Structure and Activity Analyses of *Escherichia coli* K-12 NagD Provide Insight into the Evolution of Biochemical Function in the Haloalkanoic Acid Dehalogenase Superfamily†,‡

Lee W. Tremblay,§ 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 87131-0001

Received September 11, 2005; Revised Manuscript Received November 22, 2005

ABSTRACT: The HAD superfamily is a large superfamily of proteins which share a conserved core domain that provides those active site residues responsible for the chemistry common to all family members. The superfamily is further divided into the four subfamilies I, IIA, IIB, and III, based on the topology and insertion site of a cap domain that provides substrate specificity. This structural and functional division implies that members of a given HAD structural subclass may target substrates that have similar structural characteristics. To understand the structure/function relationships in all of the subfamilies, a type IIA subfamily member, NagD from *Escherichia coli* K-12, was selected (type I, IIB, and III members have been more extensively studied). The structure of the NagD protein was solved to 1.80 Å with R_{work} = 19.8% and R_{free} = 21.8%. Substrate screening and kinetic analysis showed NagD to have high specificity for nucleotide monophosphates with $k_{cat}/K_m = 3.12 \times 10^4$ and 1.28×10^4 μM^{-1} s^{-1} for UMP and GMP, respectively. This specificity is consistent with the presence of analogues of NagD that exist as fusion proteins with a nucleotide pyrophosphatase from the Nudix family. Docking of the nucleoside substrate in the active site brings it in contact with conserved residues from the cap domain that can act as a substrate specificity loop (NagD residues 144−149) in the type IIA subfamily. NagD and other subfamily IIA and IIB members show the common trait that substrate specificity and catalytic efficiencies (k_{cat}/K_m) are low (1×10^4 M^{-1} s^{-1}) and the boundaries defining physiological substrates are somewhat overlapping. The ability to catabolize other related secondary metabolites indicates that there is regulation at the genetic level.

The haloalkanoate dehalogenase superfamily (HADSF)[1] enzymes catalyze group transfer reactions mediated by an active site aspartate nucleophile. Most members of this family are phosphotransferases that employ water as the phosphoryl group acceptor (*1*−*5*). The HADSF phosphatases are ubiquitous in nature, and within a single organism, they are quite numerous (e.g., 58 in human, 37 in *Escherichia coli*). The biochemical roles played within an organism vary from signal transduction to DNA repair to secondary metabolism. The evolution of biochemical function within this superfamily is based on the diversification of substrate recognition, which is the focus of this paper.

All HADSF members contain a highly conserved α/β core domain that supports a catalytic scaffold, comprised of four loops (hereafter referred to as motifs I−IV) (*6*) that position the residues functioning in Mg^{2+} cofactor binding, substrate binding, acid−base catalysis, and nucleophilic catalysis. Contacts between the core domain and the substrate are concentrated on the transferring phosphoryl group. Large substrates bind to the HAD core domain by docking with residues lining the entrance of the active site (*7*, *8*). Small substrates, on the other hand, enter the core domain active site and then are covered by a mobile cap domain that interacts primarily with the substrate leaving group (*9*−*12*). The cap domain is a key element in the evolution of new biochemical function within the HADSF (*13*, *14*).

The structure-based classification of the HADSF is based on presence, position, and fold of the cap domain. Type I cap domains are inserted between motifs I and II of the core domain and are α-helical in structure (*15*−*20*). Type II subfamily member cap domains are located between core domain motifs II and III and have two different α/β folds (*14*, *21*−*23*), designated types IIA and IIB. Type III HADSF members have only a core domain with a connecting loop serving in place of the cap domain (*7*, *8*, *24*, 25). The

† This work was supported by NIH Grant GM61099 (to K.N.A. and D.D.-M.) and National Institutes of Health Training Grant HL07291 (to L.W.T.).

‡ Coordinates of NagD have been deposited in the Protein Data Bank with the accession code 2C4N.

* Correspondence should be addressed to either D.D.-M. (tel, 505-277-3383; fax, 505-277-6202; e-mail, dd39@unm.edu) or K.N.A. (tel, 617-638-4398; fax, 617-638-4273; e-mail, allen@med-xtal.bu.edu).

§ Boston University School of Medicine.

|| University of New Mexico.

[1] Abbreviations: β-PGM, β-phosphoglucomutase; F16P, α-fructose 1,6-(bis)phosphate; GlcNAc, *N*-acetylglucosamine; GlcN6P, glucosamine 6-phosphate; GlcNAc1P, *N*-acetylglucosamine 1-phosphate; GlcNAc6P, *N*-acetylglucosamine 6-phosphate; G1P, glucose 1-phosphate; G6P, glucose 6-phosphate; β-G1P, β-D-glucose 1-phosphate; β-G16P, β-glucose 1,6-(bis)phosphate; α-G16P, α-glucose 1,6-(bis)-phosphate; HAD, haloalkanoic acid dehalogenase; HADSF, haloalkanoic acid dehalogenase superfamily; pNPP, *p*-nitrophenyl phosphate; NDP, nucleotide diphosphate; NMP, nucleotide monophosphate; rmsd, root mean square deviation.

10.1021/bi051842j CCC: $33.50
Published on Web 01/04/2006

majority of the HADSF members contain a cap domain. Given that the cap domain defines the substrate specificity and that the physiological substrate defines the biochemical function, it follows that understanding the relationship between cap domain structure and substrate structure is key to the sequence-based assignment of function to unknown HADSF members.

In previous work we have identified substrate specificity loops within selected HADSF type IA and type IIB enzymes as a starting point to function assignment to unassigned HADSF sequences (*13*, *14*). The substrate specificity loops position amino acid residues that contribute to the active site in the catalytically active "cap domain−core domain" closed enzyme conformation. These residues serve to provide steric constraints on substrate size and shape and define the electrostatic environment of the catalytic site as well as anchor and activate the substrate for attack by the core domain aspartate nucleophile.

The goal of the work described in this paper was to identify the substrate specificity loop(s) functioning in substrate recognition among the HADSF type IIA members. Herein, we report the structure and substrate specificity of the HADSF type IIA enzyme NagD from the *E. coli* K-12 (accession number P15302). A clue to the biochemical function of this enzyme was provided by the observation that the NagD can function independently, as it does in *E. coli*, or in combination with a MutT domain (as an N-terminal fusion), as it does in *Paenibacillus thiaminolyticus* (*26*) and *Bacillus clausii* KSM-K16 (unpublished submission to the NCBI Genome Project). MutT belongs to the Nudix enzyme superfamily (*27*−*31*), and it specifically catalyzes phosphoanhydride cleavage of nucleoside triphosphates. The coupling of NagD and MutT into a single fused protein therefore suggested that NagD might catalyze the hydrolysis of the nucleoside monophosphates liberated by the action of the MutT domain.

An additional important factor in the selection of the *E. coli* NagD target was the fact that the unpublished structures of NagD homologues from various bacterial sources (*Thermatoga maritima*, *Streptomyces pneumoniae*, *Streptococcus pyogenes* and *Enterococcus faecalis*) had been deposited in the Protein Data Bank (PDB accession numbers 1PW5, 1YDF, 1YS9, and 1YV9), as a result of the structural genomics initiative. Thus, the elucidation of structure−function elements operating in the *E. coli* homologue might provide insight into the function of these orphaned structures. In turn, these structures could provide a backdrop against which a more extensive structure−function analysis of the type IIA subfamily could be carried out.

Here we report the X-ray crystallographic structure of the *E. coli* K-12 strain NagD protein determined to 1.8 Å resolution and the steady-state kinetic constants measured for the catalyzed hydrolysis of a wide range of organophosphate metabolites. *E. coli* NagD is shown to be a nucleotide phosphatase that has modest substrate specificity and catalytic efficiency. Evidence is presented that suggests that the HADSF type IIA cap domain utilizes a single loop is substrate recognition. These findings are combined with those from earlier structure−function analyses of HADSF type IIB phosphatases (*14*, *21*−*23*) to suggest that the HADSF type II subfamily of phosphatases possesses overlapping substrate ranges. On the basis of these results a model for evolution of physiological function in the HADSF is proposed, which is that unique biochemical functions among the HADSF members are defined in part by their preferences for the physiological substrates and in part by controlled gene expression.

MATERIALS AND METHODS

Bioinformatics, Cloning, Expression, and Purification. Multiple sequence alignment analysis was carried out using ClustalW (*32*), in conjunction with the Pfam database (*33*). Sequence alignment figures were generated using ESPRIPT (*34*). Gene cloning was carried out utilizing a PCR-based strategy using the genomic DNA from the *E. coli* (O157:H7) K-12 strain (ATCC number 10798D) with PfuTurbo DNA polymerase and custom oligonucleotide primers to target the B0675 gene which encodes the protein P15302. To amplify the entry product for recombination into the Gateway pDONR221 vector (Invitrogen), a second set of oligonucleotide primers was designed (5′-GGGGACCAC-TTTGTACAAGAAAGCTGGGTTACCACGTCCGCAG-ACGTGGATCCTATTCAG and 5′-GGGGACAAGTTTGTA-CAAAAAAGCAGGCTGGTGTCAATATCTGGGTAGC-ATATGACCATT). The amplified product was introduced via homologous recombination into the pDONR201 Gateway entry vector using the BP clonase enzyme mixture (Invitrogen). The recombinant product was then transformed into DH5α competent cells and the plasmid purified using a Qiaprep spin miniprep kit (Qiagen). The gene sequence was confirmed by DNA sequencing at the Core Facility at Tufts New England Medical Center.

The recombinant NagD construct was placed in an expression vector using the homologous recombination reaction (with the LR clonase enzyme mixture) between the entry vector and the pDEST14 expression vector. The recombinant product was transformed into DH5α competent cells and the plasmid purified using a Qiaprep spin miniprep kit. The purified plasmid was used to transform BL21AI cells for expression. The resulting cells were grown overnight at 37 °C with shaking at 250 rpm in 100 mL of Luria broth (LB) containing 50 μg/mL ampicillin. Each 1 L of media was inoculated with 25 mL of culture, and the cells were grown for 2 h to OD_{600nm} ∼0.6 before induction using 5 mM L-arabinose for 3−4 h. The cells were harvested by centrifugation (6500 rpm) for 15 min to yield 2−3 g/L of pellet. The cell pellet was frozen at −80 °C overnight and then thawed and resuspended in 10 mL/g of pellet of lysis buffer containing 50 mM Hepes, pH 7.4, 1 mM EDTA, 1 mM benzamide hydrochloride, 0.05 mg/mL trypsin inhibitor type II-S (Sigma), 1 mM 1,10-phenanthroline, 0.1 mM PMSF, 5 mM DTT, and 2 μg/mL DNase. Lysozyme (1 mg/mL) was added and incubated for 1 h on ice with occasional mixing. The lysis solution was then spun at 40000 rpm for 1 h and the supernatant retained.

The supernatant was subjected to precipitation with 30% ammonium sulfate and the precipitant removed by centrifugation at 18000 rpm for 10 min. The pellet was resuspended in buffer containing 50 mM Hepes, pH 7.4, 10 mM $MgCl_2$, and 2 mM DTT (buffer A) and dialyzed overnight against this same buffer. After dialysis, 100 mM *N*-acetyl-D-glucosamine (GlcNAc) was added for increased protein solubility. The protein was gelatinous at 5−6 mg/mL and

could not be concentrated further. Purity was assessed as >90% using SDS−PAGE. All protein concentrations were determined using Bradford (*35*) reagent using a 1 mg/mL BSA standard.

Crystallization and Data Collection. The native purified protein from the pDEST14 plasmid was screened for crystallization by sparse matrix screening (*36*) with crystal screen kits I and II from Hampton Research. After 7 days at 18 °C small rods formed in the presence of 0.4 M monoammonium dihydrogen phosphate. These crystals were then improved via three rounds of microseeding and the addition of the osmolyte 100 mM sarcosine as a crystallization additive (*37*) to attain a size of approximately 1 mm × 0.3 mm × 0.3 mm.

Crystals were frozen for data collection using 100% Paratone-N (Hampton Research) as cryoprotectant in a N_2 stream cooled by liquid nitrogen at 100 K. An initial diffraction data set was collected to 2.3 Å on an R-Axis IV^{++} image plate located at Boston University School of Medicine. A second data set was collected to 1.80 Å at Brookhaven National Laboratory (BNL) on beamline X26C. All diffraction data were indexed and scaled using DENZO and SCALEPACK. The crystals indexed in space group *I*222 with unit cell dimensions $a = 53.91$ Å, $b = 94.80$ Å, and $c = 119.24$ Å.

Molecular Replacement and Model Refinement. Molecular replacement was performed utilizing the program Molrep (*38*) on the 2.3 Å data set using as the search model 1PW5, submitted to the PDB as part of the Midwest Center for Structural Genomics initiative (the NagD homologues available in the PDB are very similar in sequence and three-dimensional structure such that choice of model does not significantly change the structure solution). The program yielded a solution with an initial *R*-factor of 55.4% and a correlation of 28.3%. The peptide sequence was mutated to polyalanine, and initial refinement was performed using simulated annealing in CNS (*39*) to 3.0 Å resolution. When data were included to 2.8 Å, some side chain density appeared in the map. The correct side chain register was obtained from a protein sequence alignment of P15302 with the 1PW5 sequence using ClustalW (*32*), which was then utilized for structural prediction using SWISS-MODEL (*40*). The predicted model was overlaid with the refined polyalanine chain. With the addition of those visible side chains to the model the density was significantly improved. Several of the active site residues, particularly those coordinated with the Mg^{2+} cofactor, showed low occupancy and distorted density in this initial structure, so new crystals were grown from a fourth round of microseeding in the presence of 100 mM $MgCl_2$. These crystals were used to collect the final 1.80 Å resolution data set. Successive rounds of rebuilding and refinement were continued using the higher resolution data with the partially refined 2.3 Å model yielding a final R_{work} of 19.8% with an R_{free} (*41*) of 21.8%. To avoid model bias, the metal cofactor, phosphate ligands, and water molecules were added when R_{free} was less than 30%. The model geometry was checked using a Ramachandran plot generated with the program Coot (*42*, *43*). The data collection and refinement statistics are summarized in Table 1.

Continuous Activity Assays. All continuous assays were carried out at 25 °C in 100 mM K^+Hepes buffer (pH 7.0) containing 10 mM $MgCl_2$, unless stated otherwise. The glucose production from glucose 6-phosphate (G6P, Sigma-Aldrich) was monitored by measuring the glucose dehydrogenase catalyzed reduction of NADP. A NagD stock solution (3.4 mg/mL) was dialyzed against of 50 mM K^+Hepes containing 10 mM $MgCl_2$, 1 mM GlcNAc (added for stability, the final concentration in each kinetics assay was 2 μM), and 2 mM DTT (pH =7.4) and warmed to room temperature before use to ensure homogeneity. The initial velocity of glucose formation by dephosphorylation of G6P in reaction solutions initially containing 0.0252 μM NagD, 0.5 unit of G6P dehydrogenase, 0.2 mM NADP, 1−15 mM α-G6P, and 10 mM $MgCl_2$ in 0.5 mL of 100 mM K^+Hepes (pH 7.4, 25 °C) was determined by monitoring the absorbance increase at 340 nm ($\Delta\epsilon = 6.2$ mM^{-1} cm^{-1}). The kinetic data were analyzed using eq 1 and the computer program of Cleland (*44*):

$$V_0 = V_m[S]/(K_m + [S]) \quad (1)$$

where [S] is the substrate concentration, V_0 is the initial velocity, V_m is the maximum velocity, and K_m is the Michaelis constant for substrate. The k_{cat} was calculated from the ratio of V_{max} and the enzyme concentration.

Discontinuous Assays. All reagents in the substrate screen were obtained from Sigma-Aldrich. Initial phosphate hy-

Table 1: X-ray Crystallographic Data Collection and Refinement Statistics

Data Collection Statistics	
space group	*I*222
unit cell dimensions *a*, *b*, *c* (Å)	53.914, 94.801, 119.236
X-ray source	Synchrotron, 1.0 Å
resolution range (Å)	100−1.80
highest resolution shell (Å)	(1.86−1.80)
no. of total/unique reflections	181383/28582
completeness (%)[a]	100 (100)
$I/\sigma(I)$	15.0 (2.0)
R_{merge} (%)	10.3 (42.2)
volume fraction of protein (%)	45.60
data redundancy	6.4 (6.3)
Refinement Statistics	
resolution range (Å)	30.55−1.80
no. of protein atoms/ asymmetric unit	1912
no. of waters/ asymmetric unit	121
no. of Mg(II) ion/ asymmetric unit	1
no. of total reflections (working set/free set)	27295/2431
R_{work}/R_{free} (%)	19.8/21.8
Ramachandran plot	
residues in most favored regions (%)	98.00
residues in additionally allowed regions (%)	2.00
average *B*-factor (Å^2)	26.6
average of all amino acids	22.5
average of Mg^{2+} atoms	22.9
average of solvent	32.8
average of main chain atoms	22.5
average of side chain atoms	31.1
Luzzati coordinate error (Å)	0.20
rmsd	
bond length (Å)	0.006
dihedrals (deg)	24.50
angles (deg)	1.50
impropers (deg)	0.79

[a] Numbers in parentheses refer to statistics in highest resolution shell.

drolysis for all substrates was assessed using the Biomol Green kit (Biomol Research) to detect total phosphate release (*45*). The 0.5 mL assay mixture, containing 50 mM Hepes, pH 7.4, 5 mM $MgCl_2$, 1 mM substrate, and 1.2 μM NagD, was incubated for 10 min at 25 °C. Background phosphate levels were monitored in parallel using a control reaction which excluded the NagD enzyme. For each reaction 100 μL of sample was used in 1 mL of the Biomol reagent. After 25 min incubation at 25 °C, the absorbance at 620 nm was measured.

Steady-state kinetics were carried out using 0.0252 μM NagD with varying concentrations of substrates ($0.5-5K_m$). Since concentrated NagD tends to form a gel (see Bioinformatics, Cloning, Expression, and Purification), the protein was warmed to room temperature for use in assays to ensure stability and homogeneity. The reactions were quenched at regular time intervals to monitor the rate of the reaction. The enzyme kinetics software (KinetAsystI) was used to find the values of V_{max} and K_m.

Gel Filtration Experiments. NagD (2 mL at 0.5 mg/mL) in buffer A plus 300 mM NaCl was loaded onto a Sephacryl S200 column preequilibrated with buffer A plus 300 mM NaCl. The experiment was repeated using the same column preequilibrated in buffer A plus 300 mM NaCl and 20 mM GlcNAc. Protein elution was monitored via absorbance at 280 nm.

RESULTS AND DISCUSSION

Overall Structure. The model of NagD was refined to 1.80 Å with an *R*-factor of 19.8% and an R_{free} of 21.8%. NagD crystallized in space group *I*222 with a monomer in the asymmetric unit. The crystal packing occurs via stacked tetramers; however, the enzyme is most likely a monomer in solution, as supported by the gel filtration experiments. The protein aggregates into a gel-like matrix when purified in buffer alone but can be concentrated to >5 mg/mL in the presence of *N*-acetylglucosamine (GlcNAc). The increased solubility in the presence of GlcNAc is most likely due to interactions with the exterior of the protein as observed in NagD from *T. maritima* (accession code 1PW5). While it appears from the crystal structure that there is no conserved exterior binding site for the GlcNAc molecule on K-12 NagD (nor is it observed in the active site), it is likely that favorable interactions are still present on the protein exterior.

Like most HADSF members, the NagD protein is composed of a cap and a core domain (see Figure 1A) with the α/β fold (Rossmann type) core domain structurally conserved for positioning of the active site residues (HAD motifs I−IV) involved in the core catalytic chemistry (Figure 1B). In NagD, motif I positions the Asp9 nucleophile along with Asp11 (also denoted as the Asp-plus-2 residue) which participates in the acid/base catalysis necessary for phosphoryl transfer. Motif II contributes Thr42 which typically forms hydrogen bonds to the substrate phosphoryl group (*6*). Motif III Lys176 provides a positive charge for orientation of the Asp9 nucleophile via salt bridge formation and supplies additional electrostatic shielding for transfer of the phosphoryl group. The carboxylate residues of motif IV bind the metal (Mg^{2+}) cofactor.

The cap is located between HAD motifs II and III, defining NagD as a member of the type IIA HADSF subfamily. The

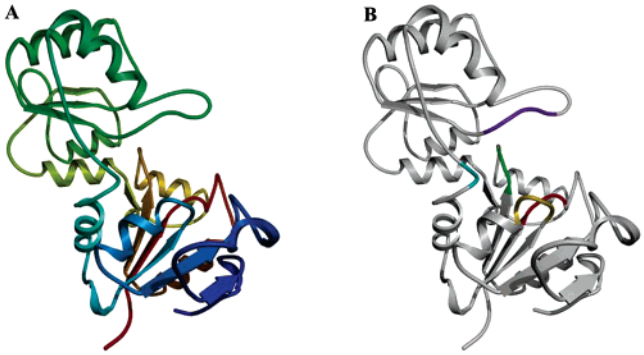

FIGURE 1: Structure of the NagD gene product depicted as a ribbon diagram (A) colored from the N-terminus (blue) to the C-terminus (red) and (B) colored by the four conserved HADSF motifs (loops) I−IV from the core domain in red, green, cyan, and gold, respectively. Additionally, the substrate specificity loop (motif V, purple) is located on the cap domain.

cap is comprised of residues 71−175 which form four α-helices and five β-strands (Figures 1 and 2A). This fold bears no resemblance to the α-helical bundle cap domain found in type I subfamily members (*15*−*20*) (Figure 2B), and its topology is distinct from the mixed α/β cap domain of type IIB subfamily members (Figure 2C) (*14*, *21*−*23*). The cap is attached to the core domain via two linkers (residues 170−72 and 171−175) comprised of conformationally flexible residues such as Thr and Gly flanked by residues important to secondary structure (such as the conserved helix breaker Pro171). Although the cap domain is inserted into the loop positioned between motifs II and III, as in all subfamily II members, the core domain retains a small insert (residues 14−21) between motifs I and II. This small insert, which forms a functional loop, is observed in other type II proteins as well as in the capless type III family members (*24*, *46*). In the type I members the insert is enlarged into an actual cap domain.

Although the structures of the cap domains in the HADSF vary (Figure 2), they conserve function, which is to desolvate the active site for catalysis and confer substrate specificity (*6*). Desolvation is achieved via the enclosure of the active site by the cap domain. In addition, the insert segment can act in conjunction with the cap domain to contribute to the water exclusive environment. Regions of the cap domain that interface with the core domain active site help to define the substrate binding site and, hence, the substrate specificity. Such a region is called the "substrate specificity loop" (see Figure 1B) (*13*, *14*), and its signature sequence motif is referred to as motif V. The substrate specificity loop contributes side chain interactions specific to the binding of the substrate leaving group, as opposed to the transferring phosphoryl group. Only in the enzyme (cap domain)−(core domain) closed conformation can this interaction occur, and thus by coupling cap closure with cap substrate interactions HADSF members are able to distinguish between the multitude of phosphate ester metabolites of the cell. In the case of the NagD structure, the Connolly surface calculation [performed with the program VOIDOO(*47*)] shows that the active site is accessible to bulk solvent. Thus, this structure represents the NagD in the cap-open conformation.

In addition, the structures of the *E. coli* NagD and those of *T. maritima*, *S. pneumoniae*, *S. pyogenes*, and *E. faecalis* NagD are similar, with rmsd values ranging from 1.55 to

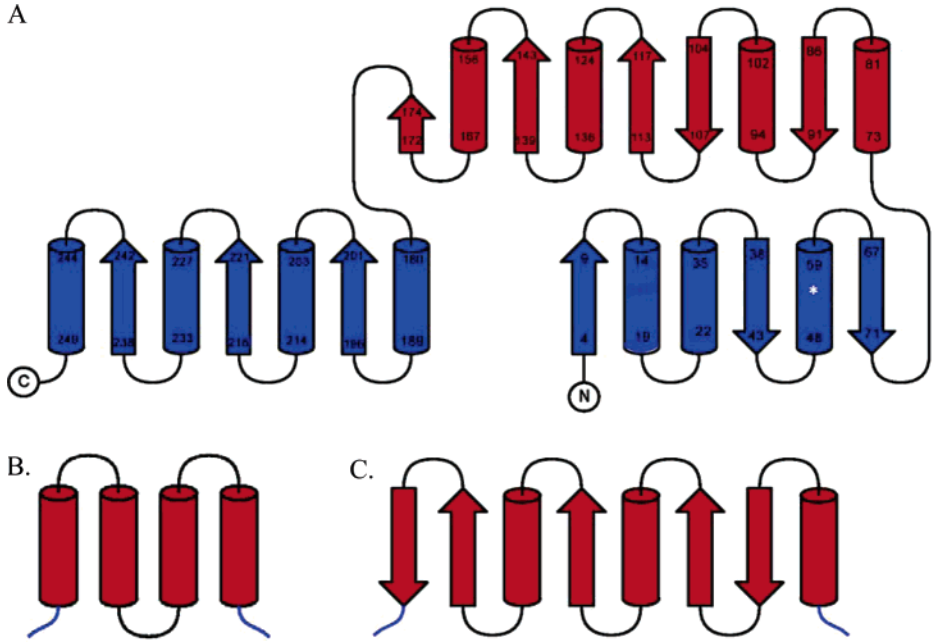

FIGURE 2: Topology diagrams of HAD family members (A) NagD cap (red) and core (blue) representing subfamily IIA and (B) the cap domains of subfamily I and (C) subfamily IIB. The position of the residue forming a salt bridge with the Asp-plus-2 position is denoted by a white asterisk.

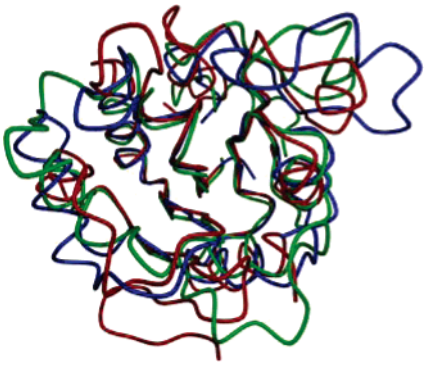

FIGURE 3: Overlay of only the core domain of *NagD* from subfamily IIA (blue), β-phosphoglucomutase from subfamily I (green), and BT4131, a fructose-6-phosphate phosphatase from subfamily IIB (red).

1.66 Å. The rmsd values are comparable when the cap and core domains are overlayed separately; thus the deviations are not due to variation in the relative orientation of the cap to the core domain. Therefore, these NagD structures also correspond to the cap-open form of the enzyme.

Active Site Structure. The cofactor binding and phosphoryl transfer sites in HADSF members are comprised mainly of residues from the core domain that, unlike the cap domain, does not differ between subclasses (Figure 3). The structure of NagD clearly shows a magnesium metal cofactor liganded within the core domain (Figure 4). The magnesium is bonded with octahedral coordination geometry (all bond lengths less than 2.3 Å) by residues Asp9, the carbonyl backbone of Asp11, and the side chain of Asp201 along with two waters, Wat47 and Wat93. Asp201 directly coordinates the Mg^{2+}, while Asp206 forms a hydrogen bond to one of the metal bound waters (Wat93).

Additional electron density was observed within bonding distance of the Mg^{2+} (2.3 Å). On the basis of the presence of phosphate in the crystallization conditions (0.4 M), a phosphate ion was assigned to this density (hereafter referred to as phosphate 2; labeled P2 in Figure 4). The phosphate ion occupies the position of the substrate phosphoryl group (the phosphate transfer site). In addition to the bond formed with the Mg^{2+} cofactor, there are multiple hydrogen bonds made to phosphate 2 originating from the backbone amide of Asp11, the side chain of Thr42 (from motif II), waters 8 and 93, and Lys176 (from motif III). A second density, similar in size, shape, and intensity, observed nearby in the active site was also modeled as a phosphate ion (Figure 4, labeled P1). This second phosphate (hereafter referred to as phosphate 1) is positioned proximal to (2.5 Å) Asp146 contributed by the cap domain along with Lys176 and Thr205 from the core. The structural results suggest that this particular region might be involved in binding the substrate leaving group in the enzyme−substrate complex, and this hypothesis is corroborated by the sequence alignment of NagD homologues (vide infra).

Phosphate 2 occupies the position of the phosphate product produced by hydrolytic cleavage of the aspartyl phosphate intermediate. In NagD, Asp9 functions as the mediator of the phosphoryl transfer from substrate to the water ligand. In the type I enzymes, acid/base catalysis is performed by a second Asp residue positioned two residues downstream from the Asp nucleophile (denoted the Asp-plus-2 residue) (*6*). On the basis of its sequence position and its placement relative to phosphate ligand 2 in the NagD structure Asp11 fulfills the role of acid/base catalyst. Asp11 places one carboxylate oxygen atom to engage in hydrogen bond formation (2.8 Å distance) with cap domain residue Arg55 and the other oxygen atom to donate a proton to the leaving group of substrate or to remove a proton from water in hydrolysis of the phosphoenzyme intermediate. Arg55 thus serves to orient the Asp11 for proton transfer during catalytic turnover.

The pairing of the Asp general acid/base with a hydrogen bond donor located in the cap domain is an invariant structural feature of the HADSF type II enzymes, which

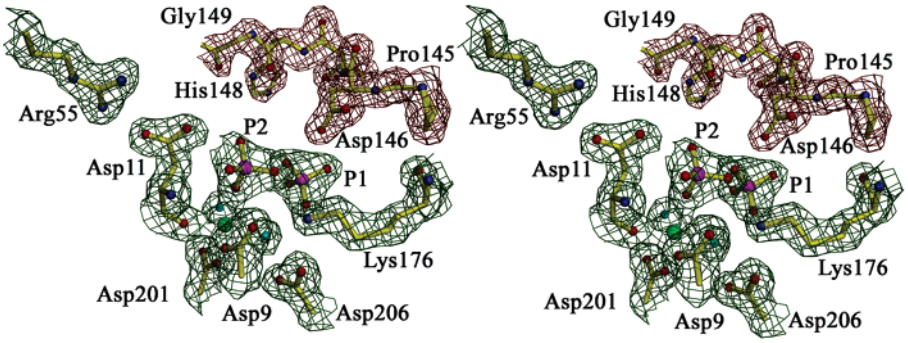

FIGURE 4: Stereoview of the active site of NagD with the $2F_o - F_c$ electron density map (green and red cages for the core can cap residues, respectively) contoured at 1σ. The Mg^{2+} cofactor is depicted as a green sphere, water molecules are depicted as cyan spheres, and the two phosphate ligands are labeled P1 and P2.

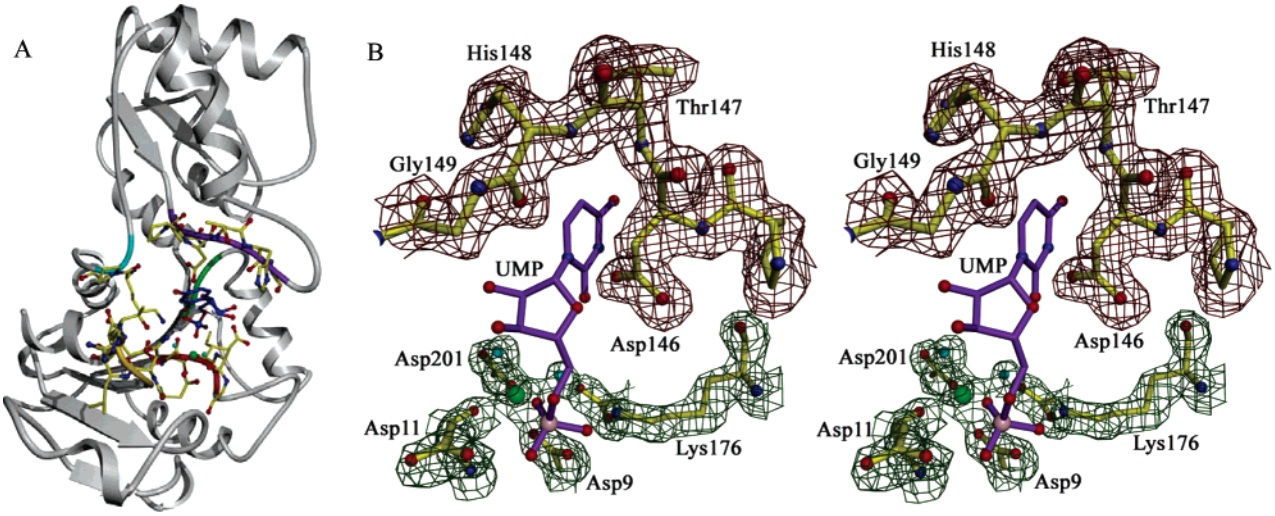

FIGURE 5: Structure of NagD with UMP (blue) manually docked in the active site depicted as (A) a loop diagram with catalytic loops colored as in Figure 1 and active site residues and putative substrate specificity residues shown as ball-and-stick models and (B) in stereo with active site residues and substrate specificity residues shown as ball-and-stick models. The Mg^{2+} cofactor is depicted as a green sphere. Electron density is depicted as described in Figure 4.

supports the catalytic role of the Asp. Sequence alignment (see Supporting Information, Figure S1) shows that the salt bridge partner is retained as either Arg or Lys in all type IIA NagD homologues except that of *S. pneumoniae*. In addition, the salt bridge is found in all of the type IIB enzymes with available structures (eight in total) to date with the exception of one enzyme (PDB accession number 1L6R) where the hydrogen bond donor is a Ser residue. The position of the salt bridge or hydrogen bond partner is also found among the type IIA and IIB members in the third helix of the core domain (see Figure 2A), a fact that hints that this feature was retained during the divergent evolution of the subclasses. Although they lack a cap domain, the type III HADSF members, represented by magnesium-dependent phosphatase I (MDP-1) and T4 polynucleotide kinase/phosphatase, also possess a salt bridge or hydrogen bond partner for the Asp-plus-2 general acid/base. Thus, the salt bridge/hydrogen bond partner serves the important role of orienting the Asp-plus 2 general acid/base within hydrogen bond distance of the leaving group of substrate in type II and type III HADSF phosphatases.

Determinants of Specificity. Based on an analysis of the strictly conserved residues of NagD enzymes from 12 different bacterial species, invariant residues were mapped to the active site region (Figure 5A). Among these are the previously mentioned core domain residues Asp9, Asp11, Thr42, Lys176, Asp206, and Asp201 comprising motifs I−IV, which function in cofactor and substrate binding and catalysis. Other strictly conserved residues located in the active site appear to be involved in substrate leaving group binding. Asp146 is contributed from the cap domain into the active site where it is held at a distance of 2.5 Å from phosphate 1 (P1 of Figure 4) by hydrogen bonds with the amide groups of Asn144 and Asn43. Notably, the loop (residues 144−149) in the cap on which Asp146 is positioned also contains the conserved residues Asn144 (except in *Thermotoga*) and Pro145. In type I and type IIB HAD members, substrate specificity is conferred via the "substrate specificity" loop found in the cap domain (*13*, *14*). Together, the structure and sequence data provide compelling evidence that the Asp146 cap domain loop imparts substrate specificity in NagD. Thus, functional homologues of NagD may be identified by the occurrence of the sequence motif NPD in the cap domain. By analogy, substrate recognition via the corresponding cap domain substrate specificity loop is anticipated for all type IIA subfamily members.

Substrate Range. The gene context of NagD within the *nagBACD* operon implicated it in the cellular processing of *N*-acetylglucosamine (GlcNAc), an essential component of cell wall biosynthesis (*48*−*53*). Previous studies had shown

Table 2: Kinetic Constants for NagD against Various Substrates[a]

substrate	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1}\,\mu M^{-1}$)
UMP	160 ± 38	5.1 ± 0.3	3.12×10^4
CMP	1470 ± 44	3.2 ± 0.3	2.13×10^3
GMP	400 ± 130	5.2 ± 0.6	1.28×10^4
AMP	840 ± 250	6.1 ± 0.7	7.29×10^3
ribose 5′-P	840 ± 124	3.4 ± 0.3	3.92×10^3
pyridoxal 5′-P	1500 ± 180	5.5 ± 0.3	3.69×10^3
G6P	5900 ± 750	24 ± 1.5	4.17×10^3
GlcN6P	6000 ± 2400	6.0 ± 1.1	1.00×10^3
glycerol 3-P	1700 ± 500	5.1 ± 0.8	3.00×10^3
dTMP	1500 ± 340	1.7 ± 0.1	1.13×10^3
dCMP	6000 ± 1700	2.6 ± 0.3	4.48×10^2
dAMP	6500 ± 1700	0.68 ± 0.08	1.05×10^2
dGMP	7600 ± 1500	0.24 ± 0.03	3.19×10^1
2dG6P	7700 ± 180	3.0 ± 0.3	3.92×10^2

[a] The following substrates were also tested but showed either no activity or had K_m values above 30 mM: GlcNAc6P; GlcNAc1P; glucosamine 1-phosphate; glucose 1-phosphate; D-fructose 6-phosphate; 2-deoxyglucose 6-phosphate; 2-deoxyribose 5′-phosphate; D-mannose 6-phosphate; D-mannose 1-phosphate; α-D-galactose 1-phosphate; fructose 6-phosphate; fructose 1,6-(bis)phosphate; fructose 2,6-(bis-)phosphate; sucrose 6-phosphate; maltose 1-phosphate; D-sorbitol 6-phosphate; trehalose 6-phosphate; ATP; ADP; UTP; UDP; glyceraldehyde 3-phosphate; glycerol 3-phosphate; thiamin phosphate, *O*-L-phosphotyrosine; *O*-L-phosphoserine; *O*-phosphoethanolamine.

that the gene product of NagA functions as a deacetylase, converting GlcNAc6P to GlcN6P. This suggested that either of these two phosphorylated compounds might be targeted substrates and that NagD may play a role in cell wall construction or disassembly.

A second clue regarding NagD function derives from the fusion between the MutT enzyme and the NagD homologue in *P. thiaminolyticus* (*26*) and *B. clausii* (unpublished submission to the NCBI Genome Project). The NagD homologue domain of this MutT-NagD fusion protein shares 28% with the *E. coli* NagD. The MutT protein is the prototypical member of the Nudix (*nu*cleoside *di*phosphates linked to moieties *X*) family of enzymes (*54*, *55*). This enzyme operates as a component of nucleotide pool cleansing by converting 8-oxo-dGTP to 8-oxo-dGMP and thereby preventing its incorporation into DNA (*27*, *28*, *30*, *31*, *56*, *57*). The NTP pyrophophosphatase activity of MutT and the fusion event between the MutT and NagD genes indicated that this NagD homologue domain, and possibly the *E. coli* NagD, might function in NDP or NMP hydrolysis.

The position of *nagD* within the *nagBACD* operon and the existence of the MutT fusion event in other bacteria pointed to two seemingly unrelated biochemical functions that are (if only by coincidence) linked by the hydrolysis of a sugar phosphate metabolite. A focused substrate screen, which took into account both lines of evidence, was applied to determine NagD substrate specificity based on experimental k_{cat} and K_m values. The results are shown in Table 2 (substrates tested that showed little to no activity are listed in the legend). NagD has the greatest activity against the 5′-phosphate substrates UMP and GMP, with K_m values of 160 and 400 μM and k_{cat} values of 5.1 and 5.2 s^{-1} for both substrates. The 5′-phosphates AMP and CMP also show activity, with K_m values of 840 and 1470 μM and k_{cat} values of 6.1 and 3.2 s^{-1}, respectively, for GMP and AMP. Overall, the substrate activity of NagD is highest for UMP and GMP, as seen by their respective k_{cat}/K_m values of 3.12×10^4 and 1.28×10^4.

The level of NagD discrimination between UMP, GMP, CMP, and AMP is judged to be low given that k_{cat}/K_m values for these substrates agree within an order of magnitude. Ribose 5′-phosphate showed reactivity with NagD but at an increased K_m of 840 μM and a decreased k_{cat} of 3.4 s^{-1}, yielding a k_{cat}/K_m value that is 3−10 times lower than for the best corresponding nucleotides. This indicates that NagD favorably interacts with the heterocyclic base of the nucleotide and is not simply binding to the sugar portion of the substrate. Although the heterocyclic base does facilitate productive binding, it is clear that NagD is not designed to target a single nucleotide.

In addition, NagD shows little to no reactivity with 2-deoxyribose 5′-phosphate and low activity toward the 2′-deoxyribose nucleoside monophosphates dTMP, dGMP, dCMP, and dAMP as judged from K_m values that are all in the low millimolar range and from the k_{cat}/K_m values that are 30−100 times lower than the k_{cat}/K_m value of the most active ribonucleotide monophosphate. To examine potential enzyme−substrate interactions and possible modes of substrate discrimination, the most active nucleotide substrate, UMP, was manually docked into the active site of NagD using the phosphate 2 ligand bound proximal to the nucleophile as a guide for placement of the corresponding 5′-phosphate. The docked structure (Figure 5B) shows polar interactions of the pyrimidine ring with the putative substrate specificity loop residue Asp146 as well as the backbone of His148. The docked structure does not show a basis for the discrimination between the 2′-deoxyibo- and ribonucleoside monophosphates. As has been noted with other members of the HADSF, substrate binding will induce cap closure. In the case of NagD structure, the cap domain is not fully closed. Therefore, we anticipate the possibility that, in the catalytically active enzyme−substrate complex, additional residues from the substrate specificity loop might come into close proximity with the substrate leaving group.

NagD showed some reactivity with other substrates, but the high K_m values rule them out as feasible physiological substrates. Most notable were pyridoxal 5-phosphate and GlcN6P. While the k_{cat} values are similar to the values for the NMPs, the K_m values of 1.5 and 6.0 mM are an order of magnitude higher than for the nucleotide substrates. Further, while activity with G6P was detected, the K_m value of 5.9 mM combined with the 10-fold decrease in substrate specificity also ruled it out as being a possible physiological substrate. The same was true for the substrate GlcNAc6P, which had an estimated K_m of >30 mM and showed little to no activity.

Homologues of the *E. coli* NagD protein are found across a variety of bacterial species, including some human pathogens such as *Staphylococcus aureus* and, *S. pneumoniae* (see Supporting Information, Figure S1). A recent investigation has shown that a disruptive insertion into the NagD gene of *S. aureus* causes reduced virulence in murine renal abscess models (*58*). The biosynthetic suppression and depletion of the intracellular nucleotide pool are standard mechanisms utilized by the sulfonamide class of synthetic drugs and hint at a possible mechanism for the observed reduced virulence resulting from the disruption of the native NagD gene. The physiological action of NagD as a NMP phosphatase could

be to maintain the cellular supply of nucleosides. The maintenance of a steady pool of the basic nucleosides for further processing and incorporation into RNA and for utilization in cell wall biosynthesis is an essential control point in the cell (*59*).

Alternatively, the nucleotide phosphohydrolase activity coupled with the sugar phosphate phosphohydrolase activity (see Table 2 and further discussion below) observed with the *E. coli* NagD suggests that the NagD domain of the MutT-NagD fusion protein is especially well equipped to perform a dual role in ADP-ribose catabolism. ADP-ribose is generated via nicotinate and nicotinamide metabolism (*60*) and can also be formed by the enzymatic removal of ADP-ribose from arginine residues in ADP-ribosylated proteins (*61*). The ADP-ribose is broken down in purine metabolism to AMP and ribose 5-phosphate by the action of the Nudix hydrolase adenosine diphosphoribose pyrophosphatase (EC 3.6.1.13) (*62*). Homologues of adenosine diphosphoribose pyrophosphatase occur in all three kingdoms (*63*). The fact that in NagD-Nudix protein the Nudix domain shares the same catalytic motif as adenosine diphosphoribose pyrophosphatase (with 27.0% overall sequence identity) implies a central role of NagD in ADP-ribose metabolism. That is, once ADP-ribose has been hydrolyzed to AMP plus ribose 5-phosphate by the Nudix hydrolase, the two products, AMP and ribose 5-phosphate, can both be dephosphorylated by NagD to generate ribose and adenosine. This role is consistent with the finding that both the nucleoside monophosphates and ribose 5-phosphate serve as good substrates for *E. coli* NagD (vide supra). The NagD-Nudix protein occurs in the two species *P. thiaminolyticus* and *B. clausii* KSM-K16 with high sequence identity (55.9%), demonstrating that these fusion proteins are true orthologues. The gene context is different for the two bacterial fusion proteins, yielding no additional clues to their common function. Indeed, although the proximity of the fusion protein from *P. thiaminolyticus* to the thiaminase I gene from a thiamin degradation pathway suggests a possible role in hydrolysis of thiamin phosphate, lack of activity toward this substrate (see Table 2, legend) rules out this option.

A second possibility connecting the NagD and Nudix functionalities in a single pathway is the use of the Nudix protein as a nucleoside triphosphate pyrophosphatase which would provide nucleoside monophosphate substrate for NagD. Liberated nucleosides would be fed directly into the purine and pyrimidine pathways.

The *E. coli* NagD, on the other hand, appears to function within the context of carbohydrate metabolism. Consistent with its occurrence in the *Nag* operon, NagD may be involved in the recycling of cell wall metabolites. Both GDP and UDP are liberated by the action of cellulose synthase (*64*) on GDP-glucose and UDP-glucose, respectively, in the formation of cellulose. Alternatively, GDP can be liberated by the action of an atypical Nudix enzyme, GDP-mannose glycosyl hydrolase which catalyzes the hydrolysis of GDP-mannose and GDP-glucose by substitution at C1 of the sugar (*65*) as part of regulation of cell wall biosynthesis. In either path, NMP would be generated by the action of a nucleoside diphosphate phosphohydrolase and it in turn hydrolyzed by NagD.

Finally, an alternative for the role of the *E. coli* NagD, implied by its fusion to MutT, is in nucleotide pool cleansing. This possibility would suggest that NagD is an essential enzyme for maintaining functional, error-free RNA in the cell (lack of activity of NagD toward deoxynucleoside phosphates rules out any role in DNA metabolism). By eliminating unwanted derivatives and/or oxygenated species of the NMPs, NagD may relieve the oxidative stress encountered by the cell. In this case, NagD would be expected to show greater specificity as reflected in k_{cat}/K_m toward the oxidized or derivative form of NMPs than for the standard NMPs. This role is not connected in any obvious manner to the *Nag* pathway, and for this reason we consider it to be an unlikely function for the *E. coli* NagD.

Specificity Overlap among Subfamily II Members. Some aspects of specificity of HADSF members can be predicted from knowledge of the subclass and oligomerization state alone. Usually, type I and type II capped proteins take small substrates and noncapped proteins (type III) take large substrates. There are, however, exceptions to this rule: *N*-acetylneuraminate-9-phosphate phosphatase (Z. Lu, unpublished results) and 3-deoxy-D-mannooctulosonate-8-phosphate phosphatase (*24*, *25*) have no cap but use an adjacent subunit core to shield the active site during turnover. On the other side of the coin, the lipid phosphatase domain of the bifunctional phosphatase−epoxide hydrolase apparently sandwiches the hydrophobic lipid tail in the interface between core and cap domains, using it in effect as a seal (*66*).

However, the general rule is that capped HADSF members will target phosphate esters in which the size of the organic substituent does not exceed the size of the cavity present in the cap-closed conformation. This concept has been used to generate a "solvent cage" from the structure of the empty active site to predict the size and shape of the substrate (*14*). Such a line of reasoning leads to the concept that the cap position and fold code for a substrate that conforms to certain structural features (size, shape, and/or polarity). In this case, members of a particular subfamily would be recruited for action on a structural subclass of substrates. A more refined version of this line of reasoning is that the position or fold of the cap might facilitate or limit coordination of domain closure to the positioning of cap and core domain residues for catalysis. In this case, some chemistries would require the use of particular cap/core combinations (i.e., subfamilies). Alternatively, the evolution of function might have been driven by other factors such as the frequency of gene duplication or the stability and plasticity of a given fold.

In principle, the hypothesis of a domain substrate code is a simple one to examine as one needs only to link structure with function. With the description of the specificity of NagD (vide supra), the substrate range of a number of HADSF members from different structural classes has now been measured. The HAD subclass I members possess fairly high specificities for substrate. These include β-phosphoglucomutase with $k_{cat}/K_m = 1.2 \times 10^6\ M^{-1}\ s^{-1}$ (for β-glucose 1-phosphate using α-glucose 1,6-(bis)phosphate as activator (*67*), phosphonoacetaldehyde hydrolase with $k_{cat}/K_m = 5.8 \times 10^5\ M^{-1}\ s^{-1}$ (*68*), and bone PHOSPHO1 with $k_{cat}/K_m = 9.9 \times 10^3\ M^{-1}\ s^{-1}$ [for phosphoethanolamine (*69*)]. In sharp contrast, the HADSF members from subfamilies IIA and IIB all have in common relatively low substrate specificity and catalytic efficiencies ($k_{cat}/K_m \leq 1 \times 10^5\ M^{-1}\ s^{-1}$), and the boundaries defining physiological substrates are somewhat

overlapping. The enzyme might show a unique preference for a particular substrate, but there is no indication that the enzyme is honed for that substrate. For instance, NagD shows preference for nucleotide monophosphates with $k_{cat}/K_m = 3.12 \times 10^4\ M^{-1}\ s^{-1}$; however, the efficiency of NagD toward ribose 5′-phosphate is only ~10-fold lower and similar to the specificity constant of the hexose-phosphate phosphatase, BT4131, for that substrate (*14*). Similarly, the protein YbiV (*22*) has $k_{cat}/K_m \sim 10^2\ M^{-1}\ s^{-1}$ for ribose 5-phosphate. In addition, NagD, BT4131, and YbiV all hydrolyze glucose 6-phosphate and pyridoxal 5-phosphate fairly efficiently ($k_{cat}/K_m \sim 10^2-10^3\ M^{-1}\ s^{-1}$). Thus, there is significant overlap in substrate specificity for enzymes which most likely exist as paralogues, given the redundancy of the occurrence of the HADSF folds in a single organism. There is not a lack of catalytic efficiency (reflected in turnover number) of any one enzyme toward a particular substrate but rather an ability to accept other substrates with similar efficiency. This differs from the concept of substrate promiscuity (*70*−*72*) in a superfamily since that phenomenon describes vestigial enzymatic activity shared by an enzyme with its evolutionary progenitor. Such activity occurs with significantly reduced efficiency compared to the preferred substrate of the enzyme in question.

What is the physiological significance of a collection of phosphatases within the same organism that are not scavengers but yet do not possess a clear definition of "substrate"? Two solutions to such a biochemical quandary are (1) regulation of the gene in question to limit the exposure of a given substrate to an enzyme with a loose specificity and (2) compartmentalization or control of the local concentration of the substrate itself. As a possible case in point, NagD in *E. coli* is under the control of a different promoter than NagA and NagB (*53*), connoting individualized control of the levels of this protein. Control of HAD paralogues on the cellular level would be required by the existence of such blurring of the lines between substrates of phosphatases that are so numerous within a single organism.

ACKNOWLEDGMENT

Data were collected at the National Synchrotron Light Source, Brookhaven National Laboratories, at Beamline X26C. We thank Dr. Annie Heroux, Brookhaven National Laboratories, for excellent technical assistance. We also thank Dr. Zhibing Lu for generous assistance with kinetics.

SUPPORTING INFORMATION AVAILABLE

One figure showing sequence alignment of NagD homologues from various bacterial species. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Aravind, L., Galperin, M. Y., and Koonin, E. V. (1998) The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold, *Trends Biochem. Sci. 23*, 127−129.
2. 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.
3. 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.
4. Collet, J. F., Stroobant, V., and Van Schaftingen, E. (2002) Evidence for phosphotransferases phosphorylated on aspartate residue in N-terminal DXDX(T/V) motif, *Methods Enzymol. 354*, 177−188.
5. Selengut, J. D. (2001) MDP-1 is a new and distinct member of the haloacid dehalogenase family of aspartate-dependent phosphohydrolases, *Biochemistry 40*, 12704−12711.
6. Allen, K. N., and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: evolution of a catalytic scaffold, *Trends Biochem. Sci. 29*, 495−503.
7. Peisach, E., Selengut, J., Dunaway-Mariano, D., and Allen, K. N. (2004) Structure of the magnesium-dependent protein tyrosine phosphatase, MDP-1, *Biochemistry 43*, 12770−12779.
8. 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 (Cambridge) 10*, 1249−1260.
9. 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.
10. 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.
11. Li, Y. F., Hata, Y., Fujii, T., Hisano, T., Nishihara, M., Kurihara, T., and Esaki, N. (1998) Crystal structures of reaction intermediates of L-2-haloacid dehalogenase and implications for the reaction mechanism, *J. Biol. Chem. 273*, 15035−15044.
12. Morais, M. C., Zhang, G., Zhang, W., Olsen, D. B., Dunaway-Mariano, D., and Allen, K. N. (2004) X-ray crystallographic and site-directed mutagenesis analysis of the mechanism of Schiff-base formation in phosphonoacetaldehyde hydrolase catalysis, *J. Biol. Chem. 279*, 9353−9361.
13. 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.
14. 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.
15. 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 (Cambridge) 9*, 65−71.
16. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2002) Caught in the act: the structure of phosphorylated β-phosphoglucomutase from *Lactococcus lactis*, *Biochemistry 41*, 8351−8359.
17. Morais, M. C., Zhang, W., Baker, A. S., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2000) The crystal structure of *Bacillus cereus* phosphonoacetaldehyde hydrolase: Insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily, *Biochemistry 39*, 10385−10396.
18. 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.
19. Ridder, I. S., Rozeboom, H. J., Kalk, K. H., Janssen, D. B., and Dijkstra, B. W. (1997) Three-dimensional structure of L-2-haloacid dehalogenase from *Xanthobacter autotrophicus* GJ10 complexed with the substrate-analogue formate, *J. Biol. Chem. 272*, 33015−33022.
20. Argiriadi, M. A., Morisseau, C., Hammock, B. D., and Christianson, D. W. (1999) Detoxification of environmental mutagens and carcinogens: structure, mechanism, and evolution of liver epoxide hydrolase, *Proc. Natl. Acad. Sci. U.S.A. 96*, 10637−10642.
21. Shin, D. H., Roberts, A., Jancarik, J., Yokota, H., Kim, R., Wemmer, D. E., and Kim, S. H. (2003) Crystal structure of a phosphatase with a unique substrate binding domain from *Thermotoga maritima*, *Protein Sci. 12*, 1464−1472.
22. Roberts, A., Lee, S. Y., McCullagh, E., Silversmith, R. E., and Wemmer, D. E. (2005) Ybiv from *Escherichia coli* K12 is a HAD phosphatase, *Proteins 58*, 790−801.
23. Kim, Y., Yakunin, A. F., Kuznetsova, E., Xu, X., Pennycooke, M., Gu, J., Cheung, F., Proudfoot, M., Arrowsmith, C. H.,

Joachimiak, A., Edwards, A. M., and Christendat, D. (2004) Structure- and function-based characterization of a new phosphoglycolate phosphatase from *Thermoplasma acidophilum*, *J. Biol. Chem. 279*, 517−526.
24. 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.
25. 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.
26. Costello, C. A., Kelleher, N. L., Abe, M., McLafferty, F. W., and Begley, T. P. (1996) Mechanistic studies on thiaminase I. Overexpression and identification of the active site nucleophile, *J. Biol. Chem. 271*, 3445−3452.
27. Fowler, R. G., and Schaaper, R. M. (1997) The role of the mutT gene of *Escherichia coli* in maintaining replication fidelity, *FEMS Microbiol. Rev. 21*, 43−54.
28. Ito, R., Hayakawa, H., Sekiguchi, M., and Ishibashi, T. (2005) Multiple enzyme activities of *Escherichia coli* MutT protein for sanitization of DNA and RNA precursor pools, *Biochemistry 44*, 6670−6674.
29. McLennan, A. G. (1999) The MutT motif family of nucleotide phosphohydrolases in man and human pathogens, *Int. J. Mol. Med. 4*, 79−89.
30. Saraswat, V., Massiah, M. A., Lopez, G., Amzel, L. M., and Mildvan, A. S. (2002) Interactions of the products, 8-oxo-dGMP, dGMP, and pyrophosphate with the MutT nucleoside triphosphate pyrophosphohydrolase, *Biochemistry 41*, 15566−15577.
31. Taddei, F., Hayakawa, H., Bouton, M., Cirinesi, A., Matic, I., Sekiguchi, M., and Radman, M. (1997) Counteraction by MutT protein of transcriptional errors caused by oxidative damage, *Science 278*, 128−130.
32. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673−4680.
33. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L. L. (2000) The Pfam protein families database, *Nucleic Acids Res. 28*, 263−266.
34. Gouet, P., Robert, X., and Courcelle, E. (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins, *Nucleic Acids Res. 31*, 3320−3323.
35. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem. 72*, 248−254.
36. Jancerik, J., and Kim, S. H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins, *J. Appl. Crystallogr. 24*, 409−411.
37. Bolen, D. W. (2004) Effects of naturally occurring osmolytes on protein stability and solubility: issues important in protein crystallization, *Methods 34*, 312−322.
38. Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 56* (Part 12), 1622−1624.
39. 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* (Part 5), 905−921.
40. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) SWISS-MODEL: an automated protein homology-modeling server, *Nucleic Acids Res. 31*, 3381−3385.
41. Brünger, A. T. (1992) The free R value: A novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472−474.
42. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2126−2132.
43. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2184−2195.
44. Cleland, W. W. (1979) Statistical analysis of enzyme kinetic data, *Methods Enzymol. 63*, 103−138.
45. Martin, B., Pallen, C. J., Wang, J. H., and Graves, D. J. (1985) Use of fluorinated tyrosine phosphates to probe the substrate specificity of the low molecular weight phosphatase activity of calcineurin, *J. Biol. Chem. 260*, 14932−14937.
46. Peisach, E., Selengut, J. D., Dunaway-Mariano, D., and Allen, K. N. (2004) X-ray crystal structure of the hypothetical phosphotyrosine phosphatase MDP-1 of the haloacid dehalogenase superfamily, *Biochemistry 43*, 12770−12779.
47. Kleywegt, G. J., and Jones, T. A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures, *Acta Crystallogr. D50*, 178−185.
48. Peri, K. G., Goldie, H., and Waygood, E. B. (1990) Cloning and characterization of the *N*-acetylglucosamine operon of *Escherichia coli*, *Biochem. Cell Biol. 68*, 123−137.
49. Souza, J. M., Plumbridge, J. A., and Calcagno, M. L. (1997) *N*-acetylglucosamine-6-phosphate deacetylase from *Escherichia coli*: purification and molecular and kinetic characterization, *Arch. Biochem. Biophys. 340*, 338−346.
50. Plumbridge, J. A., Cochet, O., Souza, J. M., Altamirano, M. M., Calcagno, M. L., and Badet, B. (1993) Coordinated regulation of amino sugar-synthesizing and -degrading enzymes in *Escherichia coli* K-12, *J. Bacteriol. 175*, 4951−4956.
51. Plumbridge, J., and Vimr, E. (1999) Convergent pathways for utilization of the amino sugars *N*-acetylglucosamine, *N*-acetylmannosamine, and *N*-acetylneuraminic acid by *Escherichia coli*, *J. Bacteriol. 181*, 47−54.
52. Yamano, N., Oura, N., Wang, J., and Fujishima, S. (1997) Cloning and sequencing of the genes for *N*-acetylglucosamine use that construct divergent operons (nagE-nagAC) from *Vibrio cholerae* non-O1, *Biosci. Biotechnol. Biochem. 61*, 1349−1353.
53. Plumbridge, J. A. (1989) Sequence of the nagBACD operon in *Escherichia coli* K12 and pattern of transcription within the nag regulon, *Mol. Microbiol. 3*, 505−515.
54. Mildvan, A. S., Xia, Z., Azurmendi, H. F., Saraswat, V., Legler, P. M., Massiah, M. A., Gabelli, S. B., Bianchet, M. A., Kang, L. W., and Amzel, L. M. (2005) Structures and mechanisms of Nudix hydrolases, *Arch. Biochem. Biophys. 433*, 129−143.
55. Bessman, M. J., Frick, D. N., and O'Handley, S. F. (1996) The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes, *J. Biol. Chem. 271*, 25059−25062.
56. Fowler, R. G., White, S. J., Koyama, C., Moore, S. C., Dunn, R. L., and Schaaper, R. M. (2003) Interactions among the *Escherichia coli* mutT, mutM, and mutY damage prevention pathways, *DNA Repair (Amsterdam) 2*, 159−173.
57. Tassotto, M. L., and Mathews, C. K. (2002) Assessing the metabolic function of the MutT 8-oxodeoxyguanosine triphosphatase in *Escherichia coli* by nucleotide pool analysis, *J. Biol. Chem. 277*, 15807−15812.
58. Begun, J., Sifri, C. D., Goldman, S., Calderwood, S. B., and Ausubel, F. M. (2005) *Staphylococcus aureus* virulence factors identified by using a high-throughput *Caenorhabditis elegans*-killing model, *Infect. Immun. 73*, 872−877.
59. Rampazzo, C., Ferraro, P., Pontarin, G., Fabris, S., Reichard, P., and Bianchi, V. (2004) Mitochondrial deoxyribonucleotides, pool sizes, synthesis, and regulation, *J. Biol. Chem. 279*, 17019−17026.
60. Ueda, K., Fukushima, M., Okayama, H., and Hayaishi, O. (1975) Nicotinamide adenine dinucleotide glycohydrolase from rat liver nuclei. Isolation and characterization of a new enzyme, *J. Biol. Chem. 250*, 7541−7546.
61. Moss, J., Jacobson, M. K., and Stanley, S. J. (1985) Reversibility of arginine-specific mono(ADP-ribosyl)ation: identification in erythrocytes of an ADP-ribose-L-arginine cleavage enzyme, *Proc. Natl. Acad. Sci. U.S.A. 82*, 5603−5607.
62. Doherty, M. D., and Morrison, J. F. (1962) The hydrolysis of adenosine diphosphate ribose by a specific phosphohydrolase of rabbit-muscle extracts, *Biochim. Biophys. Acta 65*, 364−366.
63. Sheikh, S., O'Handley, S. F., Dunn, C. A., and Bessman, M. J. (1998) Identification and characterization of the Nudix hydrolase from the archaeon, *Methanococcus jannaschii*, as a highly specific ADP-ribose pyrophosphatase, *J. Biol. Chem. 273*, 20924−20928.
64. Chambers, J., and Elbein, A. D. (1970) Biosynthesis of glucans in mung bean seedlings. Formation of beta-(1,4)-glucans from GDP-glucose and beta-(1,3)-glucans from UDP-glucose, *Arch. Biochem. Biophys. 138*, 620−631.
65. Gabelli, S. B., Bianchet, M. A., Azurmendi, H. F., Xia, Z., Sarawat, V., Mildvan, A. S., and Amzel, L. M. (2004) Structure and

mechanism of GDP-mannose glycosyl hydrolase, a Nudix enzyme that cleaves at carbon instead of phosphorus, *Structure 12*, 927−935.
66. Gomez, G. A., Morisseau, C., Hammock, B. D., and Christianson, D. W. (2004) Structure of human epoxide hydrolase reveals mechanistic inferences on bifunctional catalysis in epoxide and phosphate ester hydrolysis, *Biochemistry 43*, 4716−4723.
67. 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.
68. 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*, 144−151.
69. Roberts, S. J., Stewart, A. J., Sadler, P. J., and Farquharson, C. (2004) Human PHOSPHO1 exhibits high specific phosphoethanolamine and phosphocholine phosphatase activities, *Biochem. J. 382*, 59−65.
70. O'Brien, P. J., and Herschlag, D. (2001) Functional interrelationships in the alkaline phosphatase superfamily: phosphodiesterase activity of *Escherichia coli* alkaline phosphatase, *Biochemistry 40*, 5691−5699.
71. Joerger, A. C., Mayer, S., and Fersht, A. R. (2003) Mimicking natural evolution in vitro: An *N*-acetylneuraminate lyase mutant with an increased dihydrodipicolinate synthase activity, *Proc. Natl. Acad. Sci. U.S.A. 100*, 5694−5699.
72. Schmidt, D. M. Z., Mundorff, E. C., Dojka, M., Bermudez, E., Ness, J. E., Govindarajan, S., Babbitt, P. C., Minshull, J., and Gerlt, J. A. (2003) Evolutionary potential of $(\beta/\alpha)_8$-barrels: Functional promiscuity produced by single substitutions in the enolase superfamily, *Biochemistry 42*, 8387−8393.

BI051842J