Articles

Crystal Structure of 4-Methyl-5-β-hydroxyethylthiazole Kinase from *Bacillus subtilis* at 1.5 Å Resolution[†,‡]

Nino Campobasso, Irimpan I. Mathews, Tadhg P. Begley, and Steven E. Ealick*

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

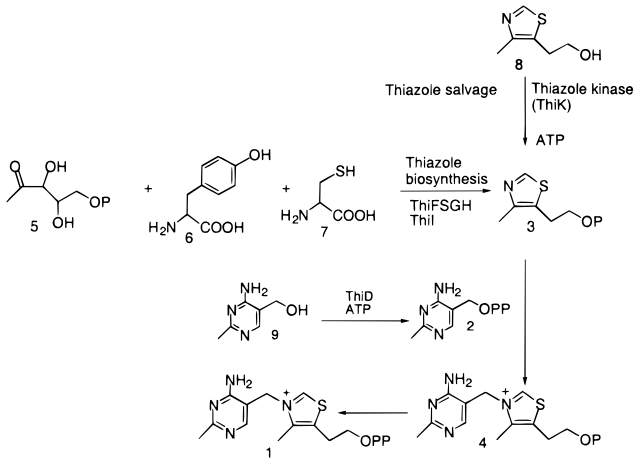
Received January 4, 2000; Revised Manuscript Received April 12, 2000

ABSTRACT: 4-Methyl-5-β-hydroxyethylthiazole kinase (ThiK) catalyzes the phosphorylation of the hydroxyl group of 4-methyl-5-β-hydroxyethylthiazole (Thz). This enzyme is a salvage enzyme in the thiamin biosynthetic pathway and enables the cell to use recycled Thz as an alternative to its synthesis from 1-deoxy-D-xylulose-5-phosphate, cysteine, and tyrosine. The structure of ThiK in the rhombohedral crystal form has been determined to 1.5 Å resolution and refined to a final *R*-factor of 21.6% (*R*-free 25.1%). The structures of the enzyme/Thz complex and the enzyme/Thz-phosphate/ATP complex have also been determined. ThiK is a trimer of identical subunits. Each subunit contains a large nine-stranded central β-sheet flanked by helices. The overall fold is similar to that of ribokinase and adenosine kinase, although sequence similarity is not immediately apparent. The area of greatest similarity occurs in the ATP-binding site where several key residues are highly conserved. Unlike adenosine kinase and ribokinase, in which the active site is located between two domains within a single subunit, the ThiK active site it formed at the interface between two subunits within the trimer. The structure of the enzyme/ATP/Thz-phosphate complex suggests that phosphate transfer occurs by an inline mechanism. Although this mechanism is similar to that proposed for both ribokinase and adenosine kinase, ThiK lacks an absolutely conserved Asp thought to be important for catalysis in the other two enzymes. Instead, ThiK has a conserved cysteine (Cys198) in this position. When this Cys is mutated to Asp, the enzymatic activity increases 10-fold. Further sequence analysis suggests that another thiamin biosynthetic enzyme (ThiD), which catalyzes the formation of 2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate by two sequential phosphorylation reactions, belongs to the same family of small molecule kinases.

Thiamin pyrophosphate **1** is the cofactor involved in the stabilization of the acyl carbanion intermediate in biosynthesis and plays a key role in carbohydrate metabolism (*1*). It is an essential component of the human diet (RDA = 1.4 mg). Annual production, by chemical synthesis, is 3300 tons (*2*).

The biosynthesis of thiamin pyrophosphate involves the separate formation of 2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate **2** and 4-methyl-5-β-hydroxyethylthiazole phosphate (Thz-P,[1] **3**) followed by their subsequent coupling catalyzed by thiamin phosphate synthase and a final phosphorylation (Scheme 1) (*3*). The de novo biosynthesis of Thz-P **3** from 1-deoxy-D-xylulose-5-phosphate **5**, cysteine **6**, and tyrosine **7** is complex and requires at least six gene products. If 4-methyl-5-β-hydroxyethylthiazole (Thz, **8**) is available from exogenous sources, this

Scheme 1



[†] This work supported by a 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 Protein Data Bank codes for 4-methyl-5-β-hydroxyethylthiazole kinase structures are ThiK, 1EKK; ThiK/Thz form I, 1EKQ; ThiK/Thz form II, 1C3Q; ThiK(C198S), 1ESJ; and ThiK(C198S)/Thz-P/MgATP, 1ESQ.

* To whom correspondence should be addressed at Cornell University. Phone: (607) 255-7961. Fax: (607) 255-4137. E-mail: see3@cornell.edu.

[1] Abbreviations: ThiK, 4-methyl-5-β-hydroxyethylthiazole kinase; Thz, 4-methyl-5-β-hydroxyethylthiazole; Thz-P, 4-methyl-5-β-hydroxyethylthiazole phosphate; CHESS, Cornell high-energy synchrotron source; APS, advanced photon source; ADSC, area detector systems corporation; RMSD, root-mean-square difference; NCS, noncrystallographic symmetry.

10.1021/bi0000061 CCC: $19.00
Published on Web 06/14/2000

complex biosynthesis is bypassed and the thiazole moiety is taken up and phosphorylated inside the cell. This phosphorylation reaction is catalyzed by 4-methyl-5-β-hydroxyethylthiazole kinase (ThiK). The gene for this enzyme has been cloned from *Bacillus subtilis*, (*4*) *Salmonella typhimurium* (*5*), and *Saccharomyces cerevisiae* (*6*) and can be identified in the genome sequence of five other microorganisms.

ThiK from *B. subtilis* is a trimer with identical subunits of approximately 29 600 Da each (284 amino acids/subunit). While the *B. subtilis* and the *S. typhimurium* enzymes are monofunctional, the *S. cerevisiae*, enzyme is bifunctional and also has thiamin phosphate synthase activity. We undertook the structural study of ThiK as part of a systematic study of the enzymes involved in thiamin biosynthesis. We hope that these studies will reveal new information about how these enzymes perform their catalytic functions. In addition, we hope to discover new information about how enzymes within a biosynthetic pathway relate to each other and to enzymes in other biochemical pathways. In this paper, we describe the structures of the homotrimeric ThiK from *Bacillus subtilis* alone, the binary complex with Thz and the ternary complex with Thz-P and ATP.

EXPERIMENTAL PROCEDURES

Protein Purification. ThiK from *B. subtilis* was cloned, overexpressed in *Escherichia coli*, and purified as described previously (*6*). After the initial structure determination, it was found that the addition of 2 mM β-mercaptoethanol was required throughout the purification to inhibit the oxidation of cysteine by the Ni^{2+} column. A 1 L culture yielded 70−90 mg of pure protein. The stock solution of ThiK, which was stored at −80 °C, was composed of 30 mg/mL protein, 50 mM Tris HCl (pH 7.5) and 2 mM dithiothreitol.

Crystallization. ThiK was crystallized in two different crystal forms using vapor diffusion equilibration. In general, rhombohedral crystals (crystal form I) were obtained in the absence of substrates, while monoclinic crystals (crystal form II) were obtained when substrates were added. Otherwise, the crystallization conditions for the two crystal forms were similar. Crystal form II could also be obtained in the absence of substrates by streak seeding into solutions containing only protein. The rhombohedral crystals belong to space group $R3$ with unit cell dimensions of $a = 77.5$ Å and $c = 230.4$ Å and two monomers. The Matthews coefficient (*7*) for crystal form I is 2.54 $Å^3$/Da, and the solvent content is 49%. The monoclinic crystals belong to space group $P2_1$ with unit cell dimensions of $a = 53.7$ Å, $b = 101.2$ Å, $c = 73.2$ Å, $\beta = 95.1°$, and three monomers per asymmetric unit. The Matthews coefficient (*7*) for crystal form II is 2.23 $Å^3$/Da, and the solvent content is 44%.

Rhombohedral crystals were obtained from hanging drops consisting of 2 μL of 30 mg/mL protein solution plus 2 μL of reservoir solution containing 100 mM Tris HCl (pH 7.5), 100 mM ammonium sulfate, and 30% (w/v) PEG 4K. After equilibration at room temperature against 1 mL of reservoir solution, crystals appeared in 5−10 days and grew to maximum dimensions of $0.6 \times 0.6 \times 0.6$ mm^3. The reservoir solution was used directly as a cryoprotectant for collecting data at low temperature. Monoclinic crystals were obtained from sitting drops consisting of 3 μL of 30 mg/mL protein solution containing 50 μM Thz, 50 μM ATP, and 100 μM $MgCl_2$ and 3 μL of reservoir solution containing 100 mM Tris HCl (pH 8.6) and 20% (w/v) PEG 8K. After equilibration at 18 °C against 1 mL of reservoir solution, crystals appeared in 2−4 days and grew to the maximum dimensions of $0.4 \times 0.2 \times 0.1$ mm^3. Cryoprotectants for the monoclinic crystals were prepared by adding 25% glycerol to thereservoir solution or by substituting PEG 4K for PEG 8K in the soaking solutions. Monoclinic crystals of a C198S mutant of ThiK were prepared using two rounds of streak seeding (*8*). The crystallization conditions were similar to those described above for crystal form II.

Preparation of ThiK Mutants. C198A, C198D, and C198S mutants were prepared using the QuikChange site-directed mutagenesis kit (Stratagene). The Oligonucleotide primers used for the C198A, C198D, and C198S mutagenesis reaction were 5′-ACAGGCGCCGGTGCCCTGCTGACTTCCGTCG-3′ and 5′-CGACGGAAGTCAGCAGGGCACCGGCGCCTGT-3′; 5′-ACAGGCGCCGGTGACCTGCTGACTTCCGTCG-3′ and 5′-CGACGGAAGTCAGCAGGTCACCGGCGCCTGT-3′; 5′-ACAGGCGCCGGTTCCCTGCTGACTTCCGTCG-3′ and 5′-CGACGGAAGTCAGCAGGGAACCGGCGCCTGT-3′, respectively, where the underlined codons indicate the site of mutation. The mutations were confirmed by DNA sequencing. The mutants were overexpressed and purified as described for the native ThiK.

Enzyme Assay. The ThiK activity of the C198 mutants was determined in triplicate under V_{max} conditions [K_m(Thz) = 34 μM, K_m(ATP) = 8.5 μM] using pyruvate kinase/lactate dehydrogenase to monitor ATP consumption (*9*). The reaction mixture (0.7 mL total volume) consisted of 5 μg of ThiK, 5 units of pyruvate kinase, 7 units of lactate dehydrogenase, 1 mM ATP, 1 mM Thz, 10 mM $MgCl_2$, 50 mM KCl, 0.2 mM NADH, and 1 mM phosphoenolpyruvate in 100 mM Tris HCl, pH 8.0. The reaction was followed by measuring the decrease in NADH absorbance at 340 nm ($\epsilon = 6220$ M^{-1} cm^{-1}).

X-ray Diffraction Data. X-ray data collection was carried out at stations A1 and F1 of the Cornell High Energy Synchrotron Source (CHESS) using CCD detectors from Area Detector Systems Corporation (ADSC), San Diego, CA, or at the Advance Photon Source (APS) Structural Biology Center beam line 19ID using a locally designed and constructed CCD detector (*10*).

The initial native data for the rhombohedral crystal form were collected at CHESS at cryogenic temperatures in two passes to increase the dynamic range. The diffraction data in the first pass were collected with 92 frames with 1.5° oscillations for 15 s per frame and with a crystal to detector distance of 69 mm. For the second pass, 30 frames of diffraction data were collected in 5° oscillations for 6 s/frame and with a crystal to detector distance of 120 mm. Heavy-atom derivatives were screened by collecting 90 frames of data in one pass with crystal to detector distance of 69 mm, an oscillation angle of 1.5°, and an exposure time of 15 s. Data were processed using DENZO (*11*) and CCP4 (*12*) programs. At this point, the low resolution data were still 20% incomplete for the resolution shell 30−8 Å, because of pixel saturations. Therefore, an additional native crystal was used to complete the low-resolution shell with data collected using a rotating anode X-ray generator and an ADSC multiwire detector. Final statistics of the merged data are given in Table 1.

Table 1: Data Collection Statistics

	form I			form II		
	wild-type	Hg	wt/Thz	wt/Thz	C198S	C198S/Thz-P/ATP
resolution (Å)[a]	1.5 (1.57−1.5)	2.2 (2.3−2.2)	2.0 (2.1−2.0)	2.0 (2.1−2.0)	1.7 (1.76−1.7)	2.5 (2.6−2.5)
space group	*R*3	*R*3	*R*3	$P2_1$	$P2_1$	$P2_1$
cell parameters						
a (Å)	77.5	77.6	78.1	53.9	53.1	54.1
b (Å)				101.2	100.9	100.8
c (Å)	230.4	230.0	232.8	73.3	73.1	72.5
β (deg)				95.3	96.0	95.1
Z (subunits)	9	9	9	6	6	6
observations	264 643	85 615	209 109	162 565	340 088	111 367
reflections	73 978	25 732	35 426	51 102	83 184	26 976
completeness (%)	89.0 (54.9)	99.2 (98.3)	99.2 (99.2)	98 (95)	98.6 (94)	99.9 (100)
R_{sym} (%)	0.071 (0.158)	0.086 (0.113)	0.098 (0.172)	0.082 (0.275)	0.042 (0.375)	0.068 (0.258)
redundancy	3.6 (1.5)	4 (2.0)	5.9 (3.7)	4.3 (3.9)	4.1 (3.7)	4.1 (4.0)
I/σ	15.0 (4.0)	26.8 (15.7)	17.3 (8.7)	18.5 (5.8)	23.0 (2.1)	18.0 (4.3)
phasing power		2.0 (1.3)				
FOM		0.50 (0.30)				
FOMdensity mod		0.89				

[a] Values in parentheses correspond to the outermost resolution shell.

Data for the ThiK complexes were collected at cryogenic temperatures at CHESS beamline F1 using an ADSC Quantum 4 detector. For the rhombohedral ThiK/Thz complex, 130 frames of 1° rotation data were collected with an exposure time of 20 s for each frame. The crystal to detector distance was 172 mm. For the $P2_1$ ThiK/Thz-P complex, 220 frames of 1° rotation data were collected with an exposure time of 30 s per frame. The crystal to detector distance was 200 mm. CHESS intensity data were integrated with the MOSFLM package and scaled with the CCP4 package (*12*). Data processing statistics are given in Table 1.

Two data sets for the C198S mutant crystals were collected at cryogenic temperature using the Advance Photon Source (APS) beam line 19ID. The first data set was collected with unliganded C198S crystals, and the second data set was collected with a C198S crystal soaked in Thz and ATP. The data were measured using a 3 × 3 mosaic CCD detector (*10*) placed 150 mm from the sample. A total of 200° of data was measured for each crystal using 1° frames measured for 30 s. Data were processed and scaled using HKL2000 (*11*). Details of the data collection statistics are given in Table 1.

Structure Determination and Refinement. The crystal structure was determined initially by single isomorphous replacement with anomalous scattering. A mercury derivative was obtained by soaking rhombohedral crystals for 16 h in the reservoir solution in the presence of 1 mM sodium ethylmercurithiosalicylate (thimersol). Four heavy-atom sites were found by inspection of difference Patterson maps. Heavy-atom parameter refinement and phase calculations were performed with the program MLPHARE (*13*). Phases were calculated to 2.2 Å and improved with solvent flattening and histogram matching by using the program DM (*14*). An initial model was built to define the molecular envelope and noncrystallographic symmetry operator (NCS). The NCS operator was refined by real-space density correlation using the program IMP from the RAVE package (*15*). Phase improvement was repeated with DM (*14*) to extend the resolution to 1.5 Å by including NCS averaging, solvent flattening, and histogram matching (Figure 1).

The atomic model was built with the program O (*16*). The resulting 1.5 Å electron density map clearly showed interpretable electron density and identifiable topology of the protein. To aid in chain tracing, the electron density map was skeletonized with RAVE (*15*). The resulting model was refined against data between 30 and 1.5 Å by simulated annealing and positional refinement using Engh and Huber (*17*) restraints as incorporated in X-PLOR 3.841 (*18*). A bulk solvent correction was also applied throughout the refinements (*19*). 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 solvent molecules. Individual atoms were assigned isotropic *B*-factors which were refined during the later stages of the refinements.

The structure of ThiK was also determined in the monoclinic crystal form by molecular replacement techniques using the program CNS (*20*) and data for the ThiK C198S mutant. A complete trimer from the rhombohedral crystal form was used as the search model. Solutions for the rotation and translation functions were easily found and further improved by simulated annealing torsion angle refinement via maximum likelihood refinement with the CNS package. Individual atoms were assigned isotropic *B*-factors and refined during the later stages of refinement.

The structures of the ThiK complexes were determined by refining the model against the data for the complex using the monoclinic or rhombohedral structure as appropriate. In all cases, the ligands were clearly defined in a difference electron density map and manually fitted. In the case of the wild-type/Thz complex only Thz was visible even though 10 mM Thz, 10 mM ATP, and 20 mM $MgCl_2$ were added to the crystals. Bond distance and angle restraints for Thz were adapted from a thiamin phosphate structure found at the Hetero-Compound Information Centre-Uppsala (*21*). Multiple rounds of simulated annealing, individual isotropic temperature factor refinement, and model improvement progressed routinely.

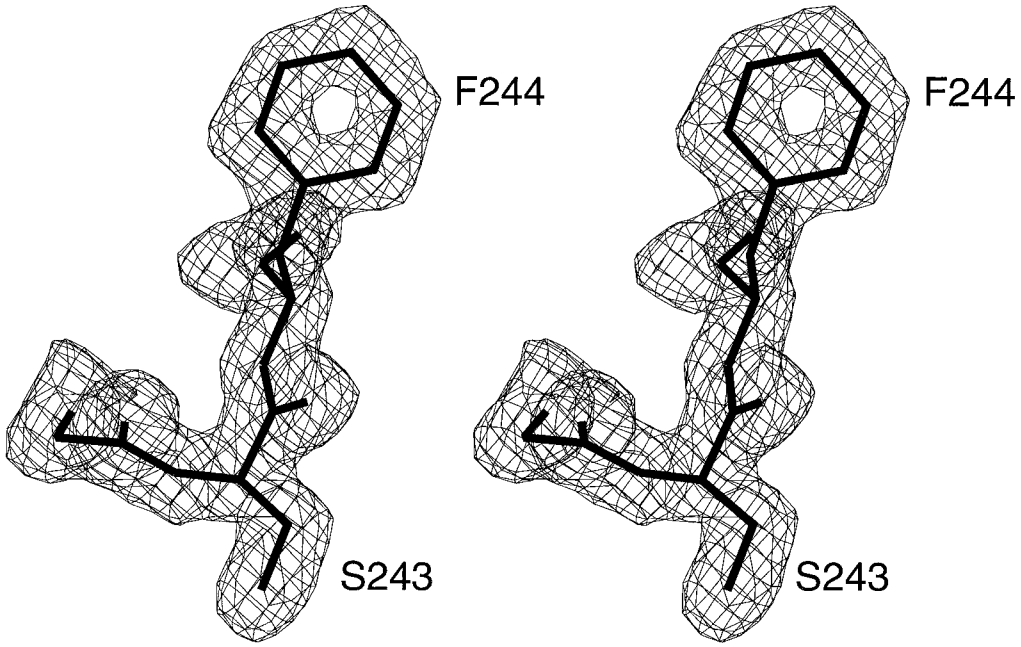


FIGURE 1: Stereoview of representative electron density and refined model of residues G241, S243, and F244 at 1.5 Å resolution. The electron density, contoured at 1σ was calculated with $2F_o - F_c$ as coefficients.

Table 2: Refinement Statistics

	form I		form II		
	wild-type	wt/Thz	wt/Thz	C198S	C198S/Thz-P/ATP
resolution (Å)	20−1.5	20−2.0	20−2.0	20−1.8	20−2.5
high res shell (Å)	1.57−1.50	2.07−2.0	2.07−2.0	1.80−1.81	2.53−2.50
R-factor (%)[a]	21.6 (30.6)	20.5 (20.9)	21.8 (26.9)	22.9 (22.9)	20.9 (24.2)
R-free (%)[a]	25.1 (29.1)	24.2 (22.0)	25.1 (28.0)	25.3 (23.1)	28.2 (31.4)
reflections (F > 1σ)	68965	35003	49532	62186	26790
no of protein atoms	3728	3728	6125	6095	5743
no. of Thz (−P) atoms		18	27		39
no. of ATP atoms					93
no. of Mg^{2+}/no. of SO_4^{-2}		−/2		−/2	6/1
no. of water molecules	258	122	208	342	179
B-factors (Å^2)					
side chain	25.3	23.4	29.2	21.1	33.7
Thz(−P)		18.6	45.8		26.0
ATP					36.9
water	30.0	25.4	31.4	25	27.7
RMSD bonds (Å)	0.009	0.006	0.006	0.005	0.007
RMSD angles (deg)	1.39	1.21	1.19	1.19	1.23
RMSD torsion angles (deg)	23.3	22.0	22.3	21.9	21.5
coordinate error (Å)	0.22	0.23	0.28	0.29	0.31

[a] Values in parentheses correspond to the outermost resolution shell.

RESULTS

Quality of the Final Models. A total of five ThiK structures have been determined: wild-type (1.5 Å), rhombohedral wild-type plus Thz (2.0 Å), monoclinic wild-type plus Thz (2.0 Å), C198S (1.8 Å), and C198plus Thz-P plus MgATP (2.5 Å). The crystallographic *R*-factors range from 20.5 to 22.9% and the coordinate errors as estimated from a Luzzati plot (*22*) range from 0.22 to 0.31 (Å). In the rhombohedral crystal form (wild-type and wild-type plus Thz), the asymmetric unit contains two subunits corresponding to the two unique 3-fold axes in space group *R*3. In the monoclinic crystal form (wild-type plus Thz, C198S, and C198S plus Thz-P and MgATP), the asymmetric unit contains a complete trimer. In the rhombohedral crystal form, residues 132−146, which are part of a flexible loop, were disordered and not visible in the electron density maps. In addition, the 12 amino acids of a polyHis tail at the N-terminus along with three residues at the carboxy terminus were not visible. In the monoclinic crystal form, several of the residues in the disordered loop were visible; however, many of the residues in the C-terminal half of the loop were still missing in each subunit. In the monoclinic crystal form, all of the polyHis tail was visible and these residues were included in the final model. Final refinement statistics for all five structures are given in Table 2.

Overall Structure of ThiK. The crystal structure of ThiK shows that the enzyme is a trimer with three identical subunits. The overall shape of the trimer is approximately that of an equilateral triangular prism, 78 Å along the edge, and 45 Å thick (Figure 2). In the rhombohedral crystals, the trimer axis is along the crystallographic 3-fold axis. Two trimers, each generated from one of the two independent

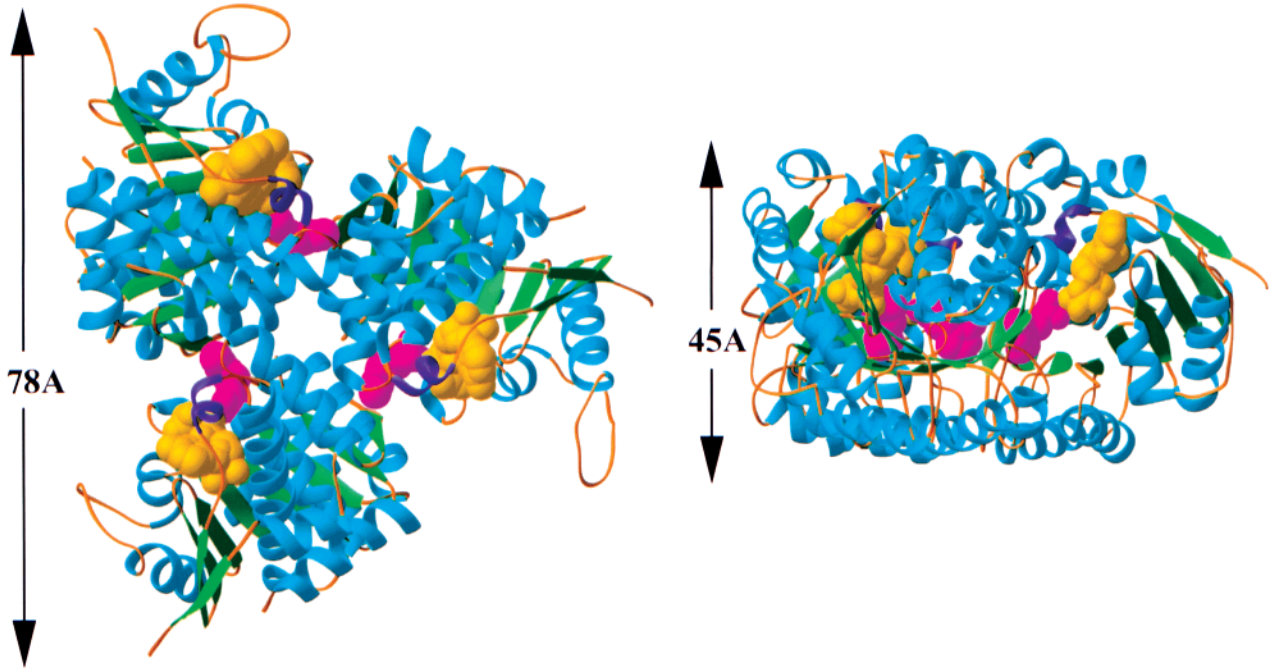


FIGURE 2: Orthogonal views of the quaternary structure of ThiK. The left view is down the 3-fold axis; the right view is rotated 90°. The figure was prepared with RIBBONS (*31*).

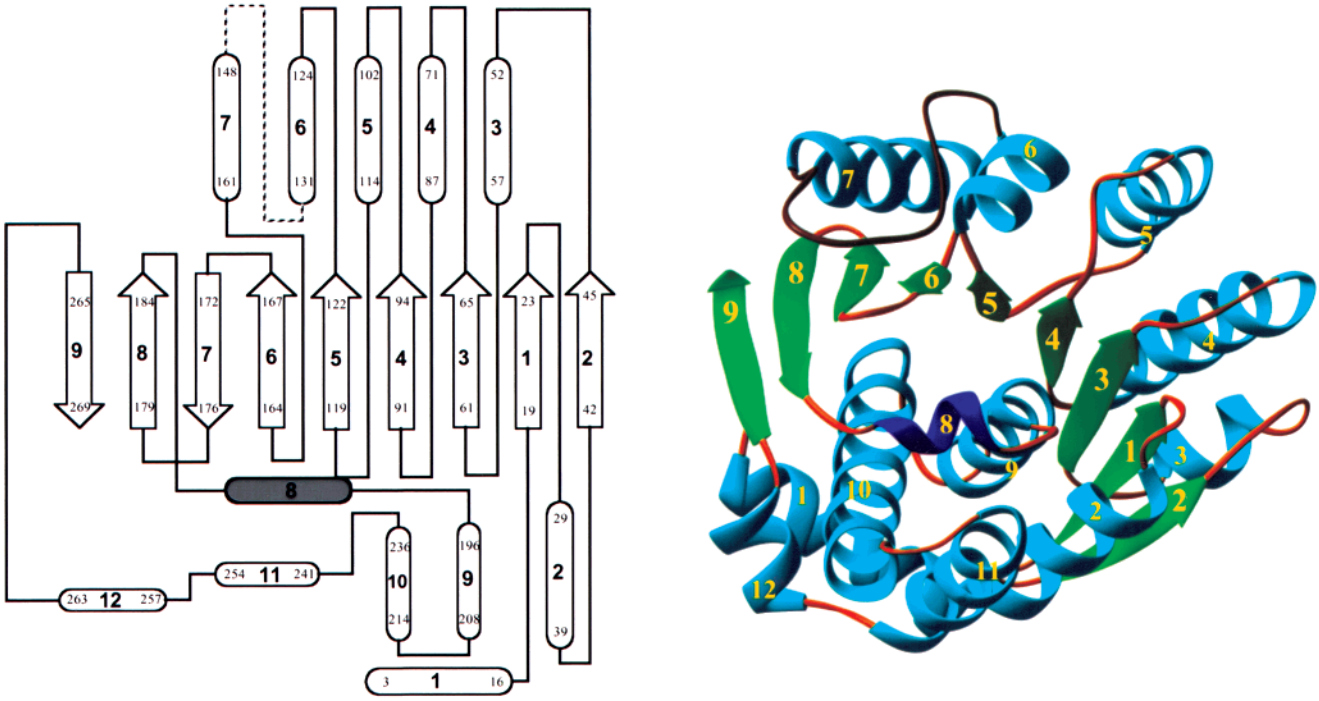


FIGURE 3: Topology and ribbon drawing of the subunit structure of ThiK. The β-strands are represented by arrows and the α-helices by cylinders. The smaller numbers indicate the beginning and ending residue number of each secondary structural element. The dotted line depicts the disordered loop. In the ribbon drawing, β-strands are depicted as flat green arrows; α-helices are depicted as cyan coils, and loops are depicted in orange. The blue helix is a 3_{10} helix. The black coil is the disorder region 132−147 modeled. The figure was prepared with RIBBONS (*31*).

subunits, are related by noncrystallographic translational symmetry. ThiK is an α/β-protein with 52% of the amino acids in α-helices and 14% in β-sheets. The remaining 34% of the structure is comprised of loops and coils. Each subunit contains a central nine-stranded mixed β-sheet that has topology $\beta2\beta1\beta3\beta4\beta5\beta6\beta7\beta8\beta9$ ($-1x$, $+2x$, $+1x$, $+1x$, $+1x$, $+1$, $+1x$, and $+1$) and 12 helices (Figure 3). The first six strands ($\beta2$, $\beta1$, $\beta3$, $\beta4$, $\beta5$, and $\beta6$) are parallel to each other, whereas strands $\beta6$, $\beta7$, $\beta8$, and $\beta9$ are antiparallel. Five helices (α3, α4, α5, α6, and α7) flank one side of the sheet and are approximately antiparallel to the strands of the β-sheet. The remaining seven helices are at the opposite side and form a bundle. Helices α2, α9, and α10 of the bundle abut the β-sheet. Viewed down the 3-fold axis, the trimer resembles a three blade propeller (Figure 2). The overall structural features for ThiK in the monoclinic form are identical to ThiK in the rhombohedral form. However, the asymmetric unit of the monoclinic forms contain a complete

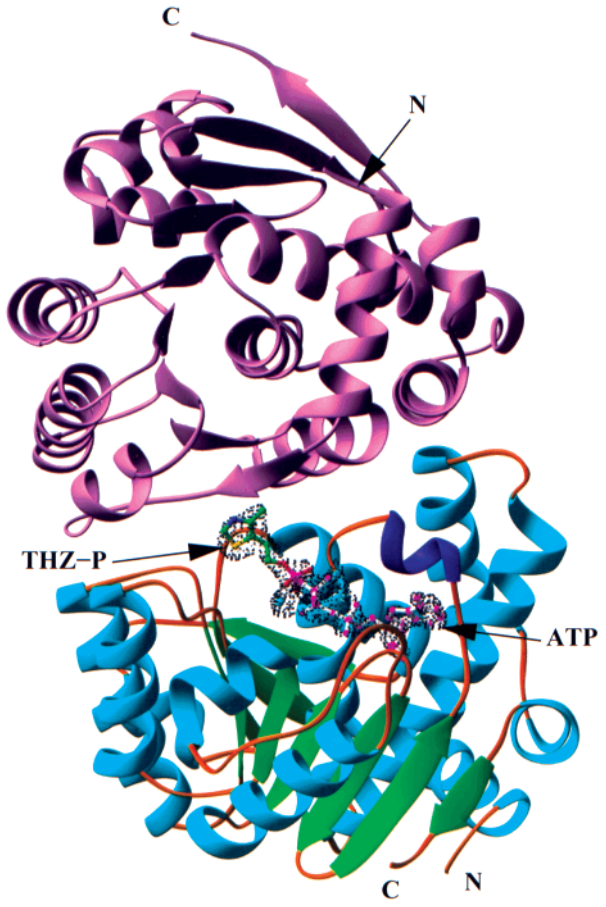


FIGURE 4: Binding site of Thz-P, ATP, and Mg^{2+} in ThiK. The difference electron density was calculated with coefficients ($F_o - F_c$) contoured at 3.0σ. It is represented as a dot surface. Substrate models were not yet included in the refinement at this stage. The lavender ribbon indicates a separate monomer to show that the substrate is at the monomer−monomer interface. The figure was prepared with RIBBONS (*31*).

trimer with the 3-fold noncrystallographic axis approximately perpendicular to the crystallographic 2-fold axis.

Active Site. The ThiK trimer contains three identical active sites. The active site is located at the interface between adjacent subunits. Each active site contains a Thz binding site and an ATP binding site as described below.

Hydroxyethylhiazole Binding. Thz or Thz-P was observed in three of the five crystal structures reported. The substrate Thz is located at the interface of two monomers (Figures 4 and 5). Interactions are contributed from highly conserved residues (Figure 6) by Asn25, Val27, Gly67, Thr68, and Thr194 from one chain and residues Ala33, Leu37 Pro43, and Met45 from a second chain. One face of the thiazole ring stacks against Val27 while the opposite face is more solvent exposed. A hydrogen bond is formed between the thiazole nitrogen atom and the amide nitrogen of Met45. The side chain of Cys198 is positioned at the bottom of the active site near to the hydroxyl group of the Thz. In the rhombohedral crystal form, Cys198 appears to be oxidized, probably to the sulfinic acid. In the monoclinic crystal form, which was analyzed at a later stage, Cys198 is not oxidized and we conclude that the oxidation is an artifact of the purification protocol and can be eliminated by adding β-mercaptoethanol during purification. The thiazole rings have the lowest thermal parameters in the rhombohedral crystal form with average *B*-factors of 17.8 and 15.6 Å^2 for the two subunits. The hydroxyl group of Thz is more flexible with *B*-factors of 27 and 24 Å^2. The average torsion angles for the hydroxymethyl group is −93° for the rotation (S1−C5−C6−C7) and 173° for the second rotation (C5−C6−C7−O7).

One of the C198S mutant crystal structures is a hybrid complex containing product Thz-P and reactant ATP. In this structure, the γ-phosphate group of ATP is swung out and away from the active site creating space for the phosphate group of Thz-P. The Thz-P binding is very similar to the Thz binding (Figure 5). In addition to the interactions previously described for Thz, additional interactions are observed between the phosphate group of Thz-P and the backbone atoms of Gly195 and Ser198. The amide group of Gly195 donates a hydrogen bond to one phosphate oxygen atom and the amide group of Ser198 donates a hydrogen bond to a different phosphate oxygen atom. Residues 195−198 are located at the N-terminus of α9; therefore, a helix

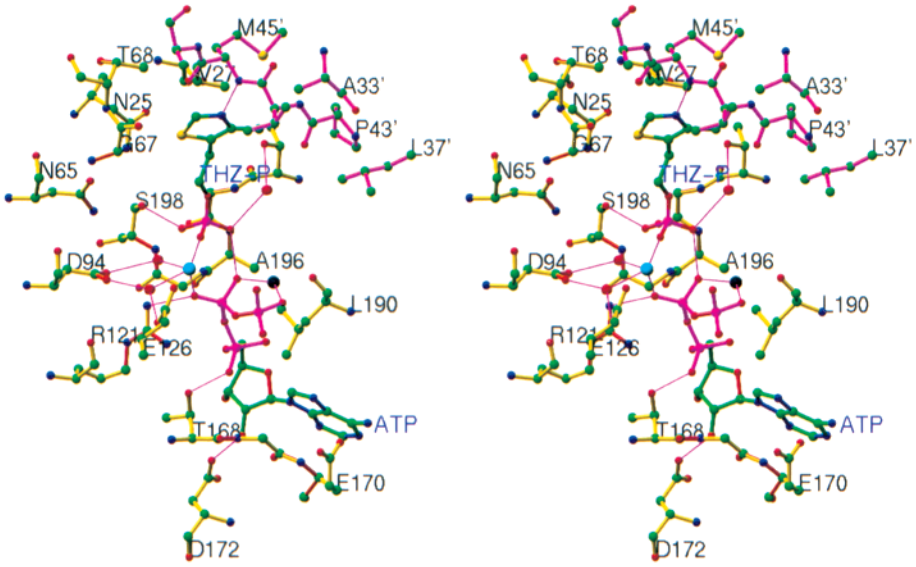


FIGURE 5: Stereoview of the active site in the ThiK C198S mutant/ATP/Thz-P complex structure. Different monomers are distinguished by bond color. Water molecules are colored as red spheres; The active site Mg^{2+} ion is colored as a cyan sphere. The additional Mg^{2+} ion is colored as a black sphere. The figure was prepared with RIBBONS (*31*).

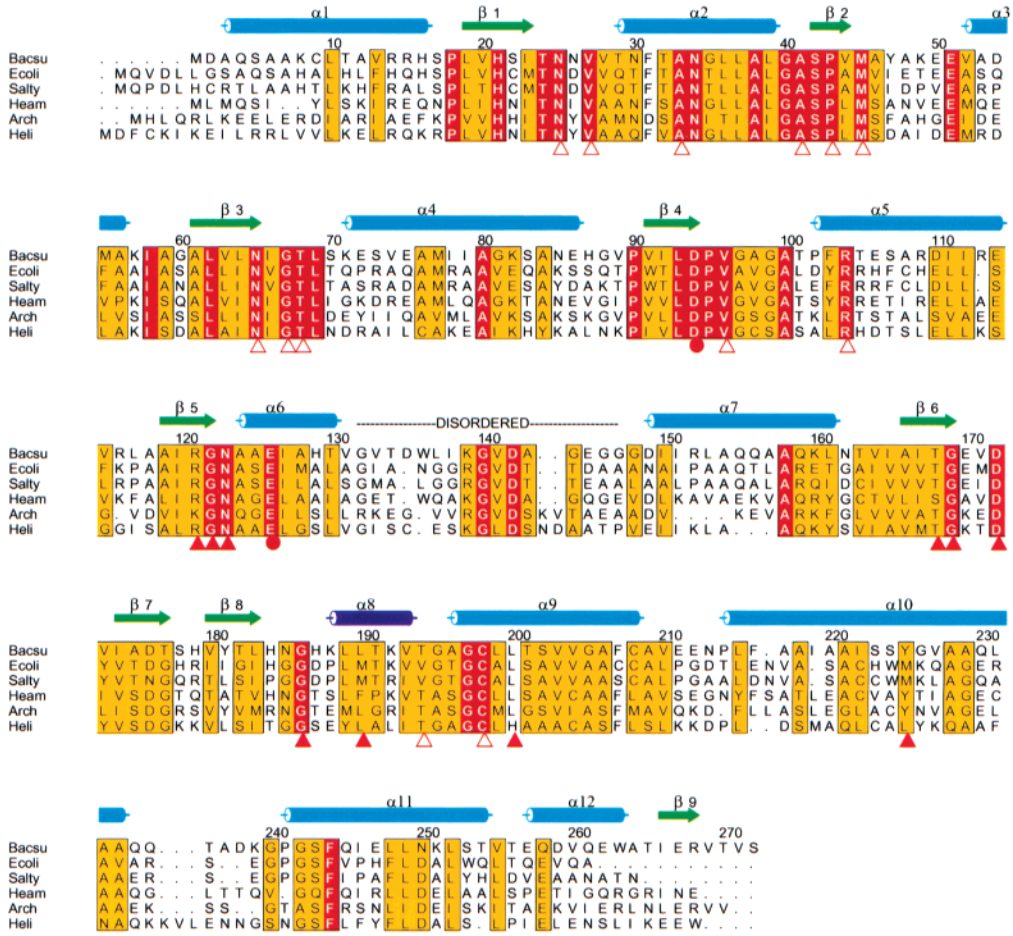


FIGURE 6: Sequence comparisons of ThiK from different organisms. Numbering follows the sequence of ThiK from *B. subtilis* for which secondary structures are shown as cylinders (helices) and arrows (strands). The blue cylinder is a 3_{10} helix. Identities among all ThiK sequences are shown with red background, positions at which amino acids with similar physical-chemical properties are conserved are shown with yellow background. Open red triangles are Thz binding residues; closed triangles are ATP binding residues; closed circles are residues that bind to Mg^{2+}. GenBank accession numbers for the various ThiK sequences are *Bacillus subtilis* (X73124), *Escherichia coli* (D90848), *Salmonella typhimurium* (U87940), *Haemophilus influenzae* (U32725), *Archaeoglobus fulgidus* (AE000960), and *Helicobacter pylori* (AE000595). The figure was prepared with ALSCRIPT (*32*).

dipole is also present to interact with the phosphate.

ATP-Binding Site. The ATP molecule stretches out along the C-terminal edge of the central β-sheet, closest to stands β5, β6, β7, and β8 (Figure 4). The adenine base is stacked between two loops. One loop contains the β-turn connecting strands β6 and β7 (residues 169−170) and the other loop contains residues 186−188 located before the 3_{10}-helix 8. There are no hydrogen bonds between the adenine base and the protein. The edge of the base is solvent exposed with hydrogen bonds between N1 and N6 and separate water molecules. The ribose moiety is in the C2′-endo conformation. The O2′ atom of the ribose moiety forms a bifurcated hydrogen bond with the two side-chain oxygen atoms of Asp172. The glycosidic torsion angle is in a typical anti conformation with a value of −128° (O4′−C1′−N9−C4).

One of the oxygen atoms of the α-phosphate is hydrogen bonded to the hydroxyl oxygen atom of Thr168. Arg121, a highly conserved residue, is hydrogen bonded to an oxygen atom of the β-phosphate. There are two Mg^{2+} ions in the ATP-binding site. Both of the Mg^{2+} ions interact with oxygen atoms of the β-phosphate. One Mg^{2+} ion also interacts with the phosphate of the product, Thz-P. In addition, this magnesium ion interacts through water molecules with two strictly conserved residues, Asp94 and Glu126. The second Mg^{2+} is bound to oxygen atoms of both the β- and γ-phosphates. However, there are no interactions with ThiK side-chain atoms. The significance of the second Mg^{2+} ion is unclear because the hybrid complex of product Thz-P and reactant ATP is not chemically relevant.

Activity of ThiK Mutants. Three ThiK mutants, C198D, C198S, and C198A, were prepared using the purification procedure described for the native enzyme. The specific activities of these mutants were 23, 0.5, and 1 μmol/mg/min, respectively. Under these assay conditions, the activity of the native enzyme was 2.6 μmol/mg/min.

DISCUSSION

Comparison with Ribokinase and Adenosine Kinase. The structure of ThiK is the third example within a recently described family of small molecule kinases (*23*, *24*). Sequence homology searches using the BLAST program found eight ThiK sequences but failed to identify any other kinases that are similar to ThiK in sequence. Structural similarity to adenosine kinase and ribokinase was found by using the DALI server (*25*), which searches for similarity based on the three-dimensional fold of a protein. The overall folds of ThiK, ribokinase, and adenosine kinase are similar,

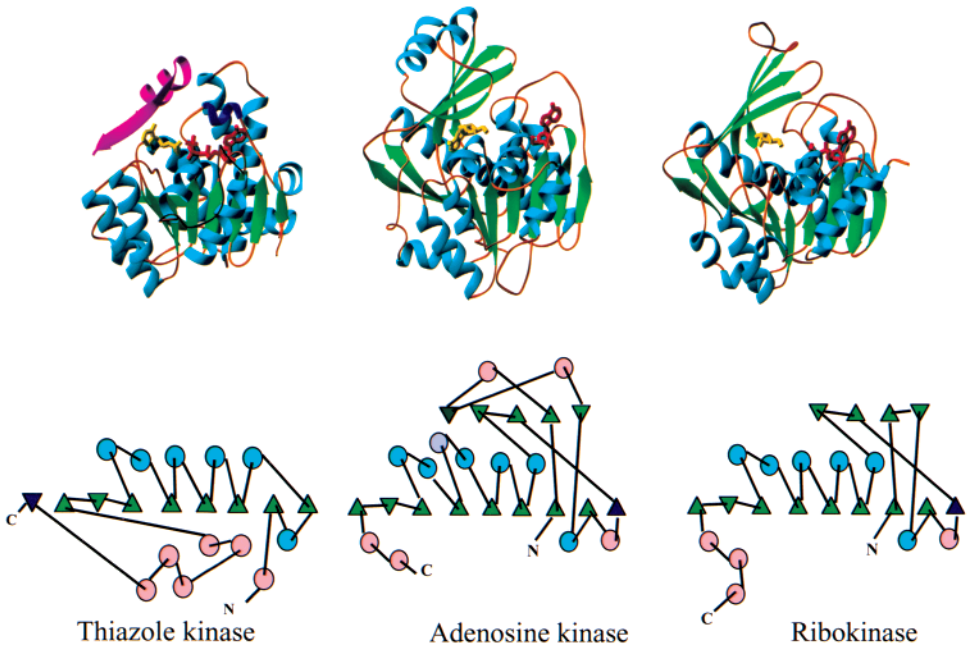


FIGURE 7: Ribbon drawings and topological comparison of ThiK, adenosine kinase, and ribokinase. Conserved secondary structural elements are colored similarly with green triangles for strands and cyan circles for helices. The substrates are colored yellow and the ATP molecules are colored red. The figure was prepared with RIBBONS (*31*).

each containing a central β-sheet flanked by several α-helices (Figure 7). Eight of the nine β-strands and six of the 12 helices in ThiK have corresponding secondary structural elements in ribokinase (*23*) and adenosine kinase (*24*). Of the 272 amino acids in ThiK, 203 of them can be superimposed on structurally equivalent residues in ribokinase with an RMSD of 3.0 Å and 199 amino acids of ThiK can be superimposed on to adenosine kinase with an RMSD of 2.9 Å. Both ribokinase and adenosine kinase have additional small β-sheets of four and five strands, respectively. In each of these structures, the small β-sheet forms a flap over the active-site region. Interestingly, a similar function to the small β-sheets in ribokinase and adenosine kinase is performed by the adjacent subunit in the ThiK trimer. When one subunit of the adenosine kinase monomer, the ribokinase dimer, and the ThiK timer is superimposed, the ATP-binding sites superimpose closely and the residues of the β-sheet flaps in ribokinase and adenosine kinase correspond to residues from the adjacent subunit of ThiK (Figure 7).

ATP-Binding Site. The ATP-binding site of ThiK has high structural similarity to that of ribokinase (*23*) and adenosine kinase (*24*). This binding fold with one mixed β-sheet is different from the commonly observed parallel five-stranded β-sheet arrangement observed in other kinase structures (*26*−*29*). In ribokinase, the ATP-binding site is located in a shallow groove of the larger domain with the adenine base stacked between the β-turn connecting strands β11 and β12 and the side chain of Val286 from α9 (*23*). In ThiK, the adenine base is stacked between the β-turn connecting the strands β6 and β7 and the main-chain atoms of residues 187 and 188. In ThiK, the ATP molecule is exposed to the solvent region, and there are no direct hydrogen-bonding interactions between the adenine base and the protein. The adenine base of ADP in ribokinase also lacks any direct hydrogen-bonding interaction with the protein atoms. However, the adenine base present in the ATP-binding region of adenosine kinase forms two hydrogen bonds between N1 and N6 of adenosine and OE1 and NE2 of a glutamine residue, respectively.

The ribose moiety assumes a C2′-endo puckering in ThiK, in contrast to the C3′-endo puckering observed in ribokinase (*23*) and adenosine kinase (*24*). This difference may result from a bifurcated hydrogen bond between O2′ and the two oxygen atoms of Asp172 in ThiK. In ThiK, one of the oxygen atoms of the α-phosphate is hydrogen bonded to the hydroxyl oxygen of Thr168. This residue corresponds to Thr223 and Thr265 in ribokinase and adenosine kinase, respectively. In ThiK, one of the oxygen atoms of the β-phosphate forms a hydrogen bond to NH2 of Arg121: however, there is no corresponding Arg or Lys in the ATP-binding site of ribokinase and adenosine kinase. In ThiK, a γ-phosphate oxygen atom makes a hydrogen bond with ND2 of the Asn123. The corresponding residue in ribokinase (Asn187) forms a hydrogen bond to one of the β-phosphate oxygen atoms.

The Mg^{2+} site in ThiK is 1.3 Å from the Mg^{2+} ion in the catalytic site of adenosine kinase when the two subunits are superimposed. In ThiK, the Mg^{2+} ion is coordinated by Asp194 and Glu126. The corresponding residues in adenosine kinase (*24*) are Asn196 and Glu226 and in ribokinase (*23*) are Asn166 and Glu190. In adenosine kinase, Asp300, which is proposed to serve as the catalytic base, also coordinates the Mg^{2+} ion. This residue corresponds to Cys198 in ThiK.

Substrate Binding. In contrast to the ATP-binding site, which is similar for ThiK, ribokinase and adenosine kinase, the substrate-binding site for ThiK is considerably different from that of the other two enzymes. In both ribokinase (*23*) and adenosine kinase (*24*) the ribose- and adenosine-binding sites are each located between two domains within the same subunit (Table 3). It is not surprising that adenosine kinase and ribokinase have similar substrate-binding sites, because in both cases, it is the 5′-hydroxyl group of ribose that is phosphorylated. Likewise, it is not surprising that the Thz-binding site in ThiK is different from the alcohol substrate-

Table 3: Residues Involved in Nucleotide Binding

ThiK	ribokinase (*23*)	adenosine kinase (*24*)	function in ThiK
Asp94	Asn166	Asn196	Mg^{2+} via water
Arg121	Thr185	Phe221	β-phosphate
Asn123	Asn187	Asn223	γ-phosphate
Glu126	Glu190	Glu226	Mg^{2+} via water
Thr168	Thr223	Thr265	α-phosphate
Gly169	Gly225	Gly267	adenine
Glu170	Ser226	Arg268	adenine
Leu190	Val286	Ile331	ribose
Thr191	Thr250	Thr295	adenine
Gly195	Ala252	Gly297	flexible loop
Ala196	Ala253	Ala298	flexible loop
Gly197	Gly254	Gly299	γ-phosphate
Cys198	Asp255	Asp300	active site base
Leu200	Phe257	Phe302	ribose

binding site of ribokinase and adenosine kinase. The common feature among all three substrates is that a primary alcohol functional group of a small molecule is the site of phosphorylation. What is remarkable is that on one hand ribokinase and adenosine kinase have evolved a flap formed from a second domain to shield the active site whereas on the other hand ThiK has utilized oligimerization of subunits to shield the active site.

An interated BLAST search using PSI−BLAST (*30*) also detected a relationship between the ThiKs, the ribokinases, and the adenosine kinases. In addition, the PSI−BLAST search revealed that ThiD, an enzyme in the thiamin biosynthetic pathway responsible for the sequential addition of two phosphate groups to the hydroxymethylpyrimidine (**9**, Scheme 1), is also probably a member of this family of small molecule kinases. X-ray crystallographic studies on ThiD are currently underway in this laboratory.

Mechanism of ThiK. Using the structure of the enzyme/ATP/Thz-P complex, it was possible to construct a model for the enzyme/Thz/ATP complex by removing the phosphate from the thiazole moiety, rotating the γ-phosphate of ATP into the active site and removing the magnesium ion that was coordinated to the ATP but not to the protein. The interactions between the enzyme and its substrates that are likely to be important for catalysis of the phosphate transfer reaction can be identified from this model and are shown in Figures 8 and 9.

The thiazole alcolol is positioned in the active site for an in-line displacement of ADP. The nucleophilicity of the alcohol may be enhanced by a proton relay system consisting of Cys198 and a magnesium bound water. The electrophilicity of the γ-phosphate of ATP is enhanced by coordination of one of the γ-phosphate oxygen atoms to the magnesium ion and by hydrogen bonding between a second phosphate oxygen and the amide protons of both Gly197 and Cys198. In addition, ADP is stabilized as a leaving group by coordination of the β-phosphate to the magnesium ion and by an electrostatic interaction with Arg121.

ThiK, ribokinase (*23*) and adenosine kinase (*24*) all utilize magnesium ion coordination and an oxyanion hole to activate the γ-phosphate of ATP for phosphate transfer. However, ribokinase and adenosine kinase do not have the equivalent of the electrostatic interaction between Arg121, and the β-phosphate found in ThiK and Asp94 of ThiK is replaced by Gln196 in adenosine kinase and Gln166 in ribokinase.

The most interesting difference between the three kinases lies in the apparent use of Cys198 as the base involved in the alcohol deprotonation by ThiK. Both ribokinase and adenosine kinase utilize conserved aspartate residues for this function (Asp255 and Asp300, respectively). In general, alcohol kinases utilize either an aspartate or a glutamate residue as the activating base and to the best of our knowledge, a cysteine/water/magnesium ion triad has not been observed in other alcohol kinases. We found that replacement of Cys198 with aspartate gave a mutant enzyme with specific activity 9-fold higher than the native enzyme. This surprising result raises the question as to why ThiK would use cysteine as the base when all of the other members of the ribokinase family utilize aspartate. One possible explanation is that base catalysis by cysteine is not critical for the phosphate transfer reaction. Two additional mutants were prepared to address this.

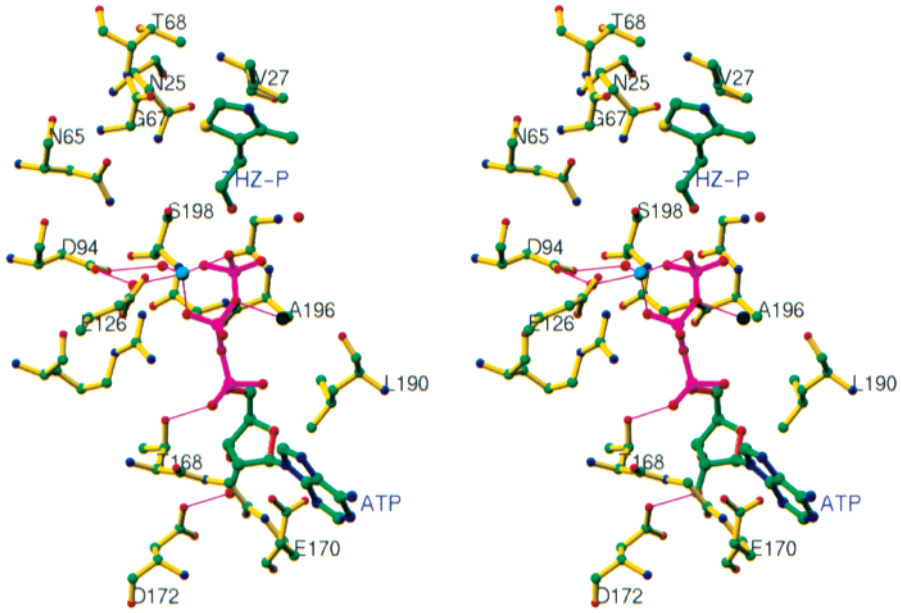


FIGURE 8: Stereoview of a modeled substrate structure. The phosphate group of Thz-P was removed and the gamma phosphate of ATP was rotated into place to display the structure before the reaction would occur. The figure was prepared with RIBBONS (*31*).

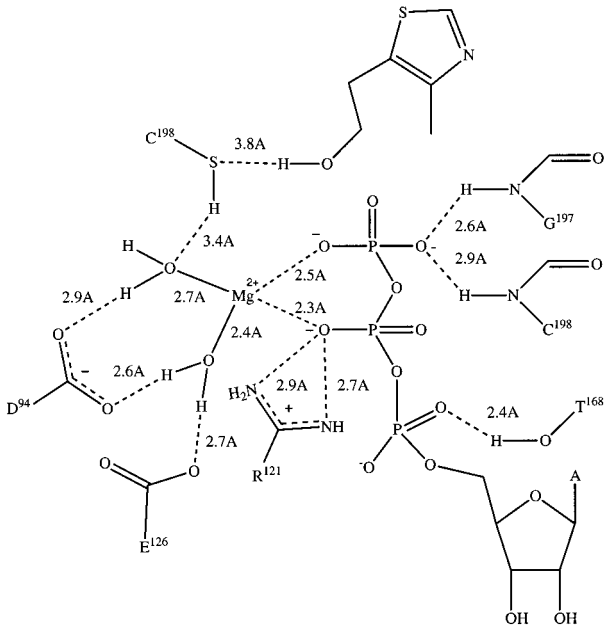


FIGURE 9: Schematic representation of the active site interactions of ThiK proposed to play a role in the catalysis of the phospho transfer reaction.

The specific activity of the C198S mutant was 20% that of the native enzyme. This suggests that Cys198 is not deprotonated in its catalytically active form because Ser has a much higher pK_a than Cys (14 vs 9). Since Ser is a very weak base, the high residual activity of the C198S mutant casts additional doubt on the importance of the Cys residue. This doubt was confirmed by analysis of the C198A mutant. The specific activity of this mutant was 40% that of the native enzyme. This suggests that an alternative base must exist. Inspection of the kinase/Thz/ATP model indicates that there are no basic amino acid side chains close to the thiazole alcohol. However, the γ-phosphate oxygen atoms of ATP are 3.1, 2.5, and 3.4 Å from the thiazole oxygen and we propose that one of these may function as the alcohol activating base.

ACKNOWLEDGMENT

The authors wish to thank Yi Zhang and Cynthia Kinsland for providing the ThiK overexpression plasmid and Jonothan Melnick for determining the K_m and k_{cat} parameters for the native enzyme. We would like to acknowledge support of the W.M. Keck Laboratory for Molecular Structure and the Lucille P. Markey Charitable Trust.

REFERENCES

1. Schowen, R. L. (1998) in *Comprehensive Biological Catalysis* (Sinnott, M., Ed.) Vol. 2, pp 217−266, Academic Press, San Diego.
2. Burdick, D. (1998) in *Kirk Othmer Encyclopedia of Chemical Technology*(Howe-Grant, M. Ed.) Vol. 25, pp 152−171, Wiley, New York.
3. Begley, T. P., Downs, D., Ealick, S., McLafferty, F., Van Loon, D., Taylor, S., Campobasso, N., Chiu, J., Kinsland, C., Reddick, J., and Xi, J. (1999) *Arch. Microbiol. 171*, 293−300.
4. Zhang, Y., Taylor, S. V., Chiu, H.-J., and Begley, T. P. (1997) *J. Bacteriol. 179*, 3030−3035.
5. Petersen, L., and Downs, D. M. (1997) *J. Bacteriol. 179*, 4894−4900.
6. Nosaka, K., Nishimura, H., Kawasaki, Y., Tsujihara, T., and Iwashima, A. (1994) *J. Biol. Chem. 269*, 30510−30516.
7. Mathews, B. W. (1968) *J. Mol. Biol. 33*, 499−501.
8. Stura, E. A., and Wilson, I. A. (1990) *Methods 1*, 38.
9. Reddick, J. J., Kinsland, C., Nicewonger, R., Christian, T., Downs, D. M., Winkler, M. E., and Begley, T. P. (1998) *Tetrahedron 54*, 15983−15991.
10. Stanton, M., Philips, W. C., O'Mara, D., Naday, I., and Westbrook, E. (1993) *Nucl. Instrum. Methods Phys. Res. A 325*, 558−567.
11. Otwinowski, Z., and Minor, W. (1997) in *Methods Enzymol.* (Sweet, R. M., and Carter, C. W., Eds.) Vol. 276, pp 307−326, Academic Press.
12. Collaborative Computing Project, No. 4. (1994) *Acta Crystallogr., Sect. D 50*, 760−763.
13. Otwinowski, Z. (1991) in *Isomorphous Replacement and Anomalous Scattering* (Wolf, W., Evans, P. R., and Leslie, A. G. W., Eds.) pp 80−88, SERC Proceedings, Daresbury Laboratories, Warrington, U.K.
14. Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, Vol. *31*, pp 34−38.
15. 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..
16. Jones, T. A., Zou, J. Y., Cowtan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A 47*, 110−119.
17. Engh, R. A., and Huber, R. (1991) *Acta Crystallogr., Sect. A 47*, 392−400.
18. Brünger, A. T. (1992) *X-PLOR Manual*, Version 3.1, Yale University Press, New Haven.
19. Jiang, J.-S., and Brünger, A. T. (1994) *J. Mol. Biol. 243*, 100−115.
20. Brünger A. T., Adams P. D., Clore G. M., Delano W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr., Sect. D 54*, 905−921.
21. Kleywegt, G. J., and Jones, T. A. (1998) *Acta Crystallogr., Sect. D 54*, 1119−1131.
22. Luzzati, P. V. (1952) *Acta Crystallogr. 5*, 802−810.
23. Sigrell, J. A., Cameron, A. D., Lones, T. A., and Mowbary, S. L. (1998) *Structure 6*, 183−193.
24. Mathews, I. I., Erion, M. D., and Ealick, S. E. (1998) *Biochemistry 37*, 15607−15620.
25. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
26. Wild, K., and Bohner, T. (1997) *Protein Sci. 6*, 2097−2106.
27. Abele, U., and Schulz, G. E. (1995) *Protein Sci. 4*, 1262−1271.
28. Stehle, T., and Schulz, G. E. (1992) *J. Mol. Biol. 224*, 1127−1141.
29. Webb, P. A., Perisic, O., Mendola, C. E., Backer, J. M., and Williams, R. L. (1995*) J. Mol. Biol. 251*, 574−587.
30. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
31. Carson, M. (1997) in *Methods in Enzymol.* (Sweet, R. M., and Carter, C. W., Eds.) Vol. 277, pp 493−505, Academic Press, New York.
32. Barton, G. J. (1993) *Protein Eng. 6*, 37−40.

BI0000061