

Divergence of Structure and Function in the Haloacid Dehalogenase Enzyme Superfamily: *Bacteroides thetaiotaomicron* BT2127 Is an Inorganic Pyrophosphatase

Hua Huang,[†] Yury Patskovsky,[‡] Rafael Toro,[‡] Jeremiah D. Farelli,[§] Chetanya Pandya,^{||} Steven C. Almo,^{*,‡} Karen N. Allen,^{*,§} and Debra Dunaway-Mariano^{*,†}

[†]Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico 87131, United States

[‡]Department of Biochemistry, Albert Einstein College of Medicine, Jack and Pearl Resnick Campus, 1300 Morris Park Avenue, Bronx, New York 10461, United States

[§]Department of Chemistry, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215-2521, United States

^{||}Bioinformatics Graduate Program, Boston University, 24 Cummington Street, Boston, Massachusetts 02215-2521, United States

S Supporting Information

ABSTRACT: The explosion of protein sequence information requires that current strategies for function assignment evolve to complement experimental approaches with computationally based function prediction. This necessitates the development of strategies based on the identification of sequence markers in the form of specificity determinants and a more informed definition of orthologues. Herein, we have undertaken the function assignment of the unknown haloalkanoate dehalogenase superfamily member BT2127 (Uniprot accession code Q8A5 V9) from *Bacteroides thetaiotaomicron* using an integrated bioinformatics–structure–mechanism approach. The substrate specificity profile and steady-state rate constants of BT2127 (with a $k_{\text{cat}}/K_{\text{m}}$ value for pyrophosphate of $\sim 1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$), together with the gene context, support the assigned in vivo function as an inorganic pyrophosphatase. The X-ray structural analysis of wild-type BT2127 and several variants generated by site-directed mutagenesis shows that substrate discrimination is based, in part, on active site space restrictions imposed by the cap domain (specifically by residues Tyr76 and Glu47).

Structure-guided site-directed mutagenesis coupled with kinetic analysis of the mutant enzymes identified the residues required for catalysis, substrate binding, and domain–domain association. On the basis of this structure–function analysis, the catalytic residues Asp11, Asp13, Thr113, and Lys147 as well the metal binding residues Asp171, Asn172, and Glu47 were used as markers to confirm BT2127 orthologues identified via sequence searches. This bioinformatic analysis demonstrated that the biological range of BT2127 orthologue is restricted to the phylum Bacteroidetes/Chlorobi. The key structural determinants in the divergence of BT2127 and its closest homologue, β -phosphoglucomutase, control the leaving group size (phosphate vs glucose phosphate) and the position of the Asp acid/base in the open versus closed conformations. HADSF pyrophosphatases represent a third mechanistic and fold type for bacterial pyrophosphatases.



Functional annotation of the rapidly growing numbers of enzyme sequences in the public databases is a challenge that should be met quickly so that the scientific community can effectively mine these databases. Ultimately, enzyme function annotation must be performed computationally because experimentally based function determinations are far too time-consuming. The Enzyme Function Initiative (<http://enzyme.function.org/>) was established in 2010, with the long-term goal of defining an overall strategy that merges experimental approaches with complementary computationally based function prediction. The work reported herein is an early product of this initiative that describes an integrated bioinformatic–structure–mechanism approach to define the function of a member of the haloalkanoate dehalogenase superfamily (HADSF). Beyond this single functional annotation, our integrated approach provides for the reliable tracking and annotation of orthologues.

The HADSF was chosen to explore methods of function assignment because of its large size, its presence in all organisms represented in the NCBI genomic database, and the diversity in the substrates upon which its members act.^{1,2} Although the HADSF members include dehalogenases, phosphonates, and phosphomutases, the vast majority are phosphatases. HADSF phosphatases can number up to 30 in bacterial cells and up to 200–300 in eukaryotic cells.³ A central question to be addressed is how many unique in vivo functions exist within this subfamily. The study of bacterial, as opposed to eukaryotic, HADSF phosphatases avoids the challenges arising from splice variants and reduces the complications associated with protein binding partners and subcellular localizations as modulators of in vivo function. Functional assignment of the

Received: July 30, 2011

Revised: September 2, 2011

Published: September 6, 2011



bacterial HADSF phosphatases thus entails identification of the physiological substrate among the pool of organophosphate metabolites that exist in the cell, followed by defining the biological context of the catalyzed reaction.

Earlier work on the HADSF that involved the screening of 23 cytoplasmic *Escherichia coli* HADSF phosphatases with an 80-compound library of known metabolites made clear the challenge of identifying the physiological substrates.⁴ Specifically, most of the phosphatases displayed activities toward multiple substrates, which in numerous instances were structurally dissimilar. In addition, there also existed some degree of overlap in the substrate ranges between phosphatases. Nevertheless, the physiological substrates of several of the 23 *E. coli* HADSF phosphatases have been convincingly identified and their *in vivo* functions assigned. Included in this group are phosphoserine phosphatase (SerB),^{4,5} histidinol phosphate phosphatase (HisB),⁶ trehalose-6-phosphate phosphatase (OtsB),⁷ 3-deoxy-D-manno-octulosonate-8-phosphate phosphatase (YrbI),⁸ and glycerol-manno-heptose 1,7-bisphosphate phosphatase (YaeD).⁹ Notably, each of these phosphatases functions in a biosynthetic pathway. The *in vivo* function assignment was facilitated by gene context as well as by a narrow substrate range coupled with a high k_{cat}/K_m value for the physiological substrate. In addition, gene knockout experiments have been used to show that the promiscuous phosphatases YniC⁴ and YjjG¹⁰ can function *in vivo* to remove toxic 2-deoxyglucose 6-phosphate and noncanonical nucleoside 5'-monophosphates, respectively. Genetic experiments have also shown that YjjG performs in the thymidine salvage pathway as a dUMP nucleotidase.¹¹

To gain insight into the question of why many HADSF phosphatases from two different species of bacteria share a common function while many do not, with the ultimate goal of facilitating orthologue assignment, we have initiated a side-by-side comparison of the HADSF phosphatases from *E. coli* and *Bacteroides thetaiotaomicron*. Even though these two species derive from different phyla, they share a common habitat, the human gut. *E. coli* possesses 23 HADSF phosphatases, and *B. thetaiotaomicron* possesses 19, two of which have been experimentally characterized. One is 3-deoxy-D-manno-octulosonate-8-phosphate phosphatase (BT1677),¹² an orthologue to *E. coli* YrbI of the 3-deoxy-D-manno-octulosonate pathway. The second is 2-keto-3-deoxy-D-glycero-D-galactonate-9-phosphate phosphatase (BT1713) of the 2-keto-3-deoxy-D-glycero-D-galactonate pathway,¹³ which, notably, is not a pathway found in *E. coli*.

In this study, we have undertaken the functional assignment of HADSF member BT2127 (Uniprot accession code Q8ASV9), a member of unknown function from *B. thetaiotaomicron*. We report the function of BT2127 as an inorganic pyrophosphatase and detail the X-ray structure and biological range, in contrast to those of the closest homologue, β -phosphoglucomutase (β -PGM), and those of a HADSF Archeal inorganic pyrophosphatase.

MATERIALS AND METHODS

General. All chemicals and buffers were purchased from Sigma-Aldrich. The sources of the gene cloning materials are as follows: primers, T4 DNA ligase, and restriction enzymes (Invitrogen); *E. coli* BL21 (DE3) competent cells and *Pfu* Turbo polymerases (Stratagene); pET14b, pET23a, and pET28a vector kits (Novagen); and Qiaprep Spin Miniprep Kit (Qiagen). DEAE Sepharose was from Amersham

Biosciences. Butyl- and Phenyl-Sepharose resins were purchased from Sigma-Aldrich, and the Ni-NTA resin was from Qiagen. Snakeskin-pleated dialysis tubing was purchased from Thermo Scientific. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) analysis was performed with a 12% acrylamide running gel and a 4% stacking gel (Bio-Rad, Hercules, CA). Protein solutions were concentrated using a 10K Amicon Ultra Centrifugal filter (Millipore). The nucleotide sequence of each cloned gene or site-directed mutation-bearing gene was determined by the Center for Genetics in Medicine, University of New Mexico. Electrospray mass spectrometry (ES-MS) determinations were conducted by the University of New Mexico Mass Spectrometry Facility. Protein concentrations were determined using the Bradford assay kit from Sigma-Aldrich with BSA standards.

Preparation of Recombinant Wild-Type and Mutant BT2127. The DNA encoding the NP₈₁₁₀₄₀ gene from *B. thetaiotaomicron* was amplified by polymerase chain reaction (PCR) using the genomic DNA from *B. thetaiotaomicron* (ATCC 29148D), *Pfu* Turbo DNA polymerase, and oligonucleotide primers (5'-GATTCCATCTAACCCACA-TATGAGAAAGAAAC and 5'-CTTTTGCATAGTAG-GATCCGTATTTATAGGT) containing restriction endonuclease cleavage sites *Nde*I and *Bam*HI. The pET-28A vector, cut by restriction enzymes *Nde*I and *Bam*HI, was ligated to the PCR product that had been purified and digested with the same restriction enzymes. The ligation product was used to transform *E. coli* BL21(DE3) competent cells that were then grown on a kanamycin-containing agar plate. A selected colony was checked for BT2127 expression, and the isolated plasmid was sequenced to verify the correct gene sequence. For BT2127 preparation, the transformed cells (9 L) were grown at 25 °C with agitation at 200 rpm in Luria broth containing 40 μ g/mL kanamycin to an OD₆₀₀ of 0.6–0.7 and then induced for 12 h at 20 °C with 0.4 mM isopropyl α -D-thiogalactopyranoside. The cells were harvested by centrifugation (7855g for 15 min at 4 °C) to yield 2.2 g/L of culture medium. The cell pellet was suspended (1 g of wet cells/10 mL) in ice-cold buffer A [50 mM Tris (pH 7.6), 5 mM MgCl₂, and 1 mM DTT]. The cell suspension was passed through a French press at 1200 psi before centrifugation at 48384g and 4 °C for 45 min. The supernatant was loaded onto a 40 cm \times 5 cm DEAE-Sepharose 50–120 column, which was eluted with a 2 L linear gradient of NaCl (from 0 to 0.5 M) in buffer A. The column fractions were analyzed by SDS–PAGE. The desired fractions were combined and loaded onto a 10 mL Ni-NTA agarose column at 4 °C. After the column had been washed with 100 mL of buffer B [50 mM NaH₂PO₄, 300 mM NaCl, and 20 mM imidazole (pH 8.0)], the enzyme was eluted with 200 mL of elution buffer C [50 mM NaH₂PO₄, 300 mM NaCl, and 250 mM imidazole (pH 8.0)]. The column fractions were analyzed by SDS–PAGE, and the desired fractions were combined and concentrated with an Amicon Ultrafiltration apparatus (PM10) before dialysis at 4 °C against buffer A. The final yield was 11 mg of BT2127/g of wet cells.

Site-directed mutagenesis was performed with a PCR-based strategy, using the BT2127-pET-28A clone as template and commercial primers. The protein variants were purified in the same manner described for wild-type BT2127.

BT2127 Molecular Mass and Determination of the Quaternary Structure. The theoretical subunit molecular mass of recombinant BT2127 was calculated by using the amino acid composition, derived from the gene sequence, and

the ExPASy Molecular Biology Server program Compute pI/MW. The subunit size of recombinant BT2127 was estimated by SDS–PAGE analysis, which included the molecular mass standards from New England Biolabs. The exact subunit mass was determined by MS-ES mass spectrometry. The molecular mass of native BT2127 was estimated by FPLC gel filtration column chromatography against protein standards (13.7–220 kDa from GE Healthcare). Elution of protein from the 1.6 cm × 60 cm Sephacryl S-200HR column (GE Healthcare) was performed at 4 °C with buffer D [50 mM HEPES and 100 mM NaCl (pH 7.5)] at a flow rate of 1 mL/min. The BT2127 molecular mass was derived from the measured elution volume by extrapolation of the plot of the elution volume of the molecular mass standard versus log molecular mass. The BT2127 native mass was also analyzed at the Howard Hughes Medical Institute Biopolymer/Keck Foundation Biotechnology Resource Laboratory at Yale University (New Haven, CT) by size exclusion chromatography coupled with online laser scattering, refractive index, and ultraviolet detection.

Kinetic Assay for β -Phosphoglucosyltransferase Activity.

Reaction solutions initially contained 7 μ M wild-type BT2127, 200 μ M β -glucose 1-phosphate, 10 μ M β -glucose 1,6-bisphosphate, 2 mM MgCl_2 , 0.4 mM NADP, and 5.4 units/mL glucose-6-phosphate dehydrogenase in 50 mM K^+ -HEPES (pH 7.5). The formation of NADPH from reduction of glucose 6-phosphate was monitored at 340 nm ($\epsilon = 6.2 \text{ mM}^{-1} \text{ cm}^{-1}$).

Determination of Steady-State Kinetic Constants.

Initial velocities for BT2127-catalyzed hydrolysis of phosphate esters and anhydrides were measured at 25 °C using assay solutions that contained 1 mM MgCl_2 , 1.0 unit/mL purine nucleoside phosphorylase, and 0.2 mM MESG in 50 mM Tris (pH 7.5). The reactions were monitored at 360 nm ($\Delta\epsilon = 9.8 \text{ mM}^{-1} \text{ cm}^{-1}$). The steady-state kinetic parameters (K_m and k_{cat}) were determined by fitting the initial velocity data measured at varying substrate concentrations (ranging from $0.5K_m$ to $5K_m$) to eq 1 using KinetAsyst I

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

where V_0 is the initial velocity, V_{max} the maximum velocity, $[S]$ the substrate concentration, and K_m the Michaelis constant for the substrate. The k_{cat} value was calculated from V_{max} and $[E]$ according to the equation $k_{\text{cat}} = V_{\text{max}}/[E]$, where $[E]$ is the enzyme concentration.

The steady-state competitive inhibition constant K_i was determined for imido diphosphate by fitting the initial velocity data, measured as a function of pyrophosphate (3, 5, 8, 10, 15, and 20 μ M) and imidodiphosphate (0, 10, 35, and 50 μ M) concentration, to eq 2 using KinetAsyst I

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

where $[I]$ is the inhibitor concentration and K_i the inhibition constant.

Determination of pH–Rate Profiles. The steady-state kinetic constants k_{cat} and K_m for BT2127-catalyzed hydrolysis of pyrophosphate were measured at 25 °C as a function of reaction solution pH. Reaction solutions initially contained pyrophosphate ($0.5K_m$ to $10K_m$), 1 mM MgCl_2 , 1.0 unit/mL purine nucleoside phosphorylase, and 0.2 mM MESG in 50 mM buffer (MES, pH 5.0–6.0; HEPES, pH 6.5–7.5; Tris, pH 8.0–8.5). The k_{cat}/K_m data were fitted with eq 3, and the k_{cat} data were fitted with eq 4.

$$\log Y = \log[C/(1 + [H]/K_a + K_b/[H])] \quad (3)$$

$$\log Y = \log[C/(1 + [H]/K_a + K_b/[H])] \quad (4)$$

where Y is the k_{cat} or k_{cat}/K_m , $[H]$ is the hydrogen ion concentration of the reaction solution, K_a and K_b are the apparent ionization constants, and C is the constant value of Y .

BT2127 Crystallization and Structure Determination.

Proteins were crystallized by the sitting-drop vapor diffusion method. In brief, the protein solutions (usually 0.3 or 1 μ L) were mixed with an equal volume of a precipitant solution and equilibrated at room temperature (294 K) against the same precipitant solution in clear tape-sealed 96-well INTELLI-plates (Art Robbins Instruments, Sunnyvale, CA). Crystallization was performed using either a TECAN crystallization robot (TECAN US, Research Triangle Park, NC) or a PHOENIX crystallization robot (Art Robbins Instruments) and three types of commercial crystallization screens: the WIZARD screen (Emerald BioSystems, Bainbridge Island, WA) and the INDEX and the CRYSTAL SCREEN I and II (both from Hampton Research, Aliso Viejo, CA). A number of conditions produced diffraction-quality protein crystals starting within 24–72 h of incubation. The crystals were harvested using cryogenic loops (Hampton Research), quickly transferred into liquid nitrogen, and stored frozen in liquid nitrogen until X-ray analysis was conducted and/or data were collected. Only in one case (PDB entry 3QU2) was the mother liquor supplemented with the cryoprotectant glycerol. Where necessary, the crystallization conditions were optimized manually using 24-well Cryschem sitting drop plates (Hampton Research). The final crystallization conditions for each X-ray crystal structure are listed in Table S1 of the Supporting Information.

The X-ray diffraction data for the majority of frozen crystals were collected at 100 K on beamline X29A (National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY) using a wavelength of either 0.93 or 1.08 Å (see Table S1 of the Supporting Information). The X-ray diffraction data for the D13A variant were collected at 100 K using an RU-200 rotating-anode X-ray generator ($\lambda = 1.5418 \text{ Å}$) coupled to a Rigaku R-Axis IV area detector. All diffraction data were processed and scaled with HKL2000.¹⁴ The first of 13 crystal structures (PDB entry 3QU2) reported here was determined by molecular replacement using coordinates for PDB entry 3DV9 (68% sequence identity) as the search model and MOLREP (the CCP4 program package suite¹⁵). The other 12 structures were determined by molecular replacement using coordinates of PDB entry 3QU2 as the search model. Each structure was refined using REFMAC 5.03 (CCP4 suite¹⁶), and the resulting models were rebuilt manually using COOT visualization and refinement software.¹⁷ The data collection and refinement statistics for these structures are listed in Tables S2–S5 of the Supporting Information.

Bioinformatic Analysis. Putative orthologues of BT2127 and the pyrophosphatase from *Thermococcus onnurineus* TON0002 were identified using BLAST searches of the NCBI microbial genome bank (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi) and selection of those sequences in which the key marker residues are conserved. The multiple-sequence alignments were made in Cobalt (<http://www.ncbi.nlm.nih.gov/tools/cobalt/>) and displayed in ESPript (<http://esprict.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>).

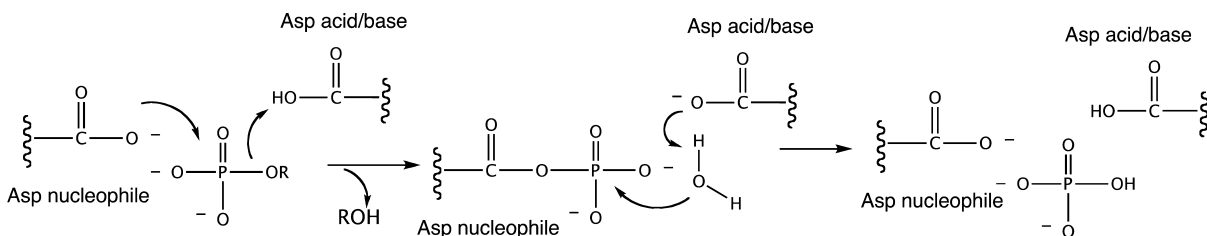
To compare biological ranges of BT2127 and β -PGM, we employed a more automated approach. BLAST searches were performed for BT2127, β -PGM, maltose phosphorylase, and trehalose phosphorylase using a locally installed copy of BLAST

Table 1. Steady-State Kinetic Constants for BT2127-Catalyzed Hydrolysis of Pyrophosphate and Selected Phosphate Monoesters at 25 °C and pH 7.5^a

substrate	k_{cat} (s ⁻¹)	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ (M ⁻¹ s ⁻¹)
pyrophosphate	0.32 ± 0.01	0.0036 ± 0.0004	9 × 10 ⁴
glycerol 1-phosphate	0.35 ± 0.02	4.8 ± 0.3	7 × 10
D-ribose 5-phosphate	0.20 ± 0.02	3.8 ± 0.3	6 × 10
fructose 6-phosphate	0.066 ± 0.002	6.4 ± 0.2	1 × 10
uridine 5-phosphate	0.15 ± 0.01	2.6 ± 0.4	6 × 10
p-nitrophenyl phosphate	0.0019 ± 0.0004	0.05 ± 0.01	4 × 10

^aSee Materials and Methods for details.

Scheme 1. HADSF Phosphatase Chemical Pathway



against the NCBI nonredundant protein database (<ftp://ftp.ncbi.nih.gov/blast/db/>) with default parameters. The resulting hits were filtered for query coverage >90 and >30% sequence identity. A reciprocal BLAST search, using an e value threshold of $1e^{-10}$, was performed for each filtered hit. A hit was considered successful if the reciprocal BLAST search identified the initial query sequence. For each successful hit, taxonomic information was downloaded from the NCBI taxonomy website (<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/>). For the selection of β -PGM orthologues, only those species with a predicted maltose or trehalose phosphorylase gene were considered as positive hits. A phylogenetic tree for the orthologues was generated using iTOL [interactive Tree Of Life (<http://itol.embl.de/>)] and visualized using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

RESULTS AND DISCUSSION

BT2127 Substrate Specificity Profile. Because the most similar characterized structural homologue of BT2127 is β -phosphoglucosyltransferase (β -PGM), the ability of purified recombinant BT2127 (see the SDS-PAGE gel in Figure S1 of the Supporting Information) to catalyze the conversion of β -glucose 1-phosphate to glucose 6-phosphate in the presence of the cofactor β -glucose 1,6-bisphosphate was tested. No activity was observed above the detection limit of one catalytic turnover per hour.

To determine the substrate range of BT2127, phosphatase activity was tested using a structurally diverse chemical library of 21 organophosphate metabolites as substrates (Table S6 and S7 of the Supporting Information). The most active substrates identified by this screen, pyrophosphate, glycerol 1-phosphate, D-ribose 5-phosphate, fructose 6-phosphate, uridine 5-phosphate, and pNPP, were analyzed to determine k_{cat} and $k_{\text{cat}}/K_{\text{m}}$ values (Table 1) at the pH optimum of 7.5 (pH-rate profile shown in Figure S2 of the Supporting Information). HADSF phosphatases are known to utilize a two-step reaction entailing transfer of a phosphoryl group from the bound substrate to the Asp nucleophile followed by hydrolysis of the aspartyl phosphate intermediate (Scheme 1). Variation in the k_{cat} value with substrate structure indicates that the first step is

rate-limiting (or at least partially rate-limiting) and that some substrates bind more productively than others. The largest k_{cat} value was measured for the substrate inorganic pyrophosphate (0.3 s^{-1}). Inorganic pyrophosphate exists in the cell as a magnesium complex that undergoes spontaneous hydrolysis to two molecules of orthophosphate at an uncatalyzed rate in solution of $2.8 \times 10^{-10} \text{ s}^{-1}$ (at 25 °C and pH 8.5).¹⁸ BT2127 thus increases the hydrolysis rate of pyrophosphate $\sim 1 \times 10^9$ -fold (a transition-state stabilization of >12 kcal/mol).

The $k_{\text{cat}}/K_{\text{m}}$ value (also known as the substrate specificity constant) is determined by the substrate binding affinity as well as by the efficiency with which the bound substrate is converted to product. Because the concentration of the substrate in the cell is likely to be subsaturating, the $k_{\text{cat}}/K_{\text{m}}$ value is most useful for identifying substrates that have physiologically relevant activities. HADSF phosphatases that target a single physiological substrate typically display a $k_{\text{cat}}/K_{\text{m}}$ value in the range of 1×10^5 to $1 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$,^{8,9,13} whereas those that function as regulators of the level and composition of organophosphate metabolite pools typically have K_{m} values in the millimolar range and $k_{\text{cat}}/K_{\text{m}}$ values in the range of 1×10^3 to $1 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$.^{19,20} The $k_{\text{cat}}/K_{\text{m}}$ value of $\sim 1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ measured for inorganic pyrophosphate is consistent with its assignment as a physiological substrate. Furthermore, the fact that the screen did not identify other substrates with activities in a range to be considered physiologically relevant suggests that BT2127 might indeed function in vivo as an inorganic pyrophosphatase.

Determination of the BT2127 Structure. The structure of BT2127 was examined to identify the structural determinants of substrate recognition. A total of 13 X-ray structure determinations were conducted with the wild-type enzyme and the D11N, D13A, D13N, E47A, E47D, and E47N variants unliganded and bound to ligands. The crystallization conditions (Table S1 of the Supporting Information) and the crystallographic data collection and refinement statistics are reported in Tables S2–S5 of the Supporting Information. The various crystallization conditions led to different crystal packing and to different ligands observed in the active site. All crystallizations were performed in the presence of 5 mM MgCl_2 ; however, only six of the structures (PDB entries 3QU2, 3QU4, 3QU9,

3QUQ, 3QUT, and 3QX7) contained Mg^{2+} within the active site. In two of the structures (PDB entries 3QU7 and 3QYP), Ca^{2+} derived from the crystallization solution replaced Mg^{2+} . This replacement has previously been observed in HADSF enzymes.²¹ Cocrystallization and crystal soaking of each enzyme were attempted with pyrophosphate as well as with its analogue, imidodiphosphate [a competitive inhibitor with a K_i of $13 \pm 1 \mu M$ (Figure S3 of the Supporting Information)]. No ligand density was observed in the imidodiphosphate-soaked crystals. Three structures derived from the pyrophosphate-soaked crystals (PDB entries 3QYP, 3QX7, and 3QU7) contained the hydrolysis product, phosphate, showing the enzyme is active under the crystallization conditions utilized.

The BT2127 protomer possesses the conserved HADSF α/β Rossmann-like catalytic domain composed of a six-stranded parallel β -sheet ($\beta 1$ and $\beta 4$ – $\beta 8$) surrounded by six helices ($\alpha 1$ – $\alpha 3$ and $\alpha 5$ – $\alpha 7$) and a four-helix cap domain (residues 19–89) inserted between $\beta 1$ and $\alpha 1$ (denoted type C1 cap domain). Well-characterized members of the C1 class include *Lactobacillus lactis* β -PGM (Z score of 10.2)²² and *Bacillus cereus* phosphonate (Z score of 7.8)²³ (Z scores calculated by DALI).²⁴ The type C1 cap domains of BT2127 and β -PGM superimpose (Figure 1) with a root-mean-square deviation



Figure 1. Superposition of BT2127 (PDB entry 3QX7) (gray) and *L. lactis* β -PGM (PDB entry 1O08) (cyan).

(rmsd) of 2.4 Å (13.4% identical cap domains), whereas the two catalytic domains superimpose with an rmsd of 1.9 Å (26.7% identical core domains). HADSF members with similar cap domain folds were identified using DALI and the cap domain (Table S8 of the Supporting Information). The closest homologue (Z score of 14.7) is the HADSF member from *Bacteroides vulgatus*, which was previously annotated “putative β -PGM” (PDB entry 3DV9). As detailed below, we posit that

this *B. vulgatus* protein is not β -PGM but a pyrophosphatase and an orthologue of BT2127.

BT2127 is a monomer in solution, as the molecular mass determined by molecular size gel filtration chromatography (27–28 kDa) matches the protomer molecular masses of 26970 Da determined by mass spectrometry (theoretical mass of BT2127 minus that of the N-terminal Met is 26983 Da) and 27 kDa determined by SDS–PAGE (Figure S1 of the Supporting Information).

Association of the Cap and Catalytic Domains. CA1 cap motion has been most thoroughly studied in *L. lactis* β -PGM for which crystallographic snapshots of the enzyme in the open and closed conformations have been obtained (Figure 2).^{25,26}

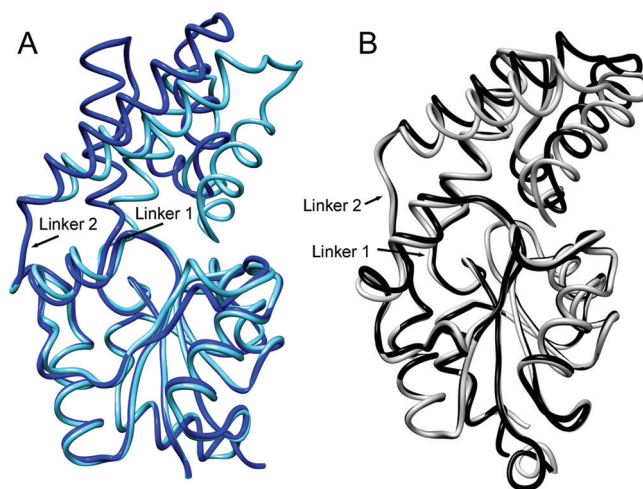


Figure 2. Backbone coil depictions of the superposition of the (A) cap-open (royal blue) and cap-closed (cyan) conformations of *L. lactis* β -PGM (PDB entries 1ZOL and 1O08, respectively) and (B) cap-open (black) and cap-closed (gray) conformations of BT2127 (PDB entries 3QUQ and 3QX7, respectively).

The structures of BT2127 (PDB entry 3QX7; wild type bound to phosphate and Mg^{2+}) reveal a cap-closed conformation that is similar to that of β -PGM (Figure 1) and a cap-open conformation (PDB entry 3QUT; D13N variant with Mg^{2+} and malate bound to the transferring phosphate site) that is not as open as that observed for β -PGM (Figure 2A,B).

In the case of β -PGM, the cap and catalytic domains associate for catalysis and dissociate to allow ligand exchange with solvent. An earlier crystallography-based analysis of the “clamlike” domain movement in β -PGM using DynDom^{27,28} indicated that the cap and catalytic domains move as rigid bodies with a large relative rotation of 26°; this movement is primarily the result of changes in backbone conformations (ψ and ϕ angles) at domain–domain linker 1 hinge residues Thr14, Asp15, Thr16, and Ala17 (Dai 2009). A similar analysis shows that the more modest domain movement observed in BT2127 is due to changes in ϕ and ψ angles at linker 1 hinge residues Ser19, Met20, and Pro21 and linker 2 hinge residues Pro87, Glu88, Ala89, Glu90, and Arg91. The rotation angle between the cap and catalytic domains in BT2127 is 16.3°, with a translation of 0.6 Å. As seen with β -PGM, the BT2127 cap and core domains must dissociate for pyrophosphate to bind to the active site and phosphate to dissociate because only in the cap-open conformation is there bulk solvent access to the active site.

To assess the relative sizes of the active site cavities of the two enzymes, Voidoo (<http://xray.bmc.uu.se/usf/voidoo.html>) was used to calculate the volume of the active sites of BT2127 (189.8 Å³) and β -PGM (285.1 Å³) in their cap-closed conformations. Next, the molecular volumes of pyrophosphate (99.4 Å³) and β -glucose 1,6-bisphosphate (242.5 Å³) were calculated using the molinspiration server (<http://www.molinspiration.com/>). Whereas pyrophosphate is small enough to fit in the active site of BT2127, β -glucose 1,6-bisphosphate is not. Models of BT2127 bound to pyrophosphate and to β -glucose 1,6-bisphosphate were generated to show the fit of the ligands in the active site cavity (Figure 3). Whereas the

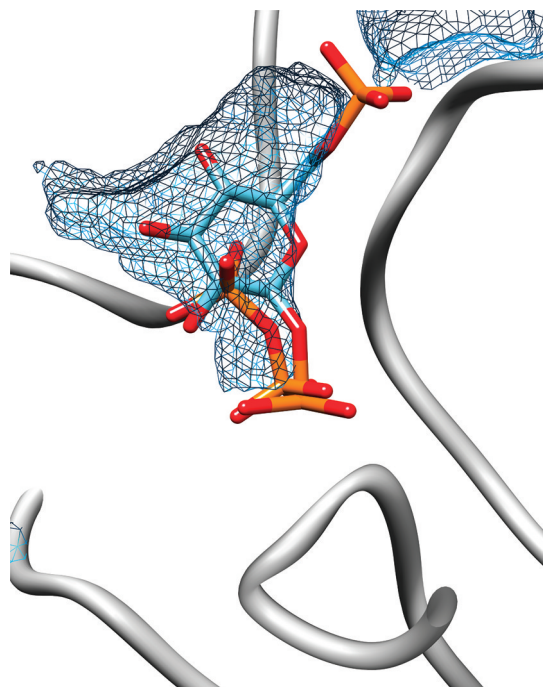


Figure 3. Active site of BT2127 in the cap-closed conformation (PDB entry 3QX7) showing the active site volume calculated in Voidoo (mesh) and the pyrophosphate and β -glucose 1,6-bisphosphate ligands (shown as sticks, phosphorus atoms colored orange) modeled in COOT.

pyrophosphate can be easily docked in a productive binding mode without steric clash, the larger β -glucose 1,6-bisphosphate ligand cannot fit without the cap dissociating to some extent from the core domain. The low-level activity that is observed with the larger substrates (Table 1 and Table S6 of the Supporting Information) might be rationalized by catalytic turnover in a conformer in which the cap domain is partially dissociated to enlarge the active site cavity.

Active Site Scaffold. The active site scaffold of the HADSF phosphotransferase catalytic domain consists of a four-loop platform on which the conserved catalytic residues are located: the Asp nucleophile and Asp acid/base on loop 1, the Thr/Ser and Lys hydrogen bond donors to the transferring phosphoryl group on loops 2 and 3, respectively, and the Asp/Glu Mg^{2+} ligand on loop 4.¹ Figure 4 shows the residues surrounding the Mg^{2+} cofactor in the BT2127–Mg complex and the interacting residues in the BT2127 E47N variant in a complex with Ca^{2+} and phosphate. As observed with other HADSF phosphotransferases, the Mg^{2+} cofactor is coordinated to the loop 1 Asp nucleophile (Asp11) carboxylate group and the Asp acid/base

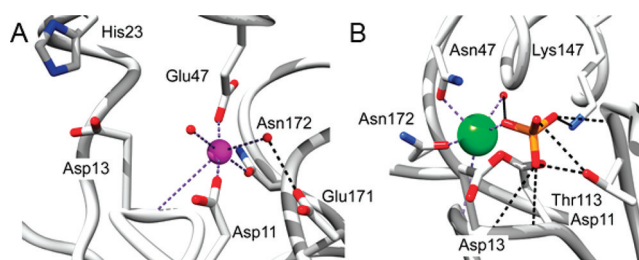


Figure 4. (A) Observed Mg^{2+} binding site structure of wild-type BT2127 bound to Mg^{2+} (magenta sphere) (PDB entry 3QUQ). (B) Phosphate binding site observed in the structure of the BT2127 E47N mutant bound with phosphate (phosphorus colored orange) and Ca^{2+} (green sphere) (PDB entry 3QYP). Water molecules are represented as red spheres, coordination bonds as dashed purple lines, and hydrogen bonds as dashed black lines.

(Asp13) backbone C=O group (2.2 Å). Loop 4 Asn172 forms a coordination bond with Mg^{2+} (2.3 Å), and loop 4 Asp171 forms a hydrogen bond (2.7 Å) to one of two waters that coordinate the Mg^{2+} . A deviation from the classical HADSF Mg^{2+} binding site²⁹ is the identity of the sixth ligand, Glu47 (2.0 Å), which is located on the specificity loop of the cap domain (Glu47–Gly48–Arg49–Thr50).^a

From the structure of the E47N variant bound to phosphate and Ca^{2+} at the catalytic site, we observe that an oxygen atom of the phosphate ligand replaces a water ligand (Figure 4), as is usually observed in the HADSF.¹² The phosphate ligand also is engaged in numerous hydrogen bond interactions with the loop 2 Thr113 and loop 3 Lys147 side chains, and loop 1 Asp13 and loop 2 Gly114 and Ser115 backbone amide N atoms. For the best examination of the key interactions in the context of the pyrophosphate substrate, a model was made of the BT2127 active site (cap-closed conformer) with pyrophosphate bound in an orientation that superimposes the transferring phosphoryl group on the phosphate ligand of the BT2127 E74N– Ca^{2+} –phosphate structure (Figure 5). Only one P–O–P conformer placed the phosphate leaving group within the vacant space defined by the Voidoo analysis (vide supra). Here it can engage in hydrogen bond formation with a water molecule that is bound to cap domain Trp27 and with the backbone amide NH group of cap domain residue Gly48 and catalytic domain Ser115. Moreover, the Asp13 acid/base is positioned to donate a hydrogen bond to a nonbridging oxygen atom, and to the bridging oxygen atom of the phosphate leaving group. Notably, at pH 7.5, the pyrophosphate is monoprotonated ($pK_a = 9.3$), and if it were to bind with the protonated phosphate group in the leaving group position, there would be no need for acid catalysis. On the other hand, we have proposed that the Asp acid/base of the HADSF phosphotransferases is protonated by the substrate phosphate ester, which binds to the active site in the monoprotonated form, and upon coordination with the Mg^{2+} , the phosphate substituent loses its proton.³⁰ Thus, for the Asp13 to be protonated, the pyrophosphate would need to be fully ionized.

Notably, via replacement of the pyrophosphate bridging oxygen atom with an NH group, as in imidodiphosphate, tight binding to BT2127 is retained ($K_i = 13 \pm 1 \mu M$) but the extent of catalytic turnover is greatly diminished (Table S2 of the Supporting Information). We infer from this observation that Asp13 is not able to align with the nonbonding electron pair of the bridging nitrogen atom for proton transfer. In contrast, several of the 23 *E. coli* HADSF phosphatases examined by

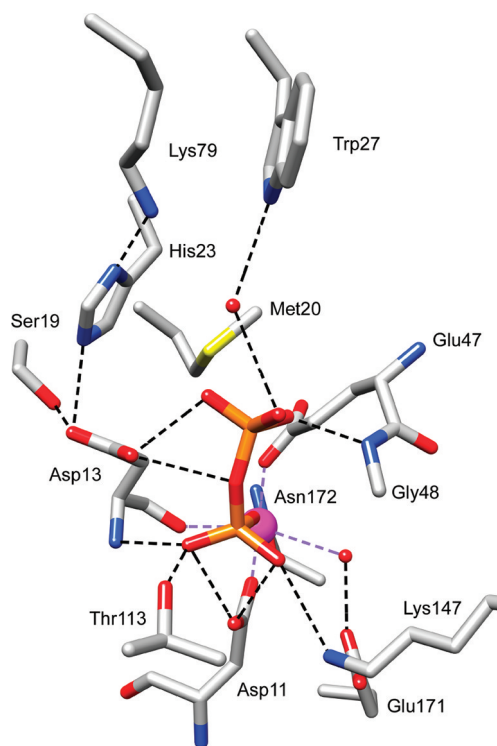


Figure 5. BT2127 active site with pyrophosphate (manually docked). Atom coloring and bonds as in Figure 4.

Kuznetsova et al.⁴ showed significant hydrolase activity with imidodiphosphate but, curiously, not with pyrophosphate.

An interesting finding is revealed by the position that the phosphate ligand assumes in the structure of the wild-type BT2127–Mg²⁺–phosphate complex (PDB entry 3QX7). Superposition of this structure with the structure of *L. lactis* β -PGM bound to β -glucose 1,6-bisphosphate places the phosphate ligand at the same position at which the non-transferring phosphoryl group (i.e., C6 phosphate) is found (Figure S4 of the Supporting Information). It is plausible that once formed, the phosphate product in the BT2127 active site might diffuse into this vacant space to make room for the water nucleophile, which will perform in-line attack at the phosphorus atom of the aspartyl phosphate intermediate in the second step of the reaction (Scheme 1). Alternatively, the movement of the phosphate into the cap domain might trigger cap dissociation.

Site-Directed Mutagenesis of BT2127 Active Site and Domain–Domain Interface Residues.

The individual contributions that the active site residues make to catalysis were evaluated by site-directed mutagenesis coupled with steady-state kinetic analysis of the mutant enzymes in catalysis of pyrophosphate hydrolysis. The structures of the BT2127 mutants showed that the wild-type native conformation, including the active site residue side chains, was retained. Therefore, we interpret the changes in catalytic activity in the mutants as resulting from changes in side chain function and not alterations of the native conformation. The mutants that did not display activity above background are listed in Table 2 as not active (NA). The mutants that exhibited detectable yet very low activity were evaluated by measuring the initial velocity of the reaction using a concentration of pyrophosphate (300 μ M) that we assume to be saturating. The ratio of the

Table 2. Steady-State Kinetic Constants for Wild-Type and Mutant BT2127-Catalyzed Hydrolysis of Pyrophosphate at 25 °C and pH 7.5^a

BT2127	k_{cat} (min ^{−1})	K_{m} (μ M)	$k_{\text{cat}}/K_{\text{m}}$ (M ^{−1} s ^{−1})
wild type	$(1.9 \pm 0.06) \times 10$	3.6 ± 0.4	8.9×10^4
D11N	NA ^b	—	—
D13A	NA ^b	—	—
D13N	NA ^b	—	—
K147A	$\sim 2 \times 10^{-3c}$	—	—
N172A	$(7.4 \pm 1.8) \times 10^{-1}$	4.5 ± 0.3	3×10^3
T113A	$(4.0 \pm 0.03) \times 10^{-1}$	5.9 ± 0.2	1×10^3
S115A	$(8.4 \pm 0.12) \times 10^{-2}$	3.6 ± 0.4	4×10^2
T50A	$(4.1 \pm 0.18) \times 10^{-1}$	2.7 ± 0.7	3×10^3
R49A	$\sim 3 \times 10^{-2c}$	—	—
K79A	$\sim 2 \times 10^{-2c}$	—	—
H23A	$(3.3 \pm 0.06) \times 10^{-1}$	$(1.2 \pm 0.08) \times 10$	5×10^2
M20A	$\sim 5.1 \times 10^{-2c}$	—	—
M20K	$(2.4 \pm 0.04) \times 10^{-1}$	$(2.5 \pm 0.3) \times 10$	2×10^2
M20L	$(5.8 \pm 0.04) \times 10^{-2}$	3.5 ± 0.1	3×10^2
E47A	$\sim 3 \times 10^{-3}$	—	—
E47Q	$(4.8 \pm 0.4) \times 10^{-1}$	6 ± 1	1×10^3
E47N	$\sim 3 \times 10^{-3c}$	—	—
E47D	$\sim 3 \times 10^{-3c}$	—	—
S80A	$(1.7 \pm 0.04) \times 10^{-1}$	7.1 ± 0.9	4.1×10^2
Q117A	$\sim 3 \times 10^{-2c}$	—	—
Y76F	$\sim 2 \times 10^{-2c}$	—	—
S19A	$\sim 2 \times 10^{-2c}$	—	—

^aSee Materials and Methods for details. ^bNA represents no activity detected above the detection limit ($k_{\text{cat}} < 0.001 \text{ min}^{-1}$). ^cThe rate was determined by measuring the initial velocity of orthophosphate formation (micromolar per minute) and dividing that value by the enzyme concentration. Reaction solutions initially contained 10 μ M enzyme, 300 μ M pyrophosphate, 1 mM MgCl₂, 1.0 unit/mL purine nucleoside phosphorylase, and 0.2 mM MESG in 50 mM Tris (pH 7.5).

initial velocity and the enzyme concentration gives an approximate k_{cat} value (i.e., turnover rate). For the mutants that exhibited higher activity, k_{cat} and K_m values were measured.

First, the impact of replacement of the key catalytic residues, Asp11 nucleophile and Asp13 acid/base, both of which are invariant, with Ala and/or Asn was tested (see the multiple-sequence alignment of putative BT2127 orthologues in Figure S5 of the Supporting Information). Not surprisingly, replacement of BT2127 Asp11 or Asp13 with Ala and/or Asn resulted in the loss of all detectable activity (Table 2). However, the observation of a phosphate ligand in the active sites of pyrophosphate-soaked crystals of BT2127 D13N indicates that a single-turnover reaction had most likely occurred during the 15 min soaking time.

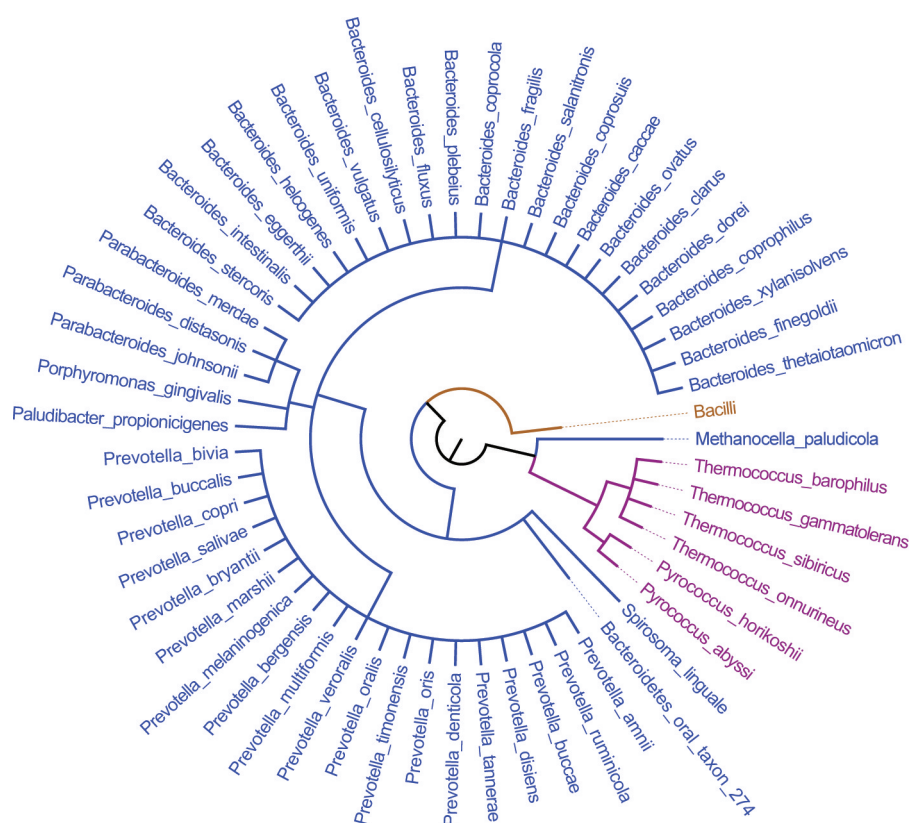
The replacement of the residues that contribute to the positioning of the Asp13 acid/base, namely, Ser19, His23, Ser115, or Met20 (each of which is invariant), with Ala reduced the turnover rate 950-, 60-, 230-, or 370-fold, respectively. Met20 was also replaced with Leu or Lys. The M20L and M20K mutants were found to have 330- and 80-fold lower turnover rates, respectively, compared to that of wild-type BT2127. Because the Met20 side chain restricts both the volume and the polarity of the region available to accommodate the substrate leaving group, we tested the activities of M20A, M20L, and M20K toward the same panel of phosphate esters that were used to screen the wild-type enzyme. The results (Table S7 of the Supporting Information) indicate that the substrate specificity of the mutants was not significantly different from that of wild-type BT2127.

The coordination of the Mg^{2+} by the loop 1 Asp nucleophile and by the backbone amide $\text{C}=\text{O}$ group of the Asp acid/base is observed in all structurally characterized Mg^{2+} -bound HADSF phosphatases.²⁹ Likewise, an Asp or Glu residue located on loop 4 invariably adds a third metal contact. In some cases, an Asp or Glu also located on loop 4 forms a water-mediated contact with the Mg^{2+} cofactor. We have not previously encountered the collaboration between a core domain loop 4 amide (Asn172) and cap domain specificity loop carboxylate (Glu47) that contributes to Mg^{2+} binding. The replacement of the stringently conserved Asn172 with Ala resulted in an only 26-fold reduction in turnover rate. In contrast, replacement of the conserved Glu47 with Ala resulted in an ~6000-fold decrease in turnover rate, and the same reduction is observed with the E47D and E47N mutants. Glu47 might be required for Mg^{2+} binding. We note that in the four structures of the three E47 mutants crystallized in the presence of Mg^{2+} the positions of the Asp11 and Asn172 side chains and the Asp13 backbone $\text{C}=\text{O}$ group are the same as in the wild type, yet the Mg^{2+} is absent. In contrast, the E47N variant crystallized in the presence of Ca^{2+} (PDB entry 3QYP) contains Ca^{2+} , which is coordinated to the loop 1 and 4 ligands as well as to Asn47. The larger ionic radius of Ca^{2+} apparently compensates for the shortened reach of the Asn side chain (Figure 4). Accordingly, the replacement of Glu47 with Gln should be conservative as it would allow Mg^{2+} coordination; this is reflected in the activity of the E47Q mutant, which has a turnover rate only 40-fold lower than that of the wild-type enzyme (Table 2). These findings suggest that E47 rather than the loop 4 Asn172 plays the key role in Mg^{2+} binding. Furthermore, the fact that the Glu47 is coordinated to the Mg^{2+} in the cap-open and cap-closed conformations might explain why the BT2127 cap domain does not dissociate from the catalytic domain to the extent observed with β -PGM.

In an examination of the ligands to the substrate, Thr113 and Lys147 activate the transferring phosphoryl group via hydrogen bond formation (Figures 4 and 5). Both are stringently conserved, and we found that the turnover rate was reduced 42-fold in the case of the T113A mutant and 9500-fold in the case of the K147A mutant. The stringently conserved cap domain residue Lys49 is not close enough to the catalytic site for direct interaction with pyrophosphate. It does, however, form a hydrogen bond with the stringently conserved cap domain residue His23, which in turn forms a hydrogen bond with the catalytic domain Asp13 acid/base (Figure 5). Whereas the turnover rate of the K49A mutant was found to be 950-fold slower than that of wild-type BT2127, that of the H23A mutant was reduced only 60-fold. Thus, Lys49 appears to contribute to catalysis in some way beyond simply serving as a hydrogen bond partner to His23 (vide infra).

Additional mutants were examined to evaluate the contributions made by residues located at the domain–domain interface, which appear from the structures to stabilize the cap-closed conformation through hydrogen bond formation with a partner located on the opposing domain (Figure S6 of the Supporting Information). First, the stringently conserved Gln117 of the catalytic domain forms a hydrogen bond with cap domain residue Ser80 (replaced with Thr, Ala, or Arg in 8 of 33 orthologous sequences). Replacement of Gln117 with Ala reduced the turnover rate 630-fold, whereas replacement of Ser80 with Ala reduced the turnover rate only 110-fold. Second, the stringently conserved Tyr76 located on interdomain linker 2 is close enough (3.7 Å) to make possible formation of a hydrogen bond with the backbone amide NH group of catalytic domain Gln117 and Gly116 (highly conserved; replaced with Ala in 1 of 33 sequences). The conservative replacement of Tyr76 with Phe reduced the turnover rate 950-fold. Third, the highly conserved Arg49 (31 of 33 sequences) and Thr50 (32 of 33 sequences) are located on the cap domain specificity loop (Glu47–Gly48–Arg49–Thr50). The Arg49 side chain projects into solvent where the nonpolar hydrocarbon region of the side chain packs against the side chains of the catalytic domain Pro148 (stringently conserved) and Leu175 (conservatively replaced with Met in 5 of 33 sequences), an interaction that might facilitate the proper positioning of Glu47 within the active site and/or influence the association of the cap and core domains. Replacement of Arg49 with Ala reduced the turnover rate 630-fold. Thr50 is seen in the crystal structures to be too far (4.3 Å) from the stringently conserved catalytic domain Gly114 $\text{C}=\text{O}$ group for hydrogen bond formation. Nevertheless, replacement of Thr50 with Ala reduced the rate of catalytic turnover 50-fold, suggesting that the solution conformation might allow hydrogen bond formation or, alternatively, that Thr50 contributes to the conformation of the specificity loop.

BT2127 Biological Range. On the basis of the structure–function analysis described in the previous sections, we selected the catalytic residues Asp11, Asp13, Thr113, and Lys147 as well as the metal binding residues Asp171, Asn172, and Glu47 to curate the list of BT2127 homologues identified by BLAST searches. An alignment of 33 putative homologues was generated (Figure S5 of the Supporting Information), which highlighted the invariant conservation of 46 residues (including the seven marker residues) of 244 residues total (19% conservation). The lowest level of pairwise sequence identity between BT2127 and any putative sequence orthologue is 47%. Many of these sequences have N-terminal extensions (~20



10

Figure 6. Phylogenetic representation of the biological range of the putative orthologues of BT2127 (blue), β -PGM (brown), and the Archeal pyrophosphatase TON0002 (purple).

amino acids) when compared to the BT2127 sequence. One of these is the orthologue from *B. vulgatus* (Uniprot accession code A6L7P8) that was also identified as the most similar structure in the DALI search for structural homologues (vide supra). The superposition of the X-ray structure (PDB entry 3DV9) with that of BT2127 (not shown) reveals that N-terminal extension forms an α -helix at the surface of the catalytic domain located at the edge of the β -sheet, which is distant from the active site and is not expected to have an impact catalysis.

Although the marker residues used in curating the BT2127 homologues are by definition invariant, 14 additional sequences in which a marker residue, Asn172, was replaced with Asp could be identified (multiple-sequence alignment shown in Figure S7 of the Supporting Information). These sequences are more divergent (28–30% pairwise sequence identity with BT2127), and all but one of them also possesses the conservative replacement of Met12 in BT2127 with Leu in the Asp11-Met12-Asp13 loop 1 catalytic motif. Although the residue that separates the Asp nucleophile and Asp acid/base on loop 1 is not known to have a direct catalytic role, it has been our observation that it is typically conserved within function families that have clear sequence homology boundaries. Moreover, the residues of the specificity loop and/or the residues supporting domain–domain interaction (such as Tyr76, Thr50, Gln117, Ser19, Met20, and Lys49) are not conserved. Nevertheless, the 14 sequences might in fact be pyrophosphatases, but such an annotation should be first

supported by an in vitro activity assay of a representative member; as such, these proteins will be termed “BT2127-like” homologues.

The biological range of the putative BT2127 orthologues is restricted to the phylum Bacteroidetes/Chlorobi (Figure 6). BT2127 orthologues are found in all *Bacteroides* species and strains represented in the gene databases (with the exception of *Bacteroides pectinophilus* that instead possesses a BT2127-like homologue), as well as in the *Parabacteroides* species and strains and *Prevotella* species and strains. *Bacteroides*, *Parabacteroides*, and *Prevotella* colonize the human cavity from mouth to colon yet occupy different regions along this environmentally stratified track. The BT2127-like homologues are also contained in the Bacteroidetes/Chlorobi phylum where they are found in select species that occupy a diverse range of habitats: *Capnocytophaga ochracea* (human mouth), *Chlorobium luteolum* and *Pelodictyon phaeoclathratiforme* (sulfide rich aquatic environments), *Croceibacter atlanticus* and *Flavobacterium bacterium* BAL38, BBFL7, and ALC-1 (seawater), *Listeria innocua*, *Listeria marthii*, *Listeria monocytogenes*, and *Listeria seeligeri* (ubiquitous in the environment), and *Paenibacillus* sp. Y412MC10 and *Paenibacillus vortex* (heterogeneous and complex environments such as soil and rhizosphere).

BT2127 in Vivo Function. The substrate specificity profile and the active site structure strongly indicate that inorganic pyrophosphate is the physiological substrate for BT2127. Independent supportive evidence of the in vivo function of BT2127 as an inorganic pyrophosphatase derives from gene

context analysis. Pyrophosphate is derived from nucleoside 5'-triphosphates as they undergo nucleotidyl transfer reactions catalyzed by ligases, synthases, nucleotidyl transferases, and DNA and RNA polymerases. The hydrolysis of the pyrophosphate product drives these reactions forward and at the same time replenishes the orthophosphate pool.

The gene neighborhoods of BT2127 and the 33 putative orthologues were examined for the bacterial genomes that have been mapped. The BT2127 gene overlaps the *fkp* gene that encodes the bifunctional L-fucose kinase/L-fucose-1-phosphate guanylttransferase. The products of the kinase-catalyzed reaction are β -L-fucose 1-phosphate (a poor substrate for BT2127 with a k_{cat} of $\sim 0.1 \text{ min}^{-1}$) and ADP, whereas the products of the transferase-catalyzed reaction are inorganic pyrophosphate and β -L-fucose-GDP. The β -L-fucose-GDP is in turn used as a precursor for the decoration of surface capsular polysaccharides and glycoproteins with L-fucose for colonizing the mammalian intestine.³¹ Thus, we posit that the biochemical function of BT2127 is inorganic pyrophosphate hydrolysis, which supports the biological function of host colonization. Intriguingly, several genes upstream of and on the same DNA strand as BT2127 is the lysyl-tRNA synthase gene *lysS* (BT2122). Thus, BT2127 might also assist protein synthesis by removal of the synthase pyrophosphate product.

Structural Determinants in the Divergence of Function in BT2127 and β -PGM. β -PGM and maltose phosphorylase or trehalose phosphorylase collaborate in the utilization of maltose or trehalose as a carbon and energy source. β -PGM and BT2127 share 30% identical sequences and a common cap domain fold (four α -helices) in addition to the highly conserved catalytic domain fold. Most significantly, they share the cap His20-Lys76 hydrogen bond diad motif that in β -PGM is responsible for the discrimination between β -glucose 1-phosphate and the solvent as reactants with the aspartyl phosphate intermediate (Figure S8 of the Supporting Information).³⁰ Briefly, in the cap-open conformation the β -PGM, the Asp acid/base side chain assumes a rotamer conformation stabilized by formation of a hydrogen bond with the backbone amide NH group and side chain of the linker hinge residue Thr117, which places it outside of the active site.²⁵ Upon substrate-induced cap closure, His20 of the His20-Lys76 diad and the β -glucose 1,6-bisphosphate bridging oxygen atom form a hydrogen bond to the Asp acid/base, placing it into position for catalysis (Figure S8 of the Supporting Information).²⁶ BT2127 conserves the His23-Lys79 diad, and in both the cap-open and -closed conformations, the Asp13 side chain is in the same conformation, engaged in a hydrogen bond with linker residue Ser15. In the cap-closed conformation, the His23-Lys79 diad is close enough to the Asp13 carboxylate for formation of a hydrogen bond with His23; however, in the cap-open conformation, it is too distant. Thus, the cap domain His-Lys diad appears to be a vestige of the common ancestry of BT2127 and β -PGM.

Examination of the biological range of β -PGM using the same approach applied to BT2127 showed that the putative β -PGM orthologues are primarily found in the phylum Firmicutes and especially in the genera *Bacilli* (as represented in Figure 6). However, they can also be found in certain species of the phyla Proteobacteria, Actinobacteria, and Bactoidetes/Chlorobi (not represented in Figure 6). Notably, we did not uncover a bacterial species that contains orthologues of both BT2127 and β -PGM. A detailed analysis of β -PGM orthologues and their biological range will be published separately.

Evolution of Pyrophosphatases within the HADSF and Other Superfamilies. Inorganic pyrophosphatases are ubiquitous and essential to cellular function. The large (660–770 amino acid) membrane-bound pyrophosphatases are limited in range to plants and certain bacteria, where they function as proton pumps. The soluble pyrophosphatases are ubiquitous and are found in two distinct fold families (type I and type II). The type I pyrophosphatases are widely distributed in all three kingdoms of life. This is an ancient family with a high degree of sequence divergence. The more recently characterized family is the type II family. Inorganic pyrophosphatases of the type II family are restricted in range to select lineages of bacteria.³²

The type I pyrophosphatases are single-domain, β -barrel proteins, whereas the type II pyrophosphatases are two-domain α/β -proteins in which the active site is located at the domain–domain interface. Type I and II pyrophosphatases employ three or four Mg^{2+} cofactors to activate the transferring phosphoryl group and water nucleophile and to stabilize the phosphate leaving group.^{32–34} HADSF pyrophosphatases represent a third mechanistic and fold type for bacterial pyrophosphatases.

Recently, a novel pyrophosphatase (TON0002) from *T. onnurineus* was discovered;³⁵ this Archeal pyrophosphatase is a member of the HADSF. Like BT2127, TON0002 possesses a C1 four- α -helix cap domain, yet this enzyme has evolved unique structural features as inferred from the structure of the close sequence homologue (60% identity) from *Pyrococcus horikoshii* (PDB entry 2OM6). The superposition of the structures of BT2127 and the *P. horikoshii* enzyme is shown in Figure 7. The two sets of α -helices that form the TON0002 cap

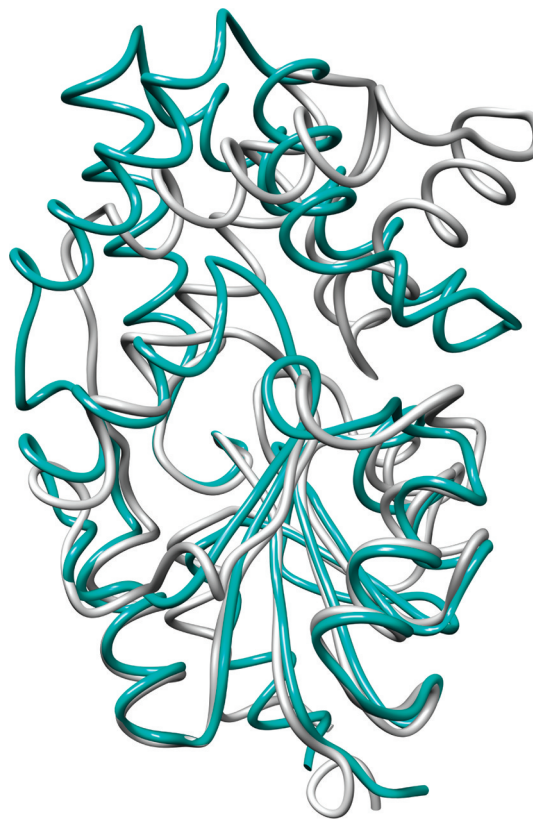


Figure 7. Superposition of the structures of BT2127 (gray) (PDB entry 3QX7) and the putative pyrophosphatase from *P. horikoshii* (teal) (PDB entry 2OM6).

domain are splayed apart, thereby expanding the area above the catalytic site. The active site is protected from bulk solvent by (1) the insertion of a hydrophobic loop (residues 123–128; viz. loop 2) from the catalytic domain between two cap α -helices and (2) the long side chains of Arg48, Lys52, and Arg55 from the cap.

A model of the *P. horikoshii* pyrophosphatase of the active site with Mg^{2+} and pyrophosphate bound was made by using the positions of the Mg^{2+} and the phosphate (to define the cofactor and transferring phosphoryl position) in the BT2127 structure and the position of the sulfate ligand in the *P. horikoshii* enzyme structure to define the position of the phosphate leaving group (Figure 8). Remarkably, the loop 1

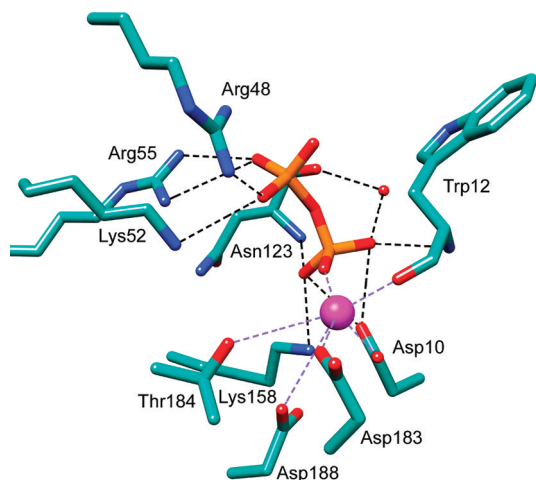


Figure 8. Active site of the putative pyrophosphatase from *T. onnurineus* BT2127 modeled with pyrophosphate (manually docked) and Mg^{2+} (derived from the superposition of BT2127 PDB entry 3QUQ). The Mg^{2+} is shown as a magenta sphere, and the water molecules are represented as red spheres. Atom coloring and bonds as in Figure 4.

Asp acid/base and loop 2 Thr/Ser are not conserved; instead, Trp12 and Gly122, respectively, are observed at these positions. The loop 1 Asp nucleophile is Asp10, and the loop 3 Lys is Lys158. The counterparts to the BT2127 loop 4 Glu171 and Asn172 are Gly182 and Asp183, respectively. In the model, the phosphate leaving group is positioned for favorable interaction with Arg48, Lys52, and Arg55 of the cap domain, whereas the transferring phosphoryl group is positioned to coordinate Mg^{2+} and to favorably interact with catalytic domain residues Asn123 (loop 2) and Lys158 (loop 3).

Because the level of sequence identity is high (60%) and because the key active site residues identified in the *P. horikoshii* enzyme structure are conserved in the *T. onnurineus* pyrophosphatase, it is reasonable to conclude that the *P. horikoshii* enzyme is also a pyrophosphatase. The active site model suggests that the backbone amide NH group of Asn123 substitutes for the conserved loop 2 Thr/Ser in activating the transferring leaving group and that cap domain residues Arg48, Lys52, and Arg55 stabilize the phosphate leaving group. In the second partial reaction (Scheme 1), the loop 1 Asp acid/base activates the water nucleophile for attack on the aspartyl phosphate. If it were to assume a side chain rotamer conformation slightly altered from that observed in the structure, the Asn123 could orient (but not deprotonate) the water nucleophile.

The biological range of the Archeal pyrophosphatase orthologues was compared to that of the BT2127 orthologues. A BLAST search conducted using the *T. onnurineus* pyrophosphatase sequence identified three orders of Archea with Archeal pyrophosphatase orthologues: Thermococcales (53–80% sequence identity), Desulfurococcales (31–41% sequence identity), and Thermoproteales (29–33% sequence identity). A sequence alignment of all homologues (all sequences conserved Trp12) was made, and those sequences that did not conserve marker residues Arg48, Lys52, Arg55, and Asn123 were omitted, producing an alignment of orthologues [orders Thermococcales and Thermoproteales (Figure S9 of the Supporting Information)]. Inspection of the gene contexts of these putative orthologues revealed that in some of the thermophiles a gene encoding DNA polymerase was found either on the same DNA strand next to pyrophosphatase gene or across from it on the opposite strand. This suggests a possible biochemical context for the pyrophosphatase, i.e., aiding the polymerase by removing its product, inorganic pyrophosphate. On the other hand, the TON0002 k_{cat}/K_m of $\sim 4 \times 10^2 M^{-1} s^{-1}$ is so low one might question the physiological relevance especially given that the *T. onnurineus* genome encodes a type I inorganic pyrophosphatase.

CONCLUSION

A BLAST search of the *B. thetaiotaomicron* genome using the *E. coli* type I inorganic pyrophosphatase and the *Bacillus subtilis* type II inorganic pyrophosphatase sequences as query failed to identify a homologue. Thus, the C1-type HADSF member BT2127 appears to be the only known cytoplasmic inorganic pyrophosphatase in *B. thetaiotaomicron*. The C1-type class appears to be the most prevalent and functionally versatile of the three HADSF classes (C0, C1, and C2).² Aside from the many phosphatases that have evolved within this class, the dehalogenase, phosphonate, β -PGM, and newly identified pyrophosphatase activities have evolved as well. Through the comparison of these structures, we gain insight into the key structural changes that might be responsible for the change in function. In the case of the inorganic pyrophosphatase BT2127 and its closest relative, β -PGM, the key structural changes are seemingly directed at how large the leaving group can be (phosphate vs glucose phosphate) and the position of the Asp acid/base in the cap-open versus cap-closed conformation. The BT2127 pyrophosphatase active site precludes binding of β -glucose 1,6-bisphosphate. The two key residues that restrict the volume of the catalytic site are cap domain residues Tyr76 and Glu47. The β -PGM activity requires the movement of the Asp acid/base out of the active site when the cap opens to allow the association of the β -glucose 1-phosphate or glucose 6-phosphate for reaction with the aspartyl phosphate. In both BT2127 and β -PGM, the Asp acid/base forms a hydrogen bond with the linker hinge residue. In β -PGM, this hydrogen bond occurs only in the cap-open conformation, and it positions the Asp acid/base outside of the active site. In the BT2127 pyrophosphatase, this hydrogen bond is present in both the cap-open and cap-closed conformations and the Asp acid/base remains in the active site. When the β -PGM Asp acid/base swings into the active site upon cap domain closure, it forms a hydrogen bond with His20 of the cap domain His20-Lys76 diad. Replacement of His20 of β -PGM with Ala reduces the k_{cat} 7000-fold, whereas replacement of Lys76 with Ala reduces the k_{cat} only 100-fold.³⁶ Replacement of the BT2127 counterparts His23 and Lys49 with Ala indicates that the formation of the

hydrogen bond with the Asp13 acid/base makes a minor contribution to the turnover rate (k_{cat} reduced 60-fold) compared to the contribution made by Lys79 (k_{cat} reduced 950-fold), which we posit might function as a docking site for the displaced phosphate as needed for the hydrolysis of the aspartyl phosphate intermediate in the second partial reaction. The pathway followed in the evolution of β -PGM and BT2127 is not clear. One might have evolved from the other, or they might have evolved separately from a common ancestor. The fact that they have assumed very specific and separate biological niches is intriguing.

From a mechanistic enzymology perspective, the function of BT2127 identifies the HADSF as a third protein fold family that has evolved a bona fide inorganic pyrophosphatase. Moreover, the catalytic mechanism employed by BT2127 is unlike the Mg^{2+} -centered catalytic mechanisms used by the type I and type II inorganic pyrophosphatases. From an orthologue tracking perspective, the *Bacteroides* inorganic pyrophosphatase BT2127 is a member of the HADSF that extends the chemistries beyond those of the HADSF phosphatase activities found in *E. coli*.

■ ASSOCIATED CONTENT

● Supporting Information

Tables S1–S8 and Figures S1–S9. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Accession Codes

The coordinates for the BT2127 X-ray crystal structures are deposited in the Protein Data Bank as entries 3QU2 (wild type), 3QXG (wild type), 3QX7 (wild type), 3QUQ (wild type), 3QU5 (D11N mutant), 3QU4 (D13A mutant), 3QU7 (D13N mutant), 3QU9 (D13N mutant), 3QUT (D13N mutant), 3QU7 (D13N mutant), 3R9K (E47D mutant), 3QYP (E47N mutant), 3QUC (E47N mutant), and 3QUB (E47A mutant).

■ AUTHOR INFORMATION

Corresponding Author

*X-ray structure determination (S.C.A. or K.N.A.) or kinetic analyses (D.D.-M.). D.D.-M.: Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, NM 87111; telephone, (505) 277-3383; e-mail, dd39@unm.edu; fax, (505) 277-6202. K.N.A.: Department of Chemistry, Boston University, 590 Commonwealth Ave., Room 299, Boston, MA 02215-2521; telephone, (617) 358-5544; e-mail, drkallen@bu.edu; fax, (202) 354-5386. S.C.A.: Department of Biochemistry, Albert Einstein College of Medicine, Jack and Pearl Resnick Campus, 1300 Morris Park Ave., Ullmann Building, Room 409, Bronx, NY 10461; telephone, (718) 430-2746; e-mail, steve.almo@einstein.yu.edu; fax, (718) 430-8565.

Funding

This work was supported by National Institutes of Health Grants U54 GM093342 (K.N.A., D.D.-M., and S.C.A.) and GM61099. Data for this study were collected at beamline X29A of the National Synchrotron Light Source, Brookhaven National Laboratory. Financial support for the operation of the beamline comes principally from NIBIB, National Institutes of Health, Grant P30-EB-09998, and the office of Basic Energy Sciences of the United States Department of Energy.

■ ABBREVIATIONS

DTT, dithiothreitol; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; Tris, tris(hydroxymethyl)aminomethane; MESG, 2-amino-6-mercapto-7-methylpurine ribonucleoside; MES, 2-(*N*-morpholino)ethanesulfonic acid; PDB, Protein Data Bank; β -PGM, β -phosphoglucosyltransferase.

■ ADDITIONAL NOTE

“The “substrate specificity loop” (first identified in phosphonate and β -PGM) is the region of the cap domain that places one or more amino acid side chains at the catalytic site. In the case of phosphonate and β -PGM, the key loop residue is a Lys that in the former forms a Schiff base adduct with the substrate and in the latter binds the substrate leaving group. In BT2127, the substrate specificity loop Glu47 forms a coordination bond with the Mg^{2+} cofactor. We do not believe that this interaction is an artifact of crystallization because it is observed in all BT2127 structures (which include multiple independent crystal forms) that possess Glu47 and cofactor. Moreover, superposition of the X-ray structure of the BT2127 orthologue from *B. vulgatus* (PDB entry 3DV9) (vide infra) shows that the homologous residue (Glu62) is likewise coordinated to the Mg^{2+} . The BT2127 Glu47– Mg^{2+} coordination is not essential for cap domain closure as indicated by the structures of the BT2127 E47 variants [PDB entries 3QUB (E47A), 3R9K (E47D), and 3QUC (E47N)], which are all observed in the cap-closed conformation.

■ REFERENCES

- (1) Allen, K. N., and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: Evolution of a catalytic scaffold. *Trends Biochem. Sci.* 29, 495–503.
- (2) 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.
- (3) Allen, K. N., and Dunaway-Mariano, D. (2009) Markers of fitness in a successful enzyme superfamily. *Curr. Opin. Struct. Biol.* 19, 658–665.
- (4) Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V., Borozan, I., Carmel, L., Wolf, Y. I., Mori, H., Savchenko, A. V., Arrowsmith, C. H., Koonin, E. V., Edwards, A. M., and Yakunin, A. F. (2006) Genome-wide analysis of substrate specificities of the *Escherichia coli* haloacid dehalogenase-like phosphatase family. *J. Biol. Chem.* 281, 36149–36161.
- (5) 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.
- (6) 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.
- (7) Joseph, T. C., Rajan, L. A., Thampuran, N., and James, R. (2010) Functional characterization of trehalose biosynthesis genes from *E. coli*: An osmolyte involved in stress tolerance. *Mol. Biotechnol.* 46, 20–25.
- (8) 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.
- (9) Wang, L., Huang, H., Nguyen, H. 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* 49, 1072–1081.

- (10) Titz, B., Hauser, R., Engelbrecher, A., and Uetz, P. (2007) The *Escherichia coli* protein YjjG is a house-cleaning nucleotidase in vivo. *FEMS Microbiol. Lett.* 270, 49–57.
- (11) Weiss, B. (2007) YjjG, a dUMP phosphatase, is critical for thymine utilization by *Escherichia coli* K-12. *J. Bacteriol.* 189, 2186–2189.
- (12) Lu, Z., Wang, L., Dunaway-Mariano, D., and Allen, K. N. (2009) Structure-function analysis of 2-keto-3-deoxy-D-glycero-D-galactononate-9-phosphate phosphatase defines specificity elements in type C0 haloalkanoate dehalogenase family members. *J. Biol. Chem.* 284, 1224–1233.
- (13) Wang, L., Lu, Z., Allen, K. N., Mariano, P. S., and Dunaway-Mariano, D. (2008) Human symbiont *Bacteroides thetaiotaomicron* synthesizes 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN). *Chem. Biol.* 15, 893–897.
- (14) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* 276, 307–326.
- (15) Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* 50, 760–763.
- (16) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* 53, 240–255.
- (17) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D* 60, 2126–2132.
- (18) Stockbridge, R. B., and Wolfenden, R. (2011) Enhancement of the rate of pyrophosphate hydrolysis by nonenzymatic catalysts and by inorganic pyrophosphatase. *J. Biol. Chem.* 286, 18538–18546.
- (19) 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.
- (20) Tremblay, L. W., Dunaway-Mariano, D., and Allen, K. N. (2006) Structure and Activity Analyses of *Escherichia coli* K-12 NagD Provide Insight Into the Evolution of Biochemical Function in the Haloalkanoic Acid Dehalogenase Superfamily. *Biochemistry* 45, 1183–1193.
- (21) Peeraer, Y., Rabijns, A., Verboven, C., Collet, J. F., Van Schaftingen, E., and De Ranter, C. (2003) High-resolution structure of human phosphoserine phosphatase in open conformation. *Acta Crystallogr. D* 59, 971–977.
- (22) 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.
- (23) 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.
- (24) Holm, L., and Sander, C. (1997) Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res.* 25, 231–234.
- (25) 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.
- (26) Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2003) The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science* 299, 2067–2071.
- (27) Hayward, S., Kitao, A., and Berendsen, H. J. (1997) Model-free methods of analyzing domain motions in proteins from simulation: A comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins* 27, 425–437.
- (28) Hayward, S., and Lee, R. A. (2002) Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Mol. Graphics Modell.* 21, 181–183.
- (29) Zhang, G., Morais, M. C., Dai, J., Zhang, W., Dunaway-Mariano, D., and Allen, K. N. (2004) Investigation of Metal Ion Binding in Phosphonoacetaldehyde Hydrolase Identifies Sequence Markers for Metal-Activated Enzymes of the HAD Enzyme Superfamily. *Biochemistry* 43, 4990–4997.
- (30) 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.
- (31) Coyne, M. J., Reinap, B., Lee, M. M., and Comstock, L. E. (2005) Human symbionts use a host-like pathway for surface fucosylation. *Science* 307, 1778–1781.
- (32) Ahn, S., Milner, A. J., Futterer, K., Konopka, M., Ilias, M., Young, T. W., and White, S. A. (2001) The “open” and “closed” structures of the type-C inorganic pyrophosphatases from *Bacillus subtilis* and *Streptococcus gordonii*. *J. Mol. Biol.* 313, 797–811.
- (33) Oksanen, E., Ahonen, A. K., Tuominen, H., Tuominen, V., Lahti, R., Goldman, A., and Heikinheimo, P. (2007) A complete structural description of the catalytic cycle of yeast pyrophosphatase. *Biochemistry* 46, 1228–1239.
- (34) Fabrichny, I. P., Lehtio, L., Tammenkoski, M., Zyryanov, A. B., Oksanen, E., Baykov, A. A., Lahti, R., and Goldman, A. (2007) A trimetal site and substrate distortion in a family II inorganic pyrophosphatase. *J. Biol. Chem.* 282, 1422–1431.
- (35) Lee, H. S., Cho, Y., Kim, Y. J., Lho, T. O., Cha, S. S., Lee, J. H., and Kang, S. G. (2009) A novel inorganic pyrophosphatase in *Thermococcus onnurineus* NA1. *FEMS Microbiol. Lett.* 300, 68–74.
- (36) Dai, J. (2005) Mechanism of β -phosphoglucomutase catalysis. Ph.D. Thesis, University of New Mexico, Albuquerque, NM.