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

4 AUTHORS, INCLUDING:

Ezra Peisach
Rutgers, The State University of New Jersey
17 PUBLICATIONS **858** CITATIONS

SEE PROFILE

Jeremy Selengut
University of Maryland, College Park
55 PUBLICATIONS **12,291** CITATIONS

SEE PROFILE

Karen N Allen
Boston University
103 PUBLICATIONS **3,198** CITATIONS

SEE PROFILE

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*

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 eukaryotic 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)[1] enzyme superfamily is a large family of enzymes, which function in catalysis of group transfer reactions utilizing an Asp carboxylate as the nucleophile (*1*). 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 phosphate/Asp nucleophile liganding residue is located on loop III. The Mg(II) cofactor Asp/Glu ligands are located on loop IV. Members of subfamily I [including ʟ-2-haloacid dehydrogenase (*3, 4*), β-phosphoglucomutase (β-PGM) (*5*), phosphonoacetaldehyde hydrolase (phosphonatase) (*6*), and phosphoserine phosphatase (*7*)] contain a sequence insert between loops I and II, while members of subfamily II (including phosphomannomutase and sucrose 6[F]-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-ᴅ-

* To whom correspondence should be addressed. K.N.A.: Department of Physiology and Biophysics, Boston University School of Medicine, 715 Albany St., Boston, MA 02118-2394; telephone, (617) 638-4398; fax, (617) 638-4273; e-mail, allen@med-xtal.bu.edu. D.D.-M.: Department of Chemistry, University of New Mexico, Albuquerque, NM 87131; telephone, (505) 277-3383; fax, (505) 277-6202; e-mail, dd39@unm.edu.

[§] Boston University School of Medicine.

[‖] The Institute for Genomic Research.

[⊥] University of New Mexico.

[1] Abbreviations: HAD, haloacid dehalogenase; IPTG, isopropyl β-ᴅ-thiogalactopyranoside; 8-KDO, 3-deoxy-ᴅ-*manno*-octulosonate 8-phosphate; LB, Luria broth; MDP-1, magnesium-dependent phosphatase-I; pNPP, *p*-nitrophenyl phosphate; PMSF, phenylmethanesulfonyl fluoride; β-PGM, β-phosphoglucomutase; phosphonatase, phosphonoacetaldehyde hydrolase; PTPs, protein tyrosine phosphatases; rms, root-mean-square deviation; SeMet, selenomethionine.
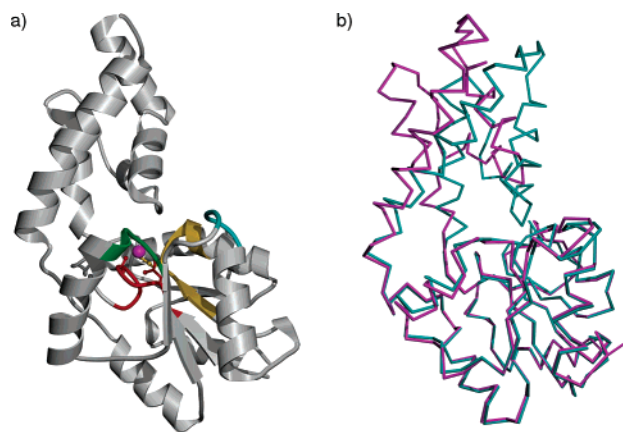
FIGURE 1: (a) Phosphonatase (PDB entry 1FEZ) (*6*) with the core loops colored. Loops I−IV are colored red, green, cyan, and gold, respectively. (b) Cap domain movement in β-PGM from open (PDB entry 1LVH) (*5*) to closed (PDB entry 1O03 (*30*). This figure was rendered in MOLSCRIPT (*41*) and POVRAY (Persistence of Vision Ray Tracer, version 3.5).

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 *Bam*HI, *Nde*I, 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 chloroamphenicol were purchased from American Bioanalytical. DTT and phenyl-methanesulfonyl fluoride (PMSF) were from Sigma, and DNAse I was from Worthington Biochemical.

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 Clontechs's high-fidelity cloning kit (Advantage HF-2) was used to remove the gene from the plasmid and introduce *Nde*I and *Bam*HI restriction sites to the 5′-end and 3′-end,

respectively. Oligonucleotides 5′-GAGATATAC<u>CATAT</u>GACGCGGCTGCCAAAGCTT-3′ for the forward primer and 5′-CAGCC<u>GGATCC</u>TTATTATCAGAGTCCAGCTT-GGGCCTT-3′ for the reverse primer were used (*Nde*I and *Bam*HI sites underlined, respectively). The PCR product was gel purified, and subjected to double-restriction enzyme digest with *Nde*I and *Bam*HI. 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 OD_{600} 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 5000g, 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 DTT, 1 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 17000g. The cloudy supernatant was removed, and 3.5 M NH_4SO_4 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 38700g, 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 DTT, 1 mM EDTA, and 1 M NH_4SO_4. Protein was eluted with a 100 mL gradient of 1 M NH_4SO_4 to buffer alone. Fractions were assayed for phosphatase activity toward *p*-nitrophenyl phosphate (*p*NPP) (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 Sephracyl S200 gel filtration column preequilibrated with 50 mM Tris (pH 8.0), 1 mM $MgCl_2$, 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), 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$\times$ 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 inoculums) were grown with shaking at 250 rpm and 37 °C. After 8 h ($A_{600} = 0.3$), 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, CaCl$_2$ 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 × 0.5 mm long for the rods. For the tungstate complex of MDP-1, protein was cocrystallized in the presence of 1 mM tungstic acid under the conditions that lead to the growth of rod-shaped crystals (see above), and the resulting crystals were soaked in 2 mM MgCl$_2$ for 15 min before being frozen for data collection.

Data Collection. Crystals were transferred to FMS oil (Hampton Research), the surface liquid was removed, and the crystal flash-frozen in a stream of N$_2$ gas at cryogenic temperatures. The DENZO/SCALEPACK package (*16*) was used for all data processing. A 2.8 Å resolution three-wavelength MAD data set was collected at beamline X12C at Brookhaven National Laboratory (Upton, NY) on a single diamond-shaped SeMet MDP-1 crystal. Subsequent data sets were collected on a Rigaku RU300 generator, equipped with Osmic mirrors and a Raxis IV^{++} area detector on native diamond-shaped crystals, and rod-shaped WO$_4^{2-}$ cocrystals soaked in Mg^{2+}. The diamond-shaped native crystals belong to space group $P4_32_12$ with one molecule per asymmetric

unit and the following unit cell dimensions: $a = b = 59.7$ Å, $c = 99.7$ Å, and $\alpha = \beta = \gamma = 90°$. The rod-shaped WO$_4^{2-}$ cocrystals soaked in MgCl$_2$ belong to space group $P2_1$ with four molecules per asymmetric unit and the following cell dimensions: $a = 68.5$ Å, $b = 57.7$ Å, $c = 76.6$ Å, $\alpha = \gamma = 90°$, and $\beta = 92.0°$. 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 $P4_32_12$ space group, and not $P4_12_12$. Polyalanine chains for this first helix and a β-sheet were initially built in O (*18*); this was followed by refinement using minimization and simulated annealing in CNS (*19*) with the MLHL 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$_4^{2-}$−Mg^{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

Table 1: Crystal Data Collection and Refinement Statistics

| | SeMet MDP-1 | | | | |
	edge	peak	remote	native	WO_4^{2-} bound
source	BNL X12C			RU300	RU300
detector	Brandeis CCD			Raxis IV++	Raxis IV++
wavelength	0.97910	0.97874	0.95004	Cu Kα	Cu Kα
cell dimensions (Å)	$a = b = 59.66$ Å			$a = b = 59.41$ Å	$a = 68.52$ Å
	$c = 99.69$ Å			$c = 99.56$ Å	$b = 57.65$ Å
					$c = 76.62$ Å
					$\beta = 91.97°$
space group	$P4_32_12$			$P4_32_12$	$P2_1$
resolution (Å) (outer shell)[a]	50−2.8 (2.9−2.8)			100−1.9 (1.97−1.9)	100−1.85 (1.92−1.85)
no. of total/unique reflections	294122/4874	294439/4851	149980/4831	153941/14718	494282/48167
completeness (%)[a]	99.9 (100.0)	99.9 (100.0)	99.1 (99.6)	99.7 (99.7)	93.8 (88.3)
$I/\sigma(I)$[a]	43.5 (16.9)	42.6 (17.3)	26.8 (9.42)	20.8 (4.6)	19.6 (2.8)
R_{merge} (%)[a,b]	6.1 (14.8)	7.1 (26.2)	6.3 (17.2)	4.4 (39.0)	5.4 (41.9)
f'/f'' (e−)	−9.2/2.6	−7.2/4.5	−2.5/3.0		
figure of merit (%)	43.1				

Refinement Statistics		
	native	WO_4^{2-} bound
no. of protein atoms per asymmetric unit	1309	5194
no. of reflections (work/free)	12812/1451	38918/4380
R_{work}/R_{free} (%)	20.0/22.5	19.1/23.6
average B-factor (Å²)		
amino acid residues	21.6	27.6
Mg(II) ions		35.1
WO_4^{2-}		75.3
acetate	45.4	
water	31.1	32.9
rms deviation		
bond lengths (Å)	0.0053	0.0053
bond angles (deg)	1.29	1.32

[a] Values in parentheses represent values for the high-resolution shell as indicated in the resolution row. [b] $R = \sum(I - \langle I \rangle)^2/\sum I^2$.

ions and 416 waters and has a final R_{work} of 19.1% and an R_{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 μL) contained 200 mM MES (pH 5.5), 10 mM $MgCl_2$, 2 mM EDTA, and 0.25−10 mM *p*NPP at 37 °C. Reactions were initiated by the addition of MDP-1 and terminated by dilution of 100 μL of reaction mix into 900 μL of 500 mM NaOH. The product formation was assessed at 405 nm ($\epsilon = 1.85 \times 10^4$ M⁻¹ cm⁻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 $MgCl_2$, and 100 mM NaCl. A Hitrap 26/60 Sephracyl S200 gel filtration column was preequilibrated 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 IPR10036 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_4^{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
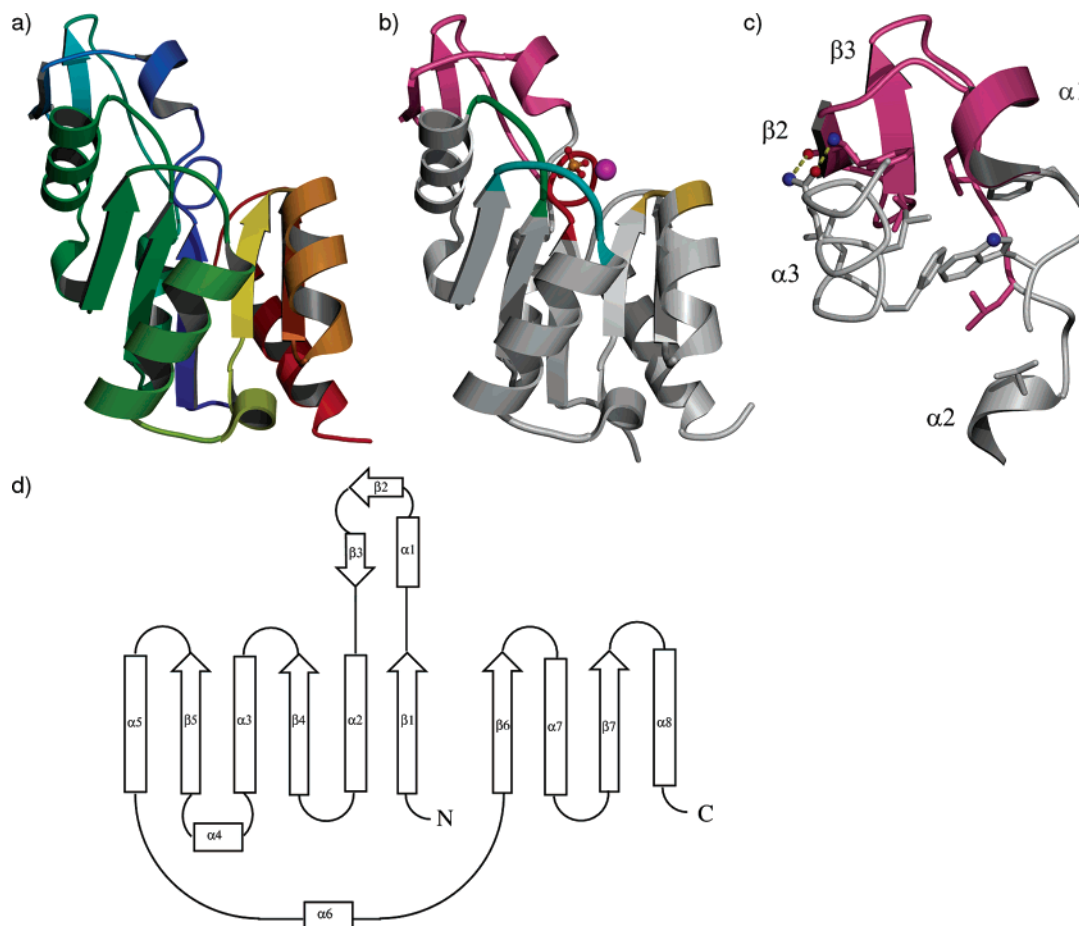
FIGURE 2: (a) Structure of unliganded MDP1 colored from the N-terminus (blue) to the C-terminus (red). The active-site Mg(II) ion is colored magenta. This figure was rendered in MOLSCRIPT (*41*) and POVRAY. (b) Ribbon diagram of tungstate-liganded MDP-1 with active-site loops I−IV colored red, green, cyan, and gold, respectively, and the insert colored pink. The tungstate is shown as an orange ball and stick, and the Mg(II) is colored magenta. (c) Hydrophilic and hydrophobic interactions between the core domain (gray) and the insert (pink). Note that the insert is approximately in the same orientation as panels a and b and that α3 is rendered as a coil for clarity. (d) Topological diagram representation of the secondary structure of MDP-1. The insert is comprised of secondary structural elements α1, β2, and β3.

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 *p*NPP at pH 5.5 and 37 °C with a k_{cat} of 38 s^{-1} and a K_m of 3.7 mM. These values yield a k_{cat}/K_m 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_4^{2-}-bound enzyme were soaked with $MgCl_2$ 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 (β1-β6) surrounded by six α-helices (α5-α10) (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_4^{2-} 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
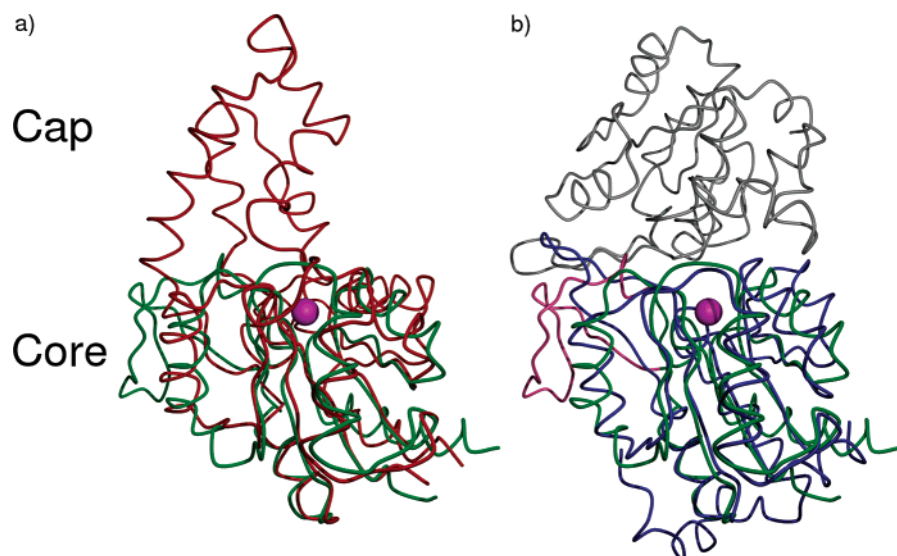
FIGURE 3: Overlay of the backbones of MDP-1 (green) with (a) phosphonatase (red) (PDB entry 1FEZ) (*6*) and (b) two monomers of the homotetramer of 8-KDO phosphatase (blue and gray) (PDB entry 1J8D) (*11*) and the insert of MDP-1 colored pink. The active-site Mg(II) of MDP-1 is shown as a magenta sphere.
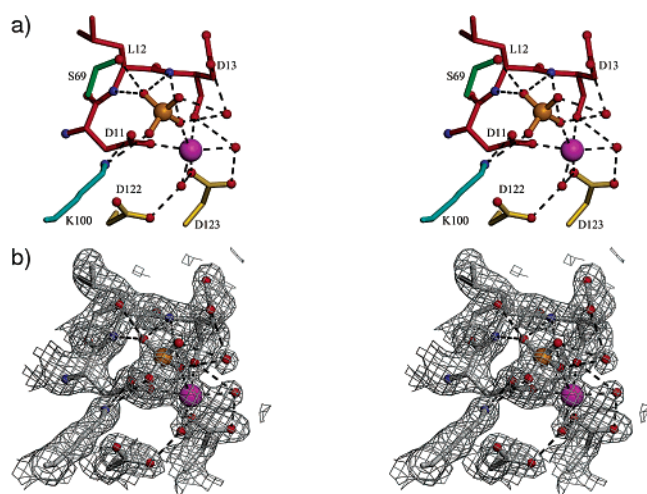

FIGURE 4: (a) Stereoimage of the phosphate-binding portion of the MDP-1 active site. Mg(II) is colored magenta, and WO_4^{2-} (representing phosphate) is colored orange. Residues from loops I−IV are colored red, green, cyan, and gold, respectively. (b) Stereoimage of a simulated annealing omit $2F_o − F_c$ electron density map contoured at 1σ.

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 carboxylate oxygens of the conserved loop IV Asp 122 (2.5 Å) and Asp 123 (3.1 Å) residues and forms a hydrogen bond to ND2 (3.2 Å) of the conserved Asn 127. The other water is held in place by interactions with one carboxylate oxygen of Asp 123 and the backbone carbonyl oxygen of Asp 13. This leaves one Mg(II) ligand available for substrate binding, which in the WO_4^{2-}-bound MDP-1 is occupied by an oxygen of WO_4^{2-} (2.5 Å) and in the unliganded structure is occupied by a water molecule. In other type I phosphotransferase superfamily members for which structures are available with either a substrate or inhibitor bound in the active site, the arrange-

ments of Mg(II)-binding ligands are very similar. In the phosphate-bound structure of phosphoserine phosphatase (PDB entry 1F5S) (*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 Lys100 (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 1F5S) (*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 phosphonatase 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 phosphonatase complexed to WO_4^{2-}, 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 phosphonatase) 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, WO_4^{2-}, is sufficient to stabilize the cap-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 WO_4^{2-}. 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 $\beta 2$ of the insert and $\alpha 3$ 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 inserted through 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 merely 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-terminal 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-

```
                              Motif I                    Insert
                          β1            α1      β2        β3
                              10      20      30        40

Mouse         1  MT................RLPKLAVFDLDYTLWPFWVDTHVDPPFH..KSSDGT.....VRDRRGQNIQ
Rat           1  MT................RLPKLAVFDLDYTLWPFWVDTHVDPPFH..KSSDGT.....VRDRRGQDIR
Human         1  MA................RLPKLAVFDLDYTLWPFWVDTHVDPPFH..KSSDGT.....VRDRRGQDVR
Tomato        1  MGD.EKVKNEAMEIMGLFQVLPRLVVFDLDYTLWPFYCECRSK..................NEKPS....
Sorghum       1  MGD.ERVKAEALQILGLFQVLPRLVVFDLDYTLWPFYCECRSK..................RDSPS....
Barley        1  MGD.KRVKAEALQILGLFQVLPRLVVFDLDYTLWPFYCECRSK..................RESPS....
Potato        1  MGD.EKVKNEAMEIMGLFQVLPRLVVFDLDYTLWPFYCECRSK..................NEKPS....
Corn          1          LQILGLFQVLPRLVVFDLDYTLWPFYCECRSK..................RDSPS....
Soybean       1   E.EKVKSEAVQIIEAFEVLPKLVVFDLDYTLWPFYCECRSK..................REIPS....
M. Truncatula 1  GESSEKAKSEAVKIIESFEVVPKLVVFDLDYTLWPFYCECRSK..................RESPS....
S. Pombe      1  M.....VKNIE.........FPKCVVFDLDYTLWPLWIDTHVTAPFKPSKNDPGV.....LIDKYGTEIC
S. Cervisiae  1  MT................GYPDVAAFDLDYTIWPCYCDTHLHGPFKPVKSSNGEVLTIICRDGYE..LT
Taurus        1  MA................RLPKLAVFDLDYTLWPFWVDTHVDPPFH..KNSDGT.....VRDRRGQSIQ
Goat          1  EQGLPPCAQPAGT..GSWQRLPKLAVFDLDYTLWPFWVDTHVDPPFH..KSSDGT.....VRDRRGQSIQ
Ciona         1  VKFVFLGMESGHGPK......PKLVVFDLDYTLWPFWVDTHFTPPFH..KDSDGN.....VLDSRQAKIL
Japan. Rice   1  MGD.ERVKAEALQILGRFEALPRLVVFDLDHTIWPLYCDCCSI..................GDSPR....

        Insert            Motif II                      Motif III
                    β4         α3        α4      β5
               α2                                        α5
               50    60        70    80        90    100

Mouse         47 LYPEVPEVLGRLQSLGVPVAAAASRTSEIQGANQLLELFDLG.........KYFIQREIYP..GSKVTHFE
Rat           47 LYPEVPEILGRLQSLGVPVAAAASRTSEIEGANQLLELFDLG.........KYFIQREIYP..GSKVTHFE
Human         47 LYPEVPEVLKRLQSLGVPGAAAASRTSEIEGANQLLELFDLF.........RYFVHREIYP..GSKITHFE
Tomato        48 MYPQAKGILHAFKDKGVNVAIASRSPTPDIANAFLQKLEIK.........SMFVAQEIFSSWTHKTEHFQ
Sorghum       48 LFRHARGIIYALKEKGIDMAIASRSPTPDKAKVFLDKLELQ.........SMFVAQEIFSSWTHKTEHFQ
Barley        48 LYRHAKGIMCALKEKGVDMAIASRSPTSDIARAFIDKLELQ.........PMFVAQEIFSSWTHKTEHFQ
Potato        48 MYPQAKGIMHAFKDKGVNVAIASRSPTPDVANAFLQKLEIK.........SMFVAQEIFSSWTHKTEHFQ
Corn          38 LFRHARGIMYALKEKGIDMAIASRSPTPDIAKVFLDKLEFQ.........SMFVAQEIFSSWTHKTEHFQ
Soybean       46 LYPHAKGILLALKEKGIDIAIASRSPTADIATAFLNKLNLT.........PMFVSQEIYSSWTHKTDHFQ
M. Truncatula 49 LYPHAMGILLALKHKGIDIAIASRSPTADIAKAFINKLGIT.........SFFVAQEIYSSWTHKTDHFQ
S. Pombe      52 FYSDITGILQELRNQKVTLCVASRTCAPKYAKQALNLMKVPIDGSLKPAIEFFTYVKAWP..GSKMDHFK
S. Cervisiae  52 IYKDIPRILGDLKDNGVKLMTASRTWAPEIAQEILKIFKVKYAGVVTPLANLFDEFQ.WGE.RSKIGHLR
Taurus        47 LYPEVPEVLERLRGLGVPIAAAASRTGEVEGANQLLELFDLV.........RYFVHREIYP..GSKVTHFE
Goat          62 LYPEVPEVLERLRGLGVPIAAAASRTGEVEGANQLLELFDLV.........RYFVHREIYP..GSKVTHFE
Ciona         58 LYPDSDDILKQLTADGYTIAVASRTSCTAEANNLLEKFDLN.........KYISHKQIYP..GCKKNHFS
Japan. Rice   48 LFRHAKGIMCALKEKGIAMAVASRSSTPDIANDFLDKLELQ.........PMFVTKEIFDSWTHKTEHFQ

                        Motif IV
                    β6          β7
               α5     α6    α7            α8
               110   120   130      140   150    160

Mouse         106 ....RLHHKTGVPFSQMVFFDDENRNIIDVGRLGVTCIHIR..DGMSLQTLTQGLETFAKAQAGL
Rat           106 ....RLRNKTGVPFSQMIFFDDENRNIIDVGKLGVTCIHIQ..DGMSLQMLTQGLETFAKAQAGL
Human         106 ....RLQQKTGIPFSQMIFFDDERRNIVDVGKLGVTCIHIQ..NGMNLQTLSQGLETFAKAQTGP
Tomato        109 ....KINRKTTVPYNEMLFFDDENRNIQAVSKMGVTSILVG..KGVNLGALRQGL
Sorghum       109 ....KIQRKTGIPYKSMLFFDDEDRNIDPVSNLGVTSVLVE..NGVNLDMFKLGLSNFATK
Barley        109 ....KIHRKTGVPYKSMLFFDDENRNIQSVSKMGVTSVLVD..NGLNLDMFKLGLSNFATTVVAP
Potato        109 ....KINRKTAVPYNEMLFFDDENRNIQAVSKMGVTSILVG..KGVNLGALRQGLSEFVKNSASV
Corn          99  ....KIQRKTGIPYKSMLFFDDENRNIEAVSKMGVTSVLVE..NGVNLDMFKLGLSNFATNCAAS
Soybean       107 ....RIHSRTGVPFNSMLFFDDENRNIQAVSKMGVTSILVG..DGVNLGSLR
M. Truncatula 110 ....KIHSATGVPFSSMLFFDDENRNIQTVSKMGVTSILVD..NGVNLGALSQGLTQFSRNWNTS
S. Pombe      120 ....EIHNESGIDYREMVFFDDESRNRE.VERLGVTFCLVP..DGLNRASFEQGINKWRKLKKTH
S. Cervisiae  120 DGLKDLYNTSDLKSKKICLFDDESRNKE.VEKYGVKFVYVR..DPENGPSWKLYQDYLSGKV
Taurus        106 ....RLQRKTGVPFSQMIFFDDEKRNIVDVSKLGVTCIHVQ..HGMSLQTLTQGLDAFTKAQAGL
Goat          121 ....RLQRKAGVPFSQMIFFDDEKRNIVDVSKLGVTCIHVQ..HGMSLQTLTQGLDAFTKAQAGL
Ciona         117 ....KFHATTGIKYEDMIFFDDEYRNVSDVSEKGVTCIFVE..HGINWKEIKDGFNKFQRNQNST
Japan. Rice   109 ....RIQRTTGIPYESMLFFDDEHRNFATVSKMGVTSILVDWDGGVNLEMFKLGLNNFAAKFAAS
```
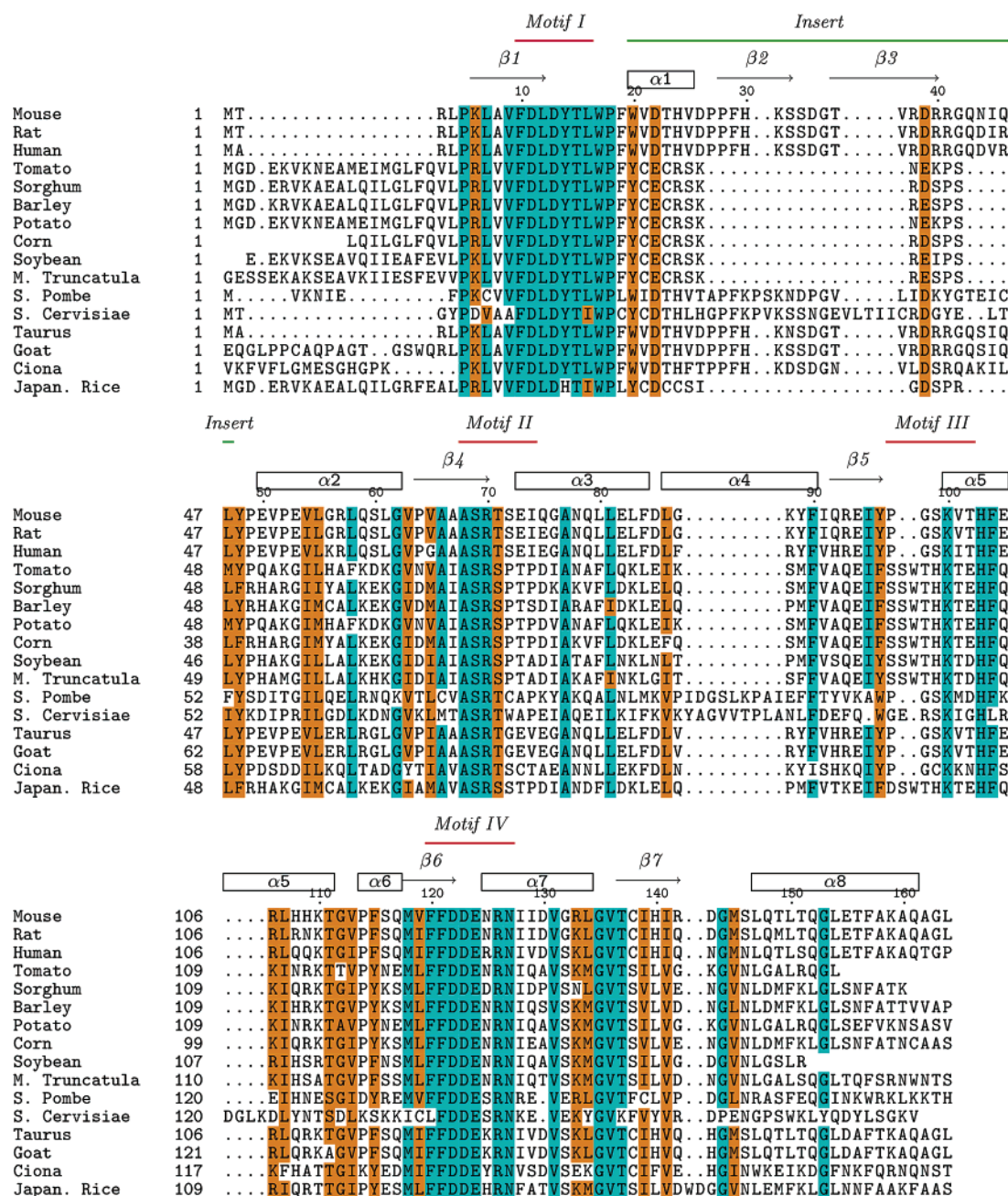
FIGURE 5: Sequence alignment of MDP-1. Loops (motifs) used to find and align superfamily members are denoted with red lines (note that one or two of the amino acids at the end of the motifs are located at the ends of α-helices or β-strands and not loops). The 28-amino acid insert is denoted with a green 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, AI781072_2 for tomato, BG933513_3 for sorghum, BF627188_3 for barley, BE339957_2 for potato, AW146994_3 for corn, BE020617_3 for soybean, BF647363_1 for *Medicago truncatula*, SPBP8B7.31 for *Schizosaccharomyces pombe*, YEW4_YEAST for *S. cerevisiae*, CB443279_2 for taurus, CD051815_1 for goat, AK114769_1 for *Ciona intestinalis*, and AAP03406.1 for *Oryza sativa* (japonica cultivar group).

nucleotide, a phosphorylated polysaccharide, or a phosphorylated protein. To examine these three possibilities, we have manually docked small sugar phosphates and phosphotyrosine 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 WO_4^{2-}. All of the ligands to the WO_4^{2-} ion from loops I–III 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.

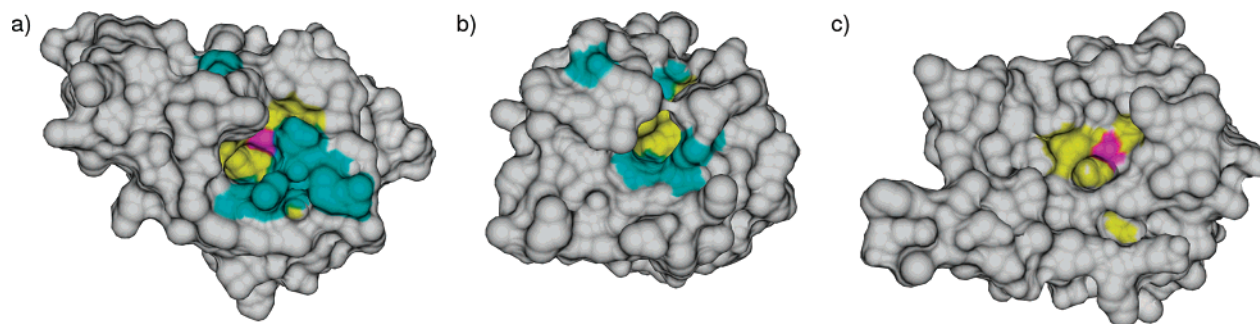Substrate specificity is expected to be derived from interactions with conserved residues at the surface of the

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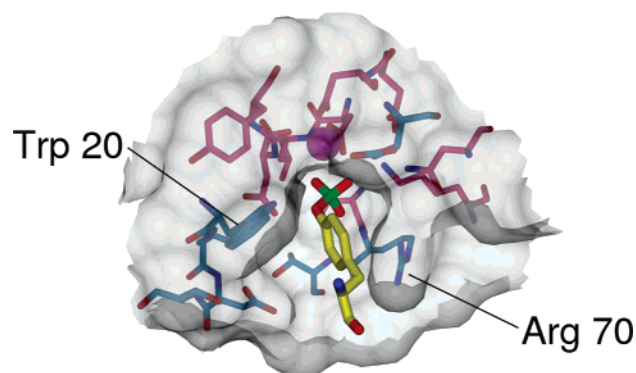
FIGURE 7: Model of phosphotyrosine docked into the active site of MDP-1. Residues that bind and interact directly with the phosphate moiety are colored pink. Residues surrounding the leaving group of the putative substrate are colored blue. Arg 70 and Trp 20 may reorient to interact with the leaving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

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.

SUPPORTING INFORMATION AVAILABLE

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

REFERENCES

1. 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.
2. 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.
3. Hisano, T., Hata, Y., Fujii, T., Liu, J. Q., Kurihara, T., Esaki, N., and 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.
4. Ridder, I. S., Rozeboom, H. J., Kalk, K. H., Janssen, D. B., and 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.
5. 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.
6. 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.
7. 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.
8. 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 phospho-

serine phosphatase: crystallographic "snapshots" of intermediate states, *J. Mol. Biol. 319*, 421−431.

9. 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.

10. 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.

11. Parsons, J. F., Lim, K., Tempczyk, 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.

12. Wu, J., and Woodard, R. W. (2003) *Escherichia coli* YrbI is 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase, *J. Biol. Chem. 278*, 18117−18123.

13. Selengut, J. D., and Levine, R. L. (2000) MDP-1: A novel eukaryotic magnesium-dependent phosphatase, *Biochemistry 39*, 8315−8324.

14. Barford, D., Das, A. K., and Egloff, M. P. (1998) The structure and mechanism of protein phosphatases: insights into catalysis and regulation, *Annu. Rev. Biophys. Biomol. Struct. 27*, 133−164.

15. Barford, D., Jia, Z., and Tonks, N. K. (1995) Protein tyrosine phosphatases take off, *Nat. Struct. Biol. 2*, 1043−1053.

16. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods Enzymol. 276*, 307−326.

17. de La Fortelle, E., and Bricogne, G. (1997) in *Methods in Enzymology* (Carter, C. W., Jr. and Sweet, R. M., Ed.) pp 472−494, Academic Press, New York.

18. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, G. J. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr. A47* (Part 2), 110−119.

19. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D54* (Part 5), 905−921.

20. Brünger, A. T., Adams, P. D., and Rice, L. M. (1997) New applications of simulated annealing in X-ray crystallography and solution NMR, *Structure 5*, 325−336.

21. Davis, I. W., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes, *Nucleic Acids Res. 32*, W615−W619.

22. Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement, *Acta Crystallogr. D56* (Part 12), 1622−1624.

23. O'Donovan, C., Martin, M. J., Gattiker, A., Gasteiger, E., Bairoch, A., and Apweiler, R. (2002) High-quality protein knowledge resource: SWISS-PROT and TrEMBL, *Briefings Bioinf. 3*, 275−284.

24. Boguski, M. S. (1995) The turning point in genome research, *Trends Biochem. Sci. 20*, 295−296.

25. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., Bucher, P., Copley, R. R., Courcelle, E., Das, U., Durbin, R., Falquet, L., Fleischmann, W., Griffiths-Jones, S., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lopez, R., Letunic, I., Lonsdale, D., Silventoinen, V., Orchard, S. E., Pagni, M., Peyruc, D., Ponting, C. P., Selengut, J. D., Servant, F., Sigrist, C. J., Vaughan, R., and Zdobnov, E. M. (2003) The InterPro Database, 2003 brings increased coverage and new features, *Nucleic Acids Res. 31*, 315−318.

26. Holm, L., and Sander, C. (1997) Dali/FSSP classification of three-dimensional protein folds, *Nucleic Acids Res. 25*, 231−234.

27. Argiriadi, M. A., Morisseau, C., Hammock, B. D., and Christianson, D. W. (1999) Detoxification of environmental mutagens and carcinogens: structure, mechanism, and evolution of liver epoxide hydrolase, *Proc. Natl. Acad. Sci. U.S.A. 96*, 10637−10642.

28. Newman, J. W., Morisseau, C., Harris, T. R., and Hammock, B. D. (2003) The soluble epoxide hydrolase encoded by EPXH2 is a bifunctional enzyme with novel lipid phosphate phosphatase activity, *Proc. Natl. Acad. Sci. U.S.A. 100*, 1558−1563.

29. Kim, Y., Yakunin, A. F., Kuznetsova, E., Xu, X., Pennycooke, M., Gu, J., Cheung, F., Proudfoot, M., Arrowsmith, C. H., Joachimiak, A., Edwards, A., and Christendat, D. (2004) Structure and function-based characterization of a new phosphoglycolate phosphatase from *Thermoplasma acidophilum*, *J. Biol. Chem. 279*, 517−526.

30. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2003) The Pentacovalent Phosphorus Intermediate of a Phosphoryl Transfer Reaction, *Science 108*, 2710.

31. 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.

32. Zhang, G., Mazurkie, A. S., Dunaway-Mariano, D., and Allen, K. N. (2002) Kinetic Evidence for a Substrate-Induced Fit in Phosphonoacetaldehyde Hydrolase Catalysis, *Biochemistry 41*, 13370−13377.

33. Grzyska, P. K., Czyryca, P. G., Golightly, J., Small, K., Larsen, P., Hoff, R. H., and Hengge, A. C. (2002) Generality of solvation effects on the hydrolysis rates of phosphate monoesters and their possible relevance to enzymatic catalysis, *J. Org. Chem. 67*, 1214−1220.

34. Wolfenden, R. (1983) Waterlogged molecules, *Science 222*, 1087−1093.

35. Warshel, A. (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites, *J. Biol. Chem. 273*, 27035−27038.

36. Andersen, J. N., Mortensen, O. H., Peters, G. H., Drake, P. G., Iversen, L. F., Olsen, O. H., Jansen, P. G., Andersen, H. S., Tonks, N. K., and Moller, N. P. (2001) Structural and evolutionary relationships among protein tyrosine phosphatase domains, *Mol. Cell. Biol. 21*, 7117−7136.

37. Zhang, M., Stauffacher, C. V., Lin, D., and Van Etten, R. L. (1998) Crystal structure of a human low molecular weight phosphotyrosyl phosphatase. Implications for substrate specificity, *J. Biol. Chem. 273*, 21714−21720.

38. Slack, D. N., Seternes, O. M., Gabrielsen, M., and Keyse, S. M. (2001) Distinct binding determinants for ERK2/p38α and JNK map kinases mediate catalytic activation and substrate selectivity of map kinase phosphatase-1, *J. Biol. Chem. 276*, 16491−16500.

39. Tootle, T. L., Silver, S. J., Davies, E. L., Newman, V., Latek, R. R., Mills, I. A., Selengut, J. D., Parlikar, B. E., and Rebay, I. (2003) The transcription factor Eyes absent is a protein tyrosine phosphatase, *Nature 426*, 299−302.

40. Allen, K. N., and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: evolution of a catalytic scaffold, *Trends Biochem. Sci. 29*, 495−503.

41. Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−950.

42. Sanner, M. F., Olson, A. J., and Spehner, J.-C. (1995) in *11th ACM Symposium on Computational Geometry*, pp C6−C7 ACM Press: New York.