1082 *Biochemistry* **2010,** *49,* 1082–1092
DOI: 10.1021/bi902019q

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Structural Determinants of Substrate Recognition in the HAD Superfamily Member D-*glycero*-D-*manno*-Heptose-1,7-bisphosphate Phosphatase (GmhB)[†,‡]

Henry H. Nguyen,[§] Liangbing Wang,[∥,@] Hua Huang,[∥] Ezra Peisach,[⊥] Debra Dunaway-Mariano,*[,∥] and Karen N. Allen*[,§,⊥]

[§]*Department of Physiology and Biophysics, Boston University School of Medicine, Boston, Massachusetts 02118-2394,* [∥]*Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico 87131, and* [⊥]*Department of Chemistry, Boston University, Boston, Massachusetts 02215.* [@]*Current address: Chengdu Di'Ao Pharmaceutical Co., Institute of Medicine, Sichuan, China.*

Received November 24, 2009; Revised Manuscript Received December 31, 2009

ABSTRACT: The haloalkanoic acid dehalogenase (HAD) enzyme superfamily is the largest family of phosphohydrolases. In HAD members, the structural elements that provide the binding interactions that support substrate specificity are separated from those that orchestrate catalysis. For most HAD phosphatases, a cap domain functions in substrate recognition. However, for the HAD phosphatases that lack a cap domain, an alternate strategy for substrate selection must be operative. One such HAD phosphatase, GmhB of the HisB subfamily, was selected for structure−function analysis. Herein, the X-ray crystallographic structures of *Escherichia coli* GmhB in the apo form (1.6 Å resolution), in a complex with Mg^{2+} and orthophosphate (1.8 Å resolution), and in a complex with Mg^{2+} and D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate (2.2 Å resolution) were determined, in addition to the structure of *Bordetella bronchiseptica* GmhB bound to Mg^{2+} and orthophosphate (1.7 Å resolution). The structures show that in place of a cap domain, the GmhB catalytic site is elaborated by three peptide inserts or loops that pack to form a concave, semicircular surface around the substrate leaving group. Structure-guided kinetic analysis of site-directed mutants was conducted in parallel with a bioinformatics study of sequence diversification within the HisB subfamily to identify loop residues that serve as substrate recognition elements and that distinguish GmhB from its subfamily counterpart, the histidinol-phosphate phosphatase domain of HisB. We show that GmhB and the histidinol-phosphate phosphatase domain use the same design of three substrate recognition loops inserted into the cap domain yet, through selective residue usage on the loops, have achieved unique substrate specificity and thus novel biochemical function.

The haloalkanoic acid dehalogenase (HAD)[1] enzyme superfamily (*7*, *8*) is Nature's major supplier of organophosphate metabolite phosphohydrolases (phosphatases). The HAD phosphatase catalytic site is housed in the Rossmann-like fold core domain and is comprised by a Mg^{2+} cofactor, an Asp nucleophile, an Asp acid/base, and two conserved hydrogen bond donors (Thr/Ser and Lys/Arg) which function in collaboration with main chain amide units to bind the substrate phosphoryl group (Figure 1) (*9*). The phosphoryl group is transferred to the Asp nucleophile to form an aspartylphosphate intermediate, which is subsequently dephosphorylated by attack of an activated water molecule (Scheme 1). The catalytic site provides only weak binding interaction with the substrate phosphoryl group in the ground state and reserves strong interaction for stabilization of the trigonal bipyramidal phosphorane-like transition states (*10*−*12*). With the exception of the catalytic Asp acid residue, which donates a hydrogen bond to the oxygen atom that bridges the organic unit of the leaving group to the phosphorus atom, there is virtually no interaction between the leaving group and the residues that form the catalytic pocket (Figure 1). Thus, the structural elements that provide the additional binding interactions necessary to secure the physiological substrate are separate from the structural elements of catalysis. Consequently, changes needed for adaptation to a novel substrate can be made to the constellation of residues that interact with the substrate leaving group (i.e., the substrate specificity residues) without perturbing the catalytic machinery. The autonomy of the structural elements of substrate recognition has allowed the successful adaptation of the HAD phosphatases to a wide variety of biochemical niches (*13*).

Inspection of the numerous HAD family sequences and structures reveals a multiplicity of ways in which the basic

[†]This work was supported by National Institutes of Health Grant GM61099 to K.N.A. and D.D.-M. Financial support for beamlines X12B, X25C, and X29 at Brookhaven National Laboratory comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the U.S. Department of Energy and the National Center for Research Resources of the National Institutes of Health.

[‡]Coordinates for the X-ray structures of *Bordatella bronchiseptica* D-*glycero*-D-*manno*-heptose-1,7-bisphosphate phosphatase in a complex with Mg^{2+} and phosphate and *Escherichia coli* D-*glycero*-D-*manno*-heptose-1,7-bisphosphate phosphatase in the apo form, in a complex with Mg^{2+} and phosphate, and in a complex with D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate have been deposited in the Protein Data Bank as entries 3L8H, 3L8E, 3L8F, and 3L8G, respectively.

*To whom correspondence should be addressed. D.D.-M.: e-mail, dd39@unm.edu; phone, (505) 277-3383; fax, (505) 277-2609. K.N.A.: phone, (617) 358-5544; fax, (617) 358-5554; e-mail, drkallen@bu.edu.

[1]Abbreviations: H1β,7bisP, D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate; H1α,7bisP, D-*glycero*-D-*manno*-heptose 1α,7-bisphosphate; GmhB, D-*glycero*-D-*manno*-heptose-1,7-bisphosphate phosphatase; HAD, haloalkanoic acid dehalogenase; HisB, bifunctional histidinol-phosphate phosphatase/imidazole-glycerol-phosphate dehydratase; SeMet, selenomethionine; KDOP-8-P, 2-keto-3-deoxyoctulosonate 8-phosphate; KDN-9-P, 2-keto-3-deoxy-D-glycero-D-galactonononate 9-phosphate; DTT, dithiothreitol; PDB, Protein Data Bank.

 Published on Web 01/05/2010

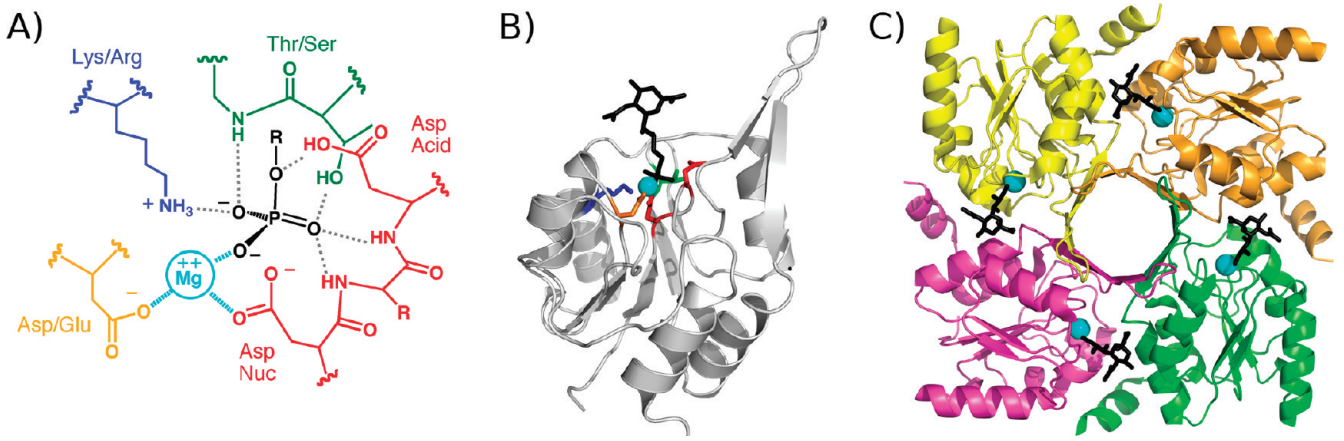
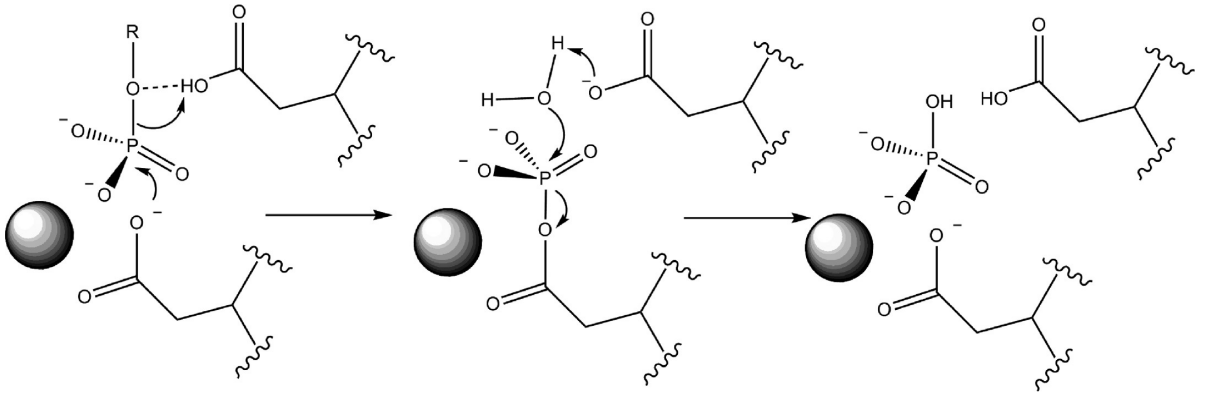


FIGURE 1: (A) Diagram of the HAD catalytic template depicting the binding interactions among the substrate (black), Mg^{2+} cofactor (cyan), and conserved catalytic residues of the HAD phosphatase active site. (B) Structure of the 2-keto-3-deoxy-D-glycero-D-galactonononate-9-phosphate phosphatase monomer bound with Mg^{2+} (cyan), vanadate, and 2-keto-3-deoxy-5-acetylamine-D-glycero-D-galactonononate 9-phosphate (black) (*23*). The conserved catalytic residues are colored as in panel A. (C) Tetramer of the 2-keto-3-deoxy-D-glycero-D-galactonononate-9-phosphate phosphatase tetramer depicted as ribbons. The subunits are individually colored and ligands colored as in panel B.

Scheme 1: General Catalytic Mechanism for Phosphohydrolase Members of the HAD Superfamily with the Mg^{2+} Cofactor Depicted as a Sphere



Rossmann-like scaffold has been elaborated, through the course of evolution, with sequence inserts that function in substrate recognition (*14*, *15*). For the vast majority of the HAD phosphatases, the added sequence forms a separate folding domain known as the cap domain. The cap domain provides a large surface area for interaction with the leaving group of the phosphorylated metabolite, which in turn contributes to the association of the cap domain and catalytic domain in preparation for catalytic turnover (Figure 2) (*16*−*18*). The "capless" HAD phosphatases, on the other hand, are well suited to function in the removal of phosphoryl groups from phosphorylated macromolecules because such substrates require a fully accessible active site (*19*−*21*), yet remarkably, some capless HAD phosphatases are known to be selective for small substrates (*22*, *23*). Structure−function analysis of these HAD members is underway in our laboratories to determine their mechanisms of substrate recognition.

D-*glycero*-D-*manno*-Heptose-1,7-bisphosphate phosphatase (GmhB) is a member of the HAD HisB subfamily. The subfamily is named after the N-terminal domain of the bifunctional histidine biosynthetic pathway enzyme histidinol-phosphate phosphatase/imidazole-glycerol-phosphate dehydratase known as HisB. The structure of the *Escherichia coli* histidinol-phosphate phosphatase domain of HisB shows that it is capless and that it possesses a novel Zn^{2+} binding loop (Figure 2) (*24*). However, because the structure is that of one single domain of a two-domain protein, the interactions between the two domains are not known and it is not certain if or how the C-terminal domain may interact with the phosphatase active site.

The fact that GmhB is capless and monomeric, and shows high catalytic efficiency ($k_{cat}/K_m \sim 1 \times 10^7$ M^{-1} s^{-1}) and substrate specificity [see the preceding paper (*25*)], suggests that it might employ a unique strategy in substrate binding compared to members of other subfamilies of capless HAD phosphatases. Herein, we report the apo and liganded structures of GmhB from *E. coli* K-12 and *Bordetella bronchiseptica* and present the results from a probe of the structural elements of substrate recognition via site-directed mutagenesis. These findings, together with information derived from bioinformatics analysis, shed new light on the structural basis for the biochemical divergence of GmhB biochemical function among the HAD phosphatases.

EXPERIMENTAL PROCEDURES

X-ray Structure Determination. The *B. bronchiseptica* GmhB (SwissProt accession number Q7WG29) was prepared as described in the preceding paper (*25*). The selenomethionine-substituted *E. coli* K-12 GmhB (SwissProt accession number P63288) was prepared using a methionine auxotroph of *E. coli* (B-834) grown in a defined medium containing 40 mg/L selenomethionine and purified using the same protocol that was used

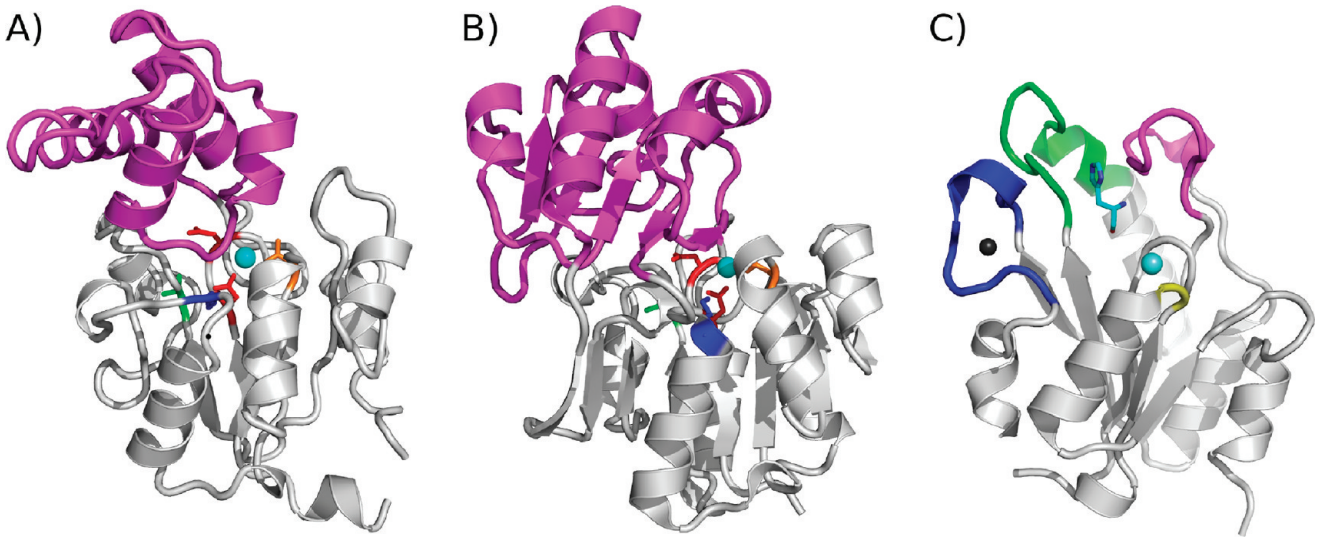


FIGURE 2: (A) Ribbon diagram of C1 HAD phosphatase Q8A5V9 (S. Almo, Albert Einstein College of Medicine, Bronx, NY, unpublished data). The Rossmann fold catalytic domain is colored gray, and conserved catalytic residues are shown as sticks and colored as in Figure 1A. The Mg^{2+} is a cyan sphere, and the cap domain is colored magenta. (B) Structure of C2 HAD phosphatase BT4131 (*39*). (C) Structure of the histidinol-phosphate phosphatase domain of *E. coli* HisB (*24*). The histidinol ligand (shown as sticks) is colored cyan. The Zn^{2+} and Mg^{2+} are depicted as black and cyan spheres, respectively. The Zn^{2+} binding loop (blue), loop 1 (magenta), loop 2 (green), and the Mg^{2+} binding loop (yellow) are colored to coordinate with the structures shown in Figures 4 and 7.

for the native enzyme (*25*). The crystallization conditions were screened at 25 °C using the Hampton Index Screen via the hanging-drop vapor-diffusion method. The optimal conditions found for *B. bronchiseptica* GmhB crystallization are 10 mg/mL protein, 10 mM fructose 1,6-bisphosphate, 0.05 M magnesium formate, 100 mM sarcosine, 5% dioxane, and 25% PEG 3350 (25 °C). The crystal was protected with paratone-N before being flash-frozen in a stream of N_2 gas at 100 K. The optimal conditions found for *E. coli* GmhB crystallization are 25 mg/mL protein, 0.1 M Tris (pH 7.5), 5 mM $MgCl_2$, and 25% PEG 3350 (25 °C). For the apo data set, the GmhB crystal was transferred to a solution consisting of 0.1 M Tris (pH 7.5), 5 mM $MgCl_2$, 35% PEG 3350, and 10% glycerol for 5 min and then transferred to the mother liquor with 20% glycerol for 5 min before being flash-frozen in a gaseous stream of N_2 at 100 K. For *E. coli* GmhB in a complex with orthophosphate, the crystal of the apoenzyme was transferred to a cryo-protecting solution containing 0.1 M Tris (pH 7.5), 5 mM $MgCl_2$, 5 mM imidodiphosphate, 25% PEG 3350, and 20% glycerol for 5 min and flash-frozen in a stream of N_2 gas at 100 K. The complex of GmhB with D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate (GmhB−Mg^{2+}−H1β,7bisP) was obtained by soaking crystals of the apoenzyme in a 10 μL drop consisting of 0.1 M Tris (pH 7.5), 5 mM $MgCl_2$, 35% PEG 3350, 30 mM D-*glycero*-D-*manno*-heptose 1β-phosphate, and 30 mM D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate for 15 min. The crystal was protected with paratone-N before being flash-frozen in a stream of N_2 gas at 100 K.

Data were also collected at the National Synchrotron Light Source at the Brookhaven National Laboratory (Upton, NY). The diffraction data set was collected on a Rigaku RU300 generator equipped with a Raxis IV^{++} area detector. All diffraction data were processed in the DENZO/SCALEPACK program package (*26*). The data set for *E. coli* apo GmhB was phased using one subunit of histidinol-phosphate phosphatase (*24*) (PDB entry 2FPR; 28% identical in sequence to GmhB) as the search model for molecular replacement in MOLREP (*27*) of the CCP4 program package. (*28*). The initial model was refined by manual rebuilding in COOT (*29*) with alternating rounds of refinement in CNS (*30*). The quality of the final model was assessed with MOLPROBITY (*31*). The *E. coli* apo GmhB structure was used as the search model for molecular replacement to phase the liganded *E. coli* GmhB structures (which crystallized in a different space group) as well as the *B. bronchiseptica* GmhB structure. Refinement of the *B. bronchiseptica* GmhB data set was debiased by including the phases from selenomethionine. The data collection and refinement statistics are summarized in Table 1.

Native Molecular Mass Determination. The molecular masses of *E. coli and B. bronchiseptica* GmhB were determined using gravity-flow gel-filtration chromatography. Specifically, the enzyme was chromatographed at 4 °C on a calibrated (Pharmacia Gel Filtration Calibration Kit) 1.5 cm × 180 cm Sephacryl S-200 column (Pharmacia) using a 50 mM K^+HEPES/5 mM $MgCl_2$/1 mM DTT mixture (pH 7.0 at 25 °C), at a flow rate of 1 mL/min, as the eluant. The molecular mass was determined from a plot of log(molecular mass) of standard proteins versus the elution volume.

Preparation of Site-Directed Mutants of E. coli GmhB. Mutagenesis was conducted using a PCR-based strategy with commercial primers and the wild-type GmhB/pET-23b plasmid serving as a template. The purified PCR products were used to transform BL21(DE3) competent cells. The mutant gene sequences were verified by DNA sequencing. The mutant proteins were prepared in the same manner that was used for wild-type GmhB (*25*).

Ligand Docking Models. Parameters for bond length, angles, and van der Waals interactions for the α- and β-anomers of D-*glycero*-D-*manno*-heptose 1,7-bisphosphate substrates were generated on the basis of ideal models. Substrates were placed in the active site such that the phosphoryl group overlaid with the phosphoryl group position in the GmhB complex. Active site waters were removed from the model. Energy minimization of the complex was conducted in CNS (*30*), utilizing only bond parametrization and van der Waals contacts.

DNA Binding Assay. DNA binding was assessed via gel-shift assays conducted according to known procedures (*32*). Solutions containing 0.5−1.0 mM GmhB and 1 molar equiv of

Table 1: Data Collection Statistics[a]

	Ec GmhB apo	Ec GmhB−Mg^{2+}−P_i	Ec GmhB−Mg^{2+}−H1β,7bisP	Bb GmhB−Mg^{2+}−P_i
space group	$P2_12_12_1$	$P2_12_12$	$P2_12_12$	$P1$
unit cell dimensions (Å)	a = 51.77	a = 64.48	a = 63.67	a = 38.1
	b = 63.97	b = 50.00	b = 51.57	b = 58.4
	c = 103.20	c = 51.88	c = 52.19	c = 85.7
X-ray source	X12B, NSLS	X25C, NSLS	R AxisIV^{++}	X29, NSLS
wavelength (Å)	0.9786	1.1000	1.5418	1.0000
resolution (Å)	1.64 (1.64−1.70)	1.8 (1.79−1.85)	2.2 (2.18−2.26)	1.7 (1.68−1.74)
total no. of reflections/no. of unique reflections	207847/42543	109143/16358	32412/9183	79124/6443
completeness (%)	99.4 (98.4)	99.9 (99.6)	97.2 (99.0)	94.6 (77.3)
$I/\sigma(I)$	16.7 (2.5)	33.5 (3.7)	10.1 (1.9)	18.4 (3.2)
R_{merge}	0.072 (0.491)	0.048 (0.501)	0.073 (0.427)	0.059 (0.253)
no. of molecules per asymmetric unit	2	1	1	4
no. of ordered/total residues	368/382	182/191	182/191	716/716
no. of nonsolvent atoms per asymmetric unit	2895	1433	1457	5386
no. of reflections (work/free)	76862/7540	26258/2844	8332/420	147262/14317
R_{work}/R_{free}	0.17/0.21	0.19/0.23	0.18/ 0.25	0.17/0.20
total no. of waters	378	110	79	666
no. of Zn^{2+}/Mg^{2+}	2/0	1/1	1/1	4/5
no./ligand	2/acetates	none	1/H1β,7bisP	5/P_i, 1/fructose
average B factor (Å^2)				
amino acid residues	15.8	31.3	31.1	19.2
Mg^{2+} ions	−	35.2	21.5	17.3
Zn^{2+} ions	12.7	34.6	33.8	20.0
water	26.0	41.2	33.6	26.1
ligands	12.7	39.2	42.9	23.1
root-mean-square deviation				
bond lengths (Å)	0.006	0.028	0.010	0.008
bond angles (deg)	1.062	2.258	1.268	1.127
Ramachandran favored/allowed (%)	98.1/100	97.2/100	96.7/100	97.6/100

[a]Values in parentheses are for the highest-resolution shell.

calf thymus DNA (Invitrogen) in a 50 mM Tris/100 mM NaCl/2 mM $MgCl_2$ mixture (pH 7.5) were incubated at 25 °C for 30 min. Loading buffer (5×) [50% glycerol and 30 mM Tris (pH 6.8)] was added, and the mixture was subjected to chromatography on a 12% polyacrylamide gel using Tris/glycine running buffer [192 mM glycine and 25 mM Tris (pH 8.3)] for 1 h at 25 °C. SDS/PAGE gels were stained with R250 Coomassie blue to visualize the protein.

Bioinformatics. Sequence homology searches against nonredundant sequence databases were conducted using PSI Blast (http://blast.ncbi.nlm). Multiple alignments were determined using COBALT (http://www.ncbi.nlm.nih.gov/tools/cobalt/).

Steady-State Kinetics. Steady-state kinetic constant determinations were conducted at 25 °C using reaction solutions initially containing GmhB, varying concentrations of D-*glycero*-D-*manno*-heptose 1,7-bisphosphate (0.5−5K_m), 1 mM $MgCl_2$, 0.2 mM 2-amino-6-mercapto-7-methylpurine ribonucleoside, 1 unit of purine nucleoside phosphorylase, and 50 mM Tris-HCl (pH 7.5). Reactions were continuously monitored for phosphate formation at 360 nm ($\Delta\varepsilon$ = 11 mM^{-1} cm^{-1}). Initial velocity data were fitted to the equation $V_0 = (V_{max}[S])/(K_m + [S])$ with KinetAsyst I. The k_{cat} value was calculated from V_{max} and [E] according to the equation $k_{cat} = V_{max}/[E]$, where [E] is the enzyme concentration.

RESULTS AND DISCUSSION

Structure of GmhB. *E. coli* GmhB and *B. bronchiseptica* GmhB were shown by size-exclusion column chromatography to have native molecular masses of ~16 and ~20 kDa, respectively, which by comparison to the respective calculated masses of 21294 and 19035 Da, indicates that both enzymes exist as monomers in solution.

X-ray structures were determined for *E. coli* GmhB in the apo form (no Mg^{2+} bound despite the inclusion of 5 mM $MgCl_2$ in the crystallization solution) at 1.6 Å resolution, in a complex with Mg^{2+} and orthophosphate at 1.8 Å resolution, and in a complex with Mg^{2+} and D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate at 2.2 Å resolution. In addition, the structure of *B. bronchiseptica* GmhB bound to Mg^{2+} and orthophosphate was determined to 1.7 Å resolution. Data collection and refinement statistics for each of the structures are reported in Table 1. All four structures reveal a Zn^{2+} bound at the predicted Zn^{2+} binding loop as evidenced by strong X-ray fluorescence at the peak wavelength (K_{edge}) of Zn^{2+}. The Zn^{2+} apparently copurified with the protein as no Zn^{2+} was added at any stage of the purification or during the crystallization procedure.

The crystalline *E. coli* GmhB−Mg^{2+}−phosphate complex was formed by soaking crystals of the apoenzyme in a buffered solution of $MgCl_2$ and imidodiphosphate. Activity assays revealed that *E. coli* GmhB hydrolyzes imidodiphosphate at a slow but detectable rate (estimated turnover rate of ~1 × 10^{-4} s^{-1}). Thus, the phosphate ligand observed in the X-ray crystal structure was generated in situ. In a similar fashion, the crystalline *B. bronchiseptica* GmhB−Mg^{2+}−phosphate complex was obtained by cocrystallization of the enzyme with $MgCl_2$ and fructose 1,6-bisphosphate. The fructose 1,6-bisphosphate is a slow substrate (k_{cat} = 0.55 ± 0.01 s^{-1}, and K_m = 680 ± 10 μM), and thus, the phosphate was generated by catalyzed hydrolysis of the fructose 1,6-bisphosphate. The crystalline *E. coli* GmhB−Mg^{2+}−D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate complex was formed by soaking crystals of the apoenzyme with $MgCl_2$

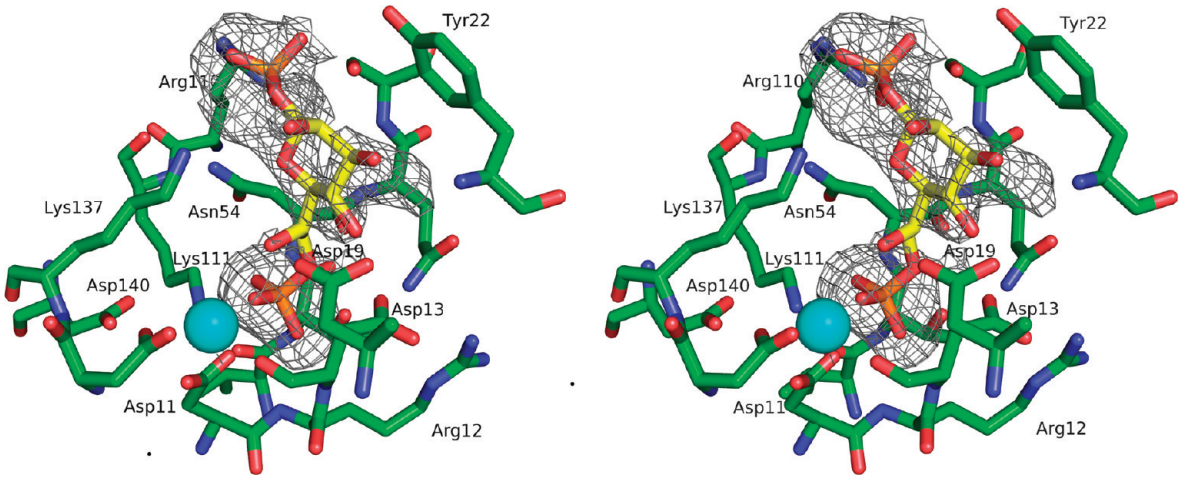


FIGURE 3: Stereoview of the active site of *E. coli* GmhB in a complex with Mg^{2+} (cyan sphere) and D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate. The simulated annealing $F_o - F_c$ electron density omit map (contoured at 2σ) is shown as a gray cage.

and a 1:1 mixture of D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate and D-*glycero*-D-*manno*-heptose 1β-phosphate (Figure 3).[2]

The *E. coli* and *B. bronchiseptica* GmhB structures reveal a single subunit protein, consistent with the molecular mass determined for the native enzymes in solution. The GmhB orthologs (Figure 4) possess the HAD enzyme superfamily catalytic domain, and as predicted from the sequence alignment data, they lack a cap domain. The GmhB backbone trace closely resembles that of the *E. coli* HisB histidinol-phosphate phosphatase domain (Figure 2) with a root-mean-square deviation (rmsd) between common Cα atoms of 1.86 Å. Superposition of the apo *E. coli* GmhB structure with the liganded structures (not shown) showed that ligand binding does not induce global or local conformational changes. The only significant change observed is the rotomer conformation of the Asp11 side chain, which in the liganded structure, coordinates to the Mg^{2+} cofactor.

Catalytic Scaffold. The catalytic scaffolds observed in the structures of the *E. coli* and *B. bronchiseptica* GmhB−Mg^{2+}−phosphate product complexes (Figure 5) comprise the Asp nucleophile (Asp11 and Asp7 in *E. coli* and *B. bronchiseptica*, respectively), Asp acid/base (Asp13 and Asp9, respectively), Thr and Lys phosphoryl-binding groups (Thr53 and Lys111 and Thr50 and Lys101, respectively), and the Asp ligand to the Mg^{2+} cofactor (Asp136 and Asp126, respectively). Together, this conserved constellation of residues comprises the phosphatase catalytic scaffold for the stabilization of the trigonal bipyramidal transition state (compare to Figure 1). In addition, the residue that performs the role of positioning the Asp acid/base in all HAD phosphatases is present in the two GmhB orthologs as Asn17 (2.7 Å from Asp13) and Asn13 (2.9 Å from Asp9) (Figure 5). The acid/base catalyst has been identified as being critical for catalysis in HAD phosphatases (*12*). The essential role of this residue is also supported in GmhB as demonstrated by the finding that the *E. coli* GmhB D13A mutant is devoid of detectable activity toward the physiological substrate D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate (Table 2).

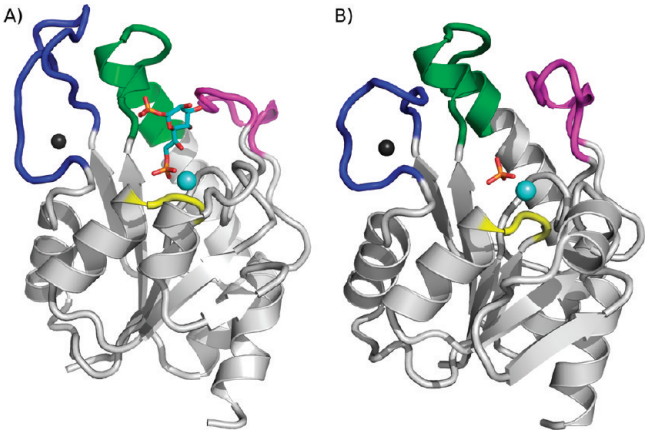


FIGURE 4: (A) Ribbon diagram of *E. coli* GmhB bound to Zn^{2+} and Mg^{2+} (black and cyan spheres, respectively) and the substrate D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate. The Zn^{2+} binding loop is colored blue, loop 2 green, loop 1 magenta, and the Mg^{2+} binding loop yellow. (B) Ribbon diagram of *B. bronchiseptica* GmhB bound to Zn^{2+} and Mg^{2+} and the product orthophosphate. The loops are colored as in Figure 2.

The phosphate ligand in the *E. coli* GmhB−Mg^{2+}−phosphate complex forms a coordination bond with Mg^{2+} (2.1 Å) and engages in a hydrogen bond interaction with the side chains of Thr53 (3.5 Å), Lys111 (2.5 Å), and Asp13 (3.4 Å) as well as the backbone amide NH group of Asp13 (3.1 Å). The phosphate ligand of the *B. bronchiseptica* GmhB−Mg^{2+}−phosphate complex forms a coordination bond with the Mg^{2+} (2.1 Å) and engages in a hydrogen bond interaction with the side chains of Thr50 (3.6 Å) and the backbone amide NH group of Asp9

[2]Structures of enzymes bound to substrate Michaelis complexes have been frequently observed by X-ray crystallography (*1*, *2*). This is seemingly in conflict with data showing that enzymes are active in the crystalline form (*3*, *4*). However, these data can be reconciled by considering the fact that the internal equilibria may differ in the crystalline enzyme versus the enzyme in solution. Because crystal structures generally do not detect species in low abundance (< 20%), a shift in the substrate:product equilibrium to 5:1 would lead to the observation of a substrate-bound structure. Alternatively, the dynamics of the protein are expected to be constrained in the crystal lattice compared to solution. To the extent that the enzyme transition state is formed by the instantaneous and optimal alignment of functional groups at the catalytic site during protein dynamic motion (*5*), the crystalline enzyme may not have sufficient mobility to attain the catalytic conformation (may not achieve sufficient energy of activation). The resulting complex does not represent an "off path" state but rather one on the reaction coordinate where collapse to the transition state may not proceed. In covalent catalysis as observed in HAD members, this may also include the case in which the nucleophile does not reach the near-attack conformation (*6*). Notably, this scenario is not similar to one in which a large conformational change such as loop closure or domain swiveling is precluded by the crystal lattice becuase the active site in such a complex would not be that of the active form in terms of bulk solvent exposure and/or amino acid composition.

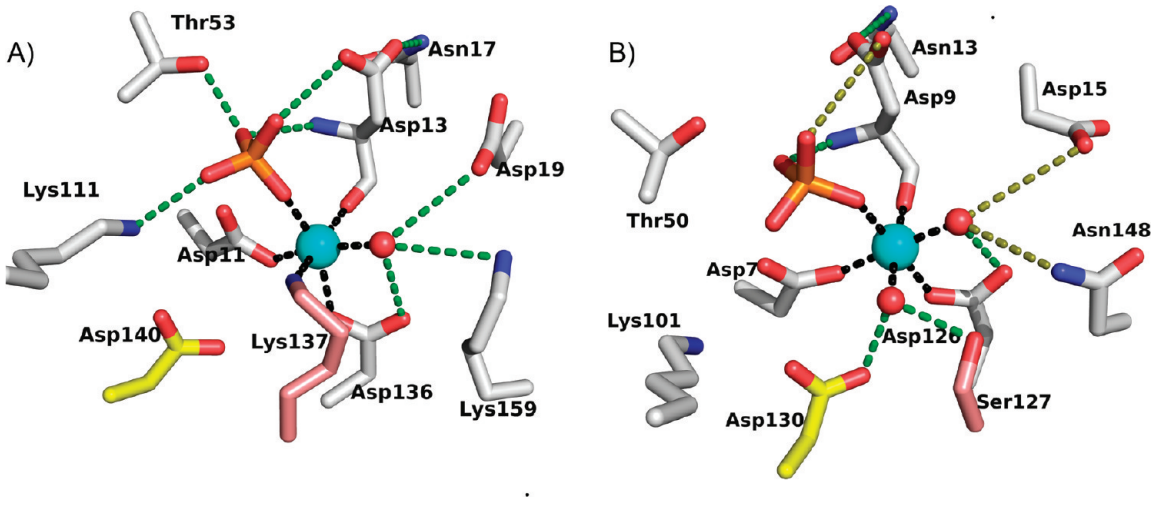


FIGURE 5: Active sites from the structures of *E. coli* GmhB−Mg^{2+}−phosphate (A) and *B. bronchiseptica* GmhB−Mg^{2+}−phosphate complexes (B). Metal coordination (black), hydrogen bonds [within 3.5 Å (green)], and long hydrogen bonds [3.6−3.9 Å (olive)] are represented by dashed lines. The Mg^{2+} is a cyan sphere, and water ligands are red spheres. The paired yellow and salmon residues occupy homologous positions in the two active sites.

Table 2: Steady-State Kinetic Constants for *E. coli* GmhB Mutant-Catalyzed Hydrolysis of the α-Anomer and/or β-Anomer of D-*glycero*-D-*manno*-Heptose 1,7-Bisphosphate in 50 mM Tris-HCl (pH 7.5, 25 °C) Containing 1 mM $MgCl_2$

GmhB	substrate anomer	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
wild type	β	35.7 ± 0.2	5.0 ± 0.1	7×10^6
wild type	α	4.6 ± 0.1	67 ± 1	7×10^4
D13A	β	$< 10^{-4}$	–	–
K137A	β	28.0 ± 0.5	39 ± 1	7×10^5
R110A	β	6.2 ± 0.1	181 ± 6	3×10^4
R110A	α	$< 10^{-4}$	–	–
C109A	β	0.34 ± 0.1	68 ± 4	5×10^3
C109A	α	0.057 ± 0.001	45 ± 2	1×10^3
C107A	β	0.63 ± 0.01	83 ± 4	8×10^3
C107A	α	0.042 ± 0.002	320 ± 20	1×10
C92A	β	0.63 ± 0.01	180 ± 6	3×10^4
C92A	α	0.0017 ± 0.0001	300 ± 20	5

(2.8 Å). The side chains of Lys101 (4.3 Å) and Asp9 (4.0 Å) do not form the expected hydrogen bonds with the phosphate ligand due to subtle differences in position of both the phosphate and side chains.

In phosphate complexes of HAD phosphatases (*33*, *34*), the six coordination sites of the Mg^{2+} cofactor are typically filled by two water ligands, one oxygen atom of the phosphate ligand, one oxygen atom from the Asp nucleophile carboxylate group, the backbone C=O group of the Asp acid/base, and one oxygen atom from the carboxylate group of an Asp/Glu residue located on the Mg^{2+} binding loop. The structure of the Mg^{2+} center of the *B. bronchiseptica* GmhB−Mg^{2+}−phosphate complex conforms to this model. Aside from the phosphate ligand, there are two water ligands (2.1 and 2.2 Å), the Asp7 carboxylate oxygen atom (2.2 Å), the Asp11 backbone carbonyl oxygen (2.1 Å), and the Asp126 carboxylate oxygen atom (2.0 Å) (Figure 5). The Mg^{2+} center of the *E. coli* GmhB−Mg^{2+}−phosphate complex, on the other hand, is unusual in that one coordination position is occupied by the side chain nitrogen atom (2.4 Å) of a Lys residue (Lys137) adjacent to the Asp136 (2.3 Å) ligand of the Mg^{2+} binding loop. The five coordination sites that remain are filled by a phosphate oxygen atom (2.0 Å), the Asp11 carboxylate oxygen atom (2.1 Å), the Asp13 backbone carbonyl oxygen (2.3 Å), and a water molecule (2.5 Å). The Lys137 residue is not stringently conserved among GmhB orthologs. In *B. bronchiseptica* GmhB, the corresponding residue is Ser127. As shown in Figure 5, Ser127 forms a hydrogen bond with the Mg^{2+} water ligand that assumes the same coordination position as Lys137 in the *E. coli* GmhB−Mg^{2+}−phosphate complex. In addition, Lys137 does not coordinate to the Mg^{2+} in the *E. coli* GmhB Michaelis complex (vide infra), nor does it make a significant contribution to catalysis (Table 2).

Binding Site for the Substrate Leaving Group: Collaboration among Three Loop Extensions. The GmhB catalytic site is elaborated by three extended peptide inserts, which pack to form a concave, semicircular surface around the substrate leaving group (Figure 6A). Together, with the bulky substrate leaving group, the three inserts shield the catalytic site from solvent [as demonstrated by using Voidoo (*35*)]. All three inserts project from the C-terminal end of the Rossmann-like fold. The first is an elongated loop, which follows the signature DxD catalytic motif containing the Asp nucleophile and Asp general acid/base residue. In the type C1 HAD phosphatases, an α-helical cap domain is inserted at this position. The second insert is a helix−turn motif that follows the conserved Thr/Ser located at the terminus of β-strand 2. The third insert is the Zn^{2+} binding loop, which precedes the conserved catalytic Lys. In the C2 HAD phosphatases, an α/β-fold cap domain is inserted at this position. Because the cap domains inserted in these positions provide substrate binding residues (*15*), it is not unreasonable to expect that the three loops in GmhB might do the same.

(*i*) *Zn^{2+} Binding Loop*. The Zn^{2+} binding loops (Figure 6B−D) of *E. coli* and *B. bronchiseptica* GmhB possess a classical **C**x**H**-(x)$_n$**C**x**C** motif that coordinates the Zn^{2+} with square planar geometry. The structure of the *E. coli* GmhB−Mg^{2+}−D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate complex shows that the Zn^{2+} does not interact with the substrate and, therefore, that it does not directly participate in catalysis. Our first impression was that the Zn^{2+} functions to stabilize the loop conformation. Consistent with this model was the finding that Ala replacement of any one of the four Zn^{2+} ligands of *E. coli* GmhB is detrimental. The His94Ala mutant proved to be insoluble, and the

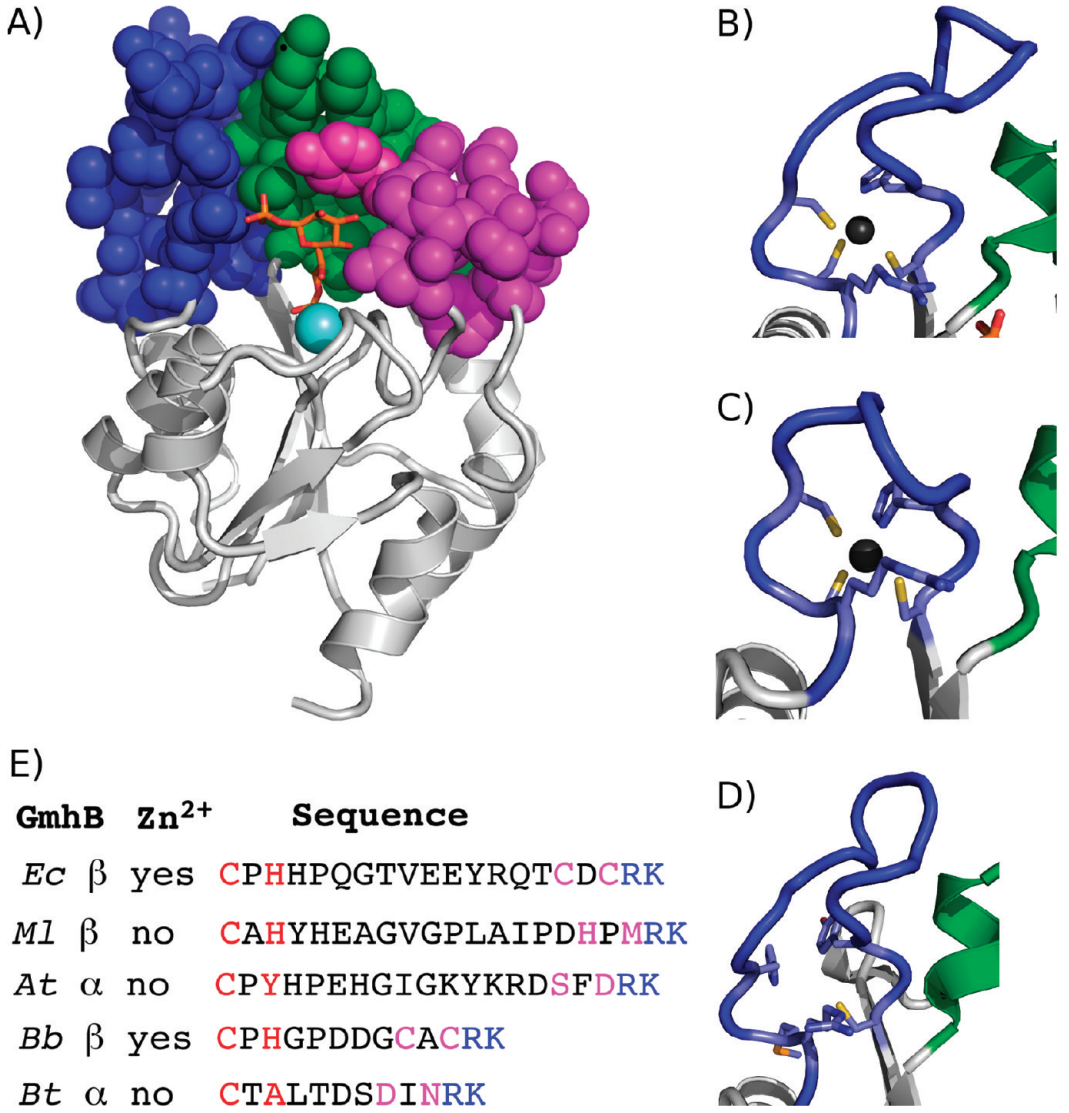


FIGURE 6: (A) Structure of the *E. coli* GmhB−Mg^{2+}−D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate complex. The Rossmann catalytic unit is depicted as a ribbon (gray) with space-filling models for the Zn^{2+} binding loop (blue), loop 2 (green), and loop 1 (magenta). (B−D) Zn^{2+} (black sphere)-bound loops of *E. coli* and *B. bronchiseptica* GmhB and the homologous, unliganded loop of *M. loti* GmhB. The Zn^{2+} ligands are shown as sticks. The stringently conserved Arg is also depicted. (E) Representative GmhB Zn^{2+} binding loops. Listed are the species [*E. coli* (Ec), *M. loti* (Ml), *A. thermoaerophilus* (At), *B. bronchiseptica* (Bb), and *B. thetaiotaomicron* (Bt)], the preferred anomer (α or β) of D-*glycero*-D-*manno*-heptose 1,7-bisphosphate, and the presence or absence of a Zn^{2+} binding motif. The first and second pairs of Zn^{2+} ligands are colored red and magenta, respectively, and the stringently conserved C(1) phosphate binding Arg and catalytic Lys residues are colored blue.

three Cys mutants, which were soluble, displayed reduced catalytic efficiencies toward catalyzed hydrolysis of the α- and β-anomers of D-*glycero*l-D-*manno*-heptose-1,7-bisphosphate (Table 2).[3]

Examination of the multiple alignment of GmhB sequences (Figure SI1 of the Supporting Information), however, revealed the unexpected; the Zn^{2+} binding loop sequence is not conserved. Moreover, for a significant fraction of the GmhB orthologs (~10%), two or more of the Zn^{2+} ligands of the **C**x**H**(x)$_n$**C**x**C** consensus sequence are absent, which is likely to eliminate binding of Zn^{2+} to the loop. In addition, the number of residues that separate the two pairs of Zn^{2+} ligands varies from 4 for the "short loop" to 12−14 for the long loop (Figure 6). *E. coli* GmhB provides an example of the **C**x**H**(x)$_n$**C**x**C** long loop (n = 12); *B. bronchiseptica* GmhB and *Bacteroides thetaiotaomicron* GmhB provide examples of the **C**x**H**(x)$_n$**C**x**C** short loop (n = 4), whereas *Aneurinibacillus thermoaerophilus* GmhB [**C**x**P**(x)$_{12}$S**x**D] and *Mesorhizobium loti* GmhB [**C**x**Y**(x)$_{14}$**H**x**M**] provide examples of the long loop, both of which have lost the capacity for Zn^{2+} binding. Each of these GmhB orthologs has been shown by in vitro kinetic analysis to be an effective catalyst of D-*glycero*-D-*manno*-heptose 1,7-bisphosphate hydrolysis. The structure of *M. loti* GmhB (PDB entry 2O2X) provides a snapshot of the "empty" loop (Figure 6). The loop is ordered, and in fact, the *B* factors for loop residues do not differ (relative to the average *B* factors of the core) from those of the *E. coli* and *B. bronchiseptica* GmhB loop residues. Furthermore, the conformation of the *M. loti* GmhB loop is similar to that of the long, Zn^{2+}-bound loop of *E. coli* GmhB (Figure 6).

[3]At first glance, this conclusion might appear to be inconsistent with the observed reduction in catalytic efficiency in the *E. coli* GmhB Zn^{2+} binding loop mutants (Table 2). However, this is not the case as the alteration of a structural motif by mutation is frequently destabilizing to the native conformation enzyme and often manifested in decreased catalytic efficiency. In this case, removal of one of the four Zn^{2+} ligands might result in misfolding (as appears to be the case for the H94A mutant which is insoluble) or destabilization of the native conformation as the Zn^{2+} ion incorporates a fourth ligand from solvent, or elsewhere in the protein.

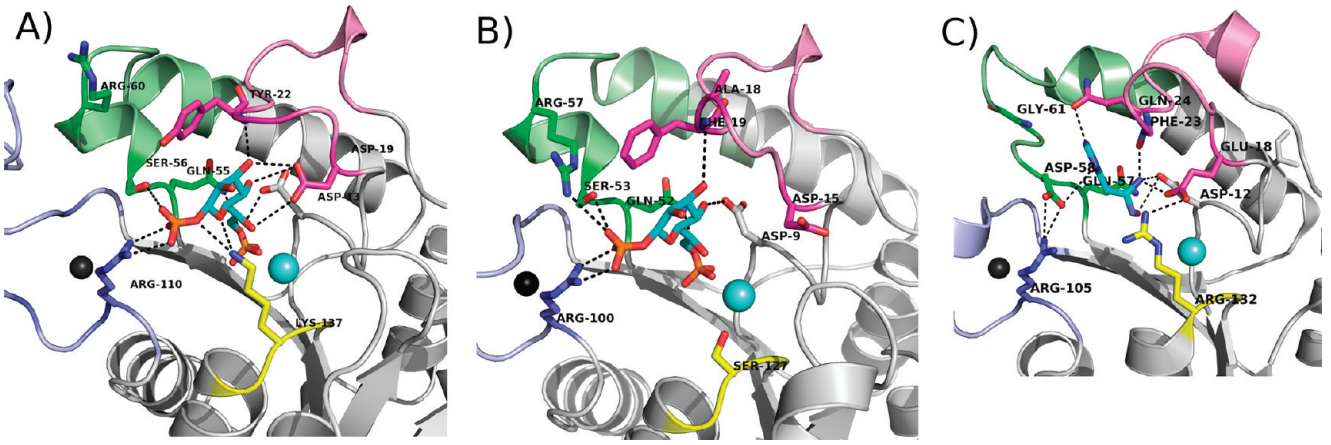


FIGURE 7: Ribbon diagrams of (A) *E. coli* GmhB, (B) *B. bronchiseptica* GmhB, and (C) *E. coli* HisB with the substrate binding residues shown as sticks. Loops are colored as in Figure 4.

Examination of the structure of *M. loti* GmhB demonstrated a possible structural feature for stabilization of the loop in the absence of the Zn^{2+} binding motif, namely, the presence of an N-terminal extension (residues 15−21) that forms an additional β-strand antiparallel to that from which the insert extends. However, inspection of the alignment of GmhB orthologs (Figure SI1 of the Supporting Information) revealed that no correlation exists between the presence and absence of an extended N-terminus and the presence or absence of a Zn^{2+} binding site. Thus, the structural significance of the presence of Zn^{2+} in some GhmB (or HisB, vide infra) orthologs, and its absence in others, is not known.

The finding that the GmhB loop exists in either a "no Zn^{2+}" or "Zn^{2+}-bound" form prompted the examination of aligned HisB sequences (Figure SI2 of the Supporting Information). Whereas the HisB orthologs in proteobacteria (the *E. coli* HisB is pictured in Figure 2) contain the Zn^{2+} binding site, those in Bacteroidetes do not. Thus, both loop types are present in both subfamilies.

Given that Zn^{2+} binding is not a determinant of catalysis, an alternate role that the Zn^{2+} might perform was considered, leading to the experimental interrogation of DNA complexation by *E. coli* and *B. bronchiseptica* GmhB. The Zn^{2+} finger is a common DNA binding motif utilized by transcriptional factors. Although the GmhB Zn^{2+} binding loop bears limited resemblance to the Zn^{2+} finger, it nevertheless seemed prudent to rule out DNA binding as a possible function. A standard DNA gel-shift assay (described in Experimental Procedures) using "generic" calf thymus DNA showed no indication of DNA binding. At present, the origin and function (if any) of the Zn^{2+} binding motif remain undefined.

The Zn^{2+} binding loop, with or without Zn^{2+}, contributes to the sequestration of the substrate. In addition, a stringently conserved Arg residue (Figure SI1 of the Supporting Information) is positioned at the base of the loop, between the **C**x**H**(x)$_n$**C**x**C** motif and the catalytic Lys [viz. **C**x**H**(x)$_n$**C**x**C***RK* (Figure 6)]. The structure of the *E. coli* GmhB−Mg^{2+}−D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate complex suggests that the Arg contributes to substrate recognition by binding the C(1) phosphate (Figures 3 and 7). Replacement of Arg110 of *E. coli* GmhB with Ala resulted in a significant reduction in the catalytic efficiency of hydrolysis of the physiological substrate D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate, reducing k_{cat} 6-fold and increasing K_m 36-fold, such that k_{cat}/K_m was reduced by 200-fold (Table 2). On the other hand, the Arg110Ala mutant showed no detectable activity toward D-*glycero*-D-*manno*-heptose 1α,7-bisphosphate, which reveals a dramatic reduction in catalytic efficiency compared to the activity of the wild-type enzyme [$k_{cat}/K_m = 7 \times 10^4\ M^{-1}\ s^{-1}$ (Table 2)]. This finding suggests that the Arg plays a critical role in preserving catalytic activity toward the α-anomer. This residue might be especially important for the GmhB orthologs that function in the D-*glycero*-D-*manno*-heptose 1α-GDP pathway, because their physiological substrate is the α-anomer. In principle, the orientation of the Arg can be directed by the backbone conformation, and it by the adjacent Zn^{2+} binding site. This raised the possibility that the Zn^{2+} binding site acts as an anomeric specificity switch. However, this model was eliminated by the finding that there is no correlation between the anomeric specificity of the GmhB ortholog and whether it possesses a Zn^{2+} binding site (Figure 6E).

(*ii*) *Loops 1 and 2*. Loop 1 [residues 18−30 (Figure 7)] and loop 2 [residues 55−64 (Figure 7)] are extensions from the core Rossmann-like fold that, as with the Zn^{2+} binding loop, are expected to provide the residues for binding the substrate leaving group. The structure of the *E. coli* GmhB−Mg^{2+}−D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate complex shows that Asp19 of loop 1 makes a hydrogen bond to the heptose C(3)OH (3.2 Å) and C(6)OH (3.6 Å) groups. This residue is conserved in ~80% of all GmhB sequences and is conservatively replaced with Glu in all other GmhB orthologs. Thus, the function of this residue in binding the substrate leaving group is conserved. Tyr22 makes a hydrogen bond to a water molecule which forms a hydrogen bond to the substrate C(1) phosphate group. Moreover, the Tyr22 ring sits over the substrate leaving group, shielding the substrate leaving group. Tyr22 is conserved in ~80% of all GmhB sequences, and in the remaining sequences, it is replaced with a Phe. This finding is consistent with the interpretation that the major role of this Tyr/Phe and the adjoining loop 1 residues is the shielding of the substrate leaving group and core active site from bulk solvent. There are no additional conserved residues on loop 1.

In loop 2, Ser56 is found in the TNQS motif wherein Thr53 is part of the HAD catalytic scaffold (vide infra). Ser56 makes two hydrogen bonds with the substrate C(1) phosphate (3.3 and 3.5 Å), while the adjacent Gln55 in the motif forms a hydrogen bond with the C(1)O. Both Ser56 and Gln55 are stringently conserved.

(*iii*) *Mg^{2+} Binding Loop*. Although technically a component of the catalytic scaffold, the *E. coli* GmhB Mg^{2+} binding loop (Figure 6) does contribute Lys137 that forms a hydrogen bond with substrate C(1) phosphate (Figure 7). However, the residue is

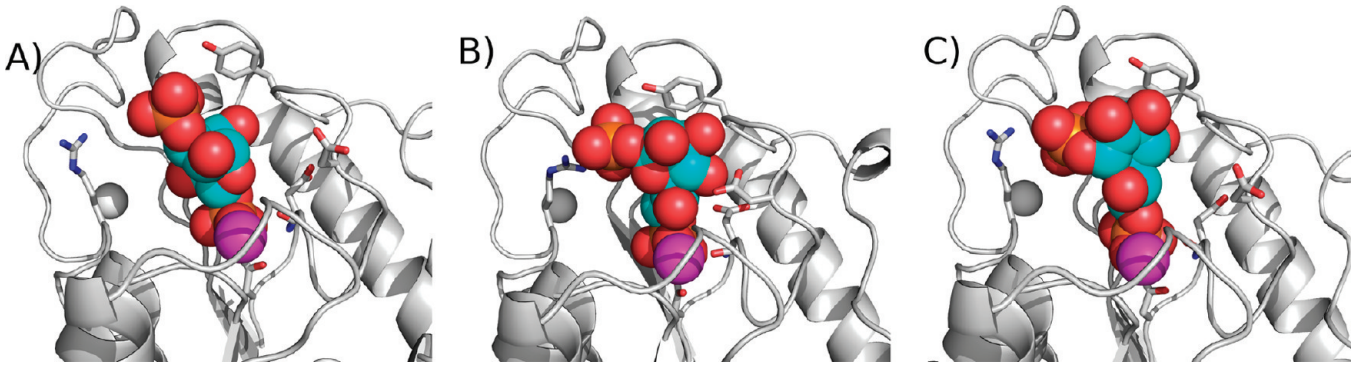


FIGURE 8: *E. coli* GmhB (A) modeled with D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate (space filling) bound in the reverse orientation [with the C(1) phosphate in the "transferring position"], (B) in the orientation experimentally determined, and (C) modeled with the α-anomer bound in the correct orientation [with the C(7) phosphate in the transferring position]. Note that the energy minimization used in generating the two models (A and C) altered the side chain conformations of some active site residues.

not stringently conserved (Figure 7 and Figure SI1 of the Supporting Information), and furthermore, its replacement with Ala results in only a small reduction in catalytic efficiency (Table 2).

Substrate Specificity. *E. coli* GmhB is highly specific for D-*glycero*-D-*manno*-heptose 1,7-bisphosphate, shows a 100:1 preference for the β-anomer over the α-anomer, removes the C(7) phosphate and not the C(1) phosphate, and does not hydrolyze the reaction product, D-*glycero*-D-*manno*-heptose 1-phosphate (*25*). To gain insight into the structural basis for this specificity, we used the structure of the *E. coli* GmhB$-Mg^{2+}-$D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate complex to generate a model in which the D-*glycero*-D-*manno*-heptose 1β,7-bisphosphate ligand is docked into the active site, in the reverse orientation, such that the C(1) phosphate rather than the C(7) phosphate is positioned for Mg^{2+} coordination and in-line attack by the Asp nucleophile (Figure 8). In general, models are not sufficiently accurate to predict the small changes in orientation in interacting groups that can result in large changes in transition-state structure and energy. However, in this case, the comparison of the experimentally observed substrate complex (Figure 8B) versus the substrate docked in the reverse orientation (Figure 8A) clearly identifies a major effect that, independent of the more subtle orientation effects, is likely to preclude hydrolysis of the C(1) phosphate. Specifically, in the reverse orientation, the leaving group does not have the correct shape to seal the active site entrance, and thus, the electrostatic stabilization of the transition state is likely to be impaired by water molecules that can solvate the catalytic site. In contrast, the model of the D-*glycero*-D-*manno*-heptose 1α,7-bisphosphate docked in the catalytic orientation (Figure 8C) predicts that the solvent seal is in place. In addition, an alteration in ring pucker brings the C(1) phosphate of the α-anomer into position to form a salt bridge with Arg110. The 100:1 anomeric preference observed for *E. coli* GmhB is likely to be determined by subtle orientation effects, which the model cannot predict.

Divergence of the HisB Histidinol-phosphate Phosphatase. We have shown (*25*) that histidinol phosphate is not a substrate for *E. coli* GmhB. Although the GmhB substrate, D-*glycero*-D-*manno*-heptose 1,7-bisphosphate, was not tested for substrate activity with *E. coli* HisB, the reported absence of activity toward other phosphate ester metabolites indicates that the histidinol-phosphate phosphatase is not promiscuous (*24*). Thus, whereas the two sequences are 34% identical, they do not share substrates. To identify the structural elements for the substrate specialization that define the respective biochemical functions of GmhB and histidinol-phosphate phosphatase, the structures of these two enzymes were compared, within the context of the residue conservation defined by multiple-sequence alignments. Accordingly, the *E. coli* histidinol-phosphate phosphatase domain of the HisB fusion protein was used as a query in a Blast search of the nonredundant protein sequence database. The sequences were aligned and curated to remove sequences that did not encode HisB (Figure SI2 of the Supporting Information). An analogous procedure was used to create the multiple-sequence alignment of the GhmB homologues identified by Blast searches with the *E. coli* and *M. loti* GmhB sequences serving as a query (Figure SI1 of the Supporting Information). No evidence for a "stand-alone" HAD family histidinol-phosphate phosphatase was found. Likewise, there was no indication that a homologue possessing a novel catalytic activity is found among the GmhB sequences.

The respective sequence families provide insight into the biological range of HisB and GmhB, which agrees with and expands upon an earlier report (*36*). The HisB is primarily found in bacterial species of the gamma subdivision of Proteobacter, yet it is also found in numerous species of the phylum Bacteroidetes/Chlorobi. HisB is believed to have evolved within the gamma subdivision of Proteobacter following its split from the beta division and is thus a relatively new enzyme (*36*). A likely scenario is that the GmhB ancestor gene underwent duplication within a gamma subdivision bacterium, and the extra copy evolved to the HisB histidinol-phosphate phosphatase gene. The HisB gene was most probably acquired by *Bacteroides* via horizontal gene transfer. GmhB has a much greater biological range, which includes a variety of species from each phylum of Bacteria as well as some species of Archaea. The HisB histidinol-phosphate phosphatase might thus be viewed as an adaptation of the GmhB to a novel biochemical niche (viz. the histidine biosynthetic pathway) within a small biological sector.

The GmhB sequences have diverged to ~30% identity, and the HisB histidinol-phosphate phosphatase sequences have diverged to ~40% identity. The two alignments are therefore reliable for use in distinguishing stringently conserved residues. First, it is noteworthy that the Arg110 residue in the *E. coli* GmhB Zn^{2+} binding loop, which plays an important role in substrate recognition (vide supra), is stringently conserved among the histidinol-phosphate phosphatase domains of HisB. In *E. coli* HisB, the Arg105 residue does not bind the substrate; instead, it forms a hydrogen bond with an Asp residue of loop 2 (Figure 7). The *E. coli* HisB Asp58 residue is stringently conserved in the

position that corresponds to *E. coli* GmhB Ser56 (Figure 7) that, like the Arg residue, binds the substrate C(1) phosphate. The counterpart to the stringently conserved GmhB "TNQS" sequence motif is the HisB TNQD motif [where T is the catalytic Thr53 in *E. coli* GmhB that is conservatively substituted with Ser in some GmhB orthologs (Figure 5) and Q is the Gln55 in *E. coli* that forms a long hydrogen bond (3.9 Å) to the substrate C(1)O group (not shown)]. The switch from Ser to Asp allows a clear distinction to be made between GmhB and histidinol-phosphate phosphatase sequences. In the *E. coli* HisB structure, the Asp58 of the TNQD motif forms a long hydrogen bond to the N1H group of the histidinol phosphate ligand and an ion pair with the Zn^{2+} binding loop Arg105 (Figure 7).

In the *E. coli* HisB structure, the stringently conserved residues Gln24 and Glu18 (of loop 1) engage the N3 atom and C(2)NH_3^+ group of the histidinol ligand in hydrogen bond formation. The homologous residues in *E. coli* GmhB, Tyr22 and Asp13, function in the respective roles of substrate leaving group desolvation and substrate binding (Figure 7). Glu18 of HisB forms an ion pair with the catalytic scaffold Arg132, yet neither it nor its *E. coli* GmhB counterpart, Lys137 (vide supra), is fully conserved.

SUMMARY AND CONCLUSION

Historically, two of the first capless members of the HAD family to be examined structurally were T4 polynucleotide kinase (*21*) and magnesium-dependent phosphatase-1 (*19*). Although these proteins have active sites that are completely solvent exposed, their large macromolecular substrates (t-RNA and protein, respectively) are expected to act like a cap by binding and protecting the phosphoryl transfer site from bulk solvent. Later, the capless HAD phosphatases, 2-keto-3-deoxyoctulosonate-8-phosphate (KDO-8-P) phosphatase (*37*, *38*) and 2-keto-3-deoxy-D-glycero-D-galactononononate-9-phosphate (KDN-9-P) phosphatase (*23*), were discovered to efficiently and specifically hydrolyze their respective (small molecule) substrates by recruiting a second subunit to act as a "cap" to the active site of the first subunit. The tetramerization of these two enzymes is mediated by a short peptide insert into the Rossmann-like fold at the position corresponding to the cap domain of the C1 HADS (Figure 1).

GmhB provides a unique solution to the problem of substrate discrimination in the capless HAD phosphatases. The GmhB structures show that the catalytic site is elaborated by three loops that form a concave, semicircular surface around the substrate leaving group. One of the three loops binds Zn^{2+} in some but not all GmhB orthologs. Each of the three loops positions stringently conserved residues that engage in hydrogen bond interaction with the substrate leaving group, D-*glycero*-D-*manno*-heptose 1-phosphate. GmhB and the histidinol-phosphate phosphatase domain of HisB share the same design, yet through selective residue usage on the three substrate recognition loops, they have achieved unique substrate specificity, and thus novel biochemical function.

ACKNOWLEDGMENT

We thank Dr. Mirek Cygler and Dr. Steven Almo for providing the coordinates of HisB and Q8A5V9, respectively, prior to publication.

SUPPORTING INFORMATION AVAILABLE

Alignment of representative GmhB and HisB sequences. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Heine, A., DeSantis, G., Luz, J. G., Mitchell, M., Wong, C. H., and Wilson, I. A. (2001) Observation of covalent intermediates in an enzyme mechanism at atomic resolution. *Science 294*, 369–374.
2. Dalby, A., Dauter, Z., and Littlechild, J. A. (1999) Crystal structure of human muscle aldolase complexed with fructose 1,6-bisphosphate: Mechanistic implications. *Protein Sci. 8*, 291–297.
3. Sygusch, J., and Beaudry, D. (1984) Catalytic activity of rabbit skeletal muscle aldolase in the crystalline state. *J. Biol. Chem. 259*, 10222–10227.
4. Schlichting, I., and Chu, K. (2000) Trapping intermediates in the crystal: Ligand binding to myoglobin. *Curr. Opin. Struct. Biol. 10*, 744–752.
5. Schramm, V. L. (2005) Enzymatic transition states: Thermodynamics, dynamics and analogue design. *Arch. Biochem. Biophys. 433*, 13–26.
6. Petrek, M., Otyepka, M., Banas, P., Kosinova, P., Koca, J., and Damborsky, J. (2006) CAVER: A new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinf. 7*, 316.
7. Collet, J. F., van Schaftingen, E., and Stroobant, V. (1998) A new family of phosphotransferases related to P-type ATPases. *Trends Biochem. Sci. 23*, 284.
8. 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.
9. Allen, K. N., and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: Evolution of a catalytic scaffold. *Trends Biochem. Sci. 29*, 495–503.
10. Smith, C. A., and Rayment, I. (1996) X-ray structure of the magnesium(II)·ADP·vanadate complex of the *Dictyostelium discoideum* myosin motor domain to 1.9 Å resolution. *Biochemistry 35*, 5404–5417.
11. 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.
12. Lu, Z., Dunaway-Mariano, D., and Allen, K. N. (2008) The catalytic scaffold of the haloalkanoic acid dehalogenase enzyme superfamily acts as a mold for the trigonal bipyramidal transition state. *Proc. Natl. Acad. Sci. U.S.A. 105*, 5687–5692.
13. Allen, K. N., and Dunaway-Mariano, D. (2009) Markers of fitness in a successful enzyme superfamily. *Curr. Opin. Struct. Biol. 19*, 658–665.
14. Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D., and Aravind, L. (2006) Evolutionary genomics of the HAD superfamily: Understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J. Mol. Biol. 361*, 1003–1034.
15. 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.
16. Zhang, G., Dai, J., Wang, L., Dunaway-Mariano, D., Tremblay, L. W., and Allen, K. N. (2005) Catalytic cycling in β-phosphoglucomutase: A kinetic and structural analysis. *Biochemistry 44*, 9404–9416.
17. Kim, H. Y., Heo, Y. S., Kim, J. H., Park, M. H., Moon, J., Kim, E., Kwon, D., Yoon, J., Shin, D., Jeong, E. J., Park, S. Y., Lee, T. G., Jeon, Y. H., Ro, S., Cho, J. M., and Hwang, K. Y. (2002) Molecular basis for the local conformational rearrangement of human phosphoserine phosphatase. *J. Biol. Chem. 277*, 46651–46658.
18. Dai, J., Finci, L., Zhang, C., Lahiri, S., Zhang, G., Peisach, E., Allen, K. N., and Dunaway-Mariano, D. (2009) Analysis of the structural determinants underlying discrimination between substrate and solvent in β-phosphoglucomutase catalysis. *Biochemistry 48*, 1984–1995.
19. 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.
20. Peisach, E., Wang, L., Burroughs, A. M., Aravind, L., Dunaway-Mariano, D., and Allen, K. N. (2008) The X-ray crystallographic structure and activity analysis of a *Pseudomonas*-specific subfamily of the HAD enzyme superfamily evidences a novel biochemical function. *Proteins 70*, 197–207.
21. Galburt, E. A., Pelletier, J., Wilson, G., and Stoddard, B. L. (2002) Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase. *Structure 10*, 1249–1260.
22. Wu, J., and Woodard, R. W. (2003) *Escherichia coli* YrbI is 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase. *J. Biol. Chem. 278*, 18117–18123.
23. Lu, Z., Wang, L., Dunaway-Mariano, D., and Allen, K. N. (2009) Structure-function analysis of 2-keto-3-deoxy-D-glycero-D-galacto-nononate-9-phosphate phosphatase defines specificity elements in

type C0 haloalkanoate dehalogenase family members. *J. Biol. Chem. 284*, 1224–1233.
24. Rangarajan, E. S., Proteau, A., Wagner, J., Hung, M. N., Matte, A., and Cygler, M. (2006) Structural snapshots of *Escherichia coli* histidinol phosphate phosphatase along the reaction pathway. *J. Biol. Chem. 281*, 37930–37941.
25. Wang, L., Huang, H., Nguyen, H., Allen, K. N., Mariano, P. S., and Dunaway-Mariano, D. (2010) Divergence of Biochemical Function in the HAD Superfamily: D-*glycero*-D-*manno*-Heptose-1,7-bisphosphate Phosphatase (GmhB). *Biochemistry* DOI 10.1021/bi902018y.
26. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol. 276*, 307–326.
27. Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D56* (Part 12), 1622–1624.
28. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760–768.
29. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.
30. 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.
31. Davis, I. W., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2004) MOLPROBITY: Structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res. 32*, W615–W619.
32. Kerr, L. D. (1995) Electrophoretic mobility shift assay. *Methods Enzymol. 254*, 619–632.
33. 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.
34. 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.
35. Kleywegt, G. J., and Jones, T. A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D50*, 178–185.
36. Brilli, M., and Fani, R. (2004) Molecular evolution of hisB genes. *J. Mol. Evol. 58*, 225–237.
37. 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.
38. Biswas, T., Yi, L., Aggarwal, P., Wu, J., Rubin, J. R., Stuckey, J. A., Woodard, R. W., and Tsodikov, O. V. (2009) The tail of KDSC: Conformational changes control the activity of a haloacid dehalogenase superfamily phosphatase. *J. Biol. Chem. 284*, 30594–30603.
39. Lu, Z., Dunaway-Mariano, D., and Allen, K. N. (2005) HAD superfamily phosphotransferase substrate diversification: Structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. *Biochemistry 44*, 8684–8696.