Crystal structure of trehalose-6-phosphate phosphatase-related protein: Biochemical and biological implications

KRISHNAMURTHY N. RAO,¹ DESIGAN KUMARAN,¹ JAYARAMAN SEETHARAMAN,¹
JEFFREY B. BONANNO,² STEPHEN K. BURLEY,³ AND
SUBRAMANYAM SWAMINATHAN¹

¹Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA

²New York Structural Biology Center, New York, New York 10027, USA

³SGX Pharmaceuticals, Inc., San Diego, California 92121, USA

(RECEIVED January 24, 2006; FINAL REVISION March 30, 2006; ACCEPTED April 4, 2006)

Abstract

We report here the crystal structure of a trehalose-6-phosphate phosphatase-related protein (T6PP) from *Thermoplasma acidophilum*, TA1209, determined by the dual-wavelength anomalous diffraction (DAD) method. T6PP is a member of the haloacid dehalogenase (HAD) superfamily with significant sequence homology with trehalose-6-phosphate phosphatase, phosphoserine phosphatase, P-type ATPases and other members of the family. T6PP possesses a core domain of known α/β -hydrolase fold, characteristic of the HAD family, and a cap domain, with a tertiary fold consisting of a four-stranded β -sheet with two α -helices on one side of the sheet. An active-site magnesium ion and a glycerol molecule bound at the interface between the two domains provide insight into the mode of substrate binding by T6PP. A trehalose-6-phosphate molecule modeled into a cage formed by the two domains makes favorable interactions with the protein molecule. We have confirmed that T6PP is a trehalose phosphatase from amino acid sequence, three-dimensional structure, and biochemical assays.

Keywords: enzymes; structure/function studies; structure; crystallography; protein structures; structural genomics; phosphatase

Supplemental material: see www.proteinscience.org

Prokaryotic and eukaryotic organisms contain a large family of magnesium-dependent phosphatase/phosphotransferase enzymes that is a subset of the larger L-2-haloacid dehalogenase (HAD) superfamily (Collet et al. 1998; Selengut and Levine 2000; Wang et al. 2001). Enzymes of this family are structurally different from the α/β -hydrolase family (pfam00561). The HAD superfam-

ily is comprised of various phosphatases, epoxide hydrolases, P-type ATPases, and L-2-haloacid dehalogenases (Koonin and Tatusov 1994). Most of them are multidomain proteins that share a phosphatase domain with an α/β -hydrolase fold, typical of the ubiquitous hydrolase family, and a cap domain.

These enzymes utilize a common mechanism and are characterized by three highly conserved sequence motifs (Collet et al. 1998; Morais et al. 2000; Wang et al. 2001). Motif I near the N terminus contains DXXX(T/V), in which the first aspartate forms a phosphoaspartate intermediate with the substrate (Collet et al. 1998; Wang et al. 2002). The third residue in this motif is also an Asp in phosphotransferases and phosphatases and plays a significant role in catalysis (Collet et al. 1998). Motif II, (S/T)GX, contains a conserved serine or threonine, which is thought to form

Reprint requests to: Subramanyam Swaminathan, Biology Department, Brookhaven National Laboratory, P.O. Box 5000, Upton, NY 11973, USA; e-mail: swami@bnl.gov; fax: (631) 344-3407.

Abbreviations: T6PP, trehalose-6-phosphate phosphatase; HAD, haloacid dehalogenase; DAD, dual wavelength anomalous diffraction; pNPP, *para*-Nitrophenyl phosphate; DTT, dithiothreitol; PSP, phosphoserine phosphatase; T6P, trehalose-6-phosphate.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.062096606>.

a hydrogen bond to a phosphoryl oxygen atom (Wang et al. 2001). Motif III, containing K(X)_{16–30}(G/S)(D/S)XXX(D/N), forms part of the active site including the residues serving as metal ligands (Morais et al. 2000; Wang et al. 2001). Depending on the length of the sequence between these motifs, the HAD superfamily is divided into three structural subgroups: The first group has an all-helical insertion domain (called cap domain) between motifs I and II. The second group contains an α/β domain that is located between motifs II and III. This group contains several proteins, including trehalose and sucrose phosphatases. The third group contains no insertion (Selengut and Levine 2000). Crystal structures of several members of the HAD superfamily have been determined (Welch et al. 1998; Ridder et al. 1999; Toyoshima et al. 2000; Wang et al. 2001; Galburt et al. 2002; Lahiri et al. 2002; Parsons et al. 2002; Rinaldo-Matthis et al. 2002; Wang et al. 2002). Phosphoserine phosphatase belongs to subgroup I (Wang et al. 2001). Subgroup II is further divided into IIA and IIB based on the topology (Lu et al. 2005). While 1PW5 is the only structural representative of subgroup IIA, T6PP, a few putative phosphatases, YbiV (Protein Data Bank [PDB] ID 1RLM; Roberts et al. 2005), TM0651 (PDB ID 1NF2; Shin et al. 2003) and TA0175 (PDB ID 1L6R; Kim et al. 2004), HAD-like hydrolase (PDB ID 1NRW), BT4131 (PDB ID 1YMQ; Lu et al. 2005), YidA (PDB ID 1RKQ), Apc014 (PDB ID 1KYT), and sucrose phosphatase (PDB ID 1TJ3; Fieulaine et al. 2005) all belong to IIB subdivision.

Based on the sequence analysis, T6PP has been classified as trehalose-6-phosphate phosphatase-related protein (Ruepp et al. 2000). Trehalose-6-phosphate phosphatases catalyze dephosphorylation of trehalose-6-phosphate (T6P) to trehalose and orthophosphate. Trehalose, a common disaccharide of bacteria, fungi, and invertebrates, is important for organism survival under stress conditions. T6PP was selected for structural studies to gain insight into the function of the protein and to define the function of other close sequence relatives. Here, we report a high-resolution (1.92 Å) X-ray crystal structure of trehalose-6-phosphate phosphatase-related protein (T6PP) from *Thermoplasma acidophilum* and confirm that this enzyme is indeed a trehalose-6-phosphate phosphatase. We analyze its function and substrate specificity based on sequence and structure comparison of this protein with other members of HAD family and by biochemical experiments.

Results and Discussion

Overall structure of T6PP

The T6PP protomer is composed of two domains, including a core α/β domain with a modified Rossmann fold of the ubiquitous hydrolase family with two extra β -strands after $\beta 3$ forming a β hairpin and a cap domain

also with an α/β fold. A ribbons diagram and a topological representation are given in Figure 1, A and B. The molecule has approximate dimensions of $55 \times 35 \times 35$ Å³. The core domain is an $\alpha\beta\alpha$ three-layer sandwich, with an eight-stranded pleated β -sheet. The order of the strands in the polypeptide chain is $\beta 5-\beta 4-\beta 3-\beta 2-\beta 1-\beta 10-\beta 11-\beta 12$, with $\beta 5\beta 4$ forming the β hairpin. Five α -helices, three on one side ($\alpha 1$, $\alpha 2$, $\alpha 7$) and two on the other ($\alpha 5$, $\alpha 6$), flank the eight-stranded β -sheet. Although the fold of the core domain is well conserved among HAD superfamily members, the cap domain structures vary in size and structure (Lu et al. 2005). The cap domain of T6PP, comprised of residues 85–156,

Figure 1. (A) Ribbons representation of structure of T6PP (hydrolase domain, blue; cap domain, red). The magnesium ion (magenta) and its coordinating residues (yellow) are shown in ball-and-stick model. The two sodium ions are shown in purple, and the two bound glycerol molecules are shown as green ball-and-stick figures. (B) Topology diagram for the structure of T6PP. β -strands are shown as arrows; α -helices, as cylinders.

displays a four-stranded β -sheet with the strand order $\beta 6$ – $\beta 7$ – $\beta 9$ – $\beta 8$, covered by two helices ($\alpha 3$, $\alpha 4$) on one side with $\alpha\beta\beta\alpha\beta\beta$ topology. The cap domain is connected to the core domain, via two hinge loops comprised of residues 79–88 and 155–160.

Subtype II family phosphatases are classified as IIA and IIB based on the size of the domain and its topology (Lu et al. 2005). While IIB has an $\alpha\beta\beta(\alpha\beta\alpha\beta)\alpha\beta\beta$ fold, IIA has an $\alpha\beta\alpha\beta\beta\alpha\beta\alpha\beta$ fold. T6PP belongs to subtype IIB, and sequence among the trehalose phosphatases suggests that they all will belong to the same structural subgroup.

Active-site architecture

The active site is located within a cavity at the interface of the core and the cap domains. The core α/β domain serves as the base and side walls of the active site, while the cap domain serves as the cover. A magnesium ion and residues belonging to the three conserved sequence motifs define the active site. Aspartate and threonine residues are clustered within the active site including aspartates 7, 9, 179, 180, and 183 and threonines 11, 45, and 182, plus Arg47 and Lys161. Superposition of active-site residues with other HAD family members demonstrates conservation of most of these active-site residues (Shin et al. 2003). An extensive H-bond network provides connections among the Mg^{2+} , the active-site residues, sodium ions, and water molecules (Fig. 2A). The guanidine group of Arg47 makes a salt bridge with the side chain of Asp9. The carbonyl oxygen of Asp9 coordinates the Mg^{2+} ion. The side chain of Asp9 (which is in motif I) also interacts with a sodium ion located near the active site. This ion-pair network may be important for the structural integrity and/or catalytic activity of T6PP. A strong negative charge distribution was observed at the active site (Supplemental Material A), which is a common feature of HAD phosphatases (Shin et al. 2003; Kim et al. 2004; Roberts et al. 2005). Asp7, the first residue of motif I of the HAD superfamily, is located at the C terminus of the first β -strand ($\beta 1$), as is typical of the active sites of the HAD enzyme family (Seal and Rose 1987; Collet et al. 1998). The Mg^{2+} ion is also coordinated by the carboxylate groups of Asp7 and Asp179 and the carbonyl oxygen of Asp9. Three water molecules complete the nearly octahedral coordination (Fig. 2A). Other catalytically important conserved residues include Thr45 from motif II and Asp180 (both interacting via water molecules), Asp183, and Lys161 from motif III. Thr45 interacts with the catalytic residue Asp9 via a water molecule bridge. The side chain of Lys161 interacts with Asp9 directly. The side chains of Asp180 and 183 interact with a water molecule coordinated to the Mg^{2+} and Lys161, respectively.

Two sodium ions are positioned about 8 Å from the magnesium position. One of the sodium ions interacts with Asp9. Presumably the sodium ion is required in this position for neutralizing the overall charge. The other

sodium ion interacts distantly via two water molecules with Asp179, which is coordinated to magnesium ion. One of the glycerol molecules lies at the interface of the two domains near the active site. The other glycerol is on the surface of the core domain away from the cap domain. All of the interactions involving the surface bound glycerol involve protein atoms. The O1 atom interacts with residues Asn61 and Thr41; O2 atom interacts with Lys36, Phe39, and Asn41; and the O3 atom interacts with Asp59.

Structural homologs

The DALI automated search for structural homologs in the PDB (Holm and Sander 1997) revealed structural similarity with several members of HAD family. Putative phosphatases from subgroup II represent the closest structural homologs to T6PP, with the following root-mean-square deviations (RMSD): 4.1 Å (203 C α pairs) for TM0651 (1NF2; Shin et al. 2003), 4.1 Å (199 C α pairs) for YbiV (1RLM; Roberts et al. 2005), 3.7 Å (190 C α pairs) for Apc014 (1KYT), 4.8 Å (194 C α pairs) for sucrose phosphatase (1TJ3; Fieulaine et al. 2005), 4.3 Å (192 C α pairs) for BT4131 (1YMQ; Lu et al. 2005), 4.0 Å (201 C α pairs) for the hydrolase (1NRW), 3.5 Å (197 C α pairs) for TA0175 (1L6R; Kim et al. 2004), and 3.7 Å (187 C α pairs) for YidA (1RKQ). Structurally homologous proteins with defined functions include deoxy-d-mannose-octulosonate 8-phosphate phosphatase (1J8D), homoserine kinase (1RKU), β -phosphoglucomutase (1O03), phosphonoacetaldehyde hydrolase (1FEZ), and phosphoserine phosphatase (PSP) (1F5S).

The core α/β Rossmann fold is common to all homolog structures. In contrast, the architecture of the cap domain of T6PP differs from those found in even the closest structural homologs, TM0651 and YbiV. The cap domain of TM0651 consists of a six-stranded β -sheet surrounded by four α -helices, whereas the cap domain of T6PP consists of a four-stranded β -sheet and two α -helices. It is remarkable that the active sites of all HAD family members are superimposable on one another, despite their belonging to different subgroups. The Mg^{2+} ion and the conserved residues are located in almost the same position in the homolog proteins. The conserved residues found on four loops of the core domain correspond to the three sequence motifs that are characteristic of the HAD superfamily. It is remarkable that the type IV β -turn found near the active site between residues Leu9 and Leu13 in T6PP forms a π -helix, as in the structures of TM0651 and PSP. In all other structural homologs, this region of the structure is a type II β -turn or type IV β -turn and not a π -helix.

T6PP was also compared with the related structures of OtsA (PDB ID 1UQT) and trehalose repressor (TreR) from *Escherichia coli* (PDB ID 1BYK; Hars et al. 1998;

Figure 2. (A) The active site of T6PP, showing a hydrogen-bonding network involving residues conserved among HAD phosphatases. Hydrogen bonds are shown as dashed lines, in black. Water molecules and sodium ions are shown as red and as purple spheres, respectively. Figure prepared by MOLSCRIPT (Kraulis 1991). (B) Stick figure stereo view of glycerol binding at the interface of the core and the cap domains. Color coding: protein, yellow; glycerol, green; hydrogen bonds, black dashed lines. Figure prepared using MOLSCRIPT (Kraulis 1991).

Gibson et al. 2004). OtsA is required for biosynthesis of trehalose-6-phosphate, and TreR, which is induced by trehalose-6-phosphate, is known to regulate transcription of the *treB* and *treC* genes (Horlacher and Boos 1997). The structural similarity of OtsA and TreR with T6PP is limited to the core domain, while the folds of cap domains are different in both size and topology. Moreover, the relative positioning of the cap domain with respect to the core domain in T6PP differs from that seen in both OtsA and TreR. The cap domains in OtsA and TreR are rotated by 120° with respect to an axis passing through the center of the core domain as compared with position of the cap domain of

T6PP. Finally, T6P binds at the interface of the core and cap domains in both T6PP and OtsA.

Sequence analysis

PSI-BLAST searches with T6PP identified several HAD hydrolases as sequence homologs (E-value > 0.002). Among the closest matches are trehalose-6-phosphate phosphatases from *Thermoplasma acidophilum* DSM 1728 and *Thermoplasma volcanium* GSS1 and trehalose-6-phosphate synthetase from *Crocospheera watsonii* WH (96%–57% identity). A list of the 10 most closely

related PSI-BLAST matches is provided as Supplemental Material B. Not surprisingly, sequence alignments of the T6PPs show homology throughout the length of the polypeptide chain and identity or close similarity among active-site residues. Trehalose-6-phosphate synthase resembles T6PP only within the C-terminal domain. A sequence alignment of phosphatases from the second subgroup family for which structures are available is given in Figure 3. Most absolutely conserved residues occur in the active-site motifs. In addition, there are strongly conserved residues in the alignment among the structural homologs, which are found at the interface of the core and the cap domains. Sequence analysis revealed potential functional relevance, for instance, conserved

substrate-binding site, among these phosphatases. Among the T6PPs, a tetrapeptide segment (residues 15–18; “IIMN” of T6PP) that does not occur in any other second subgroup family proteins appears to contribute to the specificity loop (residues 14–23).

Substrate binding

The substrate specificity for the phosphatases is thought to be dictated by the surface at the interface of the core and cap domains (Wang et al. 2001; Shin et al. 2003). Residues from the core domain (Asp7, Asp9, Pro14, Ile15, Ile16, Pro19, Glu20, Thr45–Arg47, Arg54, Phe55, Tyr65, His66, Lys161, Asp179, Asp180, Thr182,

1U02	---MSLIFLDYDGTLPVPIIMNPESYADAGLLSLISDLKER-FDTYIATGRSPEEISRF 55
1NF2	---MYRVFVFDLDGDTLLN---DNLEIS-EKDRRNIEKLSRK-CYVVFASGRMLVSTLNV 51
1RLM	--MAVKVIVTDMDDGTFLN---DAKTYNQPRFMAQYQELKKRGKIFVVASGRQYYQLISF 54
1L6R	---MIRLAADVDGDLTD---RDRLIS-TKAIESIRSAEKKGLTVSLLSGNVIPVVYAL 52
1YMQ	---MTKALFFDIDGTLVSF---ETHRIP-SSTIEALEAAHAKGLKIFATGRFP-KAITNN 52
1RKQ	MSLAIKLIAIDMDGTLN---PDHTIS-PAVKNAIAAARAGVNVVLTGRFPYAGVHNY 55
1NRW	--SNAMKLIADLDGDTLLN---SKHQVS-LENENALRQAQRDGEVNVSTGRAHFDVMSI 54
1TJ3	--MRQLLLISDLDTWVG---DQQALE---HLQEYLGDRRGNFYLAYATGRSYHSAREL 51
	* * : *
1U02	LPLDINMICYHGACSKINGQIVYNNNGSDRFLGVFDRIYEDTR--SWVSDFPGLRIYRKNL 113
1NF2	EK-KYFKRTFP--TIAYNGAIVYLPEEGVILNEKIPPEVAKDIEYIKPLN-VHWQAYID 107
1RLM	FP-ELKDEIS---FVAENGALVYEHGKQLFHGELTRHESRIVIGELLKDKQ-LNFVACGL 109
1L6R	KI-FLGINGP---VFGENGIMFDNDGSIK--KFSNENGTNKFLEEMSKRT-----SM 99
1YMQ	LS-ELQDRNLIDGYITMNGAYCFVG-EVVIKSAIPQEEVKAMAAAFCEKKG-VPCIFVEE 109
1RKQ	LK-ELHMEQPGDYCITYNGALVQKAADGSTVAQTALSYDDYRFLKLSREVGSHPHALDR 114
1NRW	FE-PLGIKT---WVISANGAVIHDPEGRLYHHETIDKKRAYDILSWLESEN-YYVEVPTG 109
1TJ3	QK-QVGLMEPDYWLTAVGSEIYHPEGLDQHWADYLSHWDQDILQAIADGF----EALKP 106
	
1U02	AVLYHL-----GLMGADMK----- 127
1NF2	DVLYSEKDNEEIKSYARHSNVDYR-VEP---NLSELVSKMGT-----KLLLI 151
1RLM	QSAYVS-ENAEAFVALMAKHVHR-LKP---VKDYQEIDDVLFK-----FSL 151
1L6R	RSILTN-----RWREASTG-----FDI 116
1YMQ	HNISVC--QPNEVMKKIPYDFLHVNVIP-----TVSFEEASNK-----EVIQ 149
1RKQ	TTLYTA--NRDISYTVH-ESFVA-TIP---LVFCEAEKMDPN-----TQFLKVMMI 159
1NRW	SAIYTP--QNGRELLDVELDRFRS-ANPEADLSVLKQAAEVQYSQSGFAYINSFQELFEA 166
1TJ3	QSPLEQ-----NP-----WKISYHLDPQ----- 124
	
1U02	--P-----KLRSRIEETARIFG-----VETYYGKMIIELRVPGVNGSAIRSVRG 170
1NF2	DTP-----ERLD-ELKEILSER-----FKDVVKVFKSFPTYLEIVPKNVDKGKALRFLRE 200
1RLM	NLP---DEQIPLVIDKLHVAL-----DGIMKPVTSFGFGFIDLIIPLGLHKANGISRLK 201
1L6R	DPE---DVDYVRKEAES-----RGFVIFYSG-YSWHLMNRGEDKAFVAVMKLKE 160
1YMQ	MTP-----FITEEEEKEVLPS-----IPTCEIGRWYPAFADVTAKGDTKQKGIIDEIIR 197
1RKQ	DEP-----AILDQAIAIRIQEV-----KEKYTVLKSAPYFLEILDKRNVNKGKGVKSLAD 208
1NRW	DEPIDFYNILGFSFFKEKLEAGWKRYEHAEDLTIVSSAEHNFELSSRKASKGQALIKRLAK 226
1TJ3	ACP-----TVIDQLTEMLKET-----GIPVQVIFSSGKDVLLFPQRSNKGKMATQYLQ 172
	
1U02	E-----RPATIAAGDDATDEAAFEANDDALTIKVGEGETHAK-----FHVADYIEMRKIL 219
1NF2	RMNWKKEEIVVFGDNENDLFMFEE--AGLRVAMENAIKVK-----EASDIVTLTNND 251
1RLM	RWDLSPQNVVAIGDSGNDAMLMK--ARYSFAMGNAAENIK-----QIARYATDDNNH 252
1L6R	MYSLEYDEILVIGDSNNDMPMFQL--PVRKACPANATDNK-----AVSDFVSDYSYG 211
1YMQ	HFGIKLEETMSFGDGGNDISMLRH--AAIGVAMQAKEDVK-----AAADYVTAIDE 248
1RKQ	VLGIKPEEIMAGDQENDIAMIEY--AGVGVAVDNAIPSVK-----EVANFVTKSNLE 259
1NRW	QLNIPLLEETAAGVDSLNDKSMLEA--AGKGVAMGNAREDIK-----SIADAVTLTNDE 277
1TJ3	HLAMEPSQTLVCGDSGNDIGLFET--SARGVIVRNAQPELLHWYDQWDSRHYRAQSSHA 230
	** * : *
1U02	KFIEMLVQKKQEGHHHHH---- 239
1NF2	SGVSYV-LERISTDCLDE----- 268
1RLM	EGALNV-IQAVLDNTYPFNS---- 271
1L6R	EEIGQT-FKHFFELM----- 224
1YMQ	DGSKA-MKHFGII----- 261
1RKQ	DGVAFV-IEKYVLNEGSHHHHH 282
1NRW	HGVAHM-MKHLL----- 288
1TJ3	GAILEA-IAHFDPLS----- 244

Figure 3. Sequence alignment of second subgroup type IIB phosphatases for which experimental structures are available, identified by their respective PDB ID codes: T6PP (1U02), TM0651 (1NF2; Shin et al. 2003), YbiV (1RLM; Roberts et al. 2005), TA0175 (1L6R; Kim et al. 2004), BT4131 (1YMQ; Lu et al. 2005), YidA (1RKQ), hydrolase (1NRW), and sucrose phosphatase (1TJ3; Fieulaine et al. 2005). Residues that form the three characterizing motifs are shown in bold. The alignment was constructed using CLUSTAL-W with the *bottom* row providing the degree of homology: (*) identical, (:) conserved, (.) semiconserved.

and Asp183) and the cap domain (Tyr109, Lys111, Leu116, His118, Tyr146, Gly148, Lys149, Ile151, Glu153, and Arg155) form a pocket near the active site. Interestingly, one of the glycerol ligands is located within this specificity pocket of T6PP (Fig. 2B). The presence of glycerol and several water molecules in this pocket suggests that substrates of substantial size could bind in this region. The glycerol hydroxyl groups form a network of hydrogen bonds with nearby water molecules. These water molecules in turn interact with several protein atoms, from residues Asp9, Ile15, Pro19, Arg47, and His118. The O1 atom of glycerol interacts with the active-site magnesium ion and Asp179 via two water molecules. We speculate that the glycerol bound in our T6PP structure mimics the position of carbohydrate substrates. Further, the volume of the cavity formed between the core and the cap domains is $\sim 930 \text{ \AA}^3$ (Binkowski et al. 2003), which is more than sufficient to accommodate phosphorylated sugar molecules, such as trehalose-6-phosphate (estimated volume = $320 \text{ \AA}^3$). Carbohydrate binding to proteins is often accompanied by stacking interaction with aromatic residues (Pratap et al. 2002; Rao et al. 2004), which suggests that Tyr23, Tyr65, Tyr78, Tyr109, Tyr146, His66, and His118 may contribute to substrate binding in T6PP. It appears likely that the pocket in the interface between the core and the cap domains in T6PP could accommodate a phosphorylated mono- or disaccharide and present the phosphate group to the appropriate active-site residues.

Modeling trehalose-6-phosphate in the putative substrate-binding site

We modeled the binding of trehalose-6-phosphate (T6P) ligand to the active site of T6PP (Fig. 4A). The modeled complex was manually placed with O using the positions of the Mg ion and the glycerol molecule in our structure, and sulfate ions in homologous structures, and then energy-minimized using CNS. T6P was found to fit well in the pocket formed by the interface of the core and the cap domain near the enzyme active site. The sugar group of T6P was located at the opening of the cleft between the two subdomains, while the phosphate group was buried inside the active site of T6PP, where it can interact with basic residues (e.g., Lys161). The phosphoryl group also interacts with OG1 atom of Thr45. Lys161 and Thr45 are the characteristic motif residues that are involved in substrate binding and catalysis as discussed below. The disaccharide sugar sits in a cage formed by the substrate specificity loops (residues 14–23 and 144–155; Fig. 4B). There are four potential hydrogen bond interactions between T6P and the protein: one from each of the three β -strands of the cap domain and one from Ile16 from the core domain.

Sequence analysis of trehalose phosphatases suggests that the putative interactions between T6P and the cap domain are also supported by other family members because the hydrogen bonding residues (Lys111, His118, Lys149, and Glu153) are conserved within trehalose phosphatase family (Supplemental Material B). Examination of structures of other phosphatases revealed a more open substrate-binding cavity at the core-cap domain interface and that may accommodate different, larger sugar phosphates. In contrast, the substrate-binding cavity of T6PP appears to fit T6P tightly, suggesting that this enzyme may be somewhat more substrate-specific. The volumes occupied by the substrate-binding cavity of TA1075 (1L6R), BT4131 (1YMQ), and T6PP are 393, 468, and $926 \text{ \AA}^3$, respectively. Interestingly, the substrates for TA1075 and BT4131 are phosphoglycolate (estimated molecular volume = $110 \text{ \AA}^3$) and monosugar phosphate (estimated molecular volume = $198 \text{ \AA}^3$), which may require smaller substrate binding pockets.

Catalytic mechanism

It is generally accepted that the HAD α/β -hydrolases utilize a covalent enzyme-substrate intermediate (Collet et al. 1998). The T6PP active site superimposes well with that of the catalytically best characterized HAD phosphatase, phosphoserine phosphatase, or PSP. The four backbone amide groups of Asp9, Gly10, Gly46, and Asp183 and the side chains of Thr45 and Lys161 have been implicated in PSP catalytic activity (Wang et al. 2001). The positions of the corresponding residues in T6PP are very similar to those in PSP, suggesting that the two enzymes employ similar catalytic mechanisms. The first aspartate (Asp7) present within motif I is believed to form a covalent intermediate with the substrate. The catalytic mechanism of HAD phosphatases involves a phospho-Asp intermediate resulting from inline nucleophilic attack of Asp7 on the phosphorus atom of the substrate. The phospho-Asp intermediate is hydrolyzed by an activated water molecule that regenerates the intact enzyme. Lys161 interacts with Asp7 and may, therefore, help neutralize the negative charge of Asp7 during the reaction. The second aspartate (Asp9) is implicated in activating the hydrolytic water molecule. Another suggested role for Asp9 is to act as a general acid catalyst for protonating the leaving-group oxygen atom following hydrolysis. Motif II presents either a serine or a threonine (Thr45 in T6PP) that interacts with the substrate phosphoryl oxygen. Motif III, proposed to accommodate much of the catalytic diversity exhibited by the HAD superfamily, provides a basic residue (Lys161 in T6PP) that interacts with the nucleophile and the substrate, and may increase the electrophilicity of the substrate. Motif III also carries an acidic residue (Asp179 in T6PP) as a ligand for the metal ion in the phosphatases and

Figure 4. (A) The substrate T6P (green) is modeled into the pocket at the interface of the core and the cap domains. The active site surrounding the magnesium ion (magenta) is also shown. Hydrogen bonds are represented by black dashed lines. (B) T6P modeled in the interface of core and cap domains. The cavity is formed primarily by the substrate specificity loop (residues 144–155, red) and another specificity loop (residues 14–23, blue).

phosphotransferases, whereas the haloacid dehalogenase, which lacks the metal, uses an asparagine residue at the corresponding location to position a hydrolytic water for attack on the ester intermediate of the reaction.

Biochemical assay results

Structural and sequence analysis suggests that the biochemical function of T6PP is unlikely to be that of a hydrolase, such as PSP, ATPase, L2-haloacid dehalogenase, phosphonoacetaldehyde hydrolase, YrbI, CheY, and CheB, because of the presence of Asp9 in the conserved

DXDGTX motif, which is found only in phosphatases and phosphotransferases. We screened T6PP for biochemical activity to investigate its ability to catalyze the release of phosphate from the T6P and pNPP. As expected, T6PP catalyzed release of phosphate from both T6P and pNPP; however, the catalytic efficiency of the enzyme markedly differed toward two substrates. The kinetic constants K_m and K_{cat} values for T6P are 2.7 mM and 10.0 sec^{-1} and those for pNPP are 17 mM and 0.8 sec^{-1} . The observed kinetic constants for T6PP using T6P as substrate reflect the high specificity typically associated with enzymatic activity (Parsons et al. 2002; Roberts et al.

2005). Similar kinetic properties were reported for other trehalose-6-phosphate phosphatases (Matula et al. 1971; Seo et al. 2000). (The kinetic constants observed for T6PP using pNPP as substrate are typical of nonspecific substrates.) Results from these assays are consistent with T6PP functioning as trehalose-6-phosphate phosphatase. When the enzyme is treated with EDTA, it loses its activity, suggesting that it is a magnesium-dependent enzyme. We have shown that the substrate-binding site is suitable for binding a disaccharide and possibly trehalose phosphate.

Biological implication

Genes encoding proteins with similar or related functions are often grouped within the genome or belong to the same operon or transcriptional unit. Such organization provides additional insights into possible biological function(s). In the Pfam database (Bateman et al. 2000), T6PP is thought to be related to the products of the *otsB* (pfam02358) and *otsA* (pfam00982) genes in *E. coli*, which reside in the same operon and are functionally well defined. Trehalose biosynthesis in *E. coli* depends on *otsB* and *otsA*, which are induced at high osmolarity conditions (Gibson et al. 2002). *otsB* codes for trehalose phosphatase, and *otsA* codes for trehalose-6-phosphate synthase. Within the *T. acidophilum* genome, the gene for T6PP is followed by a gene encoding trehalose phosphate synthase. Thus, it appears likely that T6PP is expressed in response to induction of trehalose synthesis.

Conclusions

We have reported here the crystal structure of T6PP from *T. acidophilum*, which represents the first structure of member of the trehalose-6-phosphate phosphatase family. Thus, the structure provides a framework for modeling other trehalose-6-phosphate phosphatases and planning site-directed mutagenesis experiments. Analyses of the crystal structure and available amino acid sequences and a phosphatase assay documented that T6PP is a phosphatase. Comparison with other phosphatase enzymes suggests that T6PP, like other phosphatases, may act through the formation of a phosphoaspartate intermediate. The structure of T6PP from *T. acidophilum* has been deposited with the PDB (ID 1U02).

Materials and methods

Cloning

The target gene for T6PP was amplified via PCR from *T. acidophilum* genomic DNA with the appropriate forward (ATATTCCTCGATTATGACGGCAC) and reverse (CTTGCTTT

TTCTGAACTCCTAAC) primers and Taq DNA polymerase (Qiagen) using standard methods (Deutscher 1990). Following gel purification, the PCR product was inserted into a modified pET26b vector for topoisomerase-directed cloning (Invitrogen) and designed to express the protein of interest with a C-terminal hexa-histidine tag. Transformation into BL21 (DE3) cells, followed by incubation (37°C for 60 min) yielded a culture that was plated onto LB/kanamycin and held at 37°C overnight. Twelve colonies were selected into microtiter wells containing 50 μ L of LB/kanamycin. The solutions appeared turbid after $\sim$ 4 h, and diagnostic PCR using a T7 sequencing primer and gene-specific reverse primer was conducted to confirm clones with the correct orientation. Preparations exhibiting the correct orientation were then used for DNA minipreps. Expression and solubility were tested by standard methods.

Expression and purification

A medium scale cell culture was grown by adding 500 mL LB medium (BD/Difco), 25 mL of 10% glucose (Sigma) solution, 500 μ L of 30 mg/mL kanamycin (Sigma/Aldrich) solution, and a small amount of transformed cell glycerol stock scraping to a 2 L baffled flask and shaking overnight at 250 rpm and 30°C. The expression medium was prepared by adding 25 mL of 20 $\times$ NPS buffer (BD/Difco), 10 mL of 50 $\times$ 5052 solution (BD/Difco), 500 μ L of 1 M MgSO₄, 1667 μ L of 30 mg/mL kanamycin solution, and 500 μ L of 30 mg/mL chloramphenicol solution (Sigma-Aldrich) to each 500 mL bottle of ZY media (BD/Difco). Two hundred fifty milliliters (250 mL) of this solution was added to each of eight flasks (2 L total); 5 mL of the overnight culture was added to each flask, and the cultures were held at 250 rpm and 37°C for 5 h. The temperature was then reduced to 22°C, and shaking was maintained overnight (Studier 2005).

The contents of the flasks were poured into 1 L spin bottles and spun at 6500 rpm for 10 min. After removal of the supernatant, the pellets were collected into 50 mL conical tubes (total mass 22.0 g) and frozen at -80°C . The pellet was resuspended in lysis buffer (35 mL/10 g) containing 50 μ L of protease inhibitor cocktail (Sigma) and 5 μ L benzamide (Novagen) and subjected to repeated sonication with intervals of cooling. The lysate was then clarified by centrifugation at 38,900g for 30 min. The protein was then immobilized on Ni-NTA resin (Qiagen), placed on a drip column, washed with 25 mL buffer A (50 mM Tris-HCl at pH 7.8, 500 mM NaCl, 10 mM imidazole, 10 mM methionine, and 10% glycerol), and eluted into Amicon concentrator (Millipore) with 15 mL buffer A containing 500 mM imidazole. The solution was concentrated to 6 mL, loaded onto an S200 gel filtration column, and run off with buffer containing 10 mM HEPES (pH 7.5), 150 mM NaCl, 10 mM methionine, 10% glycerol, and 5 mM DTT. The protein yield was 9.1 mg/L, which was concentrated to 2.5 mg/mL. Seleno-methionine-labeled protein was produced using M9 media with amino acids including SeMet instead of Met and purified in a similar manner.

Phosphatase assay

Initially, the phosphatase activity of T6PP was assayed against *para*-Nitrophenyl phosphate (pNPP) as substrate by continuously following production of *p*-nitrophenolate at 410 nm using Beckman DU530 spectrophotometer (Chen et al. 1956; Parsons

et al. 2002). A substrate concentration series was prepared by dissolving pNPP in a solution of 50 mM Tris-HCl (pH 8.0), 5 mM DTT, and 20 mM MgCl₂. The reaction was carried out in a total volume of 100 μ L using different substrate concentrations (5–50 mM) for 20 min at 50°C in the presence of T6PP and was stopped by diluting with 100 mM NaOH (final concentration). Reactions performed without T6PP served as control. The assay was also carried out keeping the substrate concentration fixed and varying the duration of the reaction time. The amount of reaction product released was calculated by measuring the difference in absorbance between the controls and the corresponding enzyme-containing reactions with a molar absorptive coefficient of 18,400 M⁻¹cm⁻¹ for pNPP (Parsons et al. 2002). Kinetic constants were determined from the rate of hydrolysis of the substrate over a range of concentrations using the Michaelis-Menten equation.

The trehalose-6-phosphate phosphatase activity was determined using trehalose-6-phosphate (T6P; molar absorptive coefficient 25,000 M⁻¹cm⁻¹) as substrate (1–10 mM). Release of inorganic phosphate was estimated by a published method (Chen et al. 1956). The experiment was also repeated with the enzyme treated with EDTA to remove any metal ion.

Crystallization and data collection

The preliminary screening with the native protein samples were carried out with Hampton high throughput screens (both Index and crystal screen HT) using a TECAN crystallization robot. Optimized sitting drop vapor diffusion crystallization of both the native and SeMet proteins involved mixing 2 μ L protein solution (5 mg/mL) with 2 μ L of reservoir solution, containing 30% PEG 4000 and 0.2 M MgCl₂, 0.1 M Tris-HCl (pH 8.5), and equilibration against 600 μ L of the same reservoir solution. Crystals appeared in 2 d and were frozen in liquid nitrogen using 15%–20% glycerol as cryo-protectant.

X-ray diffraction data were collected with a native protein crystal using beamline X29 at the National Synchrotron Light Source (Brookhaven National Laboratory). MAD data were collected with an SeMet crystal at X-ray wavelengths 0.9792 Å (peak), 0.9795 Å (inflection), and 0.94 Å (high energy remote). All data were processed and scaled using *HKL2000*. Crystals are in space group C2, with unit-cell parameters $a = 118.0$ Å, $b = 45.1$ Å, $c = 53.4$ Å, and $\beta = 90.7^\circ$. The Matthews unit, giving 45% solvent content in the crystal. Crystallographic statistics are given in Table 1.

Structure determination

Determination of the Se atom substructure using SOLVE was tried with three- and two-wavelength MAD data (Terwilliger and Berendzen 1999a, b). However, phasing statistics were better using peak and inflection data sets alone. Probably, the remote data set that was collected last suffered radiation damage. Accordingly, only the two-wavelength MAD data were used. The locations of all possible Se atoms identified by SOLVE were used for SHARP phase refinement (de la Fortelle and Bricogne 1997). Phases were further improved by density modification using DM (Cowtan 1994). Of a possible 239 residues (including the C-terminal hexa-histidine tag), 222 were built automatically by *ARP/wARP* (Perrakis et al. 1999). Subsequent model building and adjustments were done using *O* (Jones et al. 1991). Refinement using data collected at peak wavelength was performed with *CNS* (Brunger et al. 1998).

Table 1. Data collection and refinement statistics

Space group	C2	
Unit cell parameters	$a = 118.0$ Å, $b = 45.1$ Å, and $c = 53.4$ Å, $\beta = 90.7^\circ$	
Method: DAD	Peak	Edge
Wavelength (Å)	0.9792	0.9795
Resolution (Å)	50–1.92	50–1.97
No. of reflections	21,559	21,445
<i>R</i> -merge (%) ^a	0.07 (0.27)	0.08 (0.30)
Overall completeness (%)	99.2 (97.4)	99.0 (97.8)
Redundancy	5.7 (4.6)	5.4 (4.1)
Overall figure of merit	0.20	
After density modification	0.90	
Refinement statistics		
Resolution range (Å)	50–1.92	
No. of reflections	21,237	
<i>R</i> -factor ^b	0.197	
<i>R</i> -free ^c	0.216	
No. of protein atoms	1817	
No. of metal ions or glycerol molecules	5 (1 Mg, 2 Na, and 2 glycerol molecules)	
No. of water molecules	185	
Root-mean-square deviations		
Bond lengths (Å)	0.005	
Bond angles ($^\circ$)	1.3	

The values corresponding to the outermost shell are given within parentheses.

^a*R*-merge = $\sum_i \sum_j |I_{h,i} - \langle I_h \rangle| / \sum_i \sum_j I_{h,i}$ where $\langle I_h \rangle$ is the mean intensity of symmetry-related reflections, $I_{h,i}$.

^b*R*-factor = $\sum ||F_o| - |F_c|| / \sum |F_o|$ where F_o and F_c are the observed and the calculated structure factor amplitudes, respectively.

^c*R*-free is calculated for the 5% of the data that were withheld from refinement.

Simulated annealing refinement using slow cool protocol was followed by Powell energy minimization. In the final stages of refinement, water molecules were added to the model using residual peaks $>3\sigma$ with potential hydrogen bonding interactions from $|F_o| - |F_c|$ Fourier difference synthesis. There were three strong residual peaks that could not be attributed to water molecules. The most significant difference electron density feature, located close to Asp7, with near octahedral coordination to protein atoms and water molecules was modeled as Mg ion, since Mg was present in the crystallization condition. The remaining two difference electron density features were located 2.0–2.2 Å from protein atoms and were modeled as sodium ions consistent with their peak heights and coordination geometry and distances (Harding 2000, 2001, 2004). In addition, two glycerol molecules were included in the refinement model. The final atomic model includes residues 2–230 out of a total of 239 residues in the polypeptide chain. There was no interpretable electron density for Met1 and the 9 C-terminal residues. The quality of the model was examined using *PRO-CHECK* (Laskowski et al. 1993). Refinement statistics are given in Table 1.

Acknowledgments

Research supported by a National Institutes of Health grant (GM62529) to the NYSGXRC under the auspices of the Protein Structure Initiative (DOE Prime Contract no. DEAC02-98CH10886

to Brookhaven National Laboratory). We thank Dr. H. Robinson for data collection facilities at beamline X29, NSLS.

References

- Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., and Sonnhammer, E.L. 2000. The Pfam protein families database. *Nucleic Acids Res.* **30**: 276–280.
- Binkowski, T.A., Naghibzadeh, S., and Liang, J. 2003. CASTp: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* **31**: 3352–3355.
- 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., et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**: 905–921.
- Chen, P.S., Toribara, T.Y., and Warner, H. 1956. Microdetermination of phosphorus. *Anal. Chem.* **28**: 1756–1758.
- Collet, J.F., Stroobant, V., Pirard, M., Delpierre, G., and Schaftingen, E.V. 1998. A new class of phosphotransferases phosphorylated on an aspartate residue in an amino-terminal DXDX(T/V) motif. *J. Biol. Chem.* **273**: 14107–14112.
- Cowan, K. 1994. Joint CCP4 ESF-EACBM newsl. *Protein Crystallogr.* **31**: 34–38.
- de la Fortelle, E. and Bricogne, G. 1997. Maximum-likelihood heavy atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* **276**: 472–493.
- Deutscher, M.P. 1990. Gene expression technology. *Methods Enzymol.* **185**: 3–681.
- Dieulaire, S., Lunn, J., Borel, F., and Ferrer, J. 2005. The structure of a cyanobacterial sucrose-phosphatase reveals the sugar tongs that release free sucrose in the cell. *Plant Cell* **17**: 2049–2058.
- Galburt, E.A., Pelletier, J., Wilson, G., and Stoddard, B.L. 2002. Structure of a tRNA Repair enzyme and molecular biology workhorse: T4 polynucleotide kinase. *Structure* **10**: 1249–1260.
- Gibson, R.P., Lloyd, R.M., Charnock, S.J., and Davies, G.J. 2002. Characterization of *Escherichia coli* OtsA, a trehalose-6-phosphate synthase from glycosyltransferase family 20. *Acta Crystallogr. D Biol. Crystallogr.* **58**: 349–351.
- Gibson, R.P., Tarling, C.A., Roberts, S., Wither, S.G., and Davies, G.J. 2004. The donor subsite of trehalose-6-phosphate synthase: Binary complexes with UDP-glucose and UDP-2-deoxy-2-fluoro-glucose at 2 Å resolution. *J. Biol. Chem.* **279**: 1950–1955.
- Harding, M.M. 2000. The geometry of metal-ligand interactions relevant to proteins. II. Angles at the metal atom, additional weak metal-donor interactions. *Acta Crystallogr. D Biol. Crystallogr.* **56**: 857–867.
- . 2001. Geometry of metal-ligand interactions in proteins. *Acta Crystallogr. D Biol. Crystallogr.* **57**: 401–411.
- . 2004. The architecture of metal coordination groups in proteins. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 849–859.
- Hars, U., Horlacher, R., Boos, W., Welte, W., and Diederichs, K. 1998. Crystal structure of the effector-binding domain of the trehalose-repressor of *Escherichia coli*, a member of the LacI family, in its complexes with inducer trehalose-6-phosphate and noninducer trehalose. *Protein Sci.* **7**: 2511–2521.
- Holm, L. and Sander, C. 1997. Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res.* **25**: 231–234.
- Horlacher, R. and Boos, W. 1997. Characterization of TreR, the major regulator of the *Escherichia coli* trehalose system. *J. Biol. Chem.* **272**: 13026–13032.
- Jones, T.A., Zou, J., Cowtan, S., and Kjeldgaard, M. 1991. Improved methods in building protein models in electron density map and the location of errors in these models. *Acta Crystallogr. A* **47**: 110–119.
- Kim, Y., Yakunin, A.F., Kuznetsova, E., Xu, X., Pennycook, M., Gu, J., Cheung, F., Proudfoot, M., Arrowsmith, C.H., Joachimiak, A., et al. 2004. Structure- and function-based characterization of a new phosphoglycolate phosphatase from *Thermoplasma acidophilum*. *J. Biol. Chem.* **279**: 517–526.
- Koonin, E.V. and Tatusov, R.L. 1994. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search. *J. Mol. Biol.* **244**: 125–132.
- Kraulis, P.J. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**: 946–950.
- Lahiri, S.D., Zhang, G., Dunaway-Mariano, D., and Allen, K.N. 2002. Caught in the act: The structure of phosphorylated β -phosphoglucomutase from *Lactococcus lactis*. *Biochemistry* **41**: 8351–8359.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality for assessing the accuracy of protein structures. *J. Appl. Crystallogr.* **26**: 283–291.
- Lu, Z., Dunaway-Mariano, D., and Allen, K.N. 2005. HAD superfamily phosphotransferase substrate diversification: Structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. *Biochemistry* **44**: 8684–8696.
- Matula, M., Mitchell, M., and Elbein, A.D. 1971. Partial purification and properties of a highly specific trehalose phosphate phosphatase from *Mycobacterium smegmatis*. *J. Bacteriol.* **107**: 217–222.
- Morais, M.C., Zhang, W., Baker, A.S., Zhang, G., Dunaway-Mariano, D., and Allen, K.N. 2000. The crystal structure of *Bacillus cereus* phosphonoacetaldehyde hydrolase: Insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily. *Biochemistry* **39**: 10385–10396.
- Parsons, J.F., Lim, K., Tempezyk, A., Krajewski, W., Eisenstein, E., and Herzberg, O. 2002. From structure to function: YrbI from *Haemophilus influenzae* (HI1679) is a phosphatase. *Proteins* **46**: 393–404.
- Perrakis, A., Morris, R., and Lamzin, V.S. 1999. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**: 458–463.
- Pratap, J.V., Jeyaprakash, A.A., Rani, G.P., Sekar, K., Suroliya, A., and Vijayan, M. 2002. Crystal structures of Artocarpin, a moraceae lectin with mannose specificity, and its complex with methyl- α -D-mannose: Implications to the generation of carbohydrate specificity. *J. Mol. Biol.* **317**: 237–247.
- Rao, K.N., Suresh, C.G., Katre, U.V., Gaikwad, S.M., and Khan, M.I. 2004. Two orthorhombic crystal structures of a galactose-specific lectin from *Artocarpus hirsuta* in complex with methyl- α -D-galactose. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 1404–1412.
- Ridder, I.S., Rozeboom, H.J., Kalk, K.H., and Dijkstra, B.W. 1999. Crystal structures of intermediates in the dehalogenation of haloalkanoates by L-2-haloacid dehalogenase. *J. Biol. Chem.* **274**: 30672–30678.
- Rinaldo-Matthis, A., Rampazzo, C., Reichard, P., Bianchi, V., and Nordlund, P. 2002. Crystal structure of a human mitochondrial deoxyribonucleotidase. *Nat. Struct. Biol.* **9**: 779–787.
- Roberts, A., Lee, S., McCullagh, E., Silversmith, R., and Wemmer, D. 2005. Ybiv from *Escherichia coli* K12 is a HAD phosphatase. *Proteins* **58**: 790–801.
- Ruepp, A.G.W., Santos-Martinez, M.L., Koretke, K.K., Volker, C., Mewes, H.W., Frishman, D., Stocker, S., Lupas, A.N., and Baumeister, W. 2000. The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**: 508–513.
- Seal, S.N. and Rose, Z.B. 1987. Characterization of a phosphoenzyme intermediate in the reaction of phosphoglycolate phosphatase. *J. Biol. Chem.* **262**: 13496–13500.
- Selengut, J.D. and Levine, R.L. 2000. MDP-1: A novel eukaryotic magnesium-dependent phosphatase. *Biochemistry* **39**: 8315–8324.
- Seo, H., Koo, Y., Lim, J., Song, J., Kim, C., Kim, J.-K., Lee, J., and Choi, Y. 2000. Characterization of a bifunctional fusion enzyme between trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase of *Escherichia coli*. *Appl. Environ. Microbiol.* **66**: 2484–2490.
- Shin, D.H., Roberts, A., Jancarik, J., Yokota, H., Kim, R., Wemmer, D.E., and Kim, S.-H. 2003. Crystal structure of a phosphatase with a unique substrate binding domain from *Thermotoga maritima*. *Protein Sci.* **12**: 1464–1472.
- Studier, F.W. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**: 207–234.
- Terwilliger, T.C. and Berendzen, J. 1999a. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 849–861.
- . 1999b. Discrimination of solvent from protein regions in native Fouriers as a means of evaluating heavy-atom solutions in the MIR and MAD methods. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 501–505.
- Toyoshima, C., Nakasako, M., Nomura, H., and Ogawa, H. 2000. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **405**: 647–655.
- Wang, W., Kim, R., Jancarik, J., Yokota, H., and Kim, S.-H. 2001. Crystal structure of phosphoserine phosphatase from *Methanococcus jannaschii*, a hyperthermophile, at 1.8 Å resolution. *Structure* **9**: 65–71.
- Wang, W., Cho, H.S., Kim, R., Jancarik, J., Yokota, H., Nguyen, H.H., Grigoriev, I.V., Wemmer, D.E., and Kim, S.-H. 2002. Structural characterization of the reaction pathway in phosphoserine phosphatase: Crystallographic “snapshots” of intermediate states. *J. Mol. Biol.* **319**: 421–431.
- Welch, M., Chinardet, N., Mourey, L., Birck, C., and Samama, J.P. 1998. Structure of the CheY-binding domain of histidine kinase CheA in complex with CheY. *Nat. Struct. Biol.* **5**: 25–29.