Crystal Structure of Thiaminase-I from *Bacillus thiaminolyticus* at 2.0 Å Resolution[†,‡]

Nino Campobasso, Colleen A. Costello, Cynthia Kinsland, Tadhg P. Begley, and Steven E. Ealick*

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

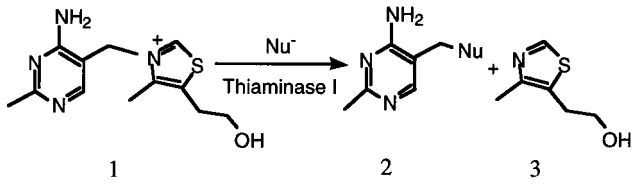
Received July 14, 1998; Revised Manuscript Received September 9, 1998

ABSTRACT: Thiaminase-I catalyzes the replacement of the thiazole moiety of thiamin with a wide variety of nucleophiles, such as pyridine, aniline, catechols, quinoline, and cysteine. The crystal structure of the enzyme from *Bacillus thiaminolyticus* was determined at 2.5 Å resolution by multiple isomorphous replacement and refined to an *R* factor of 0.195 (R_{free} = 0.272). Two other structures, one native and one containing a covalently bound inhibitor, were determined at 2.0 Å resolution by molecular replacement from a second crystal form and were refined to *R* factors of 0.205 and 0.217 (R_{free} = 0.255 and 0.263), respectively. The overall structure contains two α/β-type domains separated by a large cleft. At the base of the cleft lies Cys113, previously identified as a key active site nucleophile. The structure with a covalently bound thiamin analogue, which functions as a mechanism-based inactivating agent, confirms the location of the active site. Glu241 appears to function as an active site base to increase the nucleophilicity of Cys113. The mutant Glu241Gln was made and shows no activity. Thiaminase-I shows no sequence identity to other proteins in the sequence databases, but the three-dimensional structure shows very high structural homology to the periplasmic binding proteins and the transferrins.

Thiaminases catalyze the degradation of thiamin (vitamin B_1). This activity has been detected in bacteria, marine organisms, and plants (*1*−*4*). Although the biological function of these enzymes is unknown, it has been demonstrated that ingestion of thiaminase-containing foods by animals and humans may result in symptoms of thiamin deficiency (*3, 5*−*7*). Two thiamin-degrading enzymes have been isolated. Thiaminase-I (EC 2.5.1.2) catalyzes the reaction in which the thiazole group of thiamin is replaced by a variety of organic nucleophiles, for example, aniline, quinoline, pyridine, cysteine, dithiothreitol, and veratrylamine (Scheme 1) (*8, 9*) to give 2 and the thiazole 3. Thiaminase-II is specific for the use of water as the nucleophile. Thiaminase-I has been purified from *Bacillus thiaminolyticus* (*10*) and is a better characterized system than thiaminase-II (*8, 11, 12*). Thiaminase-I is an extracellular, 42 kDa protein. Its gene has been cloned, sequenced, and overexpressed in *Escherichia coli* (*9, 11*). Electrospray ionization Fourier transform mass spectrometry showed that the recombinant enzyme was heterogeneous at the N-terminus due to signal sequence processing at three different sites to give three species of thiaminase-I with molecular weights of 42 127, 42 197, and 42 254 Da with the ratio of 1:2:1, respectively (*13*). Amino acid sequence analysis using BLAST (*14*) showed no significant sequence similarity of thiaminase-I to other known protein sequences.

The degradation of thiamin by thiaminase-I is proposed to proceed via a double addition−elimination mechanism.

Scheme 1



This mechanism is supported by the observation of ping-pong kinetics for the reaction (*8, 15*), by the demonstration that the substitution reaction using chiral monodeuterio thiamin proceeds with retention of stereochemistry (*16, 17*) and by the mechanism-based inactivation of the enzyme with 4-amino-6-chloro-2-methylpyrimidine (*18*). Electrospray ionization Fourier transform mass spectrometry of 4-amino-6-chloro-2-methylpyrimidine inactivated enzyme showed a mass increase of 108 Da, corresponding to the addition of one pyrimidine moiety with the loss of chloride (*9*). Proteolysis of the labeled enzyme followed by mass spectrometry analysis of the resulting unfractionated mixture of peptides resulted in the identification of Cys113 as the active site nucleophile (*19*). This was further supported by the demonstration that the Cys113Ser mutant of thiaminase-I was catalytically inactive (*9*).

Despite interest in this enzyme over a period of 50 years, little was known about the details of its mechanism until recently. We report here the three-dimensional structures of thiaminase-I from two crystal forms, I and II, and of form II with inhibitor, determined by using X-ray crystallographic methods to resolutions of 2.5, 2.0, and 2.0 Å, respectively. The structure of form I was determined at room temperature and shows anisotropic protein expansion and nonuniform increase of atomic mobility with respect to the second structure (form II), which was determined at liquid nitrogen

[†] This work was supported by National Institutes of Health grants to S.E.E. (RR01646) and to T.P.B. (DK44083). S.E.E. is indebted to the W. M. Keck Foundation and the Lucille P. Markey Charitable Trust.

[‡] The Brookhaven Protein Data Bank codes for thiaminase-I are 2thi, 3thi, and 4thi.

* Corresponding author. Telephone: (607)255-7961. Fax: (607)-255-4137. E-mail: see3@cornell.edu.

10.1021/bi981673l CCC: $15.00
Published on Web 10/27/1998

Table 1: Different Crystal Forms of Thiaminase-I

	a (Å)	*b* (Å)	*c* (Å)	Spgrp[a]	no./AU[b]	Vm (Å^3 Da^{-1})
form I	87.7	120.5	76.7	$P2_12_12_1$	2	2.40
form II	85.5	117.5	36.6	$P2_12_12$	1	2.18
complex	86.7	119.3	37.3	$P2_12_12$	1	2.29
form III	114.6	123.1	92.5	$P2_12_12_1$	3	2.56

[a] Space group. [b] Number of molecules per asymmetric unit.

temperature. The third structure (form II + inhibitor) was of the enzyme inactivated with 4-amino-6-chloro-2,5-dimethylpyrimidine, also solved from crystals that were frozen at liquid nitrogen temperature. These structures provide a basis for understanding substrate binding and catalysis by thiaminase-I.

EXPERIMENTAL PROCEDURES

Protein Purification and Crystallization and Complex Formation. Thiaminase-I from *B. thiaminolyticus* was overexpressed and purified from an *E. coli* expression system, as described previously (*9*), and crystallized in three space groups (Table 1) (*20*). A 20 L culture yielded 15 mg of pure protein. The stock solution of thiaminase-I, which was stored at −80 °C, was composed of 15 mg/mL protein, 50 mM sodium phosphate (pH = 7.0), 2 mM dithiothreitol (DTT), and 2 mM ethylenediamine tetraacetic acid (EDTA). Crystallization was performed by the hanging drop method over Linbro plates by mixing 2 μL of protein solution and 2 μL of the precipitating solution. Form I crystals grew as six-sided parallelepipeds to a maximum of 1.0 mm × 0.8 mm × 0.5 mm in 10−14 days at precipitating conditions of 0.1−0.2 M ammonium sulfate, 0.1 M sodium acetate (pH 4.6−5), and 26−32% PEG2000. The stabilization buffer for crystal storage, heavy-atom soaks, and/or crystal freezing contained 0.2 M ammonium sulfate, 0.1 M sodium acetate (pH = 4.6), and 30% PEG2000. Crystal form I diffracted to 2.5 Å resolution at room temperature and belongs to space group $P2_12_12_1$ with unit cell dimensions of $a = 87.7$ Å, $b = 120.5$ Å, and $c = 76.7$ Å and two molecules per asymmetric unit. The noncystallographic twofold was approximately parallel to the twofold screw axis along *c*. When the crystals were frozen at liquid nitrogen temperature for data collection at the Cornell High Energy Synchroton Source (CHESS),[1] a phase change occurred (form II), resulting in crystals with space group $P2_12_12$ and unit cell dimensions of $a = 85.5$ Å, $b = 117.5$ Å, and $c = 36.6$ Å and one molecule per asymmetric unit. The frozen crystals diffracted to 2.0 Å. A third crystal form (III) grew from 28% PEG6000, 0.2 M sodium acetate trihydrate, and 0.1 M Tris (pH = 8.5). The space group was $P2_12_12_1$ with cell parameters $a = 114.6$ Å, $b = 123.1$ Å, and $c = 92.5$ Å. These crystals diffracted to 2.7 Å resolution at room temperature. However, freezing or soaking inhibitors cause the crystals to crack and analysis of crystal form III was not pursued further.

Form I crystals were harvested and soaked in a solution containing 10 mM 4-amino-6-chloro-2,5-dimethylpyrimidine, 0.2 M ammonium sulfate, 0.1 M sodium acetate (pH = 4.6), and 30% PEG2000 for 3 days. Soaking experiments of less than 3 days did not show any inhibitor binding; however, soaking experiments longer than 4 days destroyed the crystals.

Table 2: X-ray Data Collection Statistics

	native I	native II	complex	Hg[a]	Au[a]	U[a]
soak						
conc (mM)				1.0	2.0	1.0
time (days)				1	1	1
no. of sites				2	4	2
temperature (°C)	22	−170	−170	22	22	22
R_{sym} (%)[b]	5.0	5.6	4.0	4.4	3.8	4.3
R_{iso} (%)[c]				14	29	28
redundancy	6	6.7	5	3.6	3.1	3.0
completeness (%)	83	89	90	93	87	89
resolution range (Å)	2.5	2.0	2.0	2.7	2.7	2.7
phasing power[d]				1.3	1.3	0.70
R−Cullis[e]				0.73	0.81	0.94

[a] Abbreviations were used for the following heavy-atom reagents: sodium ethylmercurithiosalicylate (Hg), sodium gold chloride (Au) from Pfaltz & Bauer, Inc., and uranyl acetate (U). [b] $\sum |I - \langle I\rangle|/\sum I$, where I is an individual measurent of a reflection and $\langle I\rangle$ is the mean of the symmetry-related reflections. [c] $\sum |F_N - F_{PH}|/{}^1/_2\sum |F_N + F_{PH}|$, where F_N is the native structure factor amplitude and F_{PH} is the derivative strucutre factor amplitude. [d] $[\sum F_H{}^2/\sum E^2]^{1/2}$, where F_H is heavy-atom model structure factor and E is the lack of closure error. [e] $\sum |E|/\sum |F_{PH} + F_N|$ for centric reflections.

X-ray Data Collection. Native data and isomorphous derivative data from crystal form I were collected at room temperature using a Xuong-Hamlin San Diego Multiwire (SDMW) duel detector system (*21*) mounted on a Rigaku RU-200B rotating anode source generating CuKα radiation and operating at 40 kV and 100 mA. Diffraction data were collected in 0.1° oscillations for 60 s/oscillation and integrated with the SDMW software package (*22*). The resulting integrated intensities were scaled and reduced with SCALA from the CCP4 suite (*23*). Native data for structure refinement were collected from one frozen thiaminase-I crystal (form II) that diffracted to 2.0 Å resolution at CHESS, beamline A-1, ($\lambda = 0.9104$ Å), using a 1k × 1k CCD detector (*24*). The CHESS diffraction data were collected in 1.0° oscillation frames for 20 s/frame. The CHESS data were processed with DENZO (*25*) and scaled with SCALEPACK (*25*). Data from a crystal soaked with 4-amino-6-chloro-2,5-dimethylpyrimidine were collected and processed in a like manner but with a different 1k × 1k CCD (Area Detector Systems Co., San Diego, CA). The processing statistics are summarized in Table 2.

Phasing and Phase Refinement. Three isomorphous derivatives were used for phasing: sodium ethylmercurithiosalicylate (thimersol), sodium gold chloride, and uranyl acetate. All derivatives were prepared by harvesting thiaminase-I crystals and soaking them in 1 mM respective heavy-atom compound, 100 mM sodium acetate (pH = 4.6), 0.2 M ammonium sulfate, and 30% PEG4000 for 16 h. A crystal was then mounted in a glass capillary, and data were collected using the SDMW system as described above. The thimersol derivative difference Patterson map was solved by inspection and showed two positions, related by noncrystallographic symmetry. The four gold and two uranyl positions were determined by difference Fourier maps and

[1] Abbreviations: rmsd, root-mean-square difference; NCS, noncrystallographic symmetry; CHESS, Cornell high-energy synchrotron source; SDMW, San Diego Multiwire; DTT, dithiothritol; EDTA, ethylenediaminetetraacetic acid.

Table 3: Heavy-Atom Sites[a]

heavy atom	site	X^a	Y^a	Z^a	occ[b]	B^c	neighbors[d]
Hg	1a	0.059	0.230	0.065	3.0	36	Cys A113
Hg	1b	0.412	0.730	0.065	3.0	24	Cys B113
Au	1a	0.638	0.134	0.226	5.2	26	His A320
Au	2b	0.903	0.862	0.227	3.8	17	His B320
Au	2a	0.058	0.160	0.038	4.4	34	Met A159
Au	2b	0.396	0.839	0.040	4.3	34	Met B159
U	1a	0.137	0.737	0.078	1.8	23	Glu A184
U	1b	0.339	0.263	0.073	1.6	15	Glu B184

[a] *X*, *Y*, and *Z* are fractional coordinates. [b] Occ is relative occupancy. [c] *B* is temperature factor. [d] Neighbors are those residues that have atoms within 3.5 Å of the heavy-atom site.

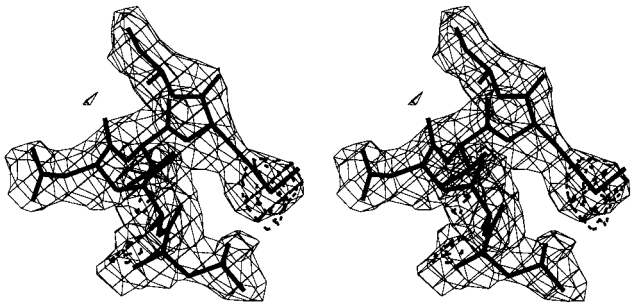


FIGURE 1: Native and anomalous electron density around residues 289−294 at 2.5 and 4.0 Å resolutions, respectively. The native electron density map (lines, 1.0σ) was calculated with F_{obs} as the coefficients and with phases after DM refinement. The anomalous electron density map (dots, 3σ) was calculated with [$F_{o(plus)} - F_{o(minus)}$] as the coefficients and with phase from DM refinement minus 90°. This figure was made with Chain (*47*).

also show noncrystallographic symmetry. All three heavy-atom derivative types bound to unique sites (Table 3). Heavy-atom parameter refinement and phase calculations were performed with the program MLPHARE (*26*). Phases were calculated to 2.7 Å resolution and improved and extended to 2.5 Å resolution with solvent flattening and histogram matching by using the program DM (*27*). An initial model was built to define the molecular envelope and the noncrystallographic symmetry operator (NCS). The NCS operator was refined by real-space density correlation using the program IMP from the RAVE package (*28*). Phase improvement and extension were repeated to include NCS averaging, solvent flattening, and histogram matching.

Model Building and Refinement. The program O (*29*) was used for displaying the electron density and building the thiaminase-I model. The resulting 2.5 Å electron density map clearly showed the topology of the protein. To aid in chain tracing, the 2.5 Å electron density map was skeletonized with RAVE (*28*). A native anomalous difference Fourier map identified sulfur atoms to further aid in chain tracing. As is shown in Figure 1, Met293 was easily identified, as were the other seven methionine residues. Of the 381 possible amino acids, 362 were built into the electron density. The missing amino acids occurred at the N- and C-termini. Poor electron density was seen at residues 260−264, but contouring at lower electron density revealed the backbone atoms for those residues. The second molecule in the asymmetric unit was generated by applying the NCS operator.

Refinement was done with the program XPLOR version 3.483 (*30*), restraining the model with Engh & Huber stereochemical parameters (*31*) and with tight NCS restraints. Data between 30 and 2.5 Å resolution were used for refinement of crystal form I. Data between 30 and 2.0 Å resolution were used for refinement of crystal form II. A third X-ray diffraction data set between 30 and 2.0 Å was used for refinement of crystal form II with the inhibitor. A bulk solvent correction was also applied throughout the refinements (*32*). During the refinements, maps with Fourier coefficients ($|F_o| - |F_c|$) and ($2|F_o| - |F_c|$) were used for model adjustments and for locating water molecules. After multiple cycles of simulated annealing, model building, and water addition in crystal form I, protein coordinates from one molecule were transformed into crystal form II for refinement to 2.0 Å resolution. Rigid body refinement was utilized to adjust the position of the thiaminase-I molecule in the form II cell. Further cycles of simulated annealing were continued to refine the model to 2.0 Å resolution and include water molecules. Restrained individual temperature factor refinement was done in both crystal form I and crystal form II.

The diffraction data with 4-amino-6-chloro-2,5-dimethylpyrimidine soaked into the crystals was nonisomorphous with crystal form II (Table 1). The protein atoms from the refined form II crystal were transformed into the new cell (form II with inhibitor). After rigid body refinement in the new cell and multiple rounds of simulated annealing and model adjustment, a positive peak (3.0 times the rmsd) was seen in a difference Fourier map that resembled 4-amino-6-chloro-2,5-dimethylpyrimidine covalently attached to the SG of Cys113 by the C6 atom. To confirm the structure of the environment around the Cys113 residue, simulated annealing omit map experiments were carried out by removing Cys113 and surrounding atoms within 5 Å.

The Cys113 residue was adjusted, and after further refinement, water molecules and restrained individual temperature factors were added. Throughout the entire refinement process, water molecules were conservatively added only if the ($|F_o| - |F_c|$) difference electron density was above 3.5σ and the hydrogen-bonding distance was between 2.5 and 3.4 Å.

Site-Directed Mutagenesis on Thiaminase-I (*Glu241Gln*). The thiaminase-I mutant was generated according to the method of Vandeyar (*9*), using the T7-Gen In Vitro mutagenesis system (USB, Cleveland, OH) with the exception that *E. coli* SURE cells (Stratagene, La Jolla, CA) were transformed by the mutagenesis reaction mixture. Single-stranded DNA was prepared from *E. coli* JM101 cells (Stratagene) transformed by a M13mp18 clone containing the 755 bp *Sal*I/*Hin*dIII fragment from pCAC303 which served as a template for mutagenesis. For the Glu241Gln mutation, the following oligonucleotide was used: 5′-GC CCA TGC GCA TCA TCG ATT GGC TGT AGC CGA TGA ACG C-3′. Mutant clones were confirmed by DNA sequencing. The *Sal*I/*Hin*dIII fragment containing the mutation was then moved to pCAC206 (*Sal*I/*Hin*dIII) to yield a representative plasmid, pCLK001.

RESULTS

Overall Structure of Thiaminase-I. The 2.5 Å resolution form I, the 2.0 Å resolution form II, and the 2.0 Å resolution form II complex refined crystal structures of thiaminase-I

Table 4: Refinement Statistics

	form I	complex	form II
resolution range (Å)	30−2.5	30−2.0	30−2.0
no. of reflection ($F/s > 2$)	23430	23256	21929
R-factor	0.195	0.217	0.205
R-free[a]	0.272	0.263	0.256
rmsd from ideal bond distances (Å)	0.007	0.009	0.008
rmsd from ideal bond angles (deg)	1.27	1.42	1.35
no. of non-hydrogen protein atoms	5718	2860	2860
rmsd for NCS molecules	0.22	N/A	N/A
no. of water	38	129	126
no. of sulfate ions[b]	1	1/2	1/2
mean *B*-factor (Å^2)	27	17	17

[a] The *R*-free refers to a random sample comprising 5% of all reflections not included in refinement. [b] The sulfate ion is along the *c*-axis in the form II and the complex structures.

have excellent quantitative parameters used to judge the quality of refined models (Table 4). The asymmetric unit of crystal form I contains two copies of thiaminase-I related by noncrystallographic symmetry, whereas crystal form II and the form II complex structures contain only one copy of thiaminase-I per asymmetric unit. The current models consist of 5720 non-hydrogen protein atoms, 38 water molecules, and one sulfate ion in form I, 2860 non-hydrogen protein atoms, 126 water molecules, and one sulfate ion in form II, and 2860 non-hydrogen protein atoms, 9 non-hydrogen inhibitor atoms, 129 water molecules, and one sulfate ion in the form II complex.

The structure of thiaminase-I is roughly ellipsoidial with dimensions 70 Å × 42 Å × 35 Å. Thiaminase-I is monomeric and has two distinct globular domains (identified as N- and C- domains), separated by a deep groove (Figure 2). Each of these domains has an α/β toplogy as shown in Figure 3. The N-domain and C-domain classifications are in accordance with those used for several structures of binding proteins with specificities for arabinose/galactose/fucose, glucose/galactose, sulfate, and maltodextrins (*15*). As will be discussed below, thiaminase-I shows very high structural similiarity to these proteins. Each domain is composed of amino acids from both the amino-terminal and the carboxy-terminal halves of the protein.

N-Domain Structure. The N-domain consists of residues 9−113 and 271−340. The C-domain consists of residues 115−268 and 352−370. Both domains are of similar size and have similar folds. Each domain has a central β-pleated sheet, flanked on both sides by α-helices (Figure 2). The β-strands are designated as capital letters in the order of appearance along the polypeptide chain, and the α-helices are designated as Roman numerials in a like manner. Overall, thiaminase-I is an α/β-protein with 45% of the amino acids in α-helix and 17% in β-sheet. The remaining 38% of the structure is comprised of loops and coils.

The N-domain contains seven α-helices and a six-stranded mixed β-sheet that has the topology BACIDJ (−1x, +2x, +2x, −1, +2) with strand I and J being antiparallel to the other four strands. The sheet is highly twisted forming what appears to be half a β-barrel. Of the seven α-helices associated with the N-domain, two (II and III) abut one face of the β-sheet, and five other α-helices (I, XI, XII, XIII, and XIV) flank the other side.

C-Domain Structure. The C-domain contains eight α-helices and a four-stranded, mixed β-sheet that has the topology HEGF (−2, −1x, +2x, +2x) with E being antiparallel to the other strands. Of the eight α-helices associated with the C-domain, five (IV, V, IX, X, and VIII) interact with the β-sheet, and three α-helices (VI, VII, and XV) cluster to form a subdomain. Helices IX and X are parallel to each other and abut the face of the C-domain β-sheet. Helices IV and V are antiparallel to each other and flank the other side of the β-sheet. The last α-helix (VIII) associated with the C-domain β-sheet lies at an edge of the β-sheet and packs alongside of strand F. The three other α-helices (VI, VII, and XV) cluster away from the β-sheet and form a subdomain.

The N-domain and the C-domain are connected by three peptide segments consisting of residues 114−115, 264−268, and 340−352. The first segment connects strand D of the N-domain to strand E of the C-domain. The second segment precedes strand I of the N-domain. The third segment, residues 340−352, packs along the interface of both domains. These three segments and the domain interface form a deep cleft.

Active Site. A cleft, approximately 15 Å deep, 12 Å wide, and 15 Å long, is located between the N and C-domains. The deep cleft is bounded by residues 16−20, 48−50, and 64−66 in the N-domain and residues 158−164, 213−224, and 238−241 in the C-domain. The interdomain connecting fragments 112−114 and 268−272 are at the base of the cleft. In the N-domain, residues 16−20 make a loop that connects β_A to α_I. Residues 48−50 make a loop that connects β_B to a coil structure that precedes β_C, and residues 64−66 make a loop that connects β_C to α_{II}. In the C-domain, residues 158−164 make a loop that connects β_F to α_{VI}. Residues 213−224 make a loop that connects α_{VIII} and α_{IX}, and residues 238−241 make a loop that connects β_G to α_X. Six tyrosine residues line the cleft along with four acidic residues. Six tyrosine residues, Tyr16, Tyr18, Tyr50, Tyr222, Tyr239, and Tyr270, form a collar of residues within the cleft. Three of the acidic residues, Asp64, Glu241, and Asp272, are located deep within the cleft and are less solvent accessible (Figure 4). The fourth acidic residue, Glu48, is at the surface. An electrostatic potential map shows a concentrated negative charge deep within the cleft; no positively charged residues are within the cleft. On the basis of the Lee and Richards (*33*) criteria, Asp64, Glu241, Asp272 have solvent accessible surface areas of 2.4, 6.3, and 4.9 Å^2, respectively (*34*). The tyrosine residues, Tyr16, Tyr18, Tyr50, Tyr222, and Tyr239, are more solvent accessible with 16.8, 107.9, 61.2, 35.6, 73.2, and 23.3 Å^2, respectively. At the bottom of the cleft is Cys113 with a solvent accessible area of 10.2 Å^2. This cysteine, the only cysteine in thiaminase-I, is proposed to be the active site nucleophile (*9, 19*). Glu241 is also at the bottom of the cleft with the carboxylate group 4.0 Å away from the sulfur atom of Cys113. Despite the low solvent accessiblity, a few ordered water molecules are seen solvating residues in the cleft.

Soaking of the mechanism-based inactivating agent, 4-amino-6-chloro-2,5-dimethylpyrimidine (*18, 19*), into the crystals of thiaminase-I resulted in covalent attachment to Cys113 (Figures 4). The resulting difference electron density map showed two peaks for the SG atom. One peak was at the 7σ level and showed additional density for the covalent adduct. The other peak was at the 3σ level with no covalent adduct. Difference electron density in the vicinity of Cys113

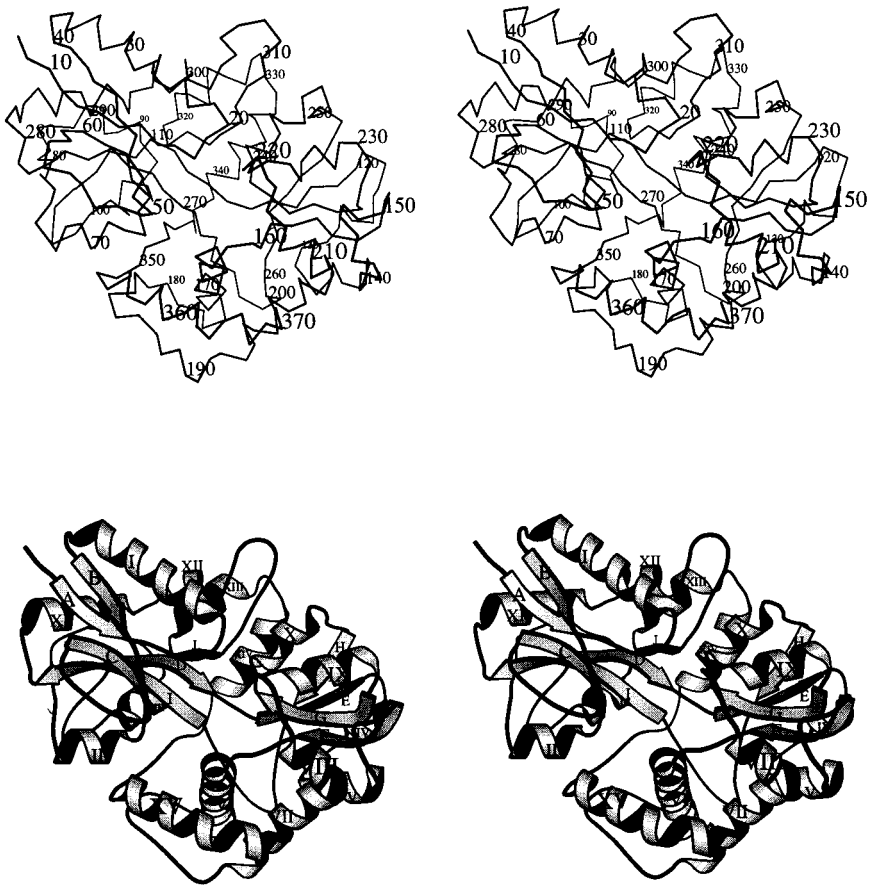


FIGURE 2: Stereo picture of the tertiary structure of thiaminase-I. The C_a backbone is traced with every 10th residue numbered. A ribbon drawing is in the same orientation as the Cα trace and with β-strands depicted as flat arrows and α-helices depicted as coils. The drawings were done with MOLSCRIPT (*48*).

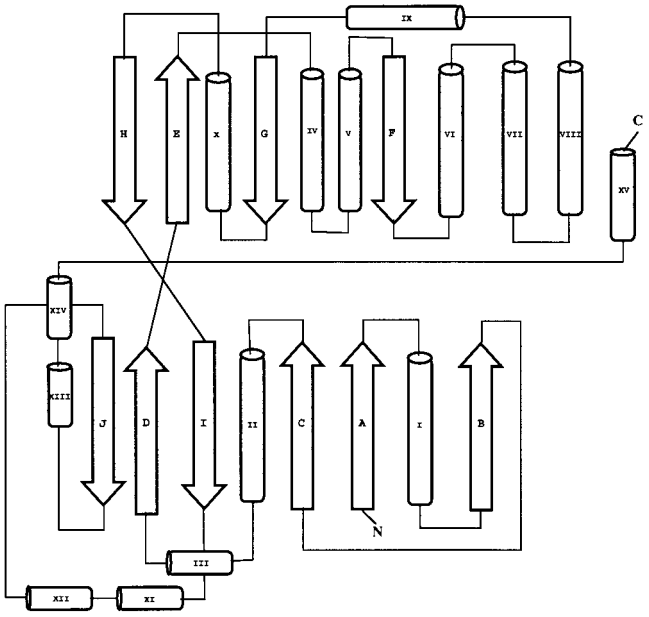


FIGURE 3: Topology of thiaminase-I. The β-strands are represented by arrows and the α-helices by cylinders.

showed positive density that resembled 4-amino-6-chloro-2,5-dimethylpyrimidine covalently attached to the sulfur atom of Cys113 via the C6 atom and a second conformation of Cys113 without any covalent adduct. The second conformation of Cys113 is identical to the conformation of Cys113 in the models without any inhibitor soaked into the crystal.

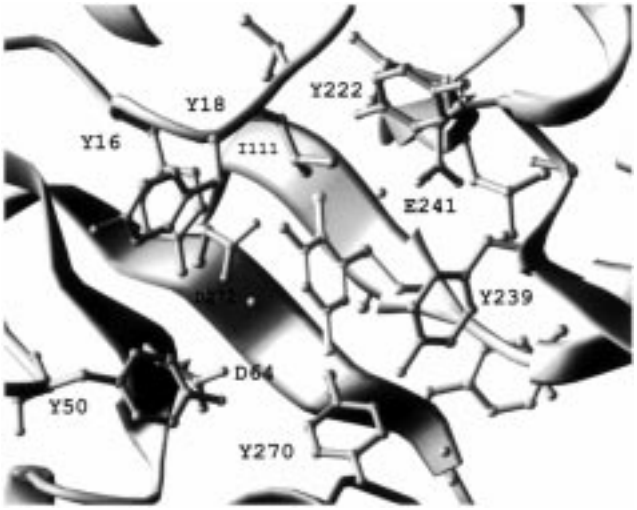


FIGURE 4: Active site structure of thiaminase-I with covalently bound inhibitor. This drawing was made with Ribbons (*49*).

With the pyrimidine moiety attached, $\chi_1 = -56°$ for Cys113, whereas in the structure without the inhibitor, $\chi_1 = 172°$.

On the basis of the Lee and Richards (*33, 34*) criteria, the solvent accessible area of 296 Å^2 of the free inhibitor was reduced to 31 Å^2 when bound to the active site. The bound pyrimidine was found to be buried deep within the cleft and held in place by a covalent bond to Cys113, one hydrogen bond between the carbonyl oxygen of Tyr18 and the 4-amino group of the pyrimidine, and van der Waals contact with a

Table 5: Structurally Similar Proteins[a]

	Z-score	rms	lali	lseq2	% id	PDB_id
D-maltodextrin-binding protein[b]						
(open form)	22	4.0	316	370	15	1omp
(closed form)		3.6	316	370	15	1mbp
domain 1 only		2.2	117	370		
domain 2 only		3.4	88	370		
spermidine/putrescine binding protein	20.6	4.3	297	232	12	1poy-2
sulfate-binding protein	12.9	4.7	254	309	13	1sbp
phosphate-binding protein	6.0	4.5	194	321	9	1quk
porphobilinogen deaminase	4.3	4.6	169	296	13	1pda
lys-, arg-, ornithine-binding protein	4.1	5.3	179	239	11	1lst
lactoferrin	3.1	5.9	155	691	7	1lcf
glucose/galactose-binding protein	2.7	4.2	126	309	13	1gca

[a] **Z-score** is the strength of structural similarity in standard deviations above expected as reported by the Dali server (http://www.embl-heidelberg.de/dali/dali.html). **rms** is the positional root-mean-square deviation of superimposed Ca atoms in Angstroms. **lali** is the total number of equivalenced residues. **lseq2** is the length of the entire chain of the equivalenced structure. **% id** is the percentage of sequence identity over equivalanced positions, and **PDB_id** is the Brookhaven protein data bank identifier. [b] The comparison of thiaminase-I to the closed form, domain 1, and domain 2 of D-maltodextrin-binding protein was done with the lsq_exp command in O (Jones, T. A).

water molecule and the hydroxyl group of residue Tyr270. Interestingly, the water molecule appears to be interacting with the π-electrons of the pyrimidine ring. The water molecule is also hydrogen bonded to Asp64 and Asp272.

Glu241Gln Mutant. The Glu241Gln mutant was purified in a way similar to that of the native recombinant enzyme. The mutation was confirmed by DNA sequencing, and the activity was measured as described previously (*9*). The Glu241Gln mutant was inactive, demonstrating that Glu241 in thiaminase-I is a critical active site residue. Interestingly, the expression level of this inactive mutant is not any better than that of the active form. Therefore, the low level of overexpression was not due to any toxic effects of thiaminase-I activity and may be due to the codon usage of the thiaminase-I gene.

DISCUSSION

Nonisomorphous Structures and Crystal Packing. Upon freezing, crystals of thiaminase-I undergo a phase change that alters the unit cell demensions and space group. At room temperature, crystal form I has unit cell dimensions of a = 87.7 Å, b = 120.5 Å, and c = 76.7 Å and belongs to spacegroup $P2_12_12_1$ with two molecules per asymmetric unit. At liquid nitrogen temperature, crystal form II of thiaminase-I has unit cell dimensions of a = 85.5 Å, b = 117.5 Å, and c = 36.6 Å and belongs to space group $P2_12_12$ with one molecule per asymmetric unit. As expected, the average temperature factor for all backbone atoms in form I is higher (24 Å^2) than the value for crystal form II (18 Å^2). However, the temperature factors do not decrease uniformly throughout the entire molecule. Instead, the average *B*-factor for the C-domain decreased by 9 Å^2, whereas the average *B*-factor for the N-domain decreased by only 3 Å^2. The inhibitor bound structure, which was determined at low temperature, showed behavior similar to that of the unbound form II structure.

At room temperature, the two molecules in the asymmetric unit are related by a noncrystallographic twofold axis and are bridged by a sulfate ion. The sulfate ion lies along the noncrystallographic twofold axis in the form-I crystal structure and is located 1.4 Å away from the crystallographic twofold screw axis along *c*. At liquid nitrogen temperature, the space group changes to $P2_12_12$ with one molecule per asymmetric unit. In this case the twofold axis along *c* becomes a crystal symmetry element, replacing the twofold screw axis in crystal form I, and the sulfate ion is on a special position. The amino acid residues involved in binding the sulfate ion are His282 and the amide protons from Glu284 and Leu285. The ND1 atom of His282 is 2.8 Å from an oxygen atom of the sulfate ion. Since the crystals were grown at a pH of 4.6, His282 is probably protonated, forming a salt bridge to the negatively charged sulfate ion. A second salt bridge is formed by the symmetry-related His282′ and second negatively charged oxygen. The respective backbone nitrogens of Glu284 and Leu285 are 2.9 Å from O3 of the sulfate ion, forming two hydrogen bonds. The symmetry-related Glu384′ and Leu285′ residues form hydrogen bonds to the O4 atom of the sulfate ion. These distances are similar in the two-crystal forms.

However, other crystal packing contacts show dramatic differences between crystal form I and crystal form II. In form I, there are 45 crystal contacts with atoms from the N-domain and 34 crystal contacts with atoms from the C-domain. In form II crystals there are 54 crystal contacts with atoms from the N-domain and 31 crystal contacts with atoms from the C-domain. Therefore, there are 9 additional contacts in the N-domain and 3 fewer in the C-domain when the crystals are frozen. The change in the number of crystal contacts in forms I and II may explain the differential change in average *B*-factors for the N-domain and C-domain.

Structural Similarity to Other Proteins. Structural similarity searches against the Protein Data Bank using the DALI algorithm (*35*) revealed a striking relationship between thiaminase-I and the group-II periplasmic binding proteins (*36*) and bilobal transferins (*37*) (Table 5). These enzymes only show 7%−15% amino acid sequence identities with thiaminase-I, yet as much as 90% of the respective amino acids can be superimposed to give an rmsd of 4.0 Å for structurally aligned Cα atoms. All of these proteins have their binding sites in a deep cleft between two domains and appear to require a hinge motion for ligand binding. For example, the maltose-binding protein (MBP) has a hinge motion of approximately 35° when the substrate binds (*38*). Interestingly, all of the proteins with structural similarity to thiaminase-I are transport proteins with the exception of porphobilinogen deaminase. Therefore, thiaminase-I is the

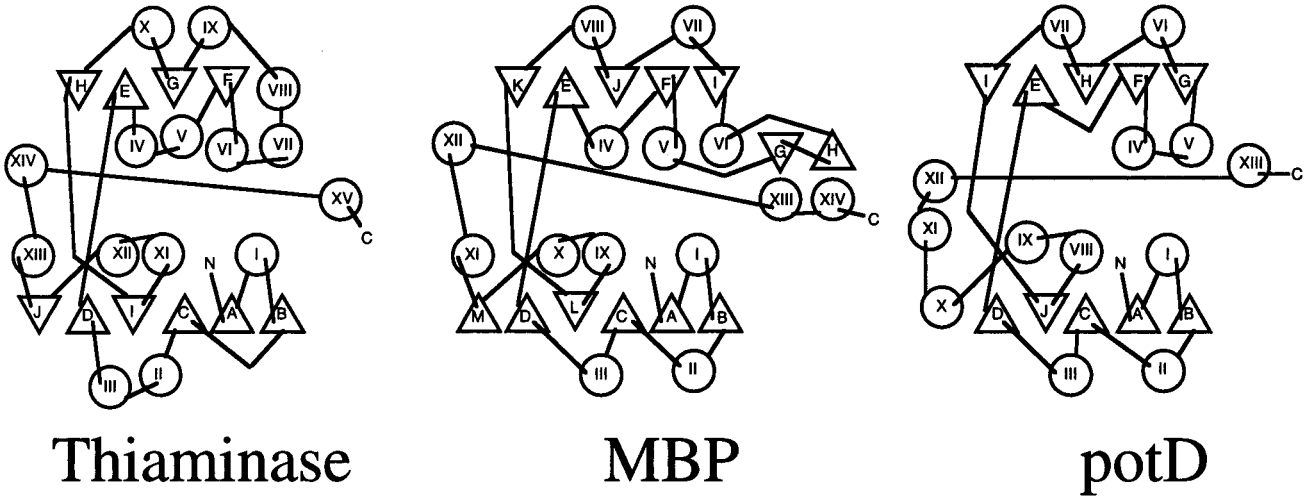


FIGURE 5: Topology drawings of thiaminase-I, maltose-binding protein, and spermidine/putrescine-binding protein. Triangles represent β-strands, and circles represent α-helices.

second enzyme to have a topology similar to the periplasmic binding proteins.

The greatest similarity occurs between the N-terminal domains of thiaminase-I and MBP (*36*) or spermidine putrescine-binding protein (potD) (*39*) (Figure 5). The respective C-domains are also similar with a few exceptions. The C-domain β-sheet of thiaminase-I has only 4 strands, whereas the β-sheets of MBP and PotD each contain 5 strands. In addition, maltose-binding protein contains an antiparallel, two-stranded β-sheet. All three proteins contain three crossover points between the N-domain and the C-domain. The structure of the liganded form of MBP and PotD shows that the ligand binds in the cleft between the two domains. Interestingly, the crystal structure of thiaminase-I is more similar to the closed form of MBP (1mbp) than to the open form (1omp). The structural similarity of thiaminase-I to the periplasmic binding proteins may provide a clue to the function of thiaminase-I. Since the biosynthesis of thiazole is complex and requires at least seven gene products in prokaryotes (based on studies in *E. coli*), thiaminase-I may have evolved from some periplasmic binding protein as a thiazole salvage enzyme.

Substrate Binding and Catalysis. The structure of thiaminase-I, labeled with 4-amino-6-chloro-2,5-dimethylpyrimidine, showed that the active site lies in a cleft between the N-domain and the C-domain (Figure 4). This thiamin analogue was previously shown to function as a mechanism-based inactivating agent and to label Cys113 with loss of chloride (Figure 6). The structure clearly shows the binding of Cys113 to the pyrimidine and identifies Glu241 as a possible active site base involved in the deprotonation of Cys113. Accordingly, the Glu241Gln mutant was inactive, supporting the assignment of an essential function to this residue. On the basis of stereoelectronic considerations, the nucleophilicity of the active site sulfur will be at a maximum when C6 of the pyrimidine, the SH bond, and the oxygen of Glu241 are collinear. In this arrangement, as electron density builds up on the sulfur due to the proton transfer, the C6 carbon of the pyrimidine is optimally positioned to react with it. This linear configuration of electrophile, nucleophile, and base is found in thioredoxin (*40*), where the electrophile is a disulfide, in the cysteine proteases (*41*), where the electrophile is a carbonyl carbon, and in thymidylate synthase

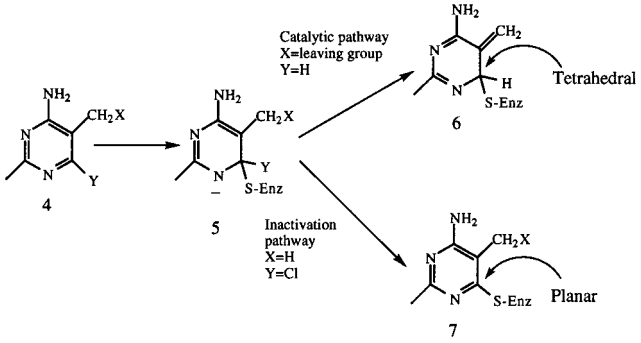


FIGURE 6: The inactivation of the enzyme with 4-amino-6-chloro-2,5-dimethylpyrimidine results in a planarized enzyme pyrimidine adduct.

(*42*), where the electrophile is the C6 carbon of deoxyuridine phosphate. We were therefore surprised to find that the angle between the electrophile, the nucleophile, and the base in the thiaminase structure was 90° rather than the expected angle of 180°.

Modeling studies suggest a possible binding geometry for thiamin. The model assumes that the pyrimidine binds similarly to the covalently bound inhibitor and that the thiazole is perpendicular to the plane of the pyrimidine as dictated by the stereoelectronic requirements of the displacement reaction. In this model, the thiazole binds in the cleft between the two domains but shows several unfavorable interactions with the aromatic residues lining the cleft. These unfavorable interactions could, however, be reduced by a hinge movement of the domains.

The nonisomorphous nature of the thiaminase-I crystals suggests the possibility for some type of structural change upon binding of inhibitor and/or freezing of crystals. However, an automatic search for hinge motions with the program VMD (*43*) showed only a small angular difference of 1.6° between the room temperature form (I) and low-temperature crystal form (II). The calculated hinge motion for the low-temperature crystal form II, with and without inhibitor, was only 0.2°. Both of these values are too small to suggest a significant hinge motion. The superposition of the three structures shows evidence of the flexible nature of the C-domain. This is in contrast to the periplasmic binding proteins for which large hinge motions have been observed

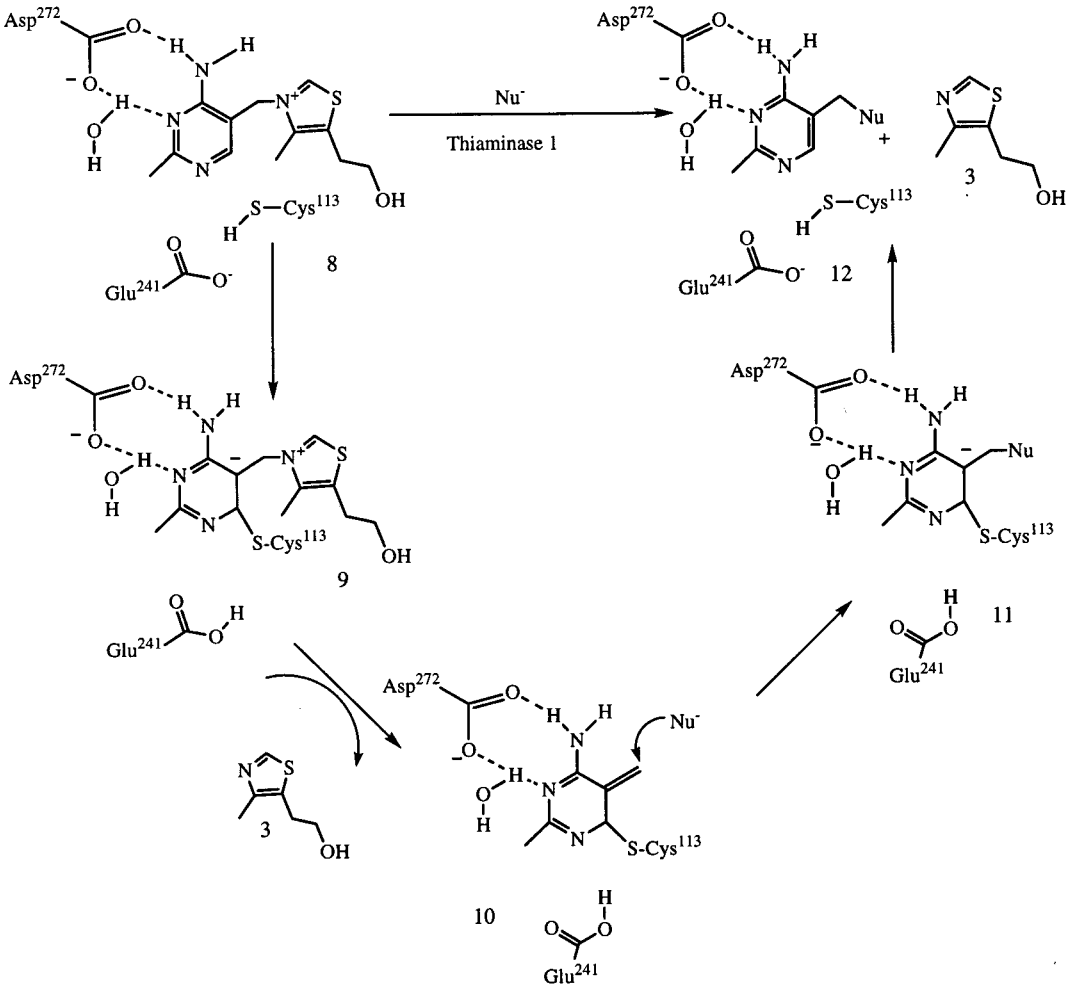


FIGURE 7: Mechanistic proposal for the thiaminase-I-catalyzed thiamin degradation reaction.

(*38*). If indeed the binding proteins and thaminase-I have a common ancestor, it is possible that the evolution of enzymatic function in thaminase-I required a rigid active site cavitity and the loss of the hinge motion required by the binding proteins. When the Cα atoms of the N-domain were superimposed, the analysis of the rms deviation of all Cα shows that the N-domain atoms had an rmsd = 0.4 Å, whereas the C-domain atoms had an average rmsd of 0.8 Å. It is possible that this flexibility is related to enzymatic activity; however, it may also result from differences in crystal-packing forces between the two crystal forms.

Catalytic Mechanism. The cleavage of thiamin by bisulfite is the closest nonenzymatic model reaction for the thiaminase-I-catalyzed reaction. For this reaction, protonation or methylation of the pyrimidine moiety at N1 increases the electrophilic character of the pyrimidine and accelerates the rate of the bisulfite addition to C6 (*44*, *45*). In the thiaminase-I/inhibitor structure, there are no acidic residues interacting with either N1 or N3 of the pyrimidine. The absence of such interactions is consistent with the pH rate profile of the enzyme which shows a broad optimum in the range 5−11 (*46*). Acid catalysis by the interaction of the pyrimidine with an un-ionized carboxylic acid is not consistent with such a pH rate profile. Asp272 and a water molecule are located close to but not interacting with the pyrimidine amino group and with N3, respectively. It is possible that these interactions occur during the catalytic reaction, but have been disrupted in the crystal structure of the thiaminase complex due to the rearomatization and resulting planarization of the pyrimidine that occurs during inactivation (Figure 6). However, these interactions are unlikely to strongly activate the pyrimidine because water is a weak acid and the water proton is also interacting with the carboxylate oxygen of Asp272. This suggests that the enzyme may utilize a strategy for pyrimidine activation different from that found in the bisulfite addition reaction.

A mechanistic proposal, based on the crystal structures and on the previous chemical studies, is outlined in Figure 7. The structure of the enzyme−inhibitor complex suggests that Cys113 is in close proximity to C6 of the pyrimidine of thiamin and that Glu241 is suitably positioned to serve as a base for the cysteine activation. Addition of Cys113 to the pyrimidine gives the zwitterionic adduct 9. Electrostatic stabilization of the negative charge on the pyrimidine by the positive charge on the adjacent thiazolium moiety is a plausible alternative to protonation for the stabilization of 9. Such an interaction is suggested by the observation that all thiamin analogues that are good substrates for the enzyme have a positively charged nitrogen attached to the pyrimidine (*8*). In addition, the k_{inact} for 4-amino-6-chloro-2-methylpyrimidine, which does not have a positive charge, is 31 000 times slower than k_{cat} for thiamin degradation by veratrylamine (*9*). To complete the reaction, cleavage of the bond linking the thiazole to the pyrimidine, followed by diffusion

of the thiazole out of the active site, would give 10. The pathway for the conversion of 10 to the final reaction products, 12 and 3, is the microscopic reverse of the pathway for the conversion of 8 to 10. Addition of the nucleophile to 10 would give 11. The location of the thiazole and the nucleophile in the large cleft between the two domains of the enzyme is consistent with the wide tolerance of thiaminase I to the structure of the nucleophile (*8*). In addition, the structure of the enzyme−inhibitor complex demonstrates that only one face of the pyrimidine is accessible to attack by the nucleophile, a result that is consistent with the stereochemistry (*17*) and the ping−pong kinetics of the reaction (*15, 46*). It is also interesting to note that the best nucleophiles for thiaminase I are either very strong nucleophiles such as thiolate and azide or amine bases that are capable of generating a zwitterionic intermediate (*8, 16*). The elimination of Cys113 from 11, catalyzed by Glu241, would complete the reaction.

To provide further evidence for the proposed catalytic mechanism, attempts have been made to model thiamin into the active site of the native enzyme. However, several possible orientations of the pyrimidine ring could be generated and it was not possible to unambiguously determine the thiamin-binding geometry. Cocrystallization experiments to resolve this issue are underway.

ACKNOWLEDGMENT

We thank Robb Nicewonger for providing samples of 4-amino-6-chloro-2,5-dimethylpyrimidine.

REFERENCES

1. Fujita, A. (1954) *Adv. Enzymol. 15*, 389−421.
2. Fujita, A. (1972) *J. Vitaminol. 18*, 67−72.
3. Evans, C. (1975) *Vitam. Horm. 33*, 467−504.
4. Murata, K. (1982) *Ann. N. Y. Acad. Sci. 378*, 146−156.
5. Hayashi, R. (1957) *Nutr. Rev. 15*, 65−67.
6. Duffy, P., Morris, H., and Neilson, G. (1981) *Am. J. Clin. Nutr. 34*, 1584−1592.
7. Earl, J. W., and McCleary, B. V. (1994) *Nature 368*, 683−684.
8. Lienhard, G. (1970) *Biochemistry 9*, 3011−3020.
9. Costello, C. A., Kelleher, N. L., Abe, M., McLafferty, F. W., and Begley, T. P. (1996) *J. Biol. Chem. 271*, 3445−3452.
10. Douthit, H., and Airth, R. (1966) *Arch. Biochem. Biophys. 113*, 331−337.
11. Abe, M., Ito, S., Kimoto, M., Hayashi, R., and Nishimune, T. (1987) *Biochim. Biophys. Acta 909*, 213−221.
12. Kimura, Y., and Iwashima, A. (1987) *Experientia 43*, 888−890.
13. Kelleher, N. L., Costello, C. A., Begley, T. P., and McLafferty, F. W. (1995) *J. Am. Soc. Mass Spectrom. 6*, 981−984.
14. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403−410.
15. Puzach, S. S., Gorbach, Z. V., and Ostrovskii, Y. M. (1984) *Biochemistry* (*Transl. of Biokhimiya (Moscow)*) *49*, 1010−1016.
16. Nicewonger, R., Rammelsberg, A., Costello, C. A., and Begley, T. P. (1995) *Bioorg. Chem. 23*, 512−518.
17. Nicewonger, R., Costello, C., and Begley, T. P. (1996) *J. Org. Chem. 61*, 4172−4174.
18. Hutter, J. A., and Slama, J. T. (1987) *Biochemistry 26*, 1969−1973.
19. Kelleher, N. L., Nicewonger, R. B., Begley, T. P., and McLafferty, F. W. (1997) *J. Biol. Chem. 272*, 32215−32220.
20. Campobasso, N., Begun, J., Costello, C. A., Begley, T. P., and Ealick, S. E. (1998) *Acta Crystallogr. D54*, 448−450.
21. Hamlin, R. (1985) *Methods Enzymol. 114*, 416−452.
22. Howard, A. J., Nielsen, C., and Xuong, N. H. (1985) *Methods Enzymol. 114*, 452−472.
23. Collaborative Computing Project, No. 4. (1994) *Acta Crystallogr. D50*, 760−763.
24. 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.
25. Otwinowski, Z., and Minor, W., (1997) *Methods Enzymol. 276*, 307−326.
26. Otwinowski, Z. (1991) pp 80−86, SERC Daresbury Laboratory, Warrington, U.K.
27. Cowtan, K. (1994) Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, Vol. 31, pp 34−38.
28. Kleywegt, G. J., and Jones, T. A. (1994) *Proceedings of the CCP4 Study Weekend.* (Bailey, S., Hubbard, R., and Waller, D., Eds.) pp 59−66, EPSRC Daresbury Laboratory, Warrington, U.K.
29. Jones, T. A., Zou, J. Y., Cowtan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
30. Brünger, A. T. (1992) X-PLOR Manual, Version 3.1, Yale University Press, New Haven and London.
31. Engh, R. A., and Huber, R. (1991) *Acta Crystallogr. A47*, 392−400.
32. Jiang, J.-S., and Brunger, A. T. (1994) *J. Mol. Biol. 243*, 100−115.
33. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.
34. Hubbard, S. J., and Thornton, J. M. (1993) NACCESS, Computer program, Department of Biochemistry and Molecular Biology, University College London.
35. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
36. Spurlino, J. C., Lu, G.-Y., and Quicho, F. A. (1991) *J. Biol. Chem. 266*, 5202−5219.
37. Sarra, R., Garratt, B., Jhoti, H., and Lindley, P. (1990) *Acta Crystallogr. B46*, 763−771.
38. Sharff, A. J., Rodseth, L. E., Spurlino, J. C., and Quicho, F. A. (1992) *Biochemistry 31*, 10657−10663.
39. Sugiyama, S., Vassylyev, D. G., Matsushima, M., Kashiwagi, K., Igarashi, K., and Morikawa, K. (1996) *J. Biol. Chem. 266*, 9519−9525,
40. Katti, S. K., Le' Master, D. M., and Eklund, H. (1990) *J. Mol. Biol. 212*, 167−184.
41. Weichsel, A., and Montfort, W. R. (1995) *Nat. Struct. Biol. 2*, 1095−1101.
42. Zhao, B., Janson, C. A., Amegadzie, B. Y., D'Alessio, K., Griffin, C., Hanning, C. R., Jones, C., Kurdyla, J., McQueney, M., Qiu, X., Smith, W. W., and Abdel-Meguid, S. S. (1997) *Nat. Struct. Biol. 4*, 109−11.
43. Humphrey, W., Dalke, A., and Schulten, K. (1996) *J. Mol. Graphics 14*, 33−38.
44. Zoltewicz, J. A. (1980) *Synthesis*, 218−219.
45. Zoltewicz, J. A., and Kauffman, G. M. (1977) *J. Am. Chem. Soc. 99*, 3134−3142.
46. Mochida, K., Ozoe, Y., Nakamura, T., Nagao, T., and Suzuki, K. (1986) *Bull. Fac. Agric., Shimane Univ. 20*, 196−201.
47. Sacks, J. (1991) Computer program, CHAIN, Baylor College of Medicine.
48. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
49. Carson, M. (1998) *Methods Enzymol.* (in press).

BI981673L