

Prospecting for Unannotated Enzymes: Discovery of a 3',5'-Nucleotide Bisphosphate Phosphatase within the Amidohydrolase Superfamily

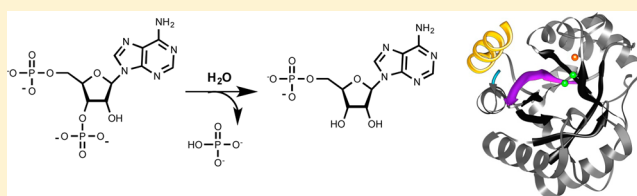
Jennifer A. Cummings,[†] Matthew Vetting,[‡] Swapnil V. Ghodge,[†] Chengfu Xu,[†] Brandan Hillerich,[‡] Ronald D. Seidel,[‡] Steven C. Almo,^{*,‡} and Frank M. Raushel^{*,†}

[†]Department of Chemistry, Texas A&M University, College Station, Texas 77843, United States

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

Supporting Information

ABSTRACT: In bacteria, 3',5'-adenosine bisphosphate (pAp) is generated from 3'-phosphoadenosine 5'-phosphosulfate in the sulfate assimilation pathway, and from coenzyme A by the transfer of the phosphopantetheine group to the acyl-carrier protein. pAp is subsequently hydrolyzed to 5'-AMP and orthophosphate, and this reaction has been shown to be important for superoxide stress tolerance. Herein, we report the discovery of the first instance of an enzyme from the amidohydrolase superfamily that is capable of hydrolyzing pAp. Crystal structures of Cv1693 from *Chromobacterium violaceum* have been determined to a resolution of 1.9 Å with AMP and orthophosphate bound in the active site. The enzyme has a trinuclear metal center in the active site with three Mn²⁺ ions. This enzyme (Cv1693) belongs to the Cluster of Orthologous Groups cog0613 from the polymerase and histidinol phosphatase family of enzymes. The values of k_{cat} and k_{cat}/K_m for the hydrolysis of pAp are 22 s⁻¹ and 1.4 × 10⁶ M⁻¹ s⁻¹, respectively. The enzyme is promiscuous and is able to hydrolyze other 3',5'-bisphosphonucleotides (pGp, pCp, pUp, and pIp) and 2'-deoxynucleotides with comparable catalytic efficiency. The enzyme is capable of hydrolyzing short oligonucleotides (pdA)₅, albeit at rates much lower than that of pAp. Enzymes from two other enzyme families have previously been found to hydrolyze pAp at physiologically significant rates. These enzymes include CysQ from *Escherichia coli* (cog1218) and Ytq1/NrnA from *Bacillus subtilis* (cog0618). Identification of the functional homologues to the experimentally verified pAp phosphatases from cog0613, cog1218, and cog0618 suggests that there is relatively little overlap of enzymes with this function in sequenced bacterial genomes.



The amidohydrolase superfamily (AHS) is a large group of enzymes that catalyze a remarkably diverse set of hydrolytic reactions. In general, these enzymatic reactions are directed at the hydrolysis of amide and ester functional groups at carbon and phosphorus centers.¹ Structural and mechanistic studies of the proteins within the AHS have shown that these enzymes utilize from zero to three divalent metal ions in the active site.^{2–4} The catalytic mechanism of hydrolysis utilizes an activated water/hydroxide for nucleophilic attack at an electrophilic carbon or phosphorus center. The enzymes in the AHS that possess a trinuclear metal center in the active site belong to the polymerase and histidinol phosphatase (PHP) family of proteins and are involved in the hydrolysis of phosphoesters.^{4,5} The available crystal structures from the PHP family indicate that these enzymes possess a distorted (β/α)₇-barrel structural fold, as opposed to the distorted (β/α)₈-barrel fold observed in most of the other members of the AHS.^{6,7}

The PHP family within the amidohydrolase superfamily is further subdivided into three Clusters of Orthologous Groups: cog1387, cog0613, and cog4464. cog4464 consists of protein tyrosine phosphatases, which have been implicated in the regulation of capsular polysaccharide assembly in Gram-positive

bacteria. Biochemical studies have provided insights into the structure and catalytically important residues of these tyrosine phosphatases.^{8,9} cog1387 consists of L-histidinol phosphate phosphatase (HPP), an enzyme from the L-histidine biosynthetic pathway, and other related enzymes, including the PHP domain of DNA polymerase X involved in DNA repair. Structural and mechanistic investigations of HPP enzymes have revealed the orientation of the substrate with respect to the metal center in the active site.⁴ Further work has also elucidated the role of the trinuclear active site in the activation of the nucleophilic hydroxide, and Lewis acid stabilization of the alcohol product.^{4,6} The mechanistic details obtained from an improved understanding of the enzymes from cog1387 should be directly applicable, in principle, to the enzymes belonging to cog0613. cog0613 consists of PHP family members that have been generically annotated as metal-dependent phosphoesterases in various online databases (NCBI and Uniprot). A sequence similarity network (SSN) map at an *E* value cutoff of

Received: December 7, 2013

Revised: January 7, 2014

Published: January 8, 2014

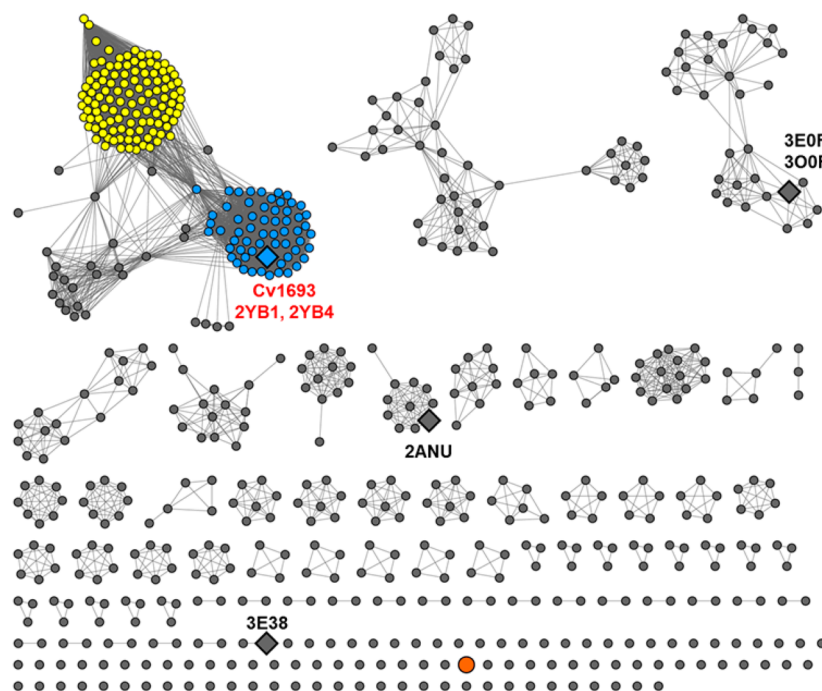


Figure 1. Sequence network map of cog0613 at an E value cutoff of 10^{-60} created using Cytoscape (<http://www.cytoscape.org>). Each node (sphere) represents a protein sequence, while each edge (line) represents those sequence pairs that are more closely related than the arbitrary E value cutoff (10^{-60}). The available crystal structures are shown as diamonds, and their respective PDB entries are given. The enzyme studied in this work, Cv1693 from *C. violaceum* ATCC 12472, and its orthologs are colored blue. Enzyme sequences colored yellow are those that have been annotated as trpH, while the single orange node represents cyclic phosphate hydrolase from *E. lenta* DSM 2243 (Elen0235), the only other enzymatic reaction known and demonstrated from cog0613.

10^{-60} of 650 nonredundant protein sequences from cog0613 is presented in Figure 1.^{10,11} There are currently three proteins from cog0613 whose crystal structures have been determined. These include Protein Data Bank (PDB) entry 3O0F (locus tag Bad1165 from *Bifidobacterium adolescentis*), PDB entry 2ANU (locus tag Tm0559 from *Thermotoga maritima*), and PDB entry 3E38 (locus tag Bvu3505 from *Bacteroides vulgatus*). The physiological substrates for these enzymes are unknown, but it has been shown that Bad1165 can hydrolyze *p*-nitrophenyl phosphate, albeit very slowly, indicating that it is a phosphoesterase. The crystal structure of this enzyme has been determined with 5'-adenosine monophosphate (5'-AMP) bound in the active site, and it has been demonstrated that this enzyme does not possess DNA polymerase or DNA proof-reading activity.¹² The only reported enzymatic function for any enzyme from cog0613 is that of cyclic phosphate dihydrolase (cPDH) for Elen0235 from *Eggerthella lenta*, which hydrolyzes the cyclic phosphate of 5-phospho-D-ribose 1,2-cyclic phosphate (PRcP) to D-ribose 5-phosphate and orthophosphate in a nonprocessive manner through D-ribose 2,5-bisphosphate as an obligatory reaction intermediate.¹³

Our search for new enzymatic functions in cog0613 focused on the characterization of proteins of unknown function contained within the two large groups of proteins (colored yellow and blue) in the SSN depicted in Figure 1. Toward that end, we have purified and determined the three-dimensional crystal structure of Cv1693 from *Chromobacterium violaceum* and have shown that this enzyme catalyzes the hydrolysis of 3',5'-adenosine bisphosphate (pAp) to 5'-AMP and orthophosphate.

■ MATERIALS AND METHODS

Materials. HEPES, 3',5'-adenosine bisphosphate (pAp), 2',5'-adenosine bisphosphate (2',5'-pAp), 3'-adenosine monophosphate (3'-AMP), 5'-adenosine monophosphate (AMP), phosphoenolpyruvate (PEP), NADH, and the enzymes myokinase, pyruvate kinase, and lactate dehydrogenase were purchased from Sigma-Aldrich. The 3'- and 5'-phosphorylated derivatives of short DNA oligonucleotides (dAdAdAdAdAp and pdAdAdAdAdA) were synthesized by Integrated DNA Technologies. The dinucleotides, 5'-phosphoadenylyl-(3'→5')-adenosine (pApA), and 5'-phosphoguanlylyl-(3'→5')-guanosine (pGpG) were purchased from Biolog Life Science Institute (Axxora, U.S. distributor). Guanosine 3',5'-bisphosphate (ppGpp) was purchased from TriLink Biotechnologies.

Synthesis of 3',5'-Nucleotide Bisphosphates. The synthesis of 2'-deoxy-3',5'-pAp, 2'-deoxy-3',5'-pCp, 2'-deoxy-3',5'-pGp, 2'-deoxy-3',5'-pIp, 2'-deoxy-3',5'-pUp, 3',5'-pTp, pCp (2',5'/3',5'-bisphosphate mixture), pGp (2',5'/3',5'-bisphosphate mixture), and pUp (2',5'/3',5'-bisphosphate mixture) were synthesized according to the method of Barrio et al.¹⁴

Cloning and Expression. A codon-optimized synthetic gene for Cv1693 (gil34497148) from *C. violaceum* was inserted into a pET30 variant by ligation-independent cloning (LIC), utilizing the synthetic gene as a template and 5'-TTAAGAA-GGAGATATACCATGGCAAACATTGACCTGCATTTTC-ATTC-3' and 5'-GATTGGAAGTAGAGTTCTCTGCATC-AGCCGGACGCAGGATACG-3' as primers.¹⁵ The resultant protein is identical to the native protein with the inclusion of a C-terminal tag (added sequence AENLYFQSHHHHHHS-HPQFEK), consisting of a tobacco etch virus (TEV) cleavage

Table 1. Data Collection Statistics for Cv1693^a

	10 mM HgCl ₂ ^b	10 mM TMLA	unliganded	Mn ²⁺ , AMP
space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
unit cell (Å)	<i>a</i> = 40.0	<i>a</i> = 40.0	<i>a</i> = 40.1	<i>a</i> = 39.9
	<i>b</i> = 46.8	<i>b</i> = 46.8	<i>b</i> = 46.5	<i>b</i> = 46.7
	<i>c</i> = 130.4	<i>c</i> = 130.8	<i>c</i> = 130.5	<i>c</i> = 131.0
resolution (Å)	40–2.4 (2.44–2.40)	40–1.9 (1.93–1.90)	40–2.20 (2.32–2.20)	40–1.9 (2.0–1.9)
completeness (%)	98.1 (92.4)	98.7 (85.1)	99.7 (100.0)	96.9 (82.6)
redundancy	6.7 (6.4)	6.2 (2.9)	3.8 (3.7)	5.3 (3.8)
mean <i>I</i> (standard deviation)	20.8 (6.5)	21.6 (3.2)	10.3 (4.8)	12.9 (4.1)
<i>R</i> _{sym}	0.096 (0.276)	0.089 (0.243)	0.098 (0.264)	0.083 (0.301)

^aData in parentheses are for the highest-resolution bin. ^bNo binding, used as the native data set.

site, a hexahistidine tag, and a STREP tag. The plasmid was transformed into ROSETTA2(plysS) and plated on LB agar plates (100 μg/mL kanamycin, 35 μg/mL chloroamphenicol, and 0.5% glucose). Ten colonies were transferred to 100 mL of LB (100 μg/mL kanamycin, 35 μg/mL chloroamphenicol, and 0.5% glucose) and grown overnight at 37 °C. The overnight culture was added to 4 L of autoinduction medium (25 mM Na₂HPO₄, 25 mM KH₂PO₄, 50 mM NH₄Cl, 5 mM Na₂SO₄, 25 mM sodium succinate, 2 mM MgSO₄, 0.5% yeast extract, 1% NZ-AMINE-AS, 0.9% glycerol, 0.45% lactose, and 0.03% glucose) distributed into eight 2 L baffled flasks.¹⁶ The flasks were shaken (250 rpm) at 30 °C for 4 h (OD < 3), and the temperature was reduced to 23 °C for an additional 16 h. The cells were harvested by centrifugation and stored at –80 °C. To increase the metal content of the isolated protein, additional expressions that were supplemented with 150 μM 2,2'-bipyridine, 1.0 mM ZnCl₂, and 1.0 mM MnCl₂ were performed prior to the temperature being reduced to 23 °C.

Purification. Thirty grams of cells were resuspended in 150 mL of buffer A [50 mM HEPES (pH 7.8), 150 mM NaCl, 20 mM imidazole, and 10% glycerol] and disrupted by sonication. The lysate was clarified by centrifugation and applied to a 10 mL Ni-Sepharose HP column equilibrated against buffer A. The column was washed with 10 column volumes of buffer A, and bound proteins were eluted with 3 column volumes of buffer B [50 mM HEPES (pH 7.8), 150 mM NaCl, 20 mM imidazole, and 10% glycerol]. Eluted proteins (~10–30 mg/mL) were pooled, aliquoted (10 mL), frozen by immersion in liquid nitrogen, and stored at –80 °C. Individual aliquots were rapidly thawed, incubated at 4 °C overnight with purified TEV protease (1/100 weight ratio), and applied in two separate runs to a Superdex 200 16/60 (GE Healthcare) column equilibrated against buffer C [10 mM HEPES (pH 7.8), 150 mM NaCl, and 10% glycerol]. Fractions from the primary peak were pooled and concentrated to 10–15 mg/mL by centrifugal ultrafiltration, aliquoted (500 μL), frozen by being immersed in liquid nitrogen, and stored at –80 °C. The purified protein contained 0.6 equiv of Mn and 2.4 equiv of Zn as determined by inductively coupled mass spectrometry (ICP-MS).

Assay for AMP. Cv1693 hydrolyzes pApA to release two molecules of 5'-adenosine monophosphate (AMP). The rate of hydrolysis of pApA was measured by coupling the formation of AMP to the oxidation of NADH. Each 250 μL assay contained 250 mM KCl, 0.7 mM PEP, 0.5 mM ATP, 5.0 mM MgCl₂, 0.2 mM NADH, myokinase (2 units), pyruvate kinase (2 units), and L-lactate dehydrogenase (3 units) in 50 mM HEPES (pH 7.5). The oxidation of NADH was followed at 340 nm with a Spectromax 384 Plus 96-well plate reader from Molecular Devices. The AMP coupling system assay was also used to

measure the initial rate of release of 5'-dAMP from the oligomer of DNA (pdApdApdApdApdA) or 2',5'-pAp.

Assay for Phosphate. The formation of phosphate was assessed using the P_i ColorLock Gold assay kit from Innova Biosciences according to the manufacturer's instructions. The enzymatic assays were conducted in 25 mM HEPES containing 250 mM KCl (pH 7.5) at 30 °C. The extinction coefficient of the phosphate–dye complex was 68000 M⁻¹ cm⁻¹ (λ = 650 nm) under the conditions used to assess the enzymatic activity. The phosphate detection assay was also used to determine if free phosphate was produced when 1.0 μM Cv1693 was incubated with 0.1 mM dAdAdAdAdAp, 3'-AMP, 3'-NADPH, CoA, 1,3-bisphosphoglycerol, 3'-phosphoadenosine 5'-phosphosulfate (PAPS), pGpG, and ppGpp for 75 min at 30 °C.

Crystallization. Cv1693 was crystallized by sitting drop vapor diffusion at 20 °C utilizing Intelliplates and 2 + 2 μL drops equilibrated against 70 μL reservoir solutions. All crystallization work used a protein that was derived from expressions that did not contain 2,2'-bipyridine, ZnCl₂, or MnCl₂. Crystals for the unliganded and Mn/AMP structures were grown from protein that included 2.0 mM EDTA during the overnight TEV cleavage reaction. Crystals utilized for phasing were grown by combining protein [10–15 mg/mL, 150 mM NaCl, 10% glycerol, 20 mM inositol 5'-monophosphate, and 10 mM HEPES (pH 7.8)] with 20% Peg2000 MME and 100 mM Bis-Tris (pH 6.0). For the unliganded (sulfate bound) structure, the crystallization conditions were Cv1693 [10–15 mg/mL, 150 mM NaCl, 10% glycerol, and 10 mM HEPES (pH 7.8)] and 25% PEG3350, 200 mM (NH₄)₂SO₄, and 100 mM Bis-Tris (pH 6.5). For the Mn²⁺–AMP complex, Cv1693 [10–15 mg/mL, 150 mM NaCl, 10% glycerol, 20 mM adenosine 5'-monophosphate, 5.0 mM MnCl₂, and 10 mM HEPES (pH 7.8)] was combined with 20% Peg2000 MME and 100 mM Bis-Tris (pH 6.0). Prior to data collection, crystals were soaked in precipitant supplemented with 20% ethylene glycol and vitrified by being immersed in liquid nitrogen. All data were collected in-house on an R-AXIS IV⁺⁺ detector with Cu Kα X-rays generated by a Rigaku RU-H3R rotating anode generator and focused with OSMIC mirrors. Data were integrated and scaled with HKL3000.¹⁷ Data statistics are listed in Table 1.

Structure Determination. To obtain experimental phases, data were collected from crystals soaked for 3 min in a cryo solution supplemented with either 50 mM trimethyllead acetate (TMLA) or 10 mM HgCl₂. The structure was determined using single isomorphous replacement with anomalous scattering (SIRAS) utilizing the TMLA data set as the anomalously scattering derivative and the HgCl₂ data set as "native". Determination of the heavy atom constellation,

Table 2. Structural Statistics for Cv1693^a

	unliganded	Mn ²⁺ , AMP
resolution (Å)	40–2.2 (2.42–2.2)	40–1.9 (2.0–1.9)
no. of unique reflections	12140 (2942)	18950 (2050)
R _{cryst} (%)	17.0 (17.7)	15.8 (19.6)
R _{free} (%; 5% of data)	23.3 (25.5)	20.7 (24.8)
contents of model		
residues 1–285 ^b	2–254, 258–288	2–254, 258–288
waters	194	271
Mn ²⁺	0	3
total no. of atoms	2340	2438
average Wilson B Factor (Å ²)	18.6	12.5
no. of TLS groups	3	4
no. of proteins/waters/Mn ²⁺ /AMP	9.4/14.2/–/–	11.5/21.1/9.0/13.1
root-mean-square deviation		
bond lengths (Å)	0.006	0.007
bond angles (deg)	1.05	1.13
MOLPROBITY statistics		
Ramachandran		
favored/outliers (%)	98.9/0.0	98.9/0.0
rotamer outliers (%)	1.9	1.4
Clashscore ^c	7.1 (97th percentile)	6.4 (93rd percentile)
overall score ^c	1.6 (97th percentile)	1.5 (96th percentile)
PDB entry	2yb4	2yb1

^aStatistics in parentheses are for the highest-resolution bin. ^bAdditional residues past position 285 are from the C-terminal tag. ^cScores are ranked according to structures of similar resolution as formulated in MOLPROBITY.

phasing, and automatic chain tracing were performed with PHENIX.¹⁸ The starting model was completed by iterative rounds of model building utilizing COOT¹⁹ followed by refinement using PHENIX. There was a single molecule per asymmetric unit with a solvent content of 37%. Waters were added to difference density map peaks ($F_o - F_c > 3.5\sigma$) located <3.2 Å from potential hydrogen bonding partners. Translation–libration–screw (TLS) refinement was included in the final rounds with the number of domains determined by PHENIX. The quality of the final model was validated with MOLPROBITY.²⁰ Data collection and refinement statistics for Cv1693 are listed in Tables 1 and 2.

Analytical Size Exclusion Chromatography. Proteins were analyzed by analytical size exclusion chromatography with a 10 mm × 300 mm SUPERDEX 200 column (GE-Healthcare) using thyroglobulin (670 kDa), δ -globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), and vitamin B12 (1.35 kDa) as molecular mass standards (Bio-Rad 151-1901).

Bioinformatics. The protein sequences from various cog's were retrieved from the NCBI (<http://www.ncbi.nlm.nih.gov>) and Microbesonline (<http://www.microbesonline.org>) databases.²¹ The redundant sequences from cog0613 obtained from the NCBI database were removed using Jalview version 2.7²² and then converted into FASTA format. Cytoscape version 2.8.2 was used to create protein sequence network diagrams,²³ while Jalview version 2.7 was used to make the amino acid sequence alignments.

RESULTS

Structure Determination of Cv1693. Crystals of Cv1693 from the original crystallization attempts were thin spindly needles that did not improve in shape upon standard grid screening protocols. A search of the PDB revealed that the structure of a distantly related protein from cog0613 (Bad1165 from *B. adolescentis*, PDB entry 3O0F, sequence identity of

25%) had previously been determined with bound adenosine monophosphate (AMP). In addition, several mono- and diphosphorylated nucleosides were found to increase the thermostability of Cv1693 by fluorescence-monitored thermal denaturation (data not shown). As such, the original Cv1693 crystallization conditions were rescreened with added inosine monophosphate (IMP), cytidine monophosphate (CMP), or AMP. In these experiments, IMP gave the best improvement in the crystals, such that several derivative data sets could be collected. The structure of Cv1693 was determined by single isomorphous replacement with anomalous dispersion from a trimethyllead acetate derivatized crystal. Analysis of the electron density maps originating from Cv1693 crystals grown in the presence of IMP with no added metal indicated low occupancy for the three metal sites and unsatisfactory density for the nucleoside. The occupancies of the metal sites are consistent with the ICP-metal analysis for the protein used in the crystallization, which exhibited a mixture of Ni²⁺ (0.6 equiv), Zn²⁺ (0.2 equiv), and Fe²⁺ (0.1 equiv).

In an attempt to improve the homogeneity of the crystals, we performed a new round of crystal screening after Cv1693 had been treated with EDTA to remove the metals. The structure of apoenzyme (no metals) Cv1693 was determined from crystallization conditions that included Peg3350 and ammonium sulfate. In this structure, there is no electron density for the metal ions, and sulfate is bound in the nucleoside 5'-phosphate binding site. The structure of the complex with Mn²⁺ and AMP was determined from crystals grown in the presence of Mn²⁺ and AMP without ammonium sulfate. The structures of Cv1693 and the Cv1693 metal–nucleoside complex are nearly identical with no large-scale structural movements [root-mean-square deviation (rmsd) of 0.22 Å, 284 common C α atoms], and only small side chain positional differences in the ligands to the metals; all subsequent discussion will pertain to the AMP-bound structure.

Structure of Cv1693. Cv1693 crystallized with a single molecule per asymmetric unit. An analysis of the intermolecular contacts within the crystal suggests that Cv1693 is monomeric in solution, consistent with analytical gel filtration results. The entire sequence of Cv1693 could be fit to the electron density except for a portion of one loop (residues 255–257). Refinement statistics are listed in Table 2. Cv1693 exhibits the (β/α)-barrel fold common to members of the amidohydrolase superfamily but with significant deviations (Figure 2).

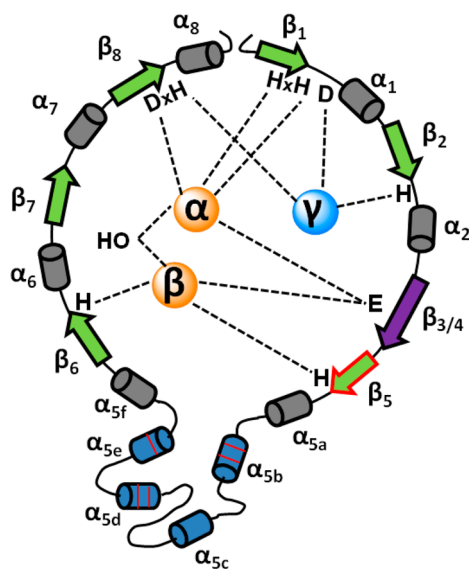


Figure 2. Schematic of the secondary structure of Cv1693. The α - and β -metal sites, which are a common feature among members of the PHP family and some amidohydrolase enzymes, are shown as orange spheres, while the γ -metal site, unique to the PHP family members, is shown as a blue sphere. The strands and helices that constitute the distorted TIM- or (β/α)-barrel are colored green and gray, respectively. The long β -strand-3/4 is colored purple, while the antiparallel β -strand 5, unique to members of cog0613 among AHS members, is highlighted with a red border. The insertion element consisting of α -helices is colored blue, and the positions of the enzyme residues seen to interact with the bound 5'-AMP are shown as red lines.

In Cv1693, the third β -strand of the barrel, β_3 , is significantly longer than the strand usually seen for other members of the AHS. Furthermore, the bridging ligand between two of the metals (M_α and M_β) is usually located at the end of β -strand 3 or 4, but in Cv1693, the bridging ligand (Glu-64) is positioned at the center of elongated β -strand 3. Therefore, in keeping with the β -strand numbering scheme for other members of the AHS used to designate key active site residues with respect to their β -strand number, we have chosen to designate β_3 as $\beta_{3/4}$. While most AHS (β/α)-barrels are composed completely of parallel β -strands, the fifth β -strand of Cv1693, β_5 , is antiparallel to $\beta_{3/4}$ and β_6 . In addition, Cv1693 lacks two (β/α)-barrel α -helices (α_3 and α_4) and includes an insertion (amino acids 83–171, α_{5b} – α_{5e}) between β_5 and α_{5f} . One of the helices, α_{5a} , assumes the typical position of α_3 , and the other, α_{5a} – α_{5e} , forms a four-helix bundle that caps the top of the (β/α)-barrel and contains several residues that contact adenosine 5'-monophosphate. These residues include Arg-99, Arg-102, Arg-134, Thr-135, and Phe-154. A stereo image of Cv1693 is presented in Figure 3.

Metal Center. Clear electron density was observed for three manganese ions (Mn_α , Mn_β , and Mn_γ) and AMP. In addition, a tetragonally shaped density feature bridging the three metal ions was interpreted as a phosphate ion assumed to originate from degradation of AMP (Figure 4). A comparison of thermal factors with coordinating protein ligands suggests the Mn^{2+} –AMP– HPO_4^{2-} species approach full occupancy in the ternary complex. Mn_α and Mn_β are bridged by contacts to Glu-64 from β_3 and an oxygen from inorganic phosphate. Mn_α is coordinated to His-7 and His-9 at the end of β -strand 1 and Asp-248 at the end of β -strand 8. Mn_β is ligated to His-75 (β_4) and His-191 (end of β_6). The third metal coordinates to Asp-14 (β_1 – α_1 loop), His-39 (β_2 – α_2 loop), and His-250 (β_8 – α_8 loop). The orthophosphate forms a bridge not only from Mn_α to Mn_β but also from Mn_α to Mn_γ , and is additionally coordinated by the guanidinium group of Arg-194 projecting from the β_6 – α_6 loop (Figure 5).

Contacts between Cv1693 and AMP. All of the contacts between Cv1693 and AMP originate from the insertion domain between β_5 and α_{5f} . Arg-99 and Arg-103 from α_{5b} and Arg-134 from α_{5d} form salt bridges to the 5'-phosphate of AMP. In the APO structure, this binding pocket was occupied by a sulfate ion. The only contact to the adenine base is a stacking interaction with Phe-154 (α_{5e}). The 2'-OH group of the ribose moiety forms a hydrogen bond to Thr-135, and the 2'-OH (2.4 Å) and 3'-OH (2.5 Å) groups are coordinated to the third metal (M_γ). Interestingly, the 3'-OH group is only 2.5 Å from an oxygen of the metal-bound orthophosphate. The contacts to the bound AMP and P_i are depicted in Figure 6.

Substrate Profile. The kinetic constants k_{cat}/K_m and k_{cat}/K_m are summarized in Table 3 for the hydrolysis of pApA and 11 3',5'-nucleotide bisphosphates. The substrates with the highest activity were N^6 -methyl-3',5'-pAp ($k_{cat}/K_m = 2.1 \times 10^6 M^{-1} s^{-1}$) and pAp ($k_{cat}/K_m = 1.4 \times 10^6 M^{-1} s^{-1}$), although Cv1693 hydrolyzed several other 3',5'-nucleotide bisphosphates with similar catalytic efficiencies. At a 2',5'-pAp concentration of 100 μM , the upper limit for substrate turnover was 0.015 s^{-1} in the presence of Cv1693. This is more than 3 orders of magnitude slower than the hydrolysis of 3',5'-pAp. In all three of the bisphosphate mixtures (pCp, pGp, and pUp), the ratios of 2',5'-pNp and 3',5'-pNp nucleotides were close to 1:1, and thus, the apparent kinetic constants are assumed to be for the 3',5'-pNp compound. However, the inhibitory properties of the 2',5'-pNp compounds have not been determined. Cv1693 was unable to hydrolyze other compounds when assayed at a concentration of 100 μM : 3'-NADPH ($\leq 8 \times 10^{-4} s^{-1}$), 3'-AMP ($\leq 2 \times 10^{-4} s^{-1}$), CoA ($\leq 7 \times 10^{-4} s^{-1}$), and a 3'-phosphorylated 5-mer of DNA dAdAdAdAdAp ($\leq 1 \times 10^{-4} s^{-1}$). The 5'-phosphorylated 5-mer of DNA, pdApdApdApdApdA, abbreviated (pdA)₅, was a poor substrate for Cv1693 with an estimated rate constant of $\sim 7 \times 10^{-3} s^{-1}$ at a concentration of 100 μM .

DISCUSSION

Substrate Determination. We have demonstrated that Cv1693 from *C. violaceum* catalyzes the efficient hydrolysis of 3',5'-adenosine bisphosphate (pAp) to 5'-AMP and orthophosphate; this reaction is presented in Scheme 1. In addition to pAp, this enzyme will also hydrolyze pCp, pUp, pTp, and pGp with nearly equivalent catalytic efficiency. However, this enzyme will not hydrolyze 2',5'-pAp, and the hydrolysis of the simple dinucleotide, pApA to 5'-AMP, is slower by 2 orders of magnitude. The successful functional annotation of Cv1693

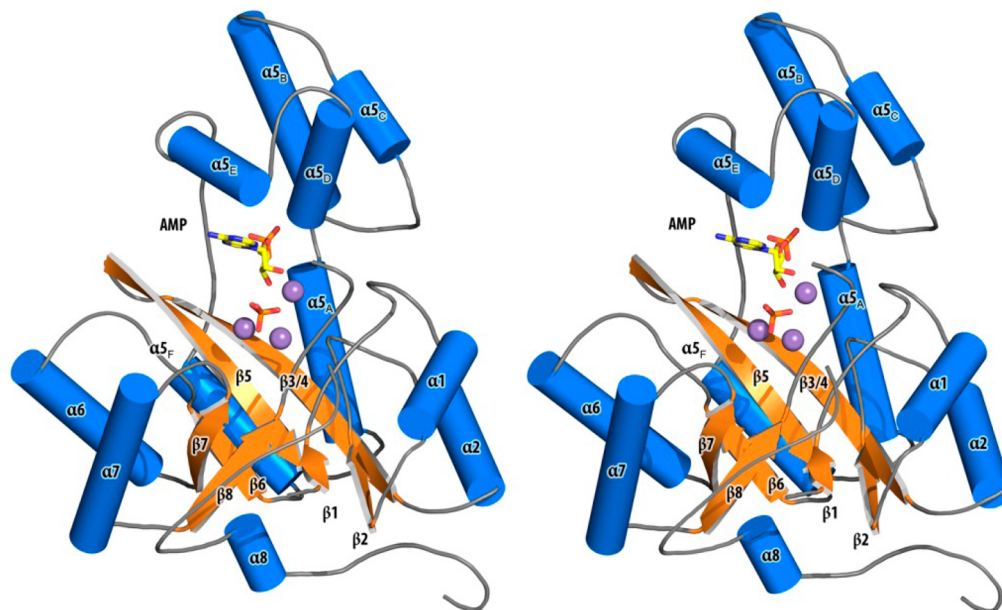


Figure 3. Stereo ribbon diagram of Cv1693. Helices are shown as blue tubes and strands as orange arrows. Three bound manganese ions are shown as maroon spheres, and bound inorganic phosphate and adenosine monophosphate are shown as sticks, colored by atom type. The insertion sequence comprises helices $\alpha 5_A$ and $\alpha 5_B$.

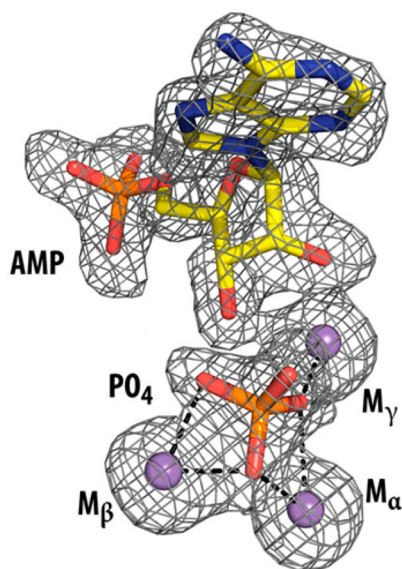


Figure 4. $F_o - F_c$ electron density kick map contoured at 2.5σ . Interactions of the bound inorganic phosphate with the three manganese ions are shown as dashed lines.

began with the determination of the three-dimensional structure of this enzyme and the recognition that another protein of unknown function, Bad1165 from *B. adolscensis*, was structurally similar (PDB entry 3O0F) despite a low level of sequence identity (25%). Because Bad1165 had been successfully crystallized with AMP, a nucleotide cocrystallization scan and an exchange of active site metals eventually led to the successful determination of the structure of Cv1693 with AMP, three manganese ions, and a phosphate ion. The numerous interactions of three arginine residues from the Cv1693 insertion domain with the 5'-phosphate of AMP suggested that the true substrate would bind an AMP-like moiety in this pocket and contribute to substrate recognition.

The initial indication that Cv1693 is a phosphatase originated from the fact that histidinol phosphate phosphatase (HPP) is also a member of the AHS from the structurally similar cog1387. Enzymes within this cog bind three divalent cations in the active site and catalyze the hydrolysis of phosphate monoesters.⁴ In addition, Bad1165 had been shown to catalyze the hydrolysis of *p*-nitrophenyl phosphate, but the hydrolysis of pNp-like substrates has apparently not been investigated.¹² Finally, the proximity of orthophosphate in the active site of Cv1693 to the 3'-OH group of the bound AMP in the ternary complex suggested that this enzyme would likely hydrolyze compounds such as 3',5'-cAMP, 3',5'-pAp, 2',5'-pAp, and 3'-AMP. The utilization of a small substrate library quickly demonstrated that Cv1693 efficiently hydrolyzes 3',5'-pAp to orthophosphate and 5'-AMP with a k_{cat}/K_m value that exceeds $10^6 \text{ M}^{-1} \text{ s}^{-1}$. The enzyme has a rather broad specificity for the identity of the aromatic base, and the 2'-hydroxyl is relatively unimportant for substrate recognition. Additionally, the enzyme can slowly hydrolyze short oligonucleotides with a terminal 5'-phosphate but will not hydrolyze short oligonucleotides without a terminal 5'-phosphate. However, the rate of hydrolysis of the phosphodiester bonds in short oligonucleotides is significantly slower than the rate of hydrolysis of the phosphomonoester bond in pAp.

Structural Comparison to Bad1165. A structure similarity search using the secondary structure similarity matching server,²⁴ with a 70% secondary structure match requirement, returned only one structural homologue: Bad1165 from *B. adolscensis* (rmsd of 1.87 Å over 255 common $C\alpha$ atoms). The three-dimensional structure of Bad1165 was determined in complex with phosphate to 2.4 Å resolution (PDB entry 3E0F) and AMP and phosphate to 1.9 Å resolution (PDB entry 3O0F).¹² On the basis of anomalous data, the metal ions in Bad1165 were modeled as Fe ions for M_α and M_β and Zn for M_γ . All of the protein-derived ligands to the three metals are conserved between Bad1165 and Cv1693, and the positions of the metal ions are very similar despite differences in

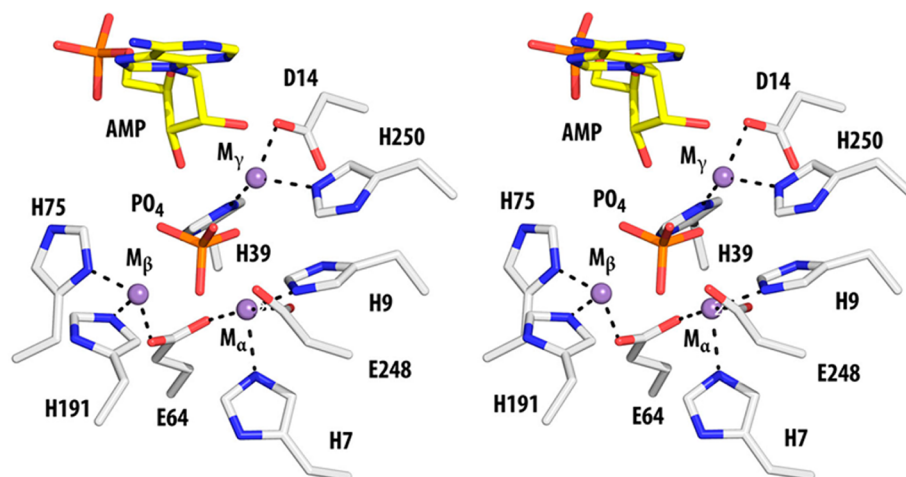


Figure 5. Stereo diagram illustrating the ligands to the three bound manganese ions. Residues colored by atom type with the adenosine monophosphate with yellow carbons and protein atoms with white carbons. Interactions between protein atoms and the manganese atoms are shown as dashed lines.

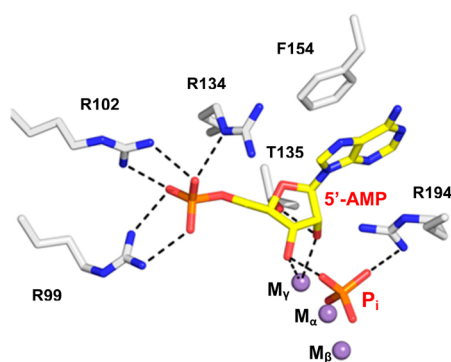


Figure 6. Interactions of AMP with protein residues, the inorganic phosphate, and the manganese ion are shown as dashed lines.

Table 3. Kinetic Constants for Cv1693 with Various Substrates

compound	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1} s^{-1}$)
pAp	22 ± 1	15.5 ± 2.4	$(1.4 \pm 0.2) \times 10^6$
2'-deoxy-pAp	7.1 ± 0.2	10.2 ± 0.9	$(7.0 \pm 0.6) \times 10^5$
pCp ^a	8.6 ± 0.3	14 ± 2	$(6.2 \pm 1.0) \times 10^5$
2'-deoxy-pCp	9.4 ± 0.3	26 ± 3	$(3.6 \pm 0.5) \times 10^5$
pGp ^a	12.3 ± 0.4	17.3 ± 2	$(7.1 \pm 0.8) \times 10^5$
2'-deoxy-pGp	4.2 ± 0.1	5.2 ± 0.4	$(7.9 \pm 0.6) \times 10^5$
pUp ^a	5.4 ± 0.2	8.8 ± 1.0	$(6.1 \pm 0.8) \times 10^5$
2'-deoxy-pUp	5.2 ± 0.2	29 ± 4	$(1.8 \pm 0.2) \times 10^5$
pTp	3.3 ± 0.1	9.3 ± 1.0	$(3.6 \pm 0.4) \times 10^5$
2'-deoxy-pTp	4.3 ± 0.2	22 ± 3	$(1.9 \pm 0.3) \times 10^5$
pApA	0.05 ± 0.01	1.6 ± 0.2	$(3.2 \pm 0.3) \times 10^4$
N ⁶ -methyl-3',5'-pAp	11.6 ± 0.4	5.4 ± 1.0	$(2.1 \pm 0.4) \times 10^6$

^aThese assays were conducted with mixtures of the 2',5'- and 3',5'-nucleotide bisphosphates.

metal content. Two other active site features are strictly conserved between these enzymes: (a) the three arginine residues originating from the insertion domain that coordinate the 5'-phosphate and (b) the phenylalanine side chain, also from the insertion domain, that stacks against the adenine ring. However, despite these similarities, the adenosine moiety takes a slightly different pose in Bad1165. In Cv1693, the ribose is in the 3'-endo conformation and the adenine is in the anti

conformation, while in Bad1165, the ribose is in the 0-4'-endo conformation and the adenine is in the syn conformation. In Cv1693, the 2'- and 3'-OH groups interact with M_γ and the 3'-OH group also interacts with the orthophosphate, while in Bad1165, the 2'- and 3'-OH groups interact with the orthophosphate and not with any of the active site metals. In Bad1165, there is a bridging water molecule between metals M_α and M_β , while in Cv1693, one of the oxygens of the orthophosphate bridges these two metals. In Bad1165, the adenine ring makes hydrogen bonds to the side chains of Arg-161 and Ser-162 from the insertion domain (not conserved in Cv1693), and to the backbone carbonyl of Gly-265 on a flexible loop between β_8 and α_8 . In Cv1693, this loop is disordered, and the adenine, because of its anti conformation, does not approach this loop. Finally, in Cv1693, Thr-135 from the insertion domain makes a hydrogen bond with the 2'-OH group, while in Bad1165, this residue is a proline. Both enzymes have a positively charged residue interacting with the bound orthophosphate; in Bad1165, this is a lysine residue from the flexible loop between β_8 and α_8 , and in Cv1693, it is an arginine residue from the loop between β_6 and α_6 . The structural overlay of Cv1693 and Bad1165 is presented in Figure 7, and the specific interactions of AMP in the active site of Bad1165 are presented in Figure 8.

Bioinformatic Analysis. The apparent physiological substrate for Cv1693 is 3',5'-adenosine bisphosphate. In the cell, the two most common sources of this substrate occur after the transfer of a sulfate from 3'-phosphoadenosine 5'-phosphosulfate (PAPS) and the transfer of the phosphopantetheine group of CoASH to the acyl-carrier protein (ACP).^{25,26} In many bacteria such as *Escherichia coli*, the pAp product is hydrolyzed to 5'-AMP and orthophosphate by an enzyme denoted CysQ.²⁷ The CysQ enzyme from *Es. coli* K12 (locus tag b4214) belongs to cog1218 and hydrolyzes pAp with a catalytic efficiency of $\sim 10^7 M^{-1} s^{-1}$.²⁸ The CysQ from *Mycobacterium tuberculosis*, Rv2131c, has a reported k_{cat}/K_m for the hydrolysis of pAp of $7 \times 10^5 M^{-1} s^{-1}$.²⁵ It has also been reported that NrnA (YtqI) orthologs from *Bacillus subtilis*, *M. tuberculosis*, *Mycobacterium pneumoniae*, and *Streptococcus mutans* can also catalyze the hydrolysis of pAp, in addition to the hydrolysis of short pieces of single-stranded RNA.^{29,30}

Scheme 1

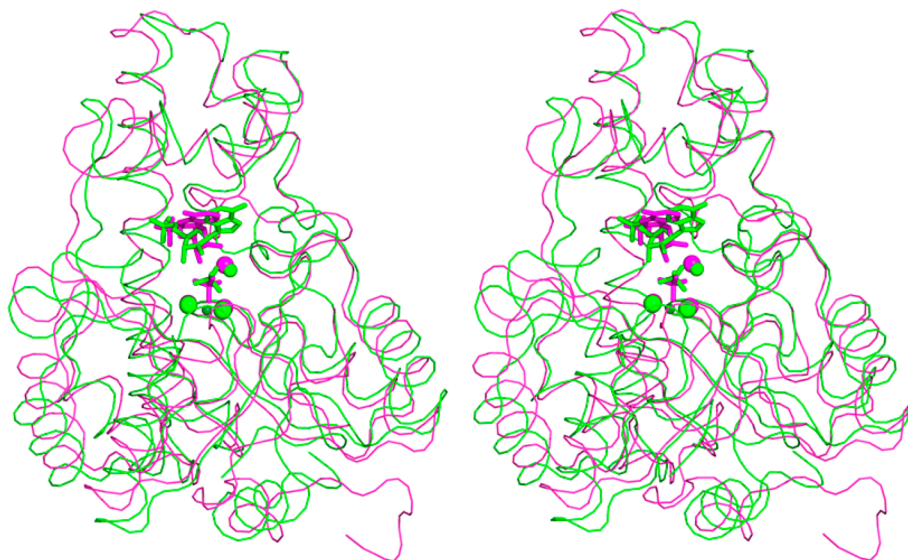
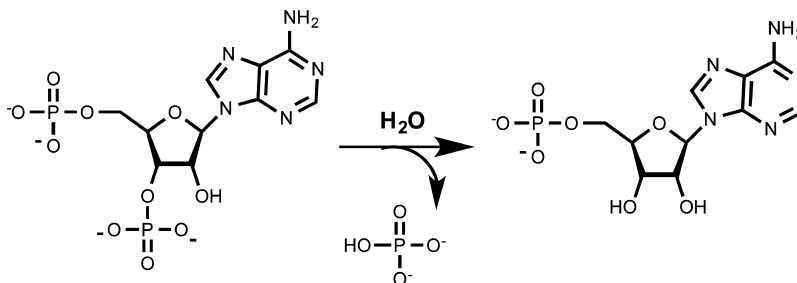


Figure 7. Stereo diagram of the superposition of Cv1693 with Bad1165. The structure of Cv1693 is shown as a purple ribbon and sticks, and the structure of Bad1165 is shown as a green ribbon and sticks.

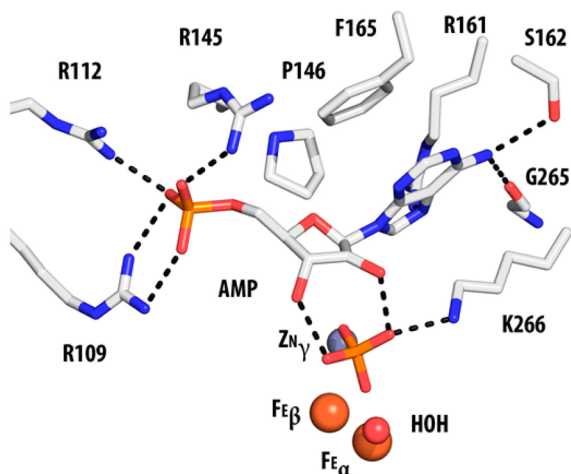


Figure 8. Interactions of AMP with Bad1165. Residues are colored by atom type, and interactions of AMP with Bad1165 are shown as dashed lines. There is a water molecule that bridges Fe_{α} and Fe_{β} .

These bifunctional enzymes (pAp phosphatase and nano RNase) belong to cog0618.

The prevalence of enzymes from different families catalyzing the same function, i.e., pAp hydrolysis, prompted us to determine the distribution of these enzymes among various organisms. Because >6000 bacterial genomes have been completely sequenced, it was more efficient to deal with a

representative subset of the sequenced bacterial genomes. We used the 1752 bacterial genomes available in the Microbesonline database (October 2013) and searched for protein sequences from cog0613, cog1218, and cog0618. This exercise retrieved 964, 830, and 856 protein sequences from three cog's, respectively. A sequence network diagram at a BLAST E value of 1×10^{-70} was constructed for each of these cog's (see Figures S1–S3 of the Supporting Information). Any two sequences connected by lines bear a level of sequence similarity greater than the BLAST E value of 1×10^{-70} , which corresponds to an overall level of sequence identity of $\geq 40\%$. We then identified the proteins that have been biochemically characterized from each of the three cog's and then annotated all protein sequences within the representative groups in the SSN to possess the same enzymatic function as the experimentally characterized enzyme. We were able to provisionally annotate 118 enzymes from cog0613, 309 from cog1218, and 283 from cog0618 as pAp phosphatases in this manner. When we compared the list of organisms to which each of these enzymes belonged, we found that $\sim 6\%$ of the organisms had apparent pAp phosphatases from two of the three cog's discussed here, while the remaining organisms had a pAp phosphatase from just one of the three cog's (see the Venn diagram in Figure S4 of the Supporting Information). There was no instance of an organism possessing pAp phosphatases from all three cog's. Further examination of the phylogenetic classification of each organism revealed that organisms possessing the pAp phosphatase from cog0613 are predom-

Cv1693	1	MAN-----IDL HFHS RTSDGALTPTEVIDRAAARAPALLALTD HD CTGGLAEAAAAAARRGIP-	58
b1266	1	MSDTNYAVI---YDL HSHT TASDGCLTPALVHRAVEMRVGTLAITD HD TAAIAPAREEISRSGLAL	65
Bad1165	1	MSHSVSYAEPQAQGWDI HCH TVFSDGTETPRTLVEQARKLGLHGVAIAD HD TTAGWDEATEASEEIGLP-	68
Elen0235	1	MIE-----DL HVHS TMSDGSDFEQVLEQAARQGVRLAFT ND TTAGLTAARELGERLGVQ-	57
Cv1693	59	-FLNGV EV SVSW--GRHT VH IVGLGIDPAEPALAAGLKSIRE GLER ARQMGASLEAAGIAGCFDGMAR	124
b1266	66	NLIPGV EI STVW--ENHE IH IVGLNIDITHPLMCEFLAQQTERR NQ RAQLIAERLEKAQIPGALEGAQR	132
Bad1165	69	-LLL GT EITAVD--EDV SVH MLAFQYDPSNEHISMFANTRAAR LR RTKRMVERLSQ-DFPITWDDVLA	133
Elen0235	58	-VVG GI EVSAYDFERGR KVH ILGLGVEEGAPALAALCGSTLER R HANSLWQLDRLVEAGYEV DVERALE	125
Cv1693	125	WCDNPE--MIS RT HFARHLV-DSGAVKDMRT VFR KYLTPGKPGYVSHQWASLEDAVGWIVGAGGM AVIA	190
b1266	133	LAQGG---AV TR GHFARFLV-ECGKASSMAD VFK KYLARGKTGYVPPQWCTIEQAIDVIHHS GGK AVLA	197
Bad1165	134	QVKEGERTTIG R PHIADALV-AAGVYETRSDA F ADAVSAKSKYYIPTSPSTHEVIAAVK GAGG VV VAA	201
Elen0235	126	LGRAST--CLY K QHLMAALTSEPYPSAAYRTLYRSL F KNGGICDRIDYVDARDAVRVV EDG GLAVLA	192
Cv1693	191	H PG R YDMGRTLIERLILD-FQAAGGQ G IEVASGSHSLDDMHK FAL HADR HGL YASSGS DFH APGEGGRD	258
b1266	198	H PG R YNLSAKW LK RLVAH-FAEHGDAMEVAQ CC QSPNERTQLAALARQHHLWAS QGS DF H QPCP-WIE	264
Bad1165	202	H AGDPQRNRLLSDEQLDAMIADGLD GL EVHWRGN PE QRERLLTAA RHD LLVT GG S DWH GK GK -PNG	269
Elen0235	193	H PG Q LDS----YDLL PD -LVE CG LGGIER F HPDHTLADHARCAELAVRYRLVCT GG S DYH G K FG R VP H	255
Cv1693	259	VGHTEDLPPICRPIWRELEARILRPAD-----	285
b1266	265	LGRKLWLPAGVEGVWQLWE---QPQNTTEREL	293
Bad1165	270	LGENLTDDDTVREILCRGVDLIGRVSSHAA--	300
Elen0235	256	VGFR--VPA-----	262

Figure 9. Primary sequence alignment of proteins from cog0613: Cv1693, pAp phosphatase from *C. violaceum* (PDB entries 2YB1 and 2YB4); b1266, TrpH from *Es. coli*; Bad1165, enzyme of unknown function from *B. adolescentis* (PDB entries 3E0F and 3O0F); and Elen0235, cyclic phosphate dihydrolase from *E. lenta*. Residues binding the metal cofactors at the active site are colored red. Residues seen interacting with the bound 5'-AMP and inorganic phosphate in the crystal structure of Cv1693 are highlighted in yellow. Thr-135 is highlighted in blue. β -Sheets that constitute the $(\beta/\alpha)_7$ -barrel are highlighted in gray.

inantly betaproteobacteria (~97%), those from cog1218 are largely gammaproteobacteria (~74%), while those from cog0618 are largely bacilli from the phylum Firmicutes (80%) (see pie charts showing the phylogenetic distribution of pAp phosphatases from the three cog's in Figures S5–S7 of the Supporting Information). It is clear that different classes of microorganisms have evolved enzymes to hydrolyze pAp from different families. This is also a strong indication that the physiological function of Cv1693 and its closely related orthologs (colored blue in Figure 1) from cog0613 is the hydrolysis of pAp.

The functional annotation of Cv1693 along with its available crystal structure provides an ideal platform for predicting new enzymatic functions in cog0613 based on the sequence–structure–function correlations. A sequence alignment of Cv1693 from *C. violaceum*, b1266 (an enzyme annotated as TrpH) from *E. coli*, Bad1165 from *B. adolescentis*, and Elen0235 (cyclic phosphate dihydrolase) from *E. lenta* is presented in Figure 9. Cv1693 and Elen0235 are the two known functional annotations from cog0613, and the crystal structures of Cv1693 and Bad1165 are available. TrpH from *E. coli* and its closely related orthologs (colored yellow in Figure 1) constitute the largest cluster of proteins of unknown function in cog0613. It is apparent from the sequence alignment that TrpH possesses all of the residues that bind the three metal ions in the active site, indicating that this enzyme likely catalyzes the hydrolysis of phosphate esters. All residues, except one, in Cv1693 that interact with 5'-AMP and orthophosphate in the crystal structure are conserved in TrpH. The sole exception is that of Thr-135 from the insertion domain in Cv1693, which interacts with the 2'-OH group and 4'-oxygen of 5'-AMP and is

replaced by Gly-142 in TrpH. However, because Cv1693 cannot distinguish between 3',5'-pNp and 2'-deoxy-3',5'-pNp, Thr-135 may not be a critical residue for substrate recognition. This leads us to the hypothesis that the substrate profile for TrpH, if distinct from that of Cv1693, will closely resemble 3',5'-pNp. If distinct, two possible substrates of TrpH may include 2'-substituted 3',5'-bisphosphonucleotides or 2',5'-bisphosphonucleotides, because the relatively small glycine residue can be envisaged to accommodate a larger functional group at the enzyme active site.

Conclusions. We have determined the crystal structure of Cv1693 from cog0613 of the amidohydrolase superfamily with 5'-AMP and orthophosphate bound in the active site. The enzyme was shown to hydrolyze the 3'-phosphate from the substrate 3',5'-pAp, and the substrate profile was shown to include various 3',5'-bisphosphonucleotides. Bioinformatic analysis showed that the vast majority of organisms that possess a closely related ortholog of Cv1693 do not possess an enzyme from cog1218 (CysQ) or cog0618 (YtqI or NrnA) that can catalyze the same reaction. This observation supports the hypothesis that pAp is the physiological substrate of Cv1693.

■ ASSOCIATED CONTENT

● Supporting Information

Sequence network maps for cog0613 (Figure S1), cog0618 (Figure S2), and cog1218 (Figure S3); Venn diagram (Figure S4) showing the number of organisms that possess pAp phosphatase from the three cog's mentioned in this study; pie charts showing the phylogenetic distribution of organisms for each of the three cog's (Figures S5–S7); and a list of enzymes identified as pAp phosphatases from each cog (Tables S1–S3),

along with the organism to which they belong and their gene identifiers. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: raushel@tamu.edu. Phone: (979) 845-3373.

*E-mail: steve.almo@einstein.yu.edu. Phone: (718) 430-2746.

Funding

This work was supported in part by the Robert A. Welch Foundation (A-840) and the National Institutes of Health (GM71790 and GM93342).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Margaret Glasner of the Department of Biochemistry and Biophysics at Texas A&M University for helpful discussions and insights into the bioinformatic analysis of pAp phosphatases. We thank Dr. Andrew W. McMillan from the Glasner lab for retrieving the list of identifiers for the pAp phosphatases presented in the Supporting Information.

REFERENCES

- (1) Holm, L., and Sander, C. (1997) An evolutionary treasure: Unification of a broad set of amidohydrolases related to urease. *Proteins: Struct., Funct., Genet.* 28, 72–82.
- (2) Seibert, C. M., and Raushel, F. M. (2005) Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry* 44, 6383–6391.
- (3) Hobbs, M. E., Malashkevich, V., Williams, H. J., Xu, C., Sauder, M., Burley, S. K., Almo, S. C., and Raushel, F. M. (2012) Structure and catalytic mechanism of LigI: Insight into the amidohydrolase enzymes of cog3618 and lignin degradation. *Biochemistry* 51, 3497–3507.
- (4) Ghodse, S. V., Fedorov, A. A., Fedorov, E. V., Hillerich, B., Seidel, R., Almo, S. C., and Raushel, F. M. (2013) Structural and mechanistic characterization of L-histidinol phosphate phosphatase from the polymerase and histidinol phosphatase family of proteins. *Biochemistry* 52, 1101–1112.
- (5) Aravind, L., and Koonin, E. V. (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res.* 26, 3746–3752.
- (6) Omi, R., Goto, M., Miyahara, I., Manzoku, M., Ebihara, A., and Hirotsu, K. (2007) Crystal structure of monofunctional histidinol phosphate phosphatase from *Thermus thermophilus* HB8. *Biochemistry* 46, 12618–12627.
- (7) Teplyakov, A., Obmolova, G., Khil, P. P., Howard, A. J., Camerini-Otero, R. D., and Gilliland, G. L. (2003) Crystal structure of the *Escherichia coli* YcdX protein reveals a trinuclear zinc active site. *Proteins* 51, 315–318.
- (8) Hagelueken, G., Huang, H., Mainprize, I. L., Whitfield, C., and Naismith, J. H. (2009) Crystal structures of Wzb of *Escherichia coli* and CpsB of *Streptococcus pneumoniae*, representatives of two families of tyrosine phosphatases that regulate capsule assembly. *J. Mol. Biol.* 392, 678–688.
- (9) Kim, H. S., Lee, S. J., Yoon, H. J., An, D. R., Kim, D. J., Kim, S.-J., and Suh, S. W. (2011) Crystal structures of YwqE from *Bacillus subtilis* and CpsB from *Streptococcus pneumoniae*, unique metal-dependent tyrosine phosphatases. *J. Struct. Biol.* 175, 442–450.
- (10) Atkinson, H. J., Morris, J. H., Ferrin, T. E., and Babbitt, P. C. (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One* 4, e4345.
- (11) Apeltsin, L., Morris, J. H., Babbitt, P. C., and Ferrin, T. E. (2011) Improving the quality of protein similarity network clustering algorithms using the network edge weight distribution. *Bioinformatics* 27, 326–333.
- (12) Han, G. W., Ko, J., Farr, C. L., Deller, M. C., Xu, Q., Chiu, H.-J., Miller, M. D., Sefcikova, J., Somarowthu, S., Beuning, P. J., Elsliger, M.-A., Deacon, A. M., Godzik, A., Lesley, S. A., Wilson, I. A., and Ondrechen, M. J. (2011) Crystal structure of a metal-dependent phosphoesterase (YP_9100281.1) from *Bifidobacterium adolescentis*: Computational prediction and experimental validation of phosphoesterase activity. *Proteins: Struct., Funct., Bioinf.* 79, 2146–2160.
- (13) Ghodse, S. V., Cummings, J. A., Williams, H. J., and Raushel, F. M. (2013) *J. Am. Chem. Soc.* 135, 16360–16363.
- (14) Barrio, J. R., Barrio, M. C. G., Leonard, N. J., England, T. E., and Uhlenbeck, O. C. (1978) *Biochemistry* 17, 2077–2081.
- (15) Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* 18, 6069–6074.
- (16) Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif.* 41, 207–234.
- (17) Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: The integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D* 62, 859–866.
- (18) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* 66, 213–221.
- (19) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D* 60, 2126–2132.
- (20) Chen, V. B., Arrendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* 66, 12–21.
- (21) Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., Friedland, G. D., Huang, K. H., Keller, K., Novichkov, P. S., Dubchak, I. L., Alm, E. J., and Arkin, A. P. (2009) MicrobesOnline: An integrated portal for comparative and functional genomics. *Nucleic Acids Res.* 38, D396–D400.
- (22) Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview version 2: A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.
- (23) Smoot, M., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011) Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* 27, 431–432.
- (24) Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* 60, 2256–2268.
- (25) Hatzios, S. K., Iavarone, A. T., and Bertozzi, C. R. (2008) Rv2131c from *Mycobacterium tuberculosis* is a CysQ 3'-phosphoadenosine-5'-phosphatase. *Biochemistry* 47, 5823–5831.
- (26) Mootz, H. D., Finking, R., and Marahiel, M. A. (2001) 4'-Phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. *J. Biol. Chem.* 276, 37289–37298.
- (27) Neuwald, A. F., Krishnan, B. R., Brikun, I., Kulakausas, S., Suziedelis, K., Tomcsanyi, T., Leyh, T. S., and Berg, D. E. (1992) CysQ, a gene needed for cysteine synthesis in *Escherichia coli* K-12 only during aerobic growth. *J. Bacteriol.* 174, 415–425.
- (28) Spielberg, B. D., Xiong, J.-P., Smith, J. J., Gu, R. F., and York, J. D. (1999) Cloning and characterization of a mammalian lithium-sensitive bisphosphate 3'-nucleotidase inhibited by inositol 1,4-bisphosphate. *J. Biol. Chem.* 274, 13619–13628.
- (29) Mechold, U., Fang, G., Ngo, S., Ogryzko, V., and Danchin, A. (2007) YtqI from *Bacillus subtilis* has both oligoribonuclease and pAp-phosphatase activity. *Nucleic Acids Res.* 35, 4552–4561.
- (30) Postic, G., Danchin, A., and Mechold, U. (2012) Characterization of NrnA homologs from *Mycobacterium tuberculosis* and *Mycoplasma pneumoniae*. *RNA* 18, 155–165.