D-AMINOACYLASES AND DIPEPTIDASES WITHIN THE AMIDOHYDROLASE SUPERFAMILY: RELATIONSHIP BETWEEN ENZYME STRUCTURE AND SUBSTRATE SPECIFICITY

A Dissertation

by

JENNIFER ANN CUMMINGS

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

DOCTOR OF PHILOSOPHY

December 2010

Major Subject: Chemistry

D-AMINOACYLASES AND DIPEPTIDASES WITHIN THE AMIDOHYDROLASE SUPERFAMILY: RELATIONSHIP BETWEEN ENZYME STRUCTURE AND SUBSTRATE SPECIFICITY

A Dissertation

by

JENNIFER ANN CUMMINGS

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

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Frank Raushel
Committee Members,	Paul Lindahl
	David Barondeau
	Gregory Reinhart
Head of Department,	David Russell

December 2010

Major Subject: Chemistry

ABSTRACT

D-Aminoacylases and Dipeptidases within the Amidohydrolase Superfamily: Relationship Between Enzyme Structure and Substrate Specificity. (December 2010)

Jennifer Ann Cummings, B.S., Southern Oregon University;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Frank Raushel

Approximately one third of the genes for the completely sequenced bacterial genomes have an unknown, uncertain, or incorrect functional annotation. Approximately 11,000 putative proteins identified from the fully-sequenced microbial genomes are members of the catalytically diverse Amidohydrolase Superfamily. Members of the Amidohydrolase Superfamily separate into 24 Clusters of Orthologous Groups (cogs). Cog3653 includes proteins annotated as *N*-acyl-D-amino acid deacetylases (DAAs), and proteins within cog2355 are homologues to the human renal dipeptidase. The substrate profiles of three DAAs (Bb3285, Gox1177 and Sco4986) and six microbial dipeptidase (Sco3058, Gox2272, Cc2746, LmoDP, Rsp0802 and Bh2271) were examined with *N*-acyl-L-, *N*-acyl-D-, L-Xaa-L-Xaa, L-Xaa-D-Xaa and D-Xaa-L-Xaa substrate libraries. The rates of hydrolysis of the library components were determined by separating the amino acids by HPLC and quantitating the products.

Gox1177 and Sco4986 hydrolyzed several *N*-acyl-D-amino acids, especially those where the amino acid was a hydrophobic residue. Gox1177 hydrolyzed L-Xaa-D-

Xaa and *N*-acetyl-D-amino acids with similar catalytic efficiencies (~10^4 $M^{-1}s^{-1}$). The best substrates identified for Gox1177 and Sco4986 were *N*-acetyl-D-Trp and *N*-acetyl-D-Phe, respectively. Conversely, Bb3285 hydrolyzed *N*-acyl-D-Glu substrates ($k_{cat}/K_m \approx 5 \times 10^6$ $M^{-1}s^{-1}$) and, to a lesser extent, L-Xaa-D-Glu dipeptides. The structure of a DAA from *A. faecalis* did not help explain the substrate specificity of Bb3285. *N*-methylphosphonate derivatives of D-amino acids were inhibitors of the DAAs examined. The structure of Bb3285 was solved in complex with the *N*-methylphosphonate derivative of D-Glu or acetate/formate. The specificity of Bb3285 was due to an arginine located on a loop which varied in conformation from the *A. faecalis* enzyme.

In a similar manner, six microbial renal dipeptidase-like proteins were screened with 55 dipeptide libraries. These enzymes hydrolyzed many dipeptides but favored L-D dipeptides. Respectable substrates were identified for proteins Bh2271 (L-Leu-D-Ala, $k_{cat}/K_m = 7.4 \times 10^4$ $M^{-1}s^{-1}$), Sco3058 (L-Arg-D-Asp, $k_{cat}/K_m = 7.6 \times 10^5$ $M^{-1}s^{-1}$), Gox2272 (L-Asn-D-Glu, $k_{cat}/K_m = 4.7 \times 10^5$ $M^{-1}s^{-1}$), Cc2746 (L-Met-D-Leu, $k_{cat}/K_m = 4.6 \times 10^5$ $M^{-1}s^{-1}$), LmoDP (L-Leu-D-Ala, $k_{cat}/K_m = 1.1 \times 10^5$ $M^{-1}s^{-1}$), Rsp0802 (L-Met-D-Leu, $k_{cat}/K_m = 1.1 \times 10^5$ $M^{-1}s^{-1}$). Phosphinate mimics of dipeptides were inhibitors of the dipeptidases. The structures of Sco3058, LmoDP and Rsp0802 were solved in complex with the pseudodipeptide mimics of L-Ala-D-Asp, L-Leu-D-Ala and L-Ala-D-Ala, respectively. The structures were used to assist in the identification of the structural determinants of substrate specificity.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Frank Raushel, for his guidance, patience and support, and my committee members, Dr. Lindahl, Dr. Barondeau, and Dr. Reinhart. Thanks to my former lab mates, Tamiko Porter and Sarah Aubert, and comrades, Eman Ghanem and LaKenya Williams for showing me the ropes during the early days. I would also like to express my appreciation for Jinny Johnson from the Protein Chemistry Laboratory at Texas A&M University for providing the amino acid analysis service and her gracious acceptance of my habitual perfectionism. Finally, I would like to thank my family and friends for their encouragement and love.

NOMENCLATURE

a.a.	One of the twenty standard amino acids except cysteine
AHS	AmidoHydrolase Superfamily
ATP	Adenosine TriPhosphate
BLAST	Basic Local Alignment Search Tool
β_1-β_8	Numbers 1 through 8 of the β-strands within the $(\beta/\alpha)_8$-barrel
Hepes	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HPLC	High-Performance Liquid Chromatography
ICP-MS	Inductively Coupled Plasma Mass Spectrometry
KEGG	Kyoto Encyclopedia of Genes and Genomes
M_α	Alpha Metal
M_β	Beta Metal
NAD^+	β-Nicotinamide Adenine Dinucleotide
NADH	β-Nicotinamide Adenine Dinucleotide reduced form
NCBI	National Center for Biotechnology Information
NYSGXRC	New York Structural Genomics Research Center
OD	Optical Density
OPA	*o*-Phthalaldehyde
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
Tris	Tris(hydroxymethyl)aminomethane
Xaa	One of the twenty standard amino acids except cysteine

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

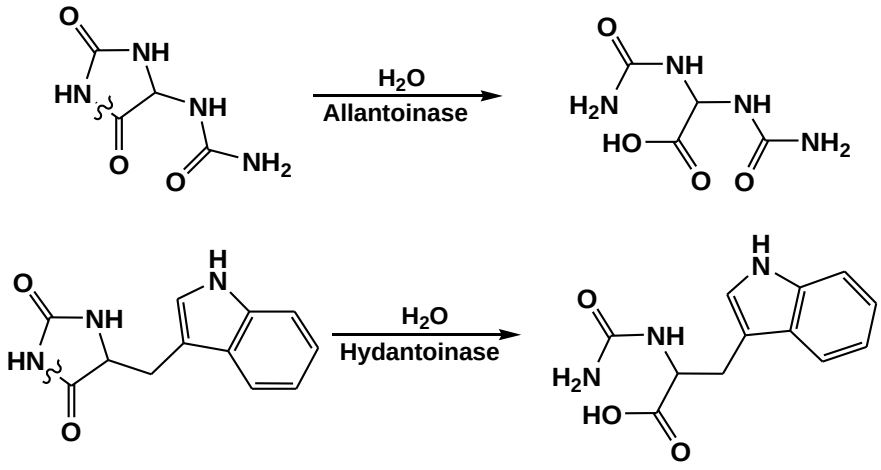
INTRODUCTION AND LITERATURE REVIEW

In the pregenomic era the problem was to answer the question "*what enzyme is carrying out a reaction of interest*?" An enzyme with a given function was identified by following an activity of interest during the purification of an enzyme of unknown amino acid sequence. The amino acid sequence of the enzyme could then be determined, and perhaps its 3D structure would also be solved in order to help explain the mode of substrate binding or aspects of the enzyme mechanism. In the genomic era, a great deal of genetic information is known and the anticipated number of open reading frames (ORFs), which code for putative proteins, now exceeds 10.5 million. With the Protein Structure Initiative (PSI), 3D structures of proteins are now being deposited at a faster rate than the determination of their function. Sources of new proteins not only arise from purifying targets in sequenced organisms, but they also come from cloning and expressing genes from environmental DNA. The problem now is to answer the question "*What is the function of a protein with a given 3D structure and/or amino acid sequence*?" The strategies for determining the substrates of enzymes include, but are not limited to, annotation based on sequence similarity, *in silico* screening of crystal structures and homology models with metabolite libraries, docking compounds of the metabolome into crystal structures, phenotypes of gene knockouts and analysis of the genome context of the target of interest.

This dissertation follows the style of *Biochemistry*.

Protein Annotation and Caveats. Functional annotation of a protein of interest based on sequence similarity to another characterized protein is the first step towards identifying its substrate. Sometimes the annotation alone is the only step required before experimentally confirming the activity of the enzyme with the substrate. Proteins within an enzyme family perform the same overall reaction and generally have greater than 30% sequence identity in a pairwise sequence alignment. Still, functionally divergent enzymes emerge as members of a given family despite their having considerably less than 30% sequence identity to the other functional members. To the other extreme, a set of proteins can have significantly more than 30% sequence identity and not perform the same overall reaction. Though a high sequence similarity between two proteins does not always translate to their accepting the same substrate, the substrates of the two enzymes may be similar in structure. For example, the *E. coli* K12 allantoinase and L-Hydantoinase from *Arthrobacter aurescens* are 38% identical, and they hydrolyze hydantoin rings with a different substituent on carbon 5 of the heterocylcic ring as depicted in **Scheme 1.1** (*1*, *2*).

Scheme 1.1 Reactions catalyzed by allantoinase and hydantoinase



There are problems associated with predicting the function of a protein based on its annotation in the protein databases. An uncharacterized protein of interest is usually given the same annotation as that of a protein with which it has the highest amino acid sequence identity (protein #2). However, protein #2 can be an uncharacterized protein annotated by the same method. There must be an origin to these annotations but locating the source can be difficult. In cases where annotations are based on amino acid sequence similarity and not on experimental evidence, the specific journal reference for the activity is rarely shown. Investigation of the journal article for the genome sequencing of the organism from which the protein target is from does not always lead one to the identity of the protein with the experimental evidence for the annotation. Protein sequences often have multiple reference numbers for the same locus tag, but different annotations can be associated with the locus tag. For example the protein with locus tag Gox1177 can have the very specific annotation of *N*-acetyl-D-glutamate

deacetylase (gi|58039630) or the broader annotation of D-aminoacylase (gi|58002044). The putative deaminase, locus tag EF1061, is annotated as a deaminase (gi|29375640) or an *N*-acyl-D-amino acid family protein (gi|29343101).

Since the top BLAST (Basic Local Alignment Search Tool) hits of a protein of interest are based on those with the highest sequence identity to the protein target, the annotation can be incorrect because a difference of as little as one key residue can alter the function of the protein or render the protein inactive. For example it was found that 1 of the 3 homologues of GTP cyclohydrolase II (GTPCII) in *Streptomyces coelicolor*, Sco6655, was not a true GTPCII enzyme but only catalyzed part of the GTPCII reaction, converting GTP to a different product than the other two homologues (*3*). Through crystallography and site-directed mutagenesis, the difference was attributed to a single amino acid change in one of the active site residues. Unfortunately, due to the inevitable lag time between discoveries in activities presented in the literature and updating of protein annotations, Sco6655 is still annotated as GTP cyclohydrolase II in the NCBI database under three reference numbers, only two of which lead to a link to this work (*3*). It is safe to assume that this is not an isolated incident and misannotations are probably more common than one would hope. The propagation of these misannotations can potentially make it difficult to draw conclusions from genome context if the proteins in the genome context are misannotated. Proteins annotated as chlorohydrolase or atrazine chlorohydrolase are a good example of a cluster of Amidohydrolase Superfamily (AHS) amino acid sequences that are overannotated.

Using Genome Context to Assign Function. The *hut*F enzyme from *Pseudomonas aeruginosa* (Pa5106), originally annotated as a chlorohydrolase, is an example of a protein which had an ambiguous annotation but whose genome context gave clues as to the likely function of the enzyme. The Pa5106 protein is part of a cluster of at least 39 amidohydrolase homologues annotated as chlorohydrolases or *hut*F. Analysis of the genome context of these proteins revealed that these homologues were usually close to other proteins annotated as proteins involved in histidine degradation, namely hutH, hutU, hutI and hutG (Pa5091). The product of the hutI reaction, *N*-formimino-glutamate, has three potential metabolic fates. However, the action of *hut*F to convert *N*-formimino-glutamate to *N*-formyl-glutamate is the only path that would warrant the need for the subsequent hutG protein, *N*-formyl-glutamate deformylase. The Pa5091 and Pa5106 proteins were experimentally confirmed to be hutG and hutF, respectively. The genome context allowed the researchers to narrow their focus to three reactions in the pursuit of assigning function to Pa5106 (*4*).

The genome context of Pa5106 is moderately organized in that the five enzymes required to degrade histidine to glutamate are located within a span of 16 genes in the genome, and, with the exception of Pa5106, all of the genes are transcribed from the same DNA strand (**Figure 1.1A**). The genes for enzymes that are metabolically related are sometimes split up into smaller clusters in remote locations of the genome in one organism and yet condensed in a different organism. In *Bacillus halodurans,* three genes which code for enzymes involved in the degradation of D-glucuronate, *uxu*AB and *uxa*C, or D-galacturonate, *uxa*ABC, are located within a span of 3 or 5 genes, respectively and

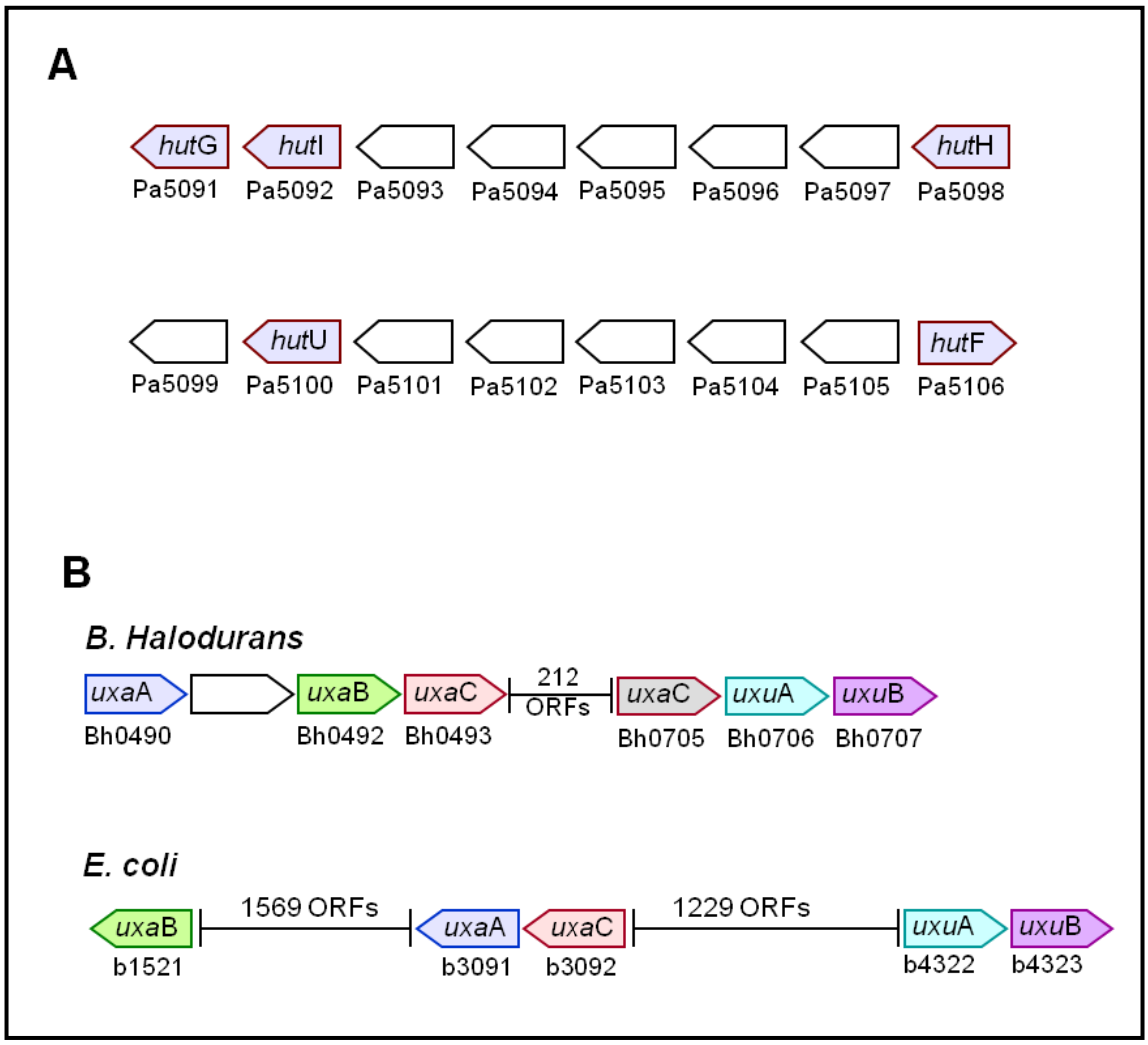


Figure 1.1: Genome context of amidohydrolase superfamily members with recently elucidated functions. **(A)** *hut*F in *P. aeruginosa* and **(B)** D-galacturonate and D-glucuronate degrading enzymes in *B. halodurans* and *E. coli*.

transcribed from the same DNA strand. However, in *E. coli*, the gene clusters *uxa*AC and *uxu*AB, and the *uxa*B gene have ~ 1000-1500 open reading frames (ORFs) between them. Furthermore, *uxa*C (b3092) catalyzes the isomerization of both D-glucuronate and D-galacturonate with similar efficiencies, but *B. halodurans* has a second *uxa*C (Bh0705), appropriately clustered with *uxu*AB, which is more specialized for the D-glucuronate than D-galacturonate (*5*). The genome contexts for the *hut* operon in *P. aeruginosa* and D-galacturonate/D-glucuronate metabolism in *B. halodurans* and *E. coli* are depicted in **Figure 1.1**. One can envision a case where a protein of unknown function is situated in the genome in a similar fashion to that of b3092 where the genome context would not be helpful in determining the protein's metabolic significance.

It would be naïve to say that the enzymatic reactions outlined in the literature and the KEGG database can define the entire proteome. Many protein sequence entries have less than 40% amino acid sequence identity to any other entries in the protein database with known activity. Begley and coworkers used the genome contexts of the TenA (*Bacillus halodurans*, *Bacillus subtilis*, *Bacillus cereus*) and YlmB (*B. halodurans*) genes from soil bacteria to find substrates for the TenA protein. They had originally assigned TenA as a thiaminase II because the enzyme hydrolyzed thiamin to hydroxypyrimidine (HMP) and hydroxyethylthiozole (*6*). A degratory function did not seem logical considering that TenA was clustered with thiamin biosynthetic enzymes and/or with putative transporters of HMP analogs, *Thi*XYZ, or fused with the biosynthetic enzyme *Thi*D. They hypothesized that soil may degrade thiamin and were able to determine that Ylmb hydrolyzed the soil-degraded product of thiamin,

formylaminopyrimidine to aminopyrimidine. By the action of TenA, aminopyrimidine was hydrolyzed to HMP which can be reused in thiamin biosynthesis. TenA hydrolyzed aminopyrimidine at a rate ~ 100 times faster vs. that of thiamin. Furthermore, the *Thi*Y protein was found to bind formylaminopyrimidine, K_d = 200 nM. This is an example where genome context and a logic-based hypothesis led to the identification of new substrates and ligands of proteins of unknown function, namely TenA, Ylmb and *Thi*Y, when crystallographic information was relatively uninformative (*7*).

Prediction of Function from Form: Proof-of-Principle. With the number of crystal structures of AHS enzymes of unknown function continually increasing, there is interest in being able to predict substrates for these proteins based on their active sites. However, there are several crystal structures of AHS proteins where the substrates are known. The structures of dihydroorotase, adenosine deaminase, *N*-acetyl-D-glucosamine-6-phosphate deacetylase, cytosine deaminase, D-hydantoinase, isoaspartyl-dipeptidase and PTE were used in proof-of-principle docking experiments. A library of ground-state and high-energy intermediates of potential AHS substrates included molecules with electrophilic centers of restricted size extracted from the KEGG database of metabolites as well as 2009 dipeptides/acyl-amino acids, 14 hydantoins and five phosphorus esters for certain enzymes. The high-energy forms were representations of the molecule following attack of the electrophilic center; for hydrolysis of an amide bond, the high-energy form would be a tetrahedral intermediate where the oxygen of the attacking hydroxyl is negatively charged and the proton of the hydroxyl is transferred to the carbonyl oxygen. Many conformations of each potential substrate were modeled into

the active sites of the seven AHS enzymes mentioned above. The high-energy intermediate forms of the known substrates for each enzyme were in the top 10-100 hits out of a pool of 3700 molecules (*8*).

A similar but less extensive docking experiment was performed using a homology model of a protein which had not been crystallized. One of the subgroups of the enolase superfamily are the muconate lactonizing enzymes (MLEs). Based on sequence identity alone, the protein Bc3071 is an L-Ala-D/L-glutamate epimerase (AEE), a division within the MLE subgroup. Another function performed by some members of the MLE subgroup is *N*-succinyl-amino acid racemization. An AEE from *B. subtilis* was the closest homolog to Bc3071 with a known structure and catalytic reaction, with 35% sequence identity to Bc3071. However, the Arg-24 residue which interacts with the glutamate carboxylate side chain and the two aspartates which hydrogen bond to the N-terminus are not present in Bc3071. Bc3071 was screened for epimerase activity on the libraries L-Xaa-L-Xaa (Xaa = all L-amino acids) and *N*-succinyl-L-Xaa for a total of 420 potential substrates. The best substrate, within this set of compounds, was *N*-succinyl-L-arginine. A homology model of Bc0371 was created based on the structure of the *B. subtilis* protein. The same library of 420 compounds was docked to the homology model. Rotomer optimization was also performed for those unconserved residues within 5 Å of each docked compound. The highest scoring ligand was also *N*-succinyl-L-arginine followed by nine other dipeptides with L-arginine or L-lysine at the C-terminus within the top 15 hits. This experiment showed that, at least with a small library of

ligands, docking experiments could accurately identify the most favorable substrates, within a library of compounds, for an enzyme based on a homology model (*9*).

Prediction of Function from Form: Experimental. The high energy intermediate docking method was applied to the crystal structure of Tm0936, an AHS enzyme similar in sequence to another protein ambiguously annotated as a chlorohydrolase. Out of the 4,207 metabolites, 9 of the 10 top hits were adenine analogues. The 5th (5-methyl-thioadenosine, MTA), 6th (S-adenosyl-L-homocysteine, SAH), 14th (adenosine) and 80th (adenosine-5-monophosphate, AMP) ranked compounds were selected as substrates and their k_{cat}/K_m values for deamination activity were in same rank order: MTA (1.4 x 10^5 $M^{-1}s^{-1}$) > SAH > Adenosine >> AMP. This MTA deaminase activity gave more meaning to the annotation of Tm0938 as a SAM-dependent methyl transferase and to other orthologues of Tm0936 that are close to genes with the annotations such as SAM-dependent methyltransferases, UBiG-methyltransferases, adenosyl homocysteinase, 5'-methyladenosine phosphorylase and MTA/SAH nucleosidase (*10*).

Activity Screening for Substrates. In silico screening of crystal structures with the library of metabolites in the KEGG database is a reasonable task. In contrast, screening a single enzyme for *activity* against the same library would be impractical. Instead, one can choose to focus on a particular group of enzymes. In doing this one can become more familiar with the mechanisms of the enzymes and the types of reactions they catalyze. For example, enzymes in the enolase superfamily are separated into three subgroups. They catalyze the same partial reaction to form an enediolate intermediate, by abstraction of a C_α proton, but have different conserved catalytic residues at varying

positions in the structure. The catalytic residues of these subgroups are known based on amino acid sequence alignments and extensive crystallographic information. What is not known for all of the enolases is their substrate specificity.

Many members of the mandelate racemase (MR) subgroup have the same conserved catalytic residues but are not expected to be true MRs based on genome context. Gerlt and coworkers identify members of the MR subgroup that are in potential operons for the metabolism of sugars. They purify the MR enzymes and screen them for dehydration, epimerization and C_α proton exchange against a library of ~ 50 mono- and diacid sugars. These libraries include additional sugars that are not in the KEGG database of metabolites. They were able to identify an enzyme which acted exclusively as a dehydratase of D-tartrate as well as another enzyme with a significant preference for the dehydration of L-fuconate (FucD) (*11, 12*). In the case of FucD, they were able to identify two other enzymes involved in the degradation of L-fucose that were in an operon with FucD. In some instances, they have also been able to crystallize the enzymes with a ligand bound and provide explanations for the substrate specificity and fate of the enediolate intermediate. Even in the absence of useful information from the genome context, it is possible to drastically reduce the pool of candidate substrates to a more manageable number. As described earlier for the Bc3071 protein, the knowledge of the types of reactions catalyzed by members of the MLE subgroup made substrate screening a viable method to determine substrate specificity.

Enzyme Superfamilies and the TIM-Barrel Fold. It is advantageous to focus on a particular enzyme superfamily. The enzymes of different families within a superfamily

can catalyze a variety of overall reactions with at least a consistent partial reaction if not by the same mechanism entirely. Enzymes within the same family catalyze the same overall reaction (*13*). According to the 1.75 June 2009 release of the Structural Classification of Proteins (SCOP), there are 1195 folds, 1962 superfamilies and 3902 families of proteins. These statistics amount to ~ 2 superfamilies and ~ 3 different functions per fold on average. However, currently 33 superfamilies, structurally differentiated into ~ 100 families, are classified as having the TIM-barrel fold. Clearly, TIM-barrel proteins are enriched with functional diversity. With the aid of tools such as amino acid sequence alignments, site-directed mutagenesis and crystallographic analysis (especially with substrates or products bound), the key residues involved in substrate binding and those required for catalysis can be proposed. The establishment of a canonical correlation between sequence, structure and function can be used to simplify the search for substrates of a protein of interest.

Amidohydrolase Superfamily: Structure. Holm and Sander recognized the marked similarities in the active site framework and protein folds revealed in the three-dimensional crystal structures of phosphotriesterase (PTE), urease (URA) and adenosine deaminase (ADD) (*14*). All three of these proteins exhibited a $(\beta/\alpha)_8$-barrel (TIM-barrel) structural fold. The unwavering characteristic residues of the three active sites were the metal ligands—two histidines in the HxH motif at the end of β-strand 1 (β_1), histidines at the ends of β-strands 5 (β_5) and 6 (β_6), and an aspartic acid at the end of β-strand 8 (β_8). The ADD enzyme bound one Zn, ligated to the β_1, β_5 and β_6 histidines, as well as the β_8 aspartic acid. Urease and PTE have binuclear metal centers with two

divalent cations, M_α, ligated to two histidines (β_1) and an aspartic acid (β_8), and, M_β, ligated to two histidines (β_5 and β_6). The two metals are bridged by a carboxylated lysine from β-strand 4 (β_4). They used the amino acid sequences of URA, PTE and ADD as seed sequences to search through the sequence space of ~ 208,000 nonredundant protein sequences. They predicted that URA, PTE, ADD, and more than 70 other proteins in eleven other families (i.e. cytosine deaminase, allantoinase, hydantoinase, dihydroorotase, chlorohydrolase, adenine deaminase, arylphosphatase, imidazolonepropionase, formylmethanofuran dehydrogenase subunit A, aminoacylases and AMP deaminase) are part of an amidohydrolase superfamily (AHS) and would share the same active site architecture and structural fold as the three seed sequences (*14*). The validity of some of these predictions are apparent in the three-dimensional structures of cytosine deaminase (*E. coli*), hydantoinase (*B. stearothermophilus*, *A. aurescens*), imidazolonepropionase (*A. tumefaciens*) and dihydroorotase (*E. coli*) (*15-19*).

The AHS is a well-characterized family of enzymes which have high catalytic diversity. These enzymes are metal-dependent hydrolases, characterized by a common $(\beta/\alpha)_8$-barrel structural fold of eight alternating α-helices and β-strands (*20*). Four conserved histidine residues and a catalytic aspartic acid residue serve as ligands to the active site metal(s). The amidohydrolase superfamily is a mechanistically diverse superfamily because AHS members cleave more than amide bonds alone. The types of reactions catalyzed by AHS members include the hydrolysis of lactones, C—N bonds, organophosphate P—O bonds, 1,2 hydrogen transfers and decarboxylations (*20-22*).

Amidohydrolase Superfamily: Active Site Metal Centers. As the number of crystal structures of AHS members increases, more variations in the mononuclear and binuclear metal centers are being discovered. To date there are ten classes of metal centers in the AHS. Eight of the ten classes contain the HxH motif, and nine have the conserved aspartic acid from β_8; however, these two histidines and/or aspartate do not always ligate a second metal (*20*). The first class is represented by such enzymes as PTE, urease, isoaspartyl dipeptidase and dihydroorotase (DHO). In binuclear enzymes, M_α and M_β are the buried and more solvent-exposed divalent cations, respectively. In other binuclear metal center classes, the bridging ligand is a glutamate (classes 2, 4 and 6) or a cysteine on β_2 (class 5). The biological function of the class two type proteins are unknown and include the PTE homology protein (PHP) and some of those annotated as tatD deoxyribonucleases (*20*). In the human renal dipeptidase (class 6), the HxH motif is replaced by an HxD motif, where M_α is ligated to a histidine and an aspartic acid at the end of β_1, and the aspartate on β_8 is not a metal ligand (*23*). Studies on the DAA from *A. faecalis* have shown that the enzyme is more active in the mononuclear form even though it can bind two metals. This DAA required the M_β, ligated by a cysteine and the third and fourth histidines, but increased Zn concentrations lead to the ligation of the second M_α; M_α was ligated by the two histidines of the HxH motif and the Asp from β_8 which created a less active binuclear enzyme (*24*). It has also been shown that high concentrations of Zn in the assay will decrease the activity of *E. coli* allantoinase even though this enzyme has highest activity with more than one equivalent of metal (*2, 25*).

There are mononuclear AHS members that only require one metal for activity. The class 3 metal center, represented by the *E. coli* cytosine deaminase, contains the four conserved histidines (β_1, β_5, and β_6) and the aspartic acid (β_8); however, the β_6 histidine is not a metal ligand (*15*). The class 4 metal center is represented by the mononuclear *N*-acetyl-D-glucosamine-6-phosphate deacetylases (AGD) from *T. maritima* and the binuclear AGD from *B. subtilis (26).* The AGD from *E. coli* represents the seventh class, where the β_1 histidines are replaced by a glutamine and asparagine, and the enzyme is functional without an HxH or HxD motif (*27*). The class 8 metal center is that of a uronate isomerase from *B. halodurans* (Bh0493) where M_α is a Zn coordinated to water, the β_8 aspartate, and the histidines from β_1 (*5*). Two uronate isomerases, Bh0493 and Tm0064, are the only structurally characterized AHS members where the histidine from β_6 is missing.

The ninth metal center is represented by one active site variety of the three AHS members annotated as a tatD deoxyribonuclease in *E. coli* (PDB ID: *1XWY*). The tatD protein has a single Zn ligated to a glutamate (β_8) and two adjacent residues at the C-terminal end of β_6, namely H152 and C153; the HxH motif and the β_5 histidine are missing. In the tatD active site there is an aspartic acid following β_8 which does not coordinate the Zn but it is within hydrogen bonding distance to a water molecule (likely the catalytic water) which is coordinated to the Zn ion.

The most recently identified tenth variety of an AHS metal center belongs to an adenine deaminase from *Agrobacterium tumefaciens*, Atu4426, with PDB code *3NQB*. The two Zn ions are bridged by a water and the glutamate at the end of β_4. The

coordination of Zn_α is completed by the two histidines from β_1 and the catalytic aspartate from β_8. The histidine from β_5 and two consecutive residues at the end of β_6, a histidine and glutamate, are the ligands unique to Zn_β. Atu4426 is congruous to tatD from *E. coli* in that the histidine from β_6 and the adjacent residue after the β_6 histidine are both metal ligands. There are crystal structures of other AHS members where a metal is not present in the active site.

Two AHS proteins, Pa2393 (PDB ID: *3B40*) and Pa5396 (PDB ID: *2I5G*), are sometimes annotated as putative dipeptidases. Pa2393 has a glutamine in place of the histidine on the loop immediately following the C-terminal end of β_5. Both proteins have the aspartic acid on β_8, but are missing the HxH or HxD motif from β_1. Technically, the active sites of Pa2393 and Pa5396 are not metal centers because neither has a metal in the active site. A water molecule is within 3-3.5 Å of the glutamine (β_5), histidine (β_6) and aspartic acid (β_8) residues; a metal may be able to occupy the position of the water. The different metal site classes are summarized in **Table 1.1**. A structural example of each of these metal centers is shown in **Figure 1.2**. It is clear that the definition of an AHS member cannot be confined to the class 1 and 3 active sites first outlined by Holm and Sander (*14*).

Table 1.1: Summary of metal centers of AHS members

CLASS NUMBER	METAL POSITION	β-strand							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	α, β	HxH			K[a]	H	H		D
2	α, β	HxH			E	H	H		D
3	α	HxH				H	h[c]		D
4	α, β or β	hxh[b]		E		H	H		D[b]
5	α, β	HxH	C			H	H		d[c]
6	α, β	HxD		E		H	H		d[c]
7	β	qxn[c]		E		H	H		d[c]
8	α	HxH				H			D
9	β						H,C		E
10	α, β	HxH			E	H	H,E		D

a The ε amino group is carboxylated
b These histidines are not metal ligands in the M_β mononuclear center
c These residues are not metal ligands

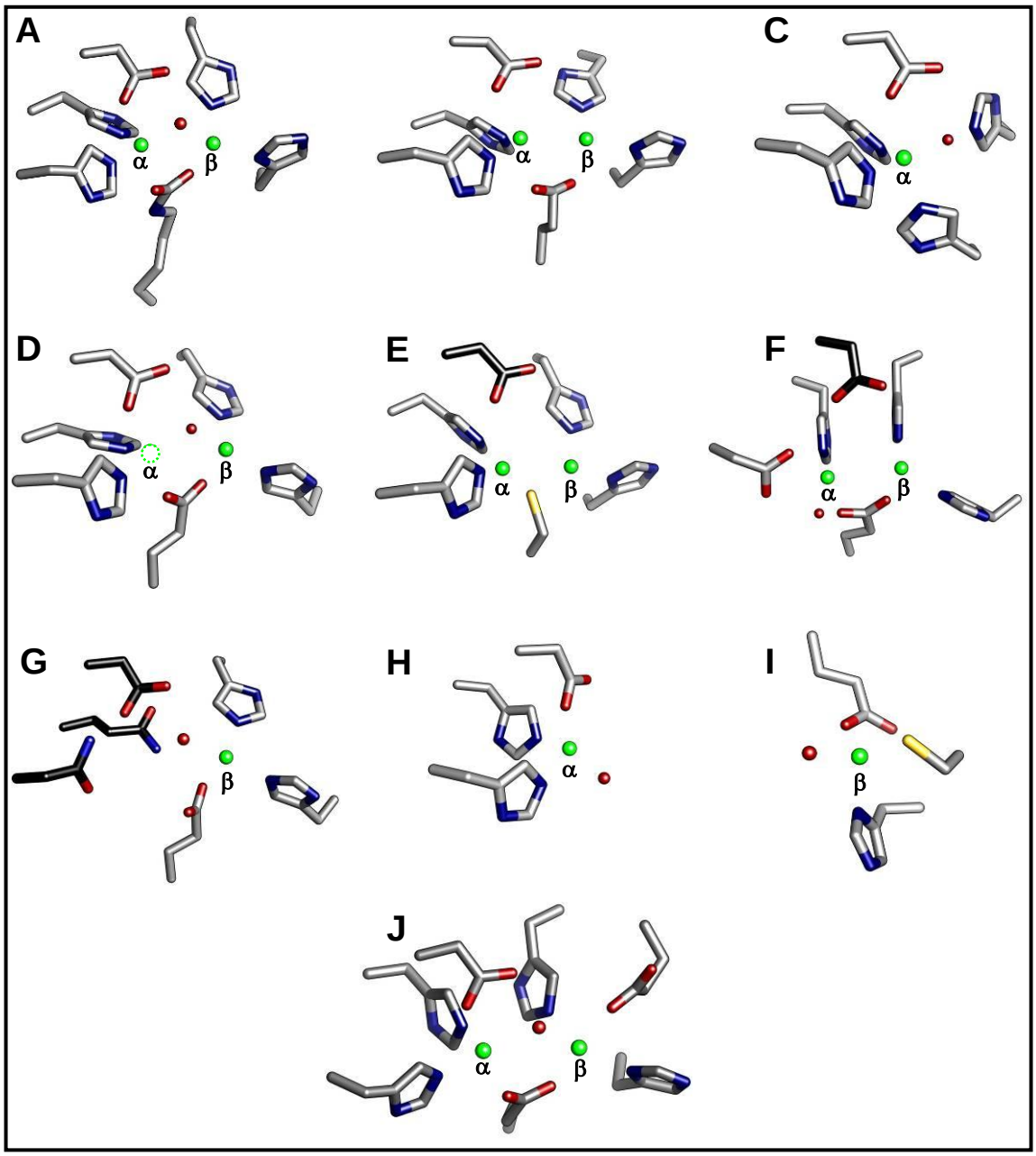


Figure 1.2: Stick representations of the ten metal class centers of the amidohydrolase superfamily. **(A)** Class 1 (dihydroorotase); **(B)** Class 2 (PTE homology protein); **(C)** Class 3 (cytosine deaminase); **(D)** Class 4 (*N*-acetyl-D-glucosamine-6-phosphate deacetylases = AGD, PDB code *2VHL*) where the open circle represents M_{α} (Fe) which is not present in the Zn-AGD from *Thermotoga maritima* (PDB code *1O12*); **(E)** Class 5 (D-aminoacylase), **(F)** Class 6 (renal dipeptidase-like), **(G)** Class 7 (AGD, PDB code *2P50*), **(H)** Class 8 (D-galacturonate isomerase), **(I)** Class 9 (tatD), **(J)** Class 10 (adenine deaminase, PDB code *3NQB*). The metals and water are represented as *green* and *red* spheres, respectively. Residues that are not metal ligands are represented with black carbons. Residues with *gray* carbons are metal ligands.

Amidohydrolase Superfamily: Roles of Divalent Cations and Enzyme Mechanism. The specific roles of M_α and M_β were studied with PTE. The Zn/Cd form of PTE is a single hybrid enzyme with Zn bound in the buried α site and Cd in the β site (*28*). The crystal structure of PTE with the bound inhibitor diisopropyl methylphosphonate (DIMP) has also been solved (*29*). Based on the crystal structure of PTE with bound DIMP and pH rate profile comparisons of Zn/Zn, Cd/Cd and Zn/Cd PTE, it has been proposed that M_α plays a dominant role in lowering the pK_a of water and M_β functions primarily to polarize the substrate (*28*). This is consistent with the crystal structure of dihydroorotase (DHO) with bound dihydroorotate where Zn_β is 2.9 Å from the carbonyl oxygen of the amide bond hydrolyzed by the enzyme (*19*). The crystal structure of isoaspartyl dipeptidase with the bound aspartate product shows how the carboxylate, formed by the cleavage of an amide bond, is positioned with respect to the two Zn ions (*30*). A catalytic mechanism was proposed for DHO where an activated water hydroxyl is the nucleophile and the β_8 aspartate bound to M_α is involved in acid/base catalysis. Aspartate acts as a base during nucleophilic attack on the carbonyl carbon of the amide bond in dihydroorotate to form a tetrahedral intermediate. The tetrahedral intermediate is stabilized by the two metals. Aspartic acid then acts as an acid during bond cleavage. The proposed catalytic mechanism for DHO is depicted in **Figure 1.3A** (*19*).

Uronate isomerase from *E. coli* is proposed to act by a different reaction mechanism as outlined in **Figure 1.3B**. The crystal structure of URI from *B. halodurans* has a single Zn present (**Figure 1.2H**), but a metal is not definitively present in the

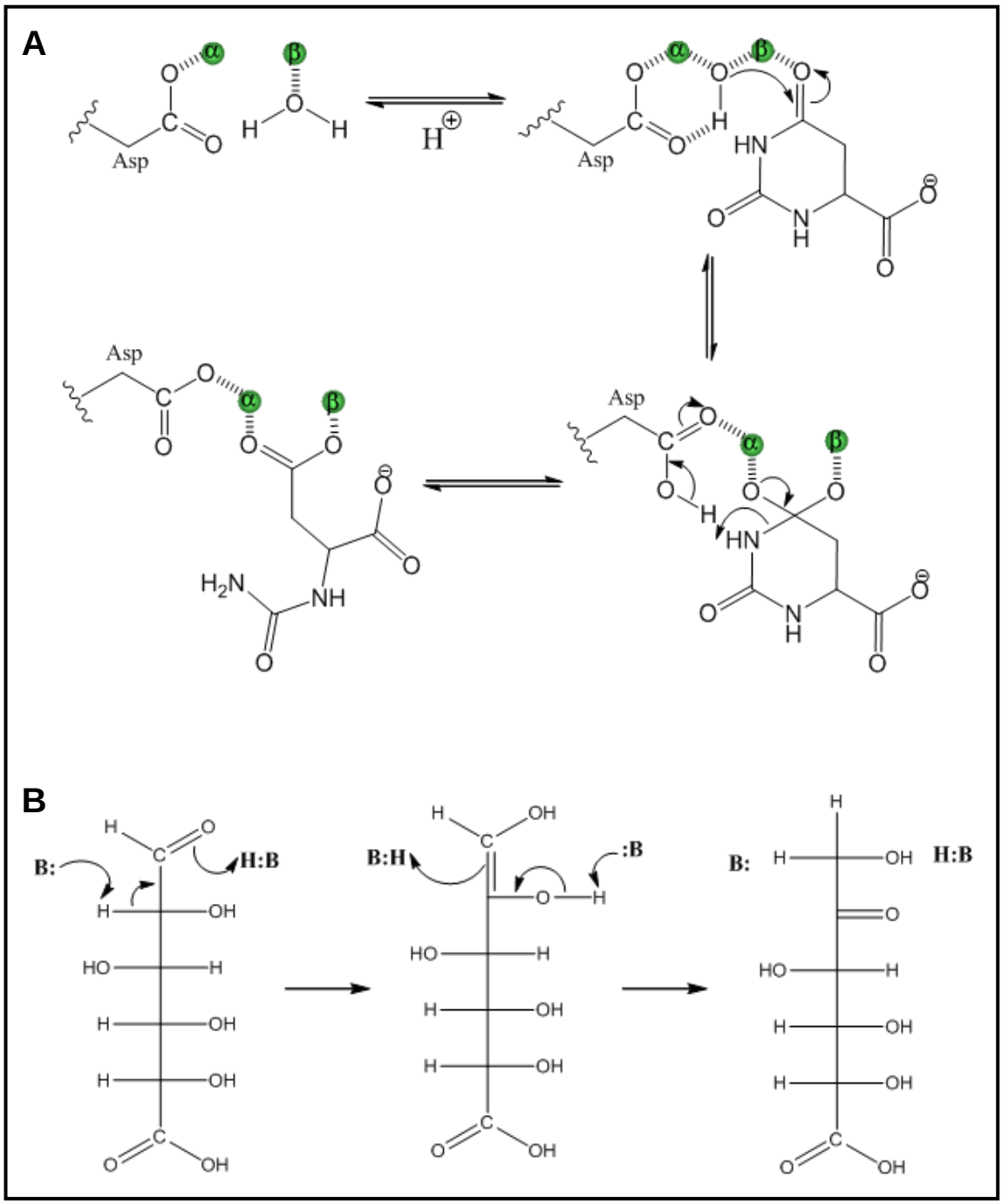


Figure 1.3: Proposed mechanism of the *E. coli* dihydroorotase and uronate isomerase.
(A) Proposed mechanism of the *E. coli* dihydroorotase. The α and β metals are represented as *green* spheres and the catalytic aspartate (Asp-250) is from β-strand 8 of the $(\beta/\alpha)_8$-barrel.
(B) Proposed catalytic mechanism of *E. coli* uronate isomerase.

uronate isomerase from *T. maritima*. Initially, it was not clear as to whether this enzyme required a metal ion for activity. The *E. coli* uronate isomerase was found to be metal-dependent, though the Zn's involvement in the 1,2-proton transfer may not be the conventional role proposed for M_α in other members of the AHS (*31*).

Members of enzyme superfamilies generally catalyze different reactions by similar reaction mechanisms. The structural fold and the fundamental active site components which allow for the chemistry to take place are present in the proteins of different enzyme families within the AHS. An amino acid sequence can be used as a tool to predict whether or not a protein is an amidohydrolase; however, without some groundwork, it is difficult to use primary structure to predict protein function. There was experimental evidence for the existence of the reactions catalyzed by the enzyme families predicted by Holm and Sander to be amidohydrolases. At that time, the bioinformatics era was beginning, a time when, typically, the 3D structure of an enzyme was determined after the substrate of the enzyme was known.

Amidohydrolase Superfamily: Functional Organization by Amino Acid Sequence. In 2004, the amino acid sequences of approximately 1,500 AHS members were extracted from ~ 200 fully sequenced genomes. The amino acid sequences of ~1,500 putative AHS members were placed into 345 separate groups defined as sets of sequences with ≥ 40% amino acid sequence identity. By 2006, this number had expanded to 5,752 putative AHS members (*32*). The amino acid sequences of 5,752 AHS members (AHs) were organized into 875 subgroups within 38 families. A graphical representation of the AHS members is shown in **Figure 1.4**, where clusters of

enzymes are separated based on a BLAST E-value cut-off of 1 x 10^{-10} (*13*). Now, in 2010, there are approximately 1200 sequenced genomes, ~ 3400 other genome sequencing projects in progress and ~ 11 million entries in the UniProtKB/TrEMBL database. As predicted from the genomic DNA sequences of the first ~ 1000 fully sequenced microbial genomes, there are ~ 11,000 genes which code for AHS proteins. Each AHS member belongs to one of 24 cogs (Clusters of Orthologous Groups). It is estimated that members of the AHS catalyze approximately 100 different reactions, but only ~ 38 have been identified thus far (*13, 21, 33*). Therefore, a significant fraction of these amino acid sequences belong to proteins or enzymes of unknown function, that is, there is no experimental evidence for their catalytic activity.

The number of potential AHS substrates is actually quite large even when only considering known metabolites. For AHS members which act by the same general mechanism as DHO, the main limitation as to the types of substrates they could accept is that the molecules have an electrophilic center which can be attacked by a hydroxyl group. The existence of uronate isomerases and their proposed reaction mechanism expands the group of potential substrates to include acid sugars. The prospect for further diversification of function within the AHS is evidenced by the gray areas in **Figure 1.4**. There are clusters of AHs that are both uncharacterized and not clustered with other proteins of known function.

Structure-Function Relationship of Broadly Annotated AHS Members. Some AHs are broadly annotated, and located within genome contexts which lend no further

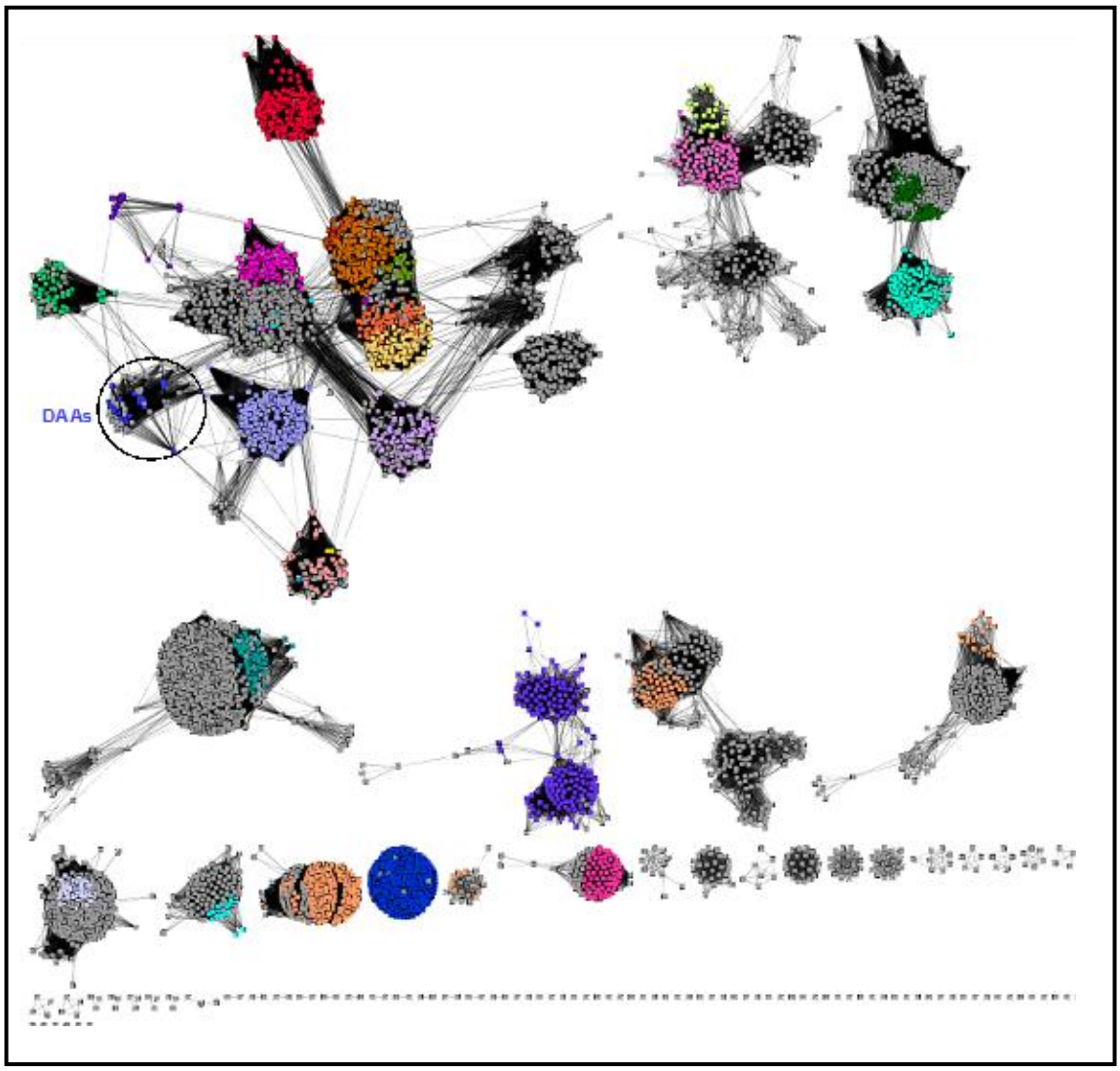


Figure 1.4: Cytoscape graphical representation of AHS members with a BLAST E-value cutoff of 1 x 10^{-10}. The D-aminoacylase cluster is circled.

information as to their function. One of these groups of AHs belongs to a set of enzymes imprecisely annotated as *N*-acetyl-D-amino acid deacetylases (DAAs) or *N*-acetyl-D-glutamate deacetylases (circled in **Figure 1.4**), but their substrates are not necessarily limited to an acetyl group as the *N*-acyl- derivative of D-amino acids (*34*). With such a broad annotation, the pool of potential substrates is greater than twenty. The biological significance of these enzymes is unclear. The genome contexts of DAAs are not helpful in establishing the substrate specificity of these enzymes. Furthermore, a majority of the DAAs which have been biochemically tested are from organisms whose genome sequence is not known.

It has been proposed that the active site of DAAs have four sites to accommodate the C_α hydrogen, *N*-acyl group, carboxy terminus and the amino acid side chain (*35*). There is one crystal structure for a DAA from *Alcaligenes faecalis*, which has a preference for *N*-acetyl-D-methionine, but a substrate, inhibitor or D-amino acid product are not bound in the active site (*36*). From the current information it is difficult to unequivocally assign which residues define the substrate specificity. There are DAAs which *preferentially* hydrolyze the *N*-acylated derivatives of D-Asp (*37*), D-Glu (*38, 39*), D-Met (*35, 40*), D-Phe (*41*), D-Val (*42*) and D-Leu (*43*). At least some of these DAAs can accept several *N*-acetyl-D-amino acids as substrates. These proteins can have high sequence identity but preferentially hydrolyze different *N*-acyl-D-amino acids. For example, *Alcaligenes xylosoxydans* subspecies A-6 has three biochemically tested DAAs with different specificities, namely *N*-acyl- derivatives of D-neutral amino acids (gi|3287876), D-aspartate (gi|3287879) and D-glutamate (gi|3287878) (*34, 37, 38, 44*).

The three DAAs from *Alcaligenes xylosoxydans* subspecies A-6 are considered paralogues which are homologous sequences that reside in the same genome. The pairwise sequence identities of these three proteins are 43-55%, and their overall sequence identity is 35%. The goal here was to purify selected DAAs, determine their substrate specificities and three-dimensional structures (in collaboration with Dr. Steve Almo's crystallography laboratory at Albert Einstein College of Medicine), and associate these differences with specific amino acid residues.

Five putative DAAs were chosen for this work, namely the putative DAAs Bb3285, Bb2785, Sco4986, bll7304 and Gox1177. These five proteins belong to four subgroups of DAAs when an E-value cutoff of 1×10^{-70} is used (as described in Chapter II). Of these, Bb3285 is the closest to the *A. faecalis* DAA of known structure (PDB ID: *1M7J*). The active site residues in the *A. faecalis* DAA are also conserved in Bb3285. Each of the other four targets were thought to have variations in active site residues based on 1) their amino acid sequence alignments within their respective subgroups of putative DAAs and 2) amino acid sequence alignments including other enzymes in the Bb3285 subgroup. Gox1177 and bll7304 were also specifically chosen from their subgroup because they were homologues of a protein thought to be involved in a small biochemical pathway.

Gerlt and coworkers discovered a pathway in *Geobacillus kaustophilus* with enzymes for the conversion of D-amino acids to L-amino acids. As outlined in **Figure 1.5**, the three enzymes in this pathway are *N*-succinyl-CoA:D-amino acid *N*-succinyltransferase (GNAT superfamily), *N*-succinylamino acid racemase (NAAAR)

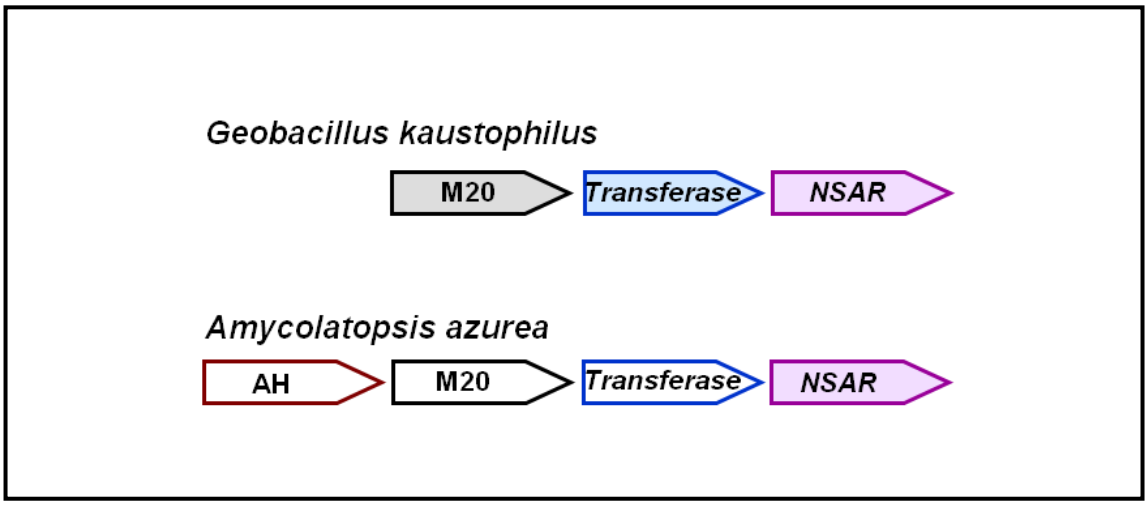


Figure 1.5: Genome context of an *N*-succinyl amino acid racemase in *Geobacillus kaustophilus* and *Amycolatopsis azurea*. *Top*: Three enzyme pathway for the conversion of D-amino acids to L-amino acids in *G. kaustophilus*. The genes code for an *N*-succinyl-CoA:D-amino Acid *N*-succinyltransferase (*blue*), *N*-succinylamino acid racemase (NSAR, *purple*) and *N*-succinyl-L-Amino acid hydrolase (M20, *gray*). *Bottom*: Genome context of the experimentally verified NSAR in *A. azurea* (*purple*). The unshaded M20 and transferase are homologues of the M20 hydrolase and transferase in *G. kaustophilus*. The putative AHS member is outlined in *red*.

and *N*-succinyl-L-amino acid hydrolase (M20 peptidase/carboxypeptidase G2 family). They noticed that in *Amycolatopsis azurea*, a GNAT superfamily protein, a homologue of the NAAAR enzyme and an M20 protein were in an operon context with a member of the amidohydrolase superfamily (*45*). The complete sequence of the gene for this AHS member in *A. azurea* is only partially sequenced, but the known portion codes for an amino acid sequence which is 38% identical to Bll7304 and Gox1177 over an alignment of 417 amino acid residues.

A second subset of AHS enzymes are putative dipeptidases/renal dipeptidase-like proteins (RDPs). Out of the ~1,500 AHS members within the first ~ 200 fully sequenced organisms, 62 amino acid sequences were organized into twelve groups of putative dipeptidases. The sequences within each group were ≥ 40% identical in amino acid sequence. Another 6 sequences had less than 40% sequence identity to any other sequence and were termed "singletons." A majority of these 68 proteins have the conserved residues outlined for the class 6 proteins in **Table 1.1**. However, a small subset of ten sequences was missing the HxD motif. The different clustering of groups of RDPs, using an E-value cutoff of 1×10^{-70}, will be discussed in Chapters III-V. The annotation "dipeptidase" is extremely broad; in only considering the twenty common amino acids and the stereochemistry at C_α, the pool of potential substrates is ~ 1,600. There is no consensus within the genome contexts of the RDPs in each group which would aid in narrowing down the list of potential substrates. A majority of these RDP annotations are given to microbial proteins which are homologous to the human RDP (hRDP).

Human RDP hydrolyzes a variety of dipeptides such as Pro-Leu, Pro-Ala, Gly-Leu, Leu-Gly, Phe-Gly, Gly-Phe, Trp-Gly, Gly-Val, Ala-Ala, Met-Met, Phe-Tyr, Gly-Gly, Gly-Ser, Gly-His and Leu-Leu (*46-48*). The crystal structure of human RDP (PDB ID: *1ITQ*) was solved prior to this work (*23*). Some, but not all, of the amino acid residues implicated as part of the active site are conserved to varying extents among the different groups of putative RDPs. One example of a microbial RDP was from *Brevibacillus borstelensis* which hydrolyzed the dipeptides L-Ala- L-Ala, D-Ala-D-Ala, L-Asp-D-Ala, L-Asp- L-Asp, and D-Ala-D-Glu, with a preference for L-Ala-D-Ala (*49*). However, the substrate specificity of RDP-like enzymes has never been tested with a large pool of potential substrates. We anticipated that RDPs would have different substrate specificities which could be linked to changes in active site residues.

In this work, the substrate specificities of selected DAAs and putative dipeptidases/RDPs were probed with substrate libraries. These libraries included *N*-acyl-amino acid (Xaa) libraries (*N*-acetyl-L-Xaa, *N*-acetyl-D-Xaa, *N*-succinyl-L-Xaa and *N*- succinyl-D-Xaa) and dipeptide libraries (L-Xaa-L-Xaa, L-Xaa-D-Xaa and D-Xaa-L-Xaa). The crystal structures of three other RDPs (PDB IDS: *2RAG, 3FDG* and *3LU2*) *without* substrate or product bound, were also solved during the time-frame of these experiments. The relationships between experimental substrate specificities, amino acid sequence alignments and crystallographic information will be discussed for both the DAA and RDP-like families of enzymes. Potential functions of these enzymes will also be discussed.

CHAPTER II

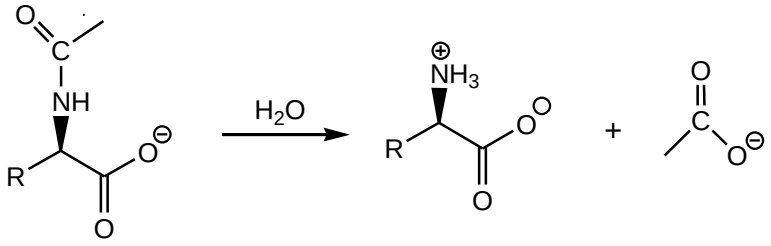
DEACYLATION OF D-AMINO ACIDS BY MEMBERS OF THE AMIDOHYDROLASE SUPERFAMILY

INTRODUCTION

The functional annotation of enzymes based solely on the amino acid sequence is a challenging task. As a consequence of the recent whole organism DNA sequencing efforts, the number of genes predicted to code for proteins far exceeds that of those which are known empirically. A majority of the completely sequenced organisms are microbes. It has been estimated that approximately one third of the encoded proteins have an *uncertain*, *unknown*, or *incorrect* functional assignment (*50*). This indicates that the breadth of the capabilities of the microbial proteome is not completely understood. In the pre-genomic era most enzymes were identified and characterized based almost entirely on the ability to follow and measure a catalytic activity. In the genomic era the ability to accurately infer the function of a protein based on its amino acid sequence is becoming increasingly important. In some cases, the amino acid sequences of uncharacterized proteins are homologous to enzymes with a known catalytic activity and specific metabolic roles. Still, the lower the sequence homology of an uncharacterized protein to any other protein with an established activity, the more difficult it is to accurately annotate its function. One approach to this problem has been to focus on the large number of enzymes of unknown function within the amidohydrolase superfamily (*4, 5, 10*).

The amidohydrolase superfamily (AHS) is a group of enzymes which accept an array of diverse substrates within the active sites of a $(\beta/\alpha)_8$-barrel structural fold (*20*). Over 11,000 protein sequences have been identified as members of the AHS among the more than 10 million gene sequences determined thus far. Members of this superfamily have been shown to catalyze the hydrolysis of organophosphate esters, lactones, and amides, in addition to decarboxylations, hydrations, and isomerization reactions (*20-22, 51*). However, a significant fraction of the members of this broad superfamily have an undefined substrate and reaction specificity (*32*). A group of homologous sequences that encode for enzymes of uncertain function has been tentatively annotated in public databases as *N*-acyl-D-amino acid deacylases. These enzymes catalyze the general reaction outlined in **Scheme 2.1**.

Scheme 2.1 General reaction catalyzed by *N*-acyl-D-amino acid deacetylases



The functional role of the *N*-acyl-D-amino acid deacylases is unclear. However, the addition of acylated D-amino acids in the growth medium of some microorganisms has been used to induce the production of these enzymes (*37, 43, 52*). D-Amino acids are associated with neurotransmission and hormone synthesis in mammals and

protection from proteolysis in bacteria (*53*). Certain organisms have the means to degrade D-amino acids (*e.g*. D-amino acid oxidase) and to prevent their incorporation into proteins by deacylating aminoacyl-tRNAs with a misincorporated D-amino acid (*53, 54*). D-Amino acids are also incorporated into key sites in some antibiotics (*55*). Enzymes have been identified which preferentially hydrolyze *N*-acylated derivatives of D-aspartate (*37*), D-glutamate (*38, 39*), D-methionine (*35, 40*), D-phenylalanine (*41*) and D-valine (*42*). These proteins can have high sequence identity to one another, but a structural explanation for the differences in the substrate specificity has not been fully elucidated.

The only structurally characterized member of this group of enzymes was a D-aminoacylase (DAA) from *Alcaligenes faecalis* DA1 (gi|28948588, PDB code *1M7J*). This enzyme can bind one or two divalent cations in the active site but only a single metal ion bound to the β-site is required for the expression of catalytic activity. The single divalent cation at the β-site is bound to Cys-96, His-220, and His-250 (*24*). It has been proposed that the nucleophilic water molecule is coordinated to the single divalent cation and hydrogen bonded to the side chain carboxylate of Asp-366 from β-strand 8. In the proposed catalytic mechanism, the carbonyl group of the amide bond is polarized via an interaction with the divalent cation bound to the β-site and the water molecule utilized for hydrolysis is activated via metal coordination and a hydrogen bonding interaction with the carboxylate group of Asp-366. Once the tetrahedral adduct is formed, the leaving group is activated via proton transfer from Asp-366 (*36*). This enzyme has been reported to preferentially hydrolyze *N*-acetyl-D-methionine (*35*).

Here, the relative substrate specificities for three putative *N*-acyl-D-amino acid deacylases from the amidohydrolase superfamily were determined (*56*). These proteins include Bb3285 from *Bordetella bronchiseptica*, Gox1177 from *Gluconobacter oxydans*, and Sco4986 from *Streptomyces coelicolor*. All of these proteins are currently annotated by NCBI as generic D-aminoacylases or more specifically as an *N*-acetyl-D-glutamate deacetylase. The protein Bb2785 from *B. bronchiseptica*, also investigated here, is annotated as a hypothetical protein but is similar in amino acid sequence to other proteins provisionally assigned as putative D-aminoacylases. A protein sequence alignment of Gox1177 with the structurally characterized D-aminoacylase from *A. faecalis* indicates that most of the residues in the active site are highly conserved although the sequence identity is only 23%. In contrast, the Bb3285 enzyme is 48% identical to the D-aminoacylase from *A. faecalis*. A sequence alignment of Gox1177, Bb2785, Bb3285, Sco4986 and the D-aminoacylase from *A. faecalis* is presented in **Figure 2.1.** Individual *N*-acyl-D-amino acids or substrate libraries containing nearly all combinations of L-Xaa-D-Xaa, *N*-acetyl-D-Xaa, *N*-acetyl-L-Xaa, *N*-succinyl-L-Xaa, and *N*-succinyl-D-Xaa were utilized to establish the substrate specificities for Bb3285, Gox1177, and Sco4986. It was demonstrated that Bb3285 preferentially hydrolyzes *N*-acetyl-D-glutamate and *N*-formyl-D-glutamate whereas Gox1177 and Sco4986 hydrolyze a rather broad spectrum of *N*-acyl-D-amino acids, including L-Xaa-D-Xaa dipeptides by Gox1177. *N*-methylphosphonate derivatives of D-amino acids were inhibitors of these enzymes with varied degrees of potency. The three-dimensional structure of Bb3285 was determined in the presence of the product, acetate, and also with the

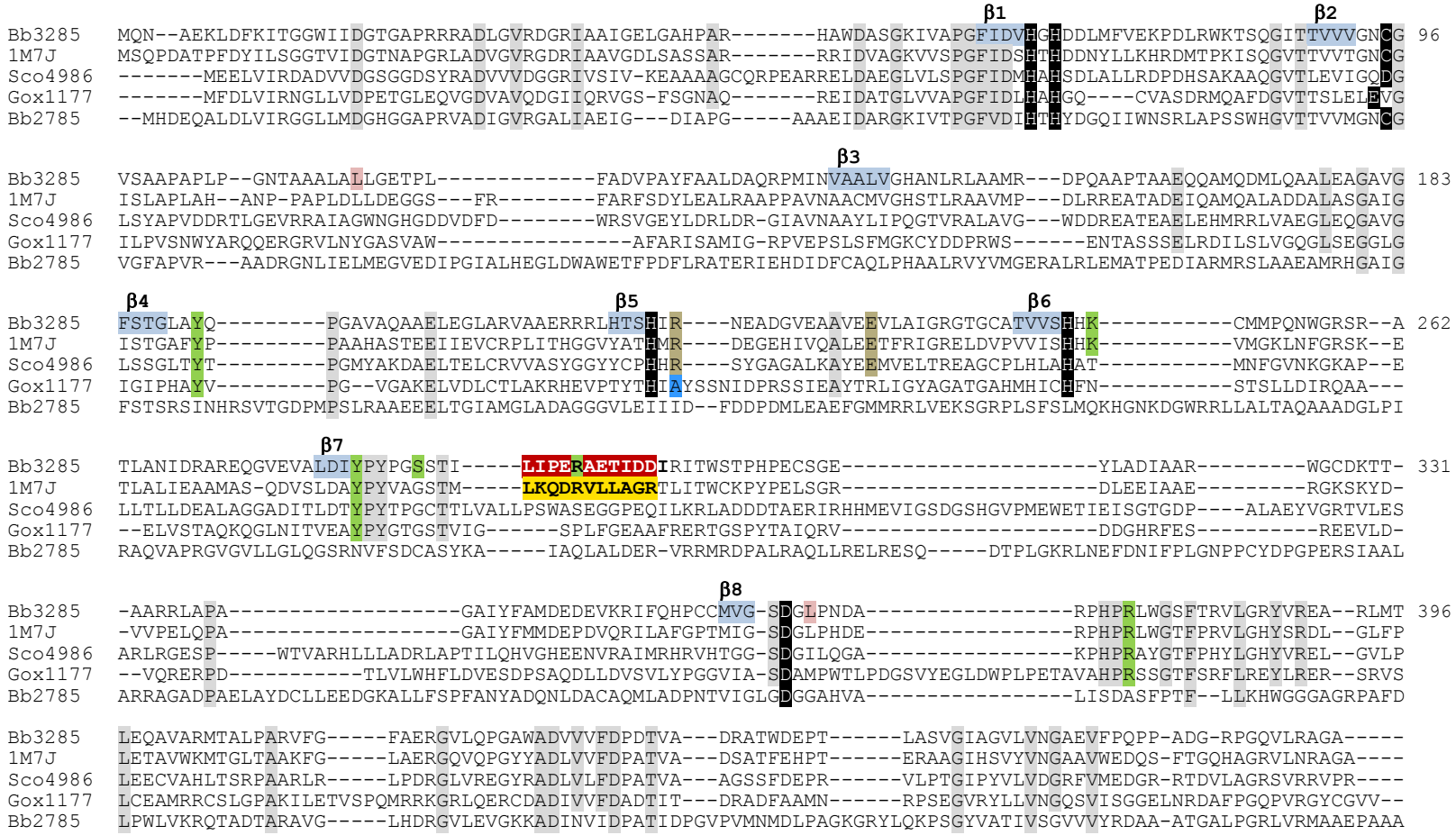


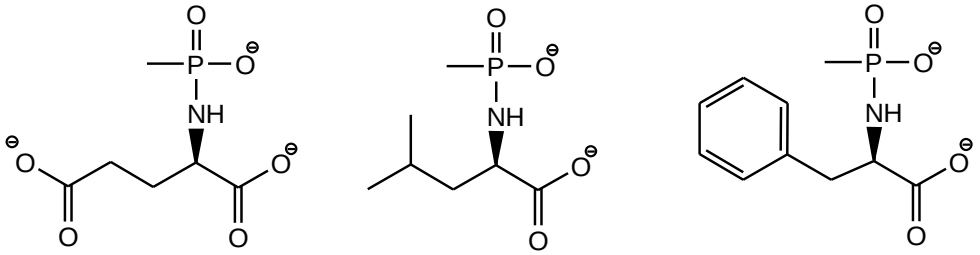
Figure 2.1: Amino acid sequence alignment of the DAA from *A. faecalis* (PDB code *1M7J*), Gox1177, Bb3285, Bb2785 and Sco4986. The catalytic aspartate and metal ligands are shaded *black*. Amino acid residues proposed to play a role in the recognition of the substrate in the active site are highlighted *green*. The variable substrate specificity loops in Bb3285 (291-301) and the DAA from *A. faecalis* (292-302) are highlighted in *red* and *yellow*, respectively. Residues which represent the β-strands of the $(\beta/\alpha)_8$-barrel are numbered and colored *light blue*. Residues conserved in the experimentally-verified enzymes are colored *gray*. Additional details are provided in the text.

methylphosphonate derivative of D-glutamate, a potent inhibitor resembling the tetrahedral intermediate formed during substrate hydrolysis (*56, 57*).

MATERIALS AND METHODS

Materials. Dr. Chengfu Xu synthesized all of the individual compounds and substrate libraries which were not purchased commercially. The *N*-acetyl derivatives of D-amino acids were synthesized as previously described (*56-58*). The syntheses of *N*-succinyl-D-Glu and *N*-formyl-D-Glu were conducted according to the method described by Sakai *et al.* (*45*). The dipeptide and *N*-acyl-amino acid libraries did not contain cysteine or D-isoleucine. The *N*-methylphosphonate derivatives of D-Glu (**1**), D-Leu (**2**) and D-Phe (**3**) (**Scheme 2.2**) and nineteen L-Xaa-D-Xaa libraries were synthesized according to the method previously described (*56, 57*). The compounds, *N*-acetyl- D/L-Asp, *N*-acetyl-D-Pro, and *N*-acetyl-D-Asn were purchased from Sigma. All other *N*-acetyl-D-amino acids were purchased from Novabiochem with the exception of *N*-acetyl-Gly (TCI America), *N*-acetyl-D-Arg (Avocado), and *N*-acetyl-D-Tyr (Bachem). Resins and protected amino acids used for solid phase peptide synthesis were purchased from

Scheme 2.2 *N*-methylphosphonate inhibitors



(1) **(2)** **(3)**

Calbiochem. ICP standards were obtained from Inorganic Ventures Inc. Enzymes and chemicals for the acetic acid coupling system assay were purchased from Megazyme. Nitrocefin was purchased from Calbiochem. Metal analyses were conducted using inductively coupled plasma mass spectrometry (ICP-MS) as previously described (*59*). Oligonucleotide syntheses and DNA sequencing were performed by the Gene Technologies Lab of Texas A&M University.

Amino Acid Analysis of N-Acyl-Xaa and Dipeptide Libraries. Amino acid analysis was performed as a service provided by the Protein Chemistry Laboratory facility at Texas A&M University's Department of Biochemistry and Biophysics. The composition of the substituted amino acid libraries was determined by amino acid analysis after acid-catalyzed hydrolysis (except for cysteine and tryptophan). The individual libraries were hydrolyzed by vapor phase 6 M HCl-2% phenol at 150 ºC for 90 minutes. The amino acids norvaline and sarcosine were added to the samples before hydrolysis as internal standards. A sample of human serum albumin was hydrolyzed under the same conditions as a control to ensure that complete hydrolysis of the mixture had occurred. Amino acid analysis was conducted with a Hewlett Packard AminoQuant System (HP 1090L). Samples were suspended in borate buffer, pH 10, and the free amino acids were derivatized with *o*-phthalaldehyde (OPA) and 9-fluoromethyl-chloroformate (Fmoc). The derivatized amino acids were separated on a 5 μm reverse phase HPLC column (Agilent 79916AA-572) using a gradient of 100% buffer A (20 mM sodium acetate/0.018% v/v triethylamine/50 mM ethylenediamine/0.3% v/v tetrahydrofuran, pH 7.2) to 60% buffer B (1:2:2 100 mM sodium acetate: acetonitrile:

MeOH) over 17 minutes at a flow rate of 0.45 mL/minute. The amino acid derivatives were detected using a photodiode array (UV-DAD) spectrophotometer or by fluorescence. The excitation/emission wavelengths for the OPA- and Fmoc- derivatized amino acids were 340/450 nm and 266/305 nm, respectively. Two sets of standards containing 2 or 5 nmol of each amino acid and the internal standards were run prior to the samples of unknown amino acid composition.

Cloning of Genes for Amidohydrolase Enzymes. The genes for five putative *N*-acyl-D-amino acid deacylases were cloned into the *Nde*I/*EcoR*I or *Nde*I/*Hind*III restriction sites of the plasmid pET-30a(+) (Novagen). The gene that encodes the protein Bb3285 (gi|33602261) was amplified from the genomic DNA of *B. bronchiseptica* (ATCC BAA-588) using the primers 5´-GCAGGAGCCATATGCAGA ACGCGGAAAAGCTGGATTTCAAGATTACCGG-3´ and 5´-CGCGGAATTCAGGC GCCGGCGCGCAGTACC-3´. The gene for the Bb2785 (gi|33601761) was also cloned from the *B. bronchiseptica* genomic DNA with the primers 5´-GCAGGAGCCATATGC ACGACGAACAGGCACTTGATCTTGTCATCCGC-3´ and 5´-CGGCAAGCTTCAG GCGGCCGCGGGCTCGGCC-3´. The gene that encodes the protein Gox1177 (gi|58039630) was amplified from the genomic DNA of *G. oxydans* using the primers 5´-GCAGGAGCCATATGTTTGATCTCGTGATCCGTAATGGTCTGCTGG-3´ and 5´-CGCGGAATTCACACCACCCCGCAATACCCACGG-3´. The genomic DNA of *G. oxydans* was a generous gift of Dr. Armin Ehrenreich from the Georg-August University, Göttingen, Germany. The gene for Sco4986 (gi|21223359) was cloned from the genomic DNA of *S. coelicolor* (ATCC BAA471-D) with the primers 5´-GCAGGAG

CCATATGGAAGAGCTGGTCATCAGGGACGCCGACGTCGTGGACGG-3´ and 5´-CGCGGAATTCAGCGCGGAACCCGGCGCACCGACCGTCCGG-3´. The gene for Bll7304 (gi|27382415) was cloned from *Bradyrhizobium japonicum* USDA 110 with the primers 5´-CGAGGAGCATTAATGCATGCTGCATCGGCATCCGCGGCGGTG-3´ and 5´CGCGGAATTCAGCCCTCGACGACCGGGCGGCGC-3´. The genomic DNA of *B. japonicum* was kindly donated by Dr. Hans-Martin Fischer from the Swiss Federal Institute of Technology, Zürich, Switzerland. The DNA polymerase chain reactions were performed using *Pfx* Platinum polymerase from Invitrogen. The sequences of the cloned genes were confirmed by DNA sequencing using the facilities from the Gene Technologies Laboratory of Texas A&M University.

Expression and Purification. E. coli BL21(DE3) cells harboring the pET-30a(+) plasmids containing the gene for the expression of Bb3285 or Bb2785 and Rosetta 2 (DE3) cells containing a pET-30a(+) plasmid for the expression of Gox1177 or Sco4986 were grown in Terrific Broth with 50 μg/mL kanamycin (additionally 25 μg/mL chloramphenicol for Rosetta 2 cells) at 30 ºC. When the optical density of the cells reached an OD_{600} of ~0.6, 2.5 mM zinc acetate was added to the culture and expression of Bb3285 was induced with 100 μM isopropyl *β*-thiogalactoside (IPTG) or 1.0 mM IPTG for the expression of Bb2785, Gox1177 and Sco4986. Following induction, the cells were grown overnight at 16-19 ºC. Cell lysis was achieved via sonication in 50 mM Tris, pH 7.5, and 100 μg/mL phenylmethanesulfonyl fluoride (PMSF) at 0 ^{o}C. The DNA present in the cell lysate was precipitated with protamine sulfate (2% w/w of cell mass) and removed by centrifugation. Solid ammonium sulfate was added slowly to the

supernatant solution to a saturation of 60% and the protein isolated after centrifugation. The enzymes were further purified by size exclusion chromatography on a 3 L (1 L for Bb3285) column of AcA34 gel filtration medium using 50 mM Tris (Hepes for Bb3285) buffer, pH 7.5, at 4 ºC. The fractions which contained the protein of the expected size were combined. The Gox1177 and Bb2785 proteins were further purified by anion exchange chromatography on a 6 mL Resource Q column from Pharmacia in 50 mM Tris, pH 7.5. The protein was eluted with a 0-1 M gradient of NaCl. The NaCl was removed in a final gel filtration step using a Superdex 200 size exclusion column in 50 mM Hepes, pH 7.5.

A total of 70 mg of Bb2785, 150 mg of Gox1177 and 142 mg of Bb3285 were isolated from 41, 36 and 66 grams of cells, respectively. The three proteins were greater than 95% pure as assessed by SDS-PAGE. The sequences of the first five amino acids of Gox1177 and Bb3285 were determined by the Protein Chemistry Laboratory at Texas A&M University. The two sequences matched those expected from the DNA sequence. The proteins Bb3285, Bb2785 and Gox1177 were found to contain 2.1, 1.8, and 1.8 equivalents of Zn as determined by ICP-MS. The protein Bll7304 was not purified since there was no expression from either BL21 (DE3) or Rosetta 2 (DE3) cells.

The Sco4986 protein was expressed in Rosetta 2 cells but the enzyme was found predominantly in the insoluble pellet after centrifugation. Rosetta 2 cells harboring the pET-30a(+) plasmid alone or the pET-30a(+)-Sco4986 plasmid were cultured under identical growth conditions and subsequently lysed via sonication in the presence of 50 mM Hepes, pH 7.5, 100 μg/mL PMSF. After centrifugation to remove insoluble

proteins, the small molecules in the lysate (MW <30 kDa) were separated via ultrafiltration using a PM-30 membrane (Millipore). The lysates were frozen and stored at -80 ºC. The protein concentration in the crude cell lysates was determined by the Bradford assay (Biorad) with 4-28 μg/mL bovine serum albumin as a protein standard.

Alternatively, Sco4986 was expressed in BL21 (DE3) cells harboring the pET-30a(+)-Sco4986 and pOFXbad-SL3 plasmids (*60*). The cells were grown under the same conditions in the presence of 50 μg/mL kanamycin and ampicillin. The pOFXbad-SL3 plasmid was a generous gift from Olivier Fayet at the Laboratory for Microbiology and Molecular Genetics in Toulouse, France. The pOFXbad-SL3 plasmid contains the genes for two molecular chaperones, groES/groEL, but overexpression of these genes was not induced. The cells expressing Sco4986 (0.2 mM IPTG) in the presence of the chaperone plasmid were used in the partial purification of Sco4986. Sco4986 was partially purified, through ion exchange, in the same fashion as Gox1177, with the exception of the sonication buffer (50 mM Hepes, 10% glycerol, 100 mM NaCl, 5 mM dithiothrietol, 100 μg/mL PMSF pH 8.5), the amount of protamine sulfate (0.4 % w/w cell mass) and the saturation level of ammonium sulfate (50-70% saturation fraction). The *N*-acetyl-D-phenylalanine deacetylase activity of the initial purification steps as well as gel filtration fractions was used to identify the protein. The ion exchange fraction of the highest activity was submitted to the Protein Chemistry Laboratory for amino acid sequence analysis. The components of the protein sample were separated by SDS-PAGE and blotted onto a PVDF membrane. The protein band which corresponded to the expected size (58 KDa) was removed and used for amino acid sequence analysis.

The amino acid sequence of this band was identical to that expected for amino acids 2-5 of Sco4986.

Crystallization and Data Collection. Crystallization screens and data collection were conducted by Alexander Fedorov and Lena Fedorov in Dr. Steve Almo's lab at Albert Einstein College of Medicine. Crystals of two different complexes were grown by the hanging drop method at room temperature for Bb3285 from *B. bronchiseptica*: (1) wild-type Bb3285, Zn^{2+}, acetate, and formate, and (2) wild-type Bb3285, Zn^{2+}, and inhibitor **1**. The crystallization conditions utilized the following conditions: for Bb3285, Zn^{2+}, acetate, and formate, the protein solution contained Bb3285 (12.6 mg/mL) in 20 mM Hepes (pH 8.0), 30 mM NaCl, and 0.05 mM $ZnCl_2$; the precipitant contained 2.0 M sodium formate and 0.1 M sodium acetate (pH 4.6). For Bb3285, Zn^{2+}, and inhibitor **1**, the protein solution contained Bb3285 (12.6 mg/mL) in 20 mM Hepes (pH 8.0), 30 mM NaCl, 0.05 mM $ZnCl_2$, and 10 mM methylphosphonate inhibitor; the precipitant contained 1.0 M ammonium phosphate, and 0.1 M sodium citrate (pH 5.6).

Prior to data collection, the crystals of two Bb3285 forms were transferred to cryoprotectant solutions composed of their mother liquids and 20% glycerol and flash-cooled in a nitrogen stream. All X-ray diffraction data sets were collected at the NSLS X4A beamline (Brookhaven National Laboratory) on an ADSC CCD detector.

Substrate Specificity of Bb3285, Gox1177, Sco4986 and Bb2785. The substrate specificities for Bb3285 and Gox1177 were determined by mixing each protein with the *N*-acetyl- and *N*-succinyl-amino acid libraries. The assays were conducted at 30 ºC in 25 mM ammonium bicarbonate, pH 8.0, and the concentration of each amino acid

derivative in these 17-20 member libraries was ~100 μM. The *N*-acetyl-D-Xaa, *N*-acetyl-L-Xaa, *N*-succinyl-D-Xaa, *N*-succinyl-L-Xaa, Gly-D-Xaa, L-Ala-D-Xaa, and L-Asp-D-Xaa libraries were incubated for 1-24 hours in the absence and presence of variable amounts of Gox1177 or Bb3285 (1-2000 nM). The reactions were quenched by removing the enzyme with a Microcon YM-10 (Millipore) membrane. A sample equivalent to ~10 μg of the initial *N*-acyl-Xaa library was dried under vacuum prior to submission to the Protein Chemistry Laboratory of Texas A&M University for determination of the liberated amino acids after the addition of each enzyme. The chromatographic peaks from the samples of unknown amino acid composition were identified by the migration time of the derivatized amino acids from known standards. The amino acids were quantified by integration of the chromatographic peaks. The relative substrate turnover rate in each library was determined by plotting the concentration of each liberated amino acid, *Q*, as a function of enzyme concentration (E_t) at a fixed period of time, fit to equation 1, where *A* is the total concentration of each substrate and *k* is the rate constant for product formation. Additionally, single 24-hour incubations of Bb3285 or Gox1177 with the Gly-D-Xaa, L-Ala-D-Xaa and L-Asp-D-Xaa libraries were prepared and submitted for amino acid analysis as described above.

$$Q = A(1 - e^{-k(E_t)}) \qquad (1)$$

Nineteen L-Xaa-D-Xaa dipeptide libraries were incubated at 30 ºC with 0-2000 nM Gox1177 or Bb3285 for 3 hours and 1 hour, respectively. The concentration of each dipeptide component in these libraries was ~100 μM in 50 mM Hepes, pH 7.5. The free amino acids were quantified with ninhydrin by measuring the change in absorbance at

507 nm and the rates of hydrolysis as a function of enzyme concentration were compared for each L-Xaa-D-Xaa library.

For the determination of the substrate specificity of Sco4986, 18 individual *N*-acetyl-D-amino acids (except proline and isoleucine) at 1.0 mM were incubated with the cell lysates from the pET-30a(+) and pET-30a(+)–Sco4986 transformations. The final protein concentration from the cell lysates was 0.4 mg/mL in 50 mM Hepes, pH 7.5 and the solutions were incubated for 14 hours at 30 ºC. In the same manner, these 18 substrates were treated with 2 μM Bb2785 for 20 hours. Under the same conditions, the rate of hydrolysis of the *N*-acetylglycine and the *N*-acetyl- derivatives of D/L-Asp, D-Cys, D-Ala, D-Glu, D-Phe, D-His, D-Lys, D-Leu, D-Met, D-Asn, D-Gln, D-Arg, D-Ser, D-Thr, D-Val, D-Trp and D-Tyr were also measured under the same conditions with 1.0 mM substrate and 0.15 mg/mL of the cell lysate containing Sco4986.

Measurement of Kinetic Constants. The deacylation of *N*-formyl-D-Glu, *N*-acetyl-D-Glu, and *N*-succinyl-D-Glu by Bb3285 was assayed by coupling the formation of D-glutamate with the ninhydrin reagent in a manner similar to that previously reported (*61*). The assays were conducted in 50 mM Hepes, pH 7.5, at 30 ^{o}C. Specifically, one volume of the enzymatic assay was quenched with four volumes (two volumes for all dipeptides) of the ninhydrin reagent. The ninhydrin reagent consisted of 0.9 % w/v ninhydrin and 1:10:80 of 1g/mL $CdCl_2$: acetic acid : EtOH. The quenched samples were heated at 85 ºC (80 ºC for all dipeptides) for 15 minutes (5 minutes for all dipeptides) and the reaction product was detected at 507 nm. The kinetic parameters for the

hydrolysis of *N*-formyl-D-Phe and the *N*-acetyl- derivatives of D-Phe, D-Tyr, D-Met and D-Trp by the Sco4986 lysate were determined using the same ninhydrin assay.

The deacetylation of the *N*-acetyl-amino acid derivatives by Gox1177 was monitored by coupling the formation of acetic acid to the formation of NADH with acetyl-CoA synthetase (1.1 U/mL, citrate synthase (2.3 U/mL) and malate dehydrogenase (16 U/mL) in the presence 5.0 mM NAD^+, 3.8 mM L-malate, 3.1 mM ATP, 3.2 mM $MgCl_2$, and 148 μM CoA in 130 mM triethylamine, pH 8.4. In the case of *N*-acetyl-D-His, *N*-acetyl-D-Thr, *N*-acetyl-D-Gln, Leu-D-Leu, Met-D-Leu and Tyr-D-Leu, the reactions were monitored by measurement of the liberated amino acid with the ninhydrin reaction. During the course of the kinetic assays of Gox1177 with dipeptide substrates, the ninhydrin reagent was kept at 0 ^oC prior to heating.

The kinetic constants were determined by a fit of the data to equation 2 where *v* is the velocity of the reaction, E_t is the total enzyme concentration, k_{cat} is the turnover number, A is the substrate concentration, and K_a is the Michaelis constant.

$$v/E_t = k_{cat}A/(K_a + A) \qquad (2)$$

Inhibition of Gox1177 and Sco4986 by N-Methylphosphonate Modified Amino Acids. The enzymatic activity of Gox1177 was monitored by the acetic acid coupling assay in the presence of 0-90 μM of *N*-methyl phosphonyl-D-Leu (**2**) at a substrate concentration of 3.0 mM *N*-acetyl-D-Leu. The partially purified lysate containing Sco4986 (0.1 mg protein/mL in the final assay) was pre-incubated with 0-5 μM of *N*-methyl phosphonyl-D-Phe (**3**) for 90 minutes at 30 ºC. The Sco4986 assays were initiated by the addition of 1.5 mM *N*-acetyl-D-Phe (5% of the total assay volume) and

the rate of D-phenylalanine formation was measured with the ninhydrin assay. The competitive inhibition constants were obtained from a fit of the data to equation 3.

$$v / E_t = k_{cat} A / (K_a (1 + (I / K_i)) + A) \qquad (3)$$

Inhibition of Bb3285 by N-Methylphosphonate Modified D-glutamate. For all inhibition studies with Bb3285, the deacylation of *N*-formyl-D-glutamate was coupled to the formation of NADH at 340 nm using 1.2 U/mL formate dehydrogenase (Sigma) in the presence of 10 mM NAD^+ (*62*). In the determination of the K_i, assay mixtures were prepared 30 and 75 minutes prior to running the assay. Inhibitor stocks were prepared by serial dilutions to make 10X inhibitor stocks ranging from 2.9-600 nM for assays. The assay mixtures contained all assay components (NAD^+, Bb3285, inhibitor **1**, and Hepes buffer, pH 7.5) except substrate in a total of 225 µL (90% assay volume). Following incubation at 30 ºC for 30 or 75 minutes, the 250 µL reaction was initiated with 25 µL *N*-formyl-D-Glu to a final concentration of 3 mM. The control reaction contained water in place of inhibitor and was used to obtain v_0, the velocity in the absence of inhibitor. Using the data from the 75-minute incubation samples, the velocity values at the different inhibitor concentrations, v_i, were divided by the value of v_0, plotted as a function of inhibitor concentration, [I], and fit to equation 4 (*57, 63*).

$$v_i/v_0 = 1 - ([E_t] + [I] + [K_i]) - (([I] + K_i + [E_t])^2 - (4[E_t][I]))^{1/2}/(2[E_t]) \qquad (4)$$

The first order rate constant for the onset of inhibition, k_{obs}, was determined by measuring the formation of NADH after the addition of Bb3285 to assay mixtures containing varying amounts of inhibitor **1** (10 to 300 nM). The concentration of Bb3285 (1 nM) and substrate (1.5 mM *N*-formyl-D-Glu) were constant. The OD_{340nm} (*P*) was

monitored continuously and the timecourse data were fit to equation 5 to obtain k_{obs} at varied concentrations of **1**, I, where t is time, and v_i and v_s are the initial and final steady velocities, respectively. If the inhibition of Bb3285 by **1** follows that of the model outlined in **Scheme 2.3**, a fit of the k_{obs} data as a function of the concentration of inhibitor to equation 6 can be used to obtain k_3 and the apparent (k_2/k_1). The data for the nonlinear relationship between k_{obs} and inhibitor concentration were fit to equation 6. The value for $(k_2/k_1)^{app}$ was adjusted for competition with substrate using equation 7. The value used for the K_m (230 μM) in equation 7 was determined using the same coupling assay (*57*).

$$P = v_s t + ((v_i - v_s)/k_{obs})(1 - e^{(-k_{obs}t)}) \qquad (5)$$

Scheme 2.3 Model for binding of inhibitor **1** to Bb3285

$$E \underset{k_2}{\overset{k_1[I]}{\rightleftharpoons}} EI \underset{k_4}{\overset{k_3}{\rightleftharpoons}} E^*I$$

$$k_3 = (k_{obs} - k_4)(1 + ((k_2/k_1)^{app}/[I])) \qquad (6)$$

$$(k_2/k_1)^{app'} = (k_2/k_1)^{app}/ (1 + ([S]/K_m)) \qquad (7)$$

Network Analysis. The amino acid sequences which are annotated as *N*-acyl-D-amino acid deacetylases make up cog3653. There are ~ 240 amino acid sequences in cog3653 that originate from the predicted protein sequences from the first ~ 1000 completely sequences microbial genomes. An all-by-all BLAST of the putative *N*-acyl-D-amino acid deacetylases (members of cog3653) was performed using the program MUSCLE (*64*). A protein similarity network based on the BLAST results was generated

using Cytoscape (*65*). The nodes (representatives of a single amino acid sequence) were arranged using the yFiles organic layout provided with Cytoscape version 2.4. Connections between nodes are shown as lines if the E-value of the best BLAST hit between two sequences is at least as good as 1×10^{-70}.

Sequence Alignment. The groups of sequences from each of clusters 1, 2, 4 and 5 were aligned separately using CLUSTALW in Biology Workbench from the San Diego Supercomputer Center (http://workbench.sdsc.edu). Sequences within clusters 1, 2 and 5 which did not have the "YPY" motif after β_7 or the nearly-conserved tyrosine from β_4, and sequences that were too short (gi|163839586) or too long (gi|32471748 and gi|94969636) were not used. A CLUSTALW profile alignment was then performed with the alignments for clusters 1 and 2. A second CLUSTALW profile alignment was conducted with the alignment of the cluster 4 sequences and the combined alignment of clusters 1 and 2. Finally, in a third CLUSTALW profile alignment, the aligned sequences from cluster 5 were aligned onto the profile alignment of clusters 1, 2 and 4. All of the amino acid sequences not displayed in **Figure 2.1** as well as extra gaps were removed.

RESULTS

Substrate Specificity of Bb3285. The determination of the substrate specificity of Bb3285 was initiated with the *N*-acetyl-D-Xaa library. Various concentrations of enzyme (0 – 0.1 μM) were mixed with a fixed concentration of the substrate library and allowed to react for 1 hour. The formation of free amino acids was measured with the ninhydrin assay and the results plotted as shown in **Figure 2.2A**. The maximum

absorbance change at 507 nm was ~0.19. If the entire substrate library had been hydrolyzed, the change in absorbance would have been ~3.5 and thus under these reaction conditions only a small fraction of the initial substrate library is hydrolyzed. A sample of the partially hydrolyzed reaction mixture was subjected to amino acid analysis to determine the identity of the *N*-acyl-amino acids that are functional substrates for Bb3285. The chromatogram is presented in **Figure 2.3** as a continuous red line. A control reaction was conducted in the absence of Bb3285 and the chromatogram is shown in **Figure 2.3** as a series of black dots. Relative to the control sample there is only a single amino acid, D-glutamate, that is formed from the hydrolysis of the *N*-acetyl-D-Xaa library after the addition of Bb3285. Therefore, the only compound to be hydrolyzed in this substrate library is *N*-acetyl-D-glutamate. The other members of this *N*-acetyl-D-amino acid library are hydrolyzed at less than 1% of the rate observed for *N*-acetyl-D-glutamate.

Bb3285 was also utilized as a catalyst for the hydrolysis of the *N*-succinyl-D-Xaa library and the results are presented in **Figure 2.2B**. Relative to the *N*-acetyl-D-Xaa library, the rate of hydrolysis is measurably slower. When the concentration of Bb3285 is increased 20-fold to 2.0 μM, the fraction of the *N*-succinyl-D-Xaa library that is hydrolyzed is approximately equal to that of the *N*-acetyl-D-Xaa library as illustrated by the red circles in **Figure 2.2B**. Under these reaction conditions there is no evidence for the hydrolysis of the *N*-acetyl-L-Xaa or *N*-succinyl-L-Xaa substrate libraries. Bb3285 was

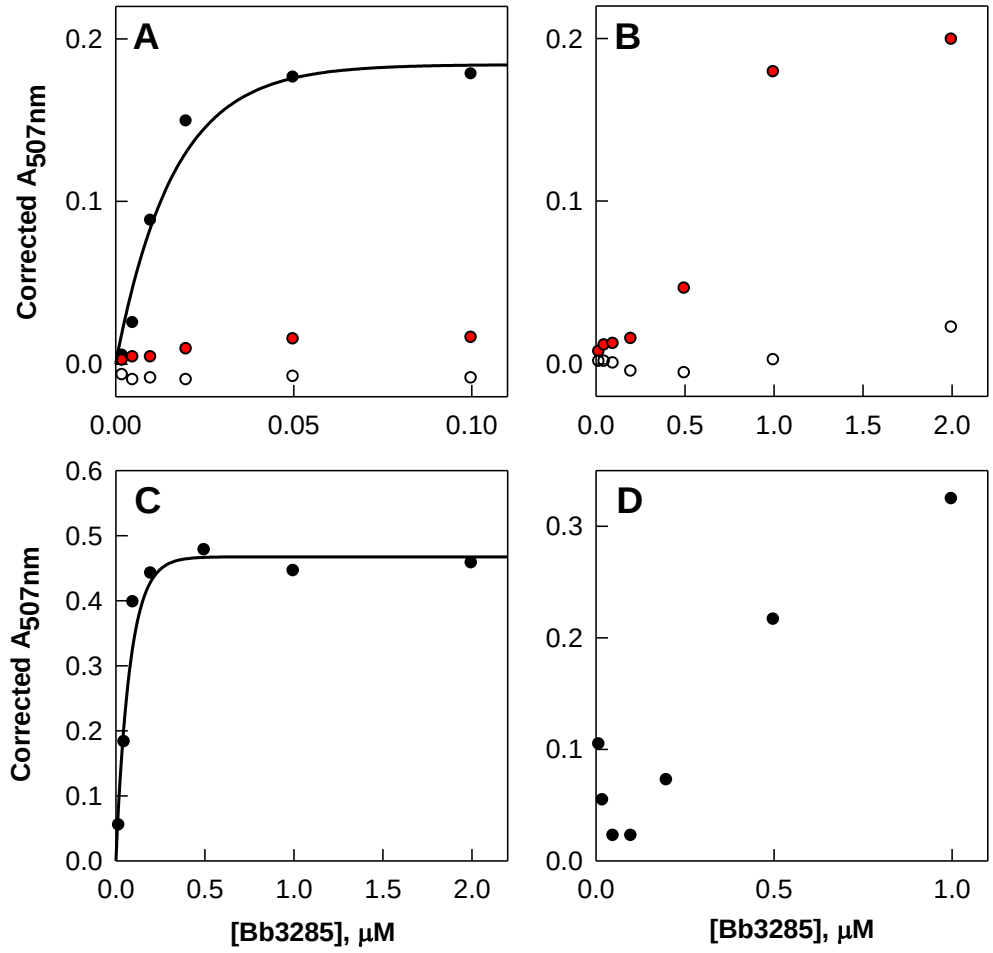


Figure 2.2: Enzyme courses for the hydrolysis of various substrate libraries by Bb3285. **(A)** *N*-acetyl-D-Xaa (*black* circles), *N*-acetyl-L-Xaa (open circles), and *N*-succinyl-D-Xaa (*red* circles). **(B)** *N*-succinyl-D-Xaa (*red* circles) and *N*-succinyl-L-Xaa (open circles). **(C)** L-Ala-D-Xaa. **(D)** L-Asp-D-Xaa. The enzyme concentration is shown along the abscissa. In panels A, B, and C, the enzyme was incubated with the substrate library for 1 hour whereas in panel D, the substrate library was incubated with the enzyme for 24 hours at 30 ^{o}C. Additional details are provided in the text.

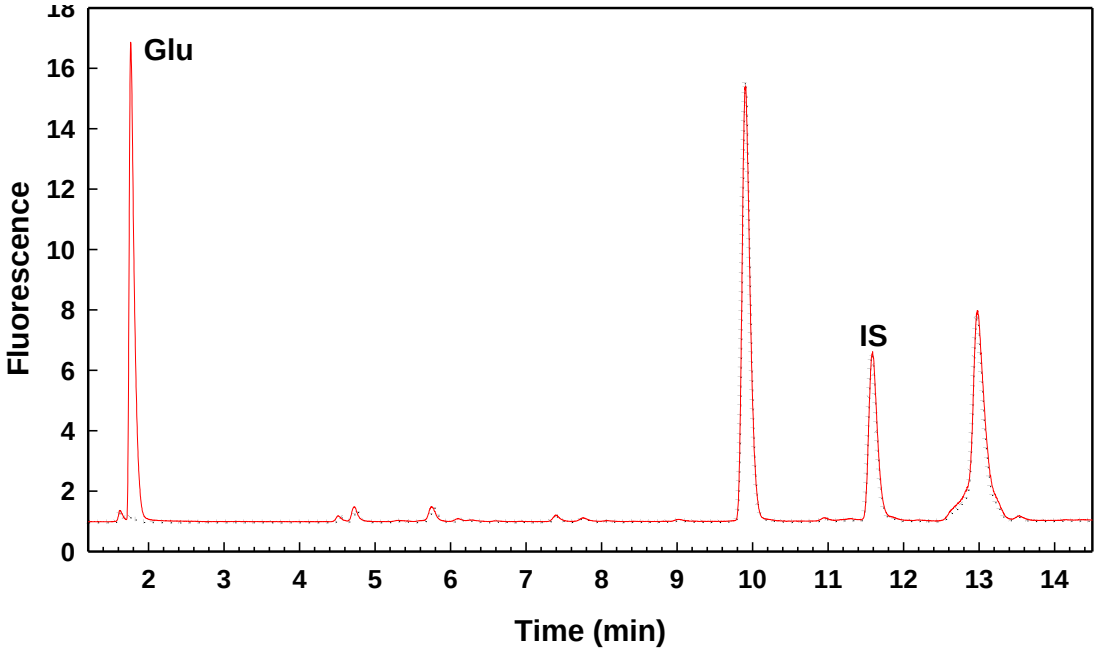


Figure 2.3: HPLC chromatogram of the *N*-acetyl-D-Xaa library treated with no enzyme (*black* dots) and 20 nM Bb3285 (*red* line) for 1 hour at 30 ºC. The OPA-derivatized D-glutamate was detected at a retention time of 1.7 minutes in the sample treated with Bb3285. The internal standard is labeled as IS. Additional details are provided in the text.

tested with two L-Xaa-D-Xaa dipeptide libraries and the results are shown in **Figures 2.2C** and **2.2D**. With the L-Ala-D-Xaa dipeptide library, 1-2 dipeptides are hydrolyzed in 1 hour at an enzyme concentration of 0.2 μM. The hydrolysis of the L-Asp-D-Xaa library is considerably slower. Amino acid analyses of these substrate libraries indicate that only those compounds with a D-glutamate at the carboxy terminus are substrates for Bb3285 (data not shown). Bb3285 was subsequently screened for dipeptidase activity towards eighteen L-Xaa-D-Xaa libraries. The relative rates of hydrolysis for these libraries are summarized in **Table 2.1**.

The kinetic constants were determined for the hydrolysis of *N*-formyl-, *N*-acetyl-, *N*-succinyl-, *N*-Met-, and *N*-Leu-derivatives of D-glutamate by Bb3285 at pH 7.5. The kinetic constants from fits of the data to equation 2 are presented in **Table 2.2**. Of the compounds tested, *N*-formyl-D-glutamate exhibited the highest values for k_{cat} (2200 s^{-1}) and k_{cat}/K_m (5.6 x 10^6 M^{-1} s^{-1}). Somewhat lower values were obtained for *N*-acetyl-D-glutamate.

Substrate Specificity of Gox1177. The substrate specificity for Gox1177 was initially interrogated with the *N*-acetyl-D-Xaa library. The enzyme (0-2 μM) was incubated for three hours with the *N*-acetyl-D-Xaa library, and the reaction was quenched by the addition of the ninhydrin reagent to determine the concentration of the free amino acids that were released following hydrolysis of the amide bond. The plot for the change in absorbance at 507 nm as a function of enzyme concentration is presented in **Figure 2.4A**. The maximum absorbance change at the highest enzyme concentrations is ~1.1. This result indicates that approximately one third of the *N*-acyl-D-amino acids in

Table 2.1: Relative rates of hydrolysis of L-Xaa-D-Xaa dipeptide libraries by Bb3285 and Gox1177

Library	Bb3285	Gox1177
L-Ala-D-Xaa	16	48
L-Arg-D-Xaa	2	3
L-Asn-D-Xaa	4	26
L-Asp-D-Xaa	1	0.1
L-Gln-D-Xaa	2	5
L-Glu-D-Xaa	1	1
L-Gly-D-Xaa	64	5
L-His-D-Xaa	2.6	22
L-Ile-D-Xaa	5	7
L-Leu-D-Xaa	86	92
L-Lys-D-Xaa	5	1
L-Met-D-Xaa	100	76
L-Phe-D-Xaa	10	71
L-Pro-D-Xaa	37	52
L-Ser-D-Xaa	4	27
L-Thr-D-Xaa	3	37
L-Trp-D-Xaa	6	80
L-Tyr-D-Xaa	7	100
L-Val-D-Xaa	7	6

Table 2.2: Kinetic parameters for Bb3285 with selected substrates at pH 7.5[a]

Substrate	k_{cat}, s^{-1}	K_m, μM	k_{cat}/K_m, $M^{-1}s^{-1}$
N-formyl-D-Glu	2200 ± 72	380 ± 35	(5.8 ± 0.6) x 10^6
N-acetyl-D-Glu	460 ± 9	88 ± 8	(5.2 ± 0.5) x 10^6
N-succinyl-D-Glu	5.8 ± 0.4	60 ± 22	(9.6 ± 3.5) x 10^4
L-Leu-D-Glu	3.6 ± 0.1	12 ± 1.4	(3.0 ± 0.4) x 10^5
L-Met-D-Glu	2.9 ± 0.1	8.4 ± 1.4	(3.4 ± 0.6) x 10^5

[a]These data were obtained from a fit of the data to equation 2.

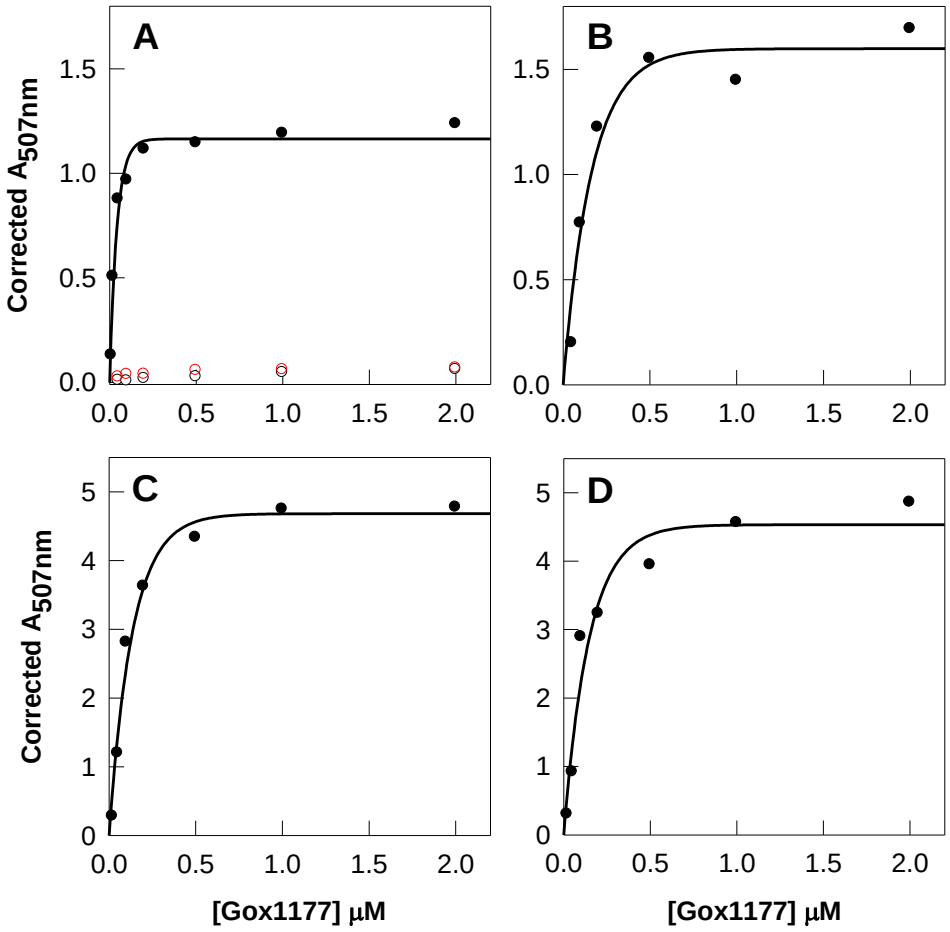


Figure 2.4: Enzyme courses for the hydrolysis of various substrate libraries by Gox1177. **(A)** *N*-acetyl-D-Xaa (*black* circles), *N*-acetyl-L-Xaa (open *black* circles) and *N*-succinyl-D-Xaa (*red* circles). **(B)** Gly-D-Xaa. **(C)** L-Ala-D-Xaa. **(D)** L-Asp-D-Xaa. In panels A, B, and C the enzyme was incubated with the substrate library for 3 hours whereas in panel D the substrate library was incubated with Gox1177 for 24 hours.

this library were hydrolyzed under these reaction conditions. A fraction of the hydrolyzed *N*-acetyl-D-Xaa library was subjected to amino acid analysis to determine the specific compounds that are substrates for Gox1177. The HPLC chromatogram is presented in **Figure 2.5** for samples that were incubated for 90 minutes with either 20 nM (red) or 200 nM (blue) Gox1177. This enzyme has substantially broader substrate specificity than does Bb3285 and shows a kinetic preference for the hydrolysis of *N*-acyl derivatives of the hydrophobic amino acids leucine, methionine, phenylalanine, tryptophan, tyrosine and valine, in addition to alanine, asparagine, glutamine, histidine, serine and threonine. The enzyme was unable to hydrolyze compounds contained within the *N*-acetyl-L-Xaa or *N*-succinyl-D-Xaa substrate libraries (**Figure 2.4A**).

The breadth of the substrate profile for Gox1177 was further examined by utilizing the dipeptide libraries Gly-D-Xaa (**Figure 2.4B**), L-Ala-D-Xaa (**Figure 2.4C**) and L-Asp-D-Xaa (**Figure 2.4D**). A significant fraction of the compounds found in these libraries were hydrolyzed as substrates by Gox1177. To determine the relative ability of Gox1177 to hydrolyze dipeptides, Gly-D-Xaa and 18 L-Xaa-D-Xaa (except L-Cys-D-Xaa) dipeptide libraries were mixed with 0-2 μM Gox1177 and the free amino acids produced after 3 hours were determined with the ninhydrin reagent. The relative rates are listed in **Table 2.1.** For the dipeptide libraries the enzyme has a preference for Tyr, Leu, Trp, Met, and Phe at the N-terminus. The kinetic parameters for the hydrolysis of L-Met-D-Leu, L-Leu-D-Leu, L-Tyr-D-Leu, and fourteen *N*-acyl-D-amino acid substrates by Gox1177 are provided in **Table 2.3**. At the C-terminal end of these *N*-acyl-D-amino acids, the preferred amino acids are leucine, tryptophan, phenylalanine, and tyrosine.

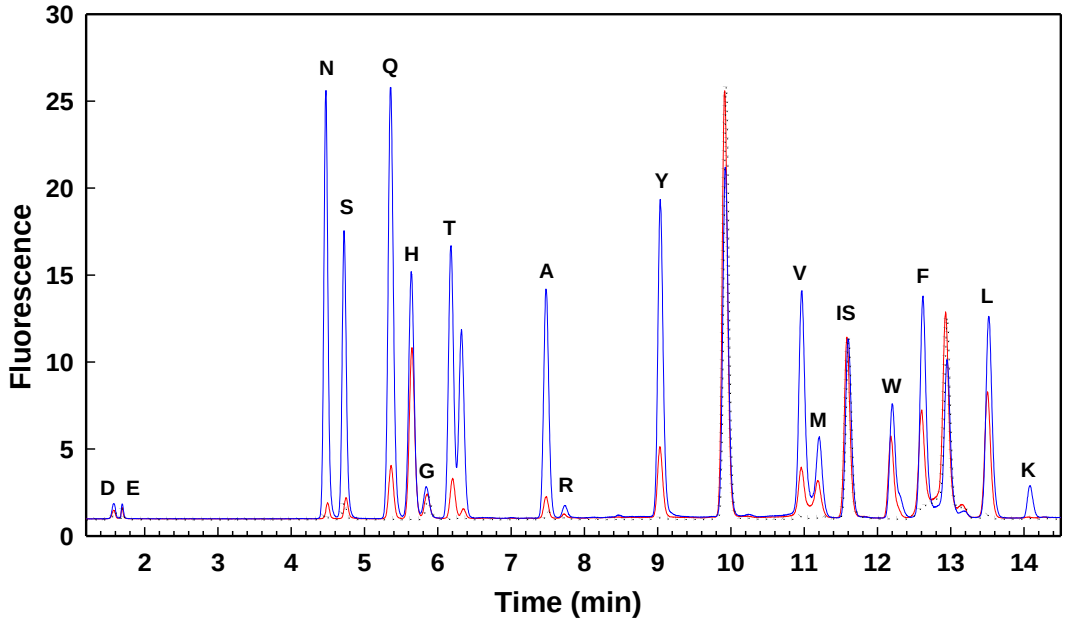


Figure 2.5: HPLC chromatograms for the hydrolysis of the *N*-acetyl-D-Xaa library treated with no enzyme (*black* dots), 20 nM Gox1177 (*red* line) and 200 nM (*blue* line) Gox1177 for 90 minutes at 30 °C. The OPA-derivatized amino acids are indicated with their single letter code. The OPA-derivatized norvaline internal standard is indicated as IS. The unlabelled peak near threonine is likely oxidized methionine. Additional details are provided in the text.

Table 2.3: Kinetic parameters for Gox1177 with selected substrates[a]

Compound	k_{cat}, s^{-1}	K_m, mM	k_{cat}/K_m, $M^{-1}s^{-1}$
N-acetyl-D-Trp	89 ± 2	2.2 ± 0.2	(4.1 ± 0.3) x 10^4
N-acetyl-D-Leu	101 ± 3	3.2 ± 0.2	(3.2 ± 0.2) x 10^4
N-acetyl-D-Phe	59 ± 2	2.5 ± 0.3	(2.4 ± 0.3) x 10^4
N-acetyl-D-Met	40 ± 2	1.8 ± 0.3	(2.2 ± 0.4) x 10^4
N-acetyl-D-Tyr	66 ± 2	7.1 ± 0.5	(9.3 ± 0.7) x 10^3
N-acetyl-D-His	nd	nd	(8.5 ± 0.2) x 10^3
N-acetyl-D-Thr	nd	nd	(6.0 ± 0.1) x 10^3
N-acetyl-D-Val	16 ± 1	4.1 ± 0.2	(3.9 ± 0.2) x 10^3
N-acetyl-D-Ala	50 ± 4	18 ± 2	(2.9 ± 0.4) x 10^3
N-acetyl-D-Gln	nd	nd	(2.7 ± 0.1) x 10^3
N-acetyl-D-Asn	nd	nd	480 ± 20
N-propionyl-D-Leu	23 ± 1	1.2 ± 0.1	(2.0 ± 0.1) x 10^4
N-formyl-D-Leu	13 ± 1	5.9 ± 0.5	(2.2 ± 0.2) x 10^3
N-succinyl-D-Leu	nd	nd	30 ± 1.4
L-Tyr-D-Leu	19 ± 1	1.3 ± 0.2	(1.5 ± 0.2) x 10^4
L-Met-D-Leu	20 ± 1	1.6 ± 0.2	(1.3 ± 0.2) x 10^4
L-Leu-D-Leu	31 ± 1	3.3 ± 0.2	(9.2 ± 0.7) x 10^3

[a]From fits of the data to equation 2. For those entries without values of k_{cat} and K_m, saturation was not achieved at concentrations up to 10 mM.

The *N*-methylphosphonate derivative of D-leucine (**2**) was tested as a competitive inhibitor of Gox1177. At pH 8.4 this compound was found to be a competitive inhibitor versus *N*-acetyl-D-leucine with a K_i value of 4.9 ± 0.1 μM from a fit of the data to equation 3.

Partial Purification of Sco4986. The SDS-PAGE gel of the partially purified Sco4986 is shown in **Figure 2.6**. The band which was used to confirm the amino acid sequence of residues 2-6 is also indicated.

Substrate Specificity of Sco4986 and Bb2785. The protein Sco4986 could not be purified, but the substrate specificity was measured using clarified cell lysates. Control experiments were conducted using cells that contained the pET-30a(+) plasmid lacking the gene for expression of Sco4986. Incubation of the nineteen *N*-acetyl-D-amino acids with the lysates of Rosetta 2 BL21 (DE3) cells harboring the pET-30a(+) or pET-30a(+)-Sco4986 plasmids was used to determine if any background D-aminoacylase activity was present in *E. coli* and if the presence of the Sco4986 gene in the plasmid resulted in the lysate having D-aminoacylase activity. There was a significant increase in the rate of formation of free amino acids when the substrates and the Sco4986 lysate were together in solution *vs*. either of these components alone. It was determined that all of the *N*-acetyl-D-amino acids were substrates for Sco4986, with the exception of the *N*-acetyl derivatives of D/L-Asp, D-Glu, D-His, D-Lys, and D-Arg. For the *N*-acetyl-derivatives of D-Ala, D/L-Cys, D-Phe, D-Leu, D-Met, D-Gln, D-Val, D-Trp and D-Tyr, the rate of hydrolysis of the substrates by the pET-30a(+) control lysate ranged from 1% to 4% of that observed for the lysate containing Sco4986. The hydrolysis of the *N*-acetyl-Gly,

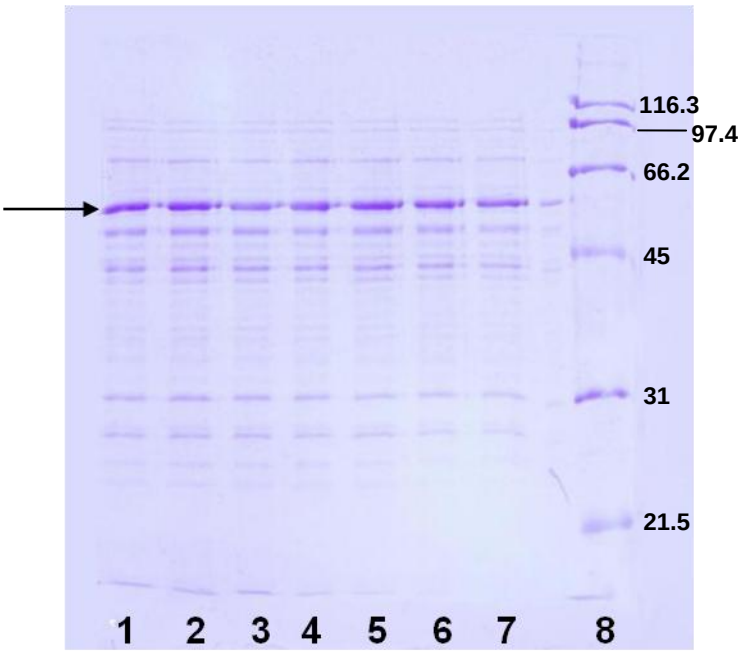


Figure 2.6: SDS-PAGE gel of partially purified Sco4986 (lanes 1-7). The band which was used for N-terminal amino acid sequencing is indicated with an arrow. Lane 8 is the molecular weight marker with the molecular weights labeled in kDa. The calculated molecular weight of Sco4986 is 58.1 kDa.

N-acetyl-D-Asn, *N*-acetyl-D-Ser and *N*-acetyl-D-Thr substrates by the control lysate was 11-12% of the rate observed for the lysate containing Sco4986. The relative rates of hydrolysis are summarized in **Table 2.4**. The best of these substrates is *N*-acetyl-D-Phe (K_m = 0.51 ± 0.04 mM, V_{max} = 0.3 ± 0.006 μmol·min^{-1}·mg^{-1}) followed by *N*-acetyl-D-Trp (K_m = 0.28 ± 0.03 mM), *N*-acetyl-D-Tyr (K_m = 9.1 ± 0.5 mM), and *N*-acetyl-D-Met (K_m = 6.2 ± 1.0 mM).

The *N*-methylphosphonate derivative of D-Phe (**3**) was found to be an inhibitor of Sco4986 with a K_i of 87 ± 4 nM. The inhibition plots for Bb3285, Gox1177 and Sco4986 with compounds **1** and **2** and **3** are presented in **Figure 2.7**. No hydrolysis could be detected upon the incubation of Bb2785 with the 18 *N*-acetyl-D-amino acid compounds.

*Inhibition of Bb3285 by Compound **1***. The *N*-methylphosphonate of D-glutamate was a tight-binding inhibitor of Bb3285 with a K_i of 460 ± 74 pM (**Figure 2.7A**). A sample of some of the time courses which were fit to equation 5 to obtain the k_{obs} values for the onset of inhibition of Bb3285 are displayed in **Figure 2.8A**. The k_{obs} values were plotted against the concentration of inhibitor (**Figure 2.8B**), and it appeared that the k_{obs} values were approaching a maximum value independent of higher inhibitor concentrations. This suggests that the EI complex could be converting to a more tightly-bound E*I complex with the forward and reverse rate constants of k_3 and k_4, respectively (**Scheme 2.3**). Fits of the k_{obs} data vs. the concentration of **1** to equation 6 (**Figure 2.8B**) gave values of $(k_2/k_1)^{app}$ = 130 ± 20 nM and k_3 = 0.017 s^{-1}. The value of $(k_2/k_1)^{app}$ adjusted for competition with the substrate was 18 ± 3 nM.

Table 2.4: Relative rates of hydrolysis of *N*-acetyl-D-amino acids by Sco4986[a]

Substrate	Relative Rate
N-acetyl-D-Phe	100
N-acetyl-D-Trp	48
N-acetyl-D-Met	30
N-acetyl-D-Tyr	25
N-acetyl-D-Val	8
N-acetyl-D-Leu	7
N-acetyl-D-Ala	7
N-acetyl-D-Ser	7
N-acetyl-D-Asn	5
N-acetyl-D-Thr	3
N-acetyl-D-Gln	3
N-acetyl-D-Gly	1

[a]pH 7.5 at 1.0 mM substrate**.**

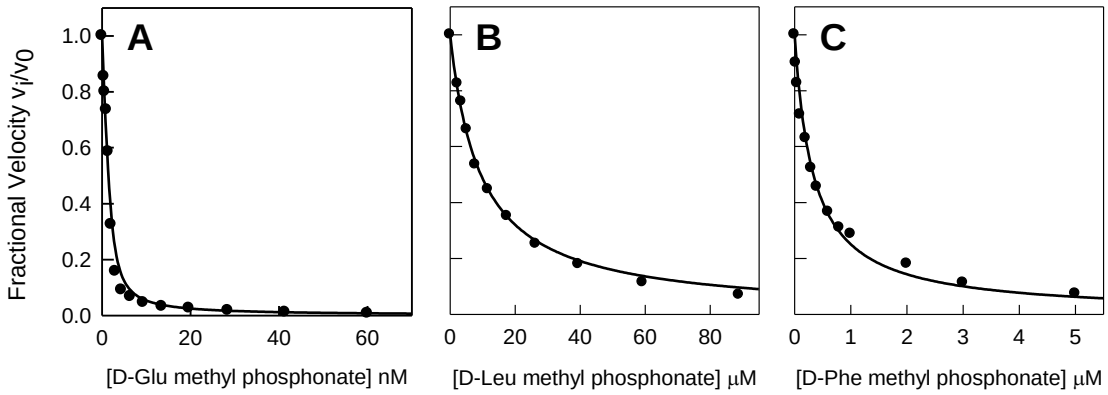


Figure 2.7: Inhibition plots of (**A**) Bb3285, (**B**) Gox1177 and (**C**) Sco4986 by the *N*-methylphosphonate derivatives of D-glutamate (**1**), D-leucine (**2**), and D-phenylalanine (**3**), respectively. **(A)** A competitive inhibition constant of 460 pM was obtained from a fit of the data to equation 4 at a substrate concentration of 3.0 mM *N*-formyl-D-Glu. **(B)** A competitive inhibition constant of 4.9 μM was obtained from a fit of the data to equation 3 at a substrate concentration of 3.0 mM *N*-acetyl-D-Leu. **(C)** A competitive inhibition constant of 87 nM was obtained from a fit of the data to equation 3 at a substrate concentration of 1.5 mM *N*-acetyl-D-Phe. Additional details are provided in the text.

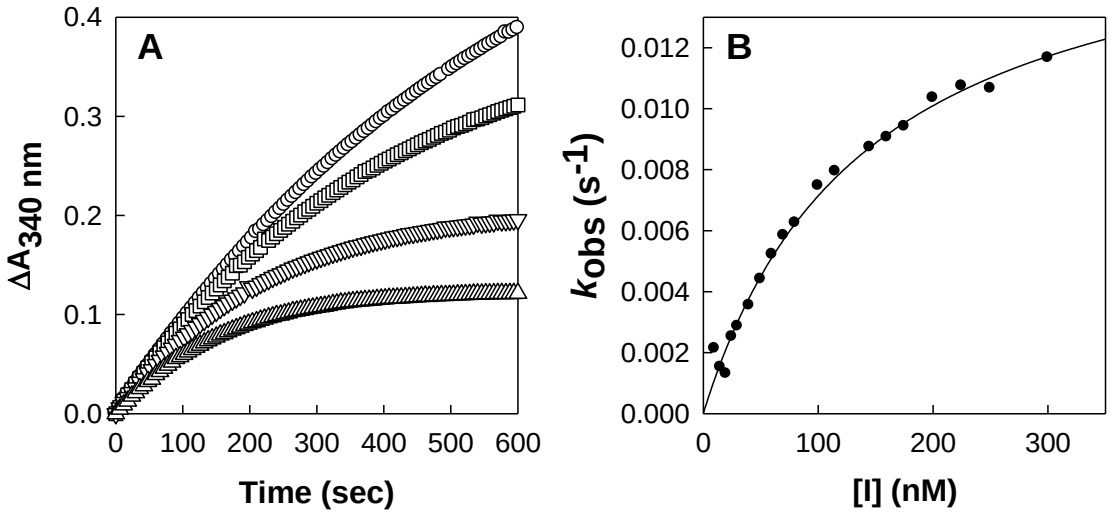


Figure 2.8 Time dependence for the onset of inhibition of Bb3285 by **1**. **(A)** Time course plots from select concentrations of **1** for the onset of inhibition of Bb3285 by **1** fit to equation 5 to determine the apparent first order rate constant. The concentrations of **1** shown are 15 nM (◯), 25 nM (□), 50 nM (▽) and 80 nM (△). **(B)** Effect of the concentration of **1** on the rate of inhibition onset. The data were fit to equation 6.

Three-Dimensional Structure of Bb3285. The crystal structure of Bb3285 was determined to a resolution of 1.5 Å as a homodimer with one zinc, one acetate and two formate molecules bound in the active site (**Figure 2.9**). The N-terminal residues 1-4 and the C-terminal residues 479-480 were disordered in both structures and were not included in the final model. In addition to the signature $(\beta/\alpha)_8$-barrel (shown in dark blue and red) two additional domains are present. The first of these domains is an insertion colored in pink (residues 287-344) between β-strand 7 and α-helix 7 of the $(\beta/\alpha)_8$-barrel. The second domain is a nine stranded β-barrel encompassing residues 5-61 (colored in teal) and residues 413-478 (colored in yellow) contributed from the N- and C-termini of the polypeptide. A long loop (residues 432-451) is inserted between the seventh and eighth β-strands of the β-barrel. The single zinc is bound in the β-metal site and is ligated by Cys-95, His-218, His-248 and the acetate product as shown in **Figure 2.10A**. In this complex one of the carboxylate oxygens from acetate is positioned 2.3 Å from the zinc and 2.6 Å away from one of the oxygens of the catalytic aspartate residue from the end of β-strand 8 (Asp-365). The other carboxylate oxygen in the acetate is 2.6 Å away from the phenolic oxygen of Tyr-190 and 2.3 Å from the zinc. The two histidines at the end of β-strand 1 (His-66 and His-68) do not ligate a second zinc in the α-metal binding site. One of the formate molecules is interacting with Lys-250, Tyr-282, and Arg-376 at distances of 2.7, 2.5, and 2.8 Å, respectively. The other formate forms a polar interaction with the side chain of Arg-295 and Ser-287.

The second Bb3285 structure had well-defined density for two Zn^{2+}, and one inhibitor molecule bound in the active site of both molecules in the asymmetric unit.

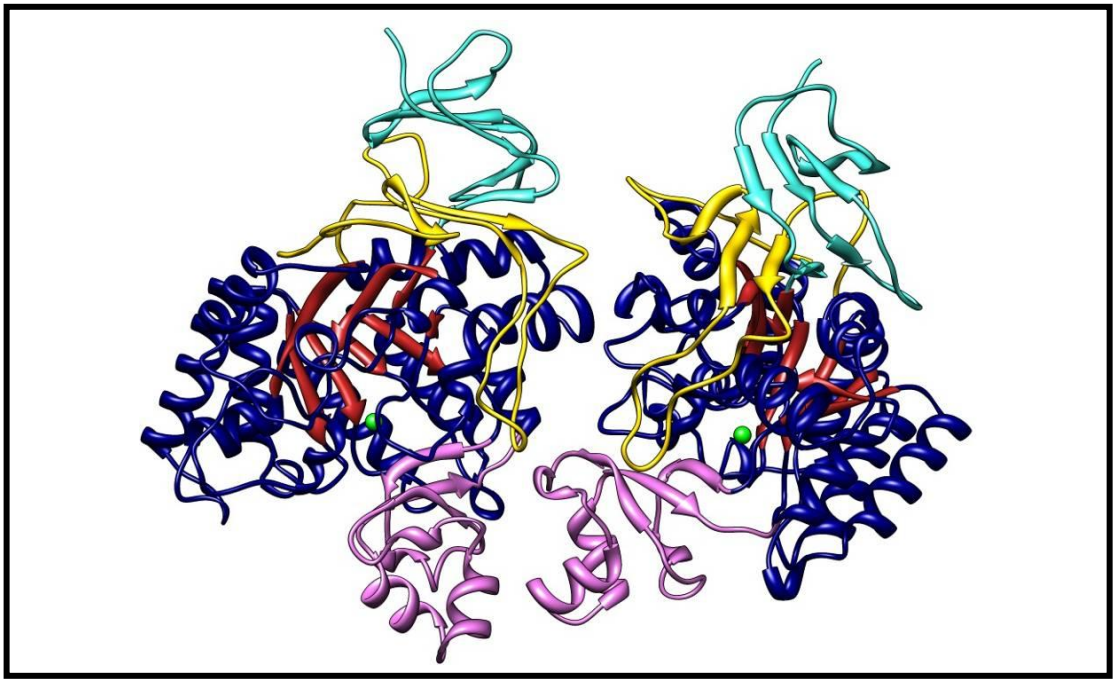


Figure 2.9: Homodimeric structure of Bb3285. The $(\beta/\alpha)_8$-barrel is colored with *red* β-strands and *dark blue* α-helices and adjoining loops. The first insertion domain (from residue 287 to 344) containing the substrate specificity loop is colored *pink*. The second insertion domain consisting of residues 5-61 and 413-478 is colored *cyan* and *yellow*, respectively. The active site Zn is represented as a *green* sphere.

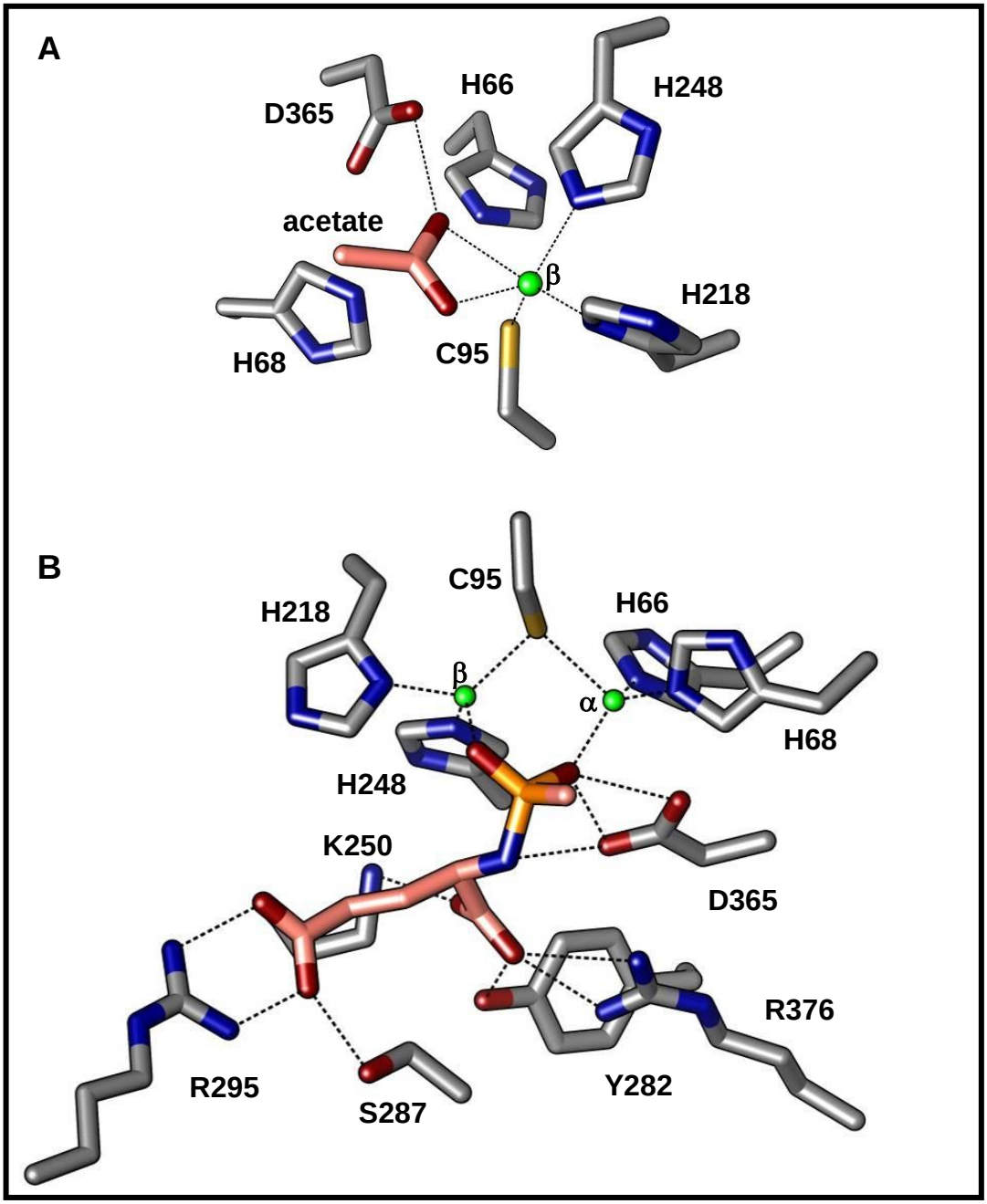


Figure 2.10: Native mononuclear metal center and binuclear inhibitor-complexed active site of Bb3285. Contacts to the metals (2.0-2.4 Å) are indicated by dashed lines. **(A)** Mononuclear Zn (*green* sphere) center with acetate (*pink* carbons) bound in the active site. **(B)** Binuclear Zn (*green* sphere) center active site of Bb3285 with bound inhibitor **1** (*pink* carbons, *orange* phosphorus). Enzyme-inhibitor contacts within 2.6-3.1 Å (3.5 Å to D365) are indicated by dashed lines. For clarity, the 2.7 Å contact between the hydroxyl group of Y190 and the phosphonyl oxygen coordinated to Zn_β is not shown. Additional details are provided in the text.

The 3-dimensional structure was solved to 1.8 Å resolution as a binuclear Zn enzyme with compound **1**, a tight-binding inhibitor of Bb3285, bound in the active site (**Figure 2.10B**). The histidine ligands, His-66 and His-68, at the end of β-strand 1 were coordinated at 2.1 and 2.0 Å to the Zn in the α-site. The Zn_β-metal was ligated by His-218 and His-248 as in the native structure. The two metals were bridged by Cys-95 (2.4 and 2.3 Å) and each phosphonyl oxygen of compound **1** to either Zn_α (2.0 Å) or Zn_β (2.0 Å). The C-terminal carboxylate of **1** was within hydrogen-bonding distance to Lys-250 (2.8 Å), Tyr-282 (2.7 Å) and Arg-376 (2.7 Å). The oxygens of the side-chain carboxylate of D-glutamate were 2.7-2.8 Å from the guanidino group of Arg-295. An oxygen of the side chain carboxylate of **1** also contacts the hydroxyl group of Ser-287 at a distance of 2.5 Å. Though not shown, for the purpose of clarity, the hydroxyl oxygen of Tyr-190 is 2.7 Å from the phosphonyl oxygen which coordinates to Zn_β. The catalytic aspartate at the end of the 8^{th} β-strand is 3.5 and 2.6 Å from the amide nitrogen of **1** and phosphonyl oxygen that is coordinated to Zn_α.

Network Analysis of N-acyl-D-amino Acid Deacylases. Approximately 240 *N*-acyl-D-amino-acid deacylase like sequences were identified in the NCBI database from the completely sequenced organisms. At an E-value cutoff of 1×10^{-70}, the sequences can be separated into 6 main clusters, as shown in **Figure 2.11**. Each node in the network represents a single amino acid sequence and each line represents a pairwise connection between two sequences with a BLAST score of $E \leq 1 \times 10^{-70}$. The lengths of lines are not meaningful, but sequences in tightly clustered groups are more similar to

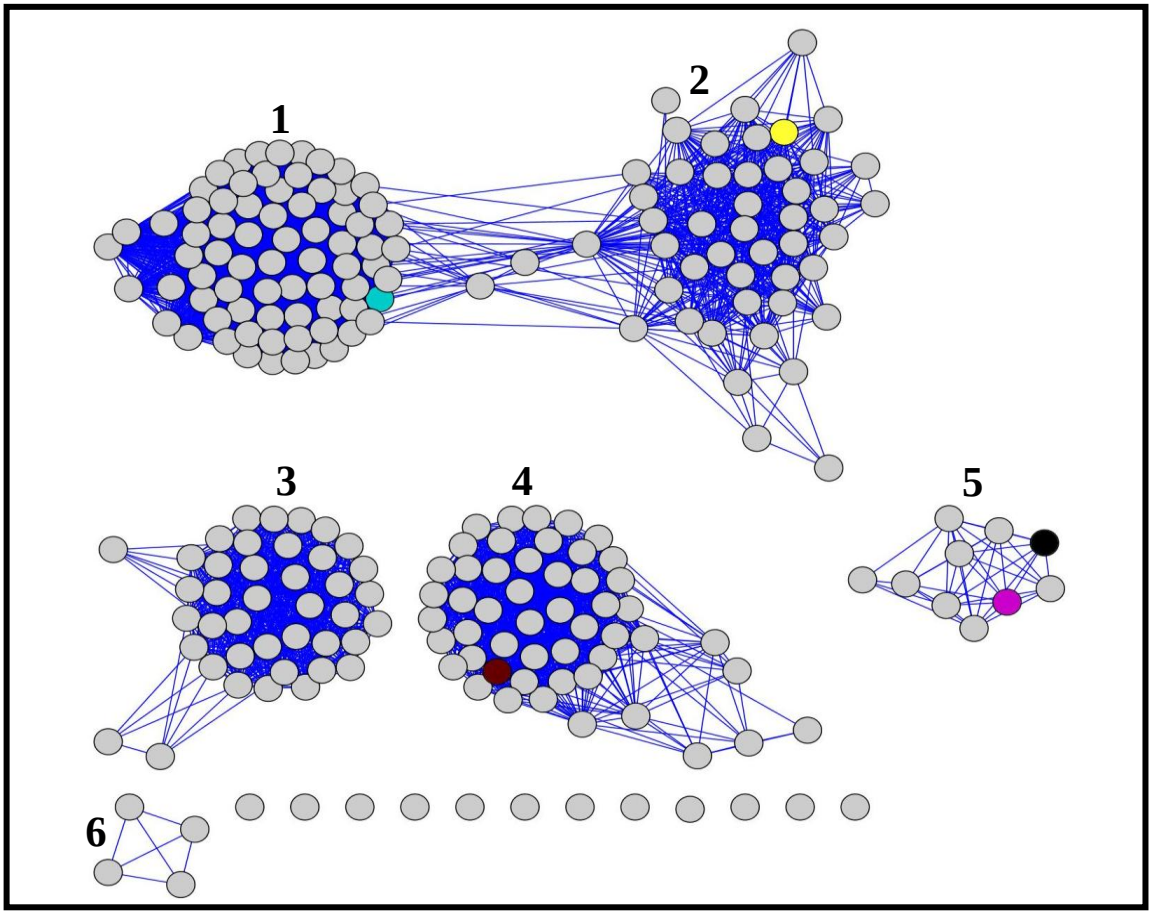


Figure 2.11: Cytoscape graphical representation of putative D-aminoacylases from cog3653. The amino acid sequences (circles) are separated into six clusters based on an e cut-off value of 1×10^{-70}. Circles which are connected by lines have an E-value $\leq 1 \times 10^{-70}$. The protein targets described in Chapter II are colored *teal* (Bb3285), *yellow* (Sco4986), *dark red* (Bb2785), *pink* (Gox1177) or *black* (Bll7304).

each other than sequences with few connections. The four proteins discussed here are highlighted in **Figure 2.11** from clusters 1, 2, 4 and 5.

With the exception of Gox1177 and Sco4986, which were functionally characterized here, all of the proteins that are known to catalyze the deacylation of *N*-acyl-D-amino acids are found in cluster 1 (*34, 35, 38, 40, 44, 56*). Cluster 4 includes the enzyme Bb2785, for which specificity could not be determined on the basis of library screening using the *N*-acetyl-D-amino acid library. From the alignment in **Figure 2.1**, it appears that Bb2785 is missing some metal ligands as well as key residues implicated in the recognition of the *N*-acyl-D-amino acid substrate. Cluster 3 does not contain characterized enzymes. As with the sequences from Cluster 4, it is difficult to say whether or not the proteins in Cluster 3 have all of the metal ligands and/or substrate-recognition residues.

DISCUSSION

Five genes coding for putative D-aminoacylases were cloned and four of these enzymes were successfully expressed in *E. coli*: Bb3285, Bb2785, Sco4986 and Gox1177. Each of these enzymes was tested as a catalyst for the hydrolysis of *N*-acyl-D/L-Xaa substrates contained within a series of well-defined libraries of *N*-substituted amino acid derivatives. Using ninhydrin and/or HPLC-based assays, it was possible to measure the rate of hydrolysis of each library component by quantifying the specific amino acids liberated within these libraries as a function of time or enzyme concentration. Each enzyme was screened against more than 400 compounds.

Specificity of Bb3285. Bb3285 exclusively hydrolyzes derivatives of D-glutamate, where this amino acid is substituted with a simple acyl group or another amino acid. This enzyme will hydrolyze *N*-acetyl-, *N*-formyl- and *N*-succinyl-derivatives of D-glutamate but not the *N*-acyl-derivatives of any other D- or L-amino acid. Bb3285 hydrolyzes a variety of L-Xaa-D-Glu dipeptides but the best substrates are *N*-formyl- and *N*-acetyl-D-Glu. The *N*-formyl-D-Glu substrate has the highest value of k_{cat} (2200 s^{-1}) but the values for k_{cat}/K_m with this substrate and *N*-acetyl-D-Glu are essentially the same. The enzyme is less stringent regarding the identity of the L-amino acid at the amino-terminus of dipeptide substrates, but it does exhibit a preference for leucine or methionine derivatives of D-Glu.

Specificity of Gox1177. The second enzyme examined in this investigation, Gox1177, was found to deacetylate a broad range of *N*-acetyl-hydrophobic D-amino acids. Changing the acyl group from *N*-acetyl- to *N*-formyl-D-Leu results in a decrease in k_{cat} and an increase in K_m. With the L-Xaa-D-Xaa dipeptides, Gox1177 has a clear preference for the larger hydrophobic and aromatic residues (Tyr, Trp, Phe, Met and Leu) at the amino terminus. The relatively high K_m values for Gox1177 with the substrates identified in this investigation calls into question whether *N*-acetyl-D-amino acids are necessarily the physiological substrates for this enzyme. Perhaps the native substrate of Gox1177 has a different functionality attached to the D-amino acid.

Specificity of Sco4986. The gene for Sco4986 was expressed in *E. coli* Rosetta 2 cells but the protein could not be purified to homogeneity because the protein was largely insoluble. However, the deacetylase activity of this protein could be detected in

whole cell lysates. The relative rates of substrate hydrolysis clearly demonstrate that Sco4986 hydrolyzes many of the *N*-acetyl-D-amino acid derivatives with a preference for hydrophobic and aromatic amino acids. The Michaelis constants for the best substrates found with Sco4986 are ~ 5-fold lower than the best substrates for Gox1177.

Activity of Bb2785. A catalytic activity for Bb2785 was not identified. Many of the residues whose side chains are expected to bind the substrate could not be identified in sequence alignments and thus if this protein has enzymatic activity, the substrates were not included in our screening libraries.

Structural Basis for Substrate Specificity. The three-dimensional X-ray structures of Bb3285 in the presence of potent inhibitor **1**, and of the acetate/formate complex, have revealed the structural determinants for enzyme specificity of this enzyme. In the complex with acetate and two formate molecules, one of the formate molecules is ion-paired with Arg-376 and Lys-250 in addition to a hydrogen bonding interaction with the phenolic side chain of Tyr-282. The arginine and tyrosine residues are fully conserved in Bb3285, Gox1177, and Sco4986 and thus these two residues are likely required for recognition of the α-carboxylate group at the C-terminus of the substrate. The arginine residue is also fully conserved among all of the enzymes that are $\geq$40% identical in amino acid sequence to Bb3285, Gox1177 or Sco4986. The tyrosine residue is semi-conserved but the substitution is limited to a histidine in these proteins. The lysine is conserved in those enzymes that are >40% identical in sequence to Bb3285 but is not conserved in those sequences that are more similar to Gox1177 or Sco4986.

The second formate in the active site of Bb3285 is ion-paired with the side chain guanidino group of Arg-295 and thus this interaction is likely required for recognition of the side chain carboxylate of the C-terminal D-glutamate. This assignment is confirmed by the structure of Bb3285 in the presence of compound **1**, a mimic of the tetrahedral reaction intermediate (**Figure 2.10B**). In Bb3285, Arg-295 is found in a loop that starts after the end of β-strand 7. This is the same loop that was proposed to serve as the specificity loop for the DAA from *A. faecalis* DA1, based upon a computational model of *N*-acetyl-D-methionine bound in the active site (*36*). The hydrophobic side chain of D-methionine was postulated to interact with Leu-298 in this enzyme. However, the corresponding residue, from a sequence alignment with Bb3285, is Glu-297 (see **Figure 2.1**) and thus it was not so clear how the side chain carboxylate of D-glutamate substrates would be able to interact with Glu-297 in Bb3285. The answer to this dilemma is found in a structural overlay of Bb3285 with the DAA from *A. faecalis* as shown in **Figure 2.12** (*36, 56*). The specificity loops in these two proteins adopt distinct conformations. In the *A. faecalis* structure, Leu-298 is pointed towards the active site and the side chain of Arg-296 is pointed away from the active site. However, in the Bb3285 structure, Arg-295 is pointed toward the active site. The conformational differences in these two loops are likely the result of a twist in the loop that is initiated by Pro-293 in Bb3285. Presumably, the side chain of *N*-acetyl-D-aspartate is not long enough to interact with Arg-295 and thus Bb3285 is unable hydrolyze substrates with a terminal D-aspartate. The D-glutamate side chain is also in contact with Ser-287, which is a glycine residue in Gox1177, Sco4986 and the DAA from *A. faecalis*. The main two

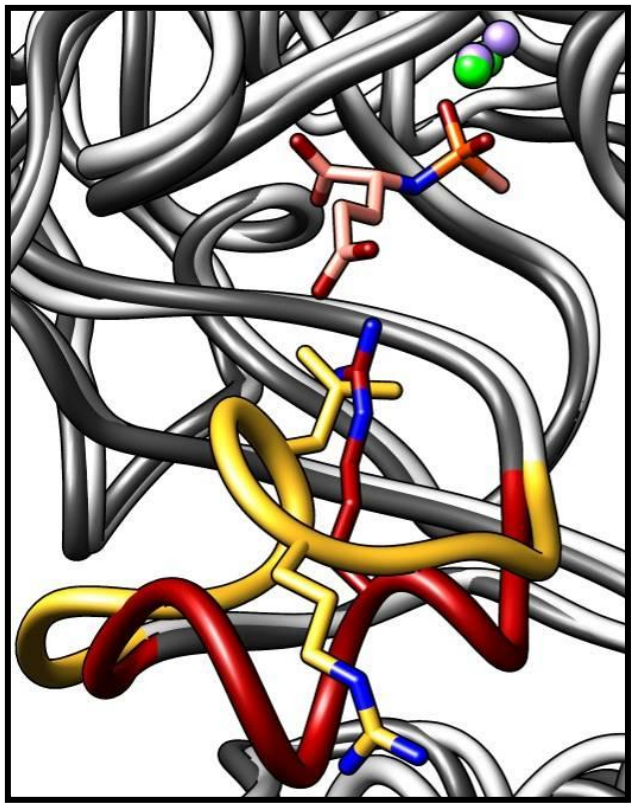


Figure 2.12: Structural overlay of Bb3285 (*dark gray* ribbon, *green* Zn) with the D-aminoacylase (PDB entry *1M7J*) from *A. faecalis* (*light gray* ribbon, *purple* Zn) showing the conformational differences between the loops in the two proteins that determine the differences in substrate specificity. The loop (colored *yellow*) from the D-aminoacylase from *A. faecalis* is for residues 292-302. In this structure, Leu-298 points toward the active site and Arg-296 points away from the active site. The loop colored *red* from Bb3285 is for residues 291-301. The phosphonamidate inhibitor is colored with *pink* carbons and an *orange* phosphorus. In this structure, Arg-295 is pointing toward the active site and the guanidino group is ion-paired with the inhibitor.

residues which contribute to the pocket which bears the methyl group of inhibitor **1** (a representative of the methyl group of the *N*-acetyl- moiety of the substrate) are Leu-367 and Leu-114.

Prediction of Functional Specificity for Uncharacterized Members of the N-acyl-D-Xaa Deacetylase Group. Conservation of functionally important residues in multiple sequence alignments is often used to infer similarity of functional characteristics. However, **Figure 2.1** is equivocal. For example, two residues in the "specificity loops" of Bb3285, Pro-293 and Arg-295, likely play a role in the exclusive preference for derivatives of D-glutamate in this protein. However, only 7 sequences in Cluster 1 of the network show conservation of these residues, suggesting that only a few of these unknowns may share this preference. Bb3285 has 47% amino acid sequence identity to the protein with locus tag Rmet_4439 (gi|94313364) which also has the equivalent residue to Pro-293 and Glu-294 in Bb3285. Rmet_4439 has a phenylalanine in place of the Arg-295 residue of Bb3285 which would be expected to drastically alter substrate preference of the enzyme from that of Bb3285.

With the addition of the Bb3285 structure there are now two DAA structures, both categorized as members of Cluster 1 (**Figure 2.11**). Unfortunately, the two substrate specificity loops (both 11 amino acids long) vary in conformation, making it difficult to predict the substrate specificity of other proteins in Cluster 1. This is because there is no reason to believe that this loop cannot have other conformations which have not been structurally characterized. Two residues within this loop in the *A. faecalis* enzyme, Lys-293 and Asp-295, are potentially correlated to its conformation.

As illustrated in **Figure 2.13A**, the free amino group of Lys-293 is 4.1 Å from Asp-295, which is 2.9-3.0 Å from the backbone nitrogens of Val-297 and Leu-298. Two DAAs from *Alcaligenes xylosoxydans*, specific for either *N*-acyl- derivatives of D-aspartate (gi|3287879) or D-neutral amino acids (gi|3287876), also have an aspartic acid at the position equivalent to Asp-295 of the *A. faecalis* enzyme (**Figure 2.13B** dark blue). The D-neutral-specific enzyme has the equivalent residue to Lys-293 as well as the active site pocket leucine equivalent to Leu-298 in the *A. faecalis* enzyme (**Figure 2.13B** purple highlight). The enzyme specific for the *N*-acyl-D-aspartate, has an arginine at both positions equivalent to Lys-293 and Leu-298 of the *A. faecalis* enzyme. Since the substrate specificities of these two enzymes relate well to the structure and substrate specificity of the DAA from *A. faecalis*, the substrate specificity loops may be in the same conformation. It could be that the Asp-295:Val-297 backbone interaction is more significant than the potential salt bridge between Lys-293 and Asp-295. There are several amino acid sequences within Cluster 1 that have the combination of a positively charged residue at the position equivalent to Lys-293 and an aspartate equivalent to Asp-295. The reverse is also true in that several sequences have lysine and aspartate residues which are in opposite positions to the Lys-293/Asp-295 combination (**Figure 2.13B**, pink shading). One can envision that these two residues may still be able to contact each other, but the positively charged residue would no longer hydrogen bond with a backbone nitrogen. DAAs specific for either *N*-acyl- derivatives of polar (uncharged) or positively charged D-Lys/Arg D-amino acids have yet to be identified in the literature. Based on the above hypothesis, there are a few putative DAAs from

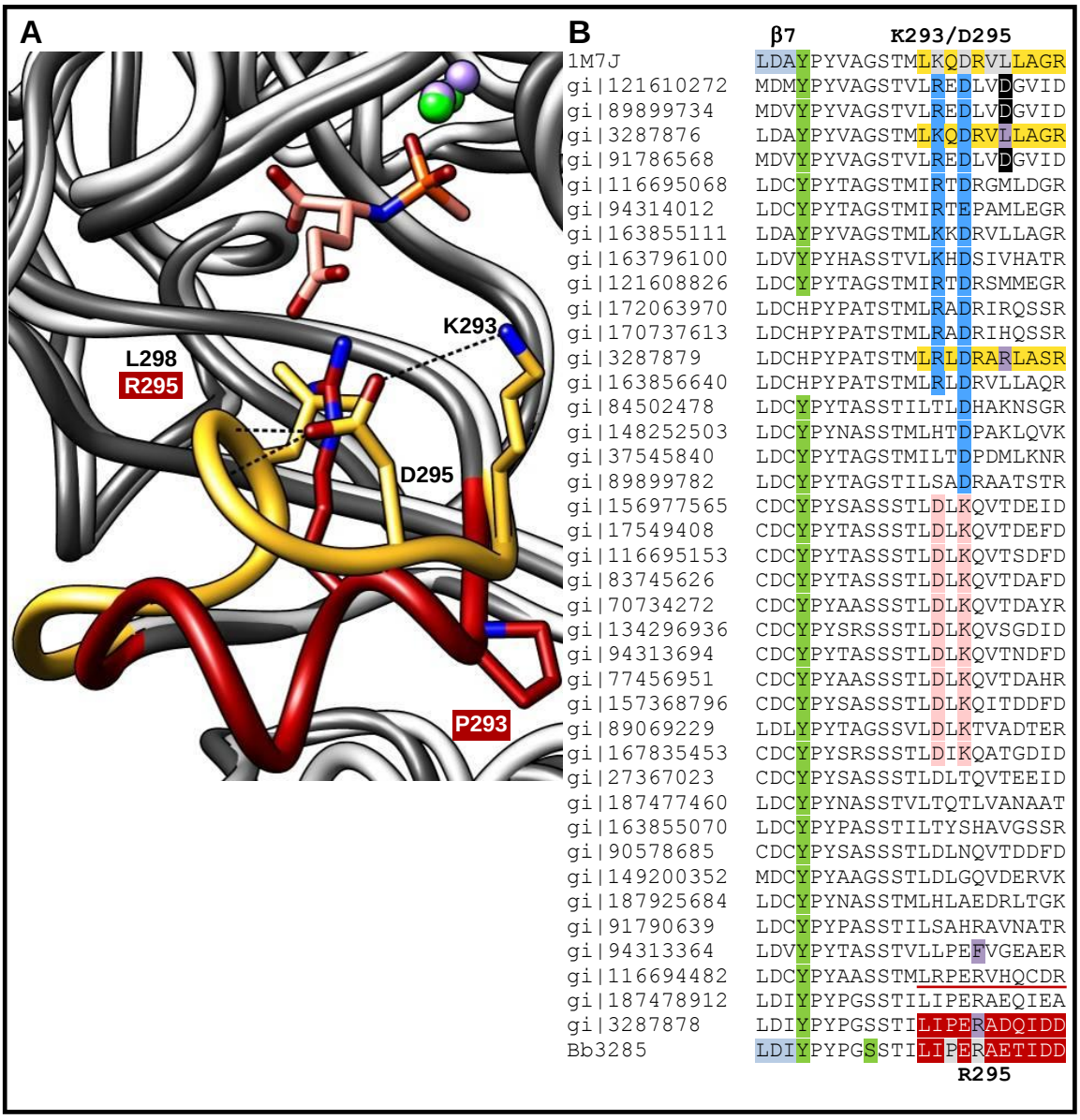


Figure 2.13: Potential determinants of the position of the substrate specificity loop in the D-aminoacylases from *A. faecalis* (*1M7J*) and Bb3285. **(A)** Representation of the superposed Bb3285 (*dark gray* ribbon except for *red* specificity loop, *red* carbons) and *1M7J* (*light gray* ribbon except for *yellow* specificity loop, *yellow* carbons). **(B)** Amino acid sequence alignment of *1M7J* sequence, Bb3285 and several proteins within 40% amino acid sequence identity to Bb3285. Those proteins which have been tested empirically have the sequence specificity loop highlighted *yellow*, if the substrate specificity loop is expected to be akin to *1M7J*, or *red* if the loop is anticipated to be more similar to Bb3285. Residues with side chains shown in panel A are shaded *gray*. Those residues which are in proteins of unknown structure, but are predicted to be facing into the active site and influencing substrate specificity, are highlighted *black* or *purple*. The potential determinants of the loop conformation in *1M7J* discussed in the text are colored *dark blue* and *pink*.

Cluster 1 which might be specific for such compounds because the equivalent residue to Leu-298 is replaced by an aspartate (gi|121610272, gi|89899734, gi|91786568, black shading in **Figure 2.13B**).

As mentioned above for proteins in Cluster 1, the ability to predict the substrate specificities of proteins within this class of enzymes is problematic even when structural data is available. The structural coverage of cog3653 has not yet been expanded to include proteins from any of the other five clusters (**Figure 2.11**). In Bb3285, the substrate specificity loop is part of an insertion region between β_7 and α_7 of the $(\beta/\alpha)_8$-barrel and there are 82 amino acid residues between the tyrosine from β_7 and the catalytic aspartate from β_8. The number of amino acids between these two residues is 87 in Gox1177 and 129 in Sco4986. From **Figure 2.1** it is clear that the alignment of this region between the proteins from Clusters 1, 2, 4 and 5 is poor. In the absence of the 3-dimensional structure of Gox1177 and Sco4986, it is difficult to comment on their substrate-specificity determinants. One observation which cannot be reconciled from the alignment in **Figure 2.1**, is the fact that the trends seen in the data presented in **Tables 2.1**, **2.2**, and **2.3**, suggest that the acyl group pocket of Gox1177 is larger than that for Bb3285.

There is also the potential for structural differences between Gox1177 and Bb3285. In Clusters 1 and 2 there is a specific conserved arginine and glutamate (**Figure 2.1**, brown shading); the arginine immediately following the histidine from β_5 (Arg-220 in Bb3285) forms a salt bridge with a glutamate (Glu-232 in Bb3285). This interaction may cause the loop after β_6 to turn away from the active site before turning

back in, making Glu-222 an accessory structural residue in contact with Lys-250. This Arg-Glu combination, highlighted in brown in **Figure 2.1**, is not expected to be present in Gox1177, though it is seen in some of the other amino acid sequences in Cluster 5. With the exception of Gox1177 and Sco4986 characterized here, all of the DAAs which have been experimentally characterized and are of known amino acid sequence are part of Cluster 1. With the exception of those enzymes specific for *N*-acyl- derivatives of D-Asp and D-Glu, these empirically-tested enzymes have broad specificity towards neutral amino acids. A network similar to the one depicted in **Figure 2.11** was created by Patricia Babbitt (University of California San Francisco) using a more stringent E-value cut-off of 10^{-135} and the sequences separated into smaller clusters. The substrate specificity of the experimentally tested DAAs could not be correlated to organismal lineage, genome context, the identity of the bridging metal ligand, or variations in sequence length and the positions of associated inserts (*56*).

There are at least 50 microbial species with multiple *N*-acyl-D-aminoacylase-like sequences. DAAs from up to three clusters have been observed in the same organism. As observed in *A. xylosoxydans*, multiple DAAs from a single organism can have varying specificities and perhaps functions. This is analogous to a recently-reported case for the uronate isomerases, another group in the AHS, where multiple representatives of this sequence group within a single organism had different substrate specificities (*5*).

Bb2785, the only protein from Cluster 4 that has been experimentally screened, showed no catalytic activity with *N*-acetyl-D-amino acid compounds. This was not surprising considering that Bb2785 appears to be missing the residues which contact the

substrate (two tyrosines, an arginine and a lysine). From the multiple-sequence alignment in **Figure 2.1**, this protein appears to be an outlier. Bb2785 also seems to be missing the last two histidine metal binding ligands. Under these circumstances, one might expect that the protein would bind one metal. However, there were 1.8 equivalents of Zn in the purified Bb2785 protein. An alignment of the amino acid sequences within Cluster 4 does show that there are histidines which are *nearly* conserved and may be candidates for the histidine ligands from β-strands 5 and 6. Bb2785 is not expected to be a D-aminoacylase based on the results outlined here.

In regards to the physiological functions of the Cluster 2 sequence group, the genome context does provide some possible clues. The gene for Sco4986 is adjacent to another open reading frame (Sco4987) that is currently annotated as a D-amino acid deaminase. The top five BLAST hits to the Sco4986 gene from different organisms show approximately 60% identity to this gene. All five are also adjacent or nearby to an ORF annotated as a D-amino acid deaminase, amino acid racemase-like protein or D-amino acid aldolase and these proteins are 38-50% identical in sequence to Sco4987. The similarities in genome context for homologues of these two genes among multiple organisms could suggest that the deacylase is liberating a free D-amino acid and then a second enzyme, the gene product of Sco4987, is racemizing or deaminating the D-amino acid. *N*-acylated-D-amino acids have been shown to be components of lipopolypeptide antibiotics, and the presence of the acyl group on such compounds can be an important component of the molecule with respect to its antibiotic potency (*66, 67*).

Acetyltransferases for D-amino acids are known and it is possible that this small cluster of genes is a simple two-enzyme system to metabolize *N*-acetyl-D-amino acids.

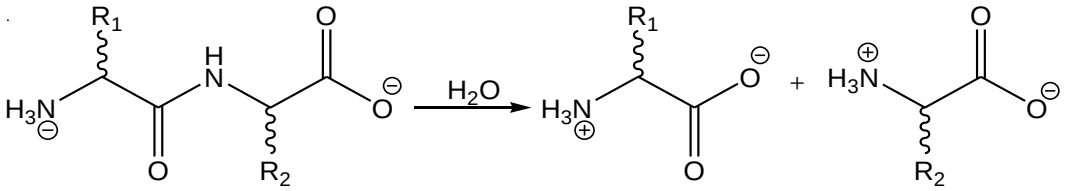
CHAPTER III

SUBSTRATE PROFILE AND STRUCTURE OF Sco3058: THE CLOSEST MICROBIAL HOMOLOGUE TO THE HUMAN RENAL DIPEPTIDASE

INTRODUCTION

The mammalian renal dipeptidase catalyzes the general reaction outlined in **Scheme 3.1** (*68-73*). This enzyme is also able to catalyze the hydrolysis of leukotriene D_4 (*48, 69, 70*) as well as the oxidized degradation product of glutathione and L-Cys-Gly (*69-71*). Human renal dipeptidase (hRDP) is the only enzyme in humans which has been shown to catalyze the hydrolysis of β-lactams such as SCH 29482, imipenem, meropenem and DA-1131 (*74, 75*). However, dipeptides are consistently better substrates for hRDP than are β-lactams. Glycine is the preferred amino acid at the C-terminus, relative to an amino acid in the L-configuration and glycyldehydrophenylalanine is a better substrate for hRDP than glycylphenylalanine (*48*). With rat renal dipeptidase, L-Ala-Gly is hydrolyzed faster than L-Ala-L-Ala and Gly-D-Ala is hydrolyzed faster than Gly-L-Ala (*70*). Mammalian renal dipeptidases and some of their microbial homologues have been assayed with various substrates, but a complete substrate specificity profile for these types of enzymes has not been determined (*48, 49, 69-71, 73, 76*).

Scheme 3.1 General reaction catalyzed by human renal dipeptidase



The crystal structure of hRDP has been solved with the inhibitor cilastatin bound in the active site (PDB code *1ITU*) (*23*). Cilastatin is considered to be a dipeptide analog, but it does not contain a free amino group at the N-terminus and it is not hydrolyzed by hRDP. The metal center in the active site consists of two Zn ions that are bridged by a molecule from solvent (water or hydroxide) and the side chain carboxylate of Glu-125. The alpha-Zn is additionally ligated by His-20 and Asp-22 whereas the beta-Zn is coordinated to His-198 and His-219. It has been suggested that the water which bridges the two metal ions attacks the carbonyl of the peptide bond to form a tetrahedral intermediate that is stabilized by an electrostatic interaction with His-152. This residue is found at the end of β-strand 4 of the $(\beta/\alpha)_8$-barrel structure and has been shown to be important for the binding of the cilastatin inhibitor (*23, 77*).

The structure of hRDP confirmed that this enzyme is a member of the amidohydrolase superfamily (AHS). This is a rather large enzyme superfamily that has been shown to catalyze the hydrolysis of ester and amide bonds found within carbohydrate, peptide, and nucleic acid containing substrates (*20*). The hRDP is contained within a large cluster of proteins designated as cog2355. This cluster of proteins is typified by an HxD motif at the end of β-strand 1 that is used to coordinate

one of the two metal ions in the active site, whereas most of the other enzymes in this superfamily have an HxH motif. In hRDP the invariant aspartate at the end of β-strand 8 is not coordinated to the alpha-metal, whereas, in nearly all of the other members of the AHS that have been structurally characterized this residue is a direct ligand to the alpha-metal ion. The proposed chemical mechanism for the hydrolysis of dipeptides by hRDP is at variance with those mechanisms of amide bond hydrolysis proposed previously for other members of the AHS (*19*, *28*, *78*). In hRDP the carbonyl group of the amide bond about to be hydrolyzed is proposed to be polarized by an electrostatic interaction with His-152 whereas in other members of the AHS this function is contributed by a direct interaction with the beta-metal ion (*23*). The renal dipeptidase-like enzyme BCS-1 from *Brevibacillus borstelensis* has a tryptophan in place of the residue equivalent to His-152. BCS-1 is reported to have respectable activity (k_{cat}/K_m = 2 x 10^5 $M^{-1}s^{-1}$), which supports the conclusion that either His-152 does not play a crucial role in the chemical mechanism or that the mechanism for substrate hydrolysis of enzymes such as BCS-1 vary from that of hRDP (*49*).

The putative dipeptidase Sco3058 from *Streptomyces coelicolor* is the closest bacterial homologue to the human enzyme from a completely sequenced microbial organism. The sequence identity is ~41%. *S. coelicolor* is a member of soil bacteria known for its versatile metabolism and production of antibiotics and pharmaceutical agents (*79*). Here, the substrate specificity of Sco3058 was unraveled by determining its catalytic activity with 55 L-Xaa-L-Xaa, L-Xaa-D-Xaa and D-Xaa-L-Xaa dipeptide libraries. The results demonstrate that this enzyme is a rather promiscuous dipeptidase

which is able to hydrolyze a large fraction of the ~1000 compounds screened. Nevertheless, Sco3058 showed some degree of specificity and the best substrate identified was L-Arg-D-Asp (k_{cat}/K_m = 7.6 x 10^5 $M^{-1}s^{-1}$). The structure of Sco3058 was determined to a resolution of 1.3 Å in the presence and absence of dipeptide mimics (*80*).

MATERIALS AND METHODS

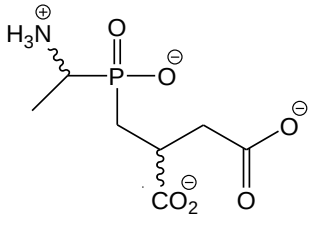
Materials. All chemicals were obtained from Sigma Aldrich, unless otherwise stated. All of the compounds and substrate libraries which were not commercially available were synthesized by Dr. Chengfu Xu. The coupling enzymes, glutamate-oxaloacetate transaminase and malate dehydrogenase, were purchased from Sigma Aldrich and Calbiochem, respectively. Glycyl-D-Leu was purchased from TCI America. Metal analyses were conducted using inductively coupled plasma mass spectrometry (ICP-MS) as previously described (*59*). ICP-MS standards were obtained from Inorganic Ventures Inc. Nitrocefin was purchased from Calbiochem. Oligonucleotide syntheses and DNA sequencing were performed by the Gene Technologies Lab of Texas A&M University.

Synthesis of Dipeptide Libraries. The L-Xaa-D-Xaa and L-Xaa-L-Xaa dipeptide libraries were synthesized as described previously (*56, 80, 81*). The 17 D-Xaa-L-Xaa libraries were synthesized in the same manner as the L-Xaa-L-Xaa libraries except that *N*-Fmoc-D-Xaa-OH amino acids were substituted for *N*-Fmoc-L-Xaa-OH amino acids. These compound libraries contain a single amino acid at the N-terminus and all of the common amino acids at the C-terminus. However, the amino acids L-Cys, D-Cys and D-

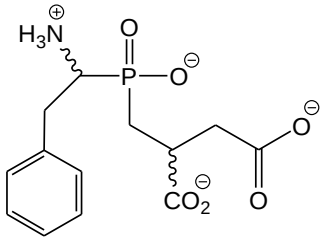
Ile were not incorporated into any of the dipeptide libraries. A sample of these libraries, expected to contain ~2 nmol of each library component, was submitted for amino acid analysis (AAA) to quantify the amount of every amino acid in these libraries except for tryptophan. The amount of tryptophan in these libraries was determined by the absorbance at 280 nm.

Synthesis of Phosphinate Pseudodipeptides. Six racemic phosphinate pseudodipeptide mimics of the putative tetrahedral intermediates for the hydrolysis of Ala-Asp (**4**), Phe-Asp (**5**) and Tyr-Asp (**6**), Ala-Ala (**7**), Leu-Ala (**8**) and Leu-Glu (**9**) were made as potential inhibitors of Sco3058. These compounds were synthesized according to previously described protocols (*82-85*). The structures of these compounds are presented in **Scheme 3.2**.

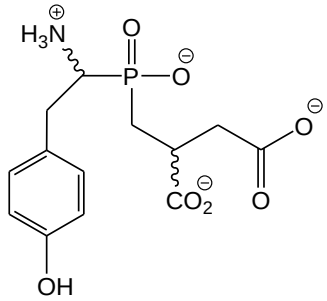
Scheme 3.2 Phosphinate pseudodipeptide mimics



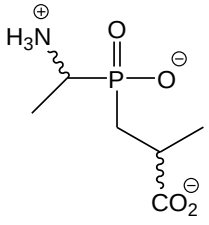
(4)



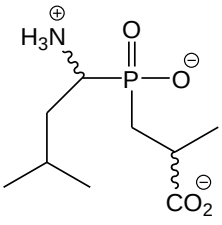
(5)



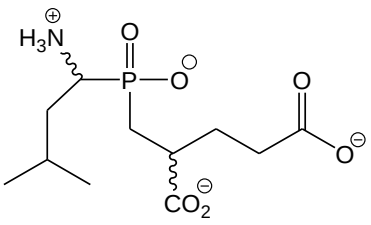
(6)



(7)



(8)



(9)

Cloning and Expression of Sco3058. The gene for Sco3058 (gi|21221500) was amplified from *Streptomyces coelicolor* A3(2) genomic DNA with the primers 5'-GCAGGAGCCATATGACATCGCTGGAGAAGGCCCGCGAGCTGCTGCGC-3' and 5'-CGCGGAATTCAGCCTTCCGGCTGCTCGGCGGCCGTGCCG-3'. The *Pfx* platinum polymerase (Invitrogen), 0.05 units/μL PCR reaction, was used according to the manufacturer's instructions except the annealing step was deleted. The PCR product was digested with *Nde*I and *EcoR*I and ligated into the pET30a(+) vector (Novagen) with T4 ligase (New England Biolabs). *E. coli* BL21 Rosetta 2 (DE3) cells were electro-transformed with the pET30a(+)-Sco3058 plasmid. Cells harboring the pET30a(+)-Sco3058 plasmid were grown at 30 ºC in Terrific Broth containing 25 μg/mL and 50 μg/mL chloramphenicol and kanamycin, respectively. When the optical density at 600 nm reached ~0.6, 2.5 mM $Zn(acetate)_2$ was added, and the expression of Sco3058 was induced with 1.0 mM isopropyl-β-thiogalactoside (IPTG). The cells were grown overnight at room temperature.

Purification of Sco3058. The cells containing Sco3058 were suspended in a purification buffer (20 mM Tris with 100 μg/mL phenylmethanesulfonylfluoride, pH 7.5) and lysed via sonication at 0 ºC. The cell lysate was obtained by centrifugation and subsequently treated with protamine sulfate (2% w/w of cell mass) and removed by centrifugation. The protein was precipitated at 60% ammonium sulfate (371 g/L at 0 ºC), isolated by centrifugation, resuspended in 20 mM Tris, pH 7.5 and purified by size exclusion chromatography using a 3 L column of AcA34 in 20 mM Tris, pH 7.5, at a flow rate of 1.0 mL/min. The fractions containing the protein of expected size, as

assessed by SDS-PAGE, were combined. Sco3058 was further purified by ion exchange chromatography with a Resource Q column (Pharmacia) using a gradient of NaCl in 50 mM Tris, pH 7.5. A final gel filtration step, with a column of Superdex 200 (Amersham Biosciences) in 50 mM Hepes, pH 7.5, was utilized to remove the salt.

Ninhydrin-Based Enzyme Assays. The measurement of dipeptidase activity was conducted in 50 mM Hepes, pH 7.5. For measurement of the inhibitory properties of the phosphinate pseudodipeptides, 0.40 mM L-Leu-D-Ala was used as the substrate and the inhibitor concentrations were varied from 0-100 μM. All assays were done at 30 ^{o}C. The formation of free amino acids was quantified using a ninhydrin-based assay as previously described (*61*). One volume of the enzymatic reaction was mixed with two volumes of ice-cold Cd-ninhydrin reagent (0.9 % w/v ninhydrin dissolved in 1:10:80 1 g/mL $CdCl_2$: acetic acid: EtOH), heated at 80 ºC for 5 min and cooled prior to reading the absorbance at 507 nm. The correlation between the absorbance at 507 nm and the concentration of free amino acids was determined for each amino acid.

β-Lactamase Assay. Sco3058 was tested for its ability to hydrolyze the *β*-lactam nitrocefin. The hydrolysis of nitrocefin was monitored by the increase in absorbance at 486 nm (ε = 20,500 $M^{-1}cm^{-1}$) in the presence of 100 mM potassium phosphate, pH 7.0. The k_{cat}/K_m was estimated by fitting the data to equation 2.

Coupled Enzyme Assays. The substrate used for the coupled enzyme assays was L-Arg-L-Asp, unless otherwise stated. The hydrolysis of L-Arg-L-Asp was monitored by coupling the formation of L-aspartate to the oxidation of NADH in a system that includes glutamate-oxaloacetate transaminase and malate dehydrogenase. The change in

the concentration of NADH was measured spectrophotometrically using a SPECTRAmax-340 plate reader (Molecular Devices Inc.) by following the decrease in absorbance at 340 nm. The standard assay conditions contained 100 mM Tris, pH 7.5, varying concentrations of L-Arg-L-Asp, 0.36 mM NADH, 7 units of glutamate-oxaloacetate transaminase, 1.0 unit of malate dehydrogenase, 3.7 mM α-ketoglutarate, 100 mM KCl and Sco3058 in a final volume of 250 μL at 30 ºC. This assay was used to determine the kinetic parameters of L-Arg-L-Asp, L-Ala-L-Asp.

N-terminal Substrate Specificity of Sco3058. Dipeptide library screens were conducted in 50 mM Hepes, pH 8.0, at 30 ºC. Enzyme-course studies (variable enzyme concentrations for a fixed time period) were performed by incubating 19 L-Xaa-L-Xaa, 19 L-Xaa-D-Xaa and 17 D-Xaa-L-Xaa libraries with 2-2000 nM Sco3058 for 6.5, 6.5, and 18 hours, respectively. For the L-Xaa-L-Xaa and L-Xaa-D-Xaa dipeptide libraries, the absorbance at 507 nm, *Q*, was plotted as a function of Sco3058 concentration, E_t, and fit to equation 1 (Chapter II). In the case of the D-Xaa-L-Xaa libraries, the free amino acids were quantified by measuring the absorbance at 507 nm and the linear rates of hydrolysis as a function of enzyme concentration were compared for each D-Xaa-L-Xaa library.

C-terminal Substrate Specificity of Sco3058. Enzyme-course assays for seven dipeptide libraries: L-Arg-L-Xaa, L-Trp-L-Xaa, L-Ala-L-Xaa, L-Met-L-Xaa, L-Glu-L-Xaa, L-Arg-D-Xaa and D-Leu-L-Xaa were analyzed by HPLC for determination of the C-terminal specificity. With the exception of the D-Leu-L-Xaa library, each library was incubated at 30 ºC for 5 hours, with and without enzyme, in 25 mM $(NH_4)HCO_3$, pH 8.0.

The D-Leu-L-Xaa library was incubated under the same conditions for 18 hours. The concentrations of enzyme used were 0-1000 nM (L-Ala-L-Xaa), 0-200 nM (L-Arg-D-Xaa), 0-250 nM (L-Arg-L-Xaa, L-Trp-L-Xaa and L-Met-L-Xaa) or 20-4000 nM (D-Leu-L-Xaa). The sample preparation and data analysis were conducted as described previously in Chapter II (*56*).

Data Analysis. The kinetic parameters, k_{cat}, K_m, and k_{cat}/K_m, were determined by fitting the initial velocity data to equation 2 (Chapter II) where ν 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. The kinetic constants for the inhibition of Sco3058 by the pseudodipeptide mimics were obtained by a fit of the data to equation 3 where K_i is the competitive inhibition constant.

Crystallization and Data Collection. Crystallization screens and data collection/analysis for the native Sco3058 and Sco3058 complexed with **4** were conducted by Alexander Fedorov and Lena Fedorov in Dr. Steve Almo's lab at Albert Einstein College of Medicine as previously described (*80*). The wild-type Sco3058·Zn^{2+} complex was crystallized in a solution containing 14.0 mg/mL Sco3058 (in 50 mM Hepes, pH 7.5), 22% polyacrylic acid (precipitant), 0.1 M Hepes pH 7.5 and 20 mM $ZnCl_2$. Crystals of the Sco3058·Zn^{2+} complex were soaked for 4 hours in a cryo-buffer consisting of 22% polyacrylic acid, 0.1 M Hepes pH 7.5, 0.5 mM $ZnCl_2$, 30% Ethylene Glycol, and 50 mM pseudodipeptide inhibitor **4** in order to obtain the ternary complex of Sco3058·Zn^{2+}·pseudodipeptide **4**.

Crystallization and data collection/analysis of the Sco3058·Zn^{2+}·citrate·glycerol complex were conducted by Tinh Nguyen as previously described (*80*). Crystals were grown by hanging drop vapor diffusion at room temperature using 0.7-0.9 M citrate, 0.1 M imidazole, pH 8.0, and 12.0 mg/mL enzyme. Data for the Sco3058·Zn^{2+}·citrate·glycerol complex were collected to a maximum resolution of 1.7 Å.

Network analysis. An all-by-all BLAST of the ~ 400 amino acid sequences within cog2355 was conducted in a similar manner as that described for the members of cog3653 in Chapter II. A protein similarity network based on the BLAST (Basic Local Alignment Search Tool) results was generated using Cytoscape (*65*).

RESULTS

Cloning, Expression and Purification of Sco3058. The gene for Sco3058 was successfully cloned into the pET30a(+) vector. The protein was expressed at moderate levels and purified to homogeneity. The identity of Sco3058 was confirmed by sequencing the first five amino acids from the N-terminus of the purified protein. The amino acid sequence (90% MTSLE and 10% TSLEK) matched that of MTSLEK predicted from the DNA sequence for this enzyme. ICP-MS confirmed the presence of 2.0 equivalents of Zn per subunit.

N-terminal Substrate Specificity of Sco3058. The substrate specificity of Sco3058 for the amino acid at the N-terminus of the various dipeptide libraries was determined by measuring the composite rate of dipeptide hydrolysis in these libraries. A typical enzyme-course plot for the hydrolysis of the L-Arg-L-Xaa library is presented in **Figure 3.1A**. An enzyme-course plot for the slower hydrolysis of the D-Xaa-L-Xaa

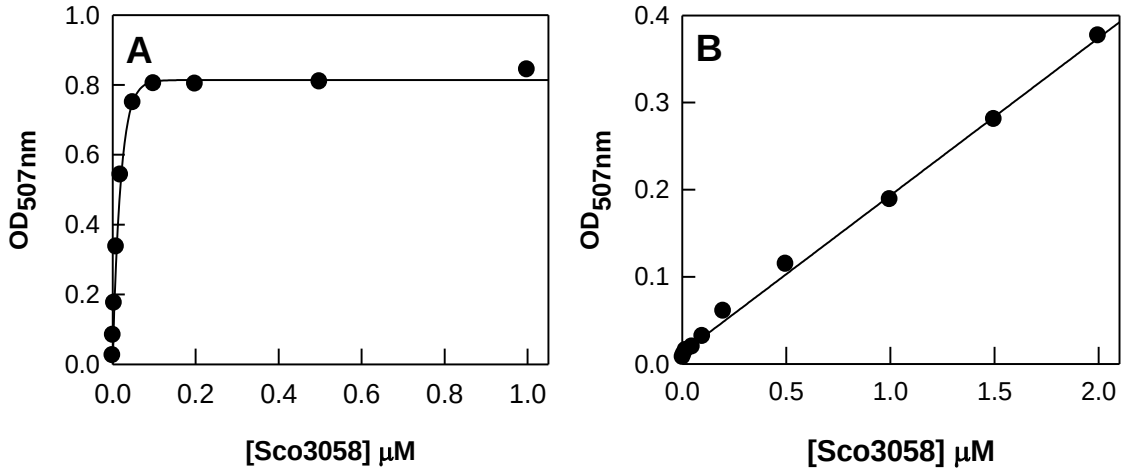


Figure 3.1: Typical enzyme course plots for the hydrolysis of dipeptide libraries. (**A)** L-Arg-L-Xaa treated with various concentrations of Sco3058. (**B)** D-Val-L-Xaa treated with various concentrations of Sco3058. Additional details are found in the text.

dipeptide libraries using D-Tyr-L-Xaa as an example is presented in **Figure 3.1B**. The relative hydrolysis rates for the L-Xaa-L-Xaa, L-Xaa-D-Xaa and D-Xaa-L-Xaa dipeptide libraries are presented in **Figures 3.2A-C**. These libraries contain a fixed amino acid at the N-terminus and 17 to 19 different amino acids at the C-terminus. For the L-Xaa-L-Xaa dipeptide libraries the N-terminal preference was for a leucine residue with slightly lower turnovers for methionine, arginine, and glutamine. For the L-Xaa-D-Xaa series of dipeptides the preference at the N-terminus is for arginine, glutamine, methionine, and leucine. With the series of D-Xaa-L-Xaa dipeptide libraries, the preferred amino acids at the N-terminus are methionine, leucine, arginine, and lysine.

C-terminal Substrate Specificity for Sco3058. The observed rates for the liberation of specific amino acids from the C-terminus of the dipeptides contained within the L-Arg-L-Xaa, L-Arg-D-Xaa, L-Ala-L-Xaa, L-Trp-L-Xaa, L-Met-L-Xaa, L-Glu-L-Xaa and D-Leu-L-Xaa are presented in **Figures 3.3A-G**, respectively. The relative rates were obtained from fits of the data to equation 1. Within each library, dipeptides that terminated with aspartate were hydrolyzed the fastest. Proline was not detected in any of the hydrolysis products and thus dipeptides that terminate in proline are not hydrolyzed. Dipeptides with L-Val, L-Thr, L-Arg or L-Lys at the C-terminus were consistently hydrolyzed more slowly than the remaining 14 L-amino acids given the same amino acid at the N-terminus.

Kinetic Constants for Selected Substrates. The kinetic constants, k_{cat}, K_m and k_{cat}/K_m, for 25 dipeptide substrates are presented in **Table 3.1**. Of the compounds tested L-Ala-L-Asp has the highest value for k_{cat} (1400 s^{-1}) and L-Arg-D-Asp has the highest

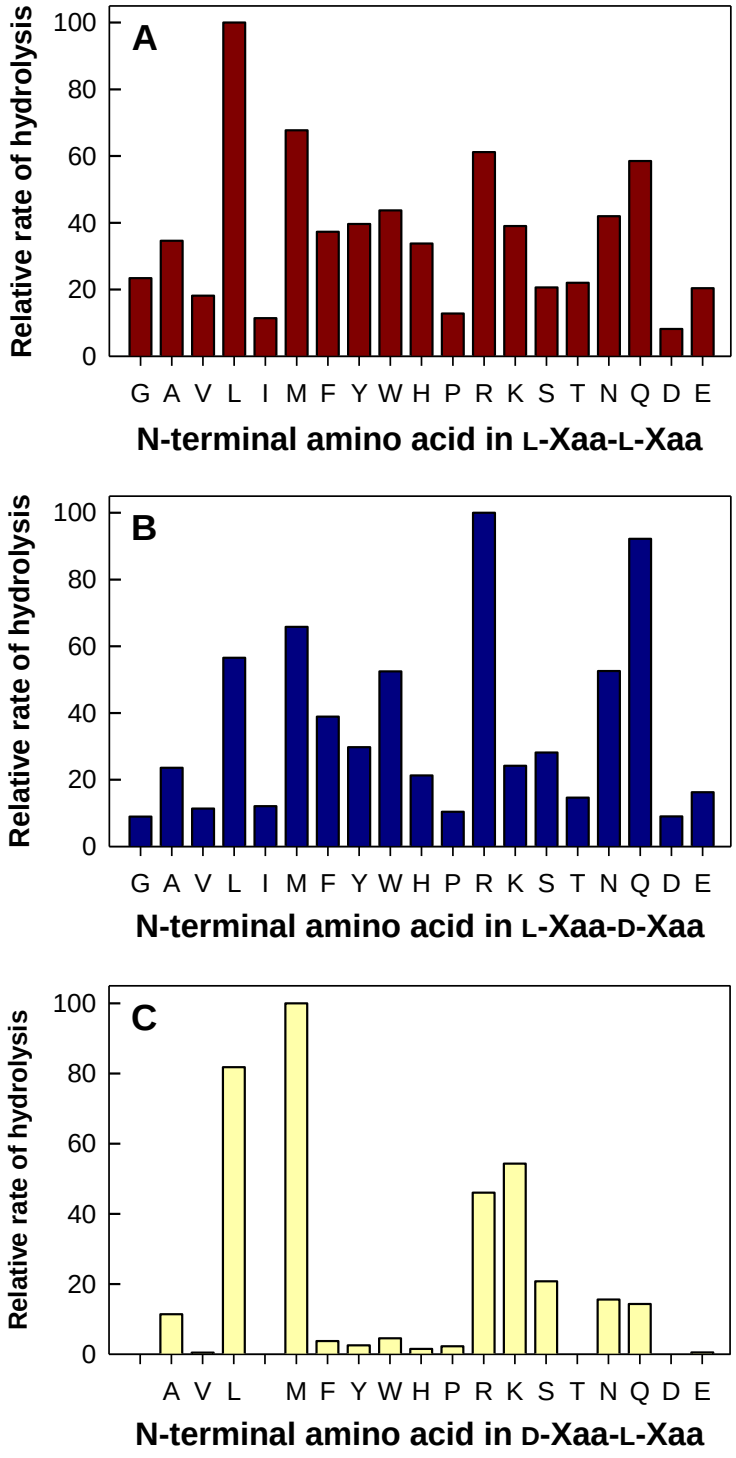


Figure 3.2: N-terminal specificity of Sco3058 against 55 dipeptide libraries. **(A)** Relative rates of hydrolysis for 19 L-Xaa-L-Xaa dipeptides. **(B)** Relative rates of hydrolysis for 19 L-Xaa-D-Xaa dipeptide libraries. **(C)** Relative rates of hydrolysis for 17 D-Xaa-L-Xaa dipeptide libraries.

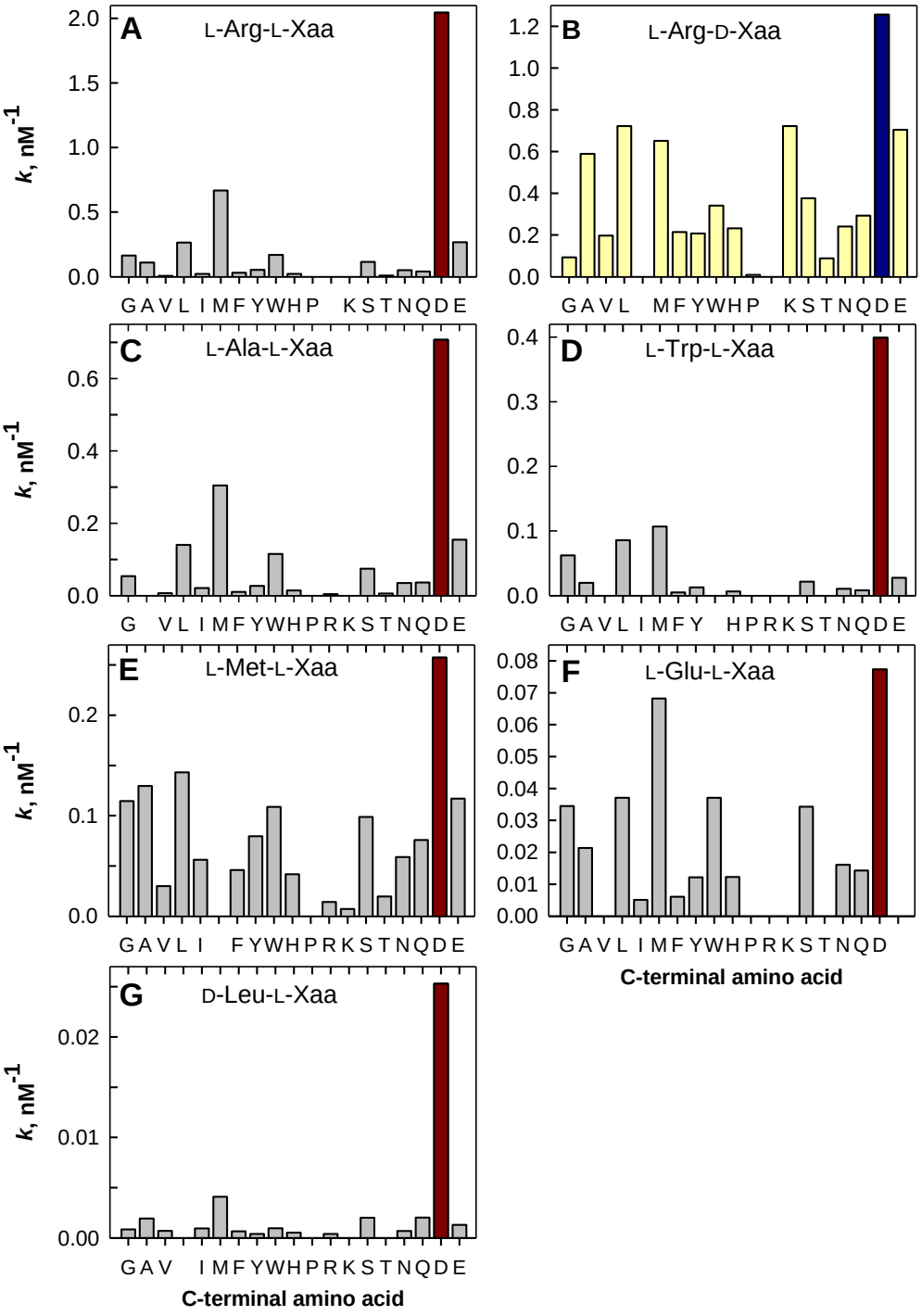


Figure 3.3: C-terminal specificity of Sco3058 with seven dipeptide libraries. The *k* (nM^{-1}) for the hydrolysis of each dipeptide within the **(A)** L-Arg-L-Xaa, **(B)** L-Arg-D-Xaa, **(C)** L-Ala-L-Xaa, **(D)** L-Trp-L-Xaa, **(E)** L-Met-L-Xaa, **(F)** L-Glu-L-Xaa and **(G)** D-Leu-L-Xaa dipeptide libraries. All L-amino acids are shaded in gray except for L-Asp (*red*). The D-amino acids are shaded in yellow except for D-Asp (*blue*).

Table 3.1: Kinetic constants for select substrates of Sco3058[a]

Substrate	k_{cat}, s^{-1}	K_m, mM	k_{cat}/K_m $M^{-1}s^{-1}$
L-Arg-D-Asp	107 ± 3.2	0.14 ± 0.02	(7.6 ± 0.9) x 10^5
L-Met-D-Glu	340 ± 4	0.89 ± 0.04	(3.8 ± 0.2) x 10^5
L-Met-D-Leu	194 ± 4	0.56 ± 0.03	(3.5 ± 0.3) x 10^5
L-Leu-D-Glu	135 ± 2	0.55 ± 0.03	(2.5 ± 0.1) x 10^5
L-Leu-D-Ala	160 ± 3	0.75 ± 0.04	(2.1 ± 0.1) x 10^5
L-Leu-D-Ser	400 ± 10	2.0 ± 0.2	(2.0 ± 0.1) x 10^5
L-Ala-D-Ala	750 ± 25	4.4 ± 0.3	(1.7 ± 0.1) x 10^5
L-Tyr-D-Leu	24 ± 0.2	0.17 ± 0.01	(1.5 ± 0.07) x 10^5
L-Met-L-Leu	310 ± 6	2.3 ± 0.1	(1.4 ± 0.1) x 10^5
L-Asn-D-Glu	288 ± 10	2.4 ± 0.2	(1.2 ± 0.1) x 10^5
L-Arg-L-Asp	515 ± 4	4.3 ± 0.1	(1.2 ± 0.4) x 10^5
L-Ala-L-Asp	1400 ± 60	13.8 ± 0.9	(1.0 ± 0.1) x 10^5
L-Leu-L-Leu	140 ± 6	1.4 ± 0.1	(1.0 ± 0.1) x 10^5
L-Leu-L-Asp	160 ± 11	2.0 ± 0.3	(8.3 ± 1.4) x 10^4
L-Leu-D-Leu	21 ± 0.3	0.37 ± 0.03	(5.6 ± 0.4) x 10^4
Gly-D-Leu			(3.3 ± 0.1) x 10^4
L-Ala-L-Glu	220 ± 14	9.5 ± 1.0	(2.4 ± 0.3) x 10^4
L-Arg-Gly			(2.2 ± 0.1) x 10^4
L-Ala-L-Ala			(1.4 ± 0.04) x 10^4
L-Ala-L-Gln	140 ± 17	16 ± 3.3	(8.4 ± 1.9) x 10^3
L-Ala-L-His			(4.5 ± 0.1) x 10^2
L-Ala-L-Arg			(1.8 ± 0.1) x 10^2
D-Ala-D-Ala			(1.3 ± 0.05) x 10^2
L-Ala-L-Lys			(7.2 ± 0.2) x 10^1
D-Ala-L-Ala			(2.2 ± 0.05) x 10^0

[a]From fits of the data to equation 2. For those entries without values of k_{cat} and K_m, saturation was not achieved at concentrations up to 5 mM.

value for k_{cat}/K_m (7.6 x 10^5 M^{-1} s^{-1}). For the four enantiomers of the dipeptide Ala-Ala, the relative rates of hydrolysis are of the order L-Ala-D-Ala > L-Ala-L-Ala > D-Ala-D-Ala > D-Ala-L-Ala. The rates of hydrolysis of *N*-formyl-D-Glu, *N*-acetyl-D-Glu and *N*-propionyl-D-Leu at a concentration of 2.0 mM were less than 2 x 10^{-3} s^{-1}. Nitrocefin was slowly hydrolyzed by Sco3058 but the value of k_{cat}/K_m was less than 0.1 $M^{-1}s^{-1}$. The phosphinate pseudodipeptide mimics **4**, **5**, **6**, **7**, **8** and **9** were found to be competitive inhibitors of Sco3058 (**Figure 3.4**) with K_i values of 390 ± 21 nM, 2.3 ± 0.3 μM, 1.1 ± 0.2 μM, 1.8 ± 0.2 μM, 270 ± 10 nM and 1.7 ± 0.1 μM, respectively.

Structure of Sco3058. The crystal structures of Sco3058, Sco3058 with glycerol/citrate bound and Sco3058 in complex with the pseudodipeptide of L-Ala-D-Asp (**4**) in the active site were determined to 1.3, 1.7 and 1.4 Å, respectively. A ribbon representation of the $(\beta/\alpha)_8$-barrel structure is presented in **Figure 3.5**. The zinc in the α-site (M_α) is coordinated to His-20 and Asp-22 from the HxD motif that originates from the end of β-strand 1. The zinc in the β-site (M_β) is ligated to the two conserved histidines (His-191 and His-212) from the ends of β-strands 5 and 6, respectively. The two metals are bridged by Glu-123 from β-strand 3 and a hydroxide/water molecule. The next nearest water molecule to either of the two metal ions is 2.4 Å away from M_β. The conserved aspartate (Asp-320) from the end of β-strand 8 is 3.6 Å from M_α and thus does not coordinate this metal. However, this residue interacts with the bridging solvent molecule at a hydrogen bonding distance of 3.0 Å. The binuclear metal center is shown in **Figure 3.6A**.

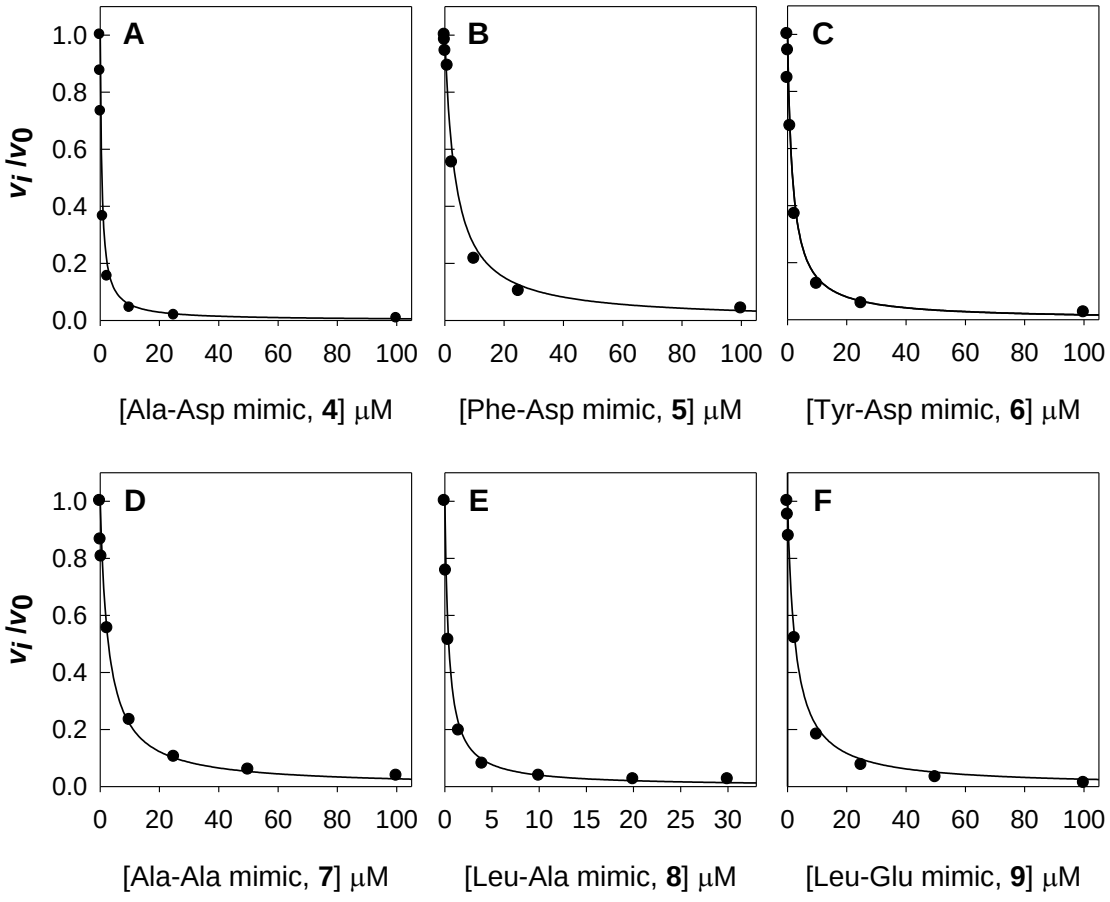


Figure 3.4: Inhibition plots for Sco3058 with pseudodipeptide mimics. The fractional velocities were plotted as a function of the concentration of compounds **4 (A)**, **5 (B)**, **6 (C)**, **7 (D)**, **8 (E)** and **9 (F)**.

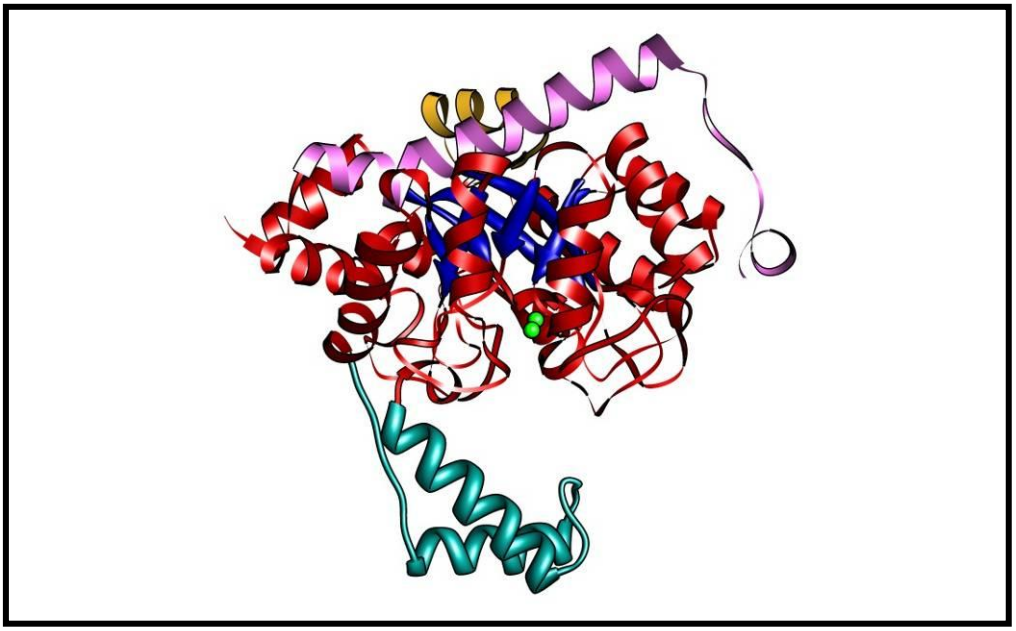


Figure 3.5: Ribbon diagram of the Sco3058 structure. The $(\beta/\alpha)_8$-barrel is colored with *blue* β-strands and *red* α-helices and loops. Insertion regions 1 (residues 98-115 between α_2 and β_3) and 2 (residues 251-295 between β_7 and α_7) are colored *yellow* and *teal*, respectively. The C-terminus (residues 351-391) is colored *pink*. The active site metals are shown in *green*.

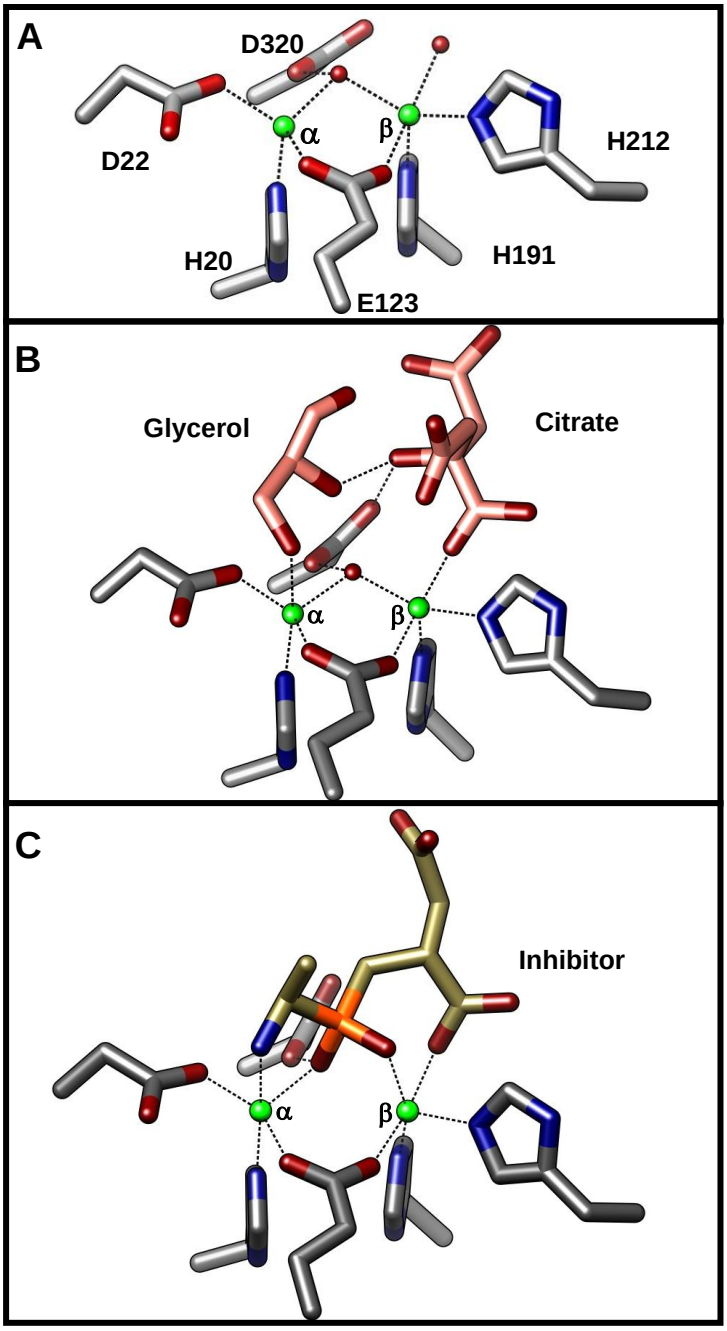


Figure 3.6: Binuclear metal center of Sco3058. Contacts between the zinc ions (*green* spheres) and water (*red* sphere) or active site residues are 1.9-2.2 Å and indicated with dashed lines. **(A)** Metal center within the native structure of Sco3058. **(B)** Structure of the binuclear metal center with citrate and glycerol (*pink* carbons) in the active site. **(C)** Structure of the binuclear metal center with the phosphinate pseudodipeptide **4** (*gold* carbons, *orange* phosphorus) bound in the active site.

The active site of Sco3058 in the presence of glycerol and citrate is shown in **Figures 3.6B** and **3.7A**. In the glycerol/citrate bound structure, the α-metal is tetrahedral in coordination while the β-metal exhibits a distorted trigonal bipyramidal geometry. In addition to the four ligands described for the native structure, the β-metal is additionally coordinated to the C3-carboxylate of citrate. This carboxylate is also ion-paired with Arg-223 and occupies the same position as the water molecule, which is 2.4 Å from M_β in the native structure. The carboxylate of citrate from the *pro*-S arm is ion-paired to His-150 while the carboxylate from the *pro*-R arm is interacting with Thr-324. The hydroxyl group attached to C3 of citrate is 2.8 Å from Asp-320 and is also hydrogen bonded with the C2 hydroxyl group of glycerol. In addition, the C1 hydroxyl group of glycerol is within hydrogen bonding distance (3.0 Å) to Asp-22 and 2.4 Å from the α-metal.

The two chiral centers that mimic the C_α stereocenters of the N- and C-terminal ends of a dipeptide are both in the R-configuration. This would correspond to the dipeptide L-Ala-D-Asp. In this complex the β-zinc is coordinated by five ligands in a distorted trigonal bipyramidal arrangement. One of these interactions originates from the phosphinate group and another is from the carboxylate that would correspond to the C-terminal carboxylate of a dipeptide substrate. The α-metal is in a distorted square pyramidal arrangement and the α-amino group of the inhibitor is 2.3 Å away from this metal ion. The C-terminal carboxylate of the inhibitor is ion-paired with the guanidino group of Arg-223. His-150 makes polar interactions with the phosphinate moiety at a distance of 2.7 Å. The side chain carboxylate of the inhibitor interacts with Thr-324 at a

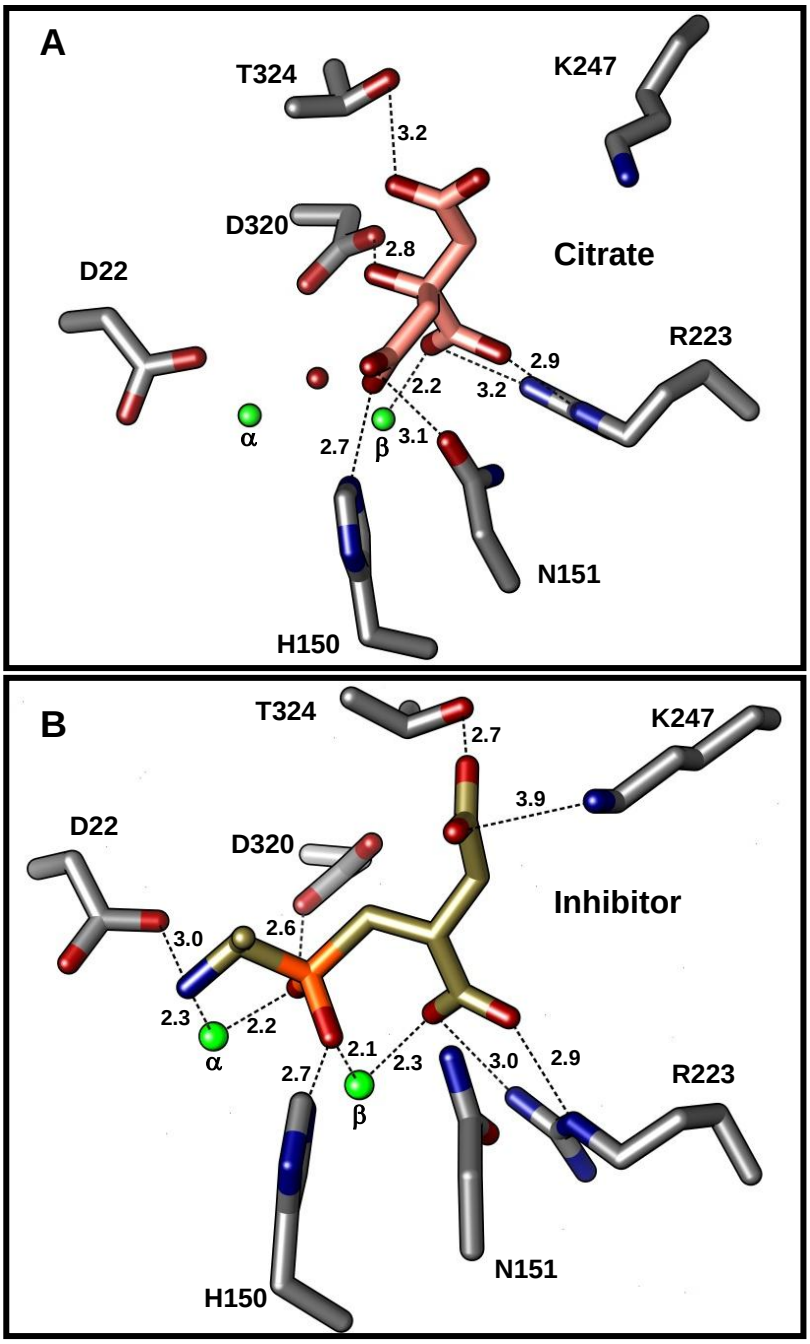


Figure 3.7: Interactions of citrate and pseudodipeptide inhibitor **4** with active site metals and residues within Sco3058. **(A)** Polar contacts between citrate (*pink* carbons) bound in the active site of Sco3058 (*gray* carbons). **(B)** Polar contacts between inhibitor **4** (*gold* carbons and *orange* phosphorus) and Sco3058 (*gray* carbons). Polar contacts are indicated with dashed lines with distances displayed in Å. The zincs and bound water molecule are represented as *green* and *red* spheres, respectively. The distance between K247 and **4** is shown for the purpose of discussion.

distance of 2.7 Å. The orientation of inhibitor **4** in the active site of Sco3058 is presented in **Figures 3.6C** and **3.7B**.

Network Analysis of cog2355. Approximately 400 microbial renal dipeptidase-like amino acid sequences were identified in the NCBI database from the completely sequenced organisms. At an E-value cutoff of 1×10^{-70}, the sequences can be separated into 27 clusters of 3 or more sequences as shown in **Figure 3.8**. Each node in the network represents a single amino acid sequence and each line represents a pairwise connection between two sequences with a BLAST score of $E \leq 1 \times 10^{-70}$. Sco3058 can be found in Cluster14.

DISCUSSION

Substrate Specificity. The substrate specificity of Sco3058 was determined using an array of dipeptide libraries. It is evident from the data presented in **Table 3.1** that this enzyme has the ability to hydrolyze a rather broad range of dipeptides having amino acids of either the D- or L-configuration. However, at the N-terminus, amino acids of the L-configuration are greatly preferred over those of the D-configuration. For example, the values of k_{cat}/K_m for L-Ala-D-Ala and L-Ala-L-Ala are 2-3 orders of magnitude greater than their D-Ala-D-Ala and D-Ala-L-Ala counterparts. Conversely, at the C-terminus, amino acids of the D-configuration are preferred over those of the L-configuration. Thus, the values of k_{cat}/K_m for L-Arg-D-Asp and L-Ala-D-Ala are about an order of magnitude greater than those of L-Arg-L-Asp and L-Ala-L-Ala, respectively. At the N-terminus there is a preference for arginine, methionine, glutamine, and leucine whereas at the C-terminus there is a clear preference for glutamate and aspartate. The

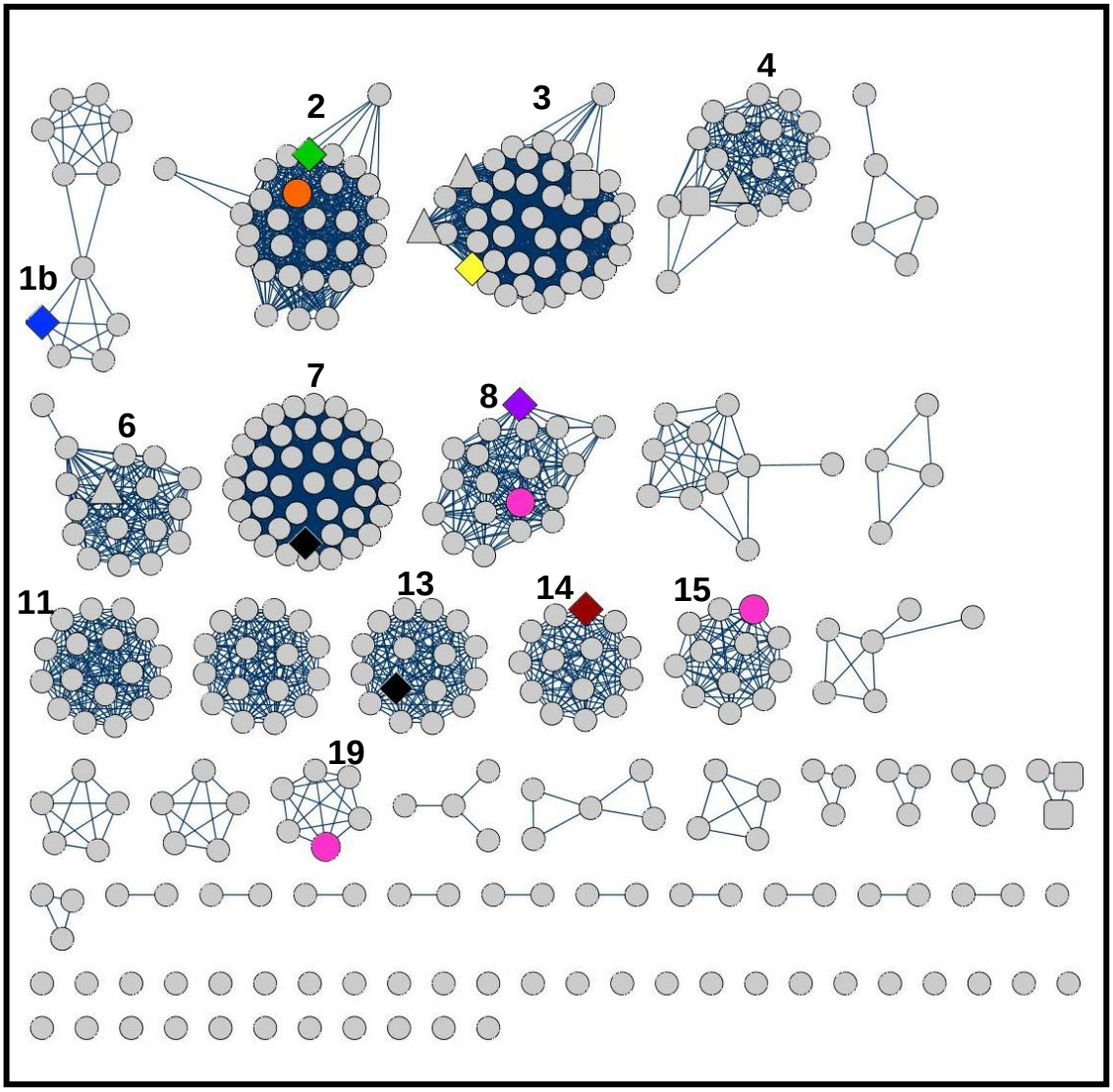


Figure 3.8: Cytoscape graphical representation of putative renal dipeptidase-like proteins from cog2355. The amino acid sequences (circles) are separated into clusters based on an E-value cut-off of 1 x 10^{-70}. Circles which are connected by lines have an E-value ≤ 1 x 10^{-70}. Proteins with a crystal structure are shaped as diamonds. The protein targets described in Chapters III, IV and V are colored *blue* (Gox2272), *green* (LmoDP), *orange* (Bh2271), *yellow* (Rsp0802), *purple* (Cc2746), and *dark red* (Sco3058). Three other targets mentioned in the text are colored *pink*. Two proteins within a smaller unstudied subset of cog2355 which have deposited structures in the PDB are colored *black* (PDB codes *2I5G* and *3B40*). Other amino acid sequences with the same shape are examples of multiple sequences within cog2355 from the same species and strain (i.e. paralogues), such as four sequences from *Sinorhizobium meliloti* or *Ruegeria pomeroyi* (triangles).

very best substrates have k_{cat}/K_m values that exceed 10^5 M^{-1} s^{-1}. *N*-acyl derivatives of D-amino acids are hydrolyzed approximately 5 orders of magnitude more slowly than are the corresponding dipeptide substrates and thus this enzyme appears to be a true dipeptidase rather than a carboxypeptidase.

Structural Determinants of Substrate Specificity. The three dimensional structure of Sco3058 was determined in the absence of ligands and also in the presence of citrate and a pseudodipeptide analog bound in the active site. The analogue of the proposed tetrahedral intermediate was synthesized without regard for the stereochemistry at the two chiral centers. The structure of the inhibitor bound in the active site is a mimic of the dipeptide L-Ala-D-Asp. As a structural mimic of the tetrahedral addition complex, this analogue is compromised since the amide nitrogen has been substituted with a methylene group to maintain hydrolytic stability of the inhibitor. Nevertheless, the structure of this complex indicates that the free α-amino group at the N-terminus of the pseudodipeptide interacts with M_α and the side chain carboxylate of Asp-22. Conversely, the free α-carboxylate at the C-terminus interacts with M_β and the side chain guanidino group of Arg-223. The phosphinate group of the inhibitor interacts with M_β and the side chain of His-150. The side chain carboxylate of the inhibitor interacts with Thr-324 and is in close proximity to the ε-amino group of Lys-247.

In an overlay of the Sco3058·citrate·glycerol and Sco3058·pseudodipeptide **4** complexes (RMSD = 0.3 Å), using MatchMaker tool in Chimera, free amino groups of the Lys-247 residues are 2.2 Å apart. The carboxylate from the *pro*-R arm of citrate is 2.7 Å from Lys-247 from the pseudodipeptide **4** complex. The hypothetical position of

the side chain of a C-terminal L-Asp was modeled with the pseudodipeptide complex by building a CH_2COOH side chain onto the chiral carbon of compound **4** that represents the C_α carbon of a dipeptide. In this model, the side chain carboxylate of the "L-Asp" side chain was 2.7 Å and 3.3 Å from the side chain nitrogens of Asn-151 and Lys-247, respectively. The oxygen and nitrogen of the side chain of Asn-151 are in opposite positions within these two structures (**Figure 3.7**). Realistically, the position of the oxygen and nitrogen atoms of the Asn-151 side chain should be positioned as they are in **Figure 3.7A** where the carbonyl is 3.1 Å from N^δ of the histidine ligand from β_5 (H191).

A similar set of interactions is apparent in the binding of citrate in the active site of Sco3058. In many ways the structure of citrate mimics the pseudodipeptide inhibitor. Thus, the carboxylate at C3 interacts with M_β and Arg-223 in much the same way as the free α-carboxylate of the inhibitor. The carboxylate from the *pro*-R arm of citrate interacts with Thr-324 similar to how the side chain carboxylate of the inhibitor interacts with this residue. The carboxylate from the *pro*-S arm of citrate interacts with His-150 in a fashion that duplicates the interactions with the phosphinate moiety with this same residue. The closest residue to the hydroxyl group of citrate is the side chain carboxylate of Asp-320. However, there are certain interactions that are missing and this may explain why the pseudodipeptide inhibitor binds with a dissociation constant that is about 4 orders of magnitude tighter than citrate (*80*).

The dipeptides L-Ala-D-Ala and D-Ala-L-Ala were docked into the Sco3058·Zn^{2+} structure by Peter Kolb from Dr. Schoichet's lab at University of California San Francisco as previously described (*80*). The dipeptides were docked as their high energy

intermediates following attack by a hydroxide, where the carbon of the amide bond is tetrahedral and bonded to a hydroxyl group and a negatively-charged oxygen. Overlays of both of these docked structures with the Sco3058·Zn^{2+}·inhibitor **4** complex are presented in **Figure 3.9**. The pose of the D-Ala-L-Ala dipeptide (**Figure 3.9A**) suggests that the free amino group of the N-terminal D-amino acid contacts the backbone of Gly-323 instead of the aspartate from β_1, and the methyl group of D-Ala is closer in position to the free amino group of inhibitor **4**. The α-carboxylates of D-Ala-L-Ala and inhibitor **4** are almost in the exact same position, but the oxygens which originate from hydroxide attack are distant from each other. Conversely, in **Figure 3.9B**, the α-carboxylates of L-Ala-D-Ala and inhibitor **4** are oriented differently, but the two phosphinate oxygens are in a very similar position to the oxygens bonded to the carbon of the scissile bond. If all D-Xaa-L-Xaa dipeptides were to orient with the C_β carbon in the same position as the methyl group modeled in **Figure 3.9A**, one would expect that the rate of hydrolysis of the D-Ala-L-Xaa library would be at least comparable to the rates observed for the D-Leu-L-Xaa and D-Met-L-Xaa libraries. The substrate profile of Sco3058 with the D-Xaa-L-Xaa was quite polarized with preferences for D-Met, D-Leu, D-Lys and D-Arg, but the reason for this is not completely clear. Still, D-Xaa-L-Xaa dipeptides are not considered to be physiologically relevant substrates of Sco3058.

Genome Context for Sco3058 and Cluster 14 Proteins in cog2355. The genome context of Sco3058 would not have been useful in predicting the general reaction of this protein or the substrate specificity. The neighboring protein in the genome, Sco3057, has ~40% amino acid sequence identity to Sco3058, but Sco3057 does not have a single

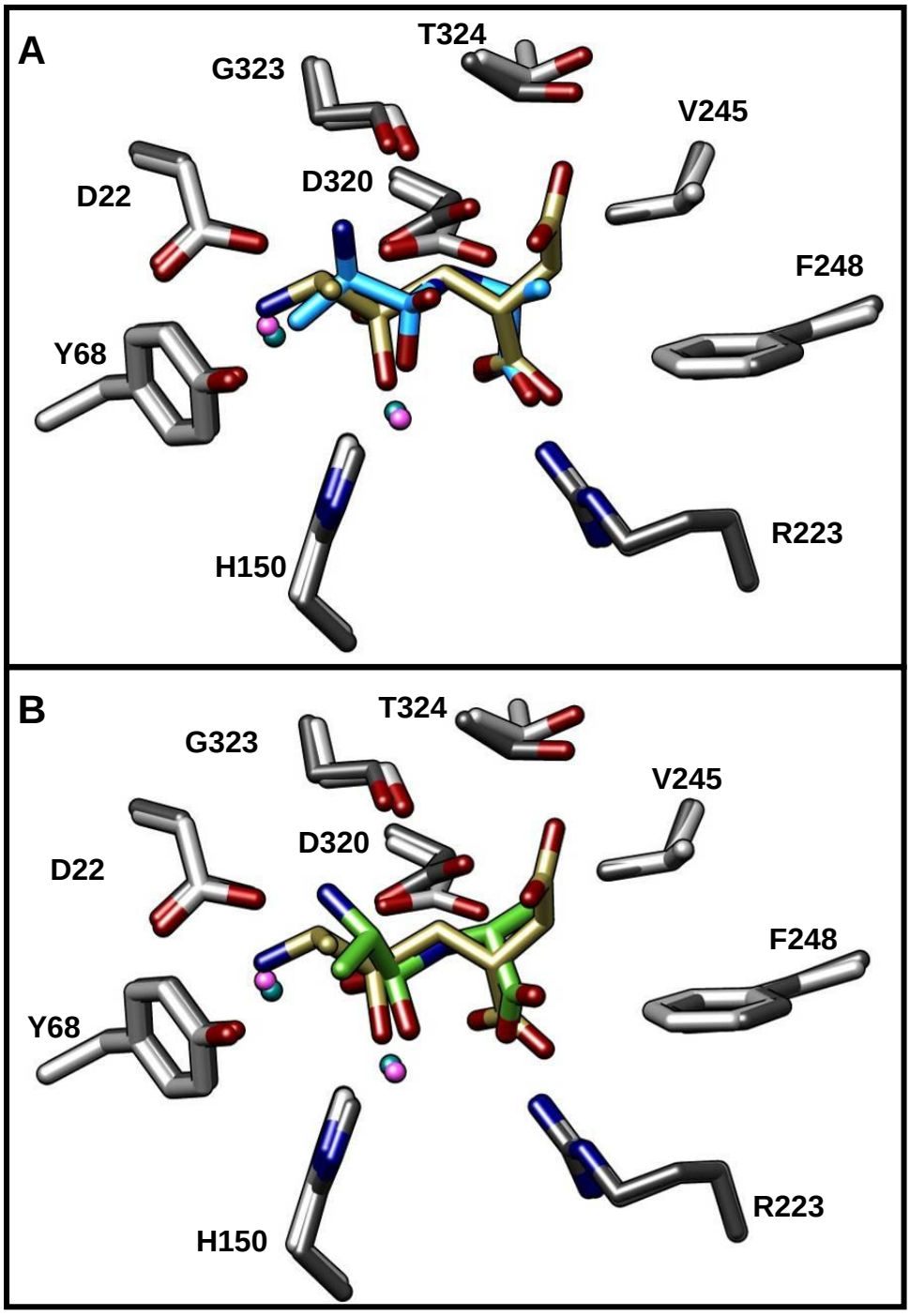


Figure 3.9: Overlay of the active sites of the L-Ala-D-Asp pseudodipeptide·Sco3058 complex (*dark gray* carbons, *pink* Zn and *gold* inhibitor carbons) and Sco3058 docked with Ala-Ala dipeptides. **(A)** The docked D-Ala-L-Ala-Sco3058 complex (*light gray* carbons, *teal* Zn, and *light blue* substrate carbons). **(B)** The docked L-Ala-D-Ala·Sco3058 complex (*light gray* carbons, *teal* Zn, and *green* substrate carbons).

metal-binding residue in common with Sco3058. The Sco3058 gene is also adjacent to *Pur*E (Sco3059) and *Pur*K (Sco3060) which code for proteins annotated as phosphoribosylaminoimidazole carboxylase catalytic and ATPase subunits, respectively. It is conceivable that Sco3058 could either hydrolyze a molecule such as 4-[(*N*-succinylamino)carbonyl]-5-aminoimidazole ribonucleotide (SAICAR) to liberate aspartate or supply aspartate for the SAICAR synthetase reaction through the cleavage of a Xaa-L-Asp dipeptide (*86*). However, the immediate proximity of *Pur*E/*Pur*K to genes which code for renal dipeptidase-like proteins is not a common feature of the genome contexts of these proteins.

To the best of my knowledge, the protein encoded by the *Sir*J gene from *Leptosphaeria maculans* (gi|46403052) is one of only two renal dipeptidase-like proteins most closely associated to Cluster 14 which have been suggested to be involved in a biosynthetic pathway based on genome context. The *Sir*J gene is clustered with genes involved in the biosynthesis of fungal phytotoxin sirodesmin (*87*). The second is the putative protein product of the *Gli*J gene, coding for a homolog of SirJ, located in a putative gene cluster for the biosynthesis of gliotoxin (*88*). Though the SirJ protein was not part of the network analysis, it has connections to some sequences in Cluster 14 with E-values of 10^{-67}–10^{-69}.

Part of the biosynthesis of sirodesmin includes the formation of a derivatized cyclic dipeptide intermediate from the amino acids L-Tyrosine and L-Serine, presumably via the putative nonribosomal peptide synthetase (NRPS) SirP. SirP was said to be missing the final cyclization domain. Prior to condensation with L-Serine, the phenol

group of L-Tyr residue is modified by an *O*-prenyltransferase coded for by the *Sir*D gene. The SirD protein only modifies L-Tyr but not the cyclic L-Tyr-L-Ser dipeptide (*89*). In general, the adenylation (A) domains of NRPSs are considered to be responsible for the selection of the amino acid with the correct side chain and chirality for the growing peptide which sidesteps the need for a substrate-specific condensation (C) domain. However, there have been observations to the contrary where the A-domain has a more relaxed specificity, and the C-domain has an editing function (*90*). The fidelity of the A and C domains of SirP has not been determined empirically with purified protein. Therefore, whether or not SirP is specific for *O*-prenyl-L-Tyr and L-Ser is not known. The biosynthetic pathway could be relying on the ability of an L-Tyr A-domain to accept an L-Tyr derivative, in which case the undesirable L-Tyr-L-Ser is likely to be synthesized and would need to be recycled. The SirJ sequence has the critical residues in common with the active dipeptidases within cog2355.

Given that sirodesmin and *cyclo*-L-Tyr-L-Ser both lack the functional groups which make polar contacts with the enzyme, namely the free amino group or the α-carboxylate, it seems unlikely that the function of SirJ would be to cleave the cyclic dipeptide or the amide bonds within the toxin. On the basis of the structures of Sco3058 complexed with citrate and inhibitor **4**, it is hypothesized that Lys-247 is considered to be a residue which can make polar contacts with the C-terminal side chain of a dipeptide. In SirJ, Lys-247 is replaced by a glutamate which could conceivably make a polar contact to the side chain of a C-terminal serine side chain. It would be reasonable to conclude that this protein would be able to hydrolyze L-Tyr-L-Ser or (*O*-prenyl-L-

Tyr)-L-Ser. It is unclear as to whether there is any other specific function of the microbial renal dipeptidase-like proteins within Cluster 14, beyond their ability to serve as true dipeptidases.

Conservation of Residues Within cog2355. There are ~400 amino acid sequences from the completely sequenced bacterial genomes which are categorized as part of cog2355. Approximately 70% of these sequences can be classified as renal dipeptidase-like because of the conservation of the complete set of metal ligands for the binuclear active site. Arginine-223 recognizes the α-carboxylate of the dipeptide, and the equivalent residue to Arg-223 is conserved in greater than 97% of the sequences found in cog2355. Within the renal dipeptidase subset of cog2355, the residue equivalent to His-150 in Sco3058 is semi-conserved with approximately 30% of the sequences having a histidine and 70% a tryptophan at this position. In the case of the renal dipeptidase-like protein from *Rhodobacter sphaeroides* (Rsp0802, PDB code *3FDG*, gi|218766909) this tryptophan residue is structurally equivalent to His-150 from Sco3058 with the nitrogen of the tryptophan side chain being in the same position as the N^{ε} of the histidine. The other residues which contact or are close to citrate or the phosphinate inhibitor (*i.e*. Thr-324, Asn-151, and Lys-247) and those which are in the vicinity of the active site (e.g. Tyr-68, Val-245, and Phe-248) are not conserved among hRDP and the cluster of renal dipeptidase-like sequences. Lys-247 is actually quite rare within cog2355. All of these residues are at the ends of or immediately follow the β-strands of the $(\beta/\alpha)_8$-barrel, which is a common location for substrate-contacting residues of AHS enzymes (*20*).

Given the lack of conservation of putative substrate recognition residues within the subset of human renal dipeptidase and its microbial homologues, there is an opportunity to identify enzymes with varying substrate specificities and define the structure-functional relationship of these enzymes. Therefore, the substrate specificity profiles and the 3-dimensional structures of several microbial renal dipeptidase-like proteins were further investigated using the same methods applied to Sco3058. Chapter IV will focus on two additional microbial dipeptidases, Gox2272 and Cc2746, from Clusters 1 and 8, respectively (**Figure 3.8**). Chapter V will address three proteins from Clusters 2 and 3 where the equivalent active site residue to His-150 in Sco3058 has been replaced by a tryptophan.

CHAPTER IV

SUBSTRATE PROFILES AND NATIVE STRUCTURES OF TWO MICROBIAL RENAL DIPEPTIDASE-LIKE PROTEINS

INTRODUCTION

Sco3058 from *S. coelicolor* is a dipeptidase which is able to hydrolyze a majority of the dipeptides to which it was exposed, though the enzyme was partial to L-Xaa-D/L-Asp substrates. A majority of the active site residues in Sco3058, besides the metal ligands, catalytic aspartate and arginine from β_6, are not conserved within the amino acid sequences found in cog2355. An amino acid sequence alignment of Sco3058 and four structurally-characterized proteins within cog2355 is displayed in **Figure 4.1**. The sequences are from *S. coelicolor* (Sco3058), *Caulobacter crescentus* (Cc2746, gi|16126978), *Gluconobacter oxydans* (Gox2272, gi|58040697), *Homo sapiens* (crystallized portion of hRDP from PDB code *1ITU*, gi|23200144) and *Pseudomonas aeruginosa* (crystallized portion of Pa2393 from PDB code *3B40*, gi|160286472). As shown for Sco3058·**4** and hRDP·cilastatin complexes, the residues highlighted in red make polar contacts to compounds bound in the proteins, namely, compound **4**, glycerol, citrate, cilastatin (**Scheme 4.1**) or a sulfate (*23, 80*).

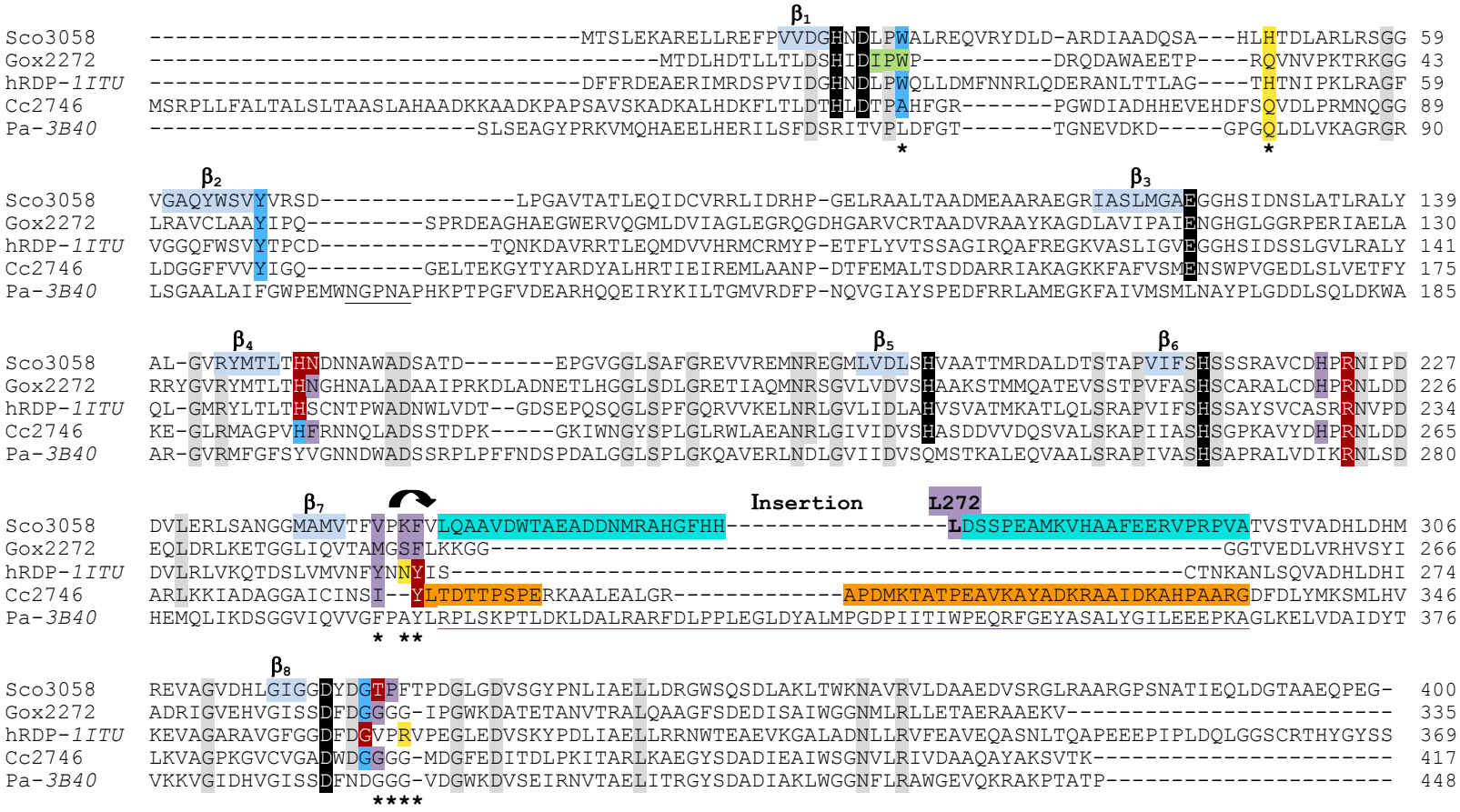
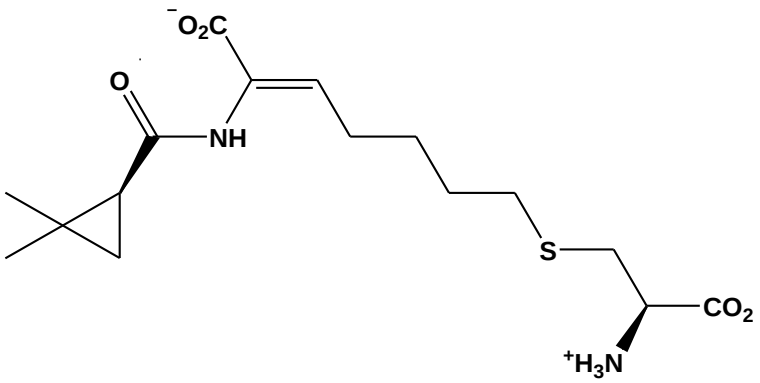


Figure 4.1: Amino acid sequence alignment of human renal dipeptidase (hRDP) and four structurally characterized microbial proteins within cog2355. The five underlined residues following β_2 and residues 1-31 of Pa-*3B40* (locus tag Pa2393) are not in the structure. The metal ligands/catalytic aspartate are shaded *black*. Residues which either make polar contacts to bound inhibitors/molecules or are predicted to contact the substrate are shaded *red*. Residues whose side chains contribute to the binding pocket of the N- (*dark blue*) or C-terminal (*purple*) sides of a dipeptide are also colored. Other conserved residues are colored *gray* and the β-strands of the $(\beta/\alpha)_8$-barrel are shaded *light blue*. Additional details are provided in the text.

Scheme 4.1 Structure of cilastatin inhibitor of human renal dipeptidase



A residue following β_8 (Sco3058-T324), two residues on a single-turn α-helix following β_7 (Sco3058-K247/F248), and the residue preceding the turn (Sco3058-V245), are part of the active site pocket for the side chain of the C-terminal amino acid of a dipeptide substrate. In hRDP, the equivalent residue to Phe-248 is tyrosine (Tyr-255) which contacts the carboxylate moiety of cilastatin equivalent to the backbone carboxylate of a dipeptide. Given that this tyrosine provides another polar contact to the α-carboxylate of cilastatin, in addition to Zn_β and an arginine, it was thought that proteins with this residue may be more effective dipeptidases than those with a phenylalanine in this position. The hRDP enzyme has reported catalytic efficiencies on the order of 10^6 $M^{-1}s^{-1}$ for the dehydrophenylalanine substrate.

The equivalent residue to Tyr-255 in hRDP is not common among the microbial proteins within cog2355, but all but one sequence in cluster 8 (**Figure 3.8**) have this residue. A protein in cluster 8 from *Xanthomonas campestris* (Xcc4228, gi|21233645) has the equivalent residue to Tyr-255 in hRDP. However, the gene for Xcc4228, cloned

into the pET30a(+) vector, failed to overexpress in *E. coli*. Cc2746 from cluster 8 is a microbial protein that also has this tyrosine residue (Tyr-284). Cc2746 was purified as a His-tag protein by NYSGXRC, and its structure has been deposited into the PDB (PDB code *2RAG*). However, in Cc2746, which lacks the single-turn α-helix (curved arrow in **Figure 4.1**) following β_7, Tyr-284 is in *nearly* the same position as Tyr-255 (hRDP) with respect to the active site.

The genomic DNA of *G. oxydans* was obtained in order to acquire the Gox1177 deacetylase (Chapter II), but Gox2272 from this organism is part of cog2355. At the time, Gox2272 had less than 40% amino acid sequence identity to any other protein from a completely sequenced organism. Gox2272 is currently part of cluster 1 (**Figure 3.8**), which can be separated into two subclusters, 1a and 1b. The six proteins in cluster 1a are putatively the fusion of two proteins from cog2355 and cog2071, but Gox2272 is from cluster 1b. Gox2272 lacks the insertion regions observed between β_7 and α_7 in Cc2746 or immediately following β_2 in Rsp0802 (PDB code *3FDG*). Both of these insertion regions were partly undefined in the structures. The residues that influence the character of the active site for the binding of the N-terminal (blue) or C-terminal (purple) portion of a dipeptide are colored in **Figure 4.1**. Between Gox2272 and Sco3058, the residues shaded blue are identical, but the purple-colored residues are different, with the exception of the asparagine following β_4.

The following describes the analysis of the crystal structures and substrate specificity profiles of Gox2272 and Cc2746 against dipeptide libraries and individual dipeptide substrates. The pseudodipeptides **4**-**9** (**Scheme 3.2**) are inhibitors of these two

enzymes. The structures of the Gox2272·**9** complex and Cc2746 with either **4**, **5** or **6** bound, were desired but not obtained during the course of this work. It was the hope that the insertion region would be completely defined in a structure of the complex of Cc2746 with an inhibitor. In Sco3058, Leu-272 is part of the same insertion region and is at the perimeter of the 10 Å radius from the C-terminal C_α of **4**; it is possible that the residues of undefined structure in Cc2746 are participating in the formation of the substrate-binding pocket. The structural overlays with other proteins within cog2355 were compared and amino acid side chains were modeled into Gox2272 and Cc2746 from the position of the inhibitor in the Sco3058·**4** complex.

MATERIALS AND METHODS

Cloning of Gox2272. The Gox2272 gene (gi|58040697) was amplified from the *Gluconobacter oxydans* genomic DNA with the primers 5'-GCAGGAGCCATATGACT GATCTGCACGACACGCTCCTGACGC-3' and 5'-CGCGGAATTCAAACCTTCTCT GCAGCCCGCTCGGCCGTCTCC-3'. The *G. oxydans* genomic DNA was kindly provided by Dr. Armin Ehrenreich from Georg-August University, Göttingen, Germany. The *Pfx* platinum polymerase (Invitrogen), 0.05 units/μL PCR reaction, was used according to the manufacturer's instructions. The PCR products were digested with *Nde*I and *EcoR*I and ligated (separately) into the pET30a(+) vector (Novagen) with T4 ligase (New England Biolabs).

Overexpression of Zn-Gox2272. E. coli BL21 cells were electrotransformed with the pET30a(+)-Gox2272 plasmid. *E. coli* cells harboring the pET30a(+)-*Gox2272* plasmid were grown at 30 ºC in Terrific Broth (TB) containing 50 μg/mL kanamycin.

When the optical density at 600 nm reached ~0.5, 2.0 mM $Zn(acetate)_2$ was added and the overexpression of Gox2272 was induced with 250 μM isopropyl-β-thiogalactoside (IPTG). Following the addition of IPTG, the cells were grown overnight at room temperature (21 ºC).

Purification of Gox2272. *E. coli* BL21-Gox2272 cells were suspended in the purification buffer (50 mM Hepes/100 μg/mL phenylmethanesulfonylfluoride, pH 7.5) and lysed via sonication at 0 ºC. The cell lysate was obtained by centrifugation (~ 28,000*g*) and treated with a protamine sulfate solution. The protamine sulfate solution was equal in volume to the cell lysate and contained a mass of protamine sulfate, equal to 2% of the starting mass of cells. The soluble portion of the cell lysate/protamine sulfate mixture was separated by centrifugation (~17,000*g*). The impurities that precipitated between zero and 60% ammonium sulfate (371 g/L at 0 ºC) were removed by centrifugation (17,000*g*). The Gox2272 protein was precipitated in the 60-75 % ammonium sulfate fraction. The precipitated proteins were removed by centrifugation, resuspended in a minimal amount of ice cold purification buffer, and eluted through a 1 L AcA34 gel filtration column at a flow rate of 1 mL/min. The fractions containing the protein of expected size, as assessed by SDS-PAGE, were combined. The purified protein was submitted to the Protein Chemistry Laboratory at Texas A&M University Department of Biochemistry for amino acid sequencing.

Purification of Mn-Gox2272. Mn-Goxx2272 was obtained by the same overexpression and purification methods as the Zn-Gox2272 with a few exceptions. During the growth of the BL21-Gox2272 cells, 200 μM bipyridyl was added when the

optical density at 600 nm was ~ 0.1-0.2. Manganese chloride was added to a final concentration of 1.0 mM instead of $Zn(acetate)_2$. During the purification, 100 μM $MnCl_2$ was added to the buffer through the resuspension of the precipitated Mn-Gox2272 with 75% saturation of $(NH_4)_2SO_4$. The metal content of purified Zn-Gox2272 and Mn-Gox2272 were determined by ICP-MS.

Overexpression of Cc2746. *E. coli* Rosetta-2 (DE3) cells (Novagen) were electrotransformed with the plasmid containing the Cc2746 gene (gi|16126978). *E. coli* cells harboring the plasmid with the Cc2746 gene were grown in a 5 mL LB/kanamycin culture at 37 °C for 8 hours and used to innoculate Terrific Broth containing 50 μg/mL kanamycin and 25 μg/mL chloramphenicol. The cells were grown in TB/kanamycin/chloramphenicol at 30 °C. When the optical density at 600 nm reached ~0.5, 1.0 mM $Zn(acetate)_2$ was added and the overexpression of Cc2746 was induced with and without 10, 20, 50, 100 μM isopropyl-β-thiogalactoside (IPTG). Following induction, the cells were grown at 20-22 °C overnight. The proteins were separated by SDS-PAGE, and it was concluded that the Cc2746 gene did not overexpress. In a similar manner, overexpression of Cc2746 was attempted in Rosetta 2 (DE3) cells with 10, 20, 20, 50, 100 and 500 μM IPTG. Following the addition of IPTG, the cells were grown overnight at ambient temperature (21 ºC). Low levels of overexpression were obtained at 20 μM IPTG, but overexpression was less at 10, 50, 100 and 500 μM IPTG; this phenomenon was observed twice. In preparation of a larger cell stock for the isolation of the protein, at the time of induction, the medium was spiked with an additional 0.4 mL each of 25 mg/mL chloramphenicol and 50 mg/mL kanamycin per

liter of culture, and 2 mM $Zn(acetate)_2$ was added (2 mL of a sterile 1.0 M Zn acetate solution).

Purification of His-tagged Cc2746 for Assays. The Rosetta 2 (DE3)-Cc2746 cells were lysed at 0 °C, via sonication, in 20 mM Tris/5 mM imidazole/0.5 M NaCl/100 μg/mL PMSF, pH 7.9. The ratio of sonication buffer to cell paste was ~ 5 mL per gram of cells. The cell lysate was harvested by centrifugation at ~ 28,000 *g*. A 25-mL Ni-Sepharose 6 Fast Flow column (Amersham Biosciences) was generated according to the manufacturer's instructions and equilibrated with the binding buffer. The sterile-filtered cell lysate (0.45 μm filter) was loaded onto the equilibrated Ni-column and eluted by gravity. The column was washed with the binding buffer (the sonication buffer without PMSF) until the OD_{280nm} of the flowthrough was less than 0.05 (~ 15 column volumes). The undesired proteins were eluted with eight column volumes of 20 mM Tris/0.5 M NaCl/100 mM imidazole, pH 7.9. The pure Cc2746 protein was eluted with ~ four column volumes of 20 mM Tris/0.5 M NaCl/200 mM imidazole, pH 7.9. The fractions containing Cc2746 were concentrated using a Centriprep ultracel YM-10 membrane (Amicon) and the imidazole was removed via gel filtration on a Superdex 200 High-load 26/60 column (Amersham Biosciences).

Synthesis of Dipeptide Libraries. The L-Xaa-D-Xaa dipeptide libraries were synthesized as described in Chapter II. The L-Xaa-L-Xaa and D-Xaa-L-Xaa libraries were synthesized as described in Chapter III.

β-Lactamase Assay. Gox2272 and Cc2746 were tested for their ability to hydrolyze the *β*-lactam nitrocefin. The hydrolysis of nitrocefin was monitored by the

increase in absorbance at 486 nm (ε = 20,500 $M^{-1}cm^{-1}$) in the presence of 100 mM potassium phosphate, pH 7.0. The values of k_{cat}, K_m and k_{cat}/K_m of Gox2272 or an estimation of the k_{cat}/K_m for Cc2746 were determined with the nitrocefin substrate by fitting the initial velocity data to equation 2.

Ninhydrin-based Enzyme Assay. Unless otherwise stated, assays were conducted at 30 °C in 50 mM Hepes, pH 7.5. The ninhydrin method was used to measure the concentration of free amino acid products at 507 nm as previously described for Sco3058 with dipeptides in Chapter III (*61*). Mn-Gox2272 was assayed with 1.0 mM L-Asn-D-Glu in the absence of additional metal in the assay and in the presence of 0.01, 0.1, 1, 5, 10, 50 and 100 μM $MnCl_2$. Other assays conducted with Mn-Gox2272 were supplemented with 20 μM $MnCl_2$.

N-terminal Substrate Specificity of Gox2272 and Cc2746. Enzyme course studies were performed by incubating the enzymes with nineteen L-Xaa-L-Xaa (0-2 μM Zn-Gox2272 for 8 hours or 0-2 μM Cc2746 for 5 hours), nineteen L-Xaa-D-Xaa (0-2 μM Zn-Gox2272 for 8 hours or 0-1.5 μM Cc2746 for 2.5 hours) and seventeen D-Xaa-L-Xaa (0-5 μM Mn-Gox2272 for 4 hours or 0-2 μM Cc2746 for 16 hours) libraries. The D-Xaa-L-Xaa screens with Mn-Gox2272 were supplemented with 20 μM $MnCl_2$. The free amino acids were measured using the ninhydrin assay as described previously. For the L-Xaa-L-Xaa and L-Xaa-D-Xaa libraries, the absorbance at 507 nm (A_{507nm}), *Q*, was plotted as a function of enzyme concentration, E_t, and fit to equation 1. The rate constants, *k* (nM^{-1}), were compared for each N-terminal amino acid within each set of libraries with the same incubation time. The hydrolysis of the D-Thr-L-Xaa and D-Val-

L-Xaa libraries by Cc2746 was too slow to fit to equation 1. The rates of hydrolysis of each of the D-Xaa-L-Xaa libraries by Mn-Gox2272 were obtained from the slope of linear relationship between the A_{507nm} and the Mn-Gox2272 concentration.

C-terminal Substrate Specificity of Gox2272. Enzyme-course assays for two dipeptide libraries, L-Asn-L-Xaa and L-Asn-D-Xaa, were analyzed by HPLC for C-terminal specificity of Zn-Gox2272. Each library was incubated at 30 ºC for 5 hours, with and without enzyme, in 25 mM $(NH_4)HCO_3$, pH 8.0. The concentrations of enzyme used were 0-2 μM (L-Asn-L-Xaa) and 0-1 μM (L-Asn-D-Xaa). The sample preparation and data analysis were performed in the same manner as described for Gox1177 and Bb3285 (Chapter II).

C-terminal substrate specificity of Cc2746. Enzyme course assays for three dipeptide libraries, L-Arg-L-Xaa and L-Arg-D-Xaa, and D-Phe-L-Xaa were analyzed by HPLC for C-terminal specificity of Cc2746. Each library was incubated at 30 ºC for 5 hours, with and without enzyme, in 25 mM $(NH_4)HCO_3$, pH 8.0. The concentrations of enzyme and incubation times used were 0-0.75 μM (L-Arg-L-Xaa, 2.5 hours), 0-0.5 μM (L-Arg-D-Xaa, 2 hours) and 0-2.5 mM (D-Phe-L-Xaa, 16 hours). The sample preparation and data analysis were performed in the same manner as described for Gox1177 and Bb3285 (Chapter II).

Kinetic and Inhibition Constants of Cc2746 and Mn-Gox2272. The kinetic constants K_m, k_{cat} and k_{cat}/K_m were assessed for the two dipeptidases with various concentrations of several dipeptide substrates. Mn-Gox2272 assays contained 20 μM $MnCl_2$. The kinetic constants were determined by a fit of the data to equation 2 where *v*

is the velocity of the reaction, E_t is the total enzyme concentration, k_{cat} is the turnover number, A is the substrate concentration, and K_a is the Michaelis constant.

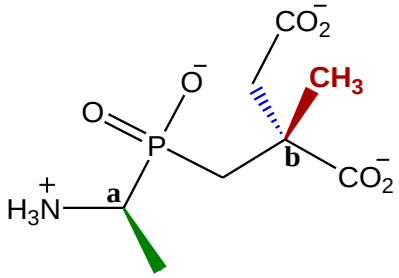
For determination of the inhibition constants of the pseudodipeptide inhibitors (**Scheme 3.2**) with Cc2746, 0.5 mM L-Leu-D-Glu (compounds **4-6** and **9**) or 0.4 mM L-Met-D-Leu (compounds **7** and **8**) substrates were used, and the assays were conducted at room temperature (25 °C). Cc2746 was assayed in the absence of inhibitor and up to 100 μM (compounds **4-8**) or 200 μM (compound **9**) inhibitor in the assay. The activity of Mn-Gox2272 (in 20 μM $MnCl_2$, substrate = 0.5 mM L-Met-D-Glu) was measured at 30 °C without inhibitor and with up to 75 μM of compounds **4-9**. The inhibition data for both enzymes were plotted in a similar fashion to **Figure 3.4**, and the K_i values were determined from fits of the data to equation 3.

Superposition of Multiple Dipeptidase Structures. The superposition of structures was performed using the MatchMaker tool in Chimera. Initial matches were made between two sets of structures. Residues in the regions with more variation in the position of the backbone were deleted and rematched. Trial and error was used, deleting sections which 1) were not present in one or the other proteins or 2) showed more variation in the backbone, and rematching the structures. The quality of the superposition of two structures was gauged by the position of the histidine ligands, the arginine at the end of β_6 (Gox2272-R222), and the histidine/tryptophan following β_4 (Gox2272-H142). In this fashion, the pairs of structures Gox2272/Cc2746, Rsp0802·**7**/Sco3058·**4** complexes, Cc2746/Pa2393, Sco3058·**4** complex/Cc2746, Rsp0802·**7**/LmoDP·**8** complexes, and Sco3058·**4**/hRDP·cilastatin were overlayed. The

pairs of overlays were used to create a master multiple overlay of the structures of these seven proteins within cog2355.

The pseudodipeptide **4** bound in the Sco3058 structure served as a starter molecule from which to model different side chains from the carbon atoms in **4** that correspond to the N- (carbon "a" in **Scheme 4.2**) and C-terminal C_α carbons of a dipeptide. A methyl group was built onto **4** in place of the hydrogen on the chiral carbon representing C_α at the C-terminus (carbon "b" in **Scheme 4.2**) of a dipeptide. **Scheme 4.1** represents the structure of **4** with this methyl group (red) added.

Scheme 4.2: Structure of phosphinate inhibitor **4** with an added methyl group (*red*) for modeling of the position of a C-terminal L-amino acid side chain.



Free amino acid molecules were introduced into the structure. To model an L-amino acid side chain at the N-terminus, the C_α–C_β bond of the inserted amino acid was superimposed onto the bond colored green in **Scheme 4.2**. Positions for different C-terminal amino acids on **4** were fashioned by manually overlaying the C_α–C_β bond of the inserted amino acid onto the bond colored red, for L-amino acids, or blue for D-amino

acids. The "Rotomers" editing tool in Chimera was used on the added amino acid to find potential contacts to the protein.

Structural Overlay of Cc2746 with Sco3058 and Pa2393. There is an insertion region between β_7 and α_7 of the $(\alpha,\beta)_8$-barrel in three of the eight proteins from cog2355 which have crystal structures. Sco3058·**4** was superimposed onto the truncated coordinates of Cc2746 where residues 282-334 of Cc2746 were deleted. Pa2393 was truncated, removing residues 101-115/298-365, and matched onto the same truncated Cc2746. The complete structures of Cc2746 and Pa2393 were then superimposed onto their respective truncated reference structures. This particular overlay was used to compare the position of the insertion region between β_7 and α_7 of the $(\alpha,\beta)_8$-barrel.

RESULTS

Cloning, Overexpression and Purification of Gox2272 and Cc2746. The gene for Gox2272 was successfully cloned into the pET30a(+) vector. The purified Zn-Gox2272 was found to contain 1.5 equivalents of Zn and 0.5 equivalents of Fe. The Mn-Gox2272 had 1.1 equivalents of Mn and 0.2 equivalents each of Zn and Fe. The purified Cc2746 had 1.3 equivalents of Zn and 0.4 equivalents of Ni.

N-terminal Specificity of Gox2272. The relative rates of hydrolysis of the L-Xaa-L-Xaa, L-Xaa-D-Xaa and D-Xaa-L-Xaa libraries are compared in **Figures 4.2A, C** and **E**. The enzyme hydrolyzed a broad spectrum of L-L, L-D and, to a lesser extent, D-L dipeptides. The L-Asn-D-Xaa library was hydrolyzed the fastest, with the L-Asn-D-Xaa library being preferred over the other L-Xaa-D-Xaa libraries. From the high absorbance values of the control it was apparent that the L-Gln-L-Xaa library had degraded. The

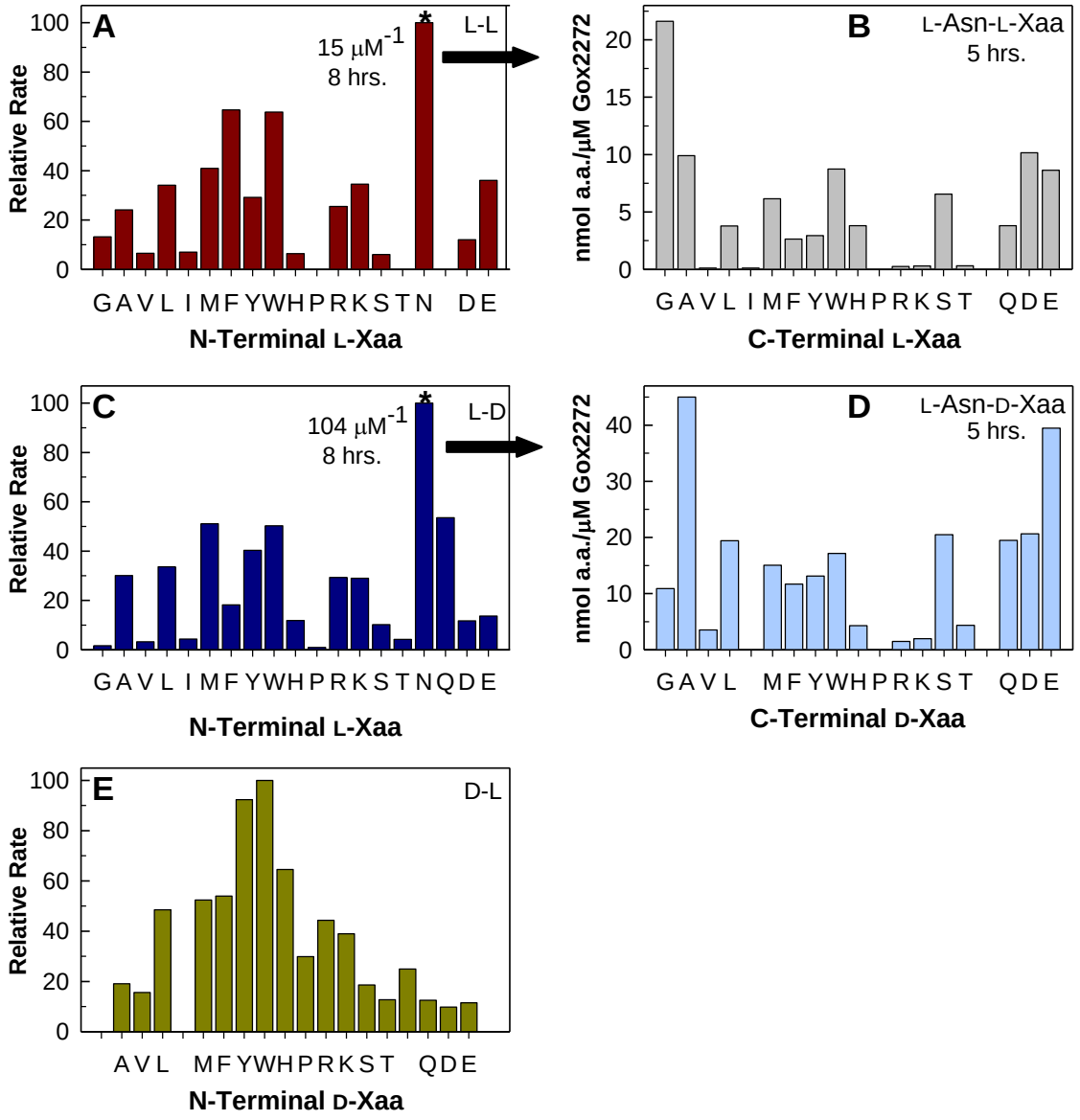


Figure 4.2: Substrate profile results for Zn-Gox2272. **(A**, **C** and **E)** Results of N-terminal specificity profile screens with 18 L-Xaa-L-Xaa (panel A, *red*), 19 L-Xaa-D-Xaa (panel C, *dark blue*) and 17 D-Xaa-L-Xaa (panel E, *olive green*) libraries. **(B, D** and **F)** Results of C-terminal specificity profile screens of three selected dipeptide libraries. The liberated C-terminal amino acids within the library marked with an asterisk on the left side panel (A and B) were quantitatively separated by HPLC to obtain the C-terminal specificity profile in the panel to its right (panels B and D). The value of *k* is given for the bar marked with an asterisk in panels A and C. The incubation times are presented for the comparison of the rates of hydrolysis of the L-Asn-L-Xaa/L-Asn-D-Xaa (panel A and C) libraries and the individual dipeptide components within these same two libraries (panels B and D). The C-terminal L-amino acids (*gray*) and D-amino acids (*light blue*) are shaded differently.

hydrolysis of L-Gln-D-Xaa was used to gauge the preference for L-Gln at the N-terminus in comparison to the other L-Xaa-D/L-Xaa libraries. Gox2272 was able to hydrolyze the D-Xaa-L-Xaa libraries to some extent, but the hydrolysis of these dipeptides was much slower. The L-Asn-L-Xaa and L-Asn-D-Xaa libraries were selected for HPLC analysis.

C-terminal Specificity of Gox2272. In preparation of the libraries for HPLC analysis, the L-Asn-L-Xaa and L-Asn-D-Xaa libraries were incubated with Gox2272 for the same amount of time so as to make the rates comparable. The rates of hydrolysis of the individual dipeptide components within these two libraries are presented **Figures 4.2B** and **4.2D**. Among the 34 substrates which can be compared within these two libraries, the best substrate was L-Asn-D-Ala, though it is only marginally preferred over L-Asn-D-Glu with a rate of hydrolysis within the standard deviation of the L-Asn-D-Ala value. Within the L-Asn-L-Xaa library, glycine was preferred over any of the L-amino acids at the C-terminus.

Kinetic and Inhibition Constants of Gox2272. The specific activities of Mn-Gox2272 are plotted in **Figure 4.3** for various concentrations of $MnCl_2$ in the assay. The kinetic constants for Mn-Gox2272 with several dipeptide substrates and nitrocefin are summarized in **Table 4.1**. The highest catalytic efficiency was observed with both the L-Asn-D-Glu and L-Met-D-Glu substrates (k_{cat}/K_m = 4.7 x 10^5 $M^{-1}s^{-1}$). The top two inhibitors (**Scheme 3.2**) were compounds **4**, the dipeptide mimic of Ala-Asp (K_i = 112 ± 9.6 nM), and **8**, the phosphinate mimic of Leu-Ala (K_i = 270 ± 6 nM). Mn-Gox2272 was also inhibited by compounds **5** (K_i = 300 ± 31 nM), **6** (K_i = 550 ± 44 nM), **7** (K_i = 560 ± 4 nM) and **9** (K_i = 350 ± 18 nM).

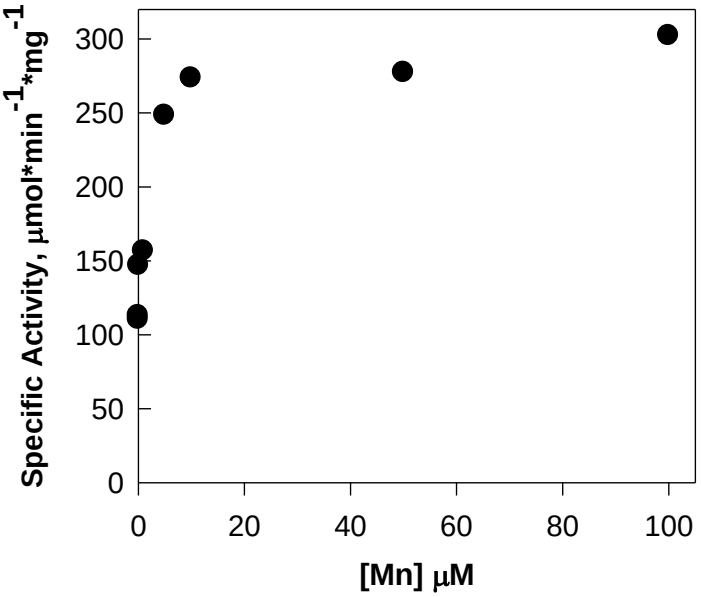


Figure 4.3: Activity of Mn-Gox2272 at various concentrations of $MnCl_2$ in the assay. The substrate was 1.0 mM L-Asn-D-Glu.

Table 4.1: Kinetic constants for select substrates of Mn-Gox2272[a]

Substrate	k_{cat}, s^{-1}	K_m, mM	k_{cat}/K_m $M^{-1}s^{-1}$
L-Asn-D-Glu	160 ± 5	0.34 ± 0.03	(4.7 ± 0.4) x 10^5
L-Met-D-Glu	270 ± 3.5	0.58 ± .03	(4.7 ± 0.2) x 10^5
L-Met-D-Leu	280 ± 7.5	2.3 ± 0.2	(1.2 ± 0.09) x 10^5
L-Ala-D-Ala	480 ± 13	4.2 ± 0.2	(1.2 ± 0.07) x 10^5
L-Arg-D-Asp	280 ± 4.2	2.4 ± 0.09	(1.2 ± 0.05) x 10^5
L-Leu-D-Ala	240 ± 5.3	2.3 ± 0.1	(1.0 ± 0.06) x 10^5
L-Leu-D-Ser	140 ± 3.9	1.5 ± 0.1	(9.5 ± 0.8) x 10^4
L-Leu-D-Glu	69 ± 1.5	0.87 ± 0.06	(7.9 ± 0.6) x 10^4
L-Arg-Gly	380 ± 52	8.6 ± 1.9	(4.4 ± 1.1) x 10^4
L-Tyr-D-Leu	140 ± 1.4	5.2 ± 0.1	(2.6 ± 0.06) x 10^4
L-Leu-D-Leu	160 ± 12	6.3 ± 0.9	(2.5 ± 0.4) x 10^4
L-Arg-L-Asp	21 ± 0.7	1.2 ± 0.1	(1.8 ± 0.2) x 10^4
L-Ala-L-Asp	27 ± 0.9	2.0 ± 0.2	(1.3 ± 0.1) x 10^4
L-Ala-L-Glu	6.6 ± 0.3	0.5 ± 09	(1.3 ± 0.2) x 10^4
L-Ala-L-Gln			(6.8 ± 0.1) x 10^3
Nitrocefin	(2.0 ± 0.2) x 10^{-3}	0.12 ± 0.004	(1.7 ± 0.07) x 10^1

[a]From fits of the data to equation 2. For those entries without values of k_{cat} and K_m, saturation was not achieved at concentrations up to 6 mM.

N-terminal Specificity of Cc2746. The rates obtained from the enzyme course data for the hydrolysis of the 54 dipeptide libraries by Cc2746 are summarized in **Figures 4.4A, C** and **E**. Cc2746 has a preference for L-Asn-, L-Met- and L-Arg-Xaa over the other L-L dipeptide libraries, and L-D dipeptides were hydrolyzed more rapidly than their L-L counterparts. The L-Arg-D-Xaa library was hydrolyzed at least 2.5-fold faster the other L-D libraries and Gly-D-Xaa. Cc2746 hydrolyzed D-Xaa-L-Xaa substrates much more slowly. Dipeptides with a D-aromatic residue were favored over other D-L dipeptides.

C-terminal Specificity of Cc2746. The slopes from the linear correlation between nmol of product and enzyme concentration were used to compare the rates of hydrolysis of the individual library components; these rates are summarized for L-Arg-L-Xaa, L-Arg-D-Xaa and D-Phe-L-Xaa in **Figures 4.4B , D** and **F**, respectively. The top six substrates within these three libraries were in the order of L-Arg-Gly > L-Arg-D-Glu > L-Arg-D-Leu ≈ L-Arg-L-Glu ≈ L-Arg-L-Asp > L-Arg-D-Ala. Within the L-Arg-L-Xaa library, glycine was preferred over any of the L-amino acids at the C-terminus.

Kinetic and Inhibition Constants of Cc2746. The kinetic constants for Cc2746 with selected dipeptide substrates and nitrocefin are summarized in **Table 4.2**. The highest catalytic efficiency was observed with the L-Met-D-Leu substrate (k_{cat}/K_m = 4.6 x 10^5 $M^{-1}s^{-1}$), but the enzyme hydrolyzed several dipeptides with comparable activity. The top two inhibitors (**Scheme 3.2**) were compounds **5**, the dipeptide mimic of Phe-Asp (K_i = 570 ± 43 nM), and **8**, the phosphinate mimic of Leu-Ala (K_i = 710 ± 78 nM). Cc2746

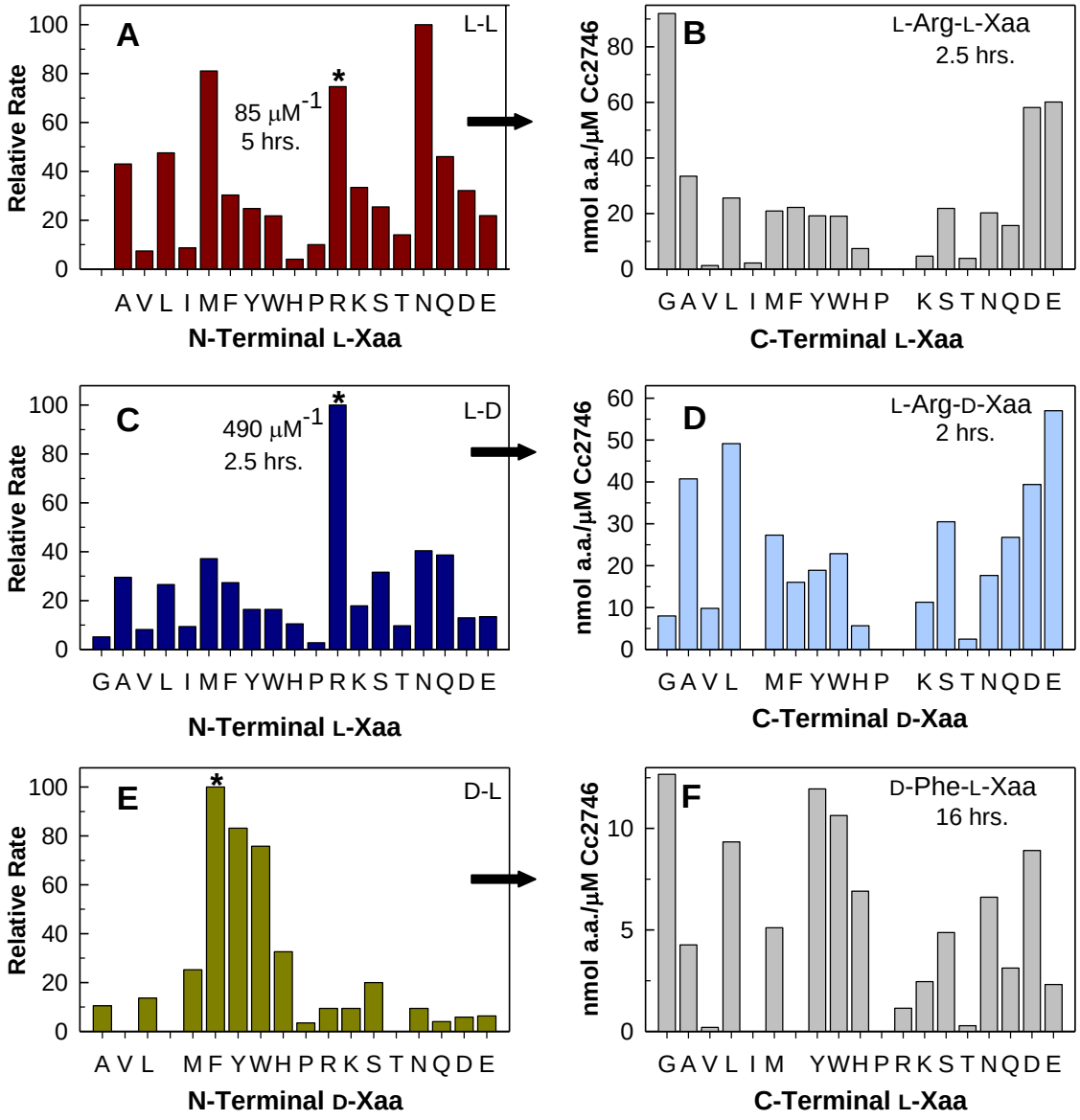


Figure 4.4: Substrate profile results for Cc2746. **(A**, **C** and **E)** Results of N-terminal specificity profile screens with 18 L-Xaa-L-Xaa (panel A, *red*), 19 L-Xaa-D-Xaa (panel C, *dark blue*) and 17 D-Xaa-L-Xaa (panel E, *olive green*) libraries. **(B, D** and **F)** Results of C-terminal specificity profile screens of three selected dipeptide libraries. The liberated C-terminal amino acids within the library marked with an asterisk on the left side panel (A, B and C) were quantitatively separated by HPLC to obtain the C-terminal specificity profile in the panel to its right (panels B and D and F). The value of *k* is given for the bar marked with an asterisk in panels A, C and E. The incubation times are presented for the comparison of the rates of hydrolysis of the L-Arg-L-Xaa/L-Arg-D-Xaa (panel A and C) libraries and the individual dipeptide components of three dipeptide libraries (panels B, D and F). The C-terminal L-amino acids (*gray*) and D-amino acids (*light blue*) are shaded differently.

Table 4.2: Kinetic constants for select substrates of Cc2746[a]

Substrate	k_{cat}, s^{-1}	K_m, mM	k_{cat}/K_m $M^{-1}s^{-1}$
L-Met-D-Leu	82 ± 2	0.18 ± 0.02	(4.6 ± 0.5) x 10^5
L-Asn-D-Glu	43 ± 1	0.15 ± 0.02	(3.0 ± 0.4) x 10^5
L-Arg-D-Asp	31 ± 1	0.13 ± 0.01	(2.4 ± 0.3) x 10^5
L-Met-D-Glu	68 ± 3	0.31 ± 0.04	(2.2 ± 0.3) x 10^5
L-Leu-D-Ala	28 ± 0.5	0.16 ± 0.01	(1.8 ± 0.1) x 10^5
L-Ala-D-Ala	132 ± 7	0.81 ± 0.1	(1.6 ± 0.2) x 10^5
L-Met-L-Leu	73 ± 3	0.49 ± 0.06	(1.5 ± 0.2) x 10^5
L-Leu-D-Ser	51 ± 2	0.33 ± 0.04	(1.5 ± 0.2) x 10^5
L-Arg-Gly	50 ± 2	0.36 ± 0.04	(1.4 ± 0.1) x 10^5
L-Leu-D-Leu	41 ± 2	0.4 ± 0.05	(1.0 ± 0.1) x 10^5
L-Leu-D-Glu	18 ± 1	0.18 ± 0.03	(1.0 ± 0.1) x 10^5
L-Ala-L-Ala	140 ± 6	1.7 ± 0.2	(8.1 ± 1) x 10^4
L-Tyr-D-Leu	18 ± 0.4	0.32 ± 0.03	(5.7 ± 0.6) x 10^4
L-Ala-L-Asp	85 ± 1	2.9 ± 0.01	(3.0 ± 0.1) x 10^4
L-Ala-L-Glu	34 ± 0.8	1.3 ± 0.1	(2.5 ± 0.2) x 10^4
D-Ala-D-Ala			(6.4 ± 0.1) x 10^1
D-Ala-L-Ala			(1.5 ± 0.2) x 10^1
Nitrocefin			< 0.1

[a]From fits of the data to equation 2. For those entries for dipeptides without values of k_{cat} and K_m, saturation was not achieved at concentrations up to 8 mM.

was also inhibited by compounds **4** (K_i = 1.8 ± 0.1 μM), **6** (K_i = 2.0 ± 0.1 μM), **7** (K_i = 1.0 ± 0.1 μM) and **9** (K_i = 20 ± 2.1 μM).

Structure of Zn-Gox2272. The structure of the Zn-Gox2272·SO_4 complex was solved to 1.8 Å resolution by collaborators at the Albert Einstein College of Medicine. The binuclear Zn center of Gox2272 is shown in **Figure 4.5**. The two metals are bridged by water (2.0-2.1 Å) and Glu-114 (2.0-2.2 Å) at the end of β-strand 3. The tetrahedral coordination of Zn_α is completed by contacts to His-14 (2.1 Å) and Asp-16 (2.1 Å) from β_1, whereas Zn_β has a trigonal bipyramidal geometry with additional contacts to a sulfate oxygen (2.3 Å), His-190 (2.1 Å) from β_5 and His-211 (2.1 Å) from β_6. The sulfate ion which coordinates Zn_β is 3.0 Å from N^ω of Arg-222, and a second sulfate oxygen is 2.8 Å from N^ε of His-142. The catalytic aspartate from β_8 is 2.8 Å from the bridging water.

Surface and stick representations of Gox2272 (with **4** positioned in the active site from the superposition of the structure onto the Sco3058·**4** complex) are pictured alongside Sco3058 in **Figure 4.6B**. The *side chains* of residues (C_α for glycine) whose side chains are within 10 Å of chiral carbon "a" (**Scheme 4.2**) *and* closer to carbon "a" than carbon "b" are colored blue. In the same fashion, the side chains within 10 Å of carbon "b" are colored purple. The N-terminal half of a dipeptide is considered to be represented by the aminoethylphosphonyl moiety of **4**, and the rest of the inhibitor molecule symbolizes the C-terminal half. The residues from Gox2272 anticipated to be in the immediate vicinity of the N-terminal half of the substrate are Asp-16, Tyr-52, His-

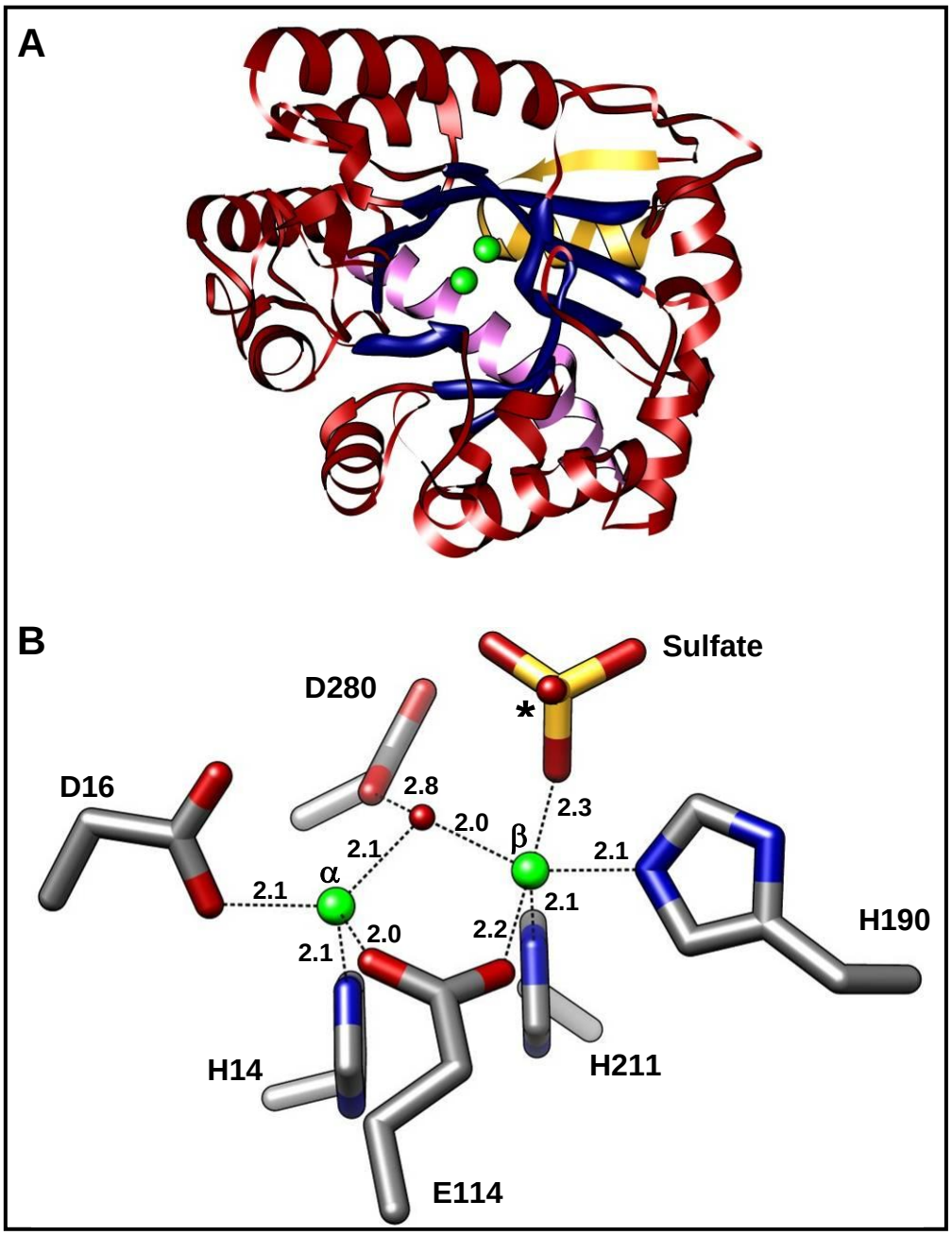


Figure 4.5: Ribbon representation and metal center of Zn-Gox2272. **(A)** Ribbon representation of ZnGox2272. The $(\beta/\alpha)_8$-barrel is shown in *blue* and *red*. The two Zn ions (*green* spheres), C-terminal α-helix (*pink*) and the insertion region between α_2 and β_3 (*yellow*) are also colored. **(B)** Metal center of Zn-Gox2272 with a tetrahedral Zn_α and pentacoordinate Zn_β Zinc ion, a bridging water (*red* sphere) and ligating sulfate (*yellow* sulfur). Oxygens and nitrogens are colored *red* and *blue*, respectively. Distances between atoms are shown in Å. Though not shown, the sulfate oxygen noted with an asterisk is 2.8 Å from His-142 from β-strand 4, and the oxygen which ligates to Zn_β is also 3.0 Å from Arg-222.

142 and Gly-283. The nearest residues to the C-terminal half of the substrate are Asn-143, Arg-222, Met-244, Ser-246, Phe-247, Asp-280 and Gly-284.

Structure of Cc2746. The three dimensional structure of the homotetrameric Cc2746 enzyme (**Figure 4.7A**) has been deposited into the PDB with a resolution of 2.0 Å. A surface representation of one subunit is shown in **Figure 4.7B**. The metal center (**Figure 4.7C**) contains two Zn ions. Zn_{α} is coordinated to His-53, $O^{\delta 1}$ and $O^{\delta 2}$ from Asp-55, and an ordered water. A chloride ion, His-229 and His-250 are unique ligands to Zn_{β}. The two metals are bridged by Glu-159. A bridging water between the two metals was not observed. On the basis of the superposition of Cc2746 onto the Sco3058·**4** complex, the ordered water coordinated to Zn_{α} is analogous to the free amino group of **4**. The chloride ion is located between the position of the two carboxylate oxygens of **4** that indicate the position of the α-carboxylate of a dipeptide.

The surface and active site stick representations of Cc2746 are shown in **Figure 4.6C** and were created for comparison to Sco3058 by the same method as Gox2272 in **Figure 4.6B**. The difference is that Cc2746 has an insertion region, colored orange, between β_7 and β_8 of the $(\beta/\alpha)_8$-barrel. The active site residues in Cc2746 expected to influence the quality of the binding pocket for the N-terminal half of the substrate are Asp-55, Pro-57, Ala-58, Tyr-98, His-186 and C_{α} from Gly-363. The residues presumed to be closest to the C-terminal half of the substrate are Phe-187, Arg-261, I-283, Tyr-284, Asp-360 and Gly-264. Though not shown in **Figure 4.6**, the backbone oxygen from the first glycine following β_8 (Sco3058-G323, Gox2272-G283, Cc2746-G363) is in proximity (2.4-3.3 Å) to the methylene group adjacent to the phosphorus of **4** that

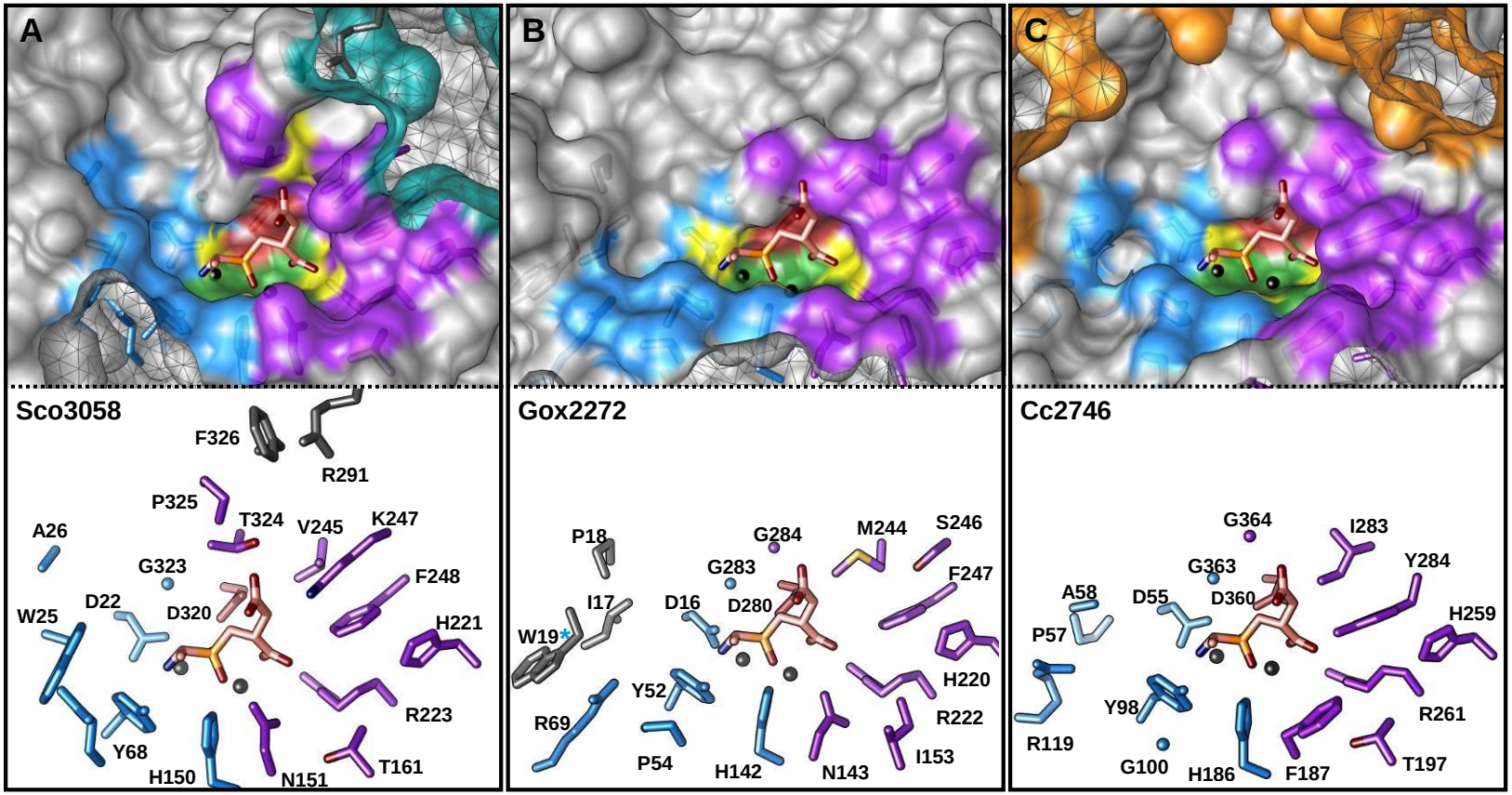


Figure 4.6: Surface and stick representations of the active sites of Sco3058·**4 (A)**, Gox2272 **(B)** and Cc2746 **(C)** in the same orientation. The placement of inhibitor **4** (*pink* carbons, *orange* phosphorus) in B and C is from the superposition of Gox2272 and Cc2746 onto the Sco3058·**4** complex. Side chains of residues which make up the pockets for the N- and C-terminal sides of a dipeptide are colored *blue* and *purple*, respectively. The aspartate from β_8 is colored *red*. Surfaces of the histidine/glutamate ligands, atoms within hydrogen-bonding distance to **4** (*yellow*), the insertion region (*teal* or *orange*), and the Zn ions (*black* spheres) are also colored. The clipped plain is depicted as a transparent mesh surface. The residues of each active site are labeled below the surface representation. Other residues discussed in the text are colored *gray*.

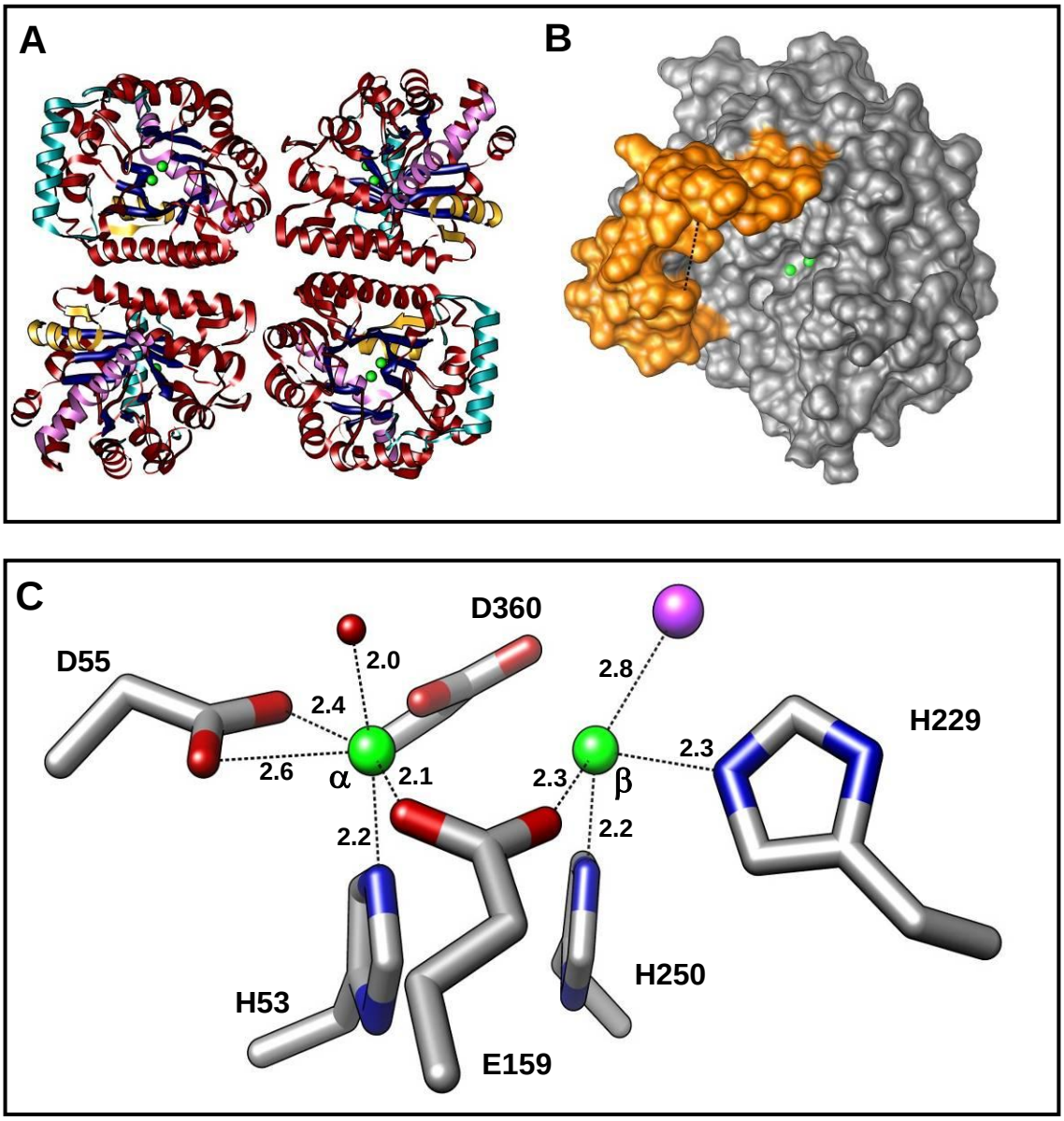


Figure 4.7: Representations of the metal center, tertiary and quaternary structure of Cc2746. Zinc ions are illustrated as *green* spheres. **(A)** Ribbon representation of tetrameric Cc2746. The $(\beta/\alpha)_8$-barrel is shown in *blue* and *red*. The insertion region (residues 285-334) is colored *teal*. The C-terminal α-helix (*pink*) and the insertion region between α_2 and β_3 (*yellow*) are also colored. **(B)** Surface representation of one subunit of Cc2746 where the insertion region (residues 285-334) is colored *orange* with *black* dashed lines depicting the missing segment of residues 294-303. **(C)** Metal center of Cc2746 with α and β Zinc ions, an additional water (*red* sphere) and chloride ion (*purple* sphere) ligands. Oxygens and nitrogens are colored *red* and *blue*, respectively. Distances between atoms are shown in Å.

symbolizes the nitrogen of the amide bond to be cleaved.

DISCUSSION

The dipeptidases Cc2746 and Gox2272 were successfully overexpressed and purified from *E. coli*. The Zn-Gox2272 protein contained a significant amount of iron, despite the extreme excess of Zn in the growth medium. A more active Mn-Gox2272 enzyme was purified by chelating Fe from the medium with bipyridyl and supplementing the culture with $MnCl_2$. The substrate profiles of the two enzymes were established by screening the proteins for activity against 54 dipeptide libraries. Both enzymes hydrolyzed a significant fraction of the dipeptides in the libraries, but had a marked preference for L-D dipeptides.

Substrate Specificity Determinants of Gox2272. Gox2272 has a moderate preference for L-Asn-Xaa dipeptides and D-Ala or D-Glu are favored over other amino acids at the C-terminus. On the basis of a model of D-Glu as the C-terminal amino acid of a dipeptide, the side chain oxygens of D-Glu were near the hydroxyl group of Ser-246 (3.3 Å) or the side chain amine of Asn-142 (3.2 Å). If the dipeptide were L-Gln-Xaa or L-Asn, the side chain heteroatoms were 2.3-2.9 Å or 2.4 Å from the phenolic oxygen of Tyr-52, respectively.

Substrate Specificity Determinants of Cc2746. For dipeptides with an L-amino acid at the C-terminus, Cc2746 has a slight preference for L-Asn-L-Xaa dipeptides, followed closely by L-Met-L-Xaa and L-Arg-L-Xaa. The enzyme had a more pronounced preference for L-Arg-D-Xaa dipeptides. Even though the Arg-Gly substrate within the L-Arg-L-Xaa library was hydrolyzed the fastest, it should be noted that Arg-

Gly was in competition with poorer L-Arg-L-Xaa substrates; the rate of Arg-Gly hydrolysis was significantly less in the presence of more favorable L-Arg-D-Xaa dipeptides. The same concept applies to differences in the rates of hydrolysis of L-Asn-Gly in the L-Asn-L-Xaa and L-Asn-D-Xaa libraries by Gox2272. When the C_α-C_β bond of L-Arg was modeled onto the C–C bond of the aminoethyl moiety of **4**, a rotomer of the L-Arg side chain was within hydrogen-bonding distance to the hydroxyl group of Tyr-98 (2.8-3.0 Å). The side chain oxygens of a modeled C-terminal D-Glu or L-Asp were 2.9 Å from the phenolic oxygen of Tyr-284.

Comparison of Insertion Regions of Cc2746, Sco3058 and Pa2393. Pa2393 is a protein within cog2355 but is missing the His and Asp ligands at the end of β_1, the histidine ligand at the end of β_5 as well as the bridging glutamate from β_3. The structure of Pa2393, with no metal bound in the protein, is available from the PDB (PDB code *3B40*). Pa2393 is similar in structure to other proteins in cog2355, has the histidine ligand from β_6, the aromatic residues following β_2 and β_7, as well as the conserved arginine which follows β_6. As shown in **Figure 4.8**, the position of this insertion with respect to the $(\beta/\alpha)_8$-barrel varies between the structures of Cc2746, Sco3058 and Pa2393. Residues 294-303 of Cc2746 are within this insertion region but are not defined in the crystal structure. Therefore, whether or not any of these ten residues in Cc2746 contribute to the formation of the substrate binding pocket is not known.

β_1-Loop/helix. Seven of the eight members of cog2355, for which there are structures, were superimposed and select parts of the structure are presented as ribbons in **Figure 4.9**. In five of the eight protein structures within cog2355 (Rsp0802 and

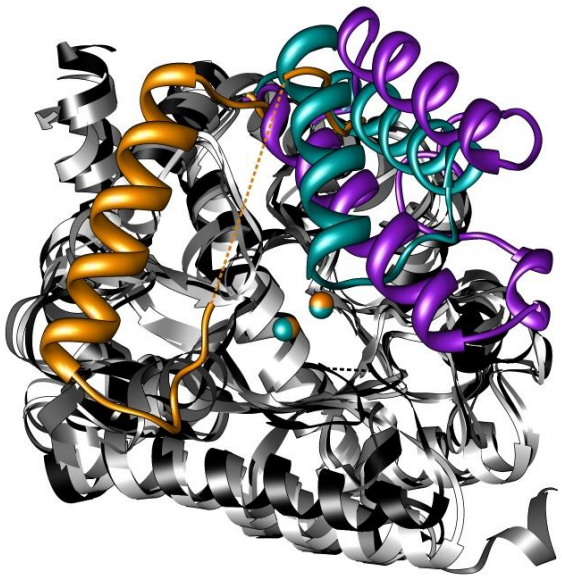


Figure 4.8: Ribbon depiction of three microbial proteins in cog2355, highlighting the different positions of the insertion region noted in the amino acid sequence alignment. Overlay of Cc2746 (*orange* and *light gray*, PDB code *2RAG*), Sco3058 (*teal* and *gray*, PDB code *3K5X*) and Pa2393 (*black* and *purple*, PDB code *3B40*). The missing segments of Cc2746 (residues 294-303) and Pa2393 (residues 106-110) are shown as *orange* and *black* dashed lines, respectively. Zn ions are represented as spheres.

LmoDP discussed Chapter V, Sco3058, hRDP, and Cc2746), there is a short α-helix (β_1-helix) and loop (β_1-loop) between the first β-strand and α-helix of the $(\beta/\alpha)_8$-barrel. The third residue after the HxD motif (A58 in Cc2746 and Trp-25 in Sco3058) is on the β_1-helix and faces into the active site; however, Gox2272 and Pa2393 only have a β_1-loop and there is not an equivalent active site residue. The number of residues between the HxD motif and a structural glutamine/histidine/aspartate on the β_1-loop is shorter for Gox2272 (16 residues) and Pa2393 (18 residues) compared to the others with the β_1-helix (20-26 residues). There is not a clear-cut explanation for the difference in secondary structure in Gox2272 and Pa2393 immediately following β_1.

Gox2272-W19 and Sco3058-W25 were thought to be equivalent to one another on the basis of the amino acid sequence alignment (**Figure 4.1**), but they are not in the same position in their respective structures (**Figure 4.6A** vs. **4.6B**). Only the C_β carbon from Gox2272-W19 (blue asterisk in **Figure 4.6B**) is within 10 Å of the chiral carbon of the aminoethyl moiety of **4**. The pocket surrounding carbon "a" (**Scheme 4.2**) is more open in Gox2272 and Cc2746 than in the Sco3058 structure, but for different reasons. In the case of Gox2272 the openness is due to W19 facing away from the active site, but the equivalent residue to Trp-25 in Sco3058 is a smaller Ala-58 residue in Cc2746.

β_8-Loop. The β_8-loop is the loop between β_8 and α_8 of the $(\beta/\alpha)_8$-barrel. The 3-4 residues immediately following β_8 are noted with asterisks in **Figure 4.1** and are colored differently for each protein in **Figure 4.9**. The position of the β_8-loop in Gox2272, Cc2746 and Pa2393 varies from that of other members of cog2355, namely, hRDP,

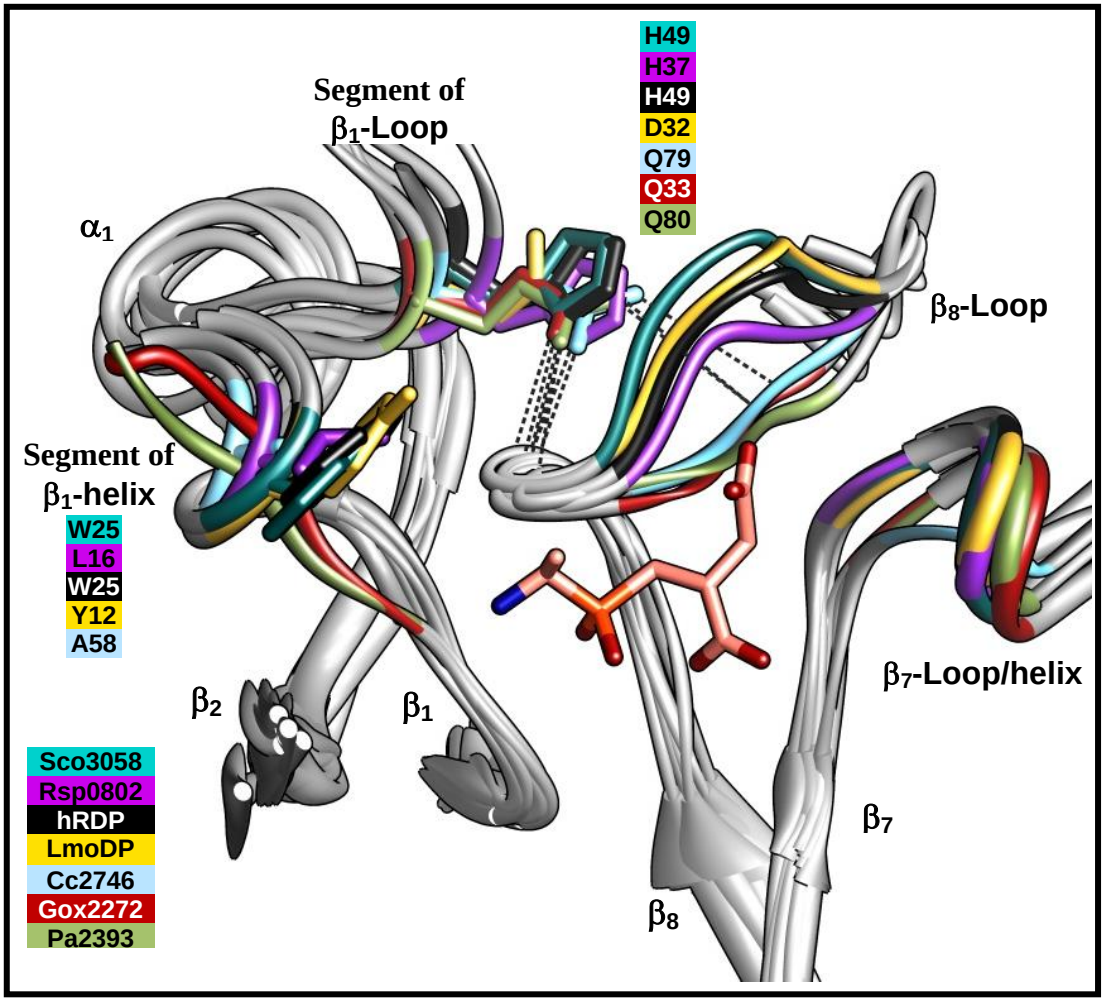


Figure 4.9: Segments of superimposed structures of seven proteins within cog2355. Ribbons are colored gray except for select residues from Sco3058 (*teal* W25, H49, V245, K247, F248, T324, P325, F326, T327), hRDP (*black* W25, H49, Y252, N254, Y255, V292, P293, R294, V295), Cc2746 (*light blue* A58, Q79, I283, Y284, G364, G365, G366), Gox2272 (*red* residues 16-20, Q33, M244, S246, F247, G284, G285, G286), Rsp0802 (*purple* L16, H37, A264, S266, F267, A308, T309, I310), LmoDP (*yellow* Y12, D32, T225, L227, F228, I265 P266, D267, H268), Pa2393 (*olive green* residues 61-64, Q80, F298, A300, Y301, G394, G395, G396) The phosphinate mimic from the Sco3058·**4** complex is shown with pink carbons and an *orange* phosphorus. The three residues which follow β_7 and contribute side chains to the binding pocket for the C-terminal amino acid side chain of a dipeptide are highlighted on the β_7-loop/helix. A major contributor to the binding pocket of the side chain of the N-terminal amino acid of the substrate is the third residue after the HxD motif from β-strand 1 and is colored on the "Segment of β_1-helix". Polar contacts between the residue on the β_1-loop and the backbone of the β_8-loop are indicated by dashed lines.

Sco3058, Rsp0802 and LmoDP. A histidine (Sco3058-H49, hRDP-H49, Rsp0802-H37), aspartate (LmoDP-D32) or glutamine (Gox2272-Q33, Cc2746-Q79 and Pa2393-Q80) residue at the end of the β_1-loop is within hydrogen-bonding distance to a backbone nitrogen of the β_8-loop (Sco3058-D322), but, a glutamine at this position makes a second contact to the backbone oxygen on the β_8-loop (Gox2272-G285).

β_7-Loop/helix. The β_7-loop/helix is a loop of three residues immediately following β_7 (residues 242-244 in Gox2272) and a subsequent single turn helix (residues 245-248 in Gox2272). The side chains of the residue which precedes the helix (Gox2272- M244) and the 2nd and 3rd residues of the helix (Ser-246 and Phe-247 in Gox2272) face towards the active site. However, Cc2746 has a loop immediately following β_7 (light blue ribbon in **Figure 4.9**), and there is not an equivalent residue to Sco3058-K247 or Gox2272-S246 in Cc2746.

In Gox2272, Cc2746 and Pa2393, the three colored residues of the β_8-loop are all glycines. One could hypothesize that if this portion of the β_8-loop is pushed downward and closer to the β_7-loop/helix by the glutamine from the β_1-loop, the side chains projecting from the first two of the three colored residues from the β_8-loop would probably interfere with the position of the active site side chains which project from the β_7-helix. In the case of Sco3058, hRDP, Rsp0802 and LmoDP, the β_8 loop is further out, extending the size of the pocket for the C-terminal side chain of a dipeptide.

Cluster 1 of cog2355. Gox2272 is part of the subset of amino acid sequences in Cluster 1 (**Figure 3.8**). Some of the proteins in this cluster are much longer in sequence

(~ 600 amino acids) compared to Gox2272 (335 amino acids). In these longer sequences, the C-terminal region belongs to cog2355. However, the ~ 250 amino acids of the N-terminal region are part of cog2071. BLAST hits to one of these longer sequences (gi|150010128), have annotations such as glutamine amidotransferase class II/dipeptidase, dipeptidase, and Class I glutamine amidotransferase. The closest amino acid sequence to gi|150010128 belonging to a protein with a structure corresponds to (Lin1909, gi|220702529, PDB code *3FIJ*). Within the *3FIJ* structure there is a mononuclear Zn center, ligated to a serine, three histidines and a water molecule, which is in close proximity to a potential Cys-His-Glu catalytic triad. A protein with these same residues is found in the *Gluconobacter oxydans* genome (locus tag Gox0800) though it is not a fusion protein. There are approximately 40 amino acid sequences which are a fusion of these two proteins, but there was not consistency in the annotations of proteins in the immediate genome context of these proteins.

Cluster 8 of cog2355. Cc2746 is in cluster 8 of cog2355. In *Stenotrophomonas maltophilia*, the dipeptidase is adjacent to a protein annotated as an amino acid permease-associated region. In *Alteromonas macleodii* (gi|196158469) *Pseudoalteromonas atlantica* (gi|109897266), the dipeptidase is adjacent to a protein annotated as a TonB receptor. Sometimes proteins annotated as TonB receptors include siderophore receptor in the annotation. Siderophore can be partly composed of L- and D-amino acids. The genome contexts of proteins within cluster 8 are not all alike and do not reveal a specific function to all members of the cluster.

Future Directions. Cc2746 and Gox2272 remain to be cocrystallized with the desired inhibitors. The human renal dipeptidase is known to hydrolyze β-lactams, albeit less efficiently than dipeptides, but there is not a structure of hRDP complexed with a β-lactam. With respect to β-lactam hydrolysis, Gox2272 was different from Sco3058 and Cc2746. The catalytic efficiency for the hydrolysis of nitrocefin by Gox2272 was at least 170-fold higher than for Sco3058 and Cc2746. The crystallization of Gox2272 with a β-lactam bound would be of interest.

There is one protein in cluster 8 which differs from the rest of the sequences in the cluster, namely the protein with locus tag PFL_1383 (gi|70728763). The equivalent active site residue to Cc2746-Y284 is an aspartate in PFL_1383. An aspartate at this position may result in 1) the enzyme favoring the hydrolysis of dipeptides with a positively charged residue at the C-terminus or 2) rejection of a negatively charged α-carboxylate at the C-terminus possibly resulting in the protein functioning as an aminopeptidase for a peptide of at least three amino acids.

CHAPTER V

THREE RENAL DIPEPTIDASE-LIKE PROTEINS WITH AN ACTIVE SITE TRYPTOPHAN

INTRODUCTION

The structure of the inhibitor complex of Sco3058 with the phosphinate mimic of L-Ala-D-Asp clarified the question as to how the α-amino group of a dipeptide was recognized by the microbial homologues of the human renal dipeptidase (hRDP). The α-amino group of the substrate is recognized by the aspartate from β-strand 1. In agreement with inhibitor complexes of Sco3058·**4** and hRDP·cilastatin, the C-terminal carboxylate of a dipeptide has electrostatic interactions with the beta metal, M_β, and the arginine on the loop following β-strand 6. The carbonyl of the amide bond makes polar contacts to M_β, and a histidine immediately following β-strand 4 ($His_{\beta4}$) (*23*, *80*).

The hRDP enzyme binds to cilastatin-Sepharose, but the non-conservative $His_{\beta4}$→Leu mutant does not (*77*). When $His_{\beta4}$ (His-150) of Sco3058 from *S. coelicolor* was mutated to an asparagine, the K_m increased by 2.7-fold, and the k_{cat} decreased by ~ 200-fold compared to wild-type (*80*). Those amino acid sequences within cog2355 which are predicted to have the complete set of metal ligands for the binuclear metal center will be termed the "binuclear subset". All of the members of the binuclear subset have either an active site tryptophan or histidine at the end of β_4 (His-150 in Sco3058), with the exception of one protein (gi|192362060) which has a tyrosine at this position. More often than not, the equivalent residue to His-150 in Sco3058 is a tryptophan, $Trp_{\beta4}$.

Of the two proteins addressed in the literature which have $Trp_{\beta 4}$, one enzyme from *Brevibacillus borstelensis* (BCS-1) was described as having good activity (L-Ala-D-Ala, k_{cat} = 760 s^{-1}, K_m = 3.7 mM, k_{cat}/K_m = 2 x 10^5 $M^{-1}s^{-1}$) whereas the second protein from *Acinetobacter calcoaceticus* (AcDP) was low in activity (Leu-D-Leu, K_m = 9.3 mM, k_{cat}/K_m = 6.6 x 10^1 $M^{-1}s^{-1}$) (*49, 76*). These K_m values are somewhat high, but the dipeptides which were selected may not have been ideal substrates (*76*). One difference between BCS-1 and AcDP is the active site residue following β-strand 2 (β_2-Tyr = Tyr-68 in Sco3058 and hRDP).

Based on amino acid sequence alignments of proteins within the binuclear subset, it is expected that the equivalent residue to β_2-Tyr is almost always an aromatic residue where the frequency of occurrence of Tyr > Phe > Trp. An overlay of Sco3058·pseudodipeptide **4** and hRDP·cilastatin complexes shows that C_3 of cilastatin, the atom in cilastatin that is closest in position to the free amino group of **4**, is 4.1 Å from the β_2-Tyr and 3.7-4.4 Å from Asp-22 (β_1). All of the proteins of known amino acid sequence that are represented in the literature have a tyrosine at this position except AcDP, which has phenylalanine (β_2-Phe). The hydroxyl group of this tyrosine residue was not expected to be important enough to account for the ~ 3-4 orders of magnitude difference in k_{cat}/K_m between BCS-1 and AcDP, but nevertheless, it is an active site residue and there seemed to be a discrepancy as to whether it was necessary to achieve optimal enzymatic activity (*49, 76*).

Attempts to obtain purified proteins with $Trp_{\beta 4}$ from *Bacillus cereus* (Bc3775, gi|30021867), *Sulfolobus sulfataricus* (Sso1864, gi|15898656) and *Chlamydophila*

pneumonia (Cpn0269, gi|15618189) were unsuccessful; the proteins were produced from recombinant DNA but were solely in the insoluble fraction of the cell lysates. The targets Bc3775 (9513b) and Sso1864 (9313a) could not be purified by the New York SGX Research Center for Structural Genomics (NYSGXRC) as well, failing at the stages of solubility and purification, respectively. However, three other proteins with $Trp_{\beta 4}$, namely Rsp0802 from *Rhodobacter sphaeroides* 2.4.1 (9523c, gi|77464378), LmoDP from *Listeria monocytogenes* serotype 4b (9513a, gi|46908635) and Bh2271 from *Bacillus halodurans* C-125 (9316a, gi|15614834), were purified at NYSGXRC and studied here. Unique genes cloned by NYSGXRC are identified by a four digit number followed by a letter.

LmoDP and Bh2271 are from Cluster 2 (**Figure 3.8**) and have a β_2-Phe residue, whereas Rsp0802 is from Cluster 3 has the β_2-Tyr residue. The native Rsp0802 structure was solved by the Midwest Center for Structural Genomics and deposited into the PDB (PDB code *3FDG*), but residues 61-65 or 61-75 were not defined in the two subunits of the homodimer. The coordinates of a close homolog to LmoDP, Lmo2462 from *Listeria monocytogenes* strain EDG-e, became available recently (PDB code *3LU2*). However, all of the predicted metal-binding residues of Lmo2462 were 3.0-3.6 Å from the two Zn ions, and the glutamate from β_3 was not bridging the zincs as a bidentate ligand.

Using the same methods outlined for Sco3058 in Chapter III, the proteins Rsp0802, LmoDP and Bh2271 were screened with dipeptide libraries in order to reveal their substrate specificity profile. From the profile, appropriate substrates were selected

to ascertain the quality of enzymes with $Trp_{\beta 4}$ and β_2-Tyr residues. Rsp0802, LmoDP and Bh2271 hydrolyzed an array of dipeptides within these libraries but showed a preference for those substrates with an L-amino acid at the N-terminus and a D-amino acid at the C-terminus. LmoDP exhibited a somewhat stricter substrate profile with a preference for L-Arg/Lys at the N-terminus and D-Ala at the C-terminus. With regard to the C-terminal preference, Bh2271 was more similar to LmoDP than Rsp0802, and dipeptides with a C-terminal D-alanine residue were also hydrolyzed faster than the other components of the L-Leu-D-Xaa library. The pseudodipeptide mimic of Leu-Ala (**8**) was found to be the most effective inhibitor of Rsp0802 and LmoDP out of the set of phosphinate compounds **4-9** (**Scheme 3.2**). In collaboration with Steve Almo's lab at Albert Einstein College of Medicine, the structures of the Rsp0802·pseudodipeptide **7** (PDB code *3LY0*) and LmoDP·pseudodipeptide **8** (PDB code *3NEH*) complexes were solved to 1.4 Å and 1.6 Å resolution, respectively. The relationship between the substrate specificity and the active site residues (real and predicted) will be discussed.

MATERIALS AND METHODS

Materials. The compounds which were purchased commercially and the precursors for dipeptide library synthesis were previously described in Chapters II and III (*56, 80*). All other purified compounds as well as the dipeptide libraries were synthesized by Dr. Chengfu Xu. Metal analyses were conducted using inductively coupled plasma mass spectrometry (ICP-MS) as previously described (*59*). ICP-MS standards were obtained from Inorganic Ventures Inc. Oligonucleotide syntheses and

DNA sequencing were performed by the Gene Technologies Lab of Texas A&M University.

Cloning and Overexpression of Rsp0802, Bh2271, LmoDP by NYSGXRC. The genes for Rsp0802, LmoDP and Bh2271 were cloned into the TOPO® vector (Invitrogen) from the genomic DNA of the aforementioned strains of *R. sphaeroides*, *L. monocytogenes* and *B. halodurans*. According to details provided in the Structural Genomics Knowledgebase, the proteins were purified with His-tags on a HiTrap Chelating column (Ni^{2+} charged) from Amersham Biosciences. Cell lysis was conducted at 4 °C in 50 mM Tris/500 mM NaCl/20 mM imidazole/0.1 % Tween 20, pH 7.5. The loading buffer was 20 mM Tris/500 mM NaCl/10 mM imidazole/10mM methionine/10% glycerol, pH 7.8, and the elution buffer was the loading buffer with 500 mM imidazole. In the final step, each protein was eluted through a gel filtration column with 10 mM Hepes/150 mM NaCl/10 mM methionine/10 % glycerol/5 mM DTT, pH 7.5. Bh2271 was overproduced in *E. coli* as a Se-Met protein.

Ninhydrin-based Enzyme Assays. Unless otherwise stated, assays were conducted at 30 °C in 50 mM Hepes, pH 7.5. With the exception of substrates L-Ala-L-Asp and L-Arg-L-Asp, the ninhydrin method was used to measure the concentration of free amino acid products at 507 nm as previously described for Sco3058 with dipeptides in Chapter III (*61, 80*). For the assay of Rsp0802 with L-Tyr-D-Leu, the heating time with the ninhydrin reagent was extended to 11 minutes. For the determination of the inhibition constants of Rsp0802 and LmoDP with the phosphinate mimic of Leu-Ala (**8**), LmoDP was assayed with 1 mM L-Ala-D-Ala and Rsp0802 with 2 mM L-Leu-D-Ala.

Both enzymes were first tested with a single concentration of 25 μM of each of the six phosphinate inhibitors (**4-9**). The concentration ranges of **8** used with LmoDP and Rsp0802 were 0-160 μM and 0-600 μM, respectively.

Aspartic Acid Coupling System Assay. The hydrolysis of L-Ala-L-Asp and L-Arg-L-Asp was measured by monitoring the decrease in NADH concentration at 340 nm. The coupling system assay contained varying concentrations of dipeptide, 3.7 mM α-ketoglutarate, 0.36 mM NADH, 100 mM KCl, 1.0 unit of malate dehydrogenase (Calbiochem) and 7 units of aspartate aminotransferase (Sigma-Aldrich) in 100 mM Hepes, pH 7.5.

Sample Preparation for HPLC Analysis. All assays used in the preparation of samples for the quantitation of the liberated C-terminal amino acids by HPLC were conducted in the same manner as described for the D-aminoacylases (Chapter II). Unless otherwise stated, the only differences from one enzyme to another are the enzyme concentration ranges and the incubation times of the libraries with the enzyme. The Rsp0802 and LmoDP enzymes were removed from the samples with a VWR® centrifugal filter (polyethersulfone membrane, MWCO = 10 KDa).

Activation of Enzymatic Activity with Divalent Cations. The buffer used in these experiments was 50 mM Hepes, pH 7.5. The substrate L-Leu-D-Ala (0.4 mM) was used to assess enzymatic activity. Zn was the most abundant metal present in the purified proteins. Rsp0802 and LmoDP were assayed in the absence and presence of varying concentrations of $ZnCl_2$ in the assay (1, 2, 5, 10, 25 and 100 μM for Rsp0802 and 0.5, 2, 5, 10 and 20 μM for LmoDP). Twenty-five μM (~ 1 mg/mL) Rsp0802 and LmoDP

were incubated with 0, 25, 50, 100 and 150 μM $ZnCl_2$ for 3 and 24 hours at 0 °C. Following incubation for 24 hours, the enzymes were assessed for their activity with and without 5 μM $ZnCl_2$ during the reaction.

The rate of hydrolysis of 0.5 mM L-Leu-D-Ala by Bh2271 was assessed in the presence of 0, 1, 5, 25 and 100 μM $ZnCl_2$, $CoCl_2$ and $MnCl_2$ in the reaction. The activity of Bh2271 could be enhanced with the addition of $ZnCl_2$, $MnCl_2$ or $CoCl_2$ in the assay. The enzyme (25 μM) was incubated with and without 100 μM $MnCl_2$ on ice for 5 minutes, 3.5 hours, and 20 hours on ice. The incubated samples were assayed at 25 °C with and without 100 μM $MnCl_2$ in the assay. The activity of Bh2271 was also assessed with 10, 25, 50 and 100 μM $MnCl_2$ to find the minimum amount needed to maximize activity.

N-terminal Specificity of 3 Dipeptidases. The 3 dipeptidases were screened for activity against 55 dipeptide libraries (19 L-Xaa-L-Xaa, 19 L-Xaa-D-Xaa, and 17 D-Xaa-L-Xaa) except for Bh2271, which was not screened with the D-Xaa-L-Xaa libraries. The concentration of each library component in the assay was ~ 100 μM, but L-Cys, D-Cys and D-Ile were not included as components of any library. For the D-Xaa-L-Xaa libraries, the incubation time of the libraries with enzyme was 16 hours and the concentration range of enzyme used was 0-2.5 μM in the presence of 5 μM $ZnCl_2$ (LmoDP) or 10 μM $ZnCl_2$ (Rsp0802). The L-Xaa-L-Xaa and L-Xaa-D-Xaa libraries were incubated with 0-2 μM Bh2271 for 18 hours. The Zn-treated LmoDP (0-2.5 μM) was incubated with the L-Xaa-L-Xaa and L-Xaa-D-Xaa libraries for 4 hours where the

total concentration of $ZnCl_2$ in the assays was 5 μM. The Zn-activated Rsp0802 was incubated with the L-Xaa-L-Xaa (0-2.5 μM Rsp0802) and L-Xaa-D-Xaa (0-1.5 μM Rsp0802) libraries for 4.5 hours in the presence of 10 μM $ZnCl_2$. The liberated free amino acids were measured via the ninhydrin method, and the absorbance at OD_{507nm}, *Q*, was plotted as a function of enzyme concentration, E_t, and fit to equation 1.

C-terminal Specificity of Rsp0802. Enzyme-course assays for five dipeptide libraries: L-Leu-D-Xaa, L-Leu-L-Xaa, L-Met-D-Xaa, L-Met-L-Xaa and D-Met-L-Xaa were analyzed by HPLC for determination of the C-terminal specificity preference. With the exception of the D-Leu-L-Xaa library, each library was incubated at 30 ºC for 5 hours, with and without Zn-activated enzyme, in 25 mM $(NH_4)HCO_3$, pH 8.0 supplemented with 10 μM $ZnCl_2$. The D-Met-L-Xaa library was incubated under the same conditions for 16 hours. The concentrations of enzyme used were 0-2 μM (L-Leu-D-Xaa), 0-2.5 μM (L-Leu-L-Xaa, L-Met-L-Xaa, and D-Met-L-Xaa) or 0-1.5 μM (L-Met-D-Xaa). The slope of the linear relationship between the quantity of each C-terminal amino acid and the concentration of enzyme was used to compare the rates of hydrolysis of the various components of the libraries. The same assay conditions, methods of sample preparation and data analysis were used in screening the C-terminal specificity of LmoDP and Bh2271.

C-terminal Specificity of LmoDP. In a similar manner to Rsp0802, LmoDP was screened for its substrate preference with respect to the identity of the C-terminal amino acid of a dipeptide. Seven dipeptide libraries were incubated in the absence and presence of the following concentration ranges of Zn-activated LmoDP: 1 nM-0.75 μM

(L-Lys-D-Xaa), 5 nM-1.0 μM (L-Lys-L-Xaa), 1 nM-0.75 μM (L-Arg-D-Xaa), 20 nM-2.5 μM (L-Arg-L-Xaa), 2 nM-1.0 μM (L-Met-D-Xaa), 20 nM-2.0 μM (L-Met-L-Xaa) and 0.1-2.5 μM (D-Trp-L-Xaa). The assays were supplemented with 5 μM $ZnCl_2$. The libraries with an N-terminal L-amino acid were incubated for 3.5 h at 30 °C, but the D-Trp-L-Xaa library was allowed to incubate for a total of 16 hours under the same conditions.

C-terminal Specificity of Bh2271. The L-Leu-D-Xaa and L-Leu-L-Xaa libraries were treated with 0-2.0 μM and 0-4 μM Bh2271 for 18 hours at 30 °C prior to sample preparation for HPLC analysis.

Crystallization and Data Collection. Crystallization screens of Rsp0802 and LmoDP were conducted by Elena Fedorov. The structure of the Rsp0802·pseudodipeptide **7** and LmoDP·pseudodipeptide **8** complexes were solved by Dr. Alexander Fedorov from Dr. Steve Almo's lab at Albert Einstein College of Medicine. The crystallization conditions described were taken from the protocols for the NYSGXRC clones 9513a and 9523c summarized on the Structural Genomic Knowledgebase. The crystallization solution of Rsp0802 contained 9.8 mg/mL Rsp0802, 20% PEG 3350, 200 mM potassium formate at 21 °C, and the cryoprotectant was 20% ethylene glycol. LmoDP crystallized at 21 °C in 100 mM Bis-Tris, pH 6.5 and 25% PEG 3350. The cryoprotectant for LmoDP was 20 % glycerol.

Data Analysis. For those assays conducted by the ninhydrin method, standards consisted of mixtures of each of the two appropriate amino acid products at equal molar ratios in the same buffer as the enzymatic assay. The standards were heated with

ninhydrin alongside each set of assays for each substrate. The initial velocity data, *v*, were fit to equation 2, where E_t is the total enzyme concentration, k_{cat} is the turnover number, A is the substrate concentration, and K_a is the Michaelis constant.

RESULTS

Purification of Rsp0802, Lmo2462 and Bh2271. Rsp0802 was found to contain 0.6 equivalents of Zn and less than 0.1 equivalents of Fe or Ni by ICP-MS. The Se-Met Bh2271 protein had 0.8 equivalents of Zn and 0.4 equivalents of Fe. The metal content of LmoDP was somewhat lower, having 0.3 equivalents of Zn and 0.15 equivalents of Ni.

Activation of Rsp0802 and LmoDP Activity with Divalent Cations. The activity of Rsp0802 and LmoDP, toward the hydrolysis of 0.4 mM L-Leu-D-Ala, increased with up to 5 μM $ZnCl_2$ added directly to the assay (**Figures 5.1A** and **5.1C**), but their activity was enhanced more by pre-incubation with $ZnCl_2$ (**Figures 5.1B** and **5.1D**). Rsp0802 and LmoDP reached maximal activity enhancement within 3 hours when soaked in 2 and 4 equivalents of $ZnCl_2$, respectively. After pre-incubating the enzymes with $ZnCl_2$, further enhancement of activity was not achieved with additional $ZnCl_2$ in the assay, and the addition of 5 μM $ZnCl_2$ to the assay resulted in less activity for LmoDP. Enzymatic assays with LmoDP and Rsp0802 were conducted after soaking the enzyme with the aforementioned optimal equivalents of $ZnCl_2$.

Activation of Bh2271 Activity with Divalent Cations. The rates of hydrolysis of L-Leu-D-Ala by Bh2271 in the presence of 0-100 μM $ZnCl_2$, $CoCl_2$ and $MnCl_2$ are

plotted in **Figure 5.2**. The activity of Bh2271 increased with up to 5 μM $ZnCl_2$ (~ 6-fold), 100 μM $CoCl_2$ (38-fold) and 100 μM $MnCl_2$ (~ 29-fold). The amount of Bh2271

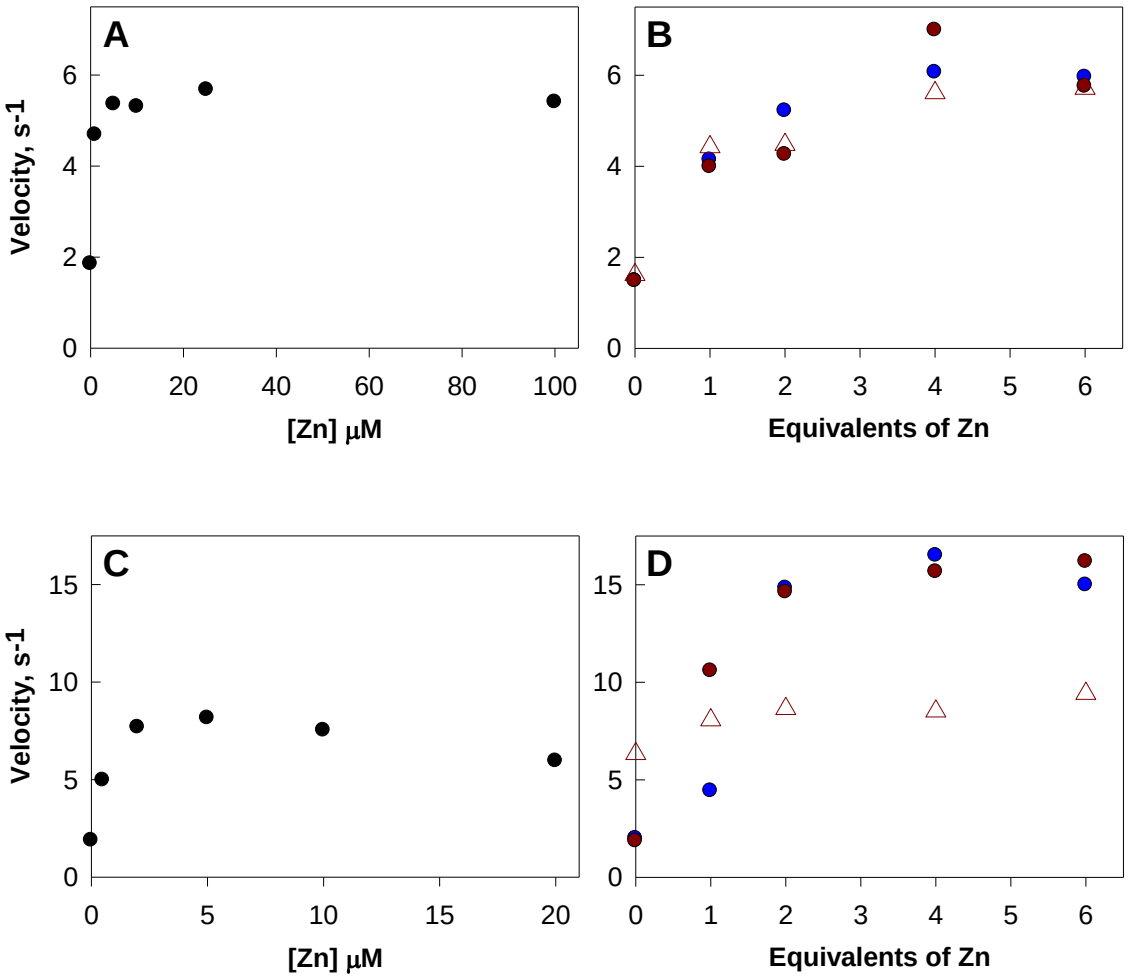


Figure 5.1: Dependence of the initial velocities of LmoDP and Rsp0802 on the presence of Zn. The substrate was 0.4 mM L-Leu-D-Ala. **(A)** Rsp0802 with various concentrations of Zn in the assay. **(B)** Rsp0802 incubated with $ZnCl_2$ for 3 hours (*blue* circles), 24 hours (*red* circles) and 24 hours with 5 μM $ZnCl_2$ supplemented in the assay (*red* triangles). **(C)** LmoDP with various concentrations of $ZnCl_2$ in the assay. **(D)** LmoDP incubated with $ZnCl_2$ for 3 hours (*blue* circles), 24 hours (*red* circles) and 24 hours with 5 μM $ZnCl_2$ supplemented in the assay (*red* triangles). The y-axis of each left plot (**A**, **C**) is scaled the same as the plot to its right (**B**, **D**).

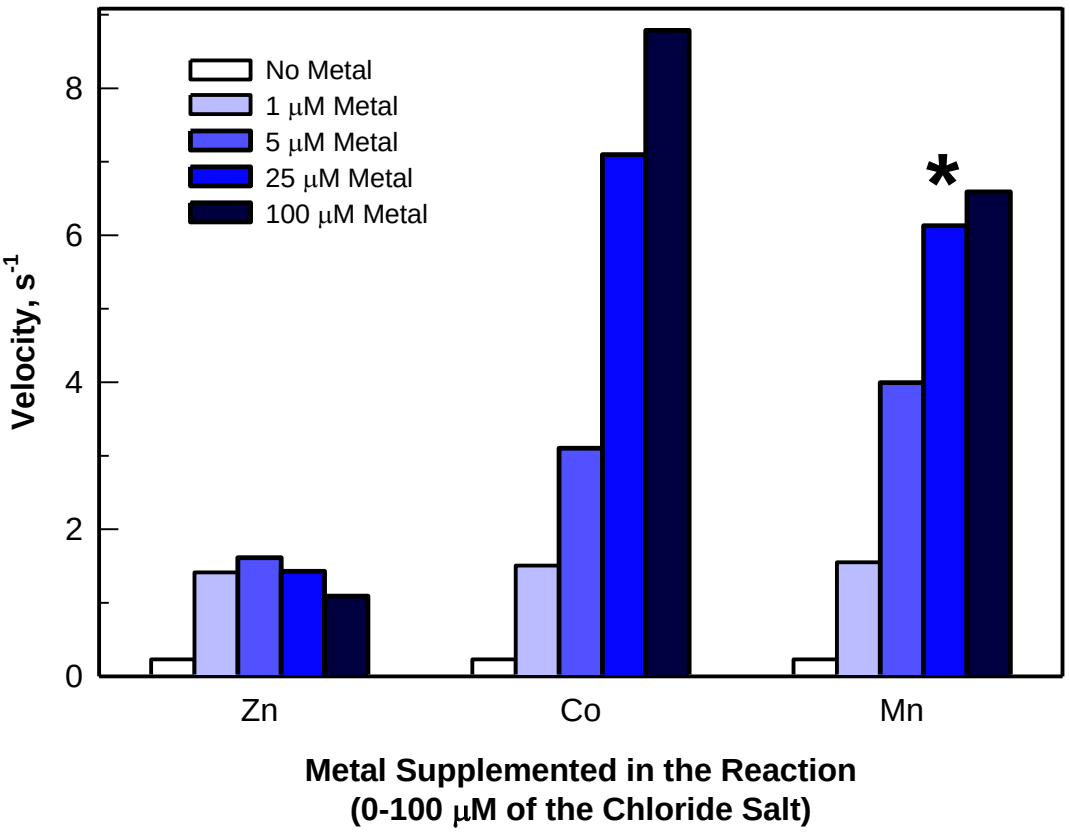


Figure 5.2: Activation of Bh2271 in the presence of various concentrations of $ZnCl_2$, $CoCl_2$ and $MnCl_2$ in the enzymatic reaction. The concentration of the substrate L-Leu-D-Ala was 0.5 mM. The condition noted with an asterisk was used in the determination of the kinetic constants with dipeptides.

protein available was minimal, and although Zn was the metal which was naturally present in purified Bh2271, 25 μM $MnCl_2$ was added to the assays with this protein because it enhanced the activity by ~ 5-fold more than $ZnCl_2$. Bh2271 was soaked with varying equivalents of $MnCl_2$ to determine if increased activity could be obtained without the need of higher concentrations in the assay. However, there was no change in the activity of the soaked enzyme between 5 minutes and 20 hours of incubation with 4 equivalents of Mn. The enzyme was tested once more with varied concentrations of $MnCl_2$ in the assay and at concentrations above 25 μM, there seemed to be a point of diminishing returns.

N-terminal Specificity of Rsp0802. As depicted in **Figure**s **5.3A-C**, Rsp0802 had a preference for dipeptides with the amino acids L-Leu, L-Met, L-Phe and L-Tyr at the N-terminus. The L-Xaa-D-Xaa libraries were hydrolyzed up to 4 times faster than their L-Xaa-L-Xaa counterparts. When the N-terminal residue of the dipeptide substrate was a D-amino acid, the preferred substrates were still those with hydrophobic side chains, namely, D-Met, D-Leu and D-Trp. Subsequently, the L-Leu-L-Xaa, L-Leu-D-Xaa, L-Met-L-Xaa and D-Met-L-Xaa libraries were chosen for quantitation of the liberated C-terminal amino acids by HPLC.

N-terminal Specificity of LmoDP. In contrast to Rsp0802, LmoDP had a marked preference for the positively charged *N*-L-Arg-Xaa and especially *N*-L-Lys-Xaa dipeptide substrates. LmoDP was also able to hydrolyze other *N*-L-Xaa dipeptides, but clearly rejected those substrates with an N-terminal L-Val, L-Ile, L-Thr, L-Pro, L-Asp and L-Glu. When a D-amino acid was at the N-terminus, *N*-D-Trp-L-Xaa substrates were

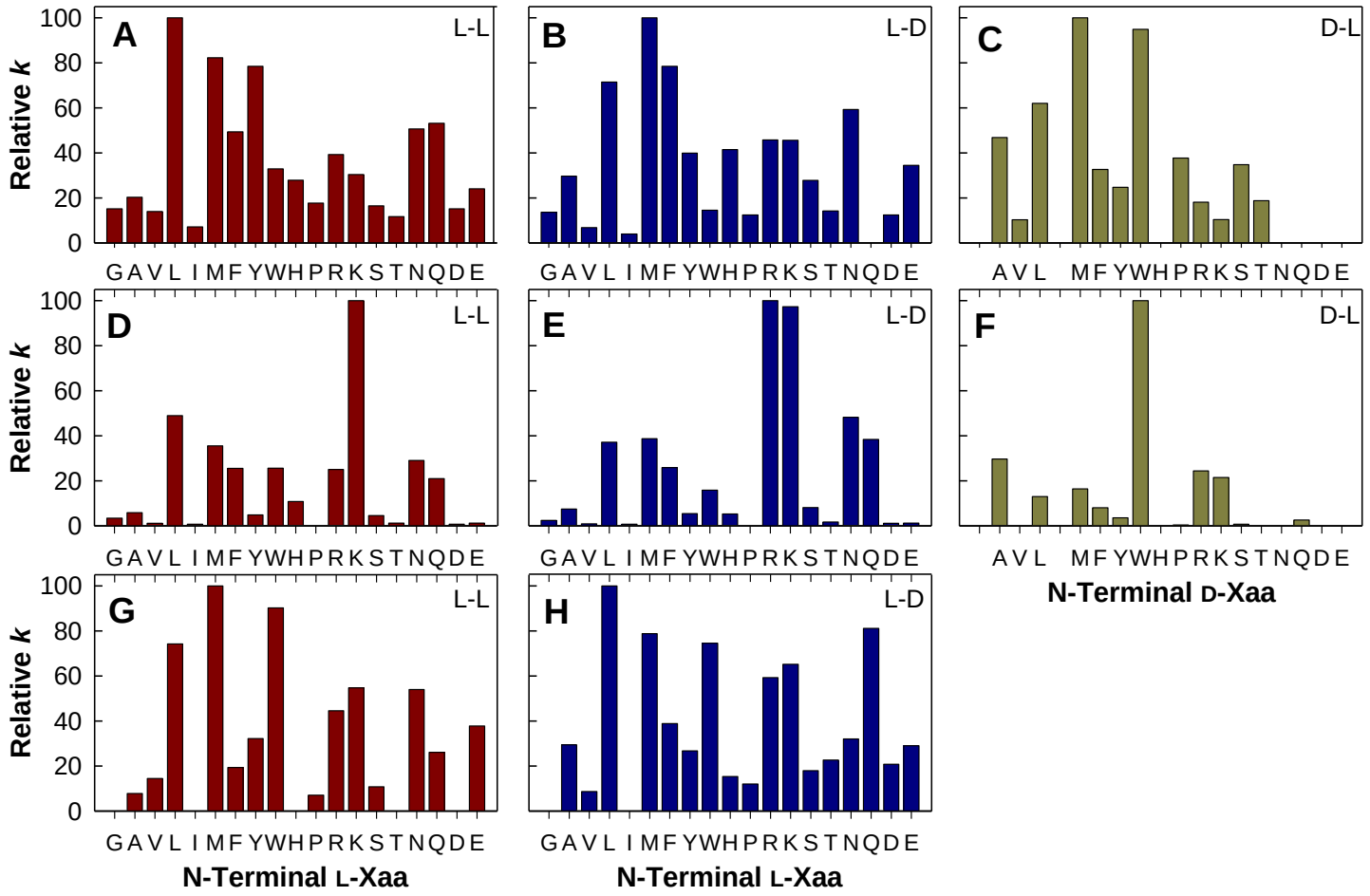


Figure 5.3: N-terminal substrate specificity profiles of Rsp0802 (panels A-C), LmoDP (panels D-F) and Bh2271 (panels G-H). The relative *k* (nM^{-1}) values for the hydrolysis of dipeptide libraries from fits of the OD_{507nm} data as a function of enzyme concentration. **(A**, **D** and **G)** *red* L-Xaa-L-Xaa. **(B**, **E** and **H)** *blue* L-Xaa-D-Xaa. **(C** and **F)** *olive* D-Xaa-L-Xaa.

hydrolyzed the fastest. The relative rates of hydrolysis of the 55 dipeptide libraries are summarized in **Figure**s **5.3 D-F**. Seven libraries, namely, L-Lys-L-Xaa, L-Lys-D-Xaa, L-Arg-L-Xaa, L-Arg-D-Xaa, L-Met-L-Xaa, L-Met-D-Xaa and D-Trp-L-Xaa were further analyzed for the C-terminal specificity of LmoDP.

N-terminal Specificity of Bh2271. The N-terminal specificity profile of Bh2271 is summarized in **Figure**s **5.3 G-H**. Bh2271 hydrolyzed *N*-L-Leu-Xaa, *N*-L-Met-Xaa and *N*-L-Trp-Xaa dipeptides slightly faster than the other libraries given the same stereochemistry of C_α at the C-terminus. Bh2271 was not screened with the seventeen D-Xaa-L-Xaa libraries because higher concentrations of enzyme are required with these libraries and there was a need to conserve limited amount of enzyme sample. The L-Leu-D-Xaa and L-Leu-L-Xaa libraries were further examined by HPLC to establish the C-terminal preference of the enzyme.

C-terminal Specificity of Rsp0802. The results of the C-terminal specificity screens of Rsp0802 against five dipeptide libraries are summarized in **Figure 5.4**. For all the libraries, dipeptides which contained a Trp, Met, Phe, Tyr or Leu at the C-terminus were hydrolyzed the fastest relative to the other dipeptides within the same library. However, the L-Xaa-D-Xaa dipeptides were preferred over their L-Xaa-L-Xaa counterparts. The dipeptide which was hydrolyzed the fastest, within the set of substrates in these five libraries, was L-Leu-D-Trp. It is expected that some of the best substrates, contained within the dipeptide libraries screened by both methods, are L-Leu-D-Trp, L-Met-D-Leu, L-Met-D-Trp, and L-Leu-D-Leu. The substrates L-Met-D-Leu, L-Leu-D-Leu and L-Tyr-D-Leu were synthesized.

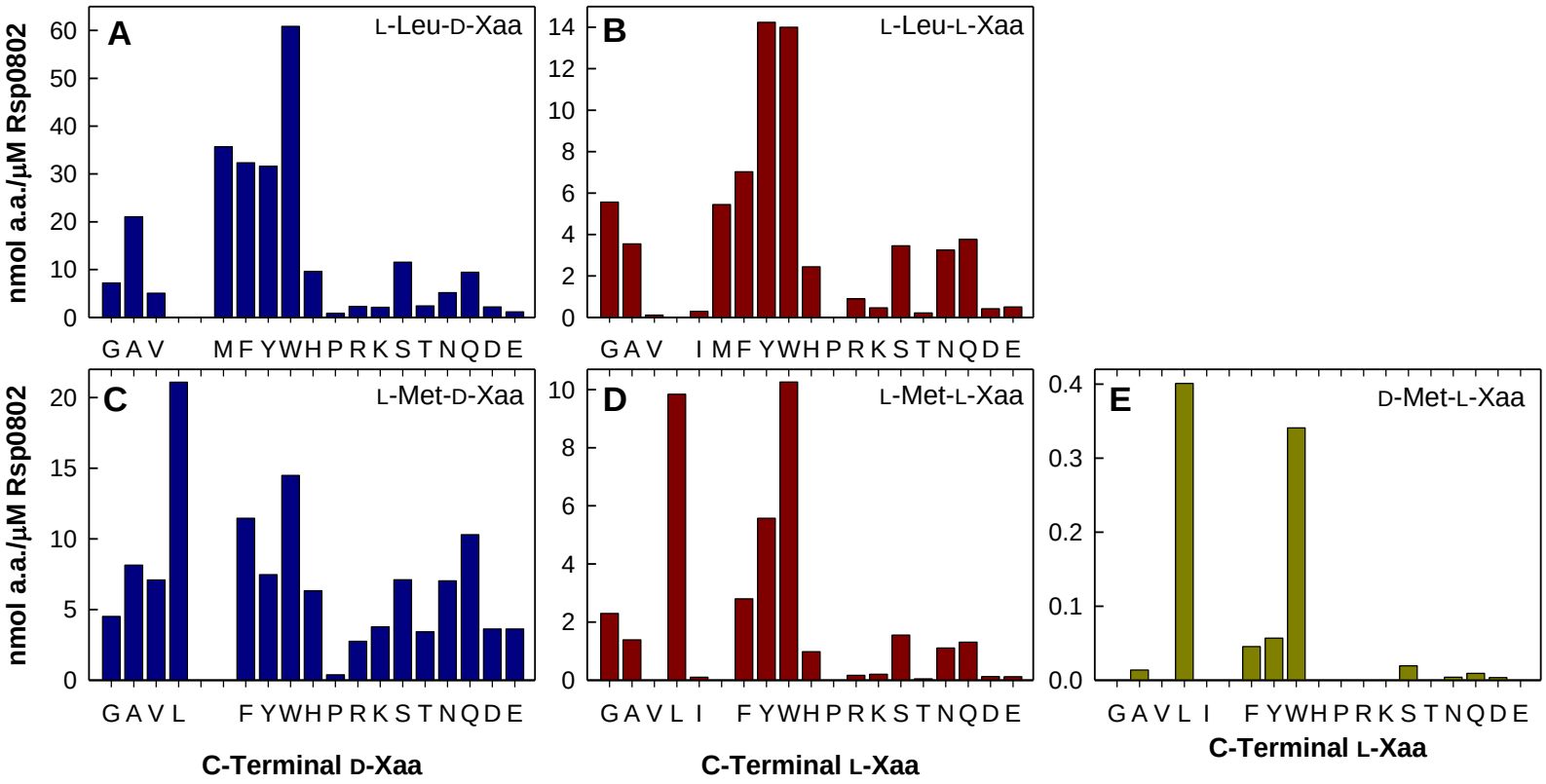


Figure 5.4: C-terminal substrate specificity profile of Rsp0802. The rates of hydrolysis of the selected dipeptide libraries **(A)** L-Leu-D-Xaa, **(B)** L-Met-D-Xaa, **(C)** L-Leu-L-Xaa, **(D)** L-Met-L-Xaa and **(E)** D-Met-L-Xaa were obtained from the slopes of the linear correlation between the nmol of liberated C-terminal amino acids (a.a.) as a function of enzyme concentration. The results presented in panels A-D are from 5-hour incubations with enzyme at 30 °C, and in panel E, the incubation time with enzyme was 16 hours under the same conditions.

Kinetic and Inhibition Constants of Rsp0802. The kinetic constants for Rsp0802 with ten selected dipeptide substrates are presented in **Table 5.1**. Of the individual dipeptides assayed, a catalytic efficiency of ~ 10^5 $M^{-1}s^{-1}$ was observed with the L-Met-D-Leu substrate (k_{cat}/K_m = 1.1 x 10^5 $M^{-1}s^{-1}$). The K_i of Rsp0802 with the pseudodipeptide of Leu-Ala (**8**) was 130 ± 7 μM.

C-terminal Specificity of LmoDP. LmoDP was screened for C-terminal specificity against seven dipeptide libraries. As illustrated in **Figure**s **5.5A-C**, within the three L-D libraries screened (L-Lys-D-Xaa, L-Arg-D-Xaa and L-Met-D-Xaa), dipeptides with a C-terminal D-Ala were hydrolyzed the fastest, relative to other *N*-L-Xaa-D-Xaa substrates within the same library. The screen of four libraries with a C-terminal L-amino acid (**Figure 5.5 D-G**) revealed that the specificity was somewhat broad with a slight preference for L-Leu, L-Ala and Gly at the C-terminus. The dipeptide substrate which had the fastest observed rate of hydrolysis was L-Lys-D-Ala, followed closely by L-Arg-D-Ala.

Kinetic and Inhibition Constants of LmoDP. The kinetic constants for LmoDP with eleven dipeptides are presented in **Table 5.2**. The best of these substrates was L-Leu-D-Ala (k_{cat}/K_m = 1.1 x 10^5 $M^{-1}s^{-1}$). The phosphinate mimic of Leu-Ala (**8**) was also an inhibitor of LmoDP (K_i = 12 ± 1 μM).

C-terminal Specificity of Bh2271. The two libraries L-Leu-L-Xaa and L-Leu-D-Xaa were treated with Bh2271 under identical conditions. Bh2271 hydrolyzed L-Leu-D-Ala the fastest within these two libraries as illustrated by the screening results summarized in **Figure 5.6**. Within the L-Leu-L-Xaa library, L-Leu-Gly was hydrolyzed

Table 5.1: Kinetic constants for select substrates of Rsp0802[a]

Substrate	k_{cat}, s^{-1}	K_m, mM	k_{cat}/K_m $M^{-1}s^{-1}$
L-Met-D-Leu	24 ± 0.3	0.22 ± 0.01	(1.1 ± 0.07) x 10^5
L-Tyr-D-Leu	5.2 ± 0.1	0.09 ± 0.01	(5.7 ± 0.3) x 10^4
L-Leu-D-Leu	5.2 ± 0.1	0.2 ± 0.01	(2.5 ± 0.2) x 10^4
L-Leu-D-Ala	55 ± 3	4.5 ± 0.5	(1.2 ± 0.1) x 10^4
L-Leu-D-Ser	76 ± 6	8.3 ± 1	(9.1 ± 1.3) x 10^3
L-Ala-D-Ala			(2.2 ± 0.1) x 10^3
L-Leu-D-Glu			(1.0 ± 0.02) x 10^3
L-Met-D-Glu			(9.5 ± 0.02) x 10^2
L-Arg-D-Asp			(9.6 ± 0.02) x 10^2
L-Arg-L-Asp			(1.9 ± 0.1) x 10^2

[a]From fits of the data to equation 2. For those entries without values of k_{cat} and K_m, saturation was not achieved at concentrations up to 10 mM.

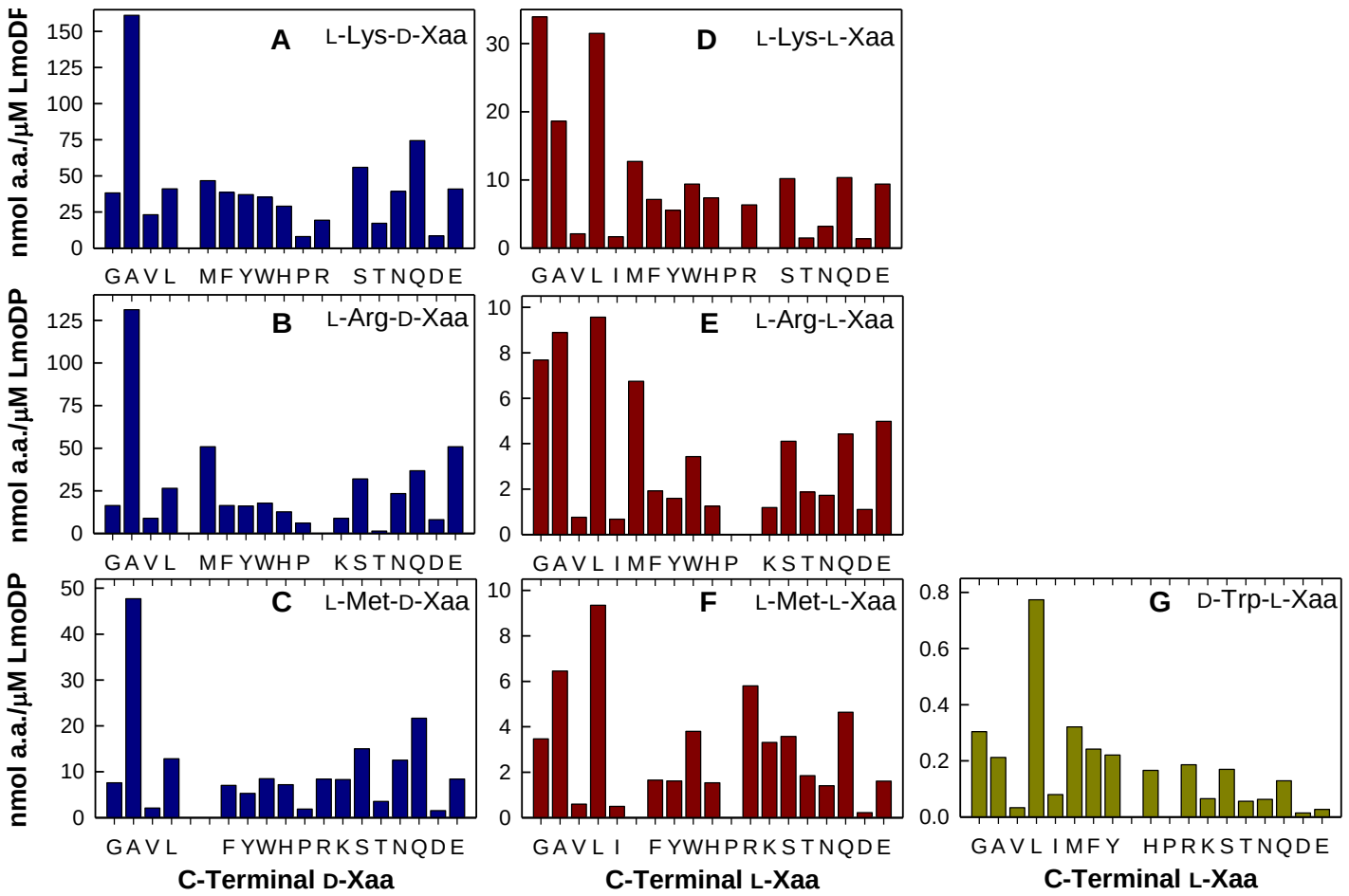


Figure 5.5: C-terminal substrate specificity profile of LmoDP. The rates of hydrolysis of the selected dipeptide libraries **(A)** L-Lys-D-Xaa, **(B)** L-Arg-D-Xaa, **(C)** L-Met-D-Xaa, **(D)** L-Lys-L-Xaa, **(E)** L-Arg-L-Xaa, **(F)** L-Met-L-Xaa and **(G)** D-Trp-L-Xaa were obtained from the slopes of the linear correlation between the nmol of liberated C-terminal amino acids (a.a) as a function of enzyme concentration. The results presented in panels A-F are from 3.5-hour incubations with enzyme at 30 °C, and in panel G, the incubation time with enzyme was 16 hours under the same conditions.

Table 5.2: Kinetic constants for select substrates of LmoDP[a]

Substrate	k_{cat}, s^{-1}	K_m, mM	k_{cat}/K_m $M^{-1}s^{-1}$
L-Leu-D-Ala	29 ± 0.6	0.27 ± 0.02	(1.1 ± 0.1) x 10^5
L-Leu-D-Ser	41 ± 1	1.1 ± 0.1	(3.6 ± 0.3) x 10^4
L-Arg-Gly	89 ± 4	2.9 ± 0.3	(3.1 ± 0.4) x 10^4
L-Met-D-Leu	98 ± 9	3.2 ± 0.3	(3.1 ± 0.3) x 10^4
L-Ala-D-Ala	58 ± 2	2.0 ± 0.2	(2.8 ± 0.2) x 10^4
L-Met-D-Glu	30 ± 2	2.6 ± 0.3	(1.2 ± 0.2) x 10^4
L-Tyr-D-Leu	3.8 ± 0.2	2.2 ± 0.2	(1.7 ± 0.2) x 10^3
L-Leu-D-Leu	112 ± 5	16 ± 1	(6.9 ± 0.6) x 10^3
L-Arg-D-Asp	42 ± 5	5.9 ± 1.2	(7.0 ± 1.7) x 10^3
L-Arg-L-Asp	1.3 ± 0.04	13 ± 1	(1.1 ± 0.08) x 10^2
L-Ala-L-Asp	0.07 ± 0.002	17 ± 1	(4.2 ± 0.3) x 10^0

[a]From fits of the data to equation 2.

the fastest, followed by L-Leu-L-Ala. Based on these results, the best substrates for Bh2271 were anticipated to be L-Leu-D-Ala, L-Met-D-Ala and L-Trp-D-Ala.

Kinetic Constants of Bh2271. The kinetic constants for Bh2271 with 8 dipeptides are presented in **Table 5.3**. The best substrate was L-Leu-D-Ala (k_{cat}/K_m = 7.4 x 10^4 $M^{-1}s^{-1}$). At one time a status report showed that Bh2271 was at the stage of optimization of crystals, but work on Bh2271 was stopped by NYSGXRC. The crystallographers in Dr. Steve Almo's lab were not in possession of the Bh2271 protein. Since Bh2271 would not be crystallized by the time of the completion of this manuscript, the enzyme was not assayed for the inhibition constants with the phosphinate compounds.

Structure of Rsp0802. The native crystal structure of Rsp0802 (PDB code *3FDG*) was solved by the Midwest Center for Structural Genomics to 1.8 Å with two Mg bound in the active site. Here, the 1.4 Å resolution X-ray crystal structure of Rsp0802 (PDB code *3LY0*) is presented in complex with a pseudodipeptide mimic of L-Ala-D-Ala (**7**) and two Zn ions (M_α and M_β) bound in the active site. Compound **7** was a mixture of four possible enantiomers but both chiral centers in the molecule bound in Rsp0802 were in the R configuration. The $(\beta/\alpha)_8$ barrel structure of the ligand-bound Rsp0802 is illustrated in **Figure 5.7**.

There are two insertion regions within the core sequence which makes up the $(\beta/\alpha)_8$ barrel. The first of these regions is inserted between the second β-strand, β_2, and second alpha-helix, α_2, of the barrel. Insertion 1 is not observed in hRDP and its other

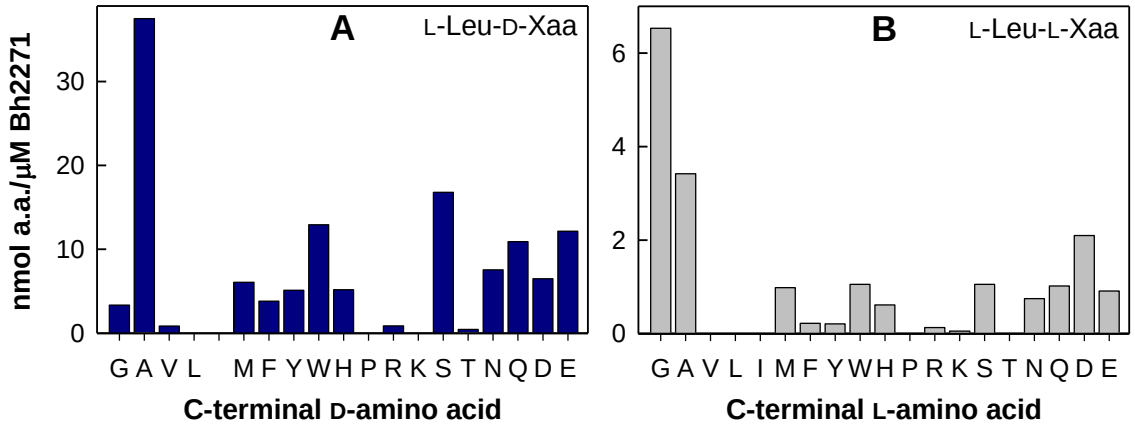


Figure 5.6: C-terminal substrate specificity profile of Bh2271. The rates of hydrolysis of the selected dipeptide libraries **(A)** L-Leu-D-Xaa and **(B)** L-Leu-L-Xaa were obtained from the slopes of the linear correlation between the nmol of liberated C-terminal amino acids (a.a) as a function of enzyme concentration. Both libraries were incubated with Bh2271 for 18 hours at 30 °C.

Table 5.3: Kinetic constants for select substrates of Bh2271[a]

Substrate	k_{cat}, s^{-1}	K_m, mM	k_{cat}/K_m $M^{-1}s^{-1}$
L-Leu-D-Ala	85 ± 3	1.1 ± 0.1	(7.4 ± 0.6) x 10^4
L-Met-D-Glu	83 ± 2	1.4 ± 0.1	(5.7 ± 0.5) x 10^4
L-Leu-D-Ser	83 ± 4	1.5 ± 0.1	(5.4 ± 0.6) x 10^4
L-Ala-D-Ala	157 ± 7	3.5 ± 0.4	(4.5 ± 0.6) x 10^4
L-Leu-D-Glu	50 ± 0.7	1.8 ± 0.06	(2.9 ± 0.1) x 10^4
L-Met-D-Leu	99 ± 4	7.4 ± 0.6	(1.3 ± 0.1) x 10^4
L-Leu-D-Leu	25 ± 0.5	7.0 ± 0.3	(3.5 ± 0.2) x 10^3
L-Tyr-D-Leu	17 ± 1	11 ± 1	(1.5 ± 0.2) x 10^3

[a]Assayed in the presence of 25 μM $MnCl_2$.

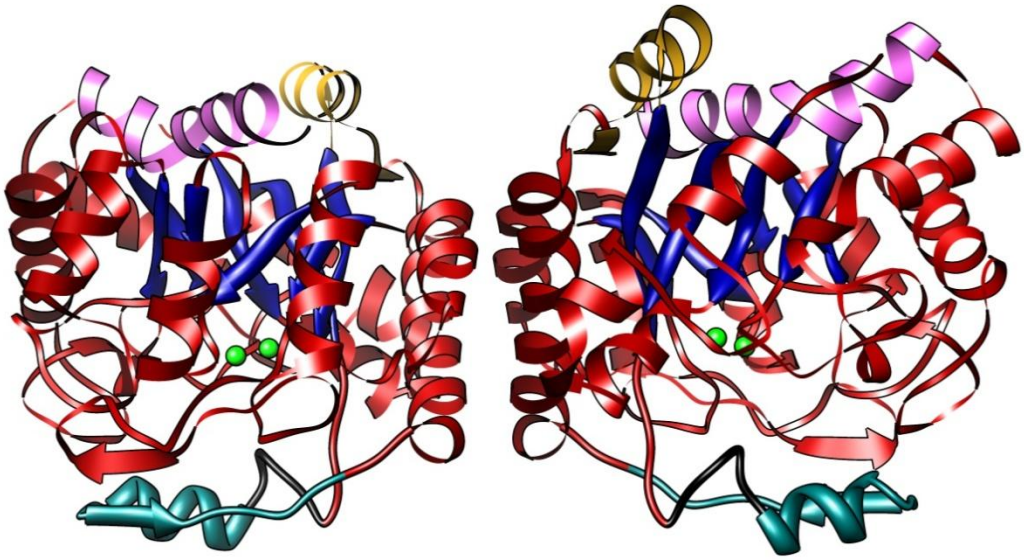


Figure 5.7: Ribbon diagram of the Rsp0802 dimer (PDB code *3LY0*). The $(\beta/\alpha)_8$-barrel is colored *red* (α-helices and loops) and *blue* (β-strands). Insertion region 1, which contains active site residues (61-84), is colored *teal* and *black*. The residues which were not defined in the PDB: *3FDG* structure (61-65) are colored *black*. Zn ions are represented as *green* spheres. The C-terminal α-helix (*pink*) and the region inserted between α_2 and β_3 of the $(\beta/\alpha)_8$-barrel (*yellow*) are also colored.

microbial homologues of known structure (PDB codes *3K5X*, *2RAG*, *1ITU*). The structure of part of insertion 1, namely residues 61-75 and 61-65 in different subunits from the PDB:*3FDG* coordinates, was undefined in the native structure but was resolved in the inhibitor-bound structure presented here. Another insertion is also present in all known structures of proteins in cog2355 and is located between α_2 and β_3. Insertion 1 is colored red and black in the ribbon diagram (**Figure 5.7**). The ordered loop at the N-terminus of insertion 1 consists of residues Gln-61, Ala-62, His-63, Asp-64 and Ala-65. Gln-61 is within hydrogen-bonding distance to His-67 from the α-helix of insertion 1. The N^{δ} of H63 is 2.8 Å from the hydroxyl group of Tyr-56, and Asp-64 forms a salt bridge with Arg-20. The insertion between β_7 and α_7, observed in Sco3058 and Cc2746, is not present in Rsp0802.

The binuclear metal center within the native Rsp0802 structure (PDB code *3FDG*) is shown in **Figure 5.8A**. The two metals, M_α and M_β, are bridged by the catalytic water molecule and Glu-137 from β-strand 3. The ligands exclusive to M_α are His-11 and Asp-13 from β-strand 1 and a water molecule. His-210, His-231 and two water molecules are ligands to M_β. The Zn_α and Zn_β ions in the Rsp0802·**7** complex (**Figure 5.8B**) ligate to the same residues as in the unbound structure, but moieties of the L-Ala-D-Ala mimic (**7**) replace the water molecules. Zn_α is ligated to the α-amino group of the dipeptide mimic, and Zn_β is coordinated to one of the phosphinate oxygens and the α-carboxylate of the inhibitor. The other phosphinate oxygen bridges the two Zn

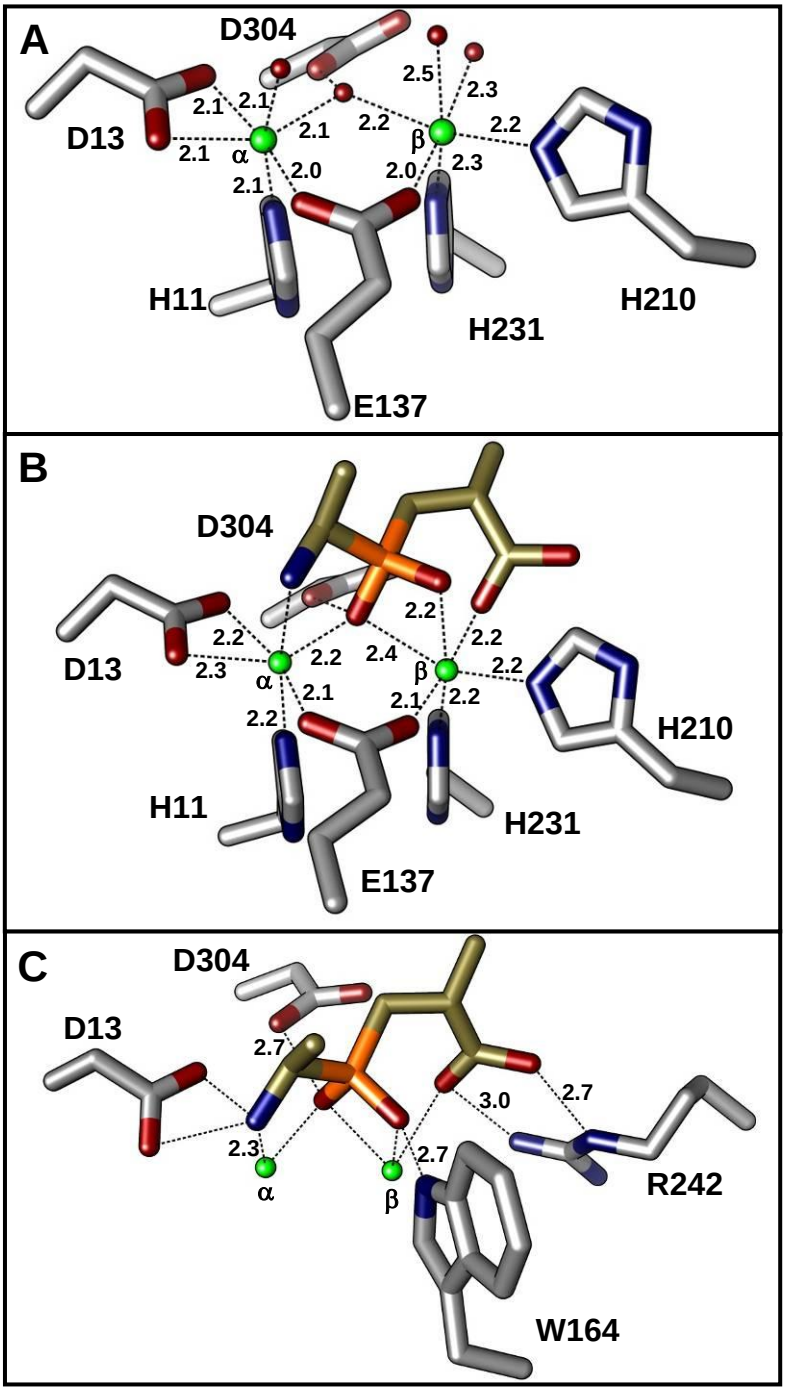


Figure 5.8: Metal center and active site of Rsp0802 in the presence and absence of inhibitor. The Zn_{α} and Zn_{β} metals are *green* spheres. The carbon atoms of Rsp0802 (*gray*) and pseudodipeptide (**7**) mimic of L-Ala-D-Ala (*gold* carbons) are colored differently. **(A)** Native structure (PDB code *3FDG*). **(B)** Metal center of the inhibitor (**7**) complex. **(C)** Other polar contacts between the inhibitor and Rsp0802 are represented as dashed lines with distances noted in Å. Distances already shown in panel **B** are not repeated in panel **C**.

ions. In both structures, the two metals appear to be hexacoordinate with square bipyramidal geometry, though the geometry of Zn_α is distorted.

The active site of the Rsp0802·**7** complex is presented in **Figure 5.8C**. The α-amino group of **7** is recognized by contacts to Zn_α (2.3 Å) and the side chain carboxylate of Asp-13 (3.1 Å to each oxygen). The N^ε and N^ω nitrogens of Arg-242 associate with both oxygens of the α-carboxylate of **7** at distances of 2.7 and 3.0 Å, respectively. One phosphinate oxygen bridges Zn_α (2.2 Å) and Zn_β (2.4 Å) and is also 2.7 Å from the catalytic Asp-304 which contacted the bridging water in the native structure. The second phosphinate oxygen has polar interactions with Zn_β (2.2 Å) and Trp-164 (2.7 Å). The inhibitor has a total of eleven polar contacts at distances of 2.2-3.1 Å.

Active Site Pocket of Rsp0802. The surface of the active site pocket of Rsp0802 is shown in **Figure 5.9A** and **5.9B** from the outside of one face and in **Figure 5.9C** and **5.9D** as a cross section. One face of the active site is composed of the metal-binding ligands (*green* side chains, **Figure 5.9A** and **C**). The hydrophobic pocket surrounding C_β of the side chain of the L-Ala moiety of the inhibitor (*blue* side chain surfaces, **Figure 5.9A** and **C**) is formed from C_α of Gly-307 and the side chains of Tyr-56, His-63, Leu-16 and Trp-164. The α-amino group, C_α and C_β of the N-terminal region of (**7**) is completely encapsulated within 6 Å from the C_α carbon. **Figures 5.9A** and **5.9C** are repeated in **Figures 5.9B** and **5.9D**, respectively, with the surface colored by element. The hydrophobic pocket for the side chain of the C-terminal amino acid is mainly composed of the side chains of Met-72, Phe-176, Ala-264, Phe-267 and Ala-308.

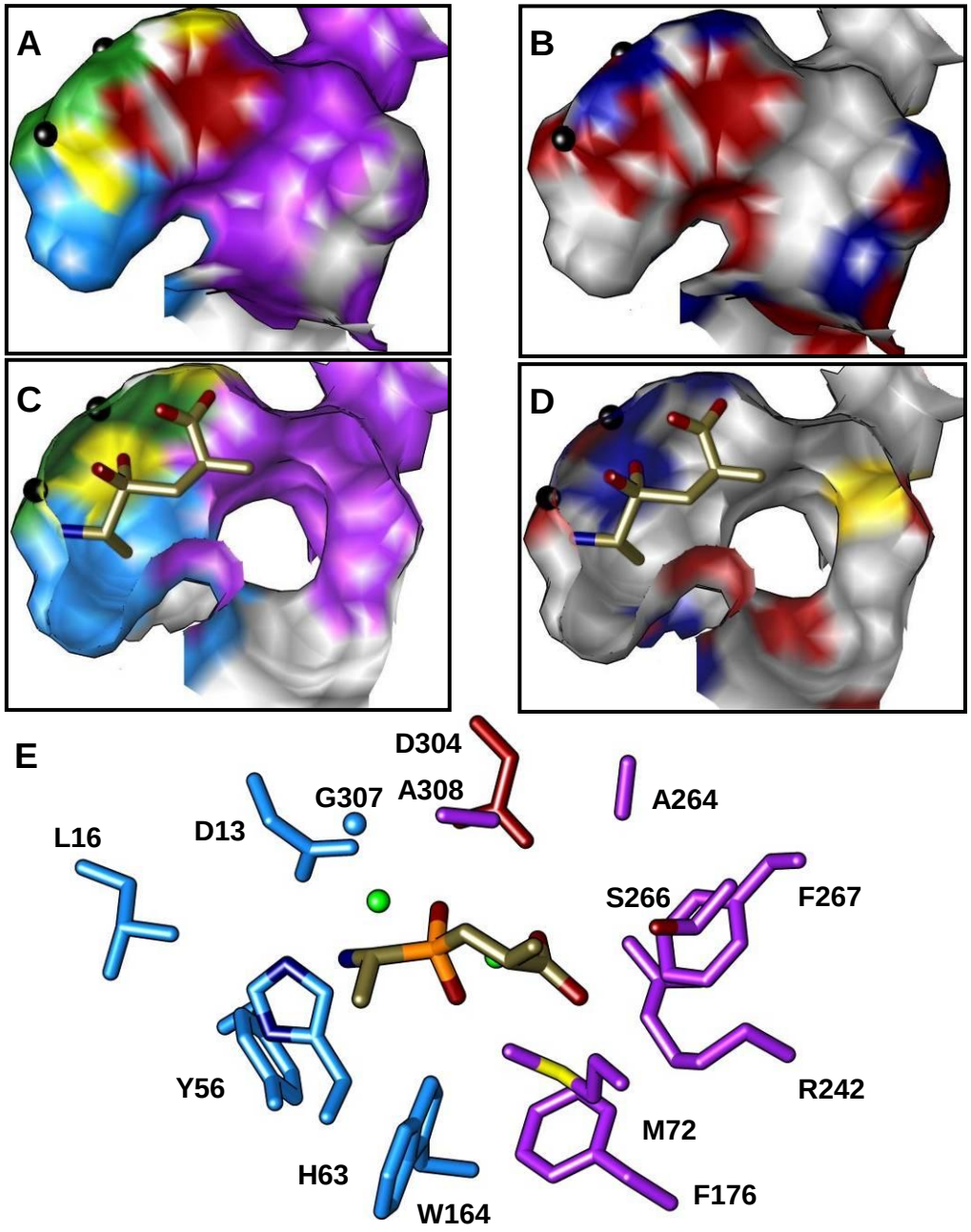


Figure 5.9: Surface and stick figure depictions of the Rsp0802 substrate binding pockets. **(A** and **C)** Surfaces of side chains of the catalytic aspartate (Asp-304, *red*), Zn ligands (*91*) other than Asp-13, residues for the pocket of the N-terminal half of a dipeptide (*blue*), residues for the pocket of the C-terminal half of the substrate (*purple*) and atoms with polar contacts to **7** (*yellow*) are color coded. All other atoms are colored gray. **(C)** Cross-section of panel A. **(B)** Same as panel A with all atoms colored by element (*blue* nitrogens, *red* oxygens, *yellow* sulfur, *gray* carbons). **(D)** Same as panel C with all atoms colored by element. **(E)** Side chains of residues are color-coded to match panels A and C. Those heteroatoms whose identities are not intuitive are colored by element.

*Structure of LmoDP·**8** Complex*. The X-ray crystal structure of LmoDP complexed with two Zn ions and the phosphinate dipeptide mimic of L-Leu-D-Ala (**8**) was solved to 1.6 Å. Like **7** in the Rsp0802·**7** complex, both of the chiral centers of inhibitor **8** identified in LmoDP active site were in the R configuration. The homodimeric structure of LmoDP is depicted in **Figure 5.10A**; Asp-269 faces in toward the active site of the opposing chain (red in **Figure 5.10A**). A salt bridge (3.9 Å) was identified between Lys-270 and Glu-145 from opposite subunits. The backbone carbonyl of Asn-232 and N^{ε} of His-201 from opposing chains were 2.8 Å apart.

The active site contains two pentacoordinate Zn ions with a distorted trigonal bipyramidal geometry that are bridged by the bidentate carboxylate side chain of Glu-105 as well as one of the phosphinate oxygens of **8**. The binuclear metal center is depicted in **Figure 5.10B**. The Zn_{α} ion is coordinated to His-7 and both of the side chain oxygens of Asp-9. The coordination sphere of Zn_{β} is completed by the non-bridging phosphinate oxygen of **8**, and the side chains of His-172 and His-192 from β-strands 5 and 6, respectively.

The free amino group of **8** is recognized by Asp-9 and Zn_{α} at distances of 3.1 and 2.7 Å, respectively. The α-carboxylate is ion-paired with N^{ε} and N^{ω} of Arg-203, both at distances of 2.9 Å, and is 2.8 Å from Zn_{β}. The phosphinate oxygen which contacts Zn_{β} is also 2.6 Å from N^{ε} of Trp-132 ($Trp_{\beta 4}$), whereas, the other phosphinate oxygen is 2.8 Å from the catalytic Asp-261 from β-strand 8. Contacts to the inhibitor not shown in **Figure 5.10B** are depicted in **Figure 5.10C**.

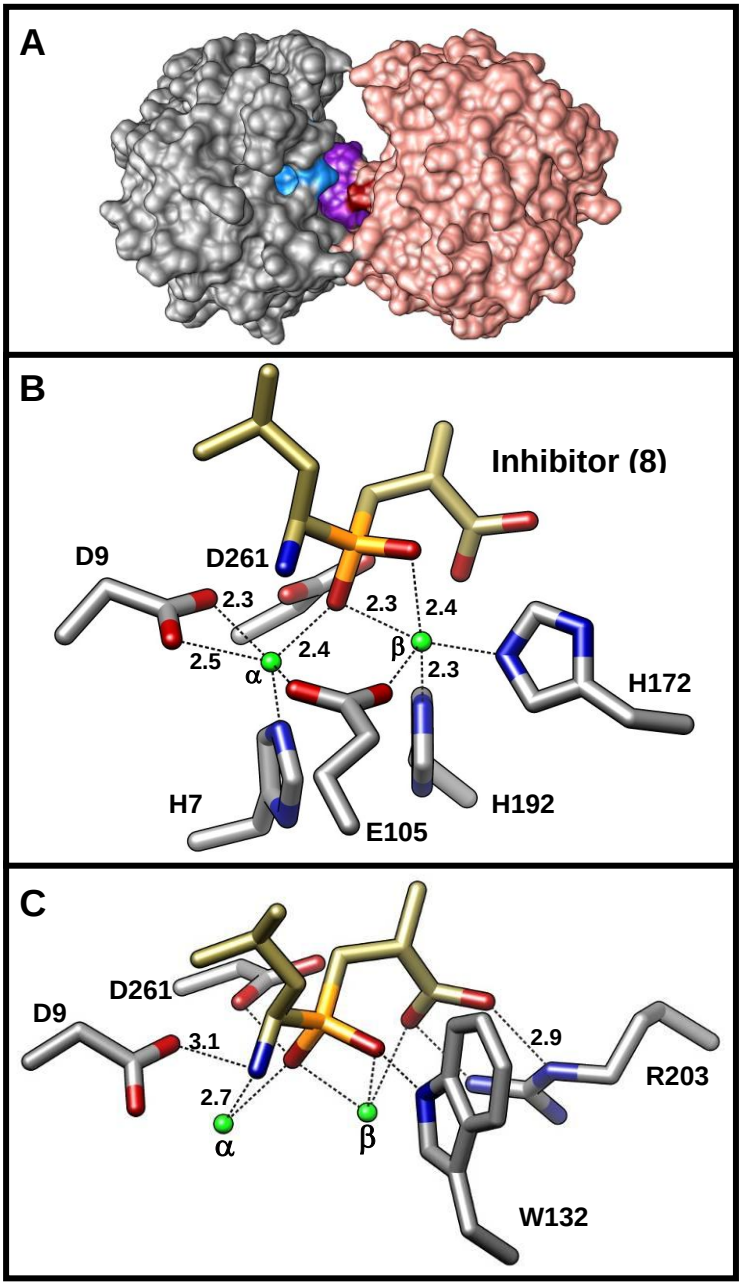


Figure 5.10: Representations of the LmoDP structure and its active site binuclear metal center complexed with inhibitor. **(A)** Surface representation of the LmoDP dimer with chains A and B colored *gray* and *pink*, respectively. The active site pockets (chain A) for the N- and C-terminal sides of a dipeptide are colored blue and purple, respectively. Asp-269 from chain B is colored *red*. **(A** and **B)** The inhibitor (*gold* carbons, *orange* phosphorus), Zn_{α}/Zn_{β} (*green* spheres) are colored the same. Polar contacts are indicated by dashed lines with distances noted in Å. **(B)** Binuclear metal center of LmoDP with inhibitor **8** bound. Those distances not labeled are 2.0-2.2 Å. **(C)** Polar contacts between LmoDP and the inhibitor. The dashed lines in panel **C** which are not labeled in panels **B** or **C** represent contacts of 2.6-2.9 Å.

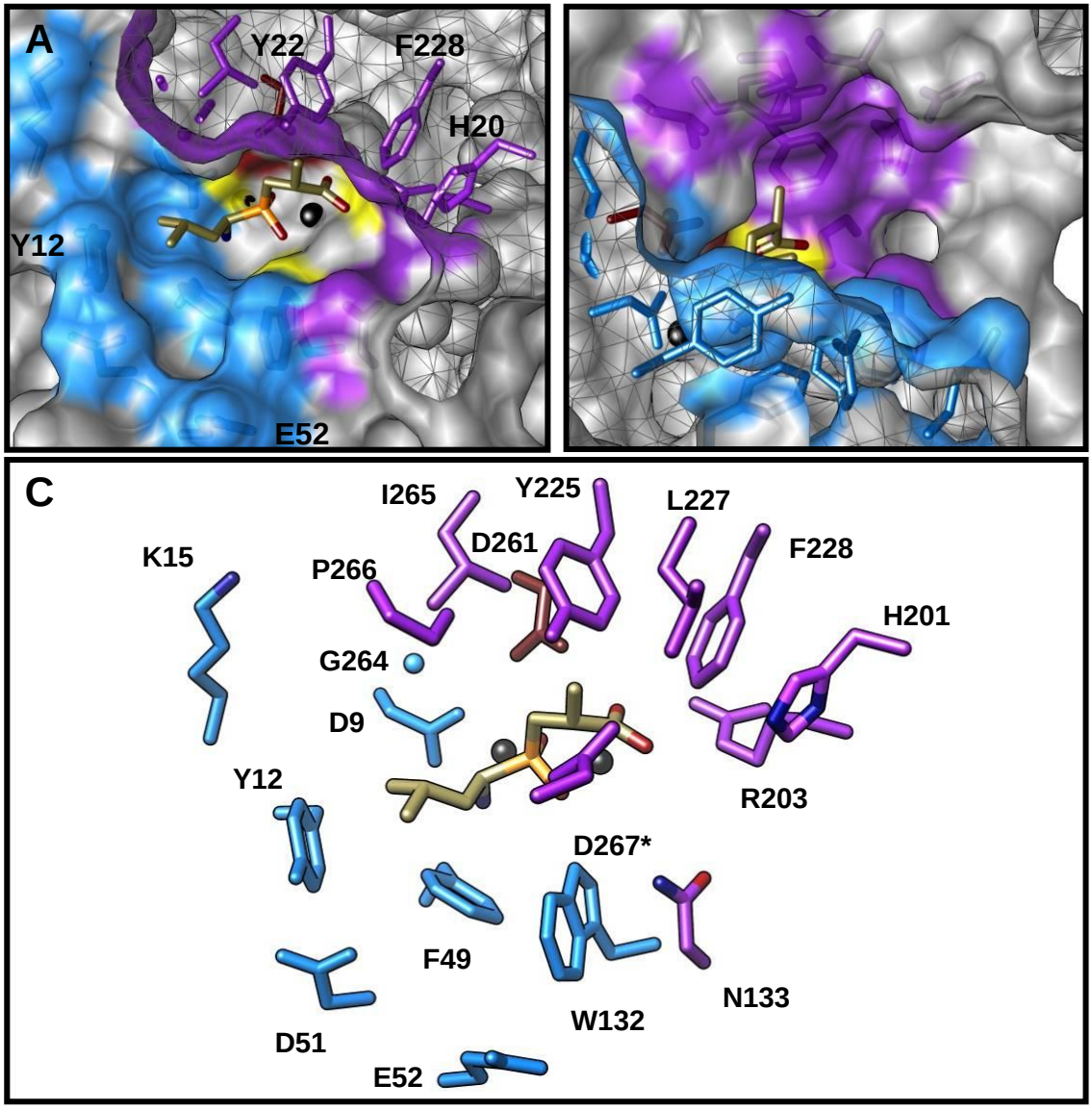


Figure 5.11: Surface and stick representations of the LmoDP substrate-binding pockets. The side chains of residues which make up the pockets for the N- and C-terminal sides of a dipeptide substrate are colored *blue* and *purple*, respectively. The catalytic Asp-261 is colored *red*. The pocket residue from the opposite chain is noted with an asterisk. **(A)** and **(B)** Surface representation of the LmoDP active site pockets with inhibitor (*gold* carbons, *orange* phosphorus) and two zincs (*black* spheres). Sections behind the clipped plane are shown with a flat mesh surface. Atoms of residues which make polar contacts to the inhibitor are colored *yellow*. Some residues are labeled to orient the reader. **(A)** Focus on the pocket for the N-terminal L-amino acid side chain of the substrate. **(B)** Highlight of the pocket for the C-terminal side chain of the substrate. The only visible parts of the inhibitor are the methylene group, α-carboxylate and the methyl group which represents a D-alanine side chain. **(C)** Residues colored in panels A & B where atoms whose identities are not intuitive are colored by element.

A surface representation of LmoDP near the active site is shown in **Figure 5.11A-B**, and stick figures of the residues colored in **Figure 5.11A-B** are labeled in **Figure 5.11C**. The crevice which makes up the pocket for the side chain of an N-terminal L-amino acid of a dipeptide is framed by C_α of Gly-264 and the side chains of Tyr-12, Phe-49, Asp-51 and Trp-132. The more solvent-exposed pocket for the side chain of the C-terminal end of the substrate is composed mainly of the side chains of Asn-133, His-201, Tyr-225, Leu-227, Phe-228 and Ile-265.

Overlay of Rsp0802, LmoDP and Lmo2462. The coordinate files for Rsp0802 (PDB code *3LY0*) and LmoDP (PDB code *3NEH*) were opened together. Residues 59-91/165-189 of Rsp0802 and 54-61/135-153 in LmoDP were deleted. The two truncated structures were superimposed using the MatchMaker tool in Chimera. The coordinates for the non-truncated structures were loaded into the same file and overlayed onto their respective truncated structures using MatchMaker.

Prior to the release of the LmoDP structure, the 2.2 Å resolution structure of native Lmo2462, which deviates from LmoDP sequence by 11 out of 308 residues, was deposited into the Protein Data Bank (PDB code *3LU2*) by the Center for Structural Genomics of Infectious Diseases. Lmo2462 and Bh2271 both have a histidine (His-225 in Lmo2462, His-224 in Bh2271) instead of Tyr-225 observed in LmoDP. In order to represent Bh2271, the Lmo2462 structure was superimposed onto the Rsp0802/LmoDP overlay using LmoDP as the reference structure. The residues in Lmo2462 were then changed to those from the Bh2271 sequence. The active site of the overlay of these

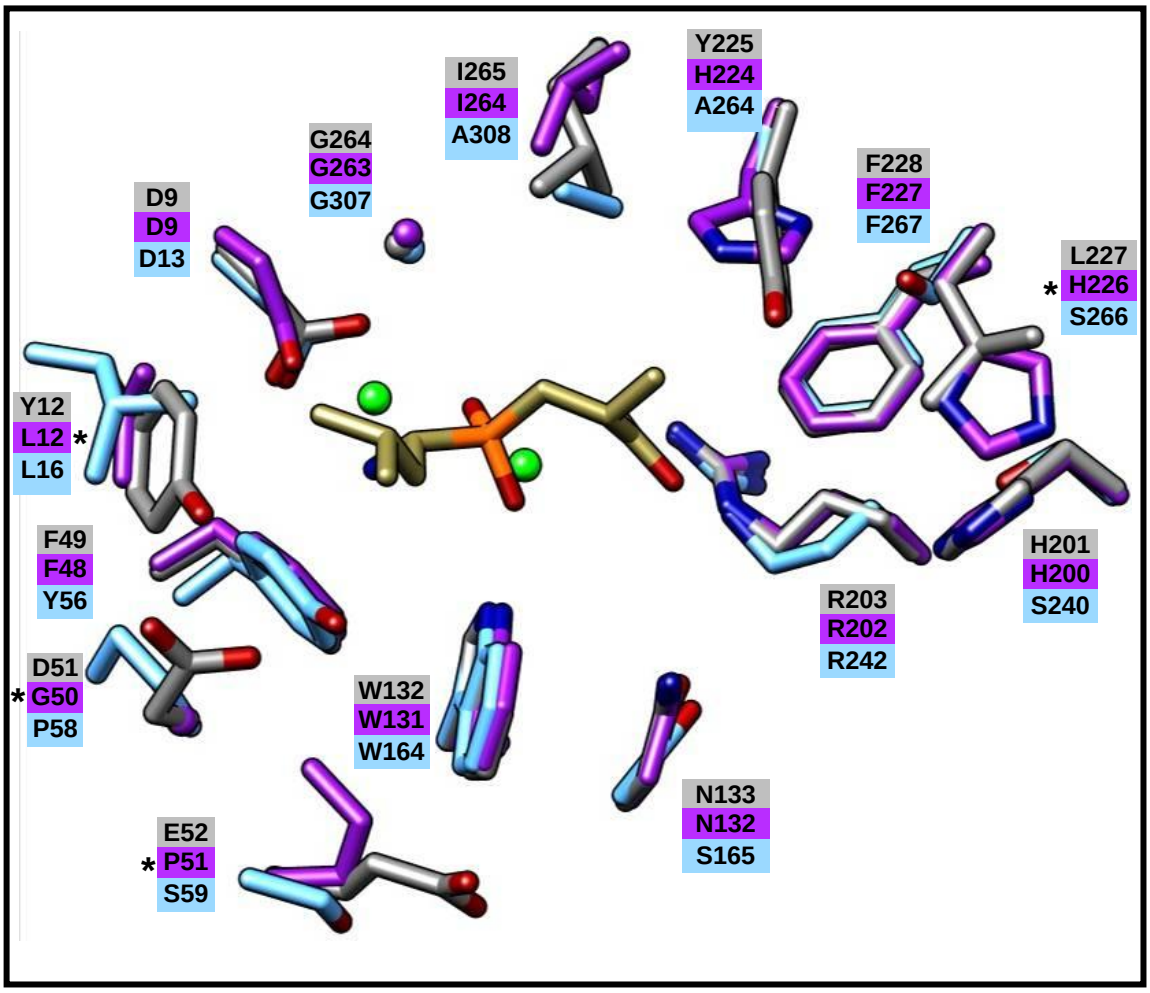


Figure 5.12: Overlay of active sites of LmoDP, Lmo2462 and Rsp0802 with residues in Lmo2462 changed to residues found in Bh2271. The carbons and residue labels from LmoDP (*gray*), Lmo2462 (*purple*) and Rsp0802 (*light blue*) are color coded. The L-Leu-D-Ala phosphinate mimic (**8**) (*gold* carbons, *orange* phosphorus) and two Zn ions (*green* spheres) from the LmoDP structure are also shown. Those residues which have been changed to match those in Bh2271 are marked by an asterisk.

three structures is shown in **Figure 5.12** with a similar orientation to **Figures 5.11A** and **5.11C**.

DISCUSSION

Amino Acid Sequence Alignment. An amino acid sequence alignment of Sco3058 and five other amino sequences from cog2355 is presented in **Figure 5.13**. The amino acid sequences originate from *R. sphaeroides* (Rsp0802), *L. monocytogenes* strain serotype 4b (LmoDP), *B. halodurans* (Bh2271), *A. calcoaceticus* (AcDP, gi|1088400), *B. borstelensis* (BCS-1, gi|9719426), *Lysinibacillus sphaericus* (Bsph4282, gi|169829713) and *S. coelicolor* (Sco3058, gi|21221500). In 2008, the genome sequence of *L. sphaericus* (formerly *Bacillus sphaericus*) was released and found to contain a gene which codes for a member of cog2355 (gi|169829713), namely Bsph4282. The calculated molecular weight of Bsph4282 is 34 kDa. A dipeptidase with an estimated molecular weight of 38 kDa (by SDS-PAGE) was isolated from *B. sphaericus*, but no part of the protein was sequenced (*92*). It is not known as to whether the isolated protein was Bsph4282, but Bsph4282 was included in the alignment for the purpose of discussion.

Substrate Specificity of Rsp0802, LmoDP and Bh2271. The three dipeptidases Rsp0802, LmoDP and Bh2271 were successfully purified by NYSGXRC, but the enzymes had less than optimal metal content. The activity of the enzymes increased with the presence of Zn, Co and Mn in the assay. The activity of LmoDP and Rsp0802 could be enhanced by pre-incubating the enzymes with $ZnCl_2$, without the need for

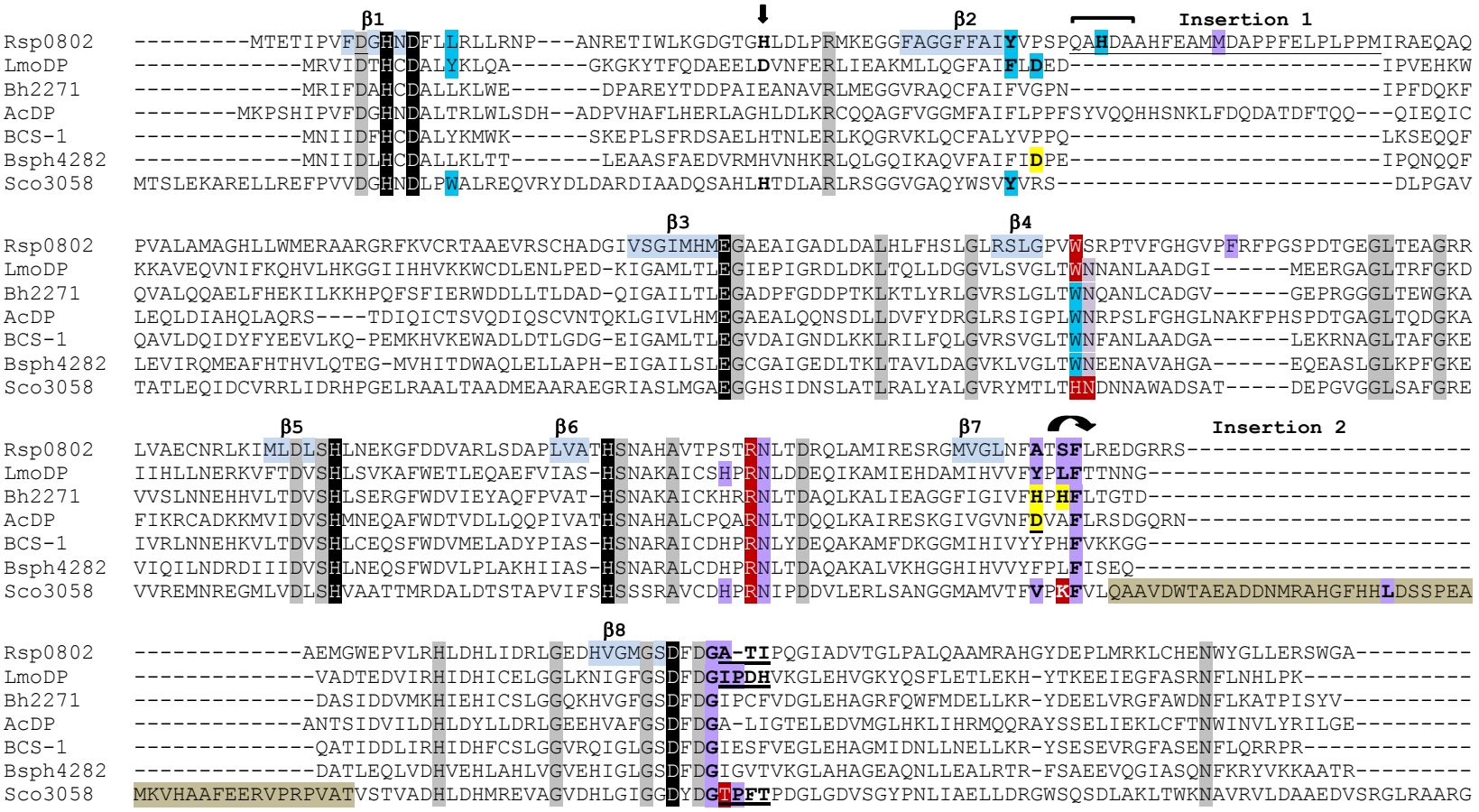


Figure 5.13: Amino acid sequence alignment of Sco3058 (residues 1-381) and select proteins from clusters 2 and 3 within cog2355. The metal ligands and catalytic aspartate are colored *black*. The residues which make up insertion region 1 (colored *teal* in the dimer figure) only observed in Rsp0802, are underlined. The second insertion region between β_7 and α_7 of Sco3058 not observed in LmoDP, Rsp0802 or hRDP is colored *brown*. Inhibitor-contacting (*red*), N-terminal substrate pocket (*medium blue*), C-terminal substrate pocket (*purple*) residues and β-strands (*light blue*) are colored coded. Other conserved residues are colored *gray*. Predicted active site residues are colored *yellow*. The β_8-loop has bold underlining and the residue which contacts the backbone of the β_8-loop is indicated by an arrow.

additional $ZnCl_2$ in the assay. The substrate specificity profile of three microbial dipeptidase homologues to the human renal dipeptidase, namely, Rsp0802, LmoDP and Bh2271, were assessed with dipeptide libraries. All three dipeptidases hydrolyzed a large fraction of the dipeptide substrates within the libraries but there were subtle differences in their N- or C-terminal specificities. Their main commonality with respect to substrate specificity was the general preference for an L-amino acid at the N-terminus and a D-amino acid at the C-terminus of the substrate.

The N-terminal specificity of LmoDP varied from Rsp0802 and Bh2271 in that LmoDP preferred the positively charged residues L-Arg and L-Lys at the N-terminus. Rsp0802 and Bh2271 were more similar in their N-terminal preference in that they were less specific than LmoDP and showed only a minor preference for the hydrophobic L-amino acids. With respect to the preference at the C-terminus of the substrate, LmoDP and Bh2271 were more similar and hydrolyzed dipeptides with a C-terminal D-Ala faster than other components of L-Xaa-D-Xaa libraries. This was in contrast to Rsp0802 which had a preference for hydrophobic D-amino acids at the C-terminus of the dipeptide substrates.

Substrate Specificity Determinants of Rsp0802. In following the trend observed for Sco3058, residues that follow β_1 (Leu-16) and β_2 (Tyr-56) contribute to the pocket for the N-terminal L-amino acid side chain. The side chains of multiple residues immediately after β_7 (Ala-264, Ser-266 and Phe-267) face into the active site pocket for the C-terminal amino acid. Two residues which contribute to the pockets for both the N- and C-terminal side chains of a dipeptide are located within insertion region 1 (His-63

and Met-72). Also, the loop following β_4 in Rsp0802 is longer than in Sco3058 and adds Phe-176 to the substrate pocket.

The structure of Rsp0802 helps explain the poor *affinity* of the *A. calcoaceticus* dipeptidase (AcDP) for the dipeptide L-Leu-D-Leu. As shown in the alignment in **Figure 5.13**, the residue which follows β_7 (Ala-264 in Rsp0802 and Tyr-225 in LmoDP) and contributes to the C-terminal D-amino acid side chain pocket, is an aspartate in AcDP (Asp-259 yellow and underlined in the alignment). The methyl group of Ala-264 is 4.6 Å from the methyl group of **7** which is representative of the side chain of D-Ala at the C-termini of a dipeptide. The k_{cat}, K_m and k_{cat}/K_m values for AcDP with Leu-D-Leu were 0.6 s^{-1}, 9.3 mM and 66 $M^{-1}s^{-1}$, respectively. All of the L-D dipeptides tested with AcDP had a C-terminal D-Ala, D-Leu or D-Phe. Given the same N-terminal glycine, the order of specific activity was Gly-D-Ala ≈ Gly-D-Leu > Gly-D-Phe (*76*). The poor *turnover* of AcDP may have been due to protein damage as AcDP does not appear to be missing any critical residues.

Substrate Specificity Determinants of LmoDP. The surface of LmoDP within ~ 10 Å of the two representative C_α carbons of **8** is shown in **Figure**s **5.11A** and **B**. A majority of the pocket immediately surrounding the N-terminal C_α are the side chains of Asp-51, the hydrophobic residues Phe-49, Trp-132 and the aromatic ring of Tyr-12. The C_α–C_β/C_β–C_γ/C_γ–C_δ bonds of L-Lys were overlayed onto the equivalent bonds of the mimic of an L-Leu side chain from **8**. This model of L-Lys at the N-terminus reveals that there are rotomers of lysine which allow the amino group of the L-Lys side chain to come within hydrogen-bonding distance to the phenolic oxygen of Tyr-12 (2.8 Å) and

the side chain carboxylate of Asp-51 (2.6 Å). The presence of the acidic Asp-51 residue at the perimeter of the active site and the hydrophobic nature of the N-terminal side chain pocket should not be conducive to allowing in dipeptides with an aspartate or glutamate at the N-terminus.

There is another example of a microbial dipeptidase in the literature which seems to have a preference for dipeptides with an N-terminal L-Lys. This enzyme was isolated from *B. sphaericus*, but the protein sequence was not determined. The enzyme had been assayed with a small set of dipeptides at single substrate concentrations of either 0.4 or 2 mM. The dipeptides L-Lys-Gly and L-Lys-D-Glu were hydrolyzed faster than the dipeptides L-Ala-Gly, L-Leu-Gly and L-Ala-D-Glu. When the C-terminal amino acid was changed to D-Glu, an N-terminal L-Lys was preferred over L-Ala by 4-fold (*92*). The genome sequence of *B. sphaericus* (*Lysinibacillus sphaericus C3-41*) has now been sequenced. When a BLAST search of the LmoDP amino acid sequence was conducted against the *B. sphaericus* genome, a hit to the protein Bsph4282 was obtained. These two proteins share 45 % amino acid sequence identity. The trend of the N-terminal amino acid preference in the order of L-Ala < L-Leu < L-Lys is similar to that seen in **Figure 5.3D-E**. Coincidentally, Bsph4282 also has the equivalent residue to Asp-51 in LmoDP. The difference between the activity of LmoDP and the dipeptidase from *B. sphaericus* was the lack of L-Xaa-L-Xaa hydrolysis activity for the *B. sphaericus* enzyme.

The reason for the rejection of dipeptides with an N-terminal isoleucine, threonine or valine by LmoDP is not inherently obvious. The space directly about C_β

may be just a little small to accommodate an R-group off of this atom. It proved difficult to find a conformation of *N*-L-Ile, *N*-L-Val- or *N*-L-Thr where the side chains did not clash with the backbone of the inhibitor or with one or more of the residues Trp-132, Phe-49, Asp-9 or Gly-264.

The binding pocket of LmoDP is shallow and more exposed to the solvent in comparison to that of Rsp0802, likely owing to the absence of insertion region 1. Though dipeptides with an N-terminal L-Lys were hydrolyzed the fastest, it is not surprising that the preference was not overwhelming. The Asp-51 residue following β_2 is the residue which is most likely the source of the preference for L-Lys and L-Arg at the N-terminus of the substrate. Ile-265, and especially Tyr-225, cause the pocket for the C-terminal D-amino acid side chain to be small and hydrophobic; L-Xaa-D-Ala is preferred over other L-Xaa-D-Xaas dipeptides.

Substrate Specificity Determinants of Bh2271. An overlay of Rsp0802, LmoDP and a hypothetical active site of Bh2271 (based on the structure of Lmo2462) is shown in **Figure 5.12**. It is not surprising that Bh2271 and Rsp0802 behaved similarly with respect to their preference of the N-terminal amino acid side chain of a dipeptide. The ordered loop in Rsp0802 (residues 61-65) does close off the active site around the N-terminal L-amino acid side chain, but the position that this loop takes could vary for larger amino acid side chains. The N-terminal side chain pocket of Bh2271 should be more open and neutral because it has Gly-50 following β_2 (Gly-50) instead of Asp-51 in LmoDP and as well as Leu-16 instead of Tyr-12 in LmoDP.

The L-Leu-Gly substrate was preferred over any L-Leu-L-Xaa substrate with L-Leu-L-Ala being the next best substrate. The reason for this is not clear. In LmoDP, the aspartate from the opposing chain faced into the active site, but was too far to be considered a potential substrate-contacting residue for dipeptides with the standard 20 amino acids. Perhaps a similar phenomenon occurs with Bh2271, but with a greater impact on the substrate binding pocket.

The residues expected to surround the D-amino acid side chain of the C-terminal amino acid are the equivalent residues to Ile-265/Phe-228 in LmoDP and His-224 (Tyr-225 in LmoDP). The substrate specificity with respect to the C-terminal amino acid would be similar since the pocket is small. Given the presence of the His-224 residue in Bh2271 it was a little surprising that L-Leu-D-Ser was not preferred over L-Leu-D-Ala in the screen or with the individual dipeptide; His-224 appears to be reasonably positioned to be within hydrogen bonding distance of a D-Ser side chain.

$Trp_{\beta 4}$ vs. $His_{\beta 4}$ and β_2-Tyr vs. β_2-Phe. From the values in **Tables 5.1** and **5.2**, Rsp0802 with a β_2-Tyr (Tyr-56) does not appear to be a better dipeptidase than LmoDP which has a β_2-Phe (Phe-49). The kinetic parameters shown in **Tables 5.1** and **5.2** reveal that k_{cat}/K_m values for LmoDP and Rsp0802 (on the order of 10^5 $M^{-1}s^{-1}$) are on a par with Sco3058, Gox2272 and Cc2746, despite the fact that they have a $Trp_{\beta 4}$ instead of $His_{\beta 4}$. The pH profiles hRDP and Sco3058 have both shown an increase in activity up to pH ~ 7.5-8 and a decrease in activity above pH ~7.5-8.5 (*80*, *93*). The decrease in activity could be due to the deprotonation of the free amino group of the substrate or the $His_{\beta 4}$ residue. In contrast to hRDP and Sco3058, the BCS-1 enzyme with a $Trp_{\beta 4}$ residue

was described as having optimal activity in the range of pH 8.5-10 (*49*). If activity loss at higher pH is due to the deprotonation of $His_{\beta4}$, it may be more beneficial to have a tryptophan at this position in order to broaden the pH optima of the enzyme. In fact, $Trp_{\beta4}$ is more common among the microbial members of cog2355 than $His_{\beta4}$.

Genome Context of cog2355-Cluster 2. The more strict N-terminal specificity of LmoDP for L-Lys/Arg is thought to be due to Asp-51. However, there was not a common factor among the genome context of LmoDP and other proteins throughout cog2355 with an aspartate at this position.

Genome Context of cog2355-Cluster 3. There are 45 amino acid sequences in Cluster 3 of cog2355 (**Figure 3.8**). The genes adjacent to members of cluster 3 in the microbial genomes are commonly annotated as transporters of dipeptides, peptides, oligopeptides, opines and amino acids. The only known biosynthetic pathway associated with proteins from Cluster 3 is for the biosynthesis of the cofactor pyrroloquinoline quinone (PQQ). The biosynthesis of PQQ has been described to involve up to six genes, namely *pqqABCDEF*. The synthesis of PQQ begins with C—C bond formation between C_γ of an L-glutamate and C_ε of an L-Tyr side chain; all of the carbon nitrogen atoms of PQQ come from glutamate and tyrosine (*94*).

The glutamate and tyrosine residues are present as part of a peptide, coded for by the *pqqA* gene, which is ~ 23-40 amino acids long, and three amino acids link the glutamate and tyrosine. However, pqqA is not required for PQQ synthesis in *Methylobacterium extorquens* and examination of the genome contexts of organisms which putatively have this gene cluster revealed that the *pqqA* gene is sometimes slightly

separated from the rest of the cluster or no gene in the genome is annotated as *pqqA* (*95*). *E. coli* is not a PQQ-producing organism, but it can synthesize PQQ when supplied with a plasmid containing the *pqqABCDE* operon genes from *Gluconobacter oxydans* 621H (*96*). Inspection of the genome contexts of 43 proteins within Cluster 3 of cog2355, showed that nine members were near, and more often than not, tightly clustered with the *pqq* genes.

In the *Rhodobacter sphaeroides* species, there are 8-9 genes between the *pqq* genes and the cog2355 dipeptidase, but there is not a protein annotated as pqqF. Similarly, in *Mycobacterium smegmatis* and *Geodermatophilus obscurus*, there are 4-5 genes between the cog2355 dipeptidase and the *pqqABCDE* cluster (no *pqqF*). It has been suggested that the pqqF protein is a peptidase which functions to hydrolyze the rest of the pqqA oligopeptide from the fused Tyr-Glu residues. Using the amino acid sequence from the *pqqC* gene of *Klebsiella pneumoniae* as a seed sequence, the *pqq* genes in *M. smegmatis* and *G. obscurus* were identified as having a *nearby*, not directly adjacent, members of cog2355 which were not part of Cluster 3, but instead were members of cluster 14. The enzyme from cog2355 may function as general dipeptidase as part of the recycling of the unused pqqA peptide, provided the peptide has been broken down into a dipeptide unit.

CHAPTER VI

CONCLUSIONS

SUMMARY

Enzymes from two of the 24 Cogs which encompass the amidohydrolase superfamily, namely cog2355 and cog3653, were selected for this work. These enzymes were broadly annotated as D-aminoacylases (DAAs) and dipeptidases. The amino acid sequences within each cog were clustered into groups with a BLAST E-value cut-off of 1×10^{-70} (**Figures 2.11** and **3.8**). With the exception of the two dipeptidases LmoDP and Bh2271, the protein targets were from different groups.

The desire was to obtain purified proteins, determine their substrate specificity and structure with bound inhibitors, and identify the substrate specificity determinants. In order to determine the substrate profiles, the proteins were screened for activity against libraries of *N*-acetyl-L-, *N*-acetyl-D-, *N*-succinyl-L- and *N*-succinyl-D- derivatives of amino acids as well as L-L, L-D and D-L dipeptide libraries. Using HPLC, the library components could be separated and quantitated to compare the rates of hydrolysis of the compounds within the libraries. Substrates with k_{cat}/K_m values on the order of $10^6 M^{-1}s^{-1}$ (*N*-acyl-D-Glu) and 10^4 $M^{-1}s^{-1}$ (four *N*-acetyl-D-amino acids) were identified for the enzymes Bb3285 and Gox1177, respectively. Dipeptide substrates with k_{cat}/K_m values of $\sim 10^4$-10^5 $M^{-1}s^{-1}$ were found for Sco3058, Cc2746, Gox2272, LmoDP, Rsp0802 and Bh2271. In collaboration with Steve Almo's lab at Albert Einstein College of Medicine

and NYSGXRC, the structures of one member of cog3653 and five members of cog2355, some in complex with phosphinate inhibitors, were solved.

D-AMINOACYLASES IN cog3653

Summary. The selected proteins from cog3653 were Bb3285, Sco4986, Gox1177, Bb2785 and Bll7304. When the amidohydrolase superfamily sequences were updated and organized by clusters with an E-value cut-off of 1×10^{-70}, the five proteins were in clusters 1 (Bb3285), 2 (Sco4986), 4 (Bb2785) and 5 (Gox1177 and Bll7304). Three of these proteins were purified and Sco4986 was partially purified. Bll7304 did not overexpress. The substrate specificity profiles of Bb3285 and Gox1177 were determined using the *N*-acyl- and *N*-L-Xaa-D-Xaa libraries. The substrate profile of Sco4986 was attained by comparing the rates of hydrolysis of *N*-acetyl-D-amino acids. The results indicate that Bb2785 is not a D-aminoacylase.

Gox1177 and Sco4986 both hydrolyze several *N*-acetyl-D-amino acids with a preference for hydrophobic side chains, whereas, Bb3285 exclusively hydrolyzes *N*-acyl- and *N*-L-Xaa- derivatives of D-Glutamate. The reason for the strict specificity of Bb3285 was clarified by the structure of the Bb3285 enzyme complexed with the *N*-methylphosphonate derivative of D-glutamate (**1**). The Bb3285 structure also revealed that the substrate specificity loop was in a different conformation from the same loop in the DAA from *A. faecalis* (*36, 56*). The substrate specificity loop is part of a larger insertion region located between β_7 and α_7 of the $(\beta/\alpha)_8$-barrel. As shown in the amino acid sequence alignment in **Figure 2.1**, the greatest variation in these sequences occurs in the region that contains this insertion domain. In the case of cog3653, the sequence

lengths varied considerably (~ 480-700 amino acids). The structural coverage of cog3653 is still limited to proteins from Cluster 1 (**Figure 2.11**).

Future Directions. If one was to continue with cog3653 in order to get the full structural coverage spanning Clusters 2-6, there might be some things to consider. With regard to Cluster 2 (target: Sco4986), the issue was obtaining the soluble protein to purify from *E. coli*. Sco4986 contains 6 cysteine residues. Perhaps it would be better to choose targets from cog3653 which have as few cysteines as possible. For example there are five putative D-aminoacylases from Cluster 2 within the *Gemmatimonas aurantiaca* genome. One of these (gi|226226566) does not have a cysteine residue while the DAA from *G. aurantiaca* that is closest to Sco4986 (gi|226227749) has only one cysteine.

It is possible that cysteine residues contribute to the problem of insoluble protein. The Bb3285, Gox1177 and Bb2785 also contained several cysteines and were purified, but Bb3285 was not soluble on the first purification attempt. However, each protein is unique. The problem of insoluble recombinant protein was also encountered with the microbial dipeptidases from cog2355 as well. There is an example in the literature where selected cysteine residues in human renal dipeptidase were mutated in order to increase the amount of soluble recombinant protein expressed in *E. coli* (*97*). NYSGXRC also attempted to purify Sco4986 but failed at the stage of expression (their vectors and growth conditions are likely both different from mine). The number of potential protein targets has increased dramatically as a result of the boom in microbial genome sequencing projects; the choices are not as limited as they were in the past.

MICROBIAL HOMOLOGUES TO THE HUMAN RENAL DIPEPTIDASE

*Summary***.** Six microbial homologues to the human renal dipeptidase were overexpressed and purified by myself (Sco3058, Gox2272 and Cc2746) or NYSGXRC (LmoDP, Rsp0802 and Bh2271). The substrate profiles were determined by screening the enzymes for activity with L-L, L-D and D-L dipeptide libraries. The enzymes exhibited subtle differences in substrate specificity though LmoDP and Sco3058 were more specific than the other four proteins. LmoDP exhibited a higher degree of specificity for L-Lys-D/L-Xaa and L-Arg-D/L-Xaa dipeptides, which was attributed to a tyrosine or aspartate following β-strands 1 and 2, respectively. The structure of the Sco3058·**4** complex suggests that the enhanced preference for L-Xaa-D/L-Asp dipeptides by Sco3058 is due to a threonine (Thr-324) and possibly a lysine residue (Lys-247). The preference of Rsp0802 for dipeptides with a hydrophobic residue at the C-terminus could also be explained by the hydrophobic pocket which harbors the C-terminal amino acid of the substrate. The structures of inhibitor-bound complexes of Gox2272 and Cc2746 were not solved during the timeframe of this work.

Future Directions. Included in this work are dipeptidases from the larger subset of cog2355. However, there is a smaller subset which includes those proteins in Clusters 4, 7, 11, and 13 (**Figure 3.8**). With the exception of two sequences in these clusters (gi|148556484 and gi|192362060), the predicted proteins in this subset do not have the predicted ligands to the alpha metal from β-strand 1(absence of the HxD motif). Proteins in this subset mainly have an LxY, LxI, LxF, HxT or RxT in lieu of the HxD motif. Two *Pseudomonas aeruginosa* proteins within this smaller subset of cog2355,

namely Pa5396 from cluster 7 (PDB code *2I5G*) and Pa2393 from cluster 13 (PDB code *3B40*), have structures deposited in the PDB.

Pa5396 and Pa2393 have a glutamine (clusters 4, 7 and 11) or leucine (cluster 13) instead of the bridging glutamate seen in the larger subset. The tryptophan/histidine from β_4, which contacts the phosphinate inhibitors, is replaced by a tyrosine in clusters 4, 7, 13, and likely a histidine or tyrosine in cluster 11. Neither Pa2393 nor Pa5396 had metal bound in the active site. The only metal ligand conserved in Pa2393 is the histidine from β_6, and Pa5396 has the two histidines from β_5 and β_6. When the glutamate which bridges the two metals in hRDP (E125) was mutated to glutamine, the enzyme retained 11% of the wild-type activity. The E125D mutant had 0.1 % the activity of the wild-type hRDP (*98*). The metal content of the wild-type E125Q/E125D mutants was not presented, but perhaps the glutamine was able to participate as ligand to at least the beta metal. Pa5396 may be able to bind metal, and perhaps the addition of metal in the medium during high levels of overexpression would allow result in a protein with metal in the active site.

Pa2393 has the β_8-loop and β_7-loop helix which contributes residues to the binding pocket of the C-terminal half of a dipeptide. The loop following β_8 in Pa5396 differs from the other seven structures. Instead of forming a portion of the substrate-binding pocket with residues on the β_8-loop, the residues following β_8 encapsulate the active site. A ribbon representation of all eight crystallized members of cog2355 is

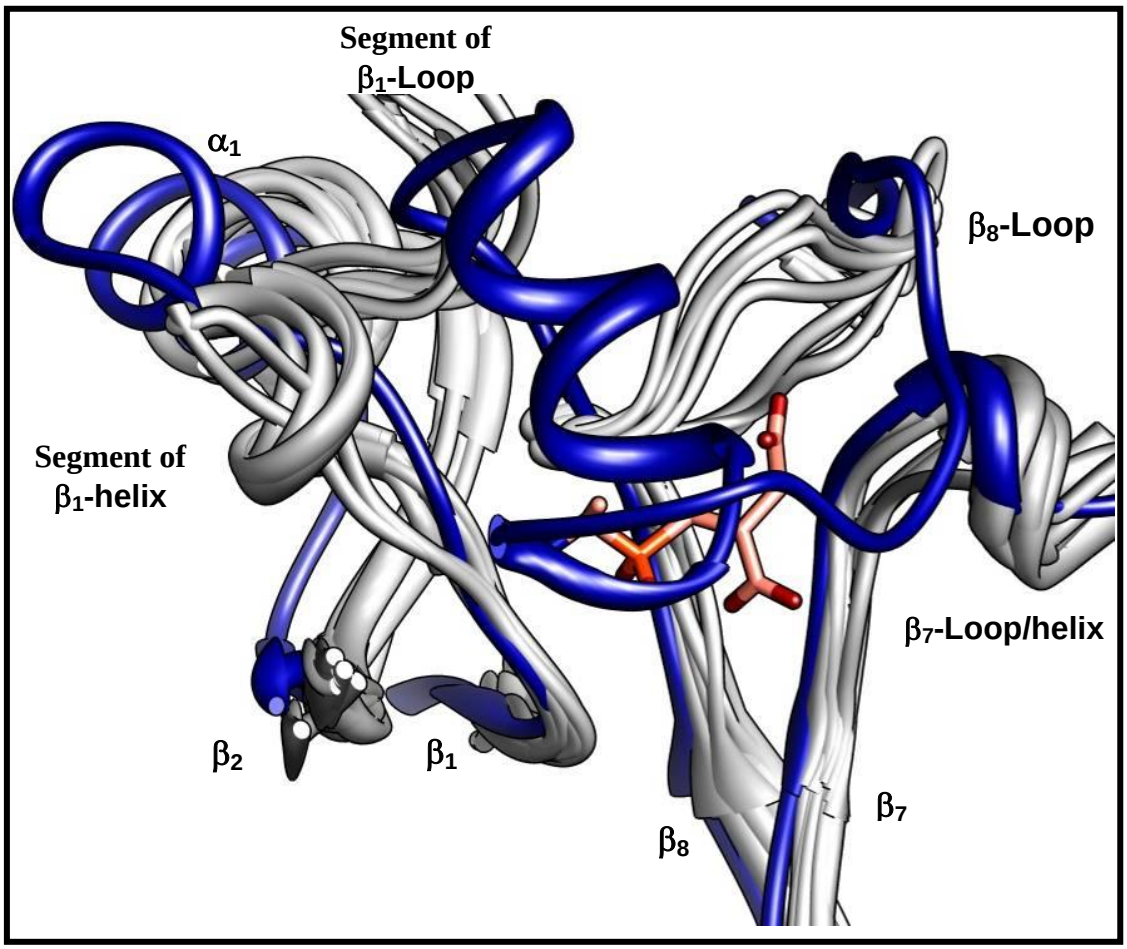


Figure 6.1: Segments of superimposed structures of the eight cog2355 proteins hRDP, Sco3058, Gox2272, Cc2746, LmoDP, Rsp0802, Pa2393 and Pa5396. Ribbons are colored *gray* for each protein except Pa5396 (*dark blue*). The phosphinate mimic from the Sco3058·**4** complex is shown with *pink* carbons and an *orange* phosphorus.

shown in **Figure 6.1**. This figure is similar to the one shown in **Figure 4.9**, but with Pa5396 added. These two proteins could be candidates for *in silico* screening with metabolites or a thermofluor assay with L- and D- amino acids. Since the active site of Pa5396 is encapsulated, it may be considered dockable for *in silico* screening. The knowledge of the structure-specificity relationship of the dipeptides described in Chapters II-V can be used to curate the docking results. The substrates for proteins within clusters 4, 7, 11 and 13 may contain an amino acid as part of the substrate.

Since their structures are already known, Pa5396 and Pa2393 would make good templates for potentially interesting rational protein evolution experiments. Perhaps by changing select residues in these two proteins, they could be transformed into dipeptidases with metal bound.

GENOME CONTEXT

A detailed search through some of the genome contexts of the microbial renal dipeptidase-like proteins within cog2355 reveals that these proteins are sometimes found near clusters of genes involved in identified or putative biosynthetic pathways. At this point in time there is no experimental evidence, at least of which I am aware, that would lead one to conjecture that the renal dipeptidase-like proteins are required in a biosynthetic pathway. However, it is hard to imagine that it is purely coincidental that the genes of cog2355 proteins are sometimes found clustered with genes that are known or predicted to be involved in biosynthetic pathways that use small peptides or oligopeptides as precursor molecules (e.g. sirodesmin, gliotoxin, pyrroloquinoline quinone) (*87*, *88*, *94*, *96*). The Pa5396 gene is clustered with genes required for the

catabolism of glycine betaine (*99*). Pa2393 gene is also called pvdM because it is located adjacent to several genes required for the biosynthesis of the siderophore pyoverdine (*100*). Pyoverdine is a derivatized oligopeptide composed of L- and D-amino acids. The enzymes in cog2355 may play a role in the recycling of unused peptide precursors. These dipeptidases are also commonly clustered tightly with proteins annotated as transporters or peptide/dipeptide transporters or D-amino acid deaminases, as was the case for some of the D-aminoacylases in Cluster 2 of cog3653.

REFERENCES

1. May, O., Siemann, M., Pietzsch, M., Kiess, M., Mattes, R., and Syldatk, C. (1998) Substrate-dependent enantioselectivity of a novel hydantoinase from *Arthrobacter aurescens* DSM 3745: purification and characterization as new member of cyclic amidases, *J. Biotechnol. 61*, 1-13.

2. Mulrooney, S. B., and Hausinger, R. P. (2003) Metal ion dependence of recombinant *Escherichia coli* allantoinase, *J. Bacteriol*. *185*, 126-134.

3. Spoonamore, J. E., and Bandarian, V. (2008) Understanding functional divergence in proteins by studying intragenomic homologues, *Biochemistry 47*, 2592-2600.

4. Marti-Arbona, R., Xu, C., Steele, S., Weeks, A., Kuty, G. F., Seibert, C. M., and Raushel, F. M. (2006) Annotating enzymes of unknown function: *N*-formimino-L-glutamate deiminase is a member of the amidohydrolase superfamily, *Biochemistry 45*, 1997-2005.

5. Nguyen, T. T., Brown, S., Fedorov, A. A., Fedorov, E. V., Babbitt, P. C., Almo, S. C., and Raushel, F. M. (2008) At the periphery of the amidohydrolase superfamily: Bh0493 from *Bacillus halodurans* catalyzes the isomerization of D-galacturonate to D-tagaturonate, *Biochemistry 47*, 1194-1206.

6. Toms, A. V., Haas, A. L., Park, J. H., Begley, T. P., and Ealick, S. E. (2005) Structural characterization of the regulatory proteins TenA and TenI from *Bacillus subtilis* and identification of TenA as a thiaminase II, *Biochemistry 44*, 2319-2329.

7. Jenkins, A. H., Schyns, G., Potot, S., Sun, G., and Begley, T. P. (2007) A new thiamin salvage pathway, *Nat. Chem. Biol. 3*, 492-497.

8. Hermann, J. C., Ghanem, E., Li, Y., Raushel, F. M., Irwin, J. J., and Shoichet, B. K. (2006) Predicting substrates by docking high-energy intermediates to enzyme structures, *J. Am. Chem. Soc. 128*, 15882-15891.

9. Song, L., Kalyanaraman, C., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Brown, S., Imker, H. J., Babbitt, P. C., Almo, S. C., Jacobson, M. P., and Gerlt, J. A. (2007) Prediction and assignment of function for a divergent *N*-succinyl amino acid racemase, *Nat. Chem. Biol. 3*, 486-491.

10. Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K., and Raushel, F. M. (2007) Structure-based activity prediction for an enzyme of unknown function, *Nature 448*, 775-779.

11. Yew, W. S., Fedorov, A. A., Fedorov, E. V., Rakus, J. F., Pierce, R. W., Almo, S. C., and Gerlt, J. A. (2006) Evolution of enzymatic activities in the enolase superfamily: L-fuconate dehydratase from *Xanthomonas campestris*, *Biochemistry 45*, 14582-14597.

12. Yew, W. S., Fedorov, A. A., Fedorov, E. V., Wood, B. M., Almo, S. C., and Gerlt, J. A. (2006) Evolution of enzymatic activities in the enolase superfamily: D-tartrate dehydratase from *Bradyrhizobium japonicum*, *Biochemistry 45*, 14598-14608.

13. Pieper, U., Chiang, R., Seffernick, J. J., Brown, S. D., Glasner, M. E., Kelly, L., Eswar, N., Sauder, J. M., Bonanno, J. B., Swaminathan, S., Burley, S. K., Zheng, X., Chance, M. R., Almo, S. C., Gerlt, J. A., Raushel, F. M., Jacobson, M. P., Babbitt, P. C., and Sali, A. (2009) Target selection and annotation for the structural genomics of the amidohydrolase and enolase superfamilies, *J. Struct. Funct. Genomics 10*, 107-125.

14. Holm, L., and Sander, C. (1997) An evolutionary treasure: unification of a broad set of amidohydrolases related to urease, *Proteins 28*, 72-82.

15. Ireton, G. C., McDermott, G., Black, M. E., and Stoddard, B. L. (2002) The structure of *Escherichia coli* cytosine deaminase, *J. Mol. Biol. 315*, 687-697.

16. Cheon, Y. H., Kim, H. S., Han, K. H., Abendroth, J., Niefind, K., Schomburg, D., Wang, J., and Kim, Y. (2002) Crystal structure of D-hydantoinase from *Bacillus stearothermophilus*: insight into the stereochemistry of enantioselectivity, *Biochemistry 41*, 9410-9417.

17. Abendroth, J., Niefind, K., May, O., Siemann, M., Syldatk, C., and Schomburg, D. (2002) The structure of L-hydantoinase from *Arthobacter aurescens* leads to an understanding of dihydropyrimidinase substrate and enantio specificity, *Biochemistry 41*, 8589-8597.

18. Tyagi, R., Kumaran, D., Burley, S. K., and Swaminathan, S. (2007) X-ray structure of imidazolonepropionase from *Agrobacterium tumefaciens* at 1.87 A resolution, *Proteins 69*, 652-658.

19. Thoden, J. B., Phillips, G. N., Jr., Neal, T. M., Raushel, F. M., and Holden, H. M. (2001) Molecular structure of dihydroorotase: a paradigm for catalysis through the use of a binuclear metal center, *Biochemistry 40*, 6989-6997.

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

21. 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.

22. Li, T., Iwaki, H., Fu, R., Hasegawa, Y., Zhang, H., and Liu, A. (2006) Alpha-amino-beta-carboxymuconic-epsilon-semialdehyde decarboxylase (ACMSD) is a new member of the amidohydrolase superfamily, *Biochemistry 45*, 6628-6634.

23. Nitanai, Y., Satow, Y., Adachi, H., and Tsujimoto, M. (2002) Crystal structure of human renal dipeptidase involved in beta-lactam hydrolysis, *J. Mol. Biol. 321*, 177-184.

24. Lai, W. L., Chou, L. Y., Ting, C. Y., Kirby, R., Tsai, Y. C., Wang, A. H., and Liaw, S. H. (2004) The functional role of the binuclear metal center in D-aminoacylase: one-metal activation and second-metal attenuation, *J. Biol. Chem. 279*, 13962-13967.

25. Kim, G. J., Lee, D. E., and Kim, H. S. (2000) Functional expression and characterization of the two cyclic amidohydrolase enzymes, allantoinase and a novel phenylhydantoinase, from *Escherichia coli*, *J. Bacteriol. 182*, 7021-7028.

26. Vincent, F., Yates, D., Garman, E., Davies, G. J., and Brannigan, J. A. (2004) The three-dimensional structure of the *N*-acetylglucosamine-6-phosphate deacetylase, NagA, from *Bacillus subtilis*: a member of the urease superfamily, *J. Biol. Chem. 279*, 2809-2816.

27. Hall, R. S., Brown, S., Fedorov, A. A., Fedorov, E. V., Xu, C., Babbitt, P. C., Almo, S. C., and Raushel, F. M. (2007) Structural diversity within the mononuclear and binuclear active sites of *N*-acetyl-D-glucosamine-6-phosphate deacetylase, *Biochemistry 46*, 7953-7962.

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

29. Benning, M. M., Hong, S. B., Raushel, F. M., and Holden, H. M. (2000) The binding of substrate analogs to phosphotriesterase, *J. Biol. Chem. 275*, 30556-30560.

30. Thoden, J. B., Marti-Arbona, R., Raushel, F. M., and Holden, H. M. (2003) High-resolution X-ray structure of isoaspartyl dipeptidase from *Escherichia coli*, *Biochemistry 42*, 4874-4882.

31. Nguyen, T. T., Fedorov, A. A., Williams, L., Fedorov, E. V., Li, Y., Xu, C., Almo, S. C., and Raushel, F. M. (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*, 8879-8890.

32. Pegg, S. C., Brown, S. D., Ojha, S., Seffernick, J., Meng, E. C., Morris, J. H., Chang, P. J., Huang, C. C., Ferrin, T. E., and Babbitt, P. C. (2006) Leveraging enzyme structure-function relationships for functional inference and experimental design: the structure-function linkage database, *Biochemistry 45*, 2545-2555.

33. Hall, R. S., Fedorov, A. A., Marti-Arbona, R., Fedorov, E. V., Kolb, P., Sauder, J. M., Burley, S. K., Shoichet, B. K., Almo, S. C., and Raushel, F. M. (2010) The hunt for 8-oxoguanine deaminase, *J. Am. Chem. Soc. 132*, 1762-1763.

34. Moriguchi, M., Sakai, K., Miyamoto, Y., and Wakayama, M. (1993) Production, purification, and characterization of D-aminoacylase from *Alcaligenes xylosoxydans* subsp. *xylosoxydans* A-6, *Biosci. Biotechnol. Biochem. 57*, 1149-1152.

35. Chen, H. P., Wu, S. H., and Wang, K. T. (1994) D-Aminoacylase from *Alcaligenes faecalis* possesses novel activities on D-methionine, *Bioorg. Med. Chem. 2*, 1-5.

36. Liaw, S. H., Chen, S. J., Ko, T. P., Hsu, C. S., Chen, C. J., Wang, A. H., and Tsai, Y. C. (2003) Crystal structure of D-aminoacylase from *Alcaligenes faecalis* DA1. A novel subset of amidohydrolases and insights into the enzyme mechanism, *J. Biol. Chem. 278*, 4957-4962.

37. Moriguchi, M., Sakai, K., Katsuno, Y., Maki, T., and Wakayama, M. (1993) Purification and characterization of novel *N*-acyl-D-aspartate amidohydrolase from *Alcaligenes xylosoxydans* subsp. *xylosoxydans* A-6, *Biosci. Biotechnol. Biochem. 57*, 1145-1148.

38. Sakai, K., Imamura, K., Sonoda, Y., Kido, H., and Moriguchi, M. (1991) Purification and characterization of *N*-acyl-D-glutamate deacylase from *Alcaligenes xylosoxydans* subsp. *xylosoxydans* A-6, *FEBS Lett. 289*, 44-46.

39. Sakai, K., Oshima, K., and Moriguchi, M. (1991) Production and characterization of *N*-acyl-D-glutamate amidohydrolase from *Pseudomonas* sp. strain 5f-1, *Appl. Environ. Microbiol. 57*, 2540-2543.

40. Lin, P. H., Su, S. C., Tsai, Y. C., and Lee, C. Y. (2002) Identification and characterization of a new gene from *Variovorax paradoxus* Iso1 encoding *N*-acyl-D-amino acid amidohydrolase responsible for D-amino acid production, *Eur. J. Biochem. 269*, 4868-4878.

41. Wakayama, M., Kitahata, S., Manoch, L., Tachiki, T., Yoshimune, K., and Moriguchi, M. (2004) Production, purification and properties of D-aminoacylase from a newly isolated *Trichoderma* sp. SKW-36, *Process Bioch. 39*, 1119-1124.

42. Kumagai, S., Kobayashi, M., Yamaguchi, S., Kanaya, T., Motohashi, R., and Isobe, K. (2004) A new D-aminoacylase from *Defluvibacter* sp. A 131-3, *J. Mol. Catal. B: Enzym. 30*, 159-165.

43. Sugie, M., and Suzuki, H. (1978) Purification and properties of D-aminoacylase of *Streptomyces olivaceus*, *Agric. Biol. Chem. 42*, 107-113.

44. Wakayama, M., Watanabe, E., Takenaka, Y., Miyamoto, Y., Tau, Y., Sakai, K., and Moriguchi, M. (1995) Cloning, Expression, and nucleotide sequence of the *N*-acyl-D-aspartate amidohydrolase gene from *Alcaligenes xylosoxydans* subsp. *xylosoxydans* A-6, *J. Ferment. Bioeng. 80*, 311-317.

45. Sakai, A., Xiang, D. F., Xu, C., Song, L., Yew, W. S., Raushel, F. M., and Gerlt, J. A. (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*, 4455-4462.

46. Priestman, D. A., and Butterworth, J. (1985) Prolinase and non-specific dipeptidase of human kidney, *Biochem J 231*, 689-694.

47. Sugiura, M., Ito, Y., Hirano, K., and Sawaki, S. (1978) Purification and properties of human kidney dipeptidases, *Biochim. Biophys. Acta 522*, 541-550.

48. Campbell, B. J., Di Shih, Y., Forrester, L. J., and Zahler, W. L. (1988) Specificity and inhibition studies of human renal dipeptidase, *Biochim. Biophys. Acta 956*, 110-118.

49. Baek, D. H., Song, J. J., Kwon, S. J., Park, C., Jung, C. M., and Sung, M. H. (2004) Characteristics of a new enantioselective thermostable dipeptidase from *Brevibacillus borstelensis* BCS-1 and its application to synthesis of a D-amino-acid-containing dipeptide, *Appl. Environ. Microbiol. 70*, 1570-1575.

50. Friedberg, I., Jambon, M., and Godzik, A. (2006) New avenues in protein function prediction, *Protein Sci. 15*, 1527-1529.

51. Roodveldt, C., and Tawfik, D. S. (2005) Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily, *Biochemistry 44*, 12728-12736.

52. Muniz-Lozano, F. E., Dominguez-Sanchez, G., Diaz-Viveros, Y., and Barradas-Dermitz, D. M. (1998) D-Aminoacylase from a novel producer: *Stenotrophomonas maltophilia* ITV-0595, *J. Ind. Microbiol. & Biotechnol. 21*, 296-299.

53. Yang, H., Zheng, G., Peng, X., Qiang, B., and Yuan, J. (2003) D-Amino acids and D-Tyr-tRNA(Tyr) deacylase: stereospecificity of the translation machine revisited, *FEBS Lett. 552*, 95-98.

54. Soutourina, O., Soutourina, J., Blanquet, S., and Plateau, P. (2004) Formation of D-tyrosyl-tRNATyr accounts for the toxicity of D-tyrosine toward *Escherichia coli*, *J. Biol. Chem. 279*, 42560-42565.

55. Baltz, R. H., Miao, V., and Wrigley, S. K. (2005) Natural products to drugs: daptomycin and related lipopeptide antibiotics, *Nat. Prod. Rep. 22*, 717-741.

56. Cummings, J. A., Fedorov, A. A., Xu, C., Brown, S., Fedorov, E., Babbitt, P. C., Almo, S. C., and Raushel, F. M. (2009) Annotating enzymes of uncertain function: the deacylation of D-amino acids by members of the amidohydrolase superfamily, *Biochemistry 48*, 6469-6481.

57. Xu, C., Hall, R., Cummings, J., and Raushel, F. M. (2006) Tight binding inhibitors of *N*-acyl amino sugar and N-acyl amino acid deacetylases, *J. Am. Chem. Soc. 128*, 4244-4245.

58. Zhang, P., Hao, Z., and Li, Y. (2002) Synthesis and steric structure of α-amino-β-lactam serivative of 1,5-benzothiazepines, *Chem. J. Chin. Univ. 23*, 2101-2105.

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

60. Castanie, M. P., Berges, H., Oreglia, J., Prere, M. F., and Fayet, O. (1997) A set of pBR322-compatible plasmids allowing the testing of chaperone-assisted folding of proteins overexpressed in *Escherichia coli*, *Anal. Biochem. 254*, 150-152.

61. Doi, E., Shibata, D., and Matoba, T. (1981) Modified colorimetric ninhydrin methods for peptidase assay, *Anal. Biochem. 118*, 173-184.

62. Lazennec, C., and Meinnel, T. (1997) Formate dehydrogenase-coupled spectrophotometric assay of peptide deformylase, *Anal. Biochem. 244*, 180-182.

63. Copeland, R. A. (2005) *Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists*, pp 111-213, John Wiley & Sons, Inc., Hoboken, NJ.

64. Edgar, R. C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity, *BMC Bioinformatics 5*, 113.

65. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res. 13*, 2498-2504.

66. Gross, H., Stockwell, V. O., Henkels, M. D., Nowak-Thompson, B., Loper, J. E., and Gerwick, W. H. (2007) The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters, *Chem. Biol. 14*, 53-63.

67. Makovitzki, A., Avrahami, D., and Shai, Y. (2006) Ultrashort antibacterial and antifungal lipopeptides, *Proc. Natl. Acad. Sci. U. S. A. 103*, 15997-16002.

68. Adachi, H., Tawaragi, Y., Inuzuka, C., Kubota, I., Tsujimoto, M., Nishihara, T., and Nakazato, H. (1990) Primary structure of human microsomal dipeptidase deduced from molecular cloning, *J. Biol. Chem. 265*, 3992-3995.

69. Adachi, H., Kubota, I., Okamura, N., Iwata, H., Tsujimoto, M., Nakazato, H., Nishihara, T., and Noguchi, T. (1989) Purification and characterization of human microsomal dipeptidase, *J. Biochem. 105*, 957-961.

70. Kozak, E. M., and Tate, S. S. (1982) Glutathione-degrading enzymes of microvillus membranes, *J. Biol. Chem. 257*, 6322-6327.

71. McIntyre, T., and Curthoys, N. P. (1982) Renal catabolism of glutathione. Characterization of a particulate rat renal dipeptidase that catalyzes the hydrolysis of cysteinylglycine, *J. Biol. Chem. 257*, 11915-11921.

72. Satoh, S., Keida, Y., Konta, Y., Maeda, M., Matsumoto, Y., Niwa, M., and Kohsaka, M. (1993) Purification and molecular cloning of mouse renal dipeptidase, *Biochim. Biophys. Acta 1163*, 234-242.

73. Watanabe, T., Kera, Y., Matsumoto, T., and Yamada, R. H. (1996) Purification and kinetic properties of a D-amino-acid peptide hydrolyzing enzyme from pig kidney cortex and its tentative identification with renal membrane dipeptidase, *Biochim. Biophys. Acta 1298*, 109-118.

74. Campbell, B. J., Forrester, L. J., Zahler, W. L., and Burks, M. (1984) Beta-lactamase activity of purified and partially characterized human renal dipeptidase, *J. Biol. Chem. 259*, 14586-14590.

75. Park, S. W., We, J. S., Kim, G. W., Choi, S. H., and Park, H. S. (2002) Stability of new carbapenem DA-1131 to renal dipeptidase (dehydropeptidase I), *Antimicrob. Agents Chemother. 46*, 575-577.

76. Adachi, H., and Tsujimoto, M. (1995) Cloning and expression of dipeptidase from *Acinetobacter calcoaceticus* ATCC 23055, *J. Biochem. 118*, 555-561.

77. Keynan, S., Hooper, N. M., and Turner, A. J. (1997) Identification by site-directed mutagenesis of three essential histidine residues in membrane dipeptidase, a novel mammalian zinc peptidase, *Biochem. J. 326 (Pt 1)*, 47-51.

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

79. Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J., and Hopwood, D. A. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), *Nature 417*, 141-147.

80. Cummings, J. A., Nguyen, T. T., Fedorov, A. A., Kolb, P., Xu, C., Fedorov, E. V., Shoichet, B. K., Barondeau, D. P., Almo, S. C., and Raushel, F. M. (2010) Structure, mechanism, and substrate profile for Sco3058: the closest bacterial homologue to human renal dipeptidase, *Biochemistry 49*, 611-622.

81. Xiang, D. F., Patskovsky, Y., Xu, C., Meyer, A. J., Sauder, J. M., Burley, S. K., Almo, S. C., and Raushel, F. M. (2009) Functional identification of incorrectly annotated prolidases from the amidohydrolase superfamily of enzymes, *Biochemistry 48*, 3730-3742.

82. Georgiadis, D., Matziari, M., and Yiotakis, A. (2001) A highly efficient method for the preparation of phosphinic pseudodipeptidic blocks suitably protected for solid-phase peptide synthesis, *Tetrahedron 57*, 3471-3478.

83. Buchardt, J., and Meldal, M. (2000) Novel methodology for the solid-phase synthesis of phosphinic peptides, *J. Chem. Soc. Perkin Trans. 1*, 3306-3310.

84. Baylis, E. K., Campbell, C. D., and Dingwall, J. G. (1984) 1-aminoalkylphosphonous acids. Part 1. lsosteres of the protein amino acids, *J. Chem. Soc. Perkin Trans. 1*, 2845-2853.

85. Liboska, R., Picha, J., Hanclova, I., Budesinsky, M., Sanda, M., and Jiracek, J. (2008) Synthesis of methionine- and norleucine-derived phosphinopeptides, *Tetrahedron Lett. 49*, 5629–5631.

86. Firestine, S. M., Poon, S. W., Mueller, E. J., Stubbe, J., and Davisson, V. J. (1994) Reactions catalyzed by 5-aminoimidazole ribonucleotide carboxylases from *Escherichia coli* and *Gallus gallus*: a case for divergent catalytic mechanisms, *Biochemistry 33*, 11927-11934.

87. Gardiner, D. M., Cozijnsen, A. J., Wilson, L. M., Pedras, M. S., and Howlett, B. J. (2004) The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans*, *Mol. Microbiol*. *53*, 1307-1318.

88. Gardiner, D. M., and Howlett, B. J. (2005) Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of *Aspergillus fumigatus*, *FEMS Microbiol. Lett*. *248*, 241-248.

89. Kremer, A., and Li, S. M. (2010) A tyrosine O-prenyltransferase catalyses the first pathway-specific step in the biosynthesis of sirodesmin PL, *Microbiology 156*, 278-286.

90. Belshaw, P. J., Walsh, C. T., and Stachelhaus, T. (1999) Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis, *Science 284*, 486-489.

91. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis, *J. Comput. Chem. 25*, 1605-1612.

92. Ciroussel, F., Vacheron, M.-J., Guinand, M., and Michel, G. (1990) Purification and properties of an LD-dipeptidase from *Bacillus sphaericus*, *Int. J. Biochem. 22*, 525.

93. Ferren, L. G., Ward, R. L., and Campbell, B. J. (1975) Monoanion inhibition and 35Cl nuclear magnetic resonance studies of renal dipeptidase, *Biochemistry 14*, 5280-5285.

94. Magnusson, O. T., Toyama, H., Saeki, M., Rojas, A., Reed, J. C., Liddington, R. C., Klinman, J. P., and Schwarzenbacher, R. (2004) Quinone biogenesis: structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone, *Proc. Natl. Acad. Sci. U. S. A. 101*, 7913-7918.

95. Toyama, H., and Lidstrom, M. E. (1998) pqqA is not required for biosynthesis of pyrroloquinoline quinone in *Methylobacterium extorquens* AM1, *Microbiology 144 (Pt 1)*, 183-191.

96. Yang, X. P., Zhong, G. F., Lin, J. P., Mao, D. B., and Wei, D. Z. (2010) Pyrroloquinoline quinone biosynthesis in *Escherichia coli* through expression of the *Gluconobacter oxydans* pqqABCDE gene cluster, *J. Ind. Microbiol. Biotechnol. 37*, 575-580.

97. O'Dwyer, R., Razzaque, R., Hu, X., Hollingshead, S. K., and Wall, J. G. (2009) Engineering of cysteine residues leads to improved production of a human dipeptidase enzyme in *E. coli*, *Appl. Biochem. Biotechnol. 159*, 178-190.

98. Adachi, H., Katayama, T., Nakazato, H., and Tsujimoto, M. (1993) Importance of Glu-125 in the catalytic activity of human renal dipeptidase, *Biochim. Biophys. Acta 1163*, 42-48.

99. Wargo, M. J., Szwergold, B. S., and Hogan, D. A. (2008) Identification of two gene clusters and a transcriptional regulator required for *Pseudomonas aeruginosa* glycine betaine catabolism, *J. Bacteriol. 190*, 2690-2699.

100. Lamont, I. L., and Martin, L. W. (2003) Identification and characterization of novel pyoverdine synthesis genes in *Pseudomonas aeruginosa*, *Microbiology 149*, 833-842.

VITA

Name:	Jennifer Ann Cummings
Address:	P. O. Box 30012 College Station, TX 77843
Email Address:	Jen_Cummings28@yahoo.com
Education:	B.S., Chemistry, Southern Oregon University, 2001 M.S., Chemistry, Texas A&M University, 2005