Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of phosphoserine aminotransferase from *Mycobacterium tuberculosis*

Fasseli Coulibaly,[a,b] Edouard Lassalle,[a] Heather M. Baker[a] and Edward N. Baker[a]*

[a]Maurice Wilkins Centre for Molecular Biodiscovery and School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand, and [b]Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia

Correspondence e-mail: ted.baker@auckland.ac.nz

Received 28 November 2011
Accepted 4 February 2012

PDB References: phosphoserine aminotransferase, 2fyf; 3vom.

Mycobacterium tuberculosis (*Mtb*), the causative agent of TB, remains a serious world health problem owing to limitations of the available drugs and the emergence of resistant strains. In this context, key biosynthetic enzymes from *Mtb* are attractive targets for the development of new therapeutic drugs. Here, the 1.5 Å resolution crystal structure of *Mtb* phosphoserine aminotransferase (MtbPSAT) in complex with its cofactor, pyridoxal 5′-phosphate (PLP), is reported. MtbPSAT is an essential enzyme in the biosynthesis of serine and in pathways of one-carbon metabolism. The structure shows that although the *Mtb* enzyme differs substantially in sequence from other PSAT enzymes, its fold is conserved and its PLP-binding site is virtually identical. Structural comparisons suggest that this site remains unchanged throughout the catalytic cycle. On the other hand, PSAT enzymes are obligate dimers in which the two active sites are located in the dimer interface and distinct differences in the MtbPSAT dimer are noted. These impact on the substrate-binding region and access channel and suggest options for the development of selective inhibitors.

1. Introduction

Mycobacterium tuberculosis (*Mtb*), the aetiological agent of human tuberculosis (TB), is a major concern for world health especially as its impact is aggravated by synergy with human immunodeficiency virus. Effective therapies against TB exist, but they rely on prolonged treatments with combinations of three or four antibiotics such as isoniazid and rifampicin that have been used for over 40 years. It is only recently that new therapeutic solutions have been proposed, with several drugs now being clinically tested to combat *Mtb* (Mdluli & Spigelman, 2006). In parallel with high-throughput screening of antibacterial compounds, a molecular and structural understanding of *Mtb* replication strategies is needed to open new avenues for the development of antibiotics that will efficiently eradicate latent infection and contain the emergence of drug-resistant strains. Biosynthetic pathways are attractive drug targets not only because many of them are essential to the survival of *Mtb*, but also because their constituent enzymes are often relatively amenable to drug design (Hasan *et al.*, 2006).

In this study, we focus on an enzyme involved in the biosynthesis of the amino acid serine. Serine can be produced by three independent routes: a phosphorylated pathway (Ichihara & Greenberg, 1955), a nonphosphorylated pathway and by formaldehyde incorporation through reverse activity of

the enzyme serine hydroxymethyltransferase (Quayle, 1972). The latter pathway is characteristic of methanotrophic and methylotrophic bacteria, whereas the nonphosphorylated pathway is essentially catabolic. In contrast, the phosphorylated pathway is generally essential for serine biosynthesis in most organisms and phosphoserine itself is also a biosynthetic precursor of L-cysteine through the cysteine synthase CysM in *Mtb* (O'Leary *et al.*, 2008).

Serine is not only an essential component of proteins and a major source of one-carbon units but is also a precursor in the biosynthesis of pyrimidine nucleotides and neurotransmitters in mammals. Although metabolic dysfunctions causing a severe deficiency in serine have been implicated in rare neurological diseases, including microcephaly, seizures and psychomotor retardation (Tabatabaie *et al.*, 2010), serine is not strictly essential in humans.

The first committed step in the phosphorylated pathway of serine biosynthesis (Fig. 1) uses D-3-phosphoglycerate, an intermediate in the glycolytic pathway, to synthesize phosphohydroxypyruvate in a NAD^+-dependent oxidation reaction catalyzed by a D-3-phosphoglycerate dehydrogenase (SerA1). Phosphohydroxypyruvate is then converted to L-3-phosphoserine, which is catalyzed by a phosphoserine aminotransferase (PSAT or SerC; EC 2.6.1.52) with glutamate as an amine donor. Finally, L-serine is produced from the 3-phosphoserine precursor by cleavage of the phosphate moiety catalyzed by a phosphoserine phosphatase. The enzymes involved in the phosphorylated pathway of *Mtb* have been identified as Rv2996c (SerA1; Dey *et al.*, 2005), Rv0728c (SerA2), Rv0505c (SerB1), Rv3042c (SerB2) and Rv0884c (SerC); three of these enzymes, SerA1, SerB2 and SerC, have been identified by transposon mutagenesis as being essential for *Mtb* growth (Sassetti *et al.*, 2003).

As a first step towards structure-based drug design targeting the biosynthesis of phosphoserine, we have determined the structure of *Mtb* phosphoserine aminotransferase (MtbPSAT) in complex with its pyridoxal 5′-phosphate (PLP) cofactor at a resolution of 1.5 Å. This mycobacterial PSAT has low overall sequence identity (∼20%) to previously characterized PSAT enzymes: the eubacterial PSATs from *Escherichia coli* (Hester *et al.*, 1999), *Bacillus circulans* (Kapetaniou *et al.*, 2006) and *B. alkalophilus* (Dubnovitsky, Kapetaniou *et al.*, 2009) and human PSAT (PDB entry 3e77; Structural Genomics Consortium, unpublished work). Nevertheless, we show that MtbPSAT shares the same fold, which is characteristic of the aspartate aminotransferase (AAT) family of PLP-dependent enzymes, and has a largely conserved active site. However, we identify differences in the substrate-binding region, particularly in the residues that bind the anionic phosphate or γ-carboxylate moieties of its successive substrates, and in the entry channel to the active site and suggest that these differences could be exploited to design mycobacterium-specific PSAT inhibitors.

2. Materials and methods

2.1. Protein expression

A culture of *E. coli* DH5α strain transformed with a pDONR221 vector containing the gene for Rv0884c of *Mtb* strain H37Rv fused to an N-terminal His_6 tag was obtained from The Institute for Genomic Research (TIGR). The construct coding for Rv0884c was transferred into a pDEST17 expression vector using the LR reaction in the Gateway system according to the manufacturer's instructions. The product of this reaction was used to transform chemically competent Top10 *E. coli* cells for amplification and to transform chemically competent *E. coli* BL21 cells for expression tests. Expression tests were carried out as described previously in a medium-throughput setup (Moreland *et al.*, 2005) utilizing auto-induction medium (Studier, 2005). Pull-down analysis with Ni–NTA resin showed that the protein was partially soluble at 291 K. Protein expression was scaled up to a 1 l culture in baffled flasks which were incubated for 3 h at 310 K after inoculation from an overnight culture and then for 21 h at 291 K. Cells were harvested by centrifugation and resuspended in 8 ml lysis buffer [50 m*M* Tris–HCl pH 9.0, 0.5 *M* NaCl, 1 mg ml^{-1} lysosyme, 100 µg ml^{-1} RNase A, 20 µg ml^{-1} DNase 1, 2 m*M* β-mercaptoethanol (β-ME)] containing a tablet of protease-inhibitor mix (Roche).

2.2. Protein purification

Cells were lysed using a cell disruptor (Constant Systems Ltd) operated at 124 MPa, followed by centrifugation for 45 min (20 000*g*, 277 K) to eliminate the insoluble fraction. The supernatant was filtered with 0.2 µm sterile filters and applied onto a HiTrap Ni–NTA column (GE Healthcare) loaded with $NiSO_4$ and equilibrated in 50 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl. Washes with increasing concentrations of imidazole (10, 50, 100 and 150 m*M* imidazole in buffer *A*, *i.e.* 50 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl) allowed the removal of most contaminants. The protein was then eluted using

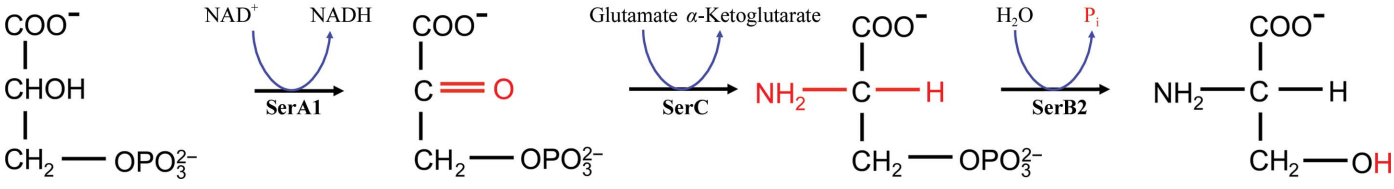


Figure 1
The phosphorylated pathway of serine biosynthesis. This pathway, in which D-3-phosphoglycerate is converted to L-serine, involves a series of three enzymes, SerA1 (Rv2996c), SerC (Rv0884c) and SerB2 (Rv3042c), in *Mtb*. SerC is a phosphoserine aminotransferase (PSAT) and is the subject of this analysis.

Table 1
Data-collection and phasing statistics.

	Native	$(CH_3)_3Pb$ derivative	K_2PtCl_4 derivative
Data collection			
Source	Rotating anode	ID29, ESRF	ID29, ESRF
Wavelength (Å)	1.5418	0.91840	1.07198
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 77.3, b = 94.3, c = 101.0	a = 77.9, b = 94.1, c = 101.1	a = 77.5, b = 94.1, c = 101.1
Resolution range† (Å)	25–2.10 (2.17–2.10)	25.0–1.60 (1.66–1.60)	25–1.50 (1.55–1.50)
Observed reflections	154987	345196	596436
Unique reflections	42955	97349	118835
Wilson B (Å^2)	21.0	14.7	12.9
Mean $I/\sigma(I)$†	9.7 (3.4)	20.6 (3.5)	17.4 (2.6)
Completeness† (%)	98.4 (99.3)	99.7 (98.9)	99.7 (98.1)
Muliplicity†	3.6 (3.5)	3.5 (3.3)	5.0 (4.7)
$R_{\rm merge}$†‡ (%)	12.1 (36.5)	5.7 (33.7)	8.6 (55.9)
Start (end) image§			
Scale factors K/B	1.0/0.0 (1.2/0.6)		1.0/0.0 (1.2/0.5)
$R_{\rm merge}$ (%)	10.9 (11.2)		9.7 (9.9)
$I/\sigma(I)$	7.3 (8.2)		6.8 (7.9)
Phasing			
Resolution† (Å)	25–2.10 (2.17–2.10)	25–2.10 (2.17–2.10)	25–2.10 (2.17–2.10)
$R_{\rm iso}$ to native¶		16.2	13.2
No. of sites		5	3
Phasing power (iso), acentric/centric†		0.80/0.70 (0.50/0.42)	1.74/1.50 (0.65/0.40)
Phasing power (ano), acentric/centric†		0.60 (0.29)	0.78 (0.33)
Overall FOM			0.40/0.34

† Values in parentheses are for the outermost shell. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement and $\langle I(hkl)\rangle$ is the weighted mean of all measurement of $I(hkl)$. § Values are the average statistics for the first ten images; values in parentheses are the average statistics for the last ten images. ¶ $R_{\rm iso} = \sum |F_{\rm PH} - F_{\rm P}|/\sum F_{\rm P}$.

buffer *A* containing 500 m*M* imidazole and dialyzed against 3 × 1 l buffer *A* with 2 m*M* β-ME. The dialysed samples were concentrated by centrifugal ultrafiltration (Vivaspin) and subjected to gel filtration on an HR 10/30 Superdex 75 column (GE Healthcare) in buffer *A* with 2 m*M* β-ME. The concentration of the final concentrated sample was estimated by spectrophotometry (NanoDrop, Thermo Fisher Scientific) assuming an extinction coefficient at 280 nm of 60 390 M^{-1} cm^{-1} as determined by the *ProtParam* web server (Gasteiger *et al.*, 2005). The final purified protein corresponds to the 376 residues of the native protein together with the following N-terminal sequence belonging to the His_6 tag: MSYYHHHHHHLESTSLYKKAGF.

2.3. Crystallization

Potential crystallization conditions were screened with a Cartesian Honeybee nanolitre dispensing robot (Genomic Solutions) using the sitting-drop method in Intelli-Plates (Genomic Solutions), with drops comprising 100 nl protein solution at 4.5 mg ml^{-1} in buffer *A* and 100 nl screening solution. The screening solutions were derived from a combination of commercial and in-house screens (Moreland *et al.*, 2005). Crystals were observed in many conditions containing phosphate and sulfate salts. After optimization, the best crystals were obtained using a screening solution comprising 2.2 *M* ammonium sulfate, 100 m*M* Tris–HCl pH 8.5. These crystals had a cuboid shape, with maximum dimensions of about 300 × 50 × 50 µm.

2.4. Data collection and processing

For data collection, crystals were transferred into a cryoprotecting solution consisting of 20% glycerol, 2.4 *M* ammonium sulfate, 50 m*M* Tris–HCl pH 8.5 for about 30 s and were then flash-cooled in liquid nitrogen. A native data set was collected in-house (Rigaku X-ray generator, MAR 345 image-plate detector) at a temperature of 100 K (Oxford Cryosystems). Diffraction data from derivative crystals were collected at 100 K using an ADSC Quantum 210 CCD detector on the ID-29 beamline at the ESRF synchrotron. Data were processed with the *HKL* suite v.1.97 using *SCALEPACK*/*DENZO* (Otwinowski & Minor, 1997). The crystals were found to belong to space group $P2_12_12_1$, with unit-cell parameters a = 77.48, b = 94.06, c = 101.06 Å, $\alpha = \beta = \gamma = 90^\circ$, and gave good-quality diffraction data to a resolution of 1.5 Å, with mean $I/\sigma(I)$ values greater than 3 in the outermost shell. The asymmetric unit was consistent with the presence of two molecules of MtbPSAT with a solvent content of 35%. Full data-collection statistics are given in Table 1.

2.5. Structure determination and refinement

A strong non-origin peak was detected in the native Patterson map with fractional coordinates (0.035, 0.000, 0.500) and an intensity of 44% of the origin peak. This indicated the presence of pseudo-translational noncrystallographic symmetry, which precluded phase determination by molecular replacement. The structure was therefore determined by multiple isomorphous replacement (MIR) using two heavy-atom derivatives (Table 1). Crystals were transferred into 10 µl of a stabilization solution (2.4 *M* ammonium sulfate, 50 m*M* Tris–HCl pH 8.5) containing 5 m*M* $KPtCl_4$ or trimethyllead acetate for 4 h and then transferred into the cryoprotecting solution for about 30 s and flash-cooled in liquid nitrogen. Heavy-atom derivative data were collected as described above. Three sites were found for the platinum derivative and two for the lead derivative using *SOLVE* (Terwilliger & Berendzen, 1999). These sites were refined in *SHARP* and minor sites were added by analysis of the residual maps (Vonrhein *et al.*, 2007).

The MIR phases were of sufficient quality for an almost complete initial model to be built using *ARP*/*wARP* (Perrakis *et al.*, 1999). The platinum-derivative data set was used for both model building and refinement as it had the highest

Table 2
Refinement statistics.

	Native MtbPSAT	MtbPSAT–$PtCl_4$
Resolution (Å)	15.0–2.10	20.0–1.50
No. of reflections (work/test)	42920/2173	112724/5973
Protein atoms (No. of residues)	5664 (736)	5750 (738)
No. of solvent molecules	711	846
Ligands	2 PLP, 1 SO_4^{2-}, 1 glycerol	2 PLP, 2 SO_4^{2-}, 7 glycerols, 6 $PtCl_4^{2-}$
TLS groups	—	4 groups for each chain: 8–98, 99–172, 173–279, 280–376
R/R_{free} (%)		
Overall	16.6/21.2	14.4/16.6
Outermost shell	18.9/27.5	20.6/24.7
R.m.s. deviations from standard values		
Bond lengths (Å)	0.010	0.013
Bond angles (°)	1.00	1.49
Average *B* factor (Å^2)		
Chain *A* (main chain/side chain)	18.1/21.4	11.7/13.8
Chain *B* (main chain/side chain)	18.3/22.1	12.0/14.3
PLP	18.9	9.6
Water	31.7	28.9
Ramachandran plot		
Most favoured regions (%)	97.5	97.8
Outliers (%)	0	0

resolution (1.5 Å) and the highest redundancy. A total of 725 residues out of 738 were built for the two chains. Subsequent $F_o - F_c$ and $2F_o - F_c$ electron-density maps revealed a large discrete piece of unmodelled density associated with each chain that was identified as belonging to the pyridoxal 5′-phosphate cofactors. Two sulfate ions, six $PtCl_4^{2-}$ ions and seven glycerol molecules were also added. The model was refined using *REFMAC*5 (Murshudov *et al.*, 2011), with rounds of refinement interspersed with manual checking, rebuilding and extension of the model using *Coot* (Emsley & Cowtan, 2004). Water molecules were added in peaks of $F_o - F_c$ density of $>3\sigma$ if within hydrogen-bonding distance of potential hydrogen-bond partners (protein or solvent). The solvent model was checked periodically during refinement and water molecules with low electron density ($<1\sigma$ in $2F_o - F_c$ maps), close contacts or no hydrogen-bond partner were removed. NCS restraints were applied in the initial cycles of rebuilding but not in the final stages of refinement. TLS groups were introduced in the final stages of refinement with four groups for each chain (residues 8–98, 99–172, 173–279 and 280–376) as suggested by the *TLS Motion Determination* server (Painter & Merritt, 2006). The final model comprises 369 residues for each molecule. Alternate conformations were modelled for the side chains of 20 residues in chain *A* (residues 13, 32, 56, 63, 117, 129, 134, 147, 190, 213, 219, 233, 242, 246, 263, 286, 290, 297, 358 and 367) and 15 residues in chain *B* (residues 24, 56, 63, 67, 114, 117, 129, 213, 219, 263, 286, 290, 297, 318 and 367). Insufficient electron density was present to model the co-ordinates of the first seven residues and residue Asp140. The final model has an *R* of 14.4% and an R_{free} of 16.6%.

To check whether the ligation state of the PLP could have been affected by radiation damage, the native structure was also refined against the home-source 2.1 Å resolution data set using *BUSTER* (Blanc *et al.*, 2004) without NCS or TLS restraints. The refined model (R = 16.6%, R_{free} = 21.2%)

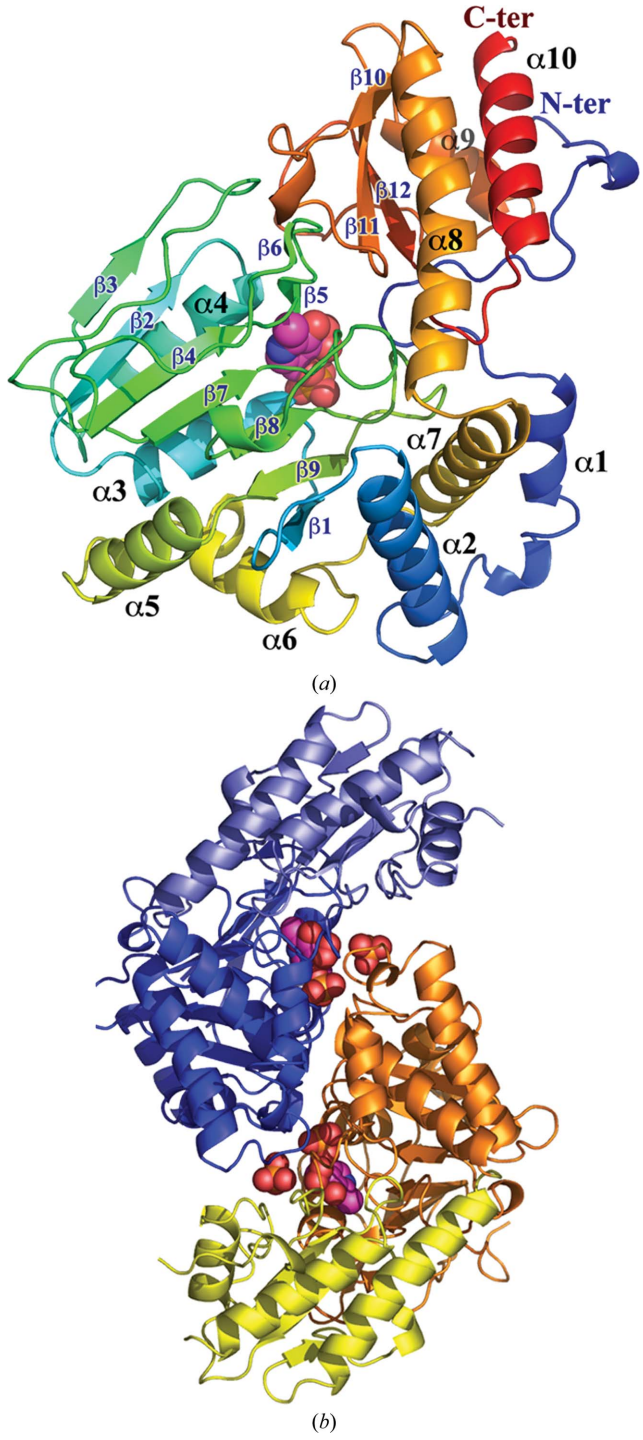


Figure 2
Structure of *Mtb* phosphoserine aminotransferase (MtbPSAT). (*a*) The MtbPSAT monomer coloured in rainbow representation from the N-terminus (N-ter) in blue to the C-terminus (C-ter) in red. Secondary-structure elements are labelled. The small domain, which includes the N-terminus plus the C-terminal ∼100 residues of the polypeptide, beginning with helix α8, is at the top of the picture. The PLP cofactor is shown in space-filling representation. (*b*) The MtbPSAT dimer, with subunits coloured with darker shades for the large domain (navy blue and orange) and lighter shades for the small domain (slate blue and yellow). The two active sites are located in the dimer interface about 19 Å apart and are indicated by the bound PLP cofactor and sulfate ion in each site shown in space-filling mode. This figure is rotated about 90° from that in (*a*) to better show the dimer.

comprised residues 9–376 for both molecules, 711 waters, two PLP molecules, a sulfate ion and a glycerol molecule. Full details of the refinement of both structures are given in Table 2.

2.6. Model quality

Model quality was assessed with *Coot* during rebuilding and with *MolProbity* (Chen *et al.*, 2010). For both structures, all residues were in the allowed regions of the Ramachandran plot, as defined by *MolProbity*, with 97.8 and 97.5% of the residues in the most favoured regions for the 1.5 and 2.1 Å resolution structures, respectively. For the high-resolution structure the *MolProbity* all-atom contact score of the final model was 3.32, placing it in the 97th percentile in the 1.25–1.75 Å resolution range, and the geometry score of 1.17 placed it in the 98th percentile. The two molecules of the dimer are very similar, with an r.m.s. difference in C^{α}-atom positions of 0.6 Å when all residues of chain *A* are superimposed on those of chain *B*. Structure factors and coordinates have been deposited in the Protein Data Bank with codes 2fyf for the 1.5 Å resolution MtbPSAT–$PtCl_4$ structure and 3vom for the 2.1 Å resolution native MtbPSAT structure.

2.7. Other calculations

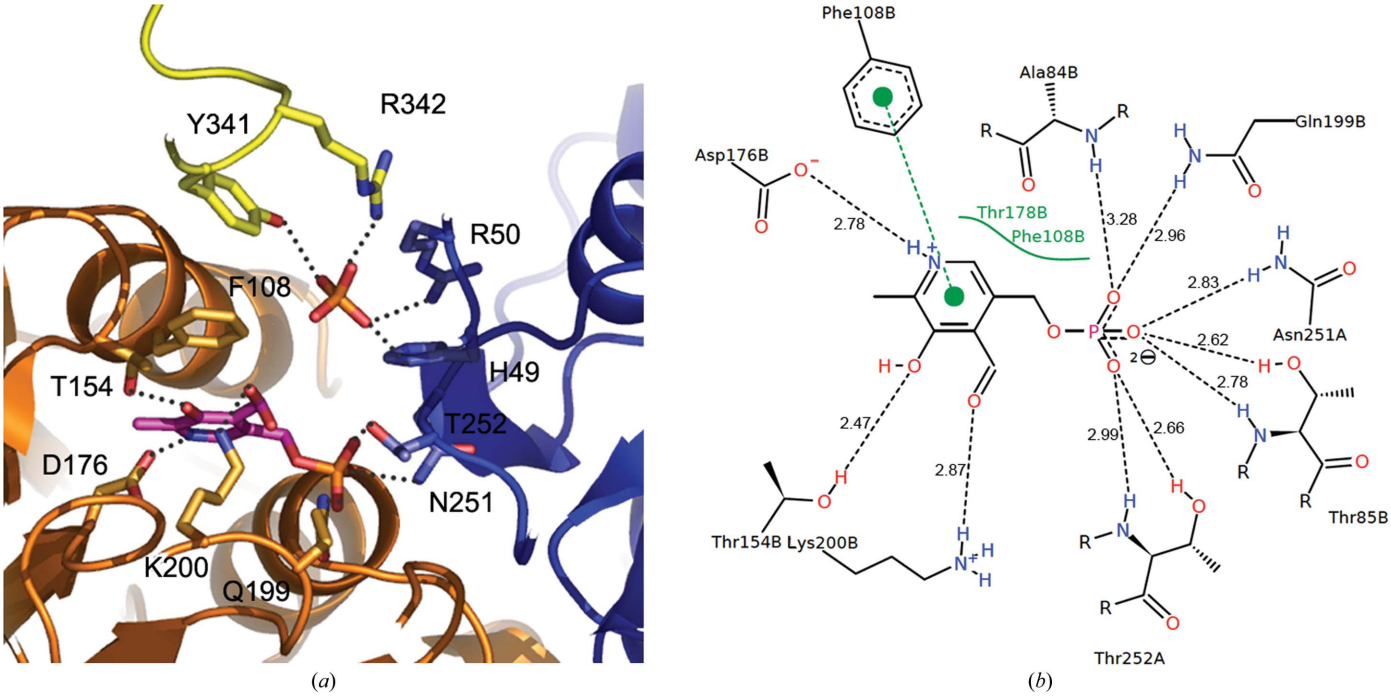
The illustrations were prepared with *PyMOL* (http://www.pymol.org).

Oligomeric assembly was investigated using *PISA* at the European Bioinformatics Institute (Krissinel & Henrick, 2007). The interaction between the PLP ligand and the protein was represented using *PoseView* (Stierand *et al.*, 2006). Structural superpositions were carried out using the programs *SSM* (Krissinel & Henrick, 2004) and *SUPERPOSE* from the *CCP*4 suite (Winn *et al.*, 2011).

3. Results and discussion

3.1. Structure determination

We have determined the crystal structure of *M. tuberculosis* phosphoserine aminotransferase (MtbPSAT), also known as SerC, at a resolution of 1.5 Å. When expressed in *E. coli* BL21 cells at 291 K the protein was only partially soluble and only about 350 µg protein could be purified per litre of culture. Nevertheless, when the His_6-tagged protein was subjected to crystallization trials using a nanolitre dispensing robot, crystals were readily obtained in conditions containing sulfate or phosphate. The structure could not be solved by molecular replacement, in part because of the low sequence identity to other structurally characterized PSAT enzymes but also because of pseudo-translational noncrystallographic symmetry. We therefore determined the structure by MIR methods and refined it at 1.5 Å resolution without NCS restraints but with TLS parameterization. The final model is of very good quality, with R and $R_{\rm free}$ of 14.4% and 16.6%, respectively, and excellent *MolProbity* scores for geometry and all-atom clash analyses (98th percentile of structures determined at 1.5 ± 0.25 Å).



(*a*) (*b*)

Figure 3
Active site of MtbPSAT. (*a*) The active-site cavity is formed between the large domains (navy blue and orange) of two monomers, with the end closed off by the small domain of one monomer (yellow). The PLP cofactor and bound sulfate ion are shown in stick mode. Hydrogen bonds to key active-site residues are shown as broken lines. (*b*) Schematic diagram of interactions in the PLP-binding site. Hydrogen bonds are shown as black broken lines, with distances in Å. The π–π interaction between PLP and Phe108, shown as a green broken line, has a centre-to-centre distance of 3.8 Å and a plane-to-plane distance of 3.7 Å. Hydrophobic contacts are represented by a green spline section highlighting the hydrophobic part of the PLP molecule and the label of the contacting residues.

Table 3
Residues contacting the PLP cofactor and the glutamate substrate analogue (AMG) in *Mtb*, *E. coli* and human PSAT.

Residues contacting the AMG substrate analogue were identified in the *E. coli* PSAT complex; for the *Mtb* and human enzymes the structurally equivalent residues are given.

Ligand	*Mtb*	*E. coli* (1bjo)	Human (3e77)
PLP	Thr85	—	Cys80
	Phe108	Phe102	Trp107
	Thr154	Thr153	Thr156
	Asp176	Asp174	Asp176
	Thr178	Ser176	Thr178
	Gln199	Gln197	Gln199
	Lys200	Lys198	Lys200
	Asn251 (*B*)	Asn239 (*B*)	Asn241 (*B*)
	Thr252 (*B*)	Thr240 (*B*)	Thr242 (*B*)
AMG γ-carboxylate	His49	His41	His44
	Arg50	Arg42	Arg45
	Tyr341	His328	His335
	Arg342	Arg329	Arg336
AMG α-carboxylate	Ser24	Ser9	—
	Arg350	Arg335	Arg342

3.2. Overall structure

MtbPSAT has a fold typical of the aspartate aminotransferase (AAT) family of PLP-dependent enzymes (Jansonius, 1998; Schneider *et al.*, 2000) organized into two tightly associated α/β-domains (Fig. 2). The larger domain comprises residues 25–271 and is organized around a seven-stranded β-sheet comprising strands $\beta1$, $\beta9$, $\beta8$, $\beta7$, $\beta4$, $\beta3$ and $\beta2$, in which all the strands are parallel except strand $\beta9$. This central β-sheet is associated with seven α-helices that pack against it and around the periphery of the domain, together with two small strands $\beta5$ and $\beta6$. The final helix, $\alpha7$, leads into the second, smaller, domain, which comprises the ~100 C-terminal residues 272–376 together with the N-terminal segment consisting of residues 8–24. This domain is built as a three-stranded antiparallel β-sheet flanked by three α-helices. The N-terminal segment (residues 8–24) follows an extended course over the surface of the small domain with no regular secondary structure.

The protein forms a stable symmetrical homodimer (Fig. 2*b*) which is maintained by extensive interactions, mostly between the large domains of the two subunits. The dimer interface is highly hydrophobic and buries a total of 2080 Å^2 of solvent-accessible surface per monomer (~15% of the total surface area), typical of an obligate dimer. The active sites, one per monomer, are marked by the presence of a bound PLP molecule whose 5′-phosphate and C4 substituents project into the dimer interface such that the phosphate moieties are ~19 Å apart and the C4 substituents are adjacent to a solvent-filled substrate-binding cavity linked by a channel to the external solution. Again, this organization is typical of enzymes of the AAT family.

3.3. Active site

Each active site contains a bound cofactor molecule modelled as PLP. The aromatic ring rests above the C-terminal end of the central strand $\beta7$ of the seven-stranded β-sheet in a pocket in which its pyridine N atom points in towards the interior of the large domain, hydrogen-bonded to the invariant Asp176, and its 5′-phosphate and C4 substituent point out into the dimer interface. The phosphate group is adjacent to the N-terminus of helix $\alpha3$, hydrogen-bonded to the main-chain amide N atoms of Ala84 and Thr85 and the side chain of Gln199 from one monomer, as well as to Asn251 and Thr252 (N and OG1) of the opposing monomer (Table 3; Fig. 3).

The C4 atom is the site of attachment of the reactive aldehyde group. In their resting state, PLP-dependent enzymes are usually joined by a covalent aldimine linkage to an essential lysine residue with a C=N bond (Jansonius, 1998). This generates the so-called internal aldimine moiety. During the reaction cycle (Kirsch *et al.*, 1984; Jansonius, 1998), this linkage, which is evidently strained (Islam *et al.*, 2000; Dubnovitsky, Ravelli *et al.*, 2009), is attacked by a substrate amino group, breaking the Lys–PLP linkage and forming an external aldimine moiety. A variety of transformations can then ensue, with the final step resulting in the regeneration of the unmodified PLP, which can then religate with the active-site lysine.

A surprising feature of the MtbPSAT structure is that PLP is not covalently linked to the active-site lysine Lys200 as was expected. Remarkably, however, the cofactor occupies exactly the same position as it does in all other PSAT enzyme structures and is in essentially the same orientation regardless of the ligation state (Fig. 4). The available structures include the human (PDB entry 3e77; Structural Genomics Consortium, unpublished work) and *E. coli* (PDB entry 1bjn; Hester *et al.*, 1999) PSATs in which PLP is bound to the active-site lysine, forming the internal aldimine moiety; structures of *B. alcalophilus* PSAT in which the Lys–PLP linkage is stretched (PDB entry 1w23; Dubnovitsky, Ravelli *et al.*, 2009); and a complex of *E. coli* PSAT in which PLP is detached from the lysine and forms an external aldimine with the substrate analogue α-methyl-L-glutamate (AMG; PDB entry 1bjo; Hester *et al.*, 1999). These all represent different stages of the reaction cycle.

Comparisons of the respective protein structures show that in no case does the PLP ring change its position or orientation (Fig. 4); when pairwise superpositions of the various structures onto MtbPSAT are carried out based only on the polypeptide structures and not PLP the r.m.s. difference in the PLP atomic positions (including all substituents except for the phosphate moiety) does not exceed 0.43 Å. Instead, movement occurs in the $\beta8$–$\beta9$ loop that carries the active-site lysine; this loop moves away as the Lys–PLP bond is broken. In MtbPSAT the C^α atom of Lys200 at the tip of this loop is displaced 1.9 Å away from the nonligated PLP when compared with human PSAT, in which Lys200 is bound to PLP. In *B. alcalophilus* PSAT the loop movement is intermediate. The PLP aromatic ring is firmly anchored in place by a strong hydrogen-bonded ion pair between its protonated pyridine N atom and the carboxylate group of the invariant Asp176 and is sandwiched between the aromatic ring of Phe108, which π-stacks against one face with a plane-to-plane distance of 3.7 Å, and the side chain of Thr178 (3.7 Å away), which packs against the other

face. Many highly conserved interactions surround this site, although individual residues can change; Phe108 is usually Trp, for example, and Thr178 is usually Ser.

Why is the PLP not bound to Lys200 in the MtbPSAT structure? Radiation damage can cause breakage of the aldimine bond (Dubnovitsky, Ravelli *et al.*, 2009), but this seems unlikely in the present case as exactly the same nonbonded structure was found when the structure was independently determined from an in-house data set at 2.1 Å resolution in which there was no significant radiation damage. This data set, comprising 180 images, showed no change in $I/\sigma(I)$ or R factor from the first image to the last and the scaling B factors were all close to 0. It is possible that the cofactor seen here is in fact PMP, a side-product of the reaction cycle in which an amino group is attached to C4′ instead of the aldehyde O atom that reacts with the lysine ε-amino group (Jansonius, 1998). Studies of a D-amino-acid aminotransferase (DAAT) from *Bacillus* species YM1 show that when the recombinant enzyme is isolated the cofactor is present in both the PLP and PMP forms, and crystal structures with both forms have been determined (Sugio *et al.*, 1995; Peisach *et al.*, 1998). In MtbPSAT, however, the C4 substituent is hydrogen-bonded (2.6 Å) to the ε-amino group of Lys200, suggesting that it is C=O rather than C—NH_2. We therefore model the cofactor as PLP. It is also possible that the Lys–PLP bond is more strained in MtbPSAT than in other PSAT enzymes, allowing more facile bond breakage. The crystal structure suggests that this could be the case, since the β8–β9 loop that carries Lys200 contains a one-residue insertion, has an altered conformation compared with the other PSATs and is significantly more constrained by hydrogen bonds.

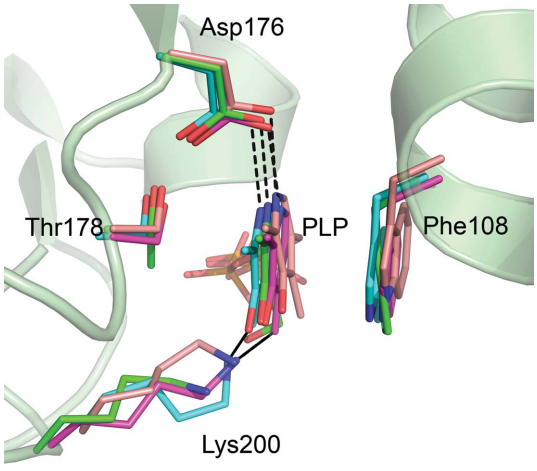


Figure 4
Conserved binding mode of PLP. The PLP binding mode is shown for MtbPSAT (this work; 1.5 Å resolution; green), human PSAT (PDB entry 3e77; 2.5 Å resolution; cyan), *B. alcalophilus* PSAT (PDB entry 1w23; 1.1 Å resolution; magenta) and the AMG complex of *E. coli* PSAT (PDB entry 1bjo; 2.8 Å resolution; wheat). The superposition was based on the respective polypeptide chains, not including the PLP (human PSAT, 318 C^{α} atoms with r.m.s. difference 1.95 Å; *B. alcalophilus* PSAT, 313 C^{α} atoms with r.m.s. difference 1.70 Å; *E. coli* PSAT, 324 C^{α} atoms with r.m.s. difference 1.88 Å). These four structures represent different stages of the reaction cycle; in human PSAT PLP forms an internal aldimine linkage with the active-site lysine, in *B. alcalophilus* PSAT the Lys–PLP bond is stretched, in the *E. coli* PSAT structure PLP is bound to AMG by an external aldimine linkage and in MtbPSAT the PLP has dissociated. Nevertheless, the PLP and surrounding structure hardly move apart from the lysine residue, which moves away as the PLP dissociates.

The substrate-binding pocket, immediately adjacent to the PLP, is located in the dimer interface, where it is flanked by residues from the large domains of two subunits and terminated by a loop from the small domain of one subunit. This pocket must progressively bind a series of substrates of similar size and shape: L-glutamate, 3-phosphohydroxypyruvate, 2-oxoglutarate and L-phosphoserine. In MtbPSAT the pocket is empty except for seven well defined water molecules and a bound sulfate ion. Residues involved in substrate recognition and catalysis can, however, be identified by comparison with the structure of the AMG complex of *E. coli* PSAT (Hester *et al.*, 1999). In the latter, the α-carboxylate group, which is common to all substrates, is bound by Ser9 and Arg335 from the small domain (Ser24 and Arg350 in MtbPSAT). At the other end of the substrate pocket, the γ-carboxylate group of glutamate is bound by residues His41, Arg42, His328 and Arg329, with the latter two residues contributed by a loop from the small domain (Hester *et al.*, 1999). Here there are distinct differences in MtbPSAT: His328 is replaced by Tyr341 and the loop has a two-residue insertion and a changed conformation. Superposition of MtbPSAT on the AMG complex of the *E. coli* enzyme shows that the sulfate ion in MtbPSAT is virtually coincident with the γ-carboxylate group. The residues that bind the sulfate ion in MtbPSAT are His49, Arg50, Tyr341 and Arg342 (Fig. 3*a*, Table 3), which are equivalent to His41, Arg42, His328 and Arg329 in *E. coli* PSAT. Interestingly, the requirement for phosphate or sulfate in the crystallization of MtbPSAT argued for a specific role for the anion, and its presence in the substrate-binding site at the dimer interface implies a stabilizing role in the active site. This presumably arises because it bridges between residues from the small domain of one subunit (Tyr341 and Arg342) and the large domain of the other subunit (His49 and Arg50), forming short geometrically favourable hydrogen bonds to both subunits.

3.4. Sequence and structure comparisons

A structure-based alignment of the MtbPSAT sequence with those of other PSATs (Fig. 5*a*) reveals strong conservation of the residues that bind the PLP cofactor. Residues binding the substrate are also generally well conserved, although the *Mtb* and archaeal enzymes have some notable differences. Thus, although residues 338–350 are strongly conserved in bacterial, mammalian, yeast and plant PSATs, the *Mtb* and archaeal enzymes have two specific differences: a tyrosine at position 341 instead of the otherwise strictly conserved histidine and a two-residue insertion prior to the invariant Arg350. These specificities are reflected in the phylogenic tree constructed from this multiple sequence alignment (Fig. 5*b*), in which MtbPSAT and the archaeal enzymes form a distinct clade that is only distantly related to the other bacterial enzymes and to human PSAT.

Structural comparison of PSATs indicates that although the overall fold is conserved between all PSATs there are several regions of difference between the *Mtb* enzyme and the other PSATs, most of which are located around the periphery of the

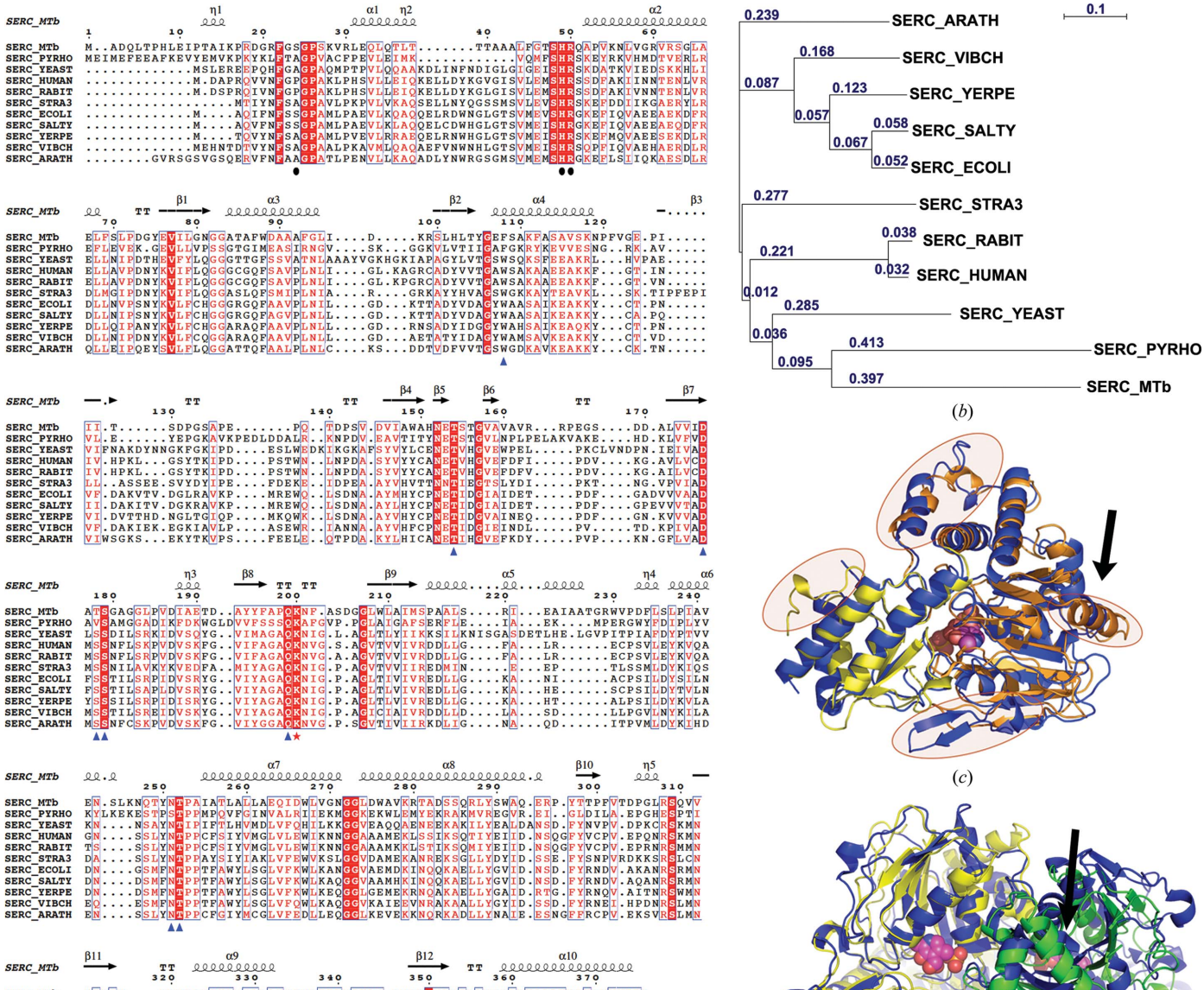
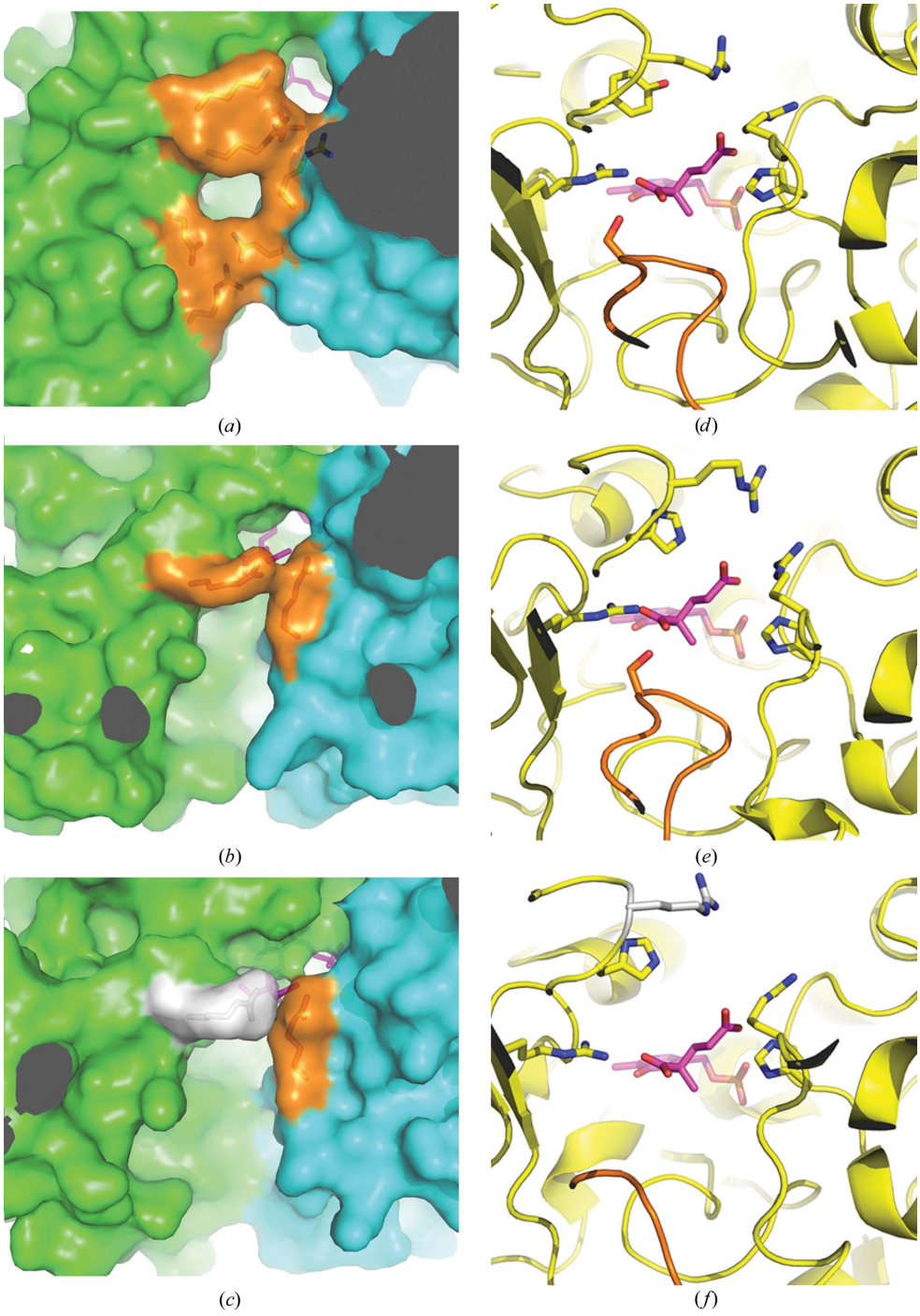


Figure 5

Sequence and structural comparisons of PSATs. (*a*) Sequence alignment of 11 PSATs. Residues on a red background are strictly conserved; residues in red font and framed in blue are similar across all sequences. Secondary-structure elements of the *Mtb* protein are indicated on top of the alignment. Triangles and dots under the alignment indicate residues implicated in PLP and substrate binding, respectively. The alignment was computed by the *Expresso* program using the following sequences retrieved from UniProtKB/Swiss-Prot and structural templates from the PDB: *Mtb* (P63514; 2fyf); *Yersina pestis* (Q8ZGB4; 3qbo; Center for Structural Genomics of Infectious Diseases, unpublished work); *Vibrio cholerae* (Q9KSU7); *Salmonella typhimurium* (P55900; 3qm2; Center for Structural Genomics of Infectious Diseases, unpublished work); *E. coli* K12 (P23721; 1bjn); *Streptococcus agalactiae* (Q8E3Y3); *Saccharomyces cerevisiae* (P33330); rabbit (P10658); human (Q9Y617; 3e77); *Arabidopsis thaliana* (Q96255; 3qbo); *Pyrococcus horikoshii* (O59033; 2dr1). The PSAT of *A. thaliana* has a 60-residue extension that was removed for clarity. (*b*) Phylogenic tree built from the structure-based sequence alignment of the PSAT proteins presented in (*a*). The tree was computed by the *Expresso* program and plotted using *NJplot*. (*c*) Superposition of MtbPSAT onto its *E. coli* homologue. The small and large domains of the *Mtb* protein are coloured yellow and orange, respectively, and the *E. coli* protein is shown in blue. The PLP molecule of MtbPSAT is shown in space-filling representation. Regions with significant structural differences are circled in orange. The arrow points to an α-helix that is only present in the *Mtb* protein. (*d*) Superposition of the *Mtb* and human PSAT dimers. MtbPSAT is coloured yellow and green for each of the dimer subunits. The human PSAT is shown in blue and the curved arrow indicates the different orientation of the subunits in the two dimers.

molecule. MtbPSAT is more compact overall than the *E. coli* enzyme; a long β-hairpin is absent because the loop 129–140 is shorter, and the loop 40–46 also adopts a more compact fold compared with *E. coli* PSAT. The only secondary-structural difference is a three-turn helix (residues 215–230 in MtbPSAT) that is a single turn in *E. coli* PSAT (Fig. 5*c*).

There are, however, significant differences in the dimer (Fig. 5*d*). Whereas the individual *Mtb* and *E. coli* PSAT subunits can be superimposed with an r.m.s. difference of 1.9 Å for the C^{α} atoms of 324 equivalent residues, their dimers agree less well, with an r.m.s. difference of 2.2 Å for 637 equivalent residues. The higher value for the dimer results from a slightly different oligomeric interface, in which the two subunits of the *Mtb* and *E. coli* PSATs are tilted with regard to each other by 10° and translated by 1.6 Å. The human PSAT (PDB entry 3e77; Structural Genomics Consortium, unpublished work) is also closer to the *E. coli* PSAT than to the *Mtb* enzyme both in sequence (45% sequence identity to the *E. coli* enzyme but only 21% sequence identity to MtbPSAT) and in structure. Notably, the human dimer is very similar to the *E. coli* dimer, with a twist of 11° and a translation of 3 Å compared with the *Mtb* dimer. This results in a much closer correspondence between the human and *E. coli* dimers (r.m.s. difference of 1.5 Å for 692 C^{α} atoms) than between the human and *Mtb* dimers (r.m.s. difference of 2.2 Å for 601 C^{α} atoms).



3.5. Active-site access

The differences between the dimer formed by MtbPSAT and the human and *E. coli* dimers result in changes in access to the active sites, which are located in the dimer interface. The latter two enzymes both have large open grooves leading to the

Figure 6
Active-site comparisons. Two views are shown for each enzyme: (*a*) and (*d*), MtbPSAT; (*b*) and (*e*), *E. coli* PSAT; (*c*) and (*f*), human PSAT. Residue Arg336 of human PSAT is shown in white in (*c*) and (*f*) because its side chain was not included in the PDB structure 3e77 and was modelled here according to the rotamer present in the *E. coli* structure. (*a*)–(*c*) show the molecular surface that defines the active-site groove in each case. The two subunits of the dimer are shown in green and cyan. The side chains of residues forming the lid over the active site (Arg342, Lys343 and Arg50 for MtbPSAT) and the bridge, present only in MtbPSAT, between the small and large domain (Arg331, Glu339 and Gln51) are shown as sticks and coloured in orange. The PLP and 2-methylaspartate molecules are shown in magenta in stick mode. (*d*)–(*f*) show an orthogonal view looking into the active site and revealing the similarity between the three proteins, except where Tyr341 in MtbPSAT is replaced by histidine in the *E. coli* and human proteins. The loop containing residue Ser24 in the *Mtb* protein, shown in orange, differs in the human protein because of a cloning artefact that introduces a shorter and nonrelevant loop.

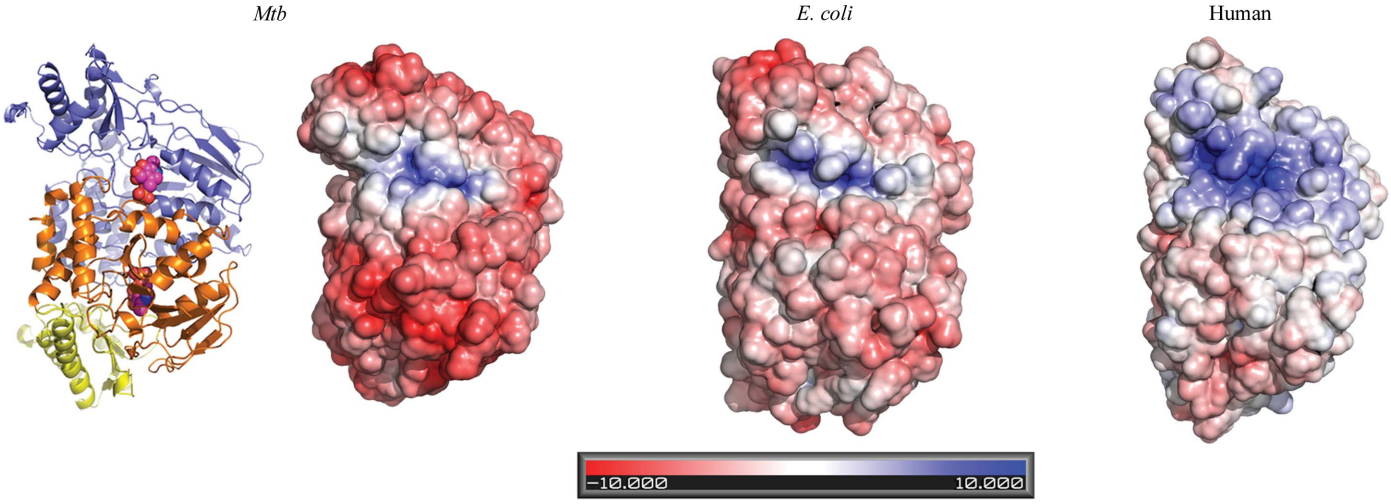


Figure 7
Surface electrostatic potential of PSATs. The solvent-accessible surface of the *Mtb*, *E. coli* and human PSATs is coloured according to the local electrostatic potential as calculated by the *Adaptive Poisson–Boltzmann Solver* program (red–white–blue scale from $-10kT$ to $+10kT$). The orientation in each case is shown by comparison with the cartoon representation of MtbPSAT shown on the left with the same colour scheme as in Fig. 2.

active site (Figs. 6*a*–6*c*), whereas in the *Mtb* structure additional contacts across the dimer interface restrict access (Figs. 6*a* and 6*d* *versus* Figs. 6*b* and 6*e*). This results from the slightly more closed MtbPSAT dimer and from the changed conformation of the loop 340–348, which contributes the sulfate-binding residues Tyr341 and Arg342. These residues, together with Arg50, which is hydrogen-bonded to the main-chain carbonyl O atoms of Thr249 and Tyr250, form a lid above the entry channel to the amino-acid-binding site. Below this, as seen in Fig. 6(*a*), residues Arg331 and Glu339 from chain *A* and Asn51 from chain *B* also bridge the small and large domains of two subunits, completing a portal that restricts access to the active site. In contrast, in the human and *E. coli* enzymes the lid seen for MtbPSAT is reduced to a single contact between an arginine of the small domain of one chain and an arginine of the large domain of the other, and the other features that constrict the channel are completely absent. The access channel in these two proteins thus becomes a deep but open solvent-exposed groove.

3.6. Surface charge

Analysis of the electrostatic potential (Fig. 7) shows that the *Mtb* protein has a highly polarized surface, with a single positively charged patch surrounding the entry to the active site and a strongly negatively charged surface at the back of the active site. This marked polarization is less strong in the *E. coli* enzyme, with an overall neutral or weakly negatively charged surface except for the active-site groove. In contrast, the human protein harbours an overall positive electrostatic potential, with an extended patch of basic residues on the protein face where the active site is located. Further away from the active site, the surfaces of the PSATs are more similar, with a highly conserved patch extending up to 22 Å away from the active-site exit. A second strongly conserved surface is found at the back of the active site, involving residues Ser285, Tyr289, Gln293 and Arg308. Residues 289 and 308 are conserved in 499 of 501 sequences (data not shown). The significance of the interaction between residues 293 and 308 is further highlighted by compensatory mutations in *Pyrococcus horikoshii* PSAT (PDB entry 2dr1; RIKEN Structural Genomics/Proteomics Initiative, unpublished work), in which the R308E mutation is compensated by a Q293R change maintaining a strong interaction between these residues.

4. Conclusions

Several key insights emerge from the structural analysis of MtbPSAT. Firstly, although this *Mtb* enzyme shows only low sequence identity to the human and eubacterial PSATs and belongs to a distinct phylogenetic clade shared with the archaeal enzymes, its PLP-binding site is highly conserved and constrained by strong specific interactions. Indeed, the discovery that PLP in this structure is not ligated to the essential active-site lysine Lys200 enables us to compare the PLP binding in structures at different stages of the reaction cycle and to show that the cofactor remains firmly anchored in this site throughout. Secondly, although many of the residues implicated in substrate binding are individually conserved in both the *Mtb* and other PSAT enzymes, there are specific differences in the *Mtb* enzyme that could be exploited for selective structure-based inhibitor design, notably the presence of Tyr341 in place of the smaller His328 of the *E. coli* enzyme and the altered loop that closes the substrate-binding site. Perhaps more significant, given the location of the active sites in the dimer interface, is the altered dimer of the *Mtb* enzyme, which results in restricted access to the active site. This could act as an anchor for inhibitors that could then be extended further into the active site.

This work was supported by the Health Research Council of New Zealand and the Foundation for Research, Science

and Technology of New Zealand (grants to ENB). FC also acknowledges support from a Career Development Award (545995) from the Australian National Health and Medical Research Council.

References

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M. & Bricogne, G. (2004). *Acta Cryst.* D**60**, 2210–2221.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Dey, S., Hu, Z., Xu, X. L., Sacchettini, J. C. & Grant, G. A. (2005). *J. Biol. Chem.* **280**, 14884–14891.

Dubnovitsky, A. P., Kapetaniou, E. G. & Papageorgiou, A. C. (2009). *Protein Sci.* **14**, 97–110.

Dubnovitsky, A. P., Ravelli, R. B., Popov, A. N. & Papageorgiou, A. C. (2009). *Protein Sci.* **14**, 1498–1507.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.

Hasan, S., Daugelat, S., Rao, P. S. S. & Schreiber, M. (2006). *PLoS Comput. Biol.* **2**, e61.

Hester, G., Stark, W., Moser, M., Kallen, J., Marković-Housley, Z. & Jansonius, J. N. (1999). *J. Mol. Biol.* **286**, 829–850.

Ichihara, A. & Greenberg, D. M. (1955). *Proc. Natl Acad. Sci. USA*, **41**, 605–609.

Islam, M. M., Hayashi, H., Mizuguchi, H. & Kagamiyama, H. (2000). *Biochemistry*, **39**, 15418–15428.

Jansonius, J. N. (1998). *Curr. Opin. Struct. Biol.* **8**, 759–769.

Kapetaniou, E. G., Thanassoulas, A., Dubnovitsky, A. P., Nounesis, G. & Papageorgiou, A. C. (2006). *Proteins*, **63**, 742–753.

Kirsch, J. F., Eichele, G., Ford, G. C., Vincent, M. G., Jansonius, J. N., Gehring, H. & Christen, P. (1984). *J. Mol. Biol.* **174**, 497–525.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Mdluli, K. & Spigelman, M. (2006). *Curr. Opin. Pharmacol.* **6**, 459–467.

Moreland, N., Ashton, R., Baker, H. M., Ivanovic, I., Patterson, S., Arcus, V. L., Baker, E. N. & Lott, J. S. (2005). *Acta Cryst.* D**61**, 1378–1385.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

O'Leary, S. E., Jurgenson, C. T., Ealick, S. E. & Begley, T. P. (2008). *Biochemistry*, **47**, 11606–11615.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.

Peisach, D., Chipman, D. M., Van Ophem, P. W., Manning, J. M. & Ringe, D. (1998). *Biochemistry*, **37**, 4958–4967.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Quayle, J. R. (1972). *Adv. Microb. Physiol.* **7**, 119–203.

Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol.* **48**, 77–84.

Schneider, G., Käck, H. & Lindqvist, Y. (2000). *Structure*, **8**, R1–R6.

Stierand, K., Maass, P. C. & Rarey, M. (2006). *Bioinformatics*, **22**, 1710–1716.

Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

Sugio, S., Petsko, G. A., Manning, J. M., Soda, K. & Ringe, D. (1995). *Biochemistry*, **34**, 9661–9669.

Tabatabaie, L., Klomp, L. W., Berger, R. & de Koning, T. J. (2010). *Mol. Genet. Metab.* **99**, 256–262.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.