X-ray Structures of Threonine Aldolase Complexes: Structural Basis of Substrate Recognition[†,‡]

Clara L. Kielkopf[§,||] and Stephen K. Burley*[,§,⊥,#]

Laboratories of Molecular Biophysics and Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021

Received May 31, 2002; Revised Manuscript Received July 22, 2002

ABSTRACT: L-Threonine acetaldehyde-lyase (threonine aldolase, TA) is a pyridoxal-5′-phosphate-dependent (PLP) enzyme that catalyzes conversion of L-threonine or L-*allo*-threonine to glycine and acetaldehyde in a secondary glycine biosynthetic pathway. X-ray structures of *Thermatoga maritima* TA have been determined as the apo-enzyme at 1.8 Å resolution and bound to substrate L-*allo*-threonine and product glycine at 1.9 and 2.0 Å resolution, respectively. Despite low pairwise sequence identities, TA is a member of aspartate aminotransferase (AATase) fold family of PLP enzymes. The enzyme forms a *222* homotetramer with the PLP cofactor bound via a Schiff-base linkage to Lys199 within a domain interface. The structure reveals bound calcium and chloride ions that appear to contribute to catalysis and oligomerization, respectively. Although L-threonine and L-*allo*-threonine are substrates for *T. maritima* TA, enzymatic assays revealed a strong preference for L-*allo*-threonine. Structures of the external aldimines with substrate/product reveal a pair of histidines that may provide flexibility in substrate recognition. Variation in the threonine binding pocket may explain preferences for L-*allo*-threonine versus L-threonine among TA family members.

Glycine performs a central biological role, both as a macromolecular building block and as a neurotransmitter in mammals. Most organisms synthesize glycine using SHMT,[1] which catalyzes formation of glycine from serine (*1*). In some cases, the SHMT pathway is supplemented by a second enzyme, threonine aldolase (TA), which is required by yeast for glycine auxotrophy (*2*). TA activity has been observed in various bacteria (*3*−*5*), fungi (*6*, *7*), and mammals (*8*). In addition to these organisms, putative homologues have been identified in nematodes, flies, and plants (Figure 1).

TA (EC 4.1.2.5, L-threonine acetaldehyde-lyase) is a pyridoxal-5′-phosphate (PLP) enzyme that catalyzes interconversion of threonine to glycine and acetaldehyde (Figure 2). High pairwise sequence identities among subclasses of L-threonine aldolases (27−52%) make them ideal targets for studying structural and functional relationships underpinning enzyme action. A number of TAs differing in stereospecificity at the threonine α- or *β*-carbon have been isolated and characterized. These variants include: (i) L-*allo*-threonine aldolase, (ii) L-threonine aldolase, (iii) low-specificity L-threonine/L-*allo*-threonine aldolase, plus corresponding classes for D-threonine/D-*allo*-threonine. The *T. maritima* enzyme described in this paper is a member of the low-specificity subfamily of L-threonine/L-*allo*-threonine aldolases, providing an opportunity for analyzing substrate preferences in some detail.

Although PLP-containing enzymes catalyze a wide variety of chemical reactions (transamination, racemization, decarboxylation, elimination reactions, and aldol cleavage), they probably share the basic reaction chemistry. As seen with other PLP enzymes, the PLP cofactor of TA is initially bound via a Schiff-base linkage to a conserved active site lysine (*4*, *5*) (shown as structure I in Figure 2). The reaction mechanisms for TA and SHMT are thought to be quite similar (*9*). Most experimental evidence suggests that enzyme catalysis begins with the threonine substrate forming an external aldimine with the PLP (II, Figure 2). Next, a retroaldol cleavage of the threonine aldimine produces acetaldehyde and the PLP−glycine quinonoid complex (III, Figure 2). Following protonation of the glycine α-carbon (IV, Figure 2), glycine is released, regenerating the internal Schiff-base with the active site lysine.

[†] This work was supported by the National Institutes of Health (P50 GM62529) and the Rockefeller University. S.K.B. is an Investigator of the Howard Hughes Medical Institute. C.L.K. is supported by an American Cancer Society Postdoctoral Research Fellowship.

[‡] Protein DataBank codes for reported structures: native, 1M6S; L-*allo*-threonine complex, 1LW4; glycine complex, 1LW5.

* Corresponding author. E-mail: stephen_burley@stromix.com. Phone: 858-558-4850, ext. 1263; Facsimile: 858-558-3402.

[§] Laboratories of Molecular Biophysics, The Rockefeller University, New York, NY 10021.

[||] Present address: Department of Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205.

[⊥] Howard Hughes Medical Institute.

[#] Present address: Structural GenomiX, Inc., 10505 Roselle St., San Diego, CA 92121.

[1] Abbreviations: apo-TA, TA internal aldimine; AATase, aspartate aminotransferase; Gly-TA, TA with bound glycine; HEPES, sodium *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; PDB, Protein DataBank; PLP, pyridoxal-5′-phosphate; MAD, multiwavelength anomalous dispersion; rmsd, root-mean-square deviations; NADH, nicotinamide adenine dinucleotide; SHMT, serine hydroxymethyltransferase; TA, threonine aldolase; TCEP, tris(carboxyethyl phosphine); THF, 5-formyltetrahydrofolate; Thr-TA, TA with bound threonine.

10.1021/bi020393+ CCC: $22.00
Published on Web 09/04/2002

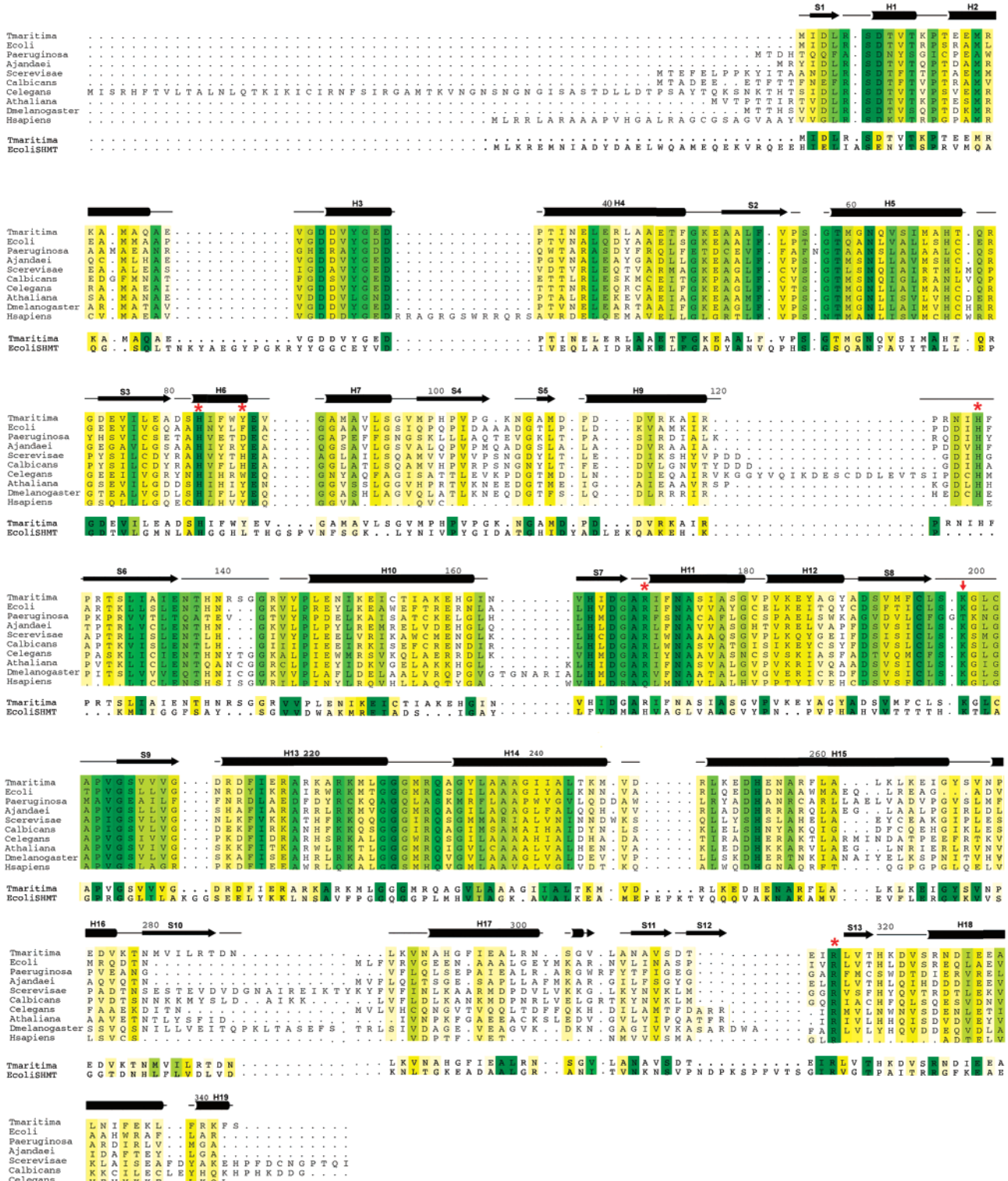

FIGURE 1: Sequence alignments of selected known and putative TAs, and structure-based alignment with *E. coli* SHMT. Sequence numbering corresponds to *T. maritima* TA (GenBank accession code gi4982321), with residues that contact the substrate/product (His83, Tyr87, *His125*, Arg171, Arg316) indicated by an asterisk. Lys199, which forms a Schiff-base with PLP, is identified with an arrow. Secondary structural elements are denoted as follows: α-helices, cylinders; *β*-strands, arrows; random coil, lines; gray circles, regions that were poorly defined in electron density maps. Sequence similarity is color-coded using a gradient from dark green (100% sequence identity) to white (<30% sequence identity).

Despite low sequence identities, PLP enzymes have been classified into families sharing similar three-dimensional folds (*10*). The amino acid sequence similarities of TAs (<15% pairwise identities) are below the critical threshold for successful structure modeling (30−35% identity), making them excellent structural genomic targets (*22*).

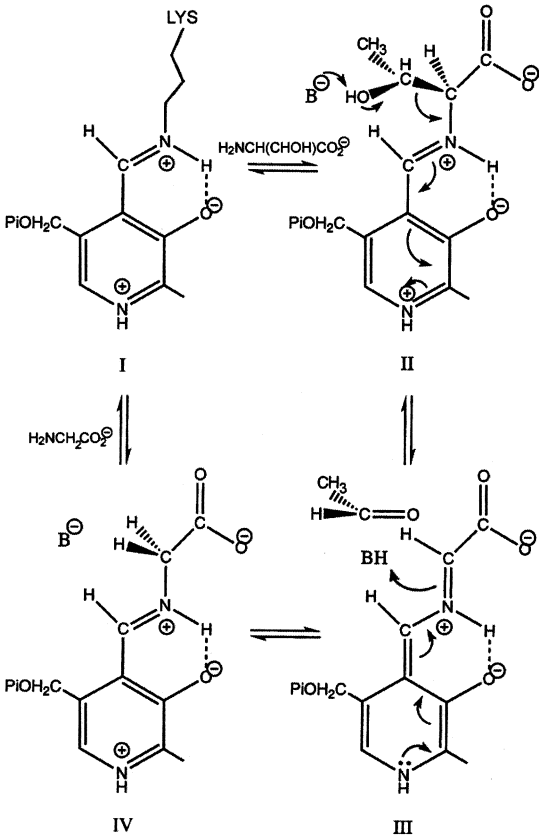

FIGURE 2: Reaction mechanism for conversion of L-threonine or L-*allo*-threonine to glycine and acetaldehyde.

Here, we describe the X-ray structures of L-threonine/L-*allo*-threonine aldolase from *T. maritima*, as the internal aldimine (apo-TA), and external aldimines with glycine (Gly-TA) or L-*allo*-threonine, converted by the enzyme from glycine plus acetaldehyde (Thr-TA). Enzyme assays established the activity and stereospecificity of the enzyme. These structural and functional studies expand our knowledge of a poorly characterized portion of protein sequence/structure space, and address fundamental questions of L-threonine aldolase mechanism and stereospecificity.

EXPERIMENTAL PROCEDURES

Protein Preparation and Crystallization. The gene encoding TA (GenBank code: gi4982321) was amplified by conventional polymerase chain reaction from *T. maritima* genomic DNA. Recombinant protein was expressed as an N-terminal hexaHis fusion protein in *E. coli* BL21(DE3) and purified to homogeneity using Ni-ion affinity, followed by anion-exchange and gel filtration chromatography after proteolytic affinity tag removal. Se-Met protein was expressed and purified using a similar protocol.

Orthorhombic crystals were obtained via hanging drop vapor diffusion at room temperature against 30% (v/v) PEG400, 200 mM $CaOAc_2$, 0.1 M HEPES, pH 7.5, and 0.5 mM TCEP [space group $P2_12_12_1$; $a = 96.0$, $b = 100.7$, $c = 150.2$ Å; four molecules per asymmetric unit]. Isomorphous Gly-TA cocrystals were grown in the presence of 20 mM glycine. To obtain the Thr-TA complex, glycine cocrystals were soaked for approximately 1 h in a crystallization solution supplemented with 3% (v/v) acetaldehyde.

Data Collection, Structure Determination, and Refinement. Se-Met MAD diffraction data (*11*) were obtained from cryogenically cooled crystals at three X-ray wavelengths in the vicinity of the Se K-absorption edge at the Advanced Photon Source, and processed with Denzo/Scalepack (*12*). Preliminary phase information was obtained using 20 Se sites located with SnB (*13*) and refined using MLPHARE (*14*). Anomalous difference Fourier syntheses revealed additional sites, giving an interpretable electron density map. Following density modification and noncrystallographic symmetry averaging with DM (*15*), automated model building with ARP/wARP (*16*) yielded 1367 of 1388 residues. The remaining residues were added manually, and the apo-TA structure was refined with CNS (*17*) at 1.8 Å resolution. Bound glycine and L-*allo*-threonine (resulting from catalysis within the acetaldehyde-soaked glycine cocrystal) were readily identified in the active site using difference electron density maps. Gly-TA and Thr-TA cocrystal structures were refined against data obtained from SeMet crystals to 2.0 and 1.9 Å resolution, respectively. Structure determination and refinement statistics are given in Table 1.

Enzyme Assays. A coupled enzyme assay adapted from a previously described method (*4*) was used to measure *T. maritima* TA activity with L-threonine and L-*allo*-threonine substrates. Reactions were measured at 22 °C in 5 mM pyridoxal-5′-phosphate, 100 mM NaCl, 2 mM $MgCl_2$, 50 mM HEPES, pH 7.5, and 0.2 mM NADH, with 100 units of yeast alcohol dehydrogenase plus appropriate amounts of enzyme and substrate. Reduction of acetaldehyde, coupled with oxidation of NADH by yeast alcohol dehydrogenase, was monitored spectrophotometrically at 340 nm. All measurements were made in triplicate.

RESULTS

Structural Overview. The crystallographic asymmetric unit contains four TA monomers that form a prolate tetramer with maximal dimensions 90 Å × 70 Å × 50 Å. Each monomer is composed of two α/β domains, with the PLP cofactor anchored via a Schiff-base linkage to Lys199 at the inter-domain cleft. The larger domain forms a seven-stranded α/β structure [residues 14−252; α1-α2-β1-α3-β2-α4-β3-α5-β4-α6-β5-α7-α8-β6-β7-α9-α10] at the core of the tetramer, with the smaller domain [residues 5−13 and 253−347; (β1)-α11-β9-α12-β10-β11-α13] at the periphery (Figure 3). The tetramer is very nearly symmetrical, with α-carbon root-mean-square deviations (rmsd) between tetramer subunits of 0.6−0.8 Å both in the presence and in the absence of ligands.

Given the similarities in spectral and catalytic properties, TAs were thought to belong to the same protein fold family of PLP enzymes as the SHMTs (*18*). Comparison of our TA structure against the Protein DataBank (www.rcsb.org/pdb) using the DALI server (*19*) detected similarity with other PLP-dependent lyases that support α,β-elimination, including aspartate aminotransferase [AATase, (*20*); PDB code 1BJW; *Z*-score 27.6; rmsd = 3.6 Å; sequence identity 14% for 313 α-carbon pairs] and SHMT [(*21*); PDB code 1BJ4; *Z*-score 25.0; rmsd 3.3 Å; sequence identity 14% for 300 α-carbon pairs]. Thus, TAs are type I PLP enzymes (*10*), of which both SHMT and AATase are particularly well characterized.

Homology Modeling and Sequence Comparisons. Homology modeling with MODWEB (*22*) demonstrated that *T.*

Table 1: Crystallographic Statistics

	Data Collection						
data set	inflection	peak 1	peak 2	remote	apo-TA	Gly-TA	Thr-TA
wavelength (Å)	0.9796	0.9794	0.9794	0.9686	0.9795	0.9690	0.9690
resolution (Å)	20.0−2.0	20.0−2.0	20.0−2.5	20.0−2.0	20.0−1.8	20.0−2.0	20.0−1.9
no. of reflections observed/unique	529752/ 168702	546169/ 172529	688141/ 96461	528208/ 169507	1036724/ 132473	903510/ 86689	1191136/ 220635
completeness overall/outer shell (%)	88.6/89.0	90.6/88.5	98.9/99.6	89.1/91.1	98.5/99.4	95.1/91.3	97.9/99.8
R_{merge} [a] overall/outer shell (%)	7.2/37.6	7.6/38.7	6.8/11.9	7.2/35.8	6.3/21.8	10.0/27.9	6.7/18.4
$\langle I/\sigma(I)\rangle$ overall/outer shell	11.7/1.9	11.3/1.7	23.7/9.4	12.0/2.1	27.2/7.8	21.4/5.1	18.0/5.6
mean FOM[b]	0.55 (20.0−2.0 Å resolution) for 88424 reflections						

		Refinement (Against All Data)		
		Apo-TA	Gly-TA	Thr-TA
number of atoms	protein	10239	10427	10355
	water	1255	883	1029
	PLP	60	75	75
	glycine	−	12	−
	threonine	−	−	24
	calcium ion	6	6	6
	chloride	6	6	6
R_{factor} [c] (%)	R_{work}	19.9	20.7	18.3
	R_{free}	21.3	23.7	20.6
rmsd	bond lengths (Å)	0.006	0.006	0.006
	bond angles (deg)	1.5	1.4	1.4
overall *G*-factor[d]		0.4	0.4	0.4

[a] $R_{merge} = \sum_{hkl}\sum_i|I(hkl)_i - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i\langle I(hkl)_i\rangle$. [b] Figure of merit calculated using MLPHARE (*14*). [c] $R_{work} = \sum_{hkl}|F_o(hkl) - F_c(hkl)|/\sum_{hkl}|F_o(hkl)|$, where F_o and F_c are observed and calculated structure factors, respectively. R_{free} is R_{work} for 7.0% of the reflections excluded from the refinement. [d] Computed with PROCHECK (*39*).

maritima TA belongs to a very large family of orthologous proteins. Although only 29 closely related protein sequences (≥30% sequence identity) could be modeled with the TA template, 2069 more distantly related sequences (≤30% sequence identity) gave informative homology models. [Models are publicly available from MODBASE (http://pipe.rockefeller.edu) via advanced search with the PDB code 1JG8.]

TAs represent a highly conserved enzyme family (Figure 1). Pairwise sequence identities with the *T. maritima* enzyme range from 25% to 48% among 'low-specificity' TAs with established activity (*A. jandaei*, *C. albicans*, *E. coli*, *P. aerugenosa*, *S. cerevisae*). Greater than 30% sequence identity was also detected with putative homologues from higher organisms (*A. thaliana*, *C. elegans*, *D. melanogaster*, *H. sapiens*). A high level of phylogenetic sequence conservation and excellent model scores [Modeler score = 1.00 (*22*)] document that all known TAs almost certainly share the same protein fold. Within the active site, the lysine responsible for Schiff-base formation with PLP is invariant. Outside the confines of the minimal tetramer defined by our structure, N- and C-terminal extensions present in related enzymes may extend the oligomerization surface or serve other, possibly regulatory, functions.

Quaternary Structure. T. maritima TA forms a homotetramer with *222* symmetry, which is consistent with gel filtration chromatography and dynamic light scattering results (data not shown). Other well-characterized TAs are also tetrameric (*3*, *5*, *6*), and it appears likely that all family members function as tetramers. The homooligomerization states of other AATase-type PLP enzymes include dimers, tetramers, and hexamers. For example, AATase and several other aminotransferases are homodimers, SHMT (*23*), tyrosine phenol-lyase (*24*), tryptophan indole-lyase (*25*), and cystathionine β-lyase (*26*) are tetramers, and ornithine aminotransferase is a hexamer.

For structurally characterized tetrameric and hexameric members of the AATase family, the interface of the AATase dimer (A/D and B/C subunits for TA) is conserved within the higher order oligomers. Likewise, the *T. maritima* TA tetramer is composed of two AATase-like dimers, with four distinct active sites. The so-called 'catalytic dimer' interfaces are more intimate than the less similar A/B, A/C, B/D, and C/D interfaces (32 Å average PLP C4−C4 distance A/D and B/C; versus 40 Å A/C and B/D; 50 Å A/B and C/D). The 'catalytic dimer' buries 3157 Å^2 of the solvent-accessible surface area, which is significantly larger than the average value (2300 Å^2) for other homodimers of comparable molecular weight (*27*). In contrast, the interface of *E. coli* SHMT dimer (*28*) is almost twice as large (6527 Å^2), with additional contributions from a 30-residue N-terminal arm and insertions in a loop near the interface (SHMT residues 45−71) (Figure 1).

Spatial organization of the catalytic dimers with respect to one another differs among the tetrameric members of the AATase fold family. The buried, solvent-accessible surface areas of the dimer−dimer interfaces of both TA and SHMT are less extensive (1935−2186 Å^2) than their respective catalytic dimers. Insertions between secondary structural elements in TA (residues 123−135) and SHMT (residues 193−197 and 242−248) appear to stabilize distinct dimer−dimer interfaces. These subtle differences suggest that modest

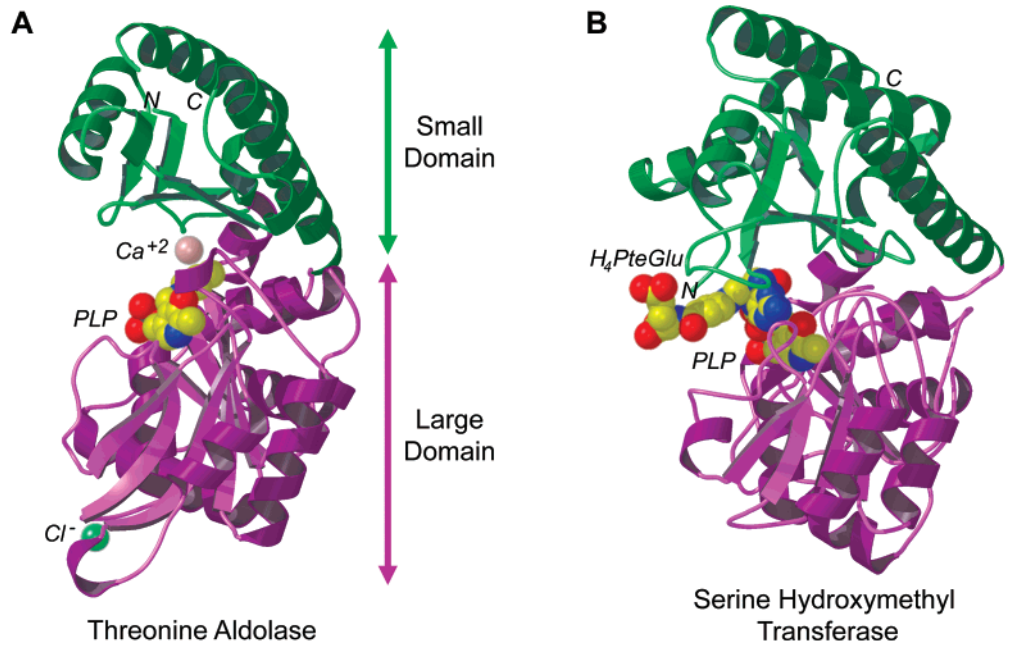

FIGURE 3: Overall structure. (A) Ribbon diagram of *T. maritima* TA with PLP cofactor and bound ions depicted as color-coded space-filled atoms (C, yellow; N, blue; O, red; Ca^{2+}, white; Cl^{-}, green). The small domain is shown in green and the large domain in purple. (B) For comparison, the structure of *E. coli* SHMT with PLP and THF cofactors is shown with a similar coloring scheme. Figures were prepared by use of Bobscript (*40*), Molscript (*41*), and Raster3D (*42, 43*).

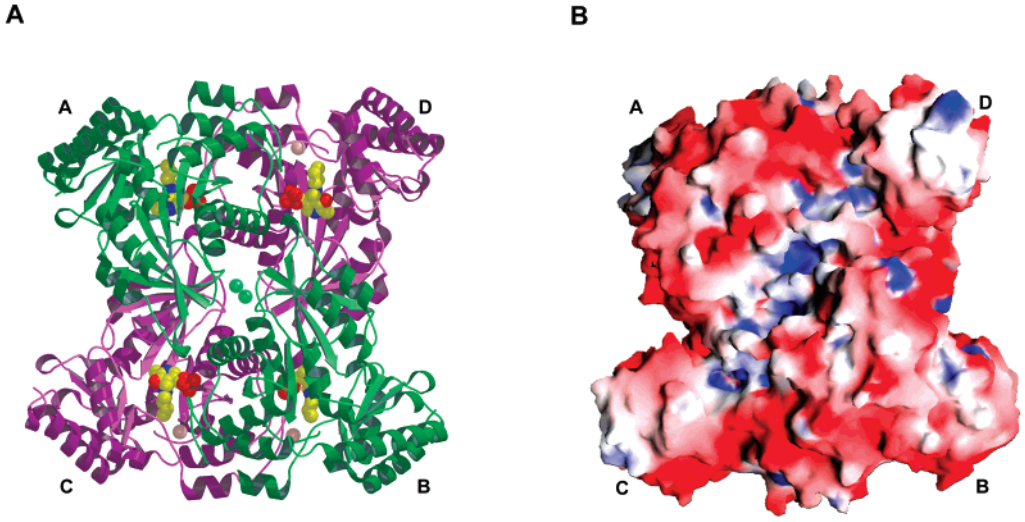

FIGURE 4: Tetramer organization. (A) Ribbon representation of the tetrameric enzyme, with monomers A and B colored in green, and monomers C and D colored in purple. Two putative chloride ions ligate the A/B and C/D subunit interfaces. (B) Solvent-accessible surface representation of the tetramer calculated using a water probe of radius 1.4 Å using the program GRASP (*44*). The surface is color-coded for electrostatic potential (color ramp from red to blue for <-10 to $>+10$ κBT, where κB is the Boltzmann constant and T is temperature).

substitutions in amino acid sequence can affect oligomerization properties. Indeed, SHMTs from different organisms form distinct tetramers (*21, 23, 28*). By analogy, sequence insertions observed among other TAs may give rise to oligomeric interfaces different from those observed for the *T. maritima* enzyme.

Despite a calculated neutral p*I*, the tetramer surface is negatively charged (Figure 4). Intersubunit surfaces are largely hydrophobic or involve hydrogen bonds between main chain atoms and water molecules. Moreover, the catalytic dimer interface is supplemented by several electrostatic interactions among side chains, for example, Arg5−Glu32, Thr8−Asp27, and Arg231−Ser198. Basic residues buried near the center of the tetramer provide binding sites for two putative chloride molecules that stabilize interactions among subunits (A/B and C/D).

Kinetic Parameters. As suggested by the high degree of sequence identity with well-characterized TAs and the presence of the PLP cofactor, coupled enzyme assays reveal that *T. maritima* TA catalyzes conversion of L-threonine or L-*allo*-threonine to acetaldehyde and glycine. The enzyme displays simple Michaelis−Menten kinetics with no apparent cooperativity between subunits (data not shown). Although compatible with either substrate, *T. maritima* TA displays a 25-fold preference for L-*allo*-threonine over L-threonine (Table 2). This L-*allo*-threonine preference is the strictest of

Table 2: Steady-State Kinetic Constants and Relative Specificities

substrate[a]	V_{max} (μM/min)	K_m (μM)
L-*allo*-threonine	27	31
L-threonine	27	752

source organism	specificity ratio $K_{m,L-allo-Thr}/K_{m,L-Thr}$	specificity residue[b]	reference
A. jandaei	infinite	Tyr	(*4*)
T. maritima	25	Tyr	this study
E. coli	15	Phe	(*3*)
S. cerevisiae	5	His	(*6*)
P. aeruginosa	0.8	Asp	(*5*)

[a] Kinetic constants for *T. maritima* TA. [b] Residue corresponding to *T. maritima* TA Tyr87.

the so-called 'low-specificity' TAs, which differ in their relative stereospecificity for the substrate β-carbon despite high pairwise sequence identities.

Comparison of Ternary Complexes. In our apo-TA structure, the PLP cofactor is anchored via a Schiff-base linkage to Lys199 (Figure 5A). The Gly-TA and Thr-TA cocrystal structures reveal either the external aldimine or the quinoid intermediates of the reaction pathway, in which glycine or L-*allo*-threonine replaces Lys199 at the C4′ atom of the PLP ring (Figure 5B,C). Although we cannot rule out larger interdomain movements in solution, bound ligands have little effect on the relative orientation of the small and large domains within each TA subunit. The overall conformations of the polypeptide chain in the presence of substrate or product are identical within the precision of the crystallographic method, displaying pairwise α-carbon rmsds of 0.1 and 0.2 Å for apo-TA versus Gly-TA and Thr-TA, respectively. Whereas polypeptide backbone atoms remain essentially unchanged in the presence of ligand, the plane of the PLP ring rotates 16° away from Lys199 (Figure 5D). The magnitude of this rotation is consistent with PLP rotations of ∼15° with respect to the internal aldimines, observed for the structures of SHMT with a reduced aldimine linkage (*23*), and external aldimine forms of SHMT (*29*) or AATase (*20, 30*).

Despite the apparent absence of positive or negative cooperativity, there is a striking asymmetry in substrate binding to individual subunits within the TA homotetramer. In both Gly-TA and Thr-TA, two of the active sites in distinct catalytic dimers appear to be fully occupied. The third active site is occupied by a mixture of the external and internal aldimines, and the fourth active site is entirely empty, showing only PLP bound to Lys199. It is remarkable that crystal structures of ternary complexes of SHMT with THF and either glycine or serine also display asymmetric ligand binding (*29, 31*). For SHMT, both halves of each catalytic dimer are either fully or weakly occupied, in contrast to equivalent asymmetric catalytic dimers observed for TA. Given that negative cooperativity was not observed during enzyme assays, it is possible that different environments within the crystal lattice dictate asymmetric ligand binding and reactivity within the tetramer.

Comparison of Gly-TA and Thr-TA revealed that the PLP cofactors and atoms common to glycine and L-*allo*-threonine display nearly identical conformations. In Gly-TA, two additional water molecules occur in the same locations as Cγ and Oγ2 of the bound L-*allo*-threonine side chain. A third water molecule positioned between the PLP phosphate and the threonine hydroxyl group is common to both Gly-TA and Thr-TA. Together, these three water molecules form a hydrogen-bond network along the exposed face of the bound glycine. In our apo-TA structure, the PLP ring is rotated toward Lys199, and five water molecules occupy the active site (two occupying the same location as the bound glycine plus three present in the Gly-TA cocrystal structure). One water molecule unique to apo-TA is sandwiched between Lys316 and PLP-O3, and the other forms a hydrogen bond with the first. Thus, binding of substrate/product displaces five ordered water molecules that recapitulate productive enzyme−ligand interactions in the apo-TA structure.

Substrate Recognition. The glycine/threonine ligands are bound in an orientation that resembles the ternary complexes of AATase and SHMT, with their amino groups slightly out of the plane of the PLP ring (Figure 5). For optimum reactivity, it is believed that the scissile bond of the external aldimine must be oriented perpendicular to the plane of the PLP ring. In the structure of Thr-TA, the Cα−Cβ bond is nearly orthogonal to the cofactor plane, forming an angle of 100°. This orientation may allow the enzyme to take advantage of stereoelectronic effects during catalysis (*32*).

As for SHMT, the ligand carboxylate groups interact with arginine and serine side chains (Arg171, Arg316, Ser6) that anchor substrate/product to the active site. The two arginines of TA provide a second positive charge to stabilize the substrate/product and intermediate resonance states along the reaction pathway via interactions with the carboxylate. Commonly observed PLP recognition strategies are also observed in our structures. PLP-N1 and PLP-O3 interact with Asp168 and Arg171, respectively, and an α-helix dipole is oriented favorably toward the PLP phosphate. A histidine, also present in SHMT but replaced with tryptophan in AATase, forms a $\pi-\pi$ arrangement with the PLP ring.

Features of the TA active site may explain substrate specificity. There are no obvious functional groups that would recognize D-threonine, consistent with classification of *T. maritima* TA as an L-threonine aldolase. Recognition of the L-*allo*-threonine hydroxyl group involves two hydrogen bonds, with His83 and a nearby water molecule. This tightly bound water molecule interacts in turn with the negatively charged PLP phosphate group. The side chain methyl group of L-*allo*-threonine is recognized via a hydrophobic contact with Tyr87. In TA, Tyr87 replaces the THF binding site of SHMT, which provides a specific hydrophobic contact (3.9 Å) with Cγ2 of L-*allo*-threonine. During SHMT catalysis, the displaced formyl group is transferred to the additional THF cofactor, whereas the TA reaction produces a freely diffusable acetaldehyde molecule. The increased complexity demanded by the SHMT reaction raises the question of whether TA is a distant ancestor of SHMT, from which the capacity to bind and perform chemistry on THF evolved.

A second histidine originating from another subunit within the tetramer, *His125* (denoted by italics), occurs in the active site of TA. Although *His125* does not interact with the bound L-*allo*-threonine in our structure, model building with L-threonine suggests that a direct hydrogen bond between *His125* and the hydroxyl group of this alternative substrate is possible (data not shown). By analogy with the known stereochemistry of SHMT catalysis (*33*), it appears likely that the hydroxyl group of L-threonine occupies a similar

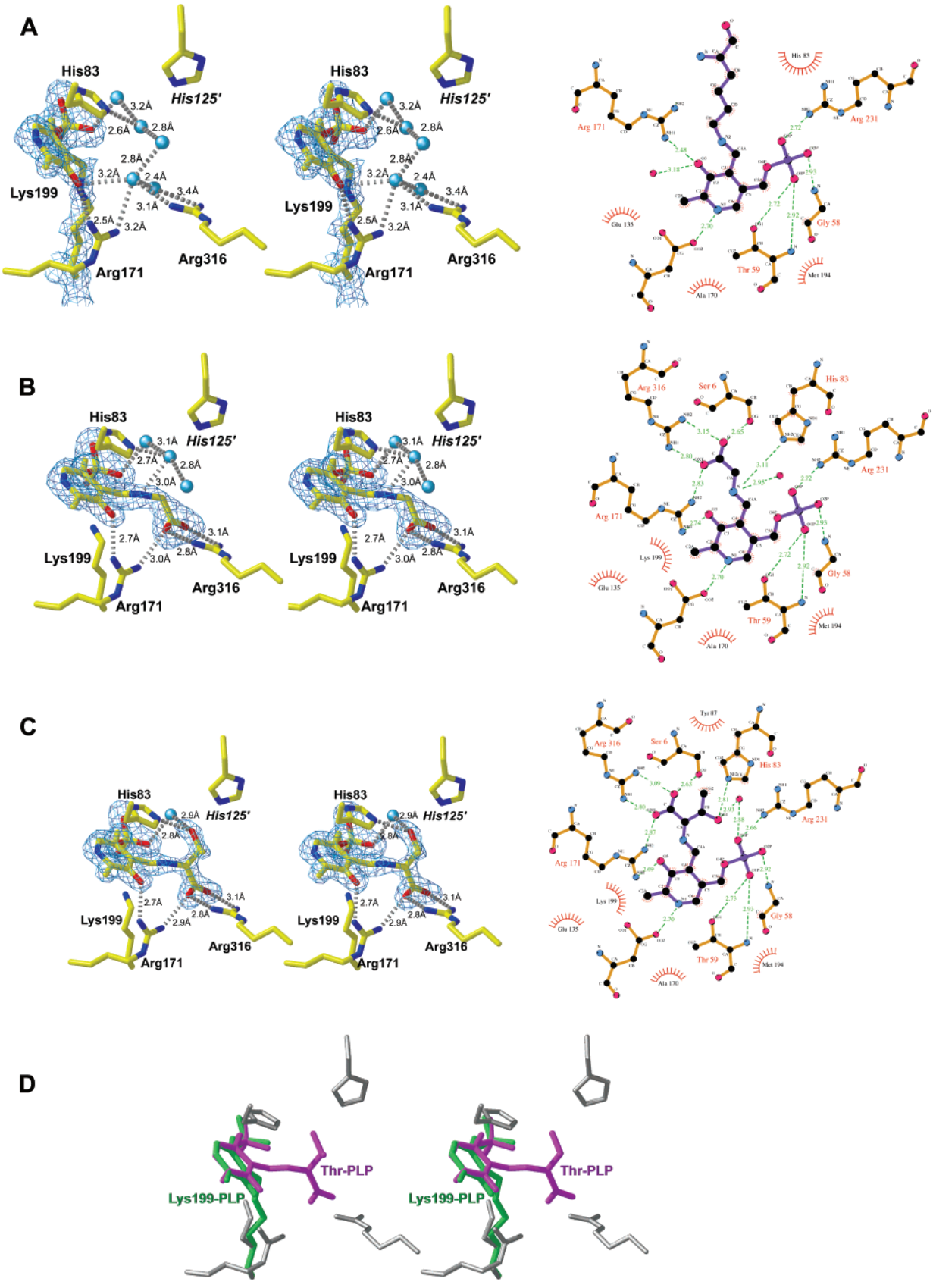

FIGURE 5: Stereoviews of the TA active site. Representative omit $|F_o| - |F_c|$ electron density maps for PLP (contoured at 3σ). (A) Internal aldimine with Lys199. (B) Complex with product, glycine. (C) Complex with substrate, L-*allo*-threonine. (D) Comparison of the PLP cofactors of the internal and external aldimines, colored purple and green, respectively. The shared atoms of glycine and L-*allo*-threonine display identical positions (for simplicity only the L-*allo*-threonine external aldimine is shown). Residues surrounding the active site are displayed in gray from the Thr-TA structure.

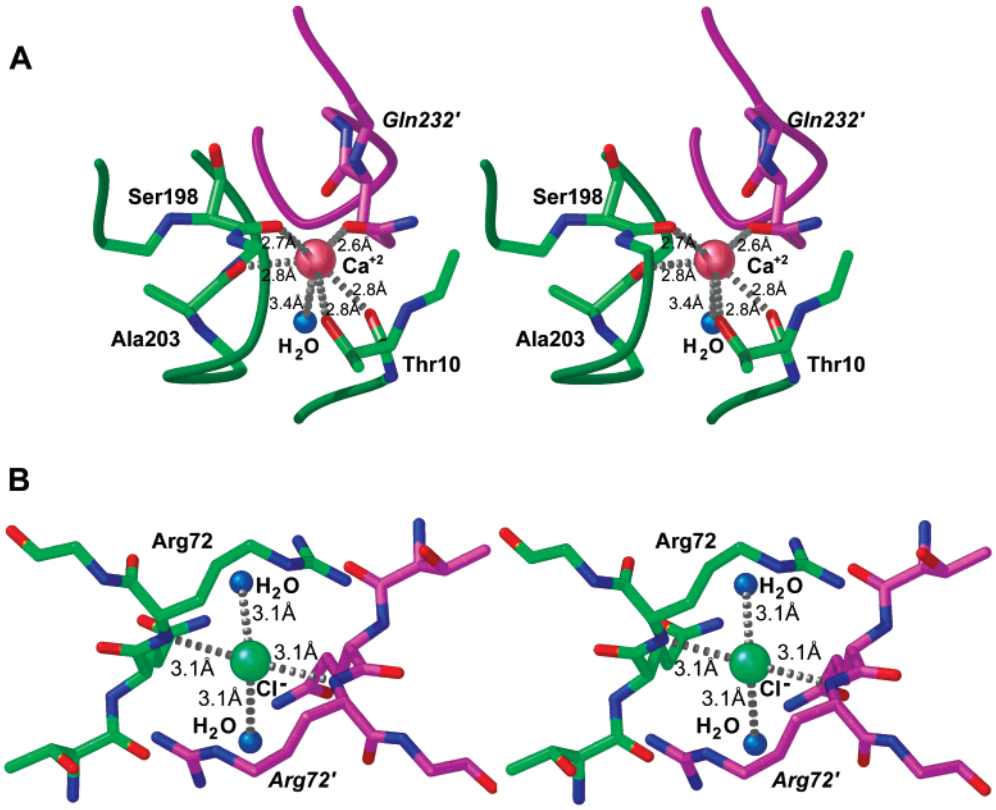

FIGURE 6: Ball-and-stick stereoviews of representative ions bound to *T. maritima* TA. (A) Ca^{2+} ion bound at the interface of the large and small domains. (B) Chloride ion coordinates subunits comprising less extensive dimer interfaces within the tetramer.

position to that observed for L-*allo*-threonine. In the absence of significant conformational changes within the tetramer, L-threonine Cγ2 would be distant from hydrophobic groups in the enzyme active site (closest distance in our model = 5 Å with Tyr87). We suggest that the presence of two catalytic histidines allows both L-*allo*-threonine and L-threonine to serve as substrates. The suboptimal interaction predicted for the methyl group of L-threonine would explain the observed 25-fold preference for L-*allo*-threonine.

Most of the active site residues that bind substrate/product in our structures are conserved among known TA homologues, including Ser6, His83, *His125*, Arg171, Lys199, and Arg316 (Figure 1). A similar human protein bearing various polypeptide chain deletions, including *His125* and Arg316, and an Arg171→Gln substitution, may have a different enzymatic function or represent a pseudogene. The only variable residue in the TA active site, Tyr87, appears to be involved in discriminating L-threonine from L-*allo*-threonine. Well-characterized TAs exhibit marked differences in stereospecificity at the substrate β-carbon. Side chain bulk at position 87 (*T. maritima* residue numbering) is correlated with specificity for L-*allo*-threonine, with larger side chains displaying higher preferences for the *allo* isomer. Indeed, *P. aerugenosa* TA, the only enzyme without a bulky aromatic amino acid at position 87, has a notable preference for L-threonine. Putative TAs from *A. thaliana*, *C. elegans*, and *D. melanogaster*, which have either Tyr or Trp residues at position 87, could be assayed for substrate preference to test our hypothesis.

Calcium Binding Sites. Given distinctive coordination geometries, low-temperature factors, and the presence of 100 mM Ca^{2+} in the crystallization buffer, several prominent peaks in $|F_o(hkl) - F_c(hkl)|$ difference Fourier syntheses for *T. maritima* TA were assigned as calcium ions. In total, six Ca^{2+} atoms were detected within the tetramer. Each subunit exhibits a bound ion pinioned between large and small domains, near the 'catalytic dimer' interface. Two additional calcium ions interact with symmetry-related molecules, and appear to stabilize the crystal lattice.

Like calcium binding in other protein structures, most of the direct Ca^{2+} ligands come from the protein, with only 1−2 water molecules per metal ion binding site (*34*). Between the domains of each subunit, the calcium ions are coordinated by main chain carbonyl oxygens of Ala203, Thr10, and Ser198, with additional contacts to water molecules and the side chain of Thr10 plus a contact with *Gln232* from an adjacent subunit (Figure 6). Direct coordination with the main chain of a residue adjacent to the conserved active site lysine (Lys199) suggests that divalent metal ion binding may affect the precise location of the PLP moiety via the polypeptide chain.

Catalytically important potassium ions have been observed in structures of other PLP enzymes in the AATase fold family. Potassium ions occur in both the tryptophanase and tyrosine phenol-lyase structures, where they play structural roles and are essential for activity (*24*, *25*). A second mode of potassium ion binding was seen in the structure of dialkylglycine decarboxylase (*35*). The interdomain calcium ion in our TA structure occurs in the same position as the potassium ions of tryptophanase and tyrosine phenol-lyase.

For tryptophanase, potassium cannot be replaced by sodium without inactivating the enzyme. The conversion of glycine and acetaldehyde to L-*allo*-threonine within the crystal provides evidence that TA is active in the presence

of calcium and sodium ions. The ionic radius of Ca^{2+} is similar to Na^+, both of which are significantly smaller than that of K^+ (114, 116, and 152 ppm, respectively). In contrast to tryptophanase, our work suggests that sodium (and not potassium) may be important for TA catalysis.

Chloride Binding Sites. In addition to the six calcium ions, two putative chloride ions were found in the TA tetramer. These anions appear to stabilize the dimer interface that contribute to tetramer formation (Figure 6). Each ion is coordinated by main chain nitrogens of Arg72 from each subunit, plus two water molecules. The positively charged arginine side chains give rise to an electrostatically favorable binding pocket for the anion. In addition to tetramer stabilization, these ions may contribute indirectly to catalysis by bringing *His125* from the neighboring subunit to the active site. Although chloride ions are thought to play inhibitory, allosteric roles for other PLP enzymes (*36*), *T. maritima* TA is active in the presence of high chloride concentrations (i.e., >250 mM).

DISCUSSION

Mechanistic Implications. Our work on TA provides structural insights into the mechanism and stereospecificity of the enzyme. With Gly-TA and Thr-TA structures in hand, active site residues can be identified. A pair of histidines (His83 and *His125*) plus bound water molecules lie close to the L-*allo*-threonine isomer at the active site. Furthermore, inspection of the active site in combination with primary sequence comparisons suggests a simple relationship between the bulkiness of the residue corresponding to *T. maritima* TA Tyr87, and the degree of stereospecificity for L-*allo*-threonine versus L-threonine.

For conversion of threonine to acetaldehyde plus the glycine quinonoid, the proposed retroaldol cleavage mechanism involves two proton-transfer steps. Initially, a catalytic base is required to abstract a proton from the threonine hydroxyl group (II, Figure 2). Subsequently, a PLP−glycine quinonoid complex is reprotonated at the α-carbon to generate the glycine external aldimine (III, Figure 2). The lysine that mediates the covalent linkage with the PLP has been suggested as a candidate base for SHMT (*31*). However, in the structure of TA, this side chain is >5 Å from the substrate/product in the external aldimine and is unlikely to participate in proton abstraction. Among TA side chains, His83 represents a likely candidate for the catalytic base that removes the proton from the L-*allo*-threonine hydroxyl group. For cleavage of L-threonine, different stereochemical requirements suggest *His125* or a water molecule activated by the negatively charged PLP-phosphate may function as the catalytic base, instead of His83. In our Gly-TA structure, water molecules in close proximity to Cα atoms of glycine represent excellent candidates for protonation of the external aldimine.

For SHMT, site-directed mutagenesis has been used to identify catalytically important residues (*37, 38*). Given the structural similarity of SHMT and TA, these earlier findings may be of some relevance to our discussion. Mutation of the SHMT residue equivalent to TA His83 inactivated the enzyme, but the interpretation of this result is not trivial, because the mutant enzyme can no longer bind PLP (*37*). Although SHMT lacks a direct equivalent of *His125*, a glutamate residue occupies a similar location in the active site (*31*). Mutations at this position in SHMT reduce enzyme activity for the serine substrate (*38*). Remarkably, retroaldol cleavage of L-*allo*-threonine by SHMT remains unaffected, which is consistent with suggestions that distinct residues contribute to recognition of different substrates by SHMT (as suggested for the histidine pair detected in our TA structure).

Kinetic studies, examining the pH dependence of the TA reaction and the relative activities of TAs with structure-based mutations, may be used to identify the catalytic base(s) within the TA class of PLP enzymes. TA plays an important role in the fundamental process of glycine biosynthesis, and possibly degradation of toxic L-*allo*-threonine. Furthermore, the enzyme has potential industrial use for resolving L-threonine from racemic starting materials. This structure provides a starting point for understanding how TA functions, and a structural rationale for future development of novel aldolases with altered substrate preferences by modification of active site residues.

ACKNOWLEDGMENT

We are grateful to L. Shapiro and K. D'Amico for X-ray measurements at the SGX-CAT (Advanced Photon Source), K. R. Rajashankar for assistance with X-ray measurements at Beamline X9A (National Synchrotron Light Source), J. B. Bonnano for expert crystallographic advice, and S. S. Ray for help with enzyme assays. We thank C. Groft, D. Jeruzalmi, D. Niessing, M. Romanowski, and J. Wedekind for useful discussions. *T. maritima* threonine aldolase represents target p044 from the New York Structural Genomics Research Consortium.

REFERENCES

1. Stauffer, G. V. (1987) in *Cellular and Molecular Biology* (Niedhart, F. C., Ed.) pp 412−418, American Society for Microbiology, Washington, DC.
2. McNeil, J. B., McIntosh, E. M., Taylor, B. V., Zhang, F. R., Tang, S., and Bognar, A. L. (1994) Cloning and molecular characterization of three genes, including two genes encoding serine hydroxymethyltransferases, whose inactivation is required to render yeast auxotrophic for glycine. *J. Biol. Chem. 269*, 9155−9165.
3. Liu, J. Q., Dairi, T., Itoh, N., Kataoka, M., Shimizu, S., and Yamada, H. (1998) Gene cloning, biochemical characterization and physiological role of a thermostable low-specificity L-threonine aldolase from *Escherichia coli*. *Eur. J. Biochem. 255*, 220−226.
4. Liu, J. Q., Dairi, T., Kataoka, M., Shimizu, S., and Yamada, H. (1997) L-*allo*-Threonine aldolase from *Aeromonas* jandaei DK-39: gene cloning, nucleotide sequencing, and identification of the pyridoxal 5′-phosphate-binding lysine residue by site-directed mutagenesis. *J. Bacteriol. 179*, 3555−3560.
5. Liu, J. Q., Ito, S., Dairi, T., Itoh, N., Kataoka, M., Shimizu, S., and Yamada, H. (1998) Gene cloning, nucleotide sequencing, and purification and characterization of the low-specificity L-threonine aldolase from *Pseudomonas* sp. strain NCIMB 10558. *Appl. Environ. Microbiol. 64*, 549−554.
6. Liu, J. Q., Nagata, S., Dairi, T., Misono, H., Shimizu, S., and Yamada, H. (1997) The GLY1 gene of *Saccharomyces cerevisiae* encodes a low-specific L-threonine aldolase that catalyzes cleavage of L-*allo*-threonine and L-threonine to glycine-expression of the gene in *Escherichia coli* and purification and characterization of the enzyme. *Eur. J. Biochem. 245*, 289−293.
7. McNeil, J. B., Flynn, J., Tsao, N., Monschau, N., Stahmann, K., Haynes, R. H., McIntosh, E. M., and Pearlman, R. E. (2000) Glycine metabolism in *Candida albicans*: characterization of the serine hydroxymethyltransferase (SHM1, SHM2) and threonine aldolase (GLY1) genes. *Yeast 16*, 167−175.

8. Roberto, P., Roberto, L., Lucia, T., John, C., Maria, P., and Enrico, M. (1991) DL-*allo*-Threonine aldolase in rat liver. *Biochem. Soc. Trans. 19*, 346−347.
9. Matthews, R. G., Drummond, J. T., and Webb, H. K. (1998) Cobalamin-dependent methionine synthase and serine hydroxymethyltransferase: targets for chemotherapeutic intervention? *Adv. Enzyme Regul. 38*, 377−392.
10. Jansonius, J. N. (1998) Structure, evolution and action of vitamin B6-dependent enzymes. *Curr. Opin. Struct. Biol. 8*, 759−769.
11. Hendrickson, W. A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science 254*, 51−58.
12. Otwinowski, Z., and Minor, W. (1997) in *Methods of Enzymology* (Carter, C. W., and Sweet, R. M., Eds.) pp 307−326, Academic Press, New York.
13. Weeks, C. M., and Miller, R. (1999) Optimizing Shake-and-Bake for proteins. *Acta Crystallogr. D 55*, 492−500.
14. Otwinowski, Z. (1991) in *Isomorphous Replacement and Anomalous Scattering* (Wolf, W., Evans, P. R., and Leslie, A. G. W., Eds.) pp 80−85, SERC Daresbury Laboratory, Warrington, U.K.
15. Cowtan, K. (1994) DM: An automated procedure for phase improvement by density modification. *Jt. CCP4 ESF-EACBM Newslett. Protein Crystallogr. 31*, 34−38.
16. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol. 6*, 458−463.
17. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D 54*, 905−921.
18. Contestabile, R., Paiardini, A., Pascarella, S., di Salvo, M. L., D'Aguanno, S., and Bossa, F. (2001) L-Threonine aldolase, serine hydroxymethyltransferase and fungal alanine racemase. A subgroup of strictly related enzymes specialized for different functions. *Eur. J. Biochem. 268*, 6508−6525.
19. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrixes. *J. Mol. Biol. 233*, 123−138.
20. Nakai, T., Okada, K., Akutsu, S., Miyahara, I., Kawaguchi, S., Kato, R., Kuramitsu, S., and Hirotsu, K. (1999) Structure of *Thermus thermophilus* HB8 aspartate aminotransferase and its complex with maleate. *Biochemistry 38*, 2413−2424.
21. Renwick, S. B., Snell, K., and Baumann, U. (1998) The crystal structure of human cytosolic hydroxymethyltransferase: a target for cancer chemotherapy. *Structure 6*, 1105−1116.
22. Marti-Renom, M. A., Stuart, A., Fiser, A., Sánchez, R., Melo, F., and Sali., A. (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct. 29*, 291−325.
23. Scarsdale, J. N., Kazanina, G., Radaev, S., Schirch, V., and Wright, H. T. (1999) Crystal structure of rabbit cytosolic serine hydroxymethyltransferase at 2.8 Å resolution: mechanistic implications. *Biochemistry 38*, 8347−8358.
24. Antson, A. A., Demidkina, T. V., Gollnick, P., Dauter, Z., von Tersch, R. L., Long, J., Berezhnoy, S. N., Phillips, R. S., Harutyunyan, E. H., and Wilson, K. S. (1993) Three-dimensional structure of tyrosine phenol-lyase. *Biochemistry 32*, 4195−4206.
25. Isupov, M. N., Antson, A. A., Dodson, E. J., Dodson, G. G., Dementieva, I. S., Zakomirdina, L. N., Wilson, K. S., Dauter, Z., Lebedev, A. A., and Harutyunyan, E. H. (1998) Crystal structure of tryptophanase. *J. Mol. Biol. 276*, 603−623.
26. Clausen, T., Huber, R., Laber, B., Pohlenz, H. D., and Messerschmidt, A. (1996) Crystal structure of the pyridoxal-5′-phosphate dependent cystathionine beta-lyase from *Escherichia coli* at 1.83 Å. *J. Mol. Biol. 262*, 202−224.
27. Jones, S., and Thornton, J. M. (1995) Protein−protein interactions: a review of protein dimer structures. *Prog. Biophys. Mol. Biol. 63*, 31−65.
28. Scarsdale, J. N., Radaev, S., Kazanina, G., Schirch, V., and Wright, H. T. (2000) Crystal structure at 2.4 Å resolution of *E. coli* serine hydroxymethyltransferase in complex with glycine substrate and 5-formyl tetrahydrofolate. *J. Mol. Biol. 296*, 155−168.
29. Trivedi, V., Gupta, A., Jala, V. R., Saravanan, P., Rao, G. S., Rao, N. A., Savithri, H. S., and Subramanya, H. S. (2002) Crystal Structure of Binary and Ternary Complexes of Serine Hydroxymethyltransferase from *Bacillus stearothermophilus*. Insights into the Catalytic Mechanism. *J. Biol. Chem. 277*, 17161−17169.
30. Okamoto, A., Higuchi, T., Hirotsu, K., Kuramitsu, S., and Kagamiyama, H. (1994) X-ray crystallographic study of pyridoxal 5′-phosphate type aspartate aminotransferases from *E. coli* in open and closed form. *J. Biochem. 116*, 95−107.
31. Szebenyi, D. M., Liu, X., Kriksunov, I. A., Stover, P. J., and Thiel, D. J. (2000) Structure of a murine cytoplasmic serine hydroxymethyltransferase quinonoid ternary complex: evidence for asymmetric obligate dimers. *Biochemistry 39*, 13313−13323.
32. Dunathan, H. C. (1971) Stereochemical aspects of pyridoxal phosphate catalysis. *Adv. Enzymol. Relat. Areas Mol. Biol. 35*, 79−134.
33. Fitzpatrick, T. B., and Malthouse, J. P. (1998) A substrate-induced change in the stereospecificity of the serine-hydroxymethyltransferase-catalysed exchange of the alpha-protons of amino acids—evidence for a second catalytic site. *Eur. J. Biochem. 252*, 113−117.
34. Katz, A. K., Glusker, J. P., Beebe, S. A., and Bock, C. W. (1996) Calcium ion coordination: a comparison with that of beryllium, magnesium and zinc. *J. Am. Chem. Soc. 118*, 5752−5763.
35. Hohenester, E., Keller, J. W., and Jansonius, J. N. (1994) An alkali metal ion size-dependent switch in the active site structure of dialkylglycine decarboxylase. *Biochemistry 33*, 13561−13570.
36. Burkhard, P., Tai, C. H., Jansonius, J. N., and Cook, P. F. (2000) Identification of an allosteric anion-binding site on *O*-acetylserine sulfhydrylase: structure of the enzyme with chloride bound. *J. Mol. Biol. 303*, 279−286.
37. Jagath, J. R., Sharma, B., Rao, N. A., and Savithri, H. S. (1997) The role of His-134, -147, and -150 residues in subunit assembly, cofactor binding, and catalysis of sheep liver cytosolic serine hydroxymethyltransferase. *J. Biol. Chem. 272*, 24355−24362.
38. Rao, J. V., Prakash, V., Rao, N. A., and Savithri, H. S. (2000) The role of Glu74 and Tyr82 in the reaction catalyzed by sheep liver cytosolic serine hydroxymethyltransferase. *Eur. J. Biochem. 267*, 5967−5976.
39. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283−291.
40. Esnouf, R. M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model. 15*, 132−134.
41. Kraulis, P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr. 24*, 946−950.
42. Merritt, E. A., and Murphy, M. E. P. (1994) Raster3D version 2.0—a program for photorealistic molecular graphics. *Acta Crystallogr. D 50*, 869−873.
43. Bacon, D. J., and Anderson, W. F. (1988) A fast algorithm for rendering space-filling molecular pictures. *J. Mol. Graph. 6*, 219−220.
44. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct., Funct., Genet. 11*, 281−296.

BI020393+