HAD Superfamily Phosphotransferase Substrate Diversification: Structure and Function Analysis of HAD Subclass IIB Sugar Phosphatase BT4131[†,‡]

Zhibing Lu,[§] Debra Dunaway-Mariano,*[,||] and Karen N. Allen*[,§]

Department of Physiology and Biophysics, Boston University School of Medicine, Boston, Massachusetts 02118-2394, and Department of Chemistry, University of New Mexico, Albuquerque, New Mexico 87131

Received January 3, 2005; Revised Manuscript Received March 21, 2005

ABSTRACT: The BT4131 gene from the bacterium *Bacteroides thetaiotaomicron* VPI-5482 has been cloned and overexpressed in *Escherichia coli*. The protein, a member of the haloalkanoate dehalogenase superfamily (subfamily IIB), was purified to homogeneity, and its X-ray crystal structure was determined to1.9 Å resolution using the molecular replacement phasing method. BT4131 was shown by an extensive substrate screen to be a broad-range sugar phosphate phosphatase. On the basis of substrate specificity and gene context, the physiological function of BT4131 in chitin metabolism has been tentatively assigned. Comparison of the BT4131 structure α/β cap domain structure with those of other type IIB enzymes (phosphoglycolate phosphatase, trehalose-6-phosphate phosphatase, and proteins of unknown function known as PDB entries 1NF2, 1NRW, and 1RKQ) identified two conserved loops (BT4131 residues 172−182 and 118−130) in the $\alpha\beta\beta(\alpha\beta\alpha\beta)\alpha\beta\beta$ type caps and one conserved loop in the $\alpha\beta\beta\alpha\beta\beta$ type caps, which contribute residues for contact with the substrate leaving group. In BT4131, the two loops contribute one polar and two nonpolar residues to encase the displaced sugar. This finding is consistent with the lax specificity BT4131 has for the ring size and stereochemistry of the sugar phosphate. In contrast, substrate docking showed that the high-specificity phosphoglycolate phosphatase (PDB entry 1L6R) uses a single substrate specificity loop to position three polar residues for interaction with the glycolate leaving group. We show how active site "solvent cages" derived from analysis of the structures of the type IIB HAD phosphatases could be used in conjunction with the identity of the residues stationed along the cap domain substrate specificity loops, as a means of substrate identification.

The haloalkanoic acid dehalogenase (HAD)[1] superfamily (*1*) consists of more than 3000 members in organisms ranging from prokaryotes to humans [for a recent review, see the work of Allen and Dunaway-Mariano (*2*)]. Although the superfamily was named after the first family member structurally characterized, 2-haloacid dehalogenase (*3*, *4*), the vast majority of known catalytic activities carried out by family members are directed at phosphoryl transfer. The phosphohydrolase activities of the ATPases and the phosphomonoesterases are the most prevalent. Diversification of the HAD superfamily (HADSF) catalytic scaffold to include phosphonate (P−C bond) hydrolysis [phosphonoacetaldehyde hydrolase (*5*)] and the transfer of phosphoryl groups between hexose hydroxyl substituents [phosphomutases (*6*)] has occurred less frequently but is, nonetheless, remarkable. Because the HAD fold is found in many different phosphohydrolases within a given organism [with 58 HAD homologues in *Homo sapiens* for instance (J. Selengut, personal communication)], it is evident that substrate diversification is the focus of the evolutionary expansion of this large superfamily. Identification of the structural features of the HAD members that underlie substrate specificity is a key to the assignment of function in this superfamily and a major goal of our work.

How then do HAD phosphohydrolases distinguish their substrates? All members of the HADSF possess an α,β-core domain consisting of a modified Rossmann fold, one face of which houses the active site. The core domain catalyzes the transfer of the phosphoryl group from a specific phosphate ester or anhydride to an active site Asp, and then to a water molecule (Figure 1A). The catalytic scaffold of the active site is formed by four loops (Figure 1B) comprised of the four consensus motifs (motifs I−IV) by which family members are recognized. The Asp nucleophile is located on loop 1. Loop 2 positions a conserved Ser/Thr that binds the substrate phosphoryl group, whereas loop 3 positions a conserved Arg/Lys that orients and shields charge in the Asp

[†] This work was supported by NIH Grant GM61099 to K.N.A. and D.D.-M.

[‡] The coordinates of the refined structure have been deposited with the Protein Data Bank as entry 1YMQ.

* To whom correspondence should be addressed. K.N.A.: Department of Physiology and Biophysics, Boston University School of Medicine, 80 East Concord 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.

[||] University of New Mexico.

[1] Abbreviations: β-PGM, β-phosphoglucomutase; HAD, haloalkanoate dehalogenase; HADSF, haloalkanoate dehalogenase superfamily; Me, metal ion; NAG6P, *N*-acetylglucosamine 6-phosphate; PNPP, *p*-nitrophenyl phosphate; phosphonatase, phosphonoacetaldehyde hydrolase; PSP, phosphoserine phosphatase.

10.1021/bi050009j CCC: $30.25
Published on Web 05/27/2005

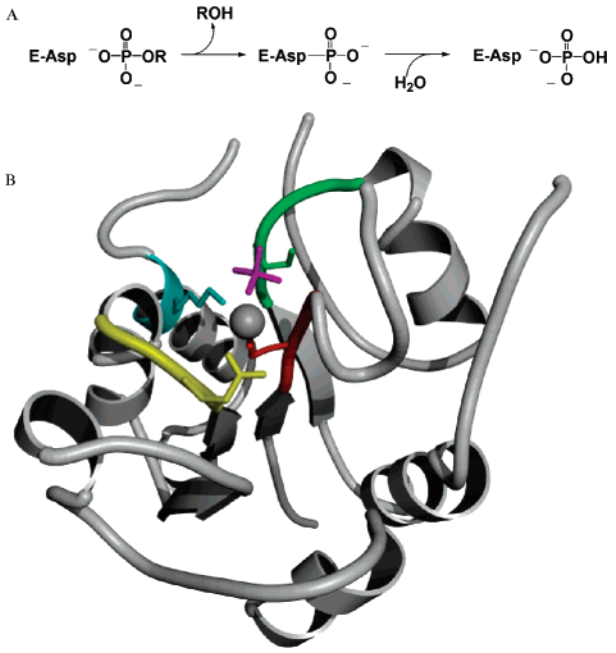

FIGURE 1: (A) Two-step reaction catalyzed by HADSF phosphohydrolases and (B) the supporting catalytic scaffold of the prototype phosphoserine phosphatase (PSP) (PDB entry 1L8L). The nucleophilic Asp (side chain shown) is located on loop 1 (red). Loop 2 (green) positions a Ser/Thr to bind substrate (phosphate group colored magenta); loop 3 (cyan) positions an Arg/Lys to orient the nucleophile, and loop 4 (gold) positions an Asp residue that coordinates the Mg^{2+} cofactor (gray sphere).

nucleophile and the phosphoryl group. Finally, loop 4 positions two or three Asp residues that bind the Mg^{2+} cofactor (in phosphohydrolases). The main roles of the core domain are thus binding, orienting, and activating the substrate phosphoryl group for transfer and the phosphoryl group acceptor for nucleophilic attack. The conserved "core residues" form the standard equipment for catalysis found in all HAD phosphotransferases examined to date.

Substrate recognition derives from structural accessorizing, which for a majority of the HAD members involves the addition of a second domain [for a discussion of substrate recognition in single-domain HAD members, see the work of Peisach et al. (*7*)]. Typically, this second domain is a smaller cap domain, the nature and location of which are used to divide the HADSF into subfamilies (*8*) (Figure 2A). Subfamily I has a small α-helical bundle cap domain located between motif I and motif II in the primary sequence; subfamily II has a mixed α/*β* domain located between motif II and motif III (Figure 2B), whereas subfamily III has no cap domain.

Studies of cap function have focused on the subfamily I enzymes, such as phosphoserine phosphatase (PSP) (*9*, *10*), phosphonoacetaldehyde hydrolase (phosphonatase) (*11*, *12*), and *β*-phosphoglucomutase (*β*-PGM) (*13*). X-ray crystallographic studies of these three enzymes have revealed two functionally important conformational states (Figure 3) that differ in the disposition of the cap domain and the core domain. In the "cap-closed" conformation, the cap domain is interfaced with the active site of the core domain. Residues from a specific cap segment termed the "substrate specificity loop" (*12*) enter the active site where they participate in substrate binding and catalysis. For instance, the substrate specificity loop of phosphonatase pictured in Figure 3 contributes the catalytic Lys53 residue that forms a Schiff base with the phosphonoacetaldehyde substrate as well as the Met49 residue, which contributes to the hydrophobic environment proposed to modulate Lys53 ionization (*11*, *14*). The cap-open conformation allows the active site access to solvent, thus facilitating substrate binding and product release. Interconversion between conformational states is based on hinge motion in the solvated domain linkers and is regulated by substrate binding (*14*).

Cap function among members of HAD subfamily II, the topic of this paper, has not been previously described. Herein, we report the structure and biochemical function of BT4131 from the human gut bacterium *Bacteroides thetaiotaomicron*. On the basis of an analysis of this enzyme, and the recently reported structures of several other HAD subfamily II enzymes [PDB entries 1NF2 (*15*), 1L6R (*16*), 1NRW, 1RKQ, and 1U02], we identify the cap-open and cap-closed conformational states, and the two substrate specificity loops, used in substrate recognition. BT4131 serves as an example of how residues stationed on substrate specificity loops can be used in conjunction with the active site solvent cage and the genome context (*17*) of the encoding gene as an inroad to the identification of biochemical function in "unknown" HAD subfamily II members.

EXPERIMENTAL PROCEDURES

Materials. Except where indicated, all chemicals were obtained from Sigma-Aldrich. The Biolmol green phosphate assay kit was purchased from Biolmol Research Laboratories Inc. Primers, T4 DNA ligase, and restriction enzymes were from Invitrogen. *Pfu*, *Pfu* Turbo polymerases, and the pET3 vector kit were from Stratagene. The Geneclean Spin Kit and the Qiaprep Spin Miniprep Kit were from Qiagen. Host cells were purchased from Novagen. Genomic DNA from *B. thetaiotaomicron* VPI-5482 (ATCC 29148) was a kind gift from J. Gordon (Washington University, St. Louis, MO).

Cloning, Expression, and Purification. The cDNA encoding the BT4131 gene from *B. thetaiotaomicron* VPI-5482 was amplified by PCR using the genomic DNA from *B. thetaiotaomicron* VPI-5482 and *Pfu* Turbo DNA polymerase. Oligonucleotide primers (5′-GGACTAAAAGGAACATATGACGAAAG and 5′-CAGATGTGCTGACCGGATCCATGAA), containing restriction endonuclease cleavage sites *Nde*I and *Bam*HI, were used in the PCRs. The pET-3a vector, cut by restriction enzymes *Nde*I and *Bam*HI, was ligated to the PCR product that had been isolated and digested with the same restriction enzymes. The ligation product (pET-3A-042) was used to transform *Escherichia coli* JM109 competent cells (Stratagene) to prepare plasmid DNA for sequencing. Plasmid DNA was purified using a Qiaprep Spin Miniprep Kit. The gene sequence was confirmed by DNA sequencing carried out by the Molecular Genetics Core Facility at the Boston University School of Medicine.

The recombinant plasmid DNA, pET-3A-042, was used to transform competent *E. coli* BL21(DE3)pLysS cells. The transformed cells (4 L) were grown at 37 °C with agitation at 250 rpm in Luria broth containing 50 *μ*g/mL ampicillin for 4−6 h to an OD_{600} of ∼0.6−1.0 and then induced for 4 h at 37 °C at a final concentration of 0.4 mM isopropyl *β*-D-thiogalactopyranoside (IPTG). The cells were harvested by centrifugation [6500 rpm (7350*g*)] for 15 min at 4 °C to

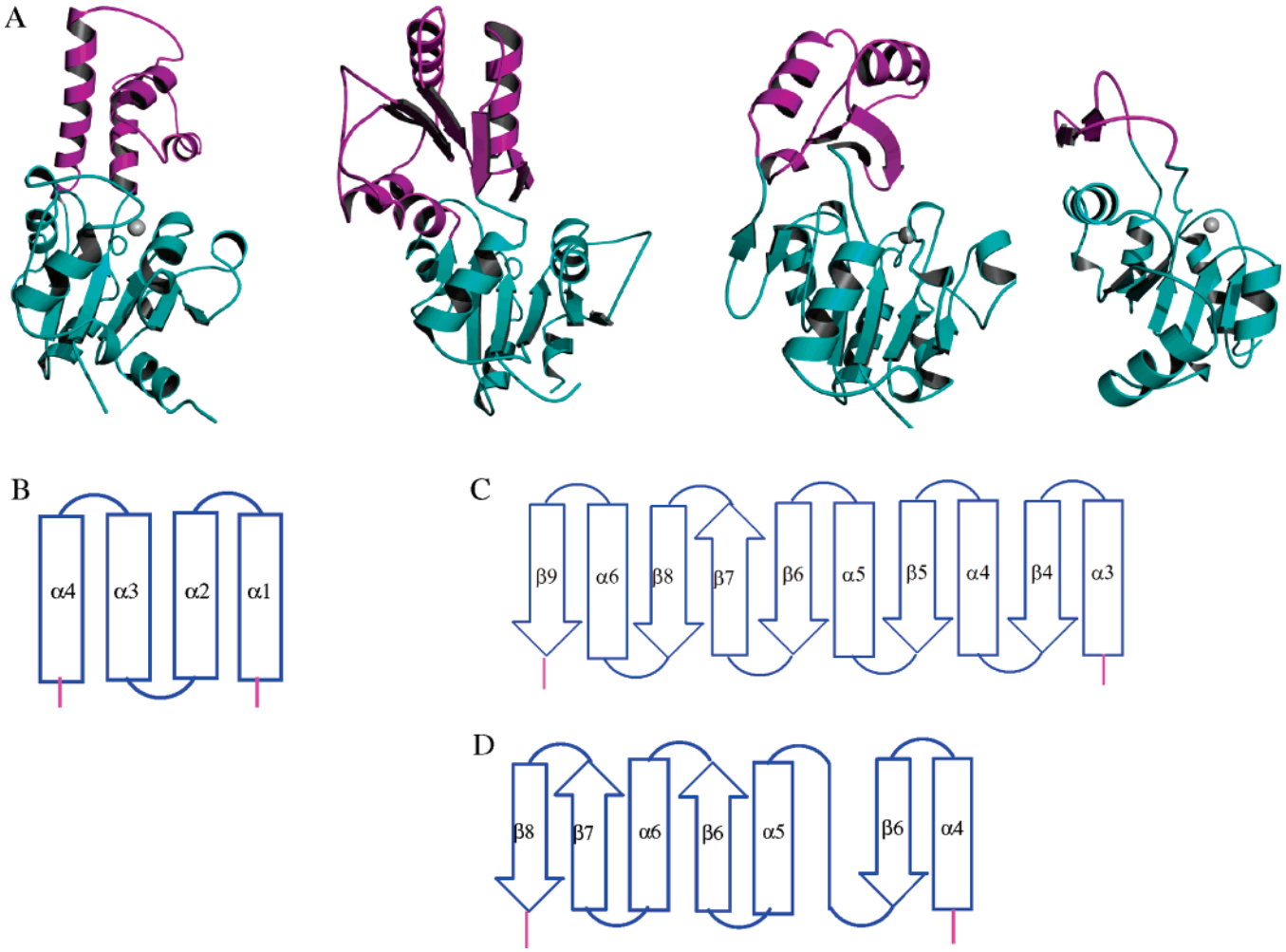

FIGURE 2: Representative structures of subfamily I (β-PGM), subfamily II type A (NagD), subfamily II type B (phosphoglycolate phosphatase), and subfamily III (magnesium-dependent phosphatase 1) (shown left to right). Subfamily members share a common core domain (cyan) but are distinguished by the presence and location of cap domains with differing topologies (magenta). (A) The entire structures are depicted as ribbon diagrams, and the cap domains of only (B) type I, (C) type IIA, and (D) type IIB are depicted as topology diagrams.

yield 2 g/L of culture medium. The cell pellet (8 g) was suspended in 80 mL of ice-cold lysis buffer consisting of 50 mM Na^+-Hepes (pH 7.5 at 25 °C) and 5 mM DTT. The cells were lysed by sonication (6 × 6 min at 50% power and 30% duty cycle) with a Branson 250 sonifier (VWR Scientific Inc.). The cell lysate was centrifuged at 4 °C for 120 min at 38384*g*. The supernatant was fractionated by ammonium sulfate-induced protein precipitation. The protein precipitated with 75−100% ammonium sulfate was harvested by centrifugation at 26895*g* for 15 min. Following dissolution in 25 mL of lysis buffer, the protein was loaded onto a 50 mL Butyl-Sepharose (Amersham Biosciences) column pre-equilibrated with 1 M ammonium sulfate in buffer A [50 mM Hepes (pH 7.5 and 25 °C) and 0.5 mM DTT]. The column was first washed with 200 mL of 1 M ammonium sulfate in buffer A and then eluted with a 1 L linear gradient of 1 to 0 M ammonium sulfate in buffer A. The column fractions were analyzed by SDS−PAGE. The desired protein eluted at ~0.1 M ammonium sulfate. The fractions were combined and concentrated at 4 °C using a 10K Amicon Ultra Centrifugal filter (Millipore) and then stored at −80 °C. The protein, shown to be homogeneous by SDS−PAGE, was obtained in a yield of 10 mg of protein/g wet cells.

Kinetic Constant Determinations. The purified recombinant enzyme was concentrated with an Amicon Ultrafiltration apparatus (PM10) or Centricon-10 (Millipore), and dialyzed against buffer A before being used in kinetic studies. The steady-state kinetic parameters (K_m and k_{cat}) of phosphorylated substrates were determined from initial reaction velocities measured at varying substrate concentrations (ranging from $0.5K_m$ to $5K_m$). The initial velocities were measured for reaction mixtures containing 5 mM $MgCl_2$ in 50 mM Hepes buffer (pH 7.0) at 25 or 37 °C. The assay methods used for the various substrates are described below. Protein concentrations were determined by the Bradford method (*18*), and absorbance measurements were performed with a Perkin-Elmer λ25 UV−vis spectrophotometer**.** Data were fitted to eq 1 with KinetAsystI

$$V_0 = V_{max}[S]/(K_m + [S]) \quad (1)$$

where V_0 is the initial velocity, V_{max} the maximum velocity, [S] the substrate concentration, and K_m the Michaelis−Menten constant for the substrate. The k_{cat} value was calculated from V_{max} and [E] according to the equation $k_{cat} = V_{max}/[E]$, where [E] is the protein subunit concentration in the assay reaction.

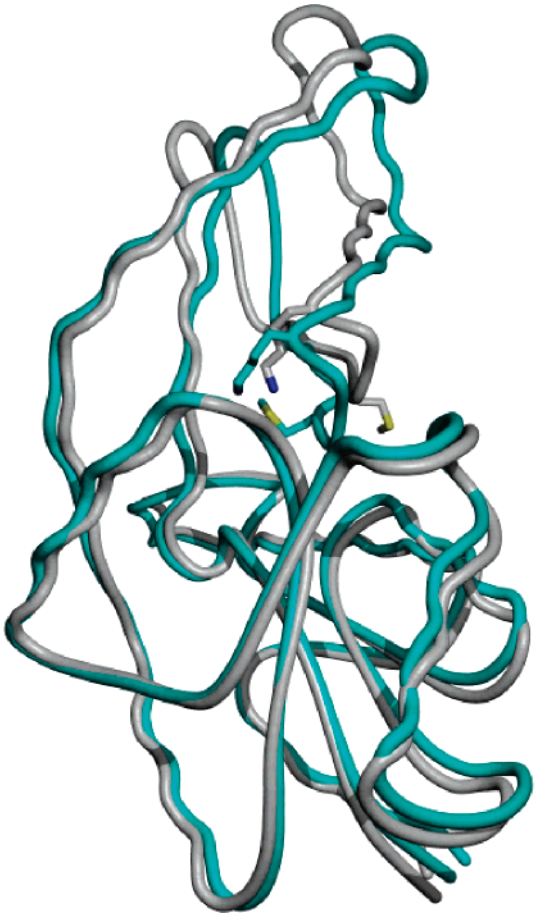

FIGURE 3: Overlay of the cap-open (gray) and cap-closed (cyan) structures of phosphonoacetaldehyde hydrolase. The Schiff base-forming Lys and essential Met contributed to the active site from the substrate specificity loop on the cap domain are depicted as sticks.

The sulfate inhibition constant was determined from initial velocity data measured in the presence and absence of sulfate and fitted to eq 2 (for competitive inhibition)

$$V_0 = V_{max}[S]/[K_m(1 + I/K_i) + [S]] \quad (2)$$

where I is the concentration of the inhibitor and K_i is the inhibition constant.

Activity Assays. The rate of *p*-nitrophenyl phosphate (PNPP) hydrolysis was determined by monitoring the increase in absorbance at 410 nm ($\Delta\epsilon = 18.4$ mM^{-1} cm^{-1}) (*15*) at 25 or 37 °C. The 0.5 mL assay mixtures contained 50 mM Hepes buffer (pH 7.0), 5 mM $MgCl_2$, and various concentrations of PNPP.

The rate of β-D-glucose 6-phosphate hydrolysis was determined by monitoring the rate of NADPH (340 nm; $\Delta\epsilon = 6.22$ mM^{-1} cm^{-1}) formation in a 0.5 mL coupled assay solution initially containing 50 mM Hepes buffer (pH 7.0 and 25 °C), 5 mM $MgCl_2$, 0.2 mM $NADP^+$, and 1 unit of glucose dehydrogenase (EC 1.1.1.119). The kinetic constants determined with this assay agreed with those determined using the fixed time phosphate assay described below.

Phosphate Assays. Phosphate ester hydrolysis for all substrates except PNPP was monitored using the Biolmol green kit to detect the total level of phosphate release. The 1 mL assay mixture, containing 50 mM Hepes buffer (pH 7.0), 5 mM $MgCl_2$, 1 mM substrate, and 1.2 μM BT4131 was incubated at 37 °C for 10 min. In parallel, the background level of phosphate was measured using a control reaction mixture, which excluded the BT4131. For analysis, 100 μL of the mixture was added to 1 mL of Biomol green reagent. After a 30 min incubation at room temperature, the absorbance of the solution at 620 nm was measured. Steady-state kinetic constant determinations were carried out using reaction solutions containing lower BT4131 concentrations (from 0.1 to 1.2 μM depending on the reactant) and varying concentrations of phosphate ester ($0.5-5K_m$).

Metal Activation. The metal cofactor specificity was examined using PNPP as a substrate. The BT4131 was first dialyzed at 4 °C against 1 L of 10 mM EDTA in 50 mM Hepes (pH 7.5) and then against 1 L of 1 mM Hepes (pH 7.0). Reaction solutions contained ~0.5 μM BT4131, 1 mM PNPP, and varying concentrations of $MgCl_2$, $MnCl_2$, $CoCl_2$, $NiCl_2$, $FeSO_4$, $ZnCl_2$, $CuBr_2$, and $CaCl_2$ in 50 mM Hepes (pH 7.0 and 25 °C). The initial velocity data were fitted with eq 1 using KinetAsystI.

Molecular Mass Determination. The theoretical molecular mass was calculated from the amino acid composition, which was derived from the gene sequence, by using the EXPASY Molecular Biology Server program Compute pI/MW (*19*). The molecular mass was measured by MALDI mass spectrometry (Dana Farber Cancer Institute, Boston, MA) and by SDS−PAGE (4% stacking gel and 12% separating gel). The molecular mass of native BT4131 was estimated by FPLC gel filtration column chromatography against protein standards (25−232 kDa from Amersham Pharmacia Biotech). The 2.5 cm × 120 cm Amersham Pharmacia Biotech Sephacryl S-200 column was eluted at 25 °C with 50 mM Hepes and 100 mM NaCl (pH 7.5). The native mass was calculated from a plot of the logarithm of the molecular mass versus elution volume from the column.

Crystallization and Diffraction Data Collection. The purified protein (~17 mg/mL), exchanged into 1 mM Hepes buffer (pH 7.5), was screened for crystallization by sparse matrix screening (*20*) with Crystal Screen Kits I and II (Hampton Research). After 10 days at room temperature (~25 °C), large orthorhombic crystals with overall dimensions of 0.5 mm × 0.4 mm × 0.4 mm appeared in the presence of 30% (w/v) polyethelene glycol 1500.

Crystals were frozen for data collection by passing them through 100% Paratone-N (Hampton Research) and then placing them directly in a stream of nitrogen gas cooled by liquid nitrogen. Diffraction data were collected at −180 °C to 1.6 Å resolution using Cu Kα radiation from a Rigaku RU-300 generator equipped with an R-Axis IV^{++} image plate located at the Boston University School of Medicine. Data were indexed and scaled using DENZO and SCALEPACK (*21*). Data collection statistics are summarized in Table 1. The crystals are orthorhombic, belonging to space group $P2_12_12_1$, with the following unit cell dimensions: $a = 49.35$ Å, $b = 55.74$ Å, and $c = 94.80$ Å. The unit cell volume is consistent with the presence of a monomer in the asymmetric unit assuming a Matthews coefficient of 2.3.

Phase Determination, Refinement, and the Final Model. The phase problem was solved using the molecular replacement method. The monomeric structure of a hypothetical phosphatase from the HADSF member whose sequence is 22.5% identical and 42.5% similar to that of BT4131 [PDB entry 1NRW (deposited as to be published)] was used

Table 1: Summarized Crystallographic Data Collection and Refinement Statistics

data collection statistics	
space group	$P2_12_12_1$
unit cell dimensions	$a = 49.35$, $b = 55.74$, $c = 94.80$
X-ray source	Cu Kα
wavelength (Å)	1.54
resolution range (Å)	94.8−1.9
no. of total/unique reflections	458 204/21 335
completeness (%)	99.7 (99.4)[a]
$I/\sigma(I)$	32.7 (4.4)[a]
R_{merge} (%)[b]	5.5 (35)[a]
volume fraction of protein (%)	62
data redundancy	4.7
refinement statistics	
no. of protein atoms	2017
no. of waters	183
no. of Mg(II) ions/no. of sulfate ions	2/1
no. of total reflections (working set/free set)	20794/2054
R_{work}/R_{free} (%)	17.2/21.0
average *B*-factor (Å^2)	27.4
protein	26.0
Mg^{2+} atoms	11.2
solvent	38.6
main chain atoms	22.3
SO_4 atoms	36.1
Luzzati coordinate error (Å)	0.17
rmsd	
bond length (Å)	0.0018
dihedrals (deg)	23.81
angles (deg)	1.71
impropers (deg)	1.27

[a] Values in parentheses represent values for the high-resolution shell (2.02−1.90 Å). [b] $R = (I - I)^2/I^2$.

as the search model. Side chains except Gly were mutated to Ala, and ligands were omitted. MOLREP in the CCP4 program suite was used to solve the rotation and translation functions, yielding a correct solution with a correlation coefficient of 31.9% and an *R*-factor 53.8% at 3.5 Å resolution. Although the *R*-factor was high, the difference in the correlation coefficient between this solution and the next best solution was large, and the resulting model gave no overlap between symmetry mates. The model obtained from Molrep was further refined using CNS (*22*).

Successive rounds of manual rebuilding were performed using the molecular graphics program O (*23*) followed by minimization and simulated annealing in CNS with data to 1.9 Å resolution. Because the initial model had 27 more amino acid residues than BT4131, these extra residues were gradually removed from the model. To avoid model bias, ligand molecules were added when R_{free} (*24*) was less than 30%. Waters were also added at this stage. The N-terminal methionine did not have clear electron density, and was removed from the model. The final model comprises 260 amino acids, two magnesium ions, one sulfate ion, and 183 water molecules, with an R_{free} of 21.0% and an R_{work} of 17.2%. Analysis of the Ramachadran plot as defined by PROCHECK (*25*) showed that 90.0% of residues fall in the most favored regions with 10.0% in the additionally allowed regions and no residues falling in the generously allowed or disallowed regions. Refinement statistics are summarized in Table 1.

Generation of an Active Site Cavity Plot and Model Docking. VOIDOO (*26*) and O were used to calculate the BT4131 active site cavity plot and to dock the potential BT4131 substrate in the calculated active site volume. For the calculation of the cavity, a probe radius of 1.4 Å was utilized. The models of the substrates used for docking were obtained from the Protein Data Bank (PDB entries 1HOT, 1FXQ, 1C9C, and 1JXA), and docking was manually performed in O. Allowed torsions were made to the ligands to give the conformer that best fit the shape of the active site volume.

RESULTS AND DISCUSSION

The goal of this work was to develop a strategy that could be used by investigators in the identification of biochemical function in unknown members of the HADSF subfamily II phosphohydrolases. The first step was to determine the structure of a subfamily II member and analyze this structure in the context of its known physiological substrate. HADSF subfamily II member BT4131 from *B. thetaiotaomicron* VPI-5482 was selected for structure−function analysis because its function as a phosphosugar phosphatase could be inferred from the genome context of its encoding gene (*27*) (see below). Thus, we had an inroad to the determination of the BT4131 physiological substrate and hence to the identification of the cap domain residues conferring substrate specificity. Several structures sharing the same cap topology (α*ββ*α*β*α*β*α*ββ*) as BT4131 had been recently deposited in the Protein Data Bank [PDB entries 1NF2 (*15*), 1L6R (*16*), 1NRW, 1RKQ, and 1U02]. Following structure−function analysis of BT4131, the structures of these homologues were analyzed to define substrate specificity determinants in the type II subfamily. Finally, the structures of the homologues were employed to explore the use of the active site solvent cage in the identification of substrate structure.

Cloning and Expression of BT4131. The BT4131 gene from *B. thetaiotaomicron* VPI-5482 was cloned and expressed in *E. coli* BL21(DE3)pLysS competent cells using the pET-3A vector. The protein was purified to homogeneity in a yield of 10 mg/g wet cells using a column chromatography-based protocol. The subunit mass of BT4131 was determined by MALDI mass spectroscopy to be 28 726 Da. The theoretical mass of the native protein was calculated as 28 856 Da, while that of the native protein without the N-terminal Met residue, 28 725 Da, agrees well with the experimental mass. The SDS−PAGE analysis gives an estimated subunit mass of 29 kDa, while the native mass determined by gel filtration analysis is 30 kDa. Thus, recombinant BT4131 is monomeric, and its N-terminal Met has been removed by post-translational modification.

Overall Structure and Fold. The final model of BT4131, including all residues except the first Met residue, was refined at 1.9 Å resolution to an R_{work} of 17.2% and an R_{free} of 21.0%. All residues are well-defined except cap domain residues 121−131, which have a high *B*-factor (57.5) compared to the average *B*-factor of the entire protein (26.0), indicative of conformational flexibility. There is one molecule in the asymmetric unit. There are two bound divalent metal cations, one which occupies the cofactor binding site found in all HAD phosphotransferases (*28*) and the other which is bound outside of the active site near the C-terminus (Figure 4A). The latter is coordinated to one oxygen atom of the Asp240 side chain, the backbone oxygen atoms of Leu219, Arg220, and Ala220, and two water molecules.

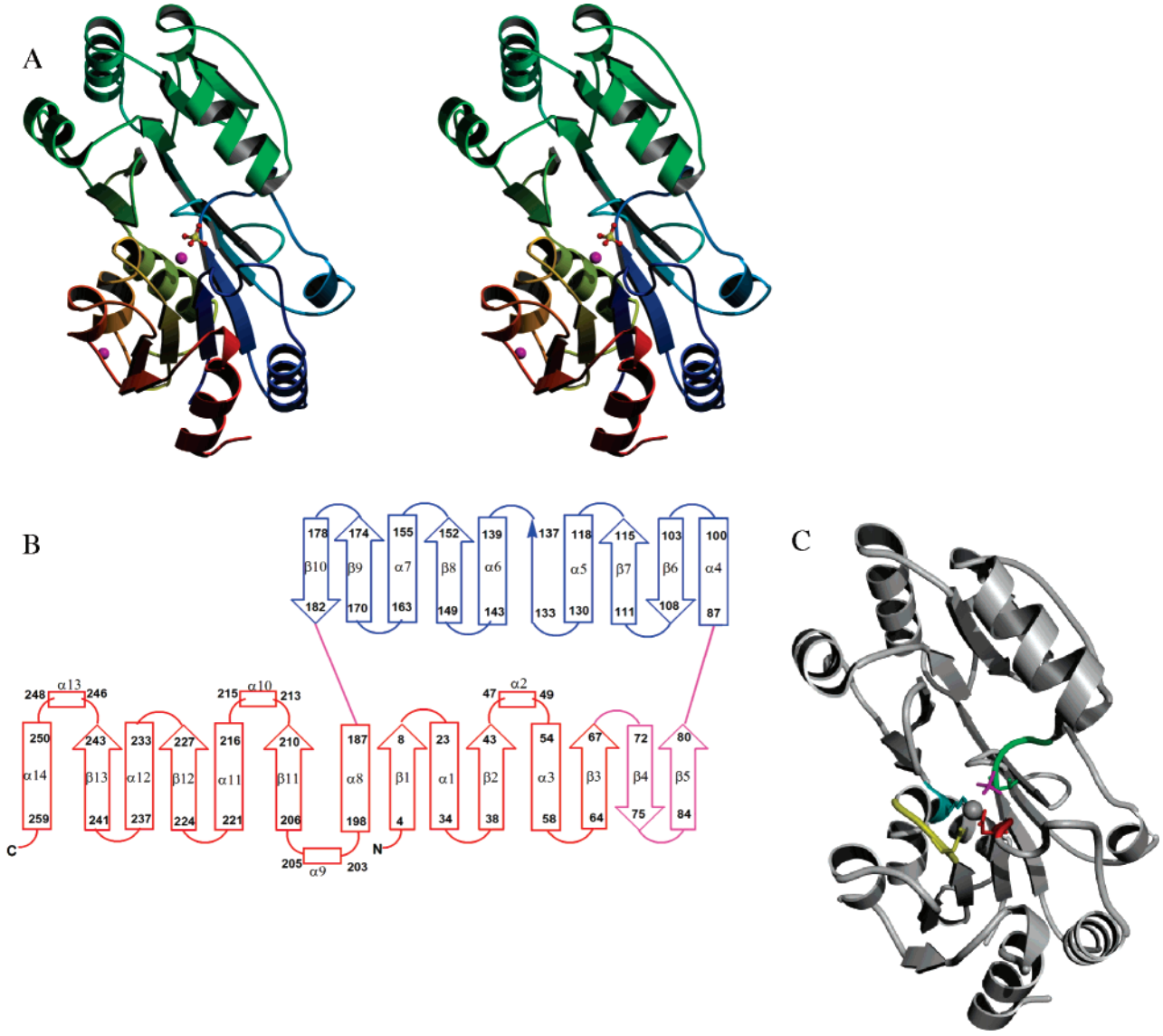

FIGURE 4: (A) Stereoview of the structure of BT4131 colored from the N-terminus (blue) to the C-terminus (red). Mg^{2+} ions are shown as magenta spheres. Sulfate ion bound in the active site is shown as a ball and stick. (B) Topological diagram of the secondary structure of BT4131. The red area represents the BT4131 core domain, the pink area the linker, and the blue area the cap domain. There is an extended segment inserted between α5 and α6 (residues 133−137) which in other subfamily II type B enzymes is a *β* strand. (C) BT4131 active site with conserved catalytic segments colored as in Figure 1B.

BT4131 is composed of two distinct domains connected by two linker regions (Figure 4A,B). The larger domain is the "conserved core" domain that houses the active site. The slightly smaller domain is the cap domain, which docks on the core domain to cover its active site. The core domain is a modified Rossmann fold comprised of a six-stranded parallel *β* sheet (*β*1−*β*3 and *β*11−*β*13) surrounded by six helices (α1, α2, and α7−α10). Two additional *β* sheets (*β*4 and *β*5) following *β*3 are part of the connection between the core domain and cap domain. The cap domain topology is of the α*ββ*(α*β*α*β*)α*ββ* type (see a further description in the section below).

The disposition of the cap domain relative to the core domain reflects a cap-closed conformation, wherein the core domain active site is sequestered from bulk solvent. In BT4131, the closed conformation might be attributed to the presence of the anionic (probably sulfate) ligand in the active site (see below). In other HAD enzymes, bound sulfate, phosphate, and tungstate ligands have been correlated with closed enzyme structures (*5*, *10*, *29*), whereas both open and closed conformations have been observed with the unliganded enzymes (*5*). It has been suggested that the anion mimics the phosphoryl group of the substrate and by shielding the positive charge of the bound Mg^{2+} facilitates cap domain−core domain association (*5*).

The α/*β* core domain is similar in structure to the core domain common to all other HADSF members. The cap domain topology places BT4131 in the HAD subfamily II type B category (subclass IIB). A search of the DALI database (*30*) identified 19 HADSF members with *Z* scores greater than 6.0. Because HADSF members share the core domain, the "hits" represent all three subfamilies. These structures share 4−24% overall amino acid sequence identity and rms deviations of 1.9−3.8 Å compared to BT4131. The HAD structures having the highest *Z* scores, however, are those which share cap topology with BT4131 and assume the same domain−domain conformation (i.e., open vs closed). The most closely related fold belongs to 1NF2

Table 2: Steady-State Kinetic Constants Measured for BT4131-Catalyzed Cleavage of PNPP at pH 7.0 and 25 °C in the Presence of Various Divalent Metal Ions

metal ion (Me)	k_{cat} (s^{-1})	K_m (μM)	k_{cat}(Me)/k_{cat}(Mg^{2+})
Mg^{2+}	0.051 ± 0.001	111 ± 8	1.0
Mn^{2+}	0.0284 ± 0.0008	9 ± 1	0.56
Co^{2+}	0.0077 ± 0.001	6.2 ± 0.3	0.15
Zn^{2+}	0.0044 ± 0.0003	13 ± 2	0.086
Ca^{2+}	0.00104 ± 0.00003	85 ± 7	0.02
Ni^{2+}	0.0017 ± 0.00003	1.5 ± 0.1	0.33
Cu^{2+}	no activation		
Fe^{2+}	no activation		

($Z = 26$), which is a subclass II type B HAD phosphatase with unknown function (*15*). The next most related structure is 1L6R ($Z = 20$), another subclass IIB HAD enzyme, which functions as a phosphoglycolate phosphatase (*16*). Both structures are, like BT4131, observed in the cap-closed conformation (see further discussion below). From the DALI search results, we anticipated that BT4131 is a phosphatase, a prediction that was confirmed by the observed catalyzed hydrolysis of PNPP (Table 2).

BT4131 Core Domain. The four-loop catalytic platform of the BT4131 core domain is shown in Figure 4C, and the residues that closely surround the bound metal ion and sulfate ligands are shown in Figure 5. In general, the structure of the core domain active site in BT4131 is essentially the same as the core domain active sites that have been reported for numerous other HADSF phosphotransferases (for example, compare the BT4131 structure in Figure 4C to the PSP structure in Figure 1B). However, there are two features of the BT4131 active site that are of special note that will be discussed here.

The first point of interest is the metal cofactor binding site. BT4131-catalyzed hydrolysis of PNPP was assessed in the presence and absence of Mg^{2+}. Whereas no activity above background was observed for metal-free BT4131 (prepared using EDTA; see Experimental Procedures) in the absence of added Mg^{2+}, the apparent k_{cat} (PNPP concentration $\sim K_m$) and K_m values (here K_m is equivalent to the dissociation constant for binding to the free enzyme in the presence of PNPP) for Mg^{2+}-activated BT4131 are 0.05 s^{-1} and 111 μM, respectively (Table 2). The purified enzyme which was not treated with EDTA retained a small amount of activity (6% of the activity measured in the presence of Mg^{2+}), indicating metal ion contamination in the enzyme preparation and/or assay solution. In fact, although divalent metal ions were not added to the crystallization buffer, electron density corresponding to a metal ion appeared in the BT4131 structure at a contour level of 11σ in an $F_o - F_c$ electron density map. The density was tentatively assigned as Mg^{2+} for two reasons. First, the anomalous peak at the metal ion binding site should show a very low-magnitude signal for magnesium (0.18 anomalous electron) and a significant peak for a transition metal such as Mn^{2+} (2.79 anomalous electrons) compared to the anomalous signal from a S atom in the protein (0.56 anomalous electron). The peak at the metal binding site is visible when the anomalous electron density map is contoured to 12.5σ, whereas the peak from the S atom of a Met residue with a low *B*-factor (14 Å^2) is visible up to 7.6σ. We assume a high total occupancy of the site (close to 100%) because of the low *B*-factors (all less than the average *B*-factor of the protein) of the liganding carboxylates (i.e., there is no evidence of a mix of occupied and unoccupied sites). Thus, from a weighted average, the amount of anomalous scattering at the metal site is consistent with 75% occupancy by a metal with the scattering properties of Mg^{2+} and 25% occupancy by a metal with the scattering properties of Mn^{2+}. Indeed, little residual density appeared after refinement of this metal in that position. Second, the kinetic constants derived from the metal ion specificity study reported in Table 2 suggested that Mg^{2+} is the physiological activator (i.e., it yields the highest activity). Of the other divalent metal cations which are activators and bind with higher affinity than Mg^{2+} (see Table 2), only Ni^{2+} has similar anomalous scattering to Mg^{2+} (but it results in one-third of the activity). Thus, although we cannot completely rule out Ni^{2+}, the metal cofactor is probably Mg^{2+}.

Regardless of the true metal ion identity, the position of the metal defines the core domain residues that bind it (Figure 5). As expected for enzyme-bound Mg^{2+}, there are six ligands with octahedral geometry. One of these ligands is assigned as a sulfate ion, and the basis for this assignment is described below. We suggest that by analogy to other HAD phosphotransferase structures, the sulfate ligand that we observe in the BT4131 active site is replaced with the phosphoryl group of the substrate in the enzyme−substrate complex (*9*, *14*, *31*). Thus, as with other HAD phosphotransferases, the Mg^{2+} cofactor serves to bind the transferring phosphoryl group. The observed coordination of the metal ion to the carboxylate of the loop 1 Asp8 nucleophile and to the backbone carbonyl of the loop 1 Asp10 acid/base residue is also common to the other HAD phosphotransferases. It is the manner in which loop 4 binds to the Mg^{2+} that makes the BT4131 catalytic site unique. In a previous analysis of loop 4-mediated metal ion binding in HAD phosphotransferases (*28*), a seven-station template that functions to position metal binding residues was described. The HAD phosphotransferases differ in the number and position of binding residues along this template as well in the type of binding (inner sphere vs outer sphere) that is involved. In BT4131, Asp211 at station 3 provides one metal ligand, Asn214 at station 6 binds the metal ion via formation of a hydrogen bond with one water ligand, and Asp215 at station 7 binds the metal via formation of a hydrogen bond with the second water ligand. This configuration is unique among those previously described (*28*) and serves as further evidence of the plasticity in configuration of the metal-binding loop of the HADSF phosphotransferases.

The assignment of the above-mentioned sulfate ligand is based on the scattering properties of sulfate and the presence of sulfate in the crystallization milieu. Specifically, electron density at a 2.5σ contour level in an $F_o - F_c$ map appeared in the enzyme active site in a region overlapping the region identified as the phosphate-binding site in other HADSF members (*5*−*7*, *10*, *29*). An anomalous difference map calculated with the Cu Kα data (1.54 Å) showed the intensity of the peak corresponding to the sulfate sulfur position is equivalent to that of a nearby sulfur atom in Met. Because no phosphate ion had been used in BT4131 purification and crystallization, but ammonium sulfate had been used in the purification, we suspected that the electron density derived from a bound sulfate ion. Sulfate ion was tested as a competitive inhibitor versus PNPP and found to have a K_i of 1.1 ± 0.1 mM, indicative of modest affinity for the

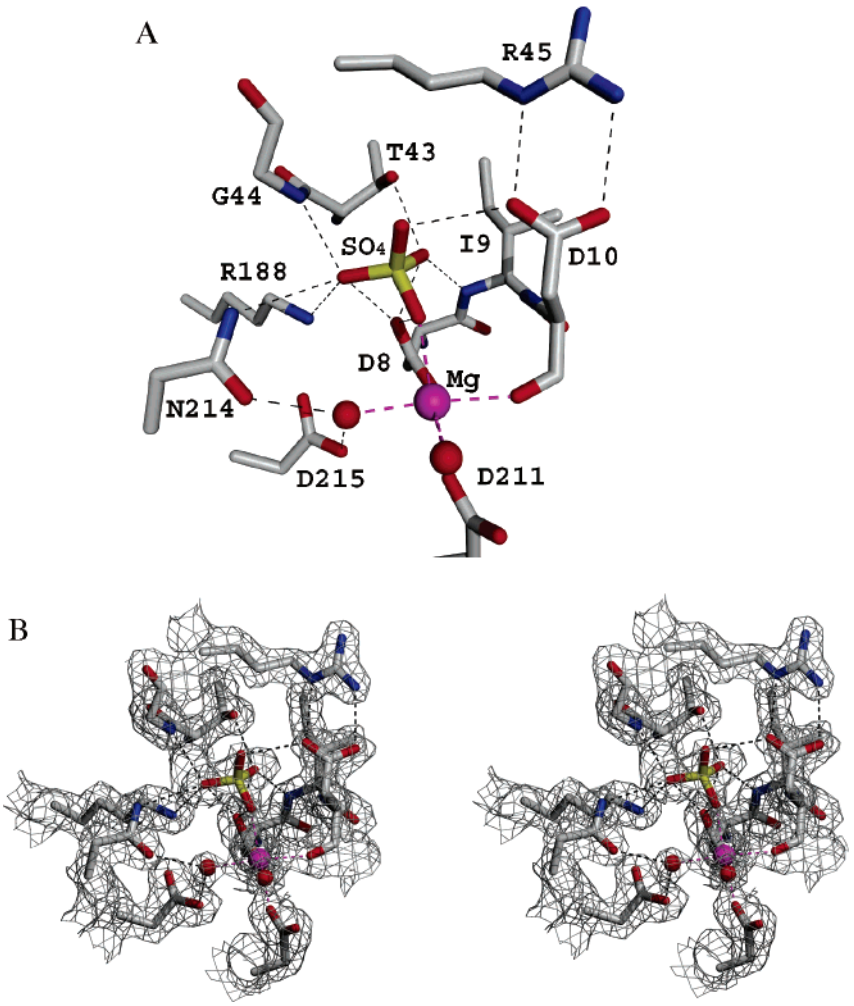

FIGURE 5: Sulfate binding site and hydrogen bond network in the BT4131 active site (A) and the $2(F_o - F_c)$ electron density map shown in stereo (B) contoured at 1.2σ. Mg^{2+} is colored magenta, and waters are colored red [one water ligand to Mg^{2+} (held by Asn214) is not shown for clarity].

phosphate-binding site. The nearest significant neighbors to the assigned sulfate ion are the Mg^{2+}, loop 1 Asp8 (the nucleophile), loop 1 Asp10 (the acid/base), loop 2 Thr43 (phosphate-binding residue), loop 3 Lys188 (phosphate-binding residue), and loop 4 Asn214 (metal-binding residue) (Figure 5). Note that Asp8 and Asp10 are proximal to the sulfate but do not participate in hydrogen bonds. This sulfate position is identical to that of sulfate or phosphate ligands in other HAD phosphotransferases (*6*, *10*, *29*).

The second feature of the BT4131 active site of special interest relates to the Asp10 acid/base residue. This residue is conserved among HAD phosphate ester hydrolases to protonate the alkoxide leaving group formed in the first partial reaction (Figure 1A). In the second partial reaction, involving the hydrolysis of the aspartyl−phosphate intermediate, it may also serve to deprotonate the water nucleophile. In the HAD subfamily I phosphotransferases examined to date, the Asp10 acid/base is positioned by the substrate (*6*, *9*, *29*). In contrast, Asp10 of BT4131 is positioned for its role in acid/base catalysis by loop 2 residue Arg45 (Figure 5). Inspection of all reported HAD phosphotransferase structures revealed that in subfamily IIB and III members, the Asp acid/base residue is bound by an electropositive residue which occupies the same spatial position as the Arg45 of BT4131 but can vary in sequence position. At present, it is not known why the Asp acid/base is "pinned down" by an active site residue in these subfamilies, but not in subfamily I, but this appears to be a general phenomenon that warrants future investigation.

Substrate Specificity and Biochemical Function. BT4131 activated with Mg^{2+} was found to catalyze the hydrolysis of PNPP with reasonable efficiency [$k_{cat}/K_m = 0.11 \times 10^3$ M^{-1} s^{-1} (Table 3)]. The physiological substrate of BT4131 was, however, unknown. On the basis of the location of the encoding gene near the chitobiase (β-*N*-acetylhexosaminidase) gene (*32*), we thought that BT4131 might be involved in sugar catabolism and thus might be a sugar phosphate phosphatase. To test this proposal, we examined the size and shape of the substrate-binding site in BT4131 as defined by the active site cavity in the catalytically active, domain−domain closed conformation of the enzyme. By using VOIDOO (*26*) to generate a surface plot of the cavity available to a probe the size of a water molecule (1.4 Å radius), a solvent cage was created. Common phosphorylated metabolites, including pyridoxal 5-phosphate and the sugar phosphates glucose 6-phosphate, arabinose 5-phosphate, and *N*-acetylglucosamine 6-phosphate (NAG6P), were evaluated for "goodness of fit" by manually docking their structures into the solvent cage. Of the metabolites screened, NAG6P gave the best fit (Figure 6) in that it filled the cage without

Table 3: Phosphorylated Compounds Tested as Substrates for BT4131 (50 mM Hepes buffer, pH 7.0, 5 mM $MgCl_2$, and 37 °C)

substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
p-nitrophenyl phosphate (PNPP)	0.083 ± 0.002	0.77 ± 0.05	0.11 × 10^3
β-D-glucose 6-phosphate	3.6 ± 0.2	10.7 ± 1.7	0.34 × 10^3
2-deoxy-D-glucose 6-phosphate	26 ± 1	3.9 ± 0.4	6.7 × 10^3
trehalose 6-phosphate	~0.06		
D-glucosamine 6-phosphate	0.27 ± 0.02	6.0 ± 0.9	0.045 × 10^3
N-acetyl-D-glucosamine 6-phosphate	3.6 ± 0.5	17 ± 5	0.21 × 10^3
D-fructose 6-phosphate	0.9 ± 0.1	4 ± 1	0.27 × 10^3
D-fructose 1,6-(bis)phopsphate	0.014 ± 0.002	4 ± 1	0.004 × 10^3
sucrose 6′-monophosphate	0.013 ± 0.001	3.1 ± 0.1	0.004 × 10^3
D-mannose 6-phosphate	1.3 ± 0.1	1.9 ± 0.3	0.7 × 10^3
D-arabinose 5-phosphate	9.6 ± 0.7	3.5 ± 0.5	3.2 × 10^3
D-ribose 5-phosphate	8.7 ± 0.4	4.9 ± 0.6	1.8 × 10^3
deoxyribose 5-phophate	1.45 ± 0.03	3.0 ± 0.1	0.48 × 10^3
ADP	0.016 ± 0.001	1.6 ± 0.4	0.01 × 10^3
DL-α-glycerol 3-phosphate	10.4 ± 0.2	8.7 ± 0.4	1.2 × 10^3
glyceraldehyde 3-phosphate	~1.0		
pyridoxal 5′-phosphate	1.75 ± 0.06	0.88 ± 0.08	2 × 10^3
sorbitol 6-phosphate	5.4 ± 0.2	6.8 ± 0.5	0.8 × 10^3
D-gluconate 6-phosphate	0.040 ± 0.005	2.2 ± 0.7	0.018 × 10^3

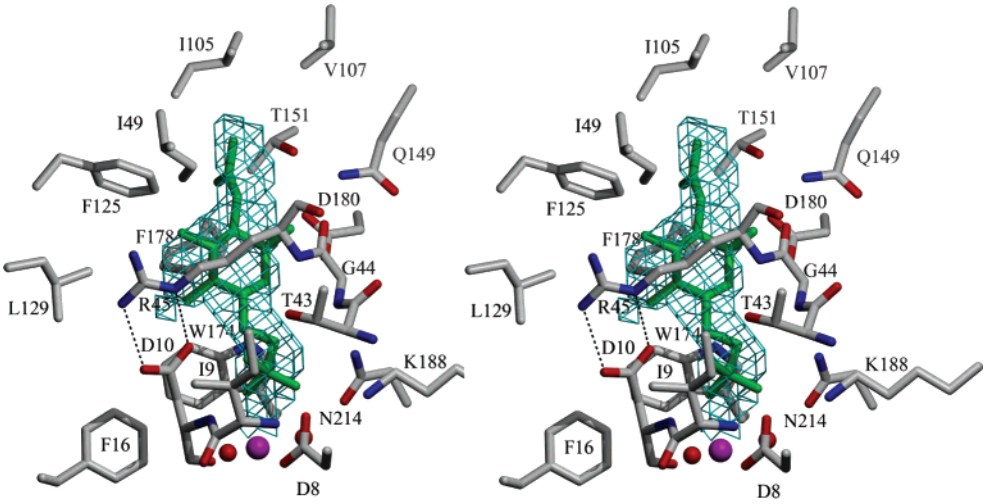

FIGURE 6: Stereodiagram of the active site of BT4131 with the substrate *N*-acetylglucosamine 6-phosphate (NAG6P, green) docked in a cage representing the accessible surface of the active site [calculated with VOIDOO (*34*)].

extending beyond it. The glucose 6-phosphate and arabinose 5-phosphate were also good matches but left small regions of the cage empty. We note that in the BT4131 structure, four water molecules occupy the space filled by the sugar motif of the docked NAG6P.

Using the results of our "virtual" substrate screen as a guide, a focused activity screen was devised. On the basis of the shape of the solvent cage, we predicted that hexose 6-phosphates or pentose 5-phosphates would be good substrates whereas ring isomers such as glucose 1-phosphate would not be. This is because direct attachment of the ring to the transferring phosphoryl group prevents the latter from extending through the narrow pocket to the Asp8 nucleophile (see Figure 6). Consistent with this prediction, we found that the cyclic sugars *β*-D-glucose 6-phosphate, 2-deoxy-D-glucose 6-phosphate, *N*-acetylglucosamine 6-phosphate, D-fructose 6-phosphate, D-mannose 6-phosphate, D-arabinose 5-phosphate, D-ribose 5-phosphate, and deoxyribose 5-phosphate are active substrates (k_{cat} = 1−26 s^{-1}; $k_{cat}/K_m \sim 1 \times 10^3$ M^{-1} s^{-1}) and that the cyclic sugars α-D-galactose 1-phosphate, α-D-glucose 1-phosphate, and α-D-(+)-mannose 1-phosphate are not (k_{cat} is under the detection limit of 0.01 s^{-1}) (Table 3). Similarly, pyridoxal phosphate, in which the phosphoryl group is extended from the planar ring by a methylene group, is significantly more reactive than PNPP, in which the phosphate group is attached directly to the ring ($k_{cat}/K_m = 2 \times 10^3$ M^{-1} s^{-1} vs $k_{cat}/K_m = 0.11 \times 10^3$ M^{-1} s^{-1}).

A second obvious feature of the substrate binding site in BT4131 is the hydrophobic nature of the residues (most of which derive from the cap domain) surrounding the sugar unit of the docked sugar−phosphate ligand (Figure 6). Because these nonpolar residues cannot offer hydrogen bonding interactions with the substrate hydroxyl groups, the interaction between the charged phosphoryl group and the electropositive groups of the core domain will be the principle source of favorable ligand binding energy (see Figure 5A). Thus, as long as the transferring phosphoryl group can access the Asp8 in the absence of a steric clash, catalysis is anticipated. Consequently, the ability of BT4131 to distinguish between most (monosaccharide) sugar phosphate substrates will be limited. Indeed, the values of the substrate specificity constant k_{cat}/K_m for D-glucose 6-phosphate, D-fructose 6-phosphate, D-mannose 6-phosphate, D-arabinose 5-phosphate, D-ribose 5-phosphate, DL-glycerol 3-phosphate, and sorbitol 6-phosphate are within a factor

of 10 in magnitude, with a mean k_{cat}/K_m value equal to 1×10^3 M^{-1} s^{-1} (Table 3).

A second corollary to the spatially confined, hydrophobic sugar-binding site is that phosphate esters that have charged substituents or that are simply too large to fit in the active site of the enzyme in the catalytically active, closed conformation will not be efficient substrates. Indeed, L-tyrosine *O*-phosphate, L-serine *O*-phosphate, ethanolamine *O*-phosphate, AMP, ATP, D-fructose 2,6-(bis)phosphate, and pyrophosphate are not active substrates, and D-gluconate 6-phosphate, D-glucosamine 6-phosphate, and D-fructose 1,6-(bis)phosphate are very poor substrates as are the disaccharides trehalose 6-phosphate and sucrose 6′-monophosphate.

On the basis of the substrate activity profile and the size, shape, and hydrophobicity of the substrate-binding site, we conclude that BT4131 might function in the catalysis of release of phosphate from hexose 6-phosphates and pentose 5-phosphates. The catalytic efficiency is low: $k_{cat}/K_m \sim 1 \times 10^3$ M^{-1} s^{-1} compared to the 1×10^6 to 1×10^8 M^{-1} s^{-1} "gold-standard" for single-substrate enzymes involved in primary metabolism (*33*). The modest catalytic efficiency and the broad substrate range are strong indicators that BT4131 functions in a secondary pathway.

Indeed, the (annotated) gene context of BT4131 in *B. thetaiotaomicron* as well as that of the ortholog BF0989 (69% identical sequences) in *Bacteroides fragilis* is suggestive of its role in recycling nitrogen and carbon from the cell wall polymer chitin. In *B. thetaiotaomicron*, neighbor gene BT4132 encodes chitobiase, an enzyme that catalyzes the hydrolysis of chitin-derived chitobiose (*N*-acetylglucosamine disaccharide) to *N*-acetylglucosamine. Another neighbor, BT4127, encodes glucosamine-6-phosphate isomerase. This enzyme recycles "NH_3" via the catalyzed conversion of glucosamine 6-phosphate and L-glutamate to fructose 6-phosphate and L-glutamine, respectively. The gene context prompts the speculation that BT4131 might function as fructose-6-phosphate phosphatase for chitin catabolism.

Enzymes that function in secondary pathways are not subjected to the selection pressure for catalytic perfection, and their k_{cat}/K_m values often reflect this. Moreover, broad substrate range is not compatible with catalytic perfection because the enzyme must accommodate a variety of substrate structures. There is in fact some evidence to suggest that the BT4131 active site accommodates rather than complements its intended sugar phosphate substrate. Whereas the optimal fit is observed with NAG6P ($k_{cat}/K_m = 0.2 \times 10^3$ M^{-1} s^{-1}), the best substrate activity is observed with 2-deoxyglucose 6-phosphate ($k_{cat}/K_m = 7 \times 10^3$ M^{-1} s^{-1}). Because 2-deoxyglucose 6-phosphate is not a known bacterial metabolite, it cannot be the physiological substrate and the 10-fold higher k_{cat} that is observed for 2-deoxyglucose 6-phosphate is not predicted by interaction between the ligand and the active site. Together, these observations suggest that the catalytic site of BT4131 is not specially tailored or honed for substrate specificity or catalytic efficiency. Thus, BT4131 appears to be an all-purpose sugar phosphate phosphatase that has been recruited for the purpose of recycling carbon and nitrogen from the cell wall polysaccharide chitin. In this capacity, exceptional speed and specificity are not required and are not features that will be cultivated by natural selection. Alternatively, it is always possible that in the panel of substrates that were tested in vitro, the true substrate has been missed. Such an omission might also be expected to result in the measurement of a relatively low k_{cat}/K_m for a number of the "nearest neighbor" substrates. However, if fructose 6-phosphate is not the "correct substrate", then it is at least a close structural analogue, and the specificity determinants found from interactions of the enzyme with this substrate model will still be valid. In the following section, we compare the structure and function of BT4131 with those of other HAD subfamily II type B enzymes to identify features of the type B cap domain that underlie catalytic efficiency and substrate specificity.

Comparison of Cap Domain Structure and Function in HAD Subfamily II Type B Phosphatases. HAD subfamily II enzymes display two topologically different cap domains indicative of two separate pathways of evolution. Subfamily II type A has the cap domain topology $\alpha\beta\alpha\beta\beta\alpha\beta\alpha\beta$, and type B has the cap domain topology $\alpha\beta\beta(\alpha\beta\alpha\beta)\alpha\beta\beta$ (see panels A and B of Figure 2). BT4131 is type B. There is only one type A structure currently reported (PDB entry 1PW5, NagD gene product from *Thermotoga maritima*), whereas several subfamily II type B structures [phosphoglycolate phosphatase, TA0175 from *Thermotoga acidophilium* (PDB entry 1L6R) (*16*), trehalose-6-phosphate phosphatase from *T. acidophilium* (PDB entry 1U02), "hypothetical phosphatase" TM0651 from *T. maritima* (PDB entry 1NF2) (*15*), "HAD-like hydrolase" from *Bacillus subtilis* (PDB entry 1NRW), and NYSGRC Target T1436 from *E. coli* (PDB entry 1RKQ)] have been reported. The type B cap domains, compared in Figure 2B, are the focus of this structure−function analysis.

As outlined in the previous discussion, the comparison of cap-open and cap-closed structures of the HADSF type I enzymes painted a picture of the catalytic cycle, starting with substrate binding (cap-open), proceeding through the chemical steps (cap-closed), and ending with product release (cap-open). Although we do not at present have a structure of BT4131 in the cap-open conformation, by overlaying the cap-closed BT4131 structure with the cap-open structure of the unknown phosphatase 1NF2, we can infer the cap domain motion required for a catalytic cycle in the type IIB HAD enzymes (Figure 7A). This analysis is allowed by the conservation of the fold in the core domains and in the cap domains of the type IIB enzymes (i.e., the gross differences between the structures are due to differences in the relative positions of the cap and core domains). The α/β fold of the cap domain of the type IIB members shows little variation in the relative orientation of structural elements within the domain (unlike the type I HADSF members where there is significant variation in the angles between the α helices in the helical bundle cap). Accordingly, cap domain movement in BT4131 is likely to occur through backbone rotation in the two solvated peptide linkers at or near residues 83−85 (linker 1, sequence SAI) and residues 184−186 (linker 2, sequence KGD). The caveat exists that a bulky residue in one structure may change the relative orientation of the core to cap and thus the analysis of the changes when comparing two different proteins. However, it is probable such changes would be more subtle than entire domain movements. Indeed, the fact that the active site in the closed form of BT4131 is completely solvent inaccessible necessitates domain−domain movement. In a more general sense, we anticipate that the same type of "hinge motion" in the interdomain linkers which

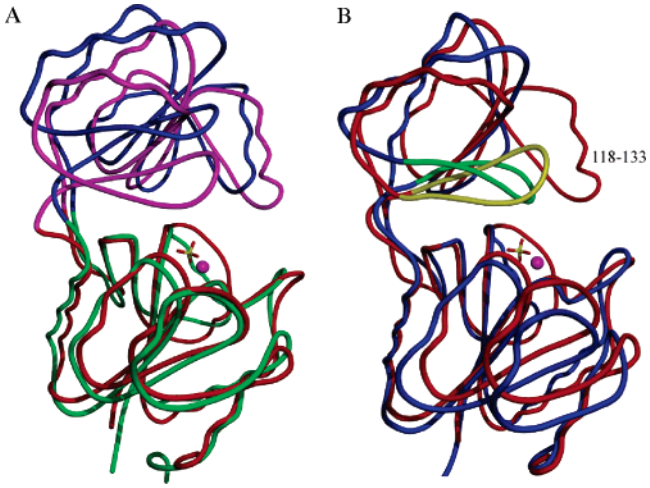

FIGURE 7: (A) Superposition of BT4131 (red and magenta for the core and cap, respectively) and 1NF2 (green and blue for the core and cap, respectively) to compare the cap-closed and -open forms, respectively, of this subgroup. (B) Superposition of the backbone of BT4131 (red) and phosphoglycolate phosphatase (1L6R, blue) with the proposed substrate specificity loops depicted in gold and green, respectively. In BT4131, the secondary substrate specificity loop originating from the cap domain (labeled 118−133) also contributes some residues to the active site and excludes bulk solvent. The Mg^{2+} cofactor and sulfate ligand from BT4131 are depicted as a ball and stick. This overlay of BT4131 and phosphoglycolate phosphatase also serves to compare the large cap and small cap topologies within subfamily II type B.

mediates cap association and dissociation in the HAD subfamily I enzymes (*9*, *11*) will also operate in the subfamily IIB enzymes.

The HAD subfamily IIB cap domains can be subclassified into the small $\alpha\beta\beta\alpha\beta\beta$ domain group or the large $\alpha\beta\beta(\alpha\beta\alpha\beta)\alpha\beta\beta$ domain group. The small domain lacks the $(\alpha\beta\alpha\beta)$ insert of the large domain and as a consequence uses a single loop to cover the core domain active site, while the larger domain uses two loops (Figure 7B). The small $\alpha\beta\beta\alpha\beta\beta$ domain group includes phosphoglycolate phosphatase (PDB entry 1LR6) and trehalose-6-phosphate phosphatase (PDB entry 1U02). The large $\alpha\beta\beta(\alpha\beta\alpha\beta)\alpha\beta\beta$ domain group includes BT4131 and unknown phosphatases 1NF2, 1NRW, and 1RKQ. These structures have been analyzed with two goals in mind: (1) to identify the structural motifs of the cap domains that contribute to the substrate-binding site of the enzyme in the cap-closed conformation and (2) to link the sequences of these motifs to substrate structure and biochemical function.

We started first with phosphoglycolate phosphatase in which high catalytic efficiency and substrate specificity are operative [$k_{cat}/K_m = 2.2 \times 10^5\ M^{-1}\ s^{-1}$ (*16*)]. The unliganded phosphoglycolate phosphatase structure reveals the cap-closed conformation. In this study, we generated a solvent cage and docked the phosphoglycolate ligand into the cage, thereby demonstrating a very good fit and identifying the residues that might, on the basis of this model of the liganded enzyme, function in substrate binding. In addition to the standard core domain interactions with the phosphoryl moiety, an interaction between the carboxylate substituent of the phosphoglycolate ligand and a cap domain His143 is evident (see Figure 8). We anticipate that His143 plays a key role in phosphoglycolate recognition. Significantly, His143 is located on the substrate specificity loop spanning residues 136−145 in the phosphoglycolate phosphatase. This loop, formed by residues 172−182 in BT4131, is present in all subfamily II type B enzymes (i.e., a structural motif common to the small and large cap domains). Notably, the specificity loop His143 of phosphoglycolate phosphatase is replaced with an Asp or Glu in all other subfamily II type B members of known structure. In the case of BT4131, the corresponding Asp180 may form a hydrogen bond with the C(1)OH group of the sugar phosphate substrate (see Figure 6). Other residues originating from the substrate specificity loop of phosphoglycolate phosphatase are Ser138 and Ser141 [corresponding to Trp174 and Phe178, respectively, in BT4131 (Figure 6)]. These polar residues contribute to the polar phosphoglycolate binding site (as Trp174 and Phe178 contribute to the nonpolar sugar phosphate binding site in BT4131).

Of the large $\alpha\beta\beta(\alpha\beta\alpha\beta)\alpha\beta\beta$ domain group, BT4131 and unknown phosphatases 1NRW and 1RKQ exist in a cap-closed conformation, which provides a "snapshot" of the composite active site formed by cap and core domain residues. Comparison of the active site solvent cages generated for 1NRW and 1RKQ shows that while 1NRW has a small cavity, 1RKQ has a cavity similar in size to that of BT4131. Both structures contain the previously noted substrate specificity loop (residues 172−182 in BT4131 numbering), hereafter termed the "primary substrate specificity loop". In addition, it is evident that a second loop (residues 118−130 in BT4131) contained within the $\alpha\beta\alpha\beta$ insert of all large domain enzymes also contributes to the active site. We denote this loop as the "secondary substrate specificity loop". 1NRW positions small polar residues Ser200 and Asn204 on the primary substrate specificity loop and Arg125 on the secondary loop (attaining the same position relative to the substrate as Phe125 in BT4131). In contrast to BT4131, 1NRW is structured to interact with a phosphate monoester that has a polar leaving group. In 1RKQ, the residue contributed to the binding site by the secondary substrate specificity loop is His129, whereas the residues contributed by the common substrate specificity loop are Ser184 and Phe188. In BT4131, the corresponding substrate specificity loop residues are Trp174 and Phe178. As with 1NRW, 1RKQ is compatible with a polar substrate. The ligand size and shape suggested by the solvent cage and the placement of the residues from the specificity loops suggest likely substrates for 1NRW and 1RKQ, which will be evaluated with a focused substrate screen.

Summary. BT4131, a representative of the HAD subfamily IIB class of phosphatases, was analyzed in terms of active site structure and substrate specificity. The potential application of the active site solvent cage in conjunction with the identity of the residues stationed along the cap domain substrate specificity loop(s) in predicting substrate structure was discovered. This "solvent cage"-based method was applied to virtual substrate screening in the known enzyme phosphoglycolate phosphatase to verify predictive power. In future studies, the method will be used to predict the substrates of unknown HAD subfamily IIB class phosphatases and their functions confirmed by applying focused substrate screens to the purified proteins.

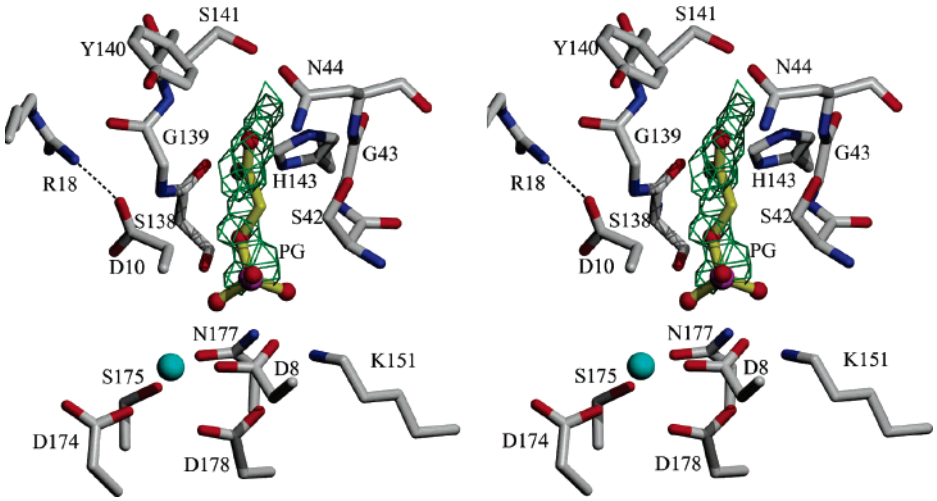

FIGURE 8: Stereoview of the active site of phosphoglycolate phosphatase (1L6R) with the substrate phosphoglycolate (yellow backbone) docked in a cage (green) representing the accessible surface of the active site [calculated with VOIDOO (*34*)].

ACKNOWLEDGMENT

We thank Dr. Jeffrey Gordon (Washington University) for kindly providing the genomic DNA from *B. thetaiotaomicron* VPI-5482. We thank Dr. Tracy Arakaki for her generous assistance with data processing, molecular replacement, and refinement. We also gratefully acknowledge Dr. Ezra Peisach for assistance with figure preparation and careful reading of the manuscript.

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. Allen, K. N., and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: Evolution of a catalytic scaffold, *Trends Biochem. Sci. 29*, 495−503.
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. 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.
6. 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.
7. Peisach, E., Selengut, J., Dunaway-Mariano, D., and Allen, K. N. (2004) Structure of the Magnesium-Dependent Protein Tyrosine Phosphatase, MDP-1, *Biochemistry 43*, 12770−12779.
8. 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.
9. Wang, W., Cho, H. S., Kim, R., Jancarik, J., Yokota, H., Nguyen, H. H., Grigoriev, I. V., Wemmer, D. E., and Kim, S. H. (2002) Structural characterization of the reaction pathway in phosphoserine phosphatase: Crystallographic "snapshots" of intermediate states, *J. Mol. Biol. 319*, 421−431.
10. 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.
11. 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.
12. Lahiri, S. D., Zhang, G., Dai, J., Dunaway-Mariano, D., and Allen, K. N. (2004) Analysis of the substrate specificity loop of the HAD superfamily cap domain, *Biochemistry 43*, 2812−2820.
13. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2003) The Pentacovalent Phosphorus Intermediate of a Phosphoryl Transfer Reaction, *Science 299*, 2067−2071.
14. Morais, M. C., Zhang, G., Zhang, W., Olsen, D. B., Dunaway-Mariano, D., and Allen, K. N. (2004) X-ray crystallographic and site-directed mutagenesis analysis of the mechanism of Schiff-base formation in phosphonoacetaldehyde hydrolase catalysis, *J. Biol. Chem. 279*, 9353−9361.
15. Shin, D. H., Roberts, A., Jancarik, J., Yokota, H., Kim, R., Wemmer, D. E., and Kim, S. H. (2003) Crystal structure of a phosphatase with a unique substrate binding domain from *Thermotoga maritima*, *Protein Sci. 12*, 1464−1472.
16. 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.
17. Aravind, L. (2000) Guilt by association: Contextual information in genome analysis, *Genome Res. 10*, 1074−1077.
18. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem. 72*, 248−254.
19. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J.-C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. (1998) Protein Identification and Analysis Tools in the ExPASy Server, in *2-D Proteome Analysis Protocols* (Link, A. J., Ed.) Humana Press, Totowa, NJ.
20. Jancerik, J., and Kim, S. H. (1991) Sparse matrix sampling: A screening method for crystallization of proteins, *J. Appl. Crystallogr. 24*, 409−411.
21. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods Enzymol. 276*, 307−326.
22. 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.
23. 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.

24. Brünger, A. T. (1992) The Free R Value: a Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures, *Nature 355*, 472−474.
25. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
26. Kleywegt, G. J., and Jones, T. A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures, *Acta Crystallogr. D50*, 178−185.
27. Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V., and Gordon, J. I. (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis, *Science 299*, 2074−2076.
28. Zhang, G., Morais, M. C., Dai, J., Zhang, W., Dunaway-Mariano, D., and Allen, K. N. (2004) Investigation of Metal Ion Binding in Phosphonoacetaldehyde Hydrolase Identifies Sequence Markers for Metal-Activated Enzymes of the HAD Enzyme Superfamily, *Biochemistry 43*, 4990−4997.
29. Rinaldo-Matthis, A., Rampazzo, C., Reichard, P., Bianchi, V., and Nordlund, P. (2002) Crystal structure of a human mitochondrial deoxyribonucleotidase, *Nat. Struct. Biol. 10*, 779−787.
30. Holm, L., and Sander, C. (1997) Dali/FSSP classification of three-dimensional protein folds, *Nucleic Acids Res. 25*, 231−234.
31. Rinaldo-Matthis, A., Rampazzo, C., Balzarini, J., Reichard, P., Bianchi, V., and Nordlund, P. (2004) Crystal Structures of the Mitochondrial Deoxyribonucleotidase in Complex with Two Specific Inhibitors, *Mol. Pharmacol. 65*, 860−867.
32. Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V., and Gordon, J. I. (2003) *Science 299*, 2074−2076.
33. Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W. H. Freeman and Co., New York.
34. Kleywegt, G. J., and Jones, T. A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures, *Acta Crystallogr. D50*, 178−185.

BI050009J