Analysis of the Substrate Specificity Loop of the HAD Superfamily Cap Domain[†,‡]

Sushmita D. Lahiri,[§] Guofeng Zhang,[#] Jianying Dai,[#] 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 September 17, 2003; Revised Manuscript Received December 23, 2003

ABSTRACT: The haloacid dehalogenase (HAD) superfamily includes a variety of enzymes that catalyze the cleavage of substrate C−Cl, P−C, and P−OP bonds via nucleophilic substitution pathways. All members possess the α/β core domain, and many also possess a small cap domain. The active site of the core domain is formed by four loops (corresponding to sequence motifs 1−4), which position substrate and cofactor-binding residues as well as the catalytic groups that mediate the "core" chemistry. The cap domain is responsible for the diversification of chemistry within the family. A tight β-turn in the helix−loop−helix motif of the cap domain contains a stringently conserved Gly (within sequence motif 5), flanked by residues whose side chains contribute to the catalytic site formed at the domain−domain interface. To define the role of the conserved Gly in the structure and function of the cap domain loop of the HAD superfamily members phosphonoacetaldehyde hydrolase and β-phosphoglucomutase, the Gly was mutated to Pro, Val, or Ala. The catalytic activity was severely reduced in each mutant. To examine the impact of Gly substitution on loop 5 conformation, the X-ray crystal structure of the Gly50Pro phosphonoacetaldehyde hydrolase mutant was determined. The altered backbone conformation at position 50 had a dramatic effect on the spatial disposition of the side chains of neighboring residues. Lys53, the Schiff Base forming lysine, had rotated out of the catalytic site and the side chain of Leu52 had moved to fill its place. On the basis of these studies, it was concluded that the flexibility afforded by the conserved Gly is critical to the function of loop 5 and that it is a marker by which the cap domain substrate specificity loop can be identified within the amino acid sequence of HAD family members.

The evolution of enzyme-mediated chemical reactions is founded on mutational change within enzyme superfamilies. The catalytic sites of superfamily members conserve residues required to catalyze the signature, or "core" chemistry specific to that superfamily, but have undergone changes at other active-site residues to extend that chemistry to new paradigms (*1*−*3*). Understanding the relationship between amino acid sequence and superfamily core chemistry provides the foundation for the simple task of sorting protein sequences of unknown function into superfamilies. The assignment of function within the superfamily poses a greater challenge, because it requires full recognition of the potential directions for expansion of core chemistry (as in branches on a tree) and sequence markers for that expansion (recognition of the specific chemical branch).

Our work with the haloalkanoate dehalogenase (HAD)[1] enzyme superfamily has focused on delineating the structural basis for new forms of catalysis within this superfamily. The goal is to establish guidelines for function assignment based on sequence data alone. At the time of this writing, the HAD enzyme superfamily contains over 2000 sequences, the vast majority of which have unknown function. Members with known function fall into one of five subfamilies: haloalkanoate dehalogenases (C−Cl bond hydrolysis), phosphonoacetaldehyde hydrolases (P−C bond hydrolysis), phosphate monoesterases, (P−OC bond hydrolysis), ATPases (P−OP bond hydrolysis), and phosphomutases (P−OC cleavage with intramolecular phosphoryl group transfer). The core chemistry catalyzed by this superfamily is group transfer, mediated by an active-site Asp carboxylate group.

The three-dimensional structure of at least one member from each subfamily is known. These are L-2-haloacid dehalogenase (*4*), phosphonatase (*1*), β-phosphoglucomutase (β-PGM) (*3*), the N-terminal domain (a lipid phosphatase)

[†] This work was supported by a grant from National Institute of Health GM16099 (to K.N.A. and D.D.-M.). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under Grant No. RR07707.

[‡] The coordinates of the refined structure have been deposited in the Protein Data Bank, accession code 1RDF.

* To whom correspondence should be addressed. Karen N. Allen, Department of Physiology and Biophysics, Boston University School of Medicine, 715 Albany Street, Boston, MA 02118-2394. Tel: 617-638-4398. Fax: 617-638-4273. E-mail: allen@med-xtal.bu.edu. Debra Dunaway-Mariano, Department of Chemistry, University of New Mexico, Albuquerque, NM, 87131. Tel: 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; DTT, dithiothreitol; G1P, β-D-glucose 1-phosphate; G6P, β-D-glucose 6-phosphate; HAD, haloacid dehalogenase; HEPES, *N*-(2-hydroxyethyl)piperzine-*N*′-2-ethanesulfonic acid; NADP, dihydronicotinamide adenine dinucleotide phosphate; NCS, noncrystallographic symmetry; phosphonatase, phosphonoacetaldehyde hydrolase; Pald, phosphonoacetaldehyde; MLF, maximum likelihood using amplitudes; rmsd, root-mean square deviation, vinyl sulfonate; vso_3.

10.1021/bi0356810 CCC: $27.50 © 2004 American Chemical Society
Published on Web 02/21/2004

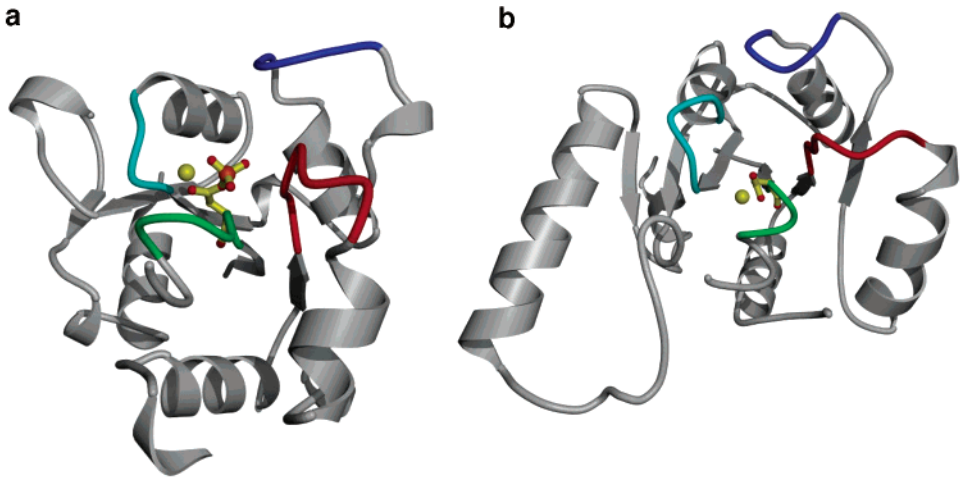

FIGURE 1: A view of (a) β-PGM and (b) phosphonatase core domains only, looking into the active site. The four conserved loops are depicted as coils with motif 1 in green, motif 2 in blue, motif 3 in cyan, and motif 4 in red. The conserved nucleophile Asp is shown in a yellow ball-and-stick model and the divalent Mg^{2+} bound to the active site is shown as a yellow ball. The phosphate atom of the phosphoaspartate species in β-PGM is colored orange.

of the bifunctional mammalian phosphatase-epoxide hydrolase (*5*), the P-domain of the calcium pump ATPase (*6*), the phosphoserine phosphatase (*7*), the 3-deoxy-D-manno-octulsonate 8-phosphate phosphatase (*8*), the deoxyribonucleotidase and the C-terminal phosphatase domain of the bifunctional T4 polynucleotide kinase (*9*). Each of these enzymes contains a highly conserved α/β core domain, in which the active site is formed by four loops (Figure 1). The four loops of the core domain contain conserved residues that orchestrate the core chemistry. The presence of the four sequence motifs in the protein sequence identifies that protein as a HAD superfamily member.

Many HAD superfamily members also possess a cap domain that acts as a dynamic lid over the core domain active site. The cap domain is formed by a sequence insert between loop 1 and loop 2 of the core domain in the HAD type I subclass (*10*) represented in Figure 2. In each of these proteins, the α-helical cap domain is linked to the core domain by two solvated linkers. A helix−turn−helix motif is poised to close over the core domain active site upon the hinge movement within the linkers (Figure 3). The turn region of this structural motif (hereafter called loop 5, or the substrate-specificity loop), comprise residues that, in the closed conformation, enter the active site and engage in substrate binding and/or catalysis. The catalytic site is thus a composite of the four loops of the core domain and the loop 5 of the cap domain. Whereas the core domain orchestrates the core chemistry, the cap domain functions in adapting that chemistry to a specific substrate.

Loop 5 retains its conformation in the cap-open and cap-closed states, even though it is loop 5 that binds the cap domain to the core domain. A common element of loop 5 within the superfamily is the conserved Gly residue located at the N-terminal end of the hairpin turn (i.e., helix2−(Gly) turn−helix3). An alignment of amino acid sequences of the helix−turn−helix motif in several HAD family members is shown in Figure 4a. Conservation of Gly at the turn can be rationalized on the basis of the strict spatial requirements of the turn region. Upstream and downstream of the Gly are residues that are stringently conserved within a given subfamily. Inspection of the enzyme structure in its liganded, closed conformation reveals the probable functions of the conserved loop 5 residues in domain−domain binding, active-site desolvation, and/or catalysis.

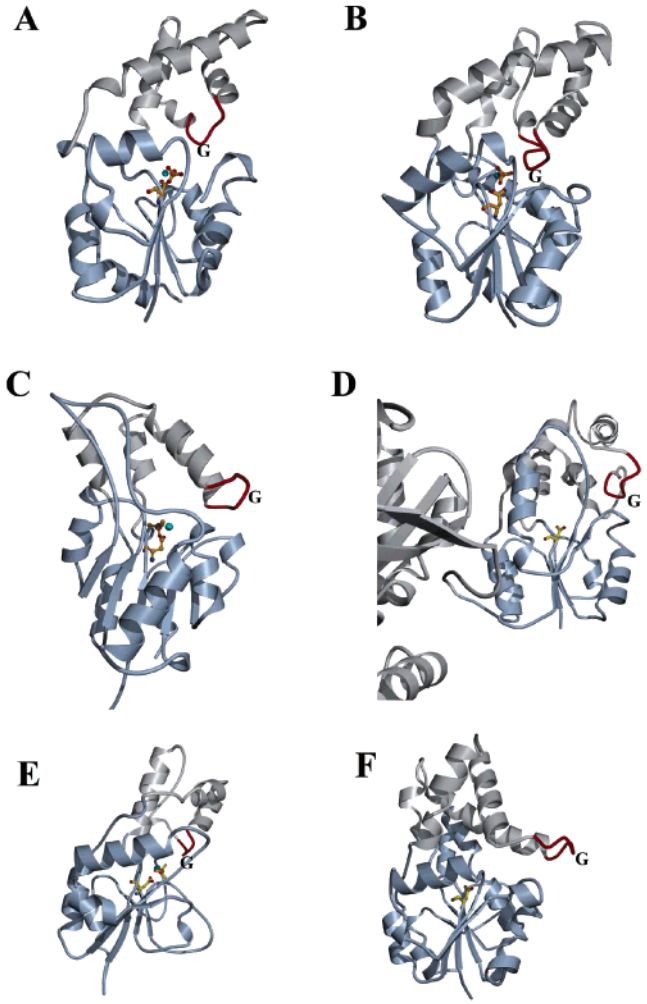

FIGURE 2: A ribbon diagram of the structures of two-domain HAD members (A) β-PGM, (B) phosphonatase, (C) phosphoserine phosphatase, (D) soluble epoxide hydrolase, (E) deoxyribo-nucleotidase, and (F) haloacid dehalogenase. The backbone of the cap domain is gray, the core is silver, and the turn region in the helix-turn-helix motif is red. The conserved Gly is marked with "G".

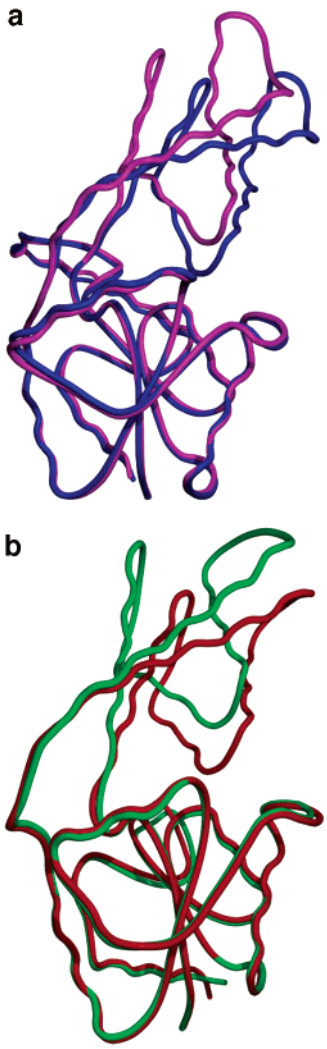

FIGURE 3: Overlay of the Cα backbone of open and closed conformations of (a) phosphonatase (open, mauve; and closed, blue) and (b) β-PGM (open, green; and closed, red).

The focus of this paper is the structure and function of loop 5 in two HAD superfamily enzymes *Bacillus cereus* phosphonoacetaldehyde hydrolase (phosphonatase) and *Lactococus lactis* β-phosphoglucomutase (β-PGM) (Figure 3). Phosphonatase catalyzes the hydrolysis of phosphonoacetaldehyde (Pald) to acetaldehyde and orthophosphate in 2-aminoethylphosphonate-degrading bacteria (*11*, *12*). β-PGM (which differs from the more familiar α-phosphoglucomutase of higher organisms) converts β-D-glucose 1-phosphate (G1P) to β-D-glucose 6-phosphate (G6P) in maltose and trehalose-degrading bacteria (*13*, *14*). The cap domains of these two enzymes are remarkably similar in fold (Figure 4b) even though they perform quite different catalytic functions.

In this study, the effects of substitution of the loop 5 essential Gly, and its flanking conserved residues, on cap-domain structure and/or function in phosphonatase and β-PGM were examined. Kinetic and X-ray crystallographic data show that Gly in the helix−turn−helix is essential for cap domain function because it determines the loop 5 conformation and flexibility, which in turn is critical to the function of essential loop catalytic and/or binding residues located on either side of it.

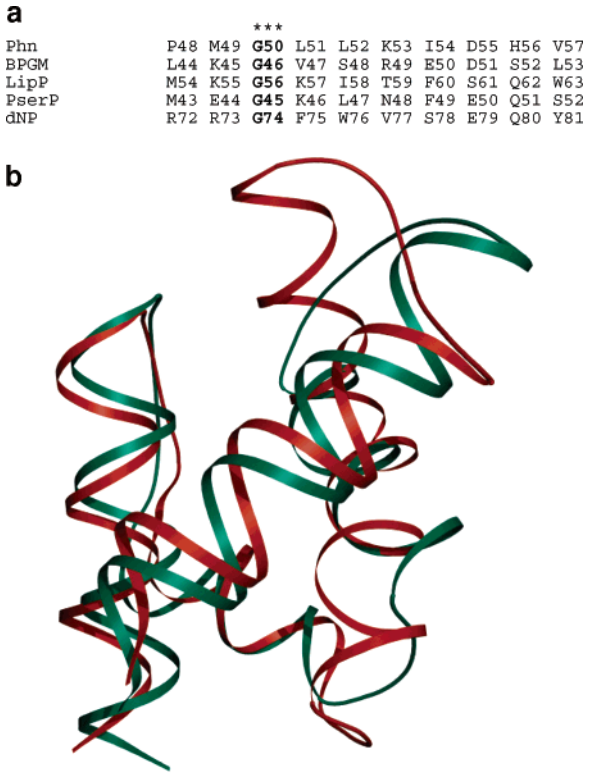

FIGURE 4: Alignment of the cap domain (a) sequences from HAD family members. Stringently conserved glycine is highlighted in bold lettering. Phn, phosphonatase; BPGM, β-phosphoglucomutase, LipP, soluble epoxide hydrolase; PserP, phosphoserine phosphatase; dNP, deoxyribo-nucleotidase. (b) Superposition of the cap domain of wild-type phosphonatase (green, accession code 1FEZ) and β-PGM structures (red, accession code 1LVH) demonstrate the similarity in the secondary fold of the cap domain.

EXPERIMENTAL PROCEDURES

Materials. All buffers used in protein purification were purchased from U.S. Biological or Sigma. DNA manipulation enzymes were purchased from New England Biolabs and used with the buffers provided. NADH, alcohol dehydrogenase, baker's yeast glucose 6-phosphate dehydrogenase, β-NADP, α-glucose 1,6-bisphosphate, and β-glucose 1-phosphate (highest grade available) were purchased from Sigma and used without further purification. Pald was synthesized according to the published method (*11*, *15*).

Preparation of Proteins. Recombinant wild-type *Bacillus cereus* phosphonatase and *Lactococus lactis* β-PGM were prepared as previously described ((*13*, *14*). The phosphonatase and β-PGM mutant proteins were generated using a PCR-based approach and the clones of the wild-type genes as templates (*13*, *14*). The primers used in the polymerase chain reaction (PCR) were synthesized by Gibco Life Technology, and the sequences of the mutant genes were confirmed by DNA sequencing by the Center for Genetics in Medicine, University of New Mexico School of Medicine, Albuquerque. The phosphonatase mutant proteins were expressed in *Escherichia coli* JM109 and purified using the published procedure for the wild-type enzyme (*15*). The β-PGM mutants proteins were expressed in *E. coli* BL21-(DE3) and purified by the published procedure for the wild-type enzyme (*16*). Protein purity was confirmed by SDS−PAGE analysis.

Kinetic Characterization of Phosphonatase and β-PGM Mutants. The K_m and V_{max} values of phosphonatase mutants were determined from the initial velocity data measured as a function of Pald concentration (0.5−10 K_m). Reaction solutions (1 mL) contained Pald, 10 mM $MgCl_2$, 0.15 mM NADH, and 10 u/mL alcohol dehydrogenase in 50 mM K^+-HEPES (pH 7.5, 25 °C). The concentration of phosphonatase ranged from 0.5 to 5 μM, depending on the activity of the protein studied. Reactions were monitored at 340 nm ($\Delta\epsilon$ = 6200 M^{-1} cm^{-1}) for the conversion of acetaldehyde and NADH to ethanol and NAD^+. The initial velocity data were analyzed using eq 1.

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

where [S] is the substrate concentration, V_0 is the initial velocity, V_{max} is the maximum velocity, and K_m is the Michaelis constant. The k_{cat} was calculated from V_{max} and the enzyme concentration using the equation $k_{cat} = V_{max}/[E]$. The enzyme concentration was determined using the Bradford method (*17*). A mutant was labeled "inactive" if the substrate turnover rate measured in the presence of 10 μM enzyme did not exceed the background rate ($k_{cat} < 1 \times 10^{-5}$ s^{-1}).

The steady-state kinetic constants of the β-PGM mutants were determined from the initial velocity data measured as a function of β-glucose 1-phosphate concentration (0.5−10 K_m). Reaction mixtures (1 mL) contained 50 mM K^+HEPES (pH 7.0, 25 °C), 2 mM $MgCl_2$, 2.5 units of glucose 6-phosphate dehydrogenase, 200 μM β-NADP, and 50 μM α-glucose 1,6-bisphosphate. The concentration of enzyme ranged from 0.5 to 2.5 μM, depending on the activity of the mutant studied. The reaction was monitored at 340 nm (NADP reduction; $\Delta\epsilon$ = 6200 M^{-1} cm^{-1}), 25 °C and the data were analyzed using eq 1.

Crystallization and Data Collection. The G50P mutant of phosphonatase was stored in 50 mM K^+HEPES, pH 7.5, 10 mM $MgCl_2$, 0.1 mM DTT buffer at 4 °C for structural studies. Before crystallization, the protein was transferred to 1 mM K^+HEPES, pH 7.5, 10 mM $MgCl_2$, 0.1 mM DTT solution and concentrated in a Millipore Ultra-free concentrator to 15 mg/mL (Bradford protein assay, BioRad (*17*)). Crystals of the mutant protein were grown by the vapor diffusion method with hanging-drop geometry using equal volumes (3 μL) each of protein solution and well solution. Attempts to grow crystals under conditions previously used for the crystallization of the wild-type enzyme (*18*) failed. Sparse matrix screening (Hampton Research) was used to identity an alternative crystallization condition. Initial crystals were obtained, after one month, from condition 4 of the commercial screen, consisting of 0.1 M Tris HCl, pH 8.5 and 2.0 M ammonium sulfate (at 18 °C). The crystals were twinned and diffracted only to 8.0 Å on a RU300 X-ray generator. Subsequently, the conditions were optimized to 0.1 M Tris HCl, pH 7.0 and 1.5 M ammonium sulfate (at 18 °C) plus 100 mM vinyl sulfonate (vso_3, an inhibitor) to produce larger, single crystals that grew within a week.

For data collection, crystals were frozen in a stream of nitrogen gas cooled to −180 °C using Paratone-N (Hampton Research) as a cryoprotectant. Data were collected to 2.8 Å resolution on the BM14C beamline at Argonne National Laboratories using a CCD detector. The details of data collection are summarized in the Table 1. The DENZO and SCALEPACK package (*19*) were used for data indexing, reduction, and scaling. Crystals are monoclinic, belonging to space group *C*2. The unit-cell dimensions were a = 181.7, b = 147.4, c = 131.3 Å, and β = 125.2°. Subsequent molecular replacement solution identified three dimers with a Matthew's coefficient of 4.1 and 70% solvent content. The unusually high solvent content was consistent with the gel-like consistency of the crystals.

Table 1: Crystallographic Data Collection and Refinement Statistics

unit-cell dimensions (Å)	a = 181.7 Å, b = 147.4 Å, c = 131.3 Å, β= 125.2°
space group	*C*2
X-ray source	APS 14-BM-C
wavelength (Å)	1.0000
resolution range (Å)	50−2.7
total/unique reflections	511820/77509
completeness (%)[a]	97.6 (95.8)
I/σ (I)	10.5 (2.5)
R_{merge}[b] (%)	0.129 (0.393)
refinement statistics	
no. of protein atoms/ asymmetric unit	12601
no. of reflections (working/free set)	44780/5041
R_{work}[d]/R_{free}[e] (%)	0.239/0.260
average B-factors (Å^2)	
amino acid residues	33.9
Mg^{2+} ions (2 total)	6.2
sulfate (2 total)	38.3
water (258 total)	36.2
r.m.s deviation	
bond length (Å)	0.008
angles (°)	1.46

[a] Completeness of the outermost shell (2.8−2.7 Å) is shown in parentheses. [b] $R_{merge} = \sum_{hkl}\sum_i | I_{hkl,i} - \langle I_{hkl}\rangle I | / \sum_{hkl}\sum_i | I_{hkl,i} |$, where $\langle I_{hkl}\rangle$ is the mean intensity of the multiple $I_{hkl,i}$ observations for symmetry related reflections. [c] Reported by SOLVE. [d] $R_{work} = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} |F_{obs}|$. [e] $R_{free} = \sum_{hkl\ \Sigma T} |F_{obs} - F_{calc}| / \sum_{hkl} |F_{obs}|$, where the test set T includes 10% of the data.

Structure Determination and Refinement. Phases were determined via the molecular replacement method (*20*) using either the "open" or "closed" conformation of the wild-type phosphonatase (accession code 1FEZ.pdb) or vso_3 bound "closed" monomer (*21*) as the model using the program *AMoRE* (*22*). The lowest R-factor and highest correlation coefficient (47.2 and 46.4, respectively) were obtained for the vso_3 bound "closed" monomer. The *AMoRE* solution was subsequently subjected to rigid-body refinement allowing independent movement of the cap and the core as implemented in the *CNS* program suite (*23*), excluding 10% of the data for statistical cross-validation purposes (calculation of R_{free}) (*24*). Because of the high solvent content, a $\sigma(I)$ = 1 cutoff was applied to the data. SigmaA weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps (*25*) were used for manual rebuilding in the molecular graphics program *O* (*26*). Multiple cycles of minimization, slow-cool torsional molecular dynamics, and group B-factor refinement using an MLF target, followed by manual rebuilding were performed until R_{free} ceased to decrease. Waters were added using the automated water-picking routine in *CNS*. The model was checked using a composite-omit map with coefficients of $2F_o - F_c$. The final model incorporated 263 out of 267 possible amino acids. Residues 1−4 were not visible in the electron density map, and were omitted from the final model.

Table 2: Kinetic Constants of Wild Type and Mutant Phosphonatase and β-PGM

enzyme	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} M^{-1})
phosphonatase			
wild type	15 ± 1	33 ± 2	5.0×10^5
M49L	0.148 ± 0.006	5100 ± 400	2.9×10^1
G50P		inactive	
G50V		inactive	
G50A	0.43 ± 0.01	11000 ± 600	3.9×10^1
K53R		inactive	
K53A		insoluble	
H56Q	0.36 ± 0.03	175 ± 10	2.1×10^3
β-PGM			
wild type	17.6 ± 0.6	4.6 ± 0.5	3.7×10^6
K45A	0.037 ± 0.002	240 ± 40	1.5×10^2
K45R	0.61 ± 0.01	8.5 ± 0.7	7.2×10^4
G46P	0.126 ± 0.003	2200 ± 100	5.7×10^1
G46V	0.0046 ± 0.0001	18.8 ± 0.8	2.6×10^2
G46A	0.018 ± 0.002	2000 ± 100	9.0
R49K	0.0580 ± 0.0009	180 ± 10	3.2×10^2
R49A	0.078 ± 0.002	13300 ± 900	5.9
S52A	9.8 ± 0.6	4.04 ± 0.04	2.5×10^6

The two strong electron densities in the active site were assigned to vso_3 and Mg^{2+} ions, respectively. The final R_{free} and R_{work} for the model were 26.0 and 23.9%, respectively. Analysis of the Ramachandran plot of the final model shows that 90.7% of the residues fall within the most favored region, and 9.3% in the generously allowed region, as defined by *PROCHECK* (*27*). Refinement statistics are compiled in Table 1. Structural figures were prepared using MOLSCRIPT (*28*) and POVRAY (Persistence of Vision Ray Tracer, version 3.01).

RESULTS AND DISCUSSION

Kinetic Properties of Loop 5 Site-Directed Mutants: Amino Acid Replacements Impaired Catalysis in Phosphonatase and β-PGM. The loop 5 residues conserved among known phosphonatase sequences are Met49 (catalysis), Gly50 (turn determinant and cap to core domain binding), Lys53 (catalysis), and His 56 (catalysis). The β-PGM counterparts are conserved residues Lys45 (cap closure via substrate binding), Gly46 (turn determinant), Arg49 (cap to core domain binding), and Ser52 (stabilization of substrate-specificity loop). The k_{cat} and k_{cat}/K_m values for wild-type phosphonatase catalyzed Pald hydrolysis (measured at pH 7.5 and 25 °C) are 15 s^{-1} and 5×10^5 M^{-1} s^{-1}, respectively. The k_{cat} and k_{cat}/K_m values for wild-type β-PGM catalyzed conversion of G1P to G6P (measured at pH 7.0 and 25 °C) are 17 s^{-1} and 4×10^6 M^{-1} s^{-1}, respectively (Table 2). The k_{cat} and k_{cat}/K_m values of the phosphonatase mutants M49L, G50A, G50V, G50P, and H56Q and of β-PGM mutants K45A, G46A, G46V, G46P, R49A, and S52A are compared to the corresponding values measured for the wild-type enzymes in Table 2.

First, the impact of mutation of phosphonatase Met49 and His56 will be addressed. In the closed conformation of phosphonatase, His56 and Met49 are located in the catalytic site where they bind a water molecule, which, in turn, is positioned to engage in hydrogen bond formation with the carbonyl oxygen of the bound Pald (as assessed from the wild-type phosphonatase−vso_3 complex (*21*)). Because this water molecule is the only acid group close to the substrate carbonyl, it may function in proton transfer in the formation and dehydration of the carbinol intermediate. By replacing His56 with Gln and Met49 with Leu, the capacity for proton transfer and water orientation, respectively, was removed while the change in the space filling (steric) properties of the native residues was minimized. The consequences of these amino acid substitutions are evident in the kinetic properties of the mutant enzymes. The k_{cat} was reduced 100-fold in the M49L mutant and the k_{cat}/K_m was reduced 17 000-fold. The k_{cat} was reduced 40-fold in the H56Q mutant and the k_{cat}/K_m was reduced 200-fold. These results provide evidence for the importance of the cap domain Met49 and His56 in phosphonatase catalysis first suggested by the wild-type enzyme X-ray structure (*1*) (see Figure 5a). In earlier work, the essential function of Lys53 in Schiff-base formation was demonstrated (*11*) (for mechanism, see Figure 5b). Loop 5 thus positions three important residues (Met49, Lys53, and His56), which, upon cap closure, enter the active site and engage in catalysis.

On the basis of the X-ray structure of the enzyme, the phosphonatase loop 5 Gly50 appears to have two essential roles. One is to facilitate the hairpin turn at the helix−turn−helix motif, and the other is to stabilize the closed conformer of the enzyme, through hydrogen-bond formation between the Gly50 backbone amide and the (conserved, motif 3) core domain Arg160. In this study, Gly50 was replaced with Ala, Val, and Pro. Only the G50A mutant displayed catalytic activity, and it was quite low. G50A has a k_{cat} value of 0.4 s^{-1} (38-fold decrease) and a k_{cat}/K_m value of 40 M^{-1} s^{-1} (10 000-fold decrease). On the basis of these results, it is evident that the Gly residue is essential for conferring flexibility at the hairpin turn position. As observed in the X-ray structure of the G50P mutant (vide infra), an enforced change in backbone conformation at the turn position has the consequence of displacement of Lys53.

β-PGM mutants K45A, R49A, and S52A were examined to evaluate the importance of the three corresponding loop residues to catalytic function. The Ser52 plays a role in the stabilization of the conformation of the substrate-specificity loop by forming a hydrogen bond between the serine hydroxyl and the backbone carbonyl of Ser48. The k_{cat} value and k_{cat}/K_m of the S52A mutant are similar to those for the wild-type enzyme. This result is consistent with the fact that Ser52 contributes merely a single hydrogen bond to the structure of the helix−turn−helix motif and does not appear to play a significant role in catalysis.

The Lys45 forms an ion pair with the Asp170 of the core domain active site (Figure 6a). This interaction is important to domain−domain closure and to the electrostatic environment at the reaction site. The k_{cat} value of the K45A mutant is 500-fold lower than that of wild-type enzyme, while the k_{cat}/K_m value is 20 000-fold lower. The side chain of Arg49 binds to the "nontransferring" phosphoryl group (positioned at the entrance of the core domain active site) of the sugar−phosphate substrate. This interaction is important to substrate binding and domain−domain closure. The k_{cat} value of the R49A mutant is 200-fold lower than that of wild-type enzyme, while the k_{cat}/K_m value is 1 000 000-fold lower. Arg49 is therefore essential to efficient β-PGM catalysis.

Kinetic analysis of the β-PGM mutants formed by Ala, Val, and Pro substitution at the Gly46 confirmed the importance of Gly at the turn position of the helix−turn−

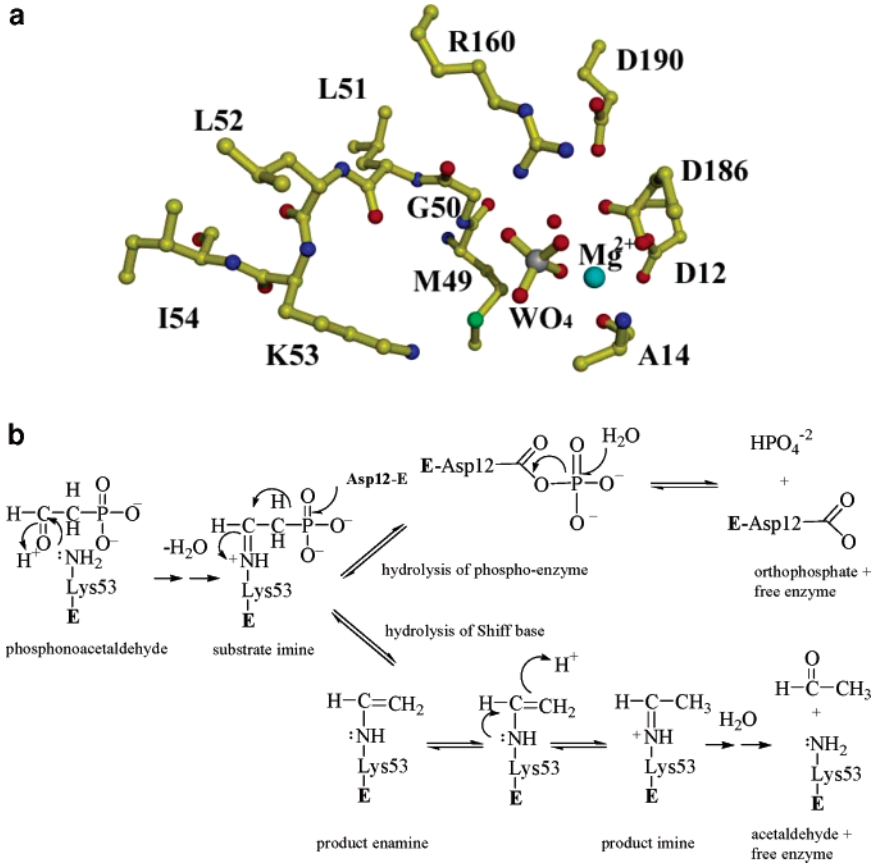

FIGURE 5: Interactions of residues of loop 5 of wild-type phosphonatase with the core domain (a) active site with tungstate bound (closed conformation) The tungsten atom is silver and Mg^{2+} is shown in as a cyan ball. (b) Chemdraw of the proposed reaction pathway of phosphonatase.

helix. The k_{cat} values of wild-type, G46A, G46V, and G46P β-PGM are 17 s^{-1}, 0.018 s^{-1} (1000-fold decrease), 0.005 s^{-1} (3000-fold decrease), and 0.13 s^{-1} (100-fold decrease), while the k_{cat}/K_m values are 4×10^6 M^{-1} s^{-1}, 9 M^{-1} s^{-1} (1 000 000-fold decrease), 3×10^2 M^{-1} s^{-1} (10 000-fold decrease) and 6×10^1 M^{-1} s^{-1} (100 000-fold decrease). Unlike the phosphonatase G50V and G50P mutants, the β-PGM G46V and G46P mutants, along with the G46A mutant retained a low level of catalytic activity. It is noteworthy that phosphonatase has positioned the essential Schiff-base forming Lys53 on loop 5, while loop 5 in β-PGM does not contain an essential catalytic residue (Figure 6b for mechanism). This may explain why phosphonatase is more sensitive to mutation at the Gly residue than is β-PGM.[2]

Structure of the Phosphonatase G50P Mutant: Amino Acid Replacement at the Hairpin Turn Disrupts Loop 5 Conformation. The crystal structure of the G50P mutant was determined to elucidate the impact of the Gly-to-Pro substitution at the helix−turn−helix cap motif and to understand the basis of substrate binding and chemical specificity imparted by the cap domain. The mutant was crystallized in the presence of Mg^{2+}, sulfate (from the crystallization conditions), and vso_3 (a competitive inhibitor with inhibition constant, $K_i = 180 \pm 20$ μM). The structure was determined to 2.8 Å resolution by molecular replacement using the wild-type phosphonatase-Mg^{2+}-tungstate structure (*1*).

Analysis of the refined final structure illustrated the same overall fold as that of the wild-type phosphonatase (Figure 7a) made of a core and a cap domain (*1*). The wild-type quaternary interactions between homodimers are maintained in the mutant structure. Hydrophobic subunit−subunit interactions are formed between the α8 helices of adjacent core domains. However, unlike the wild-type complex with tungstate, which comprises one open and one closed monomer, the G50P mutant dimers are either exclusively *closed* or *open*. No functional significance is yet attributed to this observed difference.

The "open" and "closed" mutant monomers superimpose well with the respective wild-type monomers (rmsd of Cα 1.49 and 1.21 Å, respectively). As in the wild type, the average B-factor of the cap domain is higher (B-factor = 41.9) than that of the core domain (B-factor = 26.6), indicating that the Gly50 replacement did not alter the overall flexibility of the two domains. The residues comprising the hinge/linker regions (23−26 and 90−94) between the two domains also remain unchanged (determined using the program DynDom (*29*)). In addition, calculation of the

[2] An unanticipated result from the kinetic analysis of the β-PGM G46 mutants is the ranking of catalytic efficiency. At saturating substrate concentration, the order of catalytic efficiency observed in the three β-PGM G46 mutants was β-PGM G46P > G46A > G46V, while measured at low substrate concentration, the activity order was β-PGM G46V > G46P > G46A. At first glance, these results would indicate that loop function in β-PGM is a seemingly complex interplay of steric/electrostatic interaction, because the relative activities of the mutants do not directly correlate with the degree of structural difference in Gly and its amino acid substitute. However, the activities of all three mutants are quite low, and a comparison between them (i.e., "one is less active than another") is probably not meaningful.

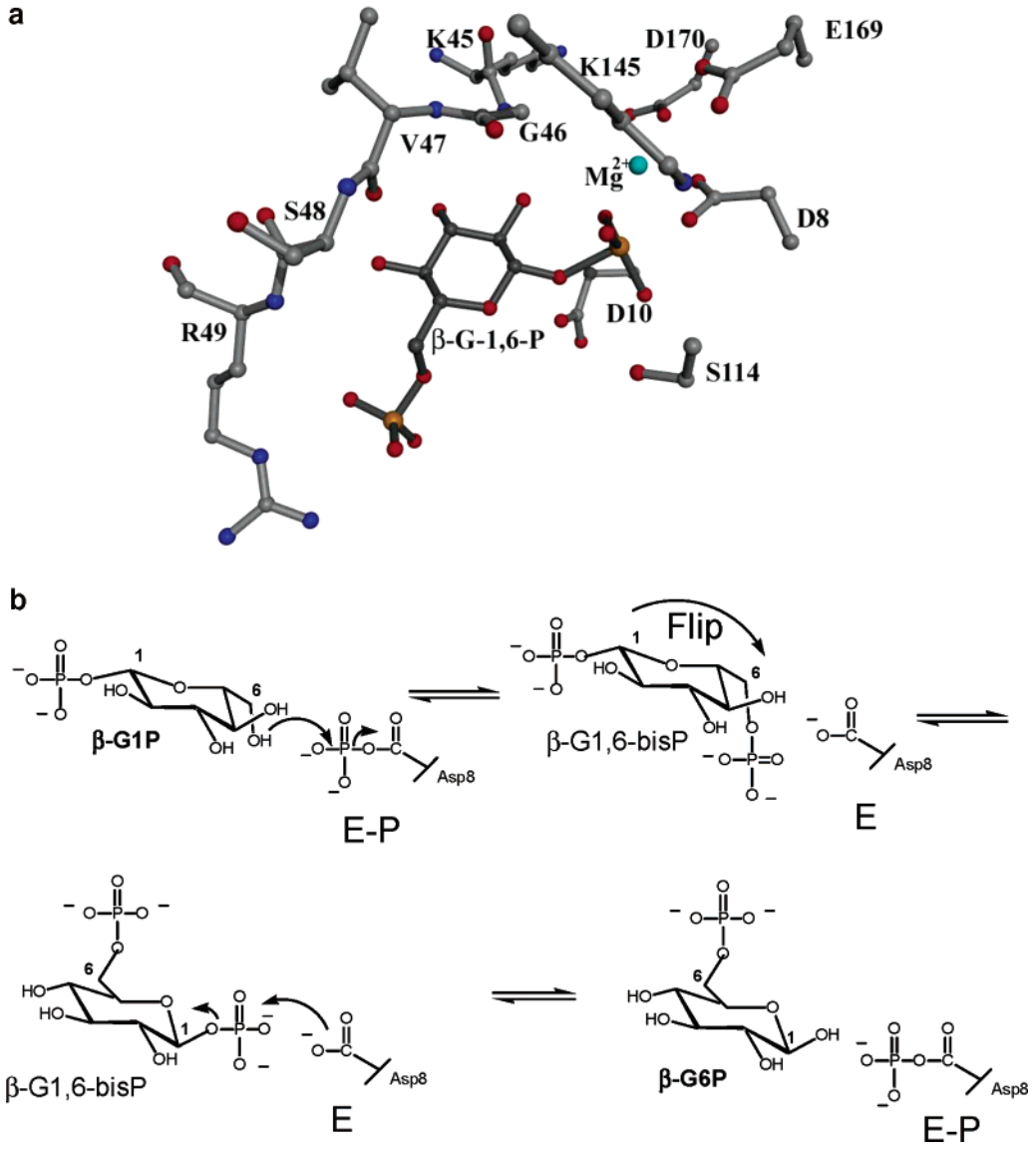

FIGURE 6: Interactions of residues of loop5 of wild-type β-PGM with the core domain. (a) Active site with intermediate (glucose 1,6-(bis)phosphate) bound (closed conformation) The phosphorus atoms are orange, and Mg^{2+} is shown as a cyan ball. (b) Chemdraw of the proposed reaction pathway of β-PGM.

Connolly surface using the program VOIDOO (*30*) shows an unaltered solvent accessibility profile for the active site; the *open* active site is solvent accessible and the *closed* is not. It was also possible, for the first time, to identify the seven C-terminal residues (261−267), which did not have distinct electron density in the wild-type phosphonatase-Mg^{2+}-tungstate structure.

At the interface of the cap and core domain lies the active site (see stereoview in Figure 8), which can easily be identified by the conserved residues and by the presence of the inhibitor vso_3 and magnesium (present in the electron density map contoured at 4.0σ). An extensive hydrogen-bond network is present between vso_3, Mg^{2+}, and the active-site residues. The coordination sphere of Mg^{2+} is maintained as in wild-type phosphonatase. The ligands include the backbone carbonyl of Ala14, carboxylates of Asp12 and Asp186, two oxygens of the sulfonate group, which mimic the oxygen atoms of orthophosphate (the product of phosphonatase catalysis), and a water molecule that is positioned by Asp190 and Thr187. The backbone amide nitrogen atoms of Ala14 and Gly127 along with a bidentate interaction from Arg160 of motif 3 position the three oxygens of vso_3. In addition, the Mg^{2+} and the water bound to Asp190 also contribute to sulfonate binding. The vinyl group is stabilized by hydrophobic interactions provided by Tyr128 and Cys22 from the core domain, and the Met49 and Leu52 from the cap domain. The interaction with Leu52 is only seen in the mutant structure and is replaced by Lys53 in the wild-type enzyme. Overall, the conformation of all critical residues of the core domain comprising the active site remain unchanged and all of the protein−ligand contacts between the wild-type enzyme core, ligand, and Mg^{2+} are preserved in the G50P mutant.

In contrast, the residues contributed by the cap domains move significantly as a result of the G50P mutation. The active-site residues from the cap domain in the wild-type structure include Cys22, Pro25, Phe29, Met49, Lys53, His56, Phe90, and Leu94. The majority of the residues originating from the cap domain are hydrophobic, and thus, stabilize the deprotonated form of the Schiff-base forming Lys and serve to provide a solvent-excluded active site in the closed form of the enzyme (*1*). Comparison of the closed conformation of wild-type phosphonatase with the closed conformation

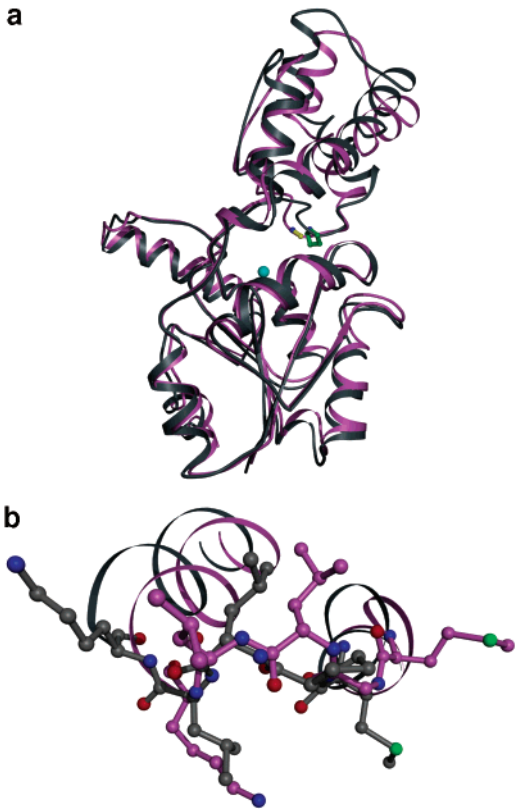

FIGURE 7: Superposition of the wild-type (gray) and G50P mutant (pink) phosphonatase structures (a) entire Cα-trace and (b) the helix−loop−helix structure of the cap domain. The amino acids at position 50, the sulfate, and the nucleophilic Asp are shown as ball-and-stick.

of G50P indicates that there are major conformational changes in these residues. In the closed conformation of wild-type phosphonatase, Lys53, surrounded by the side chains of the hydrophobic residues Pro25, Phe29, and Met49 from the cap domain and Phe90 and Leu94 from the core domain, is solvent inaccessible and positioned opposite the conserved nucleophile Asp12. The same is true in the open form, except that the active-site cleft is wider and the Lys is solvent accessible. In the mutant structure, a change in the position of Lys53 is seen, because the G50P mutation disrupts the helix−turn−helix conformation (residues 46−56) in the cap region (Figure 7b). As a result, the position of the alpha carbon of the nucleophilic Lys53 has shifted by 4 Å in the mutant enzyme as compared to the wild-type enzyme and is no longer in the active site. The side chain of Leu52 has replaced the Lys53 position, thus making the cap portion of the active site entirely hydrophobic. The Leu52 is in closer proximity to Met49 and the imidazole ring of His 56. Met49, which serves to "seal" the active site in the closed form of the enzyme, has also moved by 3.4 Å from its original position. The carbonyl oxygen of Pro48 moves within hydrogen-bonding distance (3.4 Å) of the His56 ring nitrogen, which in turn is now closer to the active site ligand by ∼2 Å. Overall, the displacement of Lys53 followed by reorientation of Leu52 has resulted in a major misalignment of the nucleophilic lysine with respect to the substrate and a disruption of the active-site environment. The hydrophobic Leu52 is accommodated the hydrophobic active site, which by design, compliments the unprotonated or neutral Lys53 side chain.

The large positional change in Lys53, Leu52, and Met49 can be attributed to a significant backbone conformational change in the helix−turn−helix region of the cap domain of G50P. In other mutants, such as G50A and G50V, the conformational change in loop 5 may not be as large as seen in the G50P mutant, but the low activity of all mutants suggests that the absence of a flexible Gly between critical active site residues drastically hampers catalysis. This is because the phi/psi angles adopted by Gly50 in the wild-type structure (53°/−116° and 132°/−37° for the open and closed conformations, respectively) would be sterically hindered for the amino acids used to replace it—Pro, Val, and Ala in this study. However, the structure of G50P demonstrates that for the Pro mutant the complete loss of activity is not due to a misfolding of the protein but it is, instead, due to a much more subtle repositioning of important residues in the active site Thus, the role of Gly of loop 5 is not simply important for the structure but to allow the flanking active site residues to be positioned appropriately for catalysis.

CONCLUSIONS

In this study, we have demonstrated that catalysis within the core domains of the two HAD family members phosphonatase and β-PGM requires the participation of loop 5 of the corresponding cap domains. On the basis of the detrimental effect of Gly substitution on loop 5 conformation and on catalytic efficiency, it is evident that the Gly is an indispensable component of this catalytic unit. Because loop 5 is tailored toward catalysis of the specialized chemistry unique to a subfamily, rather than catalysis of the core

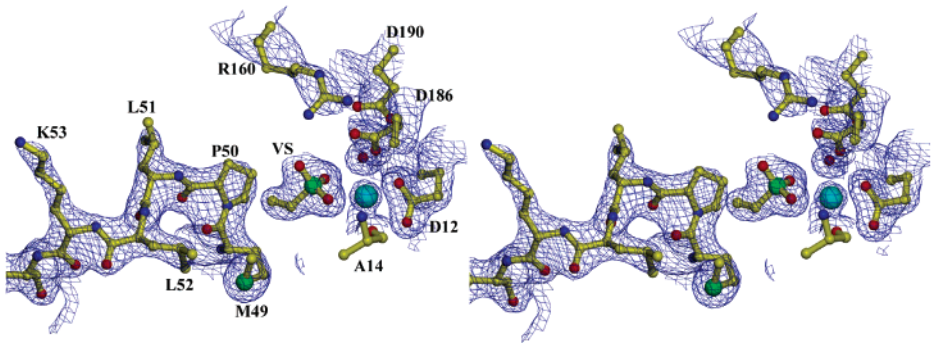

FIGURE 8: Stereodiagram of the active site of G50P with the electron density surrounding the mutation site. The active site residues and ligands vso_3 (orange) and cofactor Mg^{2+} (cyan) are depicted as ball-and -stick. The $2F_o - F_c$ electron density map (contoured at 1.5σ) is depicted as blue cages.

chemistry common to all subfamilies, its sequence (as with the rest of the cap domain sequence) is highly variable between subfamilies. Therefore, it has been difficult to recognize the loop 5 motif within sequences belonging to unknown subfamilies, based on sequence alignments alone. Using sequence motifs 1 and 2 to define the boundaries of the cap sequence segment, and using the conserved Gly to identify the loop 5 motif within this segment, it is possible to identify cap domain-derived active site residues, in the absence of a three-dimensional structure of the HAD protein. This should prove to be an effective new tool in function assignment within the vast, uncharted landscape of the HAD enzyme superfamily.

ACKNOWLEDGMENT

We wish to thank Dr. Ezra Peisach for helpful suggestions on solving the structure via molecular replacement.

REFERENCES

1. 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*, 10385−96.
2. Ridder, I. S., Rozeboom, H. J., Kalk, K. H., and Dijkstra, B. W. (1999) Crystal structures of intermediates in the dehalogenation of haloalkanoates by L-2-haloacid dehalogenase. *J. Biol. Chem. 274*, 30672−8.
3. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2002) Caught in the act: the structure of phosphorylated beta-phosphoglucomutase from *Lactococcus lactis*. *Biochemistry 41*, 8351−9.
4. 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 alpha/beta hydrolase structure that is different from the alpha/beta hydrolase fold. *J. Biol. Chem. 271*, 20322−30.
5. Argiriadi, M. A., Morisseau, C., Hammock, B. D., and Christianson, D. W. (1999) Detoxification of environmental mutagens and carcinogens: structure, mechanism, and evolution of liver epoxide hydrolase. *Proc. Natl. Acad. Sci. U.S.A. 96*, 10637−42.
6. Toyoshima, C., Nakasako, M., Nomura, H., and Ogawa, H. (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature 405*, 647−55.
7. Wang, W., Kim, R., Jancarik, J., Yokota, H., and Kim, S. H. (2001) Crystal structure of phosphoserine phosphatase from Methanococcus jannaschii, a hyperthermophile, at 1.8 Å resolution. *Structure (London) 9*, 65−71.
8. Parsons, J. F., Lim, K., Tempczyk, A., Krajewski, W., Eisenstein, E., and Herzberg, O. (2002) From structure to function: YrbI from Haemophilus influenzae (HI1679) is a phosphatase. *Proteins 46*, 393−404.
9. Galburt, E. A., Pelletier, J., Wilson, G., and Stoddard, B. L. (2002) Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase. *Structure (London) 10*, 1249−60.
10. Selengut, J. D. (2001) MDP-1 is a new and distinct member of the haloacid dehalogenase family of aspartate-dependent phosphohydrolases. *Biochemistry 40*, 12704−11.
11. Olsen, D. B., Hepburn, T. W., Moos, M., Mariano, P. S., and Dunaway-Mariano, D. (1988) Investigation of the *Bacillus cereus* phosphonoacetaldehyde hydrolase. Evidence for a Schiff base mechanism and sequence analysis of an active-site peptide containing the catalytic lysine residue. *Biochemistry 27*, 2229−34.
12. Olsen, D. B., Hepburn, T. W., Lee, S. L., Martin, B. M., Mariano, P. S., and Dunaway-Mariano, D. (1992) Investigation of the substrate binding and catalytic groups of the P−C bond cleaving enzyme, phosphonoacetaldehyde hydrolase. *Arch. Biochem. Biophys. 296*, 144−51.
13. Qian, N., Stanley, G. A., Hahn-Hagerdal, B., and Radstrom, P. (1994) Purification and characterization of two phosphoglucomutases from *Lactococcus lactis* subsp. lactis and their regulation in maltose- and glucose-utilizing cells. *J. Bacteriol. 176*, 5304−11.
14. Qian, N., Stanley, G. A., Bunte, A., and Radstrom, P. (1997) Product formation and phosphoglucomutase activities in *Lactococcus lactis*: cloning and characterization of a novel phosphoglucomutase gene. *Microbiology 143*, 855−65.
15. Baker, A. S., Ciocci, M. J., Metcalf, W. W., Kim, J., Babbitt, P. C., Wanner, B. L., Martin, B. M., and Dunaway-Mariano, D. (1998) Insights into the mechanism of catalysis by the P−C bond-cleaving enzyme phosphonoacetaldehyde hydrolase derived from gene sequence analysis and mutagenesis. *Biochemistry 37*, 9305−15.
16. Lahiri, S. D., Zhang, G., Radstrom, P., Dunaway-Mariano, D., and Allen, K. N. (2002) Crystallization and preliminary X-ray diffraction studies of beta-phosphoglucomutase from Lactococcus lactus. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 324−6.
17. 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*, 248−54.
18. Morais, M. C., Baker, A. S., Dunaway-Mariano, D., and Allen, K. N. (2000) Crystallization and preliminary crystallographic analysis of phosphonoacetaldehyde hydrolase. *Acta Crystallogr., Sect. D: Biol Crystallogr. 56*, 206−9.
19. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol. 276*, 307−26.
20. Rossmann, M. G. (1990) The molecular replacement method. *Acta Crystallogr., Sect. A: Found. Crystallogr. 46 (Pt 2)*, 73−82.
21. Morais, M. C., Zhang, W., Zhang, G., Olsen, D. B., Dunaway-Mariano, D., and Allen, K. N. (2003) X-ray Structure and Site-directed Mutagenesis Study of the Mechanism of Schiff-Base Formation in Phosphonoacetaldehyde Hydrolase Catalysis. *J. Biol. Chem.*, in press.
22. Navaza, J. (2001) Implementation of molecular replacement in AMoRe. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 57*, 1367−72.
23. 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., Sect. D: Biol. Crystallogr. 54*, 905−21.
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. Read, R. J. (1997) Model phases: probabilities and bias. *Methods Enzymol. 278*, 110−28.
26. 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., Sect. A: Found. Crystallogr. 47*, 110−9.
27. 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*, 283−91.
28. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−50.
29. Hayward, S., and Lee, R. A. (2002) Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Mol. Graphics Modell. 21*, 181−3.
30. Kleywegt, G. J., and Jones, T. A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 50*, 178−85.

BI0356810