MECHANISTIC AND FUNCTIONAL CHARACTERIZATION OF LACTONASES OF COG3618 IN THE AMIDOHYDROLASE SUPERFAMILY

A Dissertation

by

MERLIN ERIC HOBBS

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
Paul Straight
Margaret E. Glasner
Head of Department, Gregory Reinhart

May 2014

Major Subject: Biochemistry

ABSTRACT

The postgenomic era of scientific research has yielded an inundation of gene and protein sequences which are available in public databases. This torrent of sequences, literally in the millions, has altered the strategies, methodologies, and approaches taken toward function discovery and annotation. At the forefront, are interdisciplinary approaches; such as, genomic enzymology. Genomic enzymology bridges aspects of classical enzymology, structural and functional genomics, and comparative genomics. These methods also take advantage of evolutionarily related proteins, which have relatively similar sequence similarity but, yield different functions. The amidohydrolase superfamily (AHS), contains proteins of similar chemistry and topology, which are related to one another through a common ancestral progenitor. The AHS has been organized into smaller groups based only on sequence similarity, which are known as clusters of orthologous groups (COG). There are currently 24 COGs within this superfamily. Cog3618 is comprised of primarily lactonase enzymes.

The first identified member of cog3618 was LigI, which catalyzes the reversible hydrolysis of 2-pyrone-4,6-dicarboxylate. The chemical mechanism was elucidated and it was determined that this protein is the first member of the AHS that does not require a metal cofactor for catalysis. The mechanistic characteristics of LigI, along with genomic enzymology, were utilized to predicted and identify the functions of two other COG members. BmulJ_04915 and BVU_0220 were identified as L-fucono-1,5-lactonase

and L-galactono-1,5-lactonase, respectively. NMR analysis revealed that these enzymes give preference to 6-member lactones, as opposed to their more stable counterparts. In addition, it is proposed that L-galactono-1,5-lactone participates in a novel L-galactose catabolic pathway, which appears to converge with D-galacturonate degradation through the intermediate of an adjacent dehydrogenase (BVU_0222). BVU_0222 was determined to oxidize L-galactonate to D-tagaturonate, which is a pathway intermediate for the degradation of D-glucuronate.

DEDICATION

To Emiliya, the greatest result from graduate school. To Anna, for the love, support, and encouragement the past four years, especially the last few months. To my parents, who worked hard to instill the value of education. To my brother and sister, who have provided love and friendship throughout my life.

ACKNOWLEDGMENTS

This work would not have been possible without the help of many. I want to thank those who provided their valuable time and expertise. I would like to thank Dr. Frank Raushel, for his support, encouragement, and guidance. I would like to thank Dr. Howard Williams, for his assistance with NMR. I would like to thank my committee members: Dr. Tadhg Begley, Dr. Paul Straight, Dr. Margaret Glasner, and Dr. Donald Pettigrew, for providing excellent suggestions from different points of view. I would also like to thank past and present Raushel lab members, mentors, and friends.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Increased sophistication and advances in technology, more specifically biotechnology, have provided the scientific community the means to rapidly and cost efficiently sequence whole genomes of microorganisms. The past fifteen years, there has been an exponential expansion of whole genome sequences available to public databases. For example, UniProtKB/TrEMBL, which utilizes automated computational annotations, contains over 51 million sequence entries as of February 2014 up from 4 million in 2008. Increased sequence availability has led to greater understandings in metabolism, which has furthered advances in drug discovery, medicine, and synthetic biology. However, 74% of the sequences available in this automated database are annotated as having a predicted function, while 24% have annotations inferred from homology.[1] It has previously been estimated that the level of misannotation within automated public databases is >40%. UniProtKB/Swiss-Prot, a manually curated database, contains accurate annotations, however, the data base is comprised of only 1% of the available sequences.[1] The levels of misannotation are the highest for enzyme families containing multiple functions, where an enzyme family is defined as a group of evolutionarily related proteins. For example, members of the 2-pyrone-4,6-dicarboxylate lactonase (LigI) family are known to catalyze at least 5 different chemical

transformations. However, in many databases unknown members of this family are given an annotation of LigI or LigI like, regardless of sequence similarity to other functions. This is an example of database 'over-predicting' molecular function, which is thought to be one of the highest sources of error in functional annotation.[2] A consequence of the genomic era, is a paradigm shift at the core of functional annotation. Prior to this period, functions were identified and fit to genes, now genes are identified and functions are fit to them.

Interdisciplinary approaches, such as genomic enzymology, have been successful in predicting and annotating protein function in the postgenomic era.[3, 4] At the core, genomic enzymology seeks to analyze proteins based on sequence divergence from a common ancestral protein. In addition, genomic enzymology bridges elements of classical enzymology, functional, comparative, and structural genomics in order to annotate protein function. Herein, the evolutionarily related amidohydrolase superfamily (AHS) is utilized as a framework for genomic enzymology and function discovery.

Amidohydrolase Superfamily of Proteins

The term superfamily was first introduced by Margaret Dayhoff in the early 1970s, and is defined as a group of evolutionarily related proteins sharing a common ancestral progenitor.[5, 6] However, more recently it has been used to describe proteins of similar structural topology and mechanistic character, thought to originate from a

common ancestor. [3] The AHS was discovered less than 20 years ago through careful analysis of the structural similarities existing between urease, phosphotriesterase, and adenosine deaminase. [7] These structures were found to contain a distorted $(\beta/\alpha)_8$-TIM barrel housing a divalent metal center. The divalent metal ion is oriented in space through coordination to ligands that originate from the C-terminal ends the TIM barrel β-strands. Twenty years later, we have a deeper understanding of the structural, catalytic, and mechanistic diversity of this superfamily. [8] The members of the AHS are now known to house 0-3 divalent transition metal ions within the active site. [8, 9] Details of metal binding, active site geometry, and mechanism of action of the AHS are reviewed in the remaining introduction. The AHS was originally known for the hydrolysis of amide and ester bonds. However, more recently catalytic diversity among superfamily members was demonstrated with the identification of the following chemical transformations: decarboxylation, isomerization, hydration, and retroaldol cleavage.[8, 10-13]

The AHS can be organized and classified into smaller and more related groups based on sequence similarity. The cluster of orthologous groups (COG) database was initiated as an attempt to organize the inundation of protein sequences becoming available.[14] The AHS has been organized into approximately 23 separate COGS by the NCBI. COGs were originally identified through an alL-against-all sequence comparison of protein sequences of the 7 fully sequenced genomes in the late 1990s.[14,15] *Proteins*

sharing a given sequence homology were organized into clades and given a unique numerical identifier.

AHS metal center types. Members of the AHS have previously been classified based on sequence similarity (COG) and functionality (Family); herein, we have classified this superfamily on the basis of active site architecture and geometry. There are four types of AHS active sites, identified to date. These include metal independent, mononuclear, binuclear, and trinuclear metal centers.[8, 16] Active site metals are identified according to the relative site occupied, and are desginated as M_α, M_β, and M_γ. Active site metals are oriented in space through amino acid side chain ligands in the active site. The most common residues coordinating the alpha metal are an HxH motif residing at the C-terminal end of β-strand 1 and an aspartic acid from the C-terminal end of β-strand 8. Alpha metals within mononuclear sites are also coordinated by a histidine from the end of β-strand 5; while alpha metals present in binuclear sites are coordinated by additional bridging ligands. Common bridging ligands include a post-translational modified carboxylated lysine residue and an active site water. In addition to two bridging ligands, beta metals are commonly ligated by two histidines from the C-terminal ends of β-strands 5 and 6. [17]

AHS metal centers are known to bind a variety of first row transition metals, including Fe^{2+}, Mn^{2+}, Co^{2+}, and Ni^{2+}. The role each metal cation plays is dependent on the number of metals and the position which is occupied. The alpha metal has been

shown to both orient the active site water molecule, and aid in the activation of the water nucleophile. The metal ion, which acts as a Lewis acid, facilitates the ionization of the active site water to the nucleophilic hydroxide.[18] This is achieved by lowering the pKa of water to approximately 7, which is essential when performing hydrolytic reactions in biological conditions. The beta metal serves two roles; activation of the bridging water molecule, and polarizing the carbonyl and phosphoryl groups of the substrate, to facilitate nucleophilic attack. The gamma metal is proposed to promote the cleavage of C-O bonds through direct interaction with the oxygen of a leaving group.[9]

Geometry of the metal centers is dependent on two primary factors; the type of metal involved and the number of ligands originating from the protein. The overwhelming majority of metals occupying M_α are 5-coordinate with trigonal bipyramidal geometry, with few exceptions. The metal occupying the M_β has shown to be 4- or 5-coordinate with tetrahedral and trigonal bipyramidal geometry, respectively. The active site of phosphotriesterase provides an example of coordination number and geometry being metal dependent (**Figure 1.1**). The Zn/Zn PTE has coordination numbers of 5 and trigonal bipyramidal geometry, at both metal sites (**Figure 1.1A**). The beta metal is coordinated to an additional axial water molecule or oxygen atom from the bound diisopropyl methyl phosphonate inhibitor. However, in the Mn/Mn PTE the coordination number of the beta metal is increased to 6 to accommodate the preferred octahedral geometry (**Figure 1.1B**). In the crystal structure of the Mn/Mn PTE the alpha

metal retains the 5-coordinate trigonal bipyramidal; however, an active site water molecule is within 3.5Å of the M_α site, which could potential become a 6th ligand.

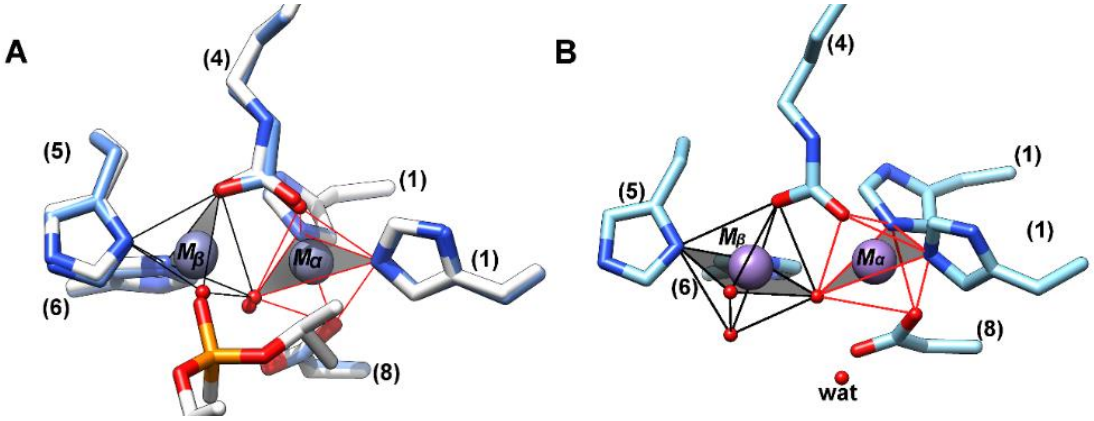


Figure 1.1. Comparison of the active sites of PTE with Zn/Zn and Mn/Mn bound. (**A**) White (pdb code 1hzy) and light blue (pdb code 1ez2 with bound diisopropyl methyl phosphonate) contain Zn^{2+} in both metal sites. (**B**) Light blue (pdb code 1i0b) contains Mn^{2+} in both metal site

Most AHS members can utilize Zn^{2+} at all metal sites. The preference of Zn^{2+} likely stems from the character of being a d^{10} transition metal. Zinc is one of the strongest first row Lewis acids; which can be taken advantage of for catalysis. As a d^{10} transition metal, Zinc, cannot undergo redox chemistry providing protection from active site oxidative damage. In addition, Zinc does not have ligand field stabilization energy, which allows freedom in coordination number and geometry. [18, 19]

The amino acid metal ligands vary, even between proteins with the same active site metal type. Therefore, the active site of the members of AHS have been classified based on the metal ligands, this grouping is known as subtype (Table 1.1).

Table 1.1: Metal Subtypes of the Amidohydrolase Superfamily

					β-strand							
Subtype	**Example**	**site type**	**Mx**	**Positions**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
I	LigI				hxh				-	h		d
II	CDA	mono	Zn/Fe	α	HXH				H	h		D
III	URI	mono	Zn	α	HxH				h	h		D
IV	LigW	mono	Mn	α	Exa				H	h		D
V	LigJ	mono	Zn	α	HxH				H	h		E
VI	RSD	mono	Zn	α	ExH				H	h		D
VII	NagA	mono	Fe	β	hxh		E		H	H		d
VIII[a]	DAA	mono	Zn	β	hxh	C			H	H		d
IX	PTE	bi	Zn, Ni	αβ	HxH			K	H	H		D
X	RDP	bi	Zn	αβ	HxD		E		H	H		d
XI	AdA	bi	Fe/Mn	αβ	HxH			E	H	H, E		D
XII	NagA	bi	Zn	αβ	HxH		E		H	H		D
XIII[b]	DAA	bi	Zn	αβ	HxH	C			H	H		D
XIV	HPP	tri	Zn,Fe	αβγ	HxH, D[c]	H[c]	E		H	H		D,H[c]

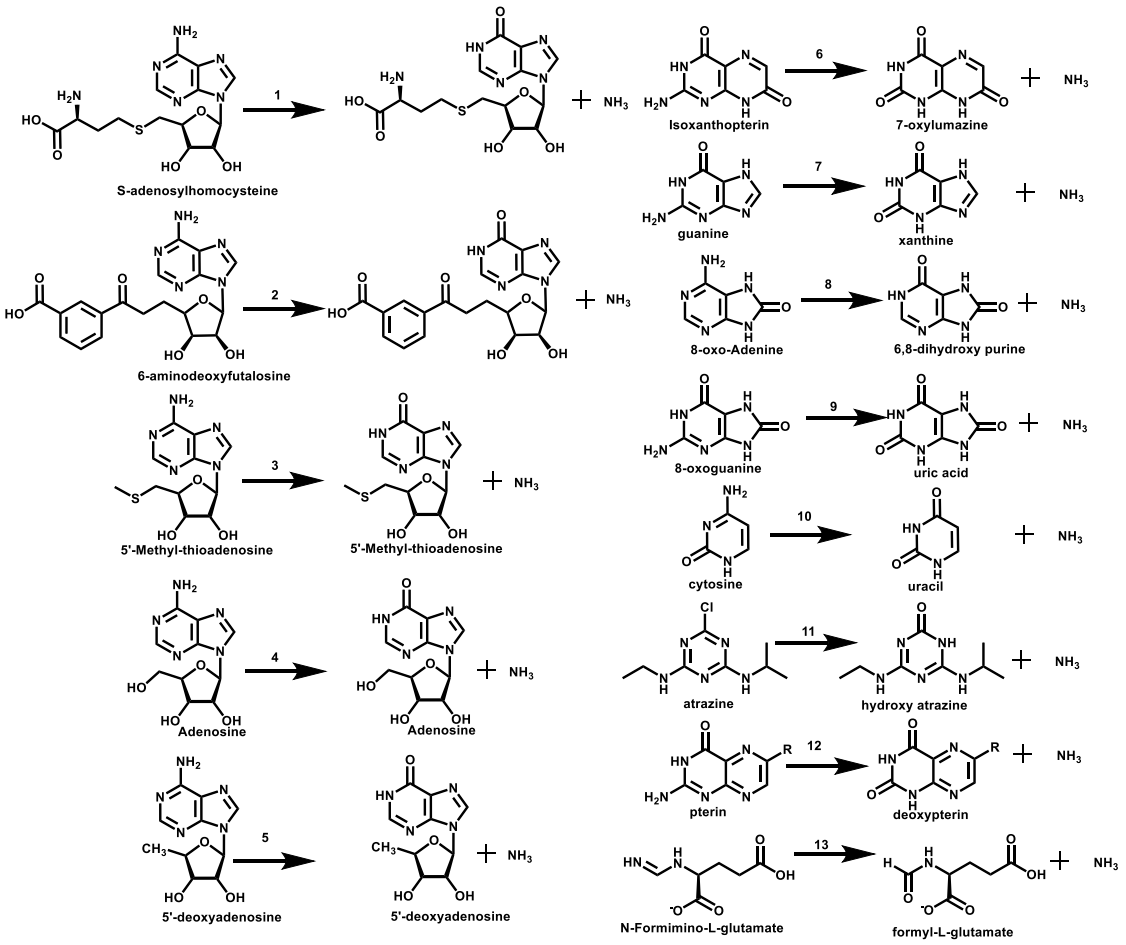
[a] These subtypes are included because the protein has shown full activity with a single metal ion and was crystalized with only the M_β site occupied. [b] These subtypes are included to accompany subtypes VII and VIII as the binuclear counterpart. [c] These residues are coordinated to the metal occupying the M_γ site.

Mononuclear Metal Centers

To date there are four COGs that are known to utilize a single divalent metal for catalysis in the AHS including: cog0402, cog1816, cog1904, and cog2159. In addition, there are two COGs in which it is unclear if the metal occupying the M_α is required for catalytic activity even though proper ligands are present. [20-22]. These COGs are cog1228 and cog1820. COGs possessing a mononuclear metal active site catalyze a diverse set of reactions; such as, deamination, isomerization, decarboxylation, C-C bond cleavage, and hydration.

Cog0402. The enzymes of cog0402 have primarily been shown to catalyze the cleavage of C-N bonds of nucleic acid bases through a deamination reaction. [17, 20, 21, 23-37] The experimentally verified functions identified within cog0402 include the following: cytosine deaminase (EC 3.5.4.1, CDA, **9**), guanine deaminase (EC 3.5.4.3, GDA,**6**), 8-oxoguanine deaminase (EC 3.5.4.32, **8**), isoxanthopterin deaminase (EC 3.5.4.X, **5**), S-adenosylhomocyteine deaminase (EC 3.5.4.28, **1**),6-aminodeoxyfutalosine deaminase (EC 3.5.4.X, **2**), atrazine chlorohydrolase (EC 3.8.1.8, **10**), pterin deaminase (EC , **11**), N-formimino-L-glutamate iminohydrolase (EC 3.5.3.13, **12**) and 8-oxo-adenosine deaminase (EC 3.5.4.X, **7**) (**Scheme1.1**). Other functions within this cog0402 are catalyzed by enzymes with promiscuous substrate profiles including: 5'-deoxyadenosine deaminase (EC 3.5.4.X, **5**), adenosine deaminase (EC 3.5.4.4, ADA, **4**), and methylthioadenosine deaminase (EC 3.5.4.X, **3**) (**Scheme 1.1**).

Scheme 1.1: Enzymatic reactions catalyzed by members of cog0402



X-ray crystal structures have been solved for 10 members of cog0402 and include: methylthioadenosine deaminase (pdb code 4gbd), S-adenosylhomocystiene deaminase (pdb code 2plm and 1p1m), 6-deoxyaminofutalosine deaminase (pdb code 3v7p and 2imr), guanine deaminase (pdb code 2i9u), isoxanthopterin deaminase (pdb code 2paj), 8-oxoguanine deaminase (pdb code 3hpa), cytosine deaminase (pdb code 1k70 and 1k6w), and N-formimino-L-glutamate deaminase (pdb code 3mdw). In addition, structures are available for the cog0402 members with promiscuous activity for adenosine like compounds; such as, methylthioadenosine deaminase, 5'-deoxyadenosine, and adenosine (pdb code 4dzh and 4f0s).

Cog0402 enzymes are of active site subtype II based on the occupation of a single divalent cation in the M_{α} position and the coordinating protein ligands. An active site schematic is provided in **Figure 1.2B**. Metal ligands include the common alpha metal HxH from β-strand 1, histidine from β-strand 5, aspartate from β-strand 8, and an active site water molecule. The alpha metal, in all available structures, has a coordination number of 5 and binds either a single Zn^{2+} or Fe^{2+} atom in trigonal bipyramidal geometry (**Figure 1.2A**). The equatorial water ligand is replaced by an oxygen molecule when the product is bound, retaining the original coordination number and active site geometry. This is clearly shown in cytosine deaminase crystalized in the presence and absence of the inhibitor 4-(R)-hydroxyL-3,4-dihydropyrimidine (**Figure 1.3**).

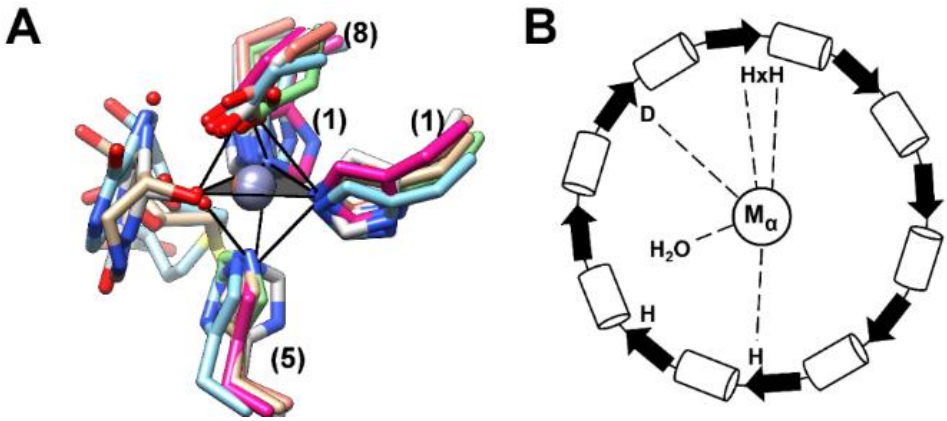


Figure 1.2. Active site architecture of members of cog0402. (**A**) Active site geometry about a single metal center, either Zn^{2+} or Fe^{2+}. Structures are colored as follows: light brown is methylthioadenosine deaminase (pdb code 4gbd), blue is guanine deaminase (pdb code 2plm), green is isoxanthopterin (pdb code 2paj), gray is cytosine deaminase (1k70), magenta is N-formimino-L-glutamate iminohydrolase (pdb code 3mdw), and yellow is a promiscuous adenosine like deaminase (pdb code 4dzh). (**B**) Active site schematic diagram for members of cog0402. Black arrows represent β-strands and white cylinders represent α-helices which make up the $(\beta/\alpha)_8$TIM barrel.

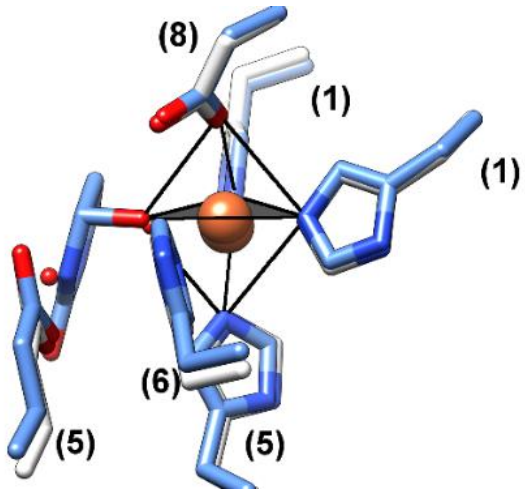
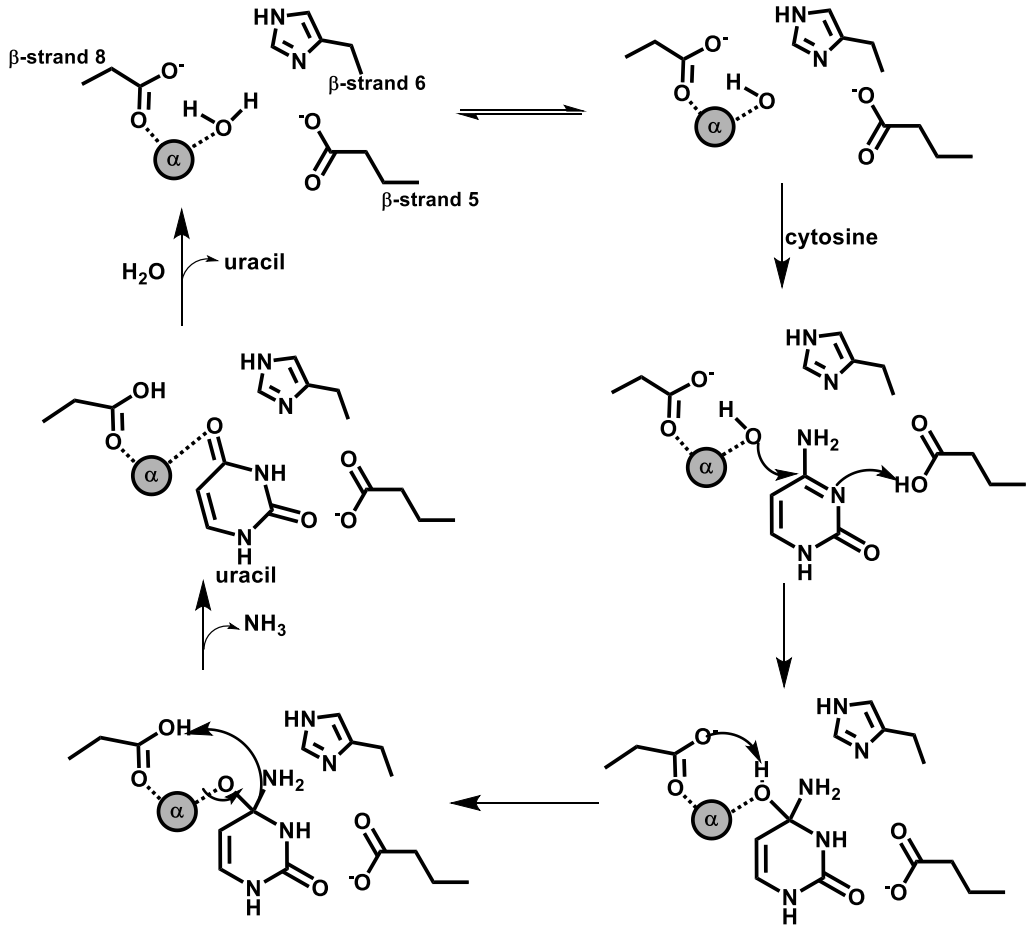


Figure 1.3. Active site geometry comparison of Fe cytosine deaminase with inhibitor. 4-(R)-hydroxyL-3, 4-dihydropyrimidine bound shown in blue (pdb code 1k70) and without inhibitor bound shown in white (pdb code 1k6w).

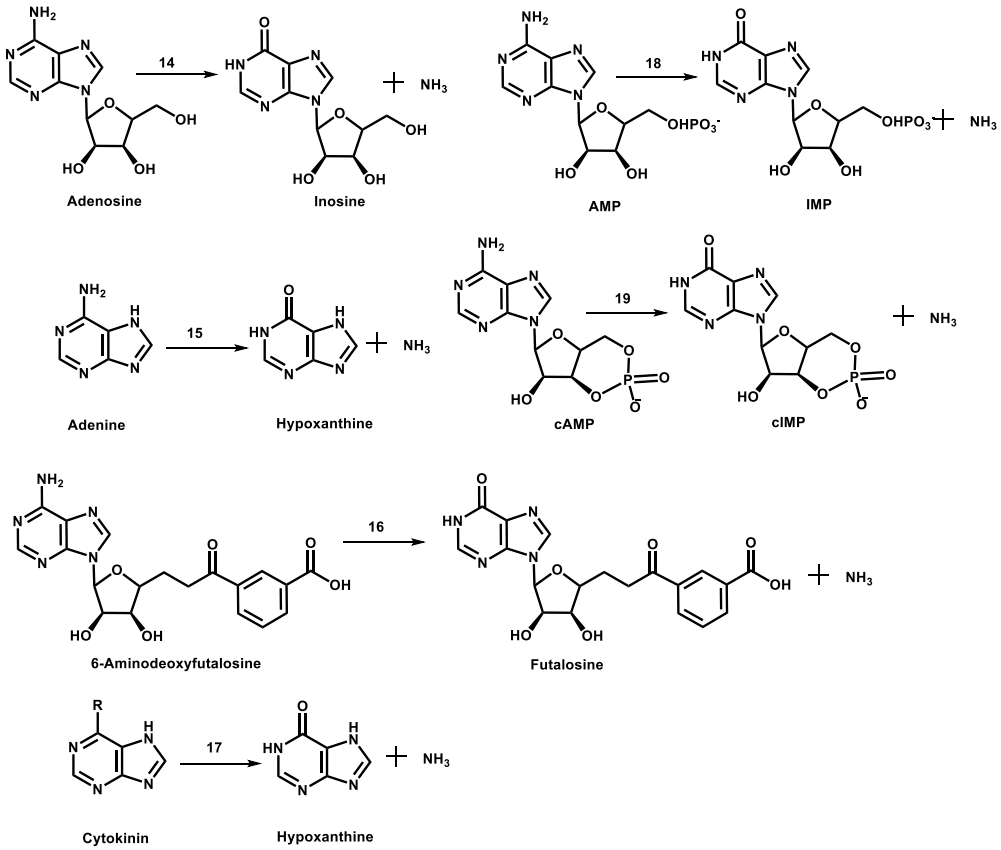
The mechanism of cytosine deaminase serves as a model for all reactions within cog0402, as the catalytic residues are conserved among all members. [26] As previously mentioned, the metal coordinating ligands consist of three histidine residues originating from β-strands 1 and 5, aspartic acid from β-strands 8, and an active site water molecule. A common theme in the catalytic mechanism of members of the AHS is the activation of an active site water molecule for nucleophilic attack. It is proposed that coordination to the active site metal ion significantly reduces the pka of water, making it readily available for deprotonation to the nucleophilic hydroxide conjugate base. In cytosine deaminase, it is proposed that a proton of the metal bound water is transferred to histidine from β-strand 6 and subsequently shuttled to a glutamate residue at the end of β-strand 5 (**Scheme 1.2**). The nucleophilic hydroxide is now available for attack on the re-face of the cytosine substrate, which forms a tetrahedral intermediate through the donation of a proton by the active site glutamate. Aspartate initiates the C-N bond cleavage through removal of a proton from the hydroxyl of the tetrahedral intermediate. The newly protonated aspartate donates a proton to the leaving group; thus, cleaving the C-N bond to form uracil.

Scheme 1.2: Proposed Mechanism of Cytosine Deaminase from cog0402



Cog1816. Similar to cog0402, members of cog1816 also catalyze C-N bond cleavage through deamination reactions. However, cog1816 members have only been observed to catalyze the deamination of adenine-like compounds.[38-41] COG members include: adenosine deaminase (EC 3.5.4.4, **14**) adenine deaminase (EC 3.5.4.2, ADA, **15**), adenosine monophosphate (AMP) deaminase (EC 3.5.4.6, **18**), cyclic adenosine monophosphate (cAMP) deaminase (EC 3.5.4.X, **19**), cytokinin deaminase (EC 3.5.4.X, **17**), and 6-aminodeoxyfutalosine deaminase (EC 3.5.4.X, **16**) (**Scheme 1.3**). Both cog0402 and cog1816 catalyze the deamination of 6-aminodeoxyfutalosine (**Scheme 1.3).** This is thought to occur through a convergent evolution event, as each protein contains no significant amount of sequence similarity. [40] X-ray crystal structures have been solved for three members of cog1816: adenine deaminase from (pdb code 3rys), AMP deaminase (PDB entry 2a3l), and adenosine deaminase (PDB entries 3mvi and 1a4m). Each of these structures contains a single Zn atom per monomeric subunit that is required for catalytic activity.

Scheme 1.3: Reactions Catalyzed by Members of cog1816.



The enzymes of cog1816, similar to cog0402, are mononuclear metal centers of subtype II, which is detailed in schematic diagram in **Figure 1.4B**. The common alpha metal ligands are conserved within all members of this cog (**Figure 1.4A and B**). Binding of the adenosine substrate appears to replace the water ligand with the nitrogen of the substrate, retaining coordination number and geometry (**Figure 1.4A**).

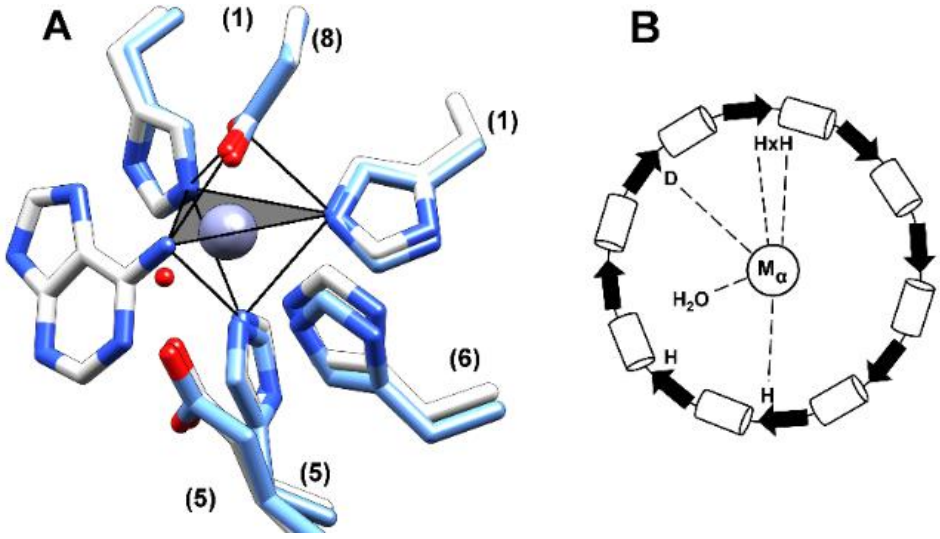
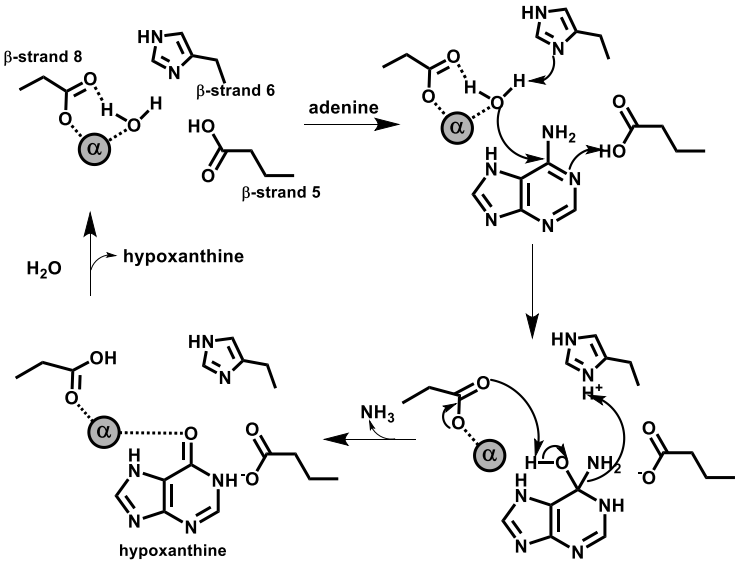


Figure 1.4. Active site architecture of adenine deaminase of cog1816. (A) Active site comparison of adenine deaminase with bound substrate shown in white (pdb code 3ou8) and without substrate shown in blue (pdb code 3rys). (B) Active site schematic diagram of the proteins of cog1816. Black arrows represent β-strands and white cylinders represent α-helices that make up the $(\beta/\alpha)_8$TIM barrel.

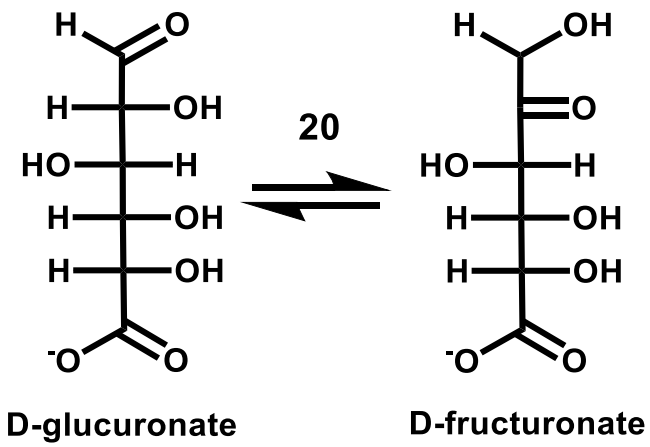
A proposed mechanism is provided in **Scheme 1.4** below. The metaL-bound water activated through deprotonation by histidine from β-strand 6; however, is not shuttled to the glutamate as shown in the mechanism of cytosine deaminase. The nucleophilic hydroxide attacks the adenine at C-6, forming a tetrahedral intermediate by donation of the proton from glutamate at N-1. This intermediate is collapsed following deprotonation of the hydroxyl by β-strand 8 aspartate. The C-N bond cleavage is finalized by the protonation of the leaving group by histidine from β-strand 6. The restoration of the active site is proposed through a proton shuttling event from the active site glutamate to histidine.

Scheme 1.4 : Proposed mechanism of adenine deaminase from cog1816



Cog1904. Cog1904 contains a single experimentally verified function, uronate isomerase (URI) (EC 5.3.1.12). URI was the first AHS member identified to not catalyze a hydrolytic reaction. [10, 42] This enzyme catalyzes the isomerization of D-glucuronate and D-galacturonate to D-fructuronate and D-tagaturonate, respectively (**Scheme 1.5**). X-ray crystal structures are available for native URI (pdb code 2q63) and with bound D-glucuronate (pdb code 3hk9).

Scheme 1.5 : Reactions Catalyzed by cog1904



URI contains a unique mononuclear metal center, subtype III, which has only been found within cog1904 (**Figure 1.5C**). A single Zn^{2+} ion occupies the M_α site through coordination to four ligands including: two histidines from HxH, aspartate from

β-strand 8, and an active site water molecule (**Figure 1.5**). The four ligands take on tetrahedral geometry about the metal ion, an unsual geometry for alpha metals within AHS. The geometry of the active site changes to trigonal bipyramidal once the substrate (D-glucuronate) binds the active site. The binding of D-glucuronate is bi-dentate replacing the active site water molecule with the oxygen from C-5, while the oxygen from C-6 becomes an axial ligand.

In the proposed mechanism, the active site water is expelled from the active site upon binding of the substrate (**Scheme 1.6**). Aspartate from β-strand 8 is positioned for the abstraction of the C-2 proton forming a cis-enediol intermediate subsequent to protonation of C-2 by an active site tyrosine. Tyrosine abstracts the proton from the hydroxyl at C-2 while a proton is donated by the active site aspartate to C-1.

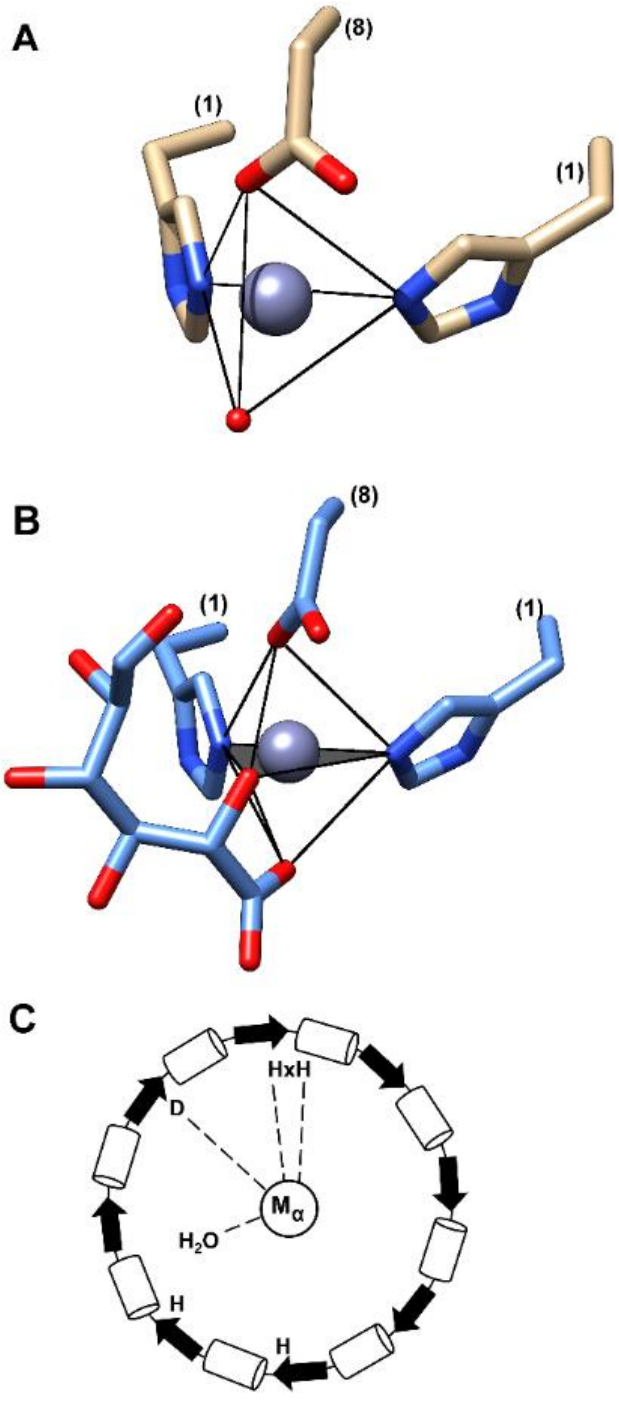
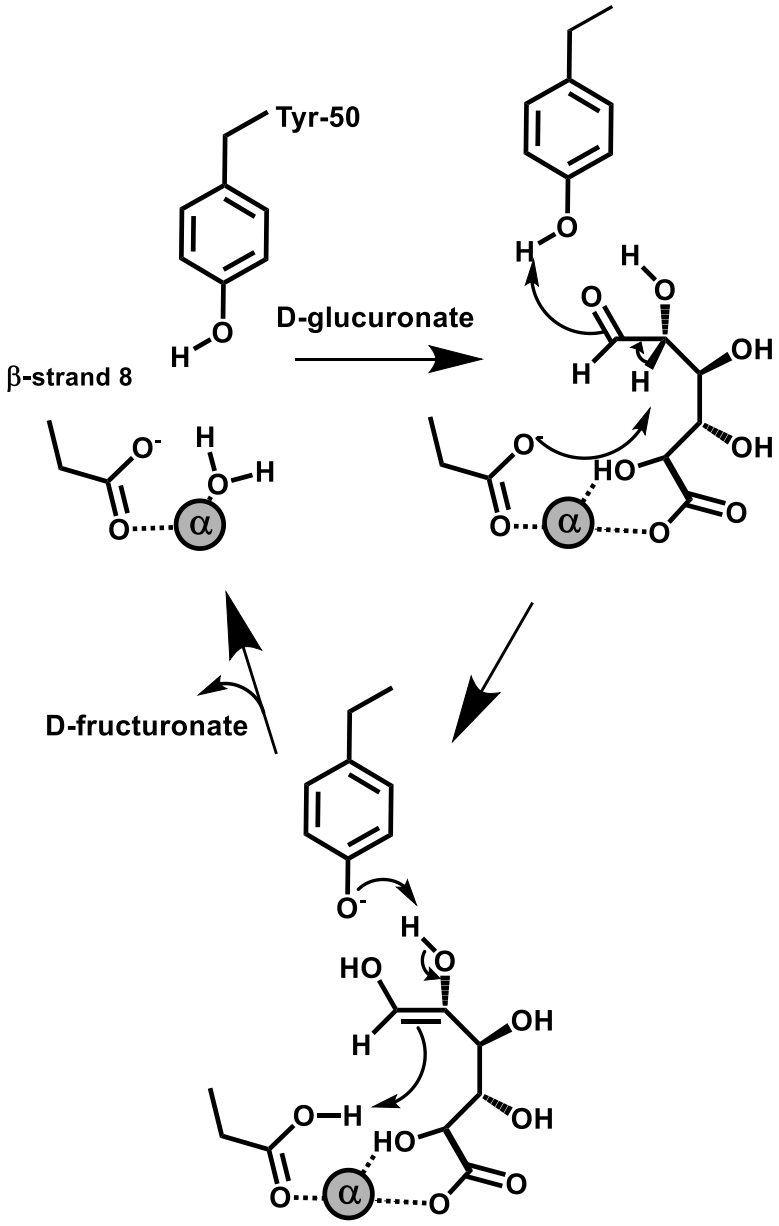


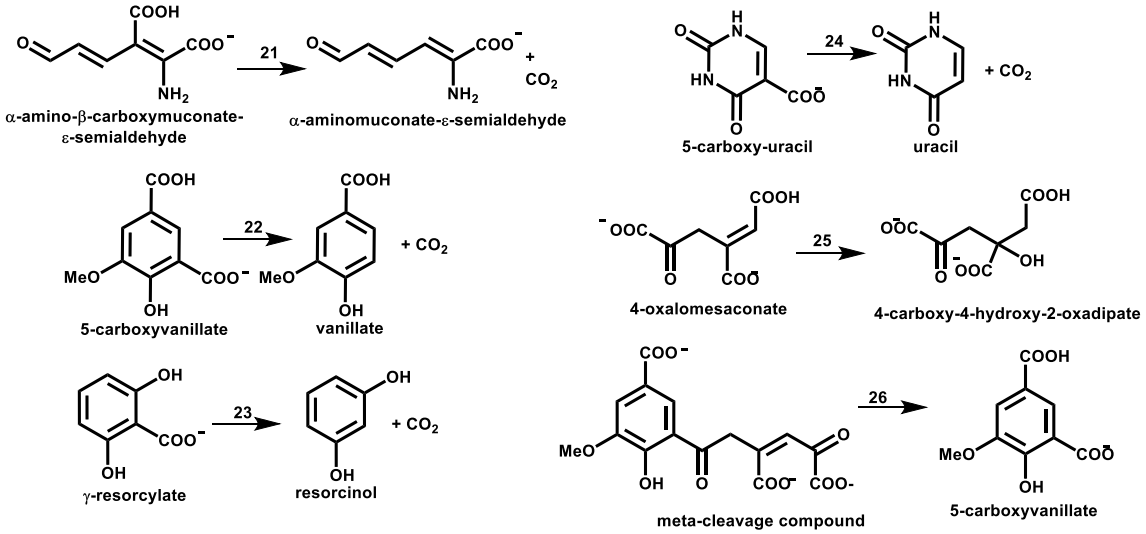
Figure 1.5. Active site analysis of URI from cog1904. (A) URI crystallized in the absence of substrate (pdb code 2q63) with tetrahedral geometry about the active site Zn^{2+}. (B) URI crystallized bound to D-glucuronate (pdb code 3hk9) with trigonal bipyramidal geometry about the active site Zn^{2+}. (C) Active site schematic diagram of URI. Black arrows represent β-strands and white cylinders represent α-helices that make up the $(\beta/\alpha)_8$TIM barrel.

Scheme 1.6: Proposed Mechanism of URI of cog1904



Cog2159. Cog2159 is the only example of a COG from the AHS whose members catalyze different chemical transformations. For example, members of this COG catalyze atleast three different chemical reactions including: non-oxidative decarboxylation, hydrolytic C-C bond cleavage, and hydration. [11-13] There are currently six experimentally verified functions which include the following: aminocarboxymuconate semialdehyde decarboxylase (ACMSD) (EC 4.1.1.45,), 5-carboxyvanillate decarboxylase (LigW) (EC 4.1.1.X,), gamma-resorcylate decarboxylase (RSD) (EC 4.1.1.X,), isoorotate decarboxylase (IDCase)(EC 4.1.1.X,), 4-oxalomesaconate hydratase (LigJ)(EC 4.2.1.83,), and 2,2',3-trihydroxy-3'-methoxy-5,5'-dicarboxybiphenyl(OH-DDVA) hydrolase (LigY) (EC 4.1.1.X,) (**Scheme 1.7**).

Scheme 1.7: Reactions Catalyzed by Members of cog2159



X-ray crystal structures are available for all of the members of cog2159 with the exception of LigY. These include native LigW (pdb code 4l5p) and with bound nitro-inhibitor (pdb code 4icm), ACMSD (pdb code 2hbv), RSD (pdb code 2dvt), IDCase (pdb code 4hk6), and LigJ (2gwg).

In addition to catalyze three types of chemistry, cog2159 also contains at least four active site subtypes (**Figure 1.6**). LigW contains a mononuclear Mn^{2+} active site of subtype IV (**Figure 1.6A**). The hallmark HxH motif is not conserved in LigW; however, all LigW enzymes conserve a mutated ExA motif at the C-terminal end of β-strand 1. The metal coordinating residues include glutamate from ExA, histidine from β-strand 5, aspartate from β-strand 8, and three active site water molecules. The additional water molecules are coordinated to provide the preferred octahedral geometry for Mn^{2+}. The geometry of the active site and coordination number are retained in the X-ray crystal structure with nitro-inhibitor bound (4-hydroxy-3-methoxy-5-nitrobenzoic acid). The inhibitor binds in bi-dentate fashion replacing two water ligands with an oxygen from the nitro group and the phenolic oxygen (**Figure 1.7**).

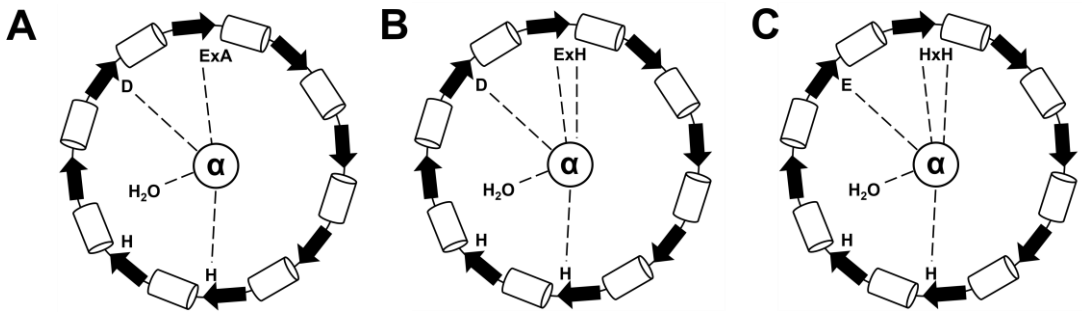


Figure 1.6. Active site schematics of subtypes IV (A), VI (B), and V(C). Black arrows represent β-strands and white cylinders represent α-helices that make up the $(\beta/\alpha)_8$TIM barrel.

The active sites of ACMSD and IDCase are both mononuclear metal centers of subtype II and conserve the common alpha metal ligands. In addition to the common ligands two active site water molecules are coordinated, which provide a coordination number of 6 and octahedral geometry.

Subtypes VI and V, which are only found in cog2159, appear to have originated through mutations of subtype II. Subtype V, found in LigJ, contains a 5-coordinate mononuclear Zn^{2+} metal in trigonal bipyramidal geometry (**Figure 1.6C**). LigJ conserves the common alpha metal ligands with the exception of aspartate at β-strand 8, which is mutated to a glutamate. Subtype VI, found in RSD, contains a single Zn^{2+} ion coordinated through interactions with a mutated ExH motif at β-strand 1, histidine at β-strand 5, aspartate at β-strand 8, and a water molecule. The active site of RSD contains a 5 coordinate Zn^{2+} with trigonal bipyramidal geometry (**Figure 1B**).

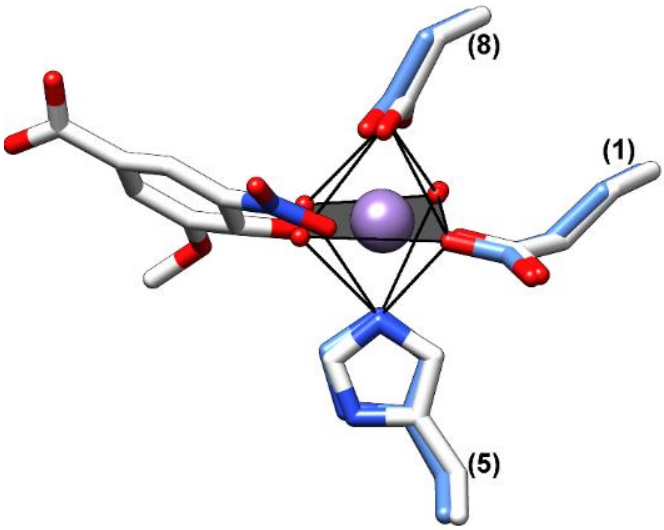
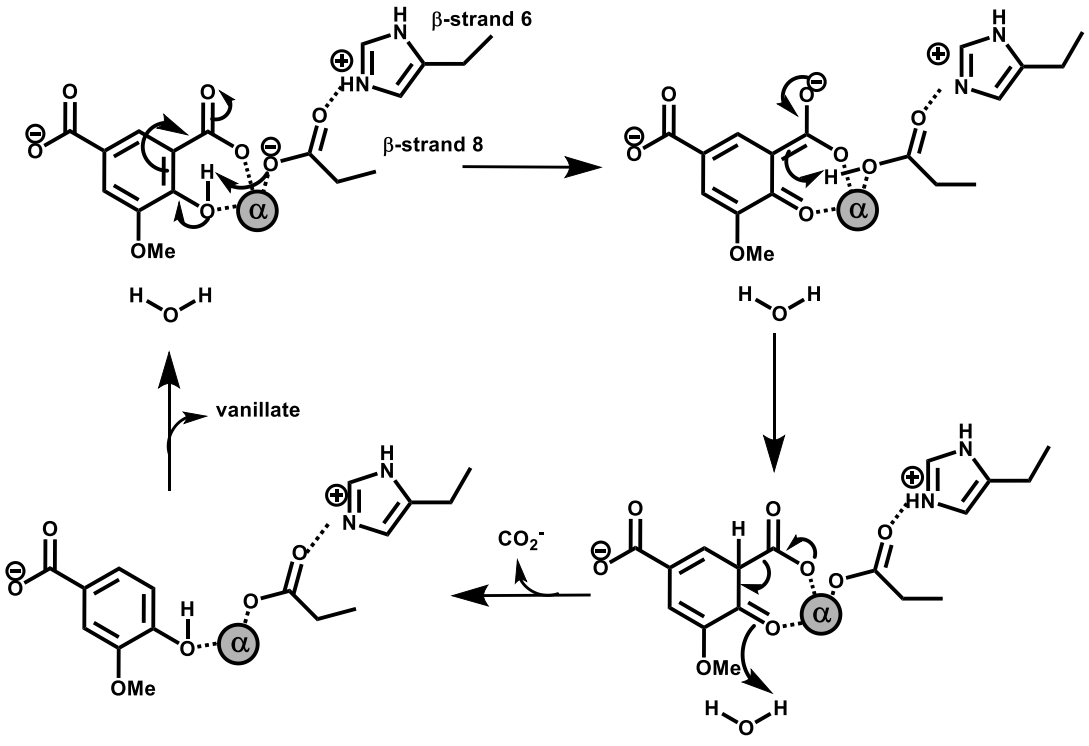


Figure 1.7. Active site comparison of holo LigW shown in blue (pdb code 4icm) and with bound nitro inhibitor shown in white (pdb code 4l5p). Mn^{2+} (purple) exhibits a coordination number of 6 and octahedral geometry.

The proposed mechanism for the reaction catalyzed by LigW is provided in **Scheme 1.8**. The binding of 5-carboxyvanilliate in the active site replaces two water ligands with oxygen atoms from the carboxyl group at C-5 and the hydroxyl group at C-6. The coordination of the phenolic proton of the substrate to the metal ion reduces the pKa, which allows the β-strand 8 aspartate to abstract the proton initiating the reaction. The electrons delocalize to the carboxyl group at position C-6. The delocalized electrons collapse allowing for aspartate to donate a proton to C-5. Finally CO_2 is removed through the protonation of the ketone functional group at C-6. Electrons are delocalized to the carboxyl functional group of the substrate with subsequent protonation across the double bond. The C-C bond is finally cleaved through the protonation of C-6 by a water molecule.

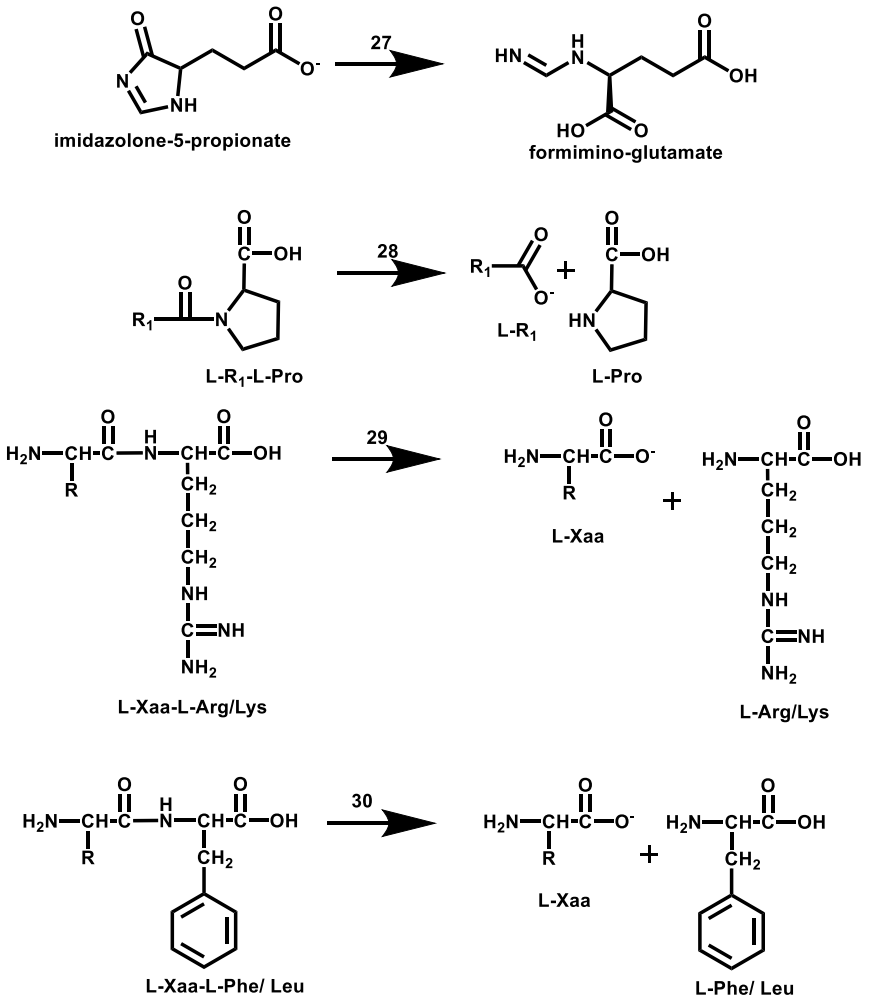
Scheme 1.8: Proposed Mechanism of LigW of cog2159



*Cog1228***.** Cog1228 is the only COG within the AHS that contains both mononuclear and binuclear metal active sites. Members of this COG have been shown to primarily catalyze the cleavage of amide (peptide) bonds. [22, 43, 44] These enzymes are imidazolone-5-propanote hydrolase (HutI, EC 3.5.2.7, **27**), prolidase (3.4.13.9,**28**), and two promiscuous carboxypeptidases Cc2672 (**29**) and Cc0300 (**30**) (EC 3.4.X.X) (**Scheme 1.9**). X-ray crystal structures are available for both active site types within cog1228

which include HutI (pdb codes 2q09 and 2oof), L-Xaa-L-Arg/Lys peptidase (pdb code 3mtw), and prolidase (pdb code 3mkv).

Scheme 1.9: Reactions catalyzed by members of cog1228



Members of cog1228 possess either a mononuclear metal center of subtype II or a binuclear metal center of subtype IX (**Figure 1.8**). HutI contains an active site of subtype II with a single Fe^{2+} ion. The active site Fe^{2+} is 5-coordinate with a typical trigonal bipyramidal geometry, utilizing the typical AHS alpha metal ligands. Active site geometry is retained when inhibitor DIP (3-(2, 5-dioxoimidazolidin-4-yl) propionic acid), which does not appear to directly interaction with the alpha Fe^{2+} (**Figure 1. 9A**).

Prolidase contains a binuclear metal active site with Zn^{2+} occupying both the M_α and M_β sites (**Figure 1.8B**). The alpha metal is oriented through ligation with the common alpha metal ligands; in addition to the bridging ligands. The bridging ligands include a posttranslationally modified carboxylated lysine residue originating from the C-terminal of β-strand 4 and an active site water molecule. The alpha metal is 5-coordinate with trigonal bipyramidal geometry (**Figure 1.9B**). The beta metal is oriented in space through ligations with the common beta metal active site ligands. These ligands include: two histidines from the C-terminal ends of β-strands 5 and 6 and the bridging ligands (**Figure 1.9B**). The beta Zn^{2+} is 4-coordinate with tetrahedral geometry (**Figure 1.9B**).

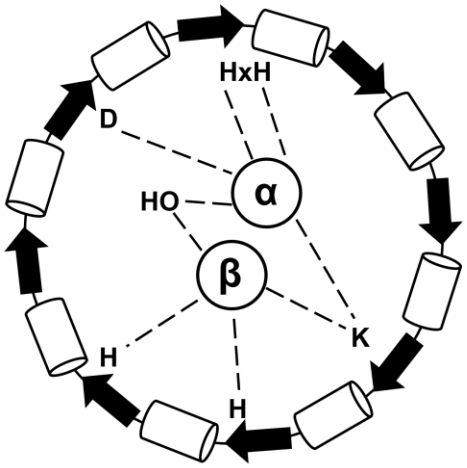


Figure 1.8. Active site schematic diagram for prolidase and promiscuous peptidases (Cc2672 and Cc0300) of subtype IX. Black arrows represent β-strands and white cylinders represent α-helices which make up the $(\beta/\alpha)_8$TIM barrel.

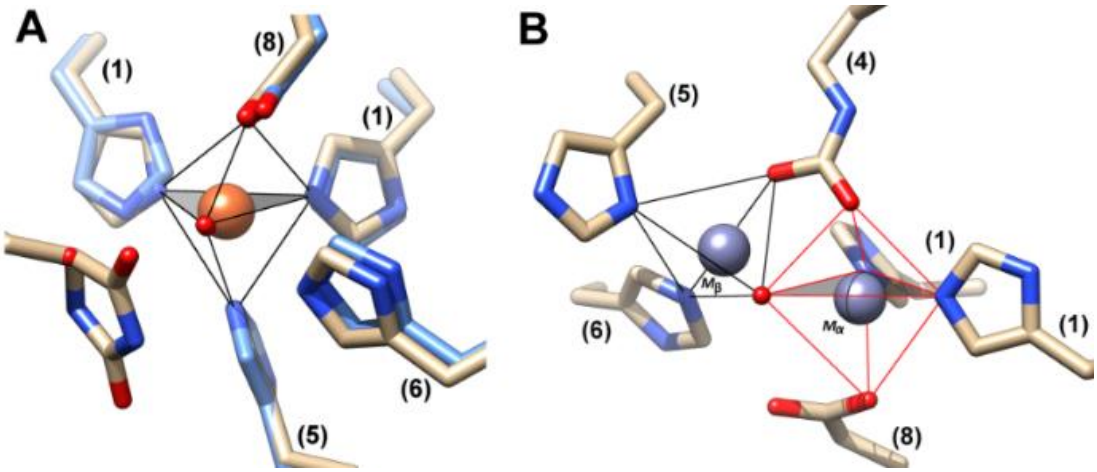
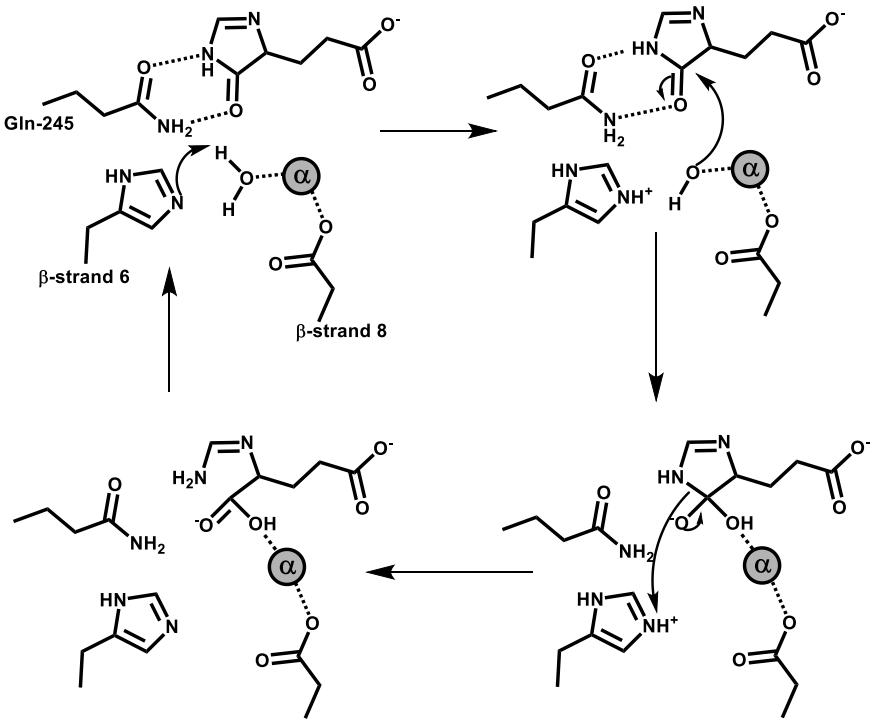


Figure 1.9. (A) Active site comparison of HutI with bound DIP inhibitor (white, pdb code 2q09) and unbound HutI (blue, 2oof). Fe^{2+} (orange) is 5 coordinate with trigonal bipyramidal geometry. (B) Active site analysis of prolidase (pdb code 3mkv). Both metal site are occupied with Zn^{2+} with the alpha metal being 5 coordinate and trigonal bipyramidal geometry and the beta metal having a coordination number of 4 and tetrahedral geometry.

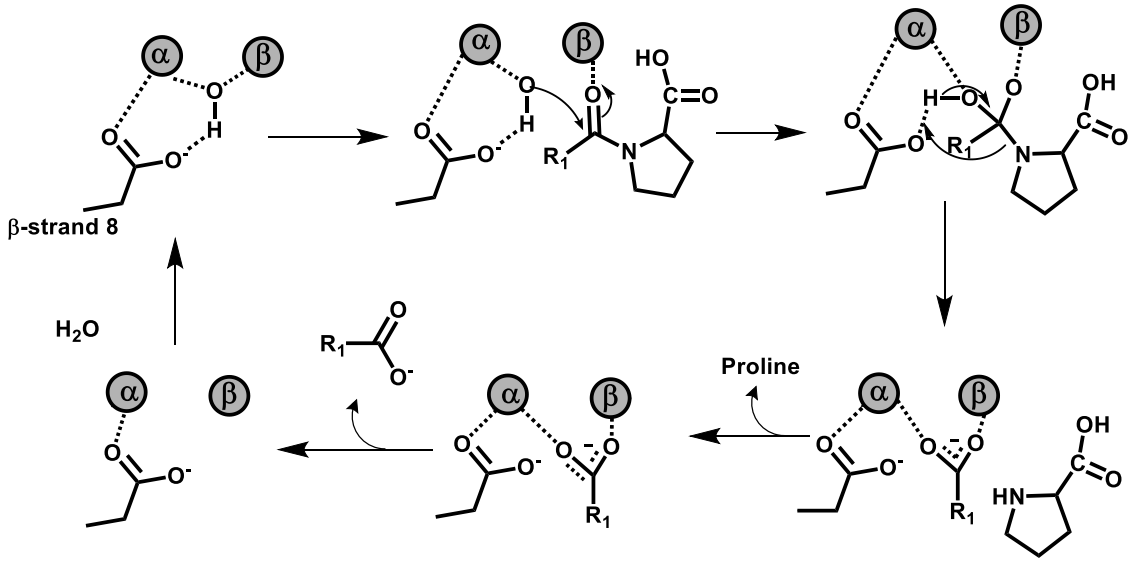
As expected, the two active site subtypes utilize different mechanism for catalysis. The proposed mechanism of HutI is presented in **Scheme 1.10**. The nucleophilic water molecule, activated by histidine from β-strand 6, attacks the amide functional group of the substrate. A tetrahedral intermediate is formed and subsequently collapsed through protonation of the leaving group the β-strand 6 histidine.

Scheme 1.10: Proposed mechanism of HutI from cog1228



The binuclear subtype IX, found in prolidase, utilize both metals for water activation (**Scheme 1.11**). The bridging hydroxide nucleophile attacks the carbonyl functional group, which is polarized through coordination with the beta metal. A subsequent tetrahedral is formed, which is then collapsed by the protonation of the leaving group nitrogen (**Scheme 1.11**).

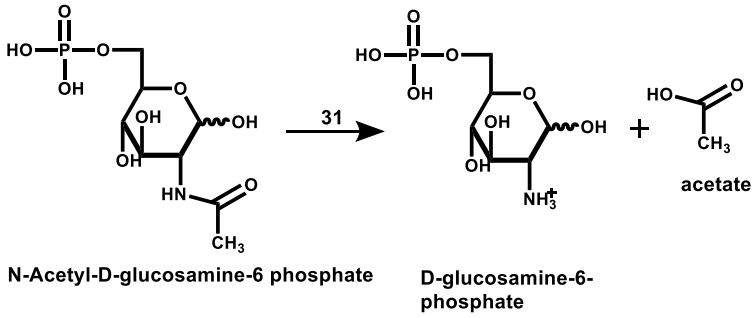
Scheme 1.11: Proposed mechanism of prolidase from cog1228



Cog1820. The single function of cog1820 is N-AcetyL-D-glucosamine-6 phosphate deacetylase (EC 3.5.1.25, **31**, NagA). [21] This enzyme catalyzes the hydrolytic

cleavage of N-AcetyL-D-glucosamine-6 phosphate to form D-glucosamine-6-phosphate and acetate (**Scheme 1.12**). The metal requirements for NagA appear to differ in different organisms. NagA from *Vibrio cholerae* (pdb code 3iv8) and *E. coli* (pdb code 2p50) were crystalized with a single divalent cation occupying the M_β site. [20] In contrast, a homologue from *Thermotoga maritima* (pdb code 1012) was crystalized with two metals occupying both the M_β and M_α sites.

Scheme 1.12: Reaction catalyzed by NagA from cog1820



NagA from *E. coli* and *V. cholerea*, both contain active sites of subtype VII, with a single Zn^{2+} occupying the M_β site (**Figure 1.10A**). The missing alpha metal in subtype VII site appears to be genuine and not an artifact of crystallization, due to the mutation of the HxH motif to QxN (**Figure 1.11A**). The beta metal is oriented in space though coordination to a glutamate from β-strand 3, two histidines from β-strands 5 and 6, and an active site water molecule (**Figure 1.10**). Conversely, NagA from *T. maritima*

contains a binuclear metal center of subtype XII (**Figure 1.10B**). The Fe^{2+} ion occupying the M_α site is 5-coordinate and oriented into position through the common alpha metal ligands and bridging ligands. The bridging ligands for subtype XII consist of a glutamate from the end of β-strands 3 and a water molecule (**Figure 1.10B**). The ligands take on a trigonal bipyramidal geometry about the alpha metal (**Figure 1.11B**). The beta Fe^{2+} is 4 coordinate with tetrahedral geometry. In addition to the common beta metal ligands the Zn^{2+} occupying M_β is bridged by the glutamate and water molecule.

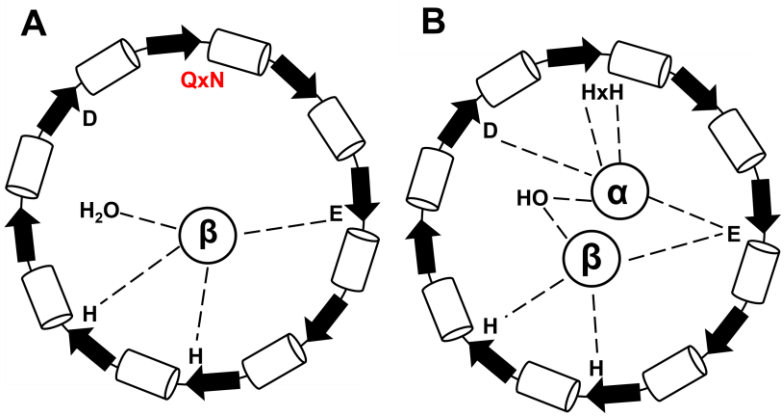


Figure 1.10. (**A**) Active site schematic diagram of NagA from *E. coli*. (**B**) Active site schematic diagram of NagA from *T. maritima*. Black arrows represent β-strands and white cylanders represent α-helices which make up the $(\beta/\alpha)_8$TIM barrel.

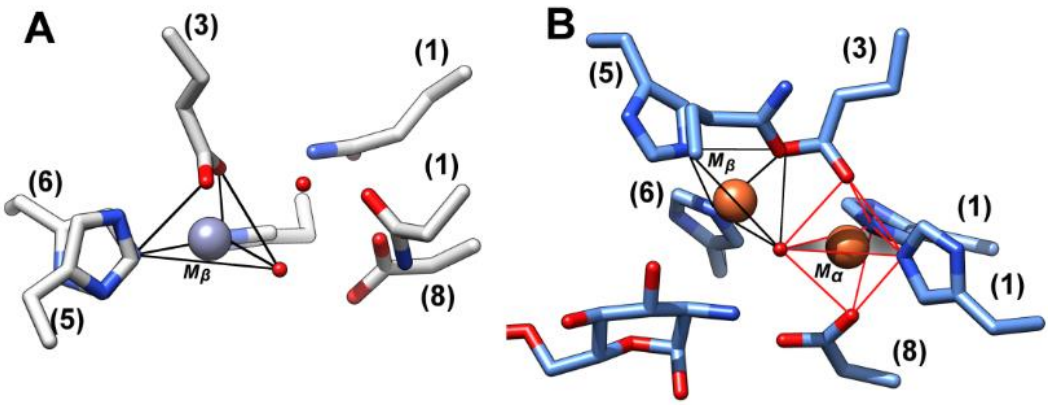


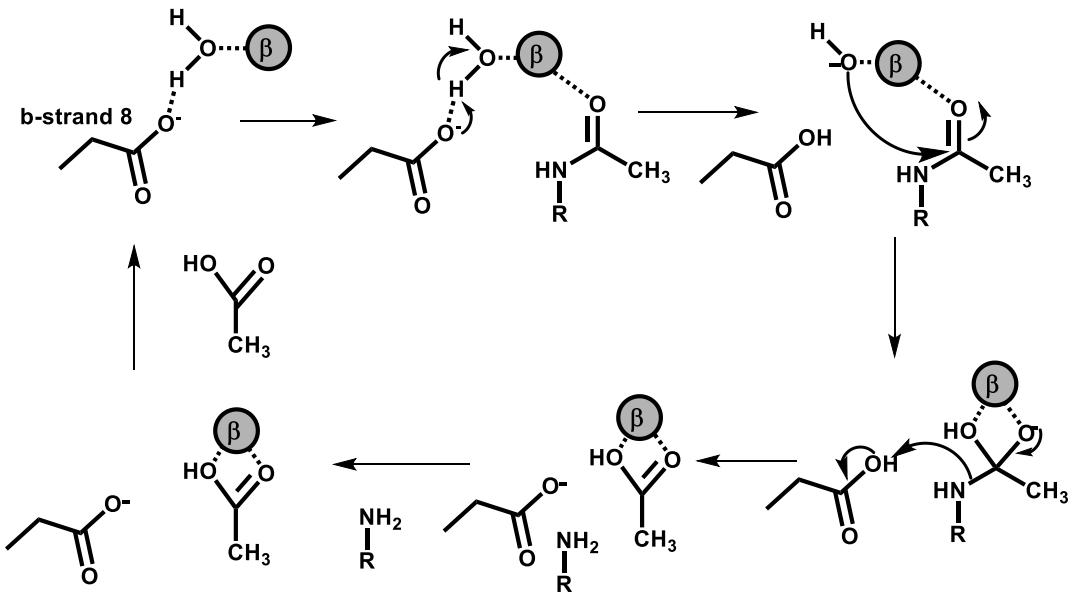
Figure 1.11. Active site comparisons of mononuclear and binuclear NagA active sites. (**A**) NagA from *E. coli* (pdb code 2p50) with single Zn^{2+} in the M_β site. (**B**) NagA from *T. maritima* containing a binuclear Fe^{2+} active site.

The chemical mechanism proposed for the mononuclear NagA of subtype VII is presented in **Scheme 1.13**. The water is activated through interactions with the beta metal and the aspartate from β-strand 8.

The chemical mechanism proposed for the mononuclear NagA involves the activation of an active site water molecule by the aspartate at the end of β-strand 8. The carbonyl of the substrate is thought to be polarized by the interactions with the beta metal and a histidine residue (His-143 not shown), which is not conserved in the binuclear metal NagA enzymes. After nucleophilic attack by the resulting hydroxide a tetrahedral intermediate is formed and subsequently collapses with the protonation of the leaving group by the active site aspartic acid (**Scheme 1.13**). The hydroxide attacks the polarized carbonyl group of the substrate to form a tetrahedral intermediate. The

C-N bond is cleaved through the collapse of the tetrahedral intermediate and protonation of the leaving group by the active site aspartate.

Scheme 1.13: Proposed mechanism for mononuclear NagA

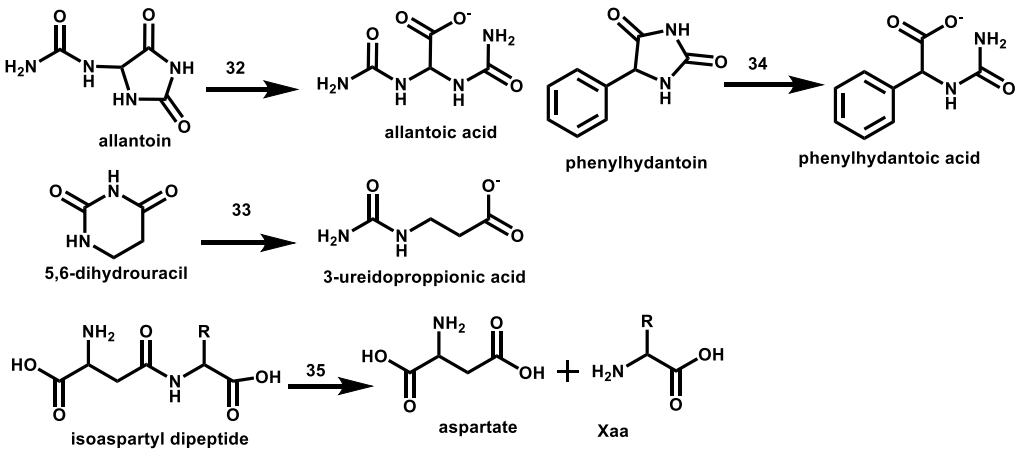


Binuclear Metal Centers

To date there are seven COGs that are known to utilize a binuclear metal center for catalysis within the AHS including: cog0044, cog0418, cog2355, cog1735, cog0804, cog3653, and cog 1001. COGs possessing a binuclear metal active site catalyze a diverse set of reactions; such as, C-N bond cleavage, deamination, C-P bond cleavage, and C-O bond cleavage. The general mechanism of action involves the activation of a nucleophilic bridging water for nucleophilic attack of the substrate, which is interacting with the beta metal.

Cog0044. Enzymes within cog0044 have been shown to primarily catalyze the hydrolytic cleavage of C-N bonds.[30, 45] The experimentally verified functions of cog include allantoinase (EC 3.5.2.5, **32**), dihydropyrimidinase (EC 3.5.2.2, **33**), phenylhydantoinase (EC 3.5.2.5, **34**), and isoaspartyl dipeptidase (EC 3.4.19.5, **35)** (**Scheme 1.14**). [46] X-ray crystal structures are available for allantoinase (pdb code 3e74), dihydropyrimidinase (pdb code 1gkp), and isoaspartyl dipeptidase (pdb codes 1onw and 1onx).

Scheme 1.14: Reactions Catalyzed by cog0044



Members of cog0044 contain a binuclear metal active site of subtype IX, which contain either Fe^{2+} or Zn^{2+} (**Figure 1.12**). The alpha metal is 5-coordinate with a distorted trigonal bipyramidal geometry (**Figure 1.13A**). The Fe^{2+} or Zn^{2+} occupying the M_α have a coordination number of 5 and a distorted trigonal bipyramidal geometry (**Figure 1.13A**). The metal occupying M_α is oriented in space through ligation to the common alpha metal ligands which include: two histidine residues from the end of β-strand 1, an aspartate from β-strand 8, and two bridging ligands. The bridging ligands are subtype dependent and vary between a glutamate, cysteine or carboxylated lysine. In subtype IX, the bridging ligands include a hydroxide/water and a carboxylated lysine residue from the end of β-strand 4 (**Figure 1.12** and **1.13**).

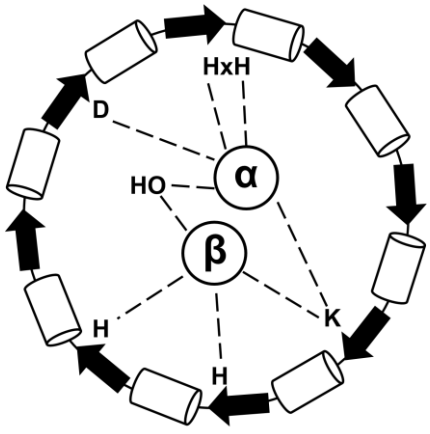


Figure 1.12: Active site schematic diagram of members of cog0044. Black arrows represent β-strands and white cylinders represent α-helices which make up the $(\beta/\alpha)_8$TIM barrel.

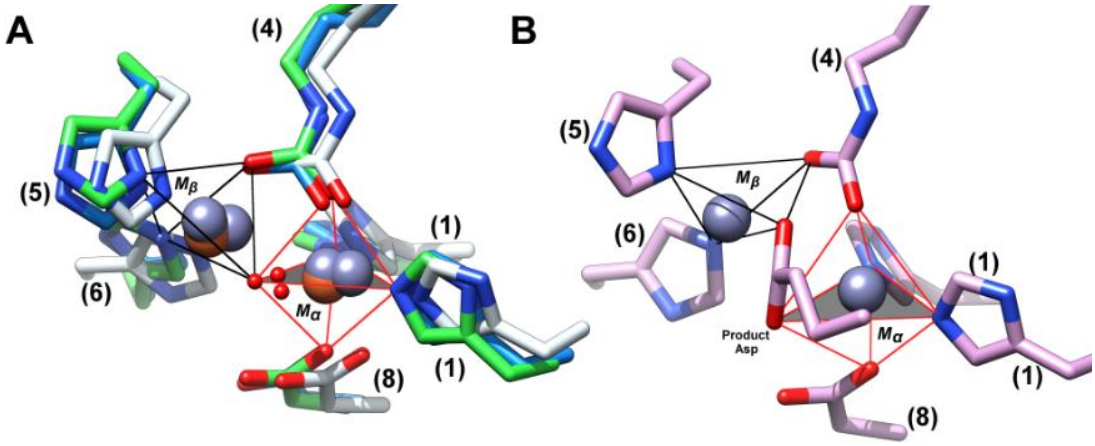
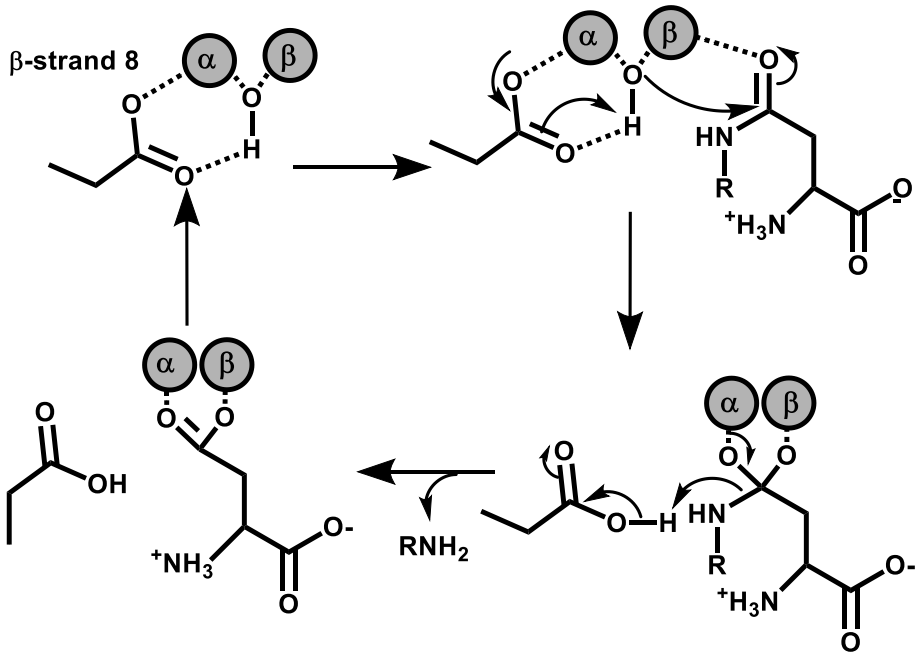


Figure 1.13: (**A**) Active site comparison of members of cog0044. IAD shown in blue (pdb code 1onw), dihydropyrimidinase shown in green (pdb code 1gkp), and allantoinase shown in white (pdb code 3e74). (**B**) Active site of IAD crystalized with the reaction product bound (aspartate).

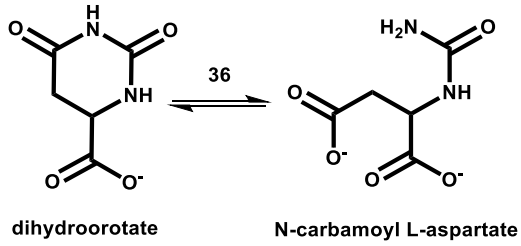
The proposed mechanism for IAD is presented in **Scheme 1.14** below. A bridging water molecule is activated for nucleophilic attack through interactions with both divalent cations in the active site. The hydroxide ion attacks the polarized carbonyl of the substrate accompanying proton transfer to the β-strand 8 aspartate. The subsequent tetrahedral intermediate collapses and C-N bond cleavage occurs subsequent to the protonation of the leaving group.

Scheme 1.15: Proposed Reaction for IAD of cog0044



*Cog0418***.** Cog0418 contains a single known function, dihydroorotase (DHO) (EC 3.5.2.3, **36**), which catalyzes the reversible cyclization of N-carbamoyL-L-aspartate to dihydroorotate (**Scheme 1.15**). [17] Crystal structures are available for DHO in the holo-form and complexed with both the product (dihydroorotate) and the substrate (N-carbamoyL-L-aspartate) (pdb codes 4lfy and 1j79).

Scheme 1.16: Reaction Catalyzed by DHO of cog0418



DHO contains a binuclear metal active site of subtype IX binding two Zn^{2+} atoms. The active site metal ions are bridged by a carboxylated lysine residue originating from the C-terminus of β-strand 4 and an active site water molecule (**Figure 1.14**). The alpha metal is oriented by the common alpha metal residues and is 5-coordinate with trigonal bipyramidal geometry (**Figure 1.14**). The beta metal, with a coordination number of 4, is bound through the common beta metal ligands with tetrahedral geometry (**Figure 1.14**). The geometries of both metals is retained when the substrate or product is bound in the active site (**Figure 1.14).**

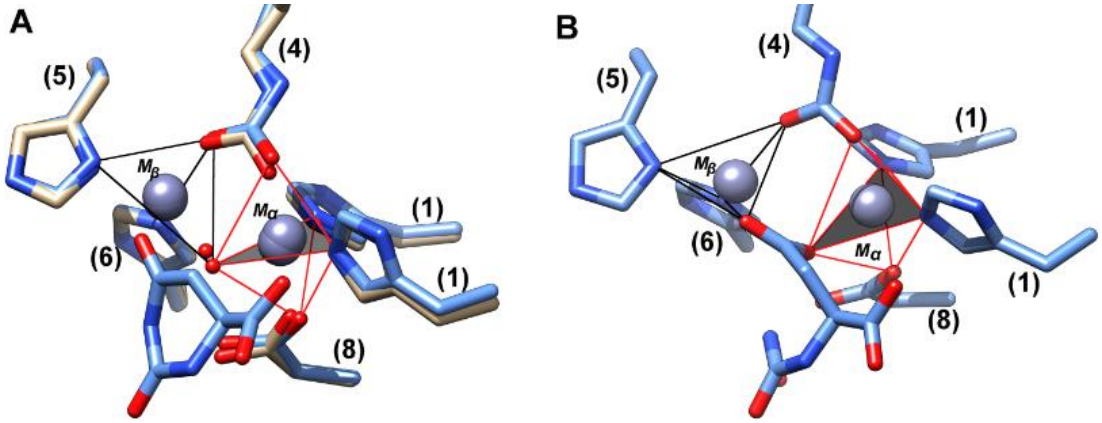
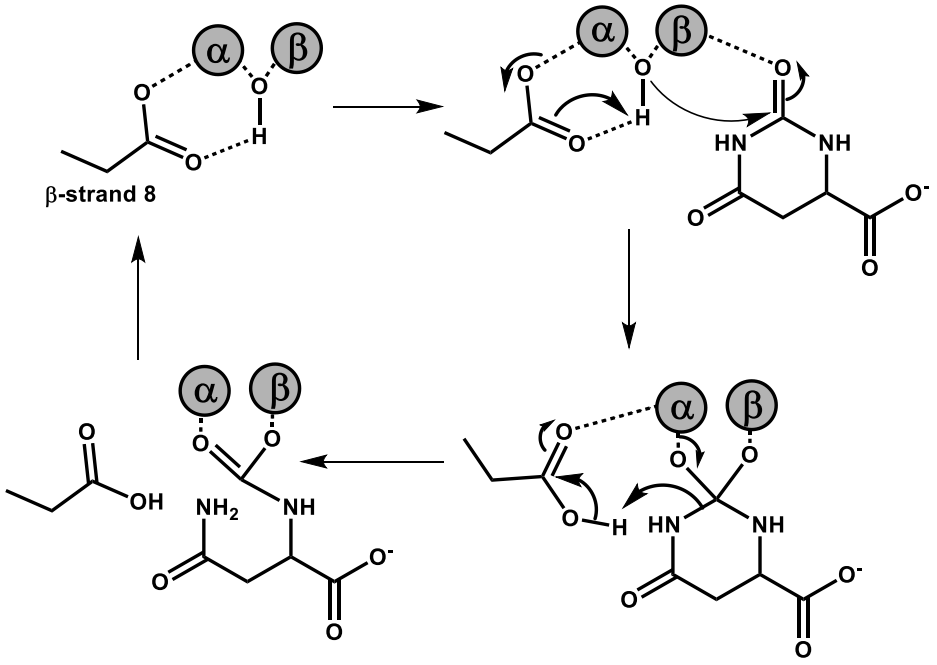


Figure 1.14. (**A**) Active site geometry analysis of DHO. With bound substrate (dihydroorotate) of the reverse reaction shown in blue (pdb code 1j79) and not bound shown in white (pdb code 4lfy). (**B**) Active site geometry analysis of DHO bound to the substrate (N-carbamoyL-L-aspartate, pdb code 1j79)

The proposed mechanism for the reverse reaction of DHO is presented in **Scheme 1.16**. The activated water molecule is poised to attack the carbonyl functional group of the substrate, which is polarized through interactions with the beta metal. A tetrahedral intermediate is formed subsequent to the nucleophilic attack, which collapses and initiates the C-N bond cleavage through the protonation of the leaving group.

Scheme 1.17 Proposed mechanism of DHO from cog0418



*Cog2355***.** The members of cog2355 have been shown to catalyze the hydrolysis of amide (peptide) bonds. Only two experimentally verified functions have been identified to date; renal dipeptidase (RDP, EC 3.4.13.19) and a promiscuous dipeptidase (Sco3058) (EC 3.4.13.19). Renal dipeptidase has been shown to catalyze the hydrolysis of various L-amino acid dipeptides, leukotriene, oxidized product of glutathione, and various β-lactams. [47, 48] The promiscuous dipeptidase, Sco3058, was shown to have activity for L-Xaa-D-Xaa dipeptides. X-ray crystal structures are available for both renal dipeptidase (pdb code 1itq) and Sco3058 (pdb codes 3itc and 3id7).

Members of cog2355 contain a binuclear metal center of subtype X binding two Zn^{2+} atoms (**Figure 1.15**). Members of cog2355 do not conserve the HxH motif from β-strand 1; however, they do contain a mutated HxD motif. The alpha metal is 5-coordinate with trigonal bipyramidal geometry, which is ligated to the HxD motif from β-strand 1, a bridging glutamate from β-strand 3, and two water molecules (**Figure 1.16A**). The axial ligands of the trigonal bipyramidal alpha metal differ as compared to other members of the AHS. Typically, the bridging residue (lysine/glutamate) and aspartate from β-strand 8 serve as axial ligands to the alpha metal. In cog2355, histidine from β-strand 1 and an active site water molecule act as the axial ligands (**Figure 1.15**). The beta metal is 4-coordiante with tetrahedral geometry, oriented in space through ligation to the bridging residues and the common beta metal ligands (**Figure 1.15** and **1.16**). When inhibitor (citrate) binds the geometry of the beta metal is changed to trigonal bipyramidal (**Figure 1.16B**).

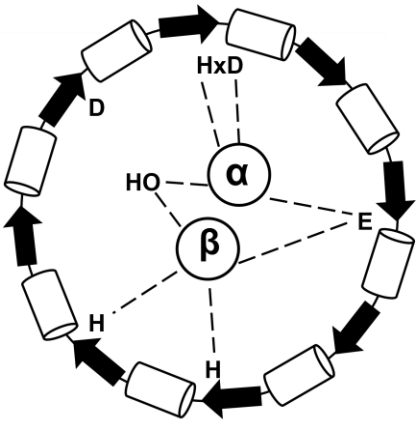


Figure 1.15. Active site schematic of members of cog2355. Black arrows represent β-strands and white cylinders represent α-helices which make up the $(\beta/\alpha)_8$TIM barrel.

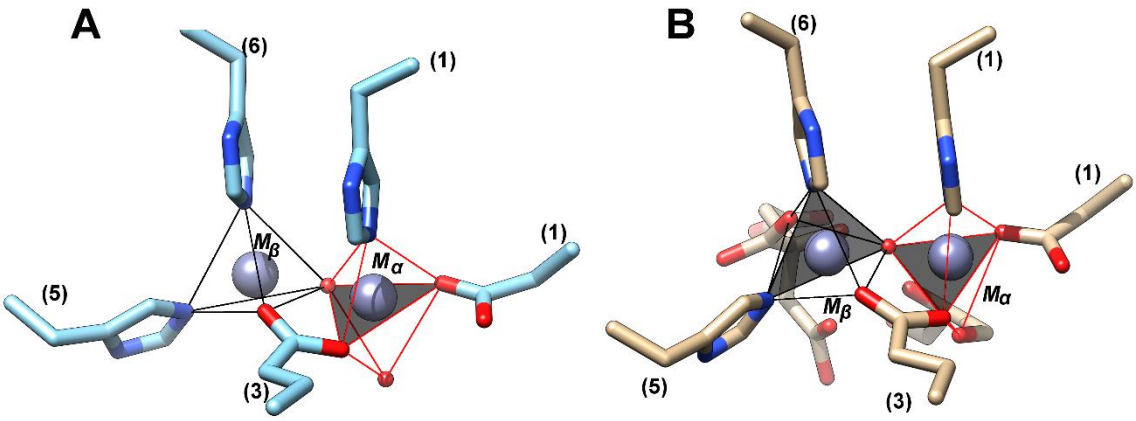
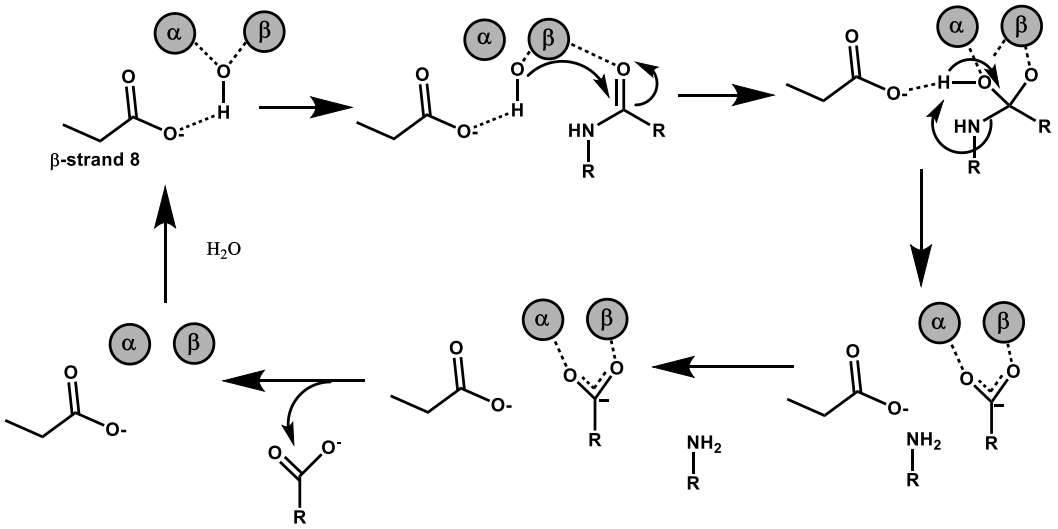


Figure 1.16. (**A**) Active site of RDP (pdb code 3id7). (**B**) Active site of Sco3058 with bound citrate.

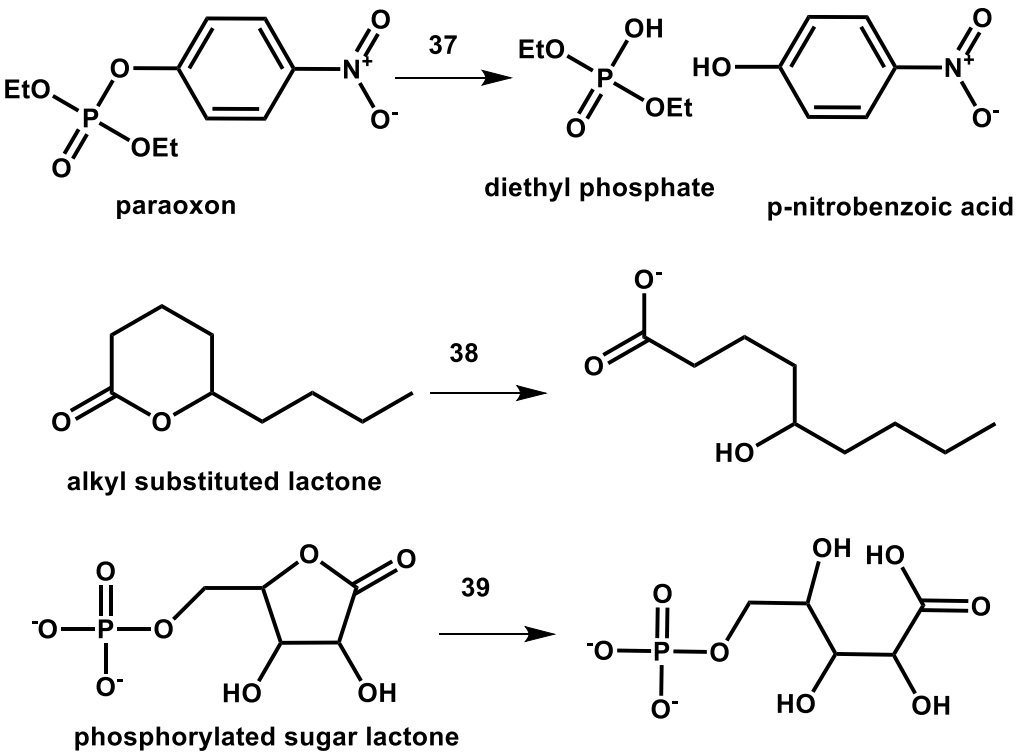
A proposed mechanism for RDP is presented in **Scheme 1.17**. An activated water molecule attacks the polarized carbonyl of the substrate. A tetrahedral intermediate is formed and collaposes through subsequent proton transfer between the leaving group and substrate, facilitated by the β-strand 8 aspartate.

Scheme 1.18 : Proposed Mechanism of RDP from cog2355.



Cog1735. Cog1735 was originally shown to catalyze the hydrolysis of P-O bonds through the discovery of phosphotriesterase (PTE, EC 3.1.8.1, **37**), which hydrolyzes organophosphates (**Scheme 1.18**).[49] More recently, it has been shown that members of cog1735 can cleave C-O bonds of lactone substrates. The enzymes Dr0930 (**38**) and Lmo2620 (**39**) have been shown to catalyze the hydrolysis of alkyl substituted lactones and phosphorylated lactones, respectively (**Scheme 1.18**).[50, 51] X-ray crystal structures are available for PTE (pdb codes 1hzy and 1ez2), Dr0930 (pdb codes 3fdk and 3htw), and Lmo2620 (pdb code 3pnz).

Scheme 1.19: Reactions catalyzed by members of cog1735



The members of cog1735 have been shown to contain a binuclear metal active site of subtype IX. The alpha metal is 5-coordinate with trigonal bipyramidal geometry, which is oriented in space by the common alpha metal ligands. The beta metal is also 5-coordinate with trigonal bipyramidal geometry (**Figure 1.17**)**.** The geometries of both metal sites is retained upon binding of a phosphonate inhibitor.

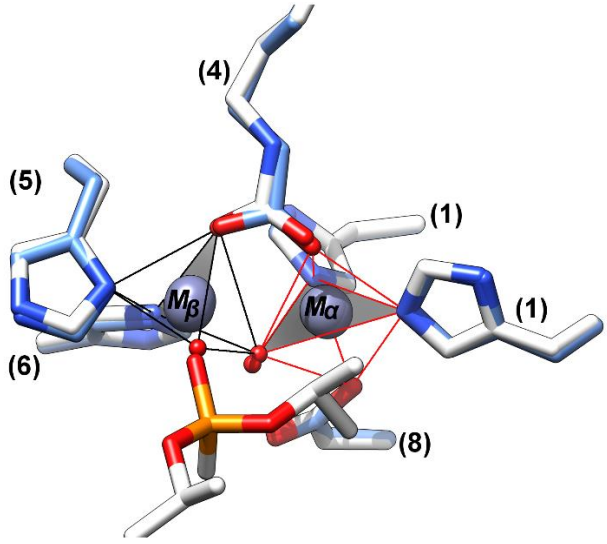
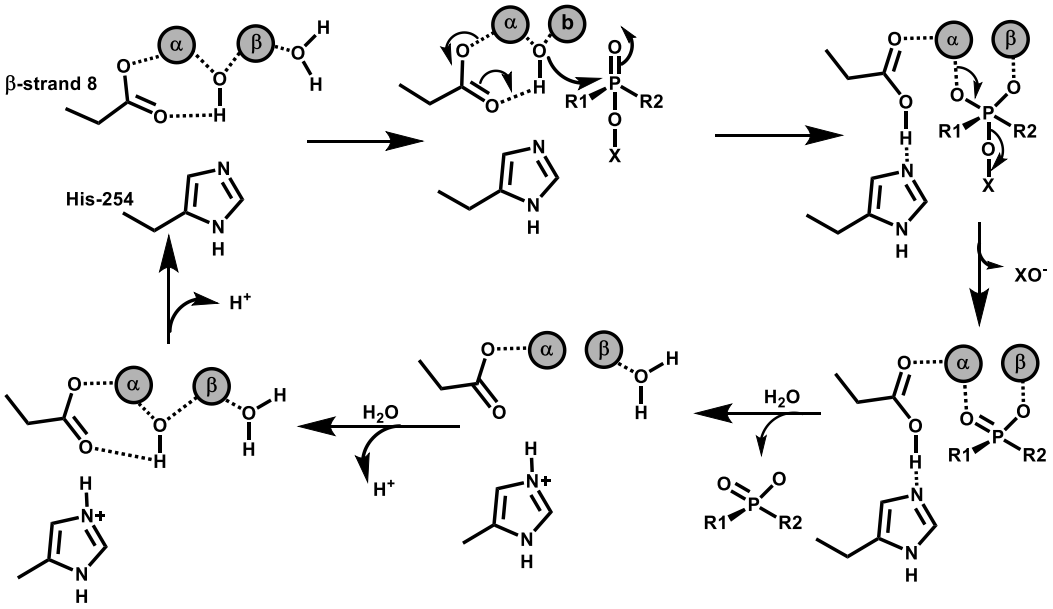


Figure 1.17. Active site comparison of PTE with bound phosphonate inhibitor shown in white (pdb code 1ez2) and without inhibitor shown in blue (pdb code 1hzy)

The proposed mechanism for PTE is provided below in **Scheme 1.19**. Once the organophosphate substrate binds a nucleophilic attack occurs at the phosphorous center by the bridging hydroxyl group forming a tetrahedral intermediate. The P-O bond is cleavage subsequent to the collapse of the intermediate. The active site is

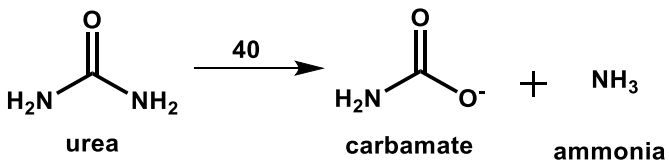
restored through the proton shuttling between the active site aspartic acid and histidine residue.

Scheme 1.20: Proposed mechanism for PTE from cog1735



Cog0804. The single function known within cog0804 is one of the founding members of the AHS, Urease (URI, EC 3.5.1.5, **40**) which catalyzes the transformation of urea to carbamate and ammonia (**Scheme 1.20**). [52] As the first protein demonstrated to have enzymatic activity and to be crystallized, there are many structures available in the protein data bank. [53] These include urease from *Bacillus pasteurii* crystalized with bound inhibitor diamidophosphate (pdb code 3ubp) and the native urease (pdb code 2ubp). Urease is the only member of the AHS to utilize Ni^{2+} cations for catalytic activity.

Scheme 1.21: Reaction catalyzed by urease from cog0804



Urease contains a binuclear metal center of subtype IX with two Ni^{2+} ions occupying both the M_α and M_β sites. The metal coordinating ligands originating from the protein active site are similar to those shown for PTE; however, the geometry of the alpha Ni^{2+} is 6 coordinate octahedral after coordinating an extra water molecule (**Figure 1.18**). There are a total of four active site water molecules within urease. Each metal is coordinated to a water ligand in addition to the bridging water molecule. The last is oriented through non-metal coordinating active site ligands. When urease or an

inhibitor (diamidophosphate) binds, the non-metal bridging water molecules are replaced (**Figure 1.18**).

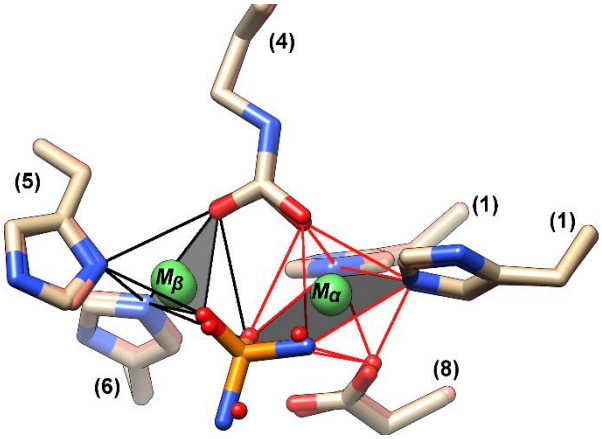
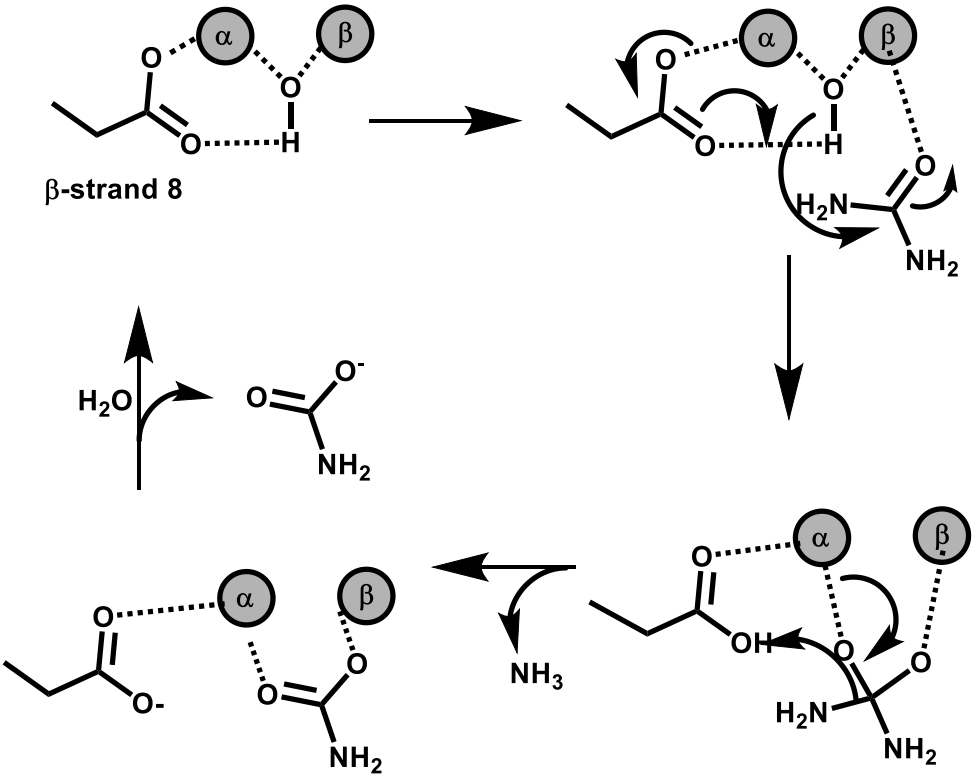


Figure 1.18. Active site analysis of native urease shown in white (pdb code 2ubp) and with bound diamidophosphate inhibitor in orange (pdb code 3ubp).

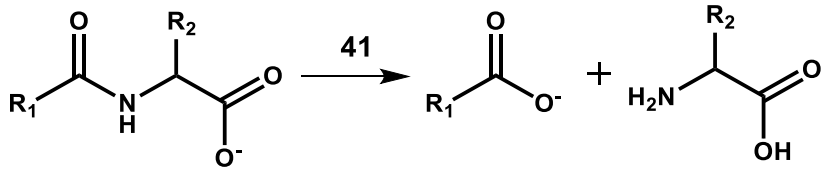
The reaction and catalytic mechanism of urease has been extensively studied since the original function discovery. [52, 54, 55] Proposed herein, urea binds the active site with the carbonyl oxygen interacting with the beta metal, ultimately polarizing the functional group. The C-NH_2 bond is simultaneously polarized through interactions with the alpha Ni^{2+}. The bridging nucleophilic hydroxide can now attack the activated urea molecule to forma tetrahedral intermediate. Protonation of the leaving group by the β-strand 8 aspartate initiates the collapse of the intermediate with subsequent C-N bond cleavage.

Scheme 1.22: Proposed Mechanism of Urease from cog0804.



Cog3653. The enzymes of cog3653 have primarily been identified as catalyzing deacylation reactions. These enzymes, D-amino acid deacylases (EC 3.5.1.14,DAA, **41**), exhibit relatively promiscuous activity for a variety of amino acid derivatives shown in **Scheme 1.22** .[56] Experimentally verified functions include BB3285 which hydrolyzes N-acyl substituted derivatives of D-glutamate and Gox1177 which preferentially hydrolyzes N-acyl substituted derivatives of hydrophobic D-amino acids. X-ray crystal structures are available for DAA from *Alcaligenes faecalis* (pdb code 1m7j) and Bb3285 (pdb code 3giq).

Scheme 1.22: Reactions Catalyzed by Members of cog3653.



Members of cog3653 contain a binuclear metal center of subtype XIII, binding two Zn^{2+} ions (**Figure 1.19**). It has been reported that DAA requires only a single metal ion for catalytic activity; however, it conserves proper ligands for a binuclear metal site and has been crystalized with metal occupying both sites. [57] If DAA utilizes only a single metal ion it is the metal occupying the beta site and be of subtype VII. The alpha metal is coordinated through the usual histidines from the HxH motif, aspartate from β-strand

8, and a bridging water molecule. Both metals are bridged by an unusual cysteine residue originating from the end of β-strand 2 (**Figure 1.20**). The alpha Zn^{2+} is 5 coordinate with a trigonal bipyramidal geometry. The beta Zn^{2+} is 4 coordinate with a tetrahedral geometry. Both metal geometries are retained in the presence of bound inhibitor (N-methylphosphonate derivative of D-glutamate).

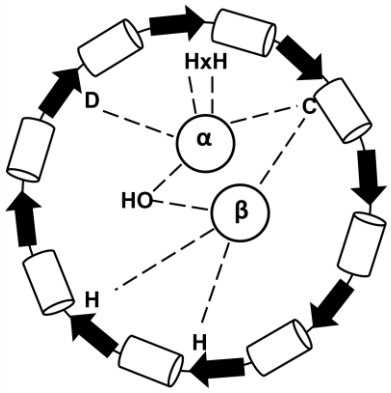


Figure 1.19. Active site schematic of the members of cog3653. Black arrows represent β-strands and white cylinders represent α-helices which make up the $(\beta/\alpha)_8$TIM barrel.

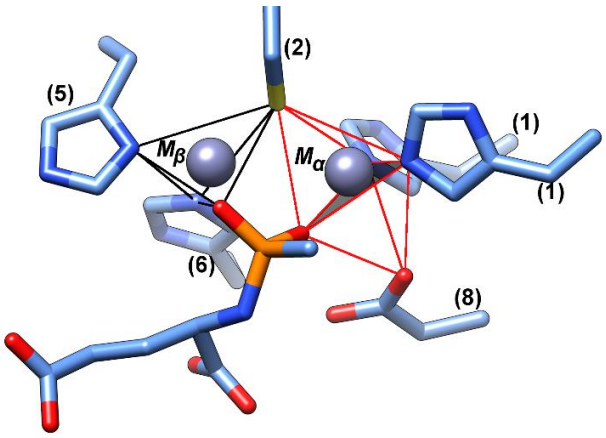
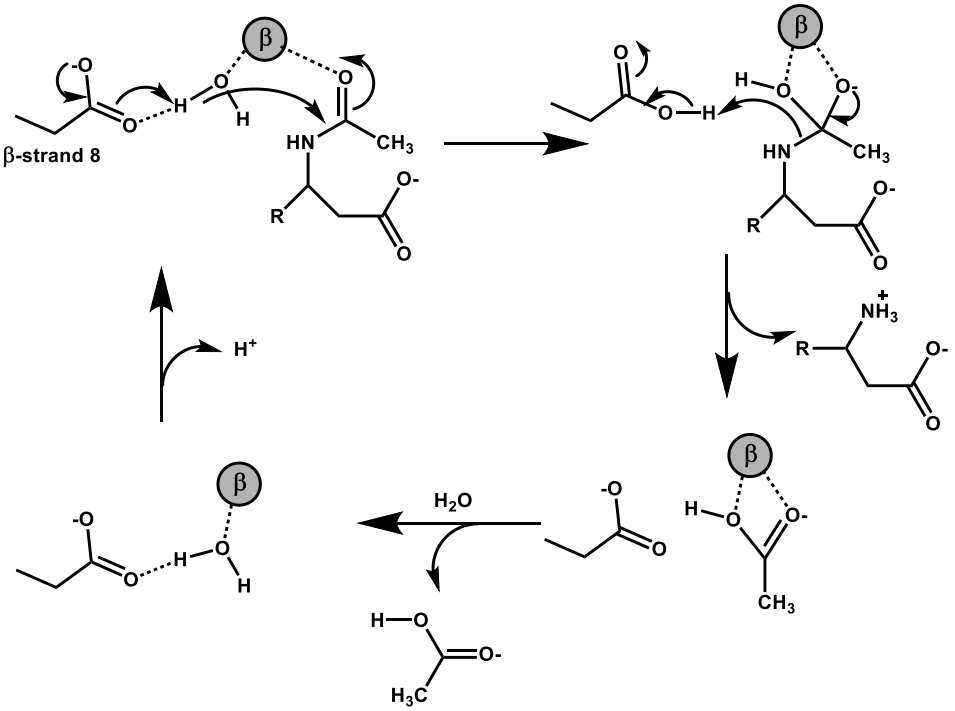


Figure 1.20. Active site of inhibitor bound DAA (Bb3285) showing the binuclear metal coordination and geometry (pdb code 3giq)

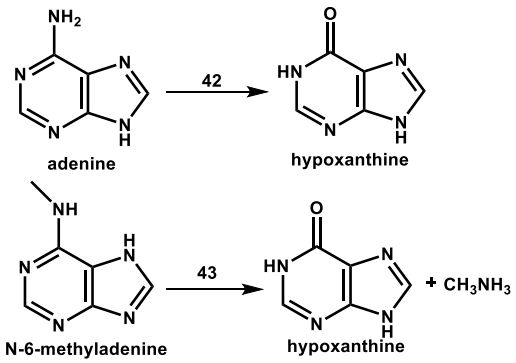
The proposed mechanism describes the catalysis of DAA utilizing a single Zn^{2+} occupying the M_β site; however, the mechanistic details would not significantly vary in the presence of a second Zn^{2+} atom (**Scheme 1.23**). The activated water molecule is positioned for attack of the polarized carbonyl of the substrate. The nucleophilic attack results in a tetrahedral intermediate forming and subsequently collapsing through the protonation of the β-strand 8 aspartate, resulting in C-N bond cleavage.

Scheme 1.23: Proposed mechanism for the reaction catalyzed by DAA utilizing a single divalent cation.



Cog1001. The enzymes within cog1001 have been shown to primarily perform deamination reactions of adenine like nucleobases; such as, adenine (AdA, EC 3.5.4.2, **42**) and N-6-methyladenine (EC 3.5.4.X, **43**) (**Scheme 1.24**).[58, 59] AdA from cog1001 was discovered prior to the identification of the mononuclear AdA from cog1816. The only X-ray crystal structures available for cog1001 are for adenine deaminase containing a di-iron or di-manganese metal center (pdb codes 3t81 and 3nqb). It is important to note the secondary activity of AdA from cog 1001. Di-iron AdA from cog1001 has been reported to have catalase activity, the first AHS member to catalyze radical chemistry.[60]

Scheme 1.24: The reactions catalyzed by members of cog1001.



Adenine deaminase contains a unique binuclear metal center of subtype XI which is not observed in other members of the AHS (**Figure 1.20**). Protein ligands for the metal occupying the M_α site are as follows: HxH from β-strand 1, aspartate from β-strand 8, an unusual bridging glutamate from β-strand 4, and a bridging water molecule. The coordination number and geometry for the alpha site are dependent on

the metal nucleus. When Fe^{2+} is occupying this site the coordination number is 5 and the geometry observed is trigonal bipyramidal; however, when Mn^{2+} is bound an extra water molecule is coordinating increasing the coordination number to 6 and the geometry to octahedral (**Figure 1.20**, **Figure 1.21**). Similarly, coordination and geometry of the M_β is metal dependent. Direct protein ligands for the beta metal in addition to the bridging ligands are two histidine residue from β-strand 5, and an HxxE motif originating from the C-terminus of β-strand 6. When Mn^{2+} is occupying this site an additional water ligand is coordinated and observing octahedral geometry (**Figure 1.21**).

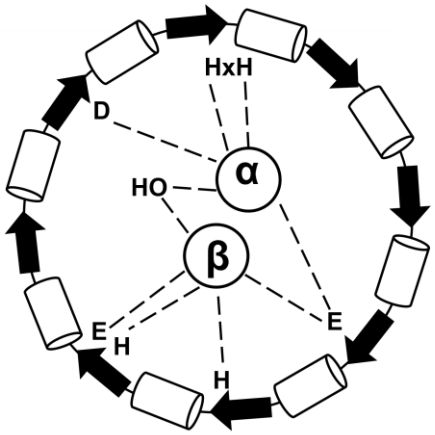


Figure 1.21. Active site schematic of AdA from cog1001. Black arrows represent β-strands and white cylinders represent α-helices which make up the $(\beta/\alpha)_8$TIM barrel.

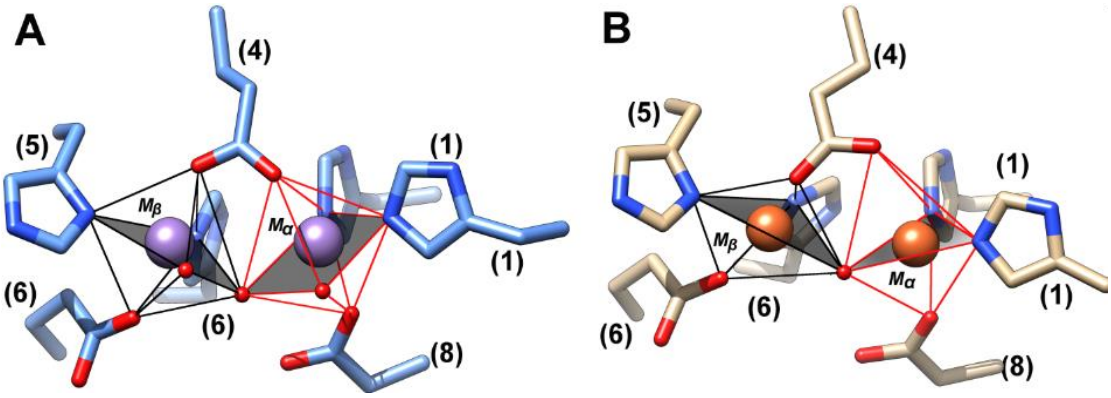
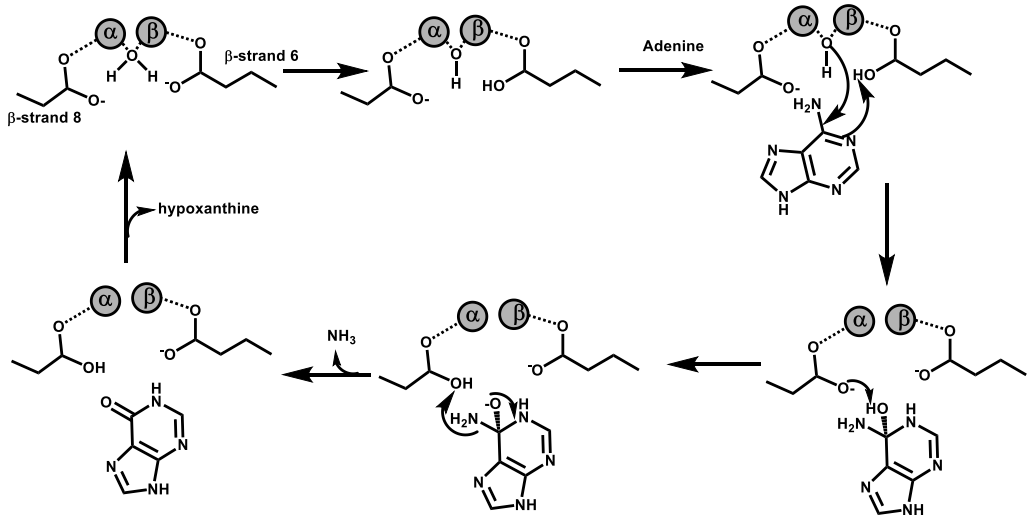


Figure 1.22. (A) Mn/Mn active site of AdA from cog1001 (pdb code 3t81). (B) Fe/Fe active site of AdA from cog1001 (pdb code 3nqb). Numbers represent the β-strand to which the residue originates.

A proposed mechanism is presented in **Scheme 1.25** below. The bridging water molecule is activated for nucleophilic attack through proton abstraction by glutamate from β-strand 6. A tetrahedral intermediate is formed after nucleophilic attack by the hydroxide ion at C-6 of the re-face of the adenine substrate. The C-N bond is cleaved through the collapse of the tetrahedral intermediate and donation of a proton to the leaving group.

Scheme 1.25: Proposed catalytic mechanism of AdA from cog1001.

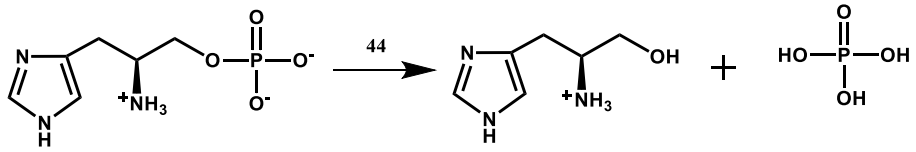


Trinuclear Metal Center

Two COGs within the AHS have been identified as contain three metal sites; cog1387 and cog0613. Both COGs have been identified to catalyze the hydrolysis of O-P bonds of phosphoesters.

Cog1387. Cog1387 contains a single enzyme of known function, histidinol phosphate phosphatase (HPP, EC 3.1.3.15, **44**). This enzyme catalyzes the dephosphorylation of L-histidinol phosphate to L-histidinol and inorganic phosphate in the L-histidine biosynthetic pathway (**Scheme 1.26**).[9] Crystal structures are available for HPP including a structure bound with L-histidinol and phosphate (pdb code 4gyf).

Scheme 1.26: Reaction catalyzed by HPP from cog1387



The active site of HPP contains a trinuclear metal center of subtype XIV with Zn^{2+} occupying all three sites. Members of this COG contain a distorted $(\beta/\alpha)_7$TIM barrel compared to the $(\beta/\alpha)_8$TIM barrel observed in other members of the AHS **(Figure 1.22**). The β-strand at position 3 is elongated and has been renumbered as β-strand 3/4 to facilitate comparison with other AHS active sites. [61] The metal occupying the M_α site is ligated through interactions with the HxH motif, a bridging glutamate at β-strand

3/4, an aspartate at β-strand 8, and a bridging water molecule (shown in red **Figure 1.23A**). In the structure with bound phosphate (pdb code 4gyf), the phosphate coordinates to the alpha metal increasing the coordination number to 6 and octahedral geometry. It is likely that the phosphate is taking the place of a water ligand; however, it is possible that the binding of phosphate/substrate induce a change in geometry (**Figure 1.23A**). The Zn^{2+} occupying the M_β site is ligated by histidine residues from β-strands 5 and 6 in addition to the bridging ligands (shown in green **Figure 1.23A**). The beta metal has a coordination number of 5 and trigonal bipyramidal geometry when coordinated to the bound phosphate molecule. The metal occupying the M_γ site (shown in blue **Figure 1.23B**) is ligated by an aspartate from β-strand 1, two histidines located at the ends of β-strands 2 and 8, and a phosphate molecule. The gamma metal is pentacoordinate with a tetrahedral geometry.

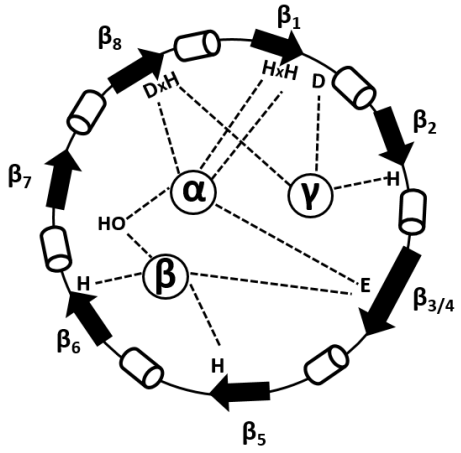


Figure 1.23. Active site schematic of HPP from cog1387. Black arrows represent β-strands and white cylinders represent α-helices which make up the (β/α)8TIM barrel.

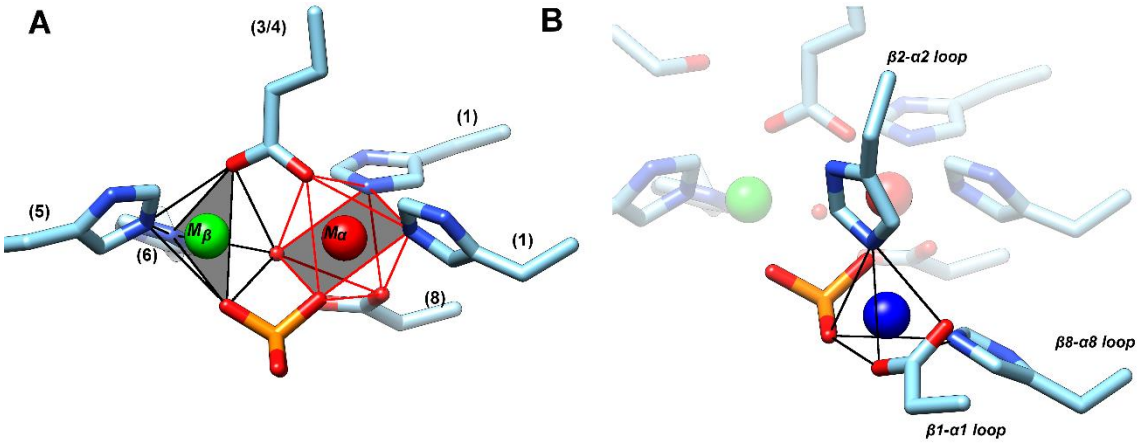
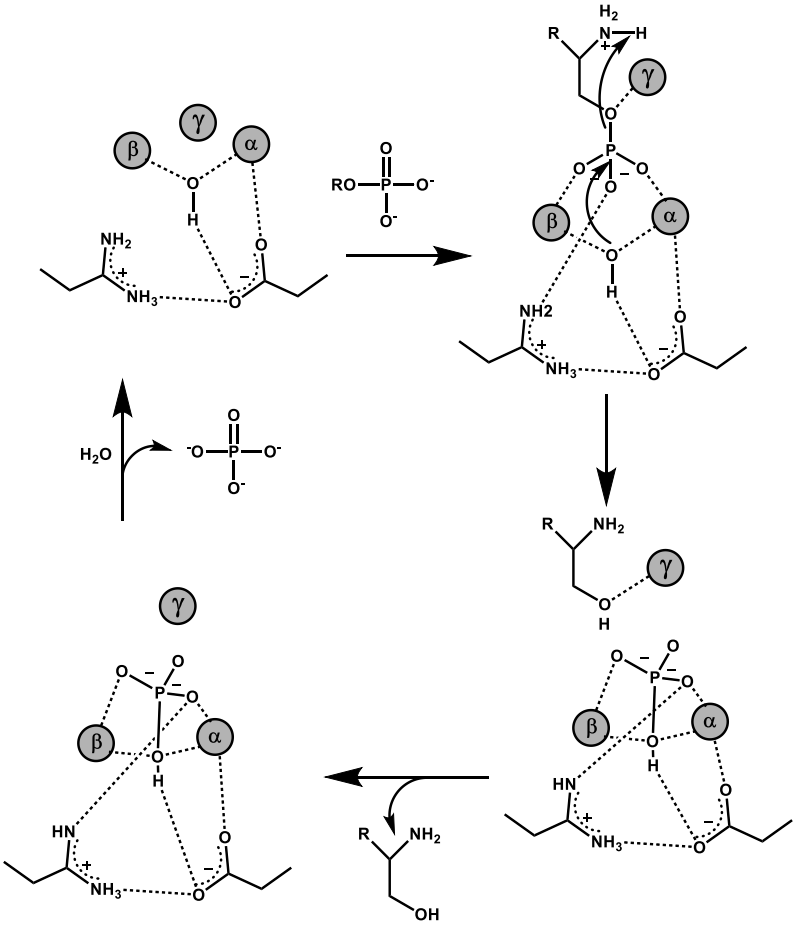


Figure 1.24 Active site analysis of HPP. Zn^{2+} cations are color coded as follows M_α is shown in red, M_β is shown in green, and M_γ is shown in blue. (**A**) M_α and M_β of HPP (pdb code 4gyf). (**B**) M_γ of HPP (pdb code 4gyf).

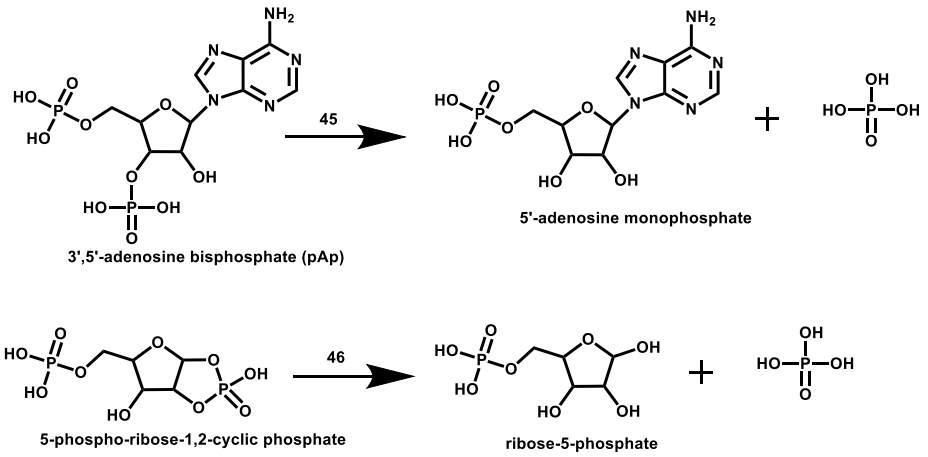
A proposed mechanism for the hydrolysis of phosphoesters is presented in **Scheme 1.27**. The substrate binds, with the phosphate end coordinating, as shown in **Figure 1.23A** and **B**, with three oxygen atoms coordinating each of the three metals. The bridging nucleophile attacks the phosphorous center with the P-O bond cleaving in a concerted manner through the protonation of the leaving group. This protonation event is thought to occur through an intramolecular mechanism with the amino group of the leaving group.

Scheme 1.27: Proposed reaction mechanism for trinuclear metal active sites of the AHS



Cog0613. The reactions catalyzed by members of cog0613 are similar to those of cog1387, both of which contain three metal binding sites. Experimentally verified functions of cog0613 include 3',5'-adenosine bisphosphate (pAp) phosphatase (EC 3.2.3.X, **45**) and 5-phospho-ribose-1,2-cyclic phosphate phosphatase (EC 3.2.3.X, **46**) (**Scheme 1.28**).[16, 61] X-ray crystal structures are only available for pAp phosphatase with full trinuclear Mn^{2+} active site and bound AMP (pdb 2yb1) and in the apo form (pdb 2yb4).

Scheme 1.28: Reactions Catalyzed by Members of cog0613.



Similar to HPP from cog1375, members of cog0613 contain a trinuclear metal active site of subtype XIV. The active site of pAp phosphatase (pdb code 2yb1) contains Mn^{2+} in each of the three metal sites. The alpha metal (shown in red **Figure 1.24**) has a coordination number of 6 with the following ligands: HxH from β-strand 1, aspartate from β-strand 8, a bridging water molecule, and a bridging glutamate from β-strand 3/4 (**Figure 1.24A**). The beta Mn^{2+} (shown in green **Figure 1.24**) is 5 coordinate with trigonal bipyramidal geometry and is ligated by two histidine residues from β-strand 5 and 6 in addition to the bridging ligands. The gamma metal (shown in blue **Figure 1.24 B**) is 6 coordinate as a result of the bound AMP. AMP exhibits bidentate coordination to the gamma metal as the 5th and 6th ligands with a distorted octahedral geometry.

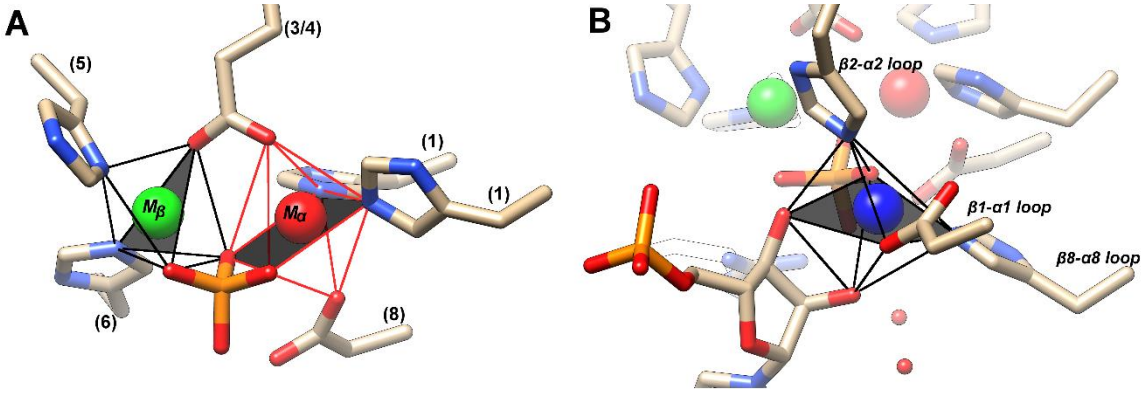


Figure 1.25. Active site analysis of pAp phosphatase (pdb code 2yb1) Mn2+ cations are color coded as follows: M_α is shown in red, M_β is shown in green, and M_γ is shown in blue. (**A**) M_α and M_β of pAp phosphatase (**B**) M_γ.

Metal Independent Centers (cog3618)

In this dissertation, the mechanism of action of the first metal independent member of cog3618 was characterized, and the mechanistic details were utilized to predict and identify the functions of two proteins of unknown function. In first part of this dissertation, the structure and catalytic mechanism of 2-pyrone-4,6-dicarboxylic acid lactonase will be addressed. The X-ray crystal structure of LigI was solved, by our collaborators in Dr. Steve Almo's laboratory at Albert Einstein College of Medicine, in the presence of bound product. The second part of this dissertation addresses the structure and discovery of the enzyme L-fucono-1,5-lactonase. Here we have identified that this enzyme prefers the less stable L-fucono-1,5-lactone as a substrate compared to the chemically stable L-fucono-1,4-lactone. The final section of this dissertation will address the In vitro characterization of an L-galactose pathway in the human gastrointestinal bacterium, *Bacteroides vulgatus*. A member of cog3618 was identified to catalyze the hydrolysis of L-galactono-1,5-lactone. In addition, two adjacent operon oxidoreductase enzymes were characterized. The first enzyme in the operon was determined to oxidize L-galactose with high specificity. The second enzyme oxidizes the product of L-galactono-1,5-lactonase to D-tagaturonate, which can then be funneled into the catabolic D-galacturonate pathway.

CHAPTER II

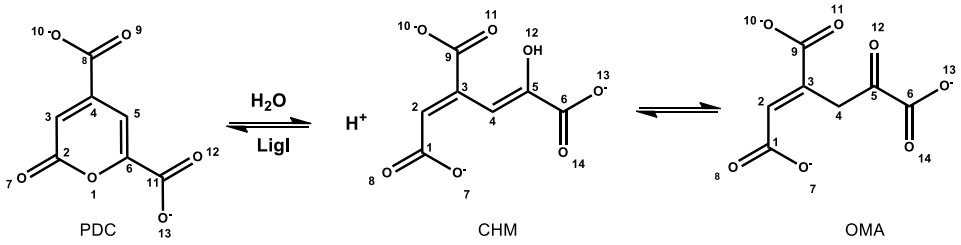
STRUCTURE AND CATALYTIC MECHANISM OF LIGI: INSIGHT INTO THE AMIDOHYDROLASE ENZYMES OF COG3618 AND LIGNIN DEGRADATION*

Lignin is the most abundant aromatic biomolecule found in nature. This compound and associated degradation products have implications for the production of biofuels and in the manufacture of adhesives and polyesters.[62] The enzyme 2-pyrone-4,6-dicarboxylate lactonase (LigI) plays an important role in the metabolism of lignin-derived aromatic compounds. In the syringate degradation pathway, lignin and tannin-derived compounds containing syringyl moieties are metabolized. Alternatively, lignin-derived compounds containing guaiacyl moieties are metabolized via the protocatechuate 4,5-cleavage pathway.[63] LigI catalyzes the reversible hydrolysis of 2-pyrone-4,6-dicarboxylate (PDC) to a mixture of 4-carboxy-2-hydroxymuconate (CHM) and 4-oxalomesaconate (OMA) as shown in **Scheme 2.1.**[63, 64] Amino acid sequence alignments have identified LigI from *Sphingomonas paucimobilis* as a member of the amidohydrolase superfamily (AHS).

---

*Reprinted with permission from "Structure and Catalytic Mechanism of LigI: Insight into the Amidohydrolase Enzymes of cog3618 and Lignin Degradation" by Merlin Eric Hobbs, Vladimir Malashkevich, Howard J. Williams, Chengfu Xu, J. Michael Sauder, Stephen K. Burley, Steven C. Almo, and Frank M. Raushel, *Biochemistry*, 2012, *51* (16), pp 3497–3507, Copyright 2012 American Chemical Society

Scheme 2.1: Reaction catalyzed by LigI



The amidohydrolase superfamily was first identified by Sander and Holm from recognition of the structural and sequence similarities of urease, phosphotriesterase, and adenosine deaminase.[7] Since the discovery of the AHS, more than 23000 proteins have been classified as members of this superfamily. Members of this superfamily catalyze a diverse set of chemical reactions, including the hydrolysis of amide or ester bonds, deamination of nucleic acids, decarboxylation, isomerization, and hydration[8]. Proteins in the AHS possess a distorted $(\beta/\alpha)_8$-TIM barrel structural fold that typically houses an active site containing one to three divalent metal ions. The active site metal ions are ligated to the protein through interactions with residues that originate from the C-terminal ends of β-strands that form the core of the β-barrel. The mononuclear metal centers are formed through ligation with a pair of histidine residues in an HxH motif from β-strand 1, a histidine from β-strand 5, and an aspartate from β-strand 8. Proteins with binuclear metal centers coordinate two divalent cations through ligation with the HxH motif from β-strand 1, a carboxylated lysine (or glutamate) from β-strand 3 or 4, two histidine residues from the ends of β-strands 5 and 6, and an aspartate from

β-strand 8. The carboxylated lysine functions as a bridge between the two divalent cations.

In contrast, LigI does not appear to be a metalloenzyme. There is no obvious residue that could bridge a binuclear metal center at the C-terminus of β-strand 3 or 4. In addition, the conserved histidine found at the C-terminal end of β-strand 5 in nearly all members of the AHS is replaced with a tyrosine in LigI.[8] These observations suggest that LigI will have an active site that is significantly different from those of the structurally characterized members of the amidohydrolase superfamily.

The AHS has been divided into 24 clusters of orthologous groups (COG) by the National Center for Biotechnology Information (NCBI).[14, 15] LigI is one of three partially characterized enzymes within cog3618, along with 4-sulfomuconolactonase (4-SML) and L-rhamnono-1,4-lactonase.[65, 66] All three of these enzymes catalyze the hydrolytic cleavage of lactones. A sequence similarity network for cog3618 is presented in **Figure 2.1.**[67, 68] The experimentally verified examples of LigI and 4-SML are found in group 3 of cog3618; these two proteins share 38% sequence identity. The L-rhamnono-1,4-lactonase is found in group 8 of cog3618 and shares 20% sequence identity with LigI.

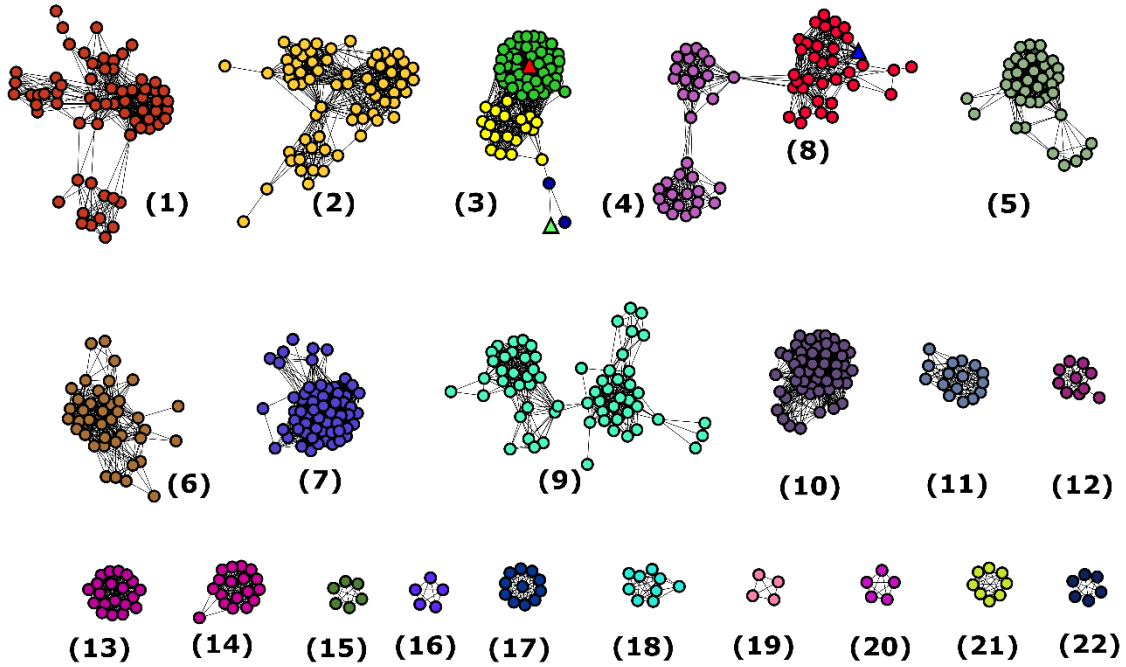


Figure 2.1. Cytoscape-generated sequence similarity network of cog3618 at a BLAST E value cutoff of 10–70. Nodes represent proteins within cog3618 connected to other proteins with an E value cutoff of $<10^{-70}$ by lines. The stringency value of 10^{-70} was arbitrarily chosen on the basis of what appears to be the presence of smaller isofunctional groups. The triangle-shaped nodes colored red and blue in group 3 represent LigI from S. paucimobilis and 4-SML from Agrobacterium tumefaciens, respectively. Circular nodes colored green in group 3 are annotated here as LigI, based on conservation of active site residues of LigI from S. paucimobilis. Nodes colored blue in group 3 are predicted to be 4-SML enzymes. The yellow nodes in group 3 are unannotated. The triangle-shaped node colored blue in group 8 is L-rhamnono-1,4-lactonase from *Azotobacter vinelandii*. The rest of the proteins are of unknown function.

Relatively little is known about the mechanism of action of LigI.[69, 70] Previously determined values of k_{cat} for the forward and reverse reactions vary from 110 to 270 s^{-1}. The K_m for PDC hydrolysis is 70 μM, and the Michaelis constant for CHM and OMA during lactone synthesis is 35 μM. Previous published reports suggested that the addition of divalent metal ions to the assays provides no rate enhancement or is inhibitory.[69] Here, we report the crystal structure of LigI in the presence and absence of the product bound in the active site. The catalytic mechanism for the hydrolysis of PDC was elucidated using pH–rate profiles and active site-directed mutants.

Materials and Methods

Materials. All chemicals and buffers were purchased from Sigma Aldrich unless otherwise specified. The synthesis of 2-pyrone-4,6-dicarboxylate was conducted according to published procedures. [63] All buffers, except for Tris-HCl, were titrated to the stated pH with NaOH.

Expression and Purification of LigI from S. paucimobilis. Original cloning and overexpression of LigI was conducted through our collaboration with Stephen Burley, at Lilly Biotechnology Center, 10300 Campus Point Drive, San Diego, California.

Construction of Mutant Enzymes. All single-site mutants of LigI were prepared according to the recommendations of the QuikChange PCR procedure manual. Protein expression and purification procedures were performed with slight modifications as described above for wild-type LigI.

Crystallization and Structure Determination. The X-ray crystal structures of LigI were solved through collaboration with Dr. Steven Almo and Dr. Vladimir Malashkevich at the Department of Biochemistry, Albert Einstein College of Medicine, at Bronx, New York. The PDB codes are 4d8l, 4di8, 4di9, and 4dia.

Enzymatic Synthesis of 4-Oxalomesaconate. OMA was synthesized enzymatically from PDC using LigI as a catalyst. The reaction was conducted in 50 mM CHES (pH 10) with 6.0 mM PDC and 10 μM LigI in a volume of 5.0 mL. The reaction was monitored spectrophotometrically at 312 nm until the reaction was complete. The enzyme was removed when the mixture was passed through a VWR centrifugal filter with a molecular mass cutoff of 3 kDa.

Measurement of Lactonase Activity. The forward and reverse reactions catalyzed by LigI were followed spectrophotometrically with a SpectraMax-340 UV–visible spectrophotometer using a differential extinction coefficient (Δε312) of 6248 M^{-1} cm^{-1} at 312 nm[62, 63]. Standard assay conditions for the hydrolysis of PDC included 50 mM BICINE (pH 8.25), varying concentrations of PDC (0–300 μM), and LigI in a final volume of 250 μL at 30 °C. For the reverse reaction, standard assay conditions included 50 mM HEPES (pH 7.5), variable concentrations of OMA (0–300 μM), and LigI in a final volume of 250 μL at 30 °C. The kinetic constants were calculated from a fit of the initial velocity data to eq 2.1 using the nonlinear least-squares fitting program in SigmaPlot 9.0

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

where v is the initial velocity, Et is the total enzyme concentration, k_{cat} is the turnover number, [A] is the substrate concentration, and K_m is the Michaelis constant. The apparent inhibition constants were determined from a fit of the data to eq 2.2.

$$v/E_t = (k_{cat}[A])/[Ka(1/Ki))+[A]] \quad (2.2)$$

Measurement of the Equilibrium Constant. The equilibrium constant for the reaction catalyzed by LigI was measured over the pH range of 6.0–9.5 at 0.25 pH unit intervals using MES (pH 6.0–6.75), HEPES (pH 7.0–8.0), BICINE (pH 8.2–9.0), and CHES (pH 9.0–9.5) to control the pH. In these measurements, LigI was incubated with 250 μM PDC in a final volume of 250 μL until the reaction reached equilibrium. The fractional concentrations of OMA and PDC at each pH were determined from the absorbance at 312 nm using a differential extinction coefficient of 6248 M–1 cm–1. The equilibrium constant was calculated from a fit of the data to eq 2.3, where y equals the mole fraction of PDC at equilibrium.

$$y = 1 + [H^+]/K_{eq} \quad (2.3)$$

pH–Rate Profiles. The pH dependence of the kinetic constants, k_{cat} and k_{cat}/K_m, was measured over the pH range of 6.5–9.75 at 0.25 pH unit intervals. The buffers used for the pH–rate profiles were MES (pH 6.0–6.75), HEPES (pH 7.0–8.0), BICINE (pH 8.25–9.0), and CHES (pH 9.0–9.75). The final pH values were recorded after the completion of

the assay. The profiles for the pH dependence of k_{cat} and k_{cat}/K_m for lactone hydrolysis and synthesis were fit to eqs 4 and 5, respectively

$$\log y = \log[\, c/(1+[H^+]/K_a)] \tag{2.4}$$

$$\log y = \log[\, c/(1+ K_b/[H^+])] \tag{2.5}$$

where c is the maximal value for either k_{cat} or k_{cat}/K_m and Ka and Kb are the ionization constants for the groups that must be unprotonated and protonated for catalytic activity, respectively.

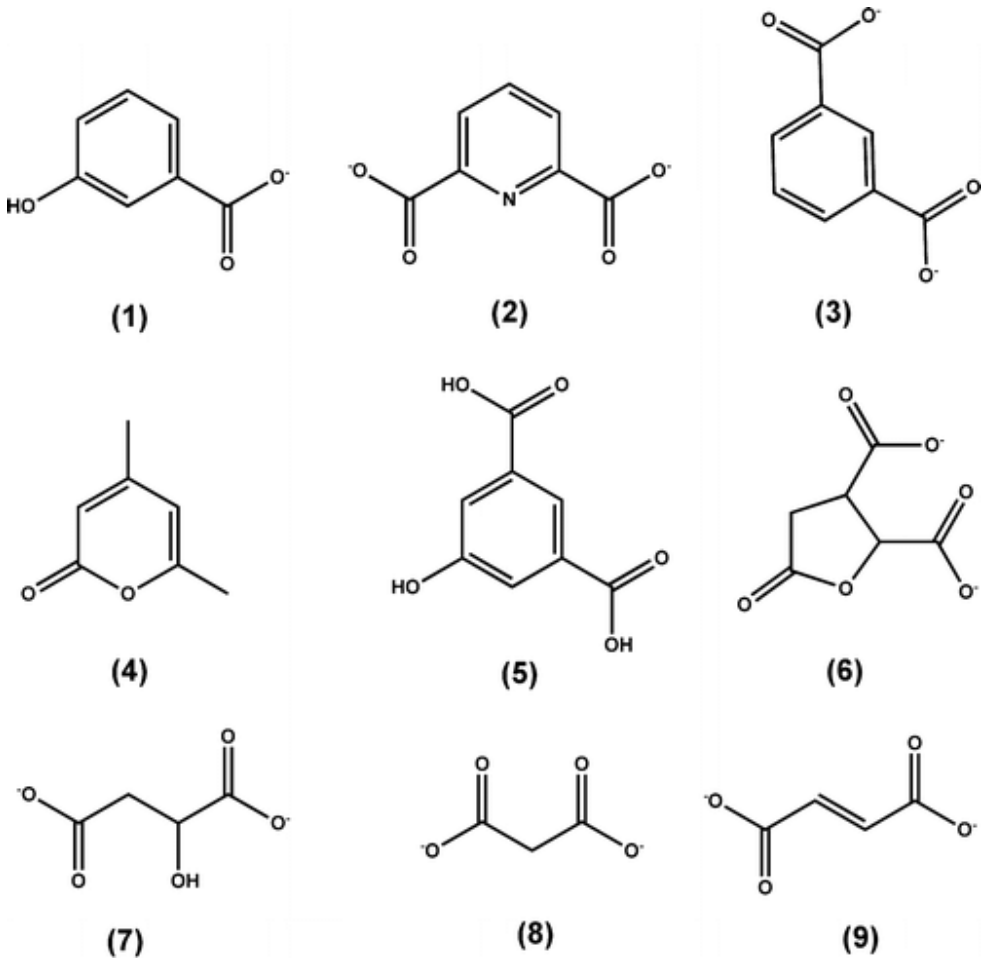
Metal Analysis. The metal content of LigI, substrates, and buffers was determined with an Elan DRC II ICP-MS instrument as previously described. [71] Protein samples for ICP-MS analysis were digested with HNO3 and then refluxed for 30 min[58]. All buffers were passed through a column of Chelex 100 (Bio-Rad) to remove trace metal contamination. EDTA, 1,1-phenanthroline, dipicolinic acid, or 2,2'-bipyridine (1.0 mM) was incubated with 1.0 μM LigI in 50 mM buffer at pH values ranging from 6 to 10 to remove divalent metal ions from LigI. The buffers for these experiments included CHES (pH 6.0), HEPES (pH 7.0), BICINE (pH 8.0), and CHES (pH 9.0 and 10.0). The effect of added divalent cations (as the chloride salts) on the catalytic activity of LigI was determined by adding Mn2+, Zn2+, Co2+, Cu2+, or Ni2+ (0–500 μM) directly to the assay mixtures. LigI (1.0 μM) was also incubated with 50–500 molar equiv of these divalent cations for 24 h at 4 °C in 50 mM HEPES (pH 7.5) and then assayed for catalytic activity.

NMR Analysis of OMA and CHM. The tautomeric distribution of OMA and CHM in solution was determined as a function of pH using heteronuclear multiple-bond correlation (HMBC) NMR spectroscopy with a Bruker Avance III 500 MHz spectrometer with an H-C-N cryoprobe using WATERGATE solvent suppression.

Hydrolysis of PDC in $H_2{}^{18}O$. The site of the hydrolysis of PDC by LigI was determined using NMR spectroscopy in 18O-labeled water. The reaction was conducted in 100 mM carbonate (pH 9.9) with 10 mM PDC using 10 µM LigI in a final volume of 1.0 mL for 1 h in 50% H218O. At the conclusion of the reaction, LigI was removed from the reaction mixture using a 3 kDa cutoff VWR centrifugal filter. D2O was added, and the ^{13}C NMR spectrum of the product was obtained.

Identification of Inhibitors. Selected compounds (**Scheme 2.2**) were evaluated as potential inhibitors of LigI on the basis of structural similarities to PDC, OMA, or CHM. The inhibition assays contained 100 µM PDC, varying concentrations of inhibitor (0–500 µM), 50 mM BICINE (pH 8.25), and LigI in a final volume of 250 µL at 30 °C.

Scheme 2.2: Potential Inhibitors of LigI



Sequence Similarity Networks. All protein sequences available through NCBI that are designated as belonging to cog3618 were obtained by a simple search using the query "cog3618". Protein sequences were then subjected to an all-by-all BLAST using the NCBI standalone BLAST program at specific stringency E values (10^{-10}, 10^{-20}, 10^{-30}, 10^{-40}, 10^{-50}, 10^{-60}, and 10^{-70}). At a stringency value of 10^{-70}, many of the groups clustered into smaller (and perhaps isofunctional) groups that were given an arbitrary identifying number.

Results

Purification and Properties of LigI from S. paucimobilis. 2-Pyrone-4,6-dicarboxylic acid lactonase from *S. paucimobilis* was cloned, expressed in E. coli, and purified to homogeneity. The purified LigI did not contain significant amounts of zinc, manganese, iron, copper, cobalt, or nickel (<0.04 equiv per monomer) when analyzed by ICP-MS. The addition of chelating agents directly to the assay solutions did not affect the rate of PDC hydrolysis. The purified LigI was further incubated with EDTA, 1,10-phenanthroline, dipicolinic acid, and 2,2'-bipyridine at pH values from 6 to 10 in an attempt to remove tightly bound metals, but the addition of these potent chelators did not change the catalytic activity of LigI. The addition of Mn^{2+}, Zn^{2+}, Co^{2+}, Cu^{2+}, or Ni^{2+} directly to the assay solutions had no effect on the catalytic activity of LigI. Therefore, LigI does not appear to require a divalent cation for enzymatic activity.

The change in absorbance when PDC is converted to OMA at pH 10.0 is presented in **Figure 2.2**. There is an absorption maximum for PDC at 312 nm that is lost upon conversion to OMA. The kinetic constants for the hydrolysis of PDC to OMA and CHM were determined at pH 8.25 to be 342 ± 25 s–1, 48 ± 11 μM, and (7.5 ± 1.0) × 10^6 M^{-1} s^{-1} (k_{cat}, K_m, and k_{cat}/K_m, respectively). The kinetic constants in the reverse reaction for the synthesis of PDC from OMA and CHM at pH 8.25 were 116 ± 6 s^{-1}, 18 ± 4 μM, and (4.9 ± 0.7) × 10^6 M^{-1} s^{-1} (k_{cat}, K_m, and k_{cat}/K_m, respectively).

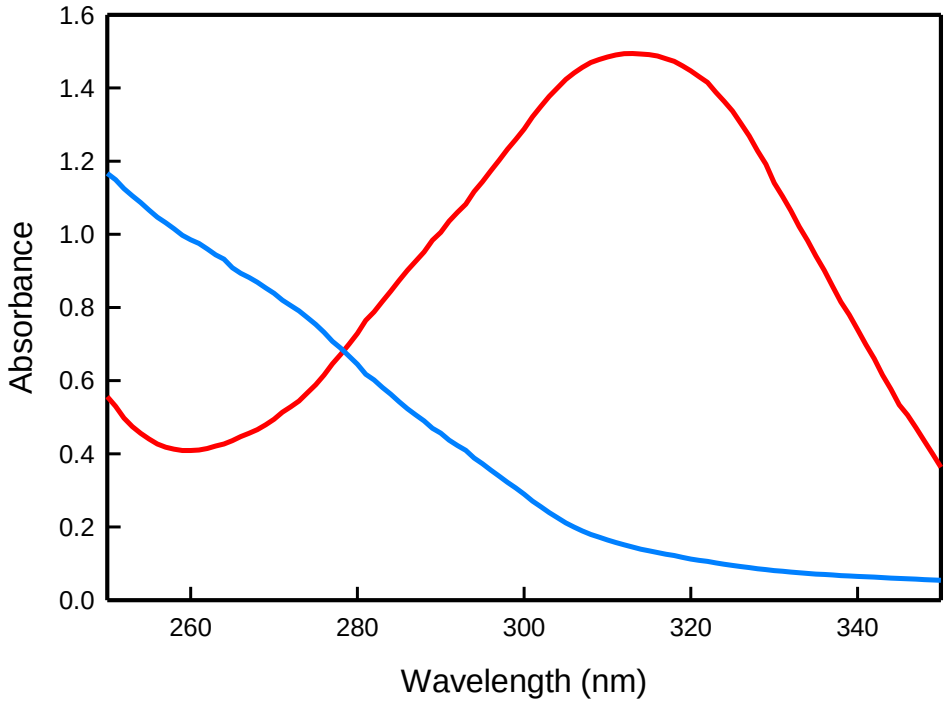


Figure 2.2.UV–visible absorbance spectrum of 0.3 mM PDC (red) and the hydrolysis product after the addition of LigI (blue) at pH 10. The maximal absorbance change is at 312 nm with a differential extinction coefficient, Δε, of 6248 M^{-1} cm^{-1}.

NMR Spectra for OMA and CHM. The product of the enzymatic reaction existed in two forms depending on the pH. At pH 10, the keto form (OMA) was predominant as shown by the presence of singlet peaks at δ 6.58 (1 H, H-2) and δ 3.92 (2 H, H2-4) in the proton NMR spectrum. The carbon spectrum showed peaks for carbonyl carbons at δ 202.0 (C-5), 175.1 (C-1), 175.0 (C-6), and 168.6 (C-9), vinyl carbons at δ 136.9 and 132.8, and an aliphatic carbon at δ 39.8 (C-4). A DEPT experiment indicated that the carbon at δ 132.8 had a single proton attached (C-2), while the carbon producing the peak at δ 39.8 had two attached protons (C-4). The vinyl carbon peak at δ 136.9 is therefore due to C-3. An HMBC experiment showed long-range couplings from the proton peak at δ 3.92 to the two vinyl carbons and carbonyl carbons at δ 202.0 and 175.0. The proton peak at δ 6.58 showed cross-peaks to the carbonyl carbon at δ 175.0, the vinyl carbon at δ 136.9, and the aliphatic carbon at δ 39.8. These data indicated that the peak at δ 175.0 correlating to both proton peaks was C-9. Because of chemical shift considerations, the peak at δ 168.6 is due to C-6, indicating that the peak at δ 175.1 comes from C-1.

At pH 6, the enol form (CHM) was predominant. Its proton spectrum consisted of two doublets at δ 7.19 (H-2) and δ 6.70 (H-4) with a coupling constant of 1.3 Hz. The carbon spectrum showed carbonyl carbon peaks at δ 170.4, 165.5, and 165.4 and vinyl peaks at δ 153.9, 152.7, 116.4, and 108.1. A DEPT experiment indicated the carbons at δ 116.4 and 108.1 each had a single attached proton. An HSQC experiment showed that the proton at δ 7.19 was attached to the carbon at δ 116.4 (C-2), and the proton at δ

6.70 was attached to the carbon at δ 108.1 (C-4). The HMBC experiment indicated the proton at δ 7.19 had long-range couplings to the carbonyl carbons at δ 170.4 (C-1) and δ 165.4 (C-6) and a vinyl carbon at δ 153.9 (C-3). The peak at δ 6.70 correlated to the carbonyl at δ 165.5 (C-6) and a vinyl carbon at δ 152.7 (C-5). The relative amounts of the two forms at given pH values as estimated by protonated vinyl carbon peak heights were 100% OMA and 0% CHM (pH 10), 90% OMA and 10% CHM (pH 9), 55% OMA and 45% CHM (pH 8), 5% OMA and 95% CHM (pH 7), and 2% OMA and 98% CHM (pH 6).

Hydrolysis of PDC in $H_2{}^{18}O$. The regiochemistry for the attack of water on PDC has not previously been determined. Water could attack C-2 with cleavage of the C-2–O-1 bond. Alternatively, water could attack C-6 with cleavage of the C-6–O-1 bond. The enzymatic reaction was conducted in 50% $H_2{}^{18}O$ to determine the electrophilic center of PDC that reacts with water. If water attacks C-2, then the C-1 carboxylate group of OMA will be labeled with 18O, whereas if the attack occurs C-6, then the C-5 carbonyl group of OMA will be labeled. Shown in **Figure 2.3A** is the ^{13}C NMR spectrum of the C-1 carboxylate of the product when the enzymatic reaction was conducted in unlabeled water at pH 9.9. The C-1 carboxylate of OMA appears at δ 175.04 and is a single resonance that is adjacent to the resonance for the C-9 carboxylate at δ 174.96. When the reaction was conducted in $H_2{}^{18}O$, the C-1 carboxylate of OMA appears as two distinct resonances with a separation of 0.03 ppm, reflecting the small chemical shift difference for OMA with a single ^{18}O label (**Figure 2.3C**). There was no change in the resonance for the C-1 carboxylate when $H_2{}^{18}O$ was added to OMA after removal of the

enzyme (**Figure 2.3B**). Therefore, under the conditions of these experiments, the attack by water/hydroxide occurs directly at C-2 of PDC.

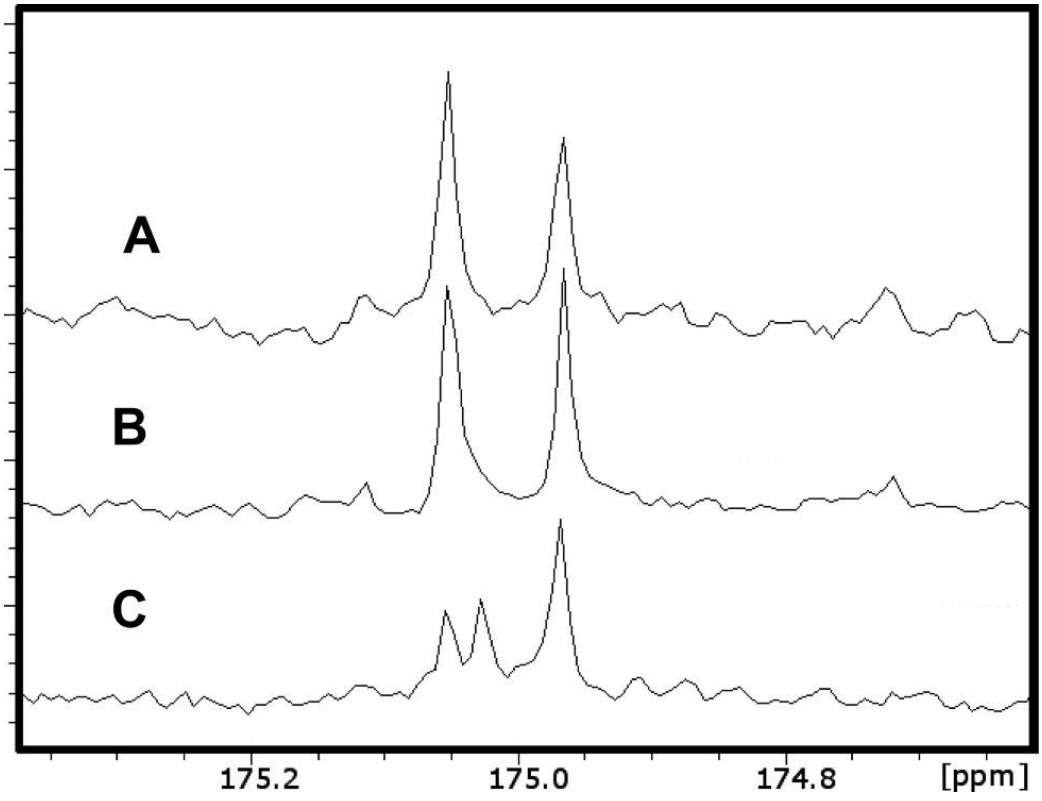


Figure2.3. (A) NMR spectrum of OMA in water at pH 9.0. (B) NMR spectrum of OMA incubated in 50% $H_2{}^{18}O$ at pH 9.0. (C) NMR spectrum of OMA produced enzymatically from PDC in $H_2{}^{18}O$ at pH 9.0.

pH-Independent Equilibrium Constant. The reaction catalyzed by LigI is reversible, and the equilibrium concentrations of PDC and OMA/CHM are dependent on the solution pH. The equilibrium constant for this reaction was determined by allowing the hydrolysis reaction to reach completion at pH values ranging from 6 to 10. The concentration of PDC was determined from the change in absorbance at 312 nm, and the results are presented in **Figure 2.4**. The equilibrium constant, from a fit of the data to eq 2, was determined to be 5.7×10^{-9} M^{-1}. At pH 8.25, the relative concentrations of PDC and the hydrolysis products are thus equal to one another. The Haldane relationship for this reaction at this pH value is presented in eq 2.6. Substitution of the k_{cat}/K_m values for the substrate and product determined at pH 8.25 gives an apparent equilibrium constant of 1.5, which is reasonably close to the experimentally determined value of 1.0.

$$K_{eq} = [OMA_{eq}]/[PDC_{eq}] = (k_{cat}/K_{PDC})/(k_{cat}/K_{OMA}) \quad (2.6)$$

Identification of Inhibitors. Nine compounds (see **Scheme 2.2**) were tested as potential inhibitors of LigI as a probe of the structural requirements for ligand binding. Of the compounds tested, only pyridine-2,4-dicarboxylic acid (2) and 5-hydroxyisophthalic acid (5) inhibited the hydrolysis of PDC. The apparent inhibition constants for pyridine-2,4-dicarboxylic acid and 5-hydroxyisophthalic acid are 75 ± 2 and 40 ± 3 μM, respectively, from fits of the data to eq 2.

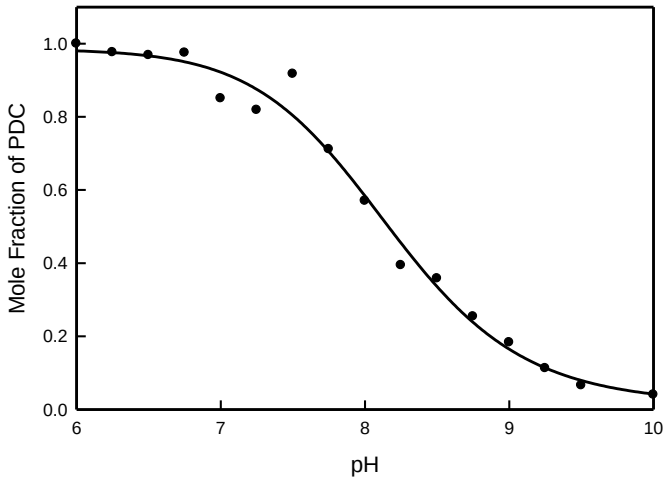


Figure 2.4. Effect of pH on the equilibrium concentrations of PDC and OMA/CHM. The line represents a fit of the data to eq 2.3. The equilibrium constant for the reaction illustrated in Scheme 1 is 5.7×10^{-9} M^{-1}.

pH–Rate Profiles. Kinetic constants for the hydrolysis and synthesis of PDC by LigI were measured as a function of pH to identify potential active site residues whose involvement in catalysis is dependent on the state of protonation. The profiles for the forward and reverse reactions are presented in **Figure 2.5**. The kinetic pKa values for the hydrolysis of PDC are 7.5 ± 0.1 for k_{cat} and k_{cat}/K_m. The kinetic pKa values for the condensation of OMA are 8.6 ± 0.1 and 7.5 ± 0.1 for k_{cat} and k_{cat}/K_m, respectively.

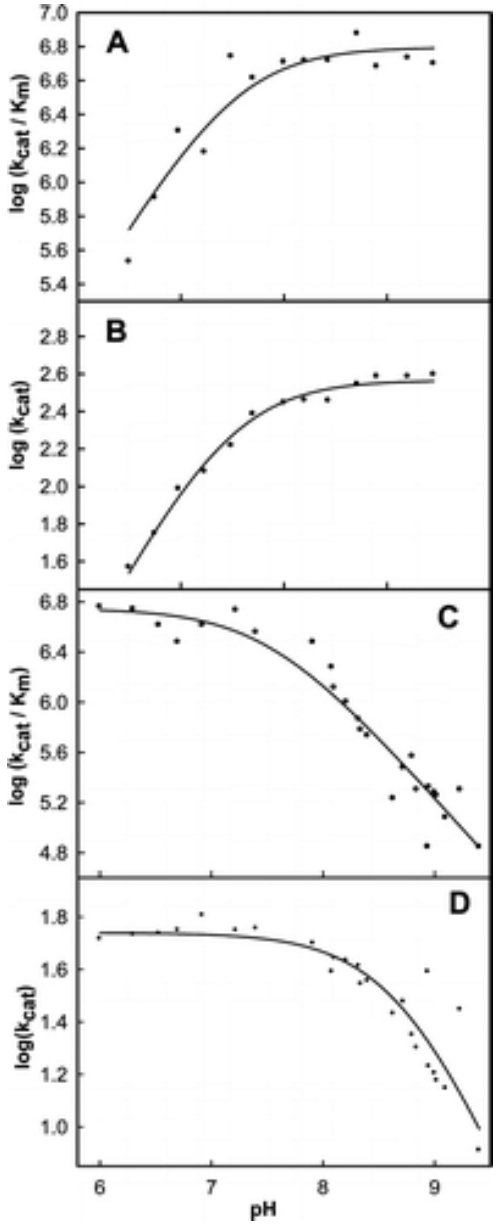


Figure 2.5. pH–rate profiles for the enzymatic hydrolysis and synthesis of PDC. (A) pH–rate profile for log(k_{cat}/K_m) for the hydrolysis of PDC. (B) pH–rate profile for log(k_{cat}) for the hydrolysis of PDC. The lines for panels A and B represent the fits of the data to eq 2.4. (C) pH–rate profile for log(k_{cat}/K_m) for the synthesis of PDC. (D) pH–rate profile for log(k_{cat}) for the synthesis of PDC. The lines for panels C and D represent the fit of the data to eq 2.5.

Structure of WilD-Type LigI. The crystal structure of native LigI was determined by single anomalous dispersion (SAD) at 2.0 Å resolution. The model contains protein residues Leu-3–Glu-296 and 225 water molecules, most of which are welL-defined in

the electron density map. The protein has a distorted TIM-barrel fold, and the wide opening on one end of the barrel suggests the entrance to the active site (**Figure 2.6A**). The closest structural homologue in the Protein Data Bank (PDB) is a putative 2-pyrone-4,6-dicarboxylate lactonase from Pseudomonas putida [PDB entry 2FFI, root-mean-square deviation (rmsd) of 2.8 Å] with a level of sequence identity of 31%.

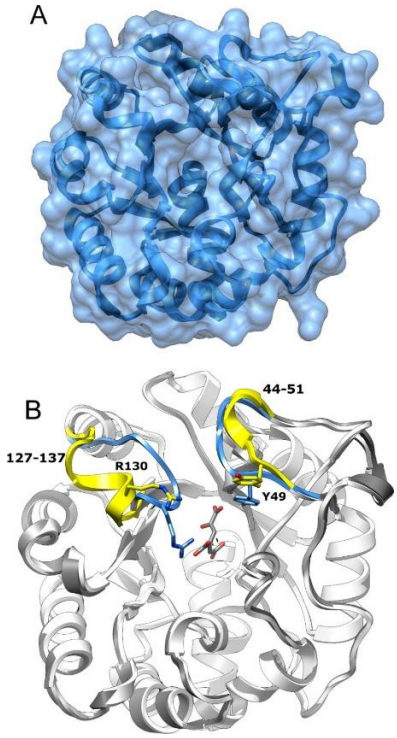


Figure 2.6. Structure of native LigI and the D248A mutant. (A) Native LigI. The narrow crevice on top of the solvent accessible surface corresponds to the active site entrance. (B) View of CHM-bound D248A LigI (white) superimposed onto the wiLD-type apo form of LigI (gray). The two loops, Phe-127–Lys-137 and Ser-44–Pro-51 that exhibit a change in conformation upon substrate binding are color-coded as follows: yellow for native LigI (apo) and blue for CHM-bound D248A.

Structure of the D248A Mutant. To reveal the structural details of the interactions of LigI with substrate, the inactive mutant D248A was produced and cocrystallized with the substrate. In both structures, the ligand is well-defined in the corresponding electron density maps. Substrate binding results in closure of the active site and large-scale rearrangement of two loops: Phe-127–Lys-137 and, to a lesser extent, Ser-44–Pro-51 (**Figure 2.6B**). As a result of this structural rearrangement, the first loop transforms from an α-helical conformation into a coiled conformation, and the guanidinium group of Arg-130 moves by 6.1 Å to directly contact the substrate C-8 carboxylic group and to close the entrance of the active site. The phenolic side chain of Tyr-49 moves 14.4 Å to be in direct contact with the C-11 carboxylate of PDC.

At pH 8.5, an almost equal mixture of the substrate and product was observed in the active site. The presence of the product can be explained by residual catalytic activity or spontaneous hydrolysis of the substrate in solution. Interestingly, at pH 6.5, only CHM is present in the active site. The electron densities for the bound substrate/product ligands in the active site of LigI at pH 8.5 and 6.5 are shown in panels A and B of **Figure 2.7**, respectively. PDC forms multiple interactions in the active site: the C-4 carboxylate forms four hydrogen bonds with Arg-130, Asn-253, and Tyr-156; the C-6 carboxylate forms four hydrogen bonds with Arg-124, Tyr-49, and Ser-77; and the C-2 carbonyl group forms four hydrogen bonds with His-31, His-180, and a water molecule. CHM occupies a position similar to that of the substrate and makes multiple interactions: the C-9 carboxylate forms four hydrogen bonds with Arg-130, Asn-253,

and Tyr-156; the C-6 carboxylate forms four hydrogen bonds with Arg-124, Tyr-49, and Ser-77; and the remaining carboxylate group (C-1) forms four hydrogen bonds with His-31, His-180, His-33, and a water molecule. The C-5 hydroxyl group of the product forms two hydrogen bonds with Arg-124. Nearby, Arg-217 forms hydrogen bonds with acetate and another water molecule (**Figure 2.8**).

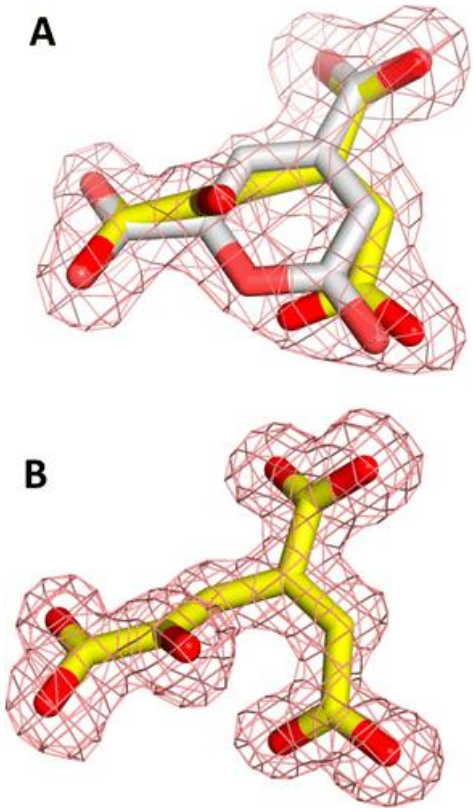


Figure 2.7. Refined electron density ($2F_o - F_c$, contoured at 1σ) of the ligand in the active site of LigI. (A) Equimolar amounts of substrate (PDC, white) and product (CHM, yellow) at pH 8.5. (B) Product only (CHM, yellow) at pH 6.5.

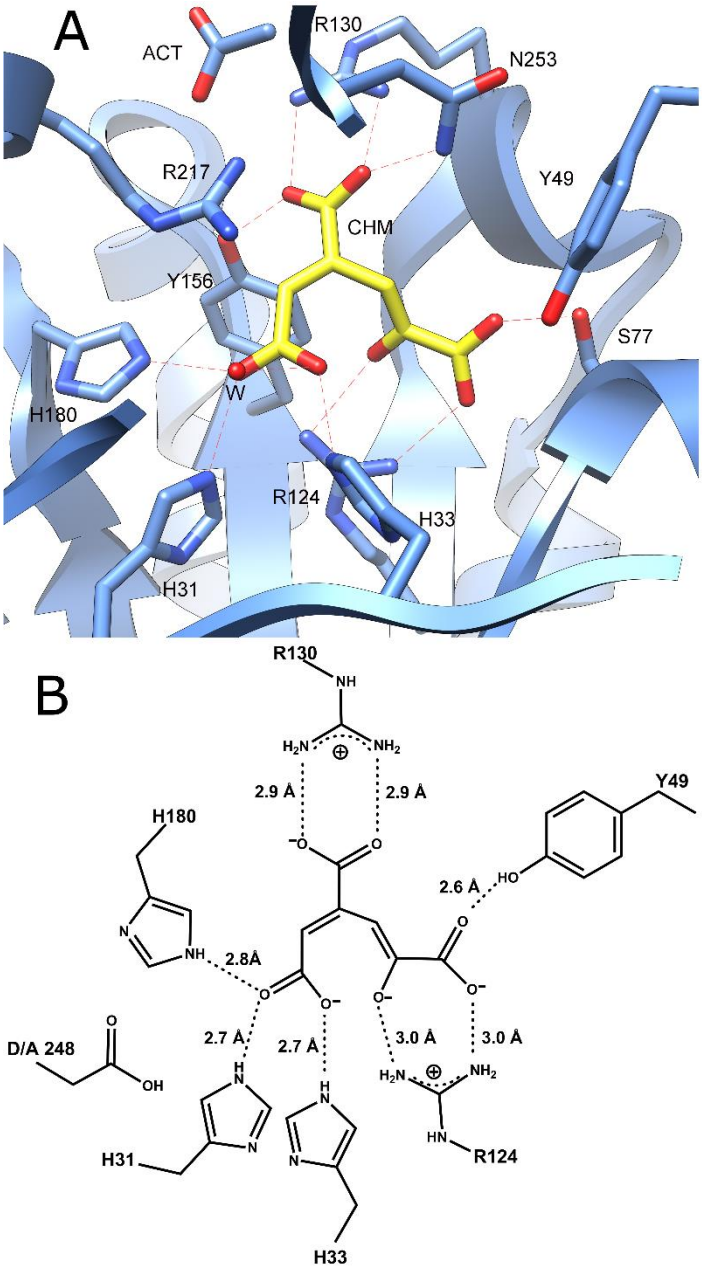


Figure 2.8. (A) CHM bound in the active site of the D248A mutant of LigI at pH 6.5. No PDC was detected at this pH. (B) Active site structure of D248A with CHM in the active site.

Structure of the D248N Mutant. The structure of the D248N mutant was determined from crystals grown at pH 4.6 in the presence of the purified product PDC. The refined electron density did not show the presence of either substrate or product in the active site. This can be explained by the reduced affinity of the active site for substrate/product under very acidic conditions. The structure of the D248N mutant is very similar to the structure of ligand-free wild-type LigI (rmsd of 0.42 Å), and the conformations of Asn-248 in the mutant enzyme and Asp-248 in the wild-type enzyme are identical.

Mutagenesis of LigI. Residues in LigI that align with the prototypical metal-coordinating residues in the AHS were mutated. These residues included His-31 and His-33 from the C-terminal end of β-strand 1, His-180 from the end of β-strand 6, and Asp-248 from the end of β-strand 8. Additional active site residues that were mutated included Tyr-156, Arg-124, Tyr-49, Arg-130, Arg-217, and Arg-183. Tyr-156 from the end of β-strand 5 replaces the conserved histidine residue that coordinates a divalent cation in those members of the AHS with either a mononuclear or binuclear metal center. Arg-124 from the end of β-strand 4 is positioned to occupy the space filled by the residue that bridges the two divalent cations in those members of the AHS with a binuclear metal center. Tyr-49, Arg-130, and Arg-183 are in the active site and interact with the substrate/product. The kinetic constants for these mutants are listed in **Table 2.1**.

Table 2.1: Kinetic Constants for LigI Variants

	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
wilD-type	342 ± 25	48 ± 11	$(7.1 \pm 1) \times 10^6$
D248A	0.0013 ± 0.0001	116 ± 18	$(1.6 \pm 0.2) \times 10^1$
D248N	0.023 ± 0.001	66 ± 13	$(2.0 \pm 0.34) \times 10^2$
R124M	0.020 ± 0.001	90 ± 15	$(2.0 \pm 0.22) \times 10^2$
R130M	5.0 ± 0.4	465 ± 104	$(1.1 \pm 0.7) \times 10^4$
H31N	0.40 ± 0.04	720 ± 122	$(5.0 \pm 0.37) \times 10^2$
R217M	0.42 ± 0.02	67 ± 10	$(6.3 \pm 0.4) \times 10^3$
H180A	8.0 ± 0.9	214 ± 44	$(3.7 \pm 0.4) \times 10^4$
H180C	32 ± 4	41 ± 7	$(7.8 \pm 0.5) \times 10^5$
H33N	3.6 ± 0.3	89 ± 20	$(4.1 \pm 0.6) \times 10^4$
Y49F	26 ± 2	217 ± 41	$(1.2 \pm 0.02) \times 10^5$
Y156F	8.7 ± 0.3	238 ± 19	$(3.7 \pm 0.02) \times 10^4$
R183M	196 ± 6	64 ± 6	$(3.0 \pm 0.02) \times 10^6$

Discussion

Potential Role of Divalent Metal Ions. Although all well-characterized amidohydrolases require one to three divalent metal ions for catalytic activity, LigI and other members of cog3618 exhibit active site differences with respect to other amidohydrolase superfamily members that argue they are not metalloenzymes. We were unable to detect bound metal ions via ICP-MS in the samples of recombinant LigI prepared for this investigation. In addition, metal ion chelators had no effect on the catalytic activity of LigI, and the addition of divalent cations to enzymatic assays had no effect on reaction rates. Therefore, we conclude that LigI from S. paucimobilis does not require a divalent cation for catalytic activity and does not bind divalent metals to any significant extent. To the best of our knowledge, this work provides the first documented example of an enzyme from the AHS that does not require a divalent cation to bind in the active site for catalytic activity. It will be of mechanistic interest to determine whether other proteins of unknown function within cog3618 require divalent cations for catalytic activity.

Active Site of LigI. Nearly all members of the AHS have a constellation of five amino acid residues that are conserved within the active site and required for metal ion binding and/or proton transfers.[8] These residues include two histidine residues from the end of β-strand 1, two additional histidine residues from the ends of β-strands 5 and 6, and a conserved aspartate residue from the end of β-strand 8. In LigI, the three

histidine residues (H31, H33, and H180) from the ends of β-strands 1 and 6 are conserved in addition to the aspartate (D248) from the end of β-strand 8. The histidine from the end of β-strand 5 is absent. Because LigI does not require the binding of divalent cations for the expression of catalytic activity, the functional role of these conserved residues must be different than they are for the rest of the amidohydrolase superfamily.

A structural alignment of the active site of LigI and phosphotriesterase (PTE), an enzyme with a prototypical binuclear metal active site, is presented in **Figure 2.9A**. The HxH motif from β-strand 1, the histidine from β-strand 6, and the aspartate from β-strand 8 of LigI align well with the corresponding residues of PTE. In addition, Tyr-156 from β-strand 5 of LigI aligns with the conserved histidine from β-strand 5 of PTE and Arg-124 from β-strand 4 of LigI aligns with the carboxylated lysine from β-strand 4 of PTE. In PTE, the carboxylated lysine serves to bridge the two divalent cations. The active site of LigI is overlaid with the active site of cytosine deaminase (CDA), a mononuclear metal active site in **Figure 2.9B**. The HxH motif from β-strand 1, the histidine from β-strand 6, and the aspartate from β-strand 8 of LigI align reasonably well with the corresponding residues of CDA. However, Tyr-156 from LigI does not align with the conserved histidine from β-strand 5 of CDA. The inability of LigI to bind a divalent metal appears to stem from the presence of the positively charged Arg-124 in the active site.

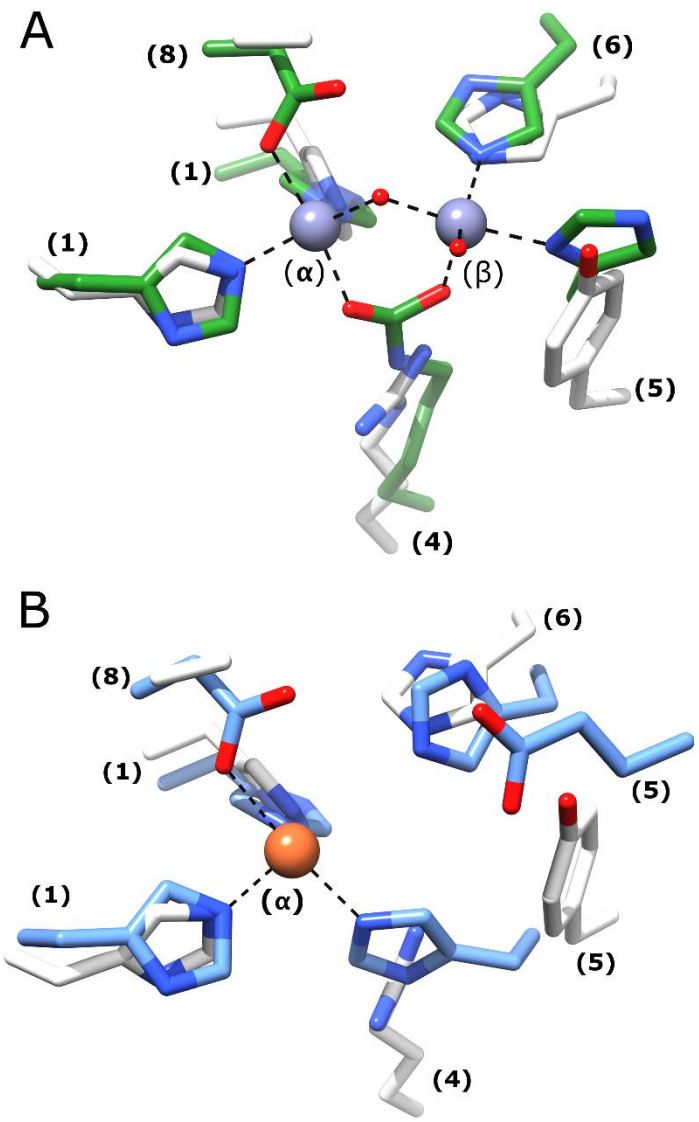
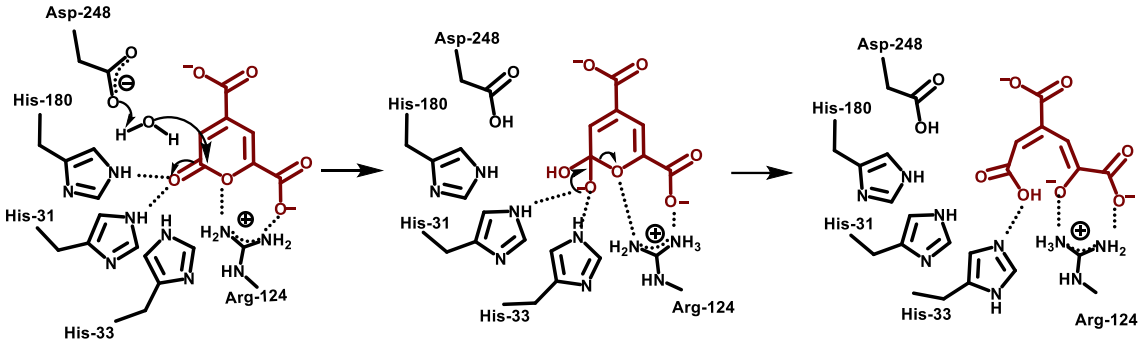


Figure 2.9. (A) Active site of D248A LigI structurally aligned with the binuclear metal center of phosphotriesterase (PTE). The active site is color-coded as follows: green for PTE (PDB entry 1hzy) and white for the D248A LigI active site. (B) Active site of D248A LigI structurally aligned with the mononuclear metal center found in the active site of cytosine deaminase. The active sites are color-coded as follows: blue for CDA (PDB entry 1k70) and white for the D248A LigI active site. Numbers in parentheses indicate the β-strand origin of each residue.

Role of Conserved Residues and Mechanism of Action. A working model for the hydrolysis of PDC by LigI is presented in **Scheme 2.3**. In this mechanism, some of the conserved AHS active site residues, which normally participate in metal coordination, have different catalytic roles in the function of this protein. The four residues in question are His-31, His-33, His-180, and Asp-248. The binding of PDC is facilitated by ion pair interactions between residues Arg-130 and Arg-124 and the C-8 and C-11 carboxylates, respectively. There is an additional electrostatic interaction between Tyr-49 and the C-11 carboxylate. The lactone carbonyl group of the substrate is polarized by electrostatic interactions with the three conserved histidine residues: His-31, His-33, and His-180. Asp-248 is positioned in the active site to deprotonate an active site water molecule for nucleophilic attack on the C-2 carbonyl group of the substrate. The orientation of Asp-248 allows for the attack on the re face of PDC, which is consistent with the relative orientation of attack determined previously for other members of the amidohydrolase superfamily.[50] After formation of the tetrahedral intermediate, the C-2–O-1 bond is cleaved with support of the electrostatic interaction of the guanidino group of Arg-124.

Scheme 2.3: Proposed Reaction Mechanism of the LigI Catalyzed Reaction



Further support for the proposed mechanism of action comes from our structure with substrate/product ligands bound to the enzyme and the catalytic properties of selected LigI mutants. The importance of D248 is confirmed by the nearly complete absence of activity for the D248A and D248N mutants. These mutants are reduced in activity by 4–5 orders of magnitude (for either k_{cat} or k_{cat}/K_m). Of the three conserved histidine residues in the active site, the mutation of His-31 results in the greatest loss of catalytic power. Mutation of this residue to asparagine reduces k_{cat}/K_m by more than 4 orders of magnitude. Mutation of either His-33 or His-180 reduces this kinetic constant by approximately 2 orders of magnitude. The mutation of Tyr-49 to phenylalanine reduces k_{cat}/K_m 60-fold, and mutation of Arg-124 reduces k_{cat}/K_m by a factor of 30000. The pH–rate profiles are consistent with an active site residue that must be deprotonated for catalytic activity. This residue is most likely assigned to either Asp-248 or His-31. Tyr-156 is 2.5 Å from the C-8 carboxylate oxygen and 3.6 Å from C-2.

CHAPTER III

DISCOVERY OF AN L-FUCONO-1,5-LACTONASE FROM COG3618 OF THE AMIDOHYDROLASE SUPERFAMILY*

L-Fucose is a widely distributed monosaccharide throughout the biosphere having functions in both plant and mammalian glycoconjugates.[72, 73] These glycoconjugates include the ABO blood group antigens in mammals, L-fucose binding lectins, glycolipids, and L-fucopyranosyl moieties in cell walls.[72] L-Fucose can function as the sole carbon source for many microbes and has been shown to be metabolized via two separate pathways as illustrated in **Figure 3.1**.[74-76] In the first pathway, L-fucose is isomerized to L-fuculose and then phosphorylated to L-fuculose 1-phosphate. It is subsequently cleaved to L-lactaldehyde and dihydroxyacetone phosphate.[75] In an alternative pathway, L-fucose is oxidized to L-fuconolactone and then hydrolyzed to L-fuconate. This product is subsequently dehydrated to 2-keto-3-deoxy-L-fuconate and then oxidized to 2,4-diketo-3-deoxy-L-fuconate. This compound is then hydrolyzed to pyruvate and L-lactate. [73] In the latter pathway, the enzymes that hydrolyze L-fuconolactone and 2,4-diketo-3-deoxy-L-fuconate have not been characterized.

---

*Reprinted with permission from "Discovery of an L-Fucono-1,5-lactonase from cog3618 of the Amidohydrolase Superfamily " by Merlin Eric Hobbs, Matthew Vetting, Howard J. Williams, Tamari Narindoshvili, Devon M. Kebodeaux, Brandan Hillerich, Ronald D. Seidel, Steven C. Almo, and Frank M. Raushel, *Biochemistry*, 2012, *51* (16), pp 3497–3507, Copyright 2012 American Chemical Society

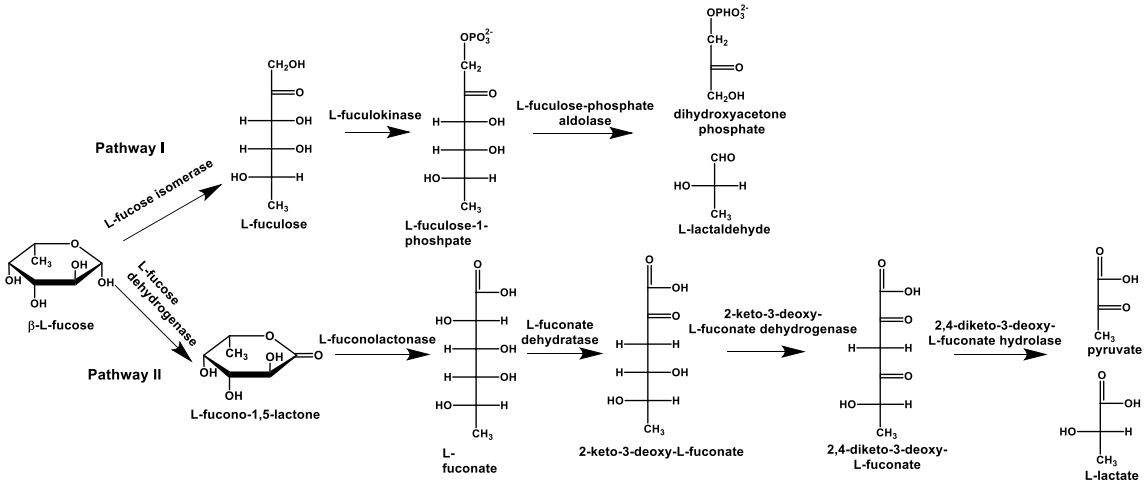


Figure 3.1. Metabolism of L-fucose. Pathway I produces dihydroxyacetone phosphate and L-lactaldehyde. Pathway II produces pyruvate and L-lactate through the oxidation of L-fucose to L-fucono-1,5-lactone.

We have determined the three-dimensional structure and functionally annotated BmulJ_04915 from *Burkholderia multivorans* ATCC 17616. Two homologues of this enzyme, Bamb_1224 from *Burkholderia ambifaria* AMMD and Patl_0798 from *Pseudoalteromonas atlantica* T6c, were also functionally annotated. These enzymes are members of the amidohydrolase superfamily (AHS) of enzymes from cog3618. Enzymes from this superfamily catalyze a diverse set of reactions, including the hydrolysis of amide or ester bonds, deamination, decarboxylation, isomerization, and hydration.[77] The structural hallmark for members of this superfamily is a distorted $(\beta/\alpha)_8$ barrel, which binds zero to three divalent metal ions within the active site. The AHS has been separated into 24 clusters of orthologous groups (COG) by NCBI.[14, 15]

The functionally characterized enzymes from cog3618 include 2-pyrone-4,6-dicarboxylate lactonase (LigI), 4-sulfomuconolactonase (4-SML), and L-rhamnono-1,4-lactonase.[65, 66, 78] A sequence similarity network for cog3618 at an *E* value cutoff of 10^{-30} is presented in **Figure 3.2A.**[68] In this figure, each node in the network represents a protein and is connected to other proteins by edges when the BLAST *E* values are smaller than 10^{-30}. At this stringency, the proteins within cog3618 split into two large groups that are arbitrarily designated here as Class I and Class II. When a more stringent*E* value of 10^{-70} is applied to the sequence similarity network, the proteins organize into smaller groups (**Figure 3.2B**), which, because of their higher level of sequence similarity, may be isofunctional. In **Figure 3.2**, the groups are assigned an arbitrary numerical identifier (i.e., 1, 2, 3, etc.). L-Rhamnono-1,4-lactonase and BmulJ_04915 map to Class I, while LigI and 4-SML map to Class II. Genomic context analysis of the enzymes within Class I identifies many proteins of unknown function that are physically located near other genes that have been previously annotated as ABC-type carbohydrate transport systems, dehydrogenases, dehydratases, and hydrolases. Therefore, many of the proteins of unknown function within Class I are likely involved in carbohydrate metabolism and are thus predicted to hydrolyze sugar lactones.

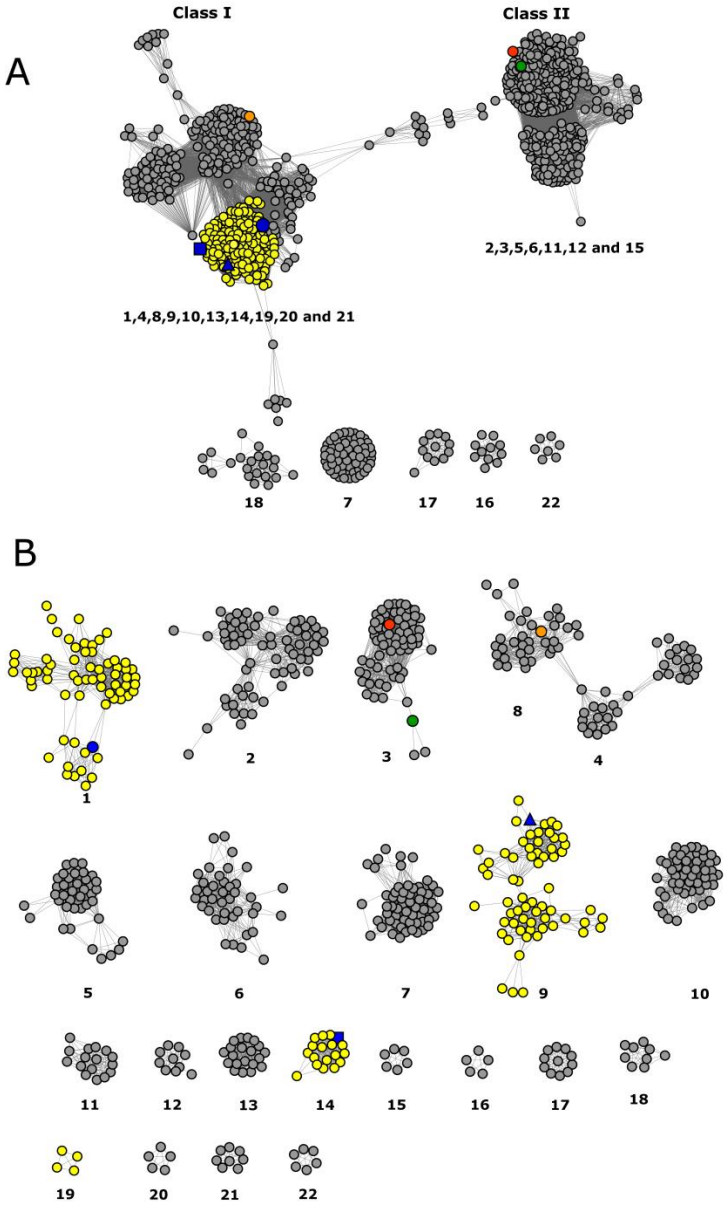


Figure 3.2. (A) cog3618 sequence similarity networks at an *E* value of 10^{-30} where each node represents a protein and an edge represents an *E* value between two proteins of $\leq 10^{-30}$. (B) cog3618 sequence similarity networks at an *E* value of 10^{-70} where each node represents a protein and an edge represents an *E* value between two proteins of $\leq 10^{-70}$. The nodes are color-coded as follows: yellow nodes for predicted L-fucono-1,5-lactonase, blue square for BmulJ_04915, blue circle for Bamb_1224, blue triangle for Patl_0798, orange node for L-rhamnono-1,4-lactonase, red node for LigI, and green node for Protein Data Bank entry 4SML

BmulJ_04915 is predicted to hydrolyze sugar lactones based on genomic context (**Figure 3.3**). The protein, BmulJ_04922, from cog1028 is 64% similar in amino acid sequence to 2-keto-3-deoxy-L-fuconate dehydrogenase (XCC4067) from *Xanthomonas campestris*. BmulJ_04920 is 60% similar in protein sequence to the known L-fuconate dehydratase from cog4948 (XCC4069)[79, 80]. However, BmulJ_04919 from cog1028 is not related to XCC4067, an enzyme that is known to catalyze the oxidation of L-fucose. The closest experimentally annotated homologue to BmulJ_04919 is L-rhamnose dehydrogenase (SKA58_03570) from *Sphingomonas* sp. SKA58.[66] In this paper, we have determined the three-dimensional structures of BmulJ_04915 and BmulJ_04919 and have shown that BmulJ_04919 catalyzes the oxidation of L-fucose to L-fucono-1,5-lactone and that BmulJ_04915 and two homologues (Bamb_1224 and Patl_0789) catalyze the hydrolysis of this product to L-fuconate.

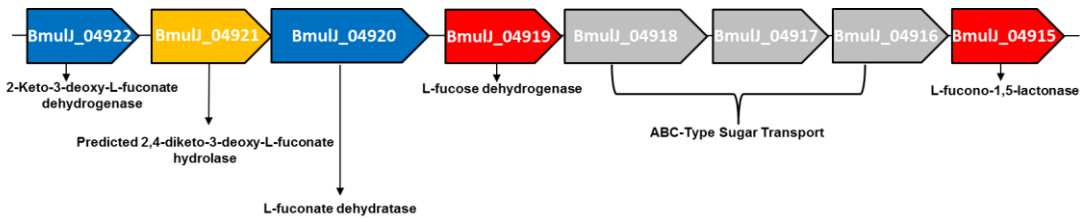


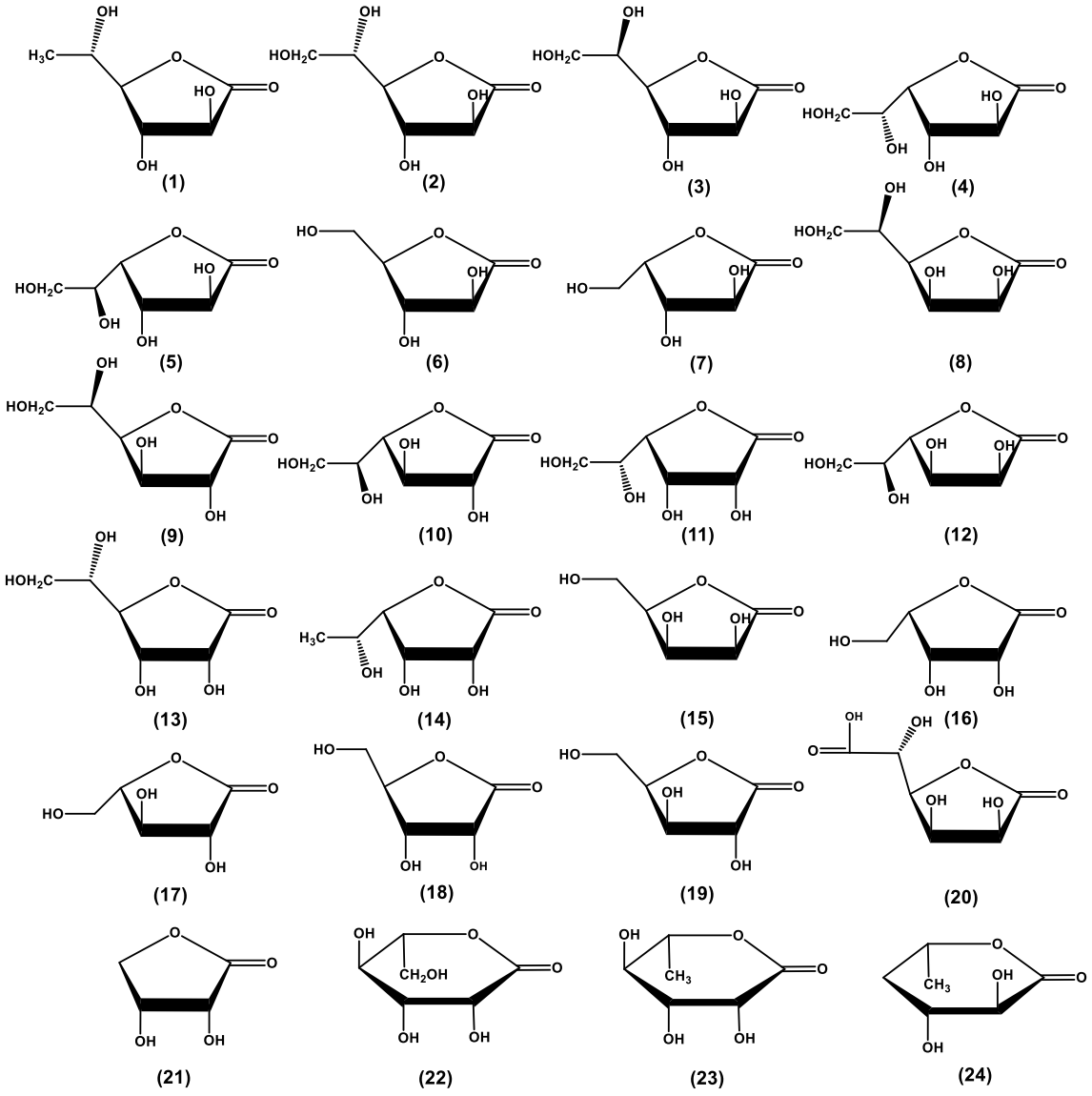
Figure 3.3. Genomic neighborhood of BmulJ_04915. Genes are color-coded as follows: red for those cloned and purified for this study, blue for those predicted from strong sequence similarity to genes encoding proteins of known function, yellow for those predicted genes based on genomic context, and gray for those predicted to function in carbohydrate transport.

Materials and Methods

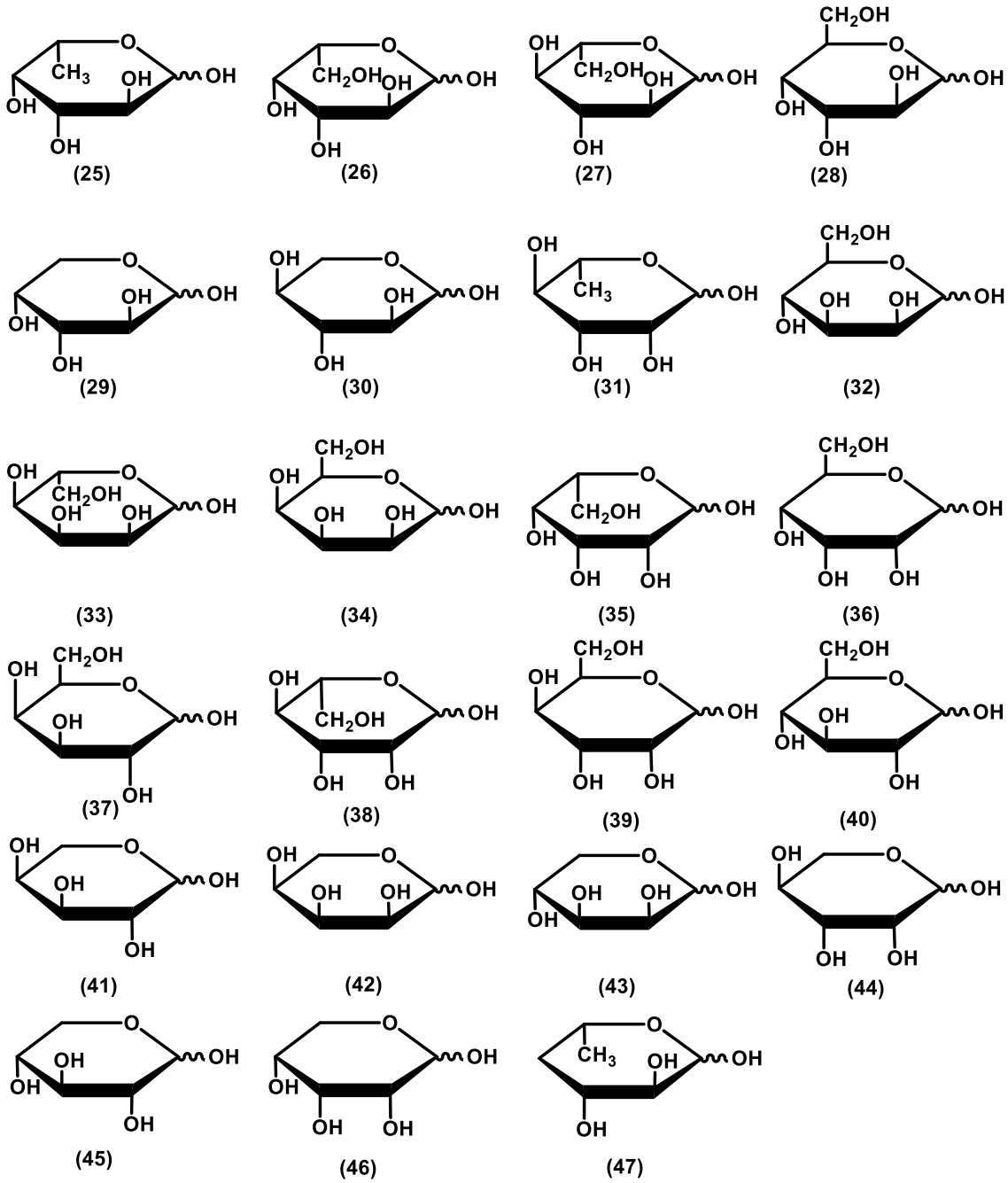
Materials. All chemicals and buffers were purchased from Sigma Aldrich unless otherwise specified. Sugar lactones that were not commercially available were synthesized according to published procedures with the exception of 4-deoxy-L-fucono-1,5-lactone, which was enzymatically synthesized.[51] The noncommercial lactones included the following: L-fucono-1,4-lactone (**1**), D-altrono-1,4-lactone (**3**),D-arabinono-1,4-lactone (**6**), L-xylono-1,4-lactone (**7**), L-mannono-1,4-lactone (**11**), D-talono-1,4-lactone (**12**), D-allono-1,4-lactone (**13**), L-rhamnono-1,4-lactone (**14**), D-lyxono-1,4-lactone (**15**),L-lyxono-1,4-lactone (**16**), L-arabinono-1,4-lactone (**17**), D-xylono-1,4-lactone (**19**), L-mannono-1,5-lactone (**22**), L-rhamnono-1,5-lactone (**23**), and 4-deoxy-L-fucono-1,5-lactone (**24**). Those sugar lactones that were available commercially included the following: L-galactono-1,4-lactone (**2**),L-glucono-1,4-lactone (**4**), D-idono-1,4-lactone (**5**), D-mannono-1,4-lactone (**8**), D-glucono-1,4-lactone (**9**), D-galactono-1,4-lactone (**10**), D-ribono-1,4-lactone (**18**), D-glucurono-6,3-lactone (**20**), and D-erythronolactone (**21**). These compounds were obtained from CarboSynth or ChromaDex. The structures of these lactones are presented in **Scheme 3.1**. The aldose sugars were obtained from either from Sigma Aldrich or Carbosynth, and the structures are presented in **Scheme 3.2**. These sugars included L-fucose (**25**), L-galactose (**26**), L-glucose (**27**), D-altrose (**28**), D-arabinose (**29**), L-xylose (**30**), L-rhamnose (**31**), D-mannose (**32**), L-allose (**33**), D-talose (**34**), L-talose (**35**), D-allose (**36**), D-galactose (**37**), L-mannose

(**38**), D-gulose (**39**), D-glucose (**40**), L-arabinose (**41**), L-ribose (**42**), D-lyxose (**43**), L-lyxose (**44**), D-xylose (**45**), D-ribose (**46**), and 4-deoxy-L-fucose (**47**).

Scheme 3.1. Lactone Substrate Library



Scheme 3.2: Sugar Substrate Library



Cloning, expression, and Purification of BmulJ_04915 and Patl_0798. Original cloning, overexpression, and purification of BmulJ_04915 and Patl_0798 were completed through collaboration with Dr. Steve Almo and Ronald Seidel from the Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York.

Cloning, Expression, and Purification of Bamb_1224. The gene for Bamb_1224 was cloned from *B. ambifaria* AMMD (gi|115351277) with the primer pair of 5'-GCAGGACCATATGACGTTCCGCATCGACGCTCATCAA-3' and 5'-GCAGGACAAGCTTTTGAGCGTGTTGATCGCCGGACGGCCGA-3'. Restriction sites for NdeI and HindIII were inserted into the forward and reverse primers, respectively. The PCR product was purified with a PCR cleanup system (Promega), doubly digested with NdeI and HindIII, and then ligated into a pET-30a(+) vector (Novagen) previously digested with NdeI and HindIII. The recombinant plasmid harboring this gene was transformed into BL21(DE3) cells (Novagen) via electroporation.

A single colony was used to inoculate 5 mL cultures containing LB and 50 μg/mL kanamycin and allowed to grow overnight at 37 °C. The overnight cultures were used to inoculate 1 L of LB medium supplemented with 50 μg/mL kanamycin. These cultures were incubated until the OD_{600}reached 0.4–0.6, and then 1.0 mM isopropyl β-thiogalactoside (IPTG) was added to induce gene expression at ambient temperature (25 °C) for an additional 18 h. Cells were harvested by centrifugation at 4 °C and then resuspended in 20 mM HEPES (pH 7.5). Approximately 0.1 mg/mL

phenylmethanesulfonyl fluoride (PMSF) was added to the cell suspension, and then the cells were lysed via multiple rounds of sonication in 50 mM HEPES (pH 7.4), 0.5 M NaCl, and 40 mM imidazole. Nucleic acids were removed from the supernatant solution by the addition of a 2% (w/v) protamine sulfate solution over the course of 30 min at 4 °C. The protamine sulfate-bound DNA was removed by centrifugation at 4 °C, and the supernatant was collected and applied to a 5 mL HisTrap HP (GE Healthcare) nickel affinity column. Protein was eluted from the nickel affinity column with 50 mM HEPES (pH 7.4), 0.5 M NaCl, and 500 mM imidazole over a gradient of 25 column volumes.

Cloning, Expression, and Purification of BmulJ_04919. The gene for BmulJ_04919 was cloned from *B. multivorans* ATCC 17616 (gi|189353674) with the primer pair of 5'-GCAGGAGCCATATGGATCTGAATCTGCAGGACAAGGTCGT3' and 5'-GCAGGAGCAAGCTTTCAGACGAGCGCACGATCGAGATGCGTAT-3'. Restriction sites for NdeI and HindIII were inserted into the forward and reverse primers, respectively. The PCR product was purified with a PCR cleanup system (Promega), doubly digested with NdeI and HindIII, and then ligated into a pET-30a(+) vector (Novagen) previously digested with NdeI and HindIII. The recombinant plasmid harboring this gene was transformed into BL21(DE3) cells (Novagen) via electroporation.

A single colony was used to inoculate 5 mL cultures containing LB and 50 μg/mL kanamycin, and the cultures were allowed to grow overnight at 37 °C. The overnight cultures were used to inoculate 1 L of LB medium supplemented with 50 μg/mL kanamycin. These cultures were incubated until the OD_{600} reached 0.4–0.6, and then

1.0 mM IPTG was added to induce gene expression at ambient temperature (25 °C) for an additional 18 h. Cells were harvested by centrifugation at 4 °C and then resuspended in 20 mM HEPES (pH 7.5). Approximately 0.1 mg/mL PMSF was added to the cell suspension, and then the cells were lysed via multiple rounds of sonication. Nucleic acids were removed from the supernatant solution by the addition of a 2% (w/v) protamine sulfate solution over the course of 30 min at 0 °C. The protamine sulfate-bound DNA was removed by centrifugation at 4 °C. Solid ammonium sulfate was added to 50% of saturation and the precipitated protein isolated by centrifugation. The protein was resuspended in a minimal amount of 20 mM HEPES (pH 7.5) and then applied to a High Load 26/60 Superdex 200 gel filitration column (GE Healthcare). Fractions containing the protein of expected size were pooled and purified further by ion exchange chromatography with a ResourceQ column (6 mL) at pH 7.5.

Crystallization and Data Collection for BmulJ_04915 and BmulJ_04919. The X-ray crystal structures of BmulJ_04919 and BmulJ_04919 were solved through collaboration with Dr. Steven Almo and Dr. Mathew Vetting at the Department of Biochemistry, Albert Einstein College of Medicine, at Bronx, New York. The PDB codes for BmulJ_04915 are 4dlf, 4dlm, 4dnm, and 4do7. PDB codes for BmulJ_04919 are 4glo, 4gvx and 4gkb.

Measurement of Enzyme Activity. The enzymatic hydrolysis of lactones was monitored with a SpectraMax340 UV–visible spectrophotometer using a colorimetric pH indicator assay at 30 °C. Protons released from the carboxylate product during lactone hydrolysis were measured using the pH indicator cresol purple or bromothymol blue.[81] Reaction mixtures measured with cresol purple contained 2.5 mM BICINE buffer (pH 8.3), 0.2 M NaCl, 0–1.0 mM lactone, 0.1 mM cresol purple, and the lactonase. Reaction mixtures measured with bromothymol blue contained 2.5 mM MOPS buffer (pH 7.1), 0.2 M NaCl, 0–0.5 mM lactone, and 0.1 mM bromothymol blue. The final concentration of DMSO was 1%. Changes in absorbance at 577 nm (ε = 1764 M^{-1} cm^{-1}) and 616 nm (ε = 1135 M^{-1} cm^{-1}) were monitored in 96-well plates for cresol purple and bromothymol blue, respectively. Background rates arising from acidification by atmospheric CO_2 were observed and subtracted from the initial rates. The dehydrogenase activity catalyzed was monitored with a SpectraMax340 UV–visible spectrophotometer by measuring the reduction of $NADP^+$ or NAD^+ at 340 nm (ε = 6220 M^{-1} cm^{-1}) at 30 °C. The assays were performed in 50 mM BICINE buffer (pH 8.0), varying concentrations of sugar substrates, and 0.5 mM $NADP^+$. Kinetic constants for $NADP^+$ and NAD^+ were determined in the same manner as described above except that the assay conditions were as follows: 50 mM BICINE buffer (pH 8.0), varying concentrations of nucleotide, and 200 µM L-fucose.

Data Analysis. The kinetic constants were determined from a fit of the initial velocity data to eq 3.1 using the nonlinear least-squares fitting program in SigmaPlot 9.0

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

where v is the initial velocity, E_t is the total enzyme concentration, k_{cat} is the turnover number, [A] is the substrate concentration, and K_m is the Michaelis constant.

Metal Analysis. The metal content of BmulJ_04915 was determined with an Elan DRC II ICP-MS instrument as previously described[71]. Protein samples for ICP-MS analysis were digested with HNO_3 and then refluxed for 30 min[58]. All buffers were passed through a column of Chelex 100 (Bio-Rad) to remove trace metal contamination. EDTA and 1,10-phenanthroline (1.0 mM) were incubated with 1.0 μM BmulJ_04915 in 50 mM buffer at pH values ranging from 6 to 10 to remove divalent metal ions. The buffers for these experiments included CHES (pH 6.0), HEPES (pH 7.0), BICINE (pH 8.0), and CHES (pH 9.0 and 10.0). The effect of added divalent cations on the catalytic activity of BmulJ_04915 was determined by adding Mn^{2+}, Zn^{2+}, Co^{2+}, Cu^{2+}, or Ni^{2+} (0–500 μM) directly to the assay mixtures. The purified enzyme was also incubated with 50–500 molar equiv of these divalent cations for 24 h at 4 °C in 50 mM HEPES (pH 7.5) and subsequently assayed for catalytic activity.

Reaction Product of BmulJ_04919 with L-Fucose. The product of the reaction catalyzed by BmulJ_04919 was determined through ^{1}H, ^{13}C heteronuclear multiple-bond correlation (HMBC) and correlation nuclear magnetic resonance (NMR) spectroscopy using a Bruker Avance III 500 MHz spectrometer with an H, C, N cryoprobe. WATERGATE solvent suppression was performed on reaction samples in 500–1000 µL of D_2O with 100 mM acetate buffer (pH 5.5), 25 mM NAD^+, 25 mM L-fucose, and 100 µM BmulJ_04919. After being incubated for 10 min, the enzyme was removed from the reaction mixture using a 3 kDa cutoff VWR centrifugal filter. NAD^+ and NADH were removed by the addition of Dowex-1 X2 anion exchange resin that was previously equilibrated with 50 mM acetate buffer (pH 5.5). The pH of the sample was then adjusted to pH 4.2 with 1 M acetate buffer (pH 4.2). L-Fucono-1,5-lactone was identified as a substrate for BmulJ_04915 using ^{1}H NMR spectroscopy. The reactions were conducted with 50 mM phosphate buffer (pH 6.5), 15 mM L-fucose, 15 mM NAD^+, 10 µM BmulJ_04919, and 10 µM BmulJ_04915 in a final reaction volume of 200 µL. ^{1}H NMR spectra were recorded over 1 min intervals with WATERGATE solvent suppression.

Enzymatic Synthesis of 4-Deoxy-L-fucono-1,5-lactone. 4-Deoxy-L-fucono-1,5-lactone (**24**) was synthesized enzymatically from 4-deoxy-L-fucose (**47**) using BmulJ_04919 as the catalyst. The reaction was conducted in 50 mM phosphate buffer (pH 7.0), 15 mM 4-deoxy-L-fucose, and 15 mM NAD^+ in a final volume of 1 mL for 20 min. The enzyme was removed from the reaction mixture using a 3 kDa cutoff VWR

centrifugal filter. NAD^+ and NADH were removed by the addition of Dowex-1 X2 anion exchange resin. The structure of the lactone was determined using HMBC and correlation NMR spectroscopy with a Bruker Avance III 500 MHz spectrometer with an H, C, N cryoprobe using WATERGATE solvent suppression

Sequence Similarity Network for cog3618. Proteins belonging to cog3618 were identified from the NCBI protein database using the query "cog3618". The proteins within cog3618 were subjected to an all-by-all BLAST at a specified *E*value (10^{-50}, 10^{-60}, etc.) using the NCBI standalone BLAST program. The BLAST files were opened and visualized in the similarity network program Cytoscape.(28)

Results

Purification of BmulJ_04915. The gene for BmulJ_04915 was cloned and expressed in *Escherichia coli*, and the protein was purified to homogeneity. The purified enzyme contained ~0.7 molar equiv of Zn^{2+}. The addition of chelating agents or divalent metal ions (Mn^{2+}, Zn^{2+}, Co^{2+}, Cu^{2+}, or Ni^{2}) directly to the assay mixture did not affect the rate of enzyme-catalyzed hydrolysis of L-fucono-1,4-lactone (**1**). The purified enzyme was incubated with EDTA or 1,10-phenanthroline at pH values ranging from 6 to 10 to chelate the metal ions bound to the protein. The addition of chelators did not diminish the catalytic activity of this enzyme, and ICP-MS analysis demonstrated that the addition of 1,10-phenanthroline removed >95% of the Zn^{2+} that was initially bound to the protein. The apoenzyme had the same catalytic activity as the as-purified

enzyme. Therefore, BmulJ_04915 does not require a divalent cation for substrate turnover.

Substrate Specificity. BmulJ_04915 is a member of cog3618 from the amidohydrolase superfamily. Other enzymes from cog3618 have been shown to catalyze the hydrolysis of lactones.[65, 66, 78] The genomic context is indicative of an enzyme that participates in the metabolism of carbohydrates, and thus, this enzyme was predicted to catalyze the hydrolysis of sugar lactones. The enzyme was screened against a small and focused library of 23 lactones as potential substrates for this enzyme (see **Scheme 3.1**). BmulJ_04915 exhibited catalytic activity for the hydrolysis of L-fucono-1,4-lactone (**1**),D-arabino-1,4-lactone (**6**), L-xylono-1,4-lactone (**7**), and L-galactono-1,4-lactone (**2**). The best substrate from this initial set of compounds is L-fucono-1,4-lactone. The kinetic constants are listed in **Table 3.1**.

Table 3.1. Catalytic Constants for BmulJ_04915, Bamb_1224, and Patl_0798

	BmulJ_04915			Bamb_1224			Patl_0798		
compound	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
At pH 8.3									
L-fucono-1,4-lactone (1)	140 ± 8	1.4 ± 0.2	1.0 (0.1) × 10^5	89 ± 4	0.9 ± 0.1	1.0 (0.1) × 105	139 ± 6	1.2 ± 0.1	1.2 (0.2) × 10^5
L-galactono-1,4-lactone (2)	32 ± 2	1.1 ± 0.2	3.0 (0.3) × 10^4	55 ± 2	1.3 ± 0.1	4.0 (0.3) × 104	119 ± 7	1.5 ± 0.2	8.0 (0.5) × 10^4
D-arabinono-1,4-lactone (6)	59 ± 2	0.7 ± 0.05	8.4 (0.5) × 10^4	19 ± 1	0.8 ± 0.1	2.4 (0.2) × 104	25 ± 3	0.75 ± 0.02	3.3 (0.2) × 10^4
L-xylono-1,4-lactone (7)	14 ± 1	1.5 ± 0.3	1.0 (0.1) × 10^4	6.0 ± 0.5	1.3 ± 0.3	5.0 (0.5) × 103	10 ± 0.3	1.5 ± 0.4	6.6 (0.2) × 10^4
4-deoxy-L-fucono-1,5-lactone (24)	ND	ND	ND	ND	ND	ND	ND	ND	ND
At pH 7.1									
L-fucono-1,4-lactone (1)	4.0 ± 0.1	0.6 ± 0.1	7 (0.1) × 10^4	4.1 ± 0.3	0.9 ± 0.1	5.0 (0.1) × 103	3.0 ± 0.2	0.6 ± 0.1	4.5 (0.5) × 10^3
L-galactono-1,4-lactone (2)	6.0 ± 0.5	1.3 ± 0.2	5.0 (0.5) × 10^3	6.2 ± 2.0	0.5 ± 0.1	4.0 (0.5) × 103	5.7 ± 0.3	2.6 ± 0.3	2.2 (0.1) × 10^3
D-arabinono-1,4-lactone (6)	2.0 ± 0.2	1.0 ± 0.3	1.6 (0.3) × 10^3	2.2 ± 0.2	0.7 ± 0.1	3.2 (0.2) × 103	1.8 ± 0.2	1.6 ± 0.3	1.0 (0.1) × 10^3
L-xylono-1,4-lactone (7)	ND	ND	1.0 (0.1) × 10^3	ND	ND	7.0 (0.3) × 102	ND	ND	1.0 (0.2) × 10^3
4-deoxy-L-fucono-1,5-lactone (24)	994 ± 64	0.13 ± 0.03	8.0 (1.0) × 10^6	1530 ± 47	0.21 ± 0.02	7.2 (0.6) × 106	1320 ± 80	0.45 ± 0.02	2.6 (0.2) × 10^6

*ND- Not determined

Structure of BmulJ_04915. The structure of BmulJ_04915 was determined by molecular replacement to 2.15 Å resolution. There is a monomer in the asymmetric unit that is consistent with the protein being a monomer in solution. The final model contains residues 3–289, all of which are well-defined in electron density. The crystal from which BmulJ_04915 was determined requires TCEP and Zn^{2+}(added prior to the knowledge that BmulJ_04915 was not a metalloenzyme) to form and was only sporadically reproducible. A well-defined Zn^{2+} ion, coordinated by Asp-176 and His-222 of one molecule and Cys-71 and Asp-72 of a crystallographically related molecule, specifies how Zn^{2+} was critical to the crystallization. BmulJ_04915 forms a distorted $(\beta/\alpha)_8$ TIM barrel with an active site at the C-terminal end of the β-barrel consistent with other members of the amidohydrolase family (**Figure 3.4A**). There are few inserted secondary structural elements, with most of the defining structural features being extended loops connecting the $(\beta/\alpha)_8$ TIM barrel elements. Residual density within the active site was modeled as a bound HEPES molecule derived from crystallization buffer. The piperazine ring lies against the face of Trp-22 (loop between β-strand 1 and α-helix 1), while the sulfonate is interacting with Arg-106 and the three active site histidine residues (His-9, His-11, and His-166) (**Figure 3.4B**). A search using the SSM server returns a number of amidohydrolases at >2.0 Å root-mean-square deviation (rmsd) and <20% sequence identity, demonstrating that the structure of BmulJ_04915 is significantly different from those previously determined.[82] The closest structure is that of the 2-pyrone-4,6-dicarboxylic acid hydrolase (LigI)

from *Sphingomonas paucimobilis* (rmsd of 2.2 Å over 229 Cα atoms, 20% sequence identity), which also catalyzes the hydrolysis of a lactone.

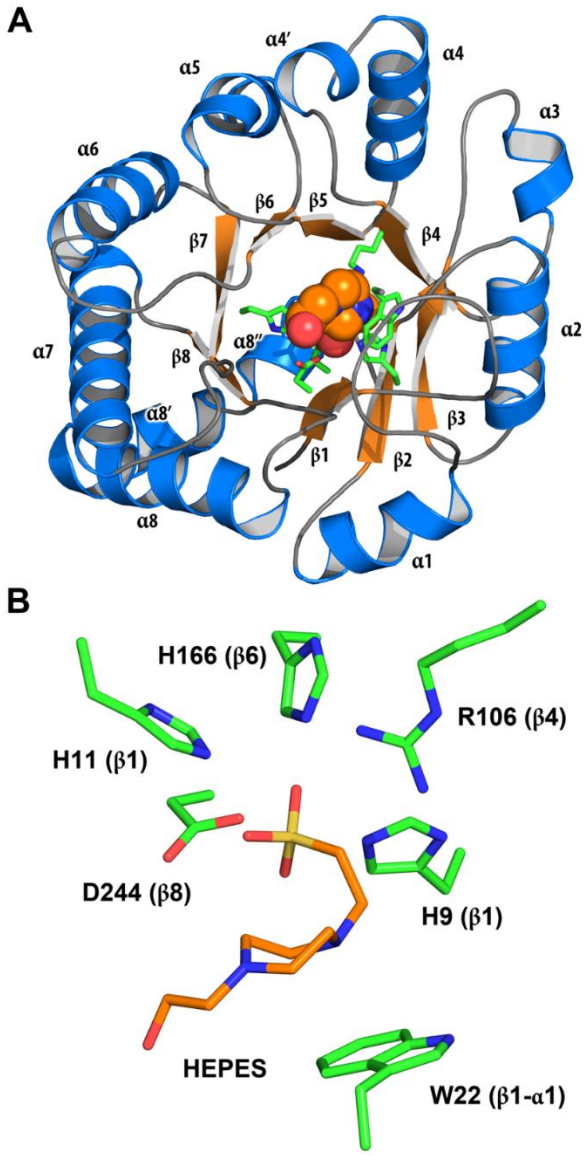


Figure 3.4. Structure of BmulJ_04915. (A) Ribbon diagram of BmulJ_04915. The HEPES molecule is shown as spheres. (B) Residues directly adjacent to bound HEPES in the active site of BmulJ_04915. Numbers in parentheses designate the approximate location in the secondary structure from which the residue originates.

Purification and Properties of BmulJ_04919. A gene for a protein currently annotated as a short chain dehydrogenase of unknown specificity (BmulJ_04919) is positioned in the same putative operon as BmulJ_04915. This gene was cloned and overexpressed and the protein purified to homogeneity. BmulJ_04919 is a member of cog1028, which includes other enzymes with the following experimentally verified functions: 2-dehydro-3-deoxy-L-fuconate dehydrogenase,(11) 2-dehydro-3-deoxy-L-rhamnonate dehydrogenase,(9)L-rhamnose dehydrogenase,(9) 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase,(30)D-gluconate dehydrogenase,(31) and sorbitoL-6-phosphate dehydrogenase.(32) The closest homologue to BmulJ_04919 with a verified catalytic activity in cog1028 appears to be L-rhamnose dehydrogenase from *Sphingomonas* sp. SKA58. BmulJ_04919 is unrelated to the putative L-fucose dehydrogenase (XCC4065 from *X. campestris*) from cog0667.(11, 33) The oxidized product of the reaction catalyzed by BmulJ_04919 is likely to be the physiological substrate for BmulJ_04915.

Substrate Specificity of BmulJ_04919. It was anticipated that BmulJ_04919 would catalyze the formation of L-fucose-1,5-lactone fromL-fucose and $NADP^+$ because the pyranose form of L-fucose is predominant in solution (**Figure 3.6A**). To determine the substrate specificity of this enzyme, a small library of pentose and hexose sugars was tested as potential substrates by monitoring the reduction of $NADP^+$ at 340 nm (**Scheme 3.2**). L-Fucose (**25**), L-galactose (**26**), and D-arabinose (**29**) were the only sugars

oxidized by this enzyme. The dehydrogenase has the highest activity with L-fucose, and the kinetic constants are listed in **Table 3.2**.

Table 3.2: Catalytic constants for BmulJ_04919 at pH 8.0

Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}\,s^{-1}$)
L-fucose (**25**)	10 ± 0.3	6.2 ± 0.8	1.5 (0.2) x 10^6
D-arabinose (**29**)	21.0 ± 0.08	94 ± 3.0	2.3 (0.1) x 10^5
L-galactose (**26**)	11 ± 0.5	285 ± 36	3.7 (0.3) x 10^4
4-deoxy-L-fucose (**47**)	26 ± 1.0	786 ± 92	4.0 (0.1) x 10^4
$NADP^+$	11 ± 0.2	4.0 ± 0.5	3.0 (0.4) x 10^6
NAD^+	15 ± 0.6	28 ± 4	5.3 (0.1) x 10^5

Identification of the Initial Reaction Product of BmulJ_04919. The only product of an L-fucose dehydrogenase that has ever been structurally characterized is L-fucono-1,4-lactone (**1**). However, it has been proposed that L-fucono-1,5-lactone is the true product of the reaction, but this compound has never been observed because of the rapid nonenzymatic conversion to L-fucono-1,4-lactone (**1**) or hydrolysis to L-fuconate. To help stabilize product formation, the reaction was conducted at pH 5.5 and then quenched at pH 4.2. The initial product was structurally characterized by NMR spectroscopy and shown to be L-fucono-1,5-lactone with the resonance for the methyl

group at 1.32 ppm, which quickly rearranges to the 1,4-lactone (1.24 ppm) at pH >5.0. For further characterization, the reaction was quenched at pH 4.2, which greatly slowed the intermolecular rearrangement.

The enzymatically produced L-fucono-1,5-lactone is rapidly converted to L-fucono-1,4-lactone (**1**). Enzyme, NAD^+, and L-fucose were added together at pH 5.5, and the reaction was monitored by ^1H NMR spectroscopy until L-fucono-1,5-lactone was the predominate product. The enzyme was removed from the reaction mixture and the pH adjusted to 6.5. The time course for the subsequent change in the ^1H NMR spectrum for the C-6 methyl groups of L-fucose (**25**) and the reaction products is presented in **Figure 3.5**. At 1.1 and 1.2 ppm, there is a pair of doublets for the α- and β-anomers of L-fucose (**25**). At 1.24 and 1.32 ppm are the two doublets for L-fucono-1,4-lactone (**1**) and L-fucono-1,5-lactone, respectively (**Figure 3.5A**). At the earliest time points, L-fucono-1,5-lactone is clearly dominant versus L-fucono-1,4-lactone (**1**). However, after 5 min, approximately half of the L-fucono-1,5-lactone has been converted to the L-fucono-1,4-lactone (**Figure 3.5B**). After 1 h, nearly all of the L-fucono-1,5-lactone was converted to L-fucono-1,4-lactone (**1**) (**Figure 3.5C**). The nonenzymatic transformation of L-fucono-1,5-lactone to L-fucono-1,4-lactone (**1**) does not utilize the hydrolyzed product, L-fuconate, as a reactive intermediate. The rate constant is approximately 0.12 min^{-1} at pH 6.5.

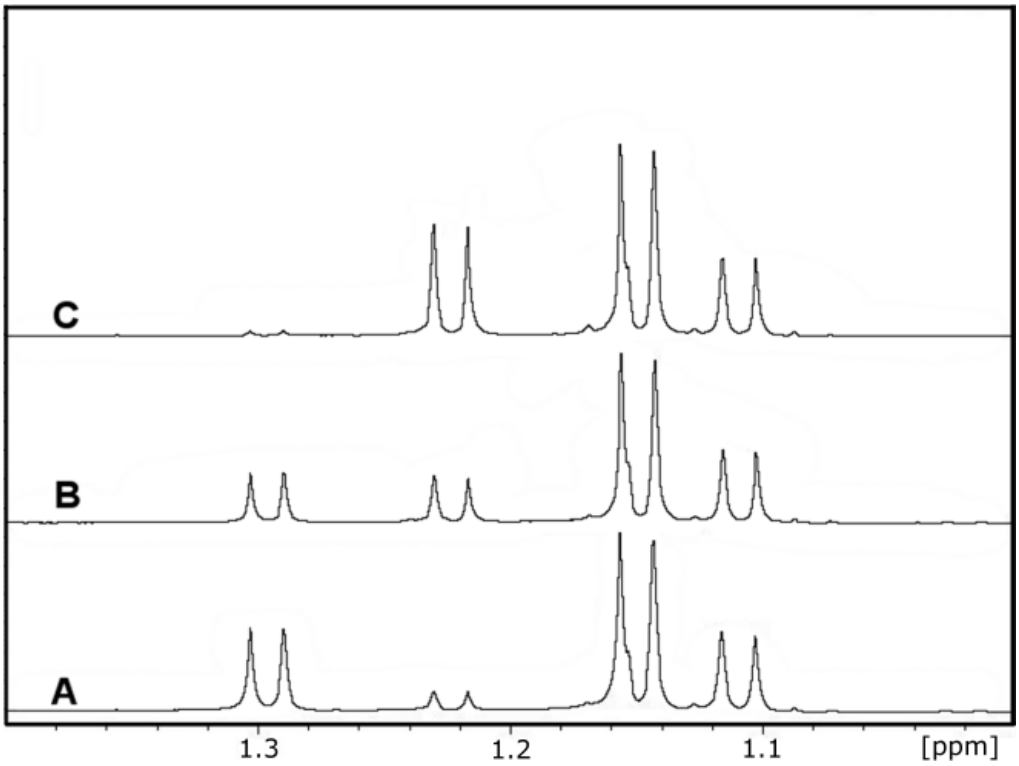


Figure 3.5. ^{1}H NMR time course for the nonenzymatic conversion of L-fucono-1,5-lactone to L-fucono-1,4-lactone. The resonances corresponding to the C-6 methyl groups of both α-L-fucose (1.12 ppm) and β-L-fucose (1.15 ppm), L-fucono-1,5-lactone (1.32 ppm), andL-fucono-1,4-lactone (1.24 ppm) are presented. (A) The major reaction product of BmulJ_04915 at pH 4.2 is shown to be L-fucono-1,5-lactone. (B) Five minutes after the reaction mixture had been adjusted to pH 6.5, the L-fucono-1,4-lactone is at a concentration equal to that of the original enzymatic product. (C) Sixty minutes after the pH of the reaction mixture had been adjusted to pH 6.5, the major peak is L-fucono-1,4-lactone.

Identification of the Physiological Substrate for BmulJ_04915. To determine whether BmulJ_04915 preferentially hydrolyzes the 1,4- or 1,5-lactone of L-fuconate, BmulJ_04919 was incubated with L-fucose (**25**) and NAD^{+}. Shown in **Figure 3.6A** is the^{1}H NMR spectrum of the methyl carbon for the α- and β-anomers of L-fucose (**25**) before the addition of BmulJ_04919. After ~5 min, the concentrations of the two lactones were approximately equal as shown in **Figure 3.6B**. The lactonase (BmulJ_04915) was

added, and the spectrum of the reaction product was monitored as a function of time. The resonances for theL-fucono-1,5-lactone disappeared at a significantly higher rate than those for L-fucono-1,4-lactone (**1**) (**Figure 3.6C**). The formation of L-fuconate was confirmed by the new resonance at approximately 1.18 ppm.

To further demonstrate that BmulJ_04915 catalyzes the hydrolysis of the 1,5-lactone faster than the 1,4-lactone, 4-deoxy-L-fucono-1,5-lactone (**24**) was synthesized enzymatically using 4-deoxy-L-fucose (**47**) as a substrate for BmulJ_04919. This compound is missing the hydroxyl group at C-4 and thus cannot form a 1,4-lactone. The product of this reaction was shown to be 4-deoxy-L-fucono-1,5-lactone (**24**) by NMR spectroscopy (**Figure 3.7A**). As no 4-hydroxyl group is available, the six-member lactone is relatively stable and undergoes slow nonenzymatic hydrolysis. A new methyl doublet appears at 1.25 ppm when 4-deoxy-L-fucose is oxidized to 4-deoxy-L-fucono-1,5-lactone by BmulJ_04919 (**Figure 3.7B**). After hydrolysis with BmulJ_04915, this resonance disappears and the methyl resonance for 4-deoxy-L-fuconate appears at 1.07 ppm (**Figure 3.7C**). The 4-deoxy-L-fucono-1,5-lactone (**24**) was isolated, and the kinetic constants for the hydrolysis of this compound by BmulJ_04915, Bamb_1224, and Patl_0798 were determined at pH 7.1. The values of k_{cat} and k_{cat}/K_m for the hydrolysis of 4-deoxy-L-fucono-1,5-lactone are approximately 250- and 115-fold greater, respectively, than those for L-fucono-1,4-lactone assayed at pH 7.1 (**Table 3.1**).

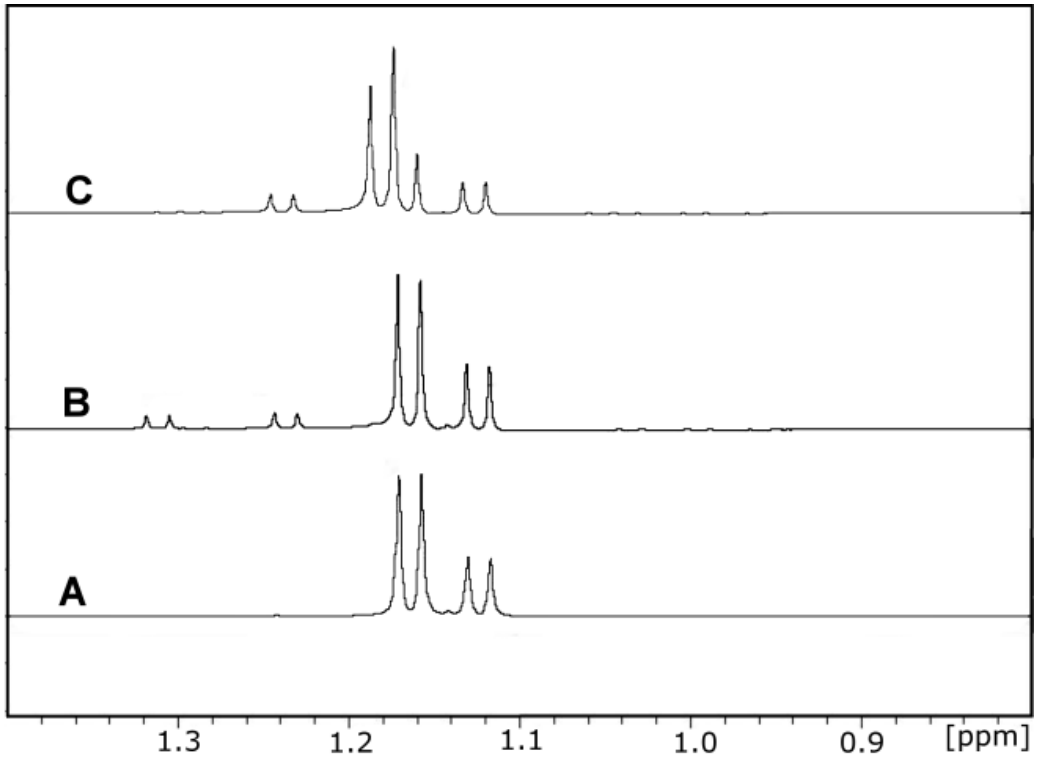


Figure 3.6. ^{1}H NMR time course of the enzymatic conversion of L-fucose to L-fucono-1,5-lactone and L-fucono-1,5-lactone to L-fuconate. The resonances corresponding to the C-6 methyl groups of both α-L-fucose (1.125 ppm) and β-L-fucose (1.15 ppm), L-fucono-1,5-lactone (1.295 ppm), L-fucono-1,4-lactone (1.225 ppm), and L-fuconate (1.182 ppm) are provided. (A) L-Fucose, the substrate for BmulJ_04919, prior to addition of the enzyme at pH 6.5. (B) Five minutes after the addition of BmulJ_04919 to L-fucose, the enzymatic product, L-fucono-1,5-lactone, and the nonenzymatic product, L-fucono-1,4-lactone, are at equal concentrations. (C) Five minutes after the addition of BmulJ_04915 to the reaction mixture, the L-fucono-1,5-lactone is no longer present in the reaction mixture and the L-fucono-1,4-lactone appears to be unchanged. L-Fuconate (1.182 ppm) appears to be the major product. The α-anomer of L-fucose appears to be unchanged; however, the concentration of the β-anomer has been reduced to at least half of the original value.

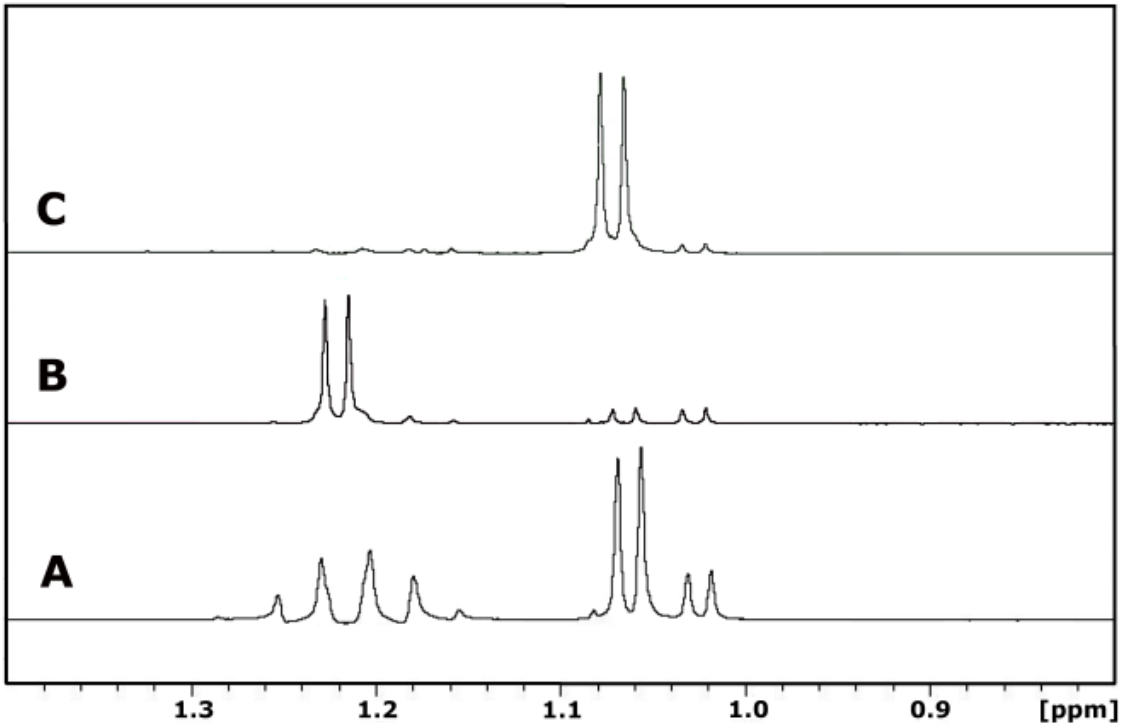


Figure 3.7. ^{1}H NMR time course of the enzymatic conversion of 4-deoxy-L-fucose to 4-deoxy-L-fucono-1,5-lactone and 4-deoxy-L-fuconate. The resonances corresponding to the C-6 methyl groups of both α-4-deoxy-L-fucose (1.02 ppm) and β-4-deoxy-L-fucose (1.06 ppm), 4-deoxy-L-fucono-1,5-lactone (1.25 ppm), and 4-deoxy-L-fuconate (1.07 ppm) are provided. (A) 4-Deoxy-L-fucose, prior to addition of enzymes at pH 6.5. (B) One minute after the addition of BmulJ_04919 to 4-deoxy-L-fucose, the enzymatic product is 4-deoxy-L-fucono-1,5-lactone. (C) One minute after the addition of BmulJ_04915 to the reaction mixture, 4-deoxy-L-fucono-1,5-lactone is no longer present in the reaction mixture and 4-deoxy-L-fuconate (1.07 ppm) appears to be the major product.

Structure of BmulJ_04919. Unliganded BmulJ_04919 crystallized in space group $P6_1$ with a tetramer per asymmetric unit and was phased by molecular replacement using a tetrameric model [PDB entry 3FTP, 3-ketoacyl(acyl carrier protein) reductase] and refined to a resolution of 1.5 Å . The NADP–L-fucose–BmulJ_04919 ternary complex crystallized in space group *C*2 with two dimers per asymmetric unit and was phased by molecular replacement using a single subunit of the unliganded structure and refined to a resolution of 1.5 Å. The density was sufficient to build the entire structure, excluding a portion of a helix–turn–helix motif (residues ~192–202) in two subunits of the unliganded structure. The subunit structure of BmulJ_04919 consists mainly of a Rossmann fold dinucleotide cofactor binding motif in which a central, twisted β-sheet consisting of seven parallel β-strands (3-2-1-4-5-6-7 strand topology) is flanked by five helices (α1−α3, α7, and α9) on one side and four helices (α3−α6) on the other (**Figures 3.8**). A structure similarity search utilizing the SSM server indicates a similarity to a number of short chain dehydrogenase/reductase (SDR) family enzymes with an rmsd of <1.5.(29) The highest scores were for human retinal short chain dehydrogenase/reductase (PDB entry1YDE, rmsd of 1.15 Å, sequence identity of 35%, *Z* score of 16.2) and *Thermoplasma acidophilum* D-aldohexose dehydrogenase, AldT (PDB entry 2DTE, rmsd of 1.31 Å, sequence identity of 32%, *Z* score of 15.3). Structurally, the largest differences between these enzymes are modest deviations in the positions of the β4−α4 loop, α4, the α7/α8 helix–turn–helix motif, and α2. For example, in AldT, α2 is converted to a loop and 10 residues are deleted compared to

the same region in BmulJ_04919.(36) Examination of the intermolecular interactions of BmulJ_04919 within the two crystal forms is consistent with BmulJ_04919 existing as a tetramer in solution. In addition, the top structural homologues as indicated by the SSM search also exist as similar tetramers in their crystal structures. The tetramer is composed of a dimer of dimers with 222 point group symmetry where the majority of interactions are between the A and B subunits and the A and D subunits. There are only minor interactions between the A and C subunits, where the C-terminus of the A subunit interacts with the C-terminus of the C subunit and α5 forming one wall of the L-fucose binding pocket. Interestingly, the C-terminus of AldT was found to cover the sugar binding pocket entirely and was postulated to stabilize the active site conformation and promote the catalytic reaction, with possibly enhanced dissociation at thermophilic temperatures[83]. In contrast, the C-terminus of BmulJ_04919 remains 6–7 Å from the sugar binding pocket and appears to be welL-anchored, suggesting it does not move during the reaction.

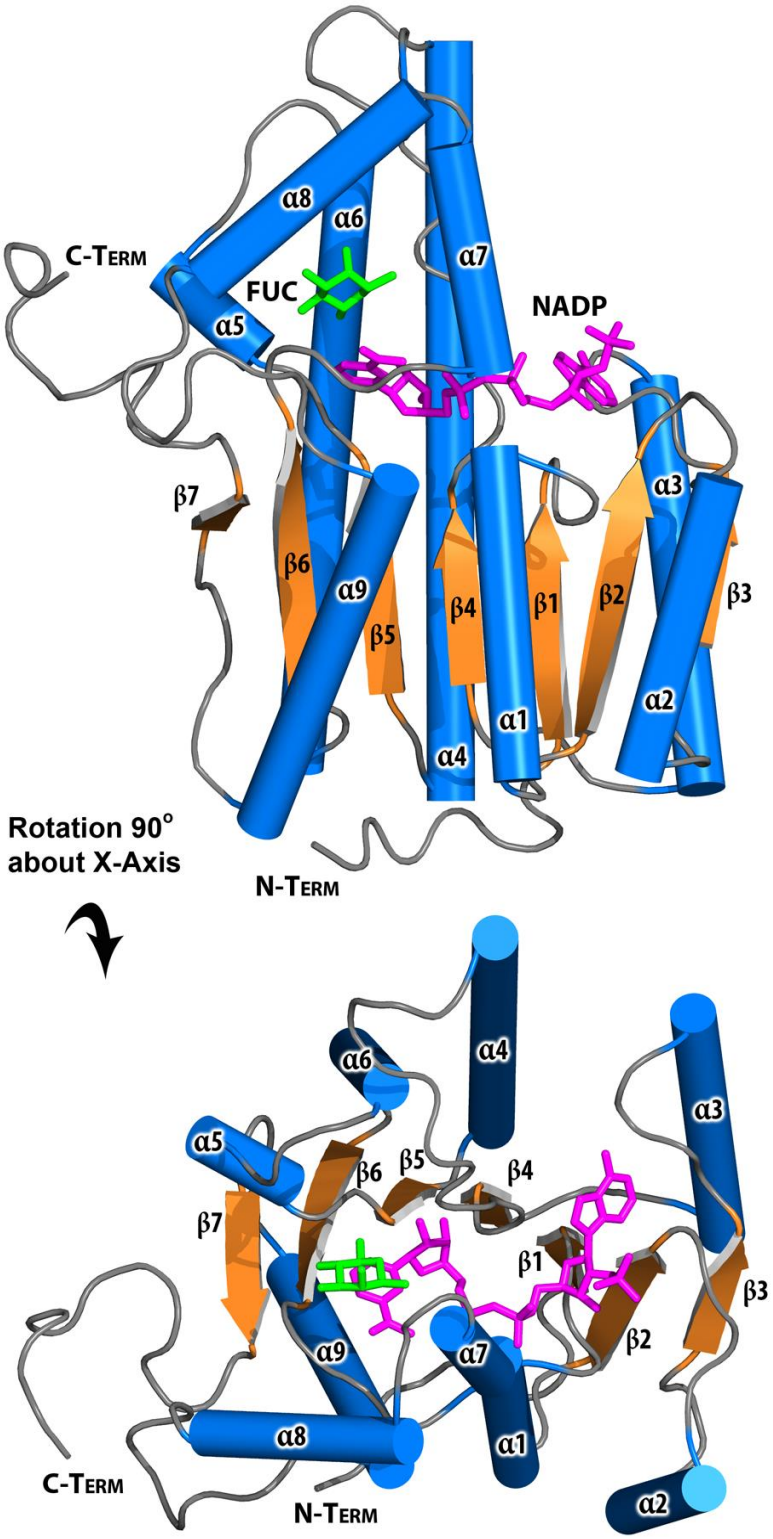


Figure 3.8. Ribbon diagram of BmulJ_04919 with bound NADP (magenta sticks) and L-fucose (green sticks).

$NADP^+$ Binding Site of BmulJ_04919. Clear and continuous electron density for $NADP^+$ was visualized in the BmulJ_04919 ternary complex (**Figure 3.9A**). As is common in SDR enzymes, the $NAD(P)^+$ is bound at the C-terminal edge of the seven-stranded β-sheet in an extended conformation (**Figure 3.9B**).[84] In addition to the general topology, the Rossman fold also includes a highly variable Gly-rich sequence essential for coordination of the cofactor pyrophosphate. In BmulJ_04919, this sequence lies between the end of β-strand 1 and the start of α-helix 1 with the sequence $G_{14}GASGIGG_{21}$. Clear binding determinants that support the preference of BmulJ_04919 for $NADP^+$ over NAD^+ can be identified. The 2'-phosphate of $NADP^+$ forms a salt bridge with the guanidinium group of Arg-39 and a hydrogen bond with the side chain of His-40, both originating from the loop between β2 and α2 (**Figure 3.9C**). Utilization of this loop for cofactor discrimination is typical for SDR enzymes.[84] Arg-39 is positioned by a hydrogen bond to Glu-63, originating from the loop between β-strand 3 and α-helix 3. The side chain of Arg-39 additionally stacks against the face of the adenosine moiety, which in combination with van der Walls contacts with Leu-64 and a hydrogen bond from Glu-63 to the adenosine amino group completes the binding pocket for the adenosine moiety.

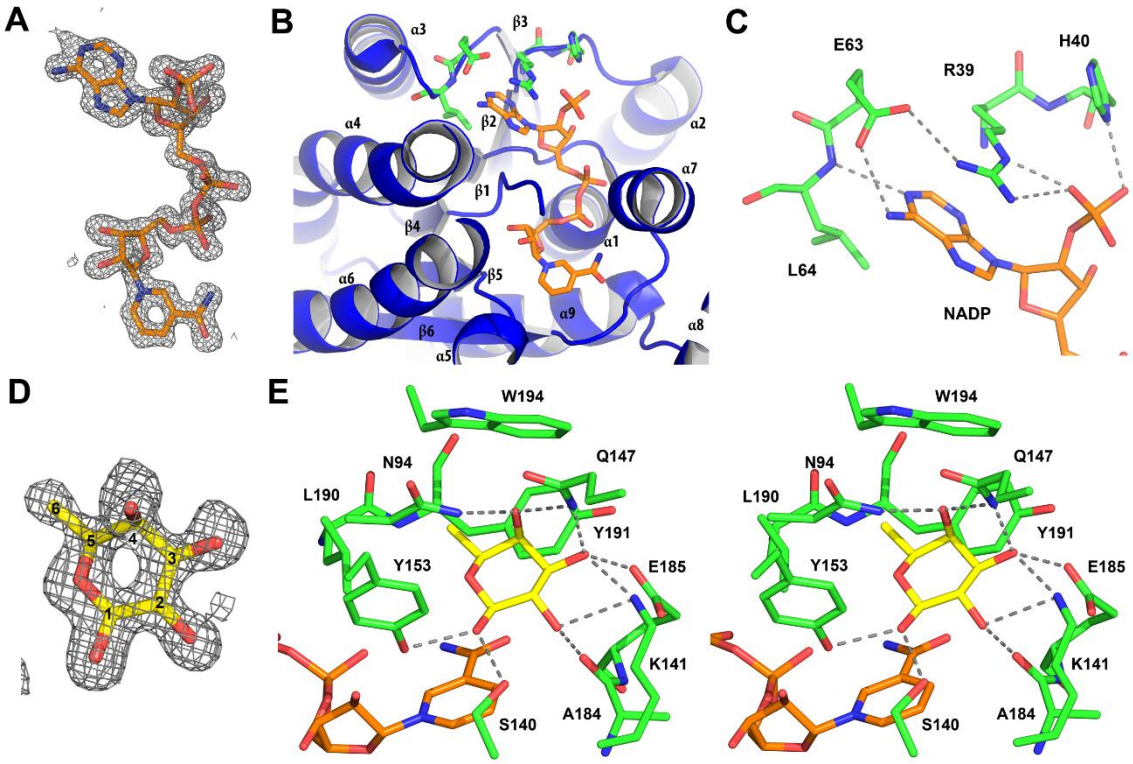


Figure 3.9. Interactions of BmulJ_04919 with ligands. (A) The 2.5σ $F_o - F_c$ kick map for $NADP^+$ bound to the $NADP^+$–L-fucose–BmulJ_04919 ternary complex. (B) Interactions of $NADP^+$ with the secondary structure elements of BmulJ_04919. (C) 2′-Adenosine phosphate binding site of BmulJ_04919. $NADP^+$ is shown as sticks with orange carbons, and residues of BmulJ_04919 adjacent to the 2′-adenosine phosphate are shown as sticks with green carbons. (D) The 2.5σ $F_o - F_c$ kick map for L-fucose bound to the $NADP^+$–L-fucose–BmulJ_04919 ternary complex. L-Fucose is shown as sticks with yellow, numbered carbons. (E) Stereoview of the interactions of L-fucose with BmulJ_04919. Protein atoms, L-fucose, and $NADP^+$ are shown as sticks with green, yellow, and white carbons, respectively.

L-Fucose Binding Site of BmulJ_04919. Clear and continuous electron density was observed for L-fucose (**Figure 3.9D**). The unliganded structure superimposes well with the ternary complex with an rmsd of 0.22 Å for 247 common Cα atoms excluding residues 192–202, which are either missing or more distant from the active site in the unliganded structure (see below). The binding site for L-fucose utilizes structural elements from the C-terminal region. L-Fucose binds in a C3-endo conformation and is highly coordinated, with each hydroxyl forming at least two hydrogen bonds to protein atoms (**Figure 3.9E**). L-Fucose is bound with C-3 positioned directly against the nicotinamide (3.3 Å) for the 4-*pro*-*S* hydride transfer typical of "classical" SDR enzymes[84]. In addition, the protein relay connecting the acid/base hydroxyL-tyrosinate (Tyr-163) to a lysine (Lys-157) to a conserved water molecule stabilized by Asn-113, again typical of classical SDR enzymes, is observed[84]. Determinants that are unique to L-fucose binding include Asn-94 from the β4−α4 loop coordinating the hydroxyl at C-4, Gln-147 from the α5−α6 loop coordinating the hydroxyl groups at positions C-3 and C-4, the backbone carbonyl of Ala-184 and the side chain of Glu-185 (β6−α7 loop) coordinating the hydroxyl groups at positions C-2 and C-3, respectively, and Lys-141 from the N-terminal end of α5 coordinating both hydroxyl groups at C-2 and C-3. In addition, the side chains of Asn-94, Leu-190, Tyr-191, and Trp-194 form a hydrophobic pocket that interacts with the methylene group of L-fucose. Indeed, the largest structural change from the apo structure to the ternary complex is the movement of the α7−α8 loop to optimize these interactions. In the apo structure, this region either is

disordered or is some 3–4 Å more distant from the L-fucose binding site. As α7 also interacts with the $NADP^+$ and most SDR reaction mechanisms proceed through an ordered "bi-bi" mechanism [$NAD(P)^+$binding first and leaving last], one can envision $NADP^+$ binding initiating the stabilization of the α7−α8 loop in a conformation competent for binding L-fucose, followed by active site closure upon the formation of the numerous interactions with the substrate. On the basis of the number and specificity of the interactions, the structure of the ternary complex suggests that BmulJ_04919 would be fairly specific for L-fucose and that these binding determinants could be utilized to model the substrate specificity of homologues of BmulJ_04919.

Discussion

Metal Content. LigI from *S. paucimobilis* was the first enzyme from cog3618 to be mechanistically and structurally characterized, and this enzyme does not require the binding of a divalent metal ion in the active site for catalytic activity.[78] BmulJ_04915, also a member of cog3618, shares ~20% identical sequence with LigI, and it is the second example of an enzyme within the AHS that does not require metal for catalytic function. We were unable to detect the binding of metals via ICP-MS in the samples of BmulJ_04915 purified in the absence of added metal, and thus, this protein does not require the metal for catalytic activity.

Active Site of BmulJ_04915. Nearly all members of the AHS are recognized by the conservation of five amino acid residue side chains that form the active site and are utilized for metal binding and/or proton transfers. These amino acid residues include two histidine residues from the C-terminal end of β-strand 1 (HxH motif), two additional histidine residues from the C-terminal ends of β-strand 5 and β-strand 6, and a conserved aspartate residue from the C-terminal end of β-strand 8.[77] In BmulJ_04915, three of the four histidine residues are fully conserved (His-9, His-11, and His-166) from the C-terminal ends of β-strand 1 and β-strand 6, in addition to the aspartate residue (Asp-244) from the C-terminal end of β-strand 8. The histidine from the end of β-strand 5 is absent and replaced with Leu-144, also at the end of β-strand 5 (**Figures 3.4B** and **3.10A**). In the absence of a metal ion bound in the active site, these four conserved residues must now acquire new functional roles.

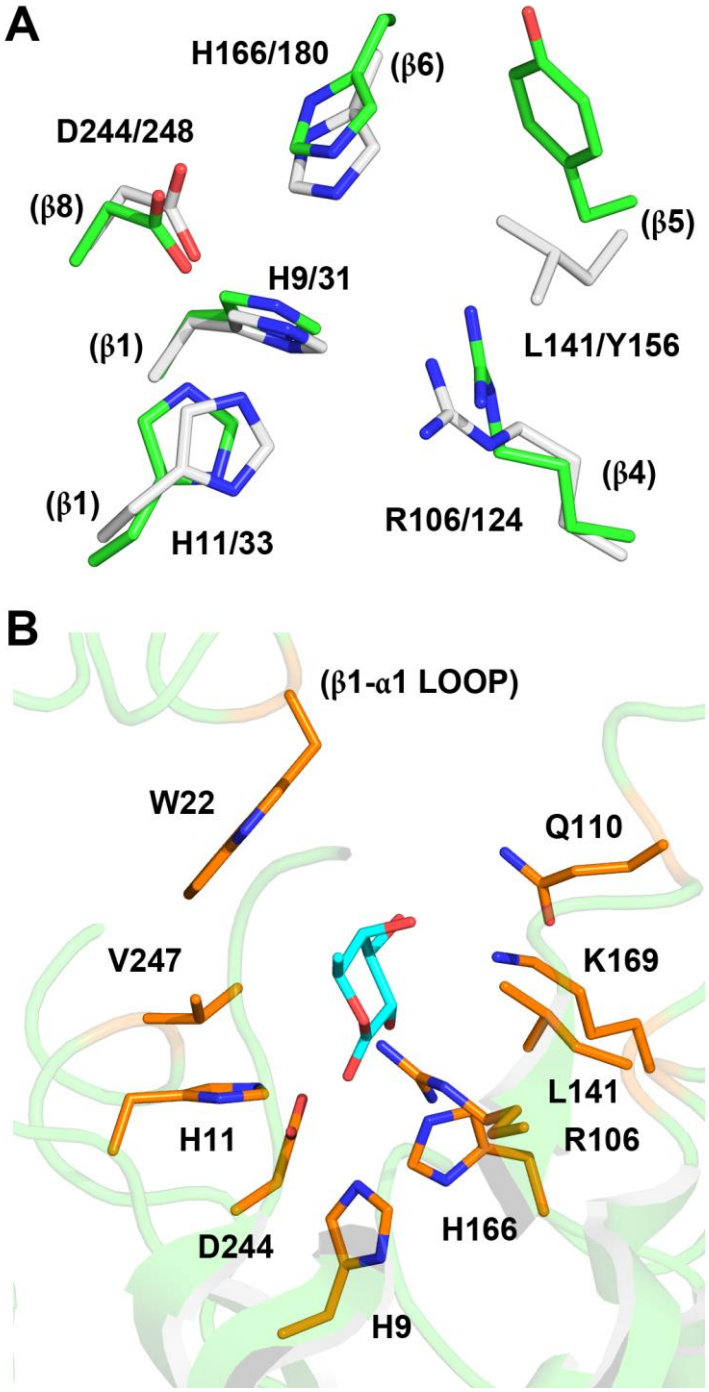


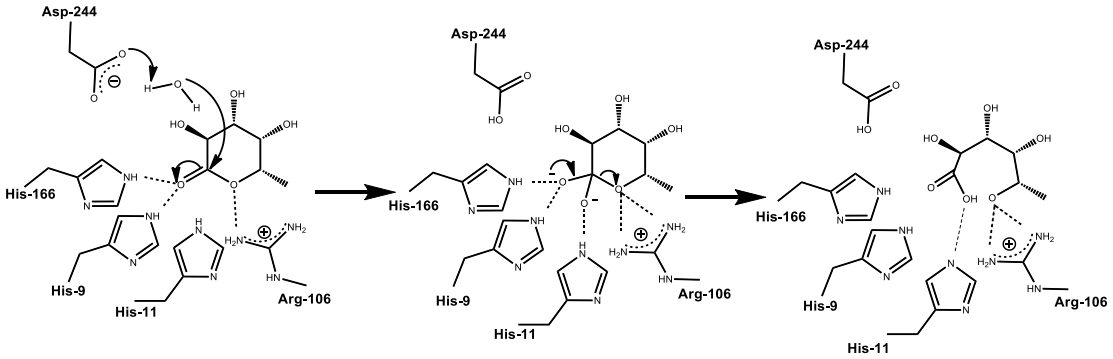
Figure 3.10. (A) Active site of BmulJ_04915 structurally aligned with that of 2-pyrone-4,6-dicarboxylic acid lactonase (LigI). The active site is color-coded as follows: green for LigI (PDB entry 4d8l) and white for BmulJ_04915. Numbers in parentheses designate the approximate locations in the secondary structure from which the residue originates. (B) L-Fucono-1,5-lactone modeled into the active site of BmulJ_04915.

A structural alignment of the LigI and BmulJ_04915 active sites reveals relatively little conservation between the two binding pockets, aside from the residues predicted to confer catalytic activity (**Figure 3.10A**). A low level of conservation of the active site can be explained by analysis of the two significantly different substrates utilized by these enzymes. LigI catalyzes the hydrolysis of a planar dicarboxylated six-membered lactone, while BmulJ_04915 catalyzes the hydrolysis of a nonplanar six-membered lactone. The structure of BmulJ_04915 could not be obtained with bound substrate or product, but it is possible to manually dock the L-fucono-1,5-lactone substrate in the active site to obtain a view of the structural determinants for substrate binding (**Figure 3.10B**). Modeling L-fucono-1,5-lactone into the active site of BmulJ_04915 was based on the structural alignment with LigI (with bound product) and the general path of the bound HEPES molecule (**Figures 3.4B** and **3.10A**). The sulfonate of the HEPES molecule bound to BmulJ_04915 is positioned like the newly formed carboxylate of the product of the reaction catalyzed by LigI (4-carboxy-2-hydroxymuconate) interacting with three histidine residues (His-9, His-11, and His-166), Arg-106, and Asp-244 (**Figure 3.4B**). In this model, Gln-110 and Lys-169 are positioned to interact with hydroxyl groups at C-3 and C-4 of L-fucono-1,5-lactone. On either face of the lactone, Leu-141 and Leu-247 constrict the active site. Finally, Trp-22 lies against the aliphatic face of the lactone (C-3–C-4–C-5) near the entrance to the active site.

Proposed Mechanism of Action. The four AHS conserved residues (His-9, His-11, His-166, and Asp-244) align with their counterparts in LigI and assume similar roles in

catalysis. The three conserved histidine residues polarize the carbonyl group of the sugar lactone substrate via electrostatic interactions.[78] The aspartate residue from β-strand 8 (Asp-244) is positioned to deprotonate an active site water molecule for nucleophilic attack on the C-1 carbonyl group of the lactone substrate. A fifth active site residue conserved within many members of cog3618 is an arginine from β-strand 4 (Arg-106). This arginine is proposed to assist in the cleavage of the lactone functional group through electrostatic interactions with the hydroxyl leaving group of the product. The proposed mechanism is presented in **Scheme 3.3**.

Scheme 3.3



Natural Substrate of BmulJ_04915. BmulJ_04915 exhibits significant catalytic activity for a series of five-membered sugar lactones (**Table 3.1**) that share the same stereochemistry at C-2 and C-3 and has the highest activity for the L-fucono-1,4-lactone. The best substrate for BmulJ_04919 is β-L-fucose, which undergoes oxidation to L-fucono-1,5-lactone and subsequent nonenzymatic transformation toL-fucono-1,4-lactone. It has been postulated that the instability of D-galactono-1,5-lactone is due to an intermolecular rearrangement that occurs in a distorted chair conformation. The rearrangement is thought to arise from the axial hydroxyl group in C-4, which is positioned for nucleophilic attack on C-1.[85] A similar rearrangement is assumed for L-fucono-1,5-lactone and is fully consistent with the NMR data reported in this paper. The ^{1}H NMR spectra show a direct conversion of L-fucono-1,5-lactone to L-fucono-1,4-lactone without transformation to L-fuconate as an intermediate. The NMR spectra clearly show that BmulJ_04915 hydrolyzes L-fucono-1,5-lactone at a rate faster than that with L-fucono-1,4-lactone (**Figure 3.6**).

It was not possible to determine the kinetic constants for the hydrolysis of L-fucono-1,5-lactone because of chemical instability, and thus, a stable mimic was utilized. 4-Deoxy-L-fucono-1,5-lactone was synthesized enzymatically and provided a pyranosyl analogue that was more stable because of the absence of the axial hydroxyl group on C-4, which allowed the kinetic constants to be determined at pH 7.1. 4-Deoxy-L-fucono-1,5-lactone was shown to be hydrolyzed by BmulJ_04915 approximately 2 orders of magnitude faster than the corresponding L-fucono-1,4-lactone. Thus, even

though L-fucono-1,5-lactone is kinetically unstable, it appears to be the natural substrate of BmulJ_04915.

The closest homologue (~27% identical sequence) of known function to BmulJ_04915 is SKA58_03595 from *Sphingomonas* sp. SKA58, which has been experimentally annotated as an L-rhamnono-1,4-lactonase.[66] Kinetic constants for L-rhamnono-1,4-lactonase have not apparently been reported, nor has a complete substrate profile been established. L-Rhamnono-1,4-lactonase is adjacent to an L-rhamnose dehydrogenase from cog1028 that shares 30% sequence identity with BmulJ_04919. Given the sequence similarities, it is predicted that these homologues of BmulJ_04915 and BmulJ_04919 will act in a similar manner to produce a 1,5-lactone, which is subsequently hydrolyzed to the acid sugar.

CHAPTER IV

CONVERGENCE OF L-GALACTOSE AND D-GALACTURONATE METABOLISM IN HUMAN GUT MICROBIOTA

The preponderance of microorganisms inhabiting the human body have been found to reside within the gastrointestinal tract. These organisms, which compose one of the most dense and diverse populations, are collectively known as the gut microbiota[86]. The gut microbiota is of particular interest for the role it plays in human metabolism and health. While the human-microbiota relationship is mostly reported as commensal, there have been reports implicating the microbiota in disease, including inflammatory bowel disease and colitis.[87] One important commensal role is the microbiota degradation of indigestible plant polysaccharides; such as pectin, xylan, and cellulose.[88] *Bacteroides* species are prevalent in the microbiota, and have been implicated as responsible, in part, for the degradation of these polysaccharides. *Bacteroides* species, including *Bacteroides vulgatus*, are known to encode an unusually high number of glycoside hydrolyases and polysaccharide lyases. *B. vulgatus*, in particular, is thought to encode the largest number of enzymes targeting pectin degradation.[86, 89, 90]

The rare sugar, L-galactose, is perhaps most notable as an intermediate in the biosynthesis of ascorbic acid in higher plant species. In the L-galactose pathway (ascorbic acid biosynthesis), L-galactose is produced through a series of reactions, beginning with D-glucose-6-phosphate which is ultimately converted to GDP-D-mannose. GDP-D-mannose is than transformed to GDP-L-galactose through epimerization, followed by the removal of GDP, yielding L-galactose phosphate, which is dephosphorylated to L-galactose.[91] L-galactose was first isolated from flaxseed oil, in 1903 by Ernest Anderson, and it was later identified to be a component of certain pectins.[92] Since then, L-galactose has been identified as a component of other biopolymers, including agar, galactagen, agaropectin, and rhamnogalacturonan II (RG-II) .[93] RG-II is a structural component of plant cell walls, which makes up approximately 10% of the total pectin in the biosphere. 6-deoxy-L-galactose, or L-fucose, is a known component of RG-II. It has recently been discovered that L-galactose is naturally substituted for L-fucose in RG-II.[93] For example, it was determined that 24% of RG-II in red wine and 14% in carrots (*Daucus carota*) has substituted L-galactose.

We have determined the function of three enzymes in a pathway for the utilization of L-galactose in *B. vulgatus* ATCC 8482: BVU_0219, BVU_0220, and BVU_0222. BVU_0220 is a member of the amidohydrolase superfamily (AHS) of enzymes from cog3618. AHS members catalyze an elaborate reaction set, including the hydrolysis of amide, ester and phosphoester bonds, deamination, isomerization, hydration, and decarboxylation.[8, 9, 11, 13, 26] Superfamily members are well known to

conserve similar structural topology, namely a distorted $(\beta/\alpha)_8$ TIM barrel, housing zero to three divalent metal ions. [78] The AHS has been organized into 24 clusters of orthologous groups (COGs), by NCBI. Organization into different COGs is based solely on protein sequence similarity.[5, 14]

Cog3618 has been shown to primarily catalyze the hydrolysis of lactones including such enzymes as, 2-pyrone-4,6-dicarboxylate lactonase (LigI), L-fucono-1,5-lactonase, 4-sulfomuconolactonase (4-SML), and L-rhamnono-1,4-lactonase.[65, 94] A sequence similarity network for cog3618 is provided at a relatively low BLAST expect (E) value 10^{-30} in **Figure 4.1**. Each node within the sequence similarity network represents a protein member of cog3618, which are connected to each other by a given sequence relationship. Here the sequence relationship is the BLAST E value of 10^{-30}. At this low stringency E value, cog3618 can be divided into two main groups, arbitrarily designated as Class I and Class II. As previously reported, many of the enzymes of Class I are encoded adjacent to genes with apparent relation to carbohydrate metabolism[94]. Also, it was predicted that enzymes within Class I will hydrolyze sugar lactone substrates.

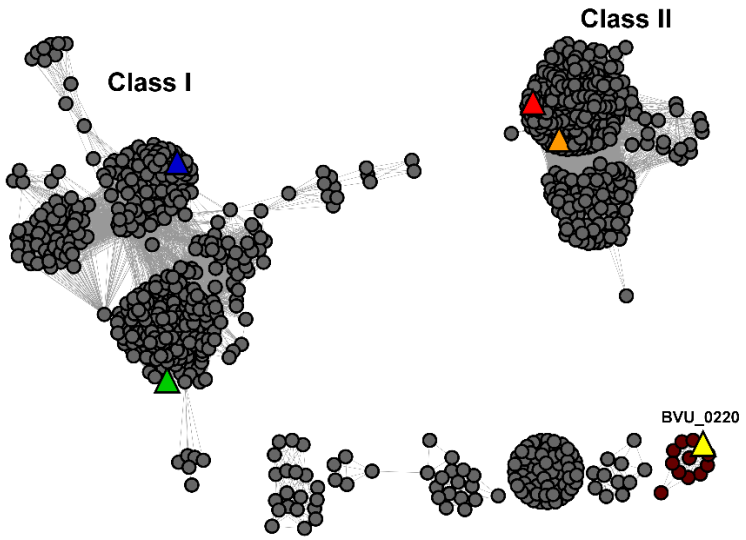


Figure 4.1. Cog3618 sequence similarity network at an E value of 10^{-30} where each node represents a protein and an edge represents an E value between two proteins of 10−30 or smaller. The triangle nodes represent an enzyme with experimentally verified function. Nodes are color coded as follows: LigI (red), 4SML (orange), L-rhamnonolactonase (blue), green (L-fucono-1,5-lactonase, and yellow BVU_0220.

At this given E value, BVU_0220 does not reside within either Class I or Class II. Nevertheless, the genomic neighborhood strongly suggests this protein will hydrolyze carbohydrate-derived lactones (**Figure 4.2**). The adjacently BVU_0219, encodes a putative oxidoreductase, which shares 30% sequence identity with L-fucose from (XCC4065) from *Xanthomonas campestris.*[95] BVU_0220 is a putative Zn^{2+} dependent dehydrogenase, which is distantly related to D-idonate dehydrogenaset[96]. Additionally, BVU_0217 and BVU_0221, encode a putative glycoside hydrolase and sugar permease, respectively. In this paper, we have identified a pathway for the utilization of L-galactose, which is postulated to funnel into uronate degradation through the reaction of BVU_0222. BVU_0222 catalyzes the oxidation of L-galactonate into the uronate degradation intermediate, D-tagaturonate. [42]

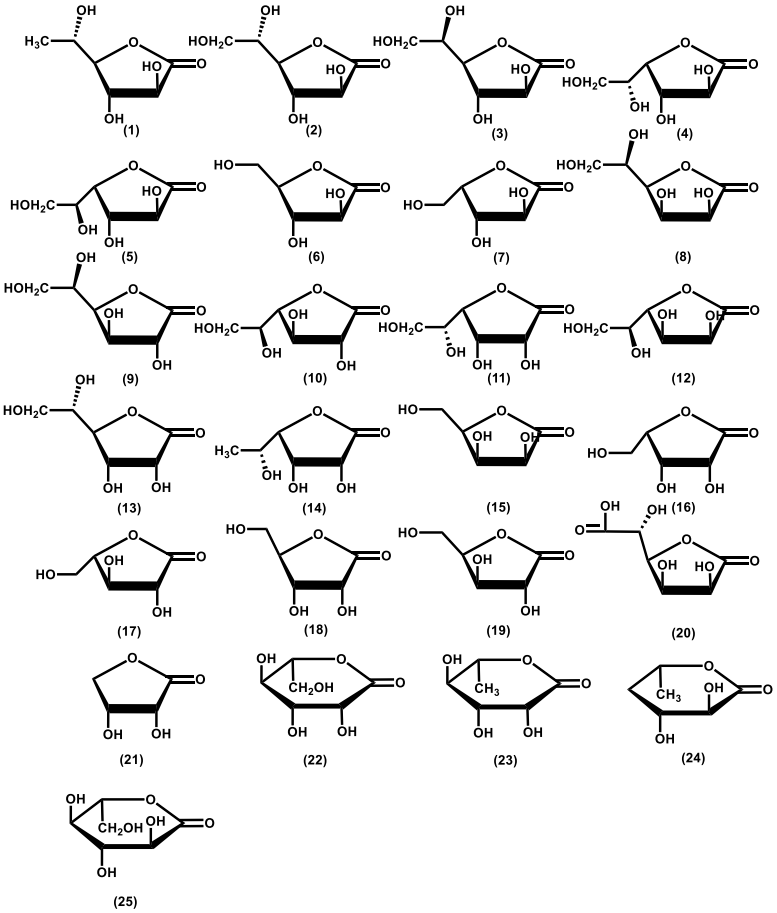
Figure 4.2. Genomic neighborhood of BVU_0220. Genes colored in maroon are the genes annoated by this study and are as follows: BVU_0219 is L-galactose dehydrognase, BVU_0220 is L-galactono-1,5-lactonase, and BVU_0222 is L-galactonate dehydrogenase. The genes in white are of unknown function. BVU_0217 is a puatative glcoside hydrolase, BVU_0218 is a putative transcription regulator, and BVU_0221 is a putative sugar permease.

Materials and Methods

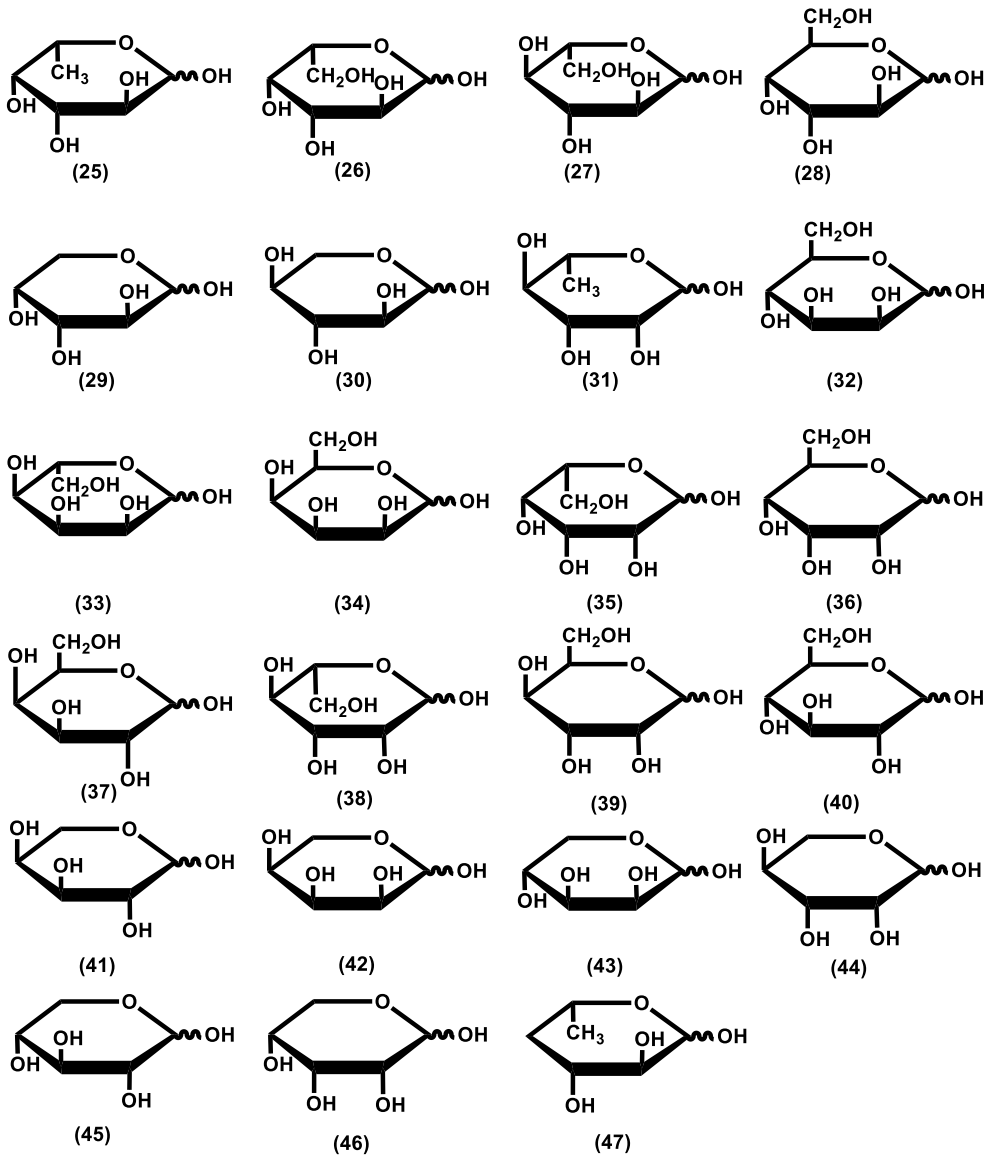
Materials. All chemicals and buffers were purchased from Sigma Aldrich unless otherwise specified. Sugar lactones that were not commercially available were synthesized according to published procedures with the exception of 4-deoxy-L-fucono-1,5-lactone, which was enzymatically synthesized.(12) The noncommercial lactones included the following: L-fucono-1,4-lactone (**1**), D-altrono-1,4-lactone (**3**),D-arabinono-1,4-lactone (**6**), L-xylono-1,4-lactone (**7**), L-mannono-1,4-lactone (**11**), D-talono-1,4-lactone (**12**), D-allono-1,4-lactone (**13**), L-rhamnono-1,4-lactone (**14**), D-lyxono-1,4-lactone (**15**),L-lyxono-1,4-lactone (**16**), L-arabinono-1,4-lactone (**17**), D-xylono-1,4-lactone (**19**), L-mannono-1,5-lactone (**22**), L-rhamnono-1,5-lactone (**23**), and 4-deoxy-L-fucono-1,5-lactone (**24**). Those sugar lactones that were available commercially included the following: L-galactono-1,4-lactone (**2**),L-glucono-1,4-lactone (**4**), D-idono-1,4-lactone (**5**), D-mannono-1,4-lactone (**8**), D-glucono-1,4-lactone (**9**), D-galactono-1,4-lactone (**10**), D-ribono-1,4-lactone (**18**), D-glucurono-6,3-lactone (**20**), D-erythronolactone (**21**), and L-glucono-1,5-lactone (**25**). These compounds were obtained from CarboSynth or ChromaDex. The structures of these lactones are

presented in **Scheme 4.1**. The aldose sugars were obtained from either from Sigma Aldrich or Carbosynth, and the structures are presented in **Scheme 4.2**. These sugars included L-fucose (**25**), L-galactose (**26**), L-glucose (**27**), D-altrose (**28**), D-arabinose (**29**), L-xylose (**30**), L-rhamnose (**31**), D-mannose (**32**), L-allose (**33**), D-talose (**34**), L-talose (**35**), D-allose (**36**), D-galactose (**37**), L-mannose (**38**), D-gulose (**39**), D-glucose (**40**), L-arabinose (**41**), L-ribose (**42**), D-lyxose (**43**), L-lyxose (**44**), D-xylose (**45**), D-ribose (**46**), and 4-deoxy-L-fucose (**47**).

Scheme 4.1. Lactone Substrate Library



Scheme 4.2: Sugar Substrate Library



Cloning, Expression, and Purification of BVU_0220. Original cloning, overexpression, and purification of BVU_0220 were completed through collaboration with Dr. Steve Almo and Ronald Seidel from the Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York.

Cloning, Expression, and Purification of BVU_0219, BVU_0222, L-Fucose Dehydrogenase and URI. The gene for BVU_0219 was cloned from *Bacteroides vulgatus* ATCC 8482 (GI|150002825) with the primer pair 5'-ATGAATTACAATGAAATGGGAAAACCGGTATGCGGGTT-3' and 5'-TGAATTAGCCCAACTGATACGCATTTGTTTTCCGATAA-3'. Restriction sites for Nhe1 and XHO1 were inserted into the forward and reverse primers. The PCR product was purified with a Promega PCR cleanup system, and subsequently digested with Nhe1 and HindIII. The resulting product was ligated into pET-24a(+) vector (Novagen). The ligation reaction was then transformed into BL21(DE3) cells via electroporation. A single colony containing the plasmid of interest was used to inoculate 5-mL overnight cultures of LB (50 µg/mL kanamycin), incubated at 37 °C. 1 L LB cultures (50 µg/mL kanamycin) inoculated with 5-mL overnight culture was grown at 37 °C until an OD_{600} of 0.4-0.6 was observed. The cultures were transferred to ambient temperature and gene induction was initiated with the addition of 1.0 mM isopropyl β-thiogalactoside (IPTG), for 18 hours. Cells were harvested via centrifugation at 4 °C and suspended in 50mM HEPES (pH 7.5), 0.25 mM NaCl, and 0.1 mg/mL of phenylmethanesulfonyl fluoride (PMSF). Resuspended cells were lysed via multiple rounds of sonication, cell debris was

removed via centrifugation at 4 °C. Nucleic acid was removed from the supernatant with the dropwise addition of 2% (w/v) of protamine sulfate at 4 °C. Protamine sulfate-bound DNA was removed by centrifugation. Ammonium sulfate was added between 30-50 % saturation and the precipitated protein was isolated by centrifugation. The protein was resuspended of 20 mM HEPES (pH 7.5) and then applied to a High Load 26/60 Superdex 200 gel filitration column (GE Healthcare). Fractions containing the protein of expected size were pooled and purified further by ion exchange chromatography with a ResourceQ column (6 mL) at pH 7.5.

The gene coding for BVU_0222 was determined to contain 19 rare codons using the Rare Codon Calculator (RaCC) (nihserver.mbi.ucla.edu/RACC/)[97]. A codon optimized gene was purchased from GenScript, which was ligated into a pUC57 cloning vector containing Nde1 and HindIII restriction sites. The plasmid harboring the gene of interest was digested with Nde1 and HindIII at 37 °C for three hours. The resulting gene was ligated into pET-32a (+) (Novagen) and transformed into BL21(DE3) cells via electroporation. Subsequent purification procedures were followed as described for BVU_0219.

Cloning and overexpression of *E. coli* uronate isomerase (URI) and *Burkholderia multivorans* L-fucose dehydrogenase (BmulJ_04915), were performed as previously described [42, 94]. Purification was conducted as described for BVU_0219, with the exception of the addition of 1mM $ZnCl_2$ to the purification buffer of URI.

Measurement of Enzyme Activity. The enzymatic hydrolysis of lactones by BVU_0220 was monitored with a SpectraMax340 UV–visible spectrophotometer using a colorimetric pH indicator assay at 30 °C, as previously described[94]. Protons released from the carboxylate product during lactone hydrolysis were measured using the pH indicator bromothymol blue. Reaction mixtures measured with bromothymol blue contained 2.5 mM MOPS buffer (pH 7.1), 0.2 M NaCl, 0–0.5 mM lactone, and 0.1 mM bromothymol blue. The final concentration of DMSO was 1%. Changes in absorbance at 616 nm (ε = 1135 M^{-1} cm^{-1}) were monitored in 96-well plates for bromothymol blue. Background rates arising from acidification by atmospheric CO_2 were observed and subtracted from the initial rates.

Dehydrogenase activity was monitored for the reduction of NAD(P) at 340 nM (ε = 6220 M^{-1} cm^{-1}) at 30 °C, with a SpectraMax340 UV–visible spectrophotometer. Reaction assays for BVU_0219 were performed in 50 mM BICINE buffer (pH 8.0), varying concentrations of substrate, and 0.5 mM NADP. Reaction assays for and BVU_0222 were similarly performed in 50 mM PHOSPHATE buffer at pH 7.0. Kinetic constants for NAD(P) were determined with 200 µM substrate and varying concentration of NAD(P) between 0-1.5mM.

Reductase activity of BVU_0222 was monitored through a coupled assay system. The coupling system consisted of 8.0 µM URI, D-galacturonate (0-4mM), 50mM PHOSPHATE buffer at pH 7.0, NADH (1.0mM), and BVU_0222. Reactions were allowed to incubate for 15 min in the absence of BVU_0222 at 30 °C. This allowed URI to

catalyze the isomerization of D-galacturonate to the desired D-tagaturonate. Reactions were initiated by the introduction of BVU_0222 and monitored through the oxidation of NADH at 340nM (ε = 6220 M^{-1} cm^{-1}) at 30 °C, with a SpectraMax340 UV–visible spectrophotometer.

Data Analysis. The kinetic constants were determined from a fit of the initial velocity data to eq 4.1 using curve fitting program in SigmaPlot 9.0

$$v/E_t = k_{cat}[A]/(K_m+[A]) \quad (4.1)$$

where v is the initial velocity, E_t is the total enzyme concentration, k_{cat} is the turnover number, [A] is the substrate concentration, and K_m is the Michaelis constant.

Metal Analysis. The metal content of BVU_0220 and BVU_0222 were determined using an Elan DRC II ICP-MS instrument as previously described (26). Protein samples for ICP-MS analysis were digested with HNO_3 and then refluxed for 30 min.(27) All buffers were passed through a column of Chelex 100 (Bio-Rad) to remove trace metal contamination. EDTA and 1,10- phenanthroline (1.0 mM) were incubated with 1.0 µM BVU_0220 in 50 mM buffer at various pH values from 6 to 10 to remove divalent metal ions, as previously described [78]. The following buffers were used: CHES (pH 6.0), HEPES (pH 7.0), BICINE (pH 8.0), and CHES (pH 9.0 and 10.0). The effect of added divalent cations on the catalytic activity of BVU_0220 was determined by adding Mn^{2+}, Zn^{2+}, Co^{2+}, Cu^{2+}, or Ni^{2+} (0–500 µM) directly to the assay mixtures. The purified

enzyme was also incubated with 50–500 molar equivalent of these divalent cations for 24 h at 4 °C in 50 mM HEPES (pH 7.5) and assayed for catalytic activity.

Reaction Product of BVU_0219 with L-galactose. The product of the reaction catalyzed by BVU_0219 was determined through NMR spectroscopy using a Bruker Avance III 500 MHz spectrometer with an H, C, N cryoprobe. ^{13}C-NMR was performed on reaction samples in 250 µL of 250 mM PHOSPHATE buffer pH 5.8 (10% D_2O), 10 mM NAD, 10 mM [^{13}C-1]-L-galactose (Omicron Biochemicals) and10-50 µM BVU_0219. After an initial 10 min incubation, the enzyme was removed with a 3 kDa cutoff VWR centrifugal filter. The pH of the sample was then adjusted to 7 to calculate the non-enzymatic transformation of L-galactono-1,5-lactone to the thermodynamically stable L-galactono-1,4-lactone.

Hydrolysis of L-galactono-1,5-lactone. L-galactono-1,5-lactone was identified as the preferred substrate for BVU_0220 through ^{13}C-NMR spectroscopy. The reactions were conducted in 250 mM PHOSPHATE buffer (pH 5.8) (10% D_2O), 10 mM NAD, 10 mM [^{13}C-1]-L-galactose, 50 µM BVU_0219. The reaction was monitored through ^{13}C-NMR until the resonances for both lactones was approximately equal. At this point 1 µM BVU_0220 was added and the spectra collected as a function of time.

Enzymatic Synthesis of L-gluconate and L-galactonate. Acid sugars of both L-galactose and L-glucose were synthesized enzymatically from L-galactono-1,4-lactone () and L-glucono-1,5-lactone () using BVU_0220 as the catalyst. Reaction conditions were

as follows: 50 mM lactone, 50 mM CARBONATE buffer at pH 9.5, and 20 μM BVU_0220 incubated overnight.

Enzymatic Synthesis of 4-Deoxy-L-fucono-1,5-lactone. 4-Deoxy-L-fucono-1,5-lactone (**24**) was synthesized enzymatically from 4-deoxy-L-fucose (**47**) using L-fucose dehydrogenase (BmulJ_04919) as the catalyst, as previously described.[94] The reaction was conducted in 50 mM phosphate buffer (pH 7.0), 15 mM 4-deoxy-L-fucose, and 15 mM NAD^+ in a final volume of 1 mL for 20 min.

Sequence Similarity Network for cog3618. Proteins belonging to cog3618 were identified from the NCBI protein database using the query "cog3618". The proteins within cog3618 were subjected to an alL-by-all BLAST at a specified *E*value (10^{-50}, 10^{-60}, etc.) using the NCBI standalone BLAST program. The BLAST files were opened and visualized in the similarity network program Cytoscape.[67]

Results

Purification of BVU_0220. The gene encoding BVU_0220 from *B.vulgatus* was cloned and overexpressed in *E. coli* (BL21(DE3)), and the protein was purified to homogeneity. Purified BVU_0220 contained ~ 0.3 molar equivalents of Mn^{2+}, 0.1 molar equivalents of Fe^{2+}, and 0.3 molar equivalents of Ni^{2+}. This metal was shown to not be required for catalytic activity through the addition of chelating agents and various first row divalent transition metals (Mn^{2+}, Zn^{2+}, Co^{2+}, Cu^{2+}, or Ni^{2}), which were added directly to the assay. The resulting reaction exhibited no significant change in the rate of lactone hydrolysis. Nor was a significant rate change observed when BVU_0220 was

incubated overnight with EDTA and 1,10-phenanthroline at pH values 6-10. From this it is apparent that BVU_0220 does not require divalent cation for activity, similar to other members of cog3618.[94]

Other members of cog3618 have been shown to hydrolyze 5- and 6-member (γ- and δ-) lactones including:L-fucono-1,5-lactone, L-rhamnono-1,4-lactone, and 2-pyrone-4,6-dicarboxylate. BVU_0220 shares approximately 28% and 15% sequence identity with L-fucono-1,5-lactonase and LigI, respectively. Even with relatively low sequence identity to other members, BVU_0220 conserves important catalytic active site residues (**Figure 4.3**). These include the HxH from β-strand 1, histidine from β-strand 6, arginine from β-strand 4, and aspartate from β-strand 8. In addition, the genomic neighborhood of BVU_0220 encodes genes indicative of carbohydrate metabolism; such as, a putative glycoside hydrolase and putative sugar permeases. Therefore, it was predicated that this enzyme would hydrolyze sugar lactones.

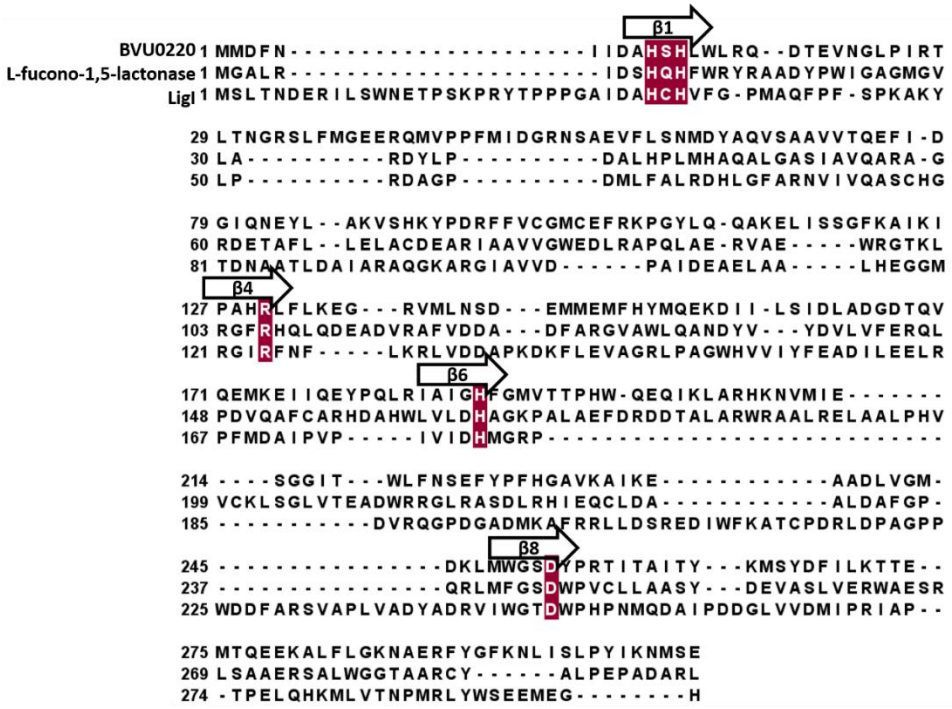


Figure 4.3. Sequence alignments of BVU_0220, LigI (GI|374074685), and BmulJ_04915 (L-fuconon-1,5-lactonase. Residues highlighted in blue are conserved active site residues important for catalysis. Black arrows represent β-strands that make up the $(\beta/\alpha)_8$-TIM barrel.

BVU_0220 was screened for activity against a library of the available 5- and 6-member sugar and sugar lactone derivatives (**Scheme 4.1**). It was found that, BVU_0220, was able to hydrolyze L-glucono-1,5-lactone (**25**), L-galactono-1,4-lactone (**2**), and L-fucono-1,4-lactone (**1**). Similar activity was found in all three substrates; however, L-glucono-1,5-lactone was found to be hydrolyzed slightly faster (**Table 4.1**).

Table 4.1. Catalytic Constants for BVU_0220 at pH 7.1

compound	k_{cat}**(s^{-1})**	K_m**(mM)**	k_{cat}/K_m **(M^{-1}s^{-1})**
L-glucono-1,5-lactone (25)	38 ± 1	0.5 ± 0.04	(8.0 ± 0.5) x 10^4
L-galactono-1,4-lactone (2)	34 ± 3	1.9 ± 0.3	(1.7 ± 0.2) x 10^4
L-fucono-1,4-lactone (1)	23 ± 2	1.3 ± 0.2	(1.8 ± 0.2) x 10^4
4-deoxy-L-fucono-1,4-lactone (24)	23 ± 1	0.05 ± 0.008	(7.0 ± 0.8) x 10^5

Purification and Properties of BVU_0219. The gene for BVU_0219 is encoded adjacent to BVU_0220 and is annotated as an oxidoreductase from cog0667. Cog0667 contains dehydrogenases; such as, L-fucose dehydrogenase and D-arabinose dehydrogenase. This enzyme also shares significant sequence identity to L-galactose dehydrogenases from two plants species; ornamental tobacco (*Nicotiana langsdorffii x Nicotiana sandera*) and *Arabidopsis thaliana* with 46 and 42% sequence identity, respectively. Based on this, it was predicted that BVU_0219 would oxidize a monosaccharide, with the initial reaction product being the physiological substrate for BVU_0220.

BVU_0219 was screened against a substrate library of commercially available pentose and hexose sugars. Initial screens were monitored spectrophotometrically by following the reduction of NAD(P) at 340 nm (**Scheme 4.2**). High catalytic specificity was displayed for L-galactose, using NADP as a redox cofactor (**Table 4.2**).

Table 4.2: Catalytic constants for BVU_0219 at pH 8.0 and BVU_0222 at pH 7.0

Enzyme	Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
BVU_0219				
	L-galactose (**26**)	21 ± 0.3	109 ± 4	(2.0 ± 0.1) x 10^5
	L-glucose (**27**)	N/D	N/D	(2.0 ± 0.2) x 10^2
	D-arabinose (**29**)	N/D	N/D	(70.5 ± 0.1) x 10^1
	L-fucose (**25**)	N/D	N/D	(34 ± 0.2) x 10^1
	D-altrose (**28**)	N/D	N/D	(13 ± 0.1) x 10^1
	$NADP^+$	22 ± 1	3.0 ± 0.5	(8.0 ± 0.4) x 10^6
	NAD^+	25 ± 1	3000 ± 200	(9.0 ± 0.1) x 10^3
BVU_0222				
	L-galactonate	0.6 ± 0.1	36 ± 3	(2.0 ± 0.12) x 10^4
	L-gluconate	0.2 ± 0.1	510 ± 70	(5.0 ± 0.6) x 10^2
	D-tagaturonate [a]	90 ± 0.6	550 ± 50	(1.6 ± 0.2) x 10^5

[a] Reverse reaction using URI coupling system
N/D- Not determined

Purification and Properties of BVU_0222. The gene for BVU_0222 is encoded adjacent to both BVU_0220 and BVU_0219, and is annotated as a Zn^{2+} dependent oxidoreductase from cog1063. Members of cog1063 include the following: L-threonine dehydrogenase, L-idonate dehydrogenase, and sorbitol dehydrogenase. BVU_02220

was determined to contain approximately 1 molar equivalent of Zn^{2+} through ICP-MS analysis. The only homologue with verified function is LgnH (GI| 425703034) from *Paracoccus sp. 43P*, which shares 46% sequence identity. This protein is reported to be encoded in an operon for the catabolism of the unnatural L-glucose. LgnH reportedly oxidizes both L-gluconate and L-galactonate, with approximately the same rate[98].

BVU_0222 was found to catalyze the oxidation of the C-5 hydroxyl group in both L-galactonate and L-gluconate. The oxidation of L-galactonate was found to be catalyzed with a k_{cat}/K_m of two orders of magnitude greater than L-gluconate (**Table 4.2)**. The oxidation of L-galactonate at C-5 yields D-tagaturonate, which was confirmed through the URI coupled assay. URI catalyzes the isomerization of D-galacturonate to D-tagaturonate, which was utilized as a substrate for the reductase reaction of BVU_0222, which yields L-galactonate.

Identification of the Initial Reaction Product of BVU_0219. Previous reports have suggested that the reaction product for the oxidation of L-galactose is L-galactono-1,4-lactone (**2**) [99-102]. Most reports of L-galactose dehydrogenase activity have been reported in plant species, where this reaction is important for the biosynthesis of ascorbic acid[91]. However, it is likely that the initial product is L-galactono-1,5-lactone, but has never been observed due to the chemical instability and rearrangement to L-galactono-1,4-lactone. To help stabilize product formation, the enzymatic oxidation was conducted at pH 5.8, as previously reported with L-fucono-1,5-lactonase[94]. The initial product of the reaction was identified through ^{13}C NMR as L-galactono-1,5-

lactone. L-galactose labeled at position C-1 was utilized as a substrate to provide strong signal for ^{13}C NMR, with few scans. The initial resonance in the carbonyl region corresponds to the carbonyl at C-1 of L-galactono-1,5-lactone , which rearranges to L-galactono-1,4-lactone with resonance at 175.9 ppm.

To determine if lactone rearrangement from the 1,5 to 1,4 was nonenzymatic; enzyme, NAD, and [^{13}C-1]-L-galactose were allowed to react at pH 5.8. The reaction was monitored via ^{13}C-NMR until L-galactono-1,5-lactone was the predominate species (174.1 ppm). The enzyme was then removed and the pH was adjusted to 7. Spectra were then collected as a function of time. The time course for the subsequent change in the C-1 of L-galactose and the reaction products are presented in **Figure 4.4B**. At 174.1 ppm and 176.0 ppm are the carbonyl resonances for L-galactono-1,5-lactone and L-galactono-1,4-lacone , respectively (**Figure 4.4B**). The initial dominant species in the reaction is clearly L-galactono-1,5-lactone, which overtime (10 min) is converted nonenzymatically to L-galactono-1,4-lactone. This nonenzymatic conversion does not occur through an acid sugar intermediate, consistent with an intramolecular rearrangement. The nonenzymatic rate of conversion was determined to be 0.2 min^{-1} ($t_{1/2}$=3min).

The ^{13}C-NMR spectrum for [^{13}C-1]-L-galactose is provided in **Figure 4.4A**. The four resonances correspond to C-1 in each of the four anomers of L-galactose. The chemical shifts of the α- and β-pyranose are 96.5 ppm and 92.3 ppm, respectively. The α- and β-furanose anomers are found at 95.0 ppm and 101.0 ppm, respectively. The

anomerization is consistent with data reported by Angyal and Pickles and is as follows: β-pyranose 64%, α-pyranose 29%, β-furanose 4% and α-pyranose 3% [103].

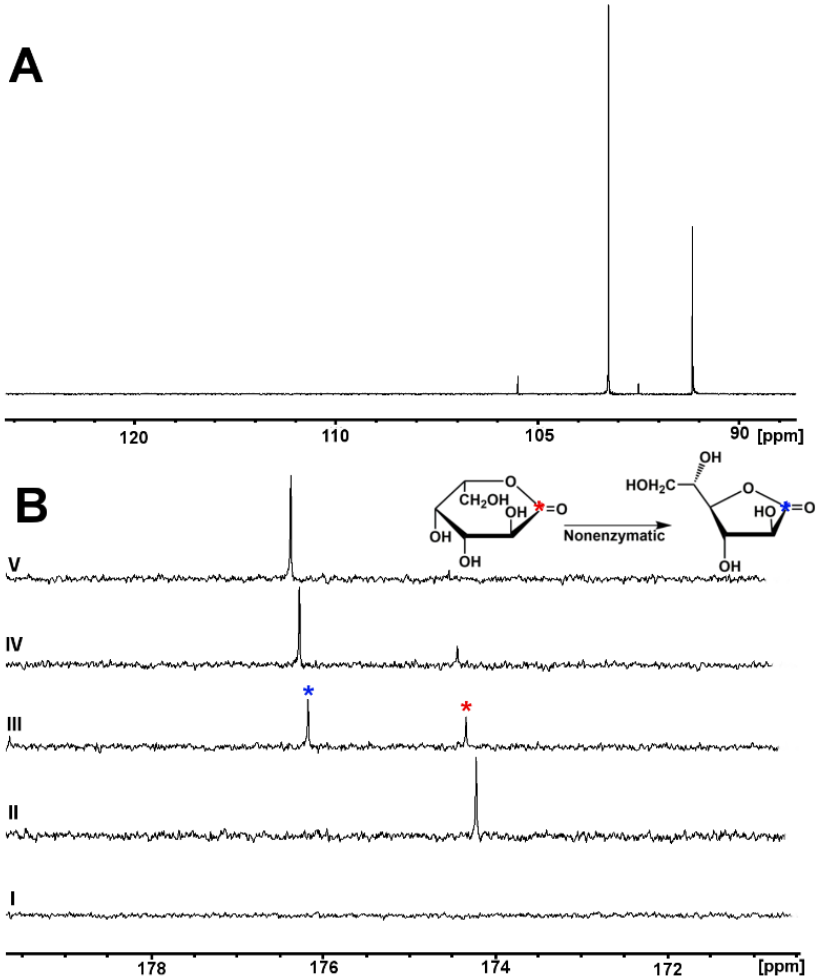


Figure 4.4: ^{13}C-NMR spectra for [^{13}C-1]-L-galactose and nonenzymatic transformation of L-galactono-1,5-lactone to L-galactono-1,4-lactone. (**A**) The four peaks correspond to the anomeric carbon for each of the four anomers of L-galactose. Chemical shifts are as follows: α-pyranose (92.3 ppm), β-pyranose (96.5 ppm), α-furanose (95.0 ppm) and β-furanose (101.0 ppm). (**B**) (I) The carbonyl region of L-galactose is presented at time 0. (II) BVU_0219 added with L-galactono-1,5-lactone as the predominate species (174.1 ppm). The enzyme was removed and the pH adjusted to 7. (III) The chemical shift for L-galactono-1,5-lactone decreasing as the chemical shift for L-galactono-1,4-lactone (175.9 ppm) is increasing in the absence of enzyme. (IV) After 10 min the predominante species is L-galactono-1,4-lactone. (V) After 15 min all of L-galctono-1,5-lactone has been converted to L-galactono-1,4-lactone.

Identification of the Physiological Substrate for BVU_0220. To determine if BVU_0220 can hydrolyze the initial product of L-galactose oxidation, BVU_0219 was incubated with [^{13}C-1]-L-galactose and NAD at pH 5.8. The reaction was monitored via ^{13}C NMR and allowed to proceed, at 15 min the resonances for both lactones was approximately equal (**Figure 4.5A**). At this point, BVU_0220 was added to the mixture and the spectra of the reaction was recorded as a function of time (**Figure 4.5B**). The resonance corresponding to L-galactono-1,5-lactone disappeared rapidly (5min) compared to the L-galctono-1,4-lactone. The resonance at 179.4 ppm corresponds to the carbonyl of the newly formed L-galactonate (**Figure 4.5C**). Similarly, when both enzymes (BVU_0219 and BVU_0220) were added together a single resonance was observed corresponding to L-galactonate (not shown). Rate constants for the hydrolysis of L-galactono-1,5-lactone and L-galactono-1,4-lactone were calculated via NMR to be 0.6 and 0.002 min^{-1}, respectively. This is a rate enhancement of approximately 300 fold when the 6-membered lactone is the substrate. The fastest rate observed was for 4-deoxy-L-fucono-1,5-lactone, further demonstrating the preference of this enzyme for 6-membered lactones.

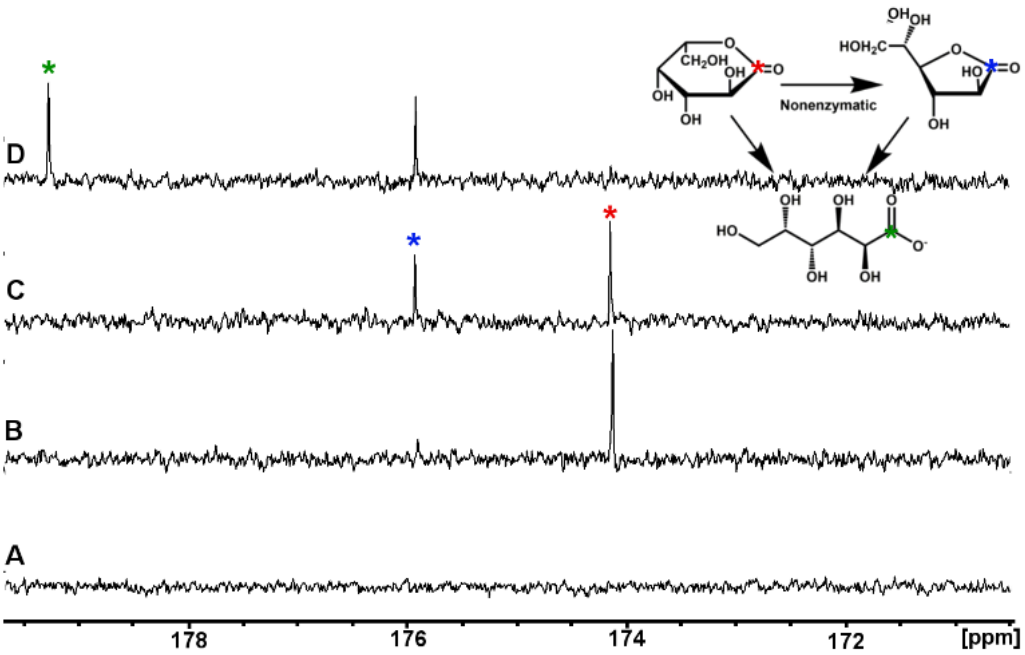


Figure 4.5: ^{13}C-NMR analysis of the reactions catalyzed by BVU_0219 and BVU_0220 utilizing [^{13}C-1]-L-galactose as substrate. (**A**) The carbonyl region of [^{13}C-1]-L-galactose. (**B**) The predominate species after the addition of BVU_0219 is L-galactono-1,5-lactone (174.1 ppm, red asterisk). (**C**) 10 min after the addition of BVU_0219 the resonances corresponding to C-1 of L-galactono-1,5-lactone and L-galactono-1,4-lactone (175.9 ppm, blue asterisk) are approximately equal. At this point BVU_0220 was added to the reaction mixture. (**D**) 5 min after the addition of BVU_0220, L-galactono-1,5-lactone has disappeared. A new peak with a chemical shift of 179.4 ppm appeared, which corresponds to L-galactonate (green asterisk).

Discussion

Metal Content. Previous members of cog3618, namely, L-fucono-1,5-lactonase and LigI have been shown to not require a divalent transition metal for catalysis[78, 94]. In addition, the enzyme 4-sulfomuconolactonase was reported to be inhibited in the presence of certain divalent metal cations[66]. BVU_0220, also a member a member of cog3618, is the third verified example of a member of the AHS that does not require a metal cation for catalytic function.

Natural Substrate of BVU_0220. BVU_0220 displays significant activity for a series of 5 and 6-membered lactones (**Table 4.1**), which share similar stereochemistry at C-2 and C-3. However, the best substrates are 6-member L-glucono-1,5-lactone and 4-deoxy-L-fucono-1,5-lactone. 4-deoxy-L-fucono-1,5-lactone is hydrolyzed with a k_{cat}/K_m an order of magnitude greater than L-glucono-1,5-lactone. The preference for 4-deoxy-L-fucono-1,5-lactone is likely due to the carbon at position C-4, which is absent in this lactone. The stereochemistry of L-glucono-1,5-lactone and the natural L-galactono-1,5-lactone at C-4 is opposite. BVU_0219 exhibits high substrate specificity for L-galactose, which is oxidized to L-galactono-1,5-lactone and rapidly converts to the thermodynamically stable 1,4-lactone. We have previously demonstrated that the chemical instability of L-fucono-1,5-lactone was due to an intramolecular rearrangement. The rearrangement is thought to occur through the axial hydroxyl group at C-4 in a distorted chair conformation[94]. The data presented herein is also consistent with an intramolecular rearrangement, as the conversion from L-galactono-

1,5-lactone to L-galactono-1,4-lactone does not proceed through an acid sugar (open chain) intermediate (**Figure 4.4B**).

Proper steady-state enzyme kinetics were not possible to determine for the hydrolysis of L-galactono-1,5-lactone, due to chemical instability. ^{13}C-NMR derived rate constants allowed for an estimate of the relative rates of hydrolysis for each lactone (**Figure 4.4B**). It was estimated that BVU_0220 can hydrolyze L-galactono-1,5-lactone at a rate at least 300 times that for L-galactono-1,4-lactone. This is the second example from cog3618, which preferentially hydrolyzes the kinetically unstable 6-member lactone. Therefore, BVU_0220, has been designated as L-galactono-1,5-lactonase.

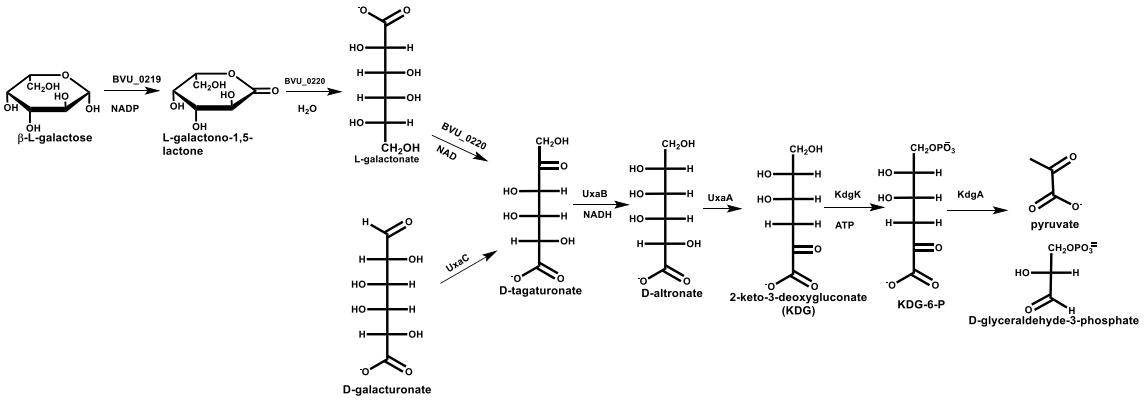
Convergence of L-galacose and Uronate Metabolism. It is well known that plant species oxidize L-galactose to L-galcatono-1,4-lactone (through the intermediate L-galactono-1,5-lactone) towards the biosynthesis of ascorbic acid[91]. However, a pathway for the utilization of L-galactose has yet to be identified. It has previously been proposed that *E. coli* contain a metabolic pathway for L-galactonate, which funnels D-tagaturonate into uronate metabolism[42]. This is a likely scenario for the utilization of L-galactose in *B. vulgatus*. Neither BVU_0219 (L-galactose dehydrogenase) or BVU_0220 (L-galactono-1,5-lactonase) are found within the genome of *E. coli*; however, a homologue of BVU_0222 (L-galactonate dehydrogenase) sharing 49% sequence identify is conserved. The relative high sequence identity suggests this protein could likely catalyze the same reaction in *E. coli.*

The uronate degradation pathway, as previously described, is provided in **Scheme 4.3**. It is likely that the pathway described here for L-galactose converges with this pathway at UxaB (D-tagaturonate reductase), which catalyzes the reduction of D-tagaturonate to D-altronate. BVU_0222 (L-galactonate dehydrogenase) was shown in this work to catalyze the oxidation of L-galactonate to D-tagaturonate, and conversely the reduction of D-tagaturonate to L-galactonate (Note: changing of carbon numbering and stereochemistry from L-sugar to D-sugar). This was shown through the utilization of the uronate degradation enzyme, uronate isomerase (URI) from *E. coli.* Sequence homology searches have found homologues of all uronate degrading enzymes within *B. vulgatus* (**Table 4.3**). Which provides further evidence for the convergence of these pathways.

Table 4.3: Homologous Uronate Degradation Enzymes in *B. vulgatus*

Organism	Gene	Function	Identity	*B. vulgatus homologue*
E. coli	UxaC	uronate isomerase (URI)	57%	BVU_3074
E. coli	UxaB	tagaturonate reductase	46%	BVU_3075
E. coli	UxaA	altronate dehydratase	46%	BVU_3053
Thermotoga maritima	TM0067	2-dehydro-3-deoxyglucokinase (KdgK)	53%	BVU_3055
E. coli	EDA	2-dehydro-3-deoxy-phosphogluconate aldolase (KdgA)	33%	BVU_3056

Scheme 4.3: Convergence of L-galactose and Uronate Metabolism



CHAPTER V

SUMMARY AND CONCLUSIONS

The 'post-genomic' age of scientific research as brought about, not only literally millions of available protein sequences, but also a new point of congestion in terms of identifying and annotating new enzymatic transformations. Recent years have seen the implementation interdisciplinary approaches, aptly termed, genomic enzymology, which has proven to be very valuable in the efforts of enzyme function discovery. The backbone of genomic enzymology is the analysis of protein sequence and enzymatic activity as a function of evolutionarily relatedness, or homology. This dissertation has explored efforts in genomic enzymology towards the discovery of novel functions within a group of evolutionarily related proteins within the AHS. Specifically, the efforts towards the discovery of new functions with cog3618 have been described.

Chapter I provided an introduction for the following chapters, in addition to a detailed analysis of the diverse active sites of the AHS. A superfamily is loosely defined as a group of evolutionarily related proteins conserving similar structural and mechanistic characteristics that share a common ancestral progenitor. The AHS is known to conserve a distorted $(\beta/\alpha)_8$-TIM barrel structural motif, which provides ligands for the coordination of metal cations. These metal centers, and the ligands orienting them in space, compose the framework of the active site. AHS member active sites can be classified according to the type of metal center; either metal

independent, mononuclear, dinuclear, or trinuclear. In this dissertation, metal centers were classified not only on the active site type, but on the metal coordinating ligands. This classification is known as 'subtype'. There are currently 14 active site subtypes for members of the AHS. The rest of this chapter was dedicated to providing detailed analysis of the active site types and subtypes that occur within each COG of the AHS. There are four dedicated mononuclear metal site COGs which includes the following: cog0402, cog1816, cog1904, and cog2159. Two additional COGs, cog1820 and cog1228, contain both mono and binuclear metal active sites. Seven COGs are dedicated binuclear metal COGS, including cog0044, cog0418, cog2355, cog1735, cog0804, cog3653, and cog 1001. More recently discovered are the trinuclear metal COGs, cog0613 and cog1387.

Chapter II describes the elucidation of the detailed catalytic mechanism of LigI, from cog3618. Elucidation of this chemical mechanism facilitated the discovery of LigI as the first member of the AHS which did not require a metal cofactor for catalysis. Metal independence was identified through X-ray crystallography, metal supplementation and chelation assays, and ICP-MS. Members of cog3618, including LigI, have been shown through sequence alignments to conserve active site residues, which have previously been implicated in metal coordination. These include two histidines of the HxH motif from the C-terminal end of β-strand 1, histidine from the end of β-strand 6, and aspartate from the end of β-strand 8. Histidine from β-strand 5 is not conserved among members of cog3618. The role these residues play in catalysis

was probed through site-directed mutagenesis. It is proposed that the three histidine residues are responsible for polarizing the carbonyl through electrostatic interactions, in addition to assisting the activation of the water molecule for attack. Aspartate from β-strand 8 is proposed to activate water for nucleophilic attack through proton abstraction.

Chapter III describes the discovery of a novel L-fucono-1,5-lactonase, which was discovered utilizing genomic enzymology techniques described in Chapter I. BmulJ_04915, which was annotated as having unknown function, was found to hydrolyze a series of 5- and 6-membered sugar lactones sharing similar stereochemistry at C-1 and C-2. The best substrate from the initial screen was L-fucono-1,4-lactone, a 5-membered lactone. However, L-fucose is known to primarily conform to a pyranose ring, not a furanose ring. To identify the natural, or preferred substrate, an adjacent gene thought to be responsible for the oxidation of L-fucose was cloned and purified. BmulJ_04919 was found to hydrolyze L-fucose to a chemically unstable L-fucono-1,5-lacotone, which rapidly rearranges (intramolecular) to the more stable L-fucono-1,4-lactone. Due to chemical instability, it was not possible to obtain proper steady-state enzyme kinetics for L-fucono-1,5-lactone, the presumed natural substrate. To overcome this, a substrate mimic, 4-deoxy-L-fucono-1,5-lactone, was enzymatically synthesized. The rates were found to be the highest with the substrate mimic; thus, reinforcing that BmulJ_04915 prefers the unstable 6-member lactone as opposed to the stable 5-member.

Chapter IV summarizes the unpublished *in vitro* analysis of a newly discovered metabolic pathway for L-galactose. BVU_0220, from cog3618, was predicted based on sequence alignments and genomic context to hydrolyze sugar lactones. This was confirmed through high-throughput substrate screening of a lactone library. This enzyme hydrolyzed L-glucono-1,5-lactone, L-galactono-1,4-lactone, and L-fucono-1,4-lactone, at similar rates. BVU_0219, an adjacently encoded enzyme, was predicted to oxidize a sugar substrate which would yield the natural substrate for the lactonase. BVU_0219, exhibited high substrate specificity for the rare sugar, L-galactose. As previously shown with BmulJ_04919, L-galactose was oxidized into an unstable 6-member ring, and subsequently rearranged into a more stable 5-member lactone. ^{13}C-NMR was used to show the preference of BVU_0220 to hydrolyze L-galacotno-1,5-lactone, as opposed to L-galactono-1,4-lactone. In addition, the preference for 6-member lactones was further reinforced through the hydrolysis of the enzymatically synthesized 4-deoxy-L-fucono-1,5-lactone. A second dehydrogenase encoded adjacently to the lactonase of interest, BVU_0222, was analyzed to determine the fate of L-galactose. This enzyme was found to oxidize the product of BVU_0220, L-galactonate, into D-tagaturonate. This was demonstrated through a coupled assay with URI, which converts D-galacturonate into D-tagaturonate. Thus, it was proposed that the apparent L-galactose pathway converges with the uronate degradation pathway.

In conclusion, this research develops and implements genomic enzymology strategies, towards the daunting task of annotating the millions of sequences available

in public databases. In this dissertation, we fully characterized the catalytic mechanism of LigI, in order to provide structure guided sequence alignments to predict the function of evolutionarily related proteins. The function of two COG members were also identified as L-fucono-1,5-lactonase and L-galactono-1,5-lactonase.

REFERENCES

[1] Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., & Bairoch, A. (2007). UniProtKB/Swiss-Prot. *Methods in Molecular Biology*, *406*, 89-112.

[2] Schnoes, A. M., Brown, S. D., Dodevski, I., Babbitt, P. C., & Valencia, A. (2009). Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Computational Biology*, *5*(12), e1000605.

[3] Gerlt, J., & Babbitt, P. (2001). Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *PLoS Computational Biology*, *70*, 209-246.

[4] Saghatelian, A., & Cravatt, B. F. (2005) The origin and evolution of protein superfamilies. Assignment of protein function in the postgenomic era. *Nature Chemical Biology*, *1*(3), 130-142.

[5] Dayhoff, M. (1976). The origin and evolution of protein superfamilies. *Federation Proceedings*, *62*, 154-161.

[6] Dayhoff, M. O., Mclaughlin, P. J., Barker, W. C., & Hunt, L. T. (1975). Evolution of sequences within protein superfamilies. *Die Naturwissenschaften*, *62*(4), 154-161.

[7] Holm, L., & Sander, C. (1997). An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins: Structure, Function, and Genetics*, *28*(1), 72-82.

[8] Seibert, C. M., & Raushel, F. M. (2005). Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry*, *44*(17), 6383-6391.

[9] Ghodge, S., Fedorov, A., Fedorov, E., Hillerich, B., Almo, S., & Raushel, F. (2013). Structural and mechanistic characterization of L-histidinol phosphate phosphatase from the polymerase and histidinol phosphatase family of proteins. *Biochemistry*, *52*, 8879-8890.

[10] Nguyen, T. T., Fedorov, A. A., Williams, L., Fedorov, E. V., Li, Y., Xu, C., et al. (2009). The mechanism of the reaction catalyzed by uronate isomerase illustrates how an isomerase may have evolved from a hydrolase within the amidohydrolase superfamily. *Biochemistry*, *48*(37), 8879-8890.

[11] Hara, H., Masai, E., Katayama, Y., & Fukuda, M. (2000). The 4-Oxalomesaconate hydratase gene, involved in the protocatechuate 4,5-cleavage pathway, Is essential to vanillate and syringate degradation in *Sphingomonas paucimobilis SYK-6. Journal of Bacteriology*, *182*(24), 6950-6957.

[12] Peng, X., Masai, E., Katayama, Y., & Fukuda, M. (1999). Characterization of the meta-cleavage compound hydrolase gene involved in degradation of the lignin-related biphenyl structure by *Sphingomonas paucimobilis SYK-6*. *Journal of Bacteriology*, *65*, 2789-2793.

[13] Peng, X., Masai, E., Kitayama, H., Harada, K., Katayama, Y., & Fukuda, M. (2002). Characterization of the 5-carboxyvanillate decarboxylase gene and its role in lignin-related biphenyl catabolism in *Sphingomonas paucimobilis SYK-6*. *Applied and Environmental Microbiology*, *68*(9), 4407-4415.

[14] Tatusov, R. L. (1997). A Genomic perspective on protein families. *Science*, *278*(5338), 631-637.

[15] Tatusov, R., Fedorova, N., Rao, B., Smirnov, S., Sverdlov, A., Vasudevan, A., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, 4, 41.

[16] Cummings, J. A., Vetting, M., Ghodge, S. V., Xu, C., Hillerich, B., Seidel, R. D., et al. (2014). Prospecting for unannotated enzymes: discovery of a 3',5'-nucleotide bisphosphate phosphatase within the amidohydrolase superfamily. *Biochemistry*, *53*(3), 591-600.

[17] Porter, T. N., Li, Y., & Raushel, F. M. (2004). Mechanism of the dihydroorotase reaction. *Biochemistry*, *43*(51), 16285-16292.

[18] Bertini, I. (2007). *Biological inorganic chemistry: structure and reactivity*. Sausalito, Calif.: University Science Books.

[19] Zastrow, M., & Pecoraro, V. (2014). Designing hydrolytic zinc metalloenzymes. *Biochemistry*, *53*, 957-978.

[20] Hall, R., Brown, S., Fedorov, A., Fedorov, E., Xu, C., Babbitt, P., et al. (2007). Structural diversity within the mononuclear and binuclear active sites of N-acetyL-D-glucosamine-6-phosphate deacetylase. *Biochemistry*, *46*, 7953-7962.

[21] Hall, R. S., Xiang, D. F., Xu, C., & Raushel, F. M. (2007). N-Acetyl-D-glucosamine-6-phosphate deacetylase: substrate activation via a single divalent metal ion. *Biochemistry*, *46*(27), 7942-7952.

[22] Tyagi, R., Eswaramoorthy, S., Burley, S. K., Raushel, F. M., & Swaminathan, S. (2008). A common catalytic mechanism for proteins of the HutI family. *Biochemistry*, *47*(20), 5608-5615.

[23] Hall, R. S., Agarwal, R., Hitchcock, D., Sauder, J. M., Burley, S. K., Swaminathan, S., et al. (2010). Discovery and structure determination of the orphan enzyme isoxanthopterin deaminase. *Biochemistry*, *49*(20), 4374-4382.

[24] Hall, R. S., Almo, S. C., Shoichet, B. K., Burley, S. K., Sauder, J. M., Kolb, P., et al. (2010). The hunt for 8-oxoguanine deaminase. *Journal of the American Chemical Society*, *132*(6), 1762-1763.

[25] Hall, R. S., Fedorov, A. A., Xu, C., Fedorov, E. V., Almo, S. C., & Raushel, F. M. (2011). Three-dimensional structure and catalytic mechanism of cytosine deaminase. *Biochemistry*, *50*(22), 5077-5085.

[26] Ireton, G. C., Mcdermott, G., Black, M. E., & Stoddard, B. L. (2002). The structure of *Escherichia coli* cytosine deaminase. *Journal of Molecular Biology*, *315*(4), 687-697.

[27] Healy, V., Mullins, L., Li, X., Hall, S., Raushel, F., & Walsh, C. (2000). D-Ala-D-X ligases: evaluation of D-alanyl phosphate intermediate by MIX, PIX and rapid quench studies. *Chemical Biology*, *7*, 505-514.

[28] Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K., et al. (2007). Structure-based activity prediction for an enzyme of unknown function. *Nature*, *448*(7155), 775-779.

[29] Xu, C., Hall, R., Cummings, J., & Raushel, F. (2006). Tight binding inhibitors of N-acyl amino sugar and N-acyl amino acid deacetylases. *Journal of American Chemical Society*, *128*, 4244-4245.

[30] Martí-Arbona, R., Fresquet, V., Thoden, J. B., Davis, M. L., Holden, H. M., & Raushel, F. M. (2005). Mechanism of the reaction catalyzed by isoaspartyl dipeptidase from. *Biochemistry*, *44*(19), 7115-7124.

[31] Marti-arbona, R., & Raushel, F. (2006). Mechanistic characterization of N-formimino-L-glutamate iminohydrolase from *Pseudomonas aeruginosa*. *Biochemistry*, *45*, 14256-14262.

[32] Martí-Arbona, R., Thoden, J. B., Holden, H. M., & Raushel, F. M. (2005). Functional significance of Glu-77 and Tyr-137 within the active site of isoaspartyl dipeptidase. *Bioorganic Chemistry*, *33*(6), 448-458.

[33] Martí-Arbona, R., Xu, C., Steele, S., Weeks, A., Kuty, G. F., Seibert, C. M., et al. (2006). Annotating enzymes of unknown function: N-formimino-D-glutamate deiminase is a member of the amidohydrolase superfamily. *Biochemistry*, *45*(7), 1997-2005.

[34] Thoden, J. B., Marti-Arbona, R., Raushel, F. M., & Holden, H. M. (2003). High-resolution x-ray structure of isoaspartyl dipeptidase from. *Biochemistry*, *42*(17), 4874-4882.

[35] Fan, H., Hitchcock, D. S., Seidel, R. D., Hillerich, B., Lin, H., Almo, S. C., et al. (2013). Assignment of pterin deaminase activity to an enzyme of unknown function guided by homology modeling and docking. *Journal of the American Chemical Society*, *135*(2), 795-803.

[36] Hitchcock, D. S., Sali, A., Shoichet, B. K., Almo, S. C., Seidel, R. D., Hillerich, B., et al. (2013). Structure-guided discovery of new deaminase enzymes. *Journal of the American Chemical Society*, *135*(37), 13927-13933.

[37] Hitchcock, D. S., Fedorov, A. A., Fedorov, E. V., Dangott, L. J., Almo, S. C., & Raushel, F. M. (2011). Rescue of the orphan enzyme isoguanine deaminase. *Biochemistry*, *50*(25), 5555-5557.

[38] Goble, A. M., Fan, H., Sali, A., & Raushel, F. M. (2011). Discovery of a cytokinin deaminase. *ACS Chemical Biology*, *6*(10), 1036-1040.

[39] Goble, A. M., Feng, Y., Raushel, F. M., & Cronan, J. E. (2013). Discovery of a camp deaminase that quenches cyclic amp-dependent regulation. *ACS Chemical Biology*, *8*(12), 2622-2629.

[40] Goble, A. M., Tanner, M. E., Fan, H., Ornelas, A., Li, X., Toro, R., et al. (2013). Deamination of 6-aminodeoxyfutalosine in menaquinone biosynthesis by distantly related enzymes. *Biochemistry*, *52*(37), 6525-6536.

[41] Goble, A., Zhang, Z., Sauder, J., Burley, S., Swaminathan, S., & Raushel, F. (2011). Pa0148 from *Pseudomonas aeruginosa* catalyzes the deamination of adenine. *Biochemistry*, *50*, 6589-6597.

[42] Williams, L., Nguyen, T., Li, Y., Porter, T. N., & Raushel, F. M. (2006). Uronate Isomerase: A nonhydrolytic member of the amidohydrolase superfamily with an ambivalent requirement for a divalent metal ion. *Biochemistry*, *45*(24), 7453-7462.

[43] Xiang, D. F., Almo, S. C., Burley, S. K., Sauder, J. M., Sisco, A. A., Fedorov, E. V., et al. (2010). Functional identification and structure determination of two novel prolidases from cog1228 in the amidohydrolase superfamily. *Biochemistry*, *49*(31), 6791-6803.

[44] Xiang, D. F., Patskovsky, Y., Xu, C., Meyer, A. J., Sauder, J. M., Burley, S. K., et al. (2009). Functional identification of incorrectly annotated prolidases from the amidohydrolase superfamily of enzymes. *Biochemistry*, *48*(17), 3730-3742.

[45] Martínez-Rodríguez, S., Martínez-Gómez, A. I., Clemente-Jiménez, J. M., Rodríguez-Vico, F., García-Ruíz, J. M., Heras-Vázquez, F. J., et al. (2010). Structure of dihydropyrimidinase from *Sinorhizobium meliloti CECT4114:* New features in an amidohydrolase family member. *Journal of Structural Biology*, *169*(2), 200-208.

[46] Kim, G. J., Lee, D. E., & Kim, H. (2000). Functional expression and characterization of the two cyclic amidohydrolase enzymes, allantoinase and a novel phenylhydantoinase, from *Escherichia coli*. *Journal of Bacteriology*, *182*(24), 7021-7028.

[47] Cummings, J. A., Almo, S. C., Barondeau, D. P., Shoichet, B. K., Fedorov, E. V., Xu, C., et al. (2010). Structure, mechanism, and substrate profile for sco3058: the closest bacterial homologue to human renal dipeptidase. *Biochemistry*, *49*(3), 611-622.

[48] Liao, R., Himo, F., Yu, J., & Liu, R. (2010). Dipeptide hydrolysis by the dinuclear zinc enzyme human renal dipeptidase: mechanistic insights from DFT calculations. *Journal of Inorganic Biochemistry*, *104*(1), 37-46.

[49] Bigley, A. N., & Raushel, F. M. (2013). Catalytic mechanisms for phosphotriesterases. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1834*(1), 443-453.

[50] Xiang, D. F., Kolb, P., Fedorov, A. A., Meier, M. M., Fedorov, L. V., Nguyen, T. T., Sterner, R., Almo, S. C., Shoichet, B. K., and Raushel, F. M. (2009) Functional annotation and three-dimensional structure of dr0930 from *Deinococcus radiodurans*, a close relative of phosphotriesterase in the amidohydrolase superfamily. *Biochemistry, 48*, 2237-2247.

[51] Xiang, D. F., Shoichet, B. K., Almo, S. C., Sterner, R., Nguyen, T. T., Fedorov, L. V., et al. (2009). Functional annotation and three-dimensional structure of dr0930 from, a close relative of phosphotriesterase in the amidohydrolase superfamily. *Biochemistry*, *48*(10), 2237-2247.

[52] Benini, S., Rypniewski, W. R., Wilson, K. S., Miletti, S., Ciurli, S., & Mangani, S. (1999). A new proposal for urease mechanism based on the crystal structures of the native and inhibited enzyme from *Bacillus pasteurii*: why urea hydrolysis costs two nickels. *Structure*, *7*(2), 205-216.

[53] Sumner, J. B., Graham, V. A., & Noback, C. V. (1924). The purification of jack bean urease. *Experimental Biology and Medicine*, *21*(8), 551-552.

[54] Benini, S., Rypniewski, W. R., Wilson, K. S., Miletti, S., Ciurli, S., & Mangani, S. (2001). Structure-based rationalization of urease inhibition by phosphate: novel insights into the enzyme mechanism. *Journal of Biological Inorganic Chemistry*, *6*(8), 778-790.

[55] Gorin, G. (1959). On the mechanism of urease action. *Biochimica Et Biophysica Acta*, *34*, 268-270.

[56] Cummings, J. A., Fedorov, A. A., Xu, C., Brown, S., Fedorov, E., Babbitt, P. C., et al. (2009). Annotating enzymes of uncertain function: the deacylation of-amino acids by members of the amidohydrolase superfamily. *Biochemistry*, *48*(27), 6469-6481.

[57] Chen, H., Wu, S., & Wang, K. (1994). D-Aminoacylase from *Alcaligenes faecalis* possesses novel activities on D-methionine. *Bioorganic & Medicinal Chemistry*, *2*(1), 1-5.

[58] Kamat, S. S., Swaminathan, S., Raushel, F. M., Lindahl, P. A., Burley, S. K., Sauder, J. M., et al. (2011). Catalytic mechanism and three-dimensional structure of adenine deaminase. *Biochemistry*, *50*, 1917-1920.

[59] Kamat, S. S., Fan, H., Sauder, J. M., Burley, S. K., Shoichet, B. K., Sali, A., et al. (2011). Enzymatic deamination of the epigenetic base-6-methyladenine. *Journal of the American Chemical Society*, *133*(7), 2080-2083.

[60] Kamat, S. S., Lindahl, P. A., Tichy, S. E., Kumaran, D., Bagaria, A., Holmes-Hampton, G. P., et al. (2011). The catalase activity of diiron adenine deaminase. *Protein Science*, *20*(12), 2080-2094.

[61] Ghodge, S. V., Cummings, J. A., Williams, H. J., & Raushel, F. M. (2013). Discovery of a cyclic phosphodiesterase that catalyzes the sequential hydrolysis of both ester bonds to phosphorus. *Journal of the American Chemical Society*, *135*(44), 16360-16363.

[62] Hishida, M., Shikinaka, K., Katayama, Y., Kajita, S., Masai, E., Nakamura, M., et al. (2009). Polyesters of 2-pyrone-4,6-dicarboxylic acid (pdc) as bio-based plastics exhibiting strong adhering properties. *Polymer Journal*, *41*(4), 297-302.

[63] Dersten, P., Dagley, S., Whittaker, J., Arciero, D., & Lipscomb, J. (1982). 2-Pyrone-4,6-dicarboxylic acid, a catabolite of gallic acids in *Pseudomonas* species. *Journal of Bacteriology*, *152*, 1154-1162.

[64] Maruyama, K. (1982). Purification and properties of 2-pyrone-4,6-dicarboxylate hydrolase. *Journal of Inorganic Biochemistry*, *93*, 557-565.

[65] Halak, S., Basta, T., Burger, S., Contzen, M., Wray, V., Pieper, D. H., et al. (2007). 4-Sulfomuconolactone hydrolases from *Hydrogenophaga intermedia S1* and *Agrobacterium radiobacter S2*. *Journal of Bacteriology*, *189*(19), 6998-7006.

[66] Watanabe, S., Saimura, M., & Makino, K. (2008). Eukaryotic and bacterial gene clusters related to an alternative pathway of nonphosphorylated L-rhamnose metabolism. *Journal of Biological Chemistry*, *283*(29), 20372-20382.

[67] Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P., & Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, *27*(3), 431-432.

[68] Atkinson, H. J., Morris, J. H., Ferrin, T. E., Babbitt, P. C., & Jordan, I. K. (2009). Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS ONE*, *4*(2), e4345.

[69] Masai, E., Momose, K., Hara, H., Nishikawa, S., Katayama, Y., & Fukuda, M. (2000). Genetic and biochemical characterization of 4-carboxy-2-hydroxymuconate-6-semialdehyde dehydrogenase and its role in the protocatechuate 4,5-cleavage pathway in *Sphingomonas paucimobilis SYK-6. Journal of Bacteriology*, *182*(23), 6651-6658.

[70] Masai, E., Shinohara, S., Hara, H., Nishikawa, S., Katayama, Y., & Fukuda, M. (1999). Genetic and biochemical characterization of a 2-pyrone-4, 6-dicarboxylic acid hydrolase involved in the protocatechuate 4, 5-cleavage pathway of Sphingomonas paucimobilis SYK-6. *Journal of Bacteriology*, *181*, 55-62.

[71] Hall, R. S., Xiang, D. F., Xu, C., & Raushel, F. M. (2007). N-Acetyl-D-glucosamine-6-phosphate Deacetylase: substrate activation via a single divalent metal ion. *Biochemistry*, *46*(27), 7942-7952.

[72] Becker, D. J., & Lowe, J. B. (2003). Fucose: biosynthesis and biological function in mammals. *Glycobiology*, *13*(7), 41R-53R.

[73] Chan, J., Nwokoro, N., & Schachter, H. (1979). L-Fucose metabolism in mammals. IV. The conversion of L-fucose to two moles of L-lactate, of L-galactose to L-lactate and glycerate, and of D-arabinose to L-lactate and glycollate. *Journal of Biological Chemistry*, *254*, 7060-7068.

[74] Yuen, R., & Schachter, H. (1972). L-Fucose metabolism in mammals. I. Pork Liver L-fuconate hydro-lyase. *Biochemistry and Cell Biology*, *50*(7), 798-806.

[75] Ghalambor, M. A., and Heath, E. C. (1962) The metabolism of L-fucose. II. The enzymatic cleavage of L-fuculose 1-phosphate, *J Biol Chem 237*, 2427-2433.

[76] Heath, E., & Ghalambor, M. (1962). Metabolism of L-fucose. I. The purification and properties of L-fuculose kinase. *Journal of Biological Chemistry*, *237*, 2423-2426.

[77] Seibert, C. M., & Raushel, F. M. (2005). Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry*, *44*(17), 6383-6391.

[78] Hobbs, M. E., Malashkevich, V., Williams, H. J., Xu, C., Sauder, J. M., Burley, S. K., et al. (2012). Structure and catalytic mechanism of LigI: Insight into the amidohydrolase enzymes of cog3618 and Lignin Degradation. *Biochemistry*, *51*(16), 3497-3507.

[79] Sakai, A., Xiang, D. F., Xu, C., Song, L., Yew, W. S., Raushel, F. M., et al. (2006). Evolution of enzymatic activities in the enolase superfamily: N-succinylamino acid racemase and a new pathway for the irreversible conversion of D- to L-amino acids. *Biochemistry*, *45*(14), 4455-4462.

[80] Yokochi, N., Nishimura, S., Yoshikane, Y., Ohnishi, K., & Yagi, T. (2006). Identification of a new tetrameric pyridoxal 4-dehydrogenase as the second enzyme in the degradation pathway for pyridoxine in a nitrogen-fixing symbiotic bacterium, *Mesorhizobium loti*. *Archives of Biochemistry and Biophysics*, *452*(1), 1-8.

[81] Chapman, E., & Wong, C. (2002). A pH sensitive colorometric assay for the high-Throughput screening of enzyme inhibitors and substrates. *Bioorganic & Medicinal Chemistry*, *10*(3), 551-555.

[82] Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D Biological Crystallography*, *60*(12), 2256-2268.

[83] Yasutake, Y., Nishiya, Y., Tamura, N., & Tamura, T. (2007). structural insights into unique substrate selectivity of *Thermoplasma acidophilum* D-aldohexose dehydrogenase. *Journal of Molecular Biology*, *367*(4), 1034-1046.

[84] Lesk, A. M. (1995). NAD-binding domains of dehydrogenases. *Current Opinion in Structural Biology*, *5*(6), 775-783.

[85] Ueberschar, K., Blachnitzky, E., & Kurz, G. (1974). Reaction mechanism of d-galactose dehydrogenases from pseudomonas saccharophila and pseudomonas fluorescens. Formation and rearrangement of aldono-1,5-lactones. *European Journal of Biochemistry*, *48*(2), 389-405.

[86] Xu, J., Knight, R. D., Minx, P., Coutinho, P. M., Henrissat, B., Martens, E. C., et al. (2007). Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biology*, *5*(7), e156.

[87] Saitoh, S., Noda, S., Aiba, Y., Takagi, A., Sakamoto, M., Benno, Y., et al. (2002). Bacteroides ovatus as the predominant commensal intestinal microbe causing a systemic antibody response in inflammatory bowel disease. *Clinical and Vaccine Immunology*, *9*(1), 54-59.

[88] Wexler, H. M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clinical Microbiology Reviews*, *20*(4), 593-621.

[89] Xu, J. (2003). A genomic view of the human-bacteroides thetaiotaomicron symbiosis. *Science*, *299*(5615), 2074-2076.

[90] Hooper, L., Midtvedt, T., & Gordon, J. (2002). How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annual Review of Nutrition*, *22*, 283-307.

[91] Imai, T., Ban, Y., Terakami, S., Yamamoto, T., & Moriguchi, T. (2009). L-Ascorbate biosynthesis in peach: cloning of six l-galactose pathway-related genes and their expression during peach fruit development. *Physiologia Plantarum*, *136*(2), 139-149.

[92] Anderson, E. (1933). The preparation of L-galactose from flaxseed mucilage. *Journal of Biological Chemistry*, *100*, 249-253.

[93] Pabst, M., Fischl, R. M., Brecker, L., Morelle, W., Fauland, A., Köfeler, H., et al. (2013). Rhamnogalacturonan II structure shows variation in the side chains monosaccharide composition and methylation status within and across different plant species. *The Plant Journal*, *76*, 61-72.

[94] Hobbs, M. E., Vetting, M., Williams, H. J., Narindoshvili, T., Kebodeaux, D. M., Hillerich, B., et al. (2013). Discovery of an L-Fucono-1,5-lactonase from cog3618 of the amidohydrolase superfamily. *Biochemistry*, *52*(1), 239-253.

[95] Yew, W. S., Fedorov, A. A., Fedorov, E. V., Rakus, J. F., Pierce, R. W., Almo, S. C., et al. (2006). Evolution of enzymatic activities in the enolase superfamily: l-fuconate dehydratase from *Xanthamonas campestris*. *Biochemistry*, *45*(49), 14582-14597.

[96] Debolt, S. (2006). L-Tartaric acid synthesis from vitamin C in higher plants. *Proceedings of the National Academy of Sciences*, *103*(14), 5608-5613.

[97] Schenk, P., Baumann, S., Mattes, R., & Steinbiss, H. (1995). Improved high-level expression system for eukaryotic genes in Escherichia coli using T7 RNA polymerase and rare ArgtRNAs. *Biotechniques*, *19*, 196-198.

[98] Shimizu, T., Takaya, N., & Nakamura, A. (2012). An L-glucose catabolic pathway in *Paracoccus Species 43P*. *Journal of Biological Chemistry*, *287*(48), 40448-40456.

[99] Zhu, Y., & Lin, E. (1987). Loss of aldehyde dehydrogenase in an Escherichia coli mutant selected for growth on the rare sugar L-galactose. *Journal of Bacteriology*, *169*, 785-789.

[100] Gatzek, S., Wheeler, G. L., & Smirnoff, N. (2002). Antisense suppression of l-galactose dehydrogenase in *Arabidopsis thaliana* provides evidence for its role in ascorbate synthesis and reveals light modulated L-galactose synthesis. *The Plant Journal*, *30*(5), 541-553.

[101] Mieda, T. (2004). Feedback inhibition of spinach L-galactose dehydrogenase by L-ascorbate. *Plant and Cell Physiology*, *45*(9), 1271-1279.

[102] Momma, M., & Fujimoto, Z. (2013). Expression, crystallization and preliminary X-ray analysis of rice-galactose dehydrogenase. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, *69*(7), 809-811.

[103] Angyal, S., Bethell, G., Cowley, D., & Pickles, V. (1976). Equilibria between pyranoses and furanoses. IV. 1-Deoxyhexuloses and 3-hexuloses. *Australian Journal of Chemistry*, *29*(6), 1239.