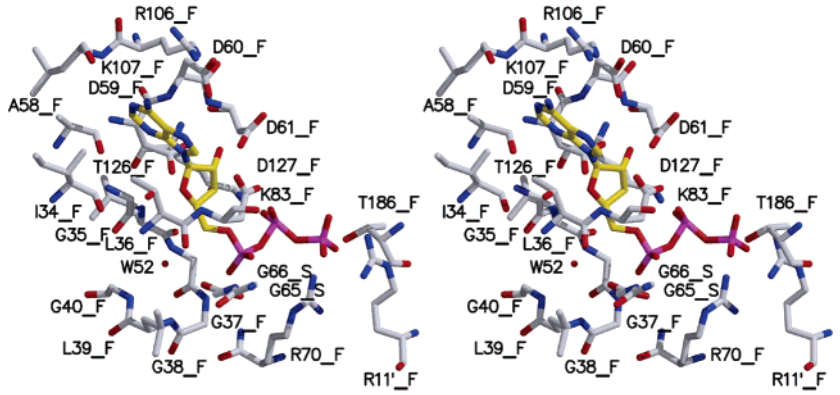



FIGURE 5: Stereoscopic model of the ThiF−ThiS−ATP complex. Oxygens are colored red, nitrogens blue, carbon atoms of ATP gold, all other carbons gray, and phosphorous atoms magenta. Water 52 is shown as a red sphere. Residues from ThiS are labeled with _S and those from ThiF with _F.

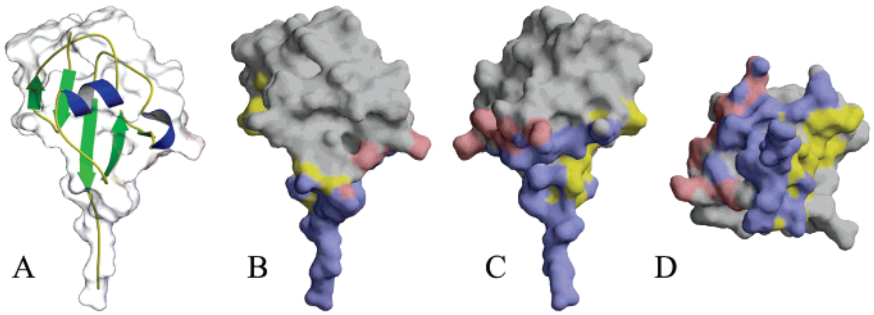



FIGURE 6: Surface representation of *B. subtilis* ThiS. (A) Ribbon diagram of ThiS. Helices are colored blue, and strands are colored green. The ThiS surface is rendered transparent. (B) Buried surfaces of the ThiS−ThiF and ThiS−ThiG complexes. The buried surface that is common to both complexes is colored blue, the surface unique to the ThiS−ThiG complex yellow, and the surface unique to the ThiS−ThiF complex red. (C) ThiS−ThiF complex rotated 180° about the vertical axis compared to panel A. (D) ThiS−ThiF complex rotated 90° about the horizontal axis compared to panel C.

## DISCUSSION

*Comparison of the ThiF−ThiS Complex to ThiF Alone*. The recently published structure of ThiF alone (*25*) superimposes with an rmsd of 0.69 Å to the ThiF of the ThiF−ThiS complex described here. The largest differences can be found between the C- and N-terminal residues as well as for the last ordered residues of the crossover loop. The Cα positions of Thr186 from the two structures differ by more then 5 Å. The predicted ThiS−ThiF model is consistent with this study with respect to the relative orientations of ThiS and ThiF. A closer examination of the crossover loop revealed that the aromatic side chain of Trp174 adopts a different orientation in the complex to avoid a close contact with Gly31 from ThiS, while the side chain of Asp176 of ThiF is oriented closer to ThiS to accept a hydrogen bond from the amide nitrogen of Gly31 of ThiS. These structural differences show that the crossover loop adopts a different conformation in the complex compared to that with ThiF alone.

*Comparison of the ThiS−ThiF Complex to the ThiS−ThiG Complex*. ThiS forms complexes with both ThiF and ThiG, and the structures of the two complexes can be used to compare the two ThiS binding modes. Because the ThiS−ThiF structure is from *E. coli* and the ThiS−ThiG structure is from *B. subtilis* (*26*), a *B. subtilis* ThiS−ThiF structure was created by homology modeling to allow a more direct comparison. The ThiS−ThiG X-ray structure was superimposed onto the ThiS−ThiF homology model using the two ThiS molecules as the reference, and the binding surfaces were calculated. The surface areas utilized by ThiS to bind ThiF or ThiG are similar in size and overlap significantly, with roughly 70% of the ThiS binding surface common to both complexes. The hydrophobic contents of the two interfaces are roughly the same. Most of the hydrophobic interactions are contributed by the C-terminus of ThiS and hydrophobic residues surrounding the base of the C-terminal extension of ThiS. In both the ThiS−ThiG and ThiS−ThiF complexes, an arginine residue (Arg230 in ThiF and Arg183 in ThiG) takes roughly the same position relative to ThiS. In the ThiS−ThiF complex, the arginine forms a salt bridge with Asp7 of ThiS, and in the ThiS−ThiG complex, the arginine forms a salt bridge with Glu35 of ThiS.

The surface areas unique to either ThiF or ThiG lie on opposite sides of ThiS (Figure 6). In the region of ThiS utilized by only ThiG, a hydrophobic pocket formed by residues Leu4, Asn5, Lys7, Val9, and Tyr25 accepts Ile149 and Phe150 from ThiG. These residues are part of an extended $\beta 2$ and $\alpha 2$ loop of the ThiG $(\beta\alpha)_8$ barrel. In the region of ThiS utilized by only ThiF, Glu44 of ThiS forms two hydrogen bonds with backbone amide groups of positions 2 and 3 of a reverse turn located just before the C-terminus of ThiF. In the homology model, aspartic acid (position 2) and alanine (position 3) occupy the corners of the turn, while in the X-ray structure of the ThiF−ThiS complex, these residues correspond to Ser238 and Gly239, respectively. In the X-ray structure, the reverse turn is type I, while in the homology model, the turn adopts a significantly different conformation. However, Glu44 is conserved

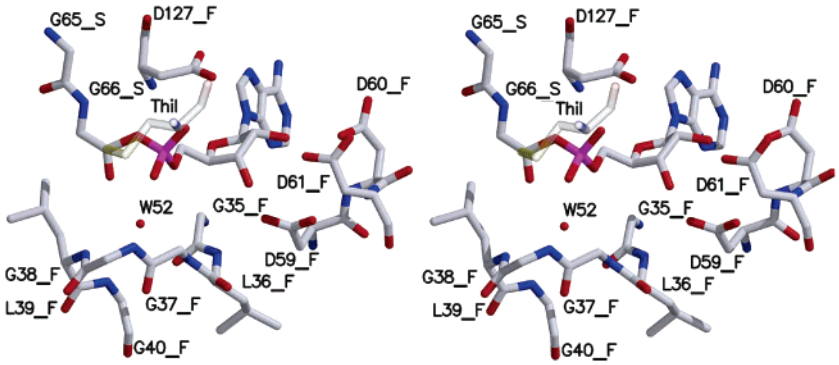



FIGURE 7: Stereoscopic model of the ThiF−ThiS−COOAMP intermediate. Oxygens are colored red, nitrogens blue, carbons gray, phosphorous atoms magenta, and sulfurs yellow. The ThiI Cys-peroxide, labeled ThiI, is modeled in the approximate orientation for nucleophilic attack and are transparent. Water 52 is shown as a red sphere. Residues from ThiS are labeled with _S and those from ThiF with _F.

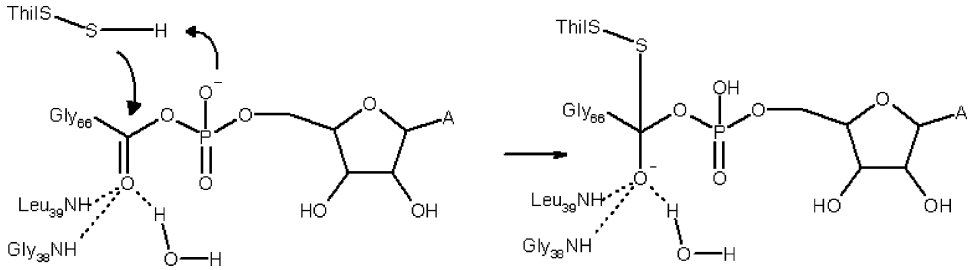



FIGURE 8: Mechanism for activation of the acyl adenylate intermediate. The water molecule is also within hydrogen bonding distances of the amide groups of Gly37 and Gly40 and the carbonyl group of Cys125.

between *E. coli* ThiS and *B. subtilis* ThiS, suggesting that this interaction may also be conserved.

In considering the homology model, it is important to note that *B. subtilis* ThiF, as well as ThiF from several related bacteria, contains an additional C-terminal domain of unknown structure and function. This roughly 90-amino acid domain has a high degree of sequence similarity within these bacterial systems. Interestingly, a homologous domain is also present in MoeB from *B. subtilis*, but not MoeB from *E. coli*. No significant sequence similarity has been detected between the extra C-terminal domains of ThiF or MoeB, and the C-terminal domains from related systems such as E1 or Ubl activating enzymes.

*Mechanistic Implications.* Our current understanding of the mechanism of thiazole formation is outlined in Scheme 2. Two pathways have been identified in bacteria. In the first pathway, which is found in *E. coli*, ThiF catalyzes the adenylation of ThiS on its carboxy terminus. The resulting acyl adenylate **2** reacts with the persulfide of ThiI **3** to form **4**. Disulfide interchange with the thiol of Cys184 of ThiF gives ThiF−ThiS acyl disulfide **5**, which has been detected as an intermediate (*6*). The subsequent steps involved in the conversion of **5** to thiazole **8** are still poorly understood (*9*). One possibility for the next step in the reaction sequence is that enamine **12**, formed between deoxy-D-xylulose 5-phosphate **6** and ThiG, reacts with **5** to form **13** (*10*).

We have carried out modeling studies to determine if the reactions involved in the conversion of **1** to **13** are consistent with the structure of the ThiF−ThiS complex. Using our ThiF−ThiS structure and the previously published ThiF−ATP structure, we generated a model for the acyl adenylate of ThiS complexed to ThiF **2** (Figure 7). This model suggests that for the conversion of **2** to **4**, one of the phosphate oxygens of the acyl adenylate is positioned to deprotonate the persulfide **3** and the carbonyl group of the acyl adenylate is activated by hydrogen bonding to the amide hydrogens of Gly38 and Leu39 and a water molecule (Figure 8). We then modeled the formation of **4** by positioning a cysteine persulfide on the reaction trajectory for nucleophilic addition to the acyl adenylate of **2** (Figure 8). This reaction occurs on an open face of ThiF that should be accessible to ThiI persulfide **3**. The ThiF−ThiS acyl disulfide **5** was generated by replacement of the ThiI cysteine with Cys184 of ThiF. This cysteine is nearly 20 Å from the α-phosphate of ATP in the ThiF−ATP structure but is on a disordered loop in our ThiF−ThiS structure. However, if we model the missing amino acids, it is not unreasonable that Cys184 could carry out a disulfide interchange with the acyl disulfide in **4**.

In the second pathway for thiazole formation, which is found in *B. subtilis*, ThiF also catalyzes the adenylation of ThiS on its carboxy terminus. The resulting acyl adenylate **2** reacts with the persulfide of IscS (not ThiI) to form an analogue of **4**. While we have been unable to detect a ThiF−ThiS acyl disulfide in this system, it is likely that **5** is formed because Cys184 is a conserved residue in ThiF for both pathways. This suggests that the acyl disulfide is rapidly reduced to thiocarboxylate **14**, which has been demonstrated to function as the sulfur source for thiazole formation in *B. subtilis* (*10*, *41*).

All ThiF proteins associated with this second pathway have an additional carboxy-terminal domain that contains a conserved cysteine (Cys261 for *B. subtilis* ThiF). This cysteine may cleave the acyl disulfide of **5** by a disulfide interchange. ThiS-thiocarboxylate **14** then adds to **15** to give **16**, which undergoes a sequence of elimination and/or dehydration reactions followed by reaction with the glycine imine **17** to give thiazole **8** (*10*, *26*). The different redox chemistry of thiazole formation in *E. coli* and *B. subtilis* probably reflects the different oxygen requirements of the two microorganisms. *E. coli* can grow readily in the presence

and absence of oxygen and therefore must be able to carry out the two-electron oxidation required for thiazole biosynthesis by an oxygen-independent pathway (i.e., thiol−disulfide chemistry). In contrast, *B. subtilis* grows very poorly in the absence of oxygen and during its normal aerobic growth uses oxygen directly to mediate the two-electron oxidation required for thiazole formation by oxidizing glycine to its imine **17**, using a flavoenzyme (*24*).

## ACKNOWLEDGMENT

We thank Leslie Kinsland for assistance in the preparation of the manuscript, Cynthia Kinsland for discussions regarding the ThiS−ThiF overexpression system, the beamline staff at CHESS beamline F2, supported by NIH grant RR01646, and the staff at NE-CAT beamline 8BM at the APS, supported by NIH grant RR15301, for assistance during data collection.

## REFERENCES

1. Jordan, F. (2003) Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions, *Nat. Prod. Rep. 20*, 184−201.
2. Martin, P. R., Singleton, C. K., and Hiller-Sturmhofel, S. (2003) The role of thiamine deficiency in alcoholic brain disease, *Alcohol Res. Health 27*, 134−142.
3. Butterworth, R. F. (2003) Thiamin deficiency and brain disorders, *Nutr. Res. Rev. 16*, 277−283.
4. 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.
5. Begley, T. P., and Ealick, S. E. (2004) Mechanistic and structural studies on thiamine biosynthetic enzymes, *Oxid. Stress Dis. 11*, 15−28.
6. Xi, J., Ge, Y., Kinsland, C., McLafferty, F. W., and Begley, T. P. (2001) Biosynthesis of the thiazole moiety of thiamin in *Escherichia coli*: Identification of an acyldisulfide-linked protein−protein conjugate that is functionally analogous to the ubiquitin/E1 complex, *Proc. Natl. Acad. Sci. U.S.A. 98*, 8513−8518.
7. Lauhon, C. T., and Kambampati, R. (2000) The iscS gene in *Escherichia coli* is required for the biosynthesis of 4-thiouridine, thiamin, and NAD, *J. Biol. Chem. 275*, 20096−20103.
8. Taylor, S. V., Kelleher, N. L., Kinsland, C., Chiu, H. J., Costello, C. A., Backstrom, A. D., McLafferty, F. W., and Begley, T. P. (1998) Thiamin biosynthesis in *Escherichia coli*. Identification of this thiocarboxylate as the immediate sulfur donor in the thiazole formation, *J. Biol. Chem. 273*, 16555−16560.
9. Leonardi, R., Fairhurst, S. A., Kriek, M., Lowe, D. J., and Roach, P. L. (2003) Thiamine biosynthesis in *Escherichia coli*: isolation and initial characterization of the ThiGH complex, *FEBS Lett. 539*, 95−99.
10. Dorrestein, P. C., Zhai, H., McLafferty, F. W., and Begley, T. P. (2004) The biosynthesis of the thiazole phosphate moiety of thiamin: The sulfur transfer mediated by the sulfur carrier protein ThiS, *Chem. Biol. 11*, 1373−1381.
11. Peapus, D. H., Chiu, H.-J., Campobasso, N., Reddick, J. J., Begley, T. P., and Ealick, S. E. (2001) Structural characterization of the enzyme−substrate, enzyme-intermediate, and enzyme−product complexes of thiamin phosphate synthase, *Biochemistry 40*, 10103−10114.
12. Webb, E., and Downs, D. (1997) Characterization of thiL, encoding thiamin-monophosphate kinase, in *Salmonella typhimurium*, *J. Biol. Chem. 272*, 15702−15707.
13. Wang, C., Xi, J., Begley, T. P., and Nicholson, L. K. (2001) Solution structure of ThiS and implications for the evolutionary roots of ubiquitin, *Nat. Struct. Biol. 8*, 47−51.
14. Appleyard, M. V., Sloan, J., Kana'n, G. J., Heck, I. S., Kinghorn, J. R., and Unkles, S. E. (1998) The *Aspergillus nidulans* cnxF gene and its involvement in molybdopterin biosynthesis. Molecular characterization and analysis of in vivo generated mutants, *J. Biol. Chem. 273*, 14869−14876.
15. Lake, M. W., Wuebbens, M. M., Rajagopalan, K. V., and Schindelin, H. (2001) Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB−MoaD complex, *Nature 414*, 325−329.
16. Hochstrasser, M. (2000) Evolution and function of ubiquitin-like protein-conjugation systems, *Nat. Cell Biol. 2*, E153−E157.
17. Rudolph, M. J., Wuebbens, M. M., Rajagopalan, K. V., and Schindelin, H. (2001) Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation, *Nat. Struct. Biol. 8*, 42−46.
18. Pickart, C. M. (2001) Mechanisms underlying ubiquitination, *Annu. Rev. Biochem. 70*, 503−533.
19. Zheng, N., Wang, P., Jeffrey, P. D., and Pavletich, N. P. (2000) Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases, *Cell 102*, 533−539.
20. 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.
21. Toms, A. V., Haas, A. L., Park, J. H., Begley, T. P., and Ealick, S. E. (2005) Structural characterization of the regulatory proteins TenA and TenI from *Bacillus subtilis* and identification of TenA as a thiaminase II, *Biochemistry 44*, 2319−2329.
22. Campobasso, N., Mathews, I. I., Begley, T. P., and Ealick, S. E. (2000) Crystal structure of 4-methyl-5-$\beta$-hydroxyethylthiazole kinase from *Bacillus subtilis* at 1.5 Å resolution, *Biochemistry 39*, 7868−7877.
23. Chiu, H. J., Reddick, J. J., Begley, T. P., and Ealick, S. E. (1999) Crystal structure of thiamin phosphate synthase from *Bacillus subtilis* at 1.25 Å resolution, *Biochemistry 38*, 6460−6470.
24. Settembre, E. C., Dorrestein, P. C., Park, J. H., Augustine, A. M., Begley, T. P., and Ealick, S. E. (2003) Structural and mechanistic studies on ThiO, a glycine oxidase essential for thiamin biosynthesis in *Bacillus subtilis*, *Biochemistry 42*, 2971−2981.
25. Duda, D. M., Walden, H., Sfondouris, J., and Schulman, B. A. (2005) Structural analysis of *Escherichia coli* ThiF, *J. Mol. Biol. 349*, 774−786.
26. Settembre, E. C., Dorrestein, P. C., Zhai, H., Chatterjee, A., McLafferty, F. W., Begley, T. P., and Ealick, S. E. (2004) Thiamin biosynthesis in *Bacillus subtilis*: Structure of the thiazole synthase/sulfur carrier protein complex, *Biochemistry 43*, 11647−11657.
27. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem. 72*, 248−254.
28. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307−326.
29. Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement, *Acta Crystallogr. D56*, 1622−1624.
30. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron density maps and the location of errors in these models., *Acta Crystallogr. A47*, 110−119.
31. 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.
32. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
33. Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structure, *J. Appl. Crystallogr. 24*, 946−950.
34. Merritt, E. A., and Bacon, D. J. (1997) Raster3D Photorealistic Molecular Graphics, *Methods Enzymol. 277*, 505−524.
35. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol. 233*, 123−138.
36. Christopher, J. A. (1998) *SPOCK*, Texas A&M University, College Station, TX.
37. Lois, L. M., and Lima, C. D. (2005) Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1, *EMBO J. 24*, 439−451.
38. Walden, H., Podgorski, M. S., Huang, D. T., Miller, D. W., Howard, R. J., Minor, D. L., Jr., Holton, J. M., and Schulman, B.

A. (2003) The structure of the APPBP1-UBA3-NEDD8-ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1, *Mol. Cell 12*, 1427−1437.
39. Richards, F. M., and Kundrot, C. E. (1988) Identification of structural motifs from protein coordinate data: Secondary structure and first-level supersecondary structure, *Proteins 3*, 71−84.
40. Ramage, R., Green, J., Muir, T. W., Ogunjobi, O. M., Love, S., and Shaw, K. (1994) Synthetic, structural and biological studies of the ubiquitin system: The total chemical synthesis of ubiquitin, *Biochem. J. 299* (Part 1), 151−158.
41. Park, J.-H., Dorrestein, P. C., Zhai, H., Kinsland, C., McLafferty, F. W., and Begley, T. P. (2003) Biosynthesis of the Thiazole Moiety of Thiamin Pyrophosphate (Vitamin B1), *Biochemistry 42*, 12430−12438.
42. Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., Vierstra, R. D., Hatfield, P. M., and Cook, W. J. (1987) Comparison of the three-dimensional structures of human, yeast, and oat ubiquitin, *J. Biol. Chem. 262*, 6396−6399.

BI051502Y