MDP-1 Is a New and Distinct Member of the Haloacid Dehalogenase Family of Aspartate-Dependent Phosphohydrolases

Jeremy D. Selengut*

Laboratory of Biochemistry, National Heart, Lung and Blood Institute, Building 50-2347, National Institutes of Health, 50 South Drive, Bethesda, Maryland 20892-8012

Received July 12, 2001; Revised Manuscript Received August 30, 2001

ABSTRACT: MDP-1 is a eukaryotic magnesium-dependent acid phosphatase with little sequence homology to previously characterized phosphatases. The presence of a conserved motif (Asp-X-Asp-X-Thr) in the N terminus of MDP-1 suggested a relationship to the haloacid dehalogenase (HAD) superfamily, which contains a number of magnesium-dependent acid phosphatases. These phosphatases utilize an aspartate nucleophile and contain a number of conserved active-site residues and hydrophobic patches, which can be plausibly aligned with conserved residues in MDP-1. Seven site-specific point mutants of MDP-1 were produced by modifying the catalytic aspartate, serine, and lysine residues to asparagine or glutamate, alanine, and arginine, respectively. The activity of these mutants confirms the assignment of MDP-1 as a member of the HAD superfamily. Detailed comparison of the sequence of the 15 MDP-1 sequences from various organisms with other HAD superfamily sequences suggests that MDP-1 is not closely related to any particular member of the superfamily. The crystal structures of several HAD family enzymes identify a domain proximal to the active site responsible for important interactions with low molecular weight substrates. The absence of this domain or any other that might perform the same function in MDP-1 suggests an "open" active site capable of interactions with large substrates such as proteins. This suggestion was experimentally confirmed by demonstration that MDP-1 is competent to catalyze the dephosphorylation of tyrosine-phosphorylated proteins.

Magnesium-dependent phosphatase-1 (MDP-1)[1] was initially identified as a component of "purified" rabbit muscle carbonic anhydrase III (CAIII) preparations. CAIII had been assigned an intrinsic phosphatase activity (*1*), but recent work has established that this activity is due entirely to trace contamination with highly active phosphatases (*2*). Sufficient rabbit MDP-1 was purified to allow N-terminal sequencing, which led to the identification, cloning, and expression of the mouse form of the enzyme in *Escherichia coli*. The lack of sequence homology between MDP-1 and any other known phosphatases prompted an investigation into the catalytic mechanism of MDP-1. Several mechanistic types were ruled out, including cysteine nucleophile, histidine nucleophile, and (non-magnesium) metal-coordinated water (hydroxide) nucleophile. Catalysis via an aspartate nucleophile was not ruled out (*3*).

* Author to whom correspondence should be addressed [telephone (301) 496-8708; fax (301) 496-0599; e-mail selengut@nih.gov].

[1] Abbreviations: MDP-1, magnesium-dependent phosphatase-1; CAIII, carbonic anhydrase isozyme III; HAD, haloacid dehalogenase; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; EDTA, ethylenediaminetetraacetic acid; EGTA, ethylene glycol bis(β-aminoethyl ether) *N*,*N*,*N*′,*N*′-tetraacetic acid; IPTG, isopropyl β-D-thiogalactopyranoside; HPLC, high-pressure liquid chromatography; PCR, Polymerase Chain Reaction; CBP, calmodulin-binding peptide; LB, Luria-Bertani; BCA, bicinchoninic acid; MES, 2-(*N*-morpholino)ethanesulfonic acid; *p*NPP, *p*-nitrophenyl phosphate; DMEM, Dulbecco's modification of Eagle's medium; HEPES, *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid; AEBSF, 4-(2-aminoethyl)benzenesulfonyl fluoride; PBS, phosphate-buffered saline; PTP-1B, protein tyrosine phosphatase 1B.

The haloacid dehalogenase (HAD) superfamily includes dehalogenases, phosphonatases, phosphomutases, phosphatases, and ATPases, all of which utilize a nucleophilic aspartate. The overall homology among these enzymes, however, is small (sequence identity <15%, *4*) and is centered on three short catalytic motifs. The sizes of these motifs are generally insufficient to allow homology search programs such as BLAST to relate functionally distinct members of the superfamily to each other with a reasonable level of confidence. Crystal structures of five members of this superfamily have been published: 2-haloacid dehalogenase (*5*, *6*), soluble epoxide hydrolase (*7*), phosphonoacetaldehyde hydrolase (a phosphonatase) (*4*), the Ca^{2+}-pumping "P-type" ATPase of sarcoplasmic reticulum (*8*), and most recently phosphoserine phosphatase (*9*). Comparison among these structures demonstrates a conserved fold and strongly suggests divergent evolution from a common ancestor. Furthermore, several structures have been published of the bacterial response regulator, CheY, which exhibits a similar fold related by a circular permutation and is phosphorylated on an aspartate correlating to the HAD superfamily nucleophile (*10*−*13*).

The large family of magnesium-dependent acid phosphatases and phosphomutases within the haloacid dehalogenase superfamily is characterized by an amino-terminal motif, DXDX(T/V), in which the initial aspartate is the nucleophile (*14*). This motif is found in MDP-1 (mouse numbering: Asp10−Thr14). A C-terminal motif, GDXXXD,

10.1021/bi011405e
Published on Web 09/27/2001

is also found with a somewhat lower degree of conservation (*15*). This sequence is not found in MDP-1. A more recent paper studying the broader HAD superfamily defines three motifs: hhhhDXXG(T/V)h, hhhh(T/S), and K(X_n)hhhh-GDXXXD, where h is a hydrophobic residue (A, C, F, G, I, L, M, V, W, or Y) (*16*). In the consensus MDP-1 sequence, three reasonable matches can be made in the proper order: hhhFDXXYTL, hhhAS, and K(X_{17})hhhFDDXXXN (mismatches are underscored). These aligned motifs are insufficient for a definitive assignment, however.

Site-specific mutagenesis studies have been carried out on phosphoserine phosphatase (*17*), Ca^{2+}-ATPase (*18*−*20*), and haloacid dehalogenase (*21*), all members of the HAD superfamily. A clear pattern in activity measurements has emerged when various residues of the conserved motifs are mutated (*17*).

Below, we use further sequence alignments and site-directed mutagenesis to establish that MDP-1 is a member of the HAD superfamily.

MATERIALS AND METHODS

Materials. Materials were purchased from the following sources: gels and buffers for SDS−PAGE from Novex (San Diego, CA); LB broth from Digene Diagnostics (Beltsville, MD); leupeptin from Bachem (King of Prussia, PA); IPTG from ICN Biomedicals (Aurora, OH); BCA protein assay reagents, bovine serum albumin standard, Tween-20, and trifluoroacetic acid from Pierce (Rockford, IL); HPLC grade acetonitrile from J. T. Baker (Philipsburg, NJ); and plasmid mini-prep and PCR purification kits from Qiagen (Valencia, CA).

Database of HAD Superfamily Sequences. We have recently constructed a greatly expanded database of HAD superfamily enzyme sequences (manuscript in preparation). In brief, the database was constructed via an iterative homology search process as follows. Sequences of known HAD superfamily enzymes were identified from literature references. The conserved motifs, hhhhDhDX(T/V), hhhh-(S/T), and hhhh(D/E/G/S)DX_{2-6}(D/E/G/N)h (where h stands for the hydrophobic residue set ACFGKLMRTVWY), were identified and manually aligned. The sequences were then truncated at the N and C termini to include only the sequence bound by the first and third motifs. One representative truncated sequence of each enzyme was used as a query for a BLAST search (www.ncbi.nlm.nih.gov/blast/blast.cgi) versus the nonredundant protein database. Results from this search were downloaded, scanned for the presence of the three target motifs, and, if any of these were evident, added to the database and aligned. Sequences diverging significantly from those already present in the database (BLAST E-value $>10^{-15}$) were used as a query in a new iteration of a BLAST search. Sequences showing high mutual homology were clustered and aligned using a multiple alignment program (CLUSTALW). If all three motifs (or close approximations thereof) were not evident as highly conserved amino acids in these clusters, the sequences were culled from the database. This process was continued until no new sequences were obtained by BLAST. At this point, sequences that were not members of a cluster or were members of a cluster containing sequences representing only a single genus were removed to a secondary database of "potential" HAD sequences. These sequences were used as seeds for searches versus various genomic databases. If matches were found such that a cluster spanning more than one genus was created, this cluster was returned to the main database.

Site-Specific Mutants. Site-specific mutants were prepared by the QuickChange mutagenesis kit (Stratagene, no. 200518) as described previously (*3*). All mutants were confirmed by electrospray mass spectroscopy of the purified proteins.

Mass Analysis. Electrospray mass spectroscopy was performed with an Agilent model G1946A instrument interfaced to a model 1100 high-pressure liquid chromatography system equipped with a Vydac narrow-bore C18 column (218TP5205, Vydac, Hesperia, CA). The initial solvent was 0.05% trifluoroacetic acid and 10% acetonitrile, and gradient elution was effected with 0.05% trifluoroacetic acid/acetonitrile at 1%/min at a flow rate of 0.2 mL/min. The column effluent was mixed in a tee with neat acetic acid delivered by another 1100 series pump (0.1 mL/min), and the mixture was introduced into the mass spectrometer (*22*).

Cloning of Saccharomyces cerevisiae MDP-1. S. cerevisiae genomic DNA was obtained from Research Genetics (no. 40802, Huntsville, AL). A ligation-independent cloning strategy was employed to express yeast MDP-1 (YER134C) as an N-terminal fusion with a calmodulin-binding peptide containing a FLAG epitope tag (kit 214405 revision 049001, Stratagene, La Jolla, CA). Primers complementary to the kit-supplied linearized cloning vector overhangs (pCAL-n-FLAG) as well as the ends of the YER134C open reading frame (omitting the ATG start codon) were obtained from Biosynthesis, Inc. (Lewiston, TX): (sense) 5′-GACGAC-GACAAGACTGGATACCCTGACGTGGC-3′; (antisense) 5′-GGAACAAGACCCGTTCAAACCTTTCCACTCAGG-TAATC-3′. PCR was carried out as follows. One hundred nanograms of yeast genomic DNA, 250 ng of sense primer, and 250 ng of antisense primer in a total of 15 μL of TE buffer were mixed with 4 μL of 10× Thermopol reaction buffer (no. 007-TDP, New England Biolabs, Beverly, MA), 0.7 μL of $Vent_R$ DNA polymerase (no. 254S, New England Biolabs), 0.7 μL of 5 mM dNTPs (an equimolar mixture of dATP, dCTP, dGTP, and dTTP), and 20.4 μL of deionized water (final volume = 50 μL). The temperature program consisted of an initial cycle of 94 °C for 2.5 min, 65 °C for 3 min, and 72 °C for 5 min, followed by 10 cycles of 94 °C for 1 min and 72 °C for 1.5 min, followed by 25 cycles of 94 °C for 0.5 min and 72 °C for 1.5 min, and finally a single step of 75 °C for 10 min. Agarose gel electrophoresis indicated a single product of the expected size (557 base pairs). Further processing of the clone to the point of transformation of BL21(DE3)-pLysS cells was carried out in a manner identical to that for the mouse MDP-1 protein (*3*).

Purification of MDP-1. BL21(DE3)-pLysS cells expressing wild-type and mutant recombinant mouse CBP-MDP-1 or yeast CBP-FLAG-MDP-1 were grown in LB medium supplemented with 50 μg/mL ampicillin and 34 μg/mL chloramphenicol to OD(600) = 0.6. IPTG was added to 1 mM, and the cells were allowed to grow an additional 3 h at 37 °C. Cells were centrifuged at 5000 rpm, and the supernatant was discarded. Cells were resuspended in an equal volume of cell lysis buffer (50 mM Tris, pH 8.0 at 25 °C, 150 mM NaCl, 0.1% Triton X-100), and aliquots were frozen in liquid nitrogen and stored at −20 °C until needed.

Cells were thawed at room temperature, and solid streptomycin sulfate was added to make a 1% solution. This was mixed thoroughly and allowed to stand on ice for 15 min. Cell debris and DNA were removed by centrifugation at 15000*g* for 5 min. The clarified cell lysate was diluted with 9 parts of calmodulin beads binding buffer [50 mM Tris, pH 8.0 at 25 °C, 150 mM NaCl, 0.01% Tween-20, 2 mM $Mg(OAc)_2$, 1 mM imidazole, 3 mM $CaCl_2$, 1 mM EDTA, 10 mM β-mercaptoethanol, 1 μg/mL leupeptin, 5 μg/mL aprotinin], and then 2 parts of a 1:1 slurry of calmodulin affinity beads (Stratagene) in binding buffer was added. The mixture was rocked at room temperature for 2−3 h. The beads were pelletted with a brief centrifugation at 5000 rpm, and the supernatant was decanted. The beads were washed three times with 5 volumes of wash buffer [50 mM Tris, pH 8.0 at 25 °C, 300 mM NaCl, 0.01% Tween-20, 2 mM $Mg(OAc)_2$, 1 mM imidazole, 3 mM $CaCl_2$, 2 mM EDTA] and eluted with three portions of 2 volumes elution buffer (50 mM Tris, pH 8.0 at 25 °C, 300 mM NaCl, 0.01% Tween-20, 2 mM EGTA).

The pooled eluted fraction was brought to 30% saturation with ammonium sulfate (on ice) and centrifuged at 20000*g* for 10 min. The supernatant was decanted and brought to 60% saturation with ammonium sulfate. The suspension was centrifuged again and the supernatant discarded. The protein was redissolved in thrombin cleavage buffer (50 mM Tris, pH 8.0 at 25 °C, 150 mM NaCl, 0.01% Tween-20, 2.5 mM $CaCl_2$) and centrifuged at 15000*g* for 5 min. The supernatant was removed to a fresh tube, and the protein concentration was determined. Five units of biotinylated thrombin (Novagen) was added for each milligram of protein and incubated at room temperature overnight. Thrombin was removed by binding to streptavidin−agarose beads (16 μL of bead slurry/unit thrombin, Novagen). These beads were removed by centrifugation, and the cleaved protein was dialyzed versus MDP-1 storage buffer (50 mM Tris, pH 8.0 at 25 °C, 50 mM NaCl, 0.01% Tween-20) using dialysis tubing with a 10000 Da molecular weight cutoff in order to remove the cleaved fusion tag (MW = 4500 Da) as well as to change the buffer. Samples were centrifuged, and the supernatant was distributed into small aliquots, which were frozen in liquid nitrogen and stored at −70 °C until use.

Protein Assay. Protein measurements were made either with the BCA assay (Pierce, 23223/4) or by observation of absorbance at 280 nm (mouse MDP-1, $\epsilon_M = 1.65 \times 10^4$; yeast MDP-1, $\epsilon_M = 3.78 \times 10^4$).

Phosphatase Assay. Assay solutions consisted of 200 mM MES, pH 5.5, 10 mM magnesium chloride, 2 mM EDTA, and 5 mM *p*NPP as substrate. The reactions were carried out in 96-well microtiter plates at 37 °C in a total volume of 40 μL. The reactions were quenched by the addition of 20 μL of 10 M NaOH. PNP anion was detected at 405 nm ($\epsilon_M = 1.85 \times 10^4$). Determination of V_{max} and K_M for yeast MDP-1 was carried out in the same manner as was previously reported for the mouse enzyme (*3*).

Protein−Tyrosine Phosphatase Blot-Overlay Assay. Peroxide-stimulated HeLa cell lysate was prepared as follows. HeLa cells grown in DMEM containing 10% fetal bovine serum were grown to 80% confluence in six-well culture dishes and then starved of serum in DMEM overnight. The culture medium was then removed and replaced with PBS buffer containing 5 mM hydrogen peroxide. After 10 min, the cells were frozen in liquid nitrogen and stored at −70 °C. The cells were thawed and lysed by the addition of 0.5 mL of lysis buffer [20 mM HEPES, pH 7.2, 5% glycerol, 1% Triton X-100, 1 mM sodium orthovanadate, 5 mM sodium fluoride also including 1 tablet of Boeringer Mannheim "complete" protease cocktail without EDTA (contains AEBSF, aprotinin and leupeptin) per 50 mL of solution], on ice for 20 min. The lysate was collected and sonicated two times for 30 s in a bath-type sonicator on ice and then centrifuged for 20 min at 4 °C at maximum speed in a tabletop centrifuge. The pellet was discarded, and the supernatant was mixed with $^1/_3$ volume of 4× NuPage SDS sample buffer and $^1/_{30}$ volume of β-mercaptoethanol.

Samples of lysate were run out on 4−12% NuPAGE SDS−PAGE electrophoresis gels (Novex, San Diego, CA) and then transferred to nitrocellulose. The nitrocellulose was washed twice with water and blocked using 3% nonfat dry milk in PBS buffer (PBS-MLK) for 20 min at room temperature with constant agitation. After blocking, the nitrocellulose was cut into strips with one lysate sample per strip. Each strip was separately incubated in phosphatase assay buffer (as above) lacking substrate with or without magnesium or added MDP-1 or PTP-1B as indicated in Figure 4 and additionally containing 0.5% nonfat dry milk (recombinant human PTP-1B was prepared by Z. Y. Zhang, Albert Einstein College of Medicine, Bronx, NY). The nitrocellulose strips were washed twice with PBS-MLK and then incubated with 0.25 μg/mL anti-phosphotyrosine antibody 4G10 (no. 05-321, Upstate Biotechnology, Lake Placid, NY) in PBS-MLK overnight at 4 °C with constant agitation. After the strips had been washed twice with water, they were incubated with alkaline phosphatase-linked goat anti-mouse antibody (AC32ML, Tropix, Bedford, MA) diluted 1:2000 in PBS-MLK for 1.5 h at room temperature. The strips were washed twice with water, once with PBS containing 0.05% Tween-20, and four times with water. The strips were finally equilibrated with chemiluminescence assay buffer (100 mM diethanolamine, pH 10.0, containing 1 mM $MgCl_2$) and incubated with a solution of $^1/_{10}$ CDP-Star (MSC-020, Tropix) and $^1/_{50}$ Nitro-Block II (LNX-200, Tropix) in assay buffer for 5 min. Bands were visualized by exposure of X-Omat AR film (Kodak, Rochester, NY) to the strips for 10 min followed by development of the film.

RESULTS AND DISCUSSION

Expansion of the MDP-1 Sequence Family. The initial report on MDP-1 (*3*) listed one partial and six full-length MDP-1 sequences gleaned from database searches (rat, human, tomato, *Arabidopsis thaliana*, *Sorghum bicolor*, *Schizosaccharomyces pombe*, and *Saccharomyces cerevisiae*) in addition to the initially isolated rabbit sequence fragments and the full-length mouse sequence, which was confirmed by expression and mass analysis. Since that time several additional partial and full-length sequences have been identified (Figure 1).

Among these sequences, the *S. cerevisiae* MDP-1 (yMDP-1) is the most divergent from the mouse sequence (29% identical, 48% conserved). To strengthen the claim that all of these sequences are, like mouse MDP-1 (mMDP-1), magnesium-dependent acid phosphatases, we have cloned and expressed yMDP-1 in a manner similar to that used to

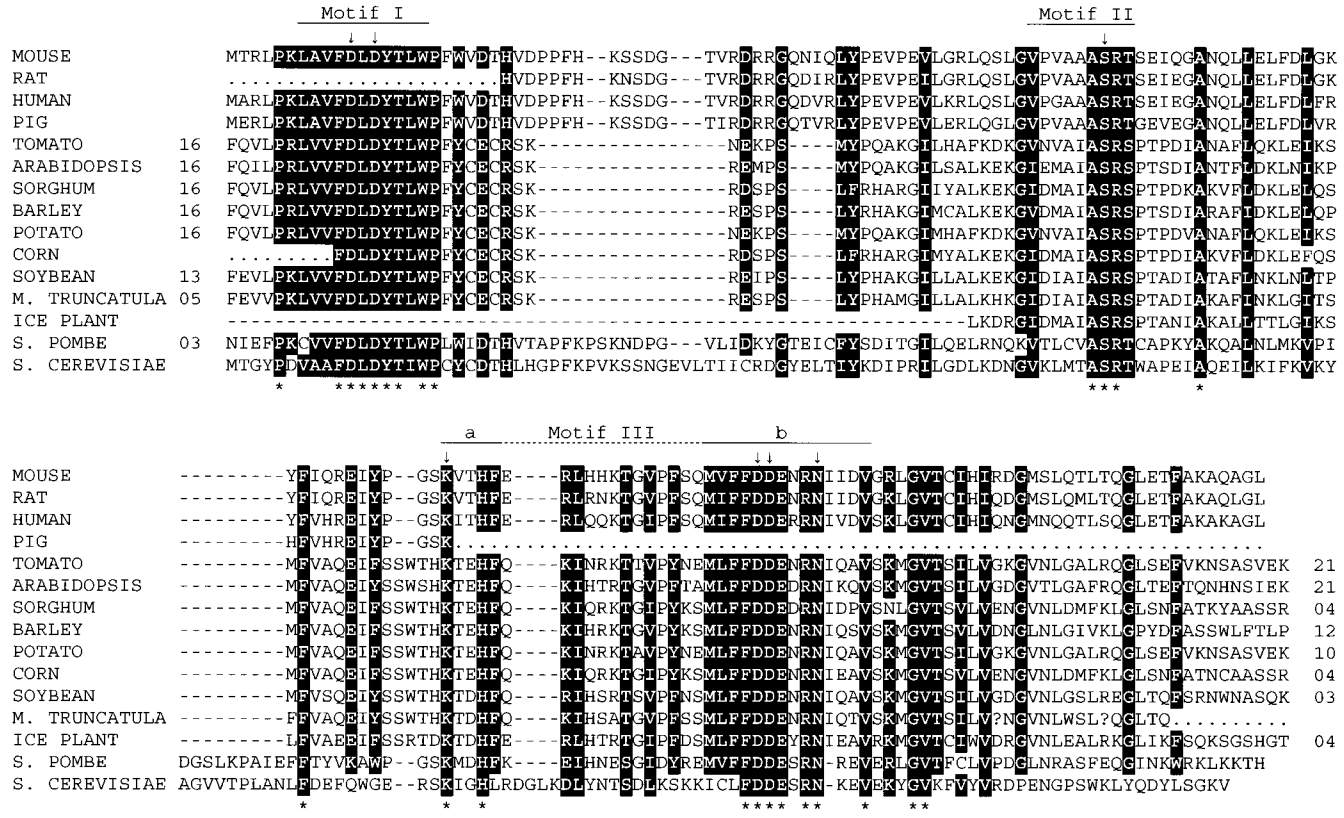

FIGURE 1: Alignment of MDP-1 sequences from various species. Numbers before and after the sequences indicate the number of known amino acids not displayed. Highly conserved residues are highlighted in black (90% conservation in the following groupings: small/hydrophobic, ACGPTV; charged, DEHKR; aromatic, FHWY; aliphatic, LIV; tiny, AGS; alcohol, ST). Invariant residues are marked (*) at the bottom. Positions of point mutations described in the text are marked (↓) at the top. Sequence sources: mouse, AAK00763; rat, AI408718 + BF559389; human, AI948475, chromosome 14; pig, AW425033; *Arabidopsis thaliana* (thale cress), AAD25851; *Sorghum bicolor* (milo), BG933513; barley, BF627188; potato, BE339957; corn, AW146994; soybean, BE020617; *Medicago truncatula* (barrel medic), BF647363; *Mesembryanthemum crystallinum* (ice plant), BE130774; *Schizosccharomyces pombe*, CAA21816; *Saccharomyces cerevisiae*, AAC03232.

prepare mMDP-1 (see Materials and Methods). Indeed, yMDP-1 exhibits kinetic parameters very similar to those of mMDP-1 with pNPP as a substrate ($V_{max} = 3\ s^{-1}$ and $K_M = 2$ mM versus 3.8 s^{-1} and 1.7 mM, respectively, for mMDP-1).

Relationship of MDP-1 to the HAD Superfamily by Sequence Comparison. The database of HAD superfamily sequences we have compiled (manuscript in preparation) excludes ATPases and contains at present 719 sequences. Among them are 269 sequences coding for, or highly similar to, known phosphatases, phosphonatases, and phosphosugar mutases (see Supporting Information). These have been separated into 15 groups: phosphoserine phosphatases, trehalose-6-phosphate phosphatases, sucrose-phosphate phosphatases, phosphomannomutases, β-phosphoglucomutases, 2-deoxyglucose-6-phoshatases and glycerol-2-phosphatases, phosphoglycolate phosphatases, bacterial class B nonspecific acid phosphatases, bacterial class C nonspecific acid phosphatases, plant class C-like nonspecific acid phosphatases, *p*-nitrophenyl phosphatases, histidinol phosphatases, enolase-phosphatases, phosphonoacetaldehyde hydrolases (phosphonatases), and DNA-5′-kinase-3′-phosphatases (Table 1). A consensus sequence was determined for each of these groups, a group of 20 representative P-type ATPases (*23*, because ATPases were excluded from our database), and the group of MDP-1 sequences. An overall consensus at the 70% level was then prepared manually by combining the 16 consensus sequences (excluding MDP-1, Table 1). This was preferable to creating a consensus from the 284 sequences directly because of the uneven representation of the various groups. The regions of high homology among these groups correspond directly to the three motifs identified by previous researchers (*16*). The agreement between each of the 16 groups and the overall consensus ranges from 75 to 95%, while the MDP-1 consensus has a 78% agreement. This level of agreement is consistent with the assignment of MDP-1 as a member of the HAD superfamily of phosphatases but is not conclusive.

To determine whether MDP-1 is properly placed within any of the 16 groups defined above, each of the groups was compared with the MDP-1 consensus individually. This does not identify any group with a particularly close relationship; on the contrary, MDP has more agreement with the overall consensus than with any of the groups which make it up.

Site-Directed Mutagenesis. The putative critical catalytic residues of the HAD superfamily hydrolases have been experimentally examined by site-directed mutagenesis studies and crystal structure determinations. Table 2 shows the results of site-directed mutagenesis experiments on mouse MDP-1 in comparison with similar experiments performed on phosphoserine phosphatase (PSPase) and the calcium-pumping P-type ATPase (Ca-ATPase; *17*−*20*).

Consistent with its assignment as the nucleophilic aspartate, conservative mutations of Asp-10 (D10N, D10E) completely eliminate enzymatic activity. The second aspartate of motif I is believed to have a structural rather than a

Table 1: Alignment of Core Motifs of Phosphatase and Phosphomutase Members of the HAD Superfamily

Name[a]	Motif[b] I	II	IIIa	IIIb	% match to Cons. / MDP-1
T6PP	zhhhDhDGTLs	hshAlloGR	KGcAl	zlalGD....DhTDEsAF	95/60
ATPase[c]	hlzXDKTGTLT	lpzhhlTGD	KXXll	VzhhGD....GhNDzzzL	92/49
PMM	lzLFDVDGTLT	IGhVGGSDh	KpYCL	IHFFGDcTXXGGNDaEIa	92/52
PSPase	hhzhDhDfTlh	hpzhlhSGG	KschX	slhlGD....GAfDlXhh	92/68
HisB	hLFIDRDGTLI	achVhloNQ	KPphh	ShVIGD....RXTDhpLA	92/51
C-NSAP	AllhDlDEThL	zplFYloNR	KXzRh	VLhhGD....NLsDFssh	92/54
pNPPase	hhlhDhDGzlh	hXhhhloNN	KPphh	hhhlGDp...hXoDlXhz	92/70
DOG6P	zhLFDhDGTll	hWAlVTSzf	KPcPc	zVVhED....APAGlhAG	92/55
PGP	hlhFDhDGTLh	hhhzlhofK	KPpzX	zlhlGD....fXXDlXzz	92/69
SPP	hlVSDLDpThV	SLLVFsTGR	KGpAL	TLzCGD....SGNDzELF	90/50
B-NSAP	AVsFDIDDTVL	DplaFlTGR	KzXhl	pIaYGD....SDsDIhAA	90/62
DEM-1	lhsFDhDsTlI	aXlVIFoNQ	KPXzG	shazGD....AAGRhXXX	88/57
β-PGM	zzlFDLDGVlT	hhhzlzSsS	KPsPc	zhzhED....sXAGlXAh	88/63
P-NSAP	zWlFDlDcThL	hchhhlosR	KSpXR	hGNXGD....QWsDLhXX	82/56
PhnX	AllhDWAGThV	lKIGSfoGY	RPhPh	hVKVsD....zhssIpEG	80/48
E-1	slLhDIEGTzs	hXlalYSSG	KpEsX	ILFLoD....XhXEhsAA	75/47
Cons.[d]	hhhhDhDsThh	hXhhhhosp	KsXhh	hhhhGD....shschXhh	
MDP-1	hzzFDLDYTLW	lshAhASRo	KhpHF	MlFFDD....EpRNhXXV	78/-

[a] Enzyme groups are listed in order of similarity to the overall consensus: T6PP, trehalose-6-phosphate phosphatase; ATPase, P-type ATPase; PMM, phosphomannomutase; PSPase, phosphoserine phosphatase; HisB, histidinol phosphatase/imidazol-glycerolphosphate dehydrogenase; C-NSAP, class C bacterial nonspecific acid phosphatase; pNPPase, *p*-nitrophenyl phosphate phosphatase; DOG6P/GPP, 2-deoxyglucose-6-phosphate phosphatase and glycerolphosphate phosphatases; PGP, phosphoglycolate phosphatase; SPP, sucrose-6^F-phosphate phosphatase; B-NSAP, class B bacterial nonspecific acid phosphatase; DEM-1, DNA-5′-kinase-3′-phosphatase; *β*-PGM, *β*-phosphoglucomutase; P-NSAP, plant class C-like nonspecific acid phosphatase; PhnX, phosphonoacetaldehyde hydrolase (phosphonatase); E-1, 2,3-diketo-5-methylthio-1-phosphopentane phosphatase (enolase-phosphatase). A listing of the sequences comprising each group can be found in the Supporting Information. [b] Consenses were calculated using the program CONSENSUS (http://www.bork.embl-heidelberg.de/cgi/consensus) at the 80% level using the following groupings: a = aromatic (FHWY), c = charged (DEHKR), f = small polar (CDNST), h = hydrophobic (ACFGHIKLMPRTVWY), l = aliphatic (ILV), o = hydroxyllic (ST), p = polar (CDEHKNQRST), s = small (ACDGNPSTV), z = small hydrophobic (ACGPTV), X = any. [c] Compiled from an alignment of 20 representative P-type ATPases (Axelsen and Palmgren, *J. Mol. Evol.* **1998**, *46*, 84, Figure 3, reference *23*). [d] Consensus of the 16 consensus sequences was determined by finding the smallest grouping which accommodates 70% of the groups at that position in the consenses.

catalytic role. The crystal structure of the *Bacillus cereus* phosphonatase, which has an alanine at this position, shows the main-chain carbonyl of this residue to coordinate the magnesium ion (*4*). Despite this, conservative mutations in PSPase (D22N), Ca-ATPase (T353S), and HAD (Y12F), which does not coordinate magnesium, all show significant reductions in activity. Mutation of this residue in MDP-1 (D12N) results in a 92% loss of activity.

The hydroxyllic residue in motif II (Ser-68 of mMDP-1) provides a hydrogen bond to the phosphate group of the substrate and may also be involved in deprotonation of water to hydrolyze the phosphorylated enzyme intermediate (*4*, *10*). The lysine residue in motif IIIa (Lys-99 of mMDP-1) provides hydrogen bonds to both the phosphate and the nucleophile (*4*, *10*). Conservative mutations of these residues (S68A, K99R) each reduce the enzymatic activity to nearly undetectable levels. Corresponding mutations on PSPase have similar effects.

Table 2: Site-Directed Mutagenesis of MDP-1 in Comparison with Other HAD Superfamily Phosphatases[a]

MDP-1[b]	% WT	PSPase[c]	% WT	Ca^{2+} ATPase[d]	% WT
motif I					
D10E	0	D20E	0	D351E	0
D10N	0	D20N	0	D351N	0
D12N	8	D22N	0	T353A	0
		D22E	50	T353S	20
motif II					
S68A	0.1	S109A	6	T625A	LE[e]
motif IIIa					
K99R	1.4	K158R	1		
motif IIIb					
D121N	0.4	G178A	74	G702A	20
D122N	0.1	D179N	0.6	D703N	<5
N126D	52	D183N	<0.4	D707N	<5

[a] Residues in equivalent positions are in the same rows. [b] This paper, see Materials and Methods for assay conditions. [c] Reference *16*. [d] References *17*−*19*. [e] LE, low expression.

As seen in the consensus sequence in Table 1, motif IIIb contains one invariant aspartate residue. In mMDP-1, conservative mutation of this residue (D122N) reduces the enzymatic activity to nearly undetectable levels, in agreement with similar effects on PSPase and Ca-ATPase. In addition, all but one of the enzymes included in Table 1 also include a second conserved aspartate or glutamate residue in motif IIIb (the exception is DEM-1, a DNA-5′-kinase/3′-phosphatase, which contains an arginine in place of the second aspartate, perhaps having something to do with the binding of DNA). Most commonly, this residue is found four amino acids downstream (+4) of the invariant aspartate but occasionally occurs in the position immediately upstream (−1, as is found in deoxyglucose-6-phosphatase, *β*-phosphoglucomutase, and four of the six phosphonatase sequences). The sequence of MDP-1 appears to conform to this latter variation of the motif, and conservative mutagenesis of the −1 aspartate (D121N) has a profound effect on enzyme activity similar to that found in the "+4" mutants of PSPase and Ca-ATPase. Mutation of the −1 residue (glycine to alanine) in PSPase and Ca-ATPase has only a moderate effect on activity, which corresponds with the effect of mutation of the +4 residue (N126D) in MDP-1. The crystal structure of the *B. cereus* phosphonatase indicates that the invariant aspartate residue and the +4 aspartate are involved in the binding of the essential magnesium ion cofactor (*4*). This structure also indicates that the −1 and +4 positions are very close in space, being at the ends of a short loop, such that placement of the aspartate in either of these positions would allow the carboxylate oxygen atoms to interact with the magnesium ion from the same angle.

Comparison of the Domain Structure of the HAD Superfamily Enzymes and MDP-1. Figure 2 shows a graphical alignment of the catalytic core region (bound by motifs I and IIIb) of MDP-1 and all of the HAD superfamily enzymes for which the biological substrate is known. An examination of the lengths of the regions between motifs I and II and between motifs II and III shows that there are three structural subgroups within the HAD superfamily. One group, consisting of phosphoserine phosphatase, *β*-phosphoglucomutase, 2-deoxyglucose-6-phosphatase, phosphoglycolate phosphatase, P-type ATPase, phosphonatase, and haloacid dehalogenase, has a large domain between motifs I and II. The

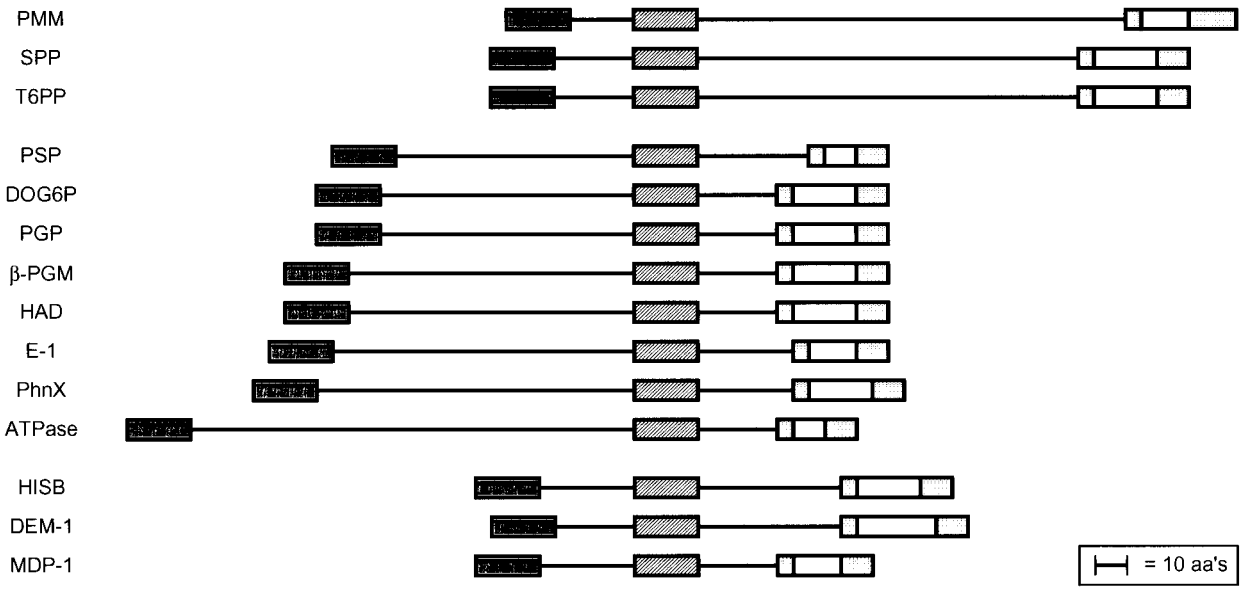

FIGURE 2: Graphical representation of the sequence regions bounded by the conserved HAD core motifs. In addition to MDP-1, only those HAD superfamily enzymes for which biological substrates have been identified are included. Motif I is represented by a gray box, motif II by a hatched box, and motifs IIIa and IIIb by stippled boxes separated by an open box. Regions between the core motifs are drawn as horizontal lines and represent the minimum size found for each enzyme. PMM, phosphomannomutase, human, AAC51368; SPP, sucrose-6^F-phosphatase, *Synechocystis* PCC6803, AAG31136; T6PP, trehalose-6-phosphatase, *Arabidopsis thaliana*, AAF87852; PSP, phosphoserine phosphatase, *Escherichia coli*, AAC77341; DOG6P, deoxyglucose-6-phosphatase, *E. coli*, BAA16129; PGP, phosphoglycolate phosphatase, *Pseudomonas aeruginosa*, AAG03455; β-PGM, β-phosphoglucomutase, *Neisseria meningitidis*, CAB85313; HAD, haloacid dehalogenase, *Pseudomonas putida*, AAB32245; E-1, 2,3-diketo-5-methylthio-1-phosphopentane phosphatase (enolase-phosphatase), *Schizosaccharomyces pombe*, CAB90135; PhnX, phosphonoacetaldehyde hydrolase (phosphonatase), *P. aeruginosa*, AAC45742; ATPase, P-type ATPase, *Synechocystis* PCC6803, D64005; HisB, histidinol phosphatase/imidazole-glycerol phosphate phosphatase, *E. coli*, CAA27151; DEM-1, DNA 5′-kinase-3′-phosphatase, human, AAD47379.

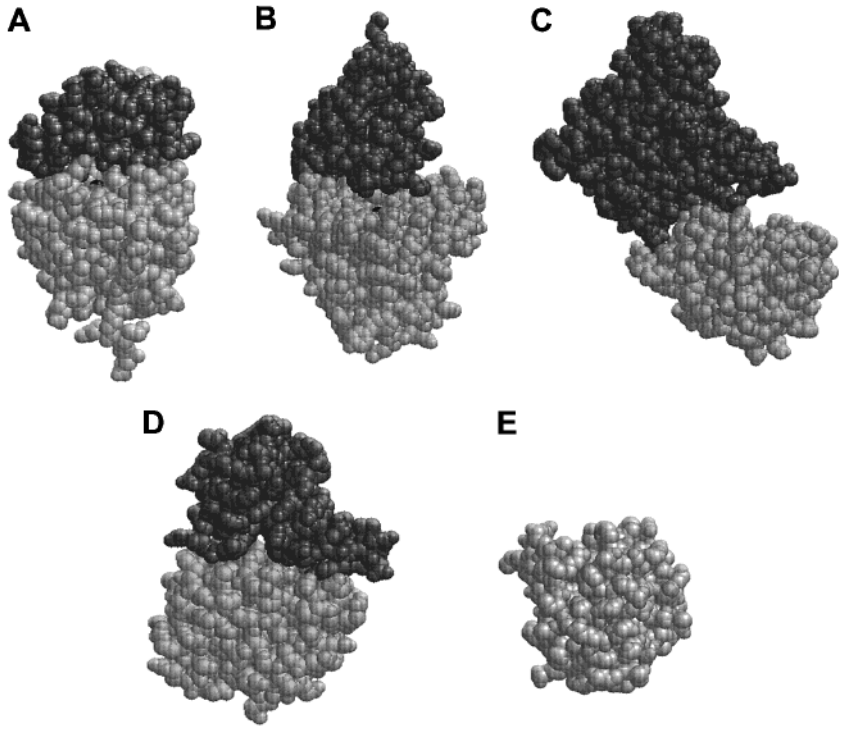

FIGURE 3: Crystal structures of representative HAD superfamily and HAD-related proteins: (A) phosphoserine phosphatase (*23*); (B) phosphonoacetaldehyde hydrolase (*4*); (C) Ca^{2+}-ATPase (M and A domains not shown, Ca^{2+}-bound, "open" state, *8*); (D) haloacid dehalogenase (*5*); (E) CheY (*26*). In each structure the core HAD catalytic domain is shown in light gray, the capping domains are shown in dark gray. Where present, the essential magnesium ion is shown in black. The structures are shown in roughly the same orientation, with the nucleophilic aspartate at the top center of the catalytic domain and the capping domain emanating from the top of the catalytic domain.

crystal structures of the latter three enzymes show that this domain serves as a "cap" positioned proximal to the active site (Figure 3). In each of these enzymes, this domain is believed to be critical for substrate binding and positioning, and, in the case of phosphonatase, catalysis (*4*−*6*, *8*). In the case of the P-type ATPase, this domain is believed to be mobile, binding ATP in the "open," calcium-bound state (Figure 3C), and presenting it to the phosphatase core domain

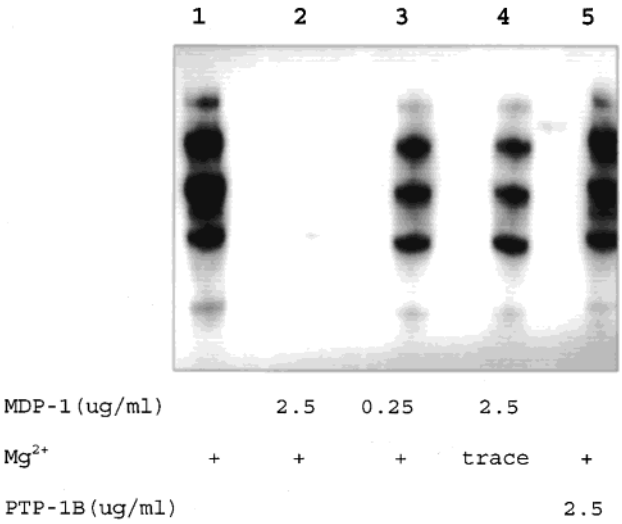

FIGURE 4: Phosphatase overlay−anti-phosphotyrosine western blot. Peroxide-stimulated HeLa cell lysate (see methods) was separated by SDS−PAGE and blotted to a nitrocellulose membrane and cut into strips. After blocking, the strips were incubated with the indicated enzyme solutions. After the strips had been washed, residual phosphotyrosine containing bands were visualized by anti-phosphotyrosine antibody western blot. Trace magnesium in lane 4 is from the 0.5% nonfat dry milk present in the enzyme incubation buffer.

for hydrolysis in the "closed" conformation. A second group, consisting of phosphomannomutase, sucrose phosphatase, and trehalose phosphatase, contains a large domain between motifs II and III. It seems likely that this domain also is positioned proximal to the active site and performs an analogous function in these enzymes, which also hydrolyze small-molecule substrates. Although no crystal structures of these enzymes are presently available, it is perhaps telling that no sequence has yet been found in which large domains are found in both positions.

The third group, consisting of histidinol phosphatase, the DEM-1 DNA kinase/phosphatase, and MDP-1, has neither of these large domains. If, as it appears from the available crystal structures, the capping domain is an integral part of the HAD superfamily enzymes, critical for recognizing and presenting the substrate to the catalytic domain, what is to be made of these three enzymes? Histidinol phosphatase (HisB) is actually a bifunctional enzyme, containing a large C-terminal domain with imidazole glycerol phosphate dehydrogenase activity. The product of this domain is a precursor of the substrate for the phosphatase, and the enzyme is known to oligomerize in solution (*24*). It is a reasonable supposition that in this case it is the dehydrogenase domain that caps the phosphatase domain (from another subunit). The DNA kinase/phosphatase, interestingly, is also a bifunctional enzyme with the kinase activity in a C-terminal domain and also containing a large N-terminal domain. Although no information is available concerning the oligomerization state of this protein or the ability of these other domains to cap the active site, it would be reasonable to assume, from what is currently known about enzymes which act on DNA, that other factors act in conjunction with the enzyme. MDP-1, on the other hand, has no N- or C-terminal domains and, at least at pH 8.0 and 200 mM NaCl on a gel filtration column, is a monomer (*3*).

This distinctive lack of ancillary domains suggests several hypotheses for MDP-1's biological activity. There may be a capping domain that associates with MDP-1 and specific small-molecule substrates *in trans* in the form of a separate protein. Alternatively, the substrate of MDP-1 may, itself, be a protein. In this case a capping domain would be unnecessary because the extended surface of the substrate would presumably dock with the surface of MDP-1 surrounding the active site. Interestingly, the CheY protein, which has a fold and catalytic machinery similar to those of the HAD superfamily enzymes, has no capping domain and accepts a phosphate from a phosphorylated protein "substrate" (*25*).

General Protein Tyrosine Phosphatase Assay. To investigate this latter possibility, the ability of MDP-1 to dephosphorylate tyrosine phosphorylated proteins was tested. We did not test MDP-1's activity toward proteins containing phosphoserine or phosphothreonine because we have previously shown that MDP-1 hydrolyzes phosphotyrosine but not phosphoserine or phosphothreonine (*3*). Hydrogen peroxide stimulation of HeLa cells results in increased tyrosine phosphorylation of a number of protein species that can be visualized by western blotting with an anti-phosphotyrosine antibody (Figure 4, lane 1). Before application of the antibody, overlay of an active protein tyrosine phosphatase may remove some or all of these phosphate groups. Because of the denaturation of the blotted proteins in this assay, any activity observed cannot be regarded as specific; rather, this is a test of the competence of an enzyme to act on protein substrates in general. MDP-1 does indeed appear able to hydrolyze phosphorylated tyrosine residues of proteins in a magnesium-dependent manner (lanes 2−4). MDP-1 appears to be considerably more active (and less discriminating) in this assay than the protein tyrosine phosphatase PTP-1B (lane 5). This is in contrast to the lack of activity we have previously reported when a suite of tyrosine-phosphorylated peptides were tested as substrates. With these substrates PTP-1B was considerably more active than MDP-1 (*3*).

Conclusion. Sequence alignments and especially site-directed mutagenesis definitively identify MDP-1 as a member of the HAD superfamily of aspartate-dependent hydrolases. Consistent with its observed enzymatic activity, the conserved motifs of MDP-1 correspond most closely to the phosphatase and phosphomutase members of the superfamily. Despite this, the homology between MDP-1 and each of the other known members of the superfamily is poor: MDP-1 appears to be a distinct enzyme presumably acting on a unique substrate. Particularly notable is MDP-1's lack of a capping domain or any other domain outside the catalytic core that defines the HAD superfamily.

Although intriguing, the protein tyrosine phosphatase assay results do not provide strong evidence that MDP-1 acts as a protein phosphatase in vivo. Until such time as a biological substrate for MDP-1 is identified, both large and small molecules must be considered as candidates.

ACKNOWLEDGMENT

I thank Dr. Rodney Levine for his essential mentoring and support and Dr. P. Boon Chock for the donation of PTP-1B.

SUPPORTING INFORMATION AVAILABLE

Listing of the species and accession number (where not from genomic data only) of every amino acid sequence used in this paper. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Koester, M. K., Pullan, L. M., and Noltmann, E. A. (1981) *Arch. Biochem. Biophys. 211*, 632−642.
2. Kim, G., Selengut, J. S., and Levine, R. L. (2000) *Arch. Biochem. Biophys. 377*, 334−340.
3. Selengut, J. S., and Levine, R. L. (2000) *Biochemistry 39*, 8315−8324.
4. Morais, M. C., Zhang, W., Baker, A. S., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2000) *Biochemistry 39*, 10385−10396.
5. Hisano, T., Hata, Y., Fujii, T., Liu, J.-Q., Kurihara, T., Esaki, N., and Soda, K. (1996) *J. Biol. Chem. 271*, 20322.
6. Ridder, I. S., Rozenboom, H. J., Kalk, K. H., Janssen, D. B., and Dijkstra, B. W. (1997) *J. Biol. Chem. 272*, 33015.
7. Argiriadi, M. A., Morisseau, C., Hammock, B. D., and Christianson, D. W. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 10637.
8. Toyoshima, C., Nakasako, M., Nomura, H., and Ogawa, H. (2000) *Nature 405*, 647−655.
9. Wang, W., Kim, R., Jancarik, J., Yokota, H., and Kim, S.-H. (2001) *Structure 9*, 65−71.
10. Ridder, I. S., and Dijkstra, B. W. (1999) *Biochem. J. 339*, 223−226.
11. Stock, A. M., and Mowbray, S. L. (1995) *Curr. Opin. Struct. Biol. 5*, 744−751.
12. Stock, A. M., Martinez-Hackert, E., Rasmussen, B. F., West, A. H., Stock, J. B., Ringe, D., and Petsko, G. A. (1993) *Biochemistry 32*, 13375−13380.
13. Halkides, C. J., McEvoy, M. M., Casper, E., Matsumura, P., Volz, K., and Dahlquist, F. W. (2000) *Biochemistry 39*, 5280−5286.
14. Collet, J.-F., Stroobant, V., Pirard, M., Delpierre, G., and Van Schaftingen, E. (1998) *J. Biol. Chem. 273*, 14107−14112.
15. Thaller, M. C., Schippa, S., and Rossolini, G. M. (1998) *Protein Sci. 7*, 1647−1652.
16. Aravind, L., Galperin, M. Y., and Koonin, E. V. (1998) *Trends Biochem. Sci. 23*, 127−129.
17. Collet, J.-F., Stroobant, V, and Van Schaftingen, E. (1999) *J. Biol. Chem. 48*, 33985−33990.
18. Clarke, D. M., Loo, T. W., and MacLennan, D. H. (1990) *J. Biol. Chem. 265*, 22223−22227.
19. Maruyama, K., and MacLennan, D. H. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3314−3318.
20. Maruyama, K., Clarke, D. M., Fujii, J., Inesi, G., Loo, T. W., and MacLennan, D. H. (1989) *J. Biol. Chem. 264*, 13038−13042.
21. Kirihara, T., Liu, J. Q., Nardi-Dei, V., Koshikawa, H., Esaki, N., and Soda, K. (1995) *J. Biochem. (Tokyo) 117*, 1317−1322.
22. Apfel, A., Fischer, S., Goldberg, G., Goodley, P. C., and Kuhlmann, F. E. (1995) *J. Chromatogr. A 712*, 177−190.
23. Axelsen, K. B., and Palmgren, M. G. (1998) *J. Mol. Evol. 46*, 84−101.
24. Staples, M. A., and Houston, L. L. (1980) *Biochim. Biophys. Acta 613*, 210−219.
25. Djordjevic, S., and Stock, A. M. J. (1998) *Struct. Biol. 124*, 189−200.
26. Volz, K., and Matsumura, P. (1991) *J. Biol. Chem. 266*, 15511.

BI011405E