MECHANISTIC CHARACTERIZATION AND FUNCTION DISCOVERY OF PHOSPHOHYDROLASE ENZYMES FROM THE AMIDOHYDROLASE SUPERFAMILY

A Dissertation

by

SWAPNIL VIJAY GHODGE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Frank M. Raushel
Committee Members,	Tadhg P. Begley
	David P. Barondeau
	Margaret E. Glasner
Head of Department,	François Gabbaï

May 2015

Major Subject: Chemistry

ABSTRACT

Rapid advances in genome sequencing technology have created a wide gap between the number of available protein sequences, and reliable knowledge of their respective physiological functions. This work attempted to bridge this gap within the confines cog1387 and cog0613, from the polymerase and histidinol phosphatase (PHP) family of proteins, which is related to the amidohydrolase superfamily (AHS). The adopted approach involved using the mechanistic knowledge of a known enzymatic reaction, and discovering functions of closely related homologs using various tools including bioinformatics and rational library screening.

L-histidinol phosphate phosphatase (HPP) catalyzes the penultimate step in the biosynthesis of the amino acid: L-histidine. Recombinant HPP from *L.lactis* was purified and its metal content and activity were optimized. Mechanistic and structural studies were conducted using pH-Rate profiles, solvent isotope, viscosity effects, site-directed mutagenesis, and X-ray crystallography. These studies, along with extensive bioinformatic analysis, helped determine the boundaries of HPP activity among closely related enzyme sequences within cog1387.

Elen0235 from cog0613 was shown to hydrolyze 5-phosphoribose-1,2-cyclicphosphate (PRcP) to ribose-5-phosphate (R5P) and inorganic phosphate (Pi), with ribose-2,5-bisphosphate as a catalytic intermediate. The mechanism involved sequential hydrolysis of the two P-O bonds of the cyclic phosphodiester, by the attack of nucleophilic hydroxide at the P-center in both steps.

Cv1693 from cog0613 was shown to hydrolyze 3′-phosphate from 3′,5′-diphosphoadenosine (pAp) and other diphosphonucleotides. However, enzymes from at least two other COGs, i.e. 1218 and 0618, are capable of hydrolyzing pAp. The authentic pAp-3′phosphatases from the three COGs under study were determined, and a thorough sequence network analysis of the respective COGs was performed. It was determined that the distribution of pAp-3′-phosphatase from each of the 3 families is largely mutually exclusive in bacteria. This work supported the hypothesis that the physiological role of the target enzyme *in vivo* is the hydrolysis of pAp.

TrpH from *Escherichia coli* was found to hydrolyze pAp efficiently. However, when the substrate profile of this enzyme was probed further, this enzyme was found capable of hydrolyzing oligonucleotides bearing a 5′-phosphoryl functional group with comparable efficiency. Further biochemical characterization revealed that this enzyme catalyzes the hydrolysis of oligoribonucleotides sequentially from 5′→3′.

DEDICATION

I dedicate this work to my family and friends, whose unwaivering faith, and constant support and encouragement, have been the pillars of my strength.

I have been able to pursue and complete this body of work only because of the efforts and sacrifices of my parents, Mr. Vijay Ghodge and Mrs. Leena Ghodge. Their guidance and support have helped me at every stage of life till now, and I am sure, will continue to do so in the future. I am indebted to my grandparents, Mr. Ganesh Nerurkar and the late Mrs. Hema Nerurkar, for instilling the values of hardwork, dedication, and perseverance from an early age. I thank my sister, Ms. Siddhi Ghodge, for being the best friend and sibling I could have ever asked for.

I have been extremely fortunate to have crossed paths with, and befriended, some of the best people one will ever meet. They have been a great source of joy, and a family away from home for the last five years. They have made this journey through graduate school thoroughly enjoyable.

ACKNOWLEDGEMENTS

I am indebted to my advisor, Dr. Frank Raushel, for his constant support, guidance, and encouragement in graduate school. I would like to thank my committee members: Dr. Tadhg Begley, Dr. David Barondeau, and Dr. Margaret Glasner, for their kind suggestions, and availability for discussion.

I am grateful to Dr. Andrew Bigley, Dr. Dao Feng Xiang, and Dr. Siddhesh Kamat for mentoring and guiding me in experimental design and techniques. I am thankful to Dr. Howard Williams for collecting the NMR data presented in Chapter III. I appreciate the work done by our collaborators Dr. Alexander Fedorov, Dr Elena Fedorov, and Dr. Mathew Vetting from the Almo lab (AECOM, Bronx, NY) in solving the crystal structures of two enzymes studied during the course of this work, and Dr. Andrew McMillan from Glasner lab (Texas A & M University) for help with the bioinformatic studies reported in chapter IV.

I am indebted to all the past and present Raushel lab members for their friendship and support. Lastly, I would like to thank all my family and friends, for constituting the support system, which has made this work possible.

NOMENCLATURE

AMP	5′-adenosine monophosphate
BLAST	Basic local alignment search tool
COG	Cluster of orthologous groups
DNA	Deoxyribonucleic acid
E-value	Expectation value
HPP	Histidinol phosphate phosphatase
NCBI	National center for biotechnology information
NMR	Nuclear magnetic resonance
pAp	3′,5′-adenosine bisphosphate
PTE	Phosphotriesterase
PHP	Polymerase and histidinol phosphatase
PSI-BLAST	Position-specific iterated BLAST
PRcP	5-phosphoribosyl 1,2-cyclic phosphate
RNA	Ribonucleic acid

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

The last two decades have seen rapid improvements in genome sequencing technology. Application of these technological advances in a systematic and efficient manner has resulted in frequent additions of new genomes into public sequence repositories, and the number of known gene sequences is increasing in an exponential manner. As of September 2014, the UniProtKB/TrEMBL protein database, contained approximately 84 million gene sequences. The corresponding number was 51 million in January 2014, and 14 million in January 2011. It is not surprising then, that a major fraction of these protein sequences do not have a reliably assigned physiological function. Biochemical determination of the protein functions, including high-throughput techniques, is relatively slow as compared to the rate of sequence deposition. Consequently, computational analysis has been used to predict the molecular function of proteins that have not been or cannot be characterized experimentally. However, an unrestricted and unverified use of such computational annotations has resulted in a high level of misannotations in public sequence databases, largely due to "overprediction" of molecular functions by computational methods.[1] Clearly, it has become indispensible to make experimental and computational efforts at protein functional annotation in a concomitant manner, since using them independently gives either inadequate or incorrect results.

Some of the strategies used for protein functional annotations are[2] : 1. sequence similarity with reference to homologous proteins (both orthologs as well as paralogs), 2. operon/metabolic context by determining the colocalized genes and their respective functions, 3. transcriptional analysis which includes chip-based and RNA-Seq technologies, 4. determination of upstream DNA motifs potentially involved in the regulation of transcription, 5. identification of coupled activities in a pathway by studying the functions and allostery of multi-domain proteins, 6. protein-protein interaction studies, and 7. phenotypes of gene deletion/ knockout mutants. Other methods used for enzyme function discovery include: 1. Assigning functions to orphan enzymes.[3] These refer to the large pool of enzyme activities that do not have an assigned protein sequence, thus providing a sound starting point to trace a known biological reaction to a protein sequence. 2. Activity-based protein profiling (ABPP).[4] This method involves the use of chemical probes that are designed to react with mechanistically related classes of enzymes. The chemical probe typically consists of a reactive group or 'warhead' that reacts irreversibly with the enzyme active site and a tag such as a fluorophore or an affinity tag. ABPP technology has proven useful as a direct tool in identifying enzymatic activities from a complex pool of proteome.

It is clear that there are several tools available to address the challenge of annotating protein functions. Careful selection and application of one or more of these tools is required for enzyme function discovery based on the target protein sequences. It is also critical to select the target pool of protein sequences on the basis of a robust system of classification to channelize efforts in a systematic manner. Accordingly, this

work has focused on discovering previously unknown enzyme functions of putative phosphoesterases within the amidohydrolase superfamily of enzymes by leveraging the mechanistic knowledge gained by studying a previously known homologous phosphatase enzyme.

The amidohydrolase superfamily (AHS) was first identified by Holm and Sanders in 1997 based on the significant resemblance between the structural features of phosphotriesterase (PTE) (PDB: 1PTA), urease (URE) (PDB: 1KRC) and adenosine deaminase (ADE) (PDB: 2ADA).[5] Subsequent studies of members of AHS have revealed fantastic insights into the reactions, the mechanisms and the breadth of substrate specificity within this superfamily.[6] Enzymes belonging to the AHS are ubiquitously found in all forms of life. Enzymes from this superfamily utilize 0-3 divalent d-block metal ions at their active sites as cofactors. These metal ion cofactors have been implicated in the activation of the nucleophilic water/hydroxide as well as the activation of the scissile bond for cleavage. The most commonly found metal ion among the characterized amidohydrolase enzymes is zinc. Other metal ions include divalent iron, manganese and nickel. Divalent cobalt metal ion can be substituted for zinc *in vitro*, with comparable or higher catalytic constants for the enzymatic reactions. The enzyme active site is located at the mouth of the beta-barrel in an overall distorted TIM-barrel or $(\beta/\alpha)_8$-barrel fold, in which eight parallel β-strands are flanked by eight α-helices on the outer faces. In general, the metal-binding residues are located at the end of β-strands while the loops that link the β-strands and α-helices bear the residues responsible for substrate specificity. Members of the AHS catalyze hydrolysis of amide,

ester, halogen or other functional groups at carbon and phosphorus centers. These include the cleavage of C-N, C-O, C-Cl, C-S, and O-P bonds. Besides, enzymes from the AHS are also known to catalyze non-hydrolytic C-C bond cleavage. Most of the characterized members of the amidohydrolase superfamily are important enzymes for histidine and tryptophan metabolism, de novo biosynthesis of purine and pyrimidine nucleotides, and biodegradation of agricultural and industrial materials including rubber chemicals, herbicides, leather, paper and others.[7]

The discovery of physiological substrates of enzymes within such a vast and diverse superfamily is a challenging proposition. In order to address this challenge, we will restrict ourselves to bacterial protein sequences, for two major reasons: 1. relative ease of obtaining bacterial proteins using recombinant overexpression in the robust *Escherichia coli* system, and 2. higher probability of colocalization of genes belonging to the same metabolic pathway in simpler bacterial genomes, which may provide a framework to select a library of potential substrates. As of 2011, the AHS consisted of over 12,000 non-redundant protein sequences from over 1200 fully sequenced bacterial genomes. These sequences have been classified into about 24 Clusters of Orthologous Groups (COG) of proteins. If a group of at least three proteins from phylogenetically distant organisms share a higher sequence similarity to each other as compared to any other protein from those organisms, then they are classified in the same COG. Thus, the COG database is an attempt to classify orthologous proteins found in fully sequenced genomes. In practical terms, enzyme sequences grouped in any given COG are likely to catalyze the same or a very similar chemical reaction. Thus, knowing even a single

enzymatic reaction from a given COG can provide a hint of the type of reaction that the other members of the COG, with no known function, will catalyze. Using the COG database, the diverse AHS can be conveniently sub-divided into several different clusters, each consisting of enzyme sequences that catalyze a similar chemical reaction.

This work has focused on the mechanistic investigation and function discovery of enzymes that were predicted to hydrolyze phosphoesters. The phosphometabolites constitute a very important part of the metabolome of any living organism. Consequently, a good understanding of the enzymes that metabolize these phosphoesters with respect to their mechanism and substrate specificity is crucial. RNA/DNA repair and synthesis, signal transduction, phospholipid metabolism, energy storage, transcription control, protein activation/deactivation, cell transformation, virulence, and several other biological phenomena rely on phosphoesterases.[8] Besides, several synthetic compounds including pesticides, insecticides, nerve agents, plasticizers, etc. have organophosphate functional group, and engineered phosphatases offer a convenient bio-processing solution to degrade these compounds under ambient and eco-friendly conditions. Clearly, the study of phosphatases within biological as well as biotechnological context is indispensible.

The classification of phosphatases has been done in several different ways historically. A simple classification suggested by Van Etten in 1982 is described here.[9] The phosphomonoesterases are divided into families based on their reaction mechanism in general, and the active site catalytic group in particular. The four broad families thus defined were histidine phosphatases, cysteine phosphatases, serine phosphatases, and

metallophosphatases. The histidine, cysteine, and serine phosphatases function by nucleophilic catalysis by the respective enzyme side chain residue, resulting in the formation of phosphohistidine, phosphocysteine, and phosphoserine intermediate respectively. The metallophosphatases utilize divalent metal ions to activate the nucleophilic water/hydroxide for a direct attack at the P-center of the phosphate being cleaved. Phosphohydrolases from the AHS utilize divalent metal ion cofactors for hydrolytic chemistry, and therefore, belong to the metallophosphatase subdivision among phosphatase enzymes.

Within the AHS, the COGs that have been known or predicted to consist of phosphoesterases are:

1. cog1735: PTE-homology proteins.

2. cog0084, cog1099, and cog1831: TatD-related DNases.

3. cog4464: Protein tyrosine phosphatases from the polymerase and histidinol phosphatase (PHP) family.

4. cog1387 and cog0613: Phosphoesterases from the PHP family.

Available details and description of each of the above COGs is presented below.

COG1735

The first enzyme function discovered from cog1735 was that of phosphotriesterase (PTE) from the organism *Pseudomonas diminuta*. Subsequent efforts to discover functions of other members of this COG have revealed that this group consists of enzyme sequences that catalyze the hydrolysis of lactones. **Figure 1.1** shows

the sequence similarity network of enzyme sequences from cog1735, while **Table 1.1** lists the biochemically characterized enzyme sequences from this COG. The presence of PTE in a cluster of homologous lactonases has provided a hint to the scientific community that this activity may have evolved from an original lactonase enzyme sequence. With the evolution of a phosphatase activity within a group of homologous lactonases, cog1735 provides an ideal platform to study the evolution of novel enzyme activities in nature.

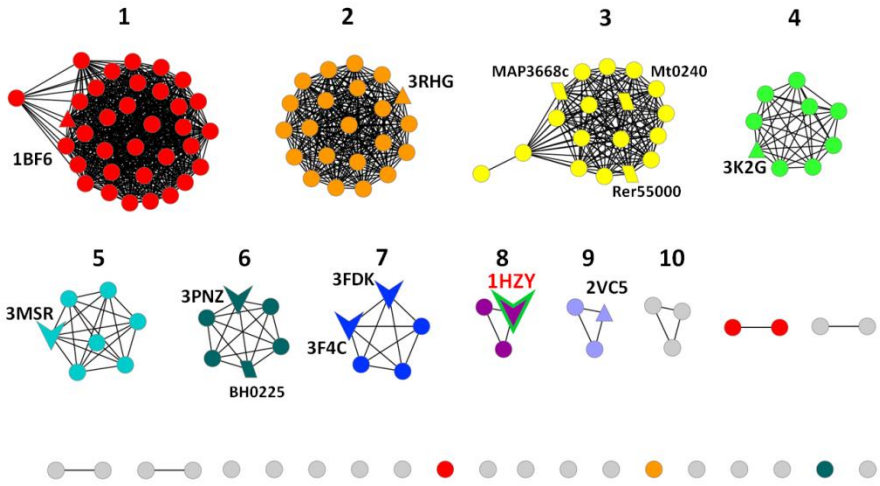


Figure 1.1. Sequence similarity network of cog1735 at a BLAST E-value of $1x10^{-80}$. The clusters were given their respective colors using an E-value of 10^{-70}, but the arbitrary numbers have be assigned to each cluster at an E-value of 10^{-80}. Each node represents a protein sequence and the line connecting two nodes (edge) represents the pairwise connection between the two sequences at a BLAST E-value of better than 10^{-80}. The individual groups of nodes are colored arbitrarily. Protein sequences whose crystal structures are available in the protein data bank, but the physiological substrates are unknown are shown as a triangle (Δ). Protein sequences whose substrates are known but the structure is unavailable are shown as parallelogram (▱). Protein sequences whose both structure and function have been characterized are shown as vee-shapes (⏛). Further details are given in **Table 1.1**.

Table 1.1. List summarizing the known information on protein sequences from cog1735.[107]

Cluster number*	Protein locus tag/name	PDB code	Function	Best substrate	k_{cat}/K_m ($M^{-1}s^{-1}$)
1	PHP**	1BF6	Unknown	n/a	n/a
2	Pmi1525	3RHG	Unknown	n/a	n/a
3	Mt0240	n/a	Quorum-quenching lactonase	N-butyryl-DL-homoserine lactone	3.0×10^4
3	Rer55000	n/a	Quorum-quenching lactonase	N-butyryl-DL-homoserine lactone	1.5×10^5
3	MAP3668c	n/a	Quorum-quenching lactonase	N-dodecanoyl-DL-homoserine lactone	1.3×10^4
4	-	3K2G	Unknown	n/a	n/a
5	Ms0025	3MSR, 3OVG	Phosphorylated sugar lactonase	D-xylono-1,4-lactone-5-phosphate	2.2×10^4
5	Mag6390	n/a	Phosphorylated sugar lactonase	D-xylono-1,4-lactone-5-phosphate	1.3×10^4
6	Lmo2620	3PNZ	Phosphorylated sugar lactonase	D-lyxono-1,4-lactone-5-phosphate	6.1×10^4
6	BH0225	n/a	Phosphorylated sugar lactonase	D-lyxono-1,4-lactone-5-phosphate	1.0×10^5
7	GsP	3F4C	Lactonase	δ-undecanoic lactone	1.5×10^4
7	Dr0930	3FDK	Lactonase	δ-nonanoic lactone	1.6×10^6
8	PTE	1HZY	Phosphotriesterase	Paraoxon	5.5×10^7
9	Sso2522	2VC5	Quorum-quenching lactonase	Thiobutyl-γ-butyrolactone	3.6×10^5

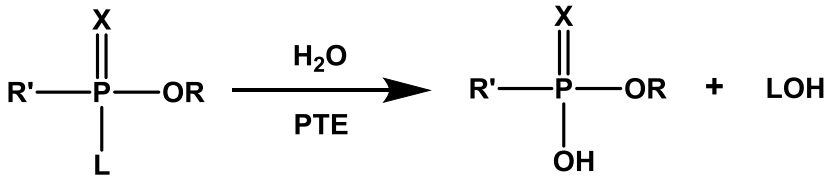
*Cluster numbers from **Figure 1.1**.

**PTE Homology Protein

A recent literature review by Bigley and Raushel provides a good compilation for understanding the studies conducted on enzymes with PTE activity.[10] Phosphotriesterase activity was first discovered in the soil bacteria *Pseudomonas diminuta* and *Flavobacterium sp.*, specimen of which were collected from soil contaminated with the organophosphate insecticide parathion. Other soil bacteria that have been found to possess a PTE ortholog from the AHS are *Agrobacterium radiobacter* (OpdA, PDB: 2R1P) and *Chryseobacterium balustinum*. Besides AHS, PTE activity has evolved among several other enzyme superfamilies, and phosphotriesterases have been identified in several bacteria, higher marine organisms as well as mammals.[11]

PTE catalyzes the hydrolysis of organophosphates as shown in **Scheme 1.1**. These organophosphate compounds are mainly comprised of insecticides and nerve agents belonging to the phosphotriester, thiophosphotriester, and phosphorothioester classes (**Figure 1.2**). The leaving group in organophosphate insecticides generally is a phenol or a thiol. It may be straight-chained, branched or aromatic. The leaving group among nerve agents may be a fluoride group (G-type nerve agents) or a branched thiol group (V-type nerve agents). The insecticide organophosphates possess an achiral phosphorus center with at least two identical ester groups, while the nerve agents possess a chiral phosphorus center by virtue of a P-CH_3 linkage. Clearly, PTE has a very broad substrate range and several studies on the structural, mechanistic and evolutionary aspects of this enzyme activity have added to our knowledge and understanding of enzyme catalysis. The best substrate for PTE is paraoxon and the catalytic rate constant

for its hydrolysis is comparable to the limits of diffusion of a substrate. In general, the best substrates for this enzyme contain electron-withdrawing phenolic leaving groups.



Scheme 1.1. General reaction catalyzed by PTE. X can be O or S atom, R is any alkyl group, R' is any alkyl or alkoxy group, and L is the leaving group, which may be an alcohol, thiol, phenol or halide.

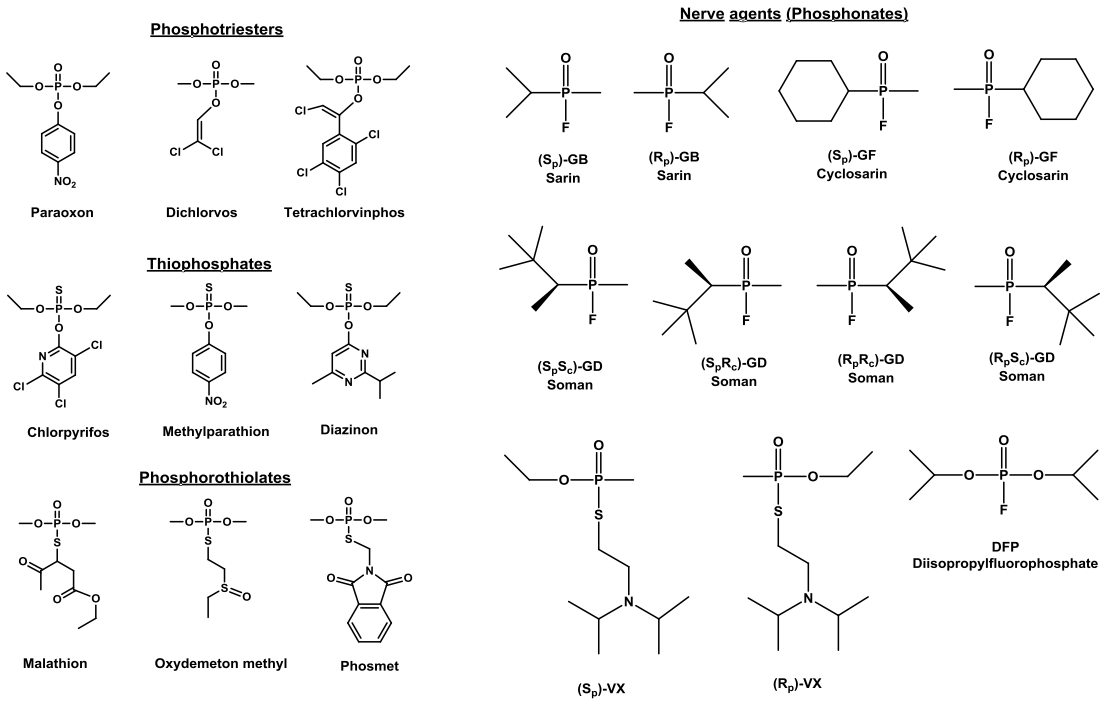


Figure 1.2. Examples of organophosphate compounds hydrolyzed by PTE.

PTE is one of the original enzyme structures that were used in the identification of the AHS. Several crystal structures of PTE and closely related orthologs have been solved. They reveal a distorted $(\beta/\alpha)_8$- or TIM-barrel fold with the active site located at the C-terminal end of the TIM-barrel (**Figure 1.3**). The active site consists of binuclear metal center consisting of divalent d-block metal ions, which are coordinated enzyme residues strictly conserved among binuclear AHS members in general, and PTE homologs in particular (**Figure 1.4**). The more buried α-metal ion is coordinated by the two histidines from β-strand **1** that constitute the HxH motif, and an aspartate from β-strand **8**. The more solvent-exposed β-metal ion is coordinated by two histidines residues, one each from β-strand **5** and β-strand **6** respectively. A hydroxide and a carboxylated lysine from β-strand **4** act as bridging ligands between the two active site metal ions. Besides, a water molecule has been observed to coordinate the β-metal ion in the absence of a substrate/inhibitor, and is displaced when the latter binds. Both α- and β-metal centers possess a trigonal bipyramidal geometrical disposition of ligands, which is retained on substrate/inhibitor binding. The native metal found in PTE is Zn^{2+}. In addition, the enzyme is active when the active site Zn^{2+} ions are replaced with Cd^{2+}, Mn^{2+}, Ni^{2+} or Co^{2+}.

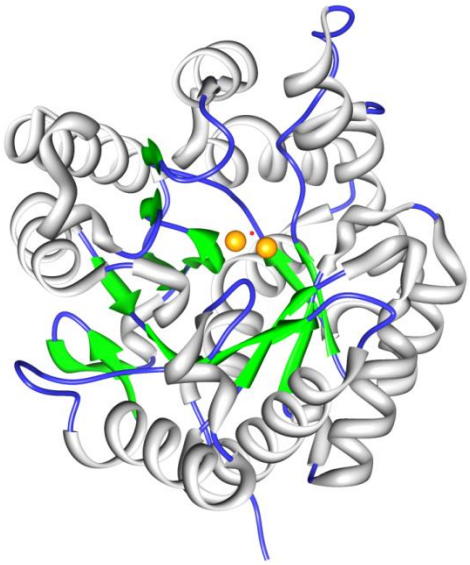


Figure 1.3. Ribbon diagram of PTE from *P.diminuta* (PDB: 1HZY). The β-strands are colored green, the α-helices are colored gray while the loops connecting a β-strand to the subsequent α-helix is shown in blue. The metal ions constituting the binuclear metal center are shown as orange spheres while the nuclephilic bridging hydroxide at the active site is shown as a red sphere.

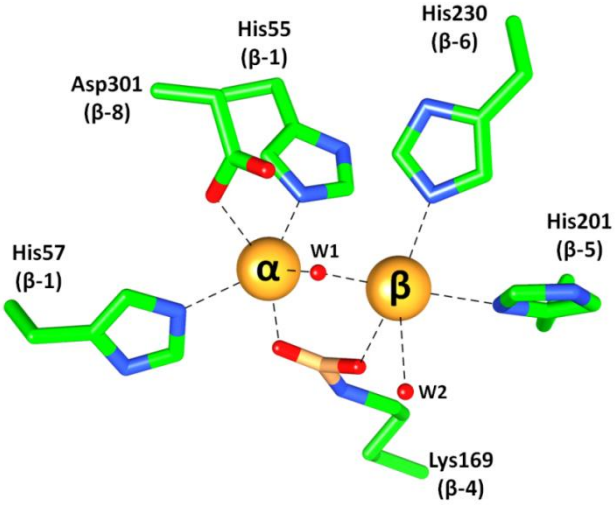


Figure 1.4. Active site structure of PTE from *P.diminuta* (1HZY). The metal ions bound at the binuclear active site (Zn^{2+} in this case) are indicated as orange spheres. The enzyme residues are shown in green while relevant water/hydroxide molecule/ion are shown in red. W1 is the nucleophilic hydroxide in the enzyme-catalyzed reaction while W2 gets displaced by the substrate.

The catalytic mechanism of PTE has been extensively probed using biochemical and physical methods. The enzyme active site consists of a dinuclear metal center with a bridging hydroxide. EPR spectroscopy of the binuclear Mn^{2+}-center of PTE showed hyperfine splitting of the Mn^{2+} signal indicating a coupling between the two metal centers through a bridging ligand.[12] pH-dependence study indicated that this coupling was lost at mildly acidic pH, while the kinetic pH rate profiles indicated that deprotonation of a residue with an apparent pKa of ~6 was essential for enzyme activity.[12-13] Besides, circular dichroism was used to show that the ligation state of the α-metal remains unchanged as trigonal bipyramidal between its resting and catalytically active state.[14] All the above evidence indicated that the bridging hydroxide was the nucleophile in the reaction catalyzed by PTE. Using ^{18}O-labelled water, it was shown that the nucleophilic attack of the bridging hydroxide was directed at the phosphorus center, and not the leaving group. Hydrolysis of a chiral substrate (O-ethyl-O-(4-nitrophenyl)-phenylphosphonothionate) by PTE produced a chiral product with an inversion of stereochemistry at the P-center, indicating that the enzymatic hydrolysis by PTE occurs by a direct attack of the bridging hydroxide at the P-center of the substrate.[15]

The direct interaction of the substrate with the metal center of PTE has been demonstrated using X-ray crystallography and verified using EPR spectroscopy, Brønsted analysis, and kinetic isotope effects. EPR studies on Mn^{2+}-PTE before and after addition of substrate analogs showed a significant shift in the g-value from g=2.2 to g=4.3 respectively.[16] Brønsted analysis was conducted to probe the transition state of the enzymatic reaction.[17] It was observed that substrates containing leaving groups with

pK_a > 8 gave a linear relationship indicating that the chemistry is fully rate limiting with these substrates. Conversely, product dissociation becomes the rate limiting step for substrates containing leaving groups with pK_a in the range of 4-7. The relatively large β_{LG} value of -2.2 indicated a late transition state. The transition state also depended on the relative hard or soft nature of the metal at the active site and whether the substrate was an oxo (hard ligand) or thiol (soft ligand) phosphate. Soft or hard metal chelates soft or hard ligands respectively, more strongly, and causes a shift to an earlier transition state, and vice versa. These results have further been verified using primary and secondary kinetic isotope effects on the hydrolysis of paraoxon and O,O-diethyl-O-(4-carbamoylphenyl) phosphate.[18] These studies also indicated an associative mechanism with a late transition state involving a pentavalent phosphorus center.

The dinuclear metal center of PTE plays a central role in the enzymatic catalysis. Point mutation of any amino acid residue ligating to either of the active site metal ions results in significant reduction of activity. The dinuclear metal center may potentially play three roles in the enzymatic hydrolysis of organophosphates. The metals may decrease the pK_a of the bound water molecule and increase the nucleophilic character of the attacking hydroxide. The metal ions may increase the polarization of the P=O bond, and thereby increase the electrophilic character of the phosphorus center. The β-metal ion has been implicated in this role based on crystallographic evidence. The metal ions may also neutralize the development of negative charge on the leaving group.

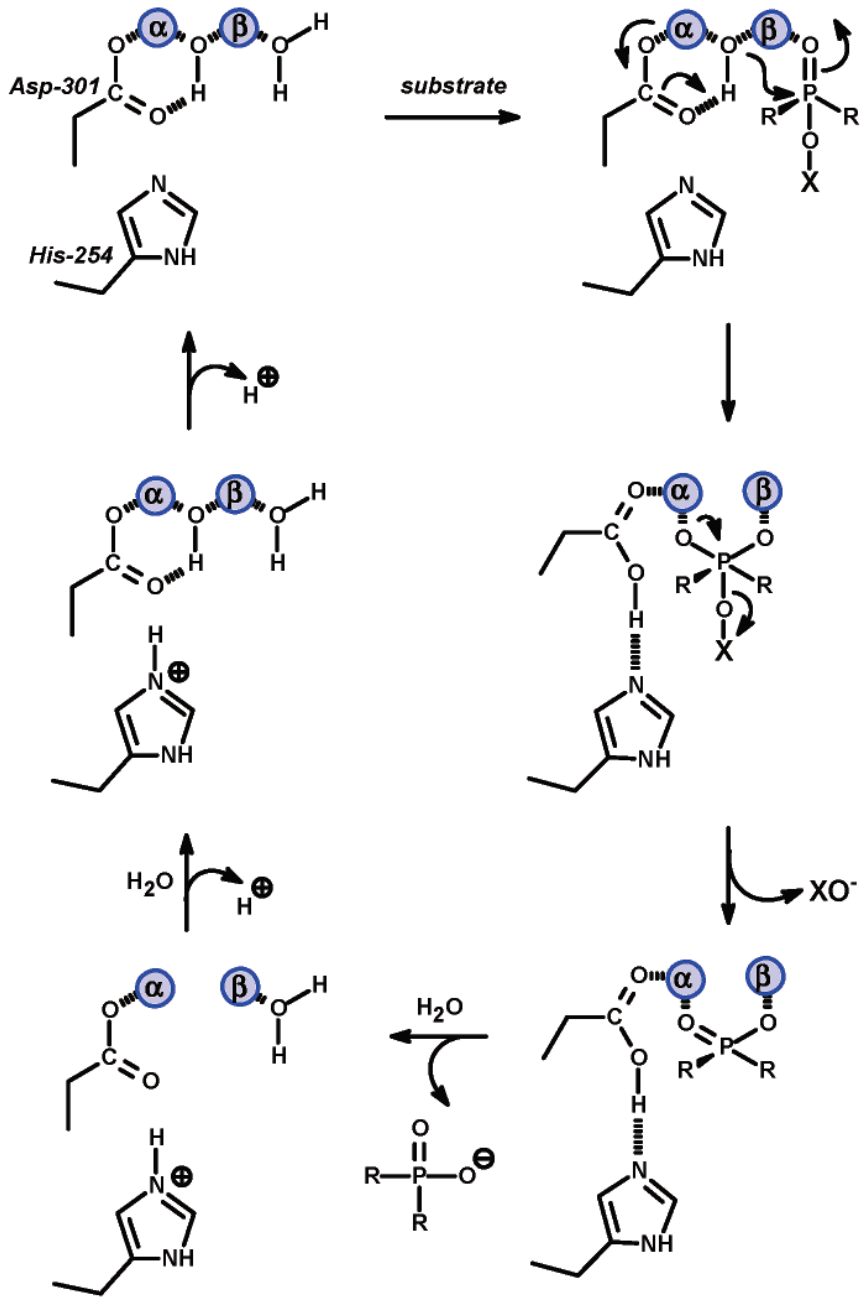
Crystal structures have demonstrated that the aspartate residue from β-strand **8**

which ligates to the α-metal ion (D301 in PTE from *P. diminuta*) also hydrogen bonds with the bridging hydroxide. Therefore, D301 has been implicated as the species that deprotonates the hydroxyl nucleophile during catalysis. Since D301 is coordinated to the α-metal ion, it has been difficult to gather clear evidence for this role. Mutation of another active site aspartate, D233, slows down the reaction of a fast substrate like paraoxon but speeds up the hydrolysis of p-chlorophenylphosphate, which has a poor leaving group. It is proposed that D233 and H254 play a role in shuttling the proton from D301 through the solvent. When the proton shuttle is disturbed, the fast reactions get slowed down but the trapped proton is able to protonate poor leaving groups, speeding up the relatively slow reactions.

Computational tools have been utilized to probe the mechanism of hydrolysis of paraoxon by PTE. A high level density functional theory (DFT) simulation involving only the bridging hydroxide, metals and metal ligands predicted a two-transition state system with a high energy intermediate.[19] The first transition state is predicted to involve substrate binding and distortion so as to bring the P-center in the proximity of the hydroxide. The high energy intermediate is predicted to be pentavalent species. In the second transition state, the hydroxide proton is transferred to D301 and the leaving group is cleaved from the P-center. The overall energy of activation calculation using DFT simulations compares favorably with the value determined experimentally. Other studies have utilized quantum mechanical/molecular mechanical (QM/MM) simulations which include the entire protein and the active site waters as well as a water shell. The first study predicted a single transition state but found a transition state plateau in the

reaction coordinate similar in nature to the intermediate state of the DFT study. Another QM/MM study predicted a series of discrete transition steps and intermediates that describe all steps of the reaction from substrate binding to final product release, using reoptimized parameters. The most interesting prediction of this study was that the highest energy barrier along the reaction pathway was for the distortion of the phosphorus center, similar to the prediction using DFT simulations.

Based on the experimental evidence described above, the catalytic mechanism for the hydrolysis of organophosphates by PTE is shown in **scheme 1.2**.[20] The proposed dissociation of the hydroxide from the β-metal ion during the course of the enzymatic reaction is supported by EPR and structural data.



Scheme 1.2. Mechanism of organophosphate hydrolysis catalyzed by PTE.[20] [Reprinted with permission from Aubert, S.D., Li, Y., and Raushel, F.M. Mechanism for the hydrolysis of organophosphates by bacterial phosphotriesterase. *Biochemistry* **2004**, *43*, 5707-5715. Copyright 2004 American Chemical Society.]

PTE hydrolyzes a broad range of organophosphate substrates. The broad specificity of this enzyme has been attributed to the non-specific interactions within the enzyme active site. The three ester groups of the substrate interact with three hydrophobic pockets on the surface of the enzyme.[21] In *P. diminuta* PTE, the leaving group pocket is made up of W131, F132, F306, and Y309. The two other ester groups bind to the large and small pockets respectively. The large pocket is made up of H254, H257, L271, and M317, while the small pocket is made up of G60, I106, L303, and S308.[22] The latter two pockets determine the stereochemical preference of the enzyme for substrates with chiral P-center, and the stereochemical selectivity of PTE can be enhanced, diminished, or reversed by modulation of the large and small pockets. The ability to manipulate the enantiomeric selectivity has made PTE a valuable tool for the isolation of enantiomerically pure compounds and selective degradation of enantiomers.[23]

COG0084, COG1099, AND COG1831

These three COGs are very closely related to each other. Cog0084 is the largest with 1259 non-redundant protein sequences while cog1099 and cog1831 are relatively small with 36 and 34 sequences respectively. These COGs have been annotated as magnesium-dependent DNases, in general. However, there is no unambiguous biochemical characterization of an enzyme function from any of these three COGs. Structures of three protein sequences from cog0084 are available in the protein data bank

(PDB: 1XWY, 1YIX, and 1ZZM). The sequence similarity networks of each of these COGs are shown in figures 1.5, 1.6, and 1.7 respectively.

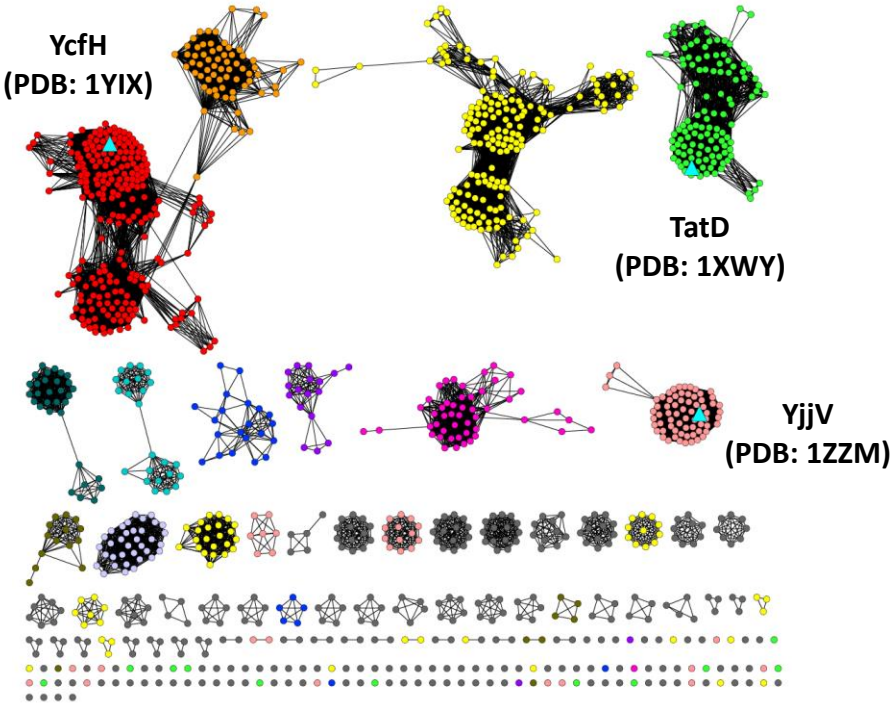


Figure 1.5. Sequence similarity network of cog0084 at a BLAST E-value of 1 x 10^{-70}. Enzyme sequences whose structures are available in the protein data bank are shown as triangles (△). The individual groups have been given arbitrary colors.

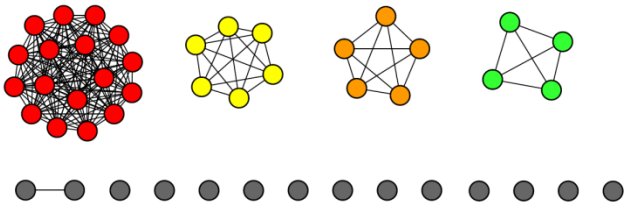


Figure 1.6. Sequence similarity network of cog1099 at a BLAST E-value of 1 x 10^{-70}. None of the proteins have been characterized from this COG. The individual groups have been given arbitrary colors.

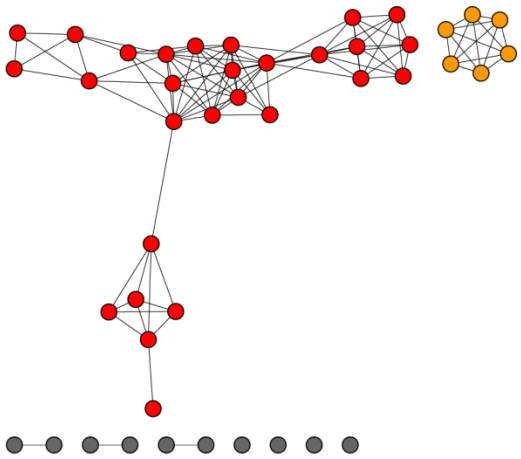


Figure 1.7. Sequence similarity network of cog1831 at a BLAST E-value of 1 x 10^{-70}. None of the proteins from this COG has been biochemically characterized. The individual groups have been given arbitrary colors.

TatD from cog0084 may have an important physiological role as its orthologs are found in several species of bacteria, fungi, plants and animals. TatD is encoded in the

Tat operon that encodes the Tat (Twin-Arginine Translocation) proteins TatA, TatB, and TatC. All three are membrane proteins that bind and export folded proteins with twin arginine-containing leader peptide from cytoplasm to periplasm. Conversely, TatD was found to be a cytosolic protein. Different studies have showed that the expression of TatD, and that of its homologs YcfH and YjjV, is not essential for the functioning of the Tat pathway in *E. coli*.[24] Other studies showed that TatD bears 3′-5′ exonuclease activity and participates in DNA fragmentation during apoptosis in *Saccharomyces cerevisiae* and *Trypanosoma brucei*.[25] Several crystal structures of TatD are available in the protein data bank, but without any biochemical evidence of its function. These include human TatD1 (PDB: 2XIO) and TatD3 (PDB: 2YIH), TatD from *S. cerevisiae* (PDB: 3E2V), as well as the TatD orthologs from *E. coli* (PDB: 1XWY (TatD), 1YIX (YcfH), and 1ZZM (YjjV)). All structures show that TatD orthologs possess the TIM-barrel fold.

Chen *et al* recently solved the structure of *E. coli* TatD with a bound fragment of DNA and demonstrated weak 3′,5′-exonuclease activity from TatD from *E. coli* (cog0084) (PDB: 4P5U, 4PE8).[26] They found that TatD was able to hydrolyze DNA in the presence of Mg^{2+} and Mn^{2+}, but not in the presence of Ca^{2+} and Zn^{2+}, or in the absence of metal ions. However, the concentrations of the metal ions used were 2 mM for Mg^{2+} and Mn^{2+}, 1 mM for Ca^{2+} and only 0.1 mM for Zn^{2+}, for unexplained reasons. The enzyme was able to fully digest single strands (~20 nt in length) of DNA as well as RNA, but not duplex strands. There were no kinetic constants or enzymatic reaction rates reported in this study. Based on the partial sensitivity of TatD-knockout strains to

high doses of H_2O_2, the paper claims that TatD may function as an exonuclease in the base excision repair process for H_2O_2-induced DNA damage. The crystal structure reported in the same study has a tri-deoxynucleotide bound at the active site. However, due to identical sequence of the first and last three nucleotides in the 8-mer DNA fragment used during crystallography, it is unclear if the tri-deoxynucleotide at the active site is the 3′-end of the unhydrolyzed or the 5′-end of the hydrolyzed original 8-mer DNA. Besides, this crystal structure does not have any metal ions bound at the active site, and therefore, may represent the inactive state of the enzyme. Based on their comparison of previously solved structures of TatD, it is apparent that the enzyme active site can bind 2-3 metal ions. It also revealed five strictly conserved potential metal binding residues (E91, H127, H152, E201, and D203).

Clearly, there is a need to gather unambiguous evidence to annotate TatD as a 3′,5′-exonuclease in a physiological context. Besides, the mechanistic details for the binding of the substrate and its hydrolysis need to be elucidated. The available structural data and the knowledge of its ability to hydrolyze fragments of DNA are a good starting point.

COG4464

This COG consists of protein tyrosine phosphatases (PTPs) from the polymerase and histidinol phosphatase (PHP) family of proteins. These enzymes are mainly found in Gram-positive bacteria while the PTPs from Gram-negative bacteria belong to the distinct Cys-based low-molecular-weight phosphotyrosine protein phosphatase

(LMPTP) family. Enzymes from cog4464 have been shown to be capable of dephosphorylating their cognate autophosphorylated protein tyrosine kinase (PTK) and the substrates of PTK. These catalytic activities enable the PTPs from cog4464 to regulate capsular polysaccharide biosynthesis (CpsB from *Streptococcus pneumoniae*) as well as polysaccharide biosynthesis, bacterial DNA metabolism, and DNA damage response (YwqE from *Bacillus subtilis*) in the respective organism.

The sequence similarity network for cog4464 is shown in **Figure 1.8**. Characterization of two enzymes has been reported in literature: YwqE from *B. subtilis* and CpsB from *S. pneumoniae*. YwqE was shown to dephosphorylate YwqD, a soluble protein kinase, and the various protein substrates phosphorylated by YwqD. One of these substrates is YwqF, a UDP-glucose dehydrogenase, which catalyzes the formation of precursors for the formation of acidic polysaccharides in *B. subtilis.*[27] Besides, YwqD and YwqE also modulate the activity of single-stranded DNA-binding proteins (SSB), which are involved in DNA metabolism and DNA damage response. CpsB was shown to dephosphorylate CpsD, its cognate protein kinase.[28] CpsCD regulate the biosynthesis of capsular biosynthesis, and consequently virulence.YwqE and CpsB share a sequence identity of 30%. Both enzymes have been shown to be metal-ion dependent phosphatases. The phosphatase activity of CpsB is enhanced in the presence of Mg^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, and Zn^{2+}, while that of YwqE in the presence of Mn^{2+}, Cu^{2+}, and Zn^{2+}, while Mg^{2+} and Ca^{2+} had no significant effect on its activity.

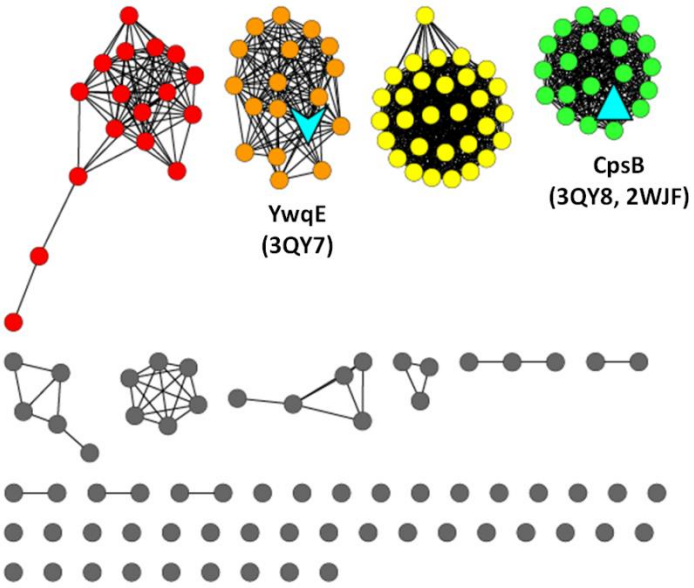
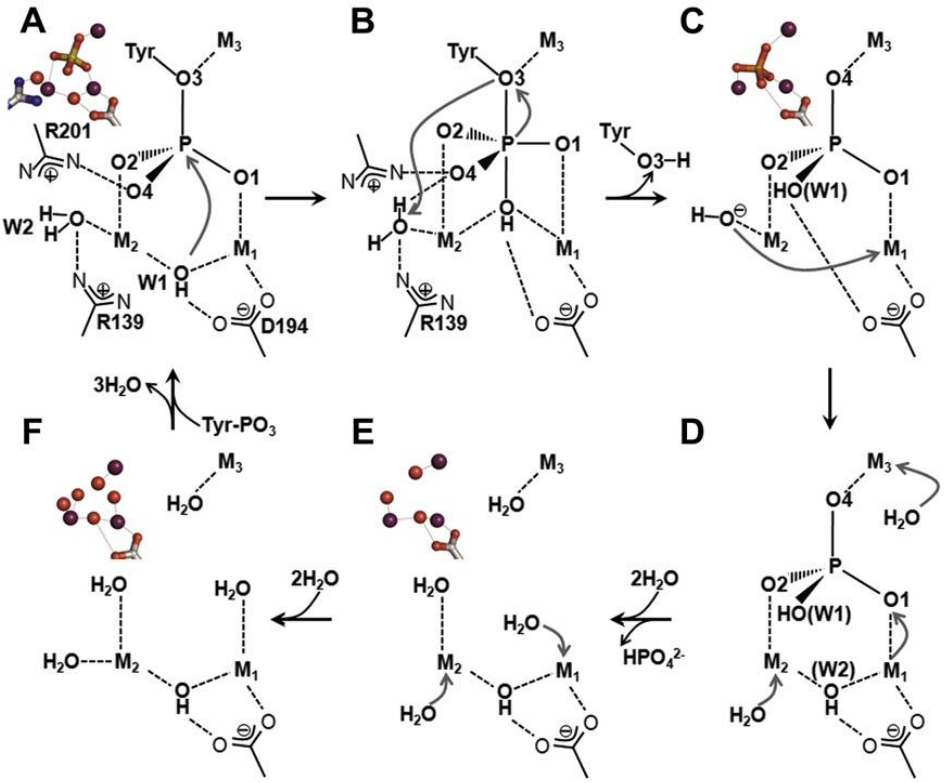


Figure 1.8. Sequence similarity network of cog4464 at a BLAST E-value of 1 x 10^{-70}. Biochemically characterized enzyme sequences are shown in cyan. CpsB from *S.pneumoniae* has been characterized structurally (△) whereas YwqE has been characterized in terms of its structure as well as kinetic rate constants (⬠).

CpsB was the first PTP from a Gram-positive bacterium whose crystal structure was solved.[29] The structures of both CpsB and YwqE showed that they belong to the PHP family of proteins and possess a distorted TIM-barrel fold. More details of the PHP family and its comparison with the binuclear AHS members are provided in chapter II. The PTP enzyme active site consists of a trinuclear metal center with the residues ligating the metal ions being strictly conserved among all orthologs of PTPs. The conformations of the protein structures with and without phosphate/sulfate bound at the active site, and modeling a phosphotyrosine substrate at the active site provided a hint that substrate specificity may be governed by two flexible loops that also act as

gatekeepers to the active site. Loop I (residues 44-51) has a consensus sequence G/P-x-Y/F, with the conserved Tyr or Phe corresponding to residue 48 in both YwqE and CpsB. Loop II has a consensus motif G/P-$x_{1\sim2}$-F-G-x_{0-1}-K/R, with the conserved Phe corresponding to Phe171 in YwqE and Phe173 in CpsB. The phosphate bound at the active site of YwqE (PDB: 3QY7) has one oxygen atom coordinating each of the active site metal and the fourth oxygen atom forming ionic interactions with a conserved arginine side chain (Arg201). Based on the crystallographic evidence of the bound phosphate, and the water molecules in the apo-structure, Kim *et al* have proposed a mechanism of hydrolysis of protein phosphotyrosine by YwqE, shown in **Scheme 1.3**.



Scheme 1.3. Proposed reaction mechanism of protein phosphotyrosine by YwqE from *B.subtilis*.[73] (A) Pre-catalytic state. M_1, M_2 and M_3 are the active site metal ions respectively. Arg139, Asp194, and Arg201 are indicated. (B) Trigonal bipyramidal transition state. The water molecule W2 is assumed to be proton donor for the leaving tyrosine. Another candidate for the role of proton donor is Arg201 of YwqE. (C) Post-catalytic state, as observed in the structure of the CpsB-PO_4^{3-} complex (PDB: 2WJF). (D) Restored hydroxo-bridged dinuclear metal center. (E and F) Metal-bound waters observed in the CpsB-apo (PDB: 2WJE) and YwqE-apo (PDB: 3QY6) structures respectively. **[**Reprinted from Journal of Structural Biology, 175, Kim, H.S., Lee, S.J., Yoon, H.J., Doo, R.A., Kim, D.J., Kim, S-J., and Suh, S.W., Crystal structures of YwqE from *Bacillus subtilis* and CpsB from *Streptococcus pneumoniae*, unique metal-dependent tyrosine phosphatases, 442-450, Copyright (2011), with permission from Elsevier.**]**

Kinetic characterization of YwqE for the dephosphorylation of *p*-nitrophenyl phosphate yielded the following kinetic constants: k_{cat}= 0.022 ± 0.001 s^{-1}, K_m= 0.345 ± 0.043 mM. The kinetic constants of dephosphorylation of a truncated mammalian tyrosine phosphopeptide KRLIEDNE(pY)TARGQ by YqwE were: k_{cat} = 0.016 ± 0.001 s^{-1}, K_m = 0.054 ± 0.008 mM. YwqE was unable to hydrolyze the phosphate group from

the serine phosphopeptide RRA(pS)VA and the threonine phosphopeptide KR(pT)IRR, which are substrates for serine and threonine phosphatases respectively. The enzyme was inhibited by vanadate with a $K_i = 14 \pm 3$ μM.[73]

COG1387 AND COG0613

Enzymes from cog1387 have been annotated as L-histidinol phosphate phosphatase-like proteins while that from cog0613 are annotated as metal-dependent phosphoesterases. These are the two COGs which are the focus of this work. Before the commencement of this study, the only known function in cog1387 was that of L-histidinol phosphate phosphatase (HPP). There were three crystal structures available: TTHA0331 from *Thermus thermophilus* (PDB: 2YXO, 2YZ5, 2Z4G), and LMOh7858 from *Listeria monocyotogenes* (PDB: 3DCP), both L-histidinol phosphate phosphatases, and B1034 from *E. coli* (PDB: 1M65, 1M68, and 1PB0), a protein of unknown function named YcdX. However, none of the structures above had any substrate-like molecule bound at the active site. The mechanism of phosphohydrolysis, the optimal conditions for enzymatic catalysis, and the structural details of substrate recognition and orientation at the active site of HPP enzymes were unknown and unexplored. There was no information available to distinguish HPP enzymes from the sequences whose functions were unknown, making it difficult to select targets for new function discovery. Several targets were computationally annotated as HPP, demonstrating the nuisance of misannotation as discussed earlier. On the other hand, there was no enzyme function or structure known from cog0613. This work attempts to bridge this gap in knowledge and

contribute towards the broad objective of developing a sequence-based strategy for annotating enzyme functions correctly. The sequence similarity network diagrams of cog1387 and cog0613 are shown in **Figures 1.9 and 1.10** respectively.

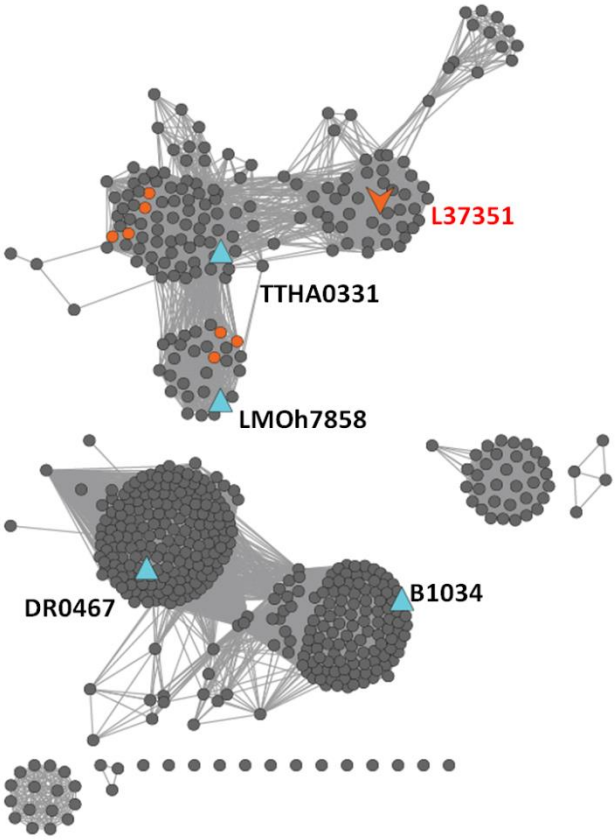


Figure 1.9. Sequence similarity network of cog1387 at an E-value of 1 x 10^{-20}. Protein sequences shown as triangles have crystal structures available in the PDB: TTHA0331 from *T.thermophilus* (PDB: 2YXO, 2YZ5, 2Z4G), LMOh7858 from *L.monocyotogenes* (PDB: 3DCP), B1034 from E. coli (PDB: 1M65, 1M68, and 1PB0), and N-terminal PHP domain of DNA polymerase X from D. radiodurans (PDB: 2W9M). L37351 was characterized and its crystal structure was solved as part of this work. Nodes colored orange were the targets of study as detailed in chapter II.

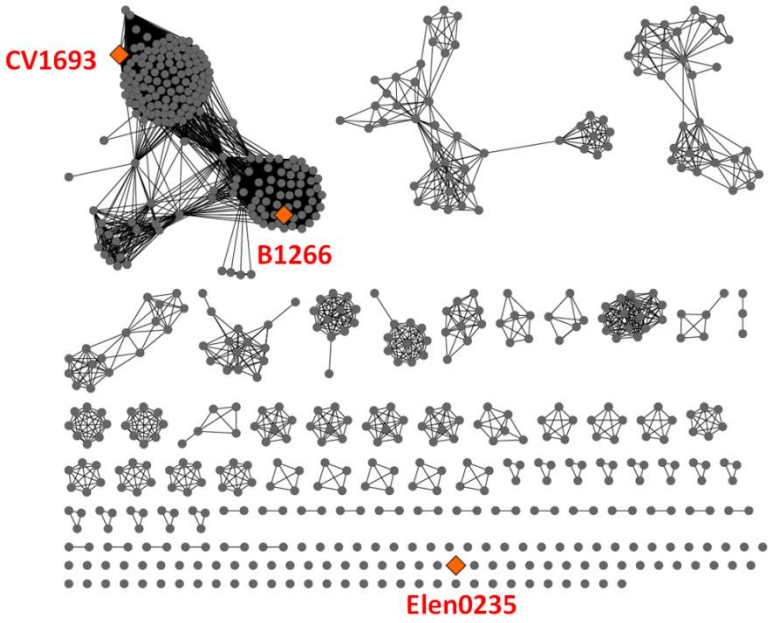


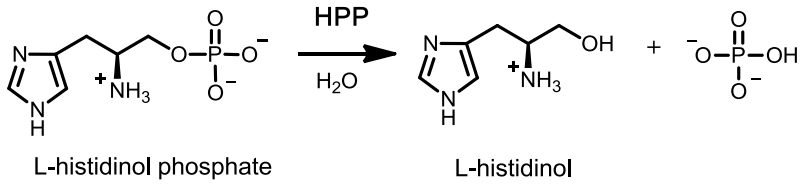
Figure 1.10. Sequence similarity network of cog0613 at a BLAST E-value of 1 x 10^{-60}. The enzyme targets characterized as a part of this work are shown in orange. Characterization of Elen0235 is described in chapter III, that of CV1693 in chapter IV, and B1266 in chapter V.

CHAPTER II

STRUCTURAL AND MECHANISTIC CHARACTERIZATION OF L-HISTIDINOL PHOSPHATE PHOSPHATASE FROM THE PHP FAMILY OF PROTEINS*

L-Histidinol phosphate phosphatase (HPP) catalyzes the penultimate step in the biosynthesis of L-histidine as illustrated in **Scheme 2.1**.[30] Two classes of HPP enzymes have been discovered. In *Escherichia coli*, for example, this enzyme evolved from the HAD superfamily of proteins and is bi-functional. The N-terminal domain has HPP activity while the C-terminal domain catalyzes the dehydration of imidazole glycerol-3-phosphate.[31-32] In other bacteria, such as *Bacillus subtilis*, HPP is a monofunctional enzyme from the polymerase and histidinol phosphatase (PHP) family of proteins.[33] In addition to HPP, the PHP family of enzymes includes the α-subunit of bacterial DNA polymerase III and family X DNA polymerases. Proteins from the PHP family have a $(\beta/\alpha)_7$-barrel structural fold and three divalent metals in the active site.[34]

*Reprinted with permission from "Structural and mechanistic characterization of L-histidinol phosphate phosphatase from the PHP family of proteins" by S. V. Ghodge, A. A. Fedorov, E. V. Fedorov, B. Hillerich, R. Seidel, S. C. Almo, and F. M. Raushel, *Biochemistry* **2013**, *52* (60), pp1101-1112. Copyright 2013 American Chemical Society.



Scheme 2.1. Reaction catalyzed by L-histidinol phosphate phosphatase.

The PHP family of proteins is also part of the amidohydrolase superfamily (AHS). Proteins from the AHS have a distorted $(\beta/\alpha)_8$-barrel structural fold and an active site that harbors a mono- or binuclear metal center.[5] In general, the metal ions within the active sites of AHS enzymes function to activate a water molecule for nucleophilic attack on ester or amide bonds. The PHP family stands apart among the amidohydrolase enzymes by having a distorted $(\beta/\alpha)_7$-barrel structural fold and an active site that contains three divalent cations. Superposition of the Cα atoms of the structurally characterized PHP and AHS enzymes demonstrates that β-strands **1-2** and **4-7** of the PHP $(\beta/\alpha)_7$-barrels overlay with β-strands **1-2** and **5-8** of the $(\beta/\alpha)_8$-barrels from the AHS. The long β-strand **3** of the $(\beta/\alpha)_7$-barrel extends from β-strand **3** to β-strand **4** of the $(\beta/\alpha)_8$-barrel enzymes.[35] It is reasonable to assume that the PHP and AHS proteins are evolutionarily related and thus the seven β-strands of the $(\beta/\alpha)_7$-barrel from the PHP family enzymes will be renumbered here to facilitate comparison with the proteins from the AHS. Therefore, the long β-strand **3** from the $(\beta/\alpha)_7$-barrel will henceforth be referred to as β-strand **3/4** while the remaining β-strands (**4-7**) will be renumbered as **5** through **8**, respectively. The third divalent metal ion in the active site of the PHP proteins will be denoted as the γ-metal. The other two metal ions in the active site will

be designated as α and β, in accordance with the binuclear metal centers from the amidohydrolase superfamily.[36]

HPP enzymes from the PHP family have all of the metal binding residues that are characteristic of the amidohydrolase superfamily.[6] There is an HxH motif at the end of β-strand **1**, two histidines at the ends of β-strands **5** and **6**, an aspartate at the end of β-strand **8** and a bridging glutamate from β-strand **3/4**. In addition to these conserved residues, an aspartate or histidine at the end of β-strand **1**, and two histidine residues at the ends of β-strands **2** and **8** ligate the third divalent cation. The close structural relationship between the PHP and AHS proteins is graphically illustrated in **Figure 2.1**.

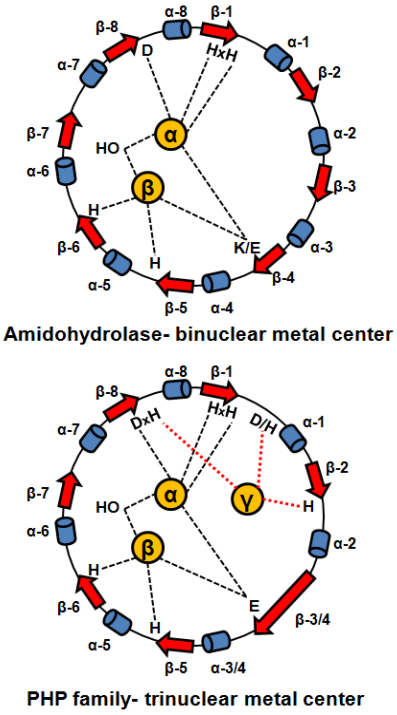


Figure 2.1. Cartoon depicting the metal ligation scheme for members of the amidohydrolase superfamily and the PHP family that bind two and three divalent metal ions respectively.

HPP and closely related homologs are grouped together in cog 1387 from the Clusters of Orthologous Groups defined by NCBI.[37] A sequence similarity network[38] for this cluster of proteins is presented in **Figure 2.2** at an E-value cutoff of 10^{-20}. The proteins (depicted in this figure as circular nodes) that function as authentic histidinol phosphate phosphatases were identified based upon whether the corresponding bacterial gene is found in an obvious operon for the biosynthesis of L-histidine. In this figure the nodes colored blue are co-localized with other genes responsible for L-histidine biosynthesis and thus have a high likelihood of being authentic HPP enzymes. Nodes that are colored green have all of the apparent sequence motifs that are characteristic of the HPP enzymes from the PHP family; however, in these organisms the genes for the proteins required for the biosynthesis of L-histidine are not clustered with one another. These proteins are annotated as HPP since these organisms do not possess a protein from the HAD superfamily that could function as HPP but they have the remaining genes needed for the biosynthesis of histidine. The red nodes represent enzymes that might function as HPP but the genomes of their respective organisms lack a majority of genes from the L-histidine biosynthetic pathway. Those nodes colored gray lack certain residues that appear critical for HPP activity, based on this investigation, while those colored orange do not catalyze the hydrolysis of histidinol phosphate.

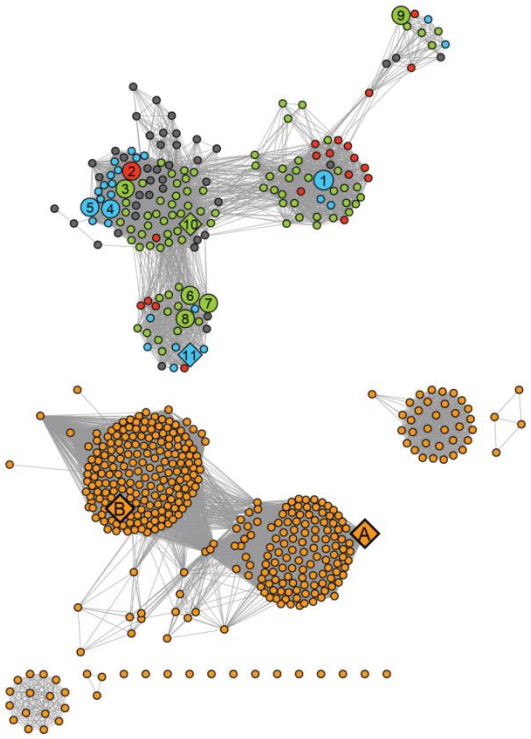


Figure 2.2. Sequence similarity network of proteins in cog 1387 at an E-value of 1 x 10^{-20} created using Cytoscape (http://www.cytoscape.org). Each node (sphere) represents a single sequence and each edge (line) represents the pairwise connection between two sequences with the most significant BLAST E-value (better than 10^{-20}). Lengths of edges are not significant, except for tightly clustered groups, which are more closely related than sequences with only a few connections. The nodes were assigned colors as follows: (blue) authentic HPP enzymes co-localized with other genes involved in the biosynthesis of L-histidine; (green) gene products that possess all the sequence motifs characteristic of the HPP enzymes, but not found co-localized with other L-histidine biosynthetic genes present in these organisms; (red) gene products that possess all the sequence motifs characteristic of an authentic HPP, but the organism lacks a majority of the genes required for L-histidine biosynthesis; (gray) protein sequences that bear significant sequence similarity to authentic HPPs, but lack certain sequence elements critical for HPP activity; (orange) proteins that are not HPP enzymes. Specific HPP enzyme referred to in this paper are indicated by numbers 1-10. These proteins are identified by their locus tags: (**1**) L37351; (**2**) MCCL_0344; (**3**) BBR47_00270; (**4**) BCE_1533; (**5**) BcerKBAB4_1335; (**6**) BSU29620; (**7**) BH3206; (**8**) GK2799; (**9**) SMU_1486c; (**10**) TTHA0331; (**11**) LMOh7858_0629. The HPP enzymes characterized in this study are shown as large spheres and enzymes whose crystal structures are available in the PDB are indicated as diamonds. Other crystal structures available for enzymes in this cog are indicated as diamonds and include (**A**) YcdX from *E. coli* (pdb: 1m65, 1m68, 1pbo) and (**B**) N-terminal PHP domain of DNA polymerase X from *D.radiodurans* (pdb: 2w9m).

Crystal structures of HPP from *Listeria monocytogenes* str. 4b h7858 (PDB id: 3dcp) and *Thermus thermophilus* HB8 (PDB id: 2z4g, 2yxo, and 2yz5) are available from the Protein Data Bank. In **Figure 2.2** these proteins are designated as **10** and **11**, respectively. Unfortunately, neither of these proteins have structures with the substrate or an inhibitor bound in the active site. In this investigation, we have determined the structure of HPP from *Lactococcus lactis* subsp. lactis Il1403 and have investigated the mechanism of action. This gene was a part of the characterized histidine biosynthetic operon in *L.lactis*[39], and hence was deemed to be a good starting point for this study. Recombinant HPP was found to bind iron and zinc in the active site. However, when the recombinant enzyme was expressed in a medium depleted of iron, the enzyme binds zinc and manganese. The zinc and manganese enzyme has significantly higher catalytic activity as compared to the iron-containing enzyme.

MATERIALS AND METHODS

All chemicals were purchased from Sigma-Aldrich, unless indicated otherwise. Pfu Turbo Polymerase, T4 DNA ligase and restriction enzymes were procured from New England Biolabs. DNA primers and Big Dye were obtained from Integrated DNA Technologies (IDT). The vector pET30a(+) was purchased from EMD4 Biosciences. *E. coli* BL21(DE3) and XL1 Blue competent cells were obtained from Stratagene. The P_i Colorlock Gold kit for the determination of inorganic phosphate was obtained from Innova Biosciences. L-Histidinol phosphate was a generous gift from Professor Debra

Dunaway-Mariano (Department of Chemistry and Chemical Biology, University of New Mexico).

Cloning, Expression and Purification of Bacterial HPP Enzymes. The genes encoding BSU29620, BH3206, GK2799, BBR47_00270, MCCL_0344, BCE_1533, SMU_1486c, BcerKBAB4_1335 were synthesized and cloned into pUC57 (GenScript). Expression constructs were generated by PCR amplification using KOD Hot Start DNA Polymerase (Novagen). The amplified fragments were cloned into the C-terminal TEV cleavable StrepII-6x-His-tag containing expression vector, CHS30, by ligation-independent cloning.[40] Expression vectors were transformed into *E. coli* BL21(DE3) containing the pRIL plasmid (Stratagene) and used to inoculate a 10 mL 2xYT culture containing 25 μg/mL kanamycin and 34 μg/mL chloramphenicol. The culture was allowed to grow overnight at 37 °C in a shaking incubator. 10 mL of the overnight culture was used to inoculate 2 L of PASM-5052 auto-induction media[41] containing 150 μM 2,2′-bipyridyl, 1 mM $ZnCl_2$, and 1 mM $MnCl_2$. The culture was placed in a LEX48 airlift fermenter and incubated at 37 °C for 4 hours and then at 22°C overnight. The culture was harvested and pelleted by centrifugation. Cells were resuspended in Lysis Buffer containing 20 mM HEPES, pH 7.5, 500 mM NaCl, 20 mM imidazole, and 10% glycerol and lysed by sonication. Lysates were clarified by centrifugation at 35,000g for 30 minutes. Proteins were purified on an AKTAxpress FPLC (GE Healthcare). Clarified lysates were loaded onto a 5 mL Strep-Tactin column (IBA), washed with 5 column volumes of Lysis Buffer, and eluted with 20 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, 10% glycerol, and 2.5 mM desthiobiotin. The

eluent was loaded onto a 1 mL His60 Ni Superflow column (Clontech), washed with 10 column volumes of Lysis Buffer, and eluted with 20 mM HEPES, pH 7.5, 500 mM NaCl, 500 mM imidazole, and 10% glycerol. The purified sample was loaded onto a HiLoad S200 16/60 PR gel filtration column which was equilibrated in 20 mM HEPES pH 7.5, 150 mM NaCl, 10% glycerol, and 5 mM DTT. Peak fractions were collected and protein was analyzed by SDS-PAGE. Samples were concentrated to 5-7 mg/mL using Amicon Ultra centrifugal filters (Millipore), snap frozen in liquid nitrogen, and stored at -80 °C.

Expression and Purification of HPP from *L. lactis*. The gene for HPP (locus tag: L37351) from *Lactococcus lactis* subsp. lactis Il1403 (gi|15673198) was initially obtained from the New York Structural GenomiX Research Consortium (NYSGXRC) as target ID: 9530a 2BCt1p1. The protein obtained by transformation of the plasmid in *E. coli* BL21(DE3) was sparingly soluble after sonication of the cells. A closer inspection of the nucleotide sequence for the gene revealed that 30 base pairs were missing. The missing base-pairs were reintroduced into the gene along with a C-terminal poly-histidine purification tag. The sub-cloning of the gene into a pET30a(+) vector was carried out using the following forward and reverse primers: 5′-AAGTATCATATGTCCCTTAAAAAATTAGATTATCATTTCCACTCTCATTTTTCGG-3′ and 5′-AGAAGAAAGCTTTTCTTTAATAGACTTTTTATTTTTATCAATTTTCATCCCACTAAAAGTTGCTAGTTCGTGAAAACC-3′. The primer pair contained restriction sites for *Nde*I and *Hin*dIII, respectively. The PCR product was amplified and isolated using

the Promega DNA purification kit. The PCR product and the vector pET30a(+) were double digested with *Nde*I and *Hin*dIII, purified using agarose gel purification, and then ligated using T4 DNA ligase.

Iron-Free Expression and Purification. The recombinant plasmid was transformed into *E. coli* BL21(DE3) cells by electroporation. A single colony was used to inoculate 5-mL starter cultures of LB medium containing 50 μg/mL kanamycin, which were incubated overnight at 37 °C. Each starter culture was used to inoculate one liter of the same medium. A modified iron-free expression protocol was adopted from a published procedure.[42] Cultures were grown at 37 °C until the OD_{600} reached 0.15 - 0.2. The temperature was reduced to 30 °C and the iron-specific chelator 2,2′-bipyridyl was added to a final concentration of 120 μM. The divalent metal ions, Zn^{2+} and/or Mn^{2+} (1:1 ratio), were added when the OD_{600} reached ~0.4 to a final concentration of 1.0 mM. Protein expression was induced with 0.25 mM isopropyl D-thiogalactopyranoside (IPTG) when the OD_{600} reached ~0.6 and then the temperature was lowered to ~15 °C. Twelve hours after the addition of IPTG, the cells were isolated by centrifugation and stored at -80 °C.

The protein purification was carried out at 4 °C using Ni-NTA affinity chromatography. Four to five grams of cells were thawed and resuspended in ~40 mL of binding buffer (20 mM HEPES, pH 7.6, 5.0 mM imidazole, and 500 mM NaCl) containing 0.5 mg/mL of the protease inhibitor phenylmethanesulfonyl fluoride (PMSF). Cells were lysed by sonication and then ~2 mg/g of cells of bovine pancreatic DNase I was added. The suspension was stirred for 15 minutes after the final sonication cycle

and the suspended cell debris was removed by centrifugation. The supernatant solution was filtered and passed through a Ni-NTA column equilibrated with the binding buffer. The column was washed with 5 column volumes of wash buffer (20 mM HEPES pH 7.6, 50 mM imidazole, and 500 mM NaCl). The protein was eluted with a solution containing 15 mM HEPES pH 7.6, 250 mM imidazole, and 350 mM NaCl. Fractions were analyzed using UV absorbance at 280 nm. The eluted protein was dialyzed against a solution containing 50 mM HEPES, pH 8.2, 250 mM NaCl, 0.1 mM $ZnCl_2$ and 0.1 mM $MnCl_2$. The protein concentration was determined by UV absorbance at 280 nm using a calculated molar extinction coefficient of 22,015 M^{-1} cm^{-1}. The protein was flash frozen using liquid nitrogen and stored at -80 °C.

Metal Content Analysis. The metal content of the proteins was determined with a Perkin Elmer DRCII inductively-coupled plasma mass spectrometer (ICP-MS) using 1% v/v nitric acid as the matrix. Concentrated protein samples were digested with nitric acid (≥ 69%, Fluka Analytical) for ~ 30 minutes to prevent protein precipitation during the measurement. The sample was subsequently diluted with deionized water to adjust the final concentrations of the protein and nitric acid to ~1 μM and 1% (v/v) respectively. Protein samples whose buffers contained divalent metal ions were first passed through a PD-10 desalting column (GE Healthcare) before subjecting them to the acid digestion.

Structure Determination of HPP from *L. lactis*. Crystals were grown by sitting drop vapor diffusion at room temperature for L-histidinol phosphate phosphatase liganded with: (a) Zn^{2+} and sulfate ion (HPP·Zn·SO_4), (b) Zn^{2+}, L-histidinol, and

phosphate ion (HPP·Zn·HOL·HPO_4), and (c) Zn^{2+}, L-histidinol, and arsenate (HPP·Zn·HAR). The crystallization conditions utilized the following conditions: (*i*) For HPP·Zn·SO_4, the protein solution contained protein (13 mg/mL) in 50 mM HEPES (pH 8.2), 300 mM NaCl, 0.5 mM $ZnCl_2$, and 0.1 mM $MnCl_2$; the precipitant contained 25% PEG 4000, 0.1 M sodium acetate (pH 4.6), and 0.2 M ammonium sulfate. Crystals appeared in 3-4 days and exhibited diffraction consistent with the space group $P2_12_12_1$, with one polypeptide chain per asymmetric unit. (*ii*) For HPP·Zn·HOL·P_i, the protein solution contained enzyme (13 mg/mL) in 50 mM HEPES (pH 8.2), 250 mM NaCl, 0.5 mM $ZnCl_2$, 0.1mM $MnCl_2$, and 100 mM L-histidinol phosphate; the precipitant contained 2.5 M NaCl, and 0.1 M imidazole (pH 8.0). Crystals appeared in 8-9 days and exhibited diffraction consistent with the space group $P2_12_12$, with one polypeptide chain per asymmetric unit. The L-histidinol phosphate hydrolyzed during the crystallization process resulting in the L-histidinol and phosphate observed in the active site. (*iii*) For HPP·Zn·HAR, the protein solution contained enzyme (13 mg/mL) in 50 mM HEPES (pH 8.2), 250 mM NaCl, 0.5 mM $ZnCl_2$, 0.1 M $MnCl_2$, 100 mM L-histidinol dihydrochloride, and 200 mM sodium arsenate; the precipitant contained 2.5 M NaCl, and 0.1 M imidazole (pH 8.0). Crystals appeared in two weeks and exhibited diffraction consistent with the space group $P2_12_12$, with one polypeptide chain per asymmetric unit.

Prior to data collection, all crystals were transferred to cryoprotectant solutions composed of their mother liquids and 20% glycerol and flash-cooled in a nitrogen stream. Data sets were collected at the NSLS X4A beamline (Brookhaven National Laboratory) on an ADSC CCD detector and at NSLS X29A beamline on the 315q CCD

detector. Diffraction intensities were integrated and scaled with programs DENZO and SCALEPACK.[43] The data collection statistics are given in **Table 2.1**.

All three HPP structures (**Table 2.1**) were determined by molecular replacement with the BALBES software suite[44]. Partially refined structures of all three HPP crystal forms (**Table 2.1**) were the outputs from BALBES. Iterative cycles of refinement were performed for each structure involving manual model rebuilding with COOT[45], refinement with PHENIX[46], and automatic model rebuilding with ARP[47]. The quality of the final structures was verified with omit maps. The stereochemistry was checked with WHATCHECK[48] and MOLPROBITY[49]. Program LSQKAB[50] was used for structural superposition. Structural figures with electron density maps were prepared with PYMOL[51].

Table 2.1. Data collection and refinement statistics for HPP complexes.

Data collection	**Structure**		
	HPP·Zn·SO_4	**HPP·Zn·HOL·HPO_4**	**HPP·Zn·HAR**
Space group	$P2_12_12_1$	$P2_12_12$	$P2_12_12$
No. of molecules in asym. unit	1	1	1
Cell dimensions			
a (Å)	51.48	85.66	85.86
b (Å)	76.49	86.68	86.62
c (Å)	78.06	45.08	45.09

Table 2.1 Continued

Data collection	**Structure**		
	HPP·Zn·SO_4	**HPP·Zn·HOL·HPO_4**	**HPP·Zn·HAR**
Resolution (Å)	1.32	1.65	1.93
No. of unique reflections	72355	40890	25741
R_{merge}	0.075	0.093	0.098
Completeness (%)	99.1	98.8	99.4
Refinement			
Resolution (Å)	25.0-1.32	25.0-1.65	25.0-1.93
R_{cryst}	0.152	0.160	0.161
R_{free}	0.162	0.186	0.195
No. of atoms			
Protein	2279	2185	2188
Waters	422	229	181
Ligand atoms	18	57	33
Bound ligands	3 ZN, 3 SO4	4 ZN, 5 Cl, 3 PEG, PO_4, 2 HOL	4ZN, 3 Cl, PEG, IMD, HAR
R.m.s deviations			
Bond lengths (Å)	0.006	0.006	0.007
Bond angles (°)	1.1	1.0	1.1
PDB entry	4GC3	3UMU	4GK8

Enzymatic Assays. The kinetic constants were determined using the P_i Colorlock Gold kit for measurement of free phosphate. The enzymatic reaction was conducted at 25 °C. A single assay consisted of 6-8 substrate concentrations with three time points taken over a period of 0-20 minutes. The color was allowed to develop for 30 minutes and the final absorbance was determined at 650 nm. The phosphate concentration of each well was calculated from a standard curve and the initial rates at each substrate concentration were determined using linear regression. The dependence of the kinetic parameters of HPP as a function of pH was determined over the pH range of 7-10. The buffers used in this study were HEPES (pH 7.0-8.4) and CHES (pH 8.6-10.0) at a final concentration of 50 mM. The assays were carried out at 22 ± 1 °C. The solvent isotope effects on k_{cat} and k_{cat}/K_m were determined using 50 mM CHES pH/pD 8.5 in 100% H_2O and D_2O. The effect of micro-viscosity on the kinetic parameters was determined using glycerol as the viscogen. The assays were performed at 25 °C and pH 8.6 using 50 mM CHES. The relative viscosities of the glycerol-water solutions were calculated according to Cheng.[52]

Site-Directed Mutagenesis. All point mutants were constructed using the standard QuikChange PCR protocol as per the manufacturer's instructions. The respective variants were expressed using the iron-free expression protocol and the growth medium was supplemented with 0.5 mM each of zinc acetate and manganese sulfate. The variants were purified using Ni-NTA affinity chromatography and assayed at pH 8.5 using 50 mM CHES as described above.

Data Analysis. Kinetic parameters, k_{cat} and k_{cat}/K_m, were obtained by fitting the initial velocity data to equation 1 using the nonlinear least-squares fitting program in SigmaPlot 10, where v is the initial velocity at substrate concentration [A], [E_t] is the enzyme concentration, k_{cat} is the turnover number, and K_m is the Michaelis constant. The pH-rate profiles were analyzed by fitting the data to equations 2 - 4, where y is the value of either k_{cat} or k_{cat}/K_m, c is the maximum value for either k_{cat} or k_{cat}/K_m and [H^+] is the proton concentration. In these equations K_a and K_b are the apparent dissociation constants for the ionizable groups.

$$v/[E_t] = k_{cat}[A] / (K_m + [A]) \quad (2.1)$$

$$\log y = \log (c/ (1 + [H^+] / K_a + K_b/[H^+])) \quad (2.2)$$

$$\log y = \log (c / (1 + [H^+] / K_a)) \quad (2.3)$$

$$\log y = \log (c / (1 + K_b / [H^+])) \quad (2.4)$$

Bioinformatics. The protein sequences assigned to cog 1387 were obtained from the NCBI database (www.ncbi.nlm.nih.gov). The redundant sequences were removed using Jalview 2.7[53] and then converted to FASTA format. The sequence similarity network diagrams at various stringency levels were created using Cytoscape 2.8.2[54]. Primary sequence alignments were made using Jalview 2.7. Genomic analyses were performed using the MicrobesOnline website for comparative genomics.

RESULTS

Optimization of Catalytic Activity. The kinetic constants for the hydrolysis of histidinol phosphate by HPP from *L. lactis*, isolated from cells grown in standard LB medium, were determined at pH 9.0 to be: k_{cat} = 1.9 s^{-1} and k_{cat}/K_m = 2.6 x 10^3 M^{-1} s^{-1}. These constants correspond reasonably well with those previously determined for HPP from *T. thermophilus* HB8[35]. However, the catalytic efficiency of ~10^3 M^{-1} s^{-1} is relatively low and the purified enzyme contained iron and zinc in a ratio of approximately 2:1. Attempts were therefore made to add, exchange, or remove metals from the protein. The enzyme did not show any significant increases in catalytic activity when Zn^{2+}, Mn^{2+}, Co^{2+} or Ni^{2+} were added to the protein at pH values ranging from 6.5 to 9.0. Incubation of the enzyme with the metal chelators EDTA or *o*-phenanthroline precipitated the protein. The low catalytic activity of the purified protein was therefore assumed to result from the presence and/or oxidation of iron in the active site.[42] Cells were subsequently grown using an iron-free expression protocol and the growth medium supplemented with Zn^{2+} and/or Mn^{2+}, since previous structural studies of PHP family enzymes have shown their affinity toward zinc and manganese ions (pdb: 1pb0- ycdX from *E. coli*, 2w9m-DNA polX from *D.radiodurans*, 2yb1-amidohydrolase from *C.violaceum*). The protein expressed and purified under the iron-free conditions with the addition of Zn^{2+} and Mn^{2+} was found to have the most catalytic activity. The values of k_{cat} and k_{cat}/K_m for the hydrolysis of histidinol phosphate by HPP increased by two orders of magnitude when the iron content of the enzyme was eliminated (**Table 2.2**).

Table 2.2. Kinetic parameters and metal content of HPP from *L. lactis.*

HPP	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	Metal content
Wild-type	174 ± 10	1.9 ± 0.2	(8.9 ± 0.3) x 10^4	Zn (4.5); Mn (0.8)
D17N	n.d.[a]	n.d.	(5.6 ± 0.1) x 10	Zn (0.5)
H42N	0.62 ± 0.04	1.3 ± 0.2	(4.8 ± 0.6) x 10^2	Zn (1.9); Fe (0.3)
E115Q	31 ± 2	1.7 ± 0.2	(1.8 ± 0.2) x 10^4	Zn (2.2); Fe (0.4); Mn (0.4)
Y117F	110 ± 3	0.74 ± 0.06	(1.5 ± 0.1) x 10^5	Zn (1.4)
Y117A	53 ± 3	5.4 ± 0.6	(9.8 ± 0.6) x 10^3	Zn (1.3); Mn (0.3)
Y157F	61 ± 3	1.7 ± 0.2	(3.6 ± 0.3) x 10^4	Zn (1.7); Fe (0.3)
R160M	0.74 ± 0.04	1.5 ± 0.2	(4.9 ± 0.5) x 10^2	Zn (1.5)
R160A	n.d.	n.d.	(1.0 ± 0.1) x 10^2	Zn (1.9); Fe (0.5)
Y161F	180 ± 6	2.3 ± 0.2	(7.9 ± 0.4) x 10^4	Zn (2.3); Fe (0.3)
Y161A	48 ± 1	2.5 ± 0.9	(2.5 ± 0.9) x 10^4	Zn (1.0); Fe (0.3)
R197M	38 ± 2	4.5 ± 0.6	(8.5 ± 0.7) x 10^3	Zn (3.3); Fe (0.6); Mn (0.5)
D228N	0.031 ± 0.002	0.14 ± 0.04	(2.2 ± 0.5) x 10^2	Zn (1.2)
H230N	n.d.	n.d.	(4.5 ± 0.1) x 10^2	Zn (1.6)

[a]nd: not determined because the enzyme did not saturate at the highest substrate concentration tested.

pH-Rate Profiles. The variation of the kinetic constants with changes in pH for the hydrolysis of histidinol phosphate by HPP from *L. lactis* was determined. The pH rate profiles are presented in **Figure 2.3**. For k_{cat}, the data indicate that a group must be unprotonated for maximum activity. The pK_a, from a fit of the data to equation 2.3, is 7.5 ± 0.1. For k_{cat}/K_m, the data indicate that an ionizable group on the substrate or

enzyme must be protonated for maximum catalytic activity with a pK_b of 9.7 ± 0.1 from a fit of the data to equation 2.4. It was not possible to measure the activity of the enzyme below pH 7 because the protein precipitated.

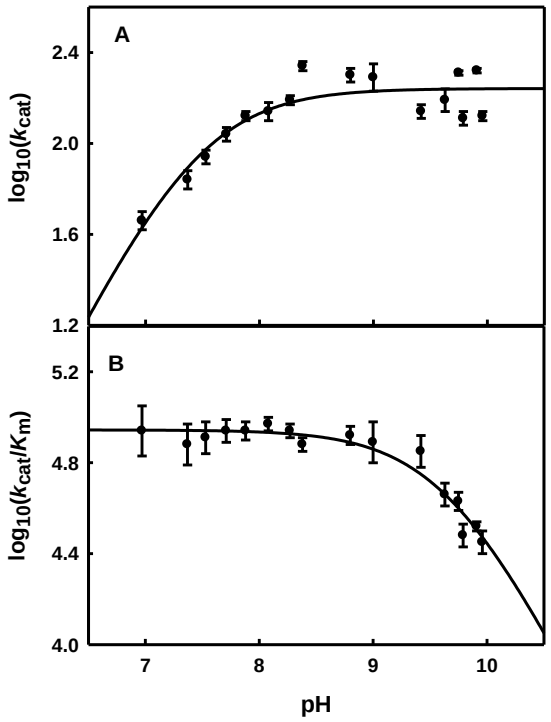


Figure 2.3. pH-Rate profile for the hydrolysis of L-histidinol phosphate by HPP from *L. lactis* at 22 ± 1 °C. (A) Variation of log k_{cat}/K_m versus pH. The solid line represents a fit of the data to equation 4. (B) Variation of log k_{cat} versus pH. The solid line represents a fit of the data to equation 3.

Solvent Isotope and Viscosity Effects. The effect of substituting H_2O with D_2O on the kinetic parameters for phosphate hydrolysis by HPP was determined at a pH/pD of 8.5. The solvent isotope effects on k_{cat} and k_{cat}/K_m were 1.4 ± 0.1 and 1.3 ± 0.1,

respectively, without accounting for the difference in the viscosities of H_2O and D_2O. The effects on the kinetic constant from changes in the micro-viscosity of the solvent on k_{cat} and k_{cat}/K_m are presented **Figure 2.4**. At pH 8.6, the slope for the k_{cat} profile is 0.1 ± 0.1 and that for k_{cat}/K_m is 0.3 ± 0.1.

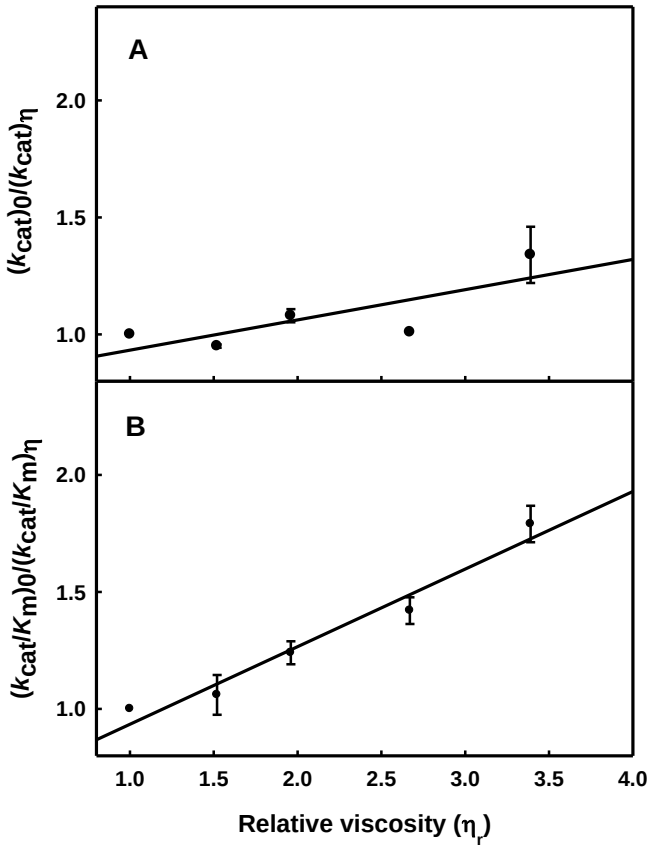


Figure 2.4: Effect of solvent viscosity on k_{cat} (A) and k_{cat}/K_m (B) for the hydrolysis of L-histidinol phosphate by HPP at pH 8.6.

Crystal Structure Determination of HPP from *L. lactis*. Three high resolution crystal structures were determined: HPP·Zn·SO_4 (PDB: 4GC3), HPP·Zn·HOL·HPO_4 (PDB: 3UMU), and HPP·Zn·HAR (PDB: 4GK8). All three structures exhibited a nearly identical overall $(\beta/\alpha)_7$-fold with r.m.s. deviation of 0.29 Å for all 264 equivalent Cα positions between structures 1 and 2; 0.27Å between structures 1 and 3; and 0.11 Å between structures 2 and 3 respectively. The HPP barrel, shown in **Figure 2.5**, has seven β-strands denoted with the amidohydrolase numbering: β-1 (residues 5-9), β-2 (36-44), β-3/4 (75-84), β-5 (103-108), β-6 (150-153), β-7 (191-196), and β-8 (223-227). The HPP β-barrel is surrounded by seven α-helices: α-1 (22-34), α-2 (56-73), α-3/4 (89-98), α-5 (128-146), α-6 (171-189), α-7 (203-218), and α-8 (236-250). Seven loops are located at C-terminal ends of the barrel: L-1 (10-21), L-2 (45-55), L-3/4 (85-88), L-5 (109-127), L-6 (154-170), L-7 (197-202), and L-8 (228-235). These loops contribute residues that form the active site of the enzyme. C-terminal chain segment 251-264 closes the opposite N-terminal end of the barrel from bulk solvent. This chain segment includes a short anti-parallel β-loop 256-263. All loops are well defined in all three liganded HPP structures. Only one non-glycine residue (His-154) lies in the disallowed regions of the Ramachandran plot in all three HPP structures (**Table 2.1**). Some active site residues may occur as real outliers in the Ramachandran plot of a high resolution crystal structure. His-154 is located in the HPP active site, serving as a ligand to the β-metal center (**Figures 2.6 and 2.7**), and belongs to the C-terminal loop L6. All three structures were produced with excess of Zn^{2+} in the cocrystallization solutions and all three metal sites refined well as Zn^{2+} ions. The sulfate ion in the HPP·Zn·SO_4 structure

is shifted 1.9 Å and oriented differently than the phosphate group in the HPP·Zn·HOL·HPO_4 structure and the arsenate group in HPP·Zn·HAR. The hydroxyl bridging the α- and β-metals in the HPP·Zn·SO_4 structure is shifted 1.1 Å relative to its position in the HPP·Zn·HOL·HPO_4 and HPP·Zn·HAR structures. Final crystallographic refinement statistics for all determined HPP structures are provided in the **Table 2.1**.

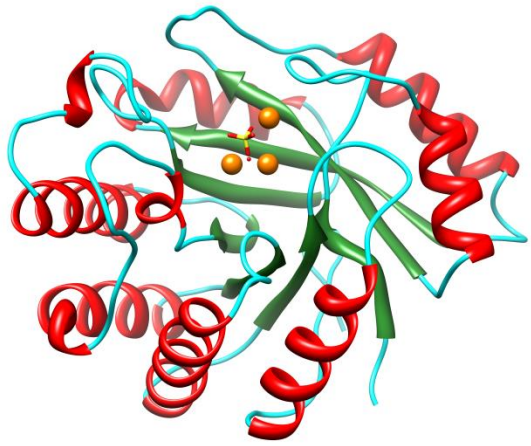


Figure 2.5. Three-dimensional structure of HPP from *L.lactis* with sulfate bound at the active site. The α-helices are shown in red, β-sheets are shown in green, loops are shown in cyan while the active site Zn^{2+} ions are shown in orange.

The active site of the HPP structures contains a trinuclear metal center with the three metal sites designated as α, β and γ in **Figure 2.6**. The α- and β-sites are located deeper inside the barrel C-terminal entrance, and γ-site is located near the surface of the monomer, closer to the bulk solvent. The Zn_α-Zn_β distance is 3.38 Å, Zn_α-Zn_γ is 4.83, and Zn_β-Zn_γ is 6.09 Å in the HPP·Zn·SO_4 structure. The corresponding distances in

HPP·Zn·HOL·HPO_4 structure are 3.65 Å, 4.53 Å, and 5.72 Å, and in HPP·Zn·HAR are 3.69 Å, 4.38 Å, and 5.71 Å. Zn ions at α and β-positions are bridged by a water molecule (or hydroxide ion), W1, and by the carboxylate of Glu-81 in a bidentate manner. His-9, His-11 and Asp-228 coordinate the α-metal center; His-109 and His-154 coordinate the β-metal center while Asp-17, His-42 and His-230 coordinate the γ-metal center. The distances between the metal ions and the respective ligands in the HPP·Zn·SO_4 structure are shown in **Figure 2.6**.

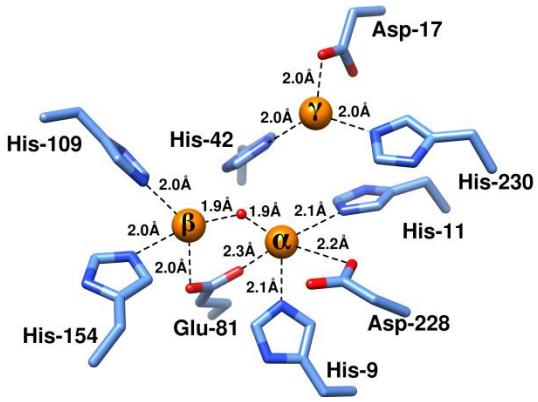


Figure 2.6. Active site structure of HPP from *L. lactis*. The three Zn^{2+} ions in the active site are shown as orange spheres and the enzyme residues serving as ligands to the metal ions are shown in light blue.

The crystalline HPP complexes HPP·Zn·HOL·HPO_4 and HPP·Zn·HAR were produced by cocrystallization with high concentrations of L-histidinol phosphate, and a mixture of L-histidinol and sodium arsenate respectively. The complex of L-histidinol and arsenate with HPP is a reasonable mimic of the orientation of the substrate at the

active site. This complex exhibited continuous electron density as illustrated in **Figure 2.7**. Clear electron density is also observed for the three metal ions in the active site.

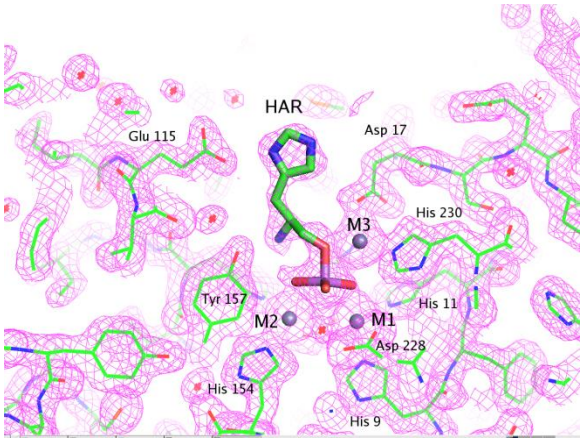


Figure 2.7. Representative electron density map for the active site of the HPP complexed with Zn^{2+} and L-histidinol arsenate and contoured at 1.5σ. The figure was produced with PyMOL. The details of the interactions between L-histidinol arsenate and the active site are described in the text.

We also searched for proteins structurally similar to HPP from *L. lactis* using DALI server.[55] The highest Z-score of 30.0 was obtained with HPP from *Thermus thermophilus* HB8 (PDB code: 2yxo; an r.m.s. deviation of 2.1 Å for 247 equivalent Cα positions, and a sequence identity of 25%). The next highest Z-score of 26.2 was obtained with HPP from *Listeria monocytogenes* str. 4b h7858 (PDB code: 3dcp; an r.m.s. deviation of 2.2 Å for 237 equivalent Cα positions, and a sequence identity 19%). Both HPP structures found by DALI have the PHP fold and exhibit the trinuclear metal center in a position similar to HPP from *L. lactis*.

Site-Directed Mutagenesis. The substitution of residues in the active site of HPP from *L. lactis* was performed to establish the requirement for the binding of the third metal, identify the role of the invariant aspartate found at the C-terminal end of β-strand **8**, and to clarify the function of residues that appear to interact with the substrate. The kinetic constants for thirteen variants constructed from ten residue positions are presented in **Table 2.2**. For the three residues that coordinate M_γ (Asp-17, His-42, and His-230), there are significant reductions in the values of k_{cat} and k_{cat}/K_m when these residues are changed to asparagine. The total metal content of the enzyme preparations for these variants is also reduced. When Asp-228 is mutated to asparagine, k_{cat} is reduced by approximately 6,000-fold. This residue coordinates M_α and has been shown in other members of the AHS to be critical for proton transfer reactions from the hydrolytic water/hydroxide in the active site.[36a, 56,57] The mutation of Glu-115, Tyr-117, Tyr-157, Tyr-161, Arg-160, and Arg-197 was designed to ascertain the functional significance of those residues that appear to interact with the substrate/product in the active site of HPP. The most significant reductions in the kinetic constants were for the mutation of Arg-160. The mutation of the three tyrosine residues to phenylalanine reduced k_{cat} by less than a factor of four. Mutation of Arg-197 and Glu-115 reduced k_{cat} by a factor of approximately 6-8 and k_{cat}/K_m by approximately 10-20.

Characterization of HPP from Other Organisms. Enzyme targets predicted to be authentic HPP enzymes from the PHP family were selected for expression, purification and characterization. The kinetic constants for these proteins for the hydrolysis of histidinol phosphate are presented in **Table 2.3**. The k_{cat} and k_{cat}/K_m values

for the hydrolysis of histidinol phosphate by these enzyme preparations are significantly lower than the kinetic constants measured for the *L. lactis* HPP. These proteins were expressed using iron-free conditions; however, the metal stoichiometry per subunit is, in general, less than three and some of the proteins were isolated with measureable amounts of iron.

Table 2.3. Kinetic parameters of L-histidinol phosphatases from cog 1387.

Locus tag	**Organism**	**k_{cat} (s^{-1})**	**K_m (mM)**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**	**Metal content per enzyme monomer**
MCCL_0344	*Macrococcus caseolyticus* JCSC5402	30 ± 2	1.5 ± 0.3	(2.0 ± 0.2) x 10^4	Mn (0.9); Fe (0.6); Zn (0.3)
BBR47_00270	*Brevibacillus brevis* NBRC 100599	7.2 ± 0.4	2.6 ± 0.3	(2.8 ± 0.2) x 10^3	Zn (1.6); Mn (0.3)
BCE_1533	*Bacillus cereus* ATCC10987	8.1 ± 0.1	1.5 ± 0.1	(5.6 ± 0.2) x 10^3	Mn (1.4); Zn (0.4)
BcerKBAB4_1335[a]	*Bacillus weihenstephanensis* KBAB4	1.2 ± 0.1	4.9 ± 0.3	(2.5 ± 0.2) x 10^2	Zn (1.5); Fe (1.2)
BcerKBAB4_1335		34 ± 2	1.3 ± 0.2	(2.6 ± 0.2) x 10^4	Mn (1.0); Zn (0.6)
BSU29620[a]	*Bacillus subtilis* subsp. subtilis str. 168	0.37 ± 0.01	0.10 ± 0.01	(3.7 ± 0.6) x 10^3	Fe (2.6); Zn (0.4)
BH3206	*Bacillus halodurans* C-125	28 ± 1	0.64 ± 0.08	(4.4 ± 0.5) x 10^4	Mn (1.1); Zn (0.3)
GK2799	*Geobacillus kaustophilus* HTA426	7.6 ± 1.4	8 ± 3	(9.0 ± 1.1) x 10^2	Mn (1.0); Zn (0.4); Fe (0.3)
SMU_1486c	*Streptococcus mutans* UA159	0.31 ± 0.01	0.58 ± 0.07	(5.3 ± 0.8) x 10^2	Zn (0.4); Fe (0.3)

[a]Enzymes were over-expressed without the addition of 2,2′-bipyridyl in the growth medium.

DISCUSSION

Tri-nuclear Metal Center. Histidinol phosphate phosphatase from the PHP family of proteins has three divalent cations in the active site (**Figure 2.6**). The recombinant enzyme expressed in *E. coli* has a propensity for the binding of iron, which in the ferrous oxidation state could occupy any of the three metal binding sites. However, the third metal site may have a higher affinity for zinc since the crystal structures of HPP that are available from *T. thermophilus* (PDB: 2yxo, 2yz5 and 2z4g) and *L. monocytogenes* (PDB: 3dcp) reportedly bind iron in the α- and β-sites and zinc in the γ-site in all but one monomer. In the investigation reported here, the catalytic power of the enzyme obtained from the iron-free expression conditions is substantially higher than it is when iron is bound in the active site. It is assumed that oxidation of iron to the ferric oxidation state is responsible for the loss of catalytic activity when iron is bound in the active site.

The three divalent metal ions in the active site of HPP play crucial roles in the recognition and hydrolysis of phosphorylated substrates. The high resolution crystal structure of HPP reveals that three out of the four oxygen atoms of the phosphate product function as direct ligands to the three metal ions in the active site. The fourth oxygen atom of the bound phosphate forms polar interactions with the guanidino side chains of Arg-160 and Arg-197. Nearly identical interactions are found in the complex with the adduct formed by arsenate and histidinol. The α- and β-metal ions also anchor the hydroxide/water molecule that is used as the nucleophile during the hydrolysis of substrate. In the structure determined with arsenate/histidinol adduct, the bridging

hydroxide is 2.7 Å from the arsenic atom. This complex mimics the Michaelis complex where the hydroxide is poised to attack the phosphorus center of the substrate. The angle of attack, defined by the geometric positioning of the bridging hydroxide, central arsenic core, and leaving-group oxygen, is ~166^{o}. In addition to these interactions, the oxygen atom of the leaving group alcohol in the arsenate/histidinol complex is 2.2 Å away from the γ-metal ion and thus, these metal ion likely functions as a Lewis acid for activation of the substrate during hydrolysis (**Figure 2.8B**).

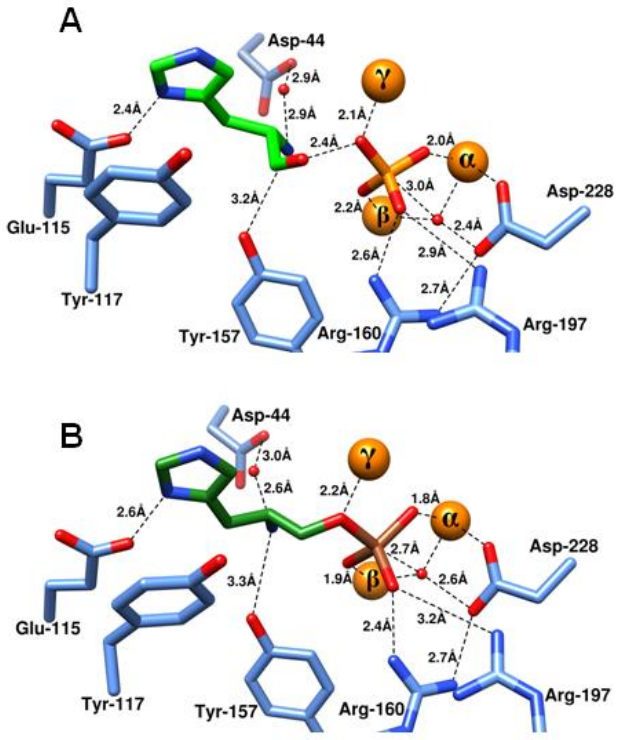


Figure 2.8. Product and inhibitory complexes in the active site of HPP. Zinc ions are presented as orange spheres while the enzyme residues are shown in blue. (A) Inorganic phosphate and L-histidinol bound at the active site of HPP from *L. lactis*. Inorganic phosphate is shown in orange and L-histidinol is shown in light green. (B) L-histidinol-arsenate ester, a substrate mimic, bound at the active site of HPP from *L.lactis*. Arsenate is shown in brown while L-histidinol is shown in dark green.

Proposed Mechanism of Action. The proposed mechanism of action for the hydrolysis of phosphate esters by HPP is presented in **Scheme 2.2**. In this scheme the substrate, histidinol phosphate, binds in the active site in a manner very much like the arsenate/histidinol adduct (**Figure 2.8B**). The phosphate moiety of the substrate forms a bridging complex with the α- and β-metal ions. The remaining non-bridge phosphoryl oxygen of the substrate makes electrostatic interactions with the side-chain guanidine groups of Arg-160 and Arg-197 that serve to make the phosphorus center more electrophilic. In addition to these interactions the leaving group oxygen of the substrate is activated by an additional contact with the γ-metal. The hydroxide that bridges the α- and β-metal ions attacks the phosphorus center of the bound substrate. This reaction is facilitated by the hydrogen bonding interaction between the bridging hydroxide and the side chain carboxylate of Asp-228. The proton from the hydroxide may be transferred to Asp-228 but this residue appears too far from the leaving group oxygen of histidinol to function as the subsequent general base. In any event, cleavage of the phosphorus-oxygen bond is enhanced by the direct electrostatic interaction of the γ-metal as a Lewis acid with the leaving group oxygen of the substrate. The most likely source of the general acid that ultimately donates a proton to the leaving group alcohol of the product is the primary amino group of the substrate itself as shown in **Figure 2.9**. The substrate amino group forms a hydrogen bond with water W2 in both structures. In the arsenate/histidinol complex the distance between the amino group and the oxygen from histidinol is ~3.1 Å. Another, less likely, possibility is the water molecule, W3, which is hydrogen bonded to Arg-197; this water was not observed in HPP.Zn.HOL.HPO_4.

However, mutation of Arg-197 to methionine reduced the atalytic activity of this enzyme by only about an order of magnitude.

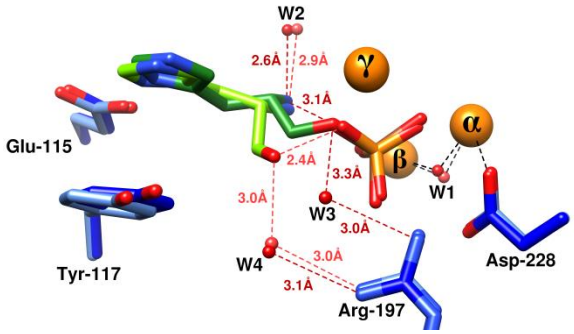
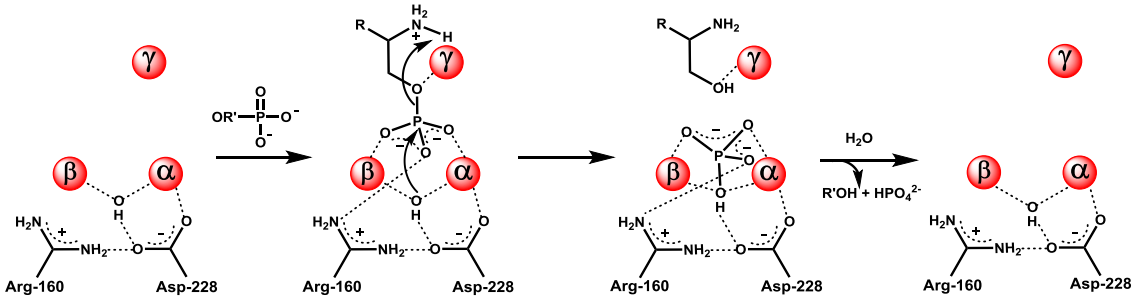


Figure 2.9. Overlay of the two structures HPP.Zn.HAR and HPP.Zn.HOL.HPO_4. The former is shown in dark shades while the latter is shown in lighter shades of colors. Metal ions are shown in orange, enzyme residues in blue, L-histidinol shown in green, water/ hydroxide shown in red while arsenate and phosphate are shown in brown and orange respectively.



Scheme 2.2. Proposed reaction mechanism of L-histidinol phosphate phosphatase from the PHP family of proteins.

Support for this proposed mechanism comes from the pH-rate profiles, catalytic properties of the active site variants, and the three-dimensional structure of the

arsenate/histidinol adduct. In the pH-rate profile for k_{cat}, there is a loss of activity at low pH with an apparent pK_a of 7.5. This ionization likely reflects the protonation of the bridging hydroxide to water. Most members of the amidohydrolase superfamily with a binuclear metal center in the active site also show a loss of activity at low pH that has been attributed to the protonation of the bridging hydroxide.[20, 58] In the pH-rate profile for k_{cat}/K_m, there is the loss of activity at high pH with an apparent pK_a of 9.7. In general, changes in activity observed in pH-rate profiles for k_{cat}/K_m reflect ionizations in free enzyme and/or free substrate and in enzyme/substrate complexes up to, and including, the first irreversible step.[59] In the active site of HPP the most likely group that accounts for this ionization is the primary amino group of the substrate. In the mechanism proposed in **Scheme 2.2**, this group ultimately serves as the general acid during the protonation of the leaving group oxygen of the substrate. An alternate mechanism may hold true, in which the P-O bond cleavage results in the formation of an alkoxide anion stabilized by the γ-metal, a Lewis acid, followed by subsequent protonation by a general acid. The best candidate for the role of general acid is the amino group of the substrate again, based on the crystallographic evidence. In the latter mechanism, the loss of activity at high pH in the k_{cat}/K_m profile may be attributed to the non-productive binding of the substrate at the active site, as the neutral amino group may coordinate to the γ-metal and decrease its Lewis acidity toward the bridging O-atom of the phosphate and/or distort the coordination geometry of the bound substrate.

Mutation of residues that bind the third metal ion, Asp-17, His-42, and His-230 leads to substantial losses in catalytic activity (**Table 2.2**). The other metal binding

residue mutated for this investigation, Asp228, results in a significant loss in catalytic activity. This residue has been proposed to help orient the bridging hydroxide and may also function in a proton transfer reaction. The mutation of residues that interact directly with the substrate, Glu-115, Arg-160, and Arg-197 produced enzymes that are less active than the wild-type enzyme but the mutation of Arg-160 had the biggest overall effect on k_{cat}/K_m. The origin of the rate limiting step in this reaction mechanism is not clear. The solvent deuterium isotope effects on either k_{cat} or k_{cat}/K_m are relatively small (1.3 to 1.4) and thus the proton transfer steps do not appear to limit the rate of the enzyme-catalyzed transformation. Relatively small effects are also found when the solvent viscosity is increased. This leaves an undefined conformational change as the most likely rate limiting step in the overall reaction mechanism.

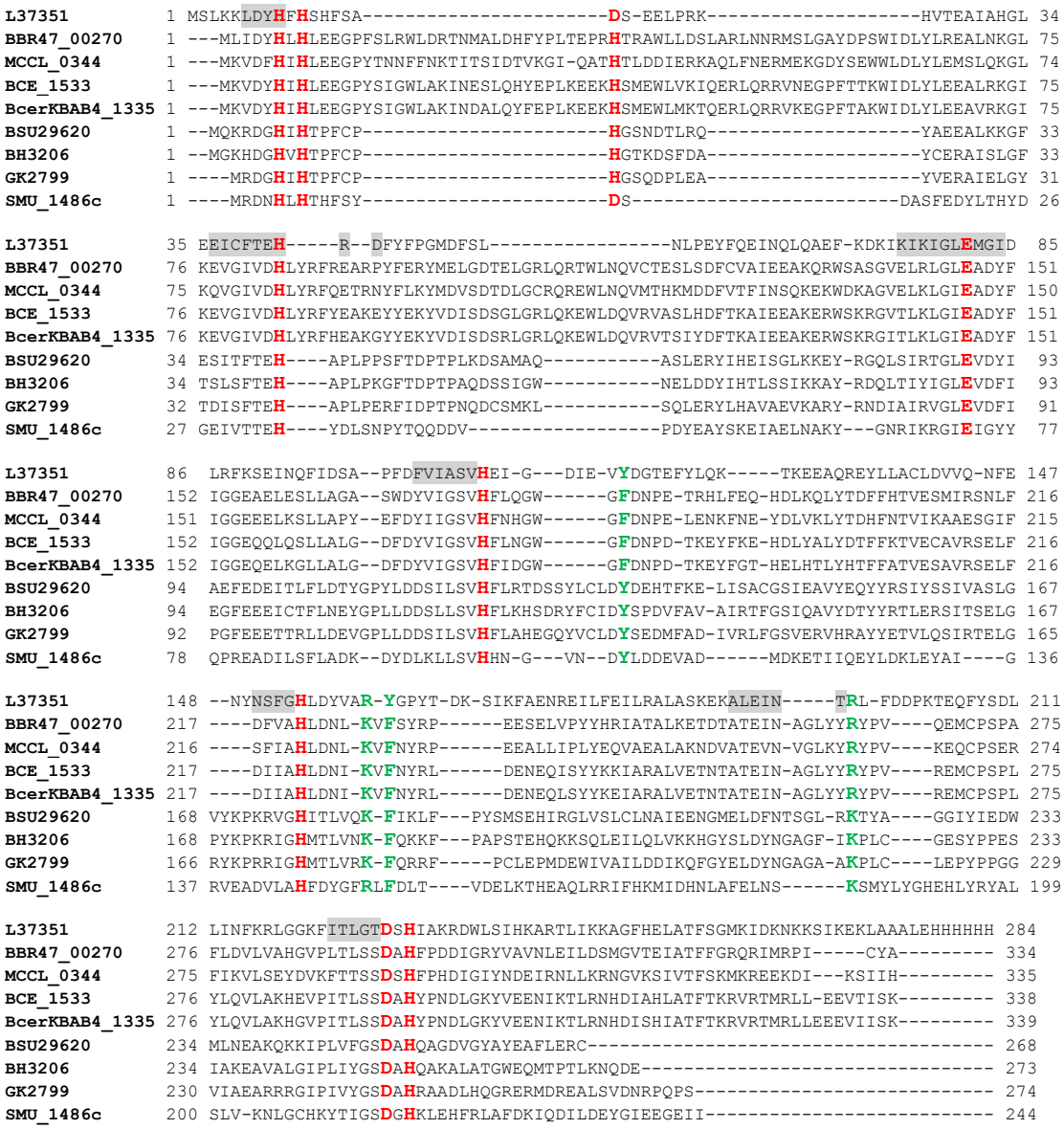


Figure 2.10. Primary sequence alignment of HPPs from various organisms. L37351 is the HPP from *L.lactis* subsp. lactis Il1403 characterized in this study. The organisms from which the other HPPs are derived are BBR47_00270: *Brevibacillus brevis* NBRC 100599, MCCL_0344: *Macrococcus caseolyticus* JCSC5402, BCE_1533: *Bacillus cereus* ATCC10987, BcerKBAB4_1335: *Bacillus weihenstephanensis* KBAB4, BSU29620: *Bacillus subtilis* subsp. subtilis str. 168, BH3206: *Bacillus halodurans* C-125, GK2799: *Geobacillus kaustophilus* HTA426, SMU_1486c: *Streptococcus mutans* UA159. The primary sequence of L37351 shown here is the actual sequence of the recombinant protein characterized in this study while the remaining sequences are from the NCBI database. Metal-binding residues are shown in red while the conserved residues interacting with the substrate are shown in green. The β-sheets observed in the available crystal structures of the three HPP enzymes are highlighted in gray.

Boundaries of HPP Orthologs. An amino acid sequence alignment of HPP from *L. lactis* and the eight other HPP enzymes that were characterized for this investigation is presented in **Figure 2.10**. Also included in this alignment are the two other HPP enzymes for which three-dimensional crystal structures are available. All nine of the residues that have been shown to coordinate one of the three divalent cations in the active site of HPP are conserved (highlighted in red). The two arginine residues in HPP from *L. lactis* that interact with the phosphate moiety of the substrate (Arg-160 and Arg-197) are also fully conserved with either a lysine or arginine at these positions, including the two structures available from *L. moncytogenes* and *T. thermophilus*. The two tyrosine residues (Tyr-117 and Tyr-157) in the active site that are located near the aromatic imidazole moiety of the substrate in the active site of HPP from *L. lactis* are also conserved with either a tyrosine or phenylalanine with the single exception of the enzyme from *T. thermophilus*. Two other residues near the histidinol binding site in the structure of HPP from *L. lactis* (Asp-44 and Glu-115) are not conserved in any of the HPP enzymes interrogated in this investigation and thus these two residues cannot be universally utilized by those enzymes that catalyze the hydrolysis of histidinol phosphate.

The conserved residues in the active site of HPP were used to identify additional sequences that likely code for proteins that catalyze the hydrolysis of histidinol phosphate. The critical residues include the nine amino acids that are responsible for the coordination of the three divalent cations. In addition to these residues there must be arginine and lysine residues shortly after the C-terminal ends of β-strands **6** and **7** and a

tyrosine or phenylalanine after the lysine/arginine at the end of β-strand **6**. These criteria were used to define the boundaries of HPP activity in those sequences represented in the sequence similarity network of **Figure 2.2.** The protein sequences that are represented by gray nodes in **Figure 2.2** lack one or more of these residues necessary for HPP activity and hence they cannot at this time be annotated as histidinol phosphate phosphatases. The physiological substrates of these proteins remain unknown. The enzymes represented by blue and green nodes are expected to catalyze the hydrolysis of histidinol phosphate to L-histidinol and phosphate. The enzymes represented by red nodes possess all the conserved residues of authentic HPP enzymes and are capable of hydrolyzing histidinol phosphate *in vitro* (see MCCL_0344, labelled 2 in **Figure 2.2**). However, the *in vivo* functions of these enzymes are not clear, as several genes that are apparently essential for the biosynthesis of L-histidine are absent in these organisms.

CONCLUSION

In summary, we have probed the enzymatic reaction mechanism of histidinol phosphate phosphatase from the PHP family of proteins, and attempted to establish the boundaries of this function amongst closely related protein sequences in cog 1387. These enzymes possess a trinuclear active site, with the hydroxide ligand bridging the α- and β-metal ions, serving as the nucleophile. The γ-metal ion serves as a Lewis acid by coordinating the bridging O-atom of the substrate L-histidinol phosphate. The solved crystal structures of *L.lactis* HPP with bound ligands, in conjunction with the sequence

alignment of selected HPP enzymes, helped identify the conserved residues involved in binding the metal cofactors and determining the substrate specificity. This study will help facilitate a better understanding of the reaction mechanism and substrate profiles of other enzymes in cog 1387, which include the PHP domain of polymerase X, and YcdX from *E. coli*.

CHAPTER III

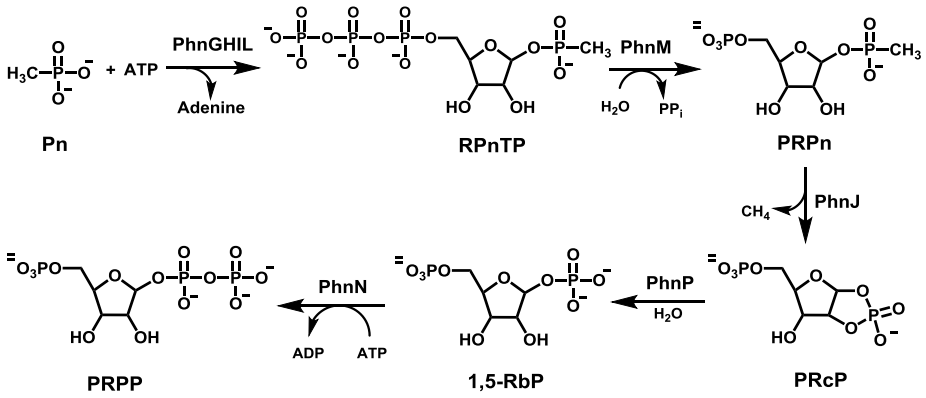
DISCOVERY OF A CYCLIC PHOSPHODIESTERASE THAT CATALYZES THE SEQUENTIAL HYDROLYSIS OF BOTH ESTER BONDS TO PHOSPHORUS*

In the absence of phosphate many bacteria can utilize organophosphonates (Pn) as a source of phosphorus.[60] In *Escherichia coli*, the metabolic machinery needed to convert organophosphonates to phosphate is governed by the 14-cistron *phnCDEFGHIJKLMNOP* operon.[61] The enzyme complex (C-P lyase) that functions to catalytically cleave the hydrolytically stable carbon-phosphorus bond of organophosphate substrates is encoded by the genes *phnGHIJKLM*. The proteins encoded by *phnCDEF* are required for the transport of phosphonate substrates while the remaining genes, *phnNOP*, are needed for the expression of three accessory enzymes. Recently, the metabolic pathway for the conversion of methyl phosphonate to 5-phospho-D-ribose-1,2-cyclic phosphate (PRcP) and methane by the C-P lyase complex in *E. coli* has been elucidated.[62] The key enzyme in this transformation is PhnJ, which converts α-D-ribose-1-phosphonate-5-phosphate (PRPn) to 5-phospho-D-ribose-1,2-cyclic phosphate (PRcP) as illustrated in **Scheme 1**.[63]

---

*Reprinted with permission from "Discovery of a cyclic phosphodiesterase that catalyzes the sequential hydrolysis of both ester bonds to phosphorus" by S. V. Ghodge, J. A. Cummings, H. J. Williams, and F. M. Raushel, *J. Am. Chem. Soc.* **2013**, *135* (44), pp16360-16363. Copyright 2013 American Chemical Society.

PRcP is subsequently hydrolyzed to D-ribose-1,5-bisphosphate (1,5-RbP) by PhnP. PhnP is a phosphodiesterase from cog1235, which is a subset of the metallo β-lactamase superfamily I enzymes.[64] The product of this reaction is then converted to 5-phosphoribosyl-1-pyrophosphate (PRPP) by the action of PhnN.[65] PhnO is an accessory enzyme, which has been shown to acetylate 1-aminoalkylphosphonic acids by acetyl CoA.[66]



Scheme 3.1. C-P lyase pathway in *E. coli.*

In an effort to further elucidate the metabolic complexities of bacterial phosphonate metabolism, we discovered that while many bacteria have a set of genes similar to that found in *E. coli* for the utilization of organophosphonates, a small cluster of organisms lack the specific gene required for the hydrolysis of PRcP; *i.e.* a homolog to PhnP. Instead, some of these bacteria possess an enzyme of unknown function from cog0613 that belongs to the polymerase and histidinol phosphatase (PHP) family of

proteins within the amidohydrolase superfamily (AHS). The structurally characterized members of the PHP family possess a distorted $(\beta/\alpha)_7$-barrel protein fold and contain a trinuclear metal center in the active site.[35, 67] The genetic architecture for a subset of these organisms, relative to that found in *E. coli*, is highlighted in **Figure 3.1**. The mechanistically characterized enzymes from the PHP family hydrolyze phosphomonoesters by activating the nucleophilic water molecule with two divalent cations, while the third metal acts as a Lewis acid to the departing alcohol leaving group.[68] It was therefore anticipated that the uncharacterized enzyme from cog0613 would act to hydrolyze PRcP to 1,5-RbP via a reaction that was homologous to that catalyzed by PhnP.

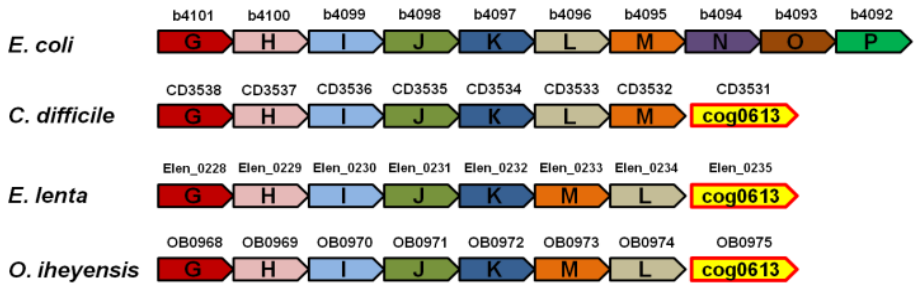


Figure 3.1. Schematic representation of organization of open reading frames encoding enzymes that constitute the C-P lyase pathway in select organisms.

MATERIALS AND METHODS

All chemicals were purchased from Sigma-Aldrich, unless indicated otherwise. Genomic DNA for *Eggerthella lenta* strain 1899B (ATCC25559) was obtained from American Type Culture Collection (ATCC). Pfu Turbo Polymerase, T4 DNA ligase and restriction enzymes were procured from New England Biolabs. DNA primers and Big Dye were obtained from Integrated DNA Technologies (IDT). The vector pET30a(+) was purchased from EMD4 Biosciences. *E. coli* BL21(DE3) and XL1 Blue competent cells were obtained from Stratagene. The P_i Colorlock Gold kit for the determination of inorganic phosphate was procured from Innova Biosciences. NMR spectra were recorded using D_2O (Cambridge Isotope Laboratories, Inc.) as the lock solvent, and a Bruker Avance III 400 MHz spectrometer with a BB probe for ^{31}P-NMR or a Bruker Avance III 500 MHz NMR spectrometer with a HCN probe for ^{1}H and ^{13}C-NMR. ^{1}H chemical shifts were measured relative to D_2O as an internal standard while ^{31}P NMR chemical shifts were recorded using 85% phosphoric acid as an external standard.

Cloning, Expression and Purification of Elen0235. The gene of interest was amplified using PCR with the following forward and reverse primer sequences respectively: 5′-GCAGGAGCCATATGATCGAAGATCTTCACGTGCATTCG-3′ and 5′-GCGCCTCGAGCGCCGGAACGCGAAACCCCACG-3′. The recognition sequences for restriction enzymes are underlined. The PCR product was purified using QIAquick (Qiagen) PCR purification kit, and double digested alongside vector pET30a(+) using *Nde*I and *Xho*I restriction enzymes. The double digested products were purified using agarose gel purification with QIAquick (Qiagen) Gel purification kit, and

ligated using T4 DNA ligase. Gene sequencing was used to verify the fidelity of the PCR amplification.

The recombinant plasmid was expressed and purified according to the procedure described elsewhere.[68] In brief, the ligated plasmid was transformed into *E. coli* BL21(DE3) cells by electroporation. 5 mL LB broth cultures containing 50 μg/mL kanamycin were inoculated with single colonies, and grown overnight. These were used to inoculate one liter of the same medium at pH 7.2, and were allowed to grow at 37 °C till the OD_{600} reached 0.15-0.2. The temperature was reduced to 30 °C and the iron-specific chelator 2,2′-bipyridyl was added to a final concentration of 150 μM. 1 mM $MnSO_4$ was added when the OD_{600} reached ~0.4, and 0.25 mM isopropyl D-thiogalactopyranoside (IPTG) when the OD_{600} reached ~0.6. The temperature was then lowered to ~15 °C, and the cells were shaken for ~16 more hours, after which, they were harvested by centrifugation and stored at -80 °C till used. The protein thus obtained had a C-terminal 6xHis-tag and its purification was carried out at 4 °C using Ni-NTA affinity chromatography. Four to five grams of cells were thawed and resuspended in ~40 mL of binding buffer (20 mM HEPES, pH 7.6, 5.0 mM imidazole, and 500 mM NaCl) containing 0.5 mg/mL of the protease inhibitor phenylmethanesulfonyl fluoride (PMSF). Cells were lysed by sonication and then ~2 mg/g of cells of bovine pancreatic DNase I was added and the suspension was stirred for 15 minutes after the final sonication cycle. After centrifugation, the supernatant solution was filtered and passed through a Ni-NTA column equilibrated with the binding buffer. The column was washed with 5 column volumes of wash buffer (20 mM HEPES pH 7.6, 50 mM imidazole, and 500 mM NaCl).

The protein was eluted with a solution containing 15 mM HEPES pH 7.6, 250 mM imidazole, and 350 mM NaCl. Fractions were analyzed using UV absorbance at 280 nm. The eluted protein was dialyzed against a solution containing 20 mM HEPES pH 8.0, 200 mM NaCl. The protein concentration was determined by UV absorbance at 280 nm using a calculated molar extinction coefficient of 20,775 M^{-1} cm^{-1} and molecular weight of 29.68 kDa. The protein was flash frozen using liquid nitrogen and stored at -80 °C.

Cloning, Expression and Purification of PhnP from *E. coli*. The PCR amplification was done from the genomic DNA of *Escherichia coli* K-12 strain using the following primer sequences: 5′-AAAAAAGGATCCATGAGCCTGACCCTCACGCTCACCGG -3′ as forward primer and 5′- AGAAGAAAGCTTTCACGCCACCCCAATCTCCATCCCATCAAACC -3′ as reverse primer. The primer pair contained restriction sites for *Nde*I and *Hind*III respectively, and the primer design facilitated the overexpression of gene product with N-terminal 6xHis-tag. The PCR product was purified with a PCR clean-up kit (Qiagen), double digested using *Bam*HI and *Hind*III, and ligated into a pET30a(+) vector previously double digested with the same set of restriction enzymes. The circular plasmid thus obtained was sequenced.

The plasmid was transformed into *E. coli* BL21(DE3) cells by electroporation. Single colonies were used to inoculate 5 mL overnight LB cultures containing 50 μg/mL kanamycin. These cultures were then transferred to 1 liter cultures of the same medium and cells were allowed to grow at 37 °C. Protein overexpression was induced by adding 0.5 mM IPTG when the OD_{600} reached ~0.6 and the temperature in the shaker-incubator

was reduced to ~20 °C. The cells were allowed to grow for additional 12-14 hours, and were isolated by centrifugation. These were then stored at -80 °C. The protein purification was carried out at room temperature using ÄKTA FPLC system (GE Healthcare) and HisTrap FF column (GE Healthcare). Cells (~5 g) were thawed and resuspended in ~ 35 mL binding buffer (50 mM HEPES pH 8.2, 150 mM NaCl, 20 mM imidazole). Cells were lysed by adding 1X BugBuster protein extraction reagent (Novagen). The cell debris was separated by centrifugation; the supernatant was filtered and passed through HisTrap (5 mL) column. The column was further washed with 8 column volumes of binding buffer, and eluted using a gradient of imidazole concentration from 20-500 mM (Elution buffer: 50 mM HEPES pH 8.2, 150 mM NaCl, 500 mM imidazole) over 20 column volumes. The excess imidazole was removed by dialysis against 20 mM HEPES pH 8.0, 150 mM NaCl, 1 mM $ZnCl_2$. The enzyme thus obtained was flash frozen using liquid nitrogen and stored at -80 °C. Enzyme concentration was determined using calculated values for molar extinction coefficient and molecular weight of 33,835 $M^{-1}cm^{-1}$ and 33.39 kDa respectively.

Synthesis of 5-phospho-D-ribose-1,2-cyclicphosphate (PRcP). The substrate for PhnP and Elen0235, PRcP, was synthesized by a slight modification of procedure used by *Hove-Jensen et al.*[65] In brief, ~17 mg 5-phosphoribosyl-1-pyrophosphate (PRPP) was dissolved in 1 mL water and barium acetate was added to a final concentration of 15 mM. The pH was adjusted to ~11.0 using 1 M NaOH and the reaction was monitored by ^{31}P-NMR spectroscopy. The conversion was complete in less than 48 hours. The product was contaminated with inorganic phosphate, and another

unidentified phosphorus containing impurity from PRPP. These were separated by anion exchange chromatography with ResourceQ-1mL column (GE Healthcare) using a gradient of 4-400 mM ammonium bicarbonate in water over 45 column volumes. The fractions containing PRcP were identified by incubating a small aliquot of the fractions with or without Elen0235 in the presence of 20 mM HEPES pH 8.0, 100 mM NaCl for one hour. The inorganic phosphate content was determined and the fractions containing inorganic phosphate released in the presence of Elen0235 were pooled together. The aqueous buffer was removed *in vacuo* using speedvac. Complete removal of ammonium bicarbonate was ensured by repeated cycles of redissolving the residue in water and drying it (500 μL x 3). The product was characterized by ^{31}P-NMR spectroscopy and ESI-MS. Mass, calculated for $C_5H_9O_{10}P_2$ = 290.9671 Da (M-H), found 290.9689.

Synthesis of ribose-2,5-bisphosphate (2,5-RbP). This synthesis was carried out according to the procedure described by Fathi and Jordan.[69] Impure or pure PRcP was hydrolyzed in the presence of 0.1 N HCl. The reaction was monitored by ^{31}P-NMR spectroscopy. After the peak corresponding to 1,2-cyclic phosphate in PRcP disappeared completely, 50 mM HEPES buffer was added to the reaction mixture and the pH was adjusted to ~8.5. 2,5-RbP was then purified by anion exchange chromatography as described for PRcP above. Product was characterized by NMR spectroscopy and ESI-MS. Mass, calculated for $C_5H_{11}O_{11}P_2$ = 308.9777 Da (M-H), found 308.9791.

Reaction Samples Analyzed by ^{31}P-NMR. Details of procedures used for carrying out reactions, referred to in the manuscript, are described here.

I) Reaction of PhnP and Elen0235 with PRcP. The substrate PRcP (1.5 mM and 3 mM of PRcP respectively) was separately incubated with 1 μM each of *E. coli* PhnP and Elen0235. The reaction mixture also contained 50 mM HEPES pH 8.5 and 0.1 M NaCl. After ~15 minutes of incubation, the reaction was quenched by addition of ~0.5 mM EDTA, and passing the reaction mixture through 10 kDa centrifugal filters. D_2O was added to a final concentration of ~17% and the reaction mixtures were immediately flash frozen using liquid nitrogen and stored at -80 °C. The reaction mixtures were thawed just before NMR analysis and warmed to room temperature in a water bath. This protocol of quenching and storage was adopted for all samples described in this study. If the reaction mixture of PRcP and Elen0235 was kept at room temperature for >30 minutes, before or after enzyme removal, 1-2 additional minor resonances developed downfield of R-5-P resonance, indicating some sort of non-enzymatic conversion/degradation of R-5-P. The same phenomenon was also observed with solution of pure R-5-P made using commercially available α-D-ribose-5-phosphate. Control reactions for PRcP with no enzyme added did not show any change in the ^{31}P-NMR spectrum for several hours, indicating that this compound is relatively stable under above conditions.

II) Verification of Products of PRcP Hydrolysis by Elen0235. 750 μL reaction mixture containing 50 mM HEPES pH 8.5, 0.1 M NaCl, 2 mM PRcP and 1 μM Elen0235 was incubated at room temperature for ~15 minutes. The reaction mixture was then passed through 10 kDa centrifugal filter to remove the enzyme. The mixture was then split into three aliquots of 250 μL each. The final NMR samples of 500 μL each

were made by adding 10 mM EDTA, 0.5 M HEPES pH 8.5, 1 M NaCl, and D_2O to final concentrations of 0.5 mM, 50 mM, 0.1 M, and ~20% respectively, besides water. In addition to these, 10 μL of 50 mM solution ribose-5-phosphate and sodium dihydrogen phosphate were added to the second and third aliquots respectively. All three samples were immediately flash frozen and stored at -80 °C till time for analysis by NMR spectroscopy.

III) Test of ribose-2,5-bisphosphate as Substrate for Elen0235. 3 mM 2,5-RbP and 1 μM Elen0235 were incubated for ~15 minutes along with 0.1 M NH_4HCO_3 pH 8.5 (NH_4OH) and 0.1 M NaCl, alongside 2,5-RbP control containing no enzyme. Reaction was quenched, diluted and frozen as described above, with the exception that NH_4HCO_3 was used instead of HEPES. Control showed no non-enzymatic hydrolysis.

IV) Test of ribose-1,5-bisphosphate as Substrate for Elen0235. 3 mM PRcP was incubated with 2 μM E. coli PhnP in a 500 μL reaction mixture containing 50 mM HEPES pH 8.5, 0.1 M NaCl for ~45 minutes. This mixture was then divided into two aliquots of 250 μL each, and Elen0235 (final concentration = 1 μM) was added to the second aliquot and the solutions were incubated at room temperature for ~45 minutes. Both solutions were then passed through 10 kDa centrifugal filters, followed by addition of 0.5 mM EDTA, 50 mM HEPES pH 8.5, 0.1 M NaCl, 20% D_2O and water and storage at -80 °C.

V) Hydrolysis of PRcP by Elen0235 in the Presence of $H_2{}^{18}O$. 3 mM PRcP and 2 μM Elen0235 were incubated in parallel reactions (500 μL each) containing 0.1 M NH_4HCO_3 pH 8.5 and 0.1 M NaCl, with one of them containing 56% $H_2{}^{18}O$ (Aldrich) by

volume. The reaction mixtures were incubated at room temperature for ~2 hours. The solutions were passed through 10 kDa filters to remove the enzyme, followed by addition of D_2O (~17%) and EDTA (~0.5 mM). Samples were analyzed on the same day without freezing.

Enzymatic Assay of Elen0235. The kinetic constants for the hydrolysis of PRcP and 2,5-RbP were determined using the P_i Colorlock Gold kit for measurement of free phosphate. The enzymatic reaction was conducted at 25 °C in 50 mM HEPES pH 8.0, 100 mM NaCl, 20 μM $MnCl_2$. A single assay consisted of 8-16 substrate concentrations with 3-4 time points taken over a period of 1-20 minutes. The color was allowed to develop for 30 minutes and the final absorbance was determined at 650 nm. The phosphate concentration of each well was calculated from a standard curve and the initial rates at each substrate concentration were determined using linear regression. The reagent contains 5 M HCl, making the final concentration of HCl to be ~1 M after quenching a reaction aliquot. PRcP was found to be labile under these conditions, and hence the stabilizer provided with the kit was added immediately (within 30 seconds) after quenching, instead of after 5 minutes, as recommended in the protocol provided by the manufacturer.

Data Analysis. Kinetic parameters, k_{cat} and k_{cat}/K_m, were obtained by fitting the initial velocity data to equation 1 using the nonlinear least-squares fitting program in SigmaPlot 10, where v is the initial velocity at substrate concentration [A], $[E_t]$ is the enzyme concentration, k_{cat} is the turnover number, and K_m is the Michaelis constant.

$$v/[E_t] = k_{cat}[A] / (K_m + [A]) \quad (3.1)$$

RESULTS AND DISCUSSION

The gene for Elen0235 (gi|257790010) was PCR-amplified by standard methods from the genomic DNA of *Eggerthella lenta* strain 1899B (ATCC 25559). The gene was subcloned into a high-copy plasmid pET30(a) to express the appropriate gene product with 6 x His-tag at the C-terminus in *E. coli*. The enzyme was purified to >95% homogeneity using Ni-NTA chromatography at 4 °C, and excess imidazole was removed by dialysis. Similarly, the recombinant enzyme for PhnP (gi|16131918) from *E. coli* was obtained and purified to homogeneity. The putative substrate, PRcP, was chemically synthesized from PRPP using a modification of a published procedure.[6] The reactions catalyzed by Elen0235 and PhnP were determined by incubating the purified enzymes with PRcP and the products of the reaction characterized by ^{31}P-NMR.

The ^{31}P-NMR spectrum of PRcP is presented in **Figure 3.2A**. The phosphate attached to the hydroxyl group at C5 resonates at 4.57 ppm, while the 1,2-cyclic phosphate resonates at 19.31 ppm. In the proton-coupled ^{31}P-NMR spectrum the cyclic phosphate appears as a doublet of doublets while the phosphate at C5 is a triplet. Shown in **Figure 3.2B** is the product of the reaction catalyzed by PhnP, D-ribose 1,5-bisphosphate. The phosphate attached to the hydroxyl group C5 resonates at 4.57 ppm, whereas the phosphate at C1 resonates at 3.02 ppm. In the proton-coupled spectrum the phosphate at C1 appears as a doublet and the phosphate at C5 appears as a triplet. All of the starting material has been consumed. The products of the reaction catalyzed by Elen0235 are clearly different from the reaction catalyzed by PhnP. The resonance that appears at 3.24 ppm is phosphate (a singlet in both the proton-coupled and decoupled

spectrum (**Figure 3.2C**). The resonance at 4.56 ppm is from the phosphate at C5 of D-ribose-5-phosphate. Clearly, both enzymes consume PRcP, but the products of hydrolysis for the two enzymes are different. The product of the reaction catalyzed by PhnP is D-ribose-1,5-bisphosphate (1,5-RbP) whereas the products of the reaction catalyzed by Elen0235 are D-ribose-5-phosphate and phosphate. The assignment of these products was confirmed by adding authentic D-ribose-5-phosphate and inorganic phosphate to the products of the reaction catalyzed by Elen0235 (**Figure 3.3**).

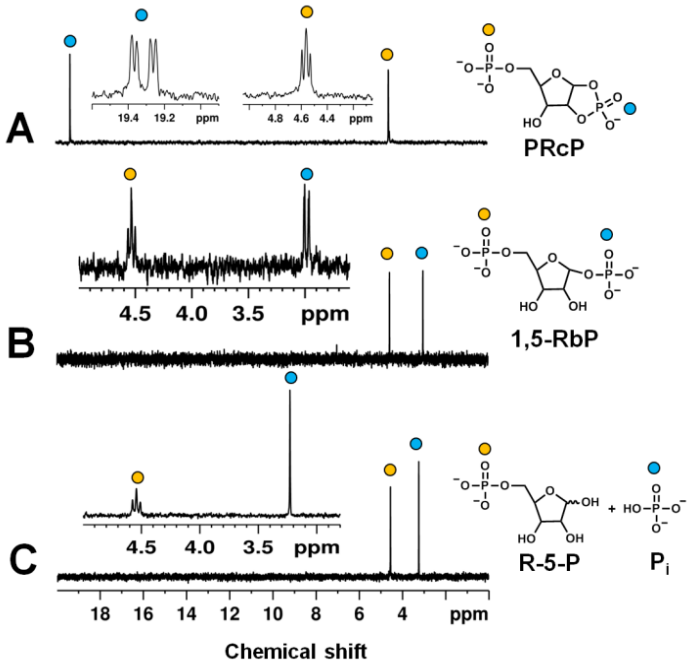


Figure 3.2. ^{31}P-NMR spectra of PRcP and products of the reactions catalyzed by PhnP and Elen0235 at pH 8.5. (A) 4 mM 5-phosphoribose-1,2-cyclic phosphate (PRcP). (B) Product of the enzymatic hydrolysis of 1 mM PRcP by PhnP from *E. coli*. (C) Products of the enzymatic hydrolysis of 2 mM PRcP by Elen0235. The insets show the respective ^{1}H-^{31}P coupled NMR spectra.

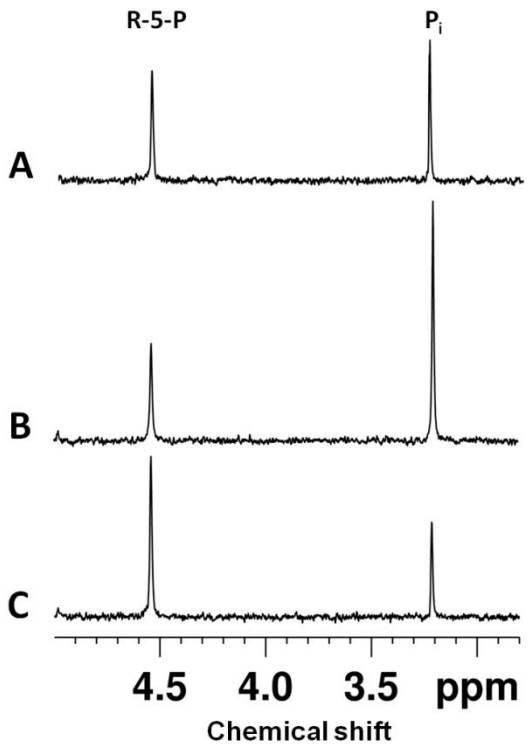


Figure 3.3. ^{31}P-NMR spectra for the verification of products of hydrolysis of PRcP by Elen0235. (A) Reaction products corresponding to 1 mM PRcP. (B) 'A' and added 1 mM NaH_2PO_4. (C) 'A' and added 1 mM α-D-ribose-5-phosphate.

Quite remarkably, the reaction catalyzed by Elen0235 with the cyclic phosphate substrate must involve two consecutive hydrolytic reactions. This observation thus requires that either D-ribose-1,5-bisphosphate (1,5-RbP) and/or D-ribose-2,5-bisphosphate (2,5-RbP) function as reaction intermediates. Ribose-1,5-bisphosphate was isolated as the product from the enzymatic hydrolysis of PRcP by PhnP. The other potential intermediate, ribose-2,5-bisphosphate, was chemically synthesized from PRcP by acid hydrolysis and then purified by anion exchange chromatography.[69] At alkaline and neutral pH, the resonances corresponding to the 2- and 5-phosphates of 2,5-RbP

largely overlap with one other (**Figure 3.4B**). Therefore, the pH was adjusted to 5.8 to enable the observation of two distinct phosphorus resonances. At this pH, 2,5-RbP shows four resonances (**Figure 3.4A**). The corresponding ^{1}H-^{31}P coupled NMR spectrum exhibits two pairs of doublets and triplets corresponding to the phosphate groups at C-2 and C-5 of the ribose ring, respectively. The relative amounts for the two species, based on the integral values from the ^{31}P-NMR spectrum, are 32% : 68%. These species represent the two anomers of 2,5-RbP since the relative percentages correspond reasonably well with the relative amounts of the α- and β-anomers of D-ribose-5-phosphate in solution (36% : 64%) determined using ^{13}C-NMR at pH 4.5.[12] When 1,5-RbP and 2,5-RbP were incubated with Elen0235 in separate reactions, there was no change in the ^{31}P-NMR spectrum of 1,5-RbP (**Figure 3.5**), while the reaction mixture with 2,5-RbP showed the formation of two new signals (**Figure 3.4C**). These two resonances coincided with the products of enzymatic hydrolysis of PRcP by Elen0235, and were confirmed to be D-ribose-5-phosphate and inorganic phosphate.

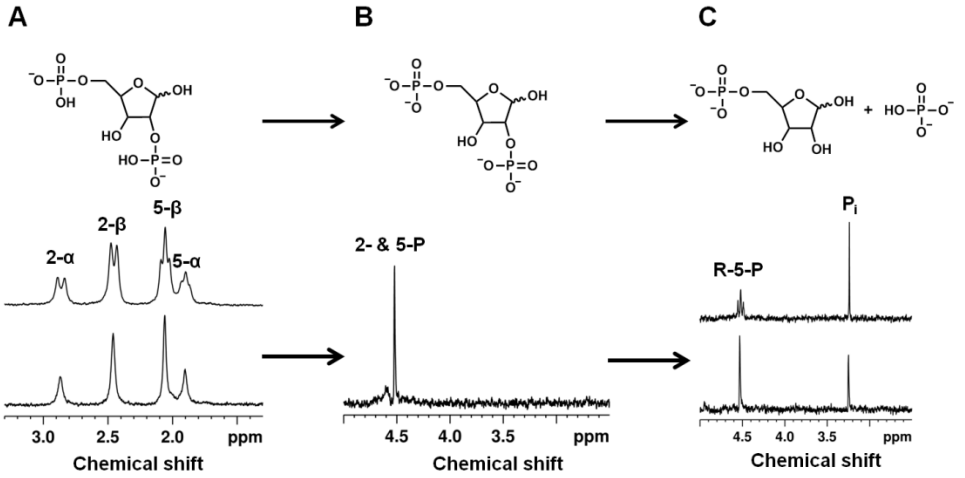


Figure 3.4. ^{31}P-NMR spectra of synthesized ribose-2,5-bisphosphate (2,5-RbP) and the products of its hydrolysis by Elen0235. (A) 4 mM 2,5-RbP at pH 5.8. (B) 2 mM 2,5-RbP at pH 8.5. (C) Products of hydrolysis of 2 mM 2,5-RbP by Elen0235 at pH 8.5. In panels A and C, ^{1}H-^{31}P coupled spectra are shown above ^{1}H-^{31}P decoupled spectra.

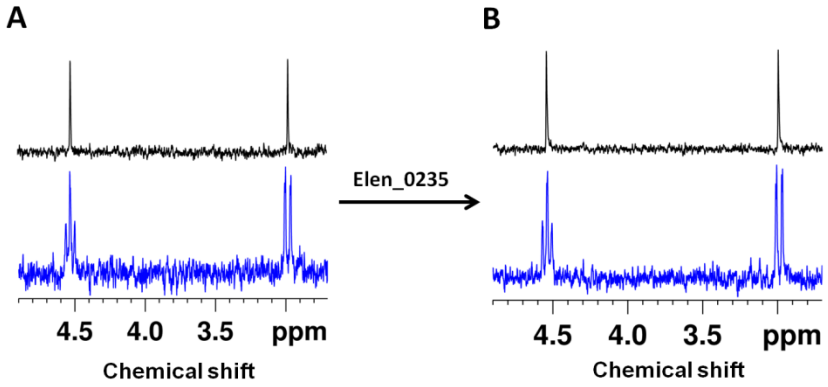


Figure 3.5: ^{31}P-NMR spectra demonstrating that Elen0235 cannot hydrolyze ribose-1,5-bisphosphate (1,5-RbP). (A) 1,5-RbP (1.5 mM) synthesized enzymatically by hydrolyzing PRcP with *E. coli* PhnP. (B) 1,5-RbP (1.5 mM) after incubation with Elen0235 (1 μM). ^{1}H-^{31}P coupled spectrum is shown in blue while decoupled spectrum is shown in black.

Enzymes from the PHP family have been shown previously to catalyze the hydrolysis of phosphomonoesters by nucleophilic attack of hydroxide at the phosphorus center.[68] With regard to the hydrolysis of 2,5-RbP by Elen0235, the reaction is expected to occur similarly. It is unclear, however, whether the initial phosphodiesterase activity with PRcP occurs by the attack of hydroxide on phosphorus or the anomeric carbon of the D-ribose ring. This issue was addressed by incubating PRcP with Elen0235 in the presence of 56% $H_2{}^{18}O$. It is known that there is a small chemical shift difference in the ^{31}P-NMR signal of phosphate esters when ^{18}O is substituted for ^{16}O.[70] For phosphate, the chemical shift difference is 0.021 ppm upfield for each ^{18}O substitution, and is additive.[71] If the hydroxide attacks at the phosphorus center in both hydrolytic steps of this reaction, then three separate resonances are expected for the inorganic phosphate product. The results are presented in **Figure 3.6** and the percentage of the phosphate species containing 0, 1, and 2 atoms of oxygen-18 are 22% : 49% : 29%. The expected percentages, based on the attack of two water molecules directly with the phosphorus center, are 19% : 49% : 31%. The results clearly demonstrate that both water molecules used in the overall reaction react directly with the phosphorus center.

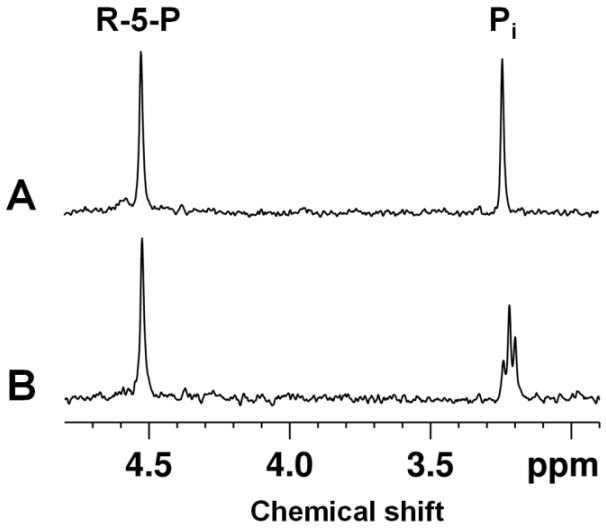


Figure 3.6. ^{31}P-NMR (^{1}H-decoupled) spectra of the reaction products from PRcP catalyzed by Elen0235 when the reaction is carried out in 100% $H_2{}^{16}O$ (panel A) and 56% $H_2{}^{18}O$ (panel B).

The kinetic parameters for the hydrolysis of PRcP and 2,5-RbP were determined by monitoring the rate of formation of inorganic phosphate. The apparent kinetic constants for the hydrolysis of PRcP at pH 8.0 and 25 °C are as follows: k_{cat} = 2.0 ± 0.2 s^{-1}; K_m = 14 ± 3 μM; and k_{cat}/K_m = 1.4 ± 0.1 x 10^5 M^{-1} s^{-1}. The kinetic constants for hydrolysis of 2,5- RbP by Elen0235 are as follows: k_{cat} = 7.4 ± 0.3 s^{-1}; K_m = 23 ± 2 μM; and k_{cat}/K_m= 3.3 ± 0.2 x 10^5 M^{-1} s^{-1}. The rate constants for the enzymatic hydrolysis of PRcP and 2,5-RbP by Elen0235 are thus very similar to one another. To determine whether or not Elen0235 catalyzed the hydrolysis of PRcP processively, 4.0 mM PRcP was incubated with 0.5 μM Elen0235 at pH 8.0 until the reaction proceeded to ~50% completion. The reaction was quenched by adding 0.5 mM EDTA, and the reaction mixture was passed through a 10 kDa molecular weight cut-off centrifugal filter to remove the enzyme. The pH of the filtrate was adjusted to 5.8 and the products of the

reaction were analyzed by ^{31}P-NMR spectroscopy to determine whether or not the intermediate, 2,5-RbP, was released into solution (**Figure 3.7**). The observed resonances were assigned by the sequential addition of authentic samples of PRcP, 2,5-RbP, R-5-P and P_i to the reaction mixture.

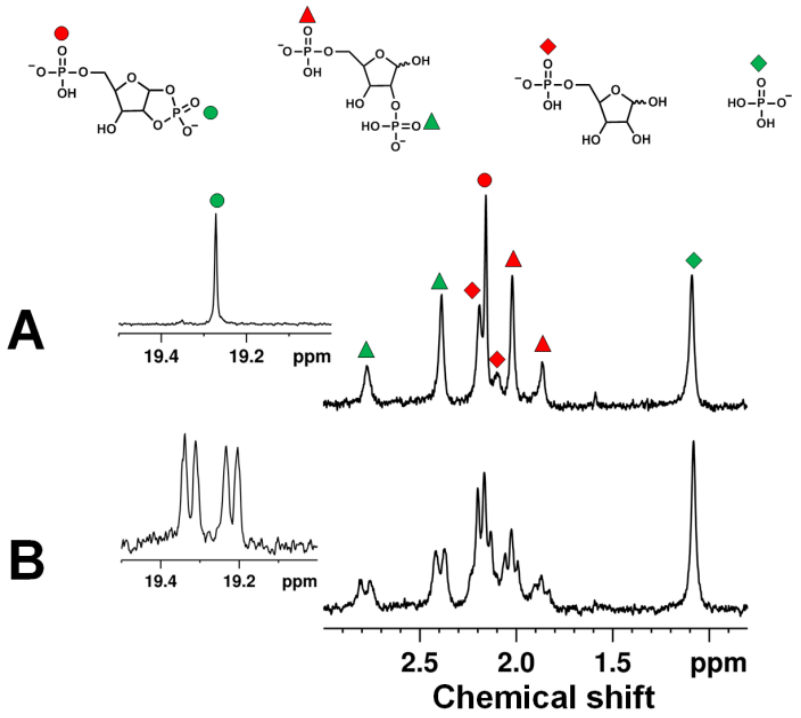
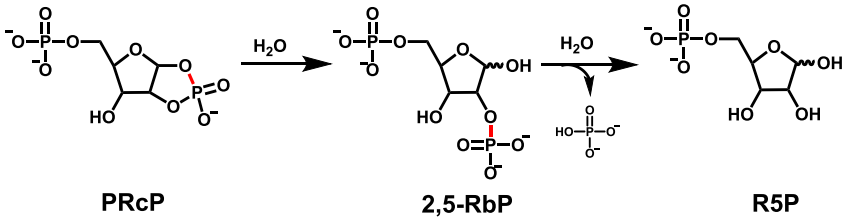


Figure 3.7. ^{31}P-NMR spectrum for the partial hydrolysis of 4 mM PRcP by 0.5 µM Elen0235. (A) ^{1}H-^{31}P decoupled spectrum. (B) ^{1}H-^{31}P coupled spectrum. The inset shows the resonance for the 1,2-cyclic phosphate group of the remaining substrate, PRcP.

The reaction products clearly show the formation of a significant amount of 2,5-RbP as a reaction intermediate. The approximate ratios of PRcP, 2,5-RbP and R5P,

calculated from the integration of the phosphorus resonances, are 22% : 39% : 39%. This observation demonstrates that the reaction mechanism is not processive and confirms that 2,5-RbP is a kinetically competent intermediate that is released from the enzyme site as the initial product of PRcP hydrolysis.

These findings raise several mechanistic questions. It is clear that this enzyme treats PRcP and 2,5-RbP as two distinct substrates of comparable affinity and efficiency. Moreover, the enzyme is able to distinguish between the P-O bonds at C1 of PRcP and 1,5-RbP. The former is rapidly hydrolyzed while hydrolysis of the latter could not be detected. This implies that the enzyme has sufficient specificity, but is surprisingly promiscuous in its ability to catalyze the hydrolysis of a phosphodiester and a phosphomonoester. There is no evidence to suggest that this is accomplished via separate active sites. The overall hydrolytic reaction from PRcP to 2,5-RbP to R-5-P and P_i is presented in **Scheme 3.2**.



Scheme 3.2. Overall reaction catalyzed by Elen0235.

Purple acid phosphatase (PAP) from pig liver and red kidney bean can catalyze the hydrolysis of ethyl *p*-nitrophenyl phosphate (EpNPP) to *p*-nitrophenol (pNP), inorganic phosphate, and ethanol.[15] PAP cannot hydrolyze bis-*p*-nitrophenyl phosphate (bis-pNPP) and the authors of this study state that the diesterase activity has no biological relevance. PAP has an Fe^{3+}-M^{2+} (M = Fe, Zn or Mn) cluster in the active site that apparently hydrolyzes the EpNPP diester in a processive manner. Elen0235 has all of the metal-binding residues that are present in other members of the PHP family and this enzyme should possess a trinuclear active site. The mechanism of hydrolysis can be envisaged similar to that proposed for L-histidinol phosphate phosphatase, another PHP family enzyme from cog1387. The α- and β-metal ions activate the nucleophilic hydroxide that bridges these two metal ions. The third metal ion (denoted as the γ-metal ion) serves as a Lewis acid by interacting with the oxygen of the leaving group alcohol. A protein sequence BLAST analysis and genomic context study of Elen0235 revealed that a limited number of organisms possess a homolog of this enzyme, as opposed to a majority of the organisms possessing the C-P lyase operon which have PhnP. A list of organisms that contain an enzyme that is predicted to hydrolyze PRcP to R5P and P_i is provided in **Table S5**. This raises the question why these organisms, which includes the highly virulent and multi-drug resistant human gut pathogen *Clostridium difficile* 630, have ribose-5-phosphate and inorganic phosphate as the terminal products of the C-P lyase pathway instead of PRPP. The enzyme PRPP synthetase converts R5P to PRPP by pyrophosphoryl transfer from ATP, and all the organisms that possess an ortholog to Elen0235 possess a copy of this enzyme.

CONCLUSION

In summary, we have discovered an enzyme that specifically hydrolyzes a cyclic phosphodiester to a vicinal diol and inorganic phosphate in a non-processive manner (**Scheme 3.2**). To the best of our knowledge, this is the first reported instance of any enzyme that can catalyze the successive hydrolysis of a phosphodiester and the resultant phosphomonoester with equal affinity and efficiency. We propose that this enzyme be called cyclic phosphate dihydrolase (cPDH) and be designated as PhnPP.

CHAPTER IV

DISCOVERY OF A 3′,5′-NUCLEOTIDE BISPHOSPHATE PHOSPHATASE WITHIN THE AMIDOHYDROLASE SUPERFAMILY*

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.[5] 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.[6, 57, 68] The catalytic mechanism of hydrolysis utilizes an activated water/hydroxide for the nucleophilic attack at an electrophilic carbon or phosphorus center. The enzymes in the AHS that possess a tri-nuclear 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.[34, 68] The available crystal structures from the PHP family indicate that these enzymes possess a distorted $(\beta/\alpha)_7$-barrel structural fold, as opposed to the distorted $(\beta/\alpha)_8$-barrel fold for most of the other members of the AHS.[35, 72]

---

*Reprinted with permission from "Prospecting for unannotated enzymes: discovery of a 3′,5′-nucleotide bisphosphate phosphatase within the amidohydrolase superfamily" by J. A. Cummings, M. Vetting, S. V. Ghodge, , C. Xu, B. Hillerich, R. D. Seidel, and F. M. Raushel, *Biochemistry* **2014**, *53* (3), 591-600. Copyright 2014 American Chemical Society.

The PHP family within the amidohydrolase superfamily consists of three Clusters of Orthologous Groups: cog0613, cog1387 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.[29, 73] 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. Further work has also elucidated the role of the tri-nuclear active site in the activation of the nucleophilic hydroxide, and Lewis acid stabilization of the alcohol product.[35, 68] The mechanistic details obtained from a greater 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 10^{-60} of 650 non-redundant protein sequences from cog0613 is presented in **Figure 4.1**.[38] There are currently three proteins from this COG whose crystal structures have been determined. These include PDB id: 3O0F (Locus tag: Bad1165 from *Bifidobacterium adolescentis*); PDB id: 2ANU (Locus tag: Tm0559 from *Thermotoga maritima*); and PDB id: 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.[74] The crystal structure of this enzyme has been solved 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-proofreading activity. 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 non-processive manner through D-ribose-2,5-bisphosphate as an obligatory reaction intermediate.[75]

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 identified in the SSN depicted in **Figure 4.1**. 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.

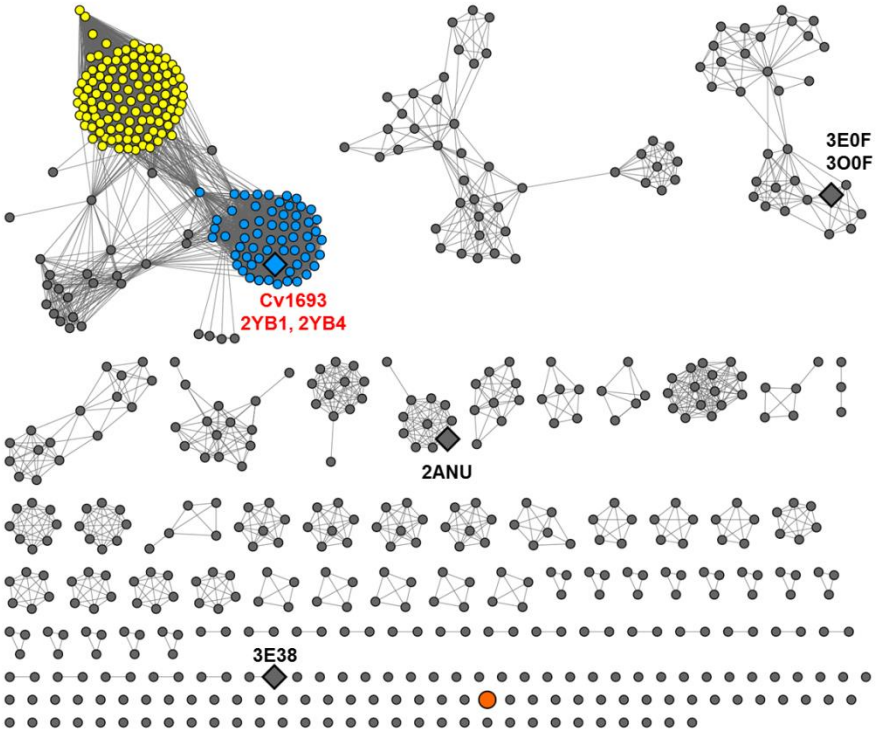


Figure 4.1. Sequence network map of cog0613 at an E-value of $1x10^{-60}$ created using Cytoscape (http://www.cytoscape.org). Each node (sphere) represents a protein sequence, while each edge (line) represents those sequence pairs that are more closely related than the arbitrary E-value cut-off (10^{-60}). The available crystal structures are shown as diamonds and their respective PDB codes are indicated. The enzyme studied in this work, Cv1693 from *Chromobacterium violaceum* ATCC 12472, and its orthologs are shown in blue. Enzyme sequences shown in yellow are those that have been annotated as trpH while the single orange node represents cyclic phosphate hydrolase from *Eggerthella lenta* DSM 2243 (Elen0235), the only other enzymatic reaction known and demonstrated from cog0613.

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′-phosphoguanylyl-(3′→5′)-guanosine (pGpG) were purchased from Biolog Life Science Institute (United States distributor Axxora). Guanosine-3′,5′-bis-diphosphate (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.*[76]

Cloning and Expression. A codon optimized synthetic gene for Cv1693 (gi|34497148) from *Chromobacterium violaceum* was inserted into a pET30 variant by ligation independent cloning (LIC), utilizing the synthetic gene as a template and 5′-TTAAGAAGGAGATATACCATGGCAAACATTGACCTGCATTTTCATTC-3′ and 5′-GATTGGAAGTAGAGGTTCTCTGCATCAGCCGGACGCAGGATACG-3′ as primers.[40] The resultant protein is identical to the native protein with the inclusion of a C-terminal tag (added sequence AENLYFQSHHHHHHWSHPQFEK), consisting of a tobacco etch virus (TEV) cleavage site, 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 auto-induction medium (25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM

Na_2SO_4, 25 mM sodium succinate, 2 mM $MgSO_4$, 0.5% yeast extract, 1% NZ-AMINE-AS, 0.9% glycerol, 0.45% lactose, 0.03% glucose) distributed into eight 2 L baffled flasks.[41] The flasks were shaken (250 rpm) at 30 °C for 4 hours (OD < 3) whereupon the temperature was reduced to 23 °C for an additional 16 hours. Following expression, the cells were isolated by centrifugation and stored at -80 °C. To increase the metal content of the isolated protein, additional expressions were performed that included 150 μM 2,2′-bipyridine, 1.0 mM $ZnCl_2$ and 1.0 mM $MnCl_2$ prior to reducing the temperature 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 cellular extract 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, with bound proteins eluted into 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 *w/w* 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 immersion in liquid nitrogen, and stored at -

80 °C. The purified protein contained 0.6 equivalents of Mn and 2.4 equivalents of Zn as determined by inductively-coupled mass spectrometry (ICP-MS).

Assay for AMP. Cv1693 hydrolyzes pApA to release 2 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 $MgCl_2$, 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 measured 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 68,000 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 of dAdAdAdAdAp, 3′-AMP, 3′-NADPH, CoA, 1,3-bisphosphoglycerol, 3′-phosphoadensosine-5′phosphosulfate (PAPS), pGpG and ppGpp for 75 minutes 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 protein that was derived from expressions that did not contain 2,2′-bipyridine/$ZnCl_2$/$MnCl_2$. 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, 10 mM HEPES, pH 7.8) with 20% Peg2000 MME, 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, 10 mM HEPES, pH 7.8) and 25% PEG3350, 200 mM $(NH_4)_2SO_4$, 100 mM BIS-TRIS, pH 6.5. For the Mn^{2+}/AMP complex, Cv1693 (10-15 mg/mL, 150 mM NaCl, 10% glycerol, 20 mM adenosine 5′-monophosphate, 5.0 mM $MnCl_2$, 10 mM HEPES, pH 7.8) was combined with 20% Peg2000 MME, 100 mM BIS-TRIS, pH 6.0. Prior to data collection, crystals were soaked in crystal growth conditions supplemented with 20% ethylene glycol and vitrified by immersion in liquid nitrogen. All data were collected in-house on an R-AXIS IV++ detector with CuKα X-rays generated by a Rigaku RU-H3R rotating anode generator and focused with OSMIC mirrors. Data were integrated and scaled with HKL3000.[77] Data statistics are listed in **Table 4.1**.

Table 4.1. Data collection statistics for Cv1693.

DATASET[a]	10 mM $HgCl_2$[b]	10 mM TMLA	Unliganded	Mn^{2+}, AMP
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit Cell (Å)	*a* = 40.0 *b* = 46.8 *c* = 130.4	*a* = 40.0 *b* = 46.8 *c* = 130.8	*a* = 40.1 *b* = 46.5 *c* = 130.5	*a* = 39.9 *b* = 46.7 *c* = 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)/sd(I)	20.8 (6.5)	21.6 (3.2)	10.3 (4.8)	12.9 (4.1)
R_{sym}	0.096 (0.276)	0.089 (0.243)	0.098 (0.264)	0.083 (0.301)

[a] Data in parenthesis is for the highest resolution bin.
[b] No binding, used as native dataset

Structure Determination. Data sets were collected from crystals soaked for 3 minutes in a cryo-solution with the addition of either 50 mM trimethyllead acetate (TMLA) or 10 mM $HgCl_2$. The structure was determined using single isomorphous replacement with anomalous scattering utilizing the TMLA dataset as the anomalously scattering derivative and the $HgCl_2$ dataset as "native". Heavy atom site determination, phasing and automatic chain tracing were performed by the program PHENIX.[46] The starting model was completed by iterative rounds of model building utilizing COOT[45] followed by refinement using PHENIX. There was a single molecule per asymmetric unit with a solvent content of 37%. Waters were added where difference map peaks (F_o

– F_c > 3.5 σ) were < 3.2 Å from potential hydrogen bonding donors. Translation-libration-screw 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.[49] Data collection and refinement statistics for Cv1693 are presented in **Tables 4.1** and **4.2**.

Table 4.2. Structure statistics for Cv1693.

Structure[a]	**Unliganded**	**Mn^{2+}, AMP**
Resolution (Å)	40-2.2 (2.42-2.2)	40-1.9 (2.0-1.9)
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^{3+}	0	3
Atoms total	2340	2438
Average B-factor ($Å^2$)		
Wilson B Factor ($Å^2$)	18.6	12.5
TLS groups (#)	3	4
Protein/Waters/Mn^{3+}/AMP	9.4 / 14.2 / - / -	11.5 / 21.1 / 9.0 / 13.1
RMSD		
Bond lengths (Å) /Angles (°)	0.006 / 1.05	0.007 / 1.13
MOLPROBITY STATISTICS		
Ramachandran Favored / Outliers (%)	98.9 / 0.0	98.9 / 0.0
Rotamer Outliers (%)	1.9	1.4

Table 4.2 Continued.

Structure[a]	Unliganded	Mn^{2+}, AMP
Clashscore [c]	7.1 (97th pctl.)	6.4 (93rd pctl.)
Overall score[c]	1.6 (97th pctl.)	1.5 (96th pctl.)
PDB ID	2YB4	2YB1

[a]Statistics in parenthesis are for the highest resolution bin.
[b]Additional residues past 285 are from C-terminal tag.
[c]Scores are ranked according to structures of similar resolution as formulated in MOLPROBITY.

Analytical Size Exclusion Chromatography. Proteins were analyzed by analytical size exclusion chromatography utilizing a 5 x 10 cm SUPERDEX 200 column (GE-Healthcare) using thyroglobulin (670 kD), δ-globulin (158 kD), ovalbumin (44 kD), myglobin (17 kD), and vitamin B12 (1.35 kD) as molecular weight standards (BioRad).

Bioinformatics. The protein sequences from various COGs were retrieved from the NCBI (http://www.ncbi.nlm.nih.gov) and Microbesonline (http://www.microbesonline.org) databases.[78] The redundant sequences from cog0613 obtained from the NCBI database were removed using Jalview 2.7[53] and then converted into FASTA format. Cytoscape 2.8.2[54] was used to create protein sequence network diagrams, while the MUSCLE program of Jalview 2.7 was used to make the protein 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 *Bifidobacterium adolescentis*; PDB id: 3O0F, sequence identity 25%) had previously been determined with bound adenosine monophosphate (AMP). In addition, several mono and diphosphorylated nucleosides were found to increase the thermo stability of Cv1693 by thermofluor analysis (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 datasets could be collected. The structure of Cv1693 was determined by single isomorphous replacement with anomalous dispersion from a trimethyl lead 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 density for the metals is consistent with the ICP-metal analysis for the protein used in the crystallization, which exhibited a mixture of Ni^{2+} (0.6 eq), Zn^{2+} (0.2 eq) and Fe^{2+} (0.1 eq).

In an attempt to improve the homogeneity of the crystals, a new round of crystal screening was performed after Cv1693 had been treated with EDTA to remove the metals. The structure of apo-enzyme (no metals) Cv1693 was determined from crystallization conditions containing 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 Mn^{2+} and AMP was solved

from crystals grown in the presence of Mn^{2+} and AMP and in conditions lacking ammonium sulfate. The structure of Cv1693 and the Cv1693 metal/nucleoside complex are nearly identical with no large-scale structural movements (RMSD 0.22, 284 common Cα) and small side chain positional differences in the ligands to the metals so 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 and this conclusion is 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 shown in **Table 4.2**. Cv1693 exhibits the (β/α)-barrel fold common to members of the amidohydrolase superfamily but with significant deviations (**Figure 4.2**). In Cv1693, the third β-strand of the barrel, $\beta_{3,}$ is significantly longer than is 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 the elongated β-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-strand **3** as β-strand **3/4**. While most AHS (β/α)-barrels are composed completely of parallel β-strands, the fifth β-strand of Cv1693, β-5, is anti-parallel to β–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 to 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 to α-5e, forms a four-helix bundle which caps the top of the (β/α)-barrel and contains several residues which 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 4.3**.

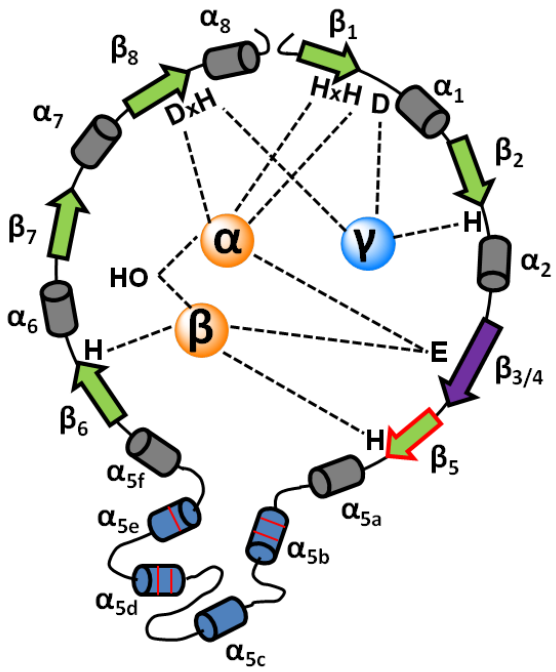


Figure 4.2: Schematic representation of the secondary structure of Cv1693. The α- and β-metal sites, which are a common feature among members of the PHP family and some amidohydrolase enzymes, are shown as orange spheres while the γ-metal site, unique to the PHP family members, is shown as a blue sphere. The strands and helices that constitute the distorted TIM- or $(\beta/\alpha)_7$-barrel are shown in green and gray respectively. The long β-strand **3/4** is shown in purple while the anti-parallel 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 shown in blue, and the positions of the enzyme residues seen interacting with the bound 5′-AMP are shown as red lines.

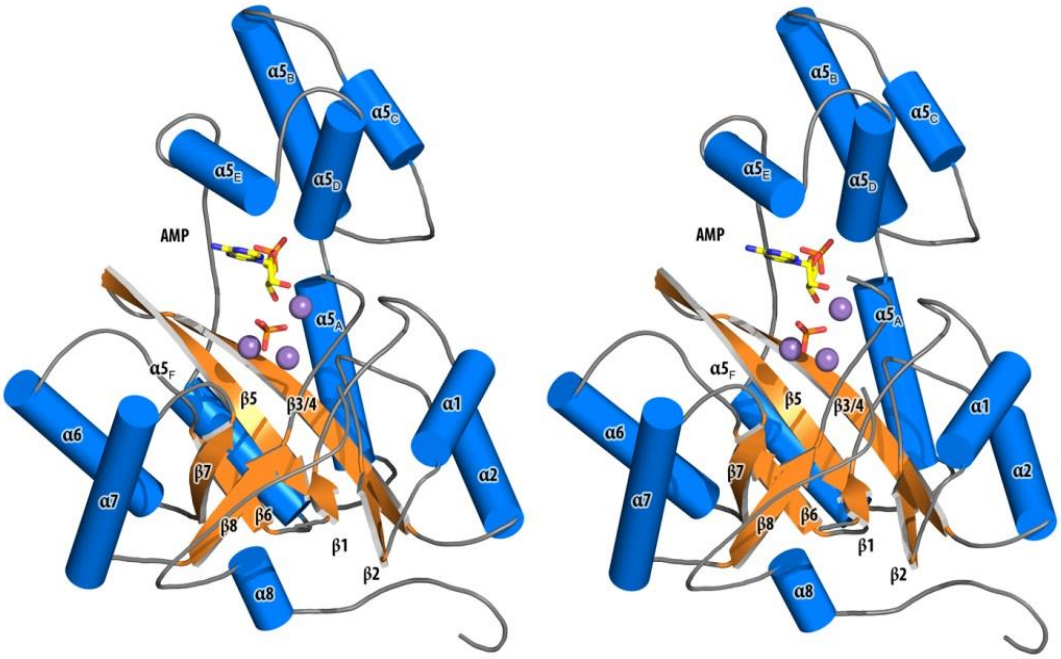


Figure 4.3: Stereo ribbon diagram of Cv1693. Helicies are shown as blue tubes, strands as orange arrows. Three bound manganese ions are shown as maroon spheres, and bound inorganic phosphate and adenosine monophosphate as sticks, colored by atom type. The insertion sequence is comprised of helices **α-5b** to **α-5e**.

Metal Center. Clear electron density was observed for three manganese ions (Mn_α, Mn_β, and Mn_γ) and AMP. In addition, a tetragonal shaped density bridging the three metal ions was interpreted as a phosphate ion, and was assumed to originate from degradation of AMP (**Figure 4.4**). A comparison of thermal factors suggests the Mn^{2+}/AMP/$HPO_4^{=}$ ligands 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** to **α-1** loop), His-39 (**β-2** to **α-2** loop) and His-

250 (**β-8** to **α-8** loop). The orthophosphate not only bridges Mn_α to Mn_β, but also Mn_α to Mn_γ, and is additionally coordinated by the guanidinium group of Arg-194 projecting from the β-6 to α-6 loop (**Figure 4.5**).

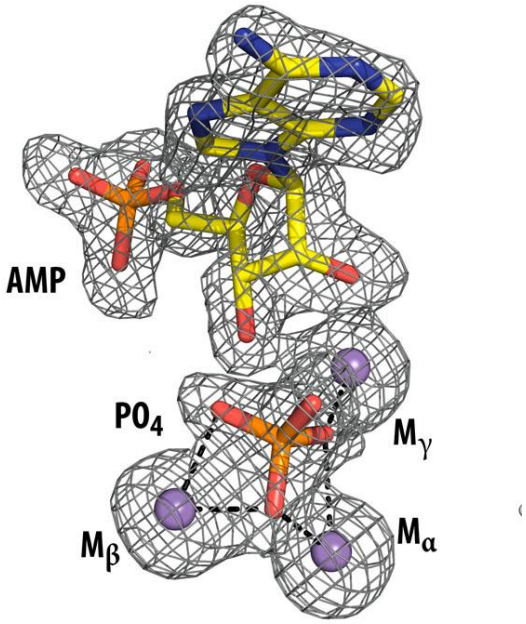


Figure 4.4. F_o-F_c electron density kick map contoured at 2.5σ. Interactions of the bound inorganic phosphate with the three manganese ions are shown as dashed lines.

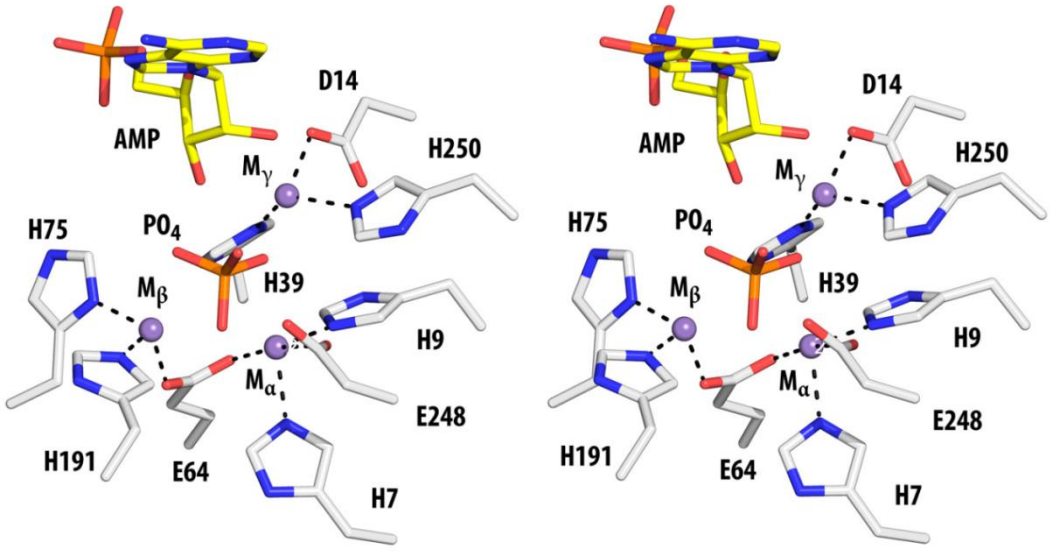


Figure 4.5. Stereo diagram illustrating the ligands to the three bound manganese ions. Residues colored by atom type with the adenosine monophosphate with yellow carbons and protein atoms with white carbons. Interactions between protein atoms and the manganese atoms are shown as dashed lines.

Contacts between Cv1693 and AMP. All of the contacts between Cv1693 and AMP originate from the insertion domain between β5-$\alpha5_f$. Arg-99 and Arg-103 from $\alpha5_b$ and Arg-134 from $\alpha5_d$ 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 ($\alpha5_e$). The ribose moiety is coordinated by a hydrogen bond to Thr-135 to the 2′-OH and coordination of the 2′-OH (2.4 Å) and 3′-OH (2.5 Å) groups 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 4.6**.

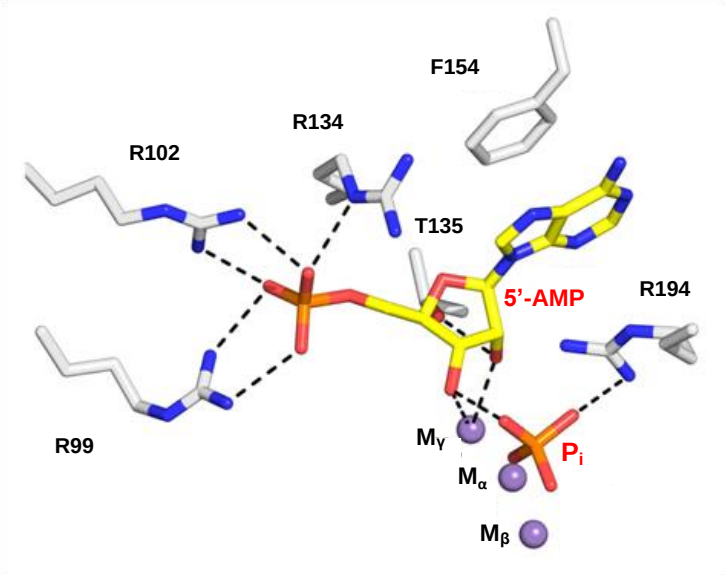


Figure 4.6. Interactions of AMP with protein residues, the inorganic phosphate, and the manganese ion are shown as dashed lines.

Substrate Profile. The kinetic constants k_{cat}, K_m and k_{cat}/K_m are summarized in **Table 4.3** for the hydrolysis of pApA and eleven 3′,5′-nucleotide bisphosphates. The substrates with the highest activity were N^6-methyl-3′,5′-pAp (k_{cat}/K_m = 2.1 x 10^6 M^{-1} s^{-1}) and pAp (k_{cat}/K_m = 1.4 x 10^6 M^{-1} s^{-1}) although Cv1693 hydrolyzed several other 3′,5′-nucleotide bisphosphates with similar catalytic efficiencies. At a concentration of 100 μM 2′,5′-pAp, the upper limit for substrate turnover was 0.015 s^{-1} in the presence of Cv1693. This is more than three 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 hydrolyzed other substrates

quite slowly when measured at a concentration of 100 μM: 3′-NADPH (~ 8 x 10^{-4} s^{-1}), 3′-AMP (≦2 x 10^{-4} s^{-1}), CoA (≤ 7 x 10^{-4} s^{-1}) and a 3′-phosphorylated 5-mer of DNA dAdAdAdAdAp (≤1 x 10^{-4} s^{-1}) and thus these compounds were not considered substrates of Cv1693. The 5′-phosphorylated 5-mer of DNA pdApdApdApdApdA, abbreviated $(pdA)_5$, was a poor substrate for Cv1693 with an estimated rate constant of ~7 x 10^{-4} s^{-1} at a concentration of 100 μM. At a concentration of 100 μM $(pdA)_5$, all of this compound was hydrolyzed to 5′-AMP in 1 hour after mixing with 1 μM Cv1693 at 30 °C.

Table 4.3. 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 ± 0.2) x 10^6
2′-deoxy-pAp	7.1 ± 0.2	10.2 ± 0.9	(7.0 ± 0.6) x 10^5
pCp[b]	8.6 ± 0.3	14 ± 2	(6.2 ± 1.0) x 10^5
2′-deoxy-pCp	9.4 ± 0.3	26 ± 3	(3.6 ± 0.5) x 10^5
pGp[b]	12.3 ± 0.4	17.3 ± 2	(7.1 ± 0.8) x 10^5
2′-deoxy-pGp	4.2 ± 0.1	5.2 ± 0.4	(7.9 ± 0.6) x 10^5
pUp[b]	5.4 ± 0.2	8.8 ± 1.0	(6.1 ± 0.8) x 10^5
2′-deoxy-pUp	5.2 ± 0.2	29 ± 4	(1.8 ± 0.2) x 10^5
pTp	3.3 ± 0.1	9.3 ± 1.0	(3.6 ± 0.4) x 10^5
2′-deoxy-pIp	4.3 ± 0.2	22 ± 3	(1.9 ± 0.3) x 10^5

Table 4.3 Continued.

Compound	k_{cat}, s^{-1}	K_m, μM	k_{cat}/K_m, $M^{-1} s^{-1}$
pApA	0.05 ± 0.01	1.6 ± 0.2	(3.2 ± 0.3) x 10^4
N^6-methyl-3′,5′-pAp	11.6 ± 0.4	5.4 ± 1.0	(2.1 ± 0.4) x 10^6

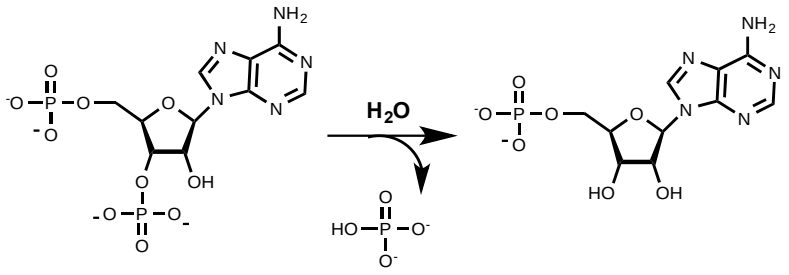
[a]These assays were conducted with mixtures of the 2′,5′ and 3′,5′ nucleotide bisphosphates.

DISCUSSION

Substrate Determination. We have demonstrated that Cv1693 from *Chromobacterium violaceum* catalyzes the efficient hydrolysis of 3′,5′-adenosine bisphosphate (pAp) to 5′-AMP and orthophosphate and this reaction is presented in Scheme 4.1. In addition to pAp, this enzyme will also hydrolyze pCp, pUp, pTp, and pGp with nearly equal 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 two orders of magnitude. The successful functional annotation of Cv1693 began with the determination of the three-dimensional structure of this enzyme and the recognition that another protein of unknown function, Bad1165 from *B. adolscentis*, was structurally quite similar (PDB id: 3O0F) despite low sequence identity (25%). Since Bad1165 had been successfully crystallized with AMP, a nucleotide cocrystallization scan and an exchange of active site metals eventually led to the successful structure determination 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 significantly 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. In addition, Bad1165 had been shown to catalyze the hydrolysis of *p*-nitrophenylphosphate, but the hydrolysis of pNp-like substrates has apparently not been investigated. Finally, the close proximity of orthophosphate in the active site of Cv1693 to the 3′-OH 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 or 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 value of k_{cat}/K_m that exceeds 10^6 M^{-1} 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 bond in short oligonucleotides is significantly slower than the rate of hydrolysis of the phosphomonoester bond in pAp.



Scheme 4.1. Reaction catalyzed by Cv1693.

Structural Comparison to Bad1165. A structure similarity search using the secondary structure similarity matching server[79], with a 70% secondary structure match requirement, returned only one structural homolog: Bad1165 from *B. adolescentis* (RMSD = 1.87 Å over 255 common Cα carbons). The three-dimensional structure of Bad1165 was determined in complex with phosphate to 2.4 Å resolution (3E0F) and AMP and phosphate to 1.9 Å resolution (3O0F). Based on 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 metal content. Two other active site features are strictly conserved between the two enzymes: (*a*) The three arginine residues originating from the insertion domain which coordinate the 5′-phosphate, and (*b*) the phenylalanine side chain, also from the insertion domain, which rests 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 adeneine is in the *syn* conformation, while in Bad1165 the ribose is

in the 0-4'*endo* conformation and the adenine is in the *anti* conformation. In Cv1693 the 2′- and 3′-hydroxyl groups interact with M_γ and the 3′-OH 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, due to its *syn* 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 flexible loop between **β-6** and **α-6**. The structural overlay of Cv1693 and Bad1165 is presented in **Figure 4.7** and the specific interactions of AMP in the active site of Bad1165 are presented in **Figure 4.8**.

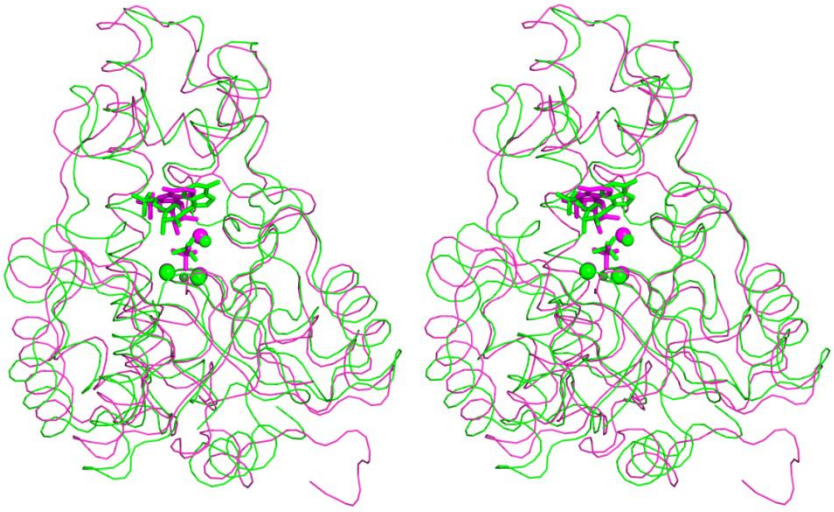


Figure 4.7. Stereo diagram of the superposition of Cv1693 with Bad1165. The structure of Cv1693 shown as a purple ribbon and sticks and the structure of Bad1165 is shown as a green ribbon and sticks.

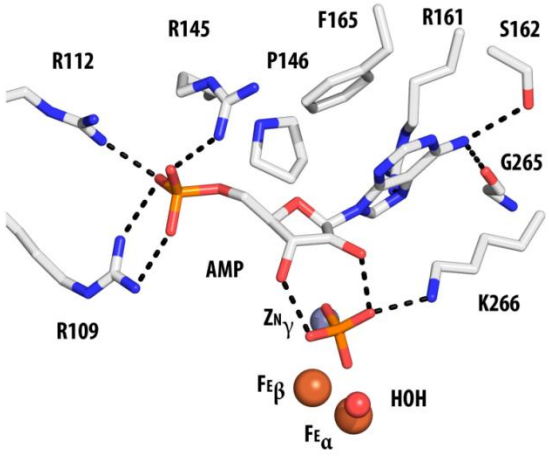


Figure 4.8. Interactions of AMP with Bad1165 in a similar orientation as seen for Cv1693 in **Figure 4.6.** Residues are colored by atom type and interactions of AMP with Bad1165 are shown as dashed lines. There is a water molecule that bridges Fe_{α} and Fe_{β}.

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 sulfate transfer from 3′-phosphoadenosine-5′phosphosulfate (PAPS) and the transfer of the phosphopantetheine group of CoASH to the acyl-carrier protein (ACP).[80] In many bacteria such as *E. coli*, the pAp product is hydrolyzed to 5′-AMP and orthophosphate by an enzyme denoted as CysQ.[81] The CysQ enzyme from *E. coli* K12 (locus tag: b4214) belongs to cog1218 and hydrolyzes pAp with a catalytic efficiency of ~10^7 M^{-1} s^{-1}.[82] The CysQ from *Mycobacterium tuberculosis*, Rv2131c, has a reported k_{cat}/K_m for the hydrolysis of pAp of 7 x 10^5 M^{-1} s^{-1}.[80a] It has also been reported that NrnA (YtqI) orthologs from *B. subtilis*, *M. tuberculosis*, *M. pneumonia* and *S. mutans* can also catalyze the hydrolysis of pAp, in addition to the hydrolysis of short pieces of single strand RNA.[83] 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. Since >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 each of the three cogs, respectively. A sequence network diagram at a BLAST E-value of 1 x 10^{-70} was constructed for each of these COGs (see **Figure 4.9-4.11**). Any two sequences

connected by lines bear a sequence similarity greater than the BLAST E-value of 1×10^{-70}, which corresponds to an overall sequence identity of ≥40%. We then identified the proteins which have been biochemically characterized from each of the three cogs 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, 288 from cog1218, and 219 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 only six organisms had an apparent pAp phosphatase from both cog0613 and cog1218 (cysQ), while the remaining organisms had a pAp phosphatase from just one of the three cogs (see Venn diagram in **Figure 4.12**). Further examination of the phylogenetic classification of each organism revealed that organisms possessing the pAp phosphatase from cog0613 are predominantly betaproteobacteria (~97%), those from cog1218 are largely gammaproteobacteria (~77%) while those from cog0618 are all bacilli from the phylum Firmicutes (100%). It is clear that different classes of microorganisms have evolved enzymes to hydrolyze pAp belong to different families. This is also a strong indication that the physiological function of Cv1693 and its closely related orthologs (shown in blue in **Figure 4.1**) from cog0613 is the hydrolysis of pAp.

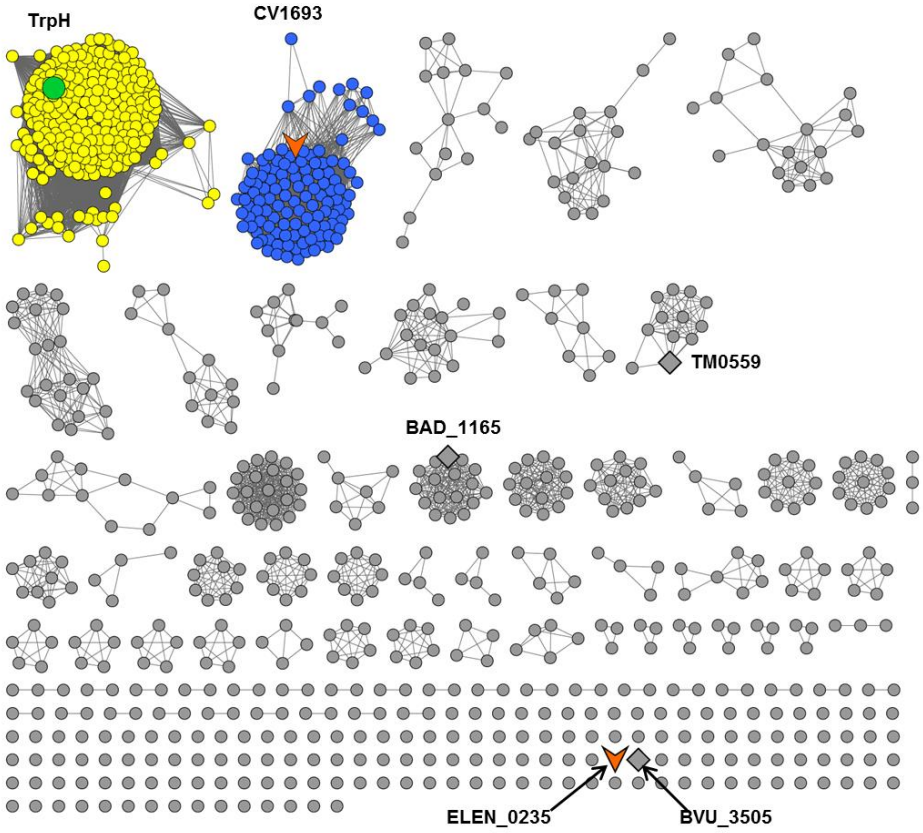


Figure 4.9. Sequence similarity network of cog0613 at E-value of 1 x 10^{-70}. Available crystal structures with unknown functions from this cog are shown as diamond-shaped nodes: TM0559 from *T. maritima* (PDB: 2ANU), Bad_1165 from *B. adolescentis* (PDB: 3E0F, 3O0F), Bvu_3505 from (3E38). The two enzymes for which physiological substrates have been discovered are shown as orange vee-shaped nodes: Elen0235 from *E. lenta* and Cv1693 from *C. violaceum* (this study). TrpH or YciV from *E. coli* (refer main text) is shown as green sphere. The arbitrary functional boundaries in cog0613 were determined based on this figure, and all enzyme sequences predicted to have the same substrate profile as CV1693 are colored blue while those predicted to be isofunctional to TrpH are colored yellow.

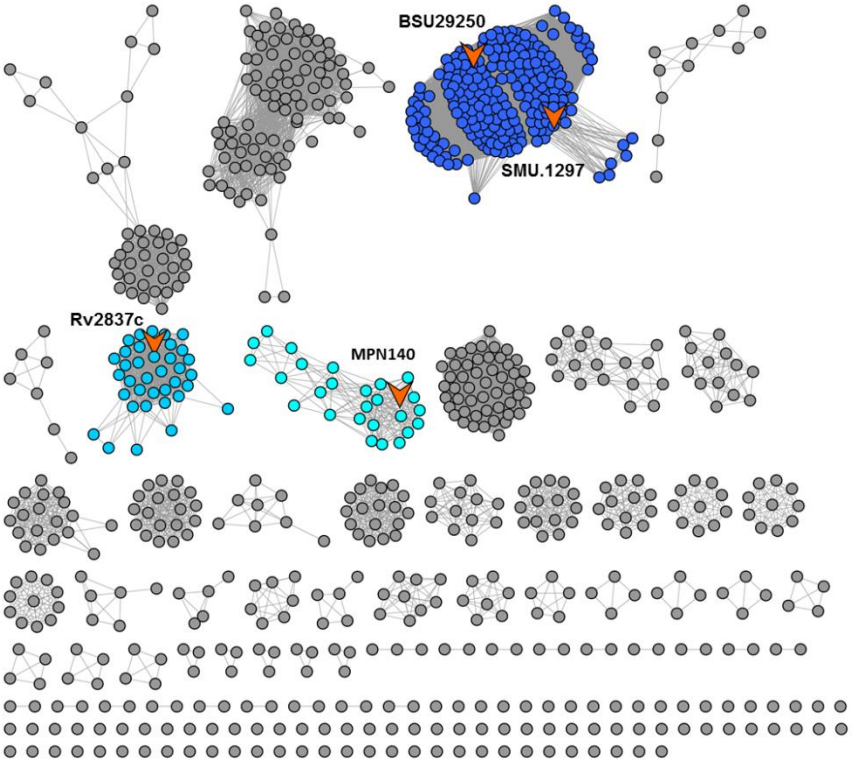


Figure 4.10. Sequence similarity network of cog0618 at E-value of 1 x 10^{-70}. The four enzymes for which physiological substrates have been discovered are shown as orange vee-shaped nodes: BSU29250 (cysQ) from *B. subtilis*, SMU.1297 from *S. mutans*, Rv2837c from *M. tuberculosis*, and MPN140 from *M. pneumonieae*. The arbitrary functional boundaries in cog0618 were determined based on this figure, and all enzyme sequences predicted to have the same substrate profile as the biochemically characterized enzymes (orange) are colored with different shades of blue.

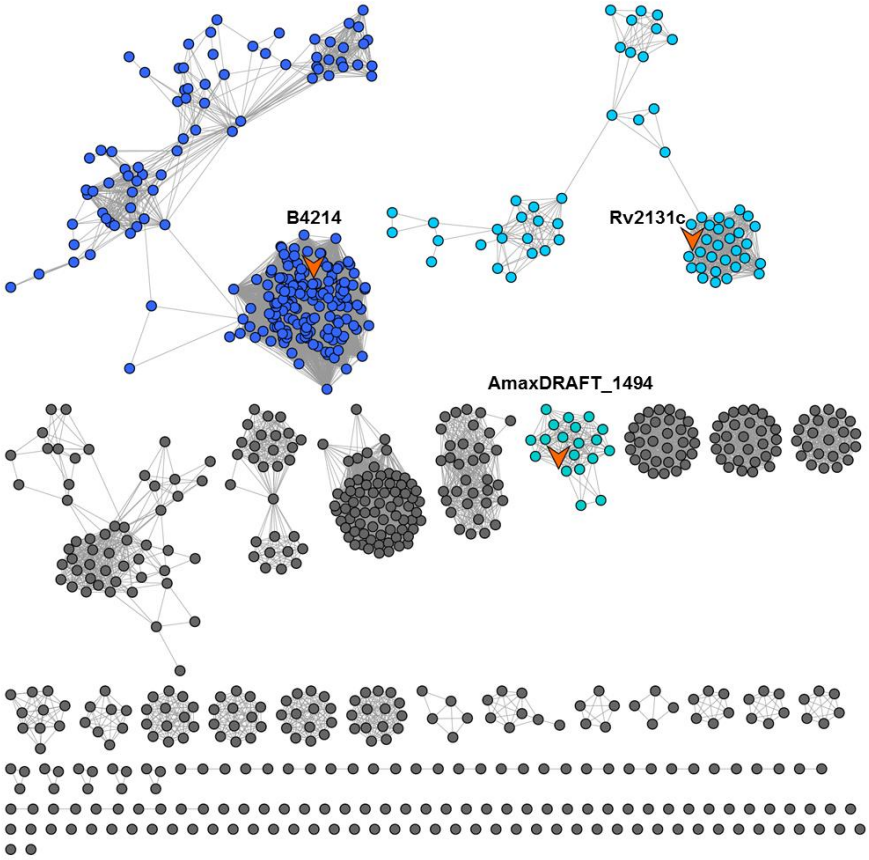


Figure 4.11. Sequence similarity network of cog1218 at E-value of 1 x 10^{-70}. The three enzymes for which physiological substrates have been discovered are shown as orange vee-shaped nodes: B4214 (cysQ) from *E. coli*, Rv2131c from *M. tuberculosis* and AmaxDRAFT_1494 from *A. maxima* (closest homologous sequence in Microbesonline database to the originally characterized halA from *A. platensis*, sequence identity = 96%). The arbitrary functional boundaries in cog1218 were determined based on this figure, and all enzyme sequences predicted to have the same substrate profile as the biochemically characterized enzymes (orange) are colored with different shades of blue.

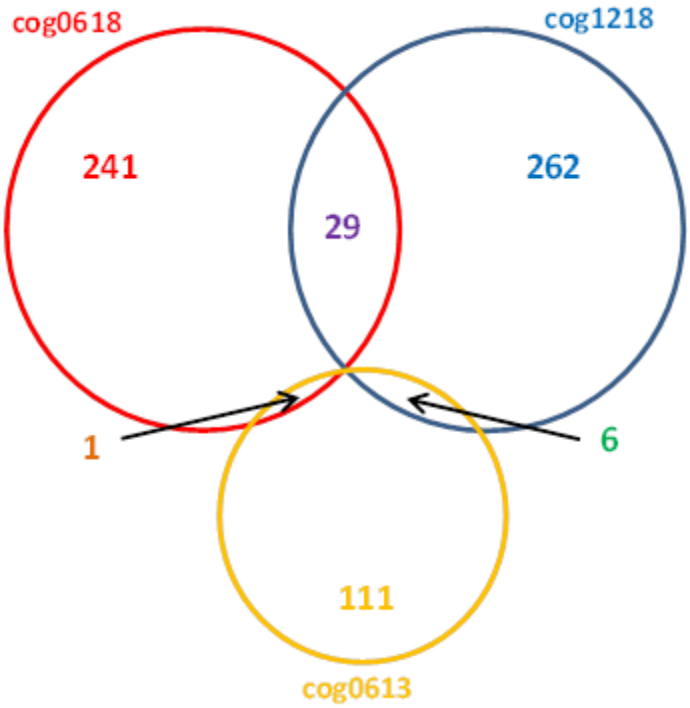


Figure 4. 12. Venn diagram showing the number of organisms that possess enzyme(s) capable of hydrolyzing 3′,5′-adenosine bisphosphate (pAp) from cogs 0613, 0618 and 1218.

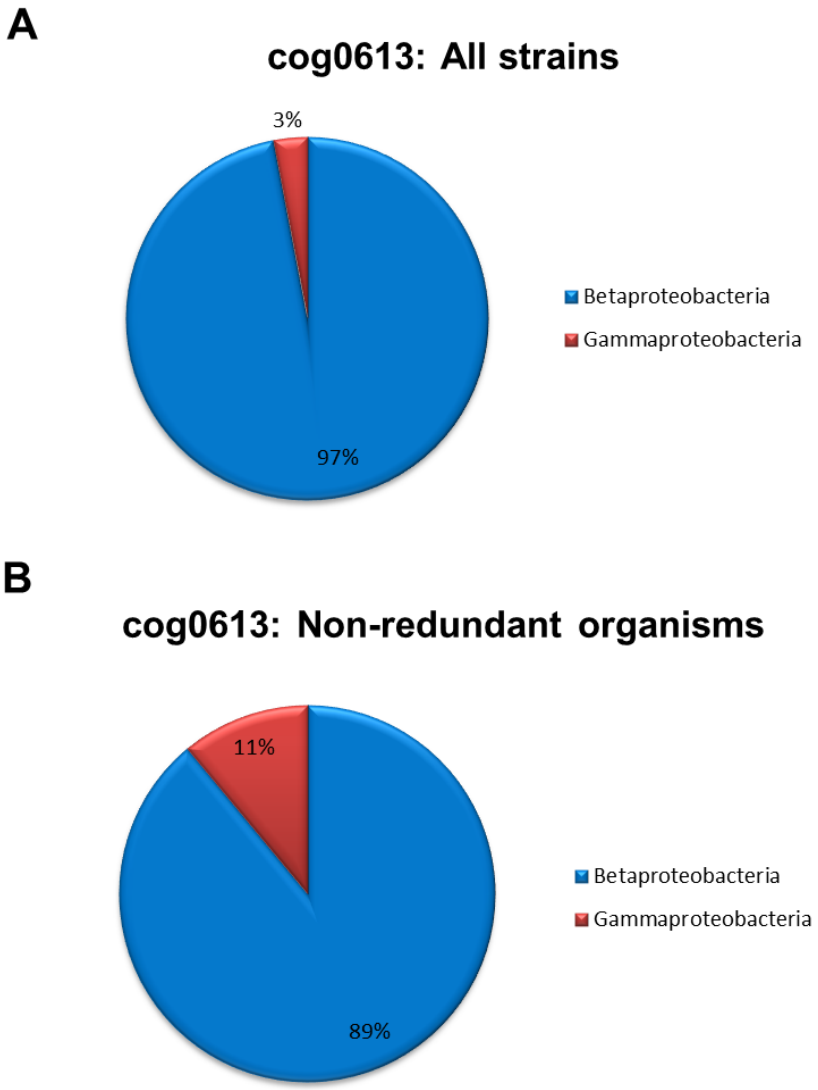


Figure 4.13. Pie chart showing the relative percentage of taxonomic classes of organisms found to possess 3′,5′-pAp phosphatase from cog0613. (A) Percentage values showing the relative number of organisms from various taxonomic classes based on all 118 strains of organisms that were found to possess an enzyme sequence bearing significant homology to Cv1693 (shown in blue in **Figure 4.9**). (B) Percentage values showing the relative number of organisms from various taxonomic classes based on 37 distinct organisms, i.e. after neglecting the redundant strains, that were found to possess an enzyme sequence bearing significant homology to CV1693.

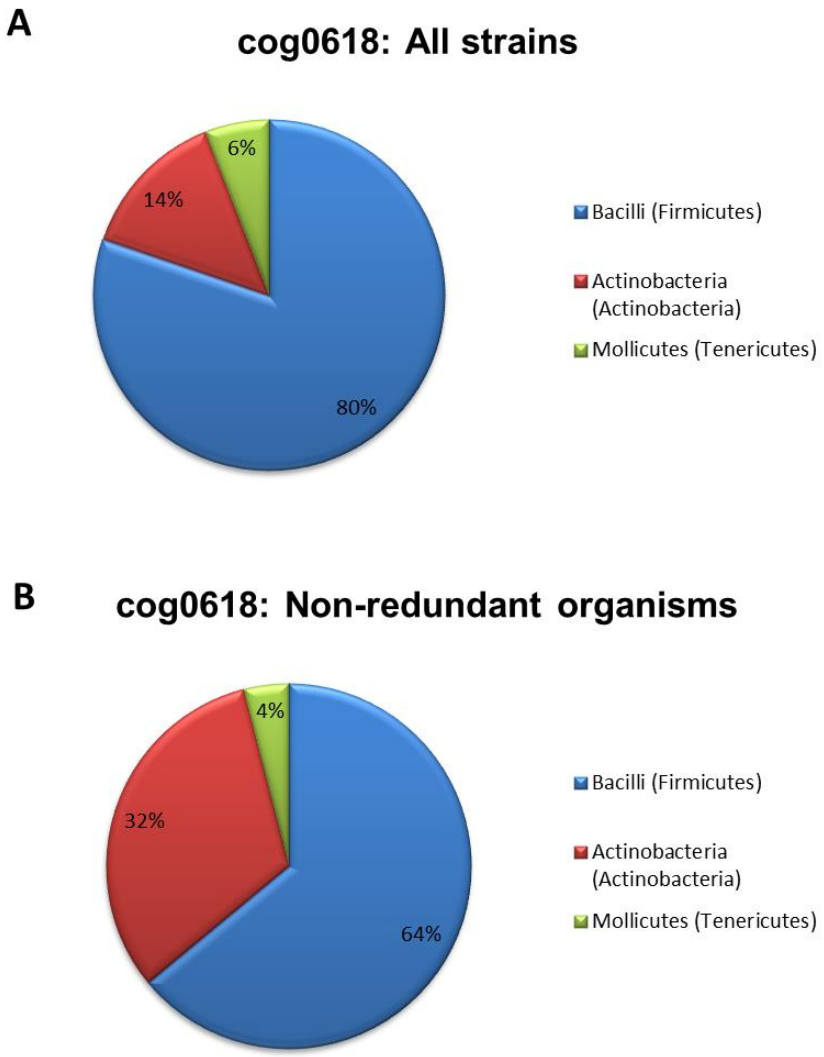


Figure 4.14. Pie chart showing the relative percentage of taxonomic classes of organisms found to possess 3′,5′-pAp phosphatase from cog0618. (A) Percentage values showing the relative number of organisms from various taxonomic classes based on all 271 strains of organisms that were found to possess the 283 enzyme sequences bearing significant homology to one among Bsu29250, Smu.1297, Rv2837c or Mpn140 (indicated in **Figure 4.10**). (B) Percentage values showing the relative number of organisms from various taxonomic classes based on 25 distinct organisms, i.e. after neglecting the redundant strains, which were found to possess an enzyme sequence bearing significant homology to a biochemically characterized pAp phosphatase mentioned above.

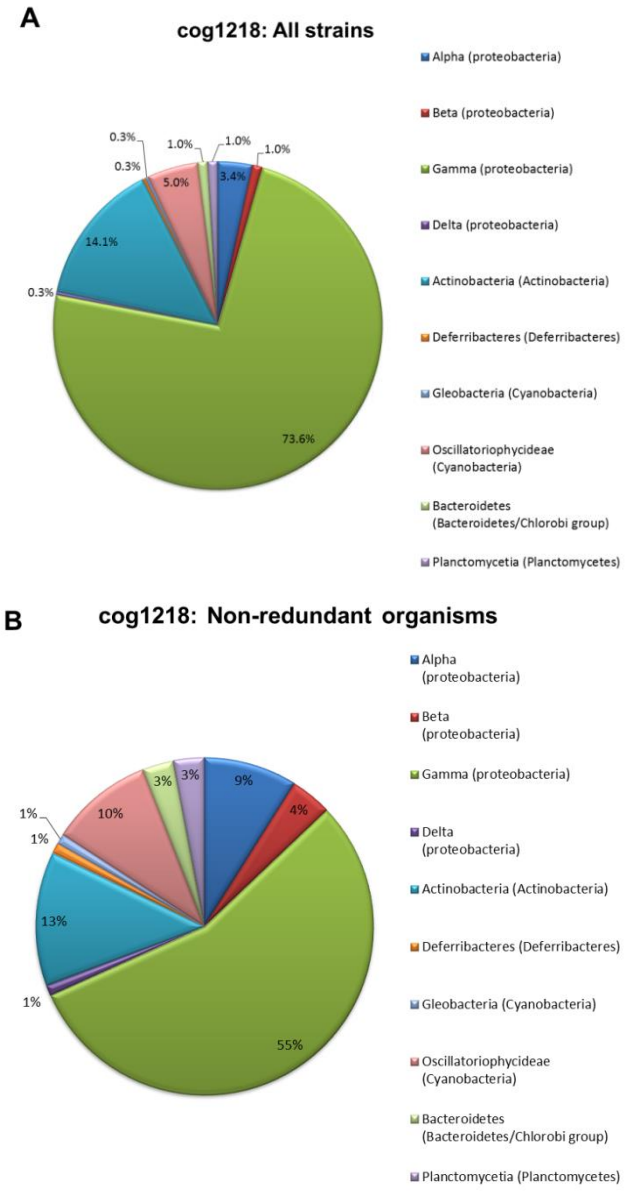


Figure 4.15. Pie chart showing the relative percentage of taxonomic classes of organisms found to possess 3′,5′-pAp phosphatase from cog1218. (A) Percentage values showing the relative number of organisms from various taxonomic classes based on all 298 strains of organisms that were found to possess the 309 enzyme sequences bearing significant homology to one among B4214, Rv2131c, or AmaxDRAFT_1494 (indicated in **Figure 4.11**). (B) Percentage values showing the relative number of organisms from various taxonomic classes based on 80 distinct organisms, i.e. after neglecting the redundant strains, which were found to possess an enzyme sequence bearing significant homology to a biochemically characterized pAp phosphatase mentioned above.

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 4.16**. 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 (shown in yellow in **Figure 4.1**) 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 is 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 of 5′-AMP, and is replaced by Gly-142 in TrpH. However, since 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, since the relatively small glycine residue can be envisaged to accommodate a larger functional group at the enzyme active site.

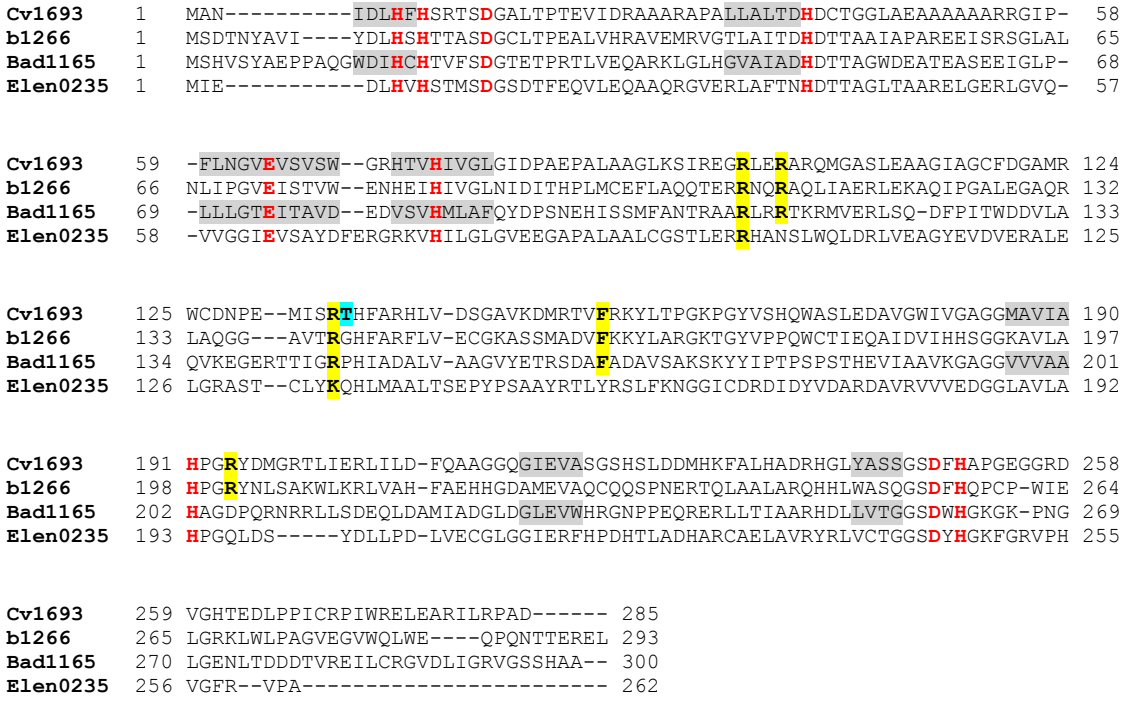


Figure 4.16. Primary sequence alignment of proteins from cog0613. Cv1693: pAp phosphatase from *Chromobacterium violaceum* (PDB: 2YB1, 2YB4); b1266: TrpH from *Escherichia coli*; Bad1165: enzyme of unknown function from *Bifidobacterium adolescentis* (PDB:3E0F, 3O0F); and Elen0235: cyclic phosphate dihydrolase from *Eggerthella lenta*. Residues binding the metal cofactors at the active site are shown in red, while residues seen interacting with the bound 5′-AMP and inorganic phosphate in the crystal structure of Cv1693 are highlighted in yellow and Thr-135 is highlighted in blue. β-sheets that constitute the $(\beta/\alpha)_7$-barrel are highlighted in gray.

CONCLUSION

We have determined the crystal structure of Cv1693 from cog0613 of the amidohydrolase superfamily with 5′-AMP and orthophosphate phosphate 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.

CHAPTER V

DISCOVERY OF A NANO-RNASE FROM *ESCHERICHIA COLI* CAPABLE OF SEQUENTIAL 5′→3′ HYDROLYSIS OF OLIGONUCLEOTIDES

Systematic efforts in high-throughput sequencing have resulted in an exponential rise in the number of known and predicted protein sequences in various databases. As of October 2014, the UniProtKB/TrEMBL protein database contained more than 86 million gene sequences. However, the process of annotating these protein sequences has not kept pace with sequencing, and consequently, there has emerged a gap between the knowledge of protein sequences, and their respective physiological substrates and functions. One of the approaches to help bridge this gap in correctly annotating protein functions is to probe the mechanistic details of a known enzymatic reaction, and then attempt to discover the best substrates of closely related homologs, using various bioinformatic tools and library screening. This work describes the use of such an approach to discover the function of a hypothetical protein named TrpH or YciV (Locus tag: b1266) from *Escherichia coli*.

TrpH belongs to the polymerase and histidinol phosphatase (PHP) family of proteins which is related to the amidohydrolase superfamily (AHS). Enzymes from the amidohydrolase superfamily are known to hydrolyze ester and amide functional groups at carbon and phosphorus centers.[6] These enzymes possess a distorted TIM-barrel ($(\beta/\alpha)_8$-barrel) fold and 0-3 divalent metal ion cofactors in the active site. The metal cofactors have been implicated in activating a water or hydroxide ligand as the

nucleophile in these hydrolytic reactions. PHP family enzymes possess a distorted $(\beta/\alpha)_7$-barrel fold and a trinuclear metal center consisting of divalent metal cofactors at the enzyme active site. These metal sites have been named α, β, and γ respectively. The PHP family within the amidohydrolase superfamily is classified into three Clusters of Orthologous Groups: cog1387, cog0613 and cog4464. The generic annotations for the three COGs, according to the NCBI database, are L-histidinol phosphate phosphatase related enzymes, metal-dependent phosphoesterases, and protein-tyrosine phosphatases respectively. We have probed the structure and mechanism of L-histidinol phosphate phosphatase (HPP) from *Lactococcus lactis*, which belongs to cog1387.[68] The structure of the holo-enzyme with a complex of L-histidinol and arsenate mimicking the substrate L-histidinol phosphate provided clear insights to the orientation of the bound substrate at the active site. In conjunction with the biochemical characterization of HPP and previous studies of other AHS enzymes, it has been proposed that the α- and β-metal ions are responsible for the activation of the bridging nucleophilic hydroxide as well as coordinating two of the four oxygen atoms of the phosphate functional group to be hydrolyzed from the organophosphoester substrate. The γ-metal ion coordinates the bridging oxygen atom of the phosphoester, acting as a Lewis acid to the leaving alcohol product of hydrolysis. The fourth oxygen atom of the phosphoester is generally stabilized by salt-bridge interactions with a positively charged amino acid side-chain like arginine or lysine. This general orientation of the phosphoester substrate at the active site is expected to be the same for all trinuclear PHP family enzymes catalyzing the hydrolysis of organophosphoesters.

TrpH belongs to cog0613 within the PHP family. The sequence similarity network diagram of cog0613 at an E-value cut-off of 1 x 10^{-60} is depicted in **Figure 5.1**. There are four proteins from cog0613 whose crystal structures are available in the Protein Data Bank (PDB): Cv1693 from *Chromobacterium violaceum* (PDB: 2YB1, 2YB4), Bad1165 from *Bifidobacterium adolescentis* (PDB: 3O0F)[74], Tm0559 from *Thermotoga maritima* (PDB: 2ANU), and Bvu3505 from *Bacteroides vulgatus* (PDB: 3E38). The only two enzymes from cog0613 whose physiological functions have been elucidated are Elen0235 from *Eggerthella lenta* and Cv1693 from *C. violaceum*. Elen0235 is a cyclic phosphodiester dihydrolase that catalyzes hydrolysis of 5-phosphoribose-1,2-cyclic phosphate to ribose-2,5-bisphosphate, followed by subsequent hydrolysis of ribose-2,5-bisphosphate to ribose-5-phosphate and inorganic phosphate (**Scheme 5.1**).[75] Cv1693 hydrolyzes the 3′-phosphate of 3′,5′-bisphosphonucleotides to give 5′-nucleotide monophosphate and inorganic phosphate (**Scheme 5.2**).[84] The sequence identities of TrpH with Elen0235 and Cv1693 are 23% and 37% respectively. Therefore, we hypothesized that the substrate for TrpH would structurally resemble the substrate for Cv1693 more than that of Elen0235. We found that TrpH can hydrolyze 3′,5′-bisphosphonucleotides efficiently. Additionally, it can hydrolyze oligonucleotides bearing a 5′-phosphoryl group with comparable efficiency as compared to 3′,5′-bisphosphonucleotides, as opposed to Cv1693. In this study, we have characterized the novel enzymatic activity of TrpH from *E. coli*.

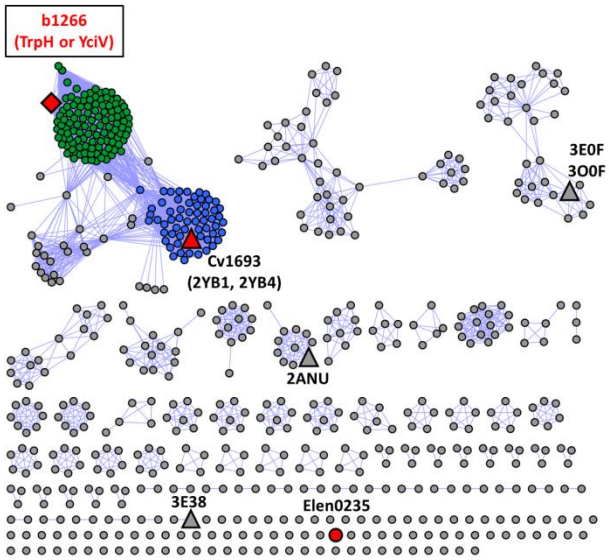
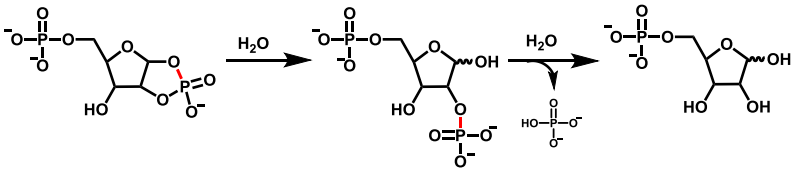
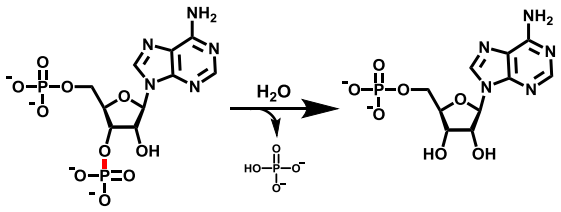


Figure 5.1. Sequence similarity network of cog0613 at an E-value cutoff of 10^{-60}. This network was made using Cytoscape (http://www.cytoscape.org). Each node represents a non-redundant protein sequence, while each blue edge (line) represents the pairs of sequences that are more closely related than the arbitrary E-value cutoff (10^{-60}).The available crystal structures are shown as triangles, and their respective PDB codes are indicated. The enzymes that have been biochemically characterized are colored red. Cv1693 is 3′,5′-adenosine diphosphate-3′-phosphatase, while Elen0235 is cyclic phosphate dihydrolase (cPDH). TrpH (b1266) is the enzyme characterized in this study. Protein sequences that bear sufficiently high sequence homology to Cv1693 (E-value<10^{-70}, which is approximately equivalent to sequence identity > 43%) are colored blue, while those that bear a similar homology to TrpH are colored green.



Scheme 5.1. Reaction catalyzed by Elen0235.



Scheme 5.2. Reaction catalyzed by Cv1693.

MATERIALS AND METHODS

All chemicals were purchased from Sigma-Aldrich, unless indicated otherwise. Genomic DNA for *Escherichia coli* K-12 substr. MG1655 (ATCC25559) was obtained from American Type Culture Collection (ATCC). Pfu Turbo Polymerase, T4 DNA ligase and restriction enzymes were procured from New England Biolabs. DNA primers and Big Dye were obtained from Integrated DNA Technologies (IDT). The vector pET30a(+) was purchased from EMD Biosciences. *E. coli* BL21(DE3) and XL1 Blue competent cells were obtained from Stratagene. SDS-PAGE gel electrophoresis was carried out using Biorad Mini-protean® TGX any-kD gels. The P_i Colorlock Gold kit for the determination of inorganic phosphate was procured from Innova Biosciences. Oligonucleotide substrates were purchased from Integrated DNA Technologies, except for p(dAp)dA, $p(dAp)_2dA$, dA(dAp), and $dA(dAp)_2$, which were obtained from Genelink. Synthesis of 3′,5′-bisphosphonucleotide substrates is described elsewhere.[76] Cloning, expression and purification of Cv1693 from *Chromobacterium violaceum* is reported in our previous work.[84] HPLC experiments were performed using GE Akta Purifier system or Biorad NGC system. UV-visible spectra were recorded using

Spectromax 384 Plus 96-well plate reader from Molecular Devices. Purified guanylate kinase from *S. cerevisiae* was a generous gift from the laboratory of Tadhg Begley (Texas A & M University).

Cloning, Expression and Purification of TrpH from *Escherichia coli*. The PCR amplification was done from the genomic DNA of *Escherichia coli* K-12 strain using the following primer sequences: 5′-GCAGGAGCCATATGAGCGACACGAATTATGCAGTGATTTACGACCTGC-3′ as forward primer and 5′-CGCGCTCGAGTAATTCCCTCTCTGTGGTGTTCTGCGGCTGTTCC-3′ as reverse primer. The primer pair contained restriction sites for *Nde*I and *Xho*I respectively. The PCR product was purified with a PCR clean-up kit (Qiagen), double digested using *Nde*I and *Xho*I, and ligated into a pET30a(+) vector previously double digested with the same set of restriction enzymes.

The recombinant plasmid was expressed and purified according to the iron-free expression procedure described elsewhere.[68] In brief, the ligated plasmid was transformed into *E. coli* BL21(DE3) cells by electroporation. LB broth cultures (5 mL) containing 50 μg/mL kanamycin were inoculated with single colonies, and grown overnight. These were used to inoculate one liter of the same medium at pH 7.2, and were allowed to grow at 37 °C till the OD_{600} reached 0.15-0.2. The temperature was reduced to 30 °C and the iron-specific chelator 2,2′-bipyridyl was added to a final concentration of 150 μM. $MnSO_4$ was added to a final concentration of 1 mM when the OD_{600} reached ~0.4, and 0.25 mM isopropyl D-thiogalactopyranoside (IPTG), when the

OD_{600} reached ~0.6. The temperature was then lowered to ~15 °C, and the cells were shaken for ~16 hours, after which, they were harvested by centrifugation and stored at -80 °C till used. For protein purification, cells (~3 g) were thawed and resuspended in ~60 mL of buffer (25 mM HEPES, pH 7.5, 250 mM KCl, 10% glycerol and 100 μM $MnCl_2$) containing 0.4 mg/mL of the protease inhibitor phenylmethanesulfonyl fluoride (PMSF). Cells were lysed by sonication and the insoluble cell debris was separated by centrifugation. The supernatant solution was treated with 10 mL of 6 mg/mL solution of protamine sulfate in the purification buffer. The precipitated DNA was separated by centrifugation after incubation for 30 minutes. Ammonium sulfate was added, first to a concentration of 35% saturation, followed by centrifugation, and then to 70%. The protein of interest salted out at 70% saturation, based on SDS-PAGE gel electrophoresis. The precipitated protein was separated by centrifugation, resuspended in 10 mL of the same purification buffer as above, and purified using gel filtration chromatography (GE Superdex 26/600 column). The fractions were analyzed using SDS-PAGE gel electrophoresis. Fractions containing pure TrpH were pooled together. The protein concentration was determined by UV absorbance at 280 nm using a calculated molar extinction coefficient of 51,450 M^{-1} cm^{-1} and molecular weight of 32.5 kDa. The protein was flash frozen using liquid nitrogen and stored at -80 °C.

Metal Content Analysis. The metal content of purified proteins was determined using Perkin Elmer DRC-II inductively-coupled mass spectrometer. Samples were run in a matrix of 1% (v/v) nitric acid. In general, concentrated samples of purified proteins were passed through a desalting column (PD-10 from GE healthcare) pre-equilibrated

with metal-free buffer. Buffer was rendered free of metal ions by treating with Chelex-100 resin (Bio-Rad). The desalted protein was then digested with nitric acid (≥69%, Fluka Analytical) by heating to 95 °C for 15-20 minutes. The digested protein was diluted using deionized water such that the final sample contained 1 μM protein and 1% (v/v) nitric acid.

Enzymatic Assays Using Inorganic Phosphate Detection. The hydrolysis of 3′-phosphate from several substrates was monitored by quantitating the release of phosphate with time using the Pi ColorLock Gold kit from Innova Biosciences according to the manufacturer's directions. The assay conditions were 50 mM HEPES pH 7.5, 250 mM KCl, 2 mM $MgCl_2$, 0.15 mM $MnCl_2$, 0.05 mg/mL BSA, and 30 °C. Assays involved withdrawing aliquots of the enzymatic reactions and quenching at specific time points using the phosphate detection kit. A set of assays consisted of 6-12 substrate concentrations with 3-4 time points taken over a period of 20-40 minutes. The color was allowed to develop for 30 minutes after quenching and the absorbance was determined at 650 nm. The phosphate concentration of each well was determined using a previously generated standard curve under the same assay conditions. The initial rates at each substrate concentration were determined by linear regression in a plot of phosphate released versus time.

Enzymatic Assays Using 5′-Nucleoside Monophosphate Detection. The release of 5′-AMP, 5′-dAMP, 5′-GMP or 5′-dGMP by enzymatic hydrolysis of oligonucleotide substrates was measured by monitoring the corresponding oxidation of NADH to NAD^+ by lactate dehydrogenase. This was achieved using a coupling system

consisting of the enzymes adenylate kinase or guanylate kinase, pyruvate kinase and lactate dehydrogenase. Each 250 μL assay contained 250 mM KCl, 2 mM $MgCl_2$, 0.15 mM $MnCl_2$, 0.7 mM PEP, 0.5 mM ATP, 0.3 mM NADH, and 20 U/mL each of adenylate kinase, pyruvate kinase, and lactate dehydrogenase. The assays were conducted at 30 °C and monitored at 340 nm.

Enzymatic Assays Using Anion Exchange Chromatography. Polynucleotide substrates were tested for hydrolysis by TrpH and Cv1693 using Resource Q (1mL) anion exchange column (GE Lifesciences) and Akta Purifier system (GE) at a wavelength of 260 nm. 10 μM oligonucleotide was incubated overnight with 1.0 μM enzyme in a 250 μL reaction containing 25 mM HEPES pH 7.5, 0.5 mM $MgCl_2$, 200 mM KCl, and 50 μM $MnCl_2$. Time-dependent hydrolysis of oligonucleotide substrates by TrpH was monitored using ResourceQ (1mL) anion exchange column and NGC system (Bio-Rad) at 255 nm. In general, the reaction was started by adding the enzyme to a solution containing substrate, and aliquots were removed at specific times. These aliquots were quenched either by injecting directly onto the anion exchange column or by mixing with hot water (50 μL aliquot added to 200 μL hot water at temperature > 90 °C), depending on the interval between two time points. Assays in which the interval between two time points was less than 10 minutes, the latter method of quenching was used.

Annealing of Two Single Strands of DNA. Two single strands of DNA consisting of complementary sequences were mixed in a solution containing 10 mM HEPES pH 7.5, 50 mM KCl, and 1 mM $MgCl_2$, to a final concentration 100 μM of each

strand. The solution was heated at 95 °C for 10 minutes, and then allowed to cool slowly to room temperature. Overnight storage was at 4 °C.

Data Analysis. Kinetic parameters, k_{cat} and k_{cat}/K_m, were obtained by fitting the initial velocity data to equation 1 using the nonlinear least-squares fitting program in SigmaPlot 10.0, where v is the initial velocity at substrate concentration [A], [E_t] is the enzyme concentration, k_{cat} is the turnover number, and K_m is the Michaelis constant.

$$v/[E_t] = k_{cat}[A] / (K_m + [A]) \quad (1)$$

RESULTS

Purification, Metal Content and Optimization of Activity of TrpH. TrpH, expressed under iron-free conditions, and in the presence of exogenously added Mn^{2+} ions before induction, was found to contain ~1 equivalent of Mn^{2+} per monomer. When 3′,5′-bisphosphoadenosine was incubated with TrpH at room temperature, inorganic phosphate was released as a function of time. The rate of hydrolysis of 3′,5′-bisphosphoadenosine by TrpH was found to depend on the concentration of Mn^{2+} added to the assay, and the enzyme showed optimal activity at [Mn^{2+}] > 100 μM. The presence of KCl in the buffers used for enzyme purification, dilution and assays, was also found indispensible for the activity of the enzyme, while the addition of bovine serum albumin (BSA) prevented loss of activity at enzyme concentrations less than ~0.5 μM (data not shown).

Preliminary Test of Oligonucleotides with TrpH and Cv1693. The substrate profile for Cv1693 was known to include 3′,5′-bisphosphonucleotides, and the most

likely physiological substrate for this enzyme was determined to be 3′,5′-bisphosphoadenosine.[84] However, it was unknown whether Cv1693 as well as TrpH were capable of hydrolyzing oligonucleotides longer than two nucleotides. Polyadenosine or polydeoxyadenosine compounds, ranging in length from 2-5, were incubated with TrpH and Cv1693 respectively. These nano-RNA or DNA compounds bore either 5′- or 3′-phosphorylation. The results indicated that TrpH completely hydrolyzed oligonucleotides containing 5′-phosphorylation to 5′-AMP or 5′-dAMP respectively. Under identical conditions and duration of incubation, Cv1693 failed to completely hydrolyze oligonucleotides containing 5′-phosphorylation. The extent of hydrolysis by Cv1693 decreased progressively with the length of the oligonucleotides tested, indicating that TrpH was more efficient than Cv1693 in hydrolyzing 5′-phosphoryl oligonucleotides longer than p(Ap)A and p(dAp)dA in general. Both enzymes, TrpH and Cv1693, did not hydrolyze oligonucleotides containing 3′-phosphorylation to a detectable extent.

Kinetic Characterization and Substrate Specificity of TrpH. The kinetic parameters for the hydrolysis of various RNA and DNA-fragments by TrpH are listed in **Table 5.1**. The enzyme showed highest catalytic activity towards 3′,5′-bisphosphoadenosine (k_{cat} = 9.9 s^{-1}, k_{cat}/K_m = 1.8 x 10^5 $M^{-1}s^{-1}$), and p(Ap)A (k_{cat} = 14 s^{-1}, k_{cat}/K_m = 9.0 x 10^4 $M^{-1}s^{-1}$). Removal of the 5′-phosphoryl group from 3′,5′-bisphosphoadenosine resulted in 200-fold reduction in catalytic efficiency, as seen from the kinetic parameters for 3′-AMP (k_{cat} = 0.22 s^{-1}, k_{cat}/K_m = 500 $M^{-1}s^{-1}$). The enzyme was able to hydrolyze 2′-phosphate from 2′,5′-bisphosphoadenosine with a catalytic

efficiency of 100 $M^{-1}s^{-1}$, which is 1000-fold less than that of 3′,5′-diphosphoadenosine. The bisphosphate mixtures for cytidine, guanosine, and uridine contained 2′,5′- and 3′,5′-bisphosphates in a ratio of close to 1:1. Therefore, the initial rates of hydrolysis determined for these substrates were assumed to be for 3′,5′-bisphosphate isomers. The variation of the identity of the base in 3′,5′-bisphosphonucleotides did not have a significant effect on the kinetic parameters of hydrolysis by TrpH. In general, nanoRNA substrates had better catalytic parameters than nanoDNA substrates of the same length. TrpH was able to completely hydrolyze $p(dAp)_9dA$ when 50 μM substrate was incubated with 1 μM enzyme overnight. The enzyme was able to hydrolyze $p(Ap)_5$ at a rate comparable to that of $p(Ap)_4A$, but unable to hydrolyze $(Ap)_4A$ at a detectable rate.

Table 5.1. Kinetic parameters for TrpH with various substrates.

Compound	k_{cat}, s^{-1}	K_m (μM)	k_{cat}/K_m $M^{-1}s^{-1}$
pAp	9.87 ± 0.51	56.0 ± 6.0	(1.8 ± 0.1) x 10^5
2′-deoxy-pAp	5.60 ± 0.20	53.0 ± 6.5	(1.1 ± 0.1) x 10^5
pCp (2′5′/3′5′ mixture)*	6.0 ± 0.30	86.4 ± 9.7	(7.0 ± 0.5) x 10^4
2′-deoxy-pCp*	1.9 ± 0.10	114 ± 10	(1.7 ± 0.1) x 10^4
pGp (2′5′/3′5′ mixture)*	5.5 ± 0.30	45.4 ± 8.6	(1.2 ± 0.2) x 10^5
2′-deoxy-pGp*	1.7 ± 0.10	30.0 ± 5.9	(5.8 ± 0.6) x 10^4
pUp (2′5′/3′5′ mixture)*	4.7 ± 0.20	58.8 ± 8.2	(8.0 ± 0.8) x 10^4
2′-deoxy-pUp*	2.1 ± 0.10	110 ± 18	(1.9 ± 0.2) x 10^4
2′-deoxy-pTp*	1.6 ± 0.10	31.9 ± 5.0	(4.9 ± 0.6) x 10^4
2′-deoxy-pIp*	3.3 ± 0.30	48.0 ± 13	(6.9 ± 1.3) x 10^4
3′-AMP	0.22 ± 0.01	420 ± 80	(5.0 ± 0.6) x 10^2
2′,5′-pAp	ND**	ND**	100 ± 10
$p(G)_2$	0.60 ± 0.07	14.0 ± 7.0	(4.1 ± 1.8) x 10^4
$p(A)_2$	14 ± 0.50	159 ± 11	(9.0 ± 0.4) x 10^4
$p(A)_3$	3.8 ± 0.10	54.8 ± 4.8	(7.0 ± 0.4) x 10^4
$p(A)_4$	3.6 ± 0.10	45.1 ± 1.5	(7.9 ± 0.2) x 10^4
$p(A)_5$	2.7 ± 0.30	58.6 ± 16	(4.6 ± 0.8) x 10^4
$p(dA)_2$	2.50 ± 0.20	149 ± 20	(1.7 ± 0.1) x 10^4
$p(dA)_3$	0.72 ± 0.04	66.9 ± 10	(1.1 ± 0.1) x 10^4
$p(dA)_4$	1.20 ± 0.06	176 ± 18	(6.8 ± 0.4) x 10^3
$p(dA)_5$	0.79 ± 0.02	275 ± 17	(2.9 ± 0.1) x 10^3

*Compounds assayed using Pi detection assay. Other substrates were assayed using the adenylate kinase/guanylate kinase-pyruvate kinase-lactate dehydrogenase coupling assay.

**Not determined.

HPLC Trace Analysis of Hydrolysis of Short Single-stranded Oligonucleotides by TrpH. $p(dAp)_4dA$ (100 μM) was incubated with TrpH (50 nM), aliquots were removed and analyzed using anion exchange chromatography. The time course showed a steady accumulation of 5′-dAMP and consumption of $p(dAp)_4dA$ (**Figures 5.2A and 5.2B**). At intermediate time points, $p(dAp)_3dA$ and $p(dAp)_2dA$ were detected, as confirmed by determining the retention volumes of the corresponding standards under the same conditions. p(dAp)dA could not be detected. After overnight incubation, $p(dA)_4dA$ was completely hydrolyzed to 5′-dAMP.

In a similar experiment, 20 μM 5′-phosphorylated DNA hexamer (5′-pdApdApdGpdCpdApdA-3′) was incubated with 200 nM TrpH, and the hydrolysis was followed using anion exchange chromatography (**Figures 5.2C and 5.2D**). The enzymatic hydrolysis resulted in a steady accumulation of deoxynucleoside monophosphates with the 5-mer, 4-mer and 3-mer intermediates detected during the assay.

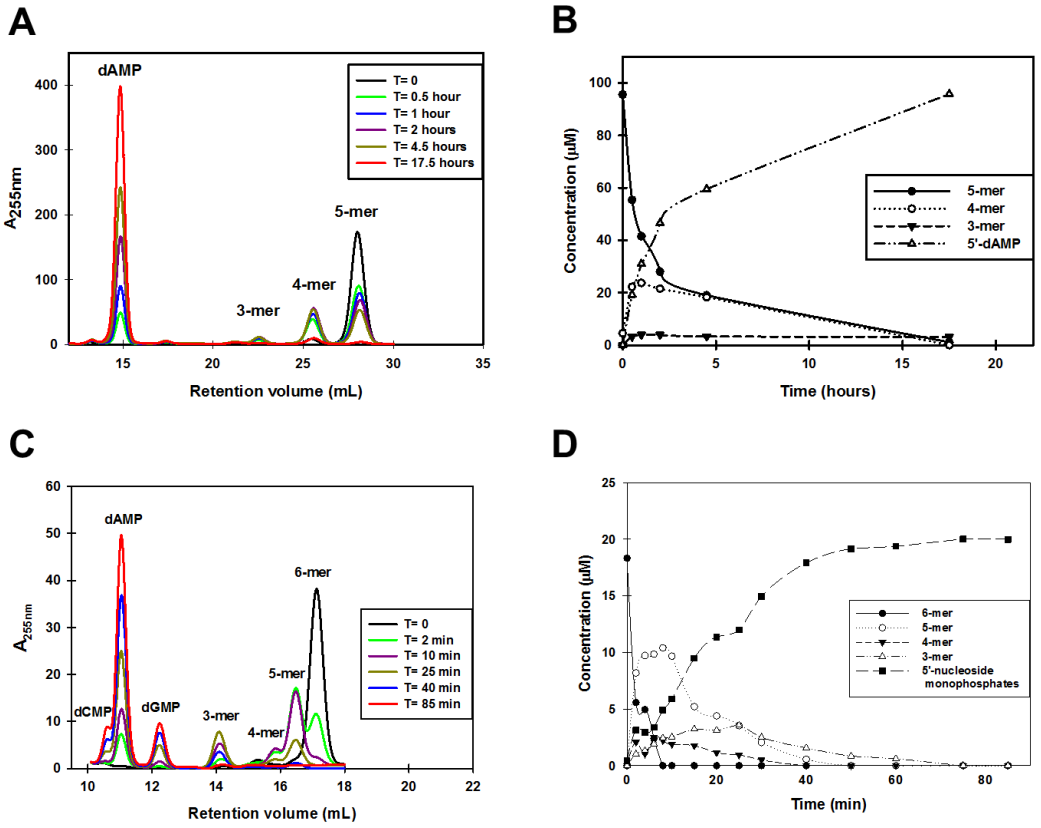


Figure 5.2. Intermediates of hydrolysis of DNA oligomers by TrpH analyzed using anion exchange chromatography. **A)** UV trace of hydrolysis of 100 μM $p(dA)_5$ by 50 nM TrpH over time. The intermediate species, $p(dA)_4$ and $p(dA)_3$ were detected, and verified by running standards separately. **B)** Plot of concentration versus time for all the species involved in the hydrolysis of $p(dA)_5$ by TrpH, shown in **A**. **C)** UV trace of hydrolysis of 20 μM p(d-AAGCAA) by 200 nM TrpH over time. The final products of hydrolysis 2′-deoxy-5′-adenosine monophosphate (dAMP), 2′-deoxy-5′-guanosine monophosphate (dAMP), and 2′-deoxy-5′-cytosine monophosphate (dAMP) were verified by running standards separately. **D)** Plot of concentration versus time for all the species involved in the hydrolysis of p(d-AAGCAA) by TrpH, shown in **C**. 5-mer is p(d-AGCAA), 4-mer is p(d-GCAA), and 3-mer is p(d-CAA).

Hydrolysis of Duplex Oligonucleotides by TrpH. DNA substrates **1-5** (**Figure 5.3**) were designed such that the 11-mer complementary sequence had a melting point of significantly higher than room temperature (T_m~68 °C).[85] The forward strand (**1**) contains 5′-phosphorylation, ensuring that the enzyme would hydrolyze it preferentially

as compared to hydrolyzing its reverse complement. For each assay, 20 μM substrate was incubated with 0.4-1.4 μM TrpH, and the initial rate of hydrolysis was determined using coupling assay described above. The specific rates of hydrolysis of 20 μM each of **1-5** were 0.10 s^{-1}, 0.076 s^{-1}, 0.0099 s^{-1}, 0.089 s^{-1}, and 0.0050 s^{-1} respectively. Overnight incubation of **1** and **3** with TrpH resulted in the complete hydrolysis of both substrates.

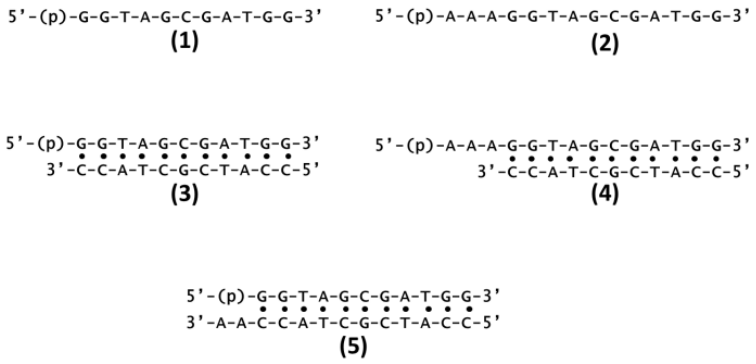


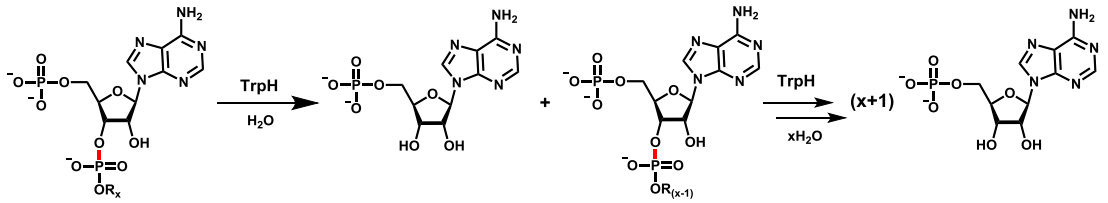
Figure 5.3. Double-stranded DNA substrates tested with TrpH. Complementary base-pairing is indicated as (●).The specific rate of hydrolysis for each substrate are given in the text.

DISCUSSION

Substrate Specificity of TrpH. TrpH efficiently catalyzes the hydrolysis of a phosphomonoester as well as a phosphodiester at the 3′-position of 3′,5′-bisphosphonucleotide to release a nucleoside monophosphate as one of the products (**Scheme 5.3**). The presence of 2′-hydroxyl functional group on the oligonucleotide substrates (RNA-derivatives) results in a 2-10 fold increase in catalytic efficiency as compared to 2′-deoxy substrates (DNA-derivatives). The variation of the identity of the

base in the 3′,5′-bisphosphonucleotide substrates did not have a significant effect on the kinetic parameters of the enzymatic hydrolysis by TrpH, as evidenced from the values reported in **Table 5.1**, as well as the hydrolysis of deoxyoligonucleotides shown in **Figures 5.2C and 5.2D**. The 5′-terminal phosphate enhances catalytic efficiency, as is evident from the 200-fold difference in the k_{cat}/K_m of hydrolysis of p(Ap) and 3′-AMP. The appearance and disappearance of oligonucleotides of intermediate length in **Figure 5.2** indicates that the enzyme hydrolyzes oligonucleotides bearing 5′-phosphate in a sequential manner resulting in an accumulation of 5′-phosphomononucleotide products. The kinetic parameters of hydrolysis, in general, tend to decrease with the length of the oligonucleotide substrates, but the enzyme is capable of hydrolyzing longer substrates (up to 14-mer tested in this study) over time. The enzyme is capable of hydrolyzing double-stranded oligonucleotides. However, the rate of hydrolysis of double-stranded DNA (**3**) was 10-fold lower than the corresponding single-stranded DNA (**1**). The enzyme was able to hydrolyze 5′- overhang on a double-stranded DNA with terminal 5′-phosphate (**4**) at a rate comparable to the corresponding single-stranded DNA (**2**). Double-stranded DNA with 3′-extension on the reverse complement (**5**) was hydrolyzed at a 2-fold lower rate as compared to **3**, and 20-fold lower rate as compared to the single-stranded DNA **1**. This suggests that TrpH may hydrolyze double-stranded DNA and RNA within the cell, albeit its primary role is likely to be hydrolysis of single-stranded nanoDNA or nanoRNA. It is unclear whether the enzyme hydrolyzes a 5′-phosphorylated double-stranded oligonucleotide by hydrolyzing the frayed end or if it can unwind the double-stranded oligonucleotide before hydrolysis at the 5′-

phosphorylated terminus. In summary, the catalytic activity of TrpH can be described as a nanoRNase, and a 5′-3′ exonuclease for 5′-phosphorylated oligonucleotides.



Scheme 5.3. Reaction catalyzed by TrpH.

Overview of Known Ribonucleases from *E. coli* and Other Organisms.[86]

RNA is found in three major forms in all living organisms: messenger (mRNA), transfer (tRNA) and ribosomal (rRNA). All organisms utilize several ribonucleases (RNases) to process RNA during its biogenesis and degradation. This processing, in turn, precisely regulates the expression of genetic information in the organism according to the changes in its environmental conditions.

In *E. coli* and other related Gram-negative organisms, RNase E, RNase G, and RNase III are the major endonucleases. RNase E cleaves RNA internally within single-stranded AU-rich regions, and shows a preference for RNAs with 5′-monophosphorylated and unpaired termini.[87] RNase G is a paralog of RNase E, and has a similar substrate specificity except that, RNase G is non-essential while RNase E is essential.[88] RNase III, unlike RNase E or RNase G, cuts RNA within double-stranded

regions to yield products that have a characteristic 2-bp overhang at the 3′-end and a 5′-monophosphate (similar to **5** shown in **Figure 5.3**).[89] The activity of these endonucleases is complemented by the exoribonucleases that degrade the products of the former to corresponding mononucleotides, which can get recycled or further processed in the cell. Polynucleotide phosphorylase (PNPase), RNase R, RNase II, and oligoribonuclease (Orn) are the exonucleases reported to be present in *E. coli*, and all of them display 3′-5′ exonuclease activity. PNPase preferentially degrades RNA with single-stranded 3′-end in a phosphorolytic manner.[90] RNase R and RNase II degrade RNA in a hydrolytic manner, and RNase R is capable of unwinding double-stranded RNA provided the 3′-end is single-stranded for initial binding of the enzyme to the RNA.[91] The end products of degradation of RNase R and RNase II were found to be 3- to 6-mers and 1- to 3-mers respectively.[91b, 92] These nanoRNAs (2-5 nt long) are degraded by Orn in *E. coli*, which has been found to be an essential gene.[93] Recently, nanoRNAs have been shown to be involved in priming transcription initiation in bacteria as they can serve as templates for RNA synthesis by DNA-dependent RNA polymerase, which explains the critical role played by Orn in maintaining cell homeostasis.[94] As widely reported till date, there is no known 5′-3′ exonuclease from *E. coli*.

In *B. subtilis* and related Gram-positive organisms[95], RNase Y and RNase J1 are the major endonucleases. Activity of RNase Y is similar to that of RNase E, and is present in organisms in which there is no homolog of RNase E.[96] RNase J was initially identified as an endonuclease[97], but later was shown to possess 5′-3′ exonuclease activity with a strong preference for 5′-monophosphorylated RNA.[98] This was the first reported

instance of 5′-3′ exonuclease among bacteria, and the exonuclease activity is similar to that of TrpH. PNPase, RNase R and RNase PH are the other exonucleases that degrade RNA in the 3′-5′ direction. The degradation of nanoRNA in *B. subtilis* and related organisms is carried out by NrnA/B enzymes, which have been shown to possess nanoRNase activity similar to Orn from *E. coli*.[83a, 99] In bacteria, the enzymes involved in the decay of RNA form a multimeric complex called an RNA degradosome. These complexes normally contain one or more ribonuclease(s), and an RNA helicase, and have been shown to be important for the efficient degradation of RNA.[100]

In eukaryotes, the degradation of RNA is triggered by the removal of the poly(A) tail at the 3′-end either by deadenylase in deadenylation-dependent pathway or an endonuclease in the deadenylation-independent pathway. Several enzymes have been annotated to carry out the two functions *in vivo*.[101] The exosome, which is a 10-12 subunit protein complex consisting of RNase PH-domain enzymes, RNA-binding proteins, RNase D-like enzymes, and RNA helicases, is responsible for 3′-5′ exonucleolytic degradation.[102] 5′-3′ degradation requires the prior removal of m7G cap at the 5′-terminus, which exposes the 5′-monophosphorylated terminus to hydrolysis by Xrn1 or Xrn2. Xrn1 is localized in the cytoplasm while Xrn2 acts in the nucleus. Both enzymes catalyze the exonucleolytic hydrolysis of 5′-monophosphorylated RNA in a processive manner to 5′-nucleoside monophosphates as final products.[103] In Archaea, RNA is degraded in the 5′-3′ direction by an ortholog of RNase J, and in the 3′-5′ direction, by the exosome. So far, there is no known endoribonuclease that plays a key role in RNA degradation in methanogenic archaea like *Sulfolobus solfataricus*.[104]

Potential Role of the TrpH Activity in *E. coli*. It is evident that 5′-3′ exonuclease activity has been found to exist in several classes of organisms. However, this activity has remained elusive in *E. coli* and related organisms. The exploration of the substrate specificity of TrpH has revealed that it is capable of hydrolyzing 5′-monophosphorylated RNA in 5′-3′ direction in a sequential manner. Why has this activity then remained unmanifested after several exhaustive studies on RNA metabolism in the *E. coli* model system? Some possible reasons may be: 1. TrpH is endogenously expressed at very low concentrations, making it incapable of taking over the function of the major exonucleases in their absence; 2. the conditions under which the expression of the TrpH gene is upregulated have not yet been identified fully or correctly; 3. the nanoRNA substrates getting hydrolyzed by TrpH may have their origins from sources other than mRNA, tRNA or rRNA degradation. One possible alternate source of 5′-phosphorylated dinucleotides is the hydrolysis of the bacterial secondary messenger (3′→5′)-cyclic-di-guanosine monophosphate (c-di-GMP). This cyclic dinucleotide has been shown to be involved in an extensive signal transduction system responsible for regulating biofilm formation, motility, virulence, the cell cycle, differentiation, and other processes. It has been demonstrated that GGDEF-domain containing diguanylate cyclases (DGCs) are responsible for its biosynthesis from two molecules of guanosine triphosphate (GTP), while the EAL-domain containing phosphodiesterases (PDEs) are responsible for its hydrolysis to $(pG)_2$.[105] The specific enzyme that catalyzes the hydrolysis of $(pG)_2$ to 5′-GMP remains unidentified.[106] The

substrate specificity of TrpH indicates that it is capable of performing the function of hydrolysis of $(pG)_2$ (k_{cat}/K_m= (4.1 ± 1.8) x 10^4 $M^{-1}s^{-1}$).

CONCLUSION

We have discovered that the enzyme sequence with the locus tag b1266 (TrpH or YciV), which had been annotated as a hypothetical protein, is capable of efficiently hydrolyzing single-stranded fragments of RNA possessing 5′-monophosphate in a 5′-3′ direction sequentially. This is the first instance of a 5′-3′ exonuclease activity reported for an enzyme sequence from *E. coli*. Considering the redundant nature of exoribonuclease substrate specificities and the data from the previous studies on RNA degradation in *E. coli*, the conditions under which TrpH may dispense the role of an exonuclease are unclear at present. Further studies should involve gene deletion and gene complementation to explore the precise physiological role of TrpH.

CHAPTER VI

SUMMARY AND CONCLUSIONS

The overall research objective of this work was to probe the reaction mechanism of an enzyme of known function, and then leverage it to discover the best substrates of homologous enzymes of unknown functions. The study commenced with the mechanistic characterization of L-histidinol phosphate phosphatase (HPP) from *Lactococcus lactis*. This reaction constitutes the penultimate step in the biosynthesis of the amino acid: L-histidine. The first stage of this study involved the optimization of metal content and activity of the recombinant enzyme. Mechanistic studies were conducted using pH-Rate profiles, solvent isotope and viscosity effects, and site-directed mutagenesis. The Almo lab at Albert Einstein College of Medicine, Bronx, was able to obtain crystal structures with and without ligands bound at the enzyme active site. These studies laid the foundation for determining the boundaries of HPP activity among closely related enzyme sequences within cog1387. Bioinformatic analysis of cog1387, and the verification of activity of eight other enzymes predicted to be L-histidinol phosphate phosphatases from this COG, enabled us to annotate about 24% of the 550 non-redundant sequences in cog1387 as authentic HPP enzymes.

The next study involved the discovery and mechanistic characterization of a novel and unique phosphodiesterase from cog0613. This gene was found localized in the *Phn* operon, enzymes from which are involved in converting organophosphonates to inorganic phosphate under conditions of phosphate starvation. The recombinant target enzyme from *Eggerthella lenta* was purified and characterized. Using exhaustive ^{31}P-

NMR spectroscopy, it was determined that this enzyme hydrolyzed 5-phosphoribose-1,2-cyclicphosphate (PRcP) to ribose-5-phosphate (R5P) and inorganic phosphate (Pi). Further experiments revealed that ribose-2,5-bisphosphate (2,5-RbP) was an intermediate in this enzymatic conversion. The mechanism involved sequential hydrolysis of the two P-O bonds of the cyclic phosphodiester, by the attack of nucleophilic hydroxide at the P-center in both steps. This is the first reported instance of a cyclic phosphodiesterase capable of hydrolyzing both ester bonds of a cyclic phosphate to P_i and a vicinal diol, and the enzyme was named cyclic phosphate dihydrolase (cPDH). The orthologs of this enzyme are found in at least 30 fully sequenced bacterial genomes, which include human gut pathogens *Clostridium difficile* and *Eggerthella lenta*.

Cv1693 from *Chromobacterium violaceum* bears a sequence identity of 29% with cPDH described above, and was shown to hydrolyze 3′-phosphate from 3′,5′-diphosphoadenosine (pAp) and other diphosphonucleotides. However, enzymes from at least two other COGs, i.e. 1218 and 0618, are capable of hydrolyzing pAp. Considering the substrate promiscuity of the target enzyme, and the presence of enzymes from other non-homologous families capable of carrying out the same molecular function, the physiological role of this enzyme was unclear. Bioinformatic analysis was used to determine the physiological relevance of this enzymatic reaction that had been characterized *in vitro*. First, the authentic pAp-3′phosphatases from the three COGs under study were determined based on literature reports, and a thorough sequence network analysis of the respective COGs was performed. Next, a list of bacteria from

the Microbesonline database was analyzed to determine the distribution of a pAp-3′-phosphatase from each of the 3 COGs in various organisms. This exercise revealed that the distribution of pAp-3′-phosphatase from each of the 3 families is largely mutually exclusive in bacteria. This work supported the hypothesis that the physiological role of the target enzyme *in vivo* is the hydrolysis of pAp, and organisms from different taxonomic phyla have evolved enzymes from different families to catalyze the same function.

The last study focused on the function discovery of TrpH from *Escherichia coli*. This enzyme shares 37% sequence identity with 3′,5′-diphosphoadenosine-3′-phosphatase (Cv1693) described above. TrpH was found to hydrolyze pAp efficiently. However, when the substrate profile of this enzyme was probed further, this enzyme was found capable of hydrolyzing oligonucleotides bearing a 5′-phosphoryl functional group with comparable efficiency. Biochemical characterization has revealed that this enzyme hydrolyzes RNA fragments better than DNA fragments. This enzyme is promiscuous with respect to the identity of the base at the 5′-end. This is the first reported oligoribonuclease from *E. coli* which catalyzes the hydrolysis of oligoribonucleotides sequentially from 5′→3′.

This work has helped us understand the mechanism of hydrolysis of trinuclear amidohydrolase enzymes from the PHP family. Besides, the successful discovery of new enzymatic functions of homologous enzymes has demonstrated the use of sequence similarity networks, and operon context analyses among bacterial genomes, as effective tools for enzyme function discovery. These studies are expected to contribute towards

increasing our understanding of the relationship between the primary sequence, structure and function of enzymes, and eventually accelerate the process of their accurate annotation.

REFERENCES

1. Schnoes, A. M., Brown,S.D.,Dodevski, I., Babbitt, P.C., Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol* **2009,** *5,* e1000605.

2. Gerlt, J. A., Allen,K.N., Almo,S.C., Armstrong,R.N., Babbitt,P.C., Cronan,J.E., Dunaway-Mariano,D., Imker,H.J., Jacobson,M.P., Minor,W., Poulter,C.D., Raushel,F.M., Sali,A., Shoichet,B.K., and Sweedler,J.V., The Enzyme Function Initiative. *Biochemistry* **2011,** *50,* 9950-9962.

3. Sorokina, M., Stam, M., Medigue, C., Lespinet, O., and Vallenet, D., Profiling the orphan enzymes. *Biology Direct* **2014,** *9,* 1-16.

4. Cravatt, B. F., Wright, A.T., and Kozarich, A.W., Activity based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* **2008,** *77,* 383-414.

5. Holm, L., and Sander, C., An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins: Struct., Funct., Genet.* **1997,** *28,* 72-82.

6. Seibert, C. M., and Raushel, F.M., Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry* **2005,** *44,* 6383-6391.

7. Liu, A., and Huo, L., Amidohydrolase superfamily. *eLS* **2014**.

8. Widlansky, T. S., and Taylor, W., Chemistry and enzymology of phosphatases. *Comprehensive natural products chemistry* **1999,** *5,* 139-162.

9. Van Etten, R. L., Human prostatic acid phosphatase: a histidine phosphatase. *Ann. NY Acad. Sci.* **1982,** *390,* 27-51.

10. Bigley, A. N., and Raushel, F.M., Catalytic mechanisms for phosphotriesterases. *Biochim. Biophys. Acta.* **2013,** *1834,* 443-453.

11. (a) Harper, L. L., McDaniel, C.S., Miller, C.E., and Wild, J.R., Dissimilar plasmids isolated from Pseudomonas diminuta MG and a Flavobacterium sp. (ATCC 27551) contain identical opd genes. *Appl. Environ. Microbiol.* **1988,** *54,* 2586-2589; (b) DeFrank, J. J., and Cheng, T.C. , Purification and properties of an organophosphorus acid anhydrase from a halophilic bacterial isolate. *J. Bacteriol.* **1991,** *173,* 1938-1943; (c) Hoskin, F. C. G., and Long, R.J., Purification of a DFP-hydrolyzing enzyme from squid head ganglion. *Arch. Biochem. Biophys.* **1972,** *150,* 548-555; (d) Furlong, C. E., Richter, R.J., Chapline, C., and Crabb, J.W., Purification of rabbit and human serum paraoxonase. *Biochemistry* **1991,** *30,* 10133-10140.

12. Samples, C. R., Howard, T., Raushel, F.M., and DeRose, V.J., Protonation of the binuclear metal center within the active site of phosphotriesterase. *Biochemistry* **2005,** *44*, 11005-11013.

13. Omburo, G. A., Kuo, J.M., Mullins, L.S., and Raushel, F.M., Characterization of the zinc binding site of bacterial phosphotriesterase. *J. Biol. Chem.* **1992,** *267*, 13278-13283.

14. Ely, F., Hadler, K.S., Mitic, N., Gahan, L.R., Ollis, D.S., Plugis, N.M., Russo, M.T., Larrabee, J.A., and Schenk, G., Electronic and geometric structures of the organophosphate-degrading enzyme from *Agrobacterium radiobacter* (OpdA). *J. Biol. Inorg. Chem.* **2011,** *16*, 777-787.

15. Lewis, V. E., Donarski, W.J., Wild, J.R., and Raushel, F.M., Mechanism and stereochemical course at the phosphorus of the reaction catalyzed by a bacterial phosphotriesterase. *Biochemistry* **1988,** *27*, 1591-1597.

16. Samples, C. R., Raushel, F.M., and DeRose, V.J., Activation of the binuclear metal center through formation of phosphotriesterase-inhibitor complexes. *Biochemistry* **2007,** *46*, 3435-3442.

17. Hong, S. B., and Raushel, F.M., Metal-substrate interactions facilitate the catalytic activity of the bacterial phosphotriesterase. *Biochemistry* **1996,** *35*, 10904-10912.

18. Caldwell, S. R., Raushel, F.M., Weiss, P.M., and Cleland, W.W., Transition-state structures for enzymatic and alkaline phosphotriester hydrolysis. *Biochemistry* **1991,** *30*, 7444-7450.

19. Wong, K. Y., and Gao, J., The reaction mechanism of paraoxon hydrolysis by phosphotriesterase from combined QM/MM simulations. *Biochemistry* **2007,** *46*, 13352-13369.

20. Aubert, S. D., Li, Y., and Raushel, F.M., Mechanism for the hydrolysis of organophosphates by bacterial phosphotriesterase. *Biochemistry* **2004,** *43*, 5707-5715.

21. Vanhooke, J. L., Benning, M.M., Raushel, F.M., and Holden, H.M., Three-dimensional structure of the zinc-containing phosphotriesterase with the bound substrate analog diethyl 4-methylbenzylphosphonate. *Biochemistry* **1996,** *35*, 6020-6025.

22. Chen-Goodspeed, M., Sogorb, M.A., Wu, F., Hong, S.B., and Raushel, F.M., Structural determinants of the substrate and stereochemical specificity oh phosphotriesterase. *Biochemistry* **2001,** *40*, 1325-1331.

23. (a) Tsai, P. C., Bigley, A.N., Li, Y., Ghanem, E., Cadieux, C.L., Kasten, S.A., Reeves, T.E., Cerasoli, D.M., and Raushel, F.M., Selective hydrolysis of organophosphate nerve agents by the bacterial phosphotriesterase. *Biochemistry* **2010,** *49*, 7978-7987; (b) Li, W. S., Li, Y., Hill, C.M., Lum, K.T., and Raushel, F.M., Enzymatic synthesis of chiral organophosphothioates from prochiral precursors. *J. Am. Chem. Soc.* **2002,** *124*, 3498-3499; (c) Li, Y., Aubert, S.D., and Raushel, F.M., Operational control of stereoselectivity during the enzymatic hydrolysis of racemic organophosphorus compounds. *J. Am. Chem. Soc.* **2003,** *125*, 7526-7527; (d) Hill, C. M., Li, W.S., Thoden, J.B., Holden, H.M., and Raushel, F.M., Enhanced degradation of chemical warfare agents through molecular engineering of the phosphotriesterase active site. *J. Am. Chem. Soc.* **2003,** *125*, 8990-8991; (e) Li, Y., Aubert, S.D., Maes, E.G., and Raushel, F.M., Enzymatic resolution of chiral phosphinate esters. *J. Am. Chem. Soc.* **2004,** *126*, 8888-8889.

24. (a) Wexler, M., Sargent, F., jack, R.L., Stanley, N.R., Bogsch, E.G., Robinson, C., Berks, B.C., and Palmer, T., TatD is a cytoplasmic protein with DNase activity. No requirement for TatD family proteins in sec-dependent protein export. *J. Biol. Chem.* **2000,** *275*, 16717-16722; (b) Lindenstrauss, U., Matos, C.F., Graubner, W., Robinson, C., and Bruser, T., Malfolded recombinant Tat substrates are Tat-independently degraded in *Escherichia coli*. *FEBS Lett.* **2010,** *584*, 3644-3648; (c) Matos, C. F. R. O., Di Cola, A., and Robinson, C., TatD is a central component of a Tat transloconinitiated quality control system for exported FeS proteins in *Escherichia coli*. *EMBO Rep* **2009,** *10*, 474-479.

25. (a) Qiu, J., Yoon, J-H., and Shen, B., Search for apoptotic nucleases in yeast: role of TatD nuclease in apoptotic DNA degradation. *J. Biol. Chem.* **2005,** *280*, 15370-15379; (b) Gannavaram, S., and Debrabant, A., Involvement of TatD nuclease during programmed cell death in the protozoan parasite *Trypanosoma brucei*. *Mol. Microbiol.* **2012,** *83*, 926-935.

26. Chen, Y.-C., Li, C-L., Hsiao, Y-Y., Duh, Y., and Yuan, H.S., Structure and function of TatD exonuclease in DNA repair. *Nucleic Acids Res.* **2014,** *42*, 10776-10785.

27. Mijakovic, I., Poncet, S., Boel, G., Maze, A., Gillet, S., Jamet, E., Decottignies, P., Grangeasse, C., Doublet, P., Le Marechal, P., and Deutscher, J., Transmembrane modulator-dependent bacterial tyrosine kinase activates UDP-glucose dehydrogenases. *EMBO J.* **2003,** *22*, 4709-4718.

28. (a) Bender, M. H., and Yother, J., CpsB is a modulator of capsule-associated tyrosine kinase activity in *Streptococcus pneumoniae*. *J. Biol. Chem.* **2001,** *276*, 47966-47974; (b) Morona, J. K., Morona, R., Miller, D.C., and Patton, J.C., Streptococcus pneumoniae capsule biosynthesis protein CpsB is a novel manganese-dependent phosphotyrosine-protein phosphatase. *J. Bacteriol.* **2002,** *184*, 577-583.

29. Hagelueken, G., Huang, H., Mainprize, I.L., Whitfield, C., and Naismith, J.H. , 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.* **2009,** *392* 678-688.

30. Alifano, P., Fani, R., Lio, P., Lazcano, A., Bazzicalupo, M., Carlomagno, M.S., Bruni, C.B. , Histidine biosynthetic pathway and genes: structure, regulation, and evolution. *Microbiol. Rev.* **1996,** *60*, 44-69.

31. Thaller, M. C., Schippa, S., Rossolini, G.M. , Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. *Protein Sci.* **1998,** *7*, 1647-1652.

32. Rangarajan, E. S., Proteau, A., Wagner, J., Hung, M-N., Matte, A., Cygler, M., Structural snapshots of *Escherichia coli* histidinol phosphate phosphatase along the reaction pathway. *J. Biol. Chem.* **2006** *281*, 37930-37941.

33. Le Coq, D., Fillinger, S., Ayemerich, S. J., Histidinol phosphate phosphatase, catalyzing the penultimate step of the histidine biosynthesis pathway, is encoded by YtvP (hisJ) in *Bacillus subtilis. J. Bacteriol.* **1999,** *181*, 3277-3280.

34. Aravind, L., Koonin, E.V., Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res.* **1998,** *26*, 3746-3752.

35. Omi R, G. M., Miyahara I, Manzoku M, Ebihara A, Hirotsu K., Crystal structure of monofunctional histidinol phosphate phosphatase from *Thermus thermophilus* HB8. *Biochemistry* **2007,** *46* (44), 12618-12627.

36. (a) Thoden, J. B., Phillips, G.N. Jr, Neal, T.M., Raushel, F.M., and Holden, H.M., Molecular structure of dihydroorotase: a paradigm for catalysis through the use of a binuclear metal center. *Biochemistry* **2001,** *40*, 6989-6997; (b) Lai, W. L., Chou, L.Y., Ting, C.Y., Kirby, R., Tsai, Y.C., Wang, A.H., and Liaw, S.H., The functional role of the binuclear metal center in D-aminoacylase: one-metal activation and second-metal attenuation. *J. Biol. Chem.* **2004,** *279*, 13962-13967; (c) Kim J, T. P., Chen SL, Himo F, Almo SC, Raushel FM., Structure of diethyl phosphate bound to the binuclear metal center of phosphotriesterase. *Biochemistry* **2008,** *47*, 9497-9504.

37. Tatusov, R. L., Galperin, M.Y., Natale, D.A., and Koonin, E.V., The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nuc. Acids. Res.* **2000,** *28*, 33-36.

38. (a) Atkinson, H. J., Morris, J.H., Ferrin, T.E., and Babbitt, P.C., Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS ONE* **2009,** *4*, e4345; (b) Apeltsin, L., Morris, J.H., Babbitt, P.C., and Ferrin, T.E., Improving the quality of protein similarity network clustering

algorithms using the network edge weight distribution. . *Bioinformatics* **2011,** *27*, 326-333.

39. Delorme, C., Ehrlich, C.D., and Renault, P., Histidine biosynthesis genes in *Lactococcus lactis* subsp. lactis. *J. Bacteriol.* **1992,** *174*, 6571-6579.

40. Aslanidis, C., and de Jong, P.J. , Ligation-independent cloning of PCR products (LIC-PCR). . *Nuc. Acids. Res.* **1990,** *18*, 6069-6074.

41. Studier, F. W., Protein production by auto-induction in high density shaking cultures. . *Protein Expr. Purif.* **2005,** *41*, 207-234.

42. Kamat SS, B. A., Kumaran D, Holmes-Hampton GP, Fan H, Sali A, Sauder JM, Burley SK, Lindahl PA, Swaminathan S, Raushel FM, Catalytic mechanism and three-dimensional structure of adenine deaminase. *Biochemistry* **2010,** *50*, 1917-1927.

43. Otwinowski, Z., and Minor, W. , Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* Carter, C. W. J., Sweet, R. M., Abelson, J. N., and Simon, M. I., Ed. Academic Press: New York, 1997.

44. Long, F., Vagin, A., Young, P., and Murshudov, G.N. , BALBES: a Molecular Replacement Pipeline. *Acta Crystallogr.* **2008,** *D64*, 125-132.

45. Emsley, P., and Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **2004,** *D60*, 2126-2132.

46. 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., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.* **2010,** *D66*, 213-221.

47. Lamzin, V. S., and Wilson, K.S. , Automated refinement of protein models. *Acta Crystallogr.* **1993,** *D49*, 129-147.

48. Hooft, R. W., Vriend, G., Sander, C., and Abola, E.E., Errors in protein structures. *Nature* **1996,** *381*, 272-273.

49. 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., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr.* **2010,** *D66*, 12-21.

50. Collaborative Computational Project, N., The CCP4 Suit: Programs for Protein Crystallography. *Acta Crystallogr.* **1994,** *50*, 760-763.

51. DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific LLC: San Carlos, CA, 2002.

52. Cheng, N.-S., Formula for the viscosity of a glycerol-water mixture. *Ind. Eng. Chem. Res.* **2008,** *47*, 3285-3288.

53. Waterhouse, A. M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J., Jalview version 2- a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009,** *25*, 1189-1191.

54. Smoot, M., Ono, K., Ruscheinski, J., Wang, P-L., and Ideker, T., Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **2011,** *27*, 431-432.

55. Holm, L., and Rosenstrom, P., DALI server: conservation mapping in 3D. *Nuc. Acids Res.* **2010,** *38*, 545-549.

56. Hall, R. S., Fedorov, A.A., Xu, C., Fedorov, E.V., Almo, S.C., and Raushel, F.M., Three-dimensional structure and catalytic mechanism of cytosine deaminase. *Biochemistry* **2011,** *50*, 5077-5085.

57. Hobbs, M. E., Malashkevich, V., Williams, H.J., Xu, C., Sauder, M., Burley, S.K., Almo, S.C., and Raushel, F.M., Structure and catalytic mechanism of LigI: Insight into the amidohydrolase enzymes of cog3618 and lignin degradation. . *Biochemistry* **2012,** *51*, 3497-3507.

58. Martí-Arbona, R., Fresquet, V., Thoden, J.B., Davis, M.L., Holden, H.M., and Raushel, F.M., Mechanism of the reaction catalyzed by isoaspartyl dipeptidase from Escherichia coli. *Biochemistry* **2005,** *44*, 7115-7124.

59. Cook, P. F., and Cleland, W.W. , In *Enzyme kinetics and mechanism* Rogers, R. L. S., Ed. Garland Science Publishing: New York, 2007; pp 325-357.

60. Chen, C. M., Ye, Q.Z., Zhu, Z.M., Wanner, B.L., and Walsh, C.T., Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the phn (psiD) genes involved in alkylphosphonate uptake and C-P lyase activity in *Escherichia coli* B. *J. Biol. Chem.* **1990,** *265*, 4461-4471.

61. Metcalf, W. W., and Wanner, B.L., Evidence for a fourteen-gene, phnC to phnP locus for phosphonate metabolism in *Escherichia coli*. *Gene* **1993,** *129*, 27-32.

62. Kamat S.S., W., H.J., Dangott, L.J., Chakrabarti, M., and Raushel, F.M., The catalytic mechanism for aerobic formation of methane by bacteria. *Nature* **2013,** *497*, 132-136.

63. Kamat, S. S., Williams, H.J., and Raushel, F.M. , Intermediates in the transformation of phosphonates to phosphate by bacteria. *Nature* **2011,** *2011*, 570-573.

64. He, S. M., Wathier, M., Podzelinska, K., Wong, M., McSorley, F.R., Asfaw, A., Hove-Jensen, B., Jia, Z., and Zechel, D.L., Structure and mechanism of PhnP, a phosphodiesterase of the carbon-phosphorus lyase pathway. *Biochemistry* **2011,** *50*, 8603-8615.

65. Hove-Jensen, B., McSorley, F.R., and Zechel, D.L. , Physiological role of phnP-specified phosphoribosyl cyclic phosphodiesterase in catabolism of organophosphonic acids by the carbon-phosphorus lyase pathway. *J. Am. Chem. Soc.* **2011,** *133*, 3617-3624.

66. Hove-Jensen, B., McSorley, F.R., and Zechel, D.L. , Catabolism and detoxification of 1-aminoalkylphosphonic acids: N-acetylation by the phnO gene product. *PLoS ONE* **2012,** *7*, e46416.

67. Leulliot, N., Cladière, L., Lecointe, F., Durand, D., Hübscher, U., and van Tilbeurgh, H., The family X DNA polymerase from Deinococcus radiodurans adopts a non-standard extended conformation. *J. Biol. Chem.* **2009,** *284*, 11992-11999.

68. Ghodge, S. V., Fedorov, A.A., Fedorov E.V., Hillerich, B., Sidel, R., Almo, S.C., and Raushel, F.M. , Structural and mechanistic characterization of L-histidinol phosphate phosphatase from the polymerase and histidinol phosphatase family of proteins. *Biochemistry* **2013,** *52*, 1101-1112.

69. Fathi, R., and Jordan, F., Certain novel ribofuranosyl phosphates derived from 5-phospho-.alpha.-D-ribofuranosyl-1-pyrophosphate: synthesis, structure, and alkaline hydrolytic reactivities. *J. Org. Chem.* **1988,** *53*, 1997-2001.

70. Cohn, M., and Hu, A., Isotopic (^{18}O) shift in ^{31}P nuclear magnetic resonance applied to a study of enzyme-catalyzed phosphate—phosphate exchange and phosphate (oxygen)—water exchange reactions. *Proc. Natl. Acad. Sci. USA* **1978,** *75*, 200-203.

71. Villafranca, J. J., and Raushel, F.M. , Biophysical applications of NMR to phosphoryl transfer enzymes and metal nuclei of metalloproteins. *Ann. Rev. Biophys. Bioeng.* **1980,** *9*, 363-392.

72. Teplyakov, A., Obmolova, G., Khil, P.P., Howard, A.J., Camerini-Otero, R.D., and Gilliland, G,L. , Crystal structure of the Escherichia coli YcdX protein reveals a trinuclear zinc active site. *Proteins: Struct., Funct., Genet.* **2003,** *51*, 315-318.

73. Kim, H. S., Lee, S.J., Yoon, H.J., An, D.R., Kim, D.J., Kim, S-J., and Suh, S.W. , Crystal structures of YwqE from *Bacillus subtilis* and CpsB from *Streptococcus*

pneumoniae, unique metal-dependent tyrosine phosphatases. . *J. Struct. Biol.* **2011** *175* 442-450.

74. 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., Crystal structure of a metal-dependent phosphoesterase (YP_910028.1) from Bifidobacterium adolescentis: Computational prediction and experimental validation of phosphoesterase activity. *Prot: Struc. Func. Bioinform.* **2011,** *79*, 2146-2160.

75. Ghodge, S. V., Cummings, J. A., Williams, H. J., and Raushel, F. M. , Discovery of a cyclic phosphodiesterase that catalyzes the sequential hydrolysis of both ester bonds to phosphorus. *J. Am. Chem. Soc.* **2013,** *135*, 16360-16363.

76. Barrio, J. R., Barrio, M.C.G, Leonard, N.J., England, T.E., and Uhlenbeck, O.C., Synthesis of modified nucleoside 3′,5′-bisphosphates and their incorporation into oligoribonucleotides with T4 RNA ligase. *Biochemistry* **1978,** *17*, 2077-2081.

77. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. , HKL-3000: the integration of data reduction and structure solution from diffraction images to an initial model in minutes. *Acta Cryst.* **2006,** *D62*, 859-866.

78. 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., MicrobesOnline: an integrated portal for comparative and functional genomics. *Nuc. Acids Res.* **2009,** *38*, 396-400.

79. Krissinel, E., and Henrick, K., Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. . *Acta Cryst.* **2004,** *D60*, 2256-2268.

80. (a) Hatzios, S. K., Iavarone, A. T., and Bertozzi, C. R. , Rv2131c from Mycobacterium tuberculosis is a CysQ 3′-phosphoadenosine-5′-phosphatase. *Biochemistry* **2008,** *47*, 5823-5831; (b) Mootz, H. D., Finking, R., and Marahiel, M. A. , 4′-Phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. *J. Biol. Chem.* **2001,** *276*, 37289-37298.

81. Neuwald, A. F., Krishnan, B.R., Brikun, I., Kulakausas, S., Suziedelis, K., Tomcsanyi, T., Leyh, T.S., and Berg, D.E. , CysQ, a gene needed for cysteine synthesis in Escherichia coli K-12 only during aerobic growth. . *J. Bacteriol.* **1992,** *174*, 415-425.

82. Spielberg, B. D., Xiong, J-P., Smith, J.J., Gu, R.F., and York, J.D. , Cloning and characterization of a mammalian Lithium-sensitive bisphosphate 3′-nucleotidase inhibited by inositol 1,4-bisphosphate. *J. Biol. Chem.* **1999,** *274*, 13619-13628.

83. (a) Mechold, U., Fang, G., Ngo, S., Ogryzko, V., and Danchin, A. , YtqI from Bacillus subtilis has both oligoribonuclease and pAp-phosphatase activity. *Nucl. Acids Res.* **2007,** *35*, 4552-4561; (b) Postic, G., Danchin, A., and Mechold, U. , Characterization of NrnA homologs from *Mycobacterium tuberculosis* and *Mycoplasma pneumoniae*. *RNA* **2012,** *18*, 155-165.

84. Cummings, J. A., Vetting, M., Ghodge, S.V., Xu, C., Hillerich, B., Seidel, R.D., Almo, S.C., and Raushel, F.M., Prospecting for unannotated enzymes: discovery of a 3′,5′-nucleotide bisphosphate phosphatase within the amidohydrolase superfamily. *Biochemistry* **2014,** *53*, 591-600.

85. SantaLucia, J. J., Allawi, H.T., and Seneviratne, P.A., Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **1996,** *35*, 3555-3562.

86. Hui, M. P., Foley, P.L., and Belasco, J.G., Messenger RNA degradation in bacterial cells. *Annu. Rev. Genet.* **2014,** *48*, 537-559.

87. McDowall, K. J., Lin-Chao, S., and Cohen, S.N., A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J. Biol. Chem.* **1994,** *269*, 10790-10796.

88. (a) Lee, K., Bernstein, J.A., and Cohen, S.N., RNase G complementation of *rne* null mutation identifies functional interrelationships with RNase E in *Escherichia coli*. *Mol. Microbiol.* **2002,** *43*, 1445-1456; (b) Deana, A., and Belasco, J.G., The function of RNase G in Escherichia coli is constrained by its amino and carboxyl termini. *Mol. Microbiol.* **2004,** *51*, 1205-1217.

89. Robertson, H. D., Webster, R.E., and Zinder, N.D., Purification and properties of ribonuclease III from *Escherichia coli*. *J. Biol. Chem.* **1968,** *243*, 82-91.

90. (a) Chen, L. H., Emory, S.A., Bricker, A.L., Bouvet, P., and Belasco, J.G., Structure and function of a bacterial mRNA stabilizer: analysis of the 5′ untranslated region of *ompA* mRNA. *J. Bacteriol.* **1991,** *173*, 4578-4586; (b) Xu, F., and Cohen, S.N., RNA degradation in Escherichia coli regulated by 3′ adenylation and 5′ phosphorylation. *Nature* **1995,** *374*, 180-183.

91. (a) Spickler, C., and Mackie, G.A., Action of Rnase II and polynucleotide phosphorylase against RNAs containing stem-loops of defined structure. *J. Bacteriol.* **2000,** *182*, 2422-2427; (b) Cheng, Z. F., and Deutscher, M.P., Purification and characterization of Escherichia coli exoribonuclease RNase R. Comparison with RNase II. *J. Biol. Chem.* **2002,** *277*, 21624-21629.

92. (a) Amblar, M., Barbas, A., Fialho, A.M., and Arraiano, C.M., Characterization of the functional domains of Escherichia coli RNase II. *J. Mol. Biol.* **2006,** *360*, 921-933; (b) Amblar, M., Barbas, A., Gomez-Puertas, P., and Arraiano, C.M., The role of the S1

domain in exoribonucleolytic activity: substrate specificity and multimerization. *RNA* **2007,** *13*, 317-327; (c) Vincent, H. A., and Deutscher, M.P., Substrate recognition and catalysis by the exoribonuclease RNase R. *J. Biol. Chem.* **2006,** *281*, 29769-29775.

93. Ghosh, S., and Deutscher, M.P., Oligoribonuclease is an essential component of the mRNA decay pathway. *Proc. Natl. Acad. Sci. USA* **1999,** *96*, 4372-4377.

94. Nickels, B. E., and Dove, S.L., NanoRNAs: A class of small RNAs that can prime transcription initiation in bacteria. *J. Mol. Biol.* **2011,** *412*, 772-781.

95. Bechhofer, D. H., *Bacillus subtilis* mRNA decay: new parts in the toolkit. *RNA* **2011,** *2*, 387-394.

96. Shahbabian, K., Jamalli, A., Zig, L., and Putzer, H., RNase Y, a novel endoribonuclease, initiates riboswitch turnover in Bacillus subtilis. *EMBO J.* **2009,** *28*, 3523-3533.

97. Even, S., Pellegrini, O., Zig, L., Labas, V., Bréchemmier-Baey, D., and Putzer, H., Ribonucleases J1 and J2: two novel endoribonucleases in *B. subtilis* with functional homology to *E. coli* RNase E. *Nucleic Acids Res.* **2005,** *33*, 2141-2152.

98. (a) Mathy, N., Benard, L., Pellegrini, O., Daou, R., Wen, T., and Condon, C., 5′-to-3′ exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5′ stability of mRNA. *Cell* **2007,** *129*, 681-692; (b) Richards, J., Liu, Q., Pellegrini, O., Celesnik, H., Yao, S., Bechhofer, D.H., Condon, C., Belasco, J.G., An RNA pyrophosphohydrolase triggers 5′-exonucleolytic degradation of mRNA in *Bacillus subtilis*. *Mol. Cell* **2011,** *43*, 940-949.

99. (a) Fang, M., Zeisberg, W.M., Condon, C., Ogryzko, V., Danchin, A., and Mechold, U., Degradation of nanoRNA is performed by multiple redundant RNases in *Bacillus subtilis*. *Nucleic Acids Res.* **2009,** *37*, 5114-5125; (b) Liu, M. F., Cescau, S., Mechold, U., Wang, J., Cohen, D., Danchin, A., Boulouis, H.J., and Biville F., Identification of a novel nanoRNase in Bartonella. *Microbiology* **2012,** *158*, 886-895.

100. (a) Py, B., Higgins, C.F., Krisch, H.M., and Carpousis, A.J., A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. *Nature* **1996,** *381*, 169-172; (b) Vanzo, N. F., Li, Y.S, Py, B., Blum, E., Higgins, C.F., Raynal, L.C., Krisch, H.M., and Carpousis, A.J., Ribonuclease E organizes the protein interactions in the *Escherichia coli* RNA degradosome. *Genes Dev.* **1998,** *12*, 2770-2781; (c) Commichau, F. M., Rothe, F.M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., Hammer, E., Völker, U., and Stülke, J., Novel activities of glycolytic enzymes in *Bacillus subtilis*: interactions with essential proteins involved in mRNA processing. *Mol. Cell Proteomics* **2009,** *8*, 1350-1360; (d) Lehnik-Habrink, M., Pförtner, H., Rempeters, L., Pietack, N., Herzberg, C., and Stülke, J., The RNA degradosome in *Bacillus subtilis*: identification of CshA as the major RNA helicase in the multiprotein complex. *Mol. Microbiol.* **2010,** *77*, 958-

971; (e) Roux, C. M., DeMuth, J.P., and Dunman, P.M., Characterization of components of the Staphylococcus aureus mRNA degradosome holoenzyme-like complex. *J. Bacteriol.* **2011,** *193*, 5520-5526.

101. Belasco, J. G., All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nat. Rev. Mol. Cell. Biol.* **2010,** *11*, 467-478.

102. Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A., RNA decay machines: the exosome. *Biochim. Biophys. Acta.* **2013,** *1829*, 552-560.

103. Miki, T. S., and Grobhans, H., The multifunctional RNase XRN2. *Biochem. Soc. Trans.* **2013,** *41*, 825-830.

104. Evguenieva-Hackenberg, E., and Bläsi, U., Attack from both ends: mRNA degradation in the crenarchaeon *Sulfolobus solfataricus*. *Biochem. Soc. Trans.* **2013,** *41*, 379-383.

105. Römling, U., Galperin, M.Y., and Gomelskyc, M., Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* **2013,** *77*, 1-52.

106. Ross, P., Aloni, Y., Weinhouse, H., Michaeli, D., Weinberger-Ohana, P., Mayer, R., and Benziman, M., Control of cellulose synthesis in *Acetobacter xylinum*. A unique guanyl oligonucleotide is the immediate activator of the cellulose synthase. *Carbohydr. Res.* **1986,** *149*, 101-117.

107. Korczynska, M., Xiang, D.F., Zhang, Z., Xu, C., Narindoshvili, T., Kamat, S.S., Williams, H.J., Chang, S.S., Kolb, P., Hillerich, B., Sauder, J.M., Burley, S.K., Almo, S.C., Swaminathan, S., Shoichet, B.K., Raushel, F.M., Functional annotation and structural characterization of a novel lactonase hydrolyzing D-xylono-1,4-lactone-5-phosphate and L-arabino-1,4-lactone-5-phosphate. *Biochemistry* **2014**, *53*, 4727-4728.

APPENDIX A

LIST OF NON-REDUNDANT PROTEIN SEQUENCES ANNOTATED AS L-HISTIDINOL PHOSPHATE PHOSPHATASES FROM COG1387

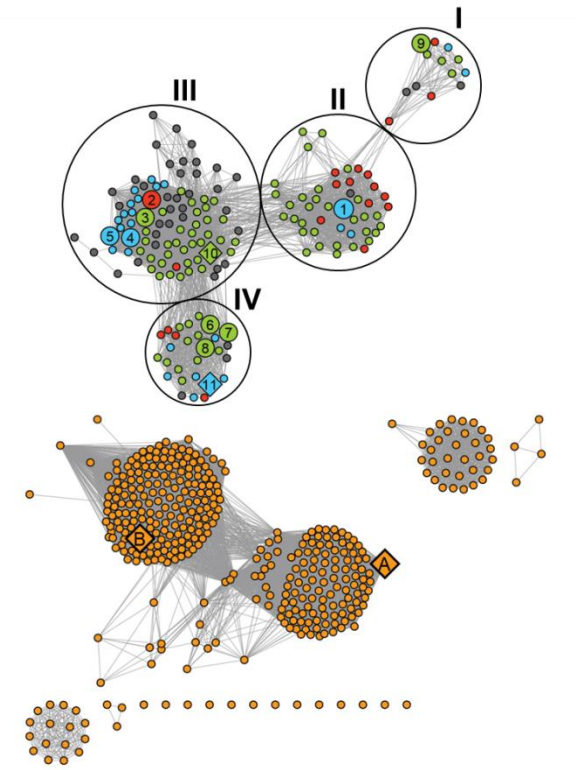


Figure S1: Sequence network diagram of cog 1387 (same as **Figure 2.2**), with HPP enzymes and close homologs grouped into four arbitrary zones I, II, III, and IV. The details of each node (protein) are provided in tables S1-S4.

Table S1: List of genes represented by blue nodes in **Figures 2.2 and S1**. These are authentic HPP enzymes which are found in the vicinity of the L-histidine biosynthesis operon on the genome of the organism.

Organism	Locus tag	GI number	Sequence identity to L37351	Zone
Bacillus anthracis str. Ames	BA_1432	30261511	22%	III
Bacillus anthracis str. 'Ames Ancestor'	GBAA_1432	47526706	22%	III
Bacillus anthracis str. Sterne	BAS1324	49184341	22%	III
Bacillus cereus 03BB102	BCA_1468	225863376	23%	III
Bacillus cereus AH187	BCAH187_A1572	217958990	21%	III
Bacillus cereus AH820	BCAH820_1504	218902621	22%	III
Bacillus cereus ATCC 10987	BCE_1533	42780607	21%	III
Bacillus cereus ATCC 14579	BC1413	161511209	21%	III
Bacillus cereus B4264	BCB4264_A1466	218234061	21%	III
Bacillus cereus E33L	BCZK1298	52143930	23%	III
Bacillus cereus G9842	BCG9842_B3878	218896444	20%	III
Bacillus cereus Q1	BCQ_1486	222095146	21%	III
Bacillus thuringiensis serovar konkukian str. 97-27	BT9727_1297	49477223	25%	III
Bacillus thuringiensis str. Al Hakam	BALH_1268	118476974	23%	III
Bacillus weihenstephanensis KBAB4	**BcerKBAB4_1335**	163939326	24%	III
Clostridium botulinum E3 str. Alaska E43	CLH_2622	188588114	31%	II
Clostridium kluyveri DSM 555	CKL_1305	153953930	31%	II
Exiguobacterium sibiricum 255-15	Exig_2369	172058376	24%	IV
Lactobacillus fermentum IFO 3956	LAF_0754	184155230	22%	IV
Lactococcus lactis subsp. cremoris MG1363	llmg_1288	125624110	80%	II
Lactococcus lactis subsp. cremoris SK11	LACR_1325	116512055	79%	II
***Lactococcus lactis* subsp. lactis Il1403**	**L37351**	**15673198**	**100%**	**II**
Listeria innocua Clip11262	LMHCC_2060	16799654	26%	IV
Listeria monocytogenes EGD-e	lmo0570	16802613	25%	IV
Listeria monocytogenes HCC23	LMHCC_2060	217965335	26%	IV
Listeria monocytogenes serotype 4b str. F2365	**LMOf2365_0599**	46906815	25%	IV
Listeria welshimeri serovar 6b str. SLCC5334	lwe0536	116871956	36%	IV
Natranaerobius thermophilus JW/NM-WN-LF	Nther_1995	188586605	25%	III
Oceanobacillus iheyensis HTE831	OB0554	23098009	33%	II

Streptococcus gordonii str. Challis substr. CH1	SGO_1401	157150565	26%	I
Streptococcus sanguinis SK36	SSA_1439	125718249	25%	I

Table S2: List of genes represented by green nodes in **Figures 2.2 and S1**. These represent authentic HPP enzymes identified based on their primary sequence. All the genes necessary for L-histidine biosynthesis are not found clustered together in these organisms.

Organism	**Locus tag**	**GI number**	**Sequence identity to L37351**	**Zone**
Akkermansia muciniphila ATCC BAA-835	Amuc_1427	187735918	22%	III
Alkaliphilus metalliredigens QYMF	Amet_3162	150390911	33%	II
Alkaliphilus metalliredigens QYMF	Amet_3545	150391281	29%	II
Anoxybacillus flavithermus WK1	Aflv_0468	212638317		IV
Arcobacter butzleri RM4018	Abu_1904	157738118	25%	III
Bacillus clausii KSM-K16	ABC0751	56962525	22%	II
Bacillus clausii KSM-K16	ABC2752	161349990	26%	IV
Bacillus halodurans C-125	**BH3206**	15615768	28%	IV
Bacillus licheniformis DSM 13 = ATCC 14580	BLi03106	52786832	29%	IV
Bacillus pumilus SAFR-032	BPUM_2607	157693363	25%	IV
Bacillus subtilis subsp. subtilis str. 168	**BSU29620**	16080014	29%	IV
Brevibacillus brevis NBRC 100599	**BBR47_00270**	226309614	22%	III
Caldicellulosiruptor bescii DSM 6725	Athe_1711	222529696	33%	II
Caldicellulosiruptor saccharolyticus DSM 8903	Csac_2251	146297249	33%	II
Campylobacter concisus 13826	CCC13826_1445	157165198	25%	III
Campylobacter curvus 525.92	CCV52592_0277	154174464	25%	III
Campylobacter fetus subsp. fetus 82-40	CFF8240_1053	118475300	24%	III
Campylobacter hominis ATCC BAA-381	CHAB381_1502	154148198	26%	III
Candidatus Desulforudis audaxviator MP104C	Daud_1449	169831605	23%	III
Carboxydothermus hydrogenoformans Z-2901	CHY_1234	78043027	28%	II
Carboxydothermus hydrogenoformans Z-2901	CHY_2084	78043205	27%	III
Clostridium acetobutylicum ATCC 824	CA_C2727	15025762	30%	II
Clostridium beijerinckii NCIMB 8052	Cbei_0975	150015861	34%	II

Clostridium beijerinckii NCIMB 8052	Cbei_2106	150016975	28%	IV
Clostridium botulinum A str. ATCC 3502	CBO2571	148380526	34%	II
Clostridium botulinum A str. Hall	CLC_2443	153937403	34%	II
Clostridium botulinum B1 str. Okra	CLD_1994	170757551	35%	II
Clostridium botulinum E3 str. Alaska E43	CLH_2953	188590530	33%	II
Clostridium botulinum F str. Langeland	CLI_2634	153940040	35%	II
Clostridium cellulolyticum H10	Ccel_1927	220929346	37%	II
Clostridium difficile 630	CD630_07370	126698317	24%	IV
Clostridium difficile 630	CD630_21910	126699807	29%	II
Clostridium kluyveri DSM 555	CKL_0268	153952905	33%	II
Clostridium kluyveri NBRC 12016	CKR_0224	219853567	34%	II
Clostridium kluyveri NBRC 12016	CKR_1201	219854544	31%	II
Clostridium novyi NT	NT01CX_0001	118444162	28%	II
Clostridium phytofermentans ISDg	Cphy_2903	160881032	30%	II
Clostridium thermocellum ATCC 27405	Cthe_0724	125973242	33%	II
Deinococcus deserti VCD115	Deide_07040	226355575	25%	III
Deinococcus geothermalis DSM 11300	Dgeo_1608	94985708	26%	III
Deinococcus radiodurans R1	DR_0470	15805497	26%	III
Desulfitobacterium hafniense DCB-2	Dhaf_2109	219668146	30%	II
Desulfitobacterium hafniense DCB-2	Dhaf_4086	219670096	27%	III
Desulfitobacterium hafniense Y51	DSY1025	89893771	29%	II
Desulfitobacterium hafniense Y51	DSY2929	89895675	27%	III
Desulfotalea psychrophila LSv54	DP2888	51246740	25%	IV
Desulfotomaculum reducens MI-1	Dred_0665	134298533	24%	III
Desulfovibrio alaskensis G20	Dde_1057	78356104	24%	II
Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_1985	220905247	22%	III
Desulfovibrio vulgaris DP4	Dvul_0754	120601804	21%	III
Desulfovibrio vulgaris str. 'Miyazaki F'	DvMF_0940	218886042	22%	III
Desulfovibrio vulgaris str. Hildenborough	DVU2490	46580894	21%	III
Dictyoglomus turgidum DSM 6724	Dtur_0185	217966622	25%	III
Geobacillus kaustophilus HTA426	**GK2799**	56421334	26%	IV
Geobacillus thermodenitrificans NG80-2	GTNG_2701	138896338	24%	IV
Helicobacter hepaticus ATCC 51449	HH1714	32267213	26%	III
Heliobacterium modesticaldum Ice1	HM1_1034	167629052	30%	II
Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293	LEUM_1558	116618652	26%	IV
Lysinibacillus sphaericus C3-41	Bsph_4155	169829588	24%	IV
Methylacidiphilum infernorum V4	Minf_0970	189218982	29%	II

Moorella thermoacetica ATCC 39073	Moth_0817	83589668	26%	III
Nautilia profundicola AmH	NAMH_1464	224373484	25%	II
Nitratiruptor sp. SB155-2	NIS_1518	152991260	21%	III
Oceanobacillus iheyensis HTE831	OB2759	23100214	28%	II
Pelotomaculum thermopropionicum SI	PTH_1872	147678207	25%	III
Rubrobacter xylanophilus DSM 9941	Rxyl_2572	108805364	21%	III
Streptococcus agalactiae A909	SAK_1303	76563046	27%	I
Streptococcus agalactiae NEM316	gbs1289	25011338	27%	I
Streptococcus mutans UA159	**SMU_1486c**	24379879	28%	I
Streptococcus thermophilus CNRZ1066	str1232	55823151	25%	I
Streptococcus thermophilus LMD-9	STER_1212	116627978	26%	I
Sulfurimonas denitrificans DSM 1251	Suden_1866	78778060	25%	III
Sulfurovum sp. NBC37-1	SUN_0354	152991950	24%	III
Symbiobacterium thermophilum IAM 14863	STH2574	51893712	24%	III
Thermotoga maritima MSB8	TM0804	15643567	27%	II
Thermotoga petrophila RKU-1	Tpet_0124	148269269	27%	II
Thermotoga sp. RQ2	TRQ2_0122	170287929	27%	II
Thermus thermophilus HB27	TTC1652	46199954	26%	III
Thermus thermophilus HB8	TTHA0331	55980300	26%	III
Wolinella succinogenes DSM 1740	WS0471	34556896	23%	III

Table S3: List of genes represented by red nodes in **Figures 2.2 and S1**. These represent HPP homologs identified based on their primary sequence, which have all the sequence features necessary for HPP activity. However, the genomes of these organisms do not possess several genes necessary for L-histidine biosynthesis, based on the current available information.

Organism	**Locus tag**	**GI number**	**Sequence identity to L37351**	**Zone**
Clostridium botulinum A3 str. Loch Maree	CLK_1956	170758750	34%	II
Clostridium botulinum B str. Eklund 17B	CLL_A3211	187934801	32%	II
Clostridium botulinum B str. Eklund 17B	CLL_A2893	187935558	32%	II
Clostridium perfringens ATCC 13124	CPF_1312	110798699	33%	IV
Clostridium perfringens ATCC 13124	CPF_2137	110800838	30%	II
Clostridium perfringens SM101	CPR_1849	110802089	30%	II
Clostridium perfringens SM101	CPR_1128	110802533	33%	IV

Clostridium perfringens str. 13	CPE1056	18310038	32%	IV
Clostridium perfringens str. 13	CPE1883	18310865	30%	II
Clostridium tetani E88	CTC01046	28210748	33%	II
Coprothermobacter proteolyticus DSM 5265	COPRO5265_0239	206895927	30%	II
Fervidobacterium nodosum Rt17-B1	Fnod_0198	154248897	30%	II
Finegoldia magna ATCC 29328	FMG_0406	169824103	27%	I
Fusobacterium nucleatum subsp. nucleatum ATCC 25586	FN0428	19703770	26%	I
Lawsonia intracellularis PHE/MN1-00	LI0725	94987167	21%	III
Leuconostoc citreum KM20	LCK_00178	170016536	25%	IV
Macrococcus caseolyticus JCSC5402	**MCCL_0344**	222150594	27%	III
Streptococcus thermophilus LMG 18311	stu1232	55821239	26%	I
Thermosipho africanus TCF52B	THA_992	217077076	28%	II
Thermosipho melanesiensis BI429	Tmel_0785	150020678	28%	II

Table S4: List of genes represented by gray nodes in **Figures 2.2 and S1**. These represent enzymes which bear reasonable sequence identity to authentic HPP enzymes but lack the primary sequence features necessary for HPP activity, as determined in this study. Hence, these enzymes must have a different physiological function.

Organism	Locus tag	GI number	Sequence identity to L37351	Zone
Agrobacterium radiobacter K84	Arad_7540	222081247	22%	III
Alkaliphilus oremlandii OhILAs	Clos_0339	158319390	28%	III
Anoxybacillus flavithermus WK1	Aflv_1054	212638895	26%	III
Bacteroides fragilis NCTC 9343	BF0879	60680430	28%	III
Bacteroides fragilis NCTC 9343	BF1597	60681093	24%	III
Bacteroides fragilis YCH46	BF0961	53712253	28%	III
Bacteroides fragilis YCH46	BF1583	53712875	24%	III
Bacteroides thetaiotaomicron VPI-5482	BT_1478	29346888	24%	III
Bacteroides thetaiotaomicron VPI-5482	BT_4257	29349665	28%	III
Brachyspira hyodysenteriae WA1	BHWA1_00960	225619891	25%	III
Brevibacillus brevis NBRC 100599	BBR47_43860	226313971	22%	III
Clostridium novyi NT	NT01CX_0128	118444797	31%	II
Clostridium perfringens ATCC 13124	CPF_1620	110798872	29%	III
Clostridium perfringens ATCC 13124	CPF_0801	110799120	28%	IV

Clostridium perfringens ATCC 13124	CPF_1229	110801334	24%	III
Clostridium perfringens SM101	CPR_1040	110803006	24%	III
Clostridium perfringens SM101	CPR_0788	110803561	33%	IV
Clostridium perfringens str. 13	CPE0804	18309786	27%	IV
Clostridium perfringens str. 13	CPE0972	18309954	24%	III
Clostridium perfringens str. 13	CPE1370	18310352	29%	III
Clostridium phytofermentans ISDg	Cphy_2094	160880232	34%	III
Desulfobacterium autotrophicum HRM2	HRM2_30690	224370152	27%	III
Desulfotalea psychrophila LSv54	DP0093	51243945	35%	III
Desulfotomaculum reducens MI-1	Dred_2577	134300416	27%	III
Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_0366	220903646	18%	III
Dictyoglomus thermophilum H-6-12	DICTH_1913	206900475	24%	III
Geobacillus kaustophilus HTA426	GK2248	56420783	24%	III
Geobacillus thermodenitrificans NG80-2	GTNG_2185	138895826	25%	III
Halorubrum lacusprofundi ATCC 49239	Hlac_1354	222479778	28%	III
Halothermothrix orenii H 168	Hore_11520	220931991	26%	III
Heliobacterium modesticaldum Ice1	HM1_2038	167630107	25%	III
Lactobacillus casei ATCC 334	LSEI_1206	116494700	25%	I
Lactobacillus casei BL23	LCABL_14250	191638198	25%	I
Lysinibacillus sphaericus C3-41	Bsph_1650	169827220	24%	III
Methylacidiphilum infernorum V4	Minf_0443	189218458	28%	III
Parabacteroides distasonis ATCC 8503	BDI_1238	150007878	32%	III
Pediococcus pentosaceus ATCC 25745	PEPE_0107	116491924	24%	I
Pseudomonas aeruginosa UCBPP-PA14	PA14_03330	116053983	23%	III
Syntrophomonas wolfei subsp. wolfei str. Goettingen	Swol_0653	114566197	27%	III
Thermoanaerobacter pseudethanolicus ATCC 33223	Teth39_0146	167036577	25%	III
Thermoanaerobacter sp. X514	Teth514_2305	167040919	25%	III
Treponema denticola ATCC 35405	TDE0796	42526308	24%	III

APPENDIX B

LIST OF PROTEIN SEQUENCES ANNOTATED AS CYCLIC PHOSPHATE DIHYDROLASE FROM COG0613

Table S5: List of orthologs of Elen0235 found using PSI-BLAST program from NCBI database, along with the respective Uniprot identifier.

Organism	Uniprot entry	GI number	Locus tag	Sequence identity to Elen0235	E-value with respect to Elen0235
Eggerthella lenta DSM2243	C8WJZ5	257790010	Elen0235	100%	0.0
Eggerthella sp. 1_3_56FAA	E5X9C5	316911385	HMPREF1023_01665	96.00%	0.0
Eggerthella sp. CAG:209	R6I7R8	524381962	BN534_00198	57.00%	5.0×10^{-99}
Cryptobacterium sp. CAG:338	R7CCY7	524668202	BN613_00910	58.00%	2.0×10^{-96}
Eggerthella sp. CAG:1427	R5FZH5	524023970	BN494_01894	56.00%	6.0×10^{-95}
Planococcus antarcticus	I4X935	388465914	A1A1_01940	54.00%	1.0×10^{-88}
Bacillus alcalophilus	J8TRD8	401725207	BalcAV_11243	52.00%	2.0×10^{-88}
Carnobacterium sp. 17-4	F4BMA2	328956846	CAR_c04960	53.00%	2.0×10^{-85}
Clostridium colicanis	N9XJP1	480655796	HMPREF1092_03059	48.00%	6.0×10^{-77}
Clostridium pasteurianum BC1	R4K585	488772205	Clopa_3535	46.00%	3.0×10^{-77}
Clostridium butyricum	N9Z8T6	480693419	HMPREF1084_00835	45.00%	6.0×10^{-74}
Kyrpidia tusciae DSM 2912	D5WWE3	295697614	Btus_3094	47.00%	4.0×10^{-70}
Paenibacillus terrae HPL-003	G7VPQ6	374324148	HPL003_21630	46.00%	1.0×10^{-69}
Bacillus nealsonii	R9C4U7	507131938	A499_08247	43.00%	9.0×10^{-68}
Bacillus sp. 2_A_57_CT2	E5WKP4	317395958	HMPREF1013_03027	45.00%	3.0×10^{-68}
Bacillus bataviensis	K6DC86	494146847	BABA_17997	43.00%	9.0×10^{-68}
Desulfosporosinus acidiphilus SJ4	I4D4K5	392425067	Desaci_1741	44.00%	5.0×10^{-67}
Collinsella tanakaei	G1WHS2	345901700	HMPREF9452_00885	44.00%	3.0×10^{-61}
Clostridium difficile CD196	C9XSD3	260211212	CD196_3320	38.00%	2.0×10^{-60}
Clostridium difficile 630	Q181D1	126701155	CD630_35310	38.00%	3.0×10^{-60}
Desulfosporosinus youngiae	H5Y4D5	374417528	DesyoDRAFT_2916	42.00%	2.0×10^{-60}
Phascolarctobacterium sp. CAG:207	R6IK46	524396277	BN533_00861	44.00%	2.0×10^{-59}
Clostridium sp. D5	F0YUY0	324033423	HMPREF0240_00949	41.00%	2.0×10^{-57}

Erysipelotrichaceae bacterium 3_1_53	E2SK82	308916212	HMPREF0983_01461	39.00%	2.0×10^{-55}
Ruminococcus sp. CAG:403	R7H452	524728808	BN645_00034	41.00%	6.0×10^{-54}
Erysipelotrichaceae bacterium 2_2_44A	G1VTZ9	345903500	HMPREF9022_03480	39.00%	8.0×10^{-53}
Erysipelotrichaceae bacterium CAG:64	R6VME1	524559624	BN746_02566	39.00%	8.0×10^{-53}
Clostridium sp. CAG:505	R7B7M8	524650905	BN684_00602	40.00%	1.0×10^{-51}
Firmicutes bacterium CAG:466	R6R1N9	524485788	BN668_00491	39.00%	2.0×10^{-49}

APPENDIX C

LIST OF NON-REDUNDANT PROTEIN SEQUENCES ANNOTATED AS 3′,5′-BISPHOSPHOADENOSINE 3′-PHOSPHATASE FROM COG0613

Table S6: List of genes from cog0613 that have been predicted to possess pAp phosphatase activity, based on this work. These enzymes have been depicted as blue nodes in **Figure 4.1** and bear homology greater than an E-value of $1x10^{-70}$ to Cv1693 from *Chromobacterium violaceum*.

VIMSS ID*	Organism	Locus tag	GI number
2089240	*Acidovorax avenae* subsp. citrulli AAC00-1	Aave_3707	120612351
2217868	*Acidovorax* sp. JS42	Ajs_3431	121595727
10342712	*Allochromatium vinosum* DSM 180	Alvin_2196	288941909
1741776	*Azoarcus* sp. BH72	azo2115	119898406
817143	*Azoarcus* sp. EbN1	ebA6127	56478936
5518654	*Bordetella avium* 197N	BAV1544	187478046
504091	*Bordetella bronchiseptica* RB50	BB2306	33601289
508080	*Bordetella parapertussis* 12822	BPP1237	33595907
512402	*Bordetella pertussis* Tohama I	BP1116	33592251
3491835	*Bordetella petrii* DSM 12804	Bpet2759	163857071
6041906	*Burkholderia ambifaria* IOP40-10	BamIOP4010DRAFT_2405	170699293
7156871	*Burkholderia ambifaria* MC40-6	BamMC406_2010	172061055
6069332	*Burkholderia ambifaria* MEX-5	BamMEX5DRAFT_6493	171322307
1889602	*Burkholderia cenocepacia* AU 1054	Bcen_5977	107028720
1892250	*Burkholderia cenocepacia* HI2424	Bcen2424_2100	116690121
5924497	*Burkholderia cenocepacia* J2315	BCAL2172	206560534
7216874	*Burkholderia cenocepacia* MC0-3	Bcenmc03_2118	170733454
2551338	*Burkholderia cenocepacia* PC184	BcenP_01002781	84355115
1918180	*Burkholderia cepacia* AMMD	Bamb_2137	115352188
2555876	*Burkholderia dolosa* AUO158	BdolA_01001300	84362213
7715695	*Burkholderia glumae* BGR1	bglu_1g24440	238027995
6051211	*Burkholderia graminis* C4D1M	BgramDRAFT_4727	170694766
2754129	*Burkholderia mallei* 2002721280	Bmal2_03002250	100265439
736528	*Burkholderia mallei* ATCC 23344	BMA1683	53723778
2414319	*Burkholderia mallei* FMH	BmalF_02003086	83618740
2348235	*Burkholderia mallei* GB8 horse 4	BmalG_01001041	67642974
2419220	*Burkholderia mallei* JHU	BmalJ_02003151	83623646
2115325	*Burkholderia mallei* NCTC 10229	BMA10299_A3132	124386635
1862581	*Burkholderia mallei* NCTC 10247	BMA10247_1458	126451172

4347847	*Burkholderia mallei* PRL-20	BmalP_010100017519	167001767
1855601	*Burkholderia mallei* SAVP1	BMASAVP1_A2185	121598609
7116299	*Burkholderia multivorans* ATCC 17616	BMULJ_02080	189350897
6414400	*Burkholderia multivorans* CGD1	BURMUCGD1_1128	221214751
6421927	*Burkholderia multivorans* CGD2	BURMUCGD2_1513	221208918
6424799	*Burkholderia multivorans* CGD2M	BURMUCGD2M_1608	221198405
4470647	*Burkholderia oklahomensis* C6786	BoklC_010100011679	167570497
4463464	*Burkholderia oklahomensis* EO147	BoklE_010100012273	167563324
7086330	*Burkholderia phymatum* STM815	Bphy_1432	186476190
7149622	*Burkholderia phytofirmans* PsJN	Bphyt_2553	187924530
1966243	*Burkholderia pseudomallei* 1106a	BURPS1106A_2624	126455438
4542662	*Burkholderia pseudomallei* 112	Bpse112_010100014269	167911655
4493744	*Burkholderia pseudomallei* 14	Bpse14_010100014688	167739311
2365103	*Burkholderia pseudomallei* 1655	Bpse1_03003662	99909875
1066820	*Burkholderia pseudomallei* 1710b	BURPS1710b_2703	76809111
4307321	*Burkholderia pseudomallei* 305	BURPS305_7296	134277773
2659092	*Burkholderia pseudomallei* 406e	Bpse4_03001459	100235223
6364987	*Burkholderia pseudomallei* 576	BUC_2888	217420514
1848679	*Burkholderia pseudomallei* 668	BURPS668_2570	126439591
4534392	*Burkholderia pseudomallei* 7894	Bpse7_010100014815	167895017
4510440	*Burkholderia pseudomallei* 9	Bpseu9_010100014729	167824922
4502197	*Burkholderia pseudomallei* 91	Bpse9_010100015413	167816527
4526108	*Burkholderia pseudomallei* B7210	BpseB_010100014233	167846439
4558885	*Burkholderia pseudomallei* BCC215	BpseBC_010100013975	167919653
4485156	*Burkholderia pseudomallei* DM98	BpseD_010100015013	167720316
742208	*Burkholderia pseudomallei* K96243	BPSL2263	53719873
7548326	*Burkholderia pseudomallei* MSHR346	GBP346_A2748	237812988
4550742	*Burkholderia pseudomallei* NCTC 13177	BpseN_010100014182	167903402
2381174	*Burkholderia pseudomallei* Pasteur	BpseP_03003442	100060803
2384924	*Burkholderia pseudomallei* S13	BpseS_03001572	100124908
1059800	*Burkholderia* sp. 383	Bcep18194_A5406	78066875
10413427	*Burkholderia* sp. CCGE1002	BC1002_1865	295676915
6334611	*Burkholderia* sp. H160	BH160DRAFT_7213	209522309
4480185	*Burkholderia thailandensis* Bt4	BthaB_010100024245	167619437
1078150	*Burkholderia thailandensis* E264	BTH_I1901	83720996
4518124	*Burkholderia thailandensis* MSMB43	Bpse38_010100011656	167837136
4451909	*Burkholderia thailandensis* TXDOH	BthaT_010100024649	167581360
4456889	*Burkholderia ubonensis* Bu	BuboB_010100015617	167586771
6818796	*Burkholderia vietnamiensis* G4	Bcep1808_2179	134296281
3361375	Burkholderia xenovorans LB400	Bxe_A1580	91784221
7523828	*Candidatus Accumulibacter phosphatis* clade IIA str. UW-1	CAP2UW1_2976	257094542
516894	***Chromobacterium violaceum* ATCC 12472**	**CV1693**	**34497148**
10457784	*Comamonas testosteroni* CNB-2	CtCNB1_0837	264676973

3036997	*Comamonas testosteroni* KF-1	CtesDRAFT_4825	118052256
5902118	*Cupriavidus taiwanensis*	RALTA_A1177	194289301
3337685	*Dechloromonas aromatica* RCB	Daro_2462	71908080
3605905	*Delftia acidovorans* SPH-1	Daci_5411	160900843
5788667	*Diaphorobacter* sp. TPSY	Dtpsy_2763	222111933
3826541	*Herminiimonas arsenicoxydans*	HEAR2181	134095365
3550174	*Janthinobacterium* sp. Marseille	mma_1277	152981327
7588717	*Laribacter hongkongensis* HLHK9	LHK_02105	226941024
5550854	*Leptothrix cholodnii* SP-6	Lcho_2582	171059263
2131490	*Methylibium petroleiphilum* PM1	Mpe_A2756	124267941
1767775	*Methylobacillus flagellatus* KT	Mfla_1877	91776229
762274	*Methylococcus capsulatus* str. Bath	MCA0359	53802475
3073336	*Methylophilales bacterium* HTCC2181	MB2181_05670	118595136
7670891	*Methylotenera mobilis* JLW8	Mmol_0689	253996062
7668144	*Methylovorus* sp. SIP3-4	Msip34_0918	253998630
818313	*Neisseria gonorrhoeae* FA 1090	NGO0081	59800542
5767051	*Neisseria gonorrhoeae* NCCP11945	NGK_0119	194097703
3543672	*Neisseria meningitidis* 053442	NMCC_0401	161869394
7761864	*Neisseria meningitidis* alpha14	NMO_0353	254804361
1791478	*Neisseria meningitidis* FAM18	NMC0394	121634259
140883	*Neisseria meningitidis* MC58	NMB1824	15677660
141796	*Neisseria meningitidis* Z2491	NMA0635	15793623
10229895	*Nitrosococcus halophilus* Nc4	Nhal_2315	292492351
1122970	*Nitrosococcus oceani* ATCC 19707	Noc_2384	77165847
368404	*Nitrosomonas europaea* ATCC 19718	NE1724	30249686
1899148	*Nitrosomonas eutropha* C71	Neut_2039	114332005
1126040	*Nitrosospira multiformis* ATCC 25196	Nmul_A2314	82703431
709433	*Polaromonas*	-	-
2001809	*Polaromonas naphthalenivorans* CJ2	Pnap_3274	121606164
1820720	*Polaromonas* sp. JS666	Bpro_1005	91786905
7350344	*Polynucleobacter necessarius* STIR1	Pnec_0994	171463678
6889789	*Polynucleobacter* sp. QLW-P1DMWA-1	Pnuc_0844	145589029
2037545	*Ralstonia eutropha* H16	H16_A1200	113867220
891559	*Ralstonia eutropha* JMP134	Reut_A1100	73540802
1773468	*Ralstonia metallidurans* CH34	Rmet_1064	94310009
7262615	*Ralstonia pickettii* 12D	Rpic12D_1078	241662682
7180316	*Ralstonia pickettii* 12J	Rpic_0982	187928075
169417	*Ralstonia solanacearum* GMI1000	RS04613	17545859
10338937	*Ralstonia solanacearum* IPO1609	RSIPO_01698	207743519
10332289	*Ralstonia solanacearum* MolK2	RSMK05533	207722876
1273606	*Rhodoferax ferrireducens* DSM 15236	Rfer_3236	89902005
10354863	*Sideroxydans lithotrophicus* ES-1	Slit_1630	291614093
7020016	*Thauera* sp. MZ1T	Tmz1t_1683	217970100

3416894	*Thiobacillus denitrificans*	Tbd_1528	74317546
7559107	*Variovorax paradoxus* S110	Vapar_4021	239816991
2068796	*Verminephrobacter eiseniae* EF01-2	Veis_1626	121608595

*VIMSS: Virtual institute for microbial stress and survival.

APPENDIX D

LIST OF NON-REDUNDANT PROTEIN SEQUENCES ANNOTATED AS 5′→3′

NANO-RNASE FROM COG0613

Table S7: List of genes from cog0613 that have been predicted to possess nanoRNase activity, based on this work. These enzymes have been depicted as green nodes in **Figure 5.1** and bear homology greater than an E-value of $1x10^{-70}$ to TrpH from *Escherichia coli*.

Organism	GI number	NCBI RefSeq
Actinobacillus pleuropneumoniae serovar 3 str. JL03	165976086	YP_001651679.1
Actinobacillus pleuropneumoniae serovar 5b str. L20	126208151	YP_001053376.1
Actinobacillus pleuropneumoniae serovar 7 str. AP76	190149980	YP_001968505.1
Aeromonas hydrophila subsp. hydrophila ATCC 7966	117620611	YP_857425.1
Aeromonas salmonicida subsp. salmonicida A449	145298416	YP_001141257.1
Aliivibrio salmonicida LFI1238	209694704	YP_002262632.1
Citrobacter koseri ATCC BAA-895	157145595	YP_001452914.1
Colwellia psychrerythraea 34H	71279433	YP_270195.1
Cronobacter sakazakii ATCC BAA-894	156933737	YP_001437653.1
Enterobacter sp. 638	146311855	YP_001176929.1
Erwinia tasmaniensis Et1/99	188533747	YP_001907544.1
Escherichia coli 536]	110641495	YP_669225.1
Escherichia coli 55989	218694837	YP_002402504.1
Escherichia coli APEC O1	117623526	YP_852439.1
Escherichia coli ATCC 8739	170020370	YP_001725324.1
Escherichia coli CFT073	26247595	NP_753635.1
Escherichia coli E24377A	157154837	YP_001462563.1
Escherichia coli ED1a	218689256	YP_002397468.1
Escherichia coli HS	157160771	YP_001458089.1
Escherichia coli IAI1	218553818	YP_002386731.1
Escherichia coli IAI39	218699973	YP_002407602.1
Escherichia coli O127:H6 str. E2348/69	215486561	YP_002328992.1
Escherichia coli O157:H7 str. EC4115	209398686	YP_002270320.1
Escherichia coli O157:H7 str. EDL933	15801913	NP_287933.1
Escherichia coli O157:H7 str. Sakai	15831092	NP_309865.1
Escherichia coli S88	218558251	YP_002391164.1

Escherichia coli SE11	209918506	YP_002292590.1
Escherichia coli SMS-3-5	170681023	YP_001743921.1
Escherichia coli str. K-12 substr. DH10B	170080943	YP_001730263.1
***Escherichia coli* str. K-12 substr. MG1655**	**16129227**	**NP_415782.1**
Escherichia coli UMN026	218704789	YP_002412308.1
Escherichia coli UTI89	91210557	YP_540543.1
Escherichia fergusonii ATCC 35469	218549043	YP_002382834.1
Haemophilus ducreyi 35000HP	33152479	NP_873832.1
Haemophilus influenzae 86-028NP	68250081	YP_249193.1
Haemophilus influenzae PittGG	148826883	YP_001291636.1
Haemophilus influenzae Rd KW20	16273309	NP_439553.1
Haemophilus parasuis SH0165	219870305	YP_002474680.1
Haemophilus somnus 129PT	113461026	YP_719093.1
Haemophilus somnus 2336	170717599	YP_001784682.1
Idiomarina loihiensis L2TR	56460848	YP_156129.1
Klebsiella pneumoniae 342	206577138	YP_002239003.1
Klebsiella pneumoniae subsp. pneumoniae MGH 78578	152969812	YP_001334921.1
Mannheimia succiniciproducens MBEL55E	52425091	YP_088228.1
Pasteurella multocida subsp. multocida str. Pm70	15602481	NP_245553.1
Pectobacterium atrosepticum SCRI1043	50121221	YP_050388.1
Photobacterium profundum SS9	54309648	YP_130668.1
Photorhabdus luminescens subsp. laumondii TTO1	37526350	NP_929694.1
Proteus mirabilis HI4320	197285201	YP_002151073.1
Psychromonas ingrahamii 37	119944819	YP_942499.1
Salmonella enterica subsp. arizonae serovar 62:z4,z23:- str. RSK2980	161503165	YP_001570277.1
Salmonella enterica subsp. enterica serovar Agona str. SL483	197251221	YP_002146305.1
Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67	62180288	YP_216705.1
Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853	198245413	YP_002215420.1
Salmonella enterica subsp. enterica serovar Enteritidis str. P125109	207856766	YP_002243417.1
Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91	205352605	YP_002226406.1
Salmonella enterica subsp. enterica serovar Heidelberg str. SL476	194450192	YP_002045768.1
Salmonella enterica subsp. enterica serovar Newport str. SL254	194443247	YP_002040975.1
Salmonella enterica subsp. enterica serovar Paratyphi A str. AKU_12601	197362277	YP_002141914.1
Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150	56413353	YP_150428.1
Salmonella enterica subsp. enterica serovar Paratyphi B str. SPB7	161613789	YP_001587754.1
Salmonella enterica subsp. enterica serovar Schwarzengrund str. CVM19633	194735330	YP_002114753.1
Salmonella enterica subsp. enterica serovar Typhi str. CT18	16760159	NP_455776.1

Salmonella enterica subsp. enterica serovar Typhi str. Ty2	29142070	NP_805412.1
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2	16765065	NP_460680.1
Serratia proteamaculans 568	157370906	YP_001478895.1
Shewanella amazonensis SB2B	119775263	YP_928003.1
Shewanella baltica OS155	126174908	YP_001051057.1
Shewanella baltica OS185	153001237	YP_001366918.1
Shewanella baltica OS195	160875911	YP_001555227.1
Shewanella baltica OS223	217972842	YP_002357593.1
Shewanella denitrificans OS217	91793801	YP_563452.1
Shewanella frigidimarina NCIMB 400	114562607	YP_750120.1
Shewanella halifaxensis HAW-EB4	167624540	YP_001674834.1
Shewanella loihica PV-4	127513178	YP_001094375.1
Shewanella oneidensis MR-1	24374542	NP_718585.1
Shewanella pealeana ATCC 700345	157961470	YP_001501504.1
Shewanella piezotolerans WP3	212634745	YP_002311269.1
Shewanella sediminis HAW-EB3	157374826	YP_001473426.1
Shewanella sp. ANA-3	117919969	YP_869161.1
Shewanella sp. MR-4	113969805	YP_733598.1
Shewanella sp. MR-7	114047034	YP_737584.1
Shewanella sp. W3-18-1	120598423	YP_962997.1
Shewanella woodyi ATCC 51908	170727253	YP_001761279.1
Shigella boydii CDC 3083-94	187733133	YP_001880096.1
Shigella boydii Sb227	82544280	YP_408227.1
Shigella dysenteriae Sd197	82776623	YP_402972.1
Shigella flexneri 2a str. 2457T	30062789	NP_836960.1
Shigella flexneri 2a str. 301	24112665	NP_707175.1
Shigella flexneri 5 str. 8401	110805269	YP_688789.1
Shigella sonnei Ss046	74312364	YP_310783.1
Sodalis glossinidius str. 'morsitans'	85059381	YP_455083.1
Vibrio campbellii ATCC BAA-1116	156975030	YP_001445937.1
Vibrio cholerae O1 biovar El Tor str. N16961	15641190	NP_230822.1
Vibrio cholerae O395	147674944	YP_001216748.1
Vibrio fischeri ES114	59711641	YP_204417.1
Vibrio fischeri MJ11	197335981	YP_002155837.1
Vibrio parahaemolyticus RIMD 2210633	28898728	NP_798333.1
Vibrio tasmaniensis LGP32	218709946	YP_002417567.1
Vibrio vulnificus CMCP6	27366332	NP_761860.1
Vibrio vulnificus YJ016	37679407	NP_934016.1
Yersinia enterocolitica subsp. enterocolitica 8081	123442467	YP_001006446.1
Yersinia pestis Angola	162421523	YP_001606746.1

Yersinia pestis Antiqua	108807564	YP_651480.1
Yersinia pestis KIM10+	22125942	NP_669365.1
Yersinia pestis Nepal516	108811840	YP_647607.1
Yersinia pestis Pestoides F	145598221	YP_001162297.1
Yersinia pseudotuberculosis IP 31758	153948700	YP_001400905.1
Yersinia pseudotuberculosis IP 32953	51596459	YP_070650.1
Yersinia pseudotuberculosis PB1/+	186895508	YP_001872620.1
Yersinia pseudotuberculosis YPIII	170024276	YP_001720781.1