Crystal Structure of Thiamin Phosphate Synthase from *Bacillus subtilis* at 1.25 Å Resolution[†,‡]

Hsiu-Ju Chiu, Jason J. Reddick, Tadhg P. Begley, and Steven E. Ealick*

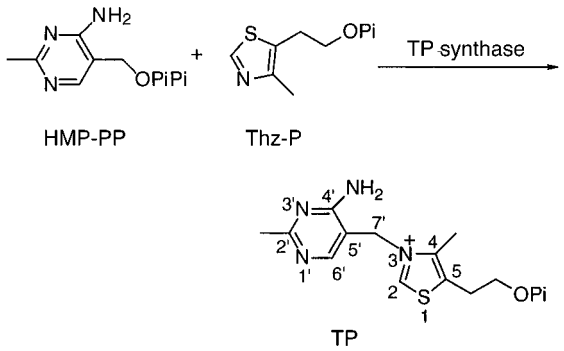
Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853

Received December 9, 1998

ABSTRACT: The crystal structure of *Bacillus subtilis* thiamin phosphate synthase complexed with the reaction products thiamin phosphate and pyrophosphate has been determined by multiwavelength anomalous diffraction phasing techniques and refined to 1.25 Å resolution. Thiamin phosphate synthase is an α/β protein with a triosephosphate isomerase fold. The active site is in a pocket formed primarily by the loop regions, residues 59−67 (A loop, joining α3 and β2), residues 109−114 (B loop, joining α5 and β4), and residues 151−168 (C loop, joining α7 and β6). The high-resolution structure of thiamin phosphate synthase complexed with its reaction products described here provides a detailed picture of the catalytically important interactions between the enzyme and the substrates. The structure and other mechanistic studies are consistent with a reaction mechanism involving the ionization of 4-amino-2-methyl-5-hydroxymethylpyrimidine pyrophosphate at the active site to give the pyrimidine carbocation. Trapping of the carbocation by the thiazole followed by product dissociation completes the reaction. The ionization step is catalyzed by orienting the C−O bond perpendicular to the plane of the pyrimidine, by hydrogen bonding between the C4′ amino group and one of the terminal oxygen atoms of the pyrophosphate, and by extensive hydrogen bonding and electrostatic interactions between the pyrophosphate and the enzyme.

Thiamin is an essential component of the human diet with a daily requirement of 1.4 mg. Thiamin deficiency results in beriberi and Wernicke's disease. This vitamin is added to many commercial foods since, unlike microorganisms, humans cannot synthesize it. Annual production, all by chemical synthesis, is on the order of 3300 tons (*1*). Thiamin-dependent enzymes play a particularly important role in carbohydrate metabolism and include transketolase, pyruvate decarboxylase, pyruvate oxidase, and acetolacetate synthase. Our understanding of this class of enzymes is well developed, where in all cases the mechanistic role of thiamin is the stabilization of the acyl carbanion (*2*). In contrast, our understanding of the mechanistic enzymology of thiamin biosynthesis is still at a relatively early stage (*3*−*6*).

In all systems examined thus far, the biosynthesis of thiamin involves the separate formation of 4-amino-2-methyl-5-hydroxymethylpyrimidine pyrophosphate (HMP-PP) and 4-methyl-5-hydroxyethylthiazole phosphate (Thz-P). These precursors are then coupled to form thiamin phosphate (TP) in a reaction catalyzed by thiamin phosphate synthase (eq 1).



(1)

Thiamin phosphate synthase has been overexpressed and purified from *Escherichia coli (7, 8*), *Bacillus subtilis (9*), and *S. cerevisiae (10*). The *E. coli* enzyme does not overexpress well due to inclusion body formation. The *B. subtilis* enzyme is small (25 277 Da), and can be overexpressed at a high level in *E. coli*. The bifunctional yeast enzyme is large (octamer of 58 058 Da subunits) and also catalyzes the phosphorylation of 4-methyl-5-hydroxyethylthiazole. Here, we report the X-ray structure of *B. subtilis* TP synthase complexed with the reaction products thiamin phosphate and pyrophosphate at 1.25 Å resolution.

MATERIALS AND METHODS

Preparation of Native and SeMet TP Synthase. Native TP synthase was overexpressed and purified as described previously (*9*). Selenomethionyl TP synthase was prepared by transforming plasmid pYZC6927 into the *E. coli* Met

[†] This work was supported by National Institutes of Health Grants to S.E.E. (RR01646 and GM46733) and to T.P.B. (DK44083) and by Hoffmann−La Roche to T.P.B.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 thiamin phosphate synthase is 2tps.

* Corresponding author: Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853. Telephone: (607)255-7961. Fax: (607)255-1227. E-mail: see3@cornell.edu.

10.1021/bi982903z CCC: $18.00
Published on Web 04/30/1999

auxotrophic DE3 strain (Novagen). A 2.5 mL starter culture of the DE3 strain containing the pYZC6927 expression plasmid was grown overnight at 37 °C in LB medium. The overnight culture was centrifuged, and the cell pellet was washed twice with fresh M9 medium containing selenomethionine (SeMet). The starter culture was diluted 1/200 (v/v) into 5 L of M9 medium supplemented with SeMet (40 mg/L) in the fermentor. Cells were grown at 37 °C to an OD_{600} of 0.6, and then expression was induced by 2 mM isopropyl-β-D-thiogalactopyranoside. Cells were grown for a further 12 h and then harvested by centrifugation (15 min, 5000*g*). SeMet TP synthase was purified in the same way as native TP synthase. Once SeMet TP synthase was eluted from a Ni^{2+}-NTA column, dithiothreitol (DTT) was added to a final concentration of 2 mM to prevent the oxidation of SeMet residues. The protein purity was analyzed by SDS−PAGE, and the protein concentration was determined using Coomassie Plus Protein Assay Reagent from Pierce (Rockford, IL). About 100 mg of pure SeMet TP synthase was obtained from a 5 L culture. The SeMet incorporation was determined to be approximately 100% by amino acid composition analysis (Cornell Biotechnology Analytical/Synthesis Facility) and electrospray ionization Fourier transform mass spectrometry.

Mutagenesis. An S130A mutant was prepared using the QuikChange site-directed mutagenesis kit (Stratagene). The oligonucleotide primers used in the mutagenesis reaction were 5′-ATTCTCGGCGTTGCTGCCCATACGATG-3′ and 5′-CATCGTATGGGCAGCAACGCCGAGAAT-3′, where the underlined codons indicate the site of the mutation. The mutation was confirmed by DNA sequencing. The S130A mutant was overexpressed and purified as described for native TP synthase.

Enzymatic Synthesis of HMP-PP. HMP-PP was prepared from pure HMP and ATP using HMP-P kinase which also has an HMP kinase activity (*11*). The reaction mixture in 1 mL of triethanolamine hydrochloride, pH 8, consisted of 200 μM HMP, 5 mM ATP, and 10 mM $MgCl_2$. The reaction was initiated by adding 20 μg of HMP-P kinase. The mixture was incubated at 25 °C for 30 min, at which point the reaction had gone to completion, and was used as a source of stock HMP-PP without further purification.

Assay for Thiamin Phosphate Synthase. TP synthase activity was measured using modifications of the thiochrome method (*9*, *12*). The assay mixture consisted of 50 mM Tris-HCl buffer, pH 7.5, containing 50 μM thiazole phosphate, 50 μM HMP-PP, and 6 mM $MgSO_4$ with a total volume of 500 μL. The reaction was initiated by the addition of 2 μg of TP synthase. Time points were taken at 30 s intervals by quenching 60 μL aliquots in 60 μL of 10% trichloroacetic acid. Quenched samples were centrifuged at 16000*g* for 2 min to remove denatured protein. The pH was adjusted by adding 100 μL of supernatant to 200 μL of 4 M potassium acetate. The sample was oxidized by the addition of 100 μL of fresh 3.8 mM potassium ferricyanide in 7 M NaOH. The mixture was vigorously mixed and, after 30 s, was quenched by adding 100 μL of fresh 0.06% H_2O_2 in saturated KH_2PO_4. The samples were diluted with 500 μL ddH_2O, and the fluorescence was measured using a Perkin-Elmer MPF-44B Fluorescence Spectrophotometer; λ(excitation) = 365 nm, λ(emission) = 444 nm. Fluorescence units were converted to micromolar thiamin phosphate by calibration using a standard curve prepared from thiamin phosphate standards oxidized by the identical procedure. All activity assays were done in triplicate to ensure reproducibility and averaged over each time-point.

Crystallization. Crystallization conditions were screened using Crystal Screen kits I and II (Hampton Research). Further optimization resulted in crystallization conditions based on a hanging-drop vapor diffusion setup at 4 °C. A drop containing 3 μL of protein solution (20 mg/mL in 50 mM Tris-HCl buffer, pH 7.5, and 1 mM DTT) plus 3 μL of reservoir solution was equilibrated against a 1 mL reservoir solution [75 mM tris-HCl, 75 mM $MgCl_2$, 21−22% (w/v) PEG4000, and 1 mM DTT, pH 7.5]. After the crystallization tray had been set up for 20−24 h, the crystallization drops became cloudy and were seeded. Two or three small crystals were placed in 40 μL of seeding solution (same as the reservoir solution except that it contained 20% PEG4000 solution) and then crushed and diluted 250−1000-fold. Square pyramid-like crystals of dimensions 0.2 mm × 0.2 mm × 0.2 mm grew in 3−4 days.

SeMet TP synthase was crystallized under similar conditions to the native protein. The SeMet TP synthase was prepared in 50 mM Tris-HCl buffer, 2 mM DTT, pH 7.5, with a concentration of 20 mg/mL. The optimal crystallization condition was based on sitting drops at 4 °C with 24−26% PEG4000, 75 mM tris-HCl, 75 mM $MgCl_2$, 2 mM β-mercaptoethanol, 5 mM β-octylglucopyranoside, pH 7.3. The PEG4000 concentration was slightly higher, and the pH was slightly lower compared to the native TP synthase crystallization. Seeding the crystallization drops after they have been set up for 20−24 h was still necessary to obtain good crystals.

X-ray Data Collection and Processing. The diffraction data from crystals of native TP synthase were measured on beamline F1 at the Cornell High Energy Synchrotron Source (CHESS) using an ADSC Quantum 4 CCD detector. For data collection, a fresh crystal was transferred into a freezing solution (75 mM Tris-HCl, 30% PEG4000, 75 mM $MgCl_2$, 10% glycerol, pH 7.5) for several minutes, mounted on a goniometer in a glass fiber loop, and flash-frozen in a cold nitrogen stream at 100 K. The crystals of native TP synthase diffracted to 1.1 Å resolution; however, the high-resolution reflections decayed quickly (after 300 s of exposure) in the X-ray beam. A data set 81.6% complete to 1.25 Å resolution was collected using a single crystal. The data were processed with MOSFLM (*13*) and scaled in SCALA of CCP4 (*14*). The crystals belonged to the tetragonal space group $P4_32_12$. The cell parameters were refined with MOSFLM to be a = 76.83 Å, c = 140.10 Å for the native data. The Matthews coefficient (*15*) was 2.05 $Å^3$/dalton, corresponding to a solvent content of 35% and two molecules in one asymmetric unit. The native data collection is summarized in Table 1.

The multiwavelength anomalous diffraction data were collected on beamline F2 at CHESS, using a 1K CCD detector (*16*). The station monochromator was set at the absorption edge of selenium (12.658 keV) using a selenium foil as a reference for energy calibration. An energy scan with the SeMet TP synthase protein was used to select appropriate wavelengths to maximize the dispersive and anomalous signals. This energy scan was performed at room temperature using several crystals packed in a capillary tube. The fluorescence spectrum was recorded using a Bicron

Table 1: X-ray Data Statistics

	inflection $\lambda1$	peak $\lambda2$	remote $\lambda3$	native
space group		$P4_32_12$		$P4_32_12$
cell dimensions: *a*, *c* (Å)		76.09, 138.65		76.83, 140.10
wavelength (Å)	0.9795	0.9792	0.9677	0.908
resolution (Å)	2.9	2.9	2.9	1.25
mosaicity (deg)	0.561	0.562	0.567	0.15
no. of unique reflections	8672	8736	8643	94169
no. of reflections (both Bijvoets)	6860	6870	6777	−
redundancy	5.3	5.3	5.3	5.1
completeness (%)[a]	89.2 (80.1)	89.1 (80.2)	89.4 (80.5)	81.6 (48.5)
R_{sym} (%)[a,b]	4.4 (5.4)	5.3 (6.9)	5.5 (7.1)	6.2 (21.9)
R_{sym} (%)[a,c]	4.0 (4.6)	4.1 (5.1)	4.0 (5.2)	−
I/σ_I[a]	28.6 (24.0)	27.9 (24.3)	27.6 (24.0)	6.8 (3.4)

[a] Values in parentheses correspond to the outermost resolution shell (all reflections are divided into 10 resolution shells, 3.0−2.9 Å for MAD data; 1.32−1.25 Å for the native). [b] $R_{sym} = (\sum|I_{hkl} - \langle I\rangle|)/(\sum I_{hkl})$, where the average intensity $\langle I\rangle$ is taken over all symmetry equivalent measurements and I_{hkl} is the measured intensity for any given reflection. [c] Values are R_{sym} when Bijvoet are considered nonequivalent.

Table 2: Analysis of MAD Signals

	observed diffraction ratios[a]			expected diffraction ratios[b]			electrons[c]	
	$\lambda1$	$\lambda2$	$\lambda3$	$\lambda1$	$\lambda2$	$\lambda3$	f'	f''
$\lambda1$ (inflection)	0.058 (0.051)	0.041	0.058	0.031	0.005	0.034	−8.335	2.083
$\lambda2$ (peak)		0.071 (0.054)	0.063		0.049	0.043	−9.571	3.387
$\lambda3$ (remote)			0.076 (0.054)			0.050	−3.707	3.656

[a] Bijvoet difference ratios, $\langle|F(hkl) - F(-h,-k,-l)|^2\rangle^{1/2}/\langle|F|^2\rangle^{1/2}$, are given in diagonal elements. The numbers in parentheses are calculated using centric reflections only and taken as a measurement of the noise in the values next to them, as they theoretically should be zero. Dispersive difference ratios, $\langle|F(hkl)_{\lambda i} - F(hkl)_{\lambda j}|^2\rangle^{1/2}/\langle|F|^2\rangle^{1/2}$, are given as off-diagonal elements. Values were calculated in WVLSCL. [b] The expected Bijvoet difference ratios are calculated as $[N_{ano}/2]^{1/2}[2f''_{max}(\lambda)/\langle|F_P|\rangle]$, and the dispersive difference ratios are calculated as $[N_{ano}/2]^{1/2}[|f'(\lambda_i) - f'(\lambda_i)/|\langle|F_P|\rangle]$, where N_{ano} is the number of anomalous scatters, $\langle|F_P|\rangle = (\sum f^2)^{1/2}$, and is estimated as 6.7*(no. of atoms)$^{1/2}$, based on the assumption that the amino acid contains 7.7 non-hydrogen atoms each with 6.7 electrons. 6.7 electrons is the effective normal atomic scattering at zero scattering angle. The expected diffraction ratios were calculated with ABSOLUTE. [c] $f'$and f'' values were initially calculated in FPRIME and refined in MADLSQ.

fluorescence detector oriented 90° to the incident X-ray beam to get the best signal-to-noise ratio. The fluorescence spectrum was used to select the inflection point ($\lambda = 0.9795$ Å), the absorption edge ($\lambda = 0.9792$ Å), and a high-energy remote ($\lambda = 0.9677$ Å). During the data collection period, reference scans were performed at the beginning and the end of each fill to check for energy drift, and recalibration was performed as necessary. Diffraction data were measured in 1° oscillation ranges for 30 s each with a crystal to detector distance of 100 mm. A total of 81 images were measured for both direct and inverse beams for each of the 3 wavelengths. All the MAD data were collected from a single crystal, and the crystal was stable over the whole data collection period. The direct beam and inverse beam diffraction data at each wavelength were processed separately using DENZO (*17*). SCALEPACK (*17*) was used to scale the inverse and direct beam data for each wavelength together. The MAD data collection statistics are summarized in Tables 1 and 2.

MAD Structure Determination. The three wavelength diffraction data were analyzed using the MADSYS suite (*18*, *19*) to extract the anomalous scattering signals. A total of 21 398 reflections to 2.9 Å were phased in MADLSQ to give structure factors $|^\circ F_Z|$ and $|^\circ F_A|$ and the phase differences $\Delta\phi$, where $|^\circ F_Z|$ represents the normal (wavelength-invariant) contribution from all atoms, $|^\circ F_A|$ represents the anomalous scattering contribution, and $\Delta\phi$ is the phase difference between $|^\circ F_Z|$ and $|^\circ F_A|$. Multiple observations of $|^\circ F_Z|$, $|^\circ F_A|$, and $\Delta\phi$ for each unique reflection were merged in MERGIT, and 6361 unique reflections with weighted $|^\circ F_Z|$, $|^\circ F_A|$, and $\Delta\phi$ were output. The derived $|^\circ F_A|$ coefficients contain both dispersive and anomalous signals and were used to solve the anomalous scattering substructure.

The SnB program package (*20*) was used to solve the anomalous scattering substructure with derived F_A data. SnB implements the Shake-and-Bake procedure which is based on minimal function defined by Hauptman (*21*−*23*). A total of 2304 F_A values [with $F_A > 5\ \sigma(F_A)$] in the resolution range of 20−4 Å were processed with the DREAR program package (*24*, *25*) to generate normalized *E* values. The 500 largest *E* values were used to generate 5000 triples. Three passes of two-step parameter shift (90° per shift) were performed for each of 500 reflections used in phase refinement. Phase refinement was alternated with real space filtering to achieve the lowest constrained minimal function value (R_{min}). The distribution of solutions as a function of R_{min} started to appear as bimodal after the first 50 random trials had been completed. A total of 500 random trials were processed, and 95 of these gave the correct positions for all 14 Se atoms (Figure 1).

The SnB solutions were cross-validated by the direct methods program SHELXS (*26*) and the automated Patterson search program SOLVE (*27*). SHELXS found 12 Se atoms with the same F_A values, and SOLVE located 13 Se atoms. The single peak wavelength anomalous data were also used to solve the Se structure (*28*), but in this case it was less successful than using derived F_A values. In this case, only 1

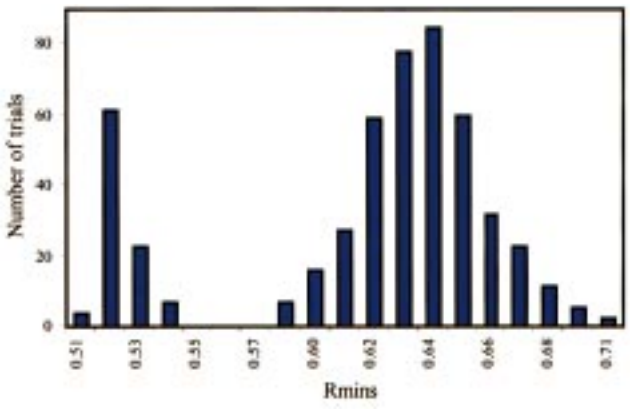


FIGURE 1: Distribution of SnB solutions as a function of R_{min} values. The distribution is bimodal, indicating that some of the randomly generated phases converged to the correct values. The correct solutions have R_{min} values between 0.51 and 0.54.

in 500 trials produced a correct solution, and only 11 of the Se were correctly identified.

Phase Refinement and Electron Density Calculation. The coordinates, occupancies, and temperature factors of the Se atoms were refined in ASLSQ using 2304 derived F_A data between 20 and 4 Å, resulting in $R = 28.9\%$. The experimental phases were calculated to 2.9 Å with the MADABCD program in the MADSYS suite. The mean figure-of-merit (FOM) was 0.526 for 8833 reflections in the resolution range between 20 and 2.9 Å. These reflections were used to calculate the experimental electron density maps. Electron density maps were calculated assuming either space group $P4_12_12$ or the enantiomeric space group $P4_32_12$. Only the map calculated in space group $P4_32_12$ showed a clear protein−solvent boundary, thus confirming $P4_32_12$ to be the correct space group assignment.

Electron density modification, including symmetry averaging, was performed with the DM program (*14*). The solvent content was assumed to be 35%. The noncrystallographic symmetry (NCS) operator was first calculated in O (*29*) using the coordinates of 14 Se atoms and then refined with IMP by real-space density correlation in the RAVE package (*30*). At this stage, the electron density clearly showed two β-barrel structures, and a protein mask for averaging was generated from the coordinates of triosephosphate isomerase (Protein Data Bank code 1BTM) superimposed onto the electron density map. The 2-fold NCS averaging, solvent flattening, and histogram matching greatly improved the experimental phases. The *R*-factor was 34.1% for the 8833 reflections between 20 and 2.9 Å, and the FOM increased from 0.526 to 0.819.

Model Building and Structure Refinement. The initial model, consisting of most of the residues between Arg19 and Phe224, was built in a 2.9 Å resolution electron density map using the computer graphics program O. At this stage, residues 1−18, 61−64, and 155−164 and 10 C-terminal residues were left unmodeled due to the poor electron density in those regions. Several amino acid residues including Lys38, Arg83, Glu84, Asp94, Leu99, Asp112, Lys116, and Arg176 were truncated to alanine. The refinement procedure consisted of a series of rigid-body refinements, simulated annealing refinements, and temperature factor refinement in X-PLOR (*31*) followed by conjugate-gradient least-squares refinement in SHELXL (*32*). Each round of refinement was

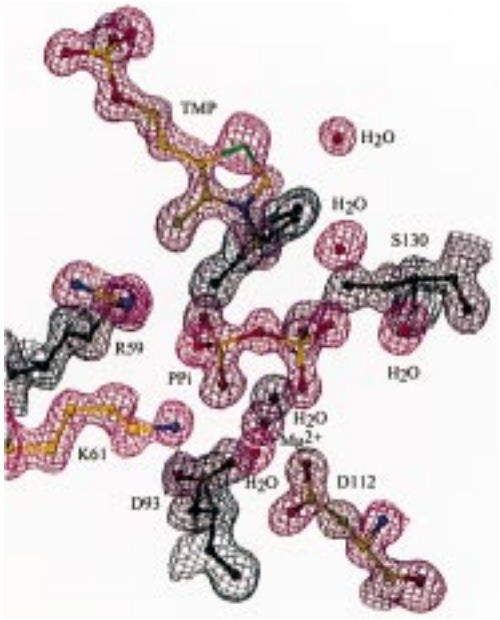


FIGURE 2: $2F_o - F_c$ electron density map around the region of pyrophosphate at the active site. The electron density map was calculated using data from 8 to 1.25 Å resolution and contoured at 2.5σ. The plot was prepared using program O (*29*).

followed by model rebuilding in O using an SGI graphics workstation. An NCS restraint of 300 kcal·mol^{-1}·Å^{-2} for main chain atoms and 100 kcal·mol^{-1}·Å^{-2} for side chain atoms was used in the X-PLOR refinements. The final refined NCS operator is defined by the spherical polar angles $\psi = 86.1°$, $\phi = 90.2°$, and $\kappa = 178.8°$. After the first round of X-PLOR refinement with 2.9 Å resolution data, the electron density of residues 225−233 was clearly revealed in an $F_o - F_c$ electron density map. In the subsequent rounds, high-resolution data were gradually included in the refinement. As the refinement proceeded, new electron density corresponding to the residues missing from the initial model was revealed. Several stretches of residues were modeled, including residues 61−64, 155−164, and 234−235. Individual atomic *B*-factors were refined when the 1.5 Å data were used, resulting in a decrease in both *R* and R_{free} (5.7% and 8.8%, respectively). The model was greatly improved, and the density for the most disordered region, residues 9−18, was clearly identified. At this stage, electron density corresponding to the enzymatic reaction products, thiamin phosphate, pyrophosphate, and Mg^{2+}, was clearly visible (Figure 2). A total of 120 well-ordered water molecules were also added, resulting in a 4% drop in both R_{free} (30.5%→26.1%) and R (27.5%→23.4%). In the latter X-PLOR refinement rounds, additional water molecules were included. Any water molecule with a high *B*-factor (≥45 Å^2) or lacking a geometrically reasonable H-bond was excluded. Two conformations of His132 were also assigned. The final *R* factor from X-PLOR refinement was 18.7% (R_{free} 21.9%). The low R_{free} value reflects the use of very high-resolution data and NCS restraints.

When the X-PLOR refinement had converged, the atomic model of TP synthase was further refined using the SHELXL package with geometric restraints tabulated by Engh and Huber (*33*). SHELXL was used for the purposes of restrained anisotropic *B*-factor refinement, refinement of disordered side chains, addition of hydrogen atoms in calculated positions,

Table 3: Refinement Statistics

Final Model	
resolution range (Å)	8−1.25
reflections used	94169
restraints used	18840
refined parameters	18915
R factor (%)[a,b]	16.8 (18.1)
R_{free} (%)[a]	20.7
Structure Description	
no. of residues (per asu)[c]	452
no. of protein atoms (per asu)[c]	3468
no. of water molecules (per asu)[c]	444
no. of TP (per asu)[c]	2
no. of PP_i (per asu)[c]	2
no. of Mg^{2+} (per asu)[c]	2
residues in Ramachandran plot (%)[d]	
most favored/additional/disallowed[e]	93.1/6.4/0.5
average *B*-factors (Å^2)	
main chain/side chain/water	12.7/18.3/27.2
TP/PP_i/Mg^{2+}	10.7/10.3/9.5
rmsd deviations from ideal values	
bond length (Å)	0.012
angle (deg)	2.344
dihedrals (deg)	23.886
impropers (deg)	2.560
rmsd NCS-related Cα (Å)	0.330

[a] $R = \sum||F_o| - |F_c||/\sum|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes; R_{free} is defined as *R* for 5% of the data omitted from the refinement. [b] *R* factor for 79 991 reflections with $F_o > 4\sigma(F_o)$; the number in parentheses is the *R* factor for all reflections. [c] asu = asymmetric unit. [d] Most favored, additional allowed, and disallowed regions are defined by PROCHECK. [e] Asp93 is the only residue in the disallowed region.

and calculation of estimated standard deviations (esd's) for the refined coordinates. The SHELXL refinement did not include NCS restraints. TP synthase was refined with the conjugate gradient least-squares option in SHELXL using 94 169 F^2 data in the resolution range from 8 to 1.25 Å and 18 840 geometrical restraints. Amino acid residues with an average *B*-factor less than 10 Å^2 were selected for aniostropic refinement. These residues were Leu25−Gly33, Ala53−Arg59, Pro88−Asn92, and Gly204−Ser212 in molecule A and Leu25−Gly33, Thr54−Gln57, Lys77−Arg83, Val87−Asp93, Ala103−Gly109, Met125−Ser130, Ala145−Leu150, Pro182−Val184, and Ala200−Val205 in molecule B. In total, 628 atoms were refined anisotropically. This reduced both *R* and R_{free} by 0.4% (*R*, 18.5%→18.1%; R_{free}, 21.8%→21.4%). A total of nine disordered amino acid side chains were refined with multiple occupancies. The occupancies of the enzymatic reaction products, thiamin phosphate, pyrophosphate, and Mg^{2+}, were also refined. Finally, hydrogen atoms were added in their calculated positions, resulting in decreases in both *R* (17.8%→16.7%) and R_{free} (21.3%→20.7%). In the final conjugate gradient least-squares refinement, all 94 169 reflections were used, and the final *R* factor was 16.8%. The SHELXL refinement statistics are summarized in Table 3.

RESULTS AND DISCUSSION

Quality of the Final Model. The final model includes 2 molecules of TP synthase (molecules A and B), each complexed with enzymatic reaction products (thiamin phosphate, pyrophosphate, and Mg^{2+}) and a total of 444 water molecules. Molecule A included residues 10−235, and molecule B consisted of residues 9−234. The occupancies of thiamin phosphate, pyrophosphate, and Mg^{2+} were estimated to be 0.95, 0.74, and 0.72, respectively, in molecule A, and 0.74, 0.60, and 0.72, respectively, in molecule B. A total of nine residues showed disordered conformations. These are Met21, Glu69, His132, Ser206, and Ser219 in molecule A and Arg13, Ser135, Ser206, and Ser219 in molecule B. There is one cis-peptide bond between Gly151 and Pro152 in both molecules. The final model was assessed using the programs PROCHECK (*34*) and O (*29*). The Ramachandran plot (*35*) showed that 93.1% of the main chain dihedral angles lie in the most favorable region and 6.4% are found in the additional favored region. Asp93, a ligand to the Mg^{2+} ion, is the only outlier. The average *B*-factor is 12.7 Å^2 for main chain atoms, 18.3 Å^2 for side chain atoms, and 27.2 Å^2 for water molecules. The *B*-factors of the TP atoms range from 7.9 to 13.0 Å^2 in molecule A, and from 8.2 to 15.6 Å^2 in molecule B. The *B*-factors of the pyrophosphate atoms range from 8.6 to 10.5 Å^2 in molecule A, and from 8.1 to 12.6 Å^2 in molecule B. The *B*-factors of Mg^{2+} are 8.5 and 10.4 Å^2 in molecules A and B, respectively. The average *B*-factor of the residues in the β-strands and α-helices are 9.6 and 17.2 Å^2, respectively. The real-space fit of the final model was calculated using the RS_FIT command in program O. An analysis of real-space fits of the final model showed that most of the residues have good correlation to the $2F_o - F_c$ electron density map. The average correlation coefficients are 0.92 for main chain atoms and 0.84 for side chain atoms. The results also indicated some flexible regions of the structures with correlation coefficients less than 0.7. These regions included residues 62−64, 84, 111, 117, 160−161, and 215 in molecule A and residues 63, 157, 160−161, and 234 in molecule B.

The coordinate error was estimated to be 0.11 Å from a Luzzati plot (*36*) based on the conventional *R*-factor. The plot was calculated using reflections with $F_o > 4\sigma$ (F_o) between 8 and 1.25 Å resolution with the program SHELX-PRO (SHELX97 Computer Program Package). We also estimated the positional esd's from the inverse of the least-squares matrix using the program SHELXL. The geometrical restraints were released and the parameters were divided into three blocks, with some overlap of parameters between consecutive blocks. Three cycles of least-squares refinement were performed in which a single block of parameters was refined for each cycle. The positional esd's for the TP atoms of molecule A range from 0.014 to 0.065 Å and from 0.021 to 0.110 Å for molecule B, where the positional esd's for the phosphorus and sulfur atoms are 0.014 and 0.018 Å, respectively (0.021 and 0.026 Å, respectively, for molecule B). The positional esd's for the pyrophosphate atoms of molecule A range from 0.017 to 0.042 Å and from 0.023 to 0.061 Å for molecule B, where the two phosphorus atoms have values of 0.017 and 0.020 Å, respectively (0.023 and 0.027 Å, respectively, for molecule B). The positional esd for the Mg^{2+} is 0.020 Å in molecule A and 0.024 Å for molecule B. The positional esd's for the key active site atoms [Gln57(Oε1), Gln57(Nε2), Arg59(Nε), Arg59(NH_2), Lys61-(Nζ), Asn92(Nδ2), Asp93(Oδ1), Asp112(Oδ2), Ser130(Oγ), Thr156(Oγ1), Thr158(Oγ1), Lys159(Nζ), Gly188(N), Ile208-(N), Ser209(N), Ser209(Oγ), and nine water molecules] ranged from 0.027 to 0.074 Å for molecule A and from 0.029 to 0.107 Å for molecule B.

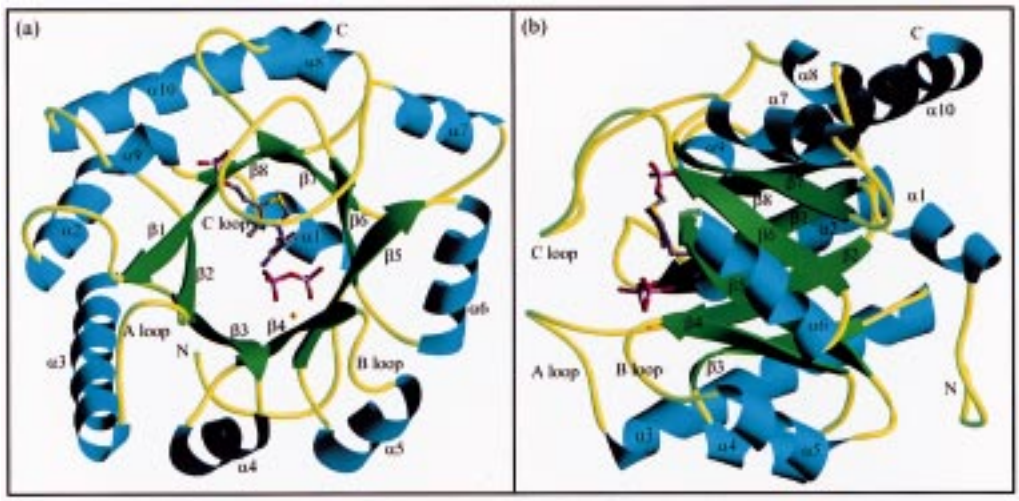


FIGURE 3: Ribbon representations of the TP synthase. The enzymatic products, thiamin phosphate and pyrophosphate, as well as a magnesium ion are shown in ball-and-stick representations with oxygen atoms in red, carbon in gray, nitrogen in blue, sulfur in yellow, phosphorus in purple, and magnesium in gold. The active site loops (A, B, and C loops) and all the secondary structure elements are labeled. (a) A view looking down the barrel. (b) A side view of the structure (i.e., rotated 90° along the vertical axis with respect to view a). Figures were created using the program RIBBONS (*37*).

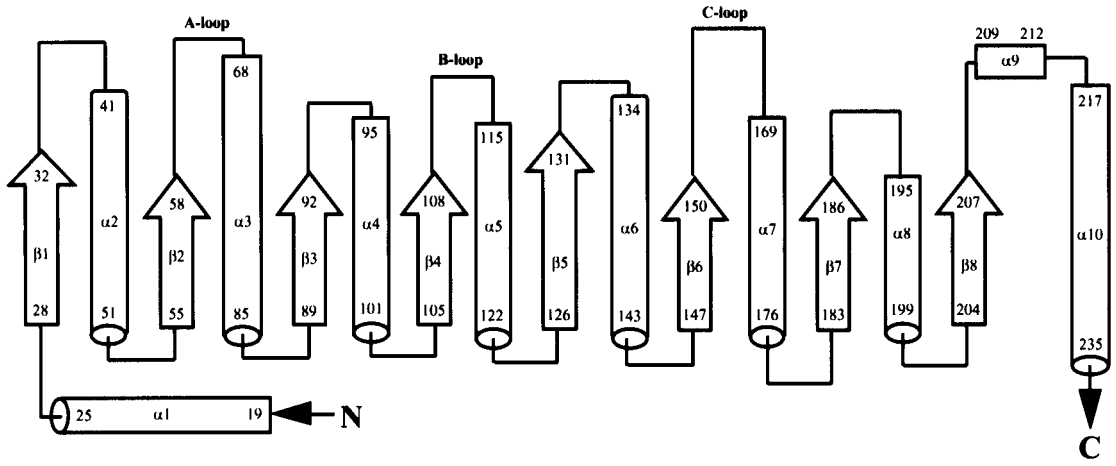


FIGURE 4: Topology diagram of TP synthase.

Overall Structure. TP synthase is an α/*β* protein with a triosephosphate isomerase fold (TIM barrel). The overall dimensions are approximately 45 × 45 × 35 Å. TP synthase has a typical TIM barrel structure with a core of eight helices (α2−α8, α10) and eight parallel *β*-strands tilted approximately 45° from the axis of the barrel as shown in Figure 3. In addition, there are two additional α-helices, α1 and α9. Helix α1 runs across the N-terminal barrel entrance, whereas the active site is located at the C-terminal barrel entrance. Helix α9 consists of only one turn and is inserted between *β*8 and α10; it is tilted approximately 20° from the axis of the TIM barrel helices. A topology diagram is shown in Figure 4. Most of the *β*-strands consist of four amino acids residues except for *β*1 and b5 which are five and six amino acids long, respectively. The lengths of the helices range from 5 to 19 amino acid residues with α8 and α9 being the shortest and α10 the longest. The second longest helix, α3, is at the interface between the 2-fold-related TP synthase molecules. TP synthase contains a mostly hydrophobic core formed by residues from the *β*-strands that point into the interior of the barrel. The side chains of these residues are divided into two layers. Residues in the top layer, which include several polar residues (Gln57, Asn92, His107, and Ser130) and Ile186, interact with thiamin phosphate and pyrophosphate through electrostatic, hydrogen bonding, and van der Waals contacts. Residues in the bottom layer including Tyr29, Ile55, Ile90, Tyr147, and Val184 form the hydrophobic floor of the active site. Tyr29 and Tyr147 provide van der Waals contacts with thiamin phosphate.

Dimer Interactions. Two symmetry-related α3 helices, one each from molecules A and B, are aligned parallel to each other along the dimer interface. In addition, residues 9−14 from the N-terminal tail and residues 99−104 from the loop between α4 and *β*4 are arranged together. Residues 9−14 are part of the His_6-affinity tag, and they provide additional contacts between the two monomers. There are a number of electrostatic and van der Waals interactions among these contacts. These include salt bridges between Asp104A(Oδ1) and Lys102B(Nζ), Lys102A(Nζ) and Asp104B(Oδ1); Arg83A(NH_2) and Leu101B(O), and Leu101A(O) and Arg83B(NH_2). Three water molecules provide hydrogen bonding bridges between the two monomers [Lys102A(Nζ)−H_2O−Lys102B(O), Lys102A(Nζ)−H_2O−Arg83B(NH_2), Asn100A(O)−H_2O−His10B(O)). The *B*-factors of these water molecules range from 14.60 to 18.0 $Å^2$.

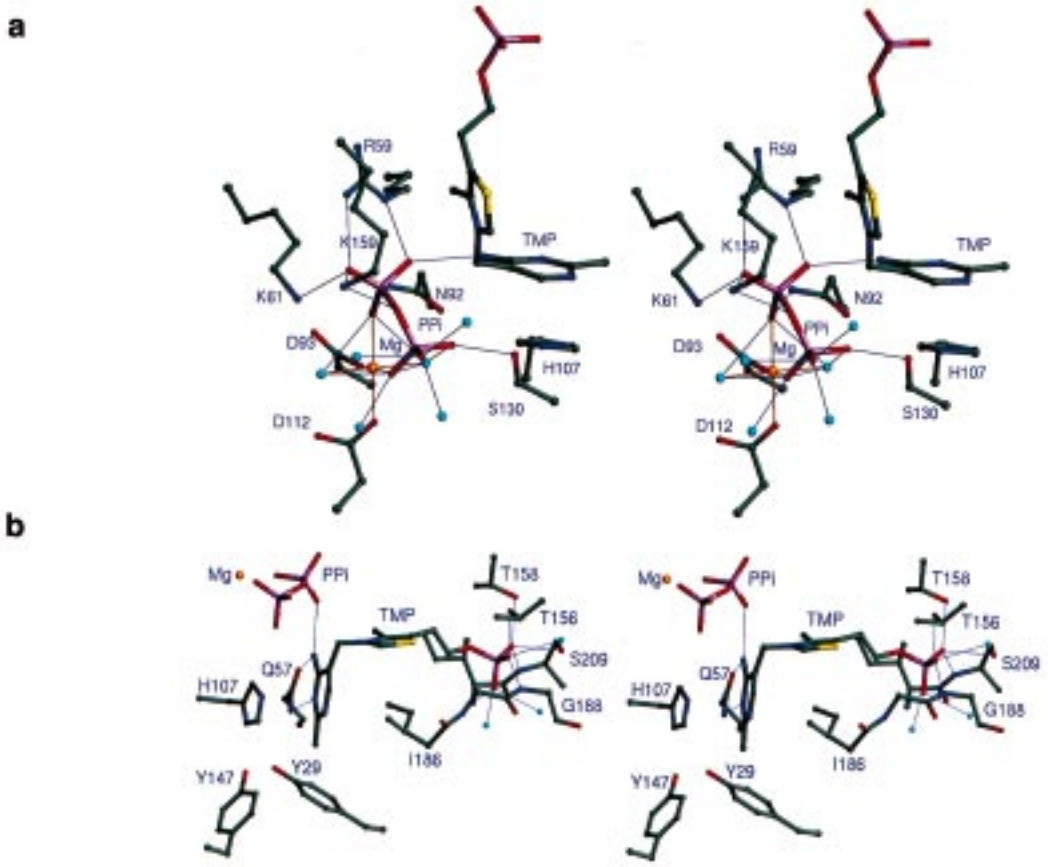


FIGURE 5: Binding site of (a) pyrophosphate and (b) thiamin phosphate in the TP synthase structure. Hydrogen bonds are indicated by blue lines, and the magnesium coordination spheres are shown as red lines. The water molecules are shown as blue spheres, and the others are colored according to the scheme described in the legend of Figure 3. The figure was produced with RIBBONS (*37*).

Crystal Packing. The two molecules (A and B) in the asymmetric unit lie side-by-side in the *ab* crystallographic plane with their barrels approximately perpendicular to the noncrystallographic 2-fold axis (approximately parallel to the crystal *c* axis). The dimers in the unit cell are arranged as four layers along the crystallographic *c* axis. In each layer, the dimers are oriented in the same direction (at 90° to the neighboring layers). Each dimer is in contact with a total of 12 adjacent dimers, 6 within the same layer and 3 from each of the layers above and below. There are two different types of packing between the dimers in the same layer. The first type is between the top and bottom of adjacent α/β barrels. Arg235 from α10 in one dimer is in close proximity to Thr67 and Arg71 from α3 in an adjacent dimer, and there are also a number of van der Waals and electrostatic interactions. The second type involves the van der Waals and electrostatic interactions between residues 164−169 (part of the active site C loop) in one dimer and equivalent residues in a symmetry-related dimer. There are also two different types of packing between adjacent layers. In the first type of packing, dimers are packed along the 2_1 axis (along *c*), where four α3 helices (one from each molecule in two adjacent dimers) are packed together, and four α5 helices from two adjacent dimers are also in close contact. The second type of packing involves the electrostatic interactions between Lys38, Gln213, and Gln215 from one dimer with Arg175, Arg176, and Gln177 from an adjacent dimer.

Active Site. The active site is located in a pocket formed primarily by loop regions 59−67 (A loop, joining α3 and β2), 109−114 (B loop, joining α5 and β4), and 151−168 (C loop, joining α7 and β6). The active site is covered by the C loop, the longest loop in the structure. The C loop acts as the front door to the active site. When the reaction is complete, the door is opened to release reaction products and to allow substrates to come in. In the crystal structure, the enzymatic reaction products, thiamin phosphate, pyrophosphate, and magnesium ion, are tightly bound to the enzyme. The refined occupancy of the reaction products in molecule A is slightly higher than in molecule B. In molecule A, the occupancies for thiamin phosphate and pyrophosphate are 0.95 and 0.74, respectively, while in molecule B the corresponding values are 0.74 and 0.60. We were surprised to find thiamin phosphate and pyrophosphate bound at the active site because the enzyme was purified in the absence of substrates and products. In fact, k_{cat} for the reaction (0.9 min^{-1}) indicates that the rate of product release is very fast compared to the purification time (~24 h). We do not yet understand this fortuitous effect. One possibility is that product release is triggered by the availability of reactants or some other unknown factor and therefore products remain bound after the final turnover.

Pyrophosphate Binding Site. The pyrophosphate is located near the three active site loops and close to the opening of the active site. The pyrophosphate is stabilized by extensive electrostatic and hydrogen bonding interactions (Figures 5a and 6a, Table 4). It has electrostatic interactions with Arg59, Lys61 from the A loop, Lys159 from the C loop, and a magnesium ion. In addition, the pyrophosphate oxygen atoms are hydrogen bonded to Ser130(Oγ), Asn(Nδ2), and six water molecules. The average *B*-factor of these six water

Table 4: Hydrogen Bonds and Electrostatic Interactions Stabilizing the Pyrophosphate and Thiamin Phosphate

		distance (Å)	
	residue (atom)	molecule A	molecule B
	Pyrophosphate		
O1A	W3027/W3029	2.74	2.70
	W3062/W3149	2.68	2.79
	W3108/W3202	2.76	2.81
O2A	Asp112 (Oδ2)	2.98	3.07
	W3020/W3022	2.92	2.86
	W3051/W3057	3.01	3.05
	W3098/W3081	2.70	2.78
O3A	Ser130 (Oγ)	2.50	2.69
	W3020/W3022	2.80	2.80
O4	Lys159 (Nζ)	3.07	3.10
O1B	Lys159 (Nζ)	2.86	2.78
	Lys61 (Nζ)	2.99	3.03
	Arg59 (NH2)	2.91	2.86
O2B	TP (N4′)	2.90	2.95
	Arg59 (Nϵ)	2.77	2.79
O3B	Asn92 (Nδ2)	3.04	3.01
	Asp93 (Oδ1)	3.11	3.02
	W3020/W3022	2.91	2.84
	W3051/W3057	3.01	3.01
	Thiamin Phosphate		
N3′	Gln57 (Nϵ2)	3.03	3.04
N4′	Gln57 (Oϵ1)	2.98	2.93
	PP_i (OB2)	2.90	2.95
O1	Gly188 (N)	2.90	2.95
	Thr156 (Oγ1)	2.69	2.74
	W3063/W3169	2.90	3.00
O2	Thr158 (Oγ1)	2.60	2.69
	Ser209 (Oγ)	2.81	2.80
	Ser209 (N)	2.99	3.09
O3	Ile208 (N)	2.93	2.91
	W3001/W3003	2.75	2.74
	W3002/W3013	2.79	2.75

molecules is 12.2 Å^2, ranging from 13.7 to 20.7 Å^2 (13.8−24.8 Å^2 for molecule B). The O(3A) of the pyrophosphate is located at a distance of 3.46 Å (3.43 Å in molecule B) away from the C7′ of the pyrimidine and is hydrogen bonded to S130(Oγ). The pyrophosphate and the thiazole moiety are positioned on opposite sides of the pyrimidine base. Atoms O(3A), C7′, and N3 are approximately collinear and located perpendicular to the plane of the pyrimidine ring. These observations suggest that the covalent bond broken during the chemical reaction was between O(3A) and C7′. Atom O(2B) of pyrophosphate is also hydrogen bonded to the C4′ amino group of thiamin phosphate with a distance of 2.90 Å (2.95 Å in molecule B).

Magnesium Binding Site. The magnesium ion is located on the edge of the active site, and is solvent-accessible (Figures 5a and 6a, Table 4). The magnesium ion is octahedrally coordinated by two aspartate side chains (Asp93, Asp112) via their Oδ atoms, two oxygen atoms from pyrophosphate, O(2A) and O(3B), which acts as a bidentate ligand, and two water molecules. The corresponding bond lengths are 2.02(0.03), 2.15(0.03), 2.09(0.03), 2.05(0.03), 2.08(0.03), and 2.10(0.03) Å, respectively. The numbers in the parentheses are the esd's from the SHELXL refinement, and the distances were averaged for the two molecules. The *B*-factors of the two water molecules are 13.7 and 16.9 Å^2, respectively. These two water molecules are also hydrogen bonded to O(2A), O(3A) and O(3B) of pyrophosphate. There is a salt bridge between Asp93(Oϵ2) and Lys61(Nζ).

Thiamin Phosphate Binding Site. The thiamin phosphate binds in a "V" conformation (Figures 5b and 6b, Table 4). This conformation is maintained primarily by van der Waals interaction between Ile186 and the pyrimidine and thiazolium rings. The "V" conformation corresponds to angles of ϕ_P and ϕ_T of −93° (−95° for molecule B) and −108° (−112° for molecule B), respectively, and has not been previously observed in any thiamin structure. The pyrimidine moiety is located at the entrance to the hydrophobic core of the barrel, in a hydrophobic pocket formed by Tyr29, His107, Tyr147, Ile186, and Ile208. It is hydrogen bonded to Gln57 (Oϵ and Nϵ) through its N3′ and N4′, packing against the imidazole ring of His107, and has van der Waals contacts with Tyr29 and Tyr147. The phosphate moiety of TP is stabilized by hydrogen bonding interactions with Thr156, Thr158, Gly188, Ile208, Ser209, and three water molecules. The two most ordered water molecules (*B*-factors 8.9 and 9.3 Å^2) in the entire structure are involved in these interactions. The *B*-factor for the third water is 17.8 Å^2. The thiazole phosphate moiety is bound in an extended conformation, lying under the active site C loop. The side chain of Lys159 runs parallel to the thiazolium ethyl group, providing van der Waals contacts. The phosphate is bound at the N-terminus of α9, forming interactions with Ile208(N) and Ser209(N and Oγ), and also contacts the C loop through hydrogen bonds to Thr156(Oγ), Thr158(Oγ). In addition, hydrogen bonding interactions form with Gly188(N) and three water molecules.

Comparison of Thiamin Binding with TPP Binding Proteins. The crystal structures of four thiamin pyrophosphate (TPP)-dependent enzymes have been determined by X-ray crystallography: transketolase (*38*, *39*), pyruvate decarboxylase (*40*, *41*), pyruvate oxidase (*42*, *43*), and benzoylformate decarboxylase (*44*). The first three of these enzymes catalyze key reactions in central metabolic pathways. The structural information provides insights on the mode of thiamin pyrophosphate binding and activation. A comparison of the crystal structures reveals that despite the lack of significant sequence similarity, the fold of the domains involved in TPP binding and the conformation of TPP in these enzymes are quite similar (*45*). The pyrimidine ring is buried in a hydrophobic pocket formed by adjacent domains. Although the general hydrophobic character of the binding site is conserved, no identities or positional conservation of these hydrophobic residues is conserved. For example, in transketolase Phe445 (His107 in TP synthase) packs against the pyrimidine ring, while no positional conservation was observed in any of the other TPP binding proteins. The N3′ and N4′ of the pyrimidine are hydrogen bonded to the protein. N3′ is hydrogen bonded to the main chain, either Leu(N) or Ile(N), and N4′ is hydrogen bonded to either Gly(O) or Ala(O) in these TPP-dependent proteins. The environment of thiazole is mostly nonpolar. In all four enzymes, the conformation of TPP is remarkably similar. TPP is bound in the "V" conformation, and the pyrimidine C4′ amino group and the thiazole methyl group are "trans". In all TPP-dependent enzymes, a bulky hydrophobic residue (Leu118 in transketolase, Ile415 in pyruvate decarboxylase, Met422 in pyruvate oxidase, and Leu403 in benzoylformate decarboxylase) intercalates between the pyrimidine and the thiazole rings of TPP. This interaction constrains the TPP in the "V" conformation and forces N4′ of the pyrimidine and C-2 of the thiazole into close proximity. One of the few conserved interactions is the hydrogen bonding between glutamate and N1′, which can tautomerize the pyrimidine

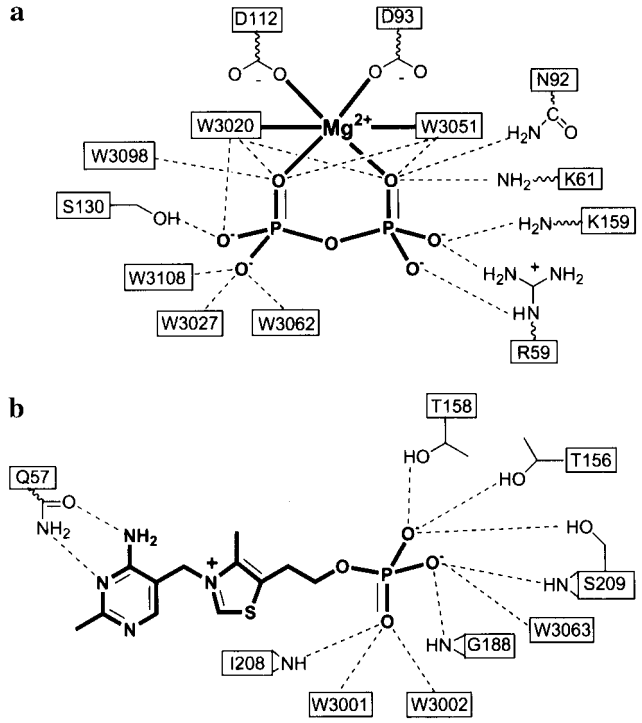


FIGURE 6: Schematic representations of the interactions between the active site residues and (a) pyrophosphate and (b) thiamin phosphate.

to produce a 4′-imino group, which functions as the base to deprotonate C-2.

In TP synthase, the thiamin phosphate is also bound in the "V" conformation, and Ile186 is intercalated between the pyrimidine and the thiazole rings. However, the C4′-amino group of the pyrimidine and the methyl group of the thiazole are "cis", and there is no acidic residue hydrogen bonded to N1′. This conformation is unique to the TP complex and reflects the fact that thiamin does not need to be activated as a cofactor at the active site of the synthase. There is no conservation of identity or position of the hydrophobic residues around the pyrimidine with the TPP-dependent enzymes. The N3′ and N4′ of the pyrimidine are hydrogen bonded to Glu57(Oϵ1 and Nϵ2). This is in contrast to TPP binding proteins in which N3′ and N4′ are bound primarily through main chain interactions. Finally, in the TPP binding proteins, the cofactor is bound at a dimer interface. TP synthase is the first example of a thiamin binding protein that utilizes a single structural domain for thiamin binding. The recently published structure of thiaminase-I from *Bacillus thiaminolyticus* (*46*) also uses a single protein subunit; however, the active site is located between two α/β domains.

Sequencing Similarity. A sequence similarity search of *B. subtilis* thiamin phosphate synthase with the BLAST network server (*47*) identified 11 additional thiamin phosphate synthase sequences and demonstrated that there are significant amino acid identities among all 12 sequences (Figure 7). All of the key active site residues identified from the *B. subtilis* structure are highly conserved, suggesting that these enzymes all have a similar active site structure and probably use the same catalytic mechanism. In addition, the amino acid residues in the β-strands pointing into the α-helices are dominated by Val, Leu, and Ile which is characteristic of the TIM barrel structure (*60*). These residues are also highly conserved, suggesting that all of the thiamin phosphate synthases identified here are α/β proteins with the triose-phosphate isomerase fold.

The *Bacillus subtilis* TenI regulates the production of several extracellular enzymes (*50*). There is significant sequence similarity between TenI and TP synthase. Most of the active site residues are conserved except the magnesium ligands, Asp93 and Asp112, and the sequence order of Glu57 and His107 is exchanged in TenI. The sequence similarity between TenI and TP synthase suggests that TenI is closely related to TP synthase and is likely to be a thiamin binding protein. One possibility is that TenI is involved in the regulation of thiamin biosynthesis.

Proposed Catalytic Mechanism. We have considered three possible mechanisms for the thiamin phosphate synthase catalyzed reaction, an S_N1 mechanism, an S_N2 mechanism, and an addition elimination mechanism (Figure 8). The addition elimination mechanism, which is precedented in the thiaminase I catalyzed displacement of the thiazole moiety of thiamin by a variety of nucleophiles, requires the addition of an active site nucleophile to C6′ of the pyrimidine (*51*, *52*). The crystal structure clearly demonstrates that such a nucleophile is not present in TP synthase, ruling out this mechanistic possibility.

In the crystal structure, the pyrophosphate and the thiazole moieties are on opposite sides of the pyrimidine, demonstrating that the substitution reaction must occur with inversion of stereochemistry. This result, however, does not differentiate between the S_N1 and the S_N2 mechanisms because the stereochemistry of the reaction is controlled by the active site structure. The pyrophosphate moiety is highly stabilized at the active site by interactions with the protein that are similar to those found in farnesyl pyrophosphate synthase (*53*), *epi*-aristolochene synthase (*54*), and pentalenene synthase (*55*). Each of these enzymes catalyzes the formation of an allylic carbocation from a pyrophosphate ester.

This suggests that the thiamin phosphate synthase catalyzed reaction may also proceed by an S_N1 mechanism. This hypothesis is supported by the facile formation of carbocations from HMP analogues. For example, we encountered great difficulty in carrying out stereocontrolled substitution at C7′ of the pyrimidine in our synthesis of chiral monodeuteriothiamin due to competing S_N1 ionization (*56*). In addition, our attempts to carry out a positional isotope exchange experiment on TP synthase failed because of facile nonenzymatic positional isotope exchange of HMP-PP during purification (*57*).

Two additional experiments have been carried out to test the hypothesis that thiamin phosphate synthase proceeds by an S_N1 mechanism. For an S_N1 mechanism, the rate of the TP synthase catalyzed reaction will be very sensitive to any changes that affect the stability of the carbocation or the pyrophosphate leaving group because the transition state is carbocation-like. In contrast, for an S_N2 mechanism, the reaction rate will be much less sensitive to these changes because the transition state will have much less carbocation character than the S_N1 transition state.

In the crystal structure, the O(3A) of the pyrophosphate is located at a distance of 3.46 Å (3.43 Å in molecule B) from the C7′ of the pyrimidine, and Ser130 is hydrogen bonded directly to the O(3A) atom. This suggests that this oxygen was originally bonded to the pyrimidine and that Ser130 activates the pyrophosphate as a leaving group.

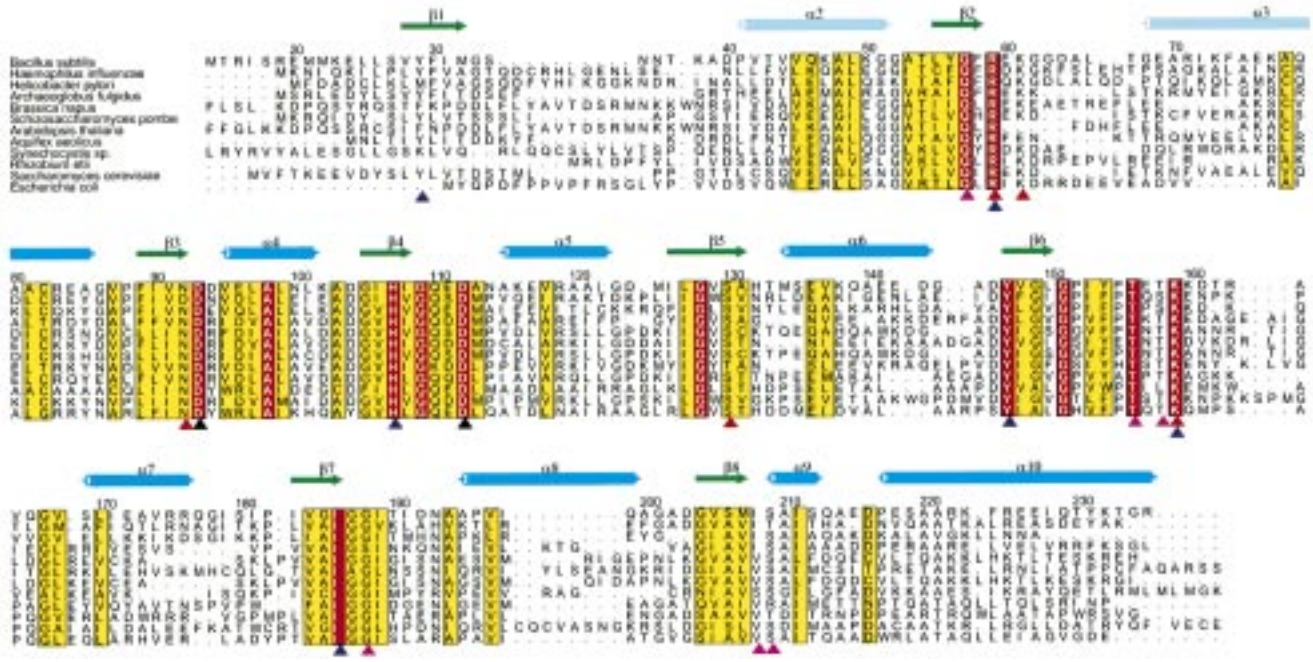


FIGURE 7: Alignment of the amino acid sequences of thiamin phosphate synthase. A sequence similarity search of *Bacillus subtilis* thiamin phosphate synthase with the BLAST server (http://www.ncbi.nlm.nih.gov/BLAST) shows that there are significant amino acid identities among 12 thiamin phosphate synthase sequences: *Helicobacter pylori* (AE000595, 41%), *Archaeoglobus fulgidus* (AE000960, 38%), *Brassica napus* (AF015310, 38%), *Haemophilus influenzae* (P71350, 37%), *Arabidopsis thaliana* (AF000657, 36%), *Rhizobium etli* (AF004408, 35%), *Schizosaccharomyces pombe* (P40386, 34%), *Aquifex aeolicus* (AE000695, 34%), *Synechocystis* PCC6803 (P72965, 34%), *Saccharomyces cerevisiae* (P41835, 32%), *Escherichia coli* (P30137, 31%). The numbers in the parentheses are accession numbers and sequence identity, respectively. The sequences of thiamin phosphate synthase were aligned using the GCG package (*48*) and plotted using the program ALSCRIPT (*49*). The sequence numbering follows the sequence of the *B. subtilis* enzyme. For the secondary structure, cylinders represent α-helices, and arrows represent *β*-strands. Residues identical in all sequences are highlighted in red and similar residues are highlighted in yellow. Residues hydrogen bonded to pyrophosphate and thiamin phosphate are shown as red and purple triangles, respectively. Residues providing hydrophobic interactions with the pyrimidine ring are indicated as blue triangles. The ligands of the Mg^{2+} are shown as black triangles.

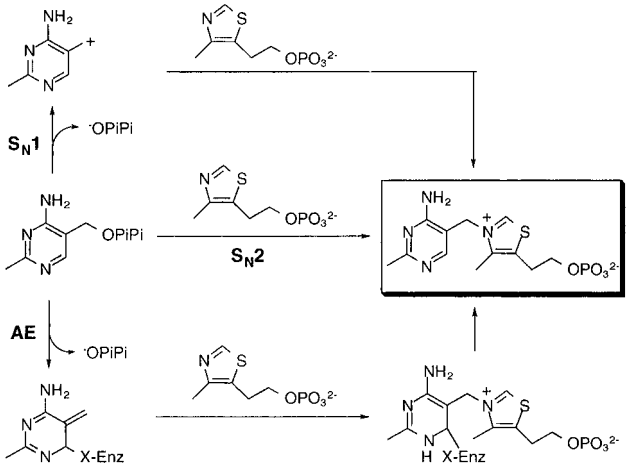


FIGURE 8: Three possible mechanisms for TP synthase (AE: addition−elimination).

Removal of this interaction provides us with a way to decrease the stability of the pyrophosphate and should have a large effect on the reaction rate for an S_N1 mechanism and a small effect on the reaction rate if the reaction follows the S_N2 pathway. We have found that the S130A mutant is inactive within the detection limits of the assay ($>10^4$-fold decrease). The structure of the S130A mutant is very similar to the native enzyme except that the reaction products are not bound (*58*). This suggests that loss of activity is not due to misfolding of the S130A mutant, but due to a true chemical effect resulting from the loss of a hydrogen bond between S130 and pyrophosphate.

The stability of the pyrimidine carbocation can be greatly reduced by replacing the electron-donating methyl group at the C2′ position with the trifluoromethyl group. The methyl/trifluoromethyl substituent effect has previously been used to probe for a carbocation intermediate in the farnesyl pyrophosphate synthase catalyzed reaction where the k_{cat} is reduced 2 800 000-fold by this substituent change (*59*). We have found that the trifluoromethyl analogue of HMP-PP is at best a poor substrate for the enzyme [$k_{cat}(CH_3) > 1000k_{cat}(CF_3)$] (*57*). The structure of the S130A mutant with the trifluoromethyl pyrimidine bound at the active site shows that the trifluormethyl substituent does not prevent this substrate analogue from binding to the active site (*58*).

Both of these observations demonstrate that the rate of the enzymatic reaction is highly sensitive to interactions that affect the stability of the products of an S_N1 ionization of HMP-PP and suggest that the enzymatic reaction proceeds via an S_N1 mechanism.

In the crystal structure, the terminal oxygen of pyrophosphate is located 2.90 Å (2.95 Å in the second molecule) from the C4′ amino group. This hydrogen bonding interaction could stabilize a pyrimidine carbocation intermediate because proton transfer from the carbocation would generate a relatively stable neutral imine.

In conclusion, the high-resolution structure of thiamin phosphate synthase complexed with its reaction products described here provides a detailed picture of the interactions between the enzyme and the substrates that is important for catalysis. The structure and other mechanistic studies are

consistent with a reaction mechanism involving the ionization of HMP-PP at the active site to give the pyrimidine carbocation. This ionization is catalyzed by orienting the C−O bond undergoing cleavage perpendicular to the plane of the pyrimidine, by hydrogen bonding between the C4′ amino group and one of the terminal oxygen atoms of the pyrophosphate, and by extensive hydrogen bonding and ionic interactions between the pyrophosphate and the enzyme. Trapping of the carbocation by the thiazole followed by product dissociation completes the reaction.

ACKNOWLEDGMENT

We thank Ulrike Breitinger for determining the initial crystallization conditions for TP synthase. We thank Neil Kelleher for performing the ESI/FT-MS experiments for both native and SeMet TP synthase.

REFERENCES

1. Burdick, D. (1998) Kirk-Othmer *Encycl. Chem. Technol. 25*, 152−171.
2. Schowen, R. L. (1998) *Thiamin-dependent enzymes. Comprehensive Biological Catalysis* (Sinnott, M., Ed.) Vol. II, pp 217−266, Academic Press, San Diego.
3. Begley, T. P., Downs, D., Ealick, S., McLafferty, F., Loon, D., Taylor, S., Campobasso, N., Chiu, H.-J., Kinsland, C., Reddick, J., and Xi, Jun. (1998) *Arch. Microbiol.* (in press).
4. Begley, T. P. (1996) *Nat. Prod. Rep. 13*, 177−185.
5. Estramareix, B., and David, S. (1996) *New J. Chem. 20*, 607−629.
6. Spenser, I. D., and White, R. L. (1996) *Angew. Chem., Int. Ed. Engl. 36*, 1032−1046.
7. Vander Horn, P., Backstrom, A. D., Stewart, V., and Begley, T. P. (1993) *J. Bacteriol. 175*, 982−992.
8. Backstrom, A. D., McMordie, A., and Begley, T. P. (1995) *J. Am. Chem. Soc. 117*, 2351−2352.
9. Zhang, Y., Taylor, S. V., Chiu, H.-J., and Begley, T. P. (1997) *J. Bacteriol. 179*, 3030−3035.
10. Nosaka, K., Nishimura, H., Kawasaki, Y., Tsujihara, T., and Iwashima, A. (1994) *J. Biol. Chem. 269*, 30510−30516.
11. Reddick, J. J., Kinsland, C., Nicewonger, R., Christian, T., Downs, D. M., Wrinkler, M. E., and Begley, T. P. (1998) *Tetrahedron 54*, 15983−15991.
12. Wyatt, D. T., Lee, M., and Hillman, R. E. (1989) *Clin. Chem. 35*, 2173−2178.
13. Leslie, A. G. W. (1990) *Molecular data processing in Crystallographic computing 5 (from Chemistry to Biology)* (Moras, D., Podjarny, A. D., and Thierry, and J. C., Eds.) pp 50−61, Oxford University Press, Oxford, U.K.
14. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
15. Mathews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
16. Tate, M. W., Eikenberry, E. F., Barna, S. L., Wall, M. E., Lowrance, J. L., and Gruner, S. M. (1995) *J. Appl. Crystallogr. 28*, 196−205.
17. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
18. Hendrickson, W. A., Smith, J., Phizackerley, R. P., and Merritt, E. A. (1988) *Proteins: Struct., Funct., Genet. 4*, 77−88.
19. Hendrickson, W. A. (1991) *Science 254*, 51−58.
20. Miller, R., Gallo, S. M., Khalak, H. G., and Weeks, C. M. (1994) *J. Appl. Crystallogr. 27*, 613−621.
21. Debaerdemaeker, T., and Woolfson, M. M. (1983) *Acta Crystallogr. A39*, 193−196.
22. Hauptman, H. A. (1988) *American Crystallography Association Meeting* Abstract R4, Philadelphia, PA.
23. Hauptman, H. A. (1991) *Crystallographic computing 5: from Chemistry to Biology* (Moras, D., Podnarny, A. D., and Thierry, J. C., Eds.) pp 324−332, Oxford University Press, Oxford, U.K.
24. Blessing, R. H. (1989) *J. Appl. Crystallogr. 22*, 396−397.
25. Blessing, R. H., Guo, D. Y., and Langs, D. A. (1997) in *Direct Methods for Solving Macromolecular Structures. NATO ASI Series* (Fortier, S., Ed.) Springer, New York.
26. Sheldrick, G. M. (1997) *Methods Enzymol. 276*, 628−641.
27. Terwilliger, T. C., and Berendezen, J. (1997) *Acta Crystallogr. D53*, 571−579.
28. Smith, G. D., Nagar, B., Rini, J. M., Hauptman, H. A., and Blessing, R. H. (1998) *Acta Crystallogr. D54*, 799−804.
29. Jones, T. A., Zhou, J.-Y. Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
30. Kleywegt, G. J., and Jones, T. A. (1997) *Acta Crystallogr. D53*, 179−185.
31. Brünger, A. T. (1996) *X-PLOR, Version 3.843, a System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
32. Sheldrick, G. M., and Schneider, T. R. (1997) *Methods Enzymol. 277*, 319−343.
33. Engh, R. A., and Huber, R. (1991) *Acta Crystallogr. A47*, 392−400.
34. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thorton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
35. Ramachandran, C., and Ramakrishnan, G. N. (1965) *Biophys. J. 5*, 909−933.
36. Luzzati, P. V. (1952) *Acta Crystallogr. 5*, 802−810.
37. Carson, M. (1991) *J. Appl. Crystallogr. 24*, 958−961.
38. Lindqvist, Y., Schneider, G., Ermler, U., and Sundström, M. (1992) *EMBO J. 11*, 2373−2379.
39. Nikkola, M., Lindqvist, Y., and Schneider, G. (1994) *J. Mol. Biol. 238*, 387−404.
40. Arjunan, P., Umland, T., Dyda, F., Swaminathan, S., Furey, W., Sax, M., Farrenkopf, B., Gao, Y., Zhang, D., and Jordan, F. (1996) *J. Mol. Biol. 256*, 590−600.
41. Dyda, F., Furey, W., Swaminathan, S., Sax, M., Farrenkopf, B., and Jordan, F. (1993) *Biochemistry 32*, 6165−6170.
42. Muller, Y. A., and Schulz, G. E. (1993) *Science 259*, 965−967.
43. Muller, Y. A., Schumacher, G., Rudolph, R., and Schulz, G. E. (1994) *J. Mol. Biol. 237*, 315−335.
44. Hasson, M. S., Muscate, A., McLeish, M. J., Polovnikova, L. S., Gerlt, J. A., Kenyon, G. L., Petsko, G. A., and Ringe, D. (1998) *Biochemistry 37*, 9918−9930.
45. Muller, Y. A., Lindqvist, Y., Furey, W., Schulz, G., Jordan, F., and Schneider, G. (1993) *Structure 1*, 95−103.
46. Campobasso, N., Costello, C. A., Kinsland, C., Begley, T. P., and Ealick, S. E. (1998) *Biochemistry 37*, 15981−15989.
47. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403−410.
48. Genetics Computer Group (1991) *Program Manual for the GCG Package*, Version 7.
49. Barton, G. J. (1993) *Protein Eng. 6*, 37−40.
50. Pang, A. S.-H., Nathoo, S., and Wong, S.-L. (1991) *J. Bacteriol, 173*, 46−54.
51. Costello, C. A., Kelleher, N. L., Abe, M., McLafferty, F. W., and Begley, T. P. (1996) *J. Biol. Chem. 271*, 3445−3452.
52. Campobasso, N., Costello, C. A., Kinsland, C. C., Begley, T. P., and Ealick, S. E. (1998) *Biochemistry 37*, 15981−15989.
53. Tarshis, L. C., Proteau, P. J., Kellogg, B. A., Sacchettini, J. C., and Poulter, C. D. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 15018−15023.
54. Starks, C. M., Back, K., Chappell, J., and Noel, J. P. (1997) *Science 277*, 1815−1820.
55. Lesburg, C. A., Zhai, G., Cane, D. E., and Christianson, D. W. (1997) *Science 277*, 1820−1824.
56. Nicewonger, R., Rammelsberg, A., Costello, C. A., and Begley, T. P. (1995) *Bioorg. Chem. 23*, 512−518.
57. Nicewonger, R. (1997) Ph.D. Thesis, Cornell University.
58. Chiu, H.-J., Reddick, J. J., Begley, T. P., and Ealick, S. E., unpublished results.
59. Poulter, C. D., and Satterwhite, D. M. (1977) *Biochemistry 16*, 5470−5478.

BI982903Z