doi:10.1016/j.jmb.2005.01.072 J. Mol. Biol. (2005) 348, 917–926

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Crystal Structure of Human E1 Enzyme and its Complex with a Substrate Analog Reveals the Mechanism of its Phosphatase/Enolase Activity

Hui Wang[1,2]†, Hai Pang[1,2]†, Mark Bartlam[1,2] and Zihe Rao[1,2]*

[1]*Laboratory of Structural Biology, School of Medicine Tsinghua University, Beijing 100084, China*

[2]*National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Science, Beijing 100101, China*

**Corresponding author*

Enolase-phosphatase E1 (MASA) is a bifunctional enzyme in the ubiquitous methionine salvage pathway that catalyzes the continuous reactions of 2,3-diketo-5-methylthio-1-phosphopentane to yield the aci-reductone metabolite using Mg^{2+} as cofactor. In this study, we have determined the crystal structure of MASA and its complex with a substrate analog to 1.7 Å resolution by multi-wavelength anomalous diffraction and molecular replacement techniques, respectively. The structures support the proposed mechanism of phosphatase activity and further suggest the probable mechanism of enolization. We establish a model for substrate binding to describe in detail the enzymatic reaction and the formation of the transition state, which will provide insight into the reaction mechanisms of other enzymes in the same family.

Keywords: enolase-phosphatase E1; methionine salvage pathway; MASA; bifunctional enzyme; phosphatase

Introduction

Enolase-phosphatase E1 (also called MASA) promotes the conversion of 2,3-diketo-5-methylthio-1-phosphopentane to 1,2-dihydroxy-3-keto-5-methylthiopentene anion (an aci-reductone) in the methionine salvage pathway.[1] *S*-Adenosylmethionine (SAM) is a critical source of activated methyl groups in all organisms, and serves also as a source of activated C3 aminopropyl units for polyamine biosynthesis during the cell cycle. The methionine salvage pathway is a ubiquitous biochemical pathway that can maintain methionine level *in vivo* by reincorporating the thiomethyl moiety of the proximate SAM and methylthioadenosine (MTA) into methionine in a wide variety of organisms, including bacteria,[2,3] protozoa,[4] plants[5] and mammals.[6,7] MASA, named after the gene *masA*, was firstly purified and characterized in the methionine salvage pathway of *Klebsiella pneumoniae*. The active bifunctional enzyme is a monomeric protein with a molecular mass of 27,000 Da. The catalytic reaction is carried out continuously by enolization and dephosphorylation. The activity of MASA requires the presence of a magnesium ion as a cofactor; however, no other prosthetic group has yet been found.[1]

Preliminary sequence analysis indicates that MASA belongs to the L-2-halo-acid dehalogenase (HAD) superfamily of hydrolases, which includes dehalogenases (C–Cl cleavage), phosphonatases (C–P cleavage), phosphotransferases and phosphatases (C–OP cleavage).[8–13] In spite of the low level of sequence identity and a diverse range of functions, the members of this superfamily share a similar tertiary structure, some highly conserved sequence motifs in the active site, and a common catalytic mechanism.[14] The structures of several members of this family are known, and include halo-acid dehalogenase,[11] phosphonatases,[15] Ca^{2+} ATPase,[16] phosphoserine phosphatase[17] and β-phosphoglucomutase.[18] Their structures consist of a typical α/β-hydrolase core domain and a small α-helical cap domain, with the active site located at the domain–domain interface. The core domain responsible for catalysis has a highly conserved tertiary structure, while the cap domain has diverse structure to accommodate the various substrates involved in the different reactions.[15] Furthermore, the catalytic mechanism of dephosphorylation in this family is believed to involve the formation of a

† H.W. & H.P. contributed equally to this work.

Abbreviations used: MASA, enolase-phosphatase E1; SAM, *S*-adenosylmethionine; HAD, L−2-halo-acid dehalogenase; MTA, methylthioadenosine; MAD, multi-wavelength anomalous diffraction; MR, molecular replacement; PSP, phosphoserine phosphatase.

E-mail address of the corresponding author: raozh@xtal.tsinghua.edu.cn

phosphoaspartate–enzyme intermediate, which could be supported by the three conserved motifs in the active-site scaffold. Motif I has the sequence DXDX(T/V), in which the first absolutely conserved aspartate residue forms an intermediate with the substrate.[10] Motif II contains a conserved serine or threonine residue that forms a hydrogen bond with the phosphoryl oxygen atom.[17] Motif III consists of a conserved lysine residue followed by a number of less conserved residues and a strictly conserved aspartate residue, which helps to coordinate the magnesium ion required for the enzyme activity.[15,17] Characteristic phosphoaspartate enzyme intermediates in this family have been proposed to function in signal transduction (two-component response regulator proteins), phosphotransfer reactions (phosphatases and mutases) and the conversion of chemical energy to ion gradients (P-type ATPases).[16,19,20] Examination of the sequence lengths of the regions between the conserved motifs shows that there are three distinct structural subgroups within the HAD family.[21] The first group, to which MASA belongs, has a large domain between motif I and motif II that is believed to be critical for substrate binding and catalysis. The other two groups contain a large domain between motif II and motif III, or have neither of these large domains.

MASA is the first bifunctional enzyme to be identified in the HAD family. Besides its function as a phosphatase, it possesses enolase activity that cannot be classified as typical enzymatic enolization. This kind of enolization requires metal ions as cofactor but the catalytic mechanism is still unclear. Here, we present the crystal structure to 1.7 Å resolution of MASA and its complex with a substrate analog, (1-oxo-heptyl)-phosphonic acid, which we have called ANA. The analog shares the same polarized phosphoric acid head group and hydrophobic carbon main-chain as the native substrate. We discuss their structural characteristics in order to elucidate the related mechanisms underlying the enzymatic reaction.

Results and Discussion

Overall structure of human MASA

A preliminary sequence search using BLAST† shows that MASA has a conserved hydrolase domain, which indicates its relationship to the hydrolase superfamily. However, the level of sequence identity between MASA and the other members of this family was too low (<16%) for the structure to be determined by molecular replacement (MR). Thus, we chose to determine the structure by multi-wavelength anomalous diffraction (MAD) using an L-SeMet-labelled protein derivative. The three-dimensional structure turns out to have a typical α/β-hydrolase fold and belongs to the HAD superfamily based on a structural homology search using the DALI server. Phosphonoacetaldehyde hydrolase (PDB entry 1FEZ) and the L-2-halo acid dehalogenase (PDB entry 1QQ5) were identified as the closest structural homologs of MASA, with DALI Z-scores higher than 10.0 and respective rmsd values of 3.9 Å and 3.4 Å. Both structures belong to the HAD superfamily and share exactly the same active sites.

The purification experiments, the structure of native MASA and its complex with substrate analog, all indicate that human MASA is a monomer in its active form. The monomer structure contains residues 4 to 257 of the total 261 amino acid residues in the protein with the dimensions of 71.3 Å×43.7 Å×30.1 Å. Three magnesium ions are displayed clearly in the interface between the two

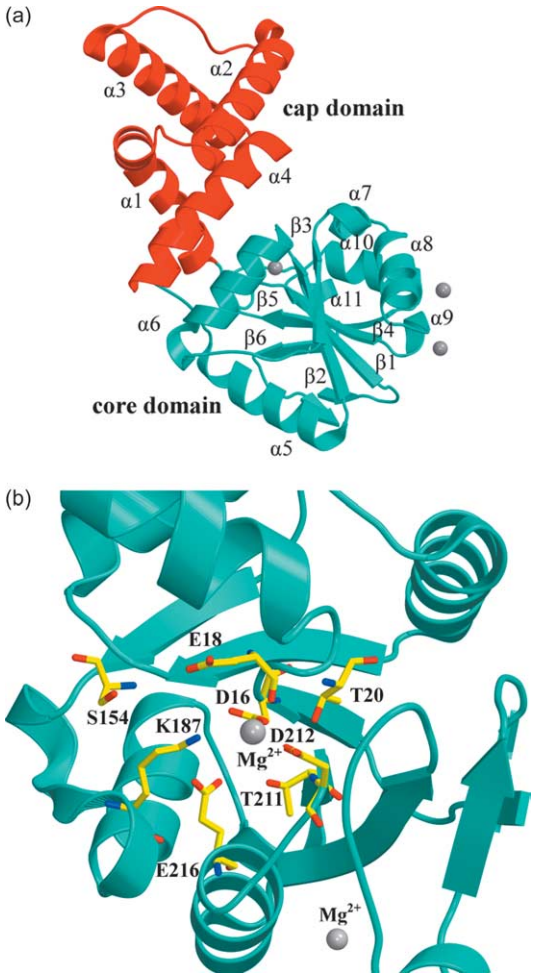


Figure 1. Structure of MASA and its active site. (a) The tertiary monomer structure is represented as a ribbon diagram. The core domain is shown in light blue and the cap domain shown in red. The three magnesium ions are shown as gray spheres. The Figure was generated using the program MOLSCRIPT.[36] (b) The active site of MASA. The magnesium ion is shown as a gray sphere and the enzyme active-site residues are shown in yellow as ball-and-stick representations. The Figure was generated using the program MOLSCRIPT.[36]

† www.ncbi.nlm.nih.gov/blast/

domains and on the molecular surface (Figure 1(b)). The structure can be formally divided into two domains (Figure 1(a)). The N-terminal 102 amino acid residues, with the exception of the β1-strand, fold into a distorted four-α-helix-bundle structure comprised of helices α1–α4, which is often referred to as the cap domain. The diversity of this domain (the cap domain shows low similarity with other known structures) suggests the variety of the activity among members of HAD family. The β1-sheet and the C-terminal 126 amino acid residues form the typical α/β-hydrolase fold of the core domain, which is known also as the Rossmann fold. This domain contains a central six-stranded, mostly parallel β-sheets (β1–β6) surrounded by seven α-helices (α5–α11). The core domain is highly conserved in this superfamily, which implies its importance for the catalytic activity. As shown in Figure 1(a), the core α/β-hydrolase domain cradles the active site and the helical cap domain covers the top, thus forming a potential pocket for the substrate binding and catalytic reaction. Two linker loops (residues 16 to 24 and residues 125 to 132) between the core and cap domains share some conserved regions with other members of the family,[22] from which further functional diversity has been shown to arise.

The active site and conserved motifs

As a member of the HAD superfamily, human MASA shares the same active site and conserved motifs with other members (Figure 2). The active site is defined by the three conserved sequence motifs, and by the bound magnesium ion in the interface between domains. A magnesium ion originating from the crystallization solution is tightly bound inside the cleft between the cap and core domains, which supports the proposition that many members of this family require divalent metals (usually a magnesium ion) for their activity.[1] In the interface of the two domains, around the cleft in which the magnesium ion is buried, several highly conserved residues form the active site (Figure 1(b)) and the motifs essential for the activity.

Motif I is located near the N terminus and contains the consensus sequence DXDX(T/V), where X can be any residue. The first aspartate residue in this motif is reported to play a critical functional role in the HAD superfamily, acting as a nucleophile in the reaction to form the phospho-aspartate enzyme intermediate.[23] In the human MASA enzyme, motif I starts with Val12 and the first conserved aspartate residue turns out to be Asp16, which is likely to be involved in the dephosphorylation as a nucleophile. The second conserved aspartate residue is replaced by a glutamate residue with a longer carbon chain at position 18, which also coordinates the magnesium ion through its main-chain carbonyl oxygen atom. Motif II has the consensus sequence (S/T)XX, in which the conserved serine/threonine residue (Ser154 in human MASA) forms a hydrogen bond with the phosphoryl oxygen atom of the substrate.[17]

Motif III is located near the C terminus and possesses the sequence K-$(X)_{18-30}$-(G/S)(D/S)XXX(D/N). In previously reported structures of the HAD superfamily members, this motif forms part of the active site and helps to coordinate the magnesium ion required for the activity (Asp212 in human MASA). Besides, the tightly conserved lysine residue (Lys187 in MASA) can form a salt-bridge with Asp16[13] in the motif I and is separated by a linker of 25 amino acid residues from

```
          Motif I          Motif II          Motif III

PGM    4 AVLFDLDGVIT   108 IKIALASAS   145 KPAPD... 165 SIGLED----SQAGIQAI
PSP    7 LILFDFDSTLV    93 YVVAVVSGG   144 KGEIL... 162 TVAVGD----GANDISMF
MASA 12 VILLDIEGTTT   148 KVYIYSSGS   187 KVESE... 207 ILFLTD    VTREASAA
PAH    8 AVIFDWAGTTV   116 IKIGSTTGY   156 RPYPW... 177 MIKVGD----TVSDMKEG
YKRX   7 FIICDFDGTLT    93 IPFYVISGG   151 KPSVI... 165 IIMIGD----SVIDVEAA
HAD    6 GIAFDLYGTLF   112 LKLAILSNG   151 KPDNR... 172 ILFVSS----NANDATGA
Subgroup 1

T6P 121 ALFLDYDGTLS   157 FPTAIISGR   301 KGKAV... 321 PIYVGD----DRTDEDAF
SP     9 MIVSDLDHTMV    44 SLLVFSTGR   174 KGQAL... 195 TLACGD----SGNDAELF
PMM    8 LCLFDVDGTLT    41 IGVVGGSDF   189 KRYCL... 204 IYFFGDKTMPGGNDHEIF
Subgroup 2

DEM 167 VAGFDLDGTLI   211 YKLVIFTNQ   260 KPVTG... 278 PISIGD----FVGDHAGR
HPIGP 5 YLFIDRDGTLI    48 YKLVMITNQ   105 KPKVK... 125 SYVIGD----RATDIQLA
Subgroup 3
```

Figure 2. Primary sequence alignment of several different members of the HAD superfamily. Three conserved domains are colored in purple (domain I), blue (domain II) and dark green (domain III). Strictly conserved residues are highlighted in red and those that share some variation are highlighted in light blue. Three subgroups of the HAD superfamily, which are different in motif distribution patterns, are shown: subgroup 1, to which MASA belongs, has a large domain between motif I and motif II; subgroup2 contains a large domain between motif II and motif III; subgroup 3 has neither of these large domains. The sequence shown are as follows: PGM, (β-phosphoglucomutase (*Lactoccocus lactis*); PSP, phosphoserine phosphatase (*Methanococcus jannaschi*); MASA, enolase/phosphatase (*Homo sapiens*); PAH, phosphonacetaldehyde hydrolase (*Bacillus cereus*); YKRX, YkrX gene product (*B. subtilis*); HAD, L-2-halo acid halogenase (*Pseudomonas* sp. YL); T6P, trehalose-6-phosphate phosphatase (*Arabidopsis thaliana*); SP, sucrose phosphatase (*A. thaliana*); PMM, phosphomannomutase (*H. sapiens*); DEM, DNA 50-kinase-30-phosphatase (*H. sapiens*); HPIGP, histidinol phosphatase/imidazole-glycerol phosphate phosphatase (*E. coli*).

conserved residues Thr211 and Asp212. The conserved G/S is replaced by Thr211 in human MASA, but residue 212 remains as aspartate. The following D/N residue of motif III is replaced conservatively by Glu216, which can stabilize the part of motif III involved in formation of the magnesium ion binding loop.

Magnesium is an essential cofactor

Three magnesium ions are found in the monomer structure of human MASA. One is buried in the active-site pocket in the interface between the two domains, where it serves as a cofactor for the enzyme activity. The other two are located on the surface of the molecule and are supposedly helpful for the packing of the crystals. Magnesium ions originate from the reservoir solution during crystallization and are absolutely required for the enzyme stability and activity. Crystals grown in the absence of magnesium ions in reservoir solution diffract poorly. Meanwhile, crystals soaked in a mother solution without magnesium ions will collapse immediately. Besides the evidence from crystallization, the magnesium ion in the active site is fixed very well by the residues and water molecules around it.

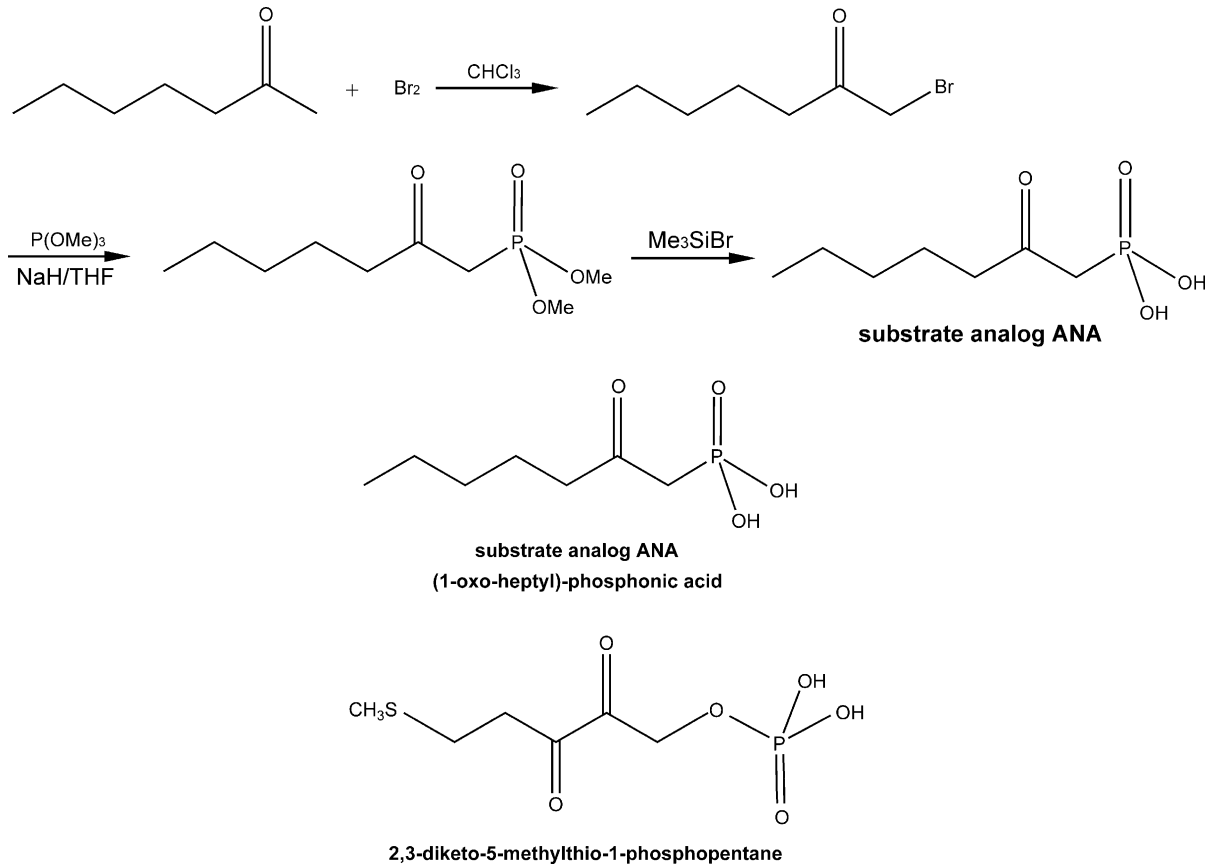
The magnesium ion bound in the active site is hexa-coordinated. The ligands are the Asp16 side-chain, the Glu18 main-chain carbonyl group, the Asp212 side-chain, and three water molecules. One of the three ligand water molecules is within hydrogen-bonding distance of the Asp16 and Glu216 side-chains, while another water molecule is within hydrogen-bonding distance of the Glu18 main-chain carbonyl group and Asp212 side-chain oxygen atoms. The third water molecule is located within hydrogen-bonding distance of the Asp16 side-chain, the Glu18 main-chain amino group, and the oxygen atom of the phosphate group of the substrate analog. The tightly bound nature of the magnesium ion suggests its importance in the catalytic reaction. Residues in motif III generate a loop structure that chelates Mg^{2+} by orientating Asp212 into the Mg^{2+}-binding site. In structural studies of other members of the HAD family, mutagenesis of the Mg^{2+} coordinating residues leads to the abolition of Mg^{2+} binding and a subsequent loss of enzyme activity.[20] In fact, the buried location of the active site and the fixed Mg^{2+} under the molecular surface indicates that the enzymatic reaction may involve some type of conformational rearrangement with regard to the Mg^{2+} cofactor and the substrate binding.

Structure of the complex with the substrate analog

The substrate analog ANA was synthesized chemically for further exploration of the structural basis for the enzymatic mechanism. The substrate, 2,3-diketo-5-methylthio-1-phosphopentane, is characterized by its phosphate group and 2-carbonyl group, which are understood to be the reactive sites in the respective processes of dephosphorylation and enolization. The analog mimics the active phosphate group and the hydrophobic carbonic chain; the 2-carbonyl group is retained but the 5-thiomethyl group is replaced by the ethyl group. In the case of phosphate cleavage during the soaking experiments, the C–OP in the substrate analog is substituted for an non-cleavable C–P bond, which causes an oxygen atom on the phosphate group to be omitted. Thus the analog is shorter than the substrate by the distance of one chemical bond between the phosphate group and the 2-carbonyl group (Scheme 1). In spite of this small dissimilarity, the analog fits the substrate pocket very well and the complex structure can partially illuminate the catalytic mechanism.

In the complex with the synthesized substrate analog, the analog sits in the cleft formed by the conserved residues in the active site (Figure 3(a)). The mimicked phosphate group is pointed towards the Mg^{2+} cofactor and the Asp16 nucleophile for dephosphorylation. The hydrophobic tail is comparatively free in the cleft, while the 2-carbonyl group related to the enolization is fixed tightly by Wat B with a hydrogen bond (1.95 Å). Wat B forms a perfect hydrogen bond with the N of the Glu18 main-chain, which could help to stabilize the carbonyl group. The substitution of Glu18 for Asp in the conserved motif I of MASA results in a longer side-chain compared with other members of HAD superfamily, and can thus provide a suitable distance (2.46 Å) for the oxygen atom of the Glu18 side-chain to attack the 2-carbonyl group. In this respect, Glu18 may act as a nucleophile in the enolization reaction. The phosphate group is stabilized by several pairs of hydrogen bonds between the hydroxyl or carbonyl group of the Pi and the conserved residues from the motifs, including Ser154 in motif II that supposedly orients the substrate, and Lys187 in motif III, which forms a salt-bridge with Asp16 to stabilize the conformation of the active site. A coordinated water molecule (Wat A) also participates in hydrogen bonding (2.77 Å) to the phosphate group.

Considering that the analog is shorter than the substrate by the distance of a chemical bond between the phosphate group and the 2-carbonyl group, the head of the phosphate group in the substrate should extend deeper into the pocket formed by Mg^{2+} and the Asp16 nucleophile. The Pi group of the phosphate may move further inside the pocket by a distance of 2–3 Å and directly bind to Mg^{2+} with its oxygen atom occupying the position of Wat A (Figure 3(b)), which will bring it to close proximity to the Asp16 nucleophile (2.41 Å in the model). The direct coordination of the phosphate group with Mg^{2+} may polarize the P–O bond and increase the electrophilicity of the phosphorus atom, as proposed previously.[24] The interaction of the side-chains of Ser154 and Lys187 with the phosphate oxygen atoms could increase the electrophilicity of the phosphorus atom and thus facilitate



Scheme 1. Synthesis of the substrate analog ANA, and a comparison between the chemical structures of the substrate and its analog.

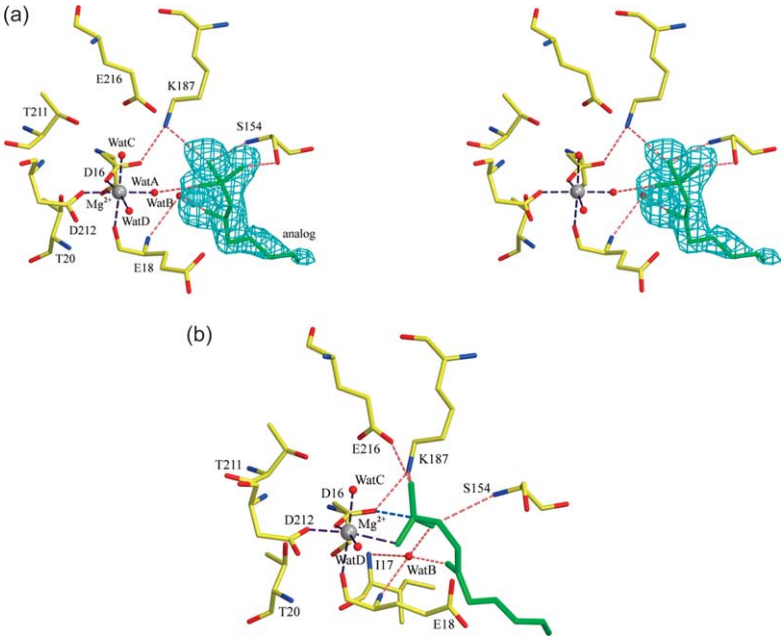


Figure 3. (a) Stereoview of the substrate analog (shown as ball-and-stick in green) binding in the active site of MASA (shown as ball-and-stick) with the corresponding electron density map (omit map contoured at 2.4 σ shown in cyan) covering the analog. The Mg^{2+} is shown as a gray sphere and the water molecules are shown as red spheres. Hydrogen bonds are represented by broken lines in pink and the hexa-coordinated bonds of the Mg^{2+} are represented by broken brown lines. The Figure was rendered using the program MOLSCRIPT.[36] (b) A model of the substrate binding implies the enzymatic mechanism. The model was based on superposition of the active sites of MASA and PSP, which indicates the position of the intermediate phosphate group. The actual binding of the substrate analog also provides clues for the substrate model location. We combine these two pieces of evidence and propose the model as follows. The substrate model is shown as ball-and-stick in green and the important enzymatic residues are shown as ball-and-stick. The Mg^{2+} is shown as a gray sphere and the water molecules are shown as red spheres. The hydrogen bonds are represented by broken lines and the hexa-coordinated bonds of Mg^{2+} are represented by broken brown lines. The probable nucleophilic attack of Asp16 is shown as a broken blue line. The Figure was rendered using the program MOLSCRIPT.[36]

nucleophilic attack on the phosphate group by Asp16 and cleavage of the P–O bond of the substrate. Furthermore, the salt-bridge between Lys187 and Asp16 may stabilize the negative charge associated with the Asp16 side-chain during the reaction.

A model for substrate-binding and implications for the enzymatic mechanism

Among the several HAD superfamily structures solved to date, human MASA is also structurally related to the phosphoserine phosphatase (PSP) (PDB entry 1F5S) with similar catalytic residues at the active site (Asp16, Ser154, Lys187 and Asp212 in MASA *versus* Asp11, Ser99, Lys144 and Asp167 in PSP), indicating a similar reaction mechanism. As previous studies have shown, the dephosphorylation reaction is initiated by nucleophilic attack on the substrate phosphate group by the side-chain carboxylate group of a strictly conserved aspartate residue, involving the Mg^{2+}-dependent formation of a phosphoryl-aspartyl enzyme intermediate. Proposed structural "snapshots" of PSP also provide a detailed understanding of this reaction pathway.[20] According to the transition state structural model of PSP (PDB entry 1L7N), we compared the active site of MASA with PSP and constructed a model (Figure 3(b)) for substrate binding in order to elucidate the mechanism of phosphatase/enolase activity of human MASA.

The phosphate group would sit in the pocket formed by Mg^{2+}, Asp16 and other conserved residues in the active site (Glu18, Ser154, Lys187 and Asp212) with one of its oxygen atoms coordinated directly with Mg^{2+}. This coordination may polarize the P–O bond and increase the electrophilicity of the phosphorus atom. The distance between Pi and the Asp16 nucleophile shortens to 2.41 Å in the model, implying the potential for nucleophilic attack by Asp16 and the existence of a pentacovalent phosphoryl-aspartyl enzyme intermediate. Hydrogen bonds are formed between oxygen atoms of the phosphate group and Lys187 (2.40 Å), Glu216 (2.53 Å), Wat B (2.00 Å) respectively to further increase the electrophilicity of the phosphorus atom to facilitate the nucleophilic attack and the cleavage of the P–O bond of substrate. The 2-carbonyl group is still fixed by Wat B *via* a hydrogen bond (2.01 Å) and Glu18, which could immobilize the main-chain of the substrate and orient the head of the phosphate group into the active site. In addition to stabilizing the phosphate group jointly, the side-chain of Glu18 is 2.46 Å from the carbon atom of the 2-carbonyl group, and is thus in a good position to act as the nucleophile for enolization involving formation of a carbon tetrahedral transition state and shift of the C=O bond to the C=C bond. The hydrogen bond between Wat A and the 2-carbonyl group could increase the electrophilicity of the carbon atom to facilitate nucleophilic attack and the shift of bonds.

On the basis of the evidence presented by the substrate-binding model, we propose that during the whole process of catalysis, residues Glu18, Ser154, Lys187, Glu216 and Wat B act as anchor residues to fasten the main-chain of the substrate and orient the head of phosphate group to the magnesium ion and the Asp16 nucleophile. The magnesium ion directly coordinates the phosphate group to make the phosphorus atom more susceptible to nucleophilic attack (Figure 4(a)). In the dephosphorylation step, the substrate binds to the Asp16 nucleophile and a proposed pentacovalent phosphoryl-aspartyl enzyme transition state is formed. Following cleavage of the P–O bond, subsequent attack by a water molecule in close proximity causes the phosphate group to break away from Asp16. The resultant trigonal bipyramidal transition state could be stabilized by all hydrogen bonds formed to fasten the phosphate group, and the salt-bridge between Asp16 and Lys187 could neutralize the negative charge on the side-chain of Asp16 during the dephosphorylation. Enolization may occur simultaneously, before or after the dephosphorylation as long as the 2-carbonyl group is fixed and the side-chain of Glu18 is in the correct position to attack the carbon atom (Figure 4(b)). An unstable carbon tetrahedral transition may be generated and the following shift of the C=O bond to the C=C bond may occur randomly. The final product of enolization should be determined by the energy decrease caused by the following automatic or catalytic hydrolysis.

Materials and Methods

Protein expression and purification

The coding sequence of human MASA was amplified from the liver library by RT-PCR.[25] The cDNA was then cloned into the bacterial expression vector pET28a (Novagen) with a His-tag on the C terminus. The recombinant plasmid was transformed into *Escherichia coli* strain BL21 (DE3) and over-expressed. The soluble His-tagged protein was purified by Ni^{2+} chelating affinity column (1.5 ml of Ni^{2+}-NTA agarose) initially, and further applied to Superdex-75 size-exclusion and MonoQ anion-exchange chromatography columns (Amersham Biosciences). The purified and concentrated protein (25 mg ml^{-1}) was stored in 20 mM Hepes (pH 7.5) at 193 K.

The L-SeMet-labelled MASA protein was also expressed in the *E. coli* strain BL21 (DE3). The cells were diluted with adaptive medium (20% (v/v) LB medium, 80% (v/v) M9 medium) and grown at 310 K to an absorbance at 600 nm (A_{600}) of 0.6–0.8. The cells were harvested and resuspended in M9 medium, transferred into restrictive medium (2% (w/v) glucose) and then cultured to an A_{600} of 0.6–0.8 before inducement. L-SeMet at 60 mg l^{-1}, lysine, threonine and phenylalanine at 100 mg l^{-1}, leucine, isoleucine and valine at 50 mg l^{-1} and 0.5 mM IPTG were added and the incubation continued at 289 K for 20 hours. The L-SeMet-labelled MASA protein was purified and stored under the same condition as the native protein.

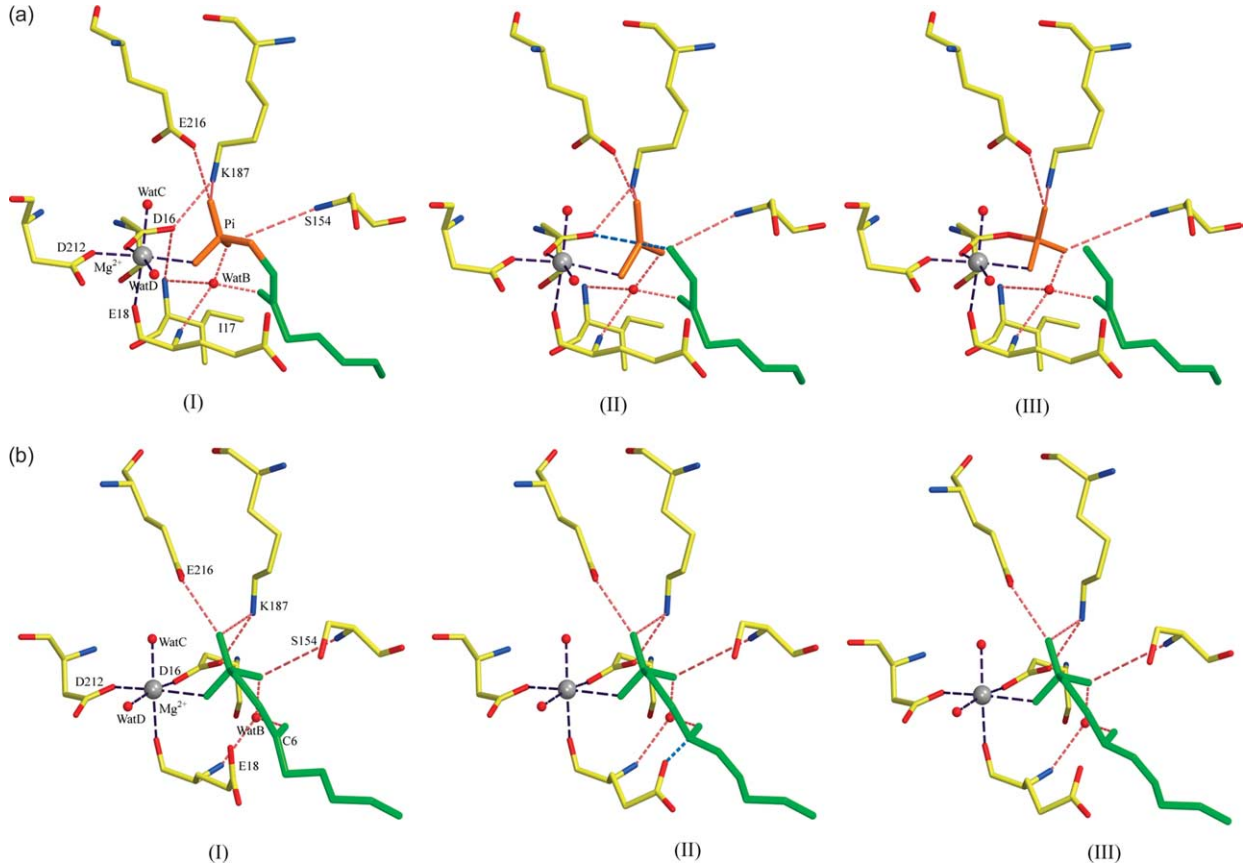


Figure 4. Structural model of human MASA phosphatase/enolase reaction cycle in the active sites. (a) Reaction cycle of phosphatase activity in the active sites. (I) The substrate-bound structural analog. (II) The phospho-aspartyl enzyme intermediate structural analog. (III) The product, Pi, bound structural analog. (b) Reaction cycle of enolase activity in the active sites. (I) The substrate-bound structural analog. (II) The unstable carbon tetrahedral transition state. (III) The final state of enolization.

Crystallization and data collection

The protein solution of MASA and L-SeMet-labelled MASA used for crystallization contained 20 mM Hepes (pH 7.5) and 25 mg ml^{-1} protein. Crystals optimized for X-ray diffraction were obtained using the hanging-drop, vapor-diffusion technique with reservoir solution containing 30% (w/v) PEG4000, 0.1 M Tris–HCl (pH 8.5), 0.2 M $MgCl_2 \cdot 6H_2O$. Protein solution (1 µl) was mixed with reservoir solution (1 µl) and equilibrated against 300 µl of reservoir solution at 291 K.

The three-wavelength MAD X-ray diffraction data sets for L-SeMet-labelled MASA protein were collected at 291 K using a MAR CCD detector and synchrotron radiation beamlines 3W and 1A at the Beijing Synchrotron Radiation Facility. The crystal was flash-frozen for data collection and diffracted to the resolution of 1.7 Å. The program automar† was used for data indexing, integrating and scaling, and determination of the unit-cell parameters (K. Bartels & C. Klein, unpublished results). Data collection statistics are summarized in Table 1. The crystal belongs to space group $P2_12_12_1$ with unit-cell parameters $a = 54.02$ Å, $b = 57.55$ Å, $c = 87.32$ Å. The Matthews coefficient suggests the presence of a monomer in an asymmetric unit with an estimated solvent content of 45.0% (v/v).[26]

† www.marresearch.com/automar

Soaking with substrate analog

The native MASA crystals were applied to substrate analog soaking. To elucidate the catalytic mechanisms of phosphatase and enolase, we synthesized the substrate analog ANA (Scheme 1) with the same polarized phosphoric acid head group and hydrophobic carbon main-chain as the native substrate 2,3-diketo-5-methylthio-1-phosphopentane. The native protein crystals were transferred to a soaking solution containing 1 mM substrate analog ANA within 1.5 µl of reservoir solution (30% PEG4000, 0.1 M Tris–HCl (pH 8.5), 0.2 M $MgCl_2 \cdot 6H_2O$) for 8–12 hours and a set of diffraction data was collected in-house at 291 K using a Rigaku RAXIS-IV++ detector and radiation wavelength of 1.5418 Å to 1.7 Å resolution. Crystals of the MASA complex with substrate analog shared the same space group and unit cell parameters with the native crystals. Data collection statistics are summarized in Table 1. The data were integrated and scaled with DENZO and SCALEPACK.[27]

Structure determination, refinement and analysis

The MASA structure was solved by the MAD technique using L-SeMet-labelled protein. The Beijing Synchrotron Radiation Facility datasets were good enough to solve the structure and able to yield clear electron density map. Five out of six expected selenium positions were determined by the program SOLVE.[28] Phases were

Table 1. X-ray data collection and structure refinement statistics of MASA protein and its complex with substrate analog

Parameter	MASA protein SeMet λ1 (edge)	SeMet λ2 (peak)	SeMet λ3 (remote)	MASA + substrate analog
A. *Data collection*				
Space group	$P2_12_12_1$			$P2_12_12_1$
Cell dimensions				
a (Å)	54.02			53.96
b (Å)	57.55			57.12
c (Å)	87.32			87.19
Wavelength (Å)	0.9800	0.9798	0.9000	1.5418
Resolution (Å)	30–1.7 (1.76–1.70)	30–1.7 (1.76–1.70)	30–1.7 (1.76–1.70)	50–1.7 (1.76–1.70)
Average $I/\sigma(I)$[a]	3.5 (0.9)	4.1 (1.1)	3.6 (1.4)	14.9 (2.1)
Total reflections	185,221	236,864	185,438	106,126
Unique reflections	30,414	30,337	30,550	30,076
Completeness (%)[a]	99.9 (98.3)	99.7 (97.5)	100 (99.8)	98.9 (98.5)
R_{merge} (%)[a,b]	9.03 (45.43)	9.11 (36.33)	9.34 (37.68)	6.4 (44.8)
B. *Structure refinement*				
Resolution (Å)	50–1.7			50–1.7
Average *B*-factor (Å^2)	20.35			44.65
R_{work}/R_{free} (%)[c]	20.8/24.1			22.5/22.8
rmsd from ideality				
Bond lengths (Å)[d]	0.012			0.012
Bond angles (deg.)[d]	1.519			1.690
Ramachandran plot				
Most favored (%)	95.2			92.6
Allowed (%)	4.8			7.0
Generously allowed (%)	0			0.4
Disallowed (%)	0			0

[a] Numbers in parentheses correspond to the highest-resolution shell.
[b] $R_{merge} = \sum_i |I_i - \langle I \rangle| / \Sigma \langle I \rangle$, where I_i is an individual intensity measurement and $\langle I \rangle$ is the average intensity for all measurements of the reflection *i*.
[c] $R_{work}/R_{free} = \sum||F_o| - |F_c|| / \sum|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.
[d] rmsd values relate to the Engh & Huber parameters.[37]

computed using SOLVE and RESOLVE.[29] The program O was used for manual rebuilding of the model,[30] and further refinements by simulated annealing, energy minimization, *B*-factor refinement and water-inclusion were performed using CNS.[31]

The structure of the complex with substrate analog was determined by molecular replacement using the native MASA structure as a model. A clear molecular replacement solution was found by cross-rotation and translation function searches performed using CNS. Manual model rebuilding was carried out by program O. PDB coordinates for the substrate analog were generated using CHEMDRAW† and inserted into the active site of MASA according to the calculated $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ difference density maps. Further refinements using CNS were alternated with manual model rebuilding based on newly calculated difference Fourier maps. The final model and the substrate analog fit the electron density map very well, with excellent spatial geometry. All residues lie in the allowed regions of the Ramachandran plot produced by PROCHECK.[32] The refinement statistics of MASA and its complex structure are summarized in Table 1.

Sequence alignment was done using CLUSTALW.[33] Figure 2 was prepared using ESPRIPT.[34] Comparison of three-dimensional structures was carried out using the DALI server.[35] Structure Figures were created using the program MOLSCRIPT.[36]

† www.cambridgesoft.com

Conclusions

MASA is the only enzyme of the HAD superfamily isolated to date for which both enolase and phosphatase activities have been established. The structure of human MASA and its complex with a substrate analog determined to 1.7 Å resolution demonstrate the proposed phosphatase mechanism, which utilizes the active-site Asp16 nucleophile to abstract the phosphate group and the existence of a pentacovalent phosphoryl-aspartyl transition state. A model for substrate binding indicates the probable enzymatic mechanism of enolization. Even though the complex structure cannot resemble the real transition state of the reaction, it provides a feasible model for understanding the catalytic pathway and will shed light on the structural studies of other members in HAD superfamily.

Protein Data Bank accession codes

The coordinates and structure factors for native MASA and its complex with a substrate analog have been deposited in the RCSB Protein Data Bank with accession codes 1WDH and 1YNS, respectively.

Acknowledgements

We thank Yi Ding, Yi Li and Xiaoai Wu for their

technical assistance. We are grateful to Zhiyong Lou, Dr Yuhui Dong and Dr Peng Liu for their kind help with data collection and processing at BSRF. We thank Dr Yiwei Liu and Hua Fu for useful discussion. This study was supported by the Ministry of Science & Technology (MOST) Human Liver Proteomics Project (Grant No. 2004CB520801), the State "863" High-Tech Project (Grant No. 2002BA711A12), "973" Project (Grant No. G1999075602) and the NSFC (Grant No. 30221003).

References

1. Myers, R. W., Wray, J. W., Fish, S. & Abeles, R. H. (1993). Purification and characterization of an enzyme involved in oxidative carbon-carbon bond cleavage reactions in the methionine salvage pathway of *Klebsiella pneumoniae*. *J. Biol. Chem.* **268**, 24785–24791.
2. Marchitto, K. S. & Ferro, A. J. (1985). The metabolism of 5′-methylthioadenosine and 5-methylthioribose 1-phosphate in *Saccharomyces cerevisiae*. *J. Gen. Microbiol.* **131**, 2153–2164.
3. Shapiro, S. K. & Schlenk, F. (1980). Conversion of 5′-methylthioadenosine into *S*-adenosylmethionine by yeast cells. *Biochim. Biophys. Acta*, **633**, 176–180.
4. Sugimoto, Y., Toraya, T. & Fukui, S. (1976). Studies on metabolic role of 5′-methylthioadenosine in *Ochromonas malhamensis* and other microorganisms. *Arch. Microbiol.* **108**, 175–182.
5. Yung, K. H., Yang, S. F. & Schlenk, F. (1982). Methionine synthesis from 3-methylthioribose in apple tissue. *Biochem. Biophys. Res. Commun.* **104**, 771–777.
6. Trackman, P. C. & Abeles, R. H. (1983). Methionine synthesis from 5′-*S*-methylthioadenosine. Resolution of enzyme activities and identification of 1-phospho-5-*S* methylthioribulose. *J. Biol. Chem.* **258**, 6717–6720.
7. Ghoda, L. Y., Savarese, T. M., Dexter, D. L., Parks, R. E., Jr, Trackman, P. C. & Abeles, R. H. (1984). Characterization of a defect in the pathway for converting 5′-deoxy-5′-methylthioadenosine to methionine in a subline of a cultured heterogeneous human colon carcinoma. *J. Biol. Chem.* **259**, 6715–6719.
8. Koonin, E. V. & 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.
9. Baker, A. S., Ciocci, M. J., Metcalf, W. W., Kim, J., Babbitt, P. C., Wanner, B. L. *et al*. (1998). Insights into the mechanism of catalysis by the P–C bond-cleaving enzyme phosphonoacetaldehyde hydrolase derived from gene sequence analysis and mutagenesis. *Biochemistry*, **37**, 9305–9315.
10. Collet, J. F., Stroobant, V., Pirard, M., Delpierre, G. & Van Schaftingen, E. (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.
11. Ridder, I. S. & Dijkstra, B. W. (1999). Identification of the Mg^{2+}-binding site in the P-type ATPase and phosphatase members of the HAD (haloacid dehalogenase) superfamily by structural similarity to the response regulator protein CheY. *Biochem. J.* **339**, 223–226.
12. Stokes, D. L. & Green, N. M. (2000). Modeling a dehalogenase fold into the 8-A density map for Ca(2+)-ATPase defines a new domain structure. *Biophys. J.* **78**, 1765–1776.
13. Hisano, T., Hata, Y., Fujii, T., Liu, J. Q., Kurihara, T., Esaki, N. & Soda, K. (1996). Crystal structure of L-2-haloacid dehalogenase from *Pseudomonas* sp. YL. An α/β hydrolase structure that is different from the α/β hydrolase fold. *J. Biol. Chem.* **271**, 20322–20330.
14. Shin, D. H., Roberts, A., Jancarik, J., Yokota, H., Kim, R., Wemmer, D. E. & Kim, S. H. (2003). Crystal structure of a phosphatase with a unique substrate binding domain from *Thermotoga maritima*. *Protein Sci.* **12**, 1464–1472.
15. Morais, M. C., Zhang, W., Baker, A. S., Zhang, G., Dunaway-Mariano, D. & 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.
16. Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. (2000). Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature*, **405**, 647–655.
17. Wang, W., Kim, R., Jancarik, J., Yokota, H. & Kim, S. H. (2001). Crystal structure of phosphoserine phosphatase from *Methanococcus jannaschii*, a hyperthermophile, at 1.8 Å resolution. *Structure (Camb)*, **9**, 65–71.
18. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D. & Allen, K. N. (2002). Caught in the act: the structure of phosphorylated β-phosphoglucomutase from *Lactococcus lactis*. *Biochemistry*, **41**, 8351–8359.
19. Welch, M., Chinardet, N., Mourey, L., Birck, C. & Samama, J. P. (1998). Structure of the CheY-binding domain of histidine kinase CheA in complex with CheY. *Nature Struct. Biol.* **5**, 25–29.
20. Wang, W., Cho, H. S., Kim, R., Jancarik, J., Yokota, H., Nguyen, H. H. *et al*. (2002). Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states. *J. Mol. Biol.* **319**, 421–431.
21. Selengut, J. D. & Levine, R. L. (2000). MDP-1: a novel eukaryotic magnesium-dependent phosphatase. *Biochemistry*, **39**, 8315–8324.
22. Zhang, Y., Heinsen, M. H., Kostic, M., Pagani, G. M., Riera, T. V., Perovic, I. *et al*. (2004). Analogs of 1-phosphonooxy-2,2-dihydroxy-3-oxo-5-(methylthio)-pentane, an acyclic intermediate in the methionine salvage pathway: a new preparation and characterization of activity with E1 enolase/phosphatase from *Klebsiella oxytoca*. *Bioorg. Med. Chem.* **12**, 3847–3855.
23. Selengut, J. D. (2001). MDP-1 is a new and distinct member of the haloacid dehalogenase family of aspartate-dependent phosphohydrolases. *Biochemistry*, **40**, 12704–12711.
24. Ridder, I. S., Rozeboom, H. J., Kalk, K. H., Janssen, D. B. & Dijkstra, B. W. (1997). Three-dimensional structure of L-2-haloacid dehalogenase from *Xanthobacter autotrophicus* GJ10 complexed with the substrate-analogue formate. *J. Biol. Chem.* **272**, 33015–33022.
25. Lloyd, B. H., Platt-Higgins, A., Rudland, P. S. & Barraclough, R. (1998). Human S100A4 (p9Ka) induces the metastatic phenotype upon benign tumour cells. *Oncogene*, **17**, 465–473.
26. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
27. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In

Macromolecular Crystallography, part A (Carter, C. W., Jr & Sweet, R. M., eds), vol. 276, pp. 307–326, Academic Press, New York.
28. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallog. sect. D*, **55**, 849–861.
29. Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallog. sect. D*, **56**, 965–972.
30. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
31. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
32. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
33. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Comput. Appl. Biosci.* **10**, 19–29.
34. Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucl. Acids Res.* **31**, 3320–3323.
35. Holm, L. & Sander, C. (1998). Protein folds and families: sequence and structure alignments. *Nucl. Acids Res.* **26**, 316–319.
36. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
37. Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A*, **47**, 392–400.

Edited by R. Huber

(Received 11 November 2004; received in revised form 26 January 2005; accepted 31 January 2005)