X-ray Crystal Structure of the Hypothetical Phosphotyrosine Phosphatase MDP-1 of the Haloacid Dehalogenase Superfamily†‡§

Ezra Peisach,Jeremy D. Selengut,Debra Dunaway-Mariano,* and Karen N. Allen*‡§

Department of Physiology and Biophysics, Boston University School of Medicine, Boston, Massachusetts 02118, The Institute for Genomic Research, Rockville, Maryland 20850, and Department of Chemistry, University of New Mexico, Albuquerque, New Mexico 87131

Received May 7, 2004; Revised Manuscript Received August 4, 2004

ABSTRACT: The haloacid dehalogenase (HAD) superfamily is comprised of structurally homologous enzymes that share several conserved sequence motifs (loops I–IV) in their active site. The majority of HAD members are phosphohydrolases and may be divided into three subclasses depending on domain organization. In classes I and II, a mobile “cap” domain reorients upon substrate binding, closing the active site to bulk solvent. Members of the third class lack this additional domain. Herein, we report the 1.9 Å X-ray crystal structures of a member of the third subclass, magnesium-dependent phosphatase-1 (MDP-1) both in its unliganded form and with the product analogue, tungstate, bound to the active site. The secondary structure of MDP-1 is similar to that of the “core” domain of other type I and type II HAD members with the addition of a small, 28-amino acid insert that does not close down to exclude bulk solvent in the presence of ligand. In addition, the monomeric oligomeric state of MDP-1 does not allow the participation of a second subunit in the formation and solvent protection of the active site. The binding sites for the phosphate portion of the substrate and Mg(II) cofactor are also similar to those of other HAD members, with all previously observed contacts conserved. Unlike other subclass III HAD members, MDP-1 appears to be equally able to dephosphorylate phosphotyrosine and closed-ring phosphosugars. Modeling of possible substrates in the active site of MDP-1 reveals very few potential interactions with the substrate leaving group. The mapping of conserved residues in sequences of MDP-1 from different euukaryotic organisms reveals that they colocalize to a large region on the surface of the protein outside the active site. This observation combined with the modeling studies suggests that the target of MDP-1 is most likely a phosphotyrosine in an unknown protein rather than a small sugar-based substrate.

The haloacid dehalogenase (HAD)‡ enzyme superfamily is a large family of enzymes, which function in catalysis of group transfer reactions utilizing an Asp carboxylate as the nucleophile (I). The vast majority of the reactions identified to date involve transfer of a phosphoryl group to water, and the HAD members which catalyze these reactions employ a Mg(II) cofactor to bridge the charged reactant and enzyme nucleophile. The phosphohydrolases fall within three structural subfamilies (I–III) based on domain organization (2). Common to all subfamilies is the highly conserved α/β core domain (illustrated in Figure 1a) that supports the four-loop catalytic scaffold. The loops contain the four motifs by which members of the superfamily are identified. An essential nucleophilic Asp is located on loop I. A conserved Ser/Thr phosphate-binding residue is located on loop II. A Lys/Arg site (N-terminal phosphotyrosine phosphatase) contain a sequence insert between loops II and III. The inserted sequence forms a cap domain whose association and dissociation with the core domain opens and closes, respectively, the catalytic site to bulk solvent (Figure 1b). The closed conformation of the cap domain is stabilized by substrate binding (8, 9). Subfamily III lacks the sequence insert and therefore the cap domain. A large substrate [as exemplified by the T4 polynucleotide 5'-kinase (N-terminal domain)-3'-phosphatase (C-terminal “HAD” domain) (10)] or an adjacent subunit [as exemplified by the 3-deoxy-d-
manno-octulosonate 8-phosphate (8-KDO) phosphatase (11, 12) in combination with a small substrate serves to occlude the active-site entrance.

In an earlier study, the subfamily III HAD enzyme, magnesium-dependent phosphatase-1 (MDP-1) from mouse, was found to exhibit Mg(II)-dependent phosphotyrosine hydrolase activity (13). This substrate activity, in combination with the sequence-based prediction of the absence of a cap domain, raised the intriguing possibility that MDP-1 represents a new class of eukaryotic phosphotyrosine phosphatases that employ an Asp rather than Cys nucleophile (14, 15). Herein, we report that MDP-1 is monomeric and that the high-resolution X-ray crystal structure reveals a classical α/β core domain without a cap domain. Using the sequences of orthologous proteins deposited in gene and protein databases, we generated an alignment to identify absolutely conserved residues. In addition to the core catalytic residues, these include surface residues that line the entrance to the active site. Together, our findings provide strong evidence in support of the functional role of MDP-1 as a protein tyrosine phosphatase.

**EXPERIMENTAL PROCEDURES**

**Materials.** Restriction enzymes BamHI, NdeI, and T4 ligase were purchased from New England Biolabs. The Advantage HF2 PCR kit was purchased from BD Biosciences Clontech. XL1-Blue BL21(DE3)pLys cells were purchased from Stratagene. Isopropyl-β-D-thiogalactopyranoside (IPTG), Luria broth (LB), ampicillin, and chloramphenicol were purchased from Sigma. Luria broth (LB), ampicillin, and chloroamphenicol were deposited in gene and protein databases, and the original mouse CBP-MDP-1 clone, described by Selengut et al. (13), was reengineered to remove the N-terminal calmodulin binding sequence as well as the thrombin and enterokinase cleavage sites, thereby restoring the natural N-terminus. PCR using BD Biosciences Clontech’s high-fidelity cloning kit (Advantage HF-2) was used to remove the gene from the plasmid and introduce NdeI and BamHI restriction sites to the 5’-and 3’-end, respectively. Oligonucleotides 5’-GAGATATACATATGACGGGGCTGCCAAAGCCTT-3’ for the forward primer and 5’-CAGCCGGATCTTTATTATCAGAGTCCAGGCTTGGCCCTT-3’ for the reverse primer were used (NdeI and BamHI sites underlined, respectively). The PCR product was gel purified, and subjected to double-restriction enzyme digest with NdeI and BamHI. A plasmid encoding pET3a was similarly digested, gel purified, and ligated to the PCR product with T4 ligase. Ligated plasmids were transformed into XL1-Blue cells and plated on media containing ampicillin (100 µg/mL) and tetracycline (12.5 µg/mL). Several colonies were grown in LB, and the plasmids were extracted and analyzed by restriction enzyme digestion for the presence of the proper insert. The inserts were subjected to DNA sequencing to verify the product. Plasmid pET3-MDP1 was subsequently transformed into BL21(DE3)pLysS cells for expression.

**Protein Expression and Purification.** An overnight culture of pET3-MDP1 in LB containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol was grown at 37 °C. The overnight culture (10 mL) was added to two 1 L flasks of LB with the same antibiotics and grown at 37 °C with shaking (at 225 rpm) until the OD600 reached 0.5. Cells were induced by the addition of IPTG to a final concentration of 0.3 mM and grown for 2.75 hours. Cells were harvested by centrifugation for 10 min at 5000 g, and the pellet was resuspended in 250 mL of 20 mM Tris-HCl (pH 8.0) and centrifuged again. Cells were quick-frozen in liquid nitrogen and stored at −80 °C.

Frozen cells were lysed by resuspension in 25 mL of lysis buffer [50 mM Tris (pH 8.0), 200 mM NaCl, 5 mM EDTA, 1 mM PMSF, and 0.01% Triton X-100]; the tip of a spatula of DNase I was added and the solution gently stirred on ice for 2 h. Following sonication at a 50% duty cycle for 30 s, the lysed cells were centrifuged for 15 min at 17000 g. The cloudy supernatant was removed, and 3.5 M NH4SO4 was added to a final concentration of 1 M and the mixture incubated overnight at 4 °C. The ammonium sulfate solution was centrifuged for 20 min at 38700 g, and the supernatant was removed and filtered through a 0.45 µm filter. Filtered protein was loaded onto a Phenyl-Sepharose-6 fast flow HR10 × 10 column (Amersham) preequilibrated with 50 mM Tris (pH 8.0), 5 mM EDTA, 1 mM PMSF, and 0.01% Triton X-100; the tip of a sp Fall of DNAase I was added and the solution gently stirred on ice for 2 h. Following sonication at a 50% duty cycle for 30 s, the lysed cells were centrifuged for 15 min at 17000 g. The cloudy supernatant was removed, and 3.5 M NH4SO4 was added to a final concentration of 1 M and the mixture incubated overnight at 4 °C. The ammonium sulfate solution was centrifuged for 20 min at 38700 g, and the supernatant was removed and filtered through a 0.45 µm filter. Filtered protein was loaded onto a Phenyl-Sepharose-6 fast flow HR10 × 10 column (Amersham) preequilibrated with 50 mM Tris (pH 8.0), 5 mM EDTA, 1 mM PMSF, and 0.01% Triton X-100; the tip of a sp

**Cloning.** The original mouse CBP-MDP-1 clone, described by Selengut et al. (13), was reengineered to remove the N-terminal calmodulin binding sequence as well as the thrombin and enterokinase cleavage sites, thereby restoring the natural N-terminus. PCR using BD Biosciences Clontech’s high-fidelity cloning kit (Advantage HF-2) was used to remove the gene from the plasmid and introduce NdeI and BamHI restriction sites to the 5’- and 3’-end, respectively. Oligonucleotides 5’-GAGATATACATATGACGGGGCTGCCAAAGCCTT-3’ for the forward primer and 5’-CAGCCGGATCTTTATTATCAGAGTCCAGGCTTGGCCCTT-3’ for the reverse primer were used (NdeI and BamHI sites underlined, respectively). The PCR product was gel purified, and subjected to double-restriction enzyme digest with NdeI and BamHI. A plasmid encoding pET3a was similarly digested, gel purified, and ligated to the PCR product with T4 ligase. Ligated plasmids were transformed into XL1-Blue cells and plated on media containing ampicillin (100 µg/mL) and tetracycline (12.5 µg/mL). Several colonies were grown in LB, and the plasmids were extracted and analyzed by restriction enzyme digestion for the presence of the proper insert. The inserts were subjected to DNA sequencing to verify the product. Plasmid pET3-MDP1 was subsequently transformed into BL21(DE3)pLysS cells for expression.

**Protein Expression and Purification.** An overnight culture of pET3-MDP1 in LB containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol was grown at 37 °C. The overnight culture (10 mL) was added to two 1 L flasks of LB with the same antibiotics and grown at 37 °C with shaking (at 225 rpm) until the OD600 reached 0.5. Cells were harvested by centrifugation for 10 min at 5000 g, and the pellet was resuspended in 250 mL of 20 mM Tris-HCl (pH 8.0) and centrifuged again. Cells were quick-frozen in liquid nitrogen and stored at −80 °C.

Frozen cells were lysed by resuspension in 25 mL of lysis buffer [50 mM Tris (pH 8.0), 200 mM NaCl, 5 mM EDTA, 1 mM PMSF, and 0.01% Triton X-100]; the tip of a spatula of DNAase I was added and the solution gently stirred on ice for 2 h. Following sonication at a 50% duty cycle for 30 s, the lysed cells were centrifuged for 15 min at 17000 g. The cloudy supernatant was removed, and 3.5 M NH4SO4 was added to a final concentration of 1 M and the mixture incubated overnight at 4 °C. The ammonium sulfate solution was centrifuged for 20 min at 38700 g, and the supernatant was removed and filtered through a 0.45 µm filter. Filtered protein was loaded onto a Phenyl-Sepharose-6 fast flow HR10 × 10 column (Amersham) preequilibrated with 50 mM Tris (pH 8.0), 5 mM EDTA, 1 mM PMSF, and 1 M NH4SO4. Protein was eluted with a 100 mL gradient of 1 M NH4SO4 to buffer alone. Fractions were assayed for phosphatase activity toward p-nitrophenyl phosphate (pNP) (see below) and analyzed by SDS–PAGE. Active fractions free of 20–30 kDa proteins were pooled, concentrated to 1.2 mL with a Millipore ultrafree centrifugal device, and loaded onto a Hitrap 26/60 Sephacryl S200 gel filtration column preequilibrated with 50 mM Tris (pH 8.0), 1 mM MgCl2, and 100 mM NaCl. Protein was eluted at a flow rate of 1.3 mL/min with the same buffer. Fractions containing phosphatase activity that were judged to be >95% pure by SDS–PAGE were pooled and dialyzed overnight against 50 mM Tris (pH 8.0), 50 mM NaCl, and 0.01% Tween 20 and then against 10 mM Tris (pH 8.0), 50 mM NaCl, and 0.01% Tween 20. The dialyzed protein was concentrated to 13.5 mg/mL in an ultrafree centrifuge filter, aliquoted, and snap-frozen in liquid nitrogen. Protein was stored at −80 °C until it was needed. The protein yield was approximately 7 mg/g of wet cells.
Expression and Purification of Selenomethionine-Substituted MDP-1. Selenomethionine (SeMet)-substituted MDP-1 was expressed in the same BL21(DE3)pLys cells as the native enzyme. A 50 mL overnight culture was grown in LB with 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. Cells were harvested by centrifugation at 2000g and suspended in 10 mL of 2 x M9 and 0.4% glucose minimal medium supplemented with 19 amino acids (no methionine) at 40 µg/mL, seleno-l-methionine at 40 µg/mL, and 1 µg/mL riboflavin and thiamine. Two 1 L cultures of minimal media (with 5 mL inocula) were grown with shaking at 250 rpm and 37 °C. After 8 h (A600 = 0.3), IPTG was added to a final concentration of 0.3 mM and induction allowed to proceed for 13 h.

Cells were harvested, resuspended in 250 mL of 20 mM Tris (pH 8), and centrifuged at 5000g for 5 min. The cell pellet was frozen in liquid nitrogen. The purification then proceeded as it did for the native protein with the following differences. After DNase I had been added, CaCl2 was added to a final concentration of 1 mM. Precipitation with ammonium sulfate was carried out on ice for 1 h instead of overnight. The Phenyl-Sepharose column was run with a 150 mL gradient. For the gel filtration column, the buffer contained 5 mM DTT and the protein was concentrated to 0.8 mL for loading on the column. The final dialysis buffer contained 1 mM DTT, and the protein was concentrated to 10 mg/mL before being frozen. The protein yield was 3.4 mg/g of wet cells.

Protein Crystallization. Frozen protein was thawed on ice and centrifuged at 15000g for 5 min. Crystallization was performed by the vapor-diffusion method with hanging-drop geometry at 18 °C. Initial crystallization conditions were achieved using Hampton Research Crystal Screen I. Conditions 41 [10% 2-propanol, 100 mM Hepes (pH 7.5), and 20% PEG 4000] and 22 [100 mM Tris (pH 8.5), 200 mM sodium acetate, and 30% PEG 4000] were found to produce crystals in two different morphologies (small diamonds and rod bundles, respectively).

Optimization of the crystal growth conditions resulted in 20–25% PEG 4000 and 150 mM sodium acetate for the diamond-shaped crystals and 5–8% PEG 3350 and 100 mM sodium acetate for the rods yielding crystals with final dimensions of 0.05 mm/edge for the diamonds and 0.1 mm wide x 0.5 mm long for the rods. For the tungstic acid isomorphous native data sets after refinement. Using the two SeMet sites for phasing, initial electron density maps showed a right-handed helix and confirmed the choice of the P4\textsubscript{2}12\textsubscript{1}2\textsubscript{1} space group, and not P4\textsubscript{1}2\textsubscript{1}2\textsubscript{1}. Polyalanine chains for this first helix and a \( \beta \)-sheet were initially built in O (18); this was followed by refinement using minimization and simulated annealing in CNS (19) with the MLH1 target function using Hendrickson–Lattman phases from SHARP. Electron density maps calculated using combined phases from SHARP and the partial model resulted in a marked improvement in map quality and allowed for rapid building of the model. Composite omit maps (20) were used to ensure proper registration and to rule out mistracing of the backbone. When 160 of 164 residues were built (\( R_{work} = 25.2\% \) and \( R_{free} = 29.7\% \)), the model was used to provide initial phases for the isomorphous high-resolution (1.9 Å) native data set. Rigid-body refinement in CNS with data to 3 Å followed by inclusion of all data to 1.9 Å with several rounds of model rebuilding, addition of water and acetate molecules, and individual-atom temperature factor refinement resulted in an \( R_{work} \) of 20.0% and an \( R_{free} \) of 22.5%. The final model comprises 162 of 164 residues (the last two are disordered), two acetates, and 99 waters and acetates. A Ramachandran plot generated in MOLPROBITY (21) showed 154 residues (96.3%) in the allowed regions, four residues (3.7%) in the generously allowed regions, and none in the disallowed regions.

MOLREP (22) was used to determine the structure of the MDP-1–WO\textsubscript{4}\textsuperscript{2–}–Mg\textsuperscript{2+} complex. Using the 1.9 Å MDP-1 structure as a search model with waters removed, MOLREP was able to automatically locate four monomers in the asymmetric unit with an overall correlation of 58.8% and an \( R_{factor} \) of 40.5%. A refinement scheme as in the native protein was followed. The final model contains two tungstate unit and the following unit cell dimensions: \( a = b = 59.7 \) Å, \( c = 99.7 \) Å, and \( \alpha = \beta = \gamma = 90^\circ \). The rod-shaped WO\textsubscript{4}\textsuperscript{2–} cocrystals soaked in MgCl\textsubscript{2} belong to space group P2\textsubscript{1}2\textsubscript{1} with four molecules per asymmetric unit and the following cell dimensions: \( a = 68.5 \) Å, \( b = 57.7 \) Å, \( c = 76.6 \) Å, \( \alpha = \gamma = 90^\circ \), and \( \beta = 92.0^\circ \). Data collection and refinement statistics are summarized in Table I.

Structure Determination and Refinement. Initial attempts to determine the structure of data to 2.8 Å collected from a native diamond-shaped crystal proved to be unsuccessful (data not reported) using several different models of HAD members with their cap domains removed. For all models, a clear rotation and translation solution was never obtained. Retrospectively, once the MDP-1 structure was determined, an attempt was made to solve the 1.9 Å data set using the same search models. Again, molecular replacement failed to identify the proper rotation solution unless the actual MDP-1 model was used. Thus, we attribute the failure of molecular replacement to the greater than 1.4 Å root-mean-square difference between MDP-1 and other HAD core structures.

The 2.8 Å SeMet data sets were used to determine the structure of MDP-1. AutoSHARP/SHARP (17) was used to locate two of the three selenium sites in the protein. These sites were consistent with anomalous Patterson maps. The third site, corresponding to the N-terminal methionine, was never visible in the model phased with the SeMet data. The N-terminal methionine was only observed in the high-resolution native data sets after refinement. Using the two SeMet sites for phasing, initial electron density maps showed a right-handed helix and confirmed the choice of the P4\textsubscript{2}12\textsubscript{1}2\textsubscript{1}2\textsubscript{1} space group, and not P4\textsubscript{1}2\textsubscript{1}2\textsubscript{1}2\textsubscript{1}. Polyalanine chains for this first helix and a \( \beta \)-sheet were initially built in O (18); this was followed by refinement using minimization and simulated annealing in CNS (19) with the MLH1 target function using Hendrickson–Lattman phases from SHARP. Electron density maps calculated using combined phases from SHARP and the partial model resulted in a marked improvement in map quality and allowed for rapid building of the model. Composite omit maps (20) were used to ensure proper registration and to rule out mistracing of the backbone. When 160 of 164 residues were built (\( R_{work} = 25.2\% \) and \( R_{free} = 29.7\% \)), the model was used to provide initial phases for the isomorphous high-resolution (1.9 Å) native data set. Rigid-body refinement in CNS with data to 3 Å followed by inclusion of all data to 1.9 Å with several rounds of model rebuilding, addition of water and acetate molecules, and individual-atom temperature factor refinement resulted in an \( R_{work} \) of 20.0% and an \( R_{free} \) of 22.5%. The final model comprises 162 of 164 residues (the last two are disordered), two acetates, and 99 waters and acetates. A Ramachandran plot generated in MOLPROBITY (21) showed 154 residues (96.3%) in the allowed regions, four residues (3.7%) in the generously allowed regions, and none in the disallowed regions.

MOLREP (22) was used to determine the structure of the MDP-1–WO\textsubscript{4}\textsuperscript{2–}–Mg\textsuperscript{2+} complex. Using the 1.9 Å MDP-1 structure as a search model with waters removed, MOLREP was able to automatically locate four monomers in the asymmetric unit with an overall correlation of 58.8% and an \( R_{factor} \) of 40.5%. A refinement scheme as in the native protein was followed. The final model contains two tungstate
The product formation was assessed at 405 nm (of 100\textsuperscript{initiated by the addition of MDP-1 and terminated by dilution -EDTA, and 0.25\textsuperscript{as a monomer with an empirical weight of 12 000). MDP-1 eluted

67 000) were used as reference standards. MDP-1 eluted

25 000), ovalbumin (MW = 45000), and albumin (MW = 67 000) were used as reference standards. MDP-1 eluted as a monomer with an empirical weight of 12 000.

ions and 416 waters and has a final $R_{\text{work}}$ of 19.1% and an $R_{\text{free}}$ of 23.6%. Positions of the two tungstate ions were also confirmed by a difference anomalous electron density map, showing that only two of the four monomers in the asymmetric unit were liganded. A Ramachandran plot generated in MOLPROBITY (21) showed 621 residues (96.7%) in the allowed regions, 21 residues (3.3%) in the generously allowed regions, and none in the disallowed regions.

An analysis of the interprotein packing between MDP-1 monomers in the two crystallographic unit cells showed different crystal contacts with no consistent oligomer unit. This is most consistent with the assignment of MDP-1 as a biological monomer.

**Phosphatase Activity Assay.** Reaction mixtures (1000 \textmu L) contained 200 mM MES (pH 5.5), 10 mM MgCl\textsubscript{2}, 2 mM EDTA, and 0.25–10 mM pNPP at 37 °C. Reactions were initiated by the addition of MDP-1 and terminated by dilution of 100 \textmu L of reaction mix into 900 \textmu L of 500 mM NaOH. The product formation was assessed at 405 nm (ε = 1.85 × 10\textsuperscript{5} M\textsuperscript{-1} cm\textsuperscript{-1}). Velocity data were plotted on an Eadie–Hofstee plot and kinetic constants calculated.

**Molecular Size Determination.** Gel filtration was carried out using the Amersham Biosciences LMW calibration kit. Proteins were resuspended in 50 mM Tris (pH 8), 1 mM MgCl\textsubscript{2}, and 100 mM NaCl. A Hitrap 26/60 Sephacryl S200 gel filtration column was prequillibrated in the same buffer. Ribonuclease A (MW = 13 700), chymotrypsinogen (MW = 25 000), ovalbumin (MW = 43 000), and albumin (MW = 67 000) were used as reference standards. MDP-1 eluted as a monomer with an empirical weight of 12 000.

**Sequence Alignments.** Database searches of the Genpept, EST, TrEMBL, and Swissprot databases (23, 24) using a combination of Psi-Blast searches and motif scans identify 16 different eukaryotic MDP-1 sequences. MDP-1 may be found in the interpro database (25) entry IPR010036 based on the TIGR01685 HMM model for this protein. This HMM identifies both eukaryote and archaea sequences. To date, only the mouse and Saccharomyces cerevisiae proteins have been cloned, and it is not known if the archaea show the same activity as eukaryotic MDP-1. We have therefore excluded these sequences as well as the eukaryotic MDP-1 sequence from Leishmania major (GenPept entry CAC14608). The Leishmania sequence is at most 25% identical with other sequences found and violates too many of the motifs used to identify MDP-1 (2).

**Substrate Modeling of Phosphotyrosine.** To examine the potential interactions of substrates in the active site of MDP-1, model compounds were manually docked into the active site. The position of the WO\textsubscript{4}\textsuperscript{2−} ion was used to place the phosphate moiety with the leaving group directed toward the active site opening on the surface. If the position of the phosphate atom is kept restrained to the expected position, the remainder of the substrate was allowed to move using energy minimization in CNS while keeping the protein coordinates fixed. During the calculation, charges were turned off such that only van der Waals interactions and geometry were used.

**RESULTS AND DISCUSSION**

**Structure Determination.** To prepare mouse MDP-1 for crystallization, the cloned gene was subcloned in pET-3 for
overexpression in *Escherichia coli*. The protein was purified to homogeneity (based on SDS-PAGE analysis) by FPLC adsorption and gel filtration chromatography. The protein molecular size was determined by gel filtration to be 12 000 Da as compared to the theoretical size of 18 600 Da which is most consistent with the interpretation that in solution MDP-1 exists as a monomer [which is also consistent with the interprotein interactions seen in the crystal forms (*vide supra*)]. The purified protein catalyzed the hydrolysis of ρNPP at pH 5.5 and 37 °C with a *k*<sub>cat</sub> of 38 s<sup>-1</sup> and a *K*<sub>m</sub> of 3.7 mM. These values yield a *k*<sub>cat</sub>/*K*<sub>m</sub> value that is comparable to that reported in the original isolation of MDP-1 from mouse muscle (13). The protein crystallized from a PEG/sodium acetate solution at pH 7.5–8.5. Crystals of the WO<sub>4</sub><sup>2-</sup>-bound enzyme were soaked with MgCl<sub>2</sub> to obtain the protein structure with the metal cofactor and phosphate analogue bound to the active site.

**Quaternary and Tertiary Structure.** The native structure, determined at 1.9 Å resolution, reveals an elliptical monomer consisting of a single domain with overall dimensions of 30 Å × 34 Å × 53 Å (Figure 2a,b). This single domain is a modified Rossmann fold comprised of a centrally located six-stranded parallel β-sheet (β<sub>1</sub>–β<sub>6</sub>) surrounded by six α-helices (α<sub>5</sub>–α<sub>10</sub>) (Figure 2d). Typical right-handed β–α–β connections are present with loops connecting the secondary structural elements (loops I–IV) for the four conserved sequence motifs shared among HAD superfamily members. Between loops I and II resides a 28-amino acid long insert (residues 20–47; see Figure 2) which resembles a highly truncated version of the cap domain found in subclass I HAD family members. A similar insert exists in the capless HAD superfamily member 8-KDO phosphatase (Figure 3) wherein it is used as an oligomerization domain. Because MDP-1 is a monomer and lacks a cap domain, the active site is fully solvent exposed. The α/β domain is similar to the core domain common to all other HAD superfamily members. A search of the DALI (26) database identified 13 subclass I and II HAD superfamily members (see the introductory section) with *Z* scores greater than 8 (next highest being 6.8), with 9–22% of the sequence being identical and rms deviations of 2.3–3.8 Å. The most closely related core domain folds belong to β-PGM (5), phosphonatase (6), L-2-haloacid dehalogenase (9), 8-KDO phosphatase (11, 12), phosphoserine phosphatase (7), the lipid phosphatase domain of the bifunctional epoxide hydrolase phosphatase (27, 28), TA0175 (29), and T4 polynucleotide kinase (10).

**Mg(II) and Phosphate-Binding Residues of the Active Site.** 

The Mg(II) and WO<sub>4</sub><sup>2-</sup> binding sites stationed on the four loops which comprise the active-site scaffold in MDP-1 are depicted in Figure 4. The Mg(II) bound in this structure...
exhibits conventional orthogonal geometry with six ligands. Direct ligands to the Mg(II) include, from loop I, one oxygen of the Asp 11 carboxylate (2.3 Å) and the backbone carbonyl oxygen of Asp 13 (2.3 Å) and, from loop four, one of the carboxylate oxygens of Asp 123 (2.2 Å). Two of the ligands are water molecules (2.2 and 2.4 Å), one of which forms hydrogen bonds with the side chains of conserved loop IV Asp 122 (3.1 Å), the loop III NZ of Lys 100 (2.9 Å), and the backbone amide nitrogens of conserved residues Leu 12, Asp 13 (loop I), and Arg 70 (loop II). The nucleophilic Asp 11 carboxylate oxygen is positioned almost equidistant from three of the oxygens of the WO$_4^{2-}$ molecule (at 2.6, 2.7, and 2.7 Å) and in the proximity of the tungsten atom (2.7 Å). Thus, by analogy, the aspartate would be in the correct position relative to the phosphorylated substrate for in-line nucleophilic attack at the phosphorus atom. This arrangement of ligands to WO$_4^{2-}$ closely mirrors that found in the structures of WO$_4^{2-}$-bound phosphonatase (PDB entry 1FEZ) (6), phosphate-bound phosphoserine phosphatase (PDB entry 1F55) (7), β-D-glucose 1,6-(bis)phosphate bound to β-PGM (PDB entry 1O03) (30), the WO$_4^{2-}$-bound structure of phosphonatase (PDB entry 1FEZ) (6), and the phosphate-bound structure of human mitochondrial deoxyribonucleotidase (PDB entry 1MH9) (31), a similar arrangement is observed with one ligand from the substrate, two ligands from loop I (the catalytic aspartate carboxylate group and the backbone carbonyl oxygen two residues downstream), and three from aspartates in loop IV (either directly or though water). The variability is in one of these water ligands to Mg(II). Usually, there is a water molecule present that bridges the loop I backbone carbonyl oxygen and the loop IV aspartate. In phosphonatase, this water is not present but instead is replaced by the side chain hydroxyl of Ser 209. Overall, MDP-1 utilizes the same constellation of amino acids for Mg(II) binding as other HAD family phosphotransferases.

In HAD phosphotransferases, the phosphate moiety of the substrate is anchored by the Mg(II) cofactor and by residues of the core domain active-site scaffold. The binding site for phosphate is occupied by WO$_4^{2-}$ in the MDP-1 structure. The WO$_4^{2-}$ is coordinated to Mg(II) (2.5 Å) and forms hydrogen bonds with the side chains of conserved loop II Ser 69 OG (3.1 Å), the loop III NZ of Lys 100 (2.9 Å), and the backbone amide nitrogens of conserved residues Leu 12, Asp 13 (loop I), and Arg 70 (loop II). The nucleophilic Asp 11 carboxylate oxygen is positioned almost equidistant from three of the oxygens of the WO$_4^{2-}$ molecule (at 2.6, 2.7, and 2.7 Å) and in the proximity of the tungsten atom (2.7 Å). Thus, by analogy, the aspartate would be in the correct position relative to the phosphorylated substrate for in-line nucleophilic attack at the phosphorus atom. This arrangement of ligands to WO$_4^{2-}$ closely mirrors that found in the structures of WO$_4^{2-}$-bound phosphonatase (PDB entry 1FEZ) (6), phosphate-bound phosphoserine phosphatase (PDB entry 1F55) (7), glucose 1,6-(bis)phosphate-bound β-PGM (PDB entry 1O03) (30), and phosphate-bound human mitochondrial deoxyribonucleotidase (PDB entry 1MH9) (31). The loop I...
amide nitrogen interactions are all retained as well as the loop II hydroxyl group of serine or threonine. In all but phosphatase there is an interaction between the substrate phosphoryl moiety and the loop III lysine or arginine and an amide nitrogen in loop II (in the 3.0 Å structure of phosphatase complexed to WO₄²⁻, these backbone interactions are lacking). Thus, MDP-1 uses the same mode of phosphate binding as other phosphotransferase HAD family members.

Acid–Base Chemistry of the Active Site. Phosphotransferases (excluding phosphotase) of the HAD family utilize the position two residues downstream from the Asp nucleophile (typically a protonated Asp) as an acid catalyst which will protonate the oxygen atom displaced from the phosphorus (in ATPases, formation of a hydrogen bond with the displaced atom suffices) (5). This is also the case with MDP-1, wherein Asp13 is positioned to protonate the oxygen atom displaced from the substrate phosphorus by Asp 11. Interaction with other regions of the leaving group serves as the basis for substrate specificity among HAD phosphotransferases. The leaving group projects into the entrance of the core domain active site (toward solvent), where it can interact with residues lining the entrance or as in the case of the subclass I and II HAD members, with residues located on the interface of the core and cap domains. In MDP-1, there is no cap domain and there is no adjacent subunit to act as an active-site cap. Therefore, residues which may participate in substrate anchoring via interaction with the leaving group moiety will necessarily be among the stringently conserved residues located at the active-site entrance (see below).

Active-Site Solvent Accessibility, Topology, and Substrate Recognition. In HAD family members that contain cap domains (subclasses I and II), substrate binding induces closure of the cap domain over the core domain active site, excluding bulk solvent (32). In fact, the binding of the phosphate analogue, WO₄²⁻, is sufficient to stabilize the closed conformation relative to the open conformation (5, 6, 8, 9, 30). In these HAD family members, binding and creation of substrate specificity is tantamount to excluding solvent. Solution studies of phosphate monoester cleavage demonstrate that an increase in the hydrolysis rate in dipolar aprotic solvents versus aqueous solvent is most consistent with a weakening of the ground-state P–O ester bond (33). In addition, it has been suggested that the removal of solvent water may affect the positions of equilibria of group transfer reactions of phosphoric acid esters, conferring rate enhancement from desolvation alone (34). Another catalytic strategy is for the active site to stabilize charges on intermediates and transition states more effectively than water (35). Catalysis will be enhanced to the extent that the intermediates and transition state are more favorably “solvated” by the active-site residues than the ground state. Thus, it is catalytically advantageous for phosphotransferases to provide a water exclusive active-site environment for phosphoryl transfer.

Given the fact that both solution studies and X-ray crystallographic results are consistent with a monomeric state for MDP-1, it is apparent that MDP-1 cannot use another monomer to close or form the active site as observed in 8-KDO phosphatase (11). The active site of apo-MDP-1 is fully exposed to solvent, and the protein structure is essentially unchanged by the binding of WO₄²⁻. How, then, does MDP-1 seal its active site for catalysis? One possibility is that the 28-amino acid insert between loops I and II (Figure 2b) undergoes a conformational change that allows it to cover the active site. Two lines of evidence argue against an active-site gating function for this structural element. First, sequence alignments of the MDP-1 sequences from other eukaryotes show that roughly half of these orthologs have significantly smaller inserts (Figure 5). If the insert played a functional role, the structure would be conserved. Second, in all structures of MDP-1, the insert is pinned to the outer surface of the protein via a series of hydrophobic and hydrogen bond interactions between β2 of the insert and α3 of the core (see Figure 2c). These interactions would be disrupted if the peptide were to cover the active site.

A second possible mechanism for active-site desolvation is substrate binding. Inspection of the topological features of the catalytic site, and its entrance (see Figure 6a), suggests that the phosphoryl group is located in a narrow passageway, the bottom of which is occupied by the Mg(II) cofactor and Asp 11 nucleophile. It could fill the entrance to the active site, provided that the leaving group of the substrate is large. If the active-site entrance is in fact used in substrate docking, the surface residues will be conserved. To test this idea, the invariant residues identified from the alignment of 16 MDP-1 sequences (Figure 5) were mapped onto the surface of the MDP-1 structure (Figure 6a). To validate such conservation analysis, the same mapping was performed on two HAD family members: the first lacking a cap domain and known to utilize a large substrate (T4 PNK, Figure 6b) and the second utilizing the cap domain to provide binding specificity for a small substrate (phosphoserine phosphatase, Figure 6c). Comparison of the two capless enzymes, MDP-1 and T4 PNK, shows that the invariant residues are not nearly confined to the sequence motifs (loops I–IV) comprising the catalytic machinery (see Figure 1S of the Supporting Information for the sequence alignment). Rather, they form a surface surrounding the active-site entrance. In contrast, for the small substrate-utilizing enzyme phosphoserine phosphatase, the invariant residues of the core domain are located solely in the conserved loops (I–IV). Substrate specificity is confined to the cap domain which includes a number of invariant residues (see Figure 2S of the Supporting Information for the sequence alignment).

Similar conservation of putative substrate-binding residues has been observed in three groups of protein tyrosine phosphatases (PTPs): the classical PTPs, the low-molecular weight PTPs, and the dual-specificity PTPs. In the classical PTPs, some residues proximal to the active site are highly conserved and are likely involved in binding of all phosphotyrosine substrates, whereas others are more variable and involve recognition of substrates between PTP subtypes (36). In the low-molecular weight PTPs, a loop, which determines the isoenzyme specificity, lines the active-site crevice (37). In the dual-specificity PTPs, a conserved cluster of residues in the noncatalytic amino-terminus domain mediates protein–protein interactions (38). On the basis of these examples, the number and location of identical residues on the surface of MDP-1 are suggestive of a role for this region in binding phosphorylated peptides and/or proteins.

Substrate Identity. The requirements for a large, phosphorylated substrate might be met by a phosphorylated poly-
nucleotide, a phosphorylated polysaccharide, or a phospho-
rylated protein. To examine these three possibilities, we have
manually docked small sugar phosphates and phosphoty-
rosine in the active site of the apo-MDP-1 structure. The
molecular surface of the MDP-1 active site conforms
optimally, in shape and electrostatic potential, to the phenyl
phosphate group of phosphotyrosine. Figure 7 shows the
MDP-1 active site into which phosphotyrosine has been
docked by overlaying the phosphate group with the bound
WO₄²⁻. All of the ligands to the WO₄²⁻ ion from loops I
are maintained. These include Asp 13, Ser 69, and Lys 100.
Asp 13 is positioned to protonate the leaving group oxygen
(2.8 Å away). With regard to the remainder of the leaving
group, the only interactions that might be predicted are a
ring stacking interaction between Trp 20 and the substrate
phenyl ring and a face-on interaction of Arg 70 with the
plane of the phenyl ring. It is interesting to note that Arg 70
is absolutely conserved among all MDP-1 sequences while
Trp 20 is sometimes replaced with Tyr, another aromatic
amino acid.

Substrate specificity is expected to be derived from
interactions with conserved residues at the surface of the

Figure 5: Sequence alignment of MDP-1. Loops (motifs) used to find and align superfamily members are denoted with red lines (note that one or two of the amino acids at the end of the motifs are located at the ends of α-helices or β-strands and not loops). The 28-amino acid insert is denoted with a green line. Secondary structural elements are labeled. Identical residues (90% level) are highlighted cyan, and similar residues are highlighted in orange. Accession numbers for sequences as follows: AAK00763.1 for mouse, XP_214200.1 for rat, NP_612485.2 for human, A1781702.2 for tomato, B267188.3 for barley, BE339957.2 for potato, AW146994.3 for corn, BE020617.3 for soybean, BF647363.1 for Medicago truncatula, SPBP873.1 for Schizosaccharomyces pombe, YE47_YEAST for S. cerevisiae, CB82279.2 for taurus, CD051815.1 for goat, AK14769.1 for Ciona intestinalis, and AAP30406.1 for Oryza sativa (japonica cultivar group).
active-site entrance. If the substrate for MDP-1 is indeed a phosphorylated protein, then the HAD family fold has been recruited to serve in signal transduction via protein dephosphorylation. Recently, Eya, a nuclear transcriptional coactivator possessing a conserved HAD core domain, has been shown to possess protein tyrosine phosphatase activity. Oblating this enzymatic activity by mutation of the HAD motifs compromises in vivo function (39). Unlike the classical protein tyrosine phosphatases which utilize an active-site Cys residue in nucleophilic catalysis (14, 15), the HAD family phosphatase is equipped with an active-site Asp to serve in nucleophilic catalysis and a Mg(II) cofactor to serve in electrophilic catalysis. Why might the HAD scaffold have been recruited for use as a phosphotyrosine phosphatase? The acidity of the Asp nucleophile eliminates the need for activation by base catalysis (as observed in classical cysteine-utilizing phosphotyrosine phosphatases) and would allow operation over a wide pH range. The high intrinsic energy of the aspartyl phosphate can be used to drive phosphoryl transfer to water in the hydrolysis step (40). At present, we hypothesize MDP-1 to be a protein tyrosine phosphatase. Ongoing studies in our laboratories are aimed at identification of the physiological substrate of MDP-1.

FIGURE 6: (a) Surface of MDP1 looking down on the active-site pocket marked by the Mg(II) ion in magenta. The conserved core residues in loops I–IV that bind and transfer phosphate are colored yellow. Identical residues in all MDP-1 sequences (Figure 5) are colored cyan. (b) Surface of the phosphatase domain of T4 polynucleotide kinase (PDB entry 1LTQ) (10) colored as in panel a using a sequence alignment (see Supporting Information) of polynucleotide kinases. (c) Surface of phosphoserine phosphatase (PDB entry 1F5S) (7) with the cap domain (residues 19–67) removed and colored as in panel a based on sequence alignment (see Supporting Information) of phosphoserine phosphatase sequences. Note that panels b and c are in an orientation identical to that of a. Surfaces calculated with MSMS (42) and figures rendered in Dino (http://www.dino3d.org).

FIGURE 7: Model of phosphotyrosine docked into the active site of MDP-1. Residues that bind and interact directly with the phosphate moiety are colored pink. Residues surrounding the leaving group of the putative substrate are colored blue. Arg 70 and Trp 20 may reorient to interact with the leaving group of the substrate. Surface calculated with MSMS (42) and figure rendered in Dino (http://www.dino3d.org).

ACKNOWLEDGMENT

Data were collected at beamline X12C of the National Synchrotron Light Source, the financial support for which comes principally from the Office of Biological and Environmental Research and of Basic Energy Sciences of the U.S. Department of Energy, and from the National Center for Research Resources of the National Institutes of Health. We thank the beamline personnel for expert assistance with data collection.

SUPPORTING INFORMATION AVAILABLE

Sequence alignments of polynucleotide 5′-kinase-3′-phosphates and phosphoserine phosphatases. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES


