MECHANISTIC STUDIES AND FUNCTION DISCOVERY OF MONONUCLEAR AMIDOHYDROLASE ENZYMES

A Dissertation

by

RICHARD STUART HALL

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 2009

Major Subject: Chemistry

MECHANISTIC STUDIES AND FUNCTION DISCOVERY OF MONONUCLEAR AMIDOHYDROLASE ENZYMES

A Dissertation

by

RICHARD STUART HALL

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 M. Raushel
Committee Members,	Paul Lindahl
	David Barondeau
	Gregory Reinhart
Head of Department,	David Russell

December 2009

Major Subject: Chemistry

ABSTRACT

Mechanistic Studies and Function Discovery of Mononuclear Amidohydrolase Enzymes.

(December 2009)

Richard Stuart Hall, B.S., The University of Arizona

Chair of Advisory Committee: Dr. Frank M. Raushel

The amidohydrolase superfamily is a functionally diverse group of evolutionarily related proteins which utilize metal cofactors in the activation of a hydrolytic water molecule and in the stabilization of the resulting tetrahedral intermediate. Members of this superfamily have been described which use one or two divalent transition metals. These metal cofactors are located in either or both of two active-site metal binding centers which are labeled as the M_α and M_β sites. The goal of this research was to elucidate the nature of the reactions catalyzed by M_α and M_β mononuclear members of the amidohydrolase superfamily. This was approached through comprehensive mechanistic evaluations of two enzymes which utilized the different metal sites. *N*-acetyl-D-glucosamine-6-phosphate deacetylase from *E. coli* (NagA) and cytosine deaminase from *E. coli* (CDA) served as models for mononuclear amidohydrolase superfamily enzymes which have evolved to utilize a single β-metal and a single α-metal for hydrolysis, respectively. This research elucidated the different properties imparted by the distinct α and β active sites and the specific interactions utilized by the enzymes for substrate binding and catalysis. These studies led to the eventual proposal of detailed

chemical mechanisms and the identification of rate determining steps. Knowledge of sequence-function relationships was applied toward the discovery of function for enzymes related to cytosine deaminase and guanine deaminase. The first group of enzymes investigated was proposed to catalyze the fourth step in riboflavin and coenzyme F420 biosynthesis in Achaea. Three putative deaminases; Mm0823 from *Methanosarcina mazei*, MmarC7_0625 from *Methanococcus maripaludis C7* and Sso0398 from *Sulfolobus solfataricus* were cloned and expressed. These proteins proved to be intractably insoluble. A second set of enzymes, Pa0142 from *Pseudomonas aeruginosa PA01* and SGX-9236e (with crystal structure PDB: 3HPA) were found to catalyze the novel deamination of 8-oxoguanine, a mutagenic product of DNA oxidation. 9236e was cloned from an unidentified environmental sample of the Sargasso Sea. The closest homolog (98% identical) is Bcep18194_A5267 from *Burkholderia sp. 383*. Additionally, it was discovered that the proteins SGX-9339a (with crystal structure PDB: 2PAJ) and SGX-9236b catalyzed the deamination of isoxanthopterin and pterin-6-carboxylate in a poorly characterized folate degradation pathway. These enzymes were also from unknown environmental samples of the Sargasso Sea. The closest homolog of 9339a (88% identical) is Bxe_A2016 from *Burkholderia xenovorans LB400*. The closest homolog of 9236b (95% identical) is Bphyt_7136 from *Burkholderia phytofirmans PsJN*.

DEDICATION

To my beautiful wife and best friend, Gwynette, you inspire me daily with your love, kindness, patience and strength, you are what I am not, and yet we are one; to my beautiful daughter Claire, you are a shooting star full of love and life, your brilliance lights the world around you; and to my son Lyon, you are like a lion, strong and brave, yet sensitive and caring. Thank you for making life wonderful, you teach me who I am and who I want to be.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Frank M. Raushel for his guidance, support and for everything he taught me. I would also like to thank my committee members Dr. Paul Lindahl, Dr. Gregory Reinhart and Dr. David Barondeau for their insight, advice and for encouraging me to find my own “Big Picture”. I am extremely appreciative to all of my colleagues from Frank Raushel’s lab over the past five years; for their assistance and tremendous friendship, especially Ricardo Marti-Arbona for being a great mentor, colleague and friend. I am indebted to my parents, Sharon and Lawrence Hall for teaching me that I can do anything I put my mind to. I would like to express my thanks to my wife Gwynette, for all of her support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION: ENZYMES AND THE AMIDOHYDROLASE SUPERFAMILY

Enzymes play an essential role in nature by catalyzing the transformation of energy and materials into a variety of useful forms. At first glance, enzymes can be viewed as mere protein. Fundamentally, enzymes are folded polymers made up of twenty different amino acids which may have acidic, basic, hydrophobic, hydrophilic, positively charged or negatively charged side chains. The true utility of enzymes is evident in the reactions they catalyze, often at rates only limited by diffusion. For example, the Cu, Zn superoxide dismutase has a reported catalytic efficiency of 2×10^9 $M^{-1}s^{-1}$ (*1, 2*). The extraordinary feat of catalytic power that can be exerted by a protein is best demonstrated with the decarboxylation of uroporphyrinogen III by human urophorphyrinogen decarboxylase (UroD). In this example a rate enhancement of over 1.2×10^{17} was measured, compared to the extrapolated un-catalyzed reaction that occurs in water (*3*). Astonishingly even greater rate enhancements may be achieved by enzymes through the use of bound metal cofactors for the activation of hydrolytic water as seen in type I protein phosphatase, fructose bisphosphatase and inositol phosphatase, where rate enhancements on the order of 10^{21} were determined (*4*). Even with diffusion limited reaction rates, enzymes are able to exert exquisite substrate specificity as well as stereospecific and stereoselective control over the chemical transformations they catalyze. These molecular machines make it possible for life to exist, providing a steady

---

This dissertation follows the style of *Biochemistry*.

stream of the energetic and structural compounds necessary for survival and reproduction.

Enzymology has been studied since the 19th century, when Louis Pasteur discovered that fermentation could be performed outside of living cells. In 1926 James Sumner purified and crystallized the first protein, jack bean urease, establishing for the first time that enzymes are well defined chemical compounds (*5*). Since the discovery of enzymes, there has always been a great deal of research and even debate over exactly how enzymes are able to perform these feats of chemistry. It was proposed by Linus Pauling in 1946 that for catalysis to occur, enzymes must have a higher affinity for the transition state of the substrate than for the ground state (*6*). If an enzyme was to bind either the substrate or the product with any great affinity, then a stable complex would result, and catalysis might occur slowly or not at all.

ENZYME CHARACTERIZATION

Database Information. Since the beginning of the human genome project in 1990, there has been an ever growing amount of gene sequence information filling the databases. According to the National Center for Biotechnology Information (NCBI) as of May 2009, there are 2538 bacterial, 108 archaeal, 149 animal, 114 fungal, 55 plant, and 52 protist genomes that have either been completely sequenced, or are currently being sequenced. Meanwhile, protein chemists are utilizing this information for the selection, production and purification of many gene products. According to the RCSB Protein Data Bank (PDB), there are 57424 protein structures posted as of May 05, 2009. 230 of those structures have been solved by electron microscopy, 7836 were solved by

NMR and 49232 structures were solved by X-ray diffraction. The result is the identification of 1283 distinct folds as defined by Structural Classification of Proteins (SCOP). Of the available structures, many proteins still do not have a functional assignment. According to the PSI Structural Genomics Knowledgebase, there are 1491 unique proteins of unknown function with available structures that were published by four of the associated research centers. Undoubtedly there are many more proteins with a known crystal structure that still have no verified function. Even with all of this knowledge, there is an even greater amount of biology still left to be discovered. The presence of incorrectly annotated enzymes in the genomic databases creates an additional layer of complexity for researchers trying to discover the function of unknown enzymes.

Function Prediction. Modern technology is unable to reliably predict protein structure without a high degree of sequence similarity to enzymes whose structures have already been determined. It is more difficult to assign function to those enzymes with no known structure, although the identification of a highly conserved cluster of enzymes related to a specific pathway may lead toward the proposal of a function. Of the proteins with known structure, it has been recently shown that functions may be proposed through the docking of high energy transition state structures to the active site of a high resolution crystal structure as was recently reported for the discovery of function for Tm0936 as an S-adenosylhomocysteine deaminase (*7*). Even with the increasingly rapid progress in computer technology for structural docking and sequence analysis, experimental verification of substrate specificity and mechanistic evaluation of active

site residues will remain an essential part of the discovery process for the foreseeable future. It is these experimental results which give bioinformatics analyses a basis for the evaluation and organization of the ever accumulating genetic information.

Superfamilies. Proteins which perform similar reactions may be grouped by sequence and structural homologies into families, while more divergent groups of related enzymes which still share some aspects of function may be grouped into superfamilies (*8*). Currently there are a total of 2033 protein superfamilies reported by the PDB. Of these, there are currently 33 superfamilies identified by SCOP which share the common TIM barrel fold (http://scop.mrc-lmb.cam.ac.uk/). Enzymes which are labeled as members of the metallo-dependent hydrolases superfamily (amidohydrolase superfamily) share the common TIM barrel fold. It is typically a $(\beta/\alpha)_8$ barrel fold comprised of eight parallel β-strands which are connected by eight α-helices. Recent discoveries have shown the presence of a $(\beta/\alpha)_7$ barrel fold as seen for histidinol phosphate phosphatase (*9*). Flexible loops connecting the carboxyl ends of the β-strands to the α-helices provide an inherent flexibility in the enzyme architecture which is capable of providing substrate binding as seen for adenosine deaminase (*10*). The conservation of several metal binding histidine residues and an aspartate provide key landmarks for the identification of enzymes within this superfamily. However, the absolute dependence of the different mechanisms on divalent transition metal cations appears to be the most salient unifying feature. Amidohydrolase enzymes are found in every organism sequenced to date and are thought to have evolved through divergent evolution from a common ancestral precursor (*10*). It also has been proposed that the

ancestral precursor may have actually evolved from the combination of two separate $(\beta/\alpha)_4$ half-barrels (*11*).

The amidohydrolase superfamily was originally identified by Holm and Sander in 1997 based on the conserved structural fold and several catalytically important residues possessed by phosphotriesterase, urease and adenosine deaminase (*12*). Currently, there are 10,045 unique sequences in the NCBI database which are labeled as members of the amidohydrolase superfamily. Out of these sequences, 41 distinct reactions have been characterized and are shown in Scheme 1.1. The proteins have been grouped into clusters of orthologous groups (COGs) which are based on consistent patterns of sequence identities, organizing them into families which are evolutionarily related. 4873 different COGs have been identified from the comparisons of conserved proteins among prokaryotic genomes and 4852 different KOGs have been identified from orthologous eukaryotic groups (*13, 14*). The COG assignments for 66 genomes are available at http://www.ncbi.nlm.nih.gov/COG/. The annotation of genes in newly sequenced genomes may be aided through comparison to the existing COG database by using the COGnitor sequence comparison tool available at the same site.

Divergent Activity within the AHS. Adenine deaminase (ADE) catalyzes the deamination of adenine to hypoxanthine (*15*) and is a member of COG 1001F (Adenine Deaminase). Adenosine deaminase (ADA) (*16, 17*) and adenosine monophosphate deaminase (AMPD) (*18*) catalyze the formation of inosine and inosine monophosphate and belong to COG 1816 F (Adenosine deaminase). Cytosine deaminase (CDA) catalyzes the deamination of cytosine to uracil (*19-23*), and belongs to COG 0402 FR

(Cytosine deaminase and related metal-dependent hydrolases). This COG also contains, guanine deaminase (GDA) (*24*), S-adenosylhomocysteine deaminase (SAHD) (*7*), *N*-formimino-L-glutamate deiminase (HutF) (*25*), atrazine chlorohydrolase (AtzA) and melamine deaminase (TriA) (*26*), s-triazine hydrolase (TrzA) (*27*), the broad specificity triazine hydrolase (TrzN) (*28*), hydroxyatrazine ethylaminohydrolase (AtzB) (*29*) and *N*-isopropylammelide isopropylaminohydrolase (AtzC) (*30*).

Dihydroorotase (DHO) is the sole member of COG 0148 F and reversibly catalyzes the interconversion of carbamoyl aspartate to dihydroorotate in the pyrimidine biosynthesis pathway (*31-33*). Dihydropyrimidinase (DHP) and the other members of COG 0044F (dihydroorotase and related cyclic amidohydrolases) are related to DHO. DHP catalyzes the hydrolysis of dihydropyrimidines to 3-ureido acids in the reductive catabolism of pyrimidines (*34*). Hydantoinase (HYD) catalyzes the hydrolysis of 5' substituted hydantoins to carbamoyl amino acids (*35, 36*). Allantoinase (ALN) belongs in COG 0044 F along with DHP and HYD. ALN is responsible for the conversion of S-allantoin into allantoate in the ureide pathway (*37*). Isoaspartyl dipeptidase (IAD) catalyzes the hydrolysis of beta-L-isoaspartyl linkages in dipeptides (*38-40*). IAD is unique in that it has an ambiguous COG assignment, aligning with either COG 0044 F or COG 1228 Q (imidazolone propionase and related amidohydrolases). Other members of COG 1228 Q include enamidase (ENA) (*41*), hydrophobic carboxypeptidase (Cc0300) (*42*), Lys/Arg carboxypeptidase (Cc3125) (*43*), and imidazolone propionase (HutI) (*44*).

Human renal dipeptidase (RDP) hydrolyzes dipeptides and is involved in the metabolism of penem and carbapenem beta-lactam antibiotics (*45*). RDP is associated

with COG 2335 E (Zn-dependent dipeptidase, microsomal dependent homolog). Lactonase (Dr0930) belongs to COG 1735 R (predicted metal-dependent hydrolase with the TIM-barrel fold) and catalyzes the hydrolysis of δ- and γ-lactones possessing alkyl substitutions on the carbon adjacent to the ring oxygen (*46*). SsoPox is a promiscuouos hyperthermophilic phosphotriesterase of COG 1735R which is found to have very high and possibly natural quorum-quenching homoserine lactonase activity (*47, 48*). Phosphotriesterase (PTE) catalyzes the hydrolysis of organophosphorus compounds (*49-51*) and also belongs to COG 1735 R. Histidinol phosphate phosphatase (HisJ) of COG 1387 (histidinol phosphatase and related hydrolases of the PHP family) catalyzes the de-phosphorylation of L-histidinol phosphate in histidine biosynthesis, forming histidinol (*9, 52*).

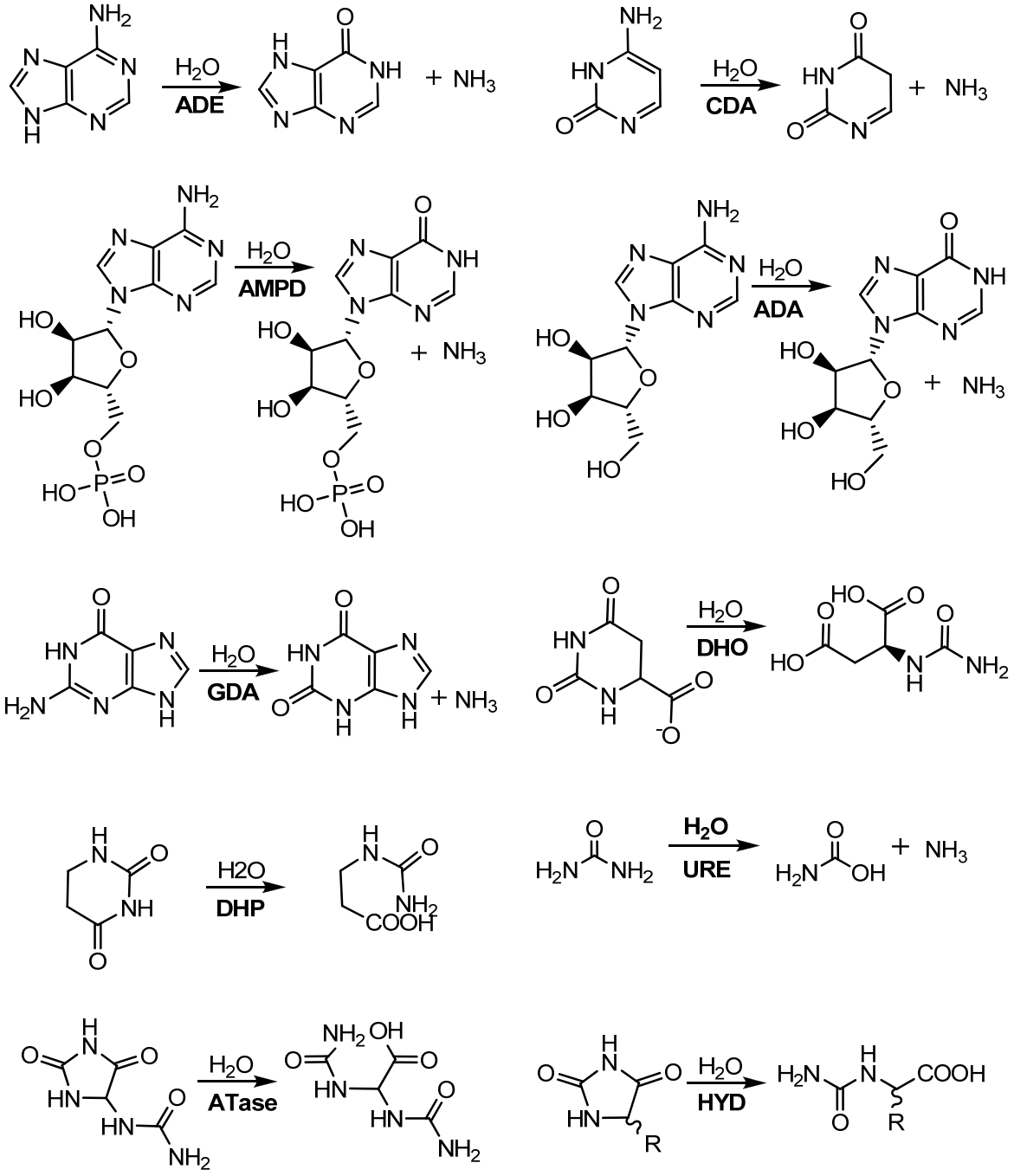
The enzyme called *meta*-cleavage compound hydrolase (LigY) is involved in the degradation of the *meta*-cleavage compound of 2, 2',3-trihydroxy-3'-methoxy-5,5'-dicarboxybiphenyl (OH-DDVA), forming 5-carboxyvanillic acid (5CVA) (*53*), and is associated with COG 2159 R (predicted metal-dependent hydrolase with the TIM-barrel fold). Several non-oxidative decarboxylases were recently discovered, also belonging to COG 2159 R which do not catalyze hydrolysis reactions. Those enzymes are 5-carboxyvannilate decarboxylase (LigW), which catalyzes the formation of vannilate from 5CVA (*54*); gamma-resorcylate decarboxylase (Rdc), which catalyzes the reversible carboxylation of resorcinol 1,3-dihydroxybenzene (*55*); alpha-amino-beta-carboxymuconate-epsilon-semialdehyde decarboxylase (ACMSD), which produces 2-aminomuconate semialdehyde (*56*); and finally uracil-5-carboxylate decarboxylase

(IDCase), which catalyzes the formation of uracil from uracil-5-carboxylate (iso-orotate) in the thymine salvage pathway (*57*).

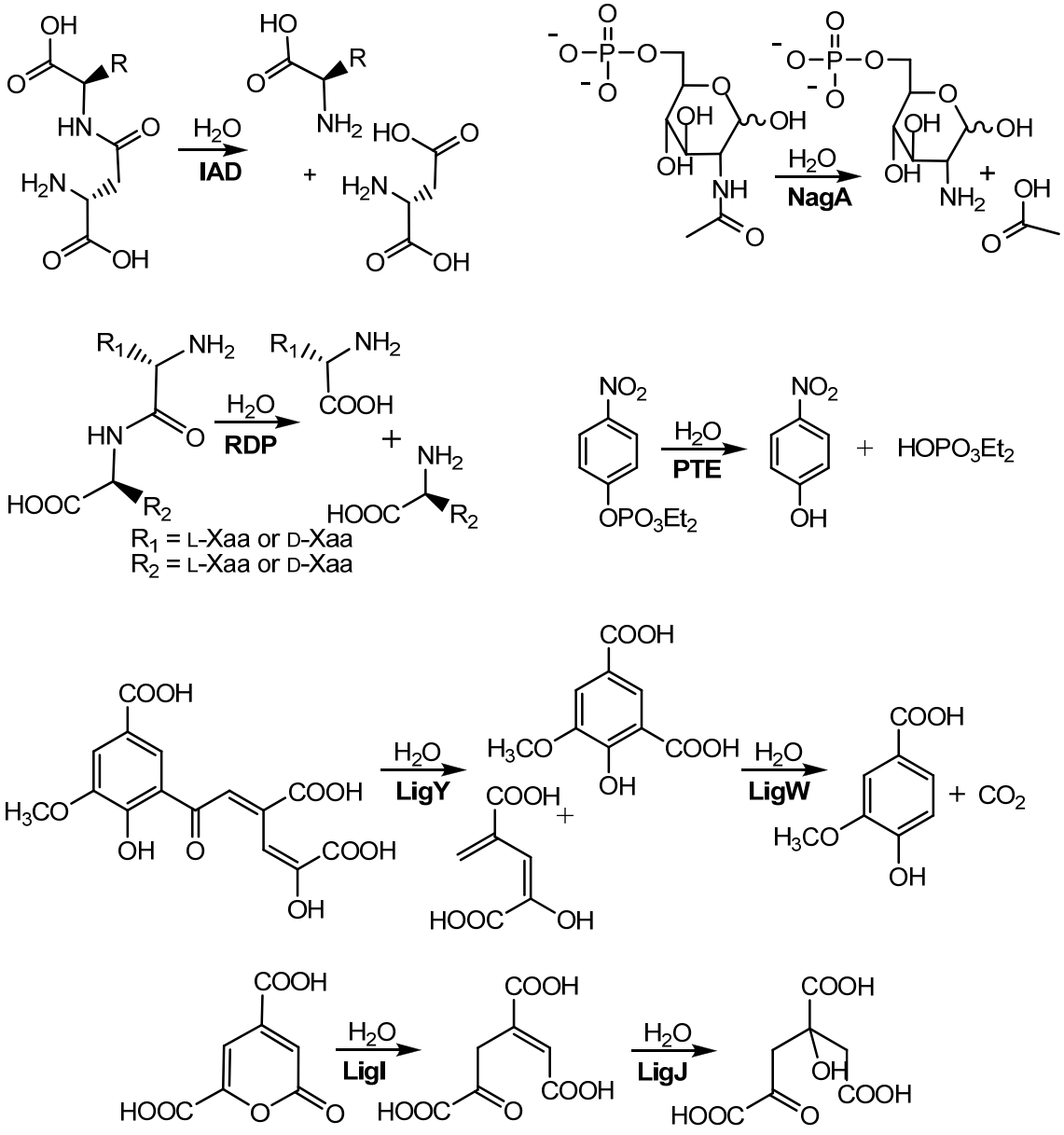
Urease (URE) is the sole representative of COG 0804 E (Urea amidohydrolase (urease) alpha subunit) and catalyzes the conversion of urea to CO_2 and 2 NH_3 molecules (*58-61*). Uronate isomerase (URI) catalyzes the isomerization of D-glucuronate to D-fructuronate and of D-galacturonate to D-tagaturonate (*62-64*). The uronate isomerase family belongs to COG 1904 G (Glucuronate isomerase) and is unique in that it is the only family within the amidohydrolase superfamily found to catalyze an isomerization reaction.

N-acetyl-D-glucosamine-6-phosphate deacetylase (NagA) catalyzes the deacetylation of N-acetylglucosamine-6-phosphate and N-acetylgalactosamine-6-phosphate in the aminosugar metabolism (*65-71*). Uniquely, NagA is the only member of COG 1820 G (N-acetylglucosamine-6-phosphate deacetylase), while all of the other identified deacetylases co-occupy COG 3653 Q (N-acyl-D-aspartate/D-glutamate deacylase). The corresponding amino acid deacetylases are *N*-acyl-D-neutral deacetylase (D-ANase) (*72*), *N*-acyl-D-methionine deacetylase (DAA) (*73*), *N*-acyl-D-glutamate deacetylase (D-AGase) (*72*), and *N*-acyl-D-aspartate deacetylase (D-AAase) (*74*). 4-oxalomesaconate hydratase (LigJ) catalyzes the conversion of 4-oxalomesaconate to 4-carboxy-4-hydroxy-2-oxo adipate (*75*) and is quite possibly the most unusual member of the amidohydrolase superfamily; because according to a sequence comparison with COGnitor, there are no COGs of related sequence.

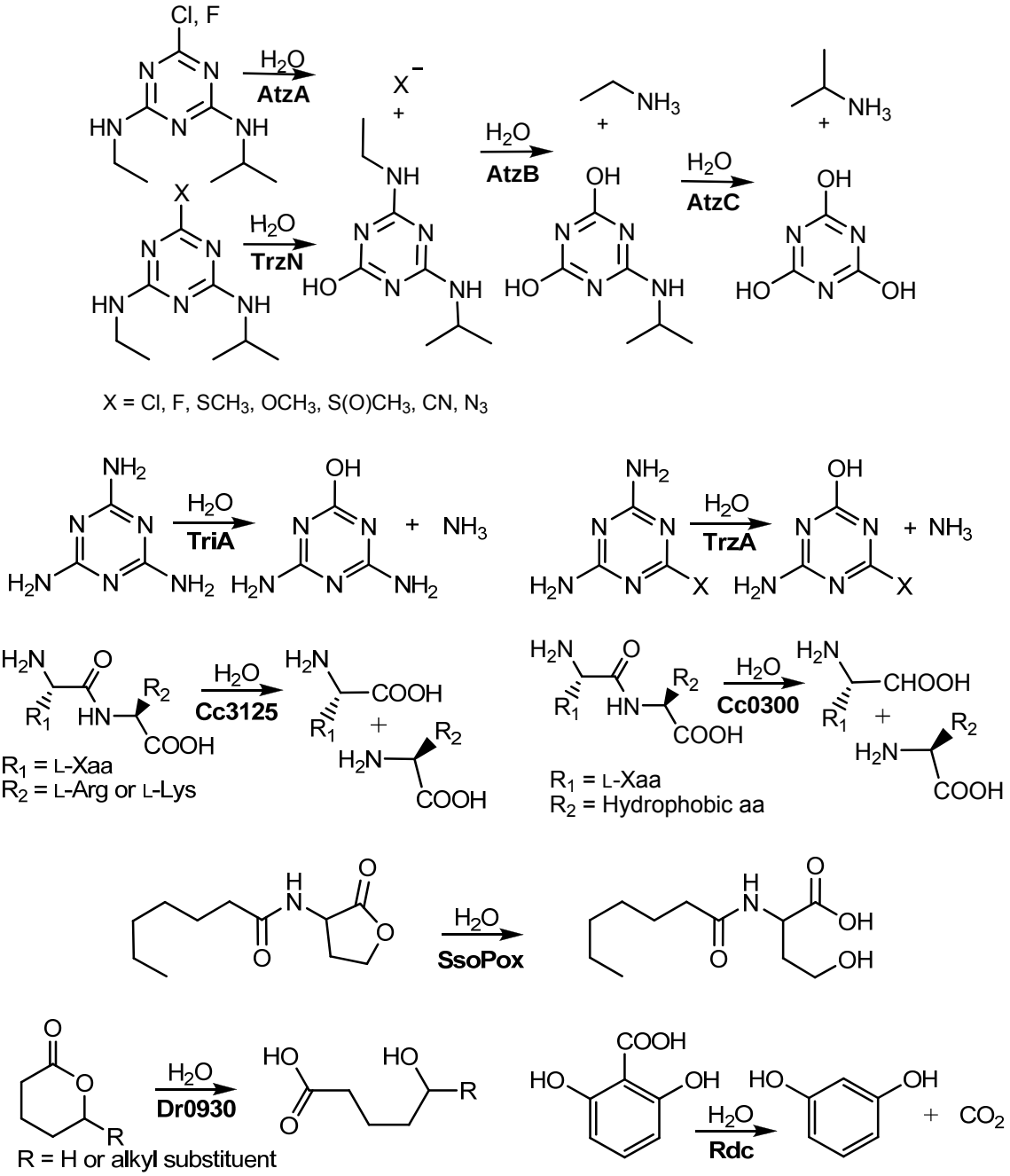
Scheme 1.1: Reactions catalyzed by members of the amidohydrolase superfamily.



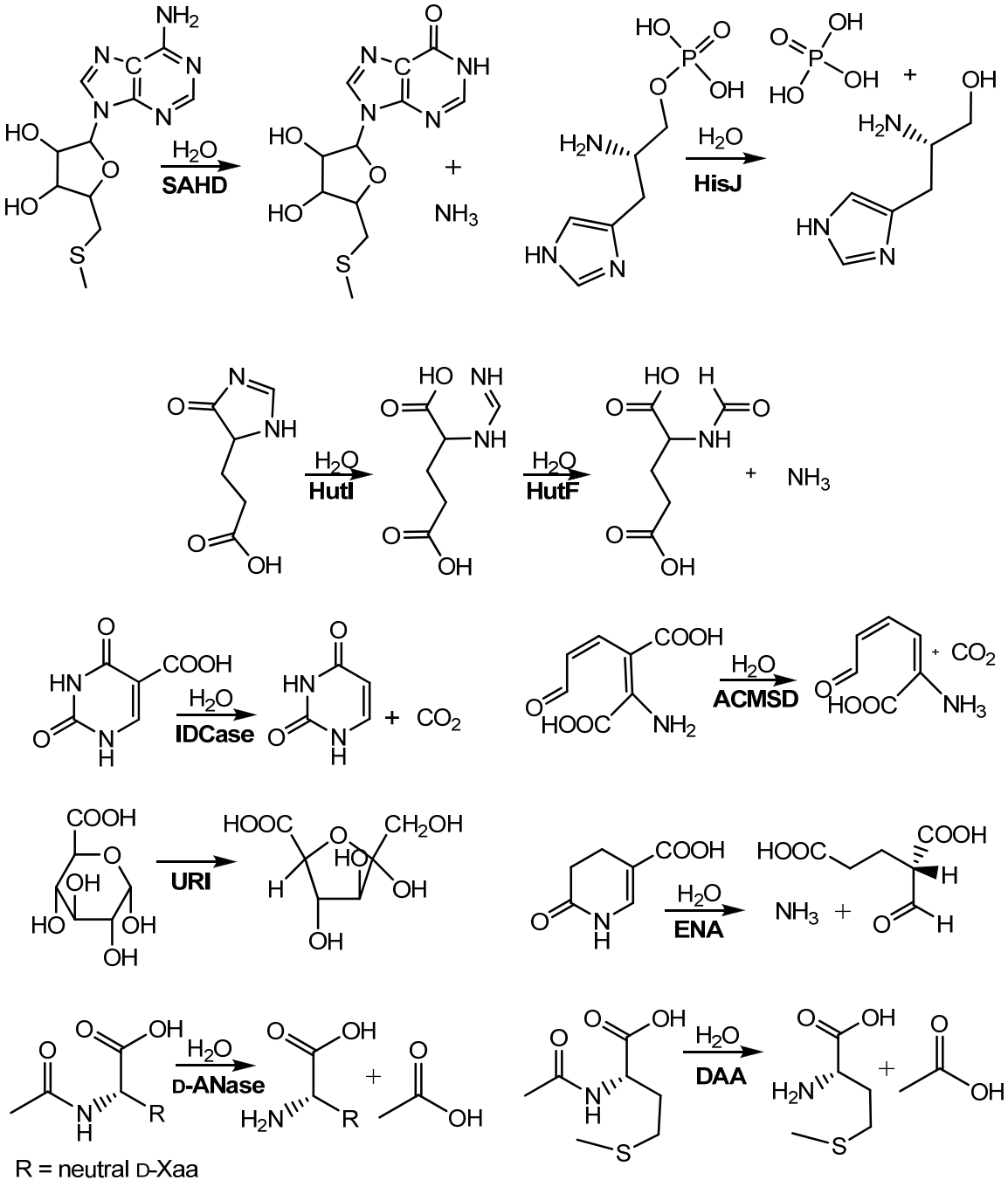
Scheme 1.1: Continued.



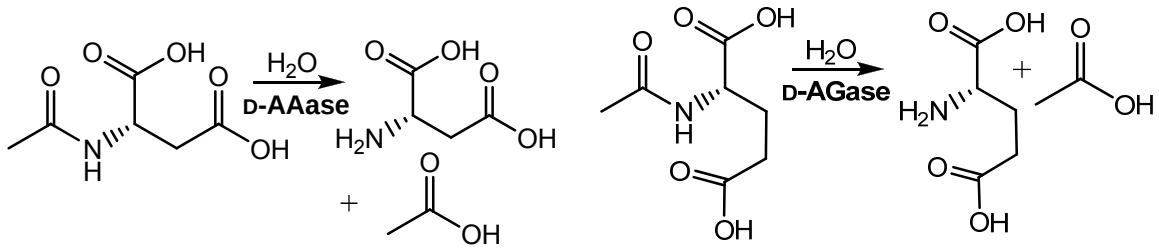
Scheme 1.1: Continued



Scheme 1.1: Continued.



Scheme 1.1: Continued.



Crystal Structures. An exhaustive search of the PDB has yielded 237 crystal structures of enzymes which belong to the AHS and is presented in Table 1.1. Structural analyses of enzymes identified as members of the AHS indicate that there are up to four metal binding positions which have been found to contain various divalent transition metal cations including Mg, Fe, Mn, Ni, Co and Zn. The four metal binding positions are named M_α and M_β, M_γ and M_δ. The M_α site is located deepest within the active site pocket, while the M_β, M_γ and M_δ sites become increasingly more solvent exposed. Historically, the three major types of enzymes within the AHS that have been characterized are the α/β-binuclear enzymes, the α-mononuclear enzymes and the β-mononuclear enzymes. The recently discovered M_γ and M_δ sites are found in a group of enzymes related to histidinol phosphate phosphatase like enzymes.

Thorough examination of the 237 crystal structures of identified members of the AHS have found that there are 17 unique types of AHS enzyme active site metal centers and they are summarized in Table 1.2. Identifications of the metal binding residues were determined with the criteria of a coordination distance of 2.5 Å or less between metal ion and amino acid side chain. Mechanistic characterizations of members of this superfamily have shown that the metals which may be utilized for catalysis are the divalent transition

metals: Mn^{2+}, Ni^{2+}, Co^{2+}, Fe^{2+}, and Zn^{2+}. While not found naturally in AHS enzymes, Cd^{2+} may be utilized by DHO (*33*) and NagA (*67*) for catalysis. The divalent transition metal cations are utilized as Lewis acids by these enzymes, effectively lowering the pK_a of a bound water molecule enabling the formation of a nucleophilic hydroxide (*76*). These metals may also provide a structural role within the enzymes, providing a nucleus around which proper folding of the enzyme can occur.

Table 1.1: All Crystal Structures of AHS Members Available in the PDB.

deposition	PDB #	annotation	organism	active site ligands
6/14/2006	2HBX	ACMSD	*P. flourescens*	1 Co
6/14/2006	2HBV	ACMSD	*P. flourescens*	1 Zn
N/A	1M7M	ADA	*H. sapiens*	Theoretical model
12/22/1992	1ADD	ADA	*M. musculus*	1 Zn, 1-deaza-adenosine
12/2/1994	2ADA	ADA	*M. musculus*	1 Zn, 6-hydroxy-7,8-dihydro purine adenosine
1/31/1998	1A4M	ADA	*M. musculus*	1 Zn, 6-hydroxy-1,6-dihydro purine nucleoside
1/31/1998	1A4L	ADA	*M. musculus*	1 Zn, 2'-Deoxycoformycin
1/10/2002	1KRM	ADA	*B. taurus*	1 Zn, 6-hydroxyl-1,6-dihydropurine riboside
12/9/2002	1NDZ	ADA	*B. taurus*	1 Zn, 1-((1R)-1-(hydroxymethyl)-3-(6-((3-(1-methyl- 1H-benzimidazol-2-yl)propanoyl)amino)-1H-indol-1-yl)propyl)-1H-IMIDAZOLE-4-carboxamide
12/9/2002	1NDY	ADA	*B. taurus*	1 Zn, 1-((1R)-1-(hydroxymethyl)-3-(1-naphthyl)propyl)- 1H-imidazole-4-carboxamide

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
12/9/2002	1NDV	ADA	*B. taurus*	1 Zn, N"-(4-(5-((1H-benzimidazol-2-ylamino)methyl)- 2-thienyl)-1,3-thiazol-2-yl)guanidine
9/8/2003	1QXL	ADA	*B. taurus*	1 Zn, 1-((1R)-1-(hydroxymethyl)-3-(1-naphthyl)propyl)- 1H-imidazole-4-carboxamide
10/3/2003	1UML	ADA	*B. taurus*	1 Zn, 1-((1R)-1-(hydroxymethyl)-3-{6-[(3-phenylpropanoyl)amino]- 1H-indol-1-yl}propyl)-1H-imidazole-4-carboxamide
10/5/2003	1O5R	ADA	*B. taurus*	1 Zn, 1-[(1R)-3-(6-{[(Benzylamino)carbonyl]amino}- 1H-indol-1-yl)-1-(hydroxymethyl)propyl]-1H-imidazole-4-carboxamide
12/14/2003	1V7A	ADA	*B. taurus*	1 Zn, 1-{(1R,2S)-2-Hydroxy-1-[2-(2-naphthyloxy)ethyl]propyl}-1H-imidazone-4-carboxamide
12/14/2003	1V79	ADA	*B. taurus*	1 Zn, 1-{(1R,2S)-1-[2-(2,3,-Dichlorophenyl)ethyl]- 2-hydroxypropyl}-1H-imidazole-4-carboxamide
4/16/2004	1VFL	ADA	*B. taurus*	1 Zn
6/22/2004	1W1I	ADA	*B. taurus*	1 Zn, Dipeptidyl peptidase (IV)
2/2/2005	1WXZ	ADA	*B. taurus*	1 Zn, 1-((1R,2S)-1-{2-[2-(4-chlorophenyl)-1,3-benzoxazol- 7-yl]ethyl}-2-hydroxypropyl)-1H-imidazole- 4-carboxamide

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
8/10/2005	2AMX	ADA	*P. yeolii*	1 Co
10/30/2006	2E1W	ADA	*B. taurus*	1 Zn, 1-{(1R,2S)-2-Hydroxy-1-[2-(1-naphthyl)ethyl]propyl}- 1H-imidazole-4-carboxamide
4/9/2007	2PGF	ADA	*P. vivax*	1 Zn, Adenosine
4/10/2007	2PGR	ADA	*P. vivax*	1 Zn, 2'-Deoxycoformycin
8/8/2007	2QVN	ADA	*P. vivax*	Guanosine-5'-monophosphate
8/20/2007	2Z7G	ADA	*B. taurus*	1 Zn (α), (2S,3R)-3-(6-amino-9H-purin-9-yl)nonan-2- ol
2/29/1996	1FKW	ADA D295E	*M. musculus*	1 Zn, Purine riboside
2/29/1996	1FKX	ADA D296A	*M. musculus*	1 Zn, 6-hydroxy-1,6-dihydro purine adenosine
8/30/1996	1UIO	ADA H238A	*M. musculus*	1 Zn, 6-hydroxy-7,8-dihydro purine adenosine/ H238A
8/30/1996	1UIP	ADA H238E	*M. musculus*	1 Zn, Purine riboside
8/17/2008	3E75	ALN	*E. coli*	2 Zn
8/17/2008	3E74	ALN	*E. coli*	2 Fe
6/25/2005	2A3L	AMPD	*A. thaliana*	1 Zn, Coformycin 5'-phosphate
8/1/2007	2QT3	AtzC	*Pseudomonas sp adp*	1 Zn (α)
4/24/2007	2PNK	BH0493	*B. halodurans*	None
6/25/2007	2QEE	BH0493	*B. halodurans*	1 Zn (α)
2/28/2009	3GGM	Bt9727_2919	*B. thuringiensis*	None
3/13/2008	3CJP	CAC3332	*C. Acetobutylicum*	2 Zn

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
10/17/2001	1K70	CDA	*E. coli*	1 Fe, 4-hydroxy-3,4-dihydro-1H-pyrimidin-2-one inhibitor
10/17/2001	1K6W	CDA	*E. coli*	1 Fe
10/31/2003	1RA5	CDA D314A	*E. coli*	1 Fe, 5-fluoro-4-(S)-hydroxy-3,4-dihydropyrimidine
10/31/2003	1R9Y	CDA D314A	*E. coli*	1 Fe
10/31/2003	1RA0	CDA D314G	*E. coli*	1 Fe, 5-fluoro-4-(S)-hydroxy-3,4-dihydropyrimidine
10/31/2003	1R9X	CDA D314G	*E. coli*	1 Fe
10/31/2003	1RAK	CDA D314S	*E. coli*	Fe, 5-fluoro-4-(S)-hydroxy-3,4-dihydropyrimidine
10/31/2003	1R9Z	CDA D314S	*E. coli*	1 Fe
1/24/2008	2VM8	Crmp-2 D-HYD like	*H. sapiens*	None
7/22/2002	1M7J	DAA	*A. faecalis*	2 Zn, Acetate
11/20/2003	1V51	DAA	*A. faecalis*	2 Zn, Acetate
11/20/2003	1RK6	DAA	*A. faecalis*	1Zn, 1 Cd, Acetate
11/20/2003	1RJP	DAA	*A. faecalis*	1 Zn, 1 Cu, Acetate
11/20/2003	1RK5	DAA D366A	*A. faecalis*	1 Zn, 1 Cu, Acetate
11/20/2003	1RJR	DAA D366A	*A. faecalis*	2 Zn, Acetate
11/20/2003	1RJQ	DAA D366A	*A. faecalis*	1 Zn, Acetate
11/20/2003	1V4Y	DAA H220A	*A. faecalis*	1 Zn, AcetateLigand free/ Mutant
7/31/2008	3E0L	Designed Ammelide Deaminase	*H. sapiens*	1 Zn (α)

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
9/17/2004	1XGE	DHO	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydro pyrimidine-4-carboxylate
10/14/2004	1XRF	DHO	*A. aeolicus*	2 Zn, Sulfate
10/15/2004	1XRT	DHO	*A. aeolicus*	2 Zn
5/4/2006	2GWN	DHO	*P. gingivalis*	2 Zn, β-mercaptoethnol, chloride, sulfate, glycerol, cacodylate ion
2/28/2007	2EG8	DHO	*E. coli*	2 Zn, 5-flouroorotate
2/28/2007	2EG7	DHO	*E. coli*	2 Zn, 2-oxo-1,2,3,6-tetrahydropyrimidine-4,6-dicarboxylate
2/28/2007	2EG6	DHO	*E. coli*	2 Zn
5/6/2007	2Z00	DHO	*T. thermophilus*	2 Zn
5/20/2008	3D6N	DHO	*A. Aeolicus*	1 Zn, Citrate
3/25/2009	3GRI	DHO	*S. Aureus*	1 Zn (α)
5/17/2007	2Z2B	DHO del. 107-116	*E. coli*	2 Zn
5/17/2007	2Z29	DHO T109A	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydro-pyrimidine-4-carboxylate
5/17/2007	2Z2A	DHO T109G	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydro pyrimidine-4-carboxylate
11/8/2006	2E25	DHO T109S	*E. coli*	2 Zn, 5-Flouroorotate
5/17/2007	2Z27	DHO T109S	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydro pyrimidine-4-carboxylate

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
5/17/2007	2Z26	DHO T110A	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydro pyrimidine-4-carboxylate
5/17/2007	2Z24	DHO T110S	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydro pyrimidine-4-carboxylate
5/17/2007	2Z25	DHO T110V	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydropyrimidine-4-carboxylate
1/5/2007	2OGJ	DHO-like	*A. tumefaciens*	2 Zn, Imidazole
1/25/2006	2FTY	DHP	*S. kluyveri*	2 Zn
1/25/2006	2FTW	DHP	*D. discoideum*	2 Zn
1/31/2006	2FVM	DHP	*S. kluyveri*	2 Zn, N-carbamoyl-β-alanine
1/31/2006	2FVK	DHP	*S. kluyveri*	2 Zn, Dihydrouracil
3/25/2008	2VR2	DHP	*H. sapiens*	2 Zn
6/3/2008	3DC8	DHP	*S. meliloti*	2 Zn
4/26/2006	2GSE	DHP-like	*H. sapiens*	None
11/11/2001	1KCX	DHP-related	*M. musculus*	None
7/30/2007	2QS8	Dipeptidase (Xaa-Pro)	*A. macleodii*	1 Mg (α) , Methionine
8/24/2006	2I5G	Dipeptidase like T56A, K300E	*P. aeruginosa*	None
1/3/2005	2BGN	Dipeptidyl dipeptidase IV and ADA	*H. sapiens and B. taurus*	Tat N-terminal nonapeptide and Zn
11/30/2008	3FEQ	eah89906	*Unidentified*	2 Zn

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
5/27/2008	2VUN	ENA	*E. barkeri*	1 Fe, 1 Zn, 1Cl
9/6/2006	2I9U	GDA	*C. acetobutylium*	1 Fe, Guanine
1/25/2007	2OOD	GDA	*B. japonicum*	1 Zn, Guanine
4/26/2007	2UZ9	GDA	*H. sapiens*	1 Zn, Xanthine
4/26/2007	2YXO	HisJ	*T. thermophilus*	2 Fe, 1Zn, Sulfate
5/3/2007	2YZ5	HisJ	*T. thermophilus*	2 Fe, 1Zn, Phosphate
6/17/2007	2Z4G	HisJ	*T. thermophilus*	2 Fe, 1Zn
10/16/2005	2BB0	HutI	*B. subtilis*	1 Zn, acetate
2/19/2006	2G3F	HutI	*B. subtilis*	1 Zn, substrate analogue imidazole-4-acetate
4/13/2006	2GOK	HutI	*A. tumefaciens*	1 Fe
1/25/2007	2OOF	HutI	*Unidentified*	1 Fe
5/9/2007	2PUZ	HutI	*A. tumefaciens*	1 Fe, N-formimino-L-glutamate product
5/21/2007	2Q09	HutI	*Unidentified*	1 Fe, 3-(2,5-Dioxo-imidazolidin-4-yl)-propionic acid inhibitor
8/20/2001	1GKR	HYD	*A. aurescens*	2 Zn
8/20/2001	1GKQ	HYD	*Thermus sp.*	2 Zn
8/20/2001	1GKP	HYD	*Thermus sp.*	2 Zn
9/25/2001	1K1D	HYD	*B. stereothermophilus*	2 Zn
12/14/2002	1NFG	HYD	*B. pickettii*	2 Zn
1/26/2005	1YNY	HYD	*Bacillus sp.*	2 Mn
3/2/2003	1ONX	IAD	*E. coli*	2 Zn, Aspartate
3/2/2003	1ONW	IAD	*E. coli*	2 Zn
6/15/2003	1POK	IAD	*E. coli*	2 Zn, Asparagine

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
6/15/2003	1PO9	IAD	*E. coli*	2 Zn
12/21/2004	1YBQ	IAD D285N	*E. coli*	2 Zn, L-β-aspartylhistidine
8/18/2005	2AQO	IAD E77Q	*E. coli*	2 Zn
8/18/2005	2AQV	IAD Y137F	*E. coli*	2 Zn
11/16/2007	3BE7	KRCP	*Unidentified*	Arginine
7/17/2008	3DUG	KRCP	*Unidentified*	2 Zn, Arginine
6/15/2007	2QAH	LigI	*S. paucimobilis*	None
5/4/2006	2GWG	LigJ	*R. palustris*	1 Zn
11/29/2005	2F6K	Lp24	*L. plantarum*	1 Mn (α)
10/15/2002	1O12	NagA	*T. maritima*	1 Fe
1/21/2005	1YMY	NagA	*E. coli*	None
2/4/2005	1YRR	NagA	*E. coli*	Phosphate
3/14/2007	2P50	NagA	*E. coli*	1 Zn
11/22/2007	2VHL	NagA	*B. subtilis*	2 Fe, glucosamine-6-phosphate,
11/22/2007	2VHL	NagA	*B. subtilis*	2 Fe, Glucosamine-6-phosphate product
9/10/2008	3EGJ	NagA	*V. cholerae*	1 Ni (β)
3/14/2007	2P53	NagA D273N	*E. coli*	1 Zn, *N*-methylphosphonamidate-glucosamine-6-phosphate inhibitor
8/23/2007	2R1O	OpdA PTE S92A	*A. tumefaciens*	1 Co, 1 Fe, Diethyl thiophosphate product, 1,2-ethanediol

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
8/23/2007	2R1P	OpdA PTE S92A, K185R	*A. tumefaciens*	1 Co, 1 Fe , Diethyl thiophosphate product, 1,2-ethanediol
8/23/2007	2R1N	OpdA PTE S92A, N265D	*A. tumefaciens*	1 Co, 1 Fe, Diethyl 4-methoxyphenyl phosphate slow substrate,
8/23/2007	2R1M	OpdA PTE S92A, N265D	*A. tumefaciens*	1 Co, 1 Fe, Diethyl phosphate product, 1,2-ethanediol
8/23/2007	2R1L	OpdA PTE S92A, N265D	*A. tumefaciens*	1 Co, 1 Fe, Diethyl thiophosphate, 1,2-ethanediol
8/23/2007	2R1K	OpdA PTE S92A, N265D	*A. tumefaciens*	1 Co, 1 Fe, Diethyl phosphate, 1,2-ethanediol
5/27/1998	1BF6	PHP	*E. coli*	2 Zn, Glycerol, (4S)-2-Methyl-2,4-pentanediol
8/11/2005	2ANU	Predicted PTE TM0559	*T. maritima*	4 Zn, 2Cl
7/7/1994	1PTA	PTE	*B. dimunita*	None
4/25/1995	1PSC	PTE	*B. dimunita*	2 Cd, Diethyl-4-methylbenzylphosphonate
2/13/1996	1DPM	PTE	*B. dimunita*	2 Zn, Diethyl-4-methylbenzylphosphonate
5/9/2000	1EZ2	PTE	*P. dimunita*	2 Zn, diisopropylmethyl phosphonate
5/9/2000	1EYW	PTE	*P. dimunita*	2 Zn, triethylphosphate
1/26/2001	1HZY	PTE	*B. dimunita*	2 Zn, Ethanediol, Formate, 2-phenyl-ethanol
1/29/2001	1I0D	PTE	*B. dimunita*	Zn, Cd, Ethanediol, Formate, 2-phenyl-ethanol
1/29/2001	1I0B	PTE	*B. dimunita*	2 Mn, Ethanediol, Formate, 2-phenyl-ethanol

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
9/1/2003	1QW7	PTE	*B. dimunita*	2 Co, Diethyl-4-methylbenzylphosphonate
9/8/2005	2D2G	PTE	*A. tumefaciens*	2 Co, Dimethylthiophosphate
9/8/2005	2D2J	PTE	*A. tumefaciens*	2 Co
9/8/2005	2D2H	PTE	*A. tumefaciens*	2 Co, Trimethylphosphate
11/2/2007	2ZC1	PTE	*D. radiodurans*	2 Co
2/20/2008	3CAK	PTE	*B. diminuta*	2 Co, Ethylphosphate
12/4/2006	2O4Q	PTE G60A	*P. dimunita*	2 Zn, Dimethylarsenate
4/8/2008	3CS2	PTE G60A	*B. diminuta*	2 Co, Dimethylarsenate (cacodylate)
4/29/2003	1P6B	PTE H254G/H257W/ L303T	*Flavobacterium sp.*	3 Zn, Diethyl-4-methylbenzylphosphate, ethyl dihydrogenphosphate
4/29/2003	1P6C	PTE H254G/H257W/ L303T	*Flavobacterium sp.*	2 Zn, Diisopropyl methyl phosphonate, Diethyl-4-methylbenzyl phosphonate
1/31/2007	2OQL	PTE H254Q/H257F	*P. dimunita*	2 Zn, 2-[Bis-(2-hydroxy-ethyl)-amino]-2-hydroxymethyl- propane-1,3-diol
8/7/2008	3E3H	PTE H254R, H275F	*B. dimunita*	2 Co, Diethyl 4-methylbenzyl phosphonate
12/18/2006	2OB3	PTE H257Y/L303T	*B. dimunita*	2 Zn
12/4/2006	2O4M	PTE I106G/F132G/H 257Y	*P. dimunita*	2 Zn, Dimethylarsenate, Acetate, Glycerol
9/13/2006	2ICS	Putative ADE	*E. faecalis*	1 Zn, Adenine
3/27/2007	2PAJ	Putative CDA / GDA	*Unidentified*	1 Zn (α)

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
12/19/2005	2FFI	Putative LigI	*P. putida*	Phosphate
10/23/2007	3B40	Putative peptidase K44Q, I62V, P64L, E69T, F73V, N79G, N195S, S203A, T215S, A216P, N314D	*P. aeruginosa*	None
3/24/2007	2P9B	Putative prolidase	*B. longum*	None
7/31/2008	3E0F	Putative PTE	*B. adolescentis*	2 Fe, 1Zn, acetate, phosphate
8/1/2006	2DVX	RDC	*Rhizobium sp.*	1 Zn, 2,3-dihydroxybenzaldehyde
8/1/2006	2DVU	RDC	*Rhizobium sp.*	1 Zn, 2,6-dihydroxybenzoate
8/1/2006	2DVT	RDC	*Rhizobium sp.*	1 Zn
2/2/2002	1ITQ	RDP	*H. sapiens*	2 Zn
2/3/2002	1ITU	RDP	*H. sapiens*	2 Zn, Cilastatin
11/25/2008	3FDG	RDP	*R. sphaeroides*	2 Mg
7/9/2002	1J6P	SAHD	*T. maritima*	1 Ni
4/12/2003	1P1M	SAHD	*T. maritima*	1 Ni, Methionine
4/20/2007	2PLM	SAHD	*T. maritima*	1 Zn , inosylhomocysteine
10/4/2006	2IMR	SAHD-like DR_0824	*D. radiodurans*	1 Zn
9/18/2007	2VC7	SsoPox	*S. solfataricus*	1 Fe, 1Co, (4S)-4-(decanoylamino)-5-hydroxy-3,4-dihydro- 2H-thiophenium
9/18/2007	2VC5	SsoPox	*S. solfataricus*	1 Fe, 1Co
11/2/2004	1XWY	TatD	*E. coli*	1 Zn

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
8/6/2008	3E2V	TatD like	*S. cerevisiae*	1 Mg
5/12/2006	2GZX	TatD MW0446	*S. aureus*	2 Ni
2/27/2009	3GG7	TatD-like	*D. radiodurans*	1 Mn (α)
3/30/2009	3GUW	TatD-like	*A. Fulgidus*	2 Zn
7/9/2002	1J6O	TatD-related	*T. maritima*	None
7/25/2007	2QPX	Unknown	*L. casei atcc 334*	2 Zn
10/8/2008	3ETK	Unknown	*P. furiosus*	2 Mg
4/21/2009	3HPA	Unknown	*Unidentified*	1 Zn (α)
11/25/2008	3FDK	Unreleased	*Unidentified*	Unknown
2/16/1995	2KAU	URE	*K. aerogenes*	2 Ni
6/20/1995	1KRA	URE	*K. aerogenes*	None
4/23/1997	1FWJ	URE	*K. aerogenes*	2 Ni
1/21/1998	1UBP	URE	*B. pasteurii*	2 Ni, β-mercaptoethanol
11/4/1998	2UBP	URE	*B. pasteurii*	2 Ni, Sulfate
12/16/1998	3UBP	URE	*B. pasteurii*	2 Ni, Diamidophosphate
2/25/1999	4UBP	URE	*B. pasteurii*	2 Ni, Acetohydroxamate
2/5/2000	1EF2	URE	*K. aerogenes*	2 Mn
11/1/2000	1E9Z	URE	*H. pylori*	2 Ni
11/1/2000	1E9Y	URE	*H. pylori*	2 Ni, Acetohydroxamate
4/9/2001	1IE7	URE	*B. pasteurii*	2 Ni, Phosphate
1/14/2004	1S3T	URE	*B. pasteurii*	2 Ni, Sulfate, Borate
3/4/2000	1EJX	URE at 100K	*K. aerogenes*	2 Ni
3/4/2000	1EJW	URE at 298K	*K. aerogenes*	2 Ni

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
4/23/1997	1FWB	URE C319A pH 6.5	*K. aerogenes*	2 Ni
4/23/1997	1FWA	URE C319A pH 7.5	*K. aerogenes*	2 Ni, Carbonate
4/23/1997	1FWC	URE C319A pH 8.5	*K. aerogenes*	2 Ni
4/23/1997	1FWD	URE C319A pH 9.4	*K. aerogenes*	2 Ni
4/23/1997	1FWF	URE C319D	*K. aerogenes*	2 Ni
4/23/1997	1FWG	URE C319S	*K. aerogenes*	2 Ni
4/23/1997	1FWH	URE C319Y	*K. aerogenes*	2 Ni
3/4/2000	1EJR	URE D221A	*K. aerogenes*	2 Ni
4/23/1997	1FWI	URE H134A	*K. aerogenes*	1 Ni (β)
6/20/1995	1KRB	URE H219A	*K. aerogenes*	2 Ni
3/4/2000	1EJS	URE H219N	*K. aerogenes*	2 Ni
3/4/2000	1EJT	URE H219Q	*K. aerogenes*	2 Ni
6/20/1995	1KRC	URE H320A	*K. aerogenes*	2 Ni, Carbon Dioxide
3/4/2000	1EJU	URE H320N	*K. aerogenes*	2 Ni
3/4/2000	1EJV	URE H320Q	*K. aerogenes*	2 Ni
2/17/1998	1A5N	URE K217A	*K. aerogenes*	2 Ni, Formate
2/17/1998	1A5M	URE K217A	*K. aerogenes*	None
2/17/1998	1A5L	URE K217C	*K. aerogenes*	None
2/17/1998	1A5O	URE K217C	*K. aerogenes*	2 Ni, Formate
2/17/1998	1A5K	URE K217E	*K. aerogenes*	None
7/2/2002	1J5S	URI	*T. maritima*	None
5/18/2007	2Q01	URI	*C. cresentus*	None

Table 1.1: Continued.

deposition	PDB #	annotation	organism	active site ligands
6/5/2007	2Q6E	URI	*B. halodurans*	1 Zn (α)
7/12/2002	1M65	YCDX	*E. coli*	1 Zn, 2 Na
7/14/2002	1M68	YCDX	*E. coli*	3 Zn, 1 Sulfate
5/14/2003	1PB0	YCDX	*E. coli*	3 Zn
1/13/2005	1YIX	YcfH TatD-like	*E. coli*	1 Zn

Diverse Active Sites within the AHS. The binuclear active site type I is the best characterized and most versatile architecture within the AHS. It contains the α-metal which is coordinated by the N^{ε} of two histidines from β-strand 1, a carboxylated lysine from strand 4, and an aspartate from strand 8. The β-metal is bound by the carboxylated lysine, the N^{δ} of a strand-5 histidine and the N^{ε} of a strand-6 histidine. From this point, unless stated otherwise, metals are coordinated to histidine through the N^{ε} atom. The type I binuclear metal is utilized by the enzymes: PTE (*51*), DHO (*31*), URE (*58*) IAD (*40*), HYD (*35, 36*), ALN (*37*), DHP (*34*), carboxypeptidase (*42*) and SsoPox (*77*).

The type II active site used by PTE from *D. radiodurans* (PDB code: 2ZC1) differs from type I only by the presence of the β-metal coordinating tyrosine located 5 residues after the last histidine from strand 1. ENA (*41*) and the enzymes of unknown function; PHP (*78*) and TatD (PDB code: 2GZX) utilize the active site type III structure. The difference between the type III and type I active site is the substitution of the bridging carboxylated lysine for a bridging glutamate. RDP (*45*) utilizes active site type IV. In that site, the M_{α} site is formed through interactions from an HxD motif from

strand 1 and a glutamate from strand 4. The M_β site consists of the bridging glutamate and the histidines from strands 5 and 6. The aspartate from strand 8 is hydrogen bonding with a bridging hydroxide, but is not coordinated to the metal.

NagA from *B. subtilis* (*69*) represents active site type V, which utilizes the strand 1 HxH motif, a bridging glutamate from strand 3 and an aspartate from strand eight to form the M_α site. The M_β site is formed by the bridging glutamate and two histidine residues from strands 5 and 6. NagA from *T. maritima* represents type VI, and has all of the same metal ligands as type V, however, the crystal structure (PDB code: 1O12) shows a single iron bound in the M_β site, coordinated to the glutamate from strand 3 and the histidines from strands 5 and 6. Instead of the α-metal, a histidine from a mobile loop after strand 3 is found nearby. It has been shown that only a single metal is necessary for catalysis (*66*) Those studies will be further discussed in Chapter III. Type VII is represented by *E. coli* NagA, which has the same active site as type VI, except it is missing the superfluous HxH motif. That enzyme was found to utilize a single metal for catalysis with the aid of the mobile loop histidine from strand 3 (*66, 67*). The characterization of that enzyme is discussed further in Capters II and III.

Type VIII provides a novel subset of the AHS and consists of DAA (PDB code: 1RJQ). It has an active site similar to the previous binuclear sites I, II and III and may bind two metals. However, there are several differences; a cysteine from strand 2 replaces the bridging carboxylate residue and only the β-metal is required for catalysis (*73*). The active site of RDC comprises type IX. In this example, the α-mononuclear active site is formed by a unique ExH motif from strand 1, a histidine from strand 5 and

an aspartate from strand 8. It should be noted that the histidine commonly seen in strand 6 is present, and it has been determined to be necessary for catalytic activity (*55*).

Deaminases dominate the type X active site structure which is defined by the strand 1 HxH motif, the strand 5 histidine and the strand 8 aspartate metal ligands. A histidine from strand 5 is also present which is involved in catalysis. Included among the members are: CDA (*23*), GDA (PDB code: 2UZ9), ADA (*17*), AMPD (*18*), Tm0936 (*7*), HutF (*79*), AtzC (PDB code: 2QT3) and even the recently discovered decarboxylase ACMSD (*56*). LigJ (PDB code: 2GWG) has a similar α-mononuclear active site, yet it is the first amidohydrolase found which does not have an aspartate at the end of strand 8. In its place is a glutamate from strand 8 which helps form the type XI active site. Uronate isomerase utilizes active site type XII, where a Zn can be seen bound to the strand 1 HxH motif and the strand-8 aspartate. Notably absent in the structure for *B. halodurans* are the histidines from strands 5 and 6. However, URI from *E. coli* and *T. maritima* have the strand-5 histidine (*62*).

Active site types XIII-XVII all share similar characteristics, as they all bind three or more active site metals and possess an overall fold which takes on a $(\beta/\alpha)_7$ barrel fold appearance, while maintaining a close relationship to the typical $(\beta/\alpha)_8$ structures. HisJ (*9*) possesses the type XIII active site, which has two Fe ions in the M_α and M_β sites, as well as a Zn in the newly defined M_γ site. The α-metal ligands come from the strand 1 HxH motif, a glutamate from strand 3, and an aspartate from strand 7 which is analogous to the strand 8 aspartate found in the other AHS enzymes. The β-metal ligands are: the bridging glutamate from strand 3 and two histidines from strands 4 and 5, which are

related to the strand 5 and 6 histidine residues found in the other AHS enzymes. The Zn binding γ-metal ligands are: a histidine located on a loop after strand 1, a histidine from strand 2 and a histidine present two amino acids after the strand-7 aspartate.

Active site type XIV with two Fe ions and a Zn ion, is found in the putative metal-dependent phosphoesterase from *B. adolescentis* ATCC 15703 (PDB code: 3E0F). The strand-1 HxH motif, a bridging strand-3 glutamate and an aspartate from strand 7 form the α-metal Fe binding site. The β-metal Fe binding site consists of the bridging glutamate from strand 3, the N^{δ} of the histidine from strand 4 and the histidine from strand 5. The γ-metal Zn binding site is made from contributions by: an aspartate from the loop after the strand 1, a histidine from strand 2 and the histidine from strand 7 found two amino acids after the aspartate.

Active site type XV is found in the YCDX protein of unknown function. In that protein the α-metal Zn binding site is formed by the HxH motif from strand 1, a bridging glutamate from strand 3 and an aspartate from strand 7. The β-metal Zn binding site has the bridging glutamate and two histidines from strands 4 and 5. Four histidines are found to bind the γ-site Zn ion in the ligand free structure (PDB code: 1PB0). They are: the histidine found six amino acids after the strand-1 HxH motif, the histidine from strand 2, a histidine from strand 6 and a histidine from strand 7, found two amino acids after the aspartate.

The final and most extreme active site type of the AHS is type XVI. This is found in the predicted metal dependent phosphoesterase Tm0559 (PDB code: 2ANU). This active site is found with 4 Zn ions. The strand-1 HxH makes up the α-site along with a bridging glutamate from strand 3 and an aspartate from strand 7. The β-site is formed by the bridging glutamate from strand 3and the N^{δ} of a strand-4 histidine. The γ-site ligands are an aspartate found five amino acids after the strand-1 HxH, a histidine from strand 2, and another histidine from strand 7 found two amino acids after the strand-7 aspartate. The rare δ-metal site is formed from interactions with two aspartates. One aspartate is found after strand 5 and the other is the bridging aspartate from strand 7. Table 1.2 summarizes these active site-variations. Histidine residues which utilize the N^{δ} atom for metal coorination are denoted with an N^{δ} subscript. All other hisidine ligands were found to coordinate the metal with the N^{ε} atom. Lower-case x is used to imply a lack of residue conservation.

Table 1.2 : Active Site Variations in the AHS.

		beta-strands							
type	positions	1	2	3	4	5	6	7	8
I	α, β	HxH			K	$H_{N\delta}$	H		D
II	α, β	HxHxxxxY			K	$H_{N\delta}$	H		D
III	α, β	HxH			E	$H_{N\delta}$	H		D
IV	α, β	HxD			E	H	H		d[b]
V	α, β	HxH		E		H	H		D
VI	β	hxh[a]		E...h[b]		H	H		d[b]
VII	β			E...h[b]		H	H		d[b]
VIII	β	hxh[a]	C			$H_{N\delta}$	H		d[b]
IX	α	ExH				H	h[b]		D
X	α	HxH				H	h[b]		D
XI	α	HxH				H	h[b]		E
XII	α	HxH							D
XIII[c]	α, β, γ	$HxH(x)_5$ H	H	E	H	H		DxH	
XIV[c]	α, β, γ	$HxH(x)_4$ D	H	E	$H_{N\delta}$	H		DxH	
XV[c]	α, β, γ	$HxH(x)_5$ H	H	E	H	H	H	DxH	
XVI[c]	α, β, γ, ε	$HxH(x)_4$ D	H	E	$H_{N\delta}$	D		DxH	

[a]Metal binding may be possible, but does not appear to be required for activity. [b]Not coordinated to a metal, but may be involved in water or substrate activation. [c]$(\beta/\alpha)_7$ fold related to the $(\beta/\alpha)_8$ fold. Additional details are provided in the text.

A general mechanism adopted from the proposed mechanisms of DHO (*31, 33*) PTE (*49, 51, 80*) and IAD (*40*) for the α/β binuclear metal assisted hydrolysis is shown below in Scheme 1.2. This is based on crystallographic evidence with substrate, product,

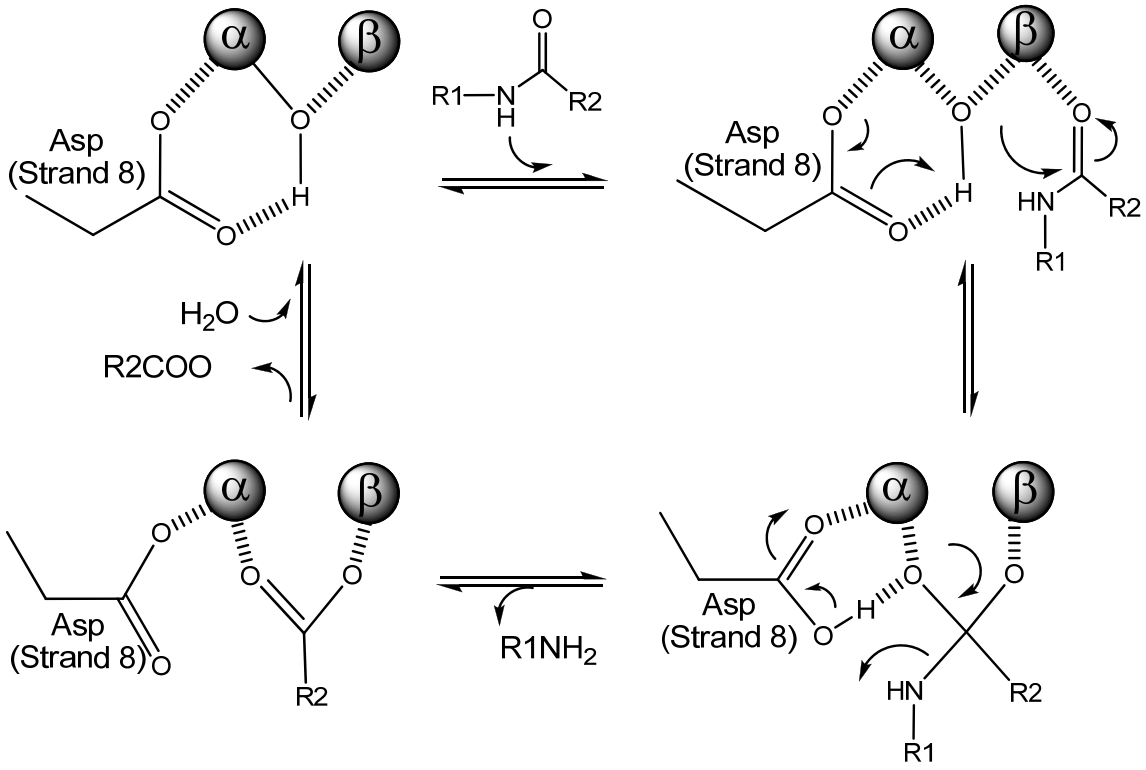
and inhibitors as well as theoretical calculations. In the resting state, two metals coordinate a bridging nucleophilic hydroxide molecule. An aspartate from β-strand eight helps orient the hydroxide, while coordinating the α-metal as well. Upon substrate coordination to the β-site metal, coordination to the hydroxide is weakened, resulting in aspartate assisted hydroxide attack at the reaction center that is followed by oxyanion formation at the β-metal. The resulting anionic tetrahedral transition-states are stabilized through a bridging complex with both metals. Collapse of the tetrahedral intermediate is facilitated through reversed flow of electrons from the oxyanion and proton transfer to the leaving group from the aspartate.

For PTE, the mechanism is slightly different. In that case, aspartate assisted hydroxide attack results in the rapid loss of the leaving group, forming an anionic tetrahedral metal bridging phosphodiester complex. The reactive state of the active site is regenerated upon dissociation of the resulting product and then the binding of another nucleophilic water molecule. QM/MM simulations have been used to suggest another mechanism for PTE (*81*). In that study, it was proposed that instead of a phosphodiester anion coordinating both metals in a bridging complex, the bridging hydroxide attack results in the rapid displacement of the leaving group product and that a single P=O oxygen is coordinated to the β-metal while the bound phosphodiester product remains neutral. The binding and deprotonation of a second water aids in the displacement of the neutral phosphodiester product, which then spontaneously deprotonates in solution. In another and very different mechanism for bacterial PTE, it was proposed, based on structural and docking data, that the bridging hydroxide acts as a base for the activation

of another hydrolytic solvent molecule which is terminally bound to the α-metal (*82*).

Differing mechanisms have also been proposed for the URE mechanism. Yamaguchi et al have utilized crystallographic, mutagenesis and pH profile studies to propose a reverse protonation mechanism which uses a histidine to protonate the leaving group amine and that no base is required to deprotonate the bridging hydroxide. They also propose that a cysteine residue is being used to bind the non-leaving group amine of urea (*83*). A different mechanism has been proposed by Benini et al, where crystal structures with bound inhibitors were used to suggest that the carbonyl oxygen coordinates to the β-Ni, while the non-leaving group amine coordinates to the α-Ni. The bridging hydroxide then becomes the acid, which protonates the leaving group amine, possibly through the aspartate (*58*). Lippard and co-workers have used synthetic dinickel urease models to suggest that an elimination mechanism might occur in urease (*84*). Quantum mechanical and molecular dynamics simulations and quantum chemical calculations of Merz and co-workers have been used to try to reconcile the opposing mechanisms of urease. From those studies, they found that both hydrolysis mechanisms, and the elimination mechanisms are possible, depending on the hydrogen bonding network present (*60, 85, 86*). Marlier and Cleland have used heavy isotope effects studies with formamide and semicarbazide substrates to support a reverse protonation mechanism which utilizes a histidine to help stabilize oxyanion formation near the β-metal in the transition state while a second histidine is used to protonate the leaving group amine (*61, 87*).

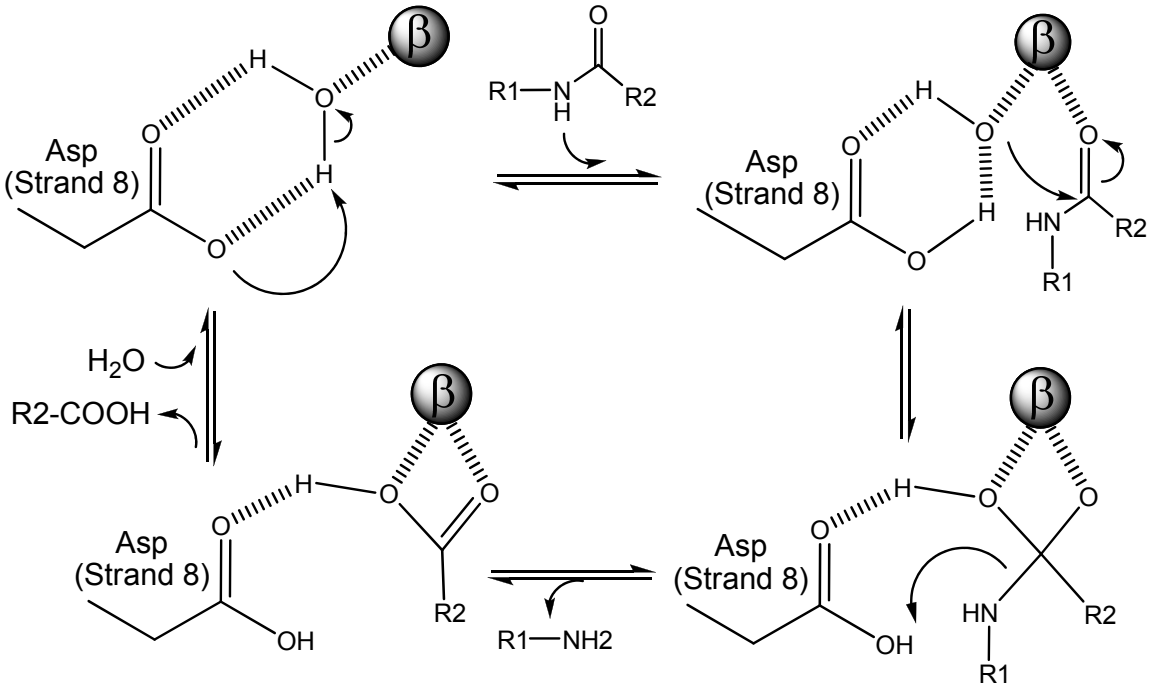
Scheme 1.2



The β-mononuclear enzymes of the amidohydrolase superfamily comprise a second sub-class of enzymes, which includes NagA from *Escherichia coli,* NagA from *Thermotoga maritima,* and D-aminoacid deacetylase (DAA). They all utilize a single metal bound only to the β-site through the use of two histidines from β-strands five and six. Each of these three enzymes represent distinct types of β-mononuclear active sites, because they possess variations in the active site residues which were previously discussed. In all of these active sites, the β-metal is held in a tetrahedral coordination state which includes a coordinated hydroxide that is hydrogen-bonded to the conserved aspartate residue from the end of β-strand eight. It is proposed for these enzymes that a fifth metal ligation site is created upon binding of substrate, creating a trigonal

bipyramidal coordination state of the active site metal. A proposed general mechanism for β-metal assisted hydrolysis of an amide bond is shown in Scheme 1.3. In this mechanism, the resting enzyme is coordinated through a glutamate from strand three, two histidines from strands five and six, in addition to a nucleophilic water molecule. An aspartate from strand eight removes a proton from the water, which is then activated for nucleophilic attack on the carbonyl carbon of the substrate. A tetrahedral intermediate is then formed and is stabilized through bi-dentate coordination to the active site metal. Collapse of the resulting tetrahedral intermediate is initiated by proton transfer from the strand-eight aspartate to the leaving group amine. Dissociation of the amine product is followed by dissociation of the newly formed carboxylate product. Regeneration of the active site may then occur upon the binding of another water molecule. This mechanism is highly analogous to that determined for the binuclear enzymes (*31, 33, 40, 49*).

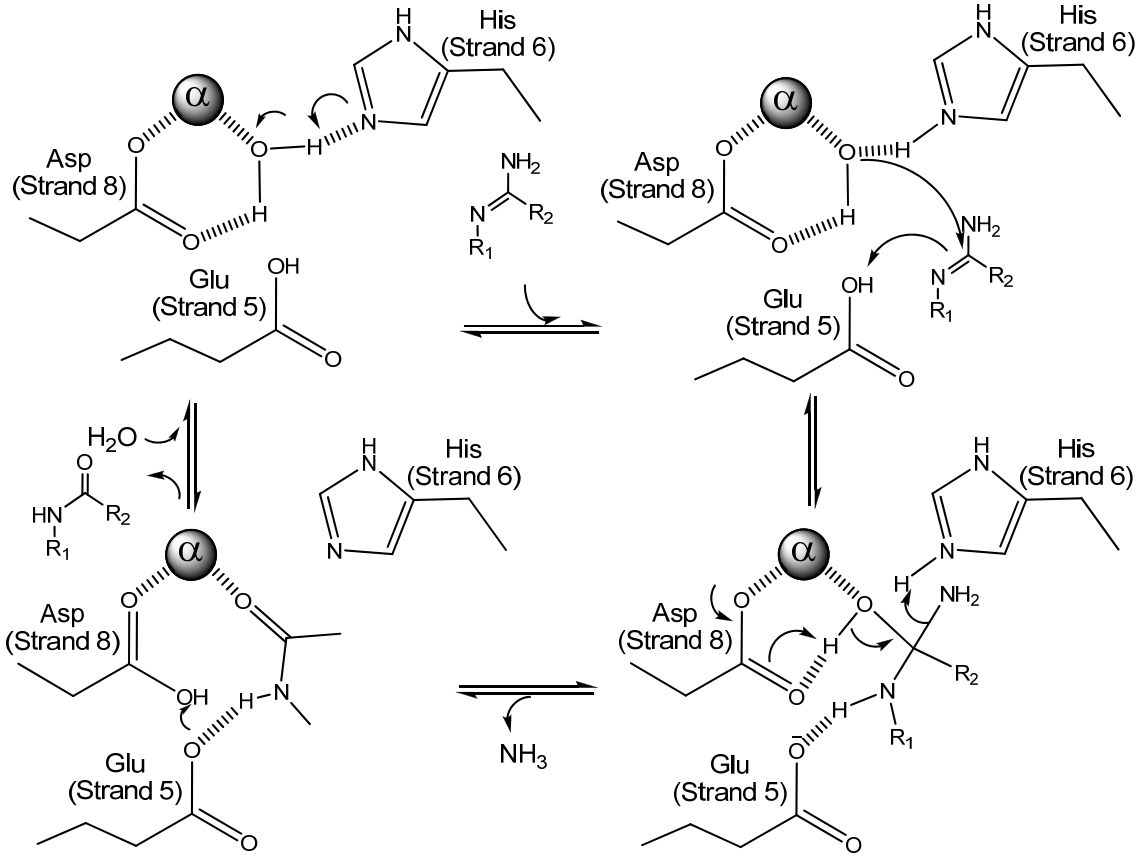
Scheme 1.3



The third class of amidohydrolase enzymes binds one metal at the α-position. This metal is ligated with the aid of the two histidine residues from β-strand one, the histidine from strand five and the aspartate from strand eight. The fifth coordination site of the divalent transition metal is occupied by a nucleophilic hydroxide which is also hydrogen-bonded to the indispensable strand-eight aspartate. Some of the members of this α-mononuclear group of amidohydrolases that have been identified are the enzymes: cytosine deaminease (CDA) (*23*), adenosine deaminase (ADA) (*17*), AMP deaminase (AMPD) (*88*), S-adenosylhomocysteine deaminase(SAHD) (*7*) and guanine deaminase (GuaD). These enzymes all share an HxxE motif found at the end of strand 5. The histidine present in this motif is analogous to the metal binding histidine from strand five that can be found in the binuclear and β-mononuclear classes of enzymes.

A general mechanism for α-metal assisted deamination of an aromatic compound is shown in Scheme 1.4. The mechanistic details of this reaction are based on the crystallographic studies of adenosine deaminase (*16, 17*), cytosine deaminase (*22, 23*) and earlier reports of QM/MM calculations for ADA, which indicated that the resting state contains a protonated glutamic acid residue and that histidine was the active base (*89*). In this proposed mechanism, the resting state of the enzyme contains a metal-coordinated nucleophilic water molecule which is also hydrogen bonded to aspartate from strand eight and histidine from strand six. Prior to, or upon substrate binding, the water is deprotonated by the hydrogen-bonding histidine. This promotes the nucleophilic attack of the resulting hydroxide anion upon the sp^2 carbon bearing the leaving group amine. Electrons move to the adjacent nitrogen N-3, abstracting a proton from the H-bonded strand-five glutamate. The movement of these electrons enables the formation and stabilization of the resulting tetrahedral transition state. A second proton abstraction from the nucleophilic water, now a transition state hydroxyl group, results in the collapse of the tetrahedral intermediate through the formation of a ketone and the release of ammonia which gets a proton from the histidine from strand six. Regeneration of the resting state occurs upon product release, which is followed by the binding of a new water molecule and proton transfer from strand-five glutamate to strand-eight aspartate.

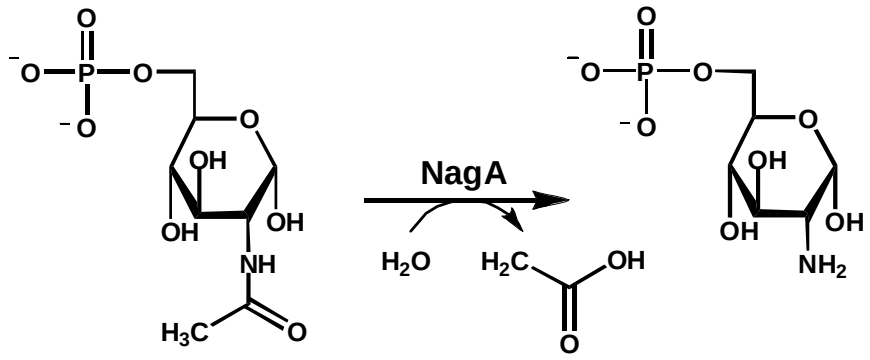
Scheme 1.4



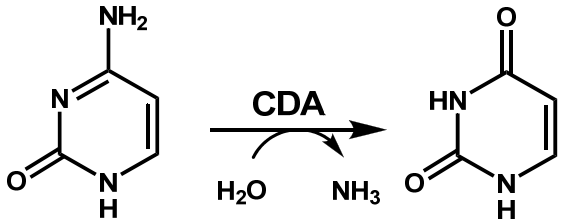
The primary focus of the research presented here was to evaluate, compare and contrast two enzymes representative of the less understood β- and α-mononuclear classes of amidohydrolase enzymes. The second goal of the research presented here was to use the structural, mechanistic and genomic information available, and apply it toward the discovery of function of previously uncharacterized enzymes. NagA and CDA were chosen as model representatives for mechanistic characterizations, based on previous reports that both of the enzymes were soluble and had only been partially characterized. NagA (a β-mononuclear enzyme) is a necessary participant in the metabolism of

aminosugars, where a number of different pathways may produce the substrate for this enzyme. NagA catalyzes the deacetylation of *N*-acetyl-D-glucosamine-6-phosphate (GlcNAc6P) to form D-glucosamine-6-phosphate (Gln6P) and acetate as shown in Scheme 1.5. CDA is an α-mononuclear enzyme and catalyzes the deamination of cytosine, forming uracil and free ammonia as shown in Scheme 1.6. This is an important reaction in the recycling and metabolism of pyrimidine nucleotides. The cytosine substrate for CDA may come from the degradation of DNA and RNA and cytidine nucleotides.

Scheme 1.5

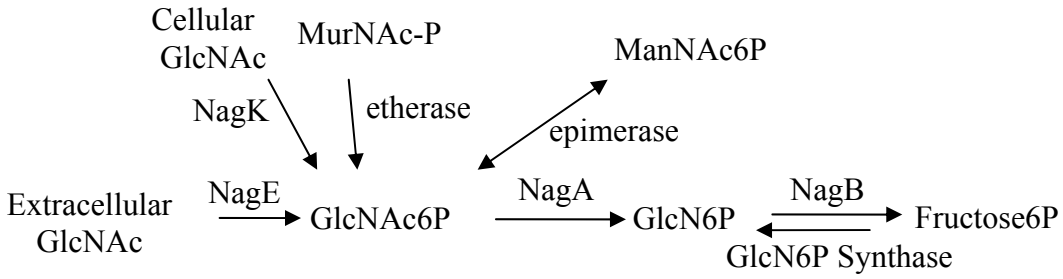


Scheme 1.6



The deacetylation of GlcNAc6P provides a source of carbon and nitrogen and prepares this substrate for entry into the glycolysis pathway. GlcNAc may be produced as a product of chitobiose degradation by the enzyme NagZ. Extracellular GlcNAc can become phosphorylated and transported into the cell with the help of a phosphoenolpyruvate utilizing phosphotransferase called NagE. Cellular GlcNAc6P can also be produced from the phosphorylation of GlcNAc from cell wall murein by a specific kinase named NagK (*90*). Another branch of the murein recycling pathway produces *N*-acetylmuramic acid-phosphate which is then modified by MurQ Etherase to form GlcNAc6P (*91*). These reactions are summarized in Scheme 1.7.

Scheme 1.7



Once in the cell, GlcNAc6P performs several functions, in addition to being deacetylated by NagA. In one function, GlcNAc6P binds to the transcription regulator NagC, preventing it from binding to its operator in the NagE-NagABCD operon. This results in the initiation of transcription for NagE, NagA, and NagB, which are sequentially responsible for the transportation and phosphorylation of GlcNAc, the deacetylation of GlcNAc6P and the deamination and isomerization of GlcN6P to form

fructose-6-phosphate. GlcNAc6P is also an allosteric K-type activator of NagB, which is the final step before entry into the glycolysis pathway. In a competing reaction, if sources of aminosugars are unavailable, GlcN6P can be produced by GlcN6P synthase from glutamine and fructose-6-P.

A previous study has analyzed the forward and reverse kinetics and mechanism of NagA from *E. coli* (*68*). In the forward reaction, it was determined that substrate inhibition is present for GlcNAc6P, while the products GlcNAc and acetate are non-competitive and competitive inhibitors, respectively. This is explained as the possible result of an ordered release of products. The substrate inhibition may arise from the addition of substrate to the enzyme with acetate still in the active site. In this study however, little attention was focused on the involvement of metal in the active site. The role of a metal cofactor is of particular interest in NagA from *E. coli*, because it contains a QxN motif at the carboxyl end of β-strand 1, which replaces the HxH M_{α} binding motif that is found in NagA from *T. maritima* and *B. subtilis*. The likely result of the QxN motif is an inability to bind metal at the α-metal binding site, resulting in a mononuclear metal center in the β-metal binding site bound with a water molecule and three amino acid ligands.

The effects of this active site deviation may be seen in Figure 1.1, which shows an alignment of the active site residues for NagA from *T. maritima* (PDB code: 1o12), *B. subtilis* (PDB code: 1un7) and *E. coli* (PDB code: 1ymy). While the active site residues are in similar positions, the α-metal binding site in *T. maritima* is unoccupied as crystallized by The Joint Center for Structural Genomics. This may be significant in

light of the lack of the α-metal and typical α-metal binding motif in the *E. coli* enzyme. The active site QxN motif is also found in NagA from *H. sapiens*. In another crystallographic study of NagA from *E. coli*, NagA was crystallized without a metal in the active site (*65*).

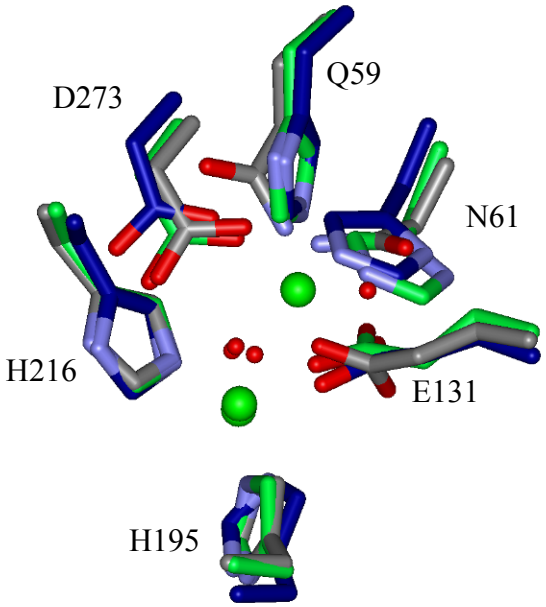


Figure 1.1: Structural alignment of the catalytic residues of NagA from *B. subtilis* (PDB code: 1un7, green carbons) , *T. maritima* (PDB code: 1o12, blue carbons) and *E. coli* (PDB code: 1ymy, grey carbons). Residue assignments correspond to *E. coli* NagA.

Prior to the work presented here, there was a lack of information pertaining to the mechanism of catalysis for NagA. It had not yet been demonstrated which active site residues were responsible for the formation of the metal center and for binding and catalysis. There was also no report on how the active site metal interacts with the substrate and how the hydrolytic water molecule becomes activated. In consideration of

the absence of the two histidine α-metal ligands from strand 1, it was anticipated that NagA from *E. coli* K-12 utilizes only the β-metal for catalysis. This body of work was aimed at achieving a better understanding of how nature has compensated for the loss of the α-metal which aids in water activation and transition state stabilization. The second chapter of this dissertation will focus on the mechanism of NagA and the third chapter will focus on the structure and evolution of the NagA active site.

The α-mononuclear cytosine deaminase from *Escherichia coli* K-12 (CDA) catalyzes an important reaction in the metabolism of pyrimidines, by performing the nucleophilic aromatic substitution by water in the deamination of cytosine to form uracil and free ammonia. Cytosine may be produced from cytidine through the hydrolytic action of a pyrimidine nucleosidase, or from CMP through a pyrimidine nucleotide N-ribosidase. Uracil produced from the deamination of cytosine may then participate in a large number of metabolic pathways, which include the production of deoxy-uridine, uridine, pseudouridine, β-alanine, or malonate and urea.

Bacteria, fungi and yeast possess a cytosine deaminase; however, mammals have only a cytidine deaminase (*92*). This aspect has been exploited through the use of 5-fluorocytosine as a fungicide and in the treatment of solid tumors, where the tumors were transfected with a gene for CDA and then treated with 5-fluourocytosine (*93*). This drug is non-toxic until it is deaminated to 5-fluorouracil, which inhibits DNA replication, thus killing the transfected cells.

A previous study of bacterial CDA found that it requires one equivalent of Fe^{2+} for maximum activity although it was also active with Mn^{2+} Co^{2+} and Zn^{2+}. The Zn^{2+}

containing enzyme was the least active, possessing only 10% of the activity seen for the Fe^{2+} enzyme at pH 7.5 (*22*). In that study, an attempt to prepare apo-CDA from enzyme which contained 0.8 equivalents of Fe and 0.2 equivalens of Zn, yielded enzyme with 0.2 equivalents of Zn. Later studies discovered that bacterial CDA oligomerizes as a homo-hexamer, with a molecular weight of about 48 kDa per subunit (*19*).

CDA utilizes a TIM barrel structural fold and coordinates Fe^{2+} in a trigonal-bipyramidal fashion through the histidine residues 62, 64 from the end of β-strand 1, His-214 from strand 5, Asp-314 from strand 8 as well as a nucleophilic water molecule. It is interesting to note that the structure of yeast CDA forms as a dimer with a 17 kDa molecular weight per subunit (*94*). The yeast CDA utilizes a half-barrel structural fold, with a tetrahedrally coordinated Zn^{2+} bound by two cysteines, a histidine and a water molecule. Bacterial CDA, a member of the amidohydrolase superfamily, is most similar in structure to adenosine deaminase (ADA), while the structure of yeast CDA is highly similar to that of cytidine deaminase (*23*). As seen from crystal structure comparisons of several deaminases including *E. coli* CDA with 4-(R)-hydroxyl-3,4-dihydropyrimidine (*23*), murine ADA with 6-hydroxyl-1,6-dihydropurine ribonucleoside (*17*), and cytidine deaminase with bound uridine (*95*), there were several unifying characteristics: a divalent metal is used in the activation of the hydrolytic water, and a glutamate residue is located close enough to protonate the endocyclic N3 of the pyrimidine ring (*23*). This may promote the enol to keto tautomerization which may be required for collapse of a metal bound tetrahedral intermediate followed by product release.

Prior to this work, there was little explanation for the apparent discrepancies seen between the activities of the Fe^{2+} and the Zn^{2+} bound enzyme forms, considering that the very similar enzymes guanine deaminase (PDB code: 2UZ9) and adenosine deaminase (*17*) were both isolated with a single bound Zn^{2+} ion. While there was already a crystal structure of this enzyme bound with a mechanism based inhibitor, it contains an Fe^{3+} metal ion. It was shown that CDA loses activity during purification and that the loss of activity was due to the oxidation of the Fe^{2+} (*23*). Therefore, some of the active site interactions shown may not be catalytically relevant, due to the charge difference. There was also very little known about the active site residues involved in substrate binding and catalysis and it was unknown how many and which active site bases are involved in the eventual abstraction of two protons from the hydrolytic water and the eventual protonation of the N-3 atom of cytosine. Prior to this work, it was also poorly understood which features of the substrate are recognized by the enzyme and are required for catalysis to occur. The fourth chapter of the present study will address the mechanism and structure of cytosine deaminase.

It was noted by White and co-workers (*96*) that a deaminase involved in the riboflavin and coenzyme F420 biosynthetic pathways of archaeal organisms has not been identified. The work presented in Chapter V describes the experiments performed in the search for the unidentified pyrimidine deaminase involved in the riboflavin biosynthesis pathway of archaea. The third protein studied in this body of work is a protein of unknown function that was selected for study based on sequence comparisons and genomic context evaluations. Genomic context evaluations suggested that the genes

of unknown function labeled Mm0823, MmarC7 0625 and Sso0398 are associated with the riboflavin biosynthesis pathway of archaea and sequence analyses assigned the proteins as chlorohydrolases related to SsnA and AtzA. Alignments with known members of the AHS show strong similarity to CDA and GDA. Taken together, these observations provide strong support that these proteins of unknown function represent the missing archaeal pyrimidine deaminase. The work presented here reports the cloning, expression and attempted purification and assay development for those three proteins.

Chapter VI of this work describes a different route toward the discovery of function of two novel 8-oxoguanine deaminases. The genes of unknown function Pa0142 and the NYSGXRC target 9236e (PDB code: 3HPA), from an environmental sample of the Sargasso Sea, appear to be located in a nucleotide salvage pathway of the bacteria, near genes which are very similar to known adenosine and guanine deaminases. Annotated as AtzB-like, the sequences of Pa0142 and 9236e are highly analogous to AtzB, GDA and to a lesser extent; CDA, SsnA and YahJ. It can be seen in the crystal structure of GDA from *H. sapiens* (PDB code: 2UZ9), and other GDA enzymes, that there is a conserved arginine which forms multiple interactions with the bound product xanthine. In these two proteins of unknown function, the residue found to be in that location is identified as a cysteine. That difference, when combined with the presence of the GDA-like protein nearby in the genome, suggests that these two unknown enzymes utilize a slightly different substrate. Pa0142 and 9236e were tested for deaminase activity against a library of pteridines, purines and pyrimidines containing primary amines. It was discovered that these enzymes catalyze the deamination of 8-oxoguanine,

guanine and isocytosine. 217 genes listed in Appendix A were identified which are predicted to deaminate 8-oxoguanine.

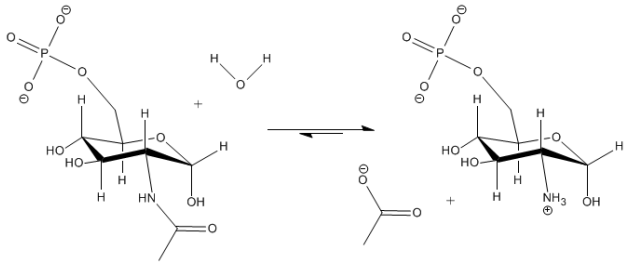
Chapter VII describes the discovery of function of two isoxanthopterin deaminases. A small group of 24 bacterial proteins which share a high degree of sequence similarity to guanine deaminases and 8-oxoguanine deaminases were identified. These enzymes share a critical difference at a location which provides substrate recognition for both guanine deaminases and the 8-oxoguanine deaminases. At this location, an arginine is conserved among guanine deaminases, a cysteine is conserved among 8-oxoguanine deaminases and in this small group of proteins a threonine is conserved. From genomic context analyses, these enzymes were found have possible pathway involvements with dioxygenases, decarboxylases and purine/pyrimidine transporter proteins. From this distinct group of enzymes, the SGX targets 9339a (PDB code: 2PAJ) and 9236b were tested for deaminase activity against a library of pteridines, purines and pyrimidines containing primary amines. From those screens, new enzymatic activities were discovered. 9339a and 9236b catalyze the deaminations of isoxanthopterin and pterine-6-carboxylate. These reactions appear to be involved in a poorly understood flavin degradation pathway which is proposed to involve decarboxylations and oxidations.

CHAPTER II

N-ACETYL-D-GLUCOSAMINE-6-PHOSPHATE DEACETYLASE: SUBSTRATE ACTIVATION VIA A SINGLE DIVALENT METAL ION*

N-acetyl-D-glucosamine-6-phosphate deacetylase (NagA) catalyzes the hydrolytic cleavage of *N*-acetyl-D-glucosamine-6-phosphate as illustrated in Scheme 2.1. The deacetylation of this compound provides a source of carbon and nitrogen by preparing this substrate for entry into the glycolytic pathway. This reaction is a key step in the catabolism of *N*-acetyl-D-glucosamine, derived from the degradation of chitin, and is an essential component of the biosynthesis of lipopolysaccharides and peptidoglycans. More recently, the reaction catalyzed by NagA has been shown to be an important step in the recycling of cell wall murein (*97*).

Scheme 2.1



* Part of this chapter is reprinted with permission from "Tight Binding Inhibitors of *N*-Acyl Amino Sugar and *N*-Acyl Amino Acid Deacetylases" by Xu, C., Hall, R., Cummings, J., and Raushel, F., 2006. *J. Am. Chem. Soc.*, 128, 4244-4245, Copyright [2006] by American Chemical Society.
Part of this chapter is reprinted with permission from "*N*-Acetyl-D-glucosamine-6-phosphate Deacetylase: Substrate Activation via a Single Divalent Metal Ion" by Hall, R. S., Xiang, D. F., Xu, C., and Raushel, F. M., 2007. *Biochemistry,*46, 7942-7952, Copyright [2007] byAmerican Chemical Society.

The purification of NagA from *Escherichia coli* was originally reported by White and Pasternak (*70*). The enzyme oligomerizes as a tetramer and each subunit contains a reactive sulfhydryl group near the active site (*68*). In the forward reaction, inhibition occurs at high substrate concentrations while the products, D-glucosamine-6-phosphate and acetate, are noncompetitive and competitive inhibitors, respectively. A phosphonate analogue of the putative tetrahedral intermediate has been synthesized that is a potent and tight-binding inhibitor (*98*). The structure of apo-NagA from *E. coli* was first reported by a team led by Steve Almo (PDB code: 1YMY) and later by Ferreira *et al*. (*65*) (PDB code: 1YRR). The protein folds as a $(\beta/\alpha)_8$-barrel and is a member of the amidohydrolase superfamily (*65*). The enzyme binds up to 1.4 equivalents of zinc and the metal can be removed with a chelating agent, resulting in the loss of activity that can be restored after reconstitution with Zn, Co, Mn or Fe (*65*).

The role of a metal cofactor is of particular interest for NagA because the two crystal structures available for the enzyme from *E. coli* have no metal bound in the active site. However, the crystal structures of NagA from *B. subtilis* and *T. maritima* have been determined with divalent cations bound within the active site (*69*). The *B. subtilis* enzyme (PDB code: 1UN7) binds two equivalents of Fe^{2+} (*69*). One of the metals is coordinated to the HxH motif at the end of β-strand 1 and an aspartic acid from the end of β-strand 8. This site has been designated as the M_α-position (*10*). The second metal (M_β-site) is coordinated to two histidine residues from the ends of β-strands 5 and 6, respectively. The two metal ions are bridged to one another by a glutamate from the end of β-strand 3 and a hydroxide (or water) from solvent.

The active site of the *T. maritima* enzyme is quite similar to that of the *B. subtilis* structure except that a single metal ion is bound to the M_{β}-site (PDB code: 1O12). The metal ligation scheme in the NagA from *E. coli* must be significantly different since the HxH motif from the end of stand 1 is replaced by QxN and thus it is highly likely that this enzyme can bind only a single divalent cation in the active site. This fundamental divergence in the active sites of NagA from these sources suggests that the mechanism for the activation of the hydrolytic water and substrate required for catalysis must differ to a significant degree.

NagA has been characterized as a member of the amidohydrolase superfamily based on an exhaustive set of sequence and structural comparisons (*10*). This superfamily is a diverse group of enzymes that catalyzes the hydrolytic cleavage of amide bonds to nucleic acids, amino acids, and sugars (*10, 12*). Within this superfamily, four metal ligation schemes have been identified. The most prevalent is a binuclear metal center where the two divalent cations are bridged by a hydroxide from solvent. An additional bridging ligand from the protein may include a carboxylated lysine from the end of β-strand 4, a glutamate from the end of β-strands three or four, or a cysteine from the end of β-strand 2 (*10, 12*). Enzymes of this type include dihydroorotase (*31*), the phosphotriesterase homology protein (*78*), renal dipeptidase (*45*) and D-amino acid deacetylase *(69)*. Many members of the amidohydrolase superfamily bind a single divalent cation, including cytosine (*23*) and adenosine deaminase (*17*). This class of enzymes binds one metal at the α-position via ligation to the HxH motif from β-strand 1, a histidine from the end of β-strand 5 and an invariant asparate from β-strand 8. The

third ligation scheme, which may include NagA from *E. coli* and *T. maritima,* utilizes a single metal bound exclusively to the β-site. The most extreme example within the amidohydrolase superfamily is uronate isomerase from *E. coli*. This enzyme catalyzes the isomerization of glucuronate to fructuronate and a metal ion is not required for catalytic activity (*63*).

There are a significant number of issues regarding the mechanism for substrate hydrolysis by NagA that are unresolved. It is unclear how the divalent cation binds within the active site of NagA and how the water molecule is activated for nucleophilic attack on the amide bond. The disparity in the metal ligation schemes between the *E. coli* and *B. subtilis* forms of NagA suggests a divergence in the evolution of this enzyme at a significant locale within the active site. In this paper, we have systematically interrogated the mechanism of substrate hydrolysis by NagA from *E. coli* with a battery of substrate analogues and active site mutants using a host of metal-substituted forms of the enzyme. We conclude from these studies that NagA from *E. coli* can bind up to one divalent cation for the activation of the hydrolytic water molecule. This water molecule is further activated by general base catalysis through the abstraction of a proton by Asp-273. There is no evidence to suggest that Glu-131 plays any role in catalysis other than to help coordinate the single divalent cation. The enzyme is rate limited by the cleavage of the amide bond.

MATERIALS AND METHODS

Materials. *N*-Acetyl-D-glucosamine-6-phosphate (**1**), *N*-acetyl-D-glucosamine-6-sulfate (**3**), and all buffers and purification reagents were purchased from Sigma-Aldrich.

The substrates *N*-acetyl-D-galactosamine-6-phosphate (**2**) *N*-thioacetyl-D-glucosamine-6-phosphate (**4**), *N*-Formyl-D-glucosamine-6-phosphate (**5**), *N*-trifluoroacetyl-D-glucosamine-6-phosphate (**6**), as well as the *N*-methylphosphonamidate-D-glucosamine-6-phosphate inhibitor (**7**) were prepared by Dr. Chengfu Xu as reported elsewhere (*67, 98*). Chromatographic columns and resins were purchased from G. E. Healthcare. Chelex 100 resin was purchased from BioRad. Slide-A-Lyzer dialysis cassettes were purchased from Pierce. ICP standards were obtained from Inorganic Ventures Inc.

Cloning, Expression, and Purification. The gene from *E. coli* encoding NagA was cloned into a pET-30a (+) expression vector through the use of the *Nde*I and *EcoR*I restriction sites by Dr. Daofeng Xiang. Mutants of NagA were prepared in accordance with procedures published in the Quikchange Site-Directed Mutagenesis Kit. Gene sequences were verified by the Gene Technologies Lab at Texas A&M University. The expression plasmids were inserted into BL21 (DE3) cells through electroporation. The transformation solution was then transferred into 1 mL of LB growth medium for incubation at 37 °C for 30 min, which was followed by plating onto LB agar containing 50 μg/mL kanamycin. The cells were allowed to grow overnight at the same temperature. Single colonies were selected for inoculation into 50 mL of LB/kanamycin for use as overnight starter cultures for inoculation into 2 L of LB/kanamycin. Prior to induction with 1.0 mM IPTG at an $A_{600\ nm}$ of 0.6, 1.0 mM $ZnCl_2$ was added to supplement the growth medium. The cells grew overnight at a temperature of 30 °C, after which they were centrifuged at 3400*g* for 12 min. The cells were then resuspended and disrupted by sonication in 10 × (v/w) 50 mM Tris buffer, pH 7.5, containing 1.0 mM

DTT and 100 μg/mL of the protease inhibitor phenylmethanesulfonyl fluoride. Insoluble cell debris was removed by centrifugation at 13900*g* for 12 min, after which 1% w/v protamine sulfate from salmon sperm was used to precipitate the nucleic acids. The precipitate was removed by centrifugation and the protein fractionated with ammonium sulfate to 50% of saturation. The precipitated protein was isolated by centrifugation at 13900*g* for 12 min. The protein was then resuspended in a minimal amount of 50 mM Tris buffer, pH 7.5, containing 1.0 mM DTT and then passed through a 0.2 µM pore filter prior to loading onto a pre-equilibrated HiLoad 26/60 Superdex 200 prep grade gel filtration column. The pooled fractions containing NagA were diluted to 50 mL in the same buffer and loaded onto a 6 mL Resource Q anion exchange column. The column was washed with several column volumes of buffer and the protein eluted with a linear gradient of 1 M NaCl, 50 mM Tris, pH 7.5. The appropriate fractions were pooled after analysis with *SDS PAGE* for purity. Protein concentrations were determined using the calculated extinction coefficient of 18,490 M^{-1} cm^{-1} at 280 nm (www.scripps.edu/_cdputnam/protcalc.html).

Measurement of Enzymatic Activity. Kinetic assays were performed at 30 °C using a 96 well quartz plate in conjunction with a SpectraMax 384-Plus spectrophotometer from Molecular Devices. The data were analyzed using Softmax Pro version 4.7.1. Extinction coefficients for the different substrates were determined from plots of absorbance versus substrate concentration before and after total enzymatic hydrolysis.

Metal Analysis. Metal determination and quantification were performed with an Elan DRC II ICP-MS from Perkin-Elmer. An analog detection mode was used with three averaged replicates per reading. External calibration standards were prepared through the serial dilution of a single 10 ppm stock mixture of Zn, Cd, Co, Cu, Mn, Ni, and Fe in 2% nitric acid. Freshly prepared standards generally contained 2, 20, and 200 ppb of the metal ions in 1% Trace Select nitric acid from Fluka, diluted in MilliQ deionized water. The masses of the isotopes detected were ^{55}Mn, ^{57}Fe, ^{59}Co, ^{60}Ni, ^{66}Zn, and ^{111}Cd. ^{115}In was used as an internal standard for ^{111}Cd whereas ^{69}Ga was used as an internal standard for all other isotopes.

Metal Chelation and Reconstitution. Apoenzyme was prepared by dialysis of NagA against buffer A (20 mM dipicolinate, 10 mM β-mercaptoethanol, 50 mM MES, pH 6.0). After 3 × 300-fold buffer changes with buffer A over 2 days, the chelator was removed by dialysis with Chelex-treated HEPES buffer, pH 8.0. For the metal reconstitution studies, 100 μL of apo-NagA at 8.23 mg/mL was mixed with 3 equiv of various metal solutions. The metal solutions consisted of freshly prepared $NiCl_2$, $CdCl_2$, $CoCl_2$, $MnCl_2$, or $ZnCl_2$ in water, and $Fe(NH_4)_2(SO_4)_2$ in 1% HCl. Reconstitution of the apoenzymes was conducted overnight at 4 °C before removal of unbound metal by passage through a PD-10 column. The PD-10 column was pretreated with dipicolinate to remove traces of unbound metal and then washed with five column volumes of metal free HEPES buffer, pH 8.0. After elution, the concentration of the metal reconstituted enzyme was determined by UV absorbance and the metal content of the samples determined by ICP-MS.

Inhibitor Studies. A methyl phosphonamidate analogue of the tetrahedral intermediate for the reaction catalyzed by NagA (**I**) was synthesized by Dr. Chengfu Xu and the structure is presented in Scheme 2. The inhibition constant for NagA was determined by incubating variable amounts of **I** with 10-45 nM of the enzyme for 30 min in 50 mM phosphate buffer, pH 7.5. The enzyme activity was determined using 0.5 mM N-acetyl-D-glucosamine-6-phosphate as the substrate. The time dependence to the onset of inhibition by **I** with NagA was determined by measuring the rate of bond cleavage as a function of time. In these experiments, 10 nM NagA was added to a solution of 0.5 mM substrate containing 0.4-2.4 μM **I**.

pH Studies. The pH dependence of the kinetic constants, k_{cat} and k_{cat}/K_m, were determined at intervals of ~0.25 pH units from pH 5.0 to 10 using 20 mM piperazine (pH 5-6.25), phosphate (pH 6.25-8), and borate (pH 8-10) with varying amounts of substrate. The pH values of the assays were determined after the assays were completed.

Solvent Viscosity Analysis. Solvent viscosity effects were analyzed for Zn-NagA with *N*-acetyl-D-glucosamine-6-phosphate as the substrate using sucrose as the micro-viscogen in 20 mM phosphate buffer, pH 7.5. The effect of sucrose was performed using 0, 10, 14, 20, 24, 27.5, 32.5 and 35% (w/w) sucrose and the corresponding relative viscosities were 1, 1.32, 1.5, 1.88, 2.2, 2.48, 3.06, and 3.42 respectively (*99, 100*).

Data Analysis. All kinetic data were fit to the corresponding equations using the nonlinear least-squares curve fitting program SigmaPlot 9.0. Simple substrate saturation curves were fit using equation 2.1 where A is the substrate concentration, v is the velocity of the reaction, k_{cat} is the turnover number and K_m is the Michaelis constant.

When substrate saturation was not achieved, k_{cat}/K_m values were determined from fits of the data with a straight line. Saturation curves showing substrate inhibition were fit using equation 2.2 were K_{is} is the apparent inhibition constant for the substrate inhibition. For the analysis of pH-rate profiles, plots of log k_{cat} and log k_{cat}/K_m vs. pH that indicated the deprotonation of a single group required for maximum activity were fit with equation 2.3, where c is the maximum activity and K_a is the acid dissociation constant. pH-rate profiles that showed the loss of catalytic activity at high and low values of pH were fit with equation 2.4, where K_b is the dissociation constant of the group that must be protonated for full activity. The pH profiles which showed that the deprotonation of two acidic groups and the protonation of one basic group are required for maximum activity were fitted to equation 2.5 where K_a is the average dissociation constant of the two acidic groups and K_b is the dissociation constant of the basic group. Equation 2.6 was used to determine the average K_a for two ionizable acidic groups and the average K_b for two ionizable basic groups.

For the evaluation of the inhibition constants (K_i) for the methyl phosphonamidate inhibitor (**I**) of NagA, eq 2.7 was fit to plots of the residual activities vs inhibitor concentrations. E_t is the total enzyme concentration, I is the inhibitor concentration, v_o is the activity of the enzyme in the absence of inhibitor, and v_i is the residual activity of the enzyme in the presence of inhibitor. For the determination of the time dependence to the onset of inhibition, time courses corresponding to the change in rate of cytosine deamination were fit with eq 2.8, where P is the product concentration, v_r and v_s are the initial and final rates of product formation, and k_{obs} corresponds to the

first order rate constants for the onset of the slower rate of product formation. The apparent second-order rate constants (appk_{obs}) for the formation of the inhibited form of Zn-CDA, was determined from the slope of the plots of k_{obs} vs the concentrations of inhibitors. Correction for the concentration of substrate according to eq 2.9, gives the second-order rate constant k_1.

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

$$v / E_t = k_{cat} A / (K_m + A + (A^2 / K_{is})) \quad (2.2)$$

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

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

$$y = \log(c / (1 + [H^+] / K_a + [H^+]^2 / K_a^2 + K_b / [H^+])) \quad (2.5)$$

$$y = \log(c / (1 + [H^+] / K_a + [H^+]^2 / K_a^2 + K_b / [H^+] + K_b^2 / [H^+]^2)) \quad (2.6)$$

$$v_i/v_o = ([E_t] - K_i - [I] + (([I] + K_i - [E_t])^2 + (4K_i[E_t]))^{1/2})/(2[E_t]) \quad (2.7)$$

$$P = v_s t + ((v_r - v_s)/k_{obs})(1\text{-}e^{(-k_{obs}t)}) \quad (2.8)$$

$$k_1 = \text{app}k_{obs} (1 + [S]/[K_m]) \quad (2.9)$$

RESULTS AND DISCUSSION

Purification and Properties of NagA. NagA was the major soluble protein after induction by IPTG as estimated from SDS-PAGE. From 3 grams of cell paste about 100 mg of homogeneous protein were isolated. Enzymatic activity was dependent upon the presence of disulfide reducing reagents such as dithiothreitol or β-mercaptoethanol during storage and purification. The native molecular weight of the purified enzyme was estimated to be 180 kDa based upon the elution volume through a calibrated gel filtration column (data not shown). The molecular weight of a single subunit from the

derived amino acid sequence of NagA is 40,951 Da. These results confirm that NagA adopts a tetrameric quaternary structure in the presence of disulfide reducing agents (*65, 68*).

Metal Dependence of Enzyme Activity. The initial purification of NagA resulted in enzyme with a turnover number of 35 s^{-1}. This protein contained 0.4 equivalents of Zn^{2+} and 0.05 equivalents of Fe^{2+} as determined by ICP-MS. NagA was subsequently expressed in LB medium supplemented with 1.0 mM $ZnCl_2$ and the protein purified from these cells contains 0.95 equivalents of Zn per subunit with a specific activity of 96 s^{-1}. The role of the bound metal on the enzymatic reaction rate was addressed by characterization of the apo-enzyme. The metal was removed from the enzyme via dialysis against 20 mM dipicolinate and the resulting protein solution was verified to be metal-free by ICP-MS. However, during the time course for measurement of catalytic activity by the apo-enzyme, the rate of substrate turnover increased with time. The activation of the apo-enzyme was apparently due to the binding of trace metals in the assay solution since the activity of the apo-enzyme in the presence of 500 μM EDTA was less than 2% of the activity exhibited by the native enzyme. These results are consistent with an absolute requirement for a bound divalent cation for the expression of catalytic activity by NagA.

Incubation of the apo-enzyme with one equivalent of Zn^{2+} for 30 minutes fully restored catalytic activity. The titration of apo-enzyme with varying amounts of $ZnCl_2$ is shown in Figure 2.1. These results demonstrate that NagA from *E. coli* has a maximum catalytic activity with a single divalent cation bound to the active site. These results are

consistent with the absence of the two histidines at the end of strand one and the determination of the x-ray structure of the Zn-NagA that shows a single Zn bound to the M_β-site (*66*). The apo-enzyme was reconstituted with three equivalents of Fe^{2+}, Mn^{2+}, Ni^{2+}, Cd^{2+}, Co^{2+} or Zn^{2+} and the metal substituted forms of NagA were found to contain approximately one equivalent of metal per protein subunit after removal of the excess metal by dialysis. The metal content and the corresponding kinetic parameters for the metal substituted variants of NagA are listed in Table 2.1.

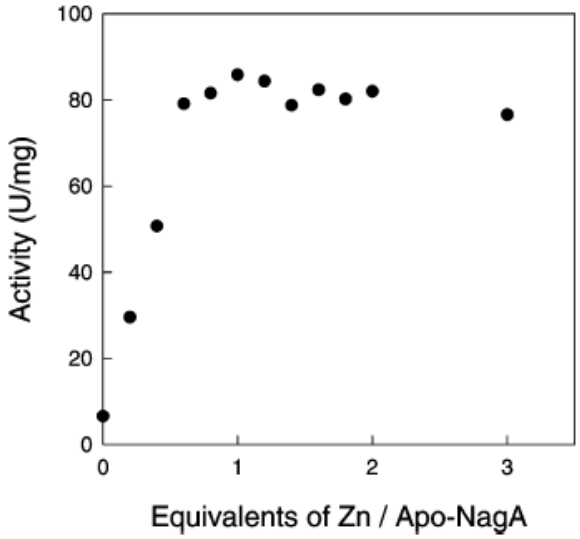


Figure 2.1: Reconstitution of apo-NagA (0.5 μM) with varying amounts of Zn. The rate of substrate hydrolysis was determined after diluting the enzyme to a concentration of 30 nM in 1.0 mM *N*-acetyl-D-glucosamine-6-phosphate, 1.0 μM EDTA and 25 mM Tris at pH 7.5 and 30 ºC.

Table 2.1: Kinetic Parameters for Metal-Reconstituted Forms of NagA[a].

Metal	Metal/subunit	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
Zn	0.9	102 ± 2	0.08 ± 0.01	$(1.2 \pm 0.2) \times 10^6$
Cd	0.8	163 ± 3	0.20 ± 0.01	$(8.2 \pm 0.4) \times 10^5$
Co	1.0	227 ± 5	0.15 ± 0.01	$(1.2 \pm 0.1) \times 10^6$
Ni	0.8	41 ± 3	0.64 ± 0.12	$(6.4 \pm 1.3) \times 10^4$
Mn	0.7	92 ± 2	0.10 ± 0.01	$(9.2 \pm 0.9) \times 10^5$
Fe	0.5	58 ± 2	0.23 ± 0.03	$(2.5 \pm 0.5) \times 10^5$

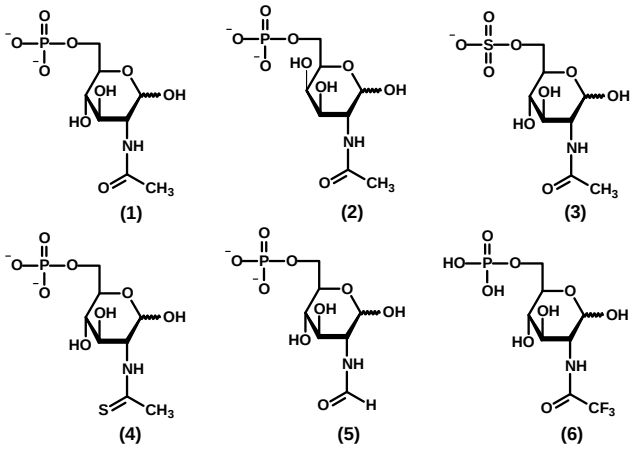
[a]The kinetic parameters were determined at pH 7.5, 30 ºC from fits of the data to eq 1 using *N*-acetyl-D-glucosamine-6-phosphate (**1**) as the substrate.

Kinetic Constants for NagA. The effect of metal substitution on the catalytic constants for NagA can be utilized to evaluate the contributions made by these divalent cations on enzymatic activity (*38, 51*). The relationship on the value of k_{cat}/K_m observed for the different metals follow the trend: Zn = Co > Mn > Cd > Fe > Ni . The Fe containing enzyme was not fully reconstituted with metal and thus the kinetic constants may be larger if the metal site was fully occupied. The values for k_{cat} follow the trend Co > Cd > Zn > Mn > Fe > Ni. The Ni enzyme has the highest K_m while the Zn enzyme has the lowest. Similar trends in the range of the catalytic constants were reported previously for the Co, Zn, Mn and Fe reconstitutions of apo-NagA from *E. coli* (*65*).

Substrate Specificity. The structures of the compounds tested as substrates for NagA are presented in Scheme 2.2. Cd- and Zn-NagA were used to address the occurrence of direct interactions of the carbonyl oxygen/sulfur and the metal ion through a comparison of the kinetic constants for *N*-acetyl-D-glucosamine-6-phosphate (**1**) and *N*-thioacetyl-D-glucosamine-6-phosphate (**4**) as substrates. The Cd-NagA catalyzes the

hydrolysis of the thioacetyl substrate about an order of magnitude faster than does the Zn-substituted enzyme (Table 2.2). The ratio of (k_{cat}/K_{thio}) / (k_{cat}/K_{acetyl}) is 0.78 (6.4×10^5 $M^{-1}s^{-1}$/8.2×10^5 $M^{-1}s^{-1}$) for the Cd-enzyme whereas the corresponding ratio for the Zn-NagA is only 0.036 (4.3×10^4 $M^{-1}s^{-1}$/1.2×10^6 $M^{-1}s^{-1}$). These results are consistent with the hard and soft interactions of metals and ligands, since the relatively soft thio-carbonyl group is potentially a better ligand for the softer cadmium ion than it is for zinc (*101*). If the softness of the carbonyl group is the only factor affecting these differences, then Mn, which is harder than Zn, should yield an even smaller ratio. However, the ratio of k_{cat}/K_m for the two substrates with Mn-NagA is essentially the same as it is for the Zn-substituted enzyme (0.050 *vs*. 0.036)).

Scheme 2.2



These results differ slightly from those observed for carboxypeptidase A, where thioamide substrate analogues were compared to the corresponding oxo-amides with the

Zn-, Cd-, and Mn-substituted enzymes (*102*). For carboxypeptidase the Zn-enzyme was the best catalyst for the hydrolysis of oxo-amides, while the Cd-enzyme was the least efficient. For the thioamide substrates, the Cd-enzyme was the best and the Mn-enzyme was the poorest catalyst. Nevertheless, the catalytic constants for the Zn- and Cd-substituted forms of NagA support the conclusion that the carbonyl group of the substrate is polarized by a direct interaction with the single divalent cation bound to the M_{β}-position in the active site.

In addition to N-acetyl-D-glucosamine-6-phosphate (**1**) and the thioacetyl analogue (**4**), NagA can hydrolyze *N*-trifluoroacetyl-D-glucosamine-6-phosphate (**6**), *N*-acetyl-D-glucosamine-6-sulfate (**3**) and *N*-formyl-D-glucosamine-6-P (**5**). The kinetic constants for NagA with these substrate analogues are listed in Table 2.2**.** The *N*-trifluoroacetyl substituted substrate is hydrolyzed 26 times faster than the natural substrate at saturating conditions but the *N*-formyl substrate is hydrolyze more slowly by a factor of 15.

Table 2.2: Kinetic Parameters for Substrates with Metal and Mutant Variant Forms of NagA[a].

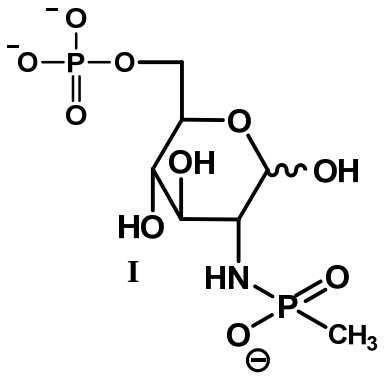
Substrate	Enzyme	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
1	Zn	102 ± 2	0.08 ± 0.01	$(1.2 \pm 0.2) \times 10^6$
1	Cd	163 ± 3	0.20 ± 0.01	$(8.2 \pm 0.4) \times 10^5$
1	Mn	92 ± 2	0.10 ± 0.01	$(9.2 \pm 0.9) \times 10^5$
2	Zn	154 ± 15	1.2 ± 0.2	$(1.2 \pm 0.2) \times 10^5$
3	Zn	64 ± 4	4.9 ± 1.0	$(1.3 \pm 0.3) \times 10^4$
3	Cd	23 ± 1	11 ± 1	$(2.1 \pm 0.2) \times 10^3$
3	Zn-K139M	60 ± 4	9.7 ± 1.3	$(6.2 \pm 0.9) \times 10$
4	Zn	10 ± 0.2	0.24 ± 0.02	$(4.3 \pm 0.4) \times 10^4$
4	Cd	128 ± 5	0.20 ± 0.03	$(6.4 \pm 1.0) \times 10^5$
4	Mn	11 ± 0.3	0.23 ± 0.03	$(4.8 \pm 0.6) \times 10^4$
5	Zn	6.9 ± 0.2	0.13 ± 0.02	$(5.3 \pm 0.8) \times 10^4$
6	Zn	2612 ± 60	0.40 ± 0.04	$(6.6 \pm 0.7) \times 10^6$

[a]The kinetic parameters were determined at pH 7.5, 30 ºC from fits of the data to equation 1.

Inhibitor Studies. The enzymatic hydrolysis of acetate N-acetylglucosamine-6-phosphate may proceed via the direct attack on the amide bond by an activated water molecule. Compounds that mimic the structure and geometric constraints of the tetrahedral intermediate have proven to be useful probes of the reaction mechanism and tight binding enzyme inhibitors (*103*). Phosphonamidates are particularly useful structural surrogates because of the tetrahedral geometry at the phosphorus center and the similarities in charge and bond lengths to those predicted to occur during the hydrolysis of amide bonds (*104-107*).

The methyl phosphonamidate transition-state analog inhibitor (**I**) was selected as a probe of the affinity for the transition state by NagA for the deacetylation of **1.** The effectiveness of the inhibitor shown in Scheme 2.3 was evaluated with the Zn^{2+} containing NagA at pH 7.5. Inhibition constants were determined through inhibition pattern fits to John Morrison's quadratic equation for tight binding inhibitors (*108, 109*) as shown in eq 2.7.

Scheme 2.3



I

The methyl phosphonamidate analog of the tetrahedral intermediate for the reaction catalyzed by NagA (**I**) is a potent and time-dependent inhibitor of NagA. After incubation of NagA with variable amounts of **I** the residual activity was plotted versus the inhibitor concentration, and the results for 25 nM NagA are presented in Figure 2.2. A fit of the data to eq 2.7 for the entire data set gave an average inhibition constant, *K*i, of 34 ± 5 nM. This inhibitor is a close mimic of the transition state for the deacetylation of **1** by NagA and as such, the inhibition constant for compound **I** is 4×10^3 -fold smaller

than the Michaelis constant for the deacetylation of **1** by NagA. This K_i value is among the lowest ever measured for the inhibition of deacetylases (*110, 111*).

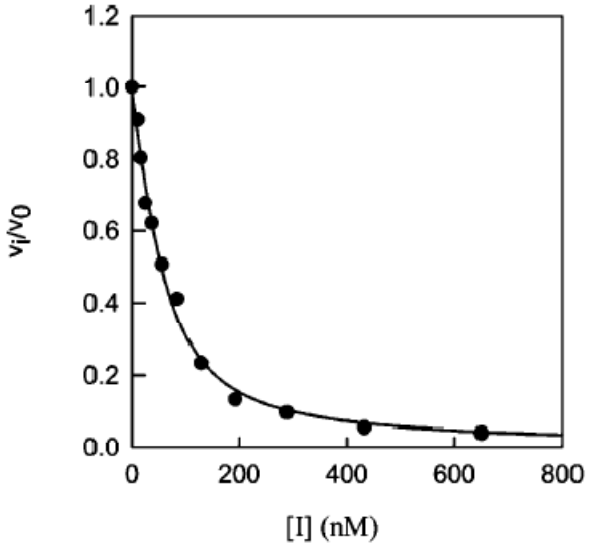


Figure 2.2: Inhibition of NagA by **I**. The kinetic data were fit to eq 2.7 to obtain a dissociation constant of 34 nM . Additional details are given in the text.

The time courses for the changes in the rates of deacetylation of **1** by NagA as seen below in Figure 2.3 were fit to eq 2.8, yielding the observed first-order rate constants (k_{obs}) for the slow onset of inhibition by **I**. The obtained rates (k_{obs}) were then plotted vs inhibitor concentrations yielding the apparent second-order rate constants (appk_{obs}) as seen below in Figure 2.4. The linearity of these plots indicate that inhibition follows a simple model for inhibitor binding shown below in Scheme 2.4.

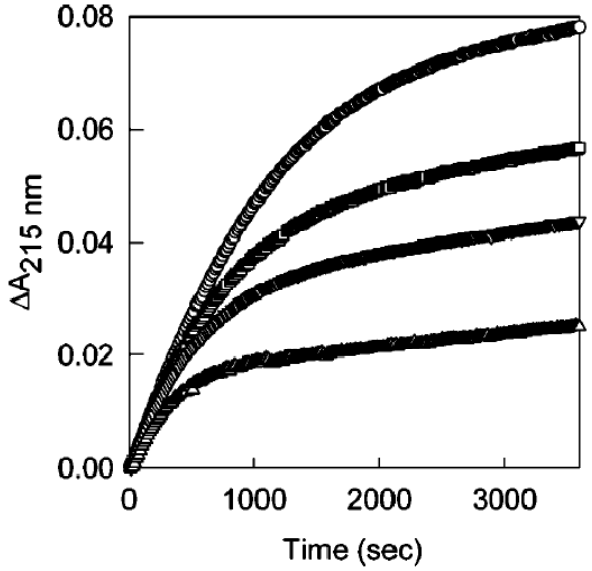


Figure 2.3: Time dependence for the onset of inhibition by 1 with NagA. The concentrations of **I** are: 0.40 μM (circle), 0.80μM (square), 1.6 μM (downward triangle), 2.4μM (upward triangle). The time courses were fit to eq 2.8 to obtain the first-order rate constants for the formation of the enzyme-inhibitor complex.

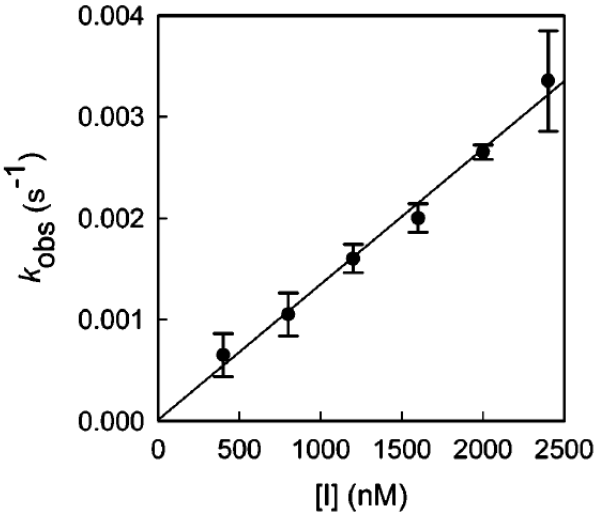


Figure 2.4: Effect of inhibitor concentration on the onset of inhibition by **I** with NagA. Additional details are provided in the text.

Scheme 2.4

$$E \underset{k_2}{\overset{k_1 I}{\rightleftharpoons}} EI$$

From the slope of the plot of k_{obs} versus the concentration of **I** in Figure 2.4, the apparent second-order rate constant is 1.3×10^3 M^{-1} s^{-1}. Correction of the apparent second-order rate for the concentration of substrate according to eq 2.9 gave a value for k_1 of 6.1×10^3 M^{-1} s^{-1} (see Scheme 2.4). The inhibitor was found to be reversibly inhibitory upon a 1:100-fold dilution of 500nM NagA which was pre-incubated with 1 μM of the inhibitor.

pH Rate Profiles. The effect of pH on the kinetic constants for the hydrolysis of substrates by NagA was measured in an attempt to identify those functional groups in the substrate and active site that are required to be in a specific state of protonation. The log k_{cat} vs. pH-profile shown in Figure 2.5 A for Zn-NagA was fit to equation 2.3 and these results indicate that the deprotonation of a single acid with a kinetic pK_a of 5.6 is required for catalytic activity in the enzyme-substrate complex. The log k_{cat}/K_m vs. pH-profile for Zn-NagA is shown in Figure 2.5 B. In this profile, catalytic activity is lost at high and low values of pH and the data were fit to equation 2.4 where two groups, in either the substrate or free enzyme, with identical pK_a values of 6.4 are required to be ionized. An additional group must be protonated with a pK_a of ~9.3. Similar results were obtained for the effect of pH on the kinetic constants for the hydrolysis of compound **1** by Cd-NagA (data not shown). From the log k_{cat} vs. pH profile, a single group must be ionized with a kinetic pK_a of 6.4. From the log k_{cat}/K_m pH-rate profile activity is lost at low and high pH with limiting slopes of 2 and -2, respectively, and the data were fit to equation 2.5.

For the log k_{cat} vs. pH profile, a single ionization is observed for either Zn- or Cd-NagA. This ionization must represent the protonation of the metal bound hydroxide that is utilized for substrate hydrolysis or, alternatively, the general base that is utilized to deprotonate the metal bound water molecule prior to substrate hydrolysis. The pK_a value is slightly higher for Cd-NagA than for Zn-NagA, as expected from the tendency of zinc to lower the pK_a of water relative to cadmium (*76*). For the log k_{cat}/K_m profiles, two ionizations are observed at low pH for both metal-substituted forms of the enzyme. One of these ionizations must be from the free enzyme and likely reflects the state of protonation for the group that must activate the metal-bound water molecule. From the x-ray crystal structure of NagA and homologous active site structures for other members of the amidohydrolase superfamily this residue must be the invariant carboxylate group, Asp-273, from the end of β-strand 8 (*12, 66*). The second ionization observed in the log k_{cat}/K_m profile could originate from other residues within the active site of NagA but it could also arise from the protonation of the substrate itself since the phosphate moiety has a pK_a of ~6.1 (*112*). If the enzyme requires (or highly prefers) a doubly ionized phosphate substituent for binding then a diminution in activity will be observed in the k_{cat}/K_m plot but not the k_{cat} vs. pH profile (*113*).

To address this issue experimentally, the pH-rate profiles for the hydrolysis of the sulfate analogue of the substrate (**3**) were measured. The sulfate derivative is a monoanion above pH 5, and thus no ionizations can originate from the protonation of this compound in the pH range available for the characterization of NagA (pH 5-10).

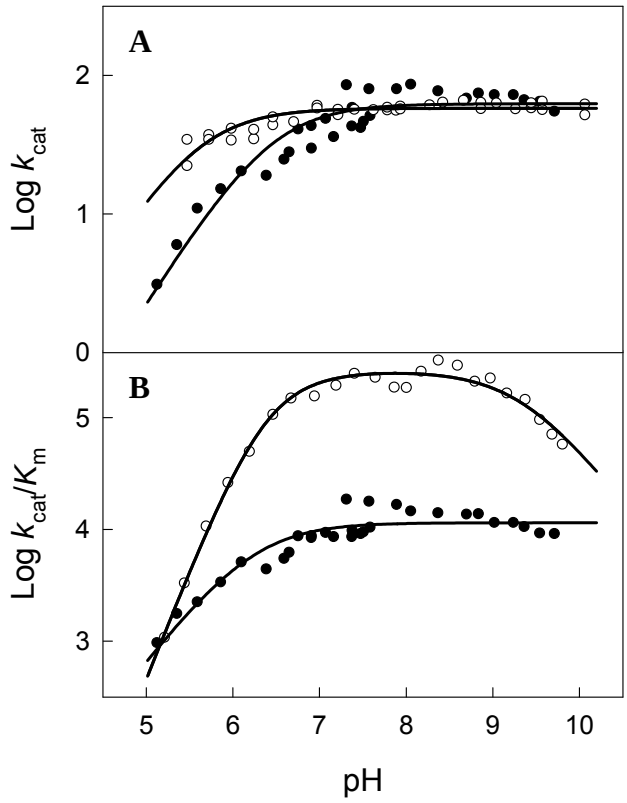


Figure 2.5: pH-rate profiles for the hydrolysis of substrates by NagA. (A) log k_{cat} vs. pH profiles for Zn-NagA with *N*-acetyl-D-glucosamine-6-phosphate (○) and *N*-acetyl-D-glucosamine-6-sulfate (●). The solid lines represent fits of the data to equation 2.3. (B) log k_{cat}/K_m vs. pH profiles for Zn-NagA with *N*-acetyl-D-glucosamine-6-phosphate (○) and *N*-acetyl-D-glucosamine-6-sulfate (●). The solid lines represent fits of the data to equations 2.5 and 2.3, respectively. Additional details are provided in the text and the dissociation constants from fits of the data to the appropriate equations are listed in Table 2.3.

Table 2.3: Kinetic pK_a Values from pH – Rate Profiles of Metal-Substituted Variants and Mutant.

Enzyme	Metal	Substrate	k_{cat} vs. pH pK_a	k_{cat}/K_m vs. pH pK_a	k_{cat}/K_m vs. pH pK_b
Wild type	Zn	**1**	5.6 ± 0.1	6.4 ± 0.1[a]	9.3 ± 0.4
Wild type	Cd	**1**	6.4 ± 0.1	6.1 ± 0.1[a]	9.1 ± 0.2[b]
K139M	Zn	**1**	5.4 ± 0.2	7.0 ± 0.1	9.1 ± 0.4
Wild type	Zn	**3**	6.5 ± 0.1	6.2 ± 0.1	-

[a]Average value of pK_a for two ionizations. [b]Average value of pK_b for two ionizations.

At saturating substrate the sulfate analogue (**3**) is hydrolyzed at approximately half the rate of the physiological substrate (**1**), although the K_m value is significantly higher (Table 2.2). For the pH-rate profiles (Figures 2.5 A and 2.5 B) a single ionization of pK_a 6.5 is observed in the log k_{cat} vs pH profiles and a *single* ionization of pK_a 6.2 is observed in the log k_{cat}/K_m profile (Table 2.3). These results are fully consistent with the conclusion that the second ionization observed in the pH rate profile for log k_{cat}/K_m with compound **1** is due to the protonation of the phosphate moiety to a monoanion. It is also curious to note that the ionization that is observed at high pH in the log k_{cat}/K_m vs pH-rate profiles for the natural substrate is not observed during the hydrolysis of the sulfate derivative. These results suggest that the monoanionic sulfate derivative does not require the interaction with an enzyme group that apparently ionizes in this pH range. The likely candidates for this group include Lys-139 and Tyr-223, based on the structure of D-glucosamine-6-phosphate bound to the enzyme from *B. subtilis* (*69*), and the inhibitor bound form of the D273N mutant from *E. coli* (*66*).

Mutation of Active Site Residues. Site specific mutants of NagA were constructed in order to ascertain the roles of specific residues in substrate recognition and catalytic function. The metal content and the kinetic parameters for the mutants of NagA constructed for this investigation are presented in Table 2.4. Residues Q59 and N61 were mutated together to a pair of histidine residues to create an HxH motif that is analogous to that found in NagA from *B. subtilis* and *T. maritima*. This alteration resulted in a decrease in k_{cat}, a large increase in K_m, and an increase in the average amount of Zn bound to the protein, suggesting that an additional metal ion can bind to this mutant. The

drop in catalytic activity may result from the binding of this second metal ion or may be due to steric crowding within the confines of an active site that has apparently evolved to operate with a single divalent cation. For the single histidine substitutions, the N61H mutant is similar in catalytic activity to the double mutant whereas the Q59H mutant is diminished by less than an order of magnitude relative to the wild type enzyme. For the substitutions at Asn-61, the replacement with an alanine is significantly less disruptive than the change to a histidine.

The X-ray structures of NagA from *E. coli* (*66*), *T. maritima* (PDB code: 1O12), and *B. subtilis* (*69*) indicate in each case that a glutamate from the end of β-strand 3 of the $(\beta/\alpha)_8$-barrel interacts with one or both of the divalent cations bound within the active site. Mutation of this residue to either glutamine or alanine results in mutant enzymes that have lost a significant amount of catalytic activity. The value of k_{cat} is reduced approximately 2 orders of magnitude, and the K_m for the E131Q mutant is increased to 3.6 mM. A significant portion of this reduction in catalytic efficiency has arisen because the metal content of the purified proteins is relatively low. Incubation of 25 μM E131A mutant with 15 equiv of $ZnCl_2$ enhanced the catalytic activity of this enzyme by a factor of 3.

In the crystal structure of the D273N mutant of NagA, the side chain of His-143 is 3.0 Å away from the phosphonate oxygen of the mimic of the tetrahedral intermediate (*66*). This result suggests that this group could facilitate the activation of the carbonyl group of the substrate in conjunction with an interaction with the metal ion bound within the active site. Mutation of H143 to an asparagine resulted in the dramatic loss of

catalytic activity and a moderate increase in the K_m, producing a 6000-fold decrease in k_{cat}/K_m. The H143Q mutant was more active, with a decrease in the value of k_{cat}/K_m of greater than 2 orders of magnitude. The H143Q mutant was isolated with 0.6 equiv of Zn, while the H143N mutant contained only 0.3 equiv. Overall; these results are consistent with the participation of H143 in the polarization of the substrate carbonyl group.

Table 2.4: Metal Content and Kinetic Parameters of NagA and Mutants[a].

Mutant	Zn/Subunit	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
Q59H/N61H	1.5	1.2 ± 0.1	5.4 ± 0.5	$(2.2 \pm 0.3) \times 10^2$
Q59H	1.3	32 ± 3	0.31 ± 0.06	$(1.0 \pm 0.2) \times 10^5$
Q59A	0.9	10 ± 0.4	0.08 ± 0.01	$(1.3 \pm 0.2) \times 10^5$
N61H	1.0	2.6 ± 0.1	1.7 ± 0.2	$(1.5 \pm 0.2) \times 10^3$
N61A	0.9	24 ± 1	0.80 ± 0.05	$(3.0 \pm 0.2) \times 10^4$
E131Q	0.1	0.7 ± 0.1	3.6 ± 1.1	$(1.9 \pm 0.6) \times 10^2$
E131A	0.2	1.9 ± 0.3	0.15 ± 0.06	$(1.3 \pm 0.6) \times 10^4$
K139M	0.9	49 ± 2	2.0 ± 0.2	$(2.5 \pm 0.3) \times 10^4$
H143Q	0.3	2.4 ± 0.1	0.33 ± 0.04	$(7.3 \pm 0.9) \times 10^3$
H143N	0.6	0.43 ± 0.02	2.1 ± 0.1	$(2.1 \pm 0.2) \times 10^2$
Y223F	0.8	167 ± 15	0.76 ± 0.11	$(2.2 \pm 0.4) \times 10^5$
H251N	0.9	6.8 ± 0.4	2.7 ± 0.4	$(2.5 \pm 0.4) \times 10^3$
D273N	0.7	< 0.02	-	-
D273A	0.8	< 0.02	-	-

[a]Kinetic parameters were determined at pH 7.5 and 30 ºC from fits to equation 1 using compound **1** as the substrate.

The most significant reduction in catalytic activity occurs with the mutation of the invariant aspartate (Asp-273) from the end of β-strand 8. In the X-ray crystal structure of NagA from *E. coli* this residue is hydrogen bonded to the lone water molecule (or hydroxide) that is coordinated to the metal ion bound to the M_β-site (*66*). Therefore, it is expected that this residue will function in catalysis by abstraction of a proton from water prior to, or concomitant with, the attack of hydroxide on the amide bond of the substrate. Within the detection limits of our assay for product formation, we were unable to measure any catalytic activity for either the D273N or D273A mutant. The loss of activity is not due to a diminished capacity to bind divalent cations since the metal content of the purified mutants was identical to that of the wild type enzyme. The catalytic properties of these mutants are fully consistent with the proposed role of this residue in the chemical mechanism as the primary general base for activation of the hydrolytic water molecule and the subsequent protonation of the leaving group amine.

The crystal structure of NagA from *B. subtilis* bound with the product, D-glucosamine-6-phosphate, indicates that an invariant arginine residue (equivalent to Arg-227 in the *E. coli* enzyme) from the adjacent subunit ion-pairs with the phosphate moiety of the substrate (*69*). It is unlikely that this arginine residue could contribute to the diminution in k_{cat}/K_m at high pH since the pK_a value for the ionization of the guanidino group is expected to exceed 12. However, in NagA from *E. coli* there are tyrosine (Y223) and lysine (K139) residues that may also be contributing to the interaction with the phosphate moiety of the substrate. Deprotonation of either of these residues at high pH could result in a drop in the value of k_{cat}/K_m. The mutation of Lys-139 to methionine

reduces the value of k_{cat} by a factor of 2 and the value of k_{cat}/K_m by a factor of ~50 when the phosphate derivative (**1)** is used as the substrate, relative to the wild type enzyme. However, there is only a 2-fold drop in k_{cat}/K_m relative to the wild type enzyme when K139M is used to hydrolyze *N*-acetyl-D-glucosamine-6-sulfate (**3**). The pH-rate profile for K139M still exhibits a drop in activity at high pH when *N*-acetyl-D-glucosamine-6-phosphate (**1**) is used as the substrate, and thus it is unlikely that the ionization state of this lysine is critical for catalytic activity. Mutation of Y223 to phenylalanine resulted in a mutant with nearly the same catalytic activity as the wild type enzyme, and thus this residue is not significant for the binding of substrate in NagA.

The structure of the D273N mutant of the *E. coli* NagA in the presence of the phosphonate inhibitor indicates that H251 interacts with the anomeric hydroxyl group at carbon 1 (*66*). The H251N mutant has a k_{cat} of 6.8 s^{-1} and a K_m of 2.7 mM. The value for k_{cat}/K_m is 1/400 of that measured for the wild type enzyme. These results confirm the importance for the role of H251 in the binding of substrate within the active site of NagA. The homologous residue was found to interact with D-glucosamine-6-phosphate in the structure of NagA from *B. subtilis*. This histidine residue (H258) was found to be 2.7 Å from the oxygen atom on the anomeric carbon in the α-conformation (*69*).

Rate Limiting Steps. Three experimental probes were marshaled with NagA to unveil the source of the rate limitation for this enzyme. The first of these experiments compared the rates of substrate hydrolysis in H_2O and D_2O. Using Zn-NagA with *N*-acetyl-D-glucosamine-6-phosphate as the substrate, initial velocity kinetics were measured at a pH/pD of 7.5 as shown in Figure 2.6. At concentrations of substrate up to

4 mM there is evidence for substrate inhibition, and the data were therefore fit to eq 2.2 to yield kinetic constants of k_{cat} (78 s^{-1}), K_m (0.14 mM), and k_{cat}/K_m (5.7 × 10^5 M^{-1} s^{-1}) for the results in H_2O. The corresponding values in D_2O under identical conditions with the same enzyme stock were the following: k_{cat} = 69 s^{-1}, K_m = 0.12 mM, and k_{cat}/K_m = 5.6 × 10^5 M^{-1} s^{-1}. The solvent isotope effect for the substitution of D_2O for H_2O on k_{cat}/K_m is 1.02, and the value on k_{cat} is 1.1. These relatively modest solvent isotope effects indicate that proton transfer is not a significant rate determining step in this transformation.

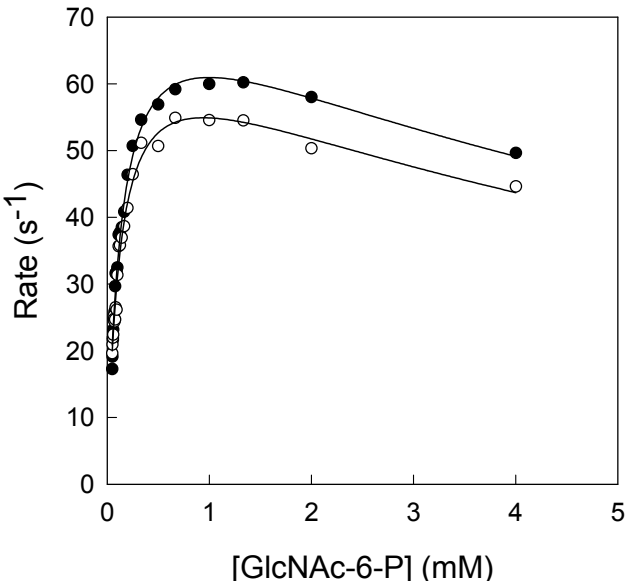


Figure 2.6: Solvent isotope effects for reactions performed in H_2O (●) and D_2O (○) at pH/pD 7.5. The solid lines represent fits of the data to equation 2. Additional details are available in the text.

The rate limitation imposed by the chemical cleavage of the amide bond was tested by employing the trifluoroacetate derivative of D-glucosamine-6-phosphate (**6**) as a substrate. Viola et al. have demonstrated that the trifluoroacetate analogue of *N*-acetyl-L-aspartate is a very good substrate for aspartoacylase from human brain (*114*). The

kinetic parameters for the hydrolysis of **6** with NagA are listed in Table 2.2. This substrate is hydrolyzed at a significantly faster rate than is the corresponding *N*-acetyl derivative. Since the pK_a of trifluoroacetic acid (0.23) is significantly lower than that of acetic acid (4.76), the carbon-nitrogen bond is weaker and the carbonyl group more electrophilic. Therefore, the rate of hydrolysis of **6** is expected to be faster than the hydrolysis of **1** if NagA is limited by the chemical step in the steady state. This assessment does not differentiate whether the rate limiting step is the formation or cleavage of the putative tetrahedral intermediate. It could be argued, however, that if NagA were limited by product release from the active site, then the dissociation of trifluoroacetate could be inherently faster than the release of acetate. This specific scenario was addressed by the utilization of changes in solvent viscosity to systematically alter the rate constants for the association and dissociation of products and substrates with NagA (*99, 100*).

The effects of solvent viscosity on the catalytic rate constants for the hydrolysis of **1** by Zn-NagA are shown in Figures 2.7 A and 2.7 B using sucrose as the micro-viscogen. The plot of relative k_{cat} vs relative viscosity was fit to a linear equation yielding a slope of 0.008 ± 0.02. The corresponding fit of data for the effect of solvent viscosity on k_{cat}/K_m gave a slope of 0.065 ± 0.049. These results indicate quite clearly that the rate of product dissociation does not limit k_{cat} for the hydrolysis of substrates by NagA. These experiments also indicate that the value of k_{cat}/K_m is not limited by the association rate constant for the formation of the NagA substrate complex. These results (together with the enhanced rate of reaction of compound **6** relative to compound **1**) are

only consistent with the conclusion that NagA is limited by either the formation or breakdown of the tetrahedral intermediate.

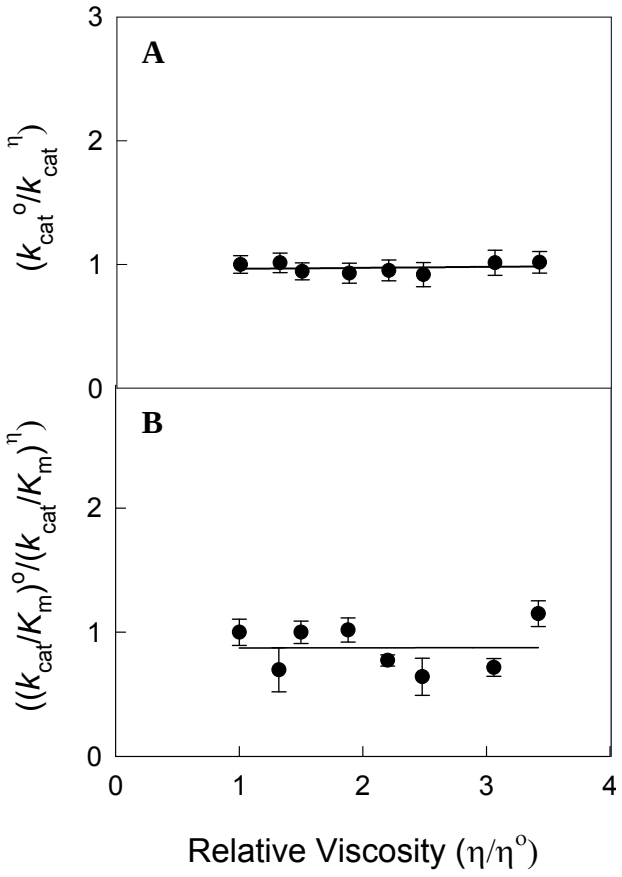


Figure 2.7: The effect of viscosity on the relative values of k_{cat} (A) and k_{cat}/K_m (B) using sucrose as the micro-viscogen. Additional details are available in the text.

Mechanism of Action. A model for the mechanism of amide bond hydrolysis by NagA from *E. coli* is presented in Figure 2.8. In this mechanism the resting state of the enzyme has a single divalent cation bound to the active site through ligation with the two histidines at the end of β-stands 5 and 6 (His-195 and His-216), and the glutamate (Glu-131) from the end of β-stand 3 (*21*). The remaining ligand is a bound water molecule

that is, in turn, hydrogen bonded to the invariant aspartate (D273) that originates from the end of β-strand 8. The reaction is initiated by the binding of the substrate in the active site and polarization of the carbonyl oxygen via ligation to the divalent cation. There is a proton transfer from the bound water molecule to Asp-273, and this is followed by nucleophilic attack on the carbonyl carbon and formation of the tetrahedral intermediate. The collapse of the tetrahedral intermediate is facilitated by proton transfer from Asp-273 to the leaving group amine of the product, D-glucosamine-6-phosphate. The products depart the active site, and another round of catalysis commences with the charging of the metal center with a molecule of water from the solvent.

The mechanism of substrate and water activation by NagA from *E. coli* must be different than the mechanism utilized by the same enzyme from *B. subtilis*. The *B. subtilis* enzyme contains a binuclear metal center whereas the *E. coli* enzyme can bind but a single divalent cation in the active site. For a binuclear metal center, the activation of the amide bond and the activation of the solvent water can be distributed separately between the two metal ions. However, in the mononuclear metal center, both functions must be borne by a single divalent cation. These results highlight the significant diversity for the evolution of function within the amidohydrolase superfamily. It is not intuitively obvious whether a binuclear or mononuclear metal center represents a more "advanced" or sophisticated site of catalytic power. However, it is tempting to speculate that the binuclear metal center currently found in NagA in some organisms is in the

process of shedding one of the divalent cations to create a fully functional active site that operates with a single divalent cation.

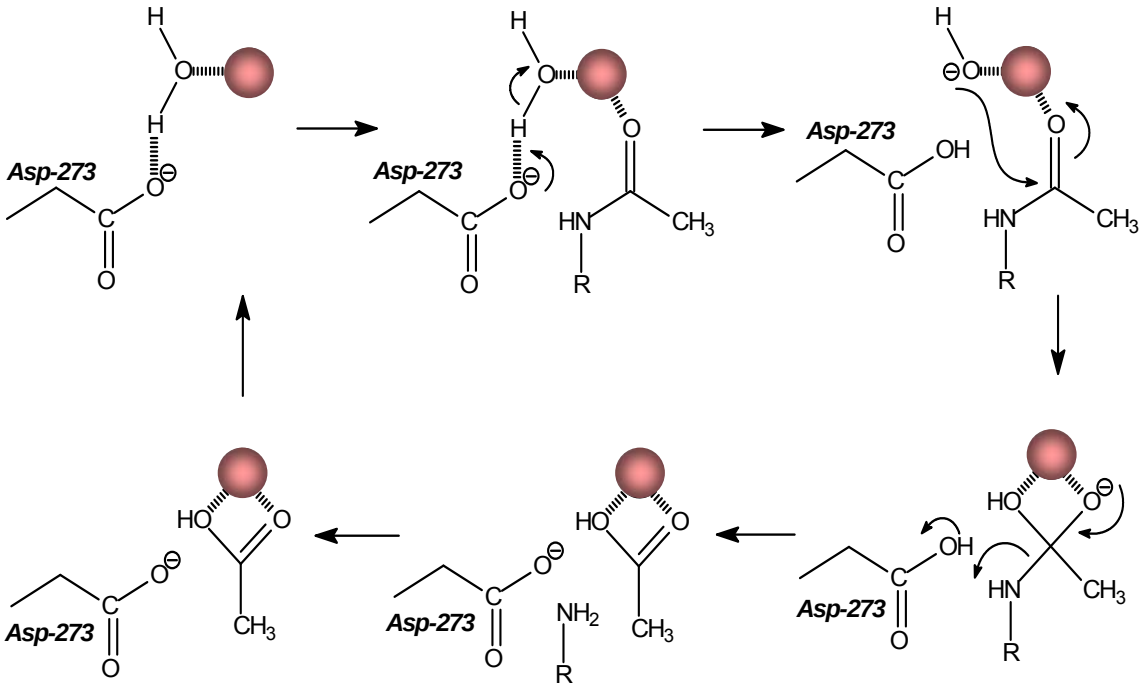


Figure 2.8: Proposed mechanism for the substrate hydrolysis by NagA from *E. coli*.

CHAPTER III

STRUCTURAL DIVERSITY WITHIN THE MONONUCLEAR AND BINUCLEAR ACTIVE SITES OF *N*-ACETYL-D-GLUCOSAMINE-6-PHOSPHATE DEACETYLASE*

NagA (E.C. 3.5.1.25) is a metal-dependent enzyme which catalyzes the deacetylation of *N*-acetyl-D-glucosamine-6-phosphate to form acetate and D-glucosamine-6-phosphate as presented in Scheme 1. NagA thus catalyzes a key step in the catabolism of *N*-acetyl-D-glucosamine from chitobiose and the recycling of cell wall murein (*90, 91, 97, 115, 116*). Over 300 sequences homologous to that of NagA from *Escherichia coli* K-12 have been identified in the current NCBI databases. Nearly all of these sequences are annotated as NagA, but some of them are annotated as *N*-acetyl-D-galactosamine-6-phosphate deacetylase (AgaA). In the genome of *E. coli* K-12 there is a deletion which eliminates the genes encoding for the *N*-acetyl-D-galactosamine (Aga) and D-galactosamine (GalN) phosphotransferase systems, while truncating the gene for AgaA. These deletions prevent the growth of *E. coli* K-12 on either *N*-acetyl-D-galactosamine or D-galactosamine (*117*).

NagA has been characterized as a member of the amidohydrolase superfamily (AHS) based on sequence and structural similarities to other enzymes within this structural fold (*12*). The enzymes in this superfamily have been shown to contain αβ-

---

*Reprinted with permission from "Structural Diversity within the Mononuclear and Binuclear Active Sites of *N*-Acetyl-D-glucosamine-6-phosphate Deacetylase" by Hall, R. S., Brown, S., Fedorov, A. A., Fedorov, E.V., Xu, C., Babbitt, P. C., Almo, S. C., and Raushel, F. M., 2007. *Biochemistry*, 46, 7953-7962, Copyright [2007] American Chemical Society.

superfamily (*10*). All proteins of the amidohydrolase superfamily possess a $(\beta/\alpha)_8$-barrel binuclear metal centers (*31, 38, 45, 50, 69, 73*), α-mononuclear metal centers (*17, 23, 79*), β-mononuclear metal centers (*10, 73*), or metal independent active sites (*63*).

X-ray crystal structures have been reported for NagA from *Bacillus subtilis* (PDB code: 1un7 (*69*)), *Thermotoga maritima* (PDB code: 1o12), and the metal-free complex of *E. coli* (PDB code: 1ymy and 1yrr (*65*)). A comparison of these three X-ray crystal structures of NagA highlights the fundamentally different active site motifs for the complexation of divalent metal ions within this enzyme. The enzyme from *B. subtilis* has an αβ-binuclear metal center with two iron atoms in the active site. The more solvent shielded metal (M_α) is coordinated to two histidine residues from β-strand 1 and an aspartate from β-strand 8 while the more solvent exposed metal (M_β) is coordinated to two histidine residues from β-strands 5 and 6. The two metal ions are bridged to one another by a glutamate from β-strand 3 and hydroxide or water from solvent (*69*). In contrast, the structure of NagA from *T. maritima* shows only a single iron bound to the M_β-site and a vacant M_α-site. However, the two histidines from β-strand 1 and the aspartate from β-strand 8 are apparently available for binding of a second metal ion. The two previously reported structures of NagA from *E. coli* were solved in the absence of any divalent cations, but it is highly probable that NagA from *E. coli* can bind only a single divalent cation at the M_β-site since the two histidines from β-strand 1, that are apparently essential for the binding of metal to this site, are absent and replaced by relatively poor metal ligands, asparagine and glutamine.

It is therefore uncertain whether the active sites in NagA from *E. coli* and/or *T. maritima* require one or two metal ions to be fully active. If NagA from either of these organisms has a different requirement for the total number of metal ions required for catalytic activity than NagA from *B. subtilis*, then the chemical mechanism for substrate hydrolysis must also differ in significant ways. The sites for the binding of metal ions in NagA are similar to the other binuclear metal centers found in the amidohydrolase superfamily (*31, 40, 50, 73*). However, in all of these other structures the histidine from β-strand 5 that coordinates M_β interacts with the metal via N^δ, whereas M_β in NagA coordinates to N^ε. The structures of the metal centers in the active sites of NagA from *B. subtilis*, *T. maritima*, and *E. coli* are presented in Figure 3.1.

The structure of the wild-type NagA from *E. coli* has now been solved, and we demonstrate that this enzyme binds but a single divalent cation exclusively to the M_β-site. In addition, the structure of the D273N mutant has been determined in the presence of a transition state inhibitor, and this structure unveils the molecular interactions between the substrate and the enzyme within the active site. An extensive bioinformatics analysis of the sequences annotated as NagA has shown that there is a correlation between the number of metal ions bound in the active site of this enzyme and the presence of a histidine or glutamine at the structural position that is equivalent to His-143 of the enzyme from *E. coli*. This residue is now postulated to facilitate the formation of the tetrahedral intermediate in the mechanism for amide bond cleavage in the active site of NagA.

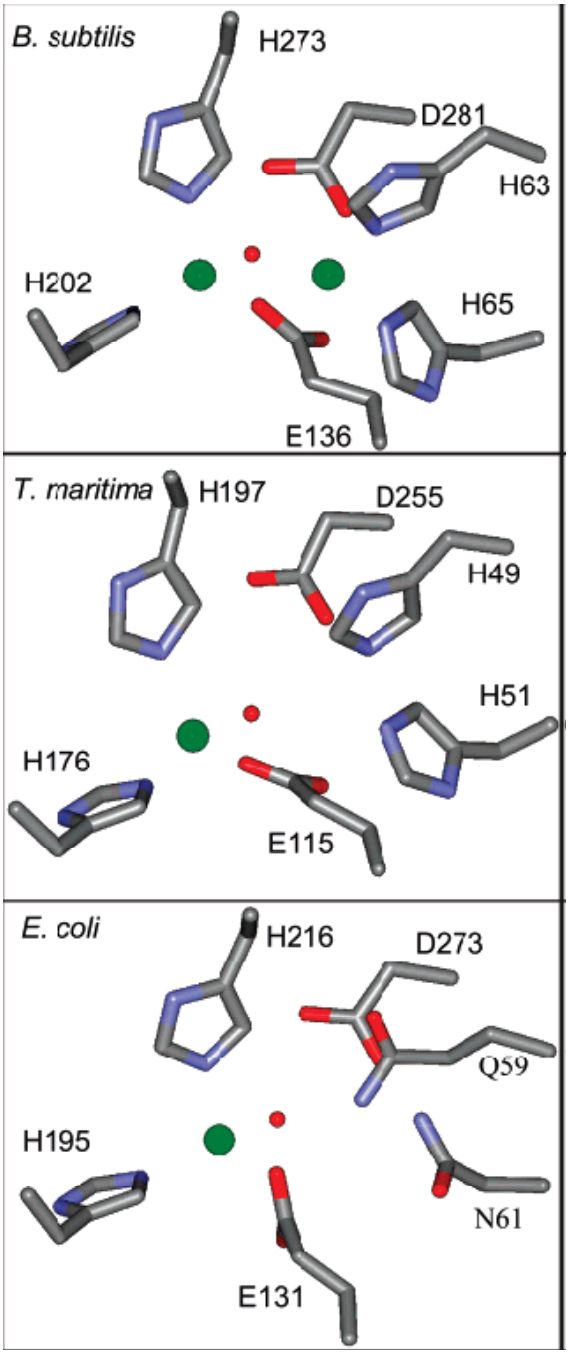
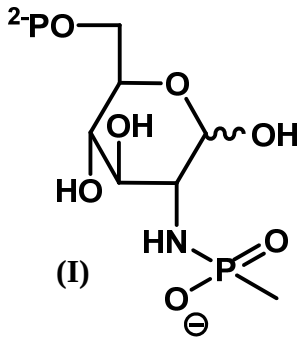


Figure 3.1: Top: The active site of NagA from *B. subtilis* with two Fe ions and a bridging water or hydroxide bound between the two metal atoms (PDB: 1un7). Middle: The active site of NagA from *T. maritima* with one Fe ion and a water molecule (PDB: 1o12). Bottom: The active site of NagA from *E. coli* with one Zn ion and a water molecule (PDB: 2p50). In these images the metal ions are represented as green spheres and the bound water or hydroxideis shown as a smaller red sphere.

MATERIALS AND METHODS

Materials. *N*-Acetyl-D-glucosamine-6-phosphate and purification reagents were purchased from Sigma-Aldrich. Chromatographic columns and resins were purchased from G. E. Healthcare. Chelex 100 resin was purchased from BioRad. Slide-A-Lyzer dialysis cassettes were purchased from Pierce. ICP standards were obtained from Inorganic Ventures Inc. The *N*-methyl phosphonate derivative of D-glucosamine-6-phosphate (**I**) was prepared according to the procedure of Xu (*98*).

Scheme 3.1



Cloning, Expression, and Purification of NagA. The wild type NagA and the D273N mutant from *E. coli* K-12 were purified as previously described (*67*). The enzyme utilized for the crystal structure of apo-NagA was expressed in LB growth medium with no additional metal. The NagA gene from *T. maritima* was amplified from genomic DNA using the nucleotide primers 5'-GCAGGAGCCATATGATTGTTGAG-AAAGTTCTGATCGTTGATCCCATCG-3' and 5'-CGCGGAATTCATCTGGATCG-AAAAACTACCTCACCC-3'. The product was gel purified and ligated through *Nde*I and *EcoR*I cloning sites into a pET-30a(+) vector incorporating kanamycin resistance and IPTG inducible overexpression. Plasmid DNA was electroporated into XL1 Blue

cells and subsequently purified with a Wizard Plus SV miniprep DNA purification system. Gene sequences of the purified plasmids were verified through PCR-based sequencing inconjunction with the Gene Technologies Laboratory of Texas A&M University.

The NagA from *T. maritima* was expressed in LB medium supplemented with 0.5 mM $ZnCl_2$. The IPTG-induced cells grew for approximately 18 h at 30 °C, after which they were centrifuged at 3400*g* for 12 min. The cells were resuspended in buffer A (50 mM Tris, pH 7.5 and 1.0 mM DTT) which was supplemented with 100 μg/mL phenylmethanesulfonyl fluoride and disrupted by sonication. Insoluble cell debris was removed by centrifugation at 13900*g* for 12 min, after which 1% w/v protamine sulfate was used to precipitate nucleic acids. The insoluble material was removed by centrifugation. The NagA from *T. maritima* was precipitated between 50 and 75% saturated ammonium sulfate. The precipitated protein was centrifuged at 13900*g* for 12 min. The protein was resuspended in a minimal volume of buffer A and loaded onto a Hiload 26/60 Superdex 200 prep grade gel filtration column. The NagA containing fractions were further purified with a Resource Q anion exchange column via elution with a linear gradient of 1.0 M NaCl in buffer A. Throughout the purification, the NagA containing fractions were identified and monitored for purity by activity analysis and with SDS PAGE. Protein concentrations of NagA from *E. coli* and *T. maritima* were determined using the calculated extinction coefficients of 18490 M^{-1} cm^{-1} and 7680 M^{-1} cm^{-1}, respectively (Protein Calculator v3.3 at http://www.scripps.edu /_cdputnam/ protcalc.html).

Measurement of Enzymatic Activity. The rate of hydrolysis of *N*-acetyl-D-glucosamine-6-phosphate was monitored by observing the cleavage of the amide bond at 215 nm using an extinction coefficient of 254 M^{-1} cm^{-1}. Kinetic assays were performed at 30 °C using a 96-well quartz plate with a SpectraMax 384-Plus spectrophotometer from Molecular Devices.

Reconstitution of NagA from T. maritima. The metal content (Cd, Co, Cu, Fe, Mn, Ni, and Zn) in the enzyme preparations was determined with ICP-MS. Metal-free NagA from *T. maritima* was prepared via dialysis of a 2.4 mg/mL enzyme solution against 10 mM dipicolinate, 30 mM MES buffer at pH 6.0 using a 10K molecular weight cutoff dialysis cassette with three buffer changes over 3 days. The dipicolinate was removed with four buffer changes against 30 mM metal-free HEPES at pH 8.0. For the metal reconstitution study, aliquots of apoenzyme were incubated with 0 to 3 equiv of $ZnCl_2$ per subunit at 4 °C for 24 h with 50 μM enzyme. The enzymatic activity of the reconstituted protein was determined after a 5000-fold dilution into a 2.0 mM *N*-acetyl-D-glucosamine-6-phosphate assay mixture containing 50 mM Tris buffer at pH 7.5 and 30 °C.

Sequence Analysis and X-Ray Crystallography. Extensive searches of the NCBI Database for NagA like sequences were collected and thoroughly analyzed by Dr. Shoshana Brown of the Dr. Patricia Babbitt Lab in the Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry at the University of California-San Francisco. Crystallization, X-ray diffraction, data collection, structure determination and refinement were performed by Drs. Alexander and Elena Fedorov of

the Dr. Steve Almo Lab in the Albert Einstein College of Medicine.

RESULTS

Sequence Analysis. The AgaA protein sequence from *E. coli* C was compared to simiar sequences in the NCBI database. Of the first 300 sequences, only six distinct sequences were annotated as AgaA. Of these six, all but one of the AgaA enzymes belong to different strains of *E. coli*. The single non-*E. coli* enzyme annotated as AgaA was located next to GlcN related enzymes. The majority of the remaining similar sequences correspond to proteins annotated as NagA. For the NagA annotated enzymes with similar sequence to AgaA, several of the organisms including *E. coli* HS and *E. coli* 101-1 contain two NagA annotated sequences with less than 30% identity. Many of these NagA annotated enzymes that are closely related to the sequence of AgaA, like that of *Vibrio vulnificus* and *Aeromonas hydrophilla,* are actually located in chromosomes next to AgaA machinery such as N-acetylgalactosamine phosphotransferase components, tagatose 1,6 bisphosphate aldolase tagatose-6-phosphate kinase and alpha galactosidase. Careful analysis of a sequence alignment of the NagA enzymes with the AgaA enzymes indicates that AgaA contains all of the important catalytic residues required for NagA from *T. maritima*. A smaller, representative alignment of the three NagA enzymes discussed here with AgaA from *E.coli C* is shown in Figure 3.2.

```
Ec AgaA   1 MSGRGRNMTHVLRARRLLTEEGWLDDHQLRIADGVIAAIEPIPAG------VTERDAELL
Bs NagA   1 -----MAESLLIKDIAIVTENEVIKNGYVGINDGKISTVSTERPKEPYSKEIQAPADSVL
Tm NagA   1 ---------MIVEKVLIVDPIDGEFTGDVEIEEGKIVKVEKR----------ECIPRGVL
Ec NagA   1 --------MYALTQGRIFTGHEFLDDHAVVIADGLIKSVCPVAELPP-EIEQRSLNGAIL

Ec AgaA  55 CPAYIDTHVHGGAGVDVMDDAPDV----LDKLAMHKAREGVGSWLPTTVTTPLNTIHAAL
Bs NagA  56 LPGMIDIHIHGGYGADTMDASFST----LDIMSSRLPEEGTTSFLATTITQEHGNISQAL
Tm NagA  42 MPGFVDPHIHGVVGADTMNCDFS-------EMEEFLYSQGVTTFLATTVSTSLEKMKEIL
Ec NagA  52 SPGFIDVQLNGCGGVQFNDTAEAVSVETLEIMQKANEKSGCTNYLPTLITTSDELMKQGV

Ec AgaA 111 KRIAQRCQRGGP---GAQVLGSYLEGPYFTPQNKGAHPPELFRELEIAELDQLIAVSQHT
Bs NagA 112 VNAREWKAAEESSLLGAELLGIHLEGPFVSPKRAGAQPKEWIRPSDVELFKKWQQEAGGL
Tm NagA  95 RKARDYILENPS----TSLLGVHLEGPYISKEKKGAHSEKHIRPPSERELSEIDSPA--
Ec NagA 112 RVMREYLAKHPN-----QALGLHLEGPWLNLVKKGTHNPNFVRKPD-AALVDFLCENADV

Ec AgaA 168 LRVVALAPEKEGALQAIRHLKQQNVKVMLGHRAATWQQTRAAFDACADRLVHCYNRMTGL
Bs NagA 172 IKIVTLAPEEDQHFELIRHLKDESIIASMGHTDADSALLSDAAKAGASHMTHLYNAMSPF
Tm NagA 148 -KMLTFAPEIE-SSELLLRLVKRDIVLSAGHSIATFEEFMKFYKEGVKRITHFPNGLKPL
Ec NagA 166 ITKVTLAPEMVP-AEVISKLANAGIVVSAGHSNATLKEAKAGFRAGITFATHLYNAMPYI

Ec AgaA 228 HHREPGIVCAGLTDKRAWLELIADGHHVHPAAMSLCCCCAKER-IVLITDAMQAAGMPDG
Bs NagA 232 HHREPGVIGTALAHDGFVTELIADGIHSHPLAAKLAFLAKGSSKLILITDSMRAKGLKDG
Tm NagA 206 HHREIGITGAGLLLDDVKLELICDGVHLSREMVKLVYKVKKANGIVLVTDSISAAGLKDG
Ec NagA 225 TGREPGLAGAILDEADIYCGIIADGLHVDYANIRNAKRLKGDK-LCLVTDATAPAGANIE

Ec AgaA 287 RYTLCGEEVQMHGGVVRTASGGLAGSTLSVDAAVRNMVELTGVTPAEAIHMASLHPARML
Bs NagA 292 VYEFGGQSVTVRGRTALLSDGTLAGSILKMNEGARHMREFTNCSWTDIANITSENAAKQL
Tm NagA 266 TTTLGDLVVKVKDGVPRLEDGTLAGSTLFFSQAVKNFRKFTGCSITELAKVSSYNSCVEL
Ec NagA 284 QFIFAGKTIYYRNGLCVDENGTLSGSSLTMIEGVRNLVEHCGIALDEVLRMATLYPARAI

Ec AgaA 347 GVDGVLGSLKPGKRASVVALDSGLHVQQIWIQGQLASF------
Bs NagA 352 GIFDRKGSVTVGKDADLVIVSSDCEVILTICRGNIAFISKEADQI
Tm NagA 326 GLDDR-GRIAEGTRADLVLLDEDLNVVMTIKEGEVVFRSR----
Ec NagA 344 GVEKRLGTLAAGKVANLTAFTPDFKITKTIVNGNEVVTQ-----
```

Figure 3.2: Protein sequence alignment of AgaA from *E. coli* C, NagA from *B. subtilis*, NagA from *T. maritima*, and NagA from *E. coli* K-12. The eight β-strands are highlighted in gray. The residue positions from the end of β-strand 1 that ligate the β-metal ion are highlighted in yellow. The residue position that corresponds to His-143 in NagA from *E. coli* K-12 is highlighted in green. Fully conserved metal binding and catalytic residues are highlighted in red. Fully conserved substrate binding residues are highlighted in blue.

Metal Dependence and Activity of NagA from T. maritima. The recombinant NagA from *T. maritima* was purified from an *E. coli* expression system and found to contain 0.25 equiv of Zn and 0.35 equiv of Fe per active site. Upon further incubation with 3 molar equiv of Zn, values of k_{cat} and K_m were determined to be 110 ± 3 s^{-1} and 0.04 ± 0.01 mM, respectively, with a value of k_{cat}/K_m of 2.8 ± 0.4 × 10^6 M^{-1} s^{-1}. At high substrate concentrations the activity diminished with an inhibition constant of 3.9 ± 0.5 mM. After removal of the bound metal, NagA from *T. maritima* lacked detectable levels of activity. The apoenzyme was reconstituted by adding up to 3 equiv of zinc per subunit. In these experiments there was a linear increase in the catalytic activity that reached a maximum at approximately 1 equiv of zinc per enzyme subunit. The change in activity as a function of added zinc is presented in Figure 3.3. These results demonstrate that NagA from *T. maritima* requires only a single divalent cation for maximum catalytic activity.

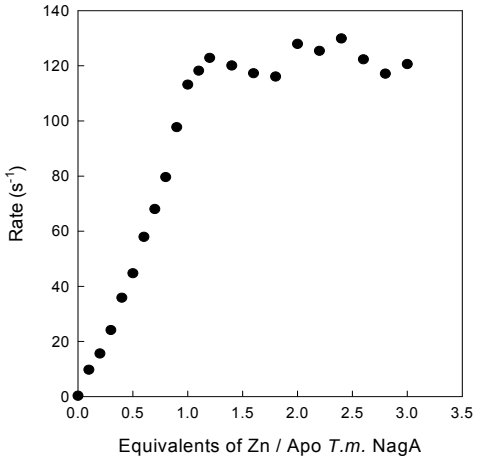


Figure 3.3: Reconstitution of apo-NagA from *T. maritima* (50 μM) with varying amounts of $ZnCl_2$. The rate of substrate hydrolysis was determined with 2.0 mM *N*-acetyl-D-glucosamine-6-phosphate. Additional details are provided in the text.

Structure of Apo-NagA. The crystal structure of the apo-NagA was determined to a resolution of 2.6 Å in the absence of any other ligands. The enzyme adopts a homotetrameric oligomerization state in solution as determined by gel filtration chromatography (*67*), SAXS (*65*) and analytical ultracentrifugation (*68*). The quaternary structure is a dimer of dimers, where each dimer pair has the active sites facing the complementary subunit as presented in Figure 3.4 A. Interactions for this active site interface occur mainly through a loop and α-helix region located after β-strand 6. Arg-227 is poised to interact with the phosphate moiety of the substrate molecule which would bind to the complementary subunit as shown previously for the NagA from *B. subtilis (69)*.

Other interactions between the two dimers occur between the β-sandwich domains formed by residues from the N- and C-termini. The tertiary fold consists of a somewhat twisted $(\beta/\alpha)_8$-barrel made of eight alternating β-strands and α-helices as shown in Figure 3.4 B. A small α-helix partially covers the N-terminus (bottom) of the barrel. The putative active site residues are located at the C-terminal end of the β- barrel. The loop region between β-strands 3 and 4, which contains residues 138-145, is conspicuously absent in the structure of the apo-NagA. This is likely due to the mobility of these residues in the absence of a bound ligand. A small β-sheet formed by residues between β-strand 8 and α-helix 9 forms a putative capping domain near the active site.

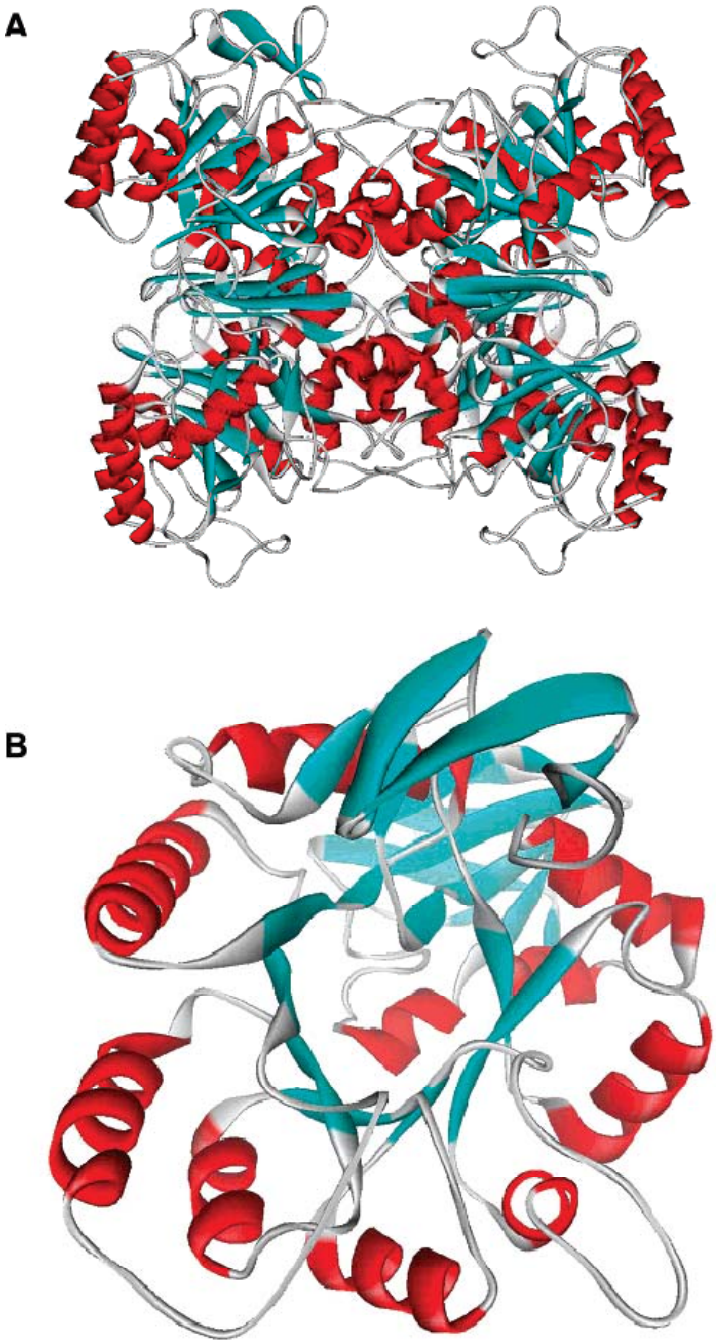


Figure 3.4: Ribbon diagram of the three-dimensional structure of NagA from *E. coli* displaying the tetrameric oligomerization (A). Ribbon diagram of an individual subunit of NagA (B).

Structure of Zn-NagA. The structure of the wild type NagA from *E. coli* was determined in the presence of zinc bound within the active site. Each subunit binds a single Zn atom in the M_β site, through electrostatic interactions with Glu-131, His-195, and His-216 as shown in Figure 3.5. The interatomic distances of Zn to these residues in the native structure are 2.1, 2.2, and 2.1 Å, respectively. Subunit A possesses two water molecules in the active site. One of these water molecules is bound directly to the Zn at a distance of 2.4 Å. This water molecule is also interacting with both oxygen atoms of Asp-273 at distances of 2.8 and 3.0 Å. A second active site water molecule is 2.9 Å from the first water molecule and is hydrogen bonded to *δ*-O of Asn-61and to one of the oxygen atoms from the carboxylate side chain of Glu-131 with distances of 2.5 and 2.7 Å, respectively. Subunit B also contains two water molecules in the active site with similar distances and geometries.

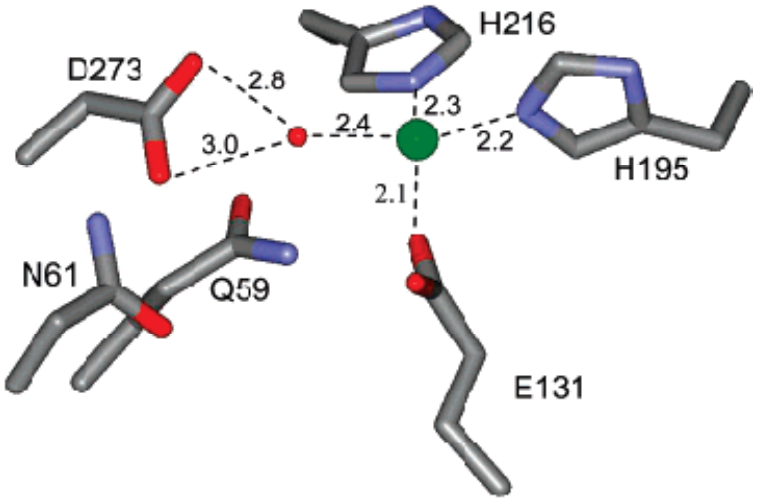


Figure 3.5: The structure of the active site of NagA from *E. coli* showing the coordination of the single Zn in the β-metal site. The zinc is represented as a green sphere, and the bound water molecule is shown as a small red sphere. The distances between the zinc and the coordinating ligand are listed in angstroms.

Structure of D273N Mutant. The zinc in the active site of the wild type NagA from *E. coli* is coordinated to a single water molecule, and this water molecule is hydrogen bonded to the side chain carboxylate of Asp-273 from β-strand 8. Mutation of this residue to asparagine results in a protein that is unable to hydrolyze the substrate (*67*). The crystal structure of the D273N mutant was determined in the presence of a transition state analogue inhibitor (**I**) (see Scheme 3.1) in an attempt to unveil the molecular interactions between the protein and substrate in a complex that cannot be hydrolyzed. The inhibitor, *N*-methylhydroxyphosphinyl-D-glucosamine-6-phosphate (**I**), is presented in Scheme 3.2 and is bound within the active site of D273N. The active site Zn in this structure is coordinated by four ligands in a distorted tetrahedral fashion. One of these interactions is from one of the side chain carboxylate oxygens of Glu-131 from β-strand 3 at a distance of 1.8 Å. There are additional interactions from His-195 from β-strand 5 and His-216 from β-strand 6 at distances of 2.1 and 2.2 Å, respectively. The fourth interaction is from one of the phosphonate oxygens of the inhibitor at a distance of 2.0 Å and is analogous to the water ligand described previously for the wild-type Zn structure. There are two other atoms near the active site Zn at slightly longer distances. These include the other oxygen from the phosphonate moiety of the inhibitor (analogous to the substrate carbonyl) and the second carboxylate oxygen of Glu-131. The oxygen from the inhibitor is positioned 3.0 Å away from the zinc and 3.0 Å from the side chain of His-143. The second carboxylate oxygen of Glu-131 is 2.9 Å away from Zn and 2.9 Å away from His-143. His-251 is hydrogen bonded to the anomeric hydroxyl of the inhibitor in the α-conformation at a distance of 2.7 Å. Two of the phosphate oxygens

from the inhibitor are positioned 2.8 and 3.0 Å from the guanidino group of Arg-227, which originates from the adjacent subunit. The X-ray crystal structure of the active site residues with bound Zn and inhibitor ligands is presented in Figure 3.6. A schematic drawing of the inhibitor-enzyme interactions is presented in Figure 3.7.

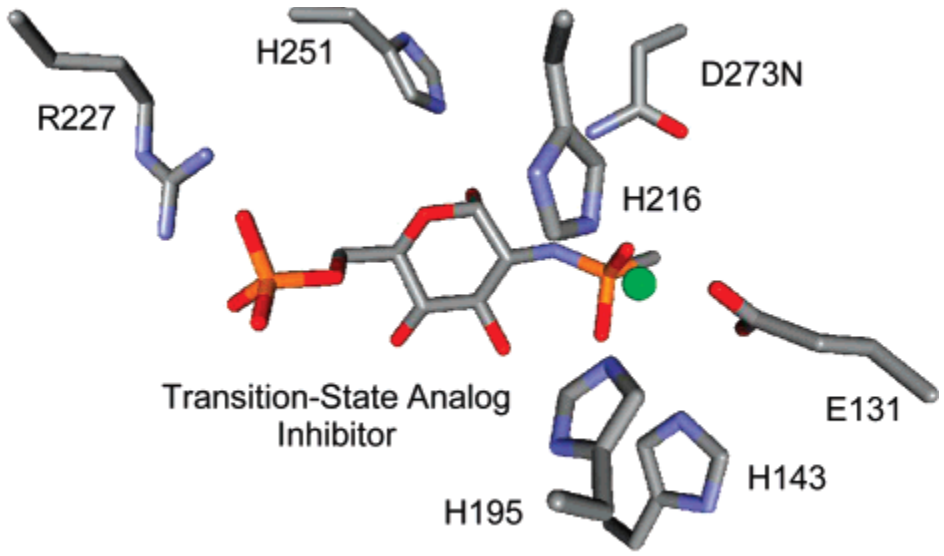


Figure 3.6: The structure of the active site of the NagA mutant D273N in the presence of the tight-binding transition state inhibitor **I.**

DISCUSSION

Differentiation of AgaA and NagA. Sequence alignment comparisons of NagA to sequences found on the NCBI database from *E. coli* K-12 indicate that there is a 31% identity between NagA and AgaA from *E. coli C,* a 31% identity between NagA from *E. coli* and from *B. subtilis*, while there is a 29% identity between NagA from *E. coli* and *T. maritima.* From sequence alignment comparisons, it is apparent that all of the important catalytic residues identified for NagA are conserved in the AgaA sequences, including

His143. AgaA possesses the HxH motif as well, making it most like the *T. marima* NagA with a 33% identity. It has been determined that NagA from *E. coli* can hydrolyze *N*-acetyl-D-galactosamine-6-phosphate nearly as well as the natural substrate *N*-acetyl-D-glucosamine-6-phosphate, making it appear that the only difference between these two enzymes is the respective operon content and transcription regulation.

T. maritima NagA Metal Requirement. The crystal structure of NagA contained only one Fe ion per active site. The crystal structures shown in Figure 3.1 and the sequence alignments shown in Figure 3.2 indicate that it possesses the same metal binding residues as NagA from *B. subtilis,* which contains two Fe ions per active site. If there was a loss of metal during purification, then it might be expected that the more solvent exposed β-metal binding site (M_β) would be vacant in the crystal structure (pdb code: 112). However, it was the deeper, more protected α-metal binding site (M_α) that was vacant in the crystal structure. We have determined that *T. maritima* NagA has similar metal content, requirement and localization compared to NagA from *E. coli*.

E. coli D273N NagA Crystal structure with Transition-State Substrate. This crystal structure of *E. coli* NagA with the bound transition state, enabled the identification of His143. This residue is found on a loop between β-strands 3 and 4 and is conserved among the predicted single metal utilizing NagA enzymes which are similar in sequence to the *E. coli* and *T. maritma* enzymes. H143 is interacting with the carbonyl oxygen of the substrate with a distance of 2.94 Å, thus His143 appears to participate with the Zn in the polarization of the amide bond. Mutagenesis of this residue, indicates that the β-Zn (3.02 Å away) is not entirely efficient at polarizing the

substrate without this histidine (*67*). The sharp angle of 82.9º formed by the carbonyl carbon, carbonyl oxygen and Zn, indicates that Zn may bear less of the electron withdrawing effect. The fact that the D273N mutant was crystallized with the transition-state intermediate bound, indicates that the Zn is capable of activating the water by itself to some extent. It is also apparent that the residue Asp273 is the only residue responsible for initiating the collapse of the tetrahedral intermediate by shuttling the proton from water to the leaving group amine.

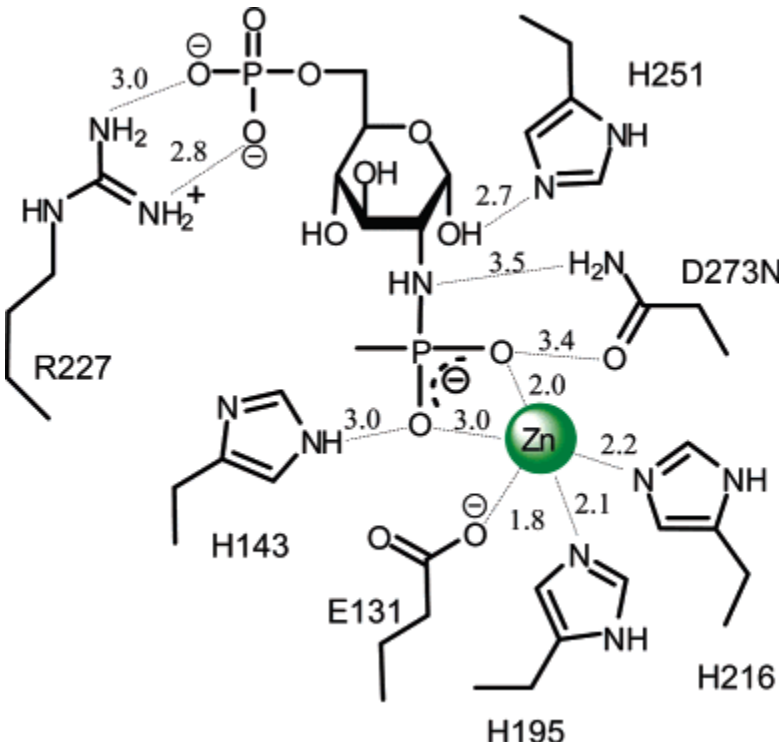


Figure 3.7: A schematic drawing of the active site of the D273N NagA mutant from *E. coli* with the bound transition state analogue inhibitor **I**. The distances between key molecular interactions are listed in angstroms.

Mechanistic Implications for One and Two Metal Systems. The substrate polarizing Hi-143 residue is conserved among all of the QxN possessing *E. coli*-like NagA enzymes as well as a small group of enzymes similar in sequence to *T. maritima* NagA. For the much larger class of *B. subtilis*-like enzymes, this residue is highly conserved as Gln. The identity of this residue may be the identifiable difference between mononuclear and binuclear mechanisms for the deacetylation of N-acetylglucosamine-6-phosphate. For without this substrate-polarizing His-143 or the α-metal, which typically has been shown to activate the hydrolytic water, the β-metal would need to bear an unprecedented catalytic role, coordinating both the water molecule and the carbonyl oxygen, for the withdrawal of electrons from both ligands which is necessary for the activation of the water molecule and polarization of the amide bond. Amazingly, these two groups of NagA enzymes have evolved to perform the deacetylation without the need of a second metal. This may prove to be a paradigm of stepwise evolution.

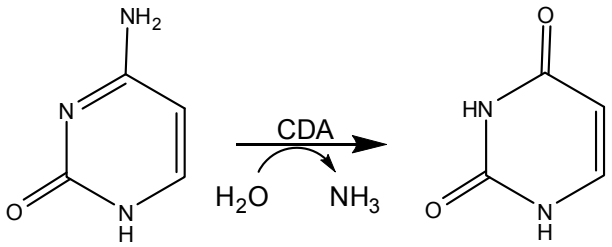
Summary. Based on these metal studies and protein sequence comparisons, we propose that the structures of NagA from *B. subtilis*, *T. maritima* and *E. coli* symbolize the evolution from the binuclear metal system seen in *B. subtilis* and similar enzymes to the mononuclear metal systems seen in *E. coli* and in higher organisms such as *H. sapiens*. We have also shown that AgaA is very similar to NagA and that the only apparent differences are the location of the genes within the chromosome. We have verified that the two NagA enzymes from *E. coli* and *T. maritima* each need only one metal for maximum activity, and that they both may only bind one metal. Therefore, the HxH α-metal binding motif used in the *B. subtilis* NagA, appears to be a vestige of evolution present in the *T. maritima* NagA. The main difference separating the single metal *T. maritima* NagA from the double metal *B. subtilis* NagA now appears to be presence of substrate-polarizing histidine residue or a Gln located between β-strands 3 and 4. It has still not been experimentally verified whether the binuclear NagA enzyme from *B. subtilis* actually needs both of the active site metals for catalysis. Further analysis of a binuclear NagA will surely provide insight into the evolution of metal dependence for this diverse set of enzymes.

CHAPTER IV

MECHANISTIC CHARACTERIZATION OF THE REACTION CATALYZED BY CYTOSINE DEAMINASE

Cytosine deaminase (CDA) catalyzes the hydrolytic deamination of cytosine, forming uracil and ammonia as shown in Scheme 4.1. Based on sequence alignments as well as the X-ray crystal structures, CDA has been characterized as a single metal dependent member of the amidohydrolase superfamily and possesses much of the typical catalytic and metal binding residues common for this superfamily. Included with these conserved residues are the HxH α-metal binding motif at the end of β-strand 1 (H62 and H64), two histidines from β-strands 5 (H215) and 6 (H247), as well as an aspartate from the end of β-strand 8 (D314). There is also a glutamate (E218) present three residues after the strand 5 histidine, whose interaction with N3 of the pyrimidine ring is of interest, as this atom presumably becomes protonated during the course of the deamination reaction. The sequence and structure of CDA from *E. coli* is highly similar to that of adenosine deaminase and guanine deaminase, while the crystal structure and sequence of *S. cerevisiae* cytosine deaminase is dissimilar and closely resembles the structure of *E. coli* cytidine deaminase.

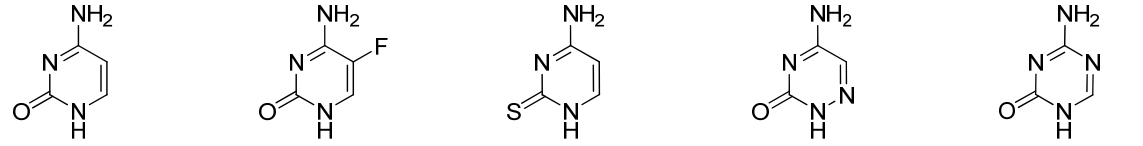
Scheme 2.1



Conspicuously absent in the human genome, CDA is a particularly promising target for drug development, because it can convert the 5-fluorocytosine to 5-fluorouracil, which inhibits replication (*93*). This effect is being exploited to selectively kill CDA containing single celled, non-eukaryotic organisms. It is also under investigation for therapeutic use in combination with gene therapy to kill cancer cells (*92*).

Previous studies of cytosine deaminase from *E. coli K-12* reported enzyme which primarily contained Fe along with some Zn. The enzyme preparations reported specific activities of 165 -185 s^{-1} at pH 7.5 and Michaelis constants of about 0.2 mM (*20, 22*). Cytosine deaminase has been reported to deaminate a number of substrates in addition to cytosine (1) including: 5-fluorocytosine (2) , 2-thiocytosine (3), 6-azacytosine (4) and 5-azacytosine (5). These are shown below in Scheme 4.2.

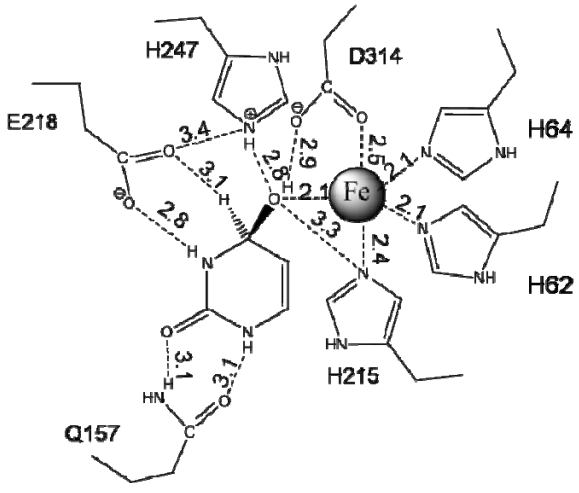
Scheme 4.2



cytosine (**1**) 5-fluorocytosine (**2**) 2-thiocytosine (**3**) 6-azacytosine (**4**) 5-azacytosine (**5**)

2-Hydroxypyrimidine is a potent mechanism-based inhibitor of CDA and has been utilized in the structural characterization of CDA through X-ray crystallography (*23*). In that X-ray crystal structure (PDB: 1K6W) a bound 2-hydroxypyrimidine can be seen with the active site Fe^{3+}, after having undergone nucleophilic attack by the active site hydroxide, resulting in the formation of 4-(R)-hydroxyl-3,4-dihydroxypyrimidine. This dihydroxy pyrimidine serves as an excellent probe of the mechanism of cytosine deamination. Fe is coordinated by His-62, His-64, His-215 and Asp-314. It can be seen that Gln-157 is participating in two hydrogen bonds with N-1 and the C-2 carbonyl oxygen of the pyrimidine. Glu-218 is interacting with N3 and is adjacent to the leaving group pocket of ammonia. His-247 is also nearby and is interacting with the 4-hydroxyl group which is a surrogate of the nucleophilic water. Asp-314 is interacting with this hydroxyl as well as the metal ion. A drawing of the observed residue-inhibitor interactions is presented below in Scheme 4.3.

Scheme 4.3



The metal cofactor utilization of *E. coli* CDA is of particular interest, because of the well documented preference for Fe^{2+} over Zn^{2+} by the enzyme (*19*). This preference is even more puzzling in light of the presence of a catalytic Zn^{2+} which is found in the active site of close structural homologs, adenosine deaminase (*16, 17*) and guanine deaminase (*24*). These enzymes are α-mononuclear metal dependent members of the amidohydrolase superfamily and catalyze similar reactions (*22*).

In spite of the large amount of structural and substrate specificity related information already reported for CDA, there are still a significant number of unresolved issues relating to the mechanism of deamination of cytosine by CDA. The nature of metal specificities remains a major unanswered question, particularly since Zn^{2+} is considered to be a superior Lewis Acid (*118*). In the existing crystal structures of wild type CDA (PDB codes: 1K6W and 1K70), CDA is crystallized with Fe^{3+} in the active site. CDA is only active with divalent transition metals, and because of the inactivity of Fe^{3+} containing enzyme (*22*), it is possible that some nearby residues may be in non-productive orientations. As a result, it is still uncertain which bases are contributing to the abstraction of the two protons from the nucleophilic water molecule and the protonation of the leaving group amine and the endocyclic N-3 atom. In this paper, these questions were addressed through the use of combinations of metal substitutions, site directed mutants, substrate analogs, pH activity profiles, solvent isotope effects, solvent viscosity effects and tight binding transition-state inhibitors.

MATERIALS AND METHODS

Materials. Reagents and buffers were obtained from Sigma-Aldrich unless otherwise noted. 3-oxauracil was purchased from Research Organics Inc. 3-methylcytosine, 4-thiouracil, 4-chlorouracil and the phosphonamidate analogs of cytosine and uracil were synthesized by Dr. Chengfu Xu. CDA knockout cells were obtained from the Keio Collection of the National Institute of Genetics, Japan. Kinetic assays were performed in a 96 well plate with a SPECTRAmax 384 Plus spectrophotometer from Molecular Devices. Protein concentrations were determined using the calculated extinction coefficient of 55190 M^{-1} cm^{-1} at 280 nm in a 1 cm path-length quartz cuvette. The extinction coefficient used was determined from the calculated protein sequence through the use of Protein Calculator v3.3 at http://www.scripps.edu/~cdputnam/protcalc.html.

Cloning, Expression, and Purification of CDA. The gene from *E. coli* encoding CDA was cloned into a pET-20a(+) expression vector. Mutants of CDA were prepared in accordance with the Quickchange mutagenesis kit instructions. Gene sequences were verified by the Gene Technologies Lab at Texas A&M University. The genes were transformed via electroporation into *E. coli K-12* CDA knockout cells which were made electro competent and then lysogenized with the λDE3 STAR lysogenization kit from Novagen.

After growing for 30 minutes at 37 °C, the cells were plated and grown overnight on LB/agarose plates enriched with 100 μg/mL ampicillin. A single colony was then selected for inoculation of 1L of LB enriched with 100 μg/mL ampicillin and 1.0 mM

$ZnCl_2$. Expression of the protein products was detected by SDS-PAGE only in the absence of any IPTG. After incubation for 18 hours with shaking at 20 °C, the cells were harvested by centrifugation at 3400g for 12 min.

The cells were then re-suspended and disrupted by sonication in 10 × (v/w) 50 mM HEPES buffer, pH 7.5, containing 1.0 mM $ZnCl_2$ and 100µg/mL of the protease inhibitor phenylmethanesulfonyl fluoride. The cell lysis solution was centrifuged at 13900g for 15 min to remove insoluble cell debris. A solution of protamine sulfate was then added drop-wise to the solution, with stirring, to a final concentration of 2% w/v, for the precipitation of nucleic acids. The solution was again centrifuged at 13900g for 15 min to remove the precipitated nucleic acids. The remaining protein was then precipitated with solid ammonium sulfate to a final concentration of 50% of saturation. The precipitated protein was then removed from solution by centrifugation at 13900g for 15 min. The pellet was then re-suspended with a minimal amount of 50 mM HEPES, pH 8.0, and then filtered with a 0.45 µm syringe filter and further purified by gel filtration with a G.E. HiLoad 26/60 Superdex 200 preparatory grade column. The CDA containing fractions which were identified by SDS- PAGE and cytosine deaminase activity, pooled and further purified using a G.E. ResourceQ anion exchange column.

Measurement of Enzymatic Activity. The rate at which cytosine deaminase converts cytosine to uracil was determined at 30 ºC using either a coupled assay or a direct assay. For the activity screening of substrate analogues with wild type and mutant enzymes, a coupled assay was used to detect the release of ammonia through the oxidation of NADH. This assay was modified from the protocol reported previously

(*79*) and typically contained 7.4 mM α-ketoglutarate, 0.4 mM NADH, 6 units of glutamate dehydrogenase in 100 mM HEPES, pH 8.5. The initial velocities were then fitted to eqn 4.1. For other activity assays at pH 7.5 a direct assay at 286 nm ($\Delta\varepsilon_{286}$ = -680 M^{-1} cm^{-1}) was used as previously reported (*22*). Mass spectrometry analysis for product verification of 3-oxauracil deamination was performed by the staff at the Laboratory for Biological Mass Spectrometry (LBMS) in the Department of Chemistry at Texas A&M University.

For the development and implementation of a direct assay for use at any given pH, unique extinction coefficients for the deamination of cytosine to uracil are assigned for each unique pH value. Changes in molar extinction coefficients for the total deamination of cytosine by CDA were determined at various pH values from 2.5-11.3. This was performed by measuring the differences in absorbance of 200 μM cytosine and of enzymatically deaminated cytosine in various100 mM buffers of 0.2 pH increments. Prior to absorbance measurements, stock solutions of substrate and enzymatic product were passed through 10,000 MWCO Centricon membranes and diluted to final concentrations of 0.2 mM in the buffers. The stock solution of cytosine consisted of 1.0 mM cytosine in 10 mM TRIS, pH 7.5. The stock solution uracil consisted of 1.0 mM cytosine reacted completely with 10 nM CDA in 10 mM TRIS buffer at pH 7.5. The buffers utilized are as follows: formate, pH 2.5-3.9; acetate, pH 4.1-5.5; MES, pH 5.7-7.1; PIPES, pH 6.7, 6.9; HEPES, pH 7.1-8.1; TAPS, pH 8.3-8.7; CHES, pH 8.9-10.1; CABS, pH 10.1-11.3. The final pH values were established through measurement with a pH meter. The differences in absorbance were plotted vs pH and fitted with eq 4.2.

The resulting pH specific constants were then used to calculate extinction coefficients for the reactions at any pH value. For pH values ranging from 5 to 7.0, constants obtained from the low-pH data obtained at 286 nm were used. At pH values from 7.0 to 8.2, constants obtained from the high pH data 286 nm were used. At pH values of 8.2 and above, the constants obtained at 298 nm were used.

Metal Analysis. Metal determination and quantification was performed with the Elan DRC II ICP-MS (inductively coupled plasma - quadrupole mass spectrometer) from Perkin Elmer. The analog detection mode is used with three averaged replicates per reading. External calibration standards were prepared through the serial dilution of a single 10 ppm stock mixture of Zn, Cd, Co, Cu, Mn, Ni and Fe in 2% nitric acid, purchased from Inorganic Ventures Inc. Freshly prepared standards generally contained 2, 20 and 200 ppb of all aforementioned analytes in 1% Trace Select Nitric Acid from Fluka that is diluted in deionized water. The isotopes of the analytes measured were ^{55}Mn, ^{56}Fe, ^{59}Co, ^{60}Ni, ^{63}Cu, ^{66}Zn and ^{111}Cd. ^{115}In was used as an internal standard for Cd and ^{69}Ga was used as an internal standard for all other analytes. The detection of ^{56}Fe was determined in the dynamic reaction cell (DRC) mode, where NH_3 gas is used to reduce interfering argon-oxide species. Protein samples were prepared for analysis by boiling in concentrated nitric acid for 30 min, followed by dilution with water to 1% nitric acid.

Metal Chelation and Reconstitution. Apo-cytosine deaminase was prepared by dialysis of 3 mL of 1 mg/mL cytosine deaminase with 3 × 300 fold buffer changes over 24 hours with 10 mM ortho-phenanthroline in 50 mM MES, pH 5.75. Ortho-

phenanthroline was then separated from the protein with a PD-10 column. Prior to use, the PD-10 column was washed with 1 column volume of a 10 mM solution of dipicolinate, and equilibrated with Chelex treated HEPES buffer, pH 8.0. The resulting apo-protein was reconstituted at 1 mg/mL overnight at 4 ºC with varying equivalents of $ZnCl_2$ and $FeSO_4$. The Fe reconstitutions were performed anaerobically under argon.

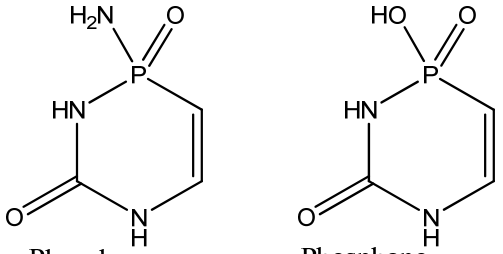
pH Dependence on Enzyme Activity. The pH dependence of k_{cat} and k_{cat}/K_m were determined for Zn^{2+} and Fe^{2+} containing CDA as well as the Zn containing mutant H247Q from pH 5.3-9.9 at ~0.2 pH unit intervals, using 75 mM MES (pH 5.3-6.5), PIPES (pH 6.7-6.9), HEPES (7.1-8.1), TAPS (8.3-8.9) and CHES (pH 9.1-9.9). The pH values of the final solutions were measured after the completion of the assays. Direct assays were utilized in which the extinction coefficient at any given pH was calculated with eq 1 using the constants obtained for that specific pH range.

Solvent Kinetic Isotope Effects and Solvent Viscosity Effects. Steady state rates were determined with the coupled assay for ammonia formation using D_2O or H_2O as the solvent and were obtained at the same time in either 100 mM pH 8.5 HEPES or pH 9.0 CHES at 30 ˚C. The proton inventories were performed in 100 mM HEPES, pH(D) 8.5 at 30 ˚C using: 0, 0.25, 0.5, 0.75 and 1.0 atom fraction deuterium as the solvent. The effects of viscosity on k_{cat} and k_{cat}/K_m of Zn^{2+} containing CDA were determined as previously described (*67*) in 100 mM MES pH 6.0, HEPES pH 7.5 and CHES pH 9.0 at 30 ˚C.

Transition-State Inhibitor Evaluations. Inhibition of Zn-CDA was evaluated using the phosphorus based analogs of cytosine and uracil shown below in Scheme 4.4.

The inhibitors were prepared according to the method of Paul Bartlett et al (*119*) by Dr. Chengfu Xu in the Frank Raushel Lab of the Chemistry Department at Texas A&M University. The inhibition constants (K_i) for Zn-CDA with the phosphonocytosine and the phosphonouracil transition state inhibitors were determined by incubating variable amounts of either inhibitor with 2.5 nM enzyme for 30-120 minutes in 100 mM HEPES buffer, pH 7.5 in a volume of 245 µL . Residual activities were then measured by initiating the reaction with 5 µL of 10 mM cytosine. Maximum inhibition was reached after 30 min. The time dependence to the onset of inhibition (k_{obs}) by the phosphonocytosine and phosphonouracil inhibitors with Zn-CDA was determined by measuring the rate of deamination as a function of time. In these experiments, 4 nM Zn-CDA was added to solutions containing either 1-15 µM phosphonocytosine or 0.96-4.33 mM phosphonouracil and 0.65 mM cytosine in 100 mM HEPES buffer at pH 7.5. The time dependence for reversal of inhibition was determined by incubating 1 µM Zn-CDA with 5-25 µM phosphonocytosine or 5-25 mM phosphonouracil in 100 mM HEPES buffer for 30 min, followed by a 100-fold dilution into 2 mM cytosine and observing the time course.

Scheme 4.4



Data Analysis. All data were analyzed through fits to the corresponding equations using the non-linear least squares fitting program SigmaPlot 10.0. For the determination of the kinetic parameters; k_{cat}, K_m and k_{cat}/K_m, saturation curves were fit using eq 2.1 where A is the substrate concentration, v is the initial velocity of the reaction, k_{cat} is the turnover number and K_m is the Michaelis constant. For the determination of the relationship between extinction coefficients and pH, plots of Δε vs pH were fit with eq 4.2, where Δε is the difference in absorbance between cytosine and uracil at a specific pH, *min* is the theoretical minimum for Δε at low pH, *max* is the theoretical maximum for Δε at high pH and Hill slope is the slope of the transition.

For the analysis of pH-rate profiles, plots of log k_{cat} and log k_{cat}/K_m vs pH that indicated the deprotonation of a single proton required for maximum activity were fit with eq 4.3, where c is the maximum activity and pK_a is the pH of inflection resulting in maximum activity. pH-rate profiles showing a loss of activity at low pH and a wave shaped loss of activity at high pH, which transitions to a steady lower level of activity, were fit with eq 4.4. In eq 4.4, pK_b is the pH where a transition to a lower level of activity occurs. pH rate profiles which showed only the wave-shaped plateau were fit ising eq 4.5.

The proton inventory which displayed inverse-concave solvent deuterium isotope effects was fit with eq 4.6, where v_n is the activity at different D_2O/H_2O ratios in solution, v_o is the activity in the absence of D_2O, n is the ratio of D_2O/H_2O, and Φ^R is the fractionation factor for the reactant-state site. Equation 4.7 was used to fit a proton inventory plot which displayed a linear inverse proton inventory.

For the evaluation of the inhibition constants (K_i) for the inhibitors of CDA, eq 2.7 was fit to plots of the residual activities vs inhibitor concentrations. E_t is the total enzyme concentration, I is the inhibitor concentration, v_o is the activity of the enzyme in the absence of inhibitor, and v_i is the residual activity of the enzyme in the presence of inhibitor. For the determination of the time dependence to the onset of inhibition, time courses corresponding to the change in rate of cytosine deamination were fit with eq 2.8, where P is the product concentration, v_r and v_s are the initial and final rates of product formation, and k_{obs} corresponds to the first order rate constants for the onset of the slower rate of product formation. The apparent second-order rate constants (appk_{obs}) for the formation of the inhibited form of Zn-CDA, was determined from the slope of the plots of k_{obs} vs the concentrations of inhibitors. Correction for the concentration of substrate according with eq 2.9, gives the second-order rate constant k_1 .

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

$$\Delta\varepsilon = min + (max - min)/(1 + 10^{(pH - pKa)} \times Hillslope) \tag{4.2}$$

$$y = \log(c/(1 + 10^{(pK\text{a-pH})})) \tag{4.3}$$

$$y = \log((c + d \times 10^{(\text{pH-p}K\text{b})})/(1 + 10^{(pK\text{a-pH})} + 10^{(\text{pH-p}K\text{b})})) \tag{4.4}$$

$$y = \log((c + d \times 10^{(\text{pH-p}K\text{b})})/(1 + 10^{(\text{pH-p}K\text{b})})) \tag{4.5}$$

$$v_n = v_0/(1 - n + n\Phi^R) \tag{4.6}$$

$$v_n = v_0(1 - n + n\Phi^T) \tag{4.7}$$

$$v_i/v_o = ([E_t] - K_i - [I] + (([I] + K_i - [E_t])^2 + (4K_i[E_t]))^{1/2})/(2[E_t]) \tag{2.7}$$

$$P = v_s t + ((v_r - v_s)/k_{obs})(1 - e^{(-k_{obs}t)}) \tag{2.8}$$

$$k_1 = \text{app}k_{obs}\,(1 + [S]/[K_m]) \tag{2.9}$$

RESULTS

Metal Dependence of CDA. Initially, wild type CDA from E. coli K-12 was expressed in BL-21 λDE3 cells and purified to homogeneity, yielding about 100 mg of pure protein from 2 liters of LB. The purified enzyme possessed a maximum activity of 64 s^{-1}. ICP-MS analysis indicated that the enzyme contained 0.56 equivalents of Zn and 0.20 equivalents of Fe. CDA was then expressed in LB enriched with 1.0 mM $ZnCl_2$ and cells were lysed via sonication in 50 mM TRIS buffer at pH 7.5 enriched with an extra 1 mM $ZnCl_2$. The protein was then purified as described earlier in the materials and methods. This enzyme preparation had a turnover rate of 132 s^{-1} with 0.79 equivalents of Zn, and 0.09 equivalents of Fe.

Ortho-phenanthroline treated apo-enzyme yielded a specific activity of 0.5 s^{-1} and ICP-MS analysis verified that there was less than 5% contamination with Fe, Zn, Mn, Ni, Cu, Cd or Co in the final apo-enzyme containing solution. 50 μM apo-CDA was then reconstituted with Zn^{2+} or Fe^{2+}. A metal titration with various equivalents of $ZnCl_2$ shows that CDA reaches maximum activity with approximately 1 equivalent of the metal as seen in Figure 4.1. The reconstitution with Fe^{2+} was performed in a septum sealed vial under N_2 with 50 μM apo-CDA and three equivalents of $FeSO_4$ in 20mM Hepes at pH 7.5.

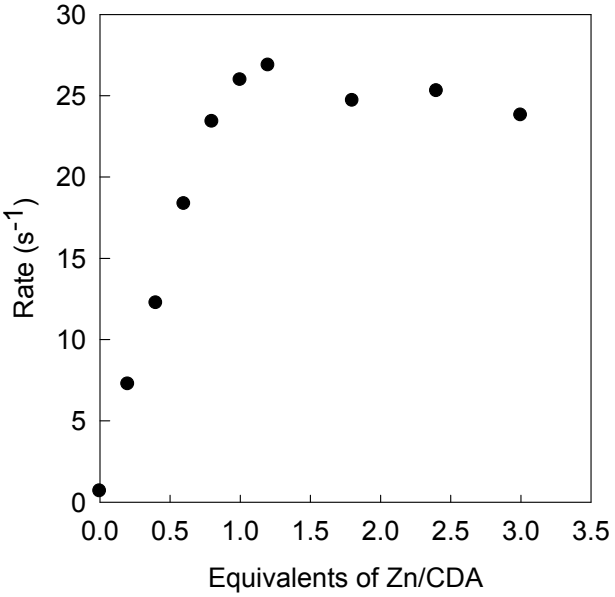


Figure 4.1: Reconstitution of apo-CDA (8.5 μM) with varying equivalents of Zn. The rates of cytosine deamination were determined after dilution of the enzyme incubations to concentrations of 50 nM in 2.0 mM cytosine and 50 mM TRIS at pH 7.5 and 30 ˚C. Additional details are provided in the text.

Mutation of Active Site Residues. Mutants of CDA were prepared in an effort to determine the roles of specific residues in substrate recognition and catalytic function. Gln-157, Glu-218, His-247 and Asp-314 were selected for mutation based on the protein-ligand interactions observed in the crystal structure of CDA with the hydroxypyrimidine mechanism based inhibitor as shown earlier. All mutations performed resulted in 0.1 % activity or less, in spite of relatively small perturbations in metal content. Only two mutants showed significant levels of activity with cytosine (> 0.005 s^{-1}). These active mutants are: H247Q, which displayed a k_{cat} of 0.18 s^{-1}, a K_m of 0.25 mM and a k_{cat}/K_m of 720 M^{-1} s^{-1}; and D314N, which has a k_{cat} of 7.5×10^{-3} s^{-1}, a K_m of 1.9 mM and a k_{cat}/K_m of 4.0 M^{-1} s^{-1}.

Table 4.1: Metal Content and Kinetic Parameters for CDA and Active-Site Mutants.

enzyme	metal content	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
WT	0.79 Zn, 0.09 Fe	132 ± 2	0.97 ± 0.03	$(1.4 \pm 0.1) \times 10^5$
Q157N	1.40 Zn, 0.04 Fe	< 0.005	-	-
Q157A	0.92 Zn, 0.04 Fe	< 0.005	-	-
E218Q	0.87 Zn, 0.15 Fe	< 0.005	-	-
E218A	0.90 Zn, 0.08 Fe	< 0.005	-	-
H247Q	0.31 Zn, 0.33 Fe	0.18 ± 0.01	0.25 ± 0.03	720 ± 95
H247N	0.53 Zn, 0.29 Fe	<0.005	-	-
H247A	0.59 Zn, 0.33 Fe	< 0.005	-	-
D314N	1.05 Zn, 0.05 Fe	$(7.5 \pm 0.2) \times 10^{-3}$	1.9 ± 0.2	4.0 ± 0.4
D314A	0.57 Zn, 0.32 Fe	< 0.005	-	-

Substrate Specificities of Zn CDA and Mutants. The library of cytosine analogues (consisting of various pyrimidines, pyridines, triazines and imidazoles) shown in Scheme 4.5 was tested for activity at concentrations of 1.0 mM, with 1 μM Zn-CDA and mutants in 75 mM Hepes, pH 7.5 and 30 ˚C. The release of ammonia from these compounds by CDA was quantified through NADH oxidation with the glutamate dehydrogenase coupled assay as described above. In that screen, it was discovered that cytosine deaminase could catalyze the deamination of creatinine (5.6 s^{-1}), isocytosine (5.1 s^{-1}) and 3-oxauracil (2.3 s^{-1}). The E218A mutant was also found to be able to catalyze the release of ammonia from 3-oxauracil (0.18 s^{-1}). The other compounds display insufficient activity (< 0.1 s^{-1}) with the wild-type and mutant enzymes to warrant further investigation. A summary of the kinetic constants determined for the active compounds are shown below in Table 4.2.

Scheme 4.3: Library of pyrimidines, pyridines, triazines and imidazoles.

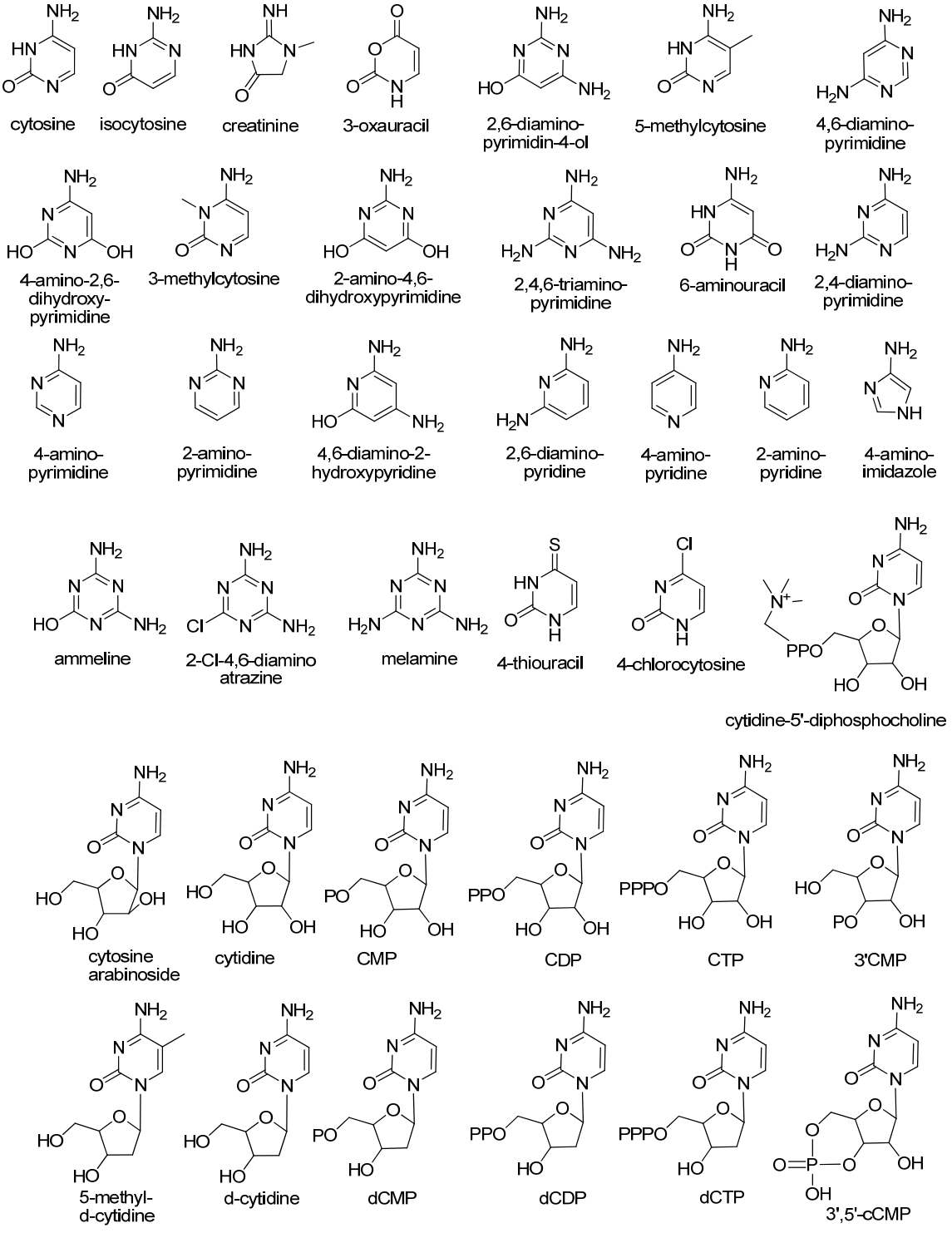
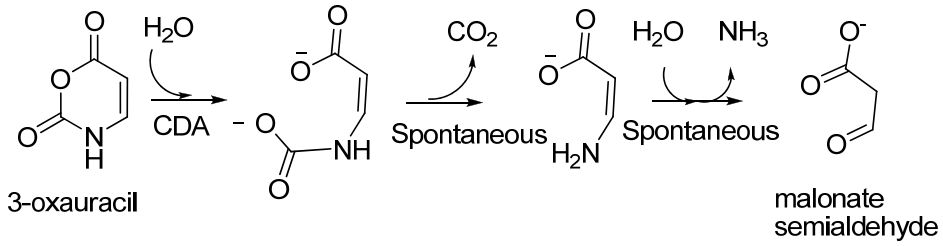


Table 4.2: Kinetic Parameters for Substrates with Zn-CDA.

Enzyme	Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
WT	cytosine	132 ± 2	0.97 ± 0.03	$(1.4 \pm 0.1) \times 10^5$
WT	isocytosine	5.1 ± 0.1	0.46 ± 0.04	$(1.1 \pm 0.1) \times 10^4$
WT	creatinine	5.6 ± 0.7	25 ± 4	$(2.2 \pm 0.5) \times 10^2$
WT	3-oxauracil	2.3 ± 0.1	4.1 ± 0.3	$(5.6 \pm 0.5) \times 10^2$
E218A	3-oxauracil	0.182 ± 0.004	0.85 ± 0.04	$(2.1 \pm 0.1) \times 10^2$

From the wild-type and E218A CDA catalyzed hydrolysis of 3-oxauracil, it was found that 1 equivalent of ammonia was rapidly released from 3-oxauracil. This deamination occurred at rates that were linearly dependent on enzyme concentration. Malonate semialdehyde was found to be the major product of this hydrolysis as verified by a prominent peak present at an m/z of 87 by negative mode ESI-MS. The proposed reaction for the hydrolysis of 3-oxauracil to malonate semialdehyde is shown below in Scheme 4.5.

Scheme 4.4



It was also determined that CDA is unable to catalyze the dethiolation of 4-thiouracil at any measurable rate ($< 6 \times 10^{-5}$ s^{-1}). This was confirmed through a lack of

changes in absorbance of the substrate which would be significant as seen from a comparison of the spectra of 4-thiouracil and the expected product uracil (data not shown). This conclusion was also verified through the use of dithionitrobenzoate (DTNB) which is used to indicate the presence of sulfide. It was also determined that CDA fails to de-chlorinate 4-chlorocytosine at a measurable rate, ($< 2.2 \times 10^{-4}$ s^{-1}). This was determined by comparing the absorbance spectra of 100 μM 4-chlorocytosine and uracil at pH 7.5 in 50 mM TRIS buffer (data not shown).

pH Studies. Profiles of pH vs. activity were determined with the Fe^{2+} and the Zn^{2+} containing enzymes using both coupled and direct assays. The resulting pH vs activity profiles were consistent, except the activity for the coupled assays dropped off with a slope of ~ -4 at pH values above 9 (data not shown). This loss of activity is due to the failure of the coupling system at higher pH values. For the development of a direct assay, the pH dependent changes in the ultraviolet absorbance spectra of both cytosine and uracil must be taken into account.

The absorbance maximum of cytosine changes from 275 nm at pH 2.5 to 267 nm at pH 11.3, while the greatest change in absorbance occurs near the pK_a of cytosine at pH 4.5. The absorbance maximum of uracil changes from 258 nm at pH 2.5 to 284 nm at pH 11.3. The greatest changes in the absorbance of uracil occur near its pK_a at pH 9.5. These ionizations are similar to earlier reported values of 4.57 and 9.26 at 30 ºC (*120*). The spectra of cytosine in buffers of pH 2.5 - 11.3 are shown in Figure 4.2 A and the spectra of the product of cytosine and CDA (uracil) in buffers of pH 2.5 - 11.3 are shown

in Figure 4.2 B. The reported ionization resonance structures for cytosine and uracil are shown below in Scheme 4.6 (*120*).

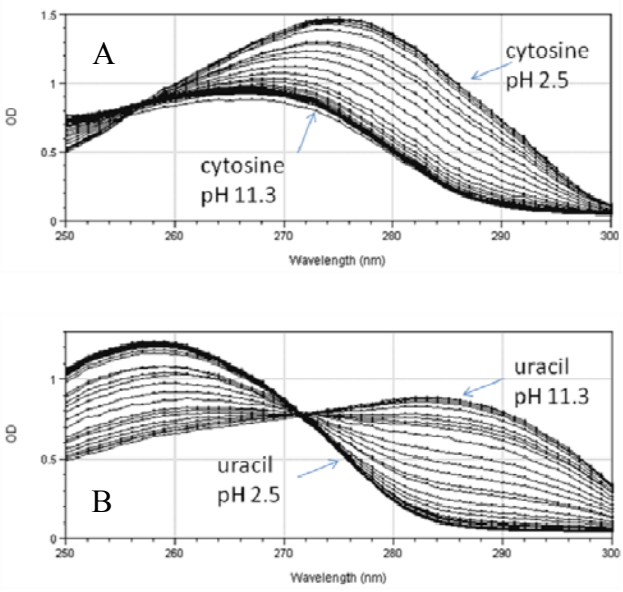
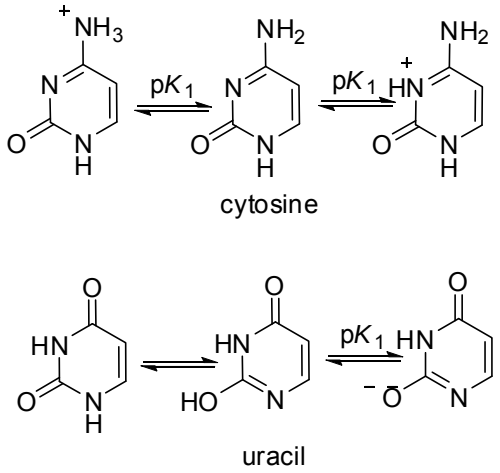


Figure 4.2: pH dependence of (A) absorbance spectra of 200 µM cytosine in various 100 mM buffers ranging from pH 2.5-11.3 at 30 ºC; and (B), absorbance spectra of 200 µM enzymatically prepared uracil in various 100 mM buffers ranging from pH 2.5-11.3 at 30 ºC.

Scheme 4.5



The differences in extinction coefficients for cytosine and uracil were determined at different pH values between 5.0 and 10.5. The resulting plots of Δε (ε cytosine - ε uracil) vs pH yielded sigmoidal plots. The pK_a values of cytosine and uracil and the pH dependent Δε values for the deamination of cytosine were calculated from eq 4.2. (see Figures 4.3, 4.4 and 4.5).

Activities measured at pH values between 5.0 -7.5 were followed at 286 nm using the constants determined from the data shown in Figure 4.3. The constants determined from the data in Figure 4.4 were used for activities measured at 286 nm within the pH range of 7.3-8.5. For pH values 8.3-11, the constants obtained from figure 4.5 were utilized for rates measured at 298 nm.

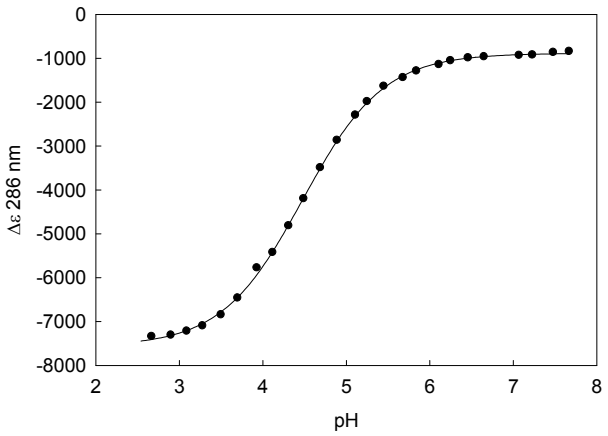


Figure 4.3: Fit of eq 4.2 to differences in absorbance between uracil and cytosine vs pH values 2.6- 7.6, as measured at 286 nm. The constants derived [$\Delta\varepsilon_{286}$ = -7566 + (-886+7566)/(1+ $10^{(pH-4.48)}$ (-0.90))] were utilized in the pH profile data from pH 5.0-7.5.

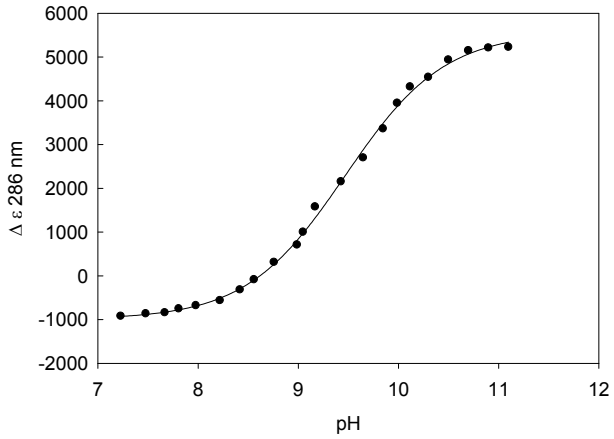


Figure 4.4: Fit of eq 4.2 to differences in absorbance between uracil and cytosine vs pH values 7.2-11.1 as measured at 286 nm. The constants derived [$\Delta\varepsilon_{286}$ = -998 + (5574+998)/(1+ $10^{(pH-9.47)}$(-0.88))] were utilized in the pH profile data from pH 7.3- 8.5.

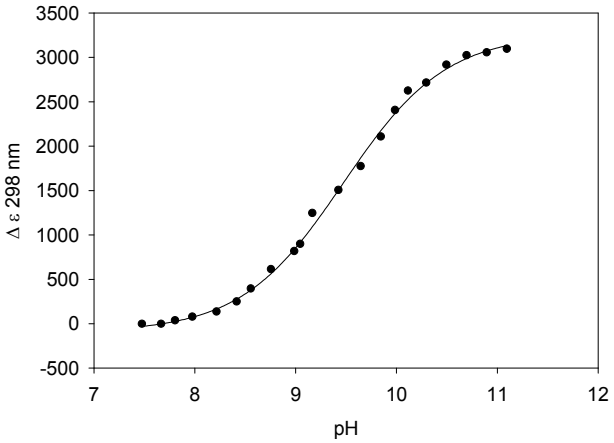


Figure 4.5: Fit of eq 4.2 to differences in absorbance between uracil and cytosine vs pH values 7.4-11.1 as measured at 298 nm. The constants derived [$\Delta\varepsilon_{298}$ = -95 + (3265+95)/(1+ $10^{(pH-9.47)}$ (0.86))] were utilized in the pH profile data from pH 7.3-10.5.

The log k_{cat} profiles for both the Zn (Figure 4.6 A) and the Fe containing CDA (Figure 4.7 A) show a loss of activity at low pH, while there is virtually no loss of activity up to pH 10. The pK_a for this loss of activity at low pH was determined to be 7.3 ± 0.1 for the Zn^{2+} enzyme, and 6.0 ± 0.1 for the Fe^{2+} enzyme as determined by fitting the data sets to eq 4.3. The log k_{cat}/K_m pH rate profiles (Figures 4.6 B and 4.7 B) show a loss of activity for Zn^{2+} CDA at low pH with a pK_a of 5.5 ± 0.2 while there is no observable loss in activity at low pH for Fe^{2+} CDA. In those k_{cat}/K_m profiles, both enzymes show a slight wave shaped plateau with a transition to a steady, slightly lower activity level at pH values above 8, resulting from small increases in K_m. The drop in k_{cat}/K_m occurs with a pK_b of pH 7.9 ± 0.4 for Zn^{2+} and 8.2 ± 0.2 for Fe^{2+} CDA. The Zn CDA dataset was fitted to eq 4.4, while the Fe CDA log k_{cat}/K_m dataset was fitted to eq 4.5. The profiles for H247Q mutant show a low level of activity which is independent of pH between pH 5 and 10 for either the log k_{cat} or the log k_{cat}/K_m as seen in Figure 4.8. These results are summarized below in Table 4.3.

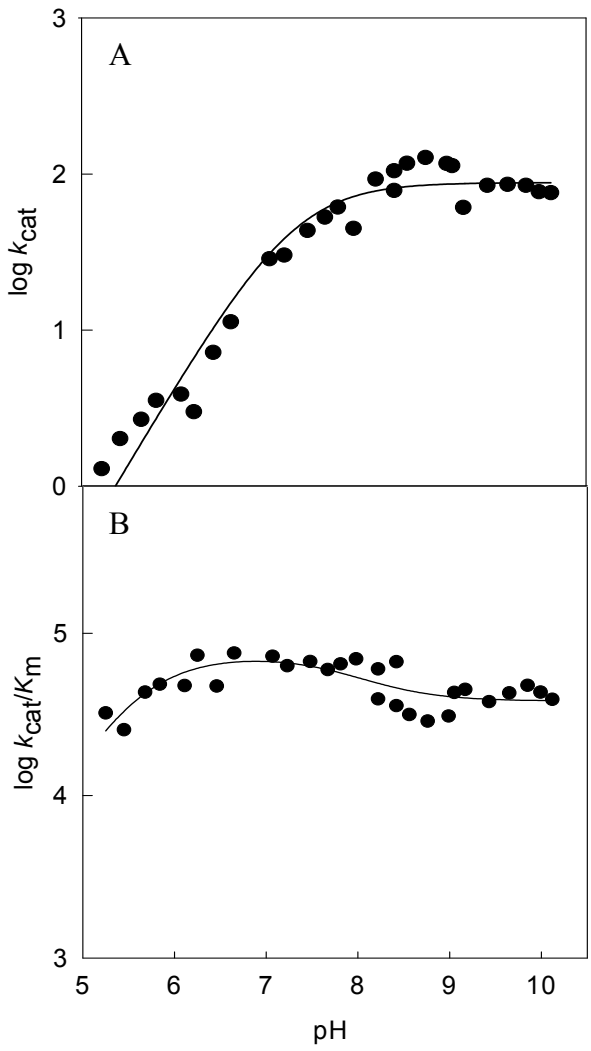


Figure 4.6: pH-rate profiles for the deamination of cytosine by Zn-CDA. (A) log k_{cat} vs pH profile for Zn-CDA. The solid line represents a fit of the data with eq 2.3 yeilding a kinetic pK_a of 7.3. (B) log k_{cat}/K_m vs pH profile for Zn-CDA. The solid line represents a fit of the data with eq 4.4 yeilding a kinetic pK_a of 5.5 and a pK_b of 7.9.

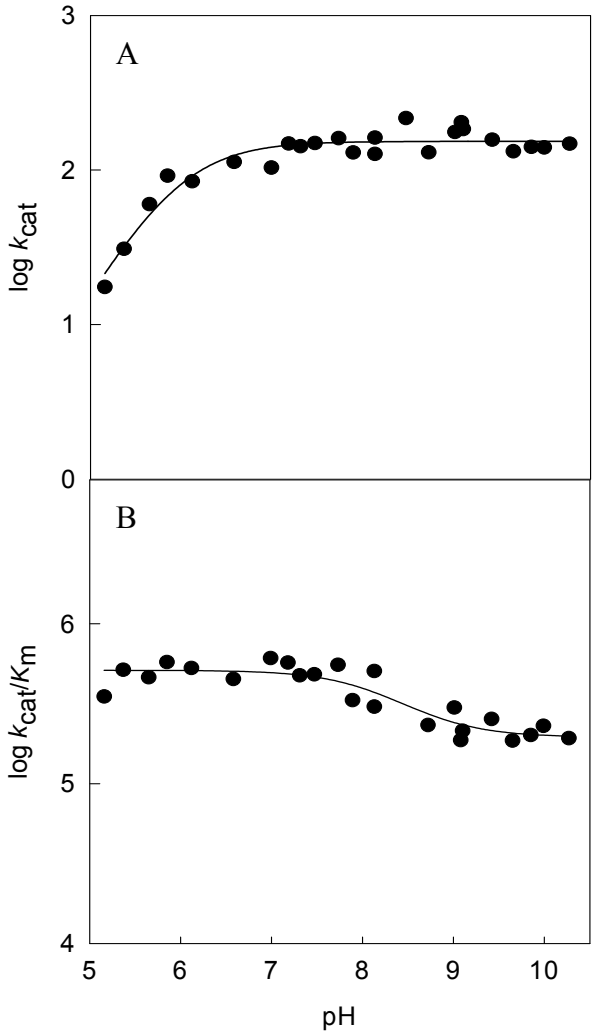


Figure 4.7 pH-rate profiles for the deamination of cytosine by Fe-CDA. (A) log k_{cat} vs pH profile for Fe-CDA. The solid line represents a fit of the data with eq 2.3 yeilding a kinetic pK_a of 6.0. (B) log k_{cat}/K_m vs pH profile for Fe-CDA. The solid line represents a fit of the data to eq 4.5 yeilding a pK_b of 8.2.

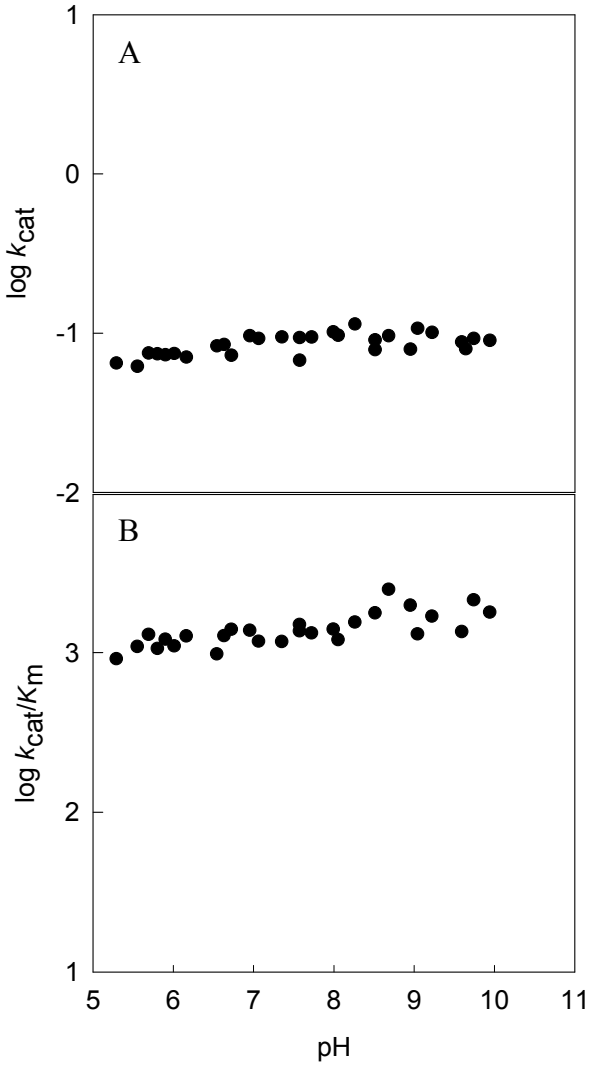


Figure 4.8: pH-rate profiles for the deamination of cytosine by Zn containing H247Q CDA. (A) log k_{cat} vs pH profile for Zn containing H247Q-CDA. (B) log k_{cat}/K_m vs pH profile for Zn containing H247Q-CDA.

Table 4.3: Kinetic pK_a Values from pH-Rate Profiles of Zn and Fe Substituted CDA and Mutant.

		pK_a		pK_b
enzyme	metal	k_{cat} vs pH	k_{cat}/K_m vs pH	k_{cat}/K_m vs pH
wild type	Zn^{2+}	7.3 ± 0.1	5.5 ± 0.2	7.9 ± 0.4
wild type	Fe^{2+}	6.0 ± 0.1	none observed	8.2 ± 0.2
H247Q	Zn^{2+}	none observed	none observed	none observed

Solvent D_2O Kinetic Isotope Effects and Proton Inventories. Solvent deuterium isotope effects are expressed as ratios of rate constants measured in H_2O (k_H) and in D_2O (k_D). The rate constants for the deamination of cytosine by Zn-CDA was found to depend highly on D_2O content, resulting in a large inverse solvent isotope effect (k_H/k_D) of 0.70 for k_{cat} and 0.43 for k_{cat}/K_m in 100 mM HEPES, pH 8.5. This pH(D) was chosen after evaluation of the pH rate profiles to avoid pK_a perturbations arising from the different isotopes. Similar effects of 0.62 and 0.40 for k_{cat} and k_{cat}/K_m respectively, were also measured at pH(D) 9.0.

For an enzyme catalyzed reaction, the ratios of the rate constants k_H and k_D share a relationship with the isotopic fractionation factors for the reactant state site (Φ^R) and the transition state site (Φ^T), where $k_H/k_D = \Phi^R/\Phi^T$. The magnitude of a fractionation factor measures the deuterium preference of a particular site compared to solution (*121*). For the deamination of cytosine by Zn-CDA, the atom ratio of deuterium in water was varied using 0, 25, 50, 75 and 100% D_2O in 100 mM HEPES buffer pH(D) 8.5 at 30 ˚C. The data for k_{cat} were best fit to eq 4.6, (the Gross-Butler equation for a loose reactant state) yielding $\Phi^R = 0.70$ +/- 0.01 as shown in figure 4.9 A. A straight line (dashed line)

would indicate a single "tight" transition site with $\Phi^T = 1.43$. Figure 4.9 B shows that the data for k_{cat}/K_m best fit to eq 4.7 (the Gross-Butler equation for a single "tight" transition state) with a $\Phi^T = 2.3$ +/- 0.1.

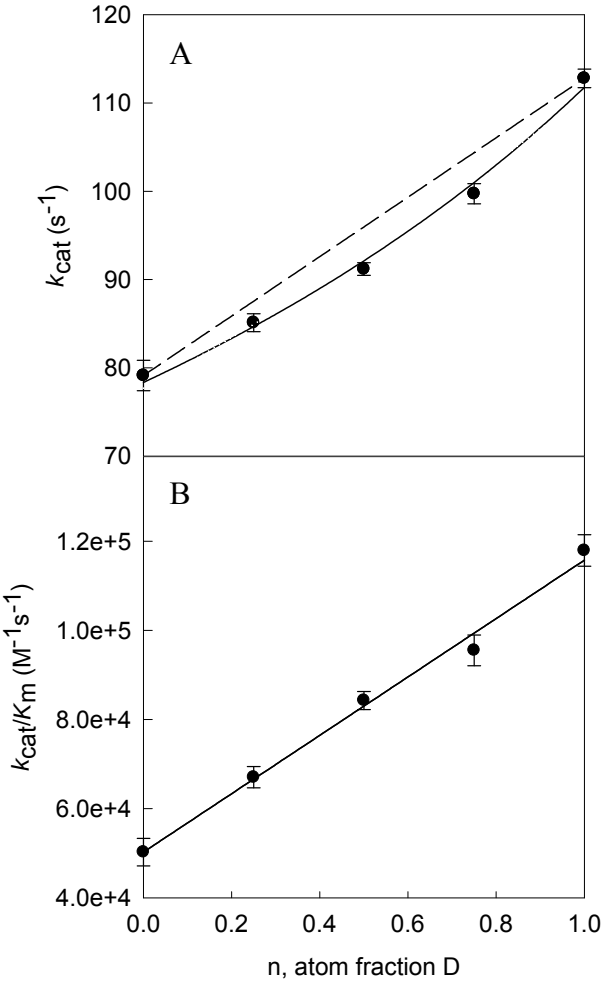


Figure 4.9: Proton inventories for the deamination of cytosine by Zn-CDA. This was performed in 100 mM HEPES pH(D) 8.5 at 30 ˚C. (A) k_{cat} vs atom fraction D fit to the Gross-Butler equation for a single loose reactant state site (eq 4.6). The straight dashed line serves to emphasize the degree of curvature. (B) k_{cat}/K_m vs atom fraction D fit to the Gross-Butler straight line equation for a single tight transition state site (eq 4.7).

Solvent Viscosity Studies: The effect of viscosity on the k_{cat} and k_{cat}/K_m for the deamination of cytosine by Zn CDA was measured with sucrose as the micro viscogen at pH 6.0, 7.5 and 9.0. At pH 6.0, a slope of 0.11 was obtained for the relative k_{cat} (Figure 4.10 A), while there was a greater effect for the relative k_{cat}/K_m (Figure 4.10 B) with a more pronounced slope of 0.27. For pH 7.5 there was a negligible slope of 0.04 for k_{cat} (Figure 4.11 A) and a larger slope of 0.37 for k_{cat}/K_m (Figure 4.11 B). However, on the k_{cat}/K_m (B) plot, a single data point at high viscosity appears to be out of place and may be pulling the slope to an artificially high value. Values obtained at pH 9.0 were less pronounced at 0.08 and 0.09 for the relative k_{cat} (Figure 4.12 A) and k_{cat}/K_m (Figure 4.12 B) plots respectively. Viscosity effects were also measured in D_2O, in an effort to see if the slight viscosity effects observed could be enhanced as a result of rate limiting step perturbation (data not shown). However, there was no increase in viscosity effects seen with D_2O.

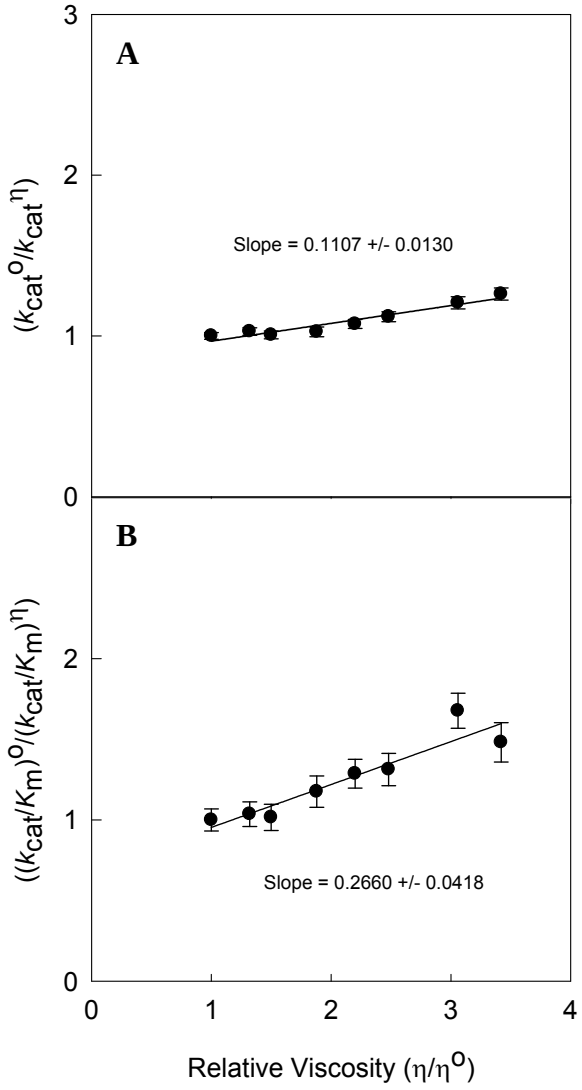


Figure 4.10: Viscosity effects on the relative values of k_{cat} (A) and k_{cat}/K_m (B) for Zn-CDA using sucrose as the micro-viscogen at pH 6.0. Additional details are available in the text.

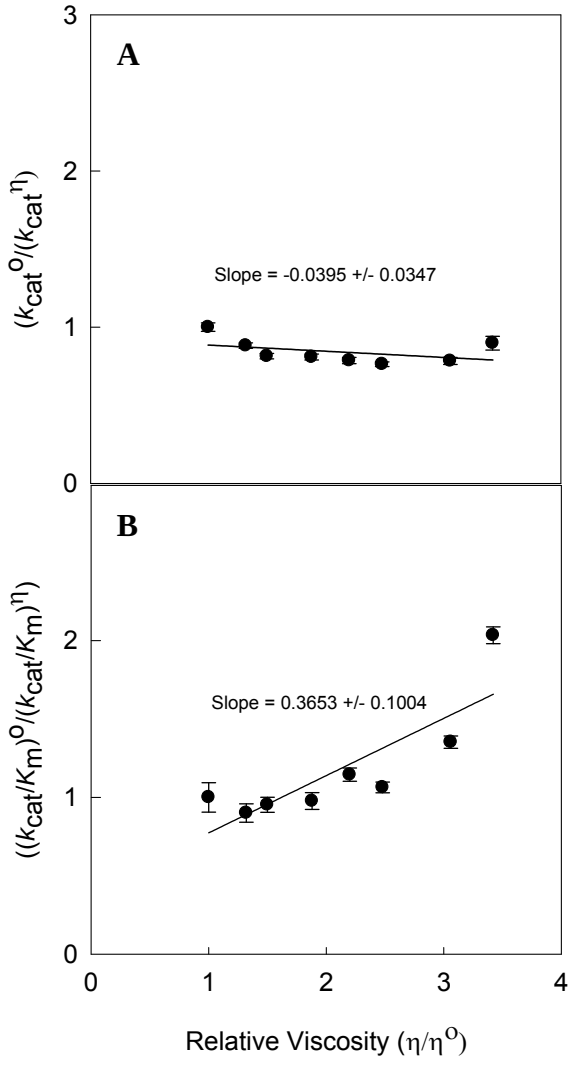


Figure 4.11: Viscosity effects on the relative values of k_{cat} (A) and k_{cat}/K_m (B) for Zn-CDA using sucrose as the micro-viscogen at pH 7.5. Additional details are available in the text.

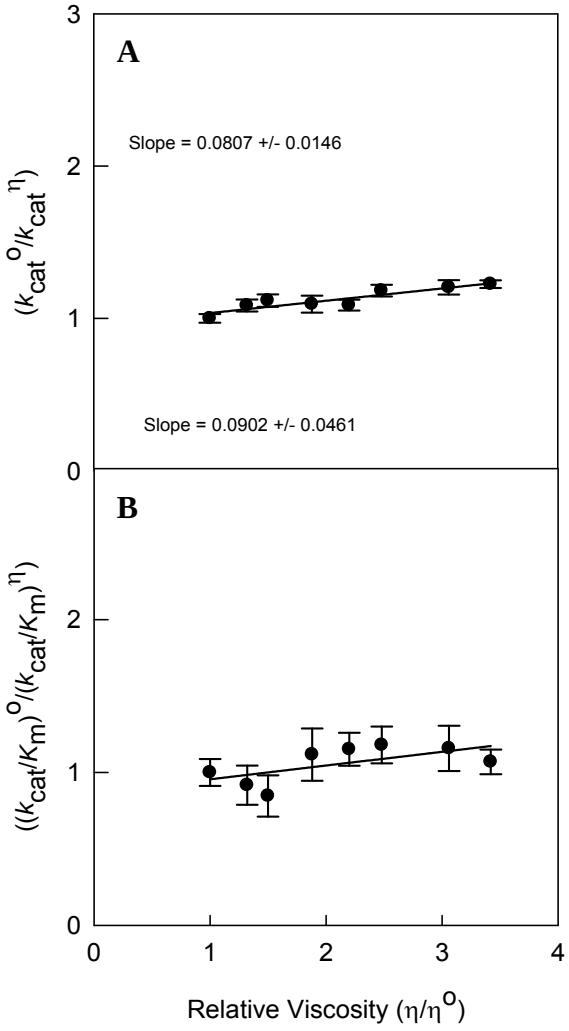


Figure 4.12: Viscosity effects on the relative values of k_{cat} (A) and k_{cat}/K_m (B) for Zn-CDA using sucrose as the micro-viscogen at pH 9.0. Additional details are available in the text.

Transition-State Inhibitors. Two phosphorus based transition-state analog inhibitors were selected as probes CDA's affinity for the transition-state of cytosine deamination (see Scheme 4.8). The effectiveness of the inhibitors was evaluated with the Zn^{2+} containing CDA at pH 7.5. Inhibition constants were determined through inhibition

pattern fits with a quadratic equation for tight binding inhibitors (*108, 109*) as shown in eq 2.8.

Scheme 4.6

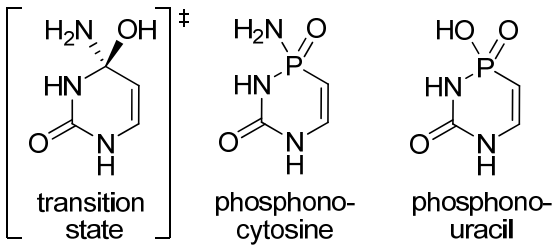


Figure 4.13 A shows the K_i determination of the phosphonocytosine inhibitor. Phosphonocytosine serves as an excellent slow, tight binding inhibitor of Zn-CDA with a K_i of 52 ± 5 nM. Figure 4.13 B shows inhibition of Zn-CDA with the phosphonouracil inhibitor at pH 7.5, where K_i was determined to be 84 ±7 μM. The inhibitor phosphonocytosine is the closest mimic of the transition state and as such, is more than a thousand times better inhibitor than the inhibitor which resembles uracil (phosphonouracil). The large difference in the inhibition constants underscores the subtleties involved in substrate binding and transition state stabilization employed by CDA.

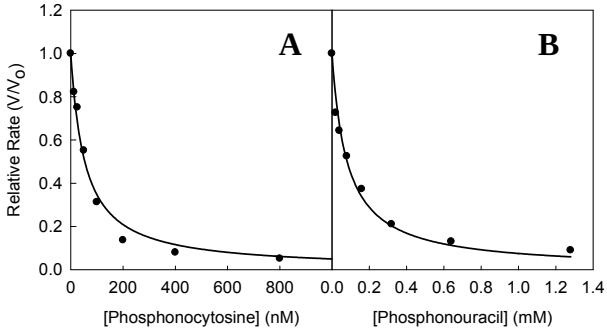


Figure 4.13: Inhibition of 2.5 nM Zn-CDA in 50 mM TRIS at pH 7.5 with (A) phosphonocytosine and (B) phophonouracil. Solid lines represent fits of the data with eq 4.8. Enzyme and inhibitor solutions were pre-incubated together for 30 min at 30 ˚C prior to initiating the reactions with 0.2 mM cytosine.

Eq 2.9 was fit to time courses for the changes in the rates of cytosine deamination, yielding the observed first-order rate constants (k_{obs}) for the slow onset of inhibition by phosphonocytosine and phosphonouracil. The obtained rates (k_{obs}) were then plotted vs inhibitor concentrations yielding the apparent second-order rate constants (appk_{obs}) as seen below in Figure 4.14 for phosphonocytosine and in Figure 4.15 for phosphonouracil. The linearity of these plots indicate that inhibition follows the simple model for inhibitor binding shown below in Scheme 4.9.

Scheme 4.9

$$E \underset{k_2}{\overset{k_1 I}{\rightleftharpoons}} EI$$

Correction for substrate concentration according to eq 2.10 yields the second order rate constant, k_1. These rate constants were determined using cytosine at a concentration equal to the K_m, therefore according to the equation, k_1 = appk_{obs}. For the inhibition of Zn-CDA by phosphonocytosine, $k_1 = 69 \pm 7$ M^{-1} s^{-1}. For the inhibition of

Zn-CDA by phosphonouracil, $k_1 = 0.18 \pm 1\ M^{-1}\ s^{-1}$. Both inhibitors were found to be reversibly inhibitory after a 100 fold dilution into 2.0 mM cytosine. Extrapolation of the linear fit in Figure 4.14 to the y-axis gives an estimate of k_2 of $(1.4 \pm 0.6) \times 10^{-4}\ s^{-1}$for phosphonocytosine. The y-intercept in Figure 4.15 yields a k_2 of $(2.5 \pm 0.4) \times 10^{-4}\ s^{-1}$ for phosphonouracil.

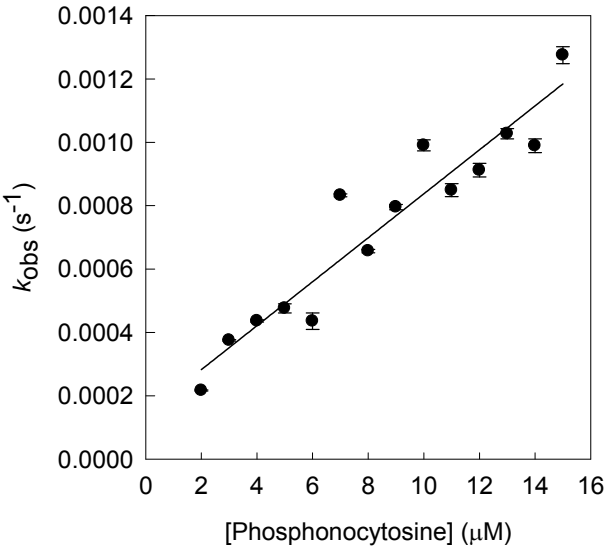


Figure 4.14: Determination of the second order rate constant (k_1) for the onset of inhibition of Zn-CDA by phosphonocytosine at pH 7.5.

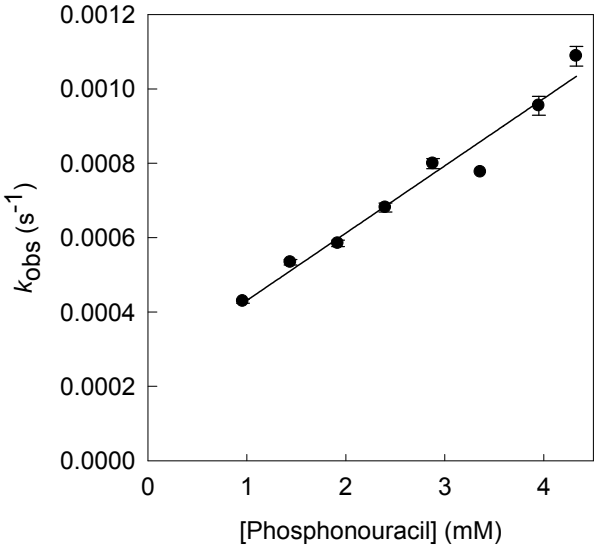


Figure 4.15: Determination of the second order rate constant (k_I) for the onset of inhibition of Zn-CDA by phosphonouracil at pH 7.5.

DISCUSSION

Metal Requirement of CDA. CDA has been described as an Fe^{2+} dependent enzyme, due to what was considered to be low activity when Zn^{2+} is bound. From the metal studies shown here and in previous studies (*22*) it is evident that cytosine deaminase has a high affinity with both Fe^{2+} and Zn^{2+}. The metal chelation and reconstitution results demonstrate the absolute requirement of divalent transition metals for CDA activity and that exactly 1 molar equivalent of Zn^{2+} provides maximum activity. This Zn^{2+} containing CDA reaches maximum activity above pH values of 7.3, while the Fe^{2+} containing enzyme possesses maximum activity down to pH 5.9.

Active Site Mutants. Mutants were expressed in CDA knockout cells to eliminate any ambiguities from slightly active mutants. As a result the Q157N, Q157A, E218Q,

E218A, H247A and D314A mutants failed to yield activity detectable by our assay methods. The activity seen for the H247Q and D314N mutants presents a dilemma because it appeared that they are essential components of catalysis. Evaluation of the crystal structure shows that His-247 and Asp-314 may both be involved in activation of the active site water. Perhaps the reason why there is some residual activity for both mutants is that the remaining residue may be able to compensate for the mutation in some unknown way.

The structural requirement of the substrate by CDA was addressed through evaluations of the substrate analogues in conjunction with site directed mutagenesis. From an examination of the only substrates displaying activity and the very similar structural patterns they possess, it appears that this enzyme maintains strict substrate specificity through the use of a small, highly cooperative active site, which is not accommodating to variations in structure.

As seen in the crystal structure of PDB 1K70, Glu-218 is hydrogen bonding with N-3 of the pyrimidine. It is expected that N-3 becomes protonated by this glutamate during the reaction. This was also concluded for ADA (*17*) and AMP deaminase (*88*) The lack of detectable activity seen for the Glu-218 mutations show that this residue is required except when no protonation is needed at N-3. 3-oxauracil served as an excellent mechanistic probe of the active-site mutants. When 3-oxauracil is used as a substrate, no protonation is needed, and hydrolysis of 3-oxauracil was found to proceed with the E218A mutant at nearly wild-type levels.

pH Studies. The pH dependencies of the extinction coefficients of cytosine and uracil were evaluated for the design of a direct assay for use during the evaluation of the pH profiles. On the extinction coefficient plots, there are two inflections which correspond to the pK_a of cytosine at pH 4.5 and the pK_a of uracil at pH 9.5. This effectively eliminated the effects of the catastrophic loss of coupling system activity at pH values greater than 9.

The pH profiles indicate that the Fe^{2+} containing enzyme is better at activating the hydrolytic water molecule, as evidenced by the broader range of activity seen for the Fe^{2+} containing enzymes where the log k_{cat} pH profiles showed a pK_a which was lower by more than 1.3 pH units as compared to the Zn^{2+} enzyme. These results are counter to previous model studies of the effectiveness of various metals toward the activation of water. In those studies Zn^{2+} is the better Lewis acid (*118*).

When His-247 is mutated to glutamine, there is a level of activity an order of magnitude lower than the lowest activities measured for wild type enzyme at pH 5.0. The lack of pH effect on the profile with H247Q and the low level of activity detected, supports that His-247 is responsible for the ionization seen in the pH profiles. His-247 seems to be an important catalytic base that needs to be de-protonated for maximum activity. Perhaps if the wild type enzyme could be measured at pH values lower than 5, the activity would level out at this mutant level of activity. If Asp-314 or Glu-218 were required to be protonated in the resting state, then it would be expected to see a loss of activity before pH 10. There was no great loss of activity at high pH for any pH profile performed. This clearly supports that both glutamate and aspartate are de-protonated in

the resting state of CDA. It may still be possible that protonated residues are stabilized by the enzyme in the active site and there is no general consensus for the protonation state of ADA. In different reports, it has been concluded either that both glutamate and aspartate are protonated (*89*) or glutamate and histidine are protonated (*17*).

Solvent Kinetic Isotope Effects and Proton Inventories. In D_2O, CDA catalyzes the deamination of cytosine at rates about twice as fast as those in water, yielding isotope effects (k_H/k_D) of 0.70 for k_{cat} and 0.43 for k_{cat}/K_m .The relatively large inverse isotope effects seen here are indicative of a compressed hydrogen bond in the rate-limiting step. Inverse isotope effects of comparable magnitudes were previously reported for AMP deaminase (*88*) and adenine deaminase (*122, 123*) where it was attributed to short hydrogen bonds between proton donors and acceptors with similar pK_a values.

The proton inventory data for k_{cat} were fit with the Gross-Butler equation showing a single proton in a loose reactant site, with $\Phi^R = 0.70$ These results are consistent with those seen for ADA, where it was concluded that a single proton site which has its fractionation factor changed during catalysis is responsible for this type of inverse isotope effects (*122*). It was concluded that the effects are likely to be the result of a proton transfer from Glu-217 to N1 of adenosine (*123*). The data for k_{cat}/K_m best fit the equation for a single proton in a tight transition state with $\Phi^T = 2.3$. If there were multiple sites experiencing a tightened potential as the transition state was formed, a line of intermediate curvature would provide a better fit (*121*), as was the observation for AMP deaminase with dAMP (*88*). In that study, it was proposed that multiple proton

transfers could occur in a relay between Zn-bound water, Asp-707, His-652, Glu-633 and the departing NH_3.

Inverse solvent isotope effects arise from pre-transition state equilibrium effects, and in this case they may be the result of a compressed hydrogen bond between Glu-218 and N3 of cytosine, which is analogous to the proposal for ADA. This could also explain why the inverse effects are much more pronounced in k_{cat}/K_m if that specific interaction were the major contributor for binding and catalysis under sub-saturating conditions, but not saturating conditions. Under saturating conditions, in mostly H_2O, the compressed hydrogen bond in question may not be so dominating a feature of the reaction coordinate, but may become more pronounced under increasing D_2O concentrations.

Viscosity Effects. The viscosity effects observed were slight and there was a slightly larger viscosity effect at pH 6 compared to pH 9, indicating that diffusion is not rate limiting at high pH, but may become more of a factor at low pH. This effect may be related to the lowering of the K_m at low pH. Small viscosity effects were seen using sucrose as a microviscogen indicating that the deamination of cytosine by Zn-CDA is not diffusion limited. It was demonstrated above that Zn-CDA turns over cytosine at rates that are about two times faster in D_2O. If the slight viscosity effects seen in H_2O were significant, they should be more pronounced in D_2O if the chemical step was no longer the rate limiting step. However, the studies carried out in D_2O failed to increase any viscosity effect, indicating that the turnover is still dictated by the chemical step. Any effect seen with the viscogen was probably not due to perturbation of diffusion and possibly arose from some low level inhibition of CDA by high levels of sucrose.

Inhibitor Studies. Phosphonocytosine proved to be an excellent inhibitor, yielding a K_i that was about a 1600 times lower than the inhibition constant of phosphonouracil and is 20,000 times lower than the Michaelis constant for cytosine. The phosphonouracil inhibitor yielded a K_i that is only 12 times lower than the Michaelis constant. The inhibition studies of these two transition state analogs show that the closer mimic of the transition state is a much more effective inhibitor of the reaction as well. From an evaluation of the individual rate constants, k_1 and k_2, it appears that there is an appreciable amount of error associated with them. This can be seen by comparing the ratios of k_2/k_1 to K_i. For phosphonocytosine, K_i = 52 nM, while k_2/k_1= 2.0 μM. For phosphonouracil, K_i = 84 μM, while k_2/k_1= 1.4 μM. These large differences most likely arise from the uncertainties in the determination of k_2, because small errors in the determination of k_1, have a larger effect on the error of k_2.

Mechanism of Cytosine Deamination. The mechanism for the deamination of cytosine by CDA has been evaluated through a host of mechanistic studies to identify the enzyme substrate interactions and their characteristics. From the studies presented in this report, several principles can be concluded. The active site metal has a large influence on the activation on the active site water, as seen from the large differences of kinetic pK_a values (1.3 pH units) observed between the Fe^{2+} and Zn^{2+} containing enzyme preparations. Steady activities up to pH 10 indicate that the catalytic residues are deprotonated. The low-level, pH independent activity seen for H247Q, indicate that His-247 is a major catalytic base responsible for pH dependent loss of activity at low pH. The activity of E218A with 3-oxauracil, shows that Glu-218 protonates N-3 of cytosine.

The extremely low activity (< 0.006 %) of the Asp-314 mutants, support the role of this residue as a major catalytic base as well.

A proposed mechanism for the deamination of cytosine by CDA is presented below in Figure 4.16. This is based on the interactions observed in the crystal structure of FeCDA (PDB: 1K6W), the crystal structure of Fe CDA with bound 2-oxopyrimidine inhibitor (PDB 1K70), and the results of these mechanistic studies. Based on experimental evidence showing activity up to a pH of 10, it is proposed that in the resting state, Glu-218, His-247 and Asp-314 are all deprotonated. A water molecule is coordinated to the divalent cation, and is hydrogen-bonded to Asp-314 and His-247 at nearly equal distances.

Prior to, or upon substrate binding, the hydrolytic water becomes deprotonated and Glu-218 becomes protonated. His-247 is reasonably positioned to perform this proton abstraction and subsequent transfer to Glu-218. The glutamate ε-O is 3.7 Å away from the histidine N^{δ}; however this distance can be decreased by axial rotation of either or both side chains. This initial deprotonation by His-247 is supported by the pH-activity profiles in several ways. First, the loss of activity which occurs at low pH has pK_a values ranging from 5.5-7.3, which are consistent with a histidine protonation. Second, with the H247N mutant, there is no discernible loss of activity at low pH.

The hydroxide then initiates the nucleophilic attack upon the sp^2 carbon bearing the leaving group amine. This is likely aided by the final deprotonation by Asp-314. N3 of cytosine then abstracts a proton from the H-bonded Glu-218. It was concluded for ADA that this interaction produces the inverse isotope effects due to the similar pK_as of the donor (~4) and acceptor (4.5), resulting in a low barrier H-bond (*123*). Given the typical H-bond distance of 2.8 Å between N3 and Glu-218, it could be possible that the solvent isotope effects arise from Zn-water deprotonation as proposed for AMP deaminase (*88*).

The resulting interactions fromGlu-218, His-247, Asp-314 and M_α**,** enable the formation and stabilization of the resulting tetrahedral transition state. Collapse of the tetrahedral intermediate is facilitated through protonation of the leaving group ammonia. While there is no obvious residue poised to perform this protonation, Asp-314 appears to be the most probable candidate at 3.6 Å away from the carbonyl carbon. Glu-218 is another possible participant for this action and is 3.7 Å away from the carbonyl carbon. Regeneration of the resting state occurs upon product release and the binding of water.

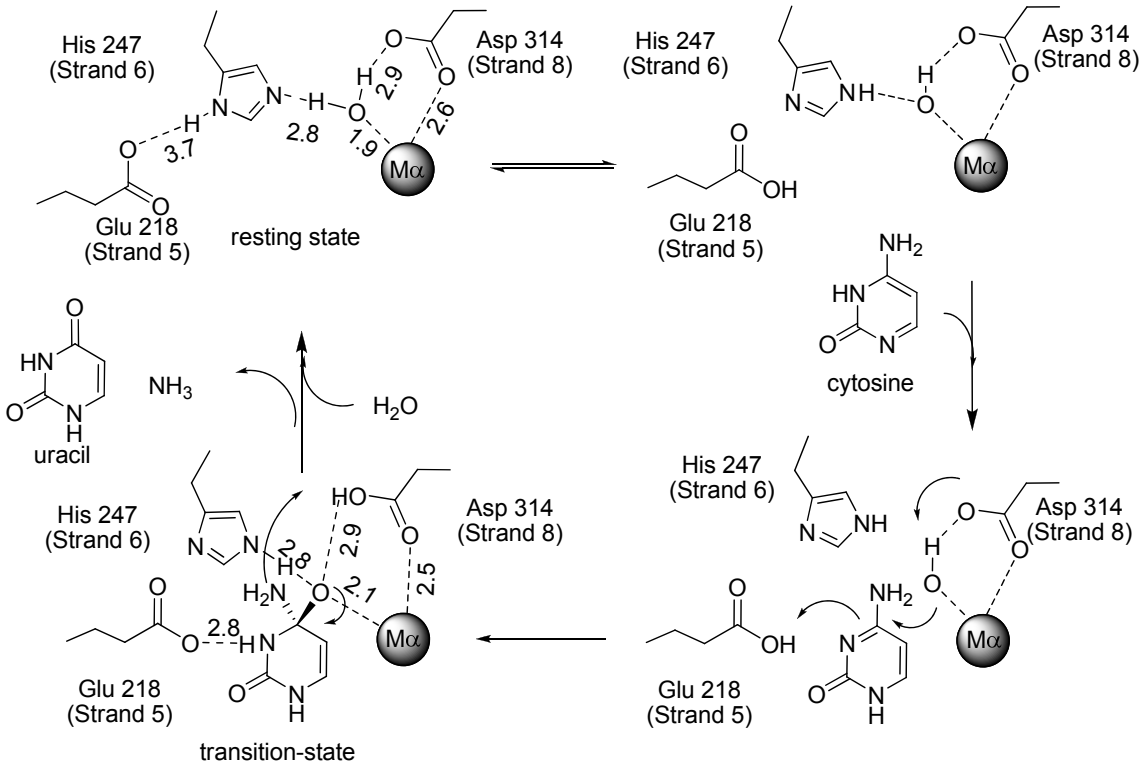


Figure 4.16: Proposed mechanism for the deamination of cytosine by CDA. Distances shown in the resting state are obtained from the crystal structure 1K6W of Fe-CDA with bound water. Distances shown in the transition state are obtained from the crystal structure 1K70 of Fe-CDA with bound mechanism based inhibitor 2-oxopyrimidine.

CHAPTER V

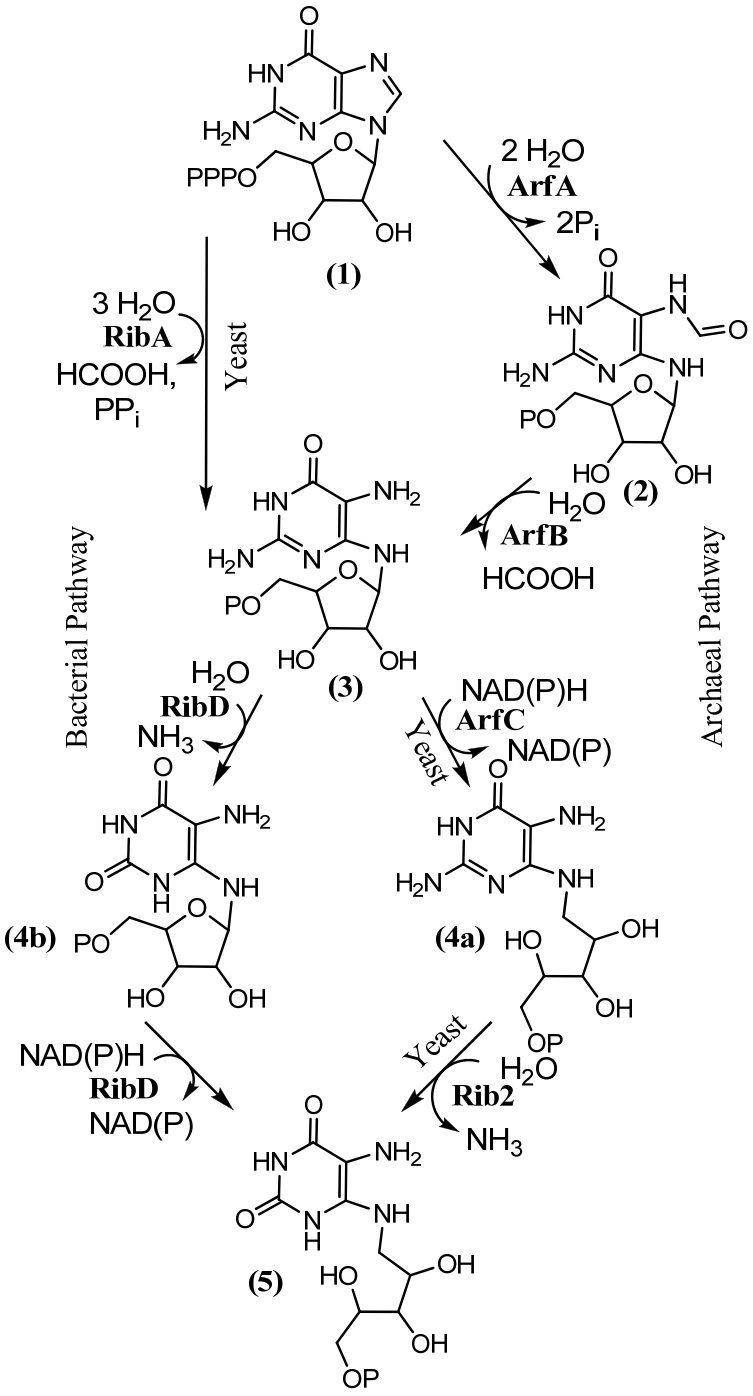
ATTEMPTS AT IDENTIFYING THE MISSING ARCHAEAL RIBOFLAVIN BIOSYNTHESIS PYRIMIDINE DEAMINASE

Riboflavin biosynthetic pathway genes are found in most plants, fungi, bacteria, and archaea but not in humans, as we rely on riboflavin absorption from our diet. The redox-active flavin cofactors produced from riboflavin are involved in many essential cellular processes (*124*). Specific examples are: one and two electron transitions, photorepair of thymidine dimers in photodamaged DNA and dehydration reactions. Flavins also function in light photoreception, bioluminescence and circadian time keeping (*125*). In archaea, the early steps of riboflavin biosynthesis may also be responsible for the biosynthesis of hydride carrier coenzyme F_{420}. Coenzyme F_{420} is essential for methanogenesis, sulfite reduction and oxygen reduction essential for archaeal survival (*126*).

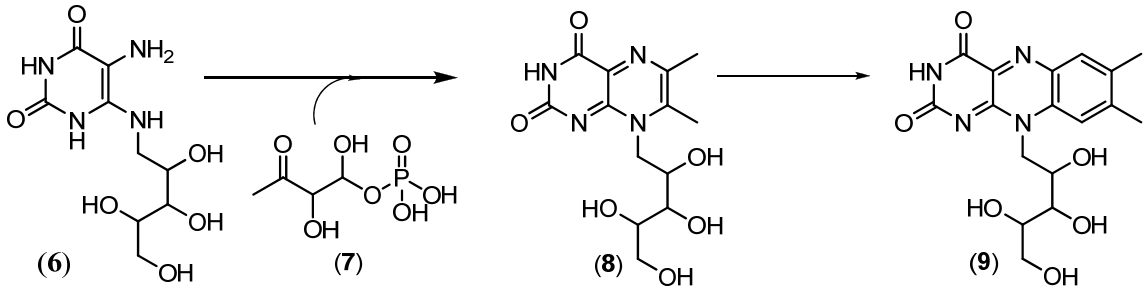
It has been discovered that there are several similar pathways which are utilized for the biosynthesis of riboflavin, depending on which type of organism is performing the synthesis. These reactions are summarized below in Scheme 5.1. For example, in bacteria and yeast, the first step in riboflavin biosynthesis is catalyzed by the Mg^{2+} dependent GTP cyclohydrolase II (RibA), which converts GTP (**1**) to 2,5-diamino-6-ribosylamino-4(3*H*)-pyrimidinone 5'-phosphate (**3**) by the release of formate and pyrophosphate. After the first step, the pathways diverge. In yeast, the second step is the NADPH dependent reduction of the sugar sidechain, producing NADP and 2,5-diamino-

6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinone 5'-phosphate (**4 a**). The third step in yeast riboflavin biosynthesis is the deamination at the C-2 of the pyrimidine by a separate enzyme, resulting in the formation of 5-amino-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinone 5'-phosphate (**5**) (*127*). In bacteria, the second and third steps are catalyzed by a bifunctional deaminase/reductase called ribD. It was determined that the deamination of **3** occurs first, producing **4 b**. This is then followed by the reduction of **4 b**, producing **5** (*124, 128*). **5** is presumably the substrate for an unidentified phosphatase which should produce 5-amino-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinone (**6**). Lumazine synthase catalyzed condensation of **6** with 3,4-dihydroxy-2-butanone 4-phosphate (**7**) yields 6,7-dimethyl-8-ribityllumazine (DMRL) (**8**), the substrate for riboflavin synthase. Riboflavin synthase then catalyzes the dismutation of DMRL, yielding riboflavin (**9**) and 5-amino-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinone (**6**). These reactions are shown below in Scheme 5.2.

Scheme 5.1


(1)
2 H2O
ArfA
2Pi
(2)
3 H2O
RibA
HCOOH,
PPi
Yeast
H2O
ArfB
HCOOH
(3)
Bacterial Pathway
Archaeal Pathway
H2O
RibD
NH3
NAD(P)H
ArfC
NAD(P)
Yeast
(4b)
(4a)
NAD(P)H
RibD
NAD(P)
Yeast
H2O
Rib2
NH3
(5)

Scheme 5.2



In archaea, riboflavin biosynthesis pathway is similar to that of yeast with a few exceptions. The first reaction is catalyzed by a different type of hydrolase, GTP hydrolase III (ArfA), which does not de-formylate the resulting pyrimidine (**2**). It was recently discovered that the release of formate from compound **2** required an additional enzyme (ArfB) which is an Fe^{2+} dependent formamide hydrolase (*126*). A reductase (ArfC) similar to the one in yeast has been identified which reduces the deformylated product **3** (*96*). The enzyme responsible for the next catalytic step in archaeal riboflavin biosynthesis remains unidentified and may catalyze the deamination of compound **4a** to compound **5.**

All known pyrimidine deaminases share a common fold, which is similar to yeast cytosine deaminase or bacterial cytidine deaminase that forms a fusion partner with the pyrimidine reductase. However, in archaea, this pyrimidine reductase has been identified and it is found in the absence of any identifiable deaminase (*96*). There is a group of three similar amidohydrolases in archaea, which are annotated as chlorohydrolase-like/ssna like/cytosine deaminase like. Though they share a generally low sequence identity, there are some common critical sequence characteristics. These

sequence similarities are shared with the SAH deaminase Tm0936 as well as other deaminases such as CDA as seen in Figure 5.1. While it is possible that any of these putative archaeal deaminases are a cytosine deaminase, a guanine deaminase or an SAH deaminase; the genomic context comparison centered on Sso0398 as shown in Figure 5.2 provides support that Sso0398 and the homologous enzymes Mm_0823, Mmp 1380, and mmarC7_0625 possibly represent the missing archaeal pyrimidine deaminase. While these enzymes are highly dissimilar from the bacterial pyrimidine deaminase, they are much like the bacterial cytosine deaminase, which catalyzes a similar deamination. This would provide an example of two different folds performing the same reaction in different organisms. This chapter describes the cloning, expression and the attempted purification of three putative deaminases from archaeal microorganisms which are predicted to be involved in riboflavin biosynthesis. Also included in this chapter are the cloning, expression and purification of the enzymes RibA and ArfC that are required to make the substrate **4a** for the missing putative deaminase. Lastly, the production of compound **4a** and the attempted detection of the production of compound **5** are discussed.

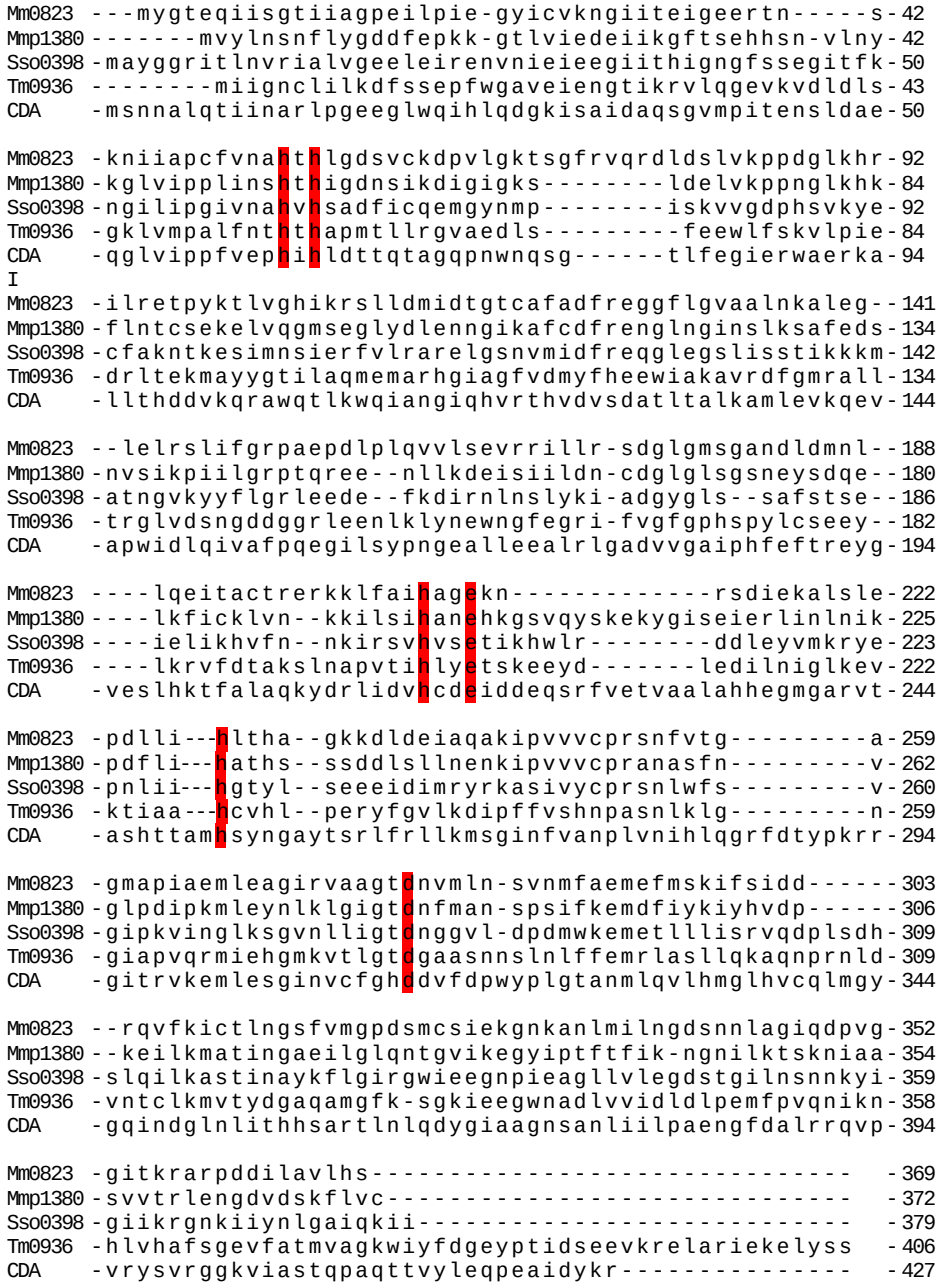


Figure 5.1: Multiple sequence alignment of three possible riboflavin deaminases from the archaea *M. mazei, M. maripaludis and S. solfataricus*, with the closely related SAH deaminase from *T. maritima* and CDA from *E. coli.*

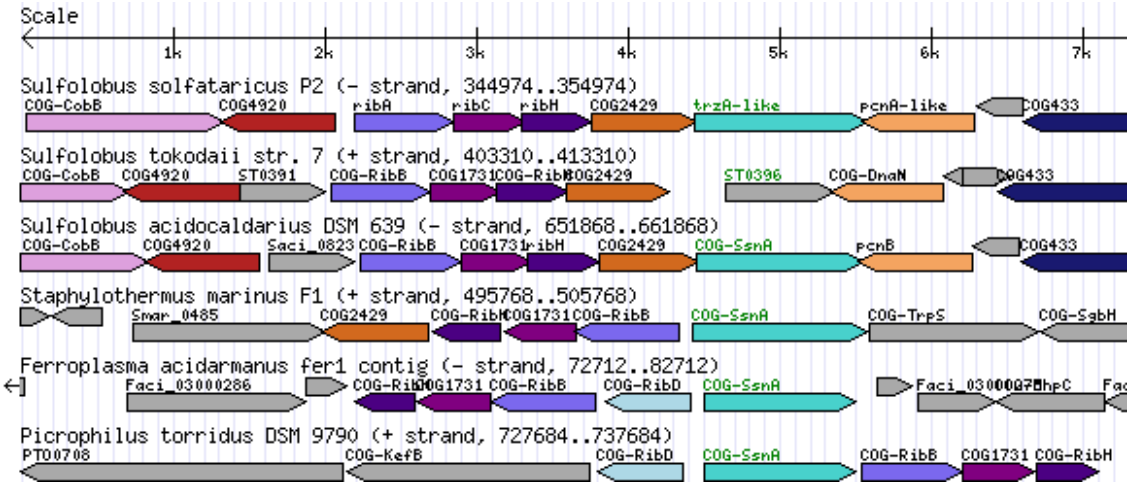


Figure 5.2: Genomic context tree of the putative deaminase from *S. solfataricus* obtained from Microbesonline.org. The center gene (teal) is the putative archaeal pyrimidine deaminases from Sulfolobus solfataricus which is labeled as trzA-like or COG-SsnA. COG2429 (dark orange) corresponds to ArfA. RibH (dark blue) corresponds to lumazine synthase. RibC/COG-1731 (purple) corresponds to riboflavin synthase. RibA/COGRibB (royal blue) corresponds to the domain responsible for synthesis of 3,4-dihydroxy-2-butanone 4-phosphate from ribulose-5-phosphate.

MATERIALS AND METHODS

Materials. All chemicals and buffers were obtained from Sigma-Aldrich unless specified otherwise. The genomic DNA for *Methanococcus maripaludis* C7, *Methanosarcina mazei* GO1 and *Sulfolobus solfataricus* were purchased from the American Type Culture Collection (ATCC). The oligonucleotide primer syntheses and DNA sequencing results were obtained from the Gene Technology Laboratory at Texas A&M University. A bacterial cell line expressing GTP cyclohydrolase II from *E. coli* (Rib A) was obtained as a frozen permanent culture containing the His6-RibA expression plasmid from the ASKA collection at the National Institute of Genetics, Japan. The pET20a(+) and pET28b(+) expression vector were acquired from Novagen along with the XL1 Blue and Rosetta PlysS DE3 cells. All restriction enzymes, as well as the pMAL-C4X maltose binding fusion protein expression vector, were purchased

from New England Bio-Labs. The Pfu-turbo DNA polymerase was purchased from Stratagene. For PCR clean-up and gel purification, the QIAquick Gel Extraction Kit was purchased from Qiagen. For miniprep purifications, the Wizard Plus SV Minipreps DNA Purification System was purchased from Promega.

Cloning and Expression and Purification. The putative pyrimidine reductase Mmp1348 from *Methanococcus maripaludis* was cloned into the pET28b(+) expression vector through the use of NcoI and BamHI restriction sites. *Methanosarcina mazei* reductase Mm0826 was successfully cloned into the pET-30b(+) expression vector through the use of NdeI and EcoRI restriction sites. The putative pyrimidine deaminase enzymes Mm0823 from *M. mazei,* MmarC7_0625 from *M. maripaludis* and Sso0398 from *S. solfataricus* were all cloned into a pET-20b(+) vector encoding for ampicillin resistance and IPTG inducible expression through the use of NdeI and XhoI restriction sites. For the PCR amplification reactions of these low GC content archaeal genes, a two- fold excess of $MgSO_4$ and the use of the polymerase PFU turbo in conjunction with 45 $^{\circ}$C and 68 $^{\circ}$C annealing and extention temperatures were successfully utilized for gene amplification. Once the sequences were verified through the use of T7 promoter and terminus primers in the forward and reverse directions respectively, the plasmids were transformed into BL21 DE3 Rosetta PlysS DE3 and Rosetta-gami DE3 cells, which provide rare codon expression and tight control over expression. For the production of C-terminal $(His)_6$ fusion tags, the stop codon was eliminated with Quickchange mutagenesis. For the production of N-$(His)_{10}$ factor Xa cleavable fusion proteins, the

genes were digested with NdeI and XhoI restriction enzymes and ligated into a pET-16b(+) vector.

For the production of RibA from *E. coli,* cells harboring the $(His)_6$RibA gene were used to inoculate a 5 mL overnight starter culture enriched with 25 μg/mL chloramphenicol. This was then grown overnight at 37 ºC. The entire overnight culture was then used to inoculate 1 L LB enriched with 25 μg/mL chloramphenicol. This was grown with shaking at 37 ºC until an absorbance at 600 nm of 0.6 was reached. At this time the heater was turned off and 1.0 mM IPTG was used to induce protein expression. The cell culture was allowed to grow overnight at room temperature with shaking after which they were centrifuged at 3400g for 12 minutes. 5 grams of the cells were then re-suspended in 50 mL buffer A (20 mM TRIS at pH 7.9, 5 mM imidazole, 500 mM NaCl and 100 μg/mL PMSF) and disrupted by sonication. Insoluble cell debris was removed by centrifugation at 25,000g for 15 minutes and then filtered with a 0.45 micron syringe filter. The entire protein solution was then loaded by gravity onto an Amersham Ni Sepharose 6 Fast Flow column pre-loaded with 1 column volume of 50 mM $NiCl_2$ and equilibrated with buffer A. After loading, the column bound protein was washed with 100 mL buffer A and eluted with a gradient of buffer B (20 mM TRIS at pH 7.9, 1 M imidazole, and 500 mM NaCl. The RibA containing fractions, as determined by SDS PAGE, were then pooled and concentrated with a 10,000 MWCO Centricon protein concentrator. The extinction coefficient of 12090 M^{-1} cm^{-1} was used to determine the final concentration. This was calculated based on the gene sequence through the use of Protein Calculator v3.3 at http://www.scripps.edu/~cdputnam/protcalc.html.

For production of the pyrimidine reductase from *Methanococcus maripaludis*, purified Mmp1348 containing pET28b(+) plasmid DNA was transformed into electro-competent Rosetta PlysS DE3 cells, followed by plating onto 50 µg/mL kanamycin containing agarose/LB plates. A single colony was used to inoculate a 5 mL overnight LB culture, which was then used to inoculate a 2 L culture. This was grown at 37 °C to an absorbance of 0.6 at 600 nm, after which 1.0 mM IPTG was added to induce expression. After 18 hours at 30 ºC, the cells were harvested by centrifugation and lysed by sonication in 50 mM HEPES, pH 8, 1.0 mM of the disulfide reducing tris-carboxyethylphosphine (TCEP) and 10 µg/mL PMSF. The nucleic acids were removed with 2% protamine sulfate, and then the protein was precipitated with ammonium sulfate at 50% of saturation. Lastly, the protein was purified by gel filtration and ion exchange chromatography. The reductase containing fractions were identified by SDS-PAGE and pooled. The sequence derived extinction coefficient of 19770 M^{-1} cm^{-1} was used to calculate the final concentration of protein.

The putative deaminases Mm0823 and MmarC7_0625 and Sso0398 were prepared with and without C-terminal $(His)_6$ fusion tags through the use of modified XhoI cloning primers lacking a stop codon and standard cloning methods described above. MmarC7_0625 was also prepared as an N-terminal $(His)_{10}$ Xa cleavable fusion protein by ligation into a pET-16b(+). Sso0398 was also cloned into the N-terminal, Factor Xa cleavable, cytoplasmic maltose binding protein fusion system pMAL-c4X through the use of XmnI and BamHI restriction sites.

The various putative deaminase expression vector constructs were expressed and purified as described above, but with additional modifications. Protein expression was induced at 17, 25, or 37 ^{o}C in BL21 λ DE3, Rosetta PLysS DE3, Rosetta-gami DE3 and BL21 λ DE3 cells harboring the PGro-7 vector encoding for the molecular chaperones GroEL and GroES. SDS-PAGE analysis of the lysis solution and lysis pellets showed clearly that every protein formed insoluble inclusion bodies. Notably, even the molecular chaperones GroEL and GroES were rendered insoluble upon expression of the putative deaminases.

High temperature expression at 45 ^{o}C was also attempted with the Sso0398 constructs in order to induce a heat shock response resulting in the production of natural chaperone proteins. This was achieved by growing the cells to an absorbance of 1.0 at 600 nm at a temperature of 37 ^{o}C, followed by the addition of 0.1 mM IPTG and subsequent incubation at 45 ^{o}C overnight. Additionally, growth and expressions were attempted for the various constructs in LB enriched with 500 mM NaCl and 1.0 mM glycine betaine at 25 ^{o}C.

The pMAL-c4X-Sso0398 fusion protein was transformed into BL21 λ DE3, Rosetta PLysS DE3 and Rosetta-Gami DE3 cells. SDS-PAGE analysis of small (5 mL) protein expression trials indicated that fusion protein expression was only observed in the Rosetta PLysS DE3 cells. An overnight culture of the Rosetta PLysS DE3 cells harboring the fusion protein was used to inoculate 1 L of rich media (LB + 0.2% glucose) enriched with 100 μg/mL ampicillin. The cells were grown with shaking at 37

ºC until an absorbance of 0.5 was reached at 600 nm. 0.3 mM IPTG was then used to induce expression of the fusion protein and the cells were grown with shaking overnight.

The cells were then harvested by centrifugation at 4000g for 20 min. 5 grams of cells were then lysed by sonication in 80 mL of 50 mM TRIS, pH 7.5 and 1 mM DTT. Insoluble debris was removed by centrifugation at 9000g for 30 min. The supernatant was then applied to 15 mL amylose resin column by gravity at a rate of 1 mL/min. The amylose resin bound fusion protein was then washed with 12 column volumes of buffer and the soluble fusion protein was eluted with 10 mM maltose containing buffer. The fractions containing the fusion protein were pooled and the concentration was determined using the entire fusion protein sequence derived extinction coefficient of 107000 M^{-1} cm^{-1}. The fusion protein was cleaved by incubating a 3 mg/mL solution of the maltose binding protein-Sso0398 fusion protein with 10 µL factor Xa in the Xa cleavage buffer, as directed by the instructions in the Factor Xa cleavage kit, overnight at 37 ºC.

Inclusion Body Extractions and Characterizations. For the attempted purification of the proteins which were expressed as inclusion bodies, the proteins were at least partially unfolded with either 9 M urea or 3.5 M guanidine HCl at pH 8.0 in 50 mM HEPES. The proteins were then dialyzed into less-denaturing levels of urea (5 M) at either 0 ^{o}C, room temperature, or 60 ^{o}C in the presence or absence of reducing agents tris-carboxyethylphosphine, DTT or BME and divalent cations, Fe^{2+}, Zn^{2+} or Mn^{2+}. The reduction in denaturant consistently resulted in the immediate precipitation of all proteins.

Dialysis at 60 ºC was attempted for the thermophilic Sso0398, which resulted in the production of a soluble and pure protein of the appropriate size. This protein was submitted for N-terminal protein sequence analysis by Dr. Larry Dangott at the Protein Chemistry Lab at Texas A&M University. The resulting N-terminal sequence was A(D)-E-I-Y-N-K-D this sequence was clear and consistent with that of outer membrane protein C from *E. coli.*

*Production of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-phosphate (**2**)and 2,5-diamino-6-ribitylamino-2,4(1H,3H)-pyrimidinone 5'-phosphate (**3**).* Typical preparation of **2** was performed as reported elsewhere (*124*) but with some modifications. Reactions containing 50 mM HEPES, pH 8, 5 mM $MgCl_2$, 1 mM GTP and 100 μM RibA and 5 mM DTT were incubated at 37 ºC for 30 min. Shifts in the absorbance spectrum from a λ_{max} of 260 nm to 296 nm confirmed the completion of the reaction. Enzyme was then removed upon passing the solution through a 30,000 MWCO Amicon membrane. The NADPH dependent reduction of **3** by Mmp1348 was monitored at 340 nm with an extinction coefficient for the formation of $NADP^+$ of 6220 M^{-1} cm^{-1}.

*Detection of 5-Amino-6-ribitylamino-2,4(1H,3H)-pyrimidinone 5'-phosphate (**4**).* Production of **5** in reactions containing compound **4a** and preparations of the putative deaminases were monitored through the coupled assay for ammonia release as described previously (*79*). Reactions containing compound **4a** and the putative deaminases were quenched with 40 mM 2,3-butanedione at 30 ºC for 30 min. The derivatized products were then separated by anion exchange in 20 mM TRIS with a gradient of 0-1M NaCl using an Amersham MonoQ anion exchange chromatography column. The products

were evaluated by comparison of UV spectra to those previously established (*124*) and submitted to the Department of Chemistry Laboratory for Biological Mass Spectrometry at Texas A&M University.

RESULTS AND DISCUSSION

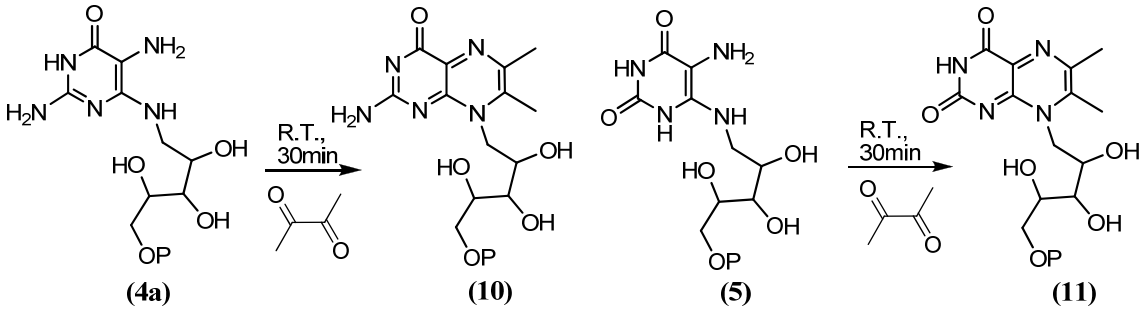
GTP cyclohydrolase II (RibA) from *E. coli* has been purified and the GTP cyclohydrolase activity was used to create the substrate (**3**) for the archaeal pyrimidine reductase (ArfB) Mmp1348. This compound was found to be stable at 0 ºC for 1 day. When flash-frozen in a dry ice / acetone bath and stored at -80 ºC, it was found to be stable for at least 1 month.

The reductase Mmp1348 has been cloned, purified and found to oxidize NADPH and NADH in the presence of the product of RibA. When combined with 120 μM of the RibA product (**3)**, Mmp1348 was found to oxidize an equal amount of NADPH. The initial rate of NADPH oxidation by Mmp1348 as determined with 120 μM of compound **3** was determined to be 0.85 s^{-1}. This rate is roughly 2.7 times faster than the k_{cat} of 19 min^{-1} that was determined for the *E. coli* pyrimidine reductase reaction of ribD with compound **4b** (*124*).

The product of this reduction reaction (**4a**) was derivatized with diacetyl to dimethylribitylpteridine-5-phosphate (**10)** shown in Scheme 5.1 and purified via anion exchange HPLC chromatography using 20 mM ammonium bicarbonate and a linear gradient of 1 M ammonium bicarbonate. The derivatized product was verified through a comparison to the published λmax of the de-phosphorylated compound **10**. The established peak absorbance wavelength values at pH 7 were 268, 325 and 414 nm (*129*)

and all agreed well with the experimental values obtained for a scan of the HPLC purified product in 20 mM TRIS buffer at pH 7.5. A spectrum of compound **10** is shown in Figure 5.3. Samples of the product were analyzed by Maldi-TOF +/- MS and Electrospray +/- MS using standard procedures by the Laboratory for Biological Mass Spectrometry in the Department of Chemistry. Results obtained failed to confirm the presence of compound **10**. These inconclusive results were possibly due to instabilities of the compound or from matrix and ionization interferences.

Scheme 5.3



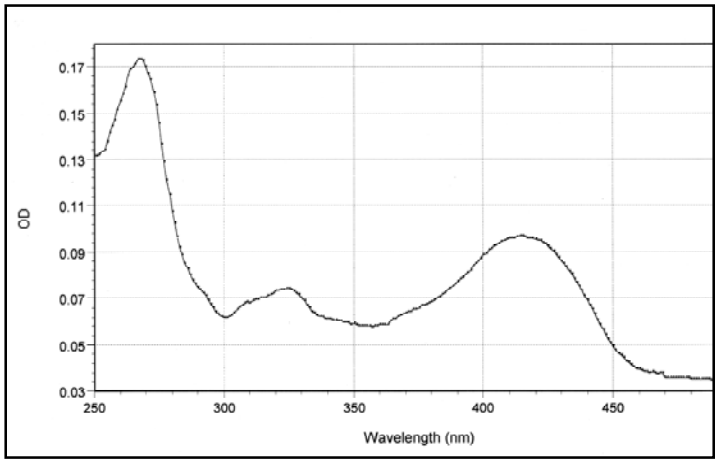


Figure 5.3: Spectrum of dimethylribitypterin-5-phosphate, pH 7.9 (compound **10**).

The putative pyrimidine deaminase enzymes from *M. mazei* and *M. maripaludis* C7 over-express well in the Rosetta PlysS DE3 cells at 17 °C with 1 mM additional $ZnCl_2$, however the protein products proved to be unyieldingly insoluble. The $(His)_{10}$ factor Xa cleavable MmarC7_0625 fusion protein did not express well in Rosetta-gami DE3 cells. The expression and purification of the protein products from inclusion bodies was attempted under the conditions listed in the Materials and Methods, but failed to yield soluble and active protein product of the proper sequence under any circumstance. The N-terminal protein sequences of the soluble protein were of the sequence DEIYNKD or AEIYNKD as verified using three different conditions. The sequence AEIYNKD is found near the N-terminus of an *E. coli* outer membrane porin C protein.

The pMAL-c4X-Sso0398 protein fusion did yield soluble fusion protein of appropriate molecular weight (85 kDa). The resulting protein solution which was purified using the amylose column was tested for metal content by ICP-MS. The protein was found to contain 0.02 eq Mn, 0.06 eq Zn and 0.08 eq Fe. It is possible that the maltose binding protein was somehow blocking the active site of Sso0398. To address this possibility, the maltose binding protein was removed from the N-terminus of Sso0398 by overnight incubation by the site-specific protease factor Xa, producing un-modified wild-type protein. Upon cleavage, some protein precipitation was visible. SDS-PAGE of the soluble cleavage products showed a single band which was consistent with either Sso0398 (42.65 kDa), the cleaved pMal protein (42.48 kDa) or factor Xa (42.32 kDa). N-Terminal sequencing of the major band indicated that the maltose binding

protein was present at greater than 95 % purity, indicating that if it was expressed, then Sso0398 had precipitated.

The crude cell extracts and the re-suspended inclusion bodies were tested for activity against **4a** with the glutamate dehydrogenase coupled assay, yet no ammonia release was detected. The HPLC purification and characterization of the derivatization products was also applied to the crude cell extracts and protein preparations in an effort to detect the slightest changes in the visible spectra expected if any amount of compound **7** were formed. Compound **4a** was treated with the putative deaminases for up to 24 hours. The subsequent derivatization and purification yielded products with spectra which are identical to those obtained without the putative deaminases (compound **10**).

In conclusion, the archaeal enzyme Mmp1348 from *Methanococcus maripaludis* annotated as topoisomerase/primase-like and dihydrofolate reductase-like by the NCBI, was successfully cloned and purified. This enzyme was effective at reducing the product of RibA as evidenced from the NADPH reduction observed. The UV/Vis absorbance spectrum of the diacetyl derivatized product, is consistent with the structure of 2-amino-6,7-dimethyl-8-ribityl lumazine 5'-phosphate (**6**). This evidence supports Mmp1348 production of **4b** from compound **3** and functioning as ArfC in archaea in the riboflavin biosynthesis pathway.

The wild type and His-tagged versions of the putative archaeal pyrimidine deaminases Mm0823, MmarC7_0625 and Sso0398, have proven to be insoluble and inactive against compound **4a**, as expressed under a wide variety of growth and purification conditions in several different strains of *E. coli*. The maltose binding protein

fusion construct prepared with Sso0398 and expressed in Rosetta pLysS cells resulted in soluble protein of the proper size that was successfully purified with an amylose resin column. The resulting fusion protein had poor metal content in spite of attempts to grow the protein in media enriched with metal and at various temperatures with various levels of inducing agent.

Visible protein precipitation upon incubation with factor Xa indicated that the deaminase was precipitating upon cleavage. This hypothesis was supported by N-terminal sequence analysis of the remaining soluble protein which showed that the maltose binding protein was the only protein present. A band of size which is consistent with the archaeal protein was visible by SDS-PAGE. Attempts at measuring deamination of **4a** were unsuccessful with all crude cell extracts and protein preparations. This was determined from the lack of ammonia detected by the glutamate dehydrogenase assay in addition to a lack of any evidence supporting the formation of compound **7**. While it is still very possible that any of these three putative deaminases are involved in riboflavin biosynthesis, they may require additional interactions with unknown proteins or cofactors for solubility and activity.

CHAPTER VI

DISCOVERY OF 8-OXOGUANINE DEAMINASE

Oxidative damage by reactive oxygen species to nucleotides, cofactors, metabolites and structural components can pose a significant problem to the metabolism of any organism if not promptly dealt with. As a result, it should not be surprising that there exist enzymes which have a specific purpose of recycling or converting the damaged compounds to more useful or common components of metabolism. Enzymes of this kind are typically not found in well established biological pathways making the purpose-driven identification of these unknown repair and salvage enzymes very challenging.

It is established that 8-oxoguanine is one of the most prevalent products of oxidative stress in cells (*130*). DFT calculations show that superoxide radical ($O^{\cdot}{}_2^-$), and singlet oxygen ($^1O_2$) readily react with guanine, producing the extremely mutagenic 8-oxoguanine (*131*). Inflammatory cytokines have been shown to increase NO levels, which has then been associated with mitochondrial guanine oxidation that may contribute significantly to neurological disorders such as multiple sclerosis (*132*). 8-oxoguanine readily pairs with adenine, producing mutations associated with many types of disease (*133*). It has been observed that there are specific enzymes (glycosylase/AP lyase) specific for the removal of oxidized guanine in DNA (*130, 134*), however the fate of the resulting 8-oxoguanine has not yet been described.

This chapter describes the discovery of function of two uncharacterized bacterial enzymes Pa0142 (gi: 9945972) from *Pseudomonas aeruginosa PA01* and the New York

SGX Research Center for Structural Genomics (NYSGXRC) protein target NYSGXRC-9236e (gi: 44264246). The crystal structure of NYSGXRC-9236e has been solved and assigned the PDB code: 3HPA. This enzyme was cloned from an environmental sample of the Sargasso Sea and the closest homolog from a sequenced organism (98% identical) is Bcep18194_A5267 from *Burkholderia sp. 383*. This is annotated by NCBI as a hydroxydechloroatrazine ethylaminohydrolase based on sequence similarities. Pa0142 has also been annotated as a hydroxydechloroatrazine ethylaminohydrolase.

Genomic context evaluations similar to the one shown below in Figure 6.1, show that Pa0142 and 9236e belong to a previously uncharacterized group of bacterial enzymes that are found near nucleotide salvage pathway enzymes including guanine deaminase, adenosine deaminase and xanthine/uracil permease. This context provides support that the unknown enzymes are involved in nucleotide metabolism, This information, when combined with the sequence similarities to known deaminases, indicated that these two enzymes may be deaminating compounds similar to pyrimidines and purines.

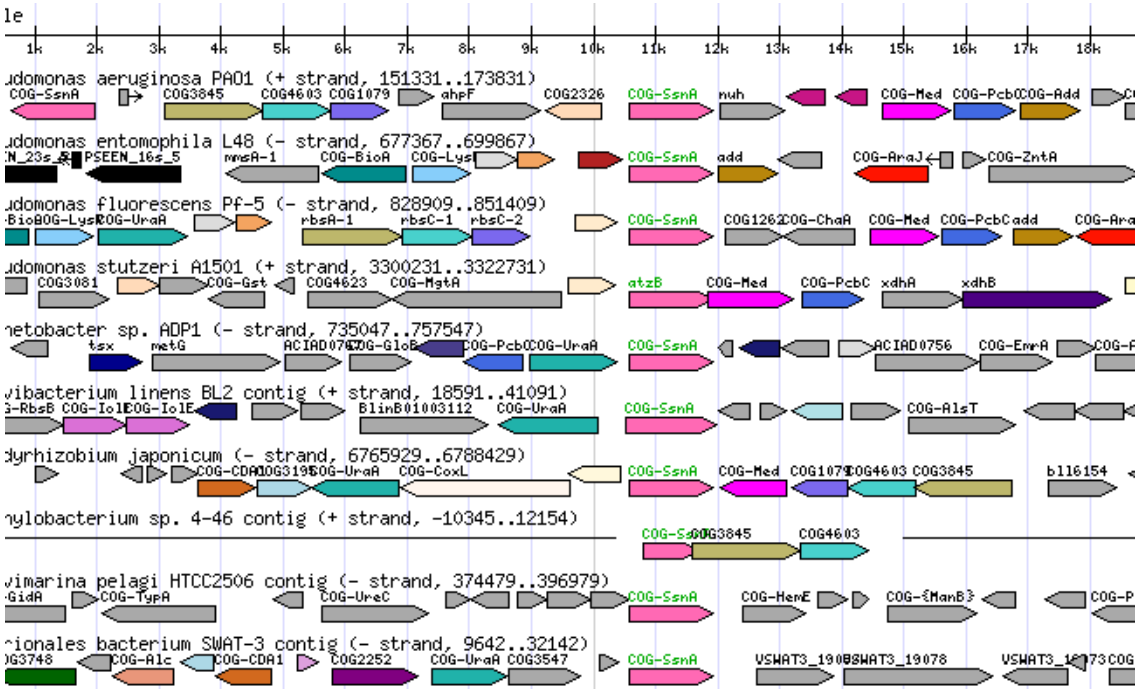


Figure 6.1: Genomic context tree of Pa0142 (center) obtained from Microbesonline.org. Nearby are nucleotide salvage enzymes including; the non-specific nucleoside hydrolase nuh (gray), xanthine/uracil permease UraA (teal), xanthine dehydrogenase xdhB (indigo), guanine deaminase guaD/COG-SsnA (pink) and adenosine deaminase add (brown).

Sequence evaluations enabled the identification of a group of genes that are similar to guanine and cytosine deaminase, yet the probable genes for guanine deaminase and cytosine deaminase are already present in these genomes. In *Pseudomonas aeruginosa*, Pa0142 is one of these genes. In fact, *P. aeruginosa* contains two additional genes Pa0134 and Pa1521, which are more like the known guanine deaminases than Pa0142. In an effort to resolve the function of these apparently redundant genes, these proteins of unknown function were selected for deaminase activity screening against a library of pteridines, purines and pyrimidines. The activity screening resulted in the discovery of activity for the nucleotides isocytosine, guanine and 8-oxoguanine.

In previous unpublished studies of Pa0142, Dr. Ricardo Marti-Arbona measured k_{cat}/K_m values of 5.7×10^3 M^{-1} s^{-1} for isocytosine, 2.0×10^3 M^{-1} s^{-1} for guanine, 2.0×10^3 M^{-1} s^{-1} for ammeline, 40 M^{-1} s^{-1} for pterine, and 30 M^{-1} s^{-1} for isoxanthopterin. In this chapter, the substrate specific activities of Pa0142 and SGX-9236e will be discussed and compared to the probable GDA (Pa0134) from *Pseudomonas aeruginosa* (gi: 15595332) and the similar *E. coli* proteins; GDA (gi: 89109663), SsnA (gi: 89109659), and YahJ (gi: 89107197).

MATERIALS AND METHODS

Materials. 8-oxoguanine was purchased from Toronto Research Chemicals Inc. All other chemicals and buffers were obtained from Sigma-Aldrich unless specified otherwise. Kinetic assays were performed in a 96 well plate with a SPECTRAmax 384 Plus spectrophotometer from Molecular Devices.

Cloning, Expression, and Purification. The unknown amidohydrolase protein NYSGXRC-9236e from an unknown organism in an environmental sample of the Sargasso Sea was cloned, expressed and purified by SGX Pharmacueticals. Pa0142 was cloned, expressed and purified by Dr. Ricardo Marti-Arbona. The *E. coli* proteins GDA (gi: 89109663), SsnA (gi: 89109659), and YahJ (gi: 89107197) were cloned, expressed and purified by Dr. Dao Feng Xiang. Pa 0134 was cloned by Jennifer Cummings.

Metal Content Analysis. Metal determination and quantification was performed with the Elan DRC II ICP-MS (inductively coupled plasma - mass spectrometer) from Perkin Elmer. The analog detection mode is used with three averaged replicates per reading. External calibration standards were prepared through the serial dilution of a single 10 ppm stock mixture of Zn, Cd, Co, Cu, Mn, Ni and Fe in 2% nitric acid, purchased from Inorganic Ventures Inc. Freshly prepared standards generally contained 2, 20 and 200 ppb of all analytes in 1% Trace Select Nitric Acid from Fluka that is diluted in deionized water. The isotopes of the analytes measured were ^{55}Mn, ^{56}Fe, ^{59}Co, ^{60}Ni, ^{63}Cu, ^{66}Zn and ^{111}Cd. ^{115}In was used as an internal standard for Cd and ^{69}Ga was used as an internal standard for all other analytes. The detection of ^{56}Fe was determined using the dynamic reaction cell (DRC) mode, using NH_3 gas to reduce the high level of interference from argon-oxide species. Protein samples were prepared for analysis by boiling in concentrated nitric acid for 30 min, followed by dilution with deionized water to 1% nitric acid.

Activity Screening. A coupled assay was used for the detection of ammonia release in the screening of the library of pteridines, purines and pyrimidines shown in Scheme 6.2. Most compounds were present at concentrations of 1 mM. The pteridine compounds and 8-oxoguanine were used at concentrations of 0.1 mM. This assay was modified from the protocol reported previously (*79*) and typically contained 7.4 mM α-ketoglutarate, 0.4 mM NADH, 6 units of glutamate dehydrogenase in 100 mM HEPES pH 8.5. Mass spectrometry analysis for reaction verification was performed by the staff

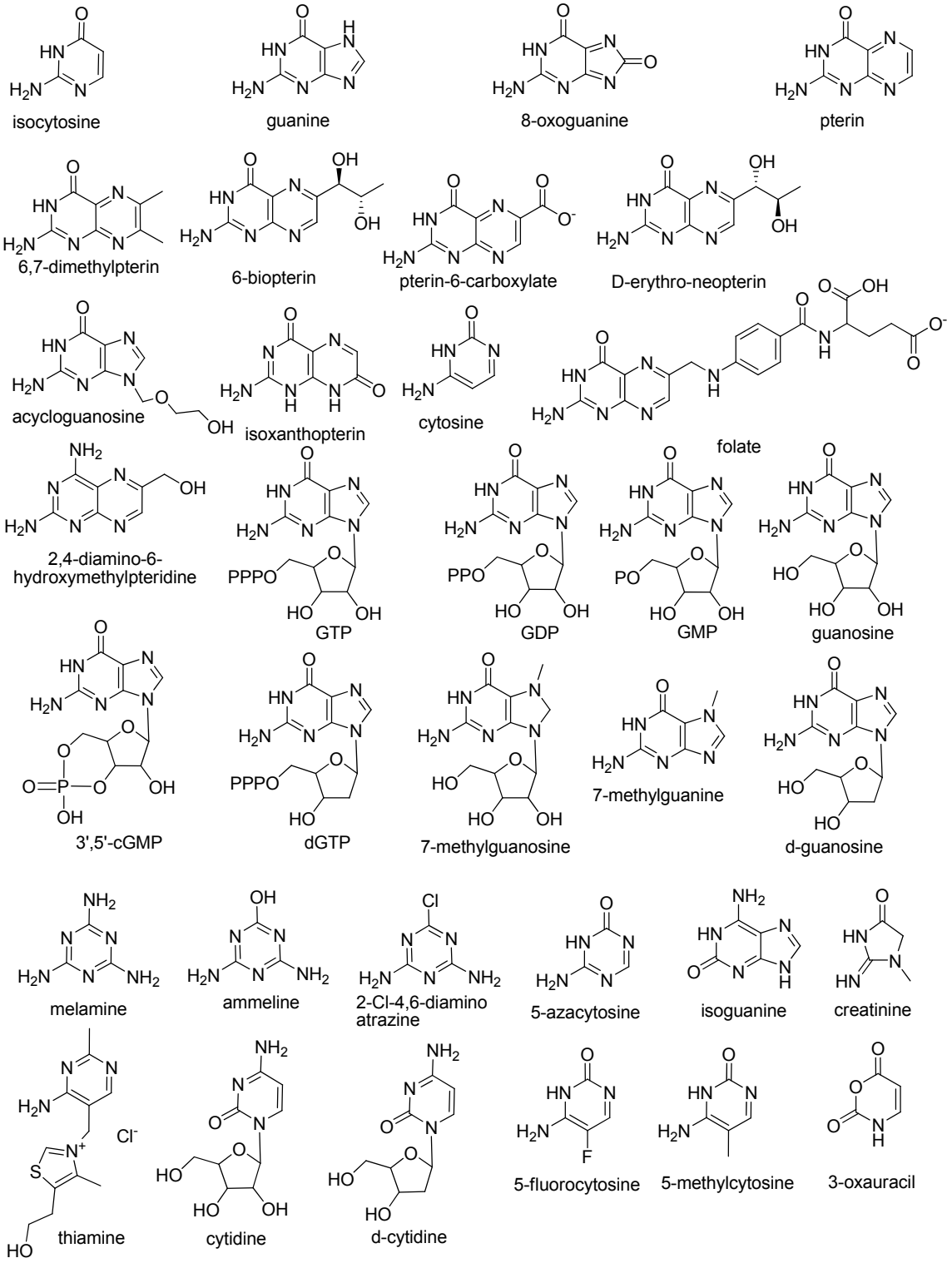
at the Laboratory for Biological Mass Spectrometry (LBMS) in the Department of Chemistry at Texas A&M University.

X-Ray Crystallography. The crystallization, X-ray diffraction, data collection and structure determination and refinement of NYSGXRC-9236e were performed by a group led by Dr. S. Almo of the Department of Biochemistry at the Albert Einstein College of Medicine. This structure was deposited in the protein data bank on 06/03/2009 and assigned the PDB code: 3HPA. The structures illustrated were created using the program WebLab Viewer Pro 4.0.

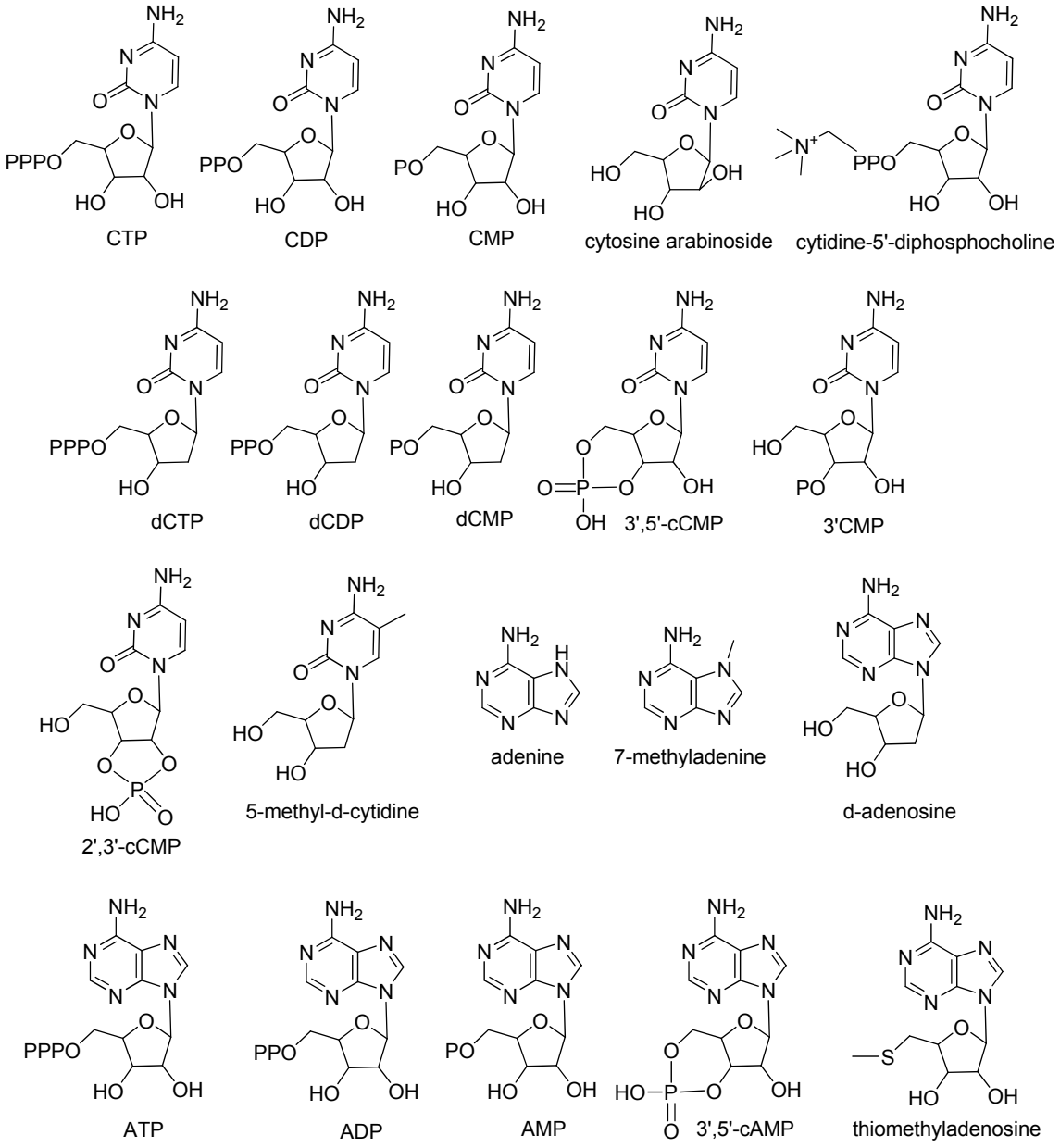
Data Analysis. The steady-state kinetic data were analyzed using Softmax Pro version 4.7.1. For the determination of the kinetic parameters; k_{cat}, K_m and k_{cat}/K_m, saturation curves were fit to equation 2.1 using the non-linear least squares fitting program SigmaPlot 10.0 where A is the substrate concentration, v is the velocity of the reaction, k_{cat} is the turnover number and K_m is the Michaelis constant.

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

Scheme 6.1


isocytosine
guanine
8-oxoguanine
pterin
6,7-dimethylpterin
6-biopterin
pterin-6-carboxylate
D-erythro-neopterin
acycloguanosine
isoxanthopterin
cytosine
folate
2,4-diamino-6-hydroxymethylpteridine
GTP
GDP
GMP
guanosine
3',5'-cGMP
dGTP
7-methylguanosine
7-methylguanine
d-guanosine
melamine
ammeline
2-Cl-4,6-diamino atrazine
5-azacytosine
isoguanine
creatinine
thiamine
cytidine
d-cytidine
5-fluorocytosine
5-methylcytosine
3-oxauracil

Scheme 6.1: Continued.



RESULTS AND DISCUSSION

Metal Content Analysis. The enzymes Pa0134, Pa0142, *E. coli* GDA, *E. coli* SsnA and *E. coli* YahJ were evaluated for metal content by ICP-MS. Due to limited protein availability and the observation of Zn in the crystal structure 3HPA, SGX-9236e was not evaluated for metal content by ICP MS. The Pa0142 preparation assayed in this manuscript was found to possess 0.25 molar equivalents of Zn and 0.05 equivalents of Fe. Pa0134 contains nearly a full metal content with 0.78 equivalents of Zn and 0.09 equivalents of Fe. GDA from *E. coli* contained 1.0 equivalents of Zn^{2+}, SsnA was found with 0.16 and 0.44 equivalents of Zn and Fe, respectively. 0.38 equivalents of Zn and 0.50 equivalents of Fe were detected in each subunit of YahJ. It is expected that each of these enzymes should require 1.0 equivalents of Zn^{2+}, Fe^{2+} or Mn^{2+} for full catalytic activity.

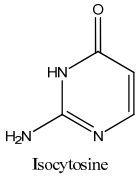
Substrate Specificity The guanine/cytosine deaminase–like enzyme SGX-9236e and Pa0142 were both screened against a library of pteridines, purines and pyrimidines. Pa0142 was previously tested by Dr. Ricardo Marti-Arbona with a smaller library of compounds that lacked 8-oxoguanine. Both enzymes were found to have significant activity with isocytosine (**1**) and guanine (**2**). Additionally, a novel enzymatic activity was discovered. These enzymes were found to have significant 8-oxoguanine (**3**) deaminase activity, 8-oxoguanine (**3**) was the best substrate for these two enzymes by about one order of magnitude, yielding k_{cat}/K_m constants of 2.0×10^4 for Pa0142 and 2.7×10^5 $M^{-1}s^{-1}$ for 9236e. It was not possible to determine the individual kinetic constants k_{cat} and K_m due to a lack of saturation seen under the limited solubility range of 8-

oxoguanine (~140 μM). Substantial activity was also measured for the significantly more soluble isocytosine (**1**) and guanine (**2**) with both enzymes. No enzymatic activity was observed for the deamination of pterin (**4**). Inspection of the active compounds in Scheme 6.2 shows that isocytosine contains all of the minimal requirements for binding and catalysis while 8-oxoguanine possibly contains additional enzyme interacting groups, making it a more effective substrate. Selected results from the activity screening are summarized in Table 6.1.

Pa0142 and SGX-9236e showed some activity against 1 mM ammeline (0.9 s^{-1} and 6 s^{-1}), as well as 5-azacytosine (0.1 s^{-1} and 0.2 s^{-1}). Due to the greater degree of activity seen with the larger compounds, in addition to the genomic context supporting purine salvage, not pesticide metabolism, it is highly likely that the isocytosine, guanine, ammeline and 5-azacytosine activity observed is merely due to enzyme promiscuity resulting from the similarities in substrates. No significant activity (< 0.05 s^{-1}) was observed for any of the other compounds in the screening library with SGX-9236e.

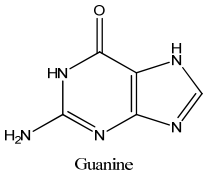
In light of the guanine deaminase activity measured here, the substrate specificity of the guanine deaminase (GDA) from *E. coli* was investigated for comparison. Of the four compounds shown in Scheme 6.2, Pa0134 (probable GDA) and *E. coli* GDA were found to only be active with guanine (**2**). Sequence evaluations of the genome of *E. coli* identified two genes, SsnA and YahJ, which possess similarities to deaminases such as ADE, CDA and GDA. However, no significant activities were found with any of the compounds tested.

Scheme 6.2



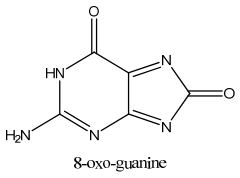
Isocytosine

(1)



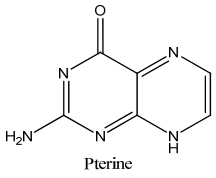
Guanine

(2)



8-oxo-guanine

(3)



Pterine

(4)

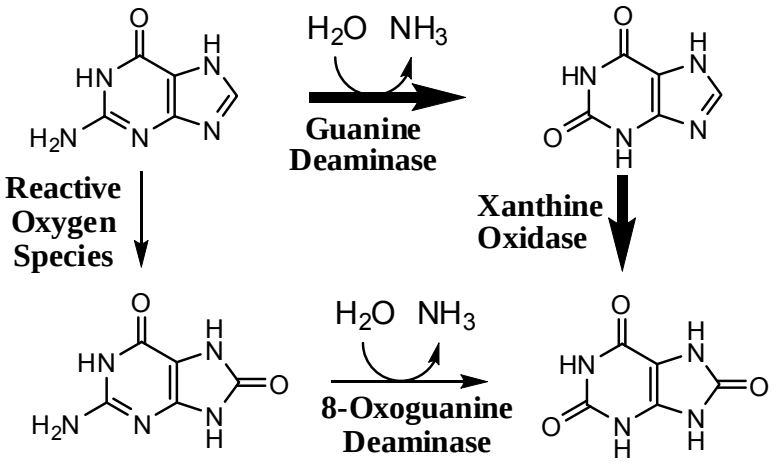
Table 6.1: Kinetic Parameters for Substrates with 8-Oxoguanine Deaminase and Related Enzymes[a].

enzyme	substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
3HPA (SGX)	**1**	62 ± 2	1.6 ± 0.2	$(3.9 \pm 0.5) \times 10^4$
	2	52 ± 3	1.0 ± 0.1	$(5.2 \pm 0.6) \times 10^4$
	3			$(2.7 \pm 0.1) \times 10^5$
	4	< 0.05		
Pa0142	**1**	3.9 ± 0.1	1.8 ± 0.1	$(2.2 \pm 0.1) \times 10^3$
	2	3.3 ± 0.6	2.6 ± 0.6	$(1.3 \pm 0.4) \times 10^3$
	3			$(2.0 \pm 0.1) \times 10^4$
	4	< 0.005		
Pa0134	**1**	< 0.005		
	2	21 ± 2	0.50 ± 0.09	$(4.2 \pm 0.9) \times 10^4$
	3	< 0.005		
	4	< 0.005		
GuaD *E. coli*	**1**	< 0.005		
	2	12.5 ± 0.2	0.154 ± 0.008	$(8.1 \pm 0.4) \times 10^4$
	3	< 0.005		
	4	< 0.005		
SsnA *E. coli*	**1**	< 0.005		
	2	< 0.005		
	3	< 0.005		
	4	< 0.005		
YahJ *E. coli*	**1**	< 0.005		
	2	< 0.005		
	3	< 0.005		
	4	< 0.005		

[a]The kinetic parameters were determined at pH 8.0, 30 ˚C in 75 mM HEPES, using the glutamate dehydrogenase coupled assay for ammonia release.

Novel Deamination of 8-Oxoguanine. In the canonical metabolism of guanine, GDA deaminates guanine, forming xanthine and ammonia. Xanthine oxidase then oxidizes xanthine, forming urate. This discovery of 8-oxoguanine deaminase introduces a never before described branch of purine metabolism. In this new path, guanine becomes oxidized by a reactive oxidizing species or possibly an unidentified enzyme. The resulting 8-oxoguanine is then deaminated by 8-oxoguanine deaminase. This pathway essentially bypasses the formation of xanthine as illustrated in Scheme 6.3. The verification of the 8-oxoguanine deamination by 8-oxoguanine deaminase has been confirmed by: monitoring the release of ammonia through the glutamate dehydrogenase coupled assay, comparing the UV spectra of the 8-oxoguanine, the 8-oxoguanine deamination product of 9236e, and authentic urate in 50 mM TRIS, pH 8.5 as shown in Figures 6.2 and 6.3. The differences in peak heights may be attributed to errors in concentration and background interferences from different impurities and buffers. The formation of urate from 8-oxoguanine was also verified by (-) ESI analysis showing a single, unambiguous peak at m/z = 167, which is consistent with that seen for the urate standard. Analysis of unreacted 8-oxoguanine showed a peak at m/z = 166, which is consistent with an 8-oxoguanine anion.

Scheme 6.3



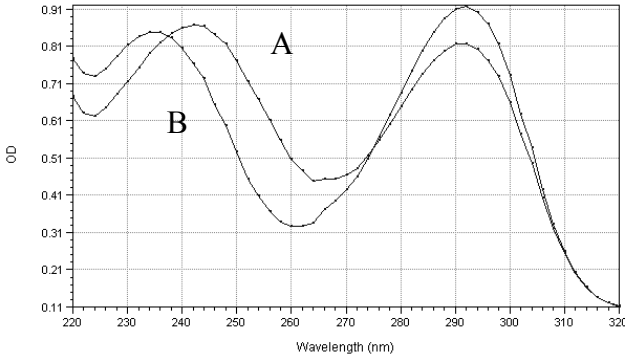


Figure 6.2: Absorbance spectra of 8-oxoguanine (A) and the urate product (B) formed by 9236e 8-oxoguanine deaminase in 50 mM TRIS pH 8.5.

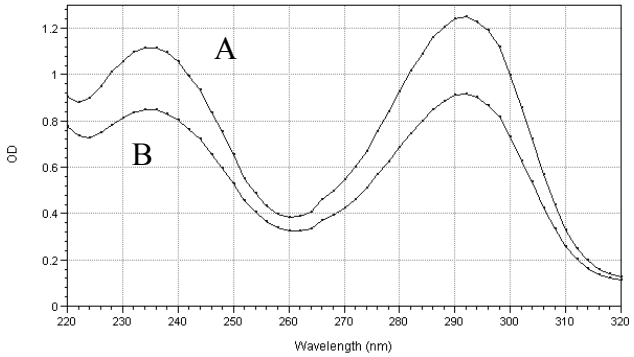


Figure 6.3: Absorbance spectra of authentic urate (A) and the product of 8-oxoguanine deamination (B) formed by 9236e 8-oxoguanine deaminase in 50 mM TRIS pH 8.5.

Crystal Structure of 8-Oxoguanine Deaminase. The NYSGXRC-9236e (8-oxoguanine deaminase) crystal structure 3HPA was deposited to the PDB on June 03, 2009 and was released 13 days later. The crystal structure shows a dimer as presented in Figure 6.4. One Zn atom is bound in each subunit and each subunit forms three distinct domains. The first ~ 80 residues on the N-terminus form a small, mainly beta roll. Starting on the beta strand just before the HxH M_α binding motif, the next 264 residues form the $(\beta/\alpha)_8$ TIM-barrel domain. This domain is common to all amidohydrolase enzymes and contains all of the metal binding residues and known catalytic residues. A ribbon diagram of the functional subunit with displayed active site residues is shown in Figure 6.5. At the C-terminus of the protein, the final 134 residues form an alpha-helix bundle, which participates in dimerization through coiled coil interactions with the adjacent subunit.

The M_α binding site is formed through coordinations by: His-93 (2.1 Å), His-95 (2.1 Å) and His-259 (2.2 Å). Asp-347 lies 2.7 Å away, which is a very long distance for a metal coordinating ligand. The organization of these ligands around the Zn ion forms an extremely distorted tetrahedral coordination. The three histidines (93, 95, and 259) each share 89-92˚ angles with each other through Zn. The His-259-Zn-Asp-347 bond angle is 167˚. A water molecule is located 3.8 Å from the Zn and the water-Zn-His-93 bond angle is 167˚. This water is presumably the hydrolytic water molecule and is H-bonded to His-296 and Asp-347 at distances of 2.8 and 2.9 Å respectively.

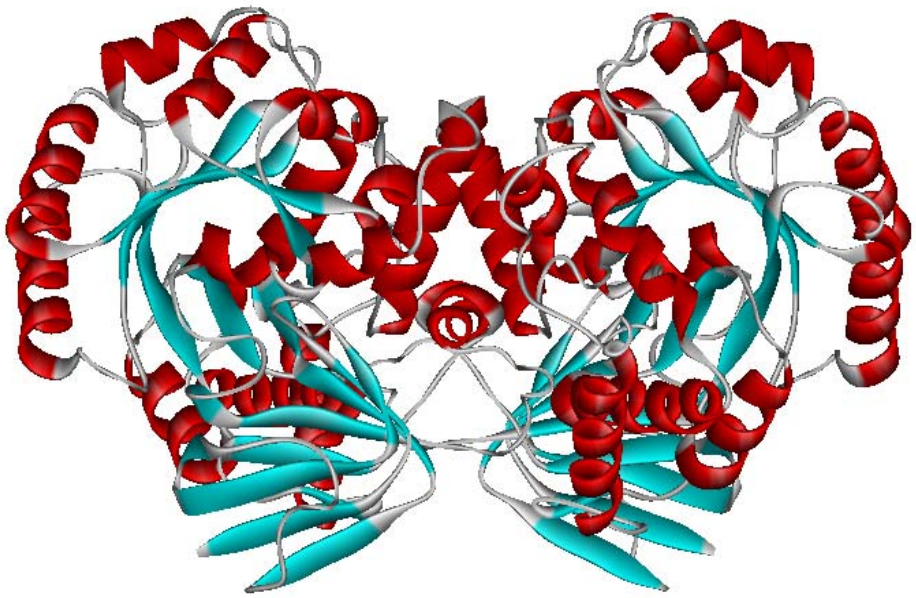


Figure 6.4: Ribbon diagram of the 3HPA (8-oxoguanine deaminase) dimer.

His-296 is 3.9 Å from the Zn ion and 171˚ opposite His-95. The arrangement of the active site residues with respect to the Zn^{2+} ion are presented in Figure 6.6. In the crystal structure of cytosine deaminase with a reported Fe^{3+} (PDB: 1K6W), both the Asp-314 and water are closer to the metal ion (2.6 and 1.9 Å), and a less distorted tetrahedral coordination is observed (*23*).

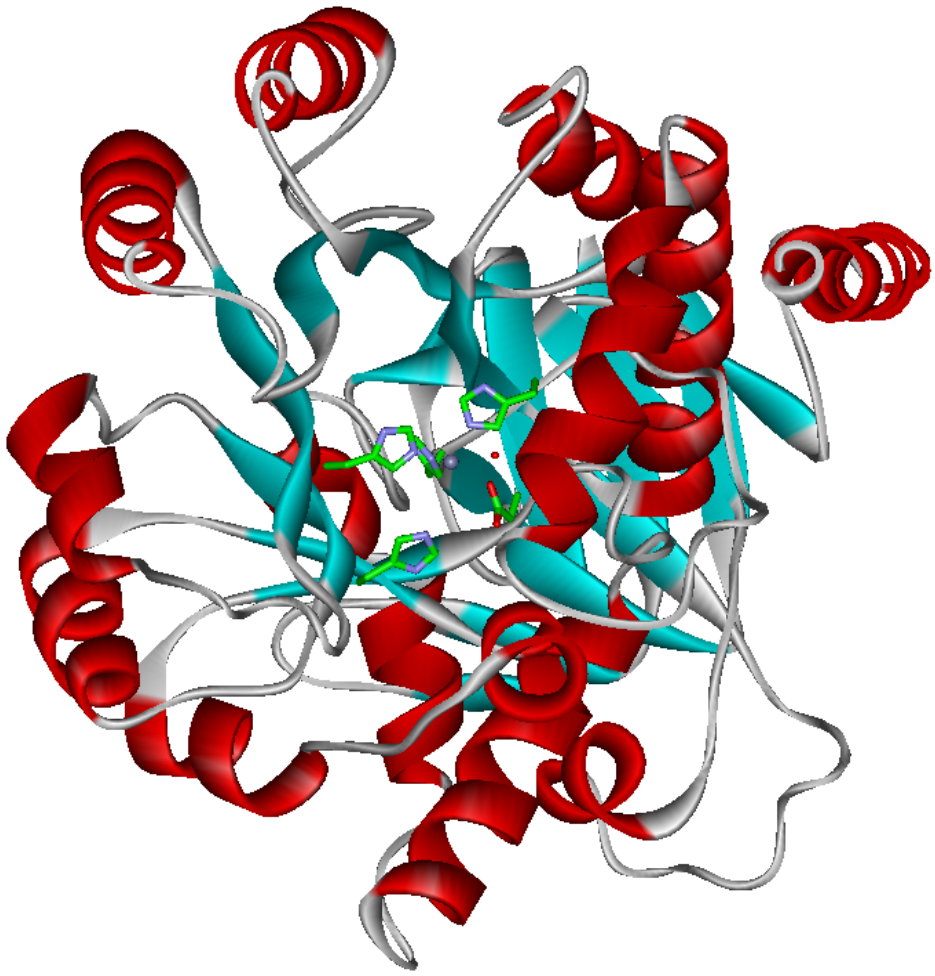


Figure 6.5: Ribbon diagram of the 3HPA subunit displaying the active site residues and the bound Zn atom. Active site residues are shown with green carbon atoms. The bound Zn atom is shown as a gray sphere and the bound water molecule is shown as a small red sphere.

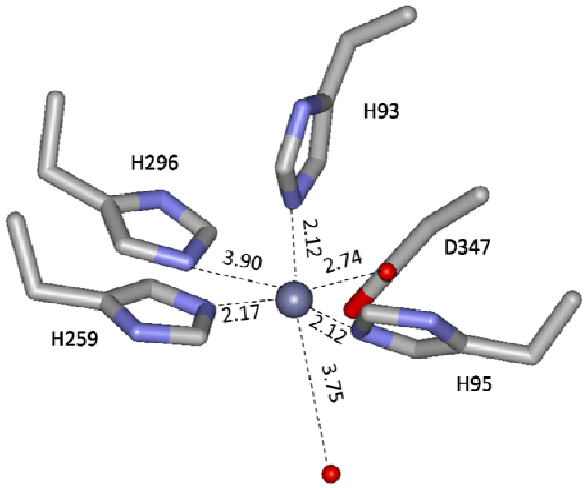


Figure 6.6: Active site of 3HPA showing interactions. Distances are presented in Å.

Sequence Analysis. Evaluation of the existing crystal structure of human GDA with bound xanthine (PDB code: 2UZ9) and the sequences of other guanine deaminases enabled the identification of two key residues which should provide substrate specificity toward guanine. Sequence comparisons of the enzymes 9236e and Pa0142 with several guanine deaminases show that these similar enzymes differ at these key locations. The presence of these differences should be an indication of a different function.

Arg-235 and Phe-236 appear to be providing substrate binding interactions and control over substrate specificity for guanine deaminase. An illustration of these interactions is shown below in Figure 6.7. Arg-235 is shown to be participating in two hydrogen bonds with the xanthine product. Phe-236 appears to be providing hydrophobic interactions for C-8 of xanthine. This residue undoubtedly improves substrate fidelity through the steric blocking of larger compounds like pterins. Arg-235

and Phe-236 are conserved throughout guanine deaminases from organisms as diverse as *E. coli*, *H. sapiens and P. aeruginosa*.

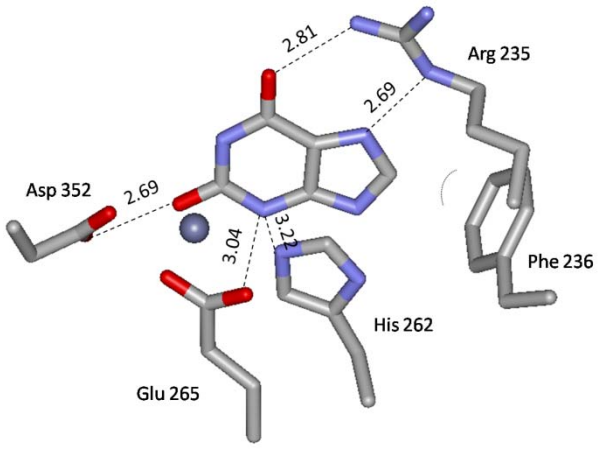


Figure 6.7: Human GDA-xanthine interactions from crystal structure PDB 2UZ9. Arg-235 is participating in multiple hydrogen bonds with the xanthine product. Phe-236 is interacting with C-8 of xanthine through hydrophobic interactions.

Sequence alignments of proteins related to Pa0142 and 9236e as shown in Figure 6.8 were used to verify that this key conserved Arg-Phe motif is missing in the group of enzymes related to Pa0142 and 9236e. Instead a Cys and a Ser have replaced the Arg and Phe residues. This substitution is found in 216 bacterial genes which all share 35 % sequence identity or greater to Pa0142. A multiple sequence alignment was constructed for the sequences of Pa0142 and SGX-9236e with GDA from *P. aeruginosa*, *H. sapiens* and *E. coli* as well as the enzymes of unknown function, SsnA and YahJ. A dendrogram of the alignment is shown in Figure 6.9. The TIM barrel β-strands identified from the structure of 3HPA are highlighted in gray. Common deaminase active site residues are highlighted in red. Conserved guanine binding residues are highlighted in green and the

proposed 8-oxoguanine deaminase specific residues are highlighted in blue. It is predicted that these enzymes containing the Cys-Ser motif at the end of β-strand 4 should demonstrate 8-oxoguanine deaminase activity. The observation of this key active site difference provides support that Pa0142 and 9236e function as 8-oxoguanine deaminases and not guanine deaminases.

```
9236e    ----mgadrrgermnleqhagarapntsssrpktllvkhadvlvtmddtr  46
Pa0142   ----------------------------------msrtwirnplaiftan  16
Pa0134   mndavpagnfvpplilhsdhggsamsssahrgrilhflgdpaklg-dkaw  49
Pa1521   ----------------------mtrnakayraailhsiadpaevgversy  28
Ec_GDA   ------------------mmsgehtlkavrgsfidvtrtidnpeeiasal  32
Hs_GDA   -----------------mcaaqmpplahifrgtfvhstwtcpmevlrdhl  33
Ec_SsnA  ------------------------------------mlilknvtavqlhp  14
Ec_YahJ  -------------------------mkitdphyyldnvlletgfdyengv  25

9236e    relrdaglyiednrivavg-psaelpeta--devldlrghlvipglvnth  93
Pa0142   gldaagglvvedgrivellgagqqpaqpc--asqfdasrhvvlpglvnth  64
Pa0134   eyfedgllwiehghvraldhatyllpqlpadlpleehpqrlllpgfvdch  99
Pa1521   eyfedglllvedgkvarlgdaetllgeig-evevfeyrdalitpgfidth  77
Ec_GDA   rfiedglllikqgkvewfgewengkhqipdtirvrdyrgklivpgfvdth  82
Hs_GDA   lgvsdsgkivfleeasqqeklakewcfkp-ceirelshheffmpglvdth  82
Ec_SsnA  akvqegvdiaiendvivaigdaltqrypd--asfkemhgrivmpgivcsh  62
Ec_YahJ  avqtrtarqtveiqdgkivalrenklhpdatlphydaggklmlpttrdmh  75

9236e    hhmyqsltravpaaqn--aelfgwltnl-ykiwahltpemievstltama  140
Pa0142   hhfyqtltrawapvvn--qplfpwlktl-ypvwarltpeklelatkvala  111
Pa0134   vhypq---lgviasyg--tqlldwlethtfpaeqrfadagyaaaqaelfl  144
Pa1521   ihfpq---tgmiasyg--eqlldwlntytfpterqfgdqahadqvaeifl  122
Ec_GDA   ihypq---semvgayg--eqllewlnkhtfpterryedleyaremsaffi  127
Hs_GDA   ihasqys--fagssid--lpllewltkytfpaehrfqnidfaeevytrvv  128
Ec_SsnA  nhfysglsrgimaniapcpdfistlknlwwrldraldeeslyysglicsl  112
Ec_YahJ  ihldktfyggpwrsln-rpagttiqdmikleqkmlpelqpytqeraekli  124
I
9236e    ellqsgcttssdhlyiypn---gsrlddsigaaqrigmrfhasrg--ams  185
Pa0142   elllsgcttaadhhylfpgg-leqaidvqagvveelgmramltrg--sms  158
Pa0134   dellrhgtttalvfgtvha----vsaeaffqaaqkrrlrmiagkv-----  185
Pa1521   qellrngtttalvfgsvhr----qsveslfeaarrldlrliagkv-----  163
Ec_GDA   kqllrngtttalvfgtvhp----qsvdalfeaashinmrmiagkv-----  168
Hs_GDA   rrtlkngtttacyfatiht----dssllladitdkfgqrafvgkvc-----  170
Ec_SsnA  eaiksgctsvidhhaspayig--gslstlrdaflkvglramtcfet-----  156
Ec_YahJ  dllqskgttiarshcniep---------vsglknlqnlqavlarr-----  160

9236e    vgqrdgglppdsvverepdilrdtqrlietyhdegryamlrvvvapcsp-  234
Pa0142   lgekdgglppqqtvqeaetiladserliaryhqrgdgarvqialapcsp-  207
Pa0134   ---lmdrnappalcdtaasgyaesralierwhgng---rlqyavtprfap  229
Pa1521   ---mmdrnapdyltdtaessyrdskalierwhgqg---rllyavtprfap  207
Ec_GDA   ---mmdrnapdylldtaessyhqskelierwhkng---rllyaitprfap  212
Hs_GDA   ---mdlndtfpeyketteesiketerfvsemlqkn-ysrvkpivtprfs-  215
Ec_SsnA  ---tdrnngikelqegveenir-farlideakkatsepylveahigahap  202
Ec_YahJ  ---qagfeceivafpqhglllskseplmreamqag---ahyvggldptsv  204
```

Figure 6.8: Multiple sequence alignment of the 8-oxoguanine deaminases (9236e and Pa0142) with guanine deaminases (Pa0134, Pa1521 , Ec_GDA and Hs_GDA) and the proteins of unknown function SsnA and YahJ.

```
9236e   fsvsrdlmrdaavlareygvslhthlaenvndiaysrekfg--mtpaeya 282
Pa0142  fsvtpeimrasaevaarhdvrlhthlaetldeedfclqrfg--lrtvdyl 255
Pa0134  tsspeqlaaaarlldeypgvylhthlsenlkevawvgelfpqaqdyldvy 279
Pa1521  tstaeqldmaarllrehpgvylhthlsenlkeiewvkelfpersgyldvy 257
Ec_GDA  tsspeqmamaqrlkeeypdtwvhthlcenkdeiawvkslypdhdgyldvy 262
Hs_GDA  lscsetlmgelgniaktrdlhiqshisenrdeveavknlypsyknytsvy 265
Ec_SsnA ftvpdaglemlreavkatgrglhihaaedlydvsyshhwyg--kdllarl 250
Ec_YahJ dgameksldtmfqialdydkgvdihlhettpagvaainymvetvektpql 254

9236e   edlgwvghdvwhahcvqlddagiglfartgtgvahcpcsnmrlasgiapv 332
Pa0142  dsvgwlgprtwlahgihfnaeeirrlgeagtgichcpssnmrlasgicpt 305
Pa0134  hraglvgersvfahgihlserecrclahknaalahcpssnlfigsglfdl 329
Pa1521  dhhgllgprsvfahgvhlcdgecqrlaetgsavafcptsnlflgsglfdl 307
Ec_GDA  hqygltgkncvfahcvhleekewdrlsetkssiafcptsnlylgsglfnl 312
Hs_GDA  dknnlltnktvmahgcylsaeelnvfhergasiahcpnsnlslssgflnv 315
Ec_SsnA aqfdlidsktlvahglylskdditllnqrdaflvhnarsnmnnhvgynhh 300
Ec_YahJ kgkltishafalat---lneqqvdelanrmvvqqisiastvpigtlhmpl 301

9236e   kkmrlagvpvglgvdgsasndgaqmvaevrqalllqrvgfgpda---mta 379
Pa0142  veleaagapiglgvdgsasndasnmilearqalylqrlrygaer---itp 352
Pa0134  graqqygirvgigsdvggg---tslsllanladaykiqqlrgts---ldp 373
Pa1521  pklerykvkvglgtdvgag---tsfsqlqslneaykvmqlqgar---ldp 351
Ec_GDA  kkawqkkvkvgmgtdigag---ttfnmlqtlneaykvlqlqgyr---lsa 356
Hs_GDA  levlkhevkiglgtdvaggysysmldairravmvsnillinkvneksltl 365
Ec_SsnA lsdirn---lalgtddgigs----dmfeemkfaffkhrdaggplwp----d 339
Ec_YahJ kqlhdkgvkvmtgtdsvidhwspyglgdmlekanlyaqlyirpn------ 345

9236e   realeiatlggakvlnrdd-igalkpgmaadfaafdlrqplfagalhdpv 428
Pa0142  elalgwatrgsarllgrsd-igelapgkqadlalfkldelrfs-gshdpl 400
Pa0134  fqalylatlggaraldldglvgnflpgreadfvaldl--------aatpm 415
Pa1521  fkslylatlggaralelddrigsfatsneadfvvldy--------hatpl 393
Ec_GDA  yeafylatlggakslglddlignflpgkeadfvvmep--------tatpl 398
Hs_GDA  kevfrlatlggsqalgldgeignfevgkefdailinp------kasdspi 409
Ec_SsnA sfakaltngnelmsrnfgakfglleagykadlticdynsptplladniag 389
Ec_YahJ -----eqnlsrslflatgdvlplnekgervwpkaqdd----------as 379

9236e   aalvfcapsqtaytvvngkvvv-regrlatldlppvierhnalahalvea 477
Pa0142  salllcaadradrvmvggawrv-vdgavegldlaaliarhraaasaliag 449
Pa0134  iaqrmehargladtlfvlntlg-ddravaetwvmgerrhvkg-------- 456
Pa1521  lsyrlsqagslaerlfaltilg-ddrtvketfaagrsvhrrd-------- 434
Ec_GDA  qqlrydnsvslvdklfvmmtlg-ddrsiyrtyvdgrlvyern-------- 439
Hs_GDA  dlfygdffgdiseaviqkflylgddrnieevyvggkqvvpfsssv----- 454
Ec_SsnA hiafgmgsgsvhsvmvngvmvyedrqfnfdcdsiyaqarkaaasmwrrmd 439
Ec_YahJ fvlvdascsaeavarisprtatfhkgqlvwgsvag--------------- 414

9236e   ar-                                                 479
Pa0142  ---                                                 449
Pa0134  ---                                                 456
Pa1521  ---                                                 434
Ec_GDA  ---                                                 439
Hs_GDA  ---                                                 454
Ec_SsnA ala                                                 442
Ec_YahJ ---                                                 414
```

Figure 6.8: Continued.

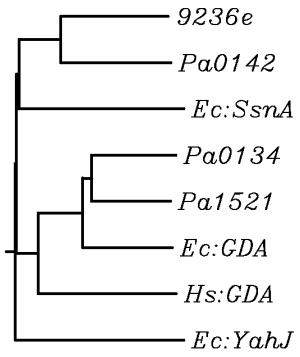


Figure 6.9: Relationship dendrogram obtained from the sequence alignments of 8-oxoguanine deaminases, guanine deaminases and the proteins of unknown function, SsnA and YahJ.

The active sites of guanine deaminase (PDB: 2UZ9) and 9236e (PDB: 3HPA) have been super-imposed by tethering the metal coordinating N^{ε} atoms and Zn atoms between the two structures. A view of this alignment is shown in Figure 6.10. The active site residues and xanthine molecule corresponding to 2UZ9 are shown with gray carbons and the residues corresponding to 3HPA are shown with green carbons. This structural approximation indicates that Cys-232 is reasonably positioned for recognition of the 8-oxo group of the substrate 8-oxoguanine.

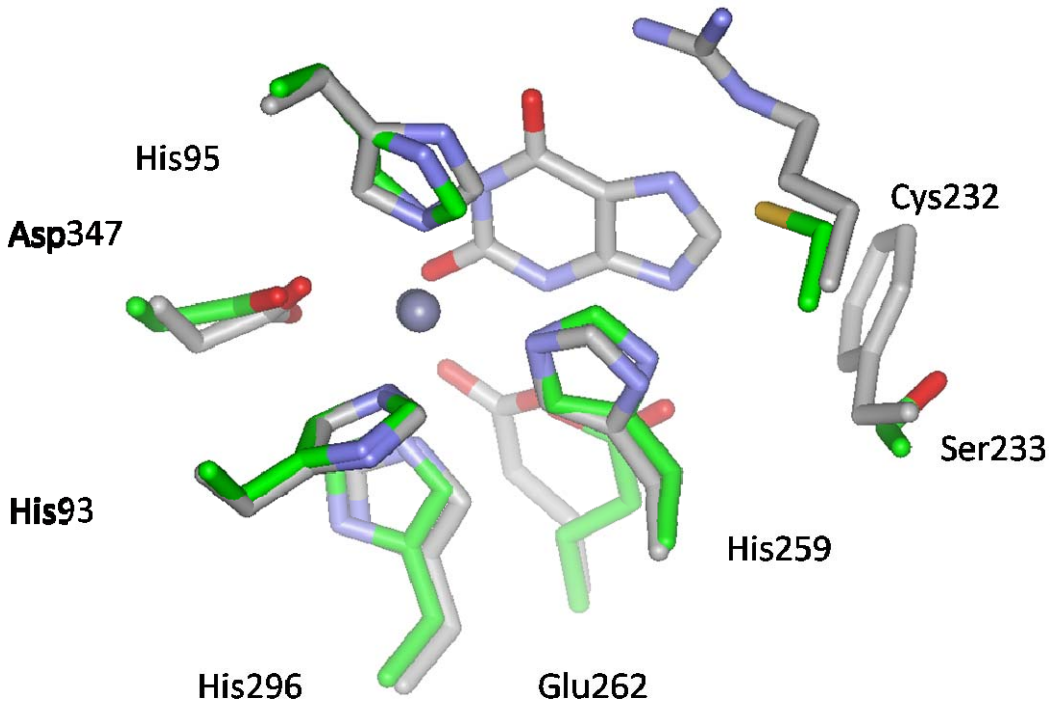


Figure 6.10: Active site overlay of 3HPA and 2UZ9 with bound xanthine. Gray carbons correspond to the guanine deaminase crystal structure. Green carbons correspond to the 8-oxoguanine deaminase crystal structure.

Conclusions. Pa0142 and 9236e were found to be active with isocytosine, guanine and 8-oxoguanine. 8-oxoguanine was the best substrate for these two enzymes by an order of magnitude. Anything larger, such as pterin, yielded very poor activity. The genuine guanine deaminases tested possessed activity only with guanine. To the best of our knowledge, this is the first description of any enzyme involved in the deamination of 8-oxoguanine.

Sequence alignments of proteins related to Pa0142 and 9236e were used to identify two key conserved residues (Cys-232 and Ser-234 from 9236e) found in 217 predominantly proteobacterial genes down to 35% identity, which should possess 8-

oxoguanine deaminase activity. The complete list of these enzymes is presented in Appendix A. This enzymatic function which is unique to proteobacteria and thermotogae, demonstrates the diverse catalytic repertoire utilized by bacteria. The closest homolog of 8-oxoguanine deaminase in *E. coli* is the protein of unknown function SsnA.

It is interesting to note that the gene for 8-oxoguanine deaminase occurs predominantly in the microorganisms identified as α, β and γ-proteobacteria. Pathogenic microorganisms like many of these carrying an 8-oxoguanine deaminase may be especially prone to oxidative stress as a result of the environment inside the host in which they may live. It has been shown that all antibodies can efficiently kill bacteria through the catalysis of the addition of singlet oxygen to water, forming the transiently stable oxidant H_2O_3 in the process (*135*). This oxidant has been shown to be an effective reactive oxygen species that is capable of forming 8-oxoguanine (*136*). It appears that these bacteria utilize a novel pathway in the metabolism of 8-oxoguanine, converting it into uric acid and ammonia, which can be further metabolized via formation of allantoin upon oxidation with uricase.

CHAPTER VII

DISCOVERY OF ISOXANTHOPTERIN DEAMINASE

Pteridines, such as tetrahydrofolate (THF) and tetrahydrobiopterin, are vital cofactors for various cellular functions such as single carbon transfers and activation of oxygen. Folates, required in the diet of animals, can be synthesized by bacteria, fungi and plants (*137*). It is known that folates are easily decomposed to pterin-6-carboxylate and *para*-amino-benzoic acid glutamate upon exposure to ultraviolet radiation (*138*). In the shallow sea water of coral reef, significant concentrations of the decomposition products of folates and pteridines have been observed. It has also been determined that 6-biopterin is extremely sensitive to solar radiation, producing pterin-6-carboxylate as the main product and further radiation of pterin-6-carboxylate has resulted in the observation of pterin (*139*).

In 1949 it was determined that pterin is readily oxidized to isoxanthopterin by a xanthine oxidase found in cream (*140*). The presence of an isoxanthopterin deaminase in the proteobacteria *Alcaligenes faecalis* was first observed by Walter McNutt in 1963. Cells of this organism were found to utilize isoxanthopterin as a sole source of carbon and nitrogen. McNutt was able to monitor the disappearance of isoxanthopterin and concomitant appearance of 7-oxylumazine in the cellular extracts (*141*). Another isoxanthopterin deaminase was observed in extracts of adult and larval silkworms (*142*). Recent estimates suggest that folate in plants can break down at 10% per day and evidence is growing for a folate salvage pathway in plants (*137*). The identification of the details of the complex metabolic pathways involving the salvage pathways of

pteridines have remained elusive, partly due to the very small amounts that are found in natural sources (*143*).

This study describes the discovery of function of two bacterial enzymes involved in pteridine deamination that are related in sequence to guanine deaminase and the newly identified 8-oxoguanine deaminase. The two enzymes of this study are the New York SGX Research Center for Structural Genomic protein targets: NYSGXRC-9236b and 9339a. The X-ray crystal structure of 9339a has been solved and assigned the PDB code: 2PAJ. These two proteins were isolated from environmental samples of the Sargasso Sea under the Global Ocean Sampling Project. The closest homolog of 9339a (88% identical) is Bxe_A2016 from *Burkholderia xenovorans LB400* and is annotated as an ATZ/TRZ-like amidohydrolase. The closest homolog of 9236b (95% identical) is Bphyt_7136 from *Burkholderia phytofirmans PsJN.* This protein is also annotated as an ATZ/TRZ-like amidohydrolase.

The sequence-based chlorohydrolase annotation of these two genes is misleading, because the results of a BLAST search of the NCBI protein database indicates that these genes show high similarity to various deaminases such as 8-oxoguanine deaminase, guanine deaminase and SAH deaminase. The presence of various transport proteins such as xanthine/uracil permease found in genomic context evaluations like that shown in Figure 7.1 can be used to implicate the involvement of these unknown genes in some sort of salvage pathway for aromatic heterocycles. The oxidases nearby could also be participating in the catabolism of such compounds. These observations led us to the hypothesis that these “chlorohydrolases” were somehow

involved in the deamination of aromatic compounds. Subsequent screening of these enzymes against select aromatic compounds containing primary amines enabled the identification and characterization of the genes responsible for deaminating isoxanthopterin and pterin-6-carboxylate.

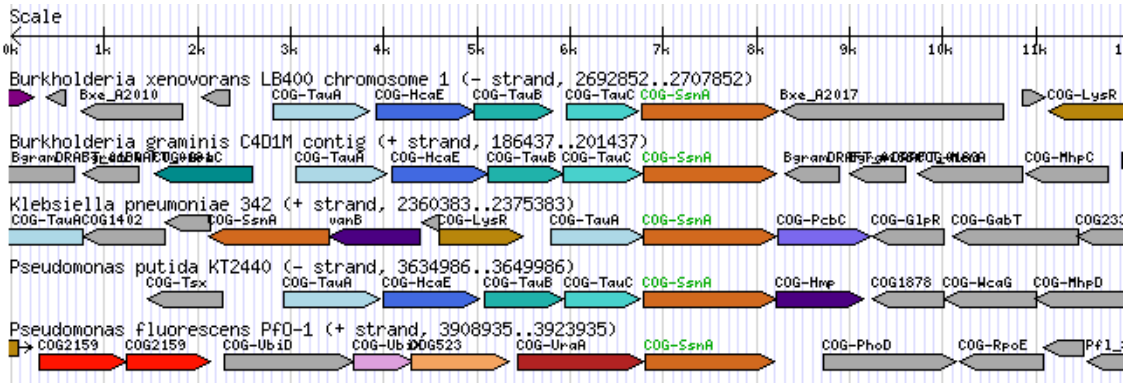


Figure 7.1: Genomic context tree of NYSGXRC-9339a (orange, COG-SsnA) obtained from microbesonline.org. Nearby are decarboxylase-like proteins (gray, COG-UbiD) and (lavender, UbiX), possible oxidoreductase/dioxygenases proteins (blue, COG-PcbC), (dark blue,COG-FabG), (indigo,COG-HhmP, COG-VanB) and (light blue,COG-HcaE), possible hydratases (red, COG2159), a deaminase like protein (orange, COG-SsnA), xanthine/uracil permease (dark red COG-UraA), and the nitrate/sulfonate/carbonate transporter subunits (pastel blue, COG-TauA; teal, TauB and sky,TauC).

MATERIALS AND METHODS

Materials. All chemicals and buffers were obtained from Sigma-Aldrich unless specified otherwise. 8-Oxoguanine was purchased from Toronto Research Chemicals Inc. Kinetic assays were performed in a 96-well plate with a SPECTRAmax 384 Plus spectrophotometer from Molecular Devices.

Cloning, Expression, Purification and X-ray Crystallography. The environmental sample proteins SGX 9339a (gi: 44585104) PDB: 2PAJ; and SGX 9236b (gi: 44611670), were cloned, expressed and purified by NYSGXRC and assigned the target

labels SGX-9339a and SGX-9236b. The crystallization, X-ray diffraction, data collection, structure determination and refinement of SGX-9339a were perfomed by a group led by Dr. S. Swaminathan of the Biology Department at Brookhaven National Laboratory. This structure was assigned the PDB code 2PAJ.

Measurement of Enzymatic Activity. Initial screening assays were performed in a 96-well UV/Vis microplate at 30 °C, using a coupled assay for the detection of ammonia release against a library of pteridines, purines and pyrimidines. Most compounds were utilized at a concentration of 1.0 mM, but the pteridine compounds and 8-oxoguanine were used at a concentration of 0.1 mM. The coupling system was modified from the protocol reported previously (*79*) and typically contained 7.4 mM α-ketoglutarate, 0.4 mM NADH, 6 units of glutamate dehydrogenase in 100mM Hepes, pH 8.5. For direct assays, average extinction coefficients for the deamination of isoxanthopterin ($\Delta\varepsilon_{350\ nm}$ = 7654 M^{-1} cm^{-1}) and pterin-6-carboxylate ($\Delta\varepsilon_{264\ nm}$ = 8080 M^{-1} cm^{-1}) were determined from slopes of the difference in absorbance, measured before and after total enzymatic deamination, over a range of substrate concentrations from 0.004-0.116 mM. Mass spectrometry analysis for reaction verification was performed by the staff at the Laboratory for Biological Mass Spectrometry (LBMS) in the Department of Chemistry at Texas A&M University.

Data Analysis. Sequence alignments were created using the ClustalW program at http://align.genome.jp/sit-bin/clustalw. Alignments were formated using the Multiple Align Show program at http://www.bioinformatics.org/sms/. Steady-state kinetic data were analyzed using the program Softmax Pro version 4.7.1. For the determination of

the kinetic parameters: k_{cat}, K_m and k_{cat}/K_m; saturation curves were fit with equation 2.1 using the non-linear least squares fitting program SigmaPlot 10.0 where *A* is the substrate concentration, *v* is the velocity of the reaction, k_{cat} is the turnover number and K_m is the Michaelis constant.

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

RESULTS

Crystal Structure. The X-ray structure for 9339a was solved to a resolution of 2.7 Å and posted to the PDB on March 27, 2007. The structure of 9339a was solved as a monomer and assigned the pdb code 2PAJ. A ribbon diagram of 2PAJ is shown in Figure 7.2. The alpha helices are displayed in red and the beta strands are shown in teal blue. The bound zinc cation is represented by the gray sphere and the bound water molecule is presented as a smaller red sphere. The active site residues are displayed with green carbons, blue nitrogens and red oxygen atoms.

2PAJ possesses 3 distinct domains. At the N-terminus, ~70 residues form a small predominantly beta structure that is homologous to the N-terminal domain found in the urease subunit c (*144*). The next ~276 residues form a $(\beta/\alpha)_8$ TIM-barrel domain containing much of the same catalytic residues and metal binding ligands that have been shown to be important for cytosine deaminase (*23*), and adenosine deaminase (*17*). This domain of 2PAJ contains one zinc atom in the active site which is coordinated by His-76 (2.29 Å), His-78 (2.05 Å), His-248 (2.05 Å) and Asp-336 (2.03 Å), in a distorted tetrahedral coordination scheme. A water molecule is found nearby at 2.66 Å from the Zn, as well as His-285 which is 2.85 Å away. The active site interactions and distances

are shown in Figure 7.3. Structural data for a section of the protein containing Glu-251 (a landmark residue for deaminases) to Tyr-263 is missing. At the C-terminus is a mainly alpha helix domain of the up-down architecture that is composed of the final 156 residues including the terminal $(His)_6$ purification tag.

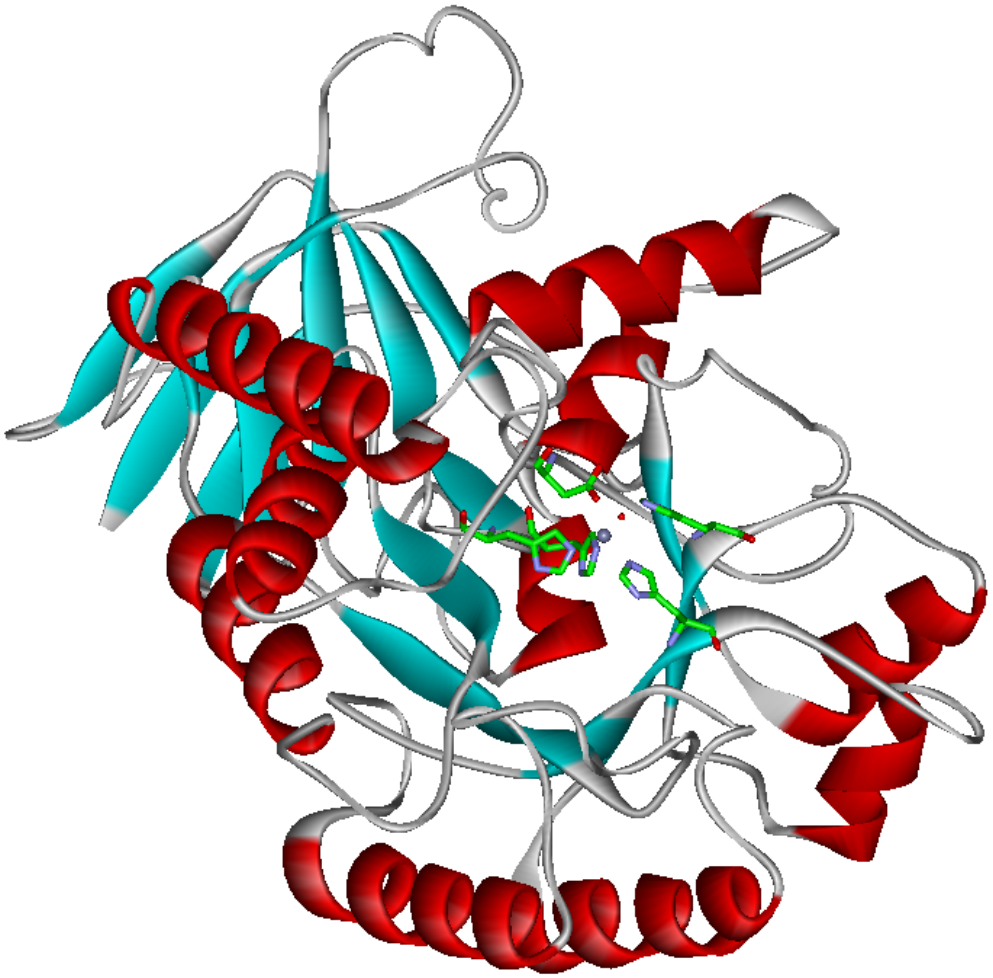


Figure 7.2: Ribbon diagram of the 2PAJ monomer displaying the active site residues and the bound Zn atom. Active site residues are shown with green carbon atoms. The bound Zn atom is shown as a gray sphere and the bound water molecule is shown as a small red sphere.

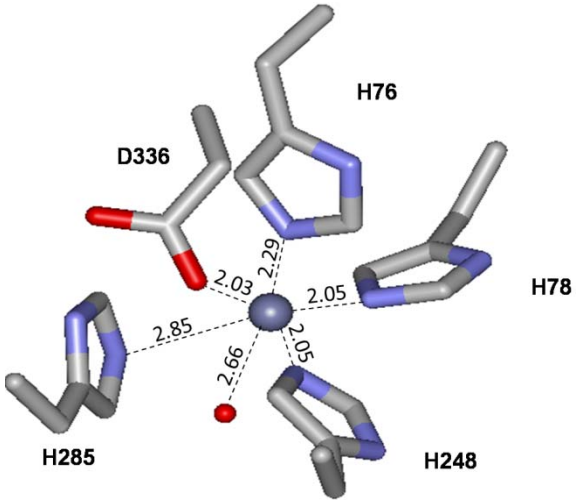


Figure7.3: Active site interactions of 2PAJ. Distances are listed in (Å).

A structural alignment of the active sites of human guanine deaminase (PDB: 2UZ9) and 9339a (PDB: 2PAJ), was constructed by tethering the metal atoms and the homologous metal-coordinating nitrogen atoms of the three histidines between the two structures and superimposing the two molecules, as shown in Figure 7.4. Based on this active site alignment, it can be seen that in 2PAJ, the Arg-Phe dyad, implicated in substrate binding of guanine deaminases, has been replaced for a novel Thr-Thr motif. The water molecule present in 2PAJ is found to occupy a position near the carbonyl carbon of the product xanthine. There are noticeable differences in the positioning of the histidines from β-strand 6 and the aspartates from strand 8.

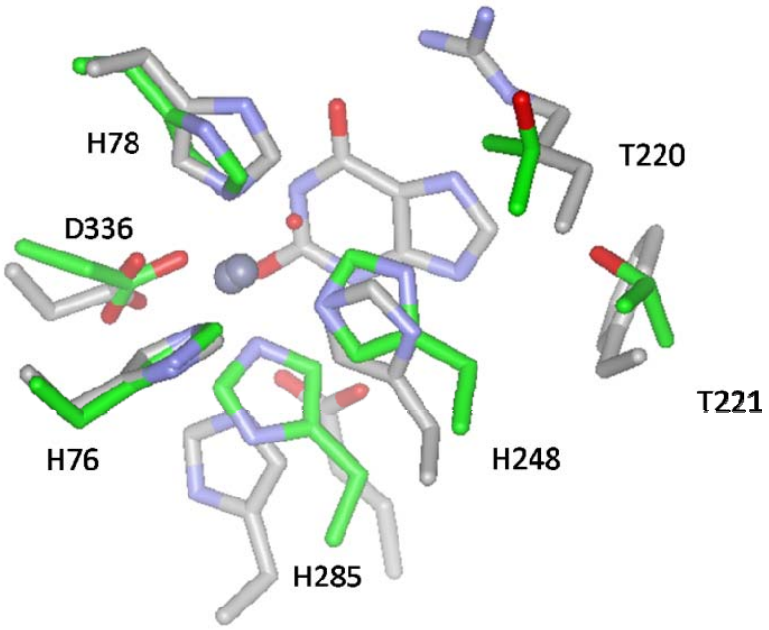
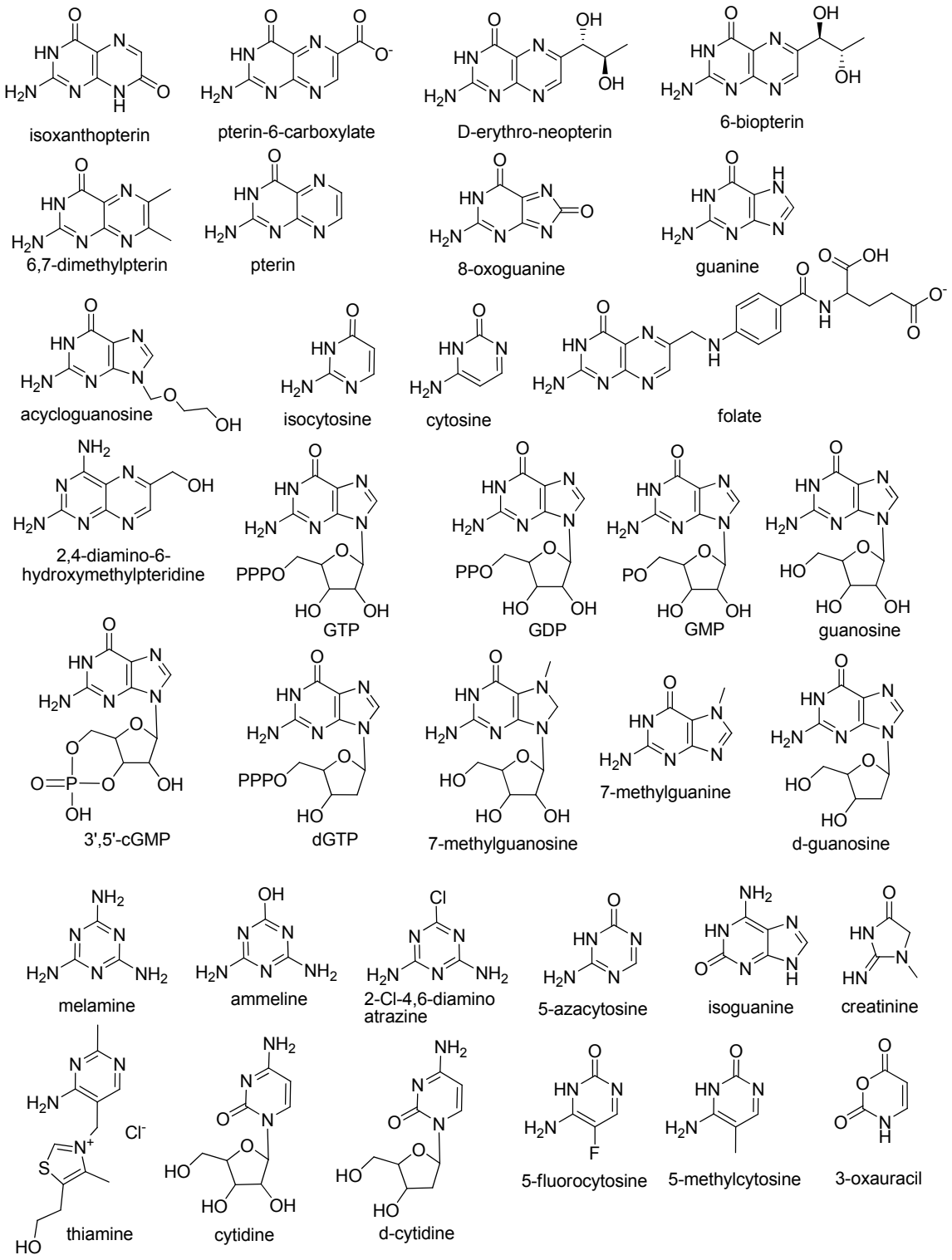


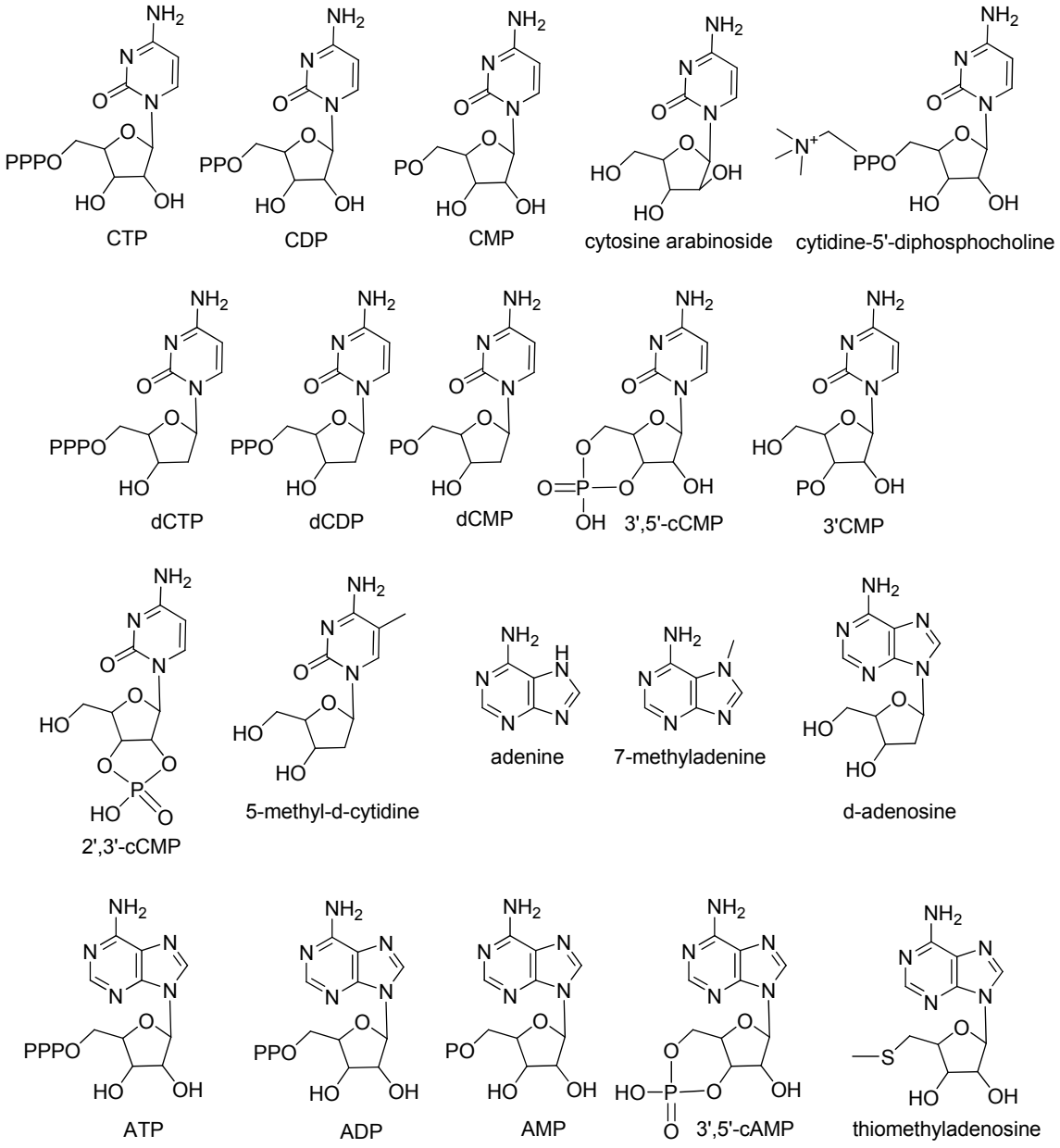
Figure 7.4: Active site overlay of 2PAJ (green carbons) and 2UZ9 with bound xanthine (gray carbons). Residue numbering corresponds to 2PAJ.

Substrate Specificity. SGX 9339a (gi: 44585104) with the X-ray crystal structure (PDB: 2PAJ), and SGX 9236b (gi: 44611670) were tested for ammonia release against the library of pteridines, purines and pyrimidines shown in Scheme 7.1. From those initial activity screens with 5 μg/mL (~100 nM) of deaminase in each assay, significant ammonia release was detected for both 9339a and 9236b with isoxanthopterin and pterin-6-carboxylate as the substrates. The reactions catalyzed by these enzymes are shown in Scheme 7.2. Significant rates of ammonia release were not detected for any of the other compounds. With the enzyme concentration present in these initial assay conditions, activity was determined to be significant if the relative rate of ammonia release was greater than 0.1 s^{-1}.

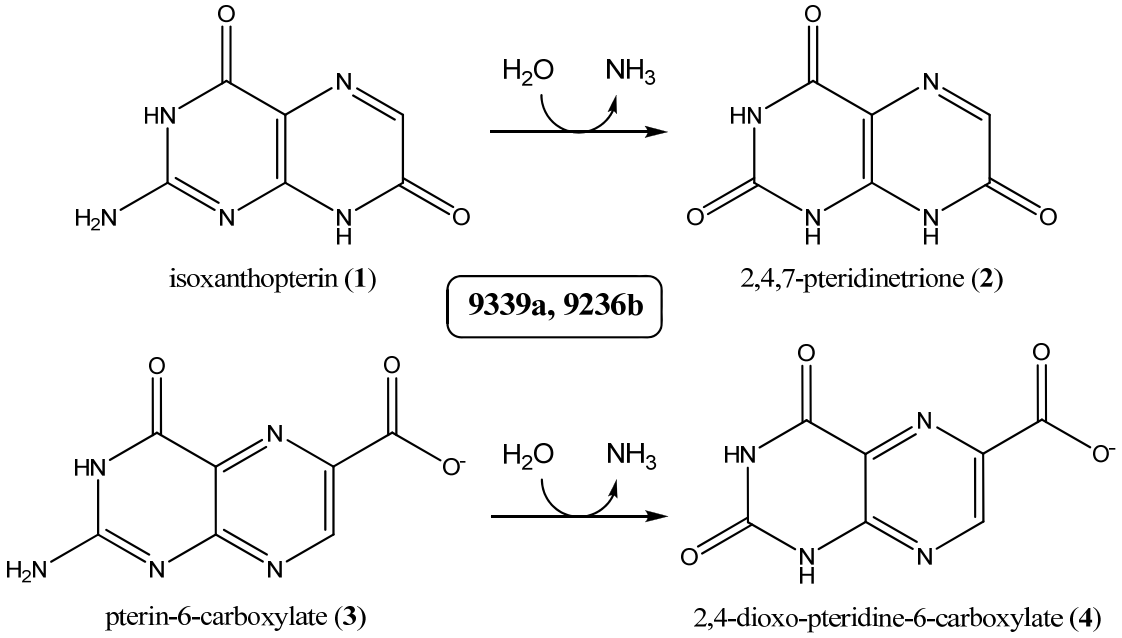
Scheme 7.1: Compounds utilized in the initial activity screens.



Scheme 7.1: Continued.



Scheme 7.2: Reactions catalyzed by 9339a and 9236b.



Absorbance spectra of isoxanthopterin (**1**) and the product of its deamination by 9339a, 2,4,7-pteridinetrione (7-oxylumazine) (**2**), are shown in Figure 7.5. In 50 mM TRIS at pH 8.5, the UV spectrum of isoxanthopterin has two peaks. The major peak has a λ_{max} at 333 nm and the smaller peak has a λ_{max} at 281 nm. The spectrum for the 7-oxylumazine (**2**) product has two peaks with λ_{max} at 326 nm and 274 nm. Kinetic parameters were determined for both enzymes with isoxanthopterin as the substrate using the direct assay at 350 nm with an extinction coefficient of 7650 M^{-1} cm^{-1}. 9339a has a k_{cat} of 1.0 s^{-1}, a K_m of 8 μM and a k_{cat}/K_m of 1.3×10^5 M^{-1} s^{-1} with isoxanthopterin (**1**). 9236b has a k_{cat} of 0.58 s^{-1}, a K_m of 24 μM and a k_{cat}/K_m of 2.4×10^4 M^{-1} s^{-1} with isoxanthopterin. The enzymatic deamination of isoxanthopterin by 9339a, producing 7-oxylumazine was verified by –ESI MS where a peak of m/z = 178 was observed for isoxanthopterin. This changed to a strong peak of m/z = 179 after addition of enzyme.

For the deamination of pterin-6-carboxylate a peak of m/z = 206 was observed. Upon addition of enzyme, a peak of m/z = 207 was observed.

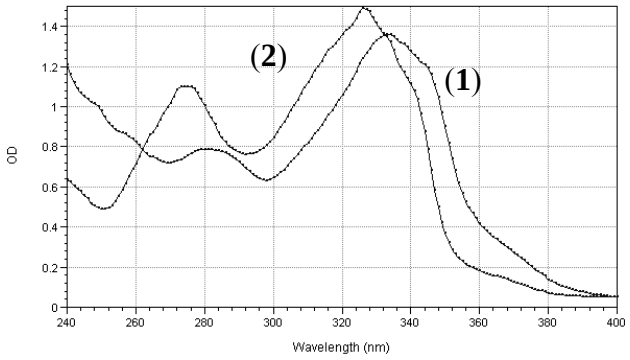


Figure 7.5: Spectra of isoxanthopterin (**1**) and 2,4,7-trioxopteridine (**2**) produced by 9339a in 50 mM TRIS, pH 8.5. Additional information is provided in the text.

Pterin-6-carboxylate (**3**) was the second best substrate tested for both 9339a and 9236b. Absorbance spectra of **3** and the deamination product, 2,4-dioxo-pteridine-6-carboxylate (**4**) are shown in Figure 7.6. The spectrum of **3** has two major peaks with λ_{max} values of 264 nm and 361 nm. There are two distinct peaks in the spectrum of **4** at λ_{max} values of 347 and 287 nm. Direct assays, measured at a wavelength of 264 nm with an extinction coefficient of 8080 M^{-1} cm^{-1}, yielded a k_{cat} of 1.01 s^{-1}, a K_m of 0.14 mM and a k_{cat}/K_m of 7.2×10^4 M^{-1} s^{-1} for 9339a with **3**. When tested with **3**, 9236b displayed a k_{cat} of 0.27 s^{-1}, a K_m of 0.09 mM and a k_{cat}/K_m of 3.1×10^3 M^{-1} s^{-1}. The kinetic parameters obtained for 9339a and 9236b with compounds **1** and **3** are summarized in Table 7.1.

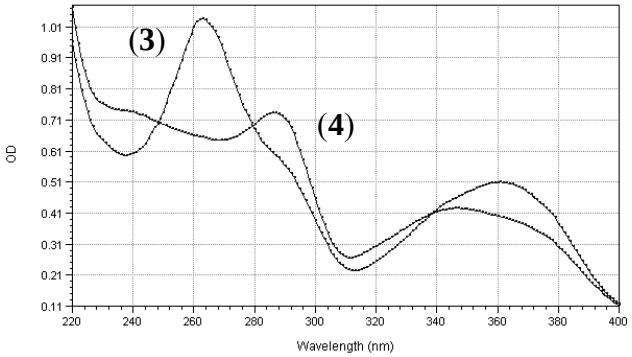


Figure 7.6: Spectra of pterin-6-carboxylate (**3**) and 2,4-dioxopteridine-6-carboxylate (**4**), produced by 9339a in 50 mM TRIS, pH 8.5. Additional information is provided in the text.

Table 7.1: Kinetic Parameters for Isoxanthopterin and Pterin-6-carboxylate Deamination with 9339a and 9236b.

enzyme	substrate	k_{cat} (s^{-1})	K_m mM	k_{cat}/K_m (M^{-1} s^{-1})
9339a	isoxanthopterin	1.05 ± 0.03	0.008 ± 0.001	$(1.3 \pm 0.2) \times 10^5$
	pterin-6-carboxylate	1.01 ± 0.10	0.14 ± 0.02	$(7.2 \pm 0.2) \times 10^3$
9236b	isoxanthopterin	0.91 ± 0.06	0.05 ± 0.007	$(1.8 \pm 0.2) \times 10^4$
	pterin-6-carboxylate	0.28 ± 0.05	0.09 ± 0.03	$(3.1 \pm 0.4) \times 10^3$

Sequence Analysis. The amino acid sequences of the two isoxanthopterin deaminases NYSGXRC-9339a and 9236b were aligned to the related 8-oxoguanine deaminases Pa0142 and NYSGXRC-9236e as well as the characterized guanine deaminases from *H. sapiens* and *E. coli.* The sequence alignment is shown in Figure 7.7. The eight TIM-barrel β-strands were identified from an analysis of the crystal structure of 9339a (PDB: 2PAJ) and are highlighted in gray. Conserved metal binding and catalytic residues which are currently observed in all AHS deaminases are highlighted in red. Conserved substrate binding residues found in guanine deaminases are highlighted

in green. The residues which have been implicated in the previous chapter of this work for the substrate binding of 8-oxoguanine deaminase are highlighted in blue. The proposed substrate binding residues specific for the binding of isoxanthopterin and pterin-6-carboxylate are highlighted in yellow.

```
9339a   ----------------------mttydtqpstlirnaaaimtggrgtaddps-30
9236b   --------------------------mqtealliknpaavmsglrg---daa-23
9236e   -mgadrrgermnleqhagarapntsssrpktllvkhadvlvtmddtr-rel-49
Pa0142  ------------------------------msrtwirnplaiftan-gld-19
Ec_GDA  ----------------mmsgehtlkavrgsfidvtrtidnpeeiasalrfi-35
Hs_GDA  ---------------mcaaqmpplahifrgtfvhstwtcpmevlrdhllgvs-37

9339a   -rvpgpdirivgdtidaig-alaprpge--tivdatdcviypawvnthhhl-77
9236b   -ragqvdlrirngrieaiapnleprade--rvidarscvvypgwvnthhhl-71
9236e   -rdaglyiednrivavg-psaelpetad--evldlrghlvipglvnthhhm-96
Pa0142  -aagglvvedgrivellgagqqpaqpca--sqfdasrhvvlpglvnthhhf-67
Ec_GDA  -edglllikqgkvewfgewengkhqipdtirvrdyrgklivpgfvdthihy-85
Hs_GDA  -dsgkivfleeasqqeklakewcfkpce--irelshheffmpglvdthiha-85

9339a   -fgsllkgepagldatltpwlaatpyrfralfderrfrlaariglielars-127
9236b   -fgnllkavpsginadlqewlaavpyprlarftpelarvaarlgmaellls-121
9236e   -yqsltravpaaqnaelfgwltnl-ykiwahltpemievstltamaellqs-145
Pa0142  -yqtltrawapvvnqplfpwlktl-ypvwarltpeklelatkvalaellls-116
Ec_GDA  -pqsemvg---aygeqllewlnkhtfpterryedleyaremsaffikqllr-132
Hs_GDA  -sqysfag--ssidlpllewltkytfpaehrfqnidfaeevytrvvrrtlk-133

9339a   -gcatvadhnyvyypgmpfdssailfeeaeklglrfvllrggatqtr-qle-176
9236b   -gvttcadhhylyhadgntetgdllfdeaaefgmrfvlcrggalqaagdhp-171
9236e   -gcttssdhlyiypn---gsrlddsigaaqrigmrfhasrg--amsvgqrd-190
Pa0142  -gcttaadhhylfpggl-eqaidvqagvveelgmramltrg--smslgekd-163
Ec_GDA  -ngtttalvfgtvhp----qsvdalfeaashinmrmiagkv--------mm-170
Hs_GDA  -ngtttacyfatiht----dssllladitdkfgqrafvgkv-------cmd-172

9339a   -adlptalrpetldayvadierlaaryhdaspramrrvvmapttvlysisp-226
9236b   -gfskvalkpetldqmladierlksryhdagdasmrrvvvaptttptfslpp-221
9236e   -gglppdsvverepdilrdtqrlietyhdegryamlrvvvapcsp-fsvsr-239
Pa0142  -gglppqqtvqeaetiladserliaryhqrgdgarvqialapcsp-fsvtp-212
Ec_GDA  -drnapdylldtaessyhqskelierwhkng---rllyaitprfaptsspe-217
Hs_GDA  -lndtfpeyketteesiketerfvsemlqkn-ysrvkpivtprfs-lscse-220

9339a   -remretaavarrlglrmhshlsetvgyqdsaysmygksp--vafcgehdw-274
9236b   -ellpevaraarrmglrlhthlsettryadfckerfnklp--vefvadhew-269
9236e   -dlmrdaavlareygvslhthlaenvndiaysrekfgmtp--aeyaedlgw-287
Pa0142  -eimrasaevaarhdvrlhthlaetldeedfclqrfglrt--vdyldsvgw-260
Ec_GDA  -qmamaqrlkeeypdtwvhthlcenkdeiawvkslypdhdgyldvyhqygl-267
Hs_GDA  -tlmgelgniaktrdlhiqshisenrdeveavknlypsyknytsvydknnl-270

9339a   -lgsdvwyahlvkvdadeiallaqtgtgvahcpqsngrlgsgicpvremad-324
9236b   -lgpdvwfahlvhlqpseiamlaetgsgcshcpvsnarlgsgiapapqmaa-319
9236e   -vghdvwhahcvqlddagiglfartgtgvahcpcsnmrlasgiapvkkmrl-337
Pa0142  -lgprtwlahgihfnaeeirrlgeagtgichcpssnmrlasgicptvelea-310
Ec_GDA  -tgkncvfahcvhleekewdrlsetkssiafcptsnlylgsglfnlkkawq-317
Hs_GDA  -ltnktvmahgcylsaeelnvfhergasiahcpnsnlslssgflnvlevlk-320
```

Figure 7.7: Sequence alignment of two isoxanthopterin deaminases (9339a and 9236b), two 8-oxoguanine deaminases (Pa0142 and 9236e) and two guanine deaminases (Ec_GDA and Hs_GDA). Additional details are provided in the text.

```
9339a   -agvpvsigvdgaasneaadmisevhmtwlaqrarlgmlaqpayrggsfeg-374
9236b   -agvpmslavdgvasnesgsmtneahfawlvhraaq----------------354
9236e   -agvpvglgvdgsasndgaqmvaevrqalllqrvgf----------------372
Pa0142  -agapiglgvdgsasndasnmilearqalylqrlry----------------345
Ec_GDA  -kkvkvgmgtdigagttfnmlqtlneaykvlqlqgy----------------352
Hs_GDA  -hevkiglgtdvaggysysmldairravmvsnillin-----------kv-358

9339a   -gagaasiaevihwgtaggarvmglde-vgkvavgyaadiavyrlddpryf-423
9236b   -gasattveevihwgsaggagvlglda-vgtlevgkaadfvlydvsdlrfn-403
9236e   -gpdamtarealeiatlggakvlnrdd-igalkpgmaadfaafdlrqplfa-421
Pa0142  -gaeritpelalgwatrgsarllgrsd-igelapgkqadlalfkldelrfs-394
Ec_GDA  ----rlsayeafylatlggakslgllddlignflpgkeadfvvmeptatplq-399
Hs_GDA  -neksltlkevfrlatlggsqalgldgeignfevgkefdailinpkasdsp-408

9339a   -glhdpaigpvasggrpsvmalfsagkrvvvddliegvdikelggearrvv-473
9236b   -gfhdvavapvtagepar vrynvvngrvvvdngvipgldlerlrheaaegv-453
9236e   -galhdpvaalvfcapsqtaytvvngkvvvregrlatldlppvierhnala-471
Pa0142  --gshdplsalllcaadradrvmvggawrvvdgavegldlaaliarhraaa-443
Ec_GDA  -qlrydnsvslvd---klfvmmtlgddrsiyrtyvdgrlvyern--------439
Hs_GDA  -idlfygdffgdiseaviqkflylgddrnieevyvggkqvvpfsssv-----454

9339a   -rellrevvv                                           -482
9236b   -kqllad---                                           -459
9236e   -halveaar-                                           -479
Pa0142  -saliag---                                           -449
Ec_GDA  ----------                                           -439
Hs_GDA  ----------                                           -454
```

Figure 7.7: Continued.

DISCUSSION

Sequence comparisons with guanine deaminases and 8-oxoguanine deaminases were used to differentiate a small group of enzymes which might catalyze a new reaction. Guanine deaminases utilize an Arg-Phe dyad for the recognition of guanine, as observed in the crystal structure of human guanine deaminase with bound xanthine (PDB: 2UZ9). It was also noted in the previous chapter of this work that 8-oxoguanine deaminases appear to utilize a Cys-Ser motif for recognition of 8-oxoguanine. In this small group of very similar enzymes, two threonine residues appear to form a Thr-Thr substrate recognition dyad in the same spot, as shown from the active site overlay with guanine deaminase and with the sequence alignments shown above.

The conservation of the Thr-Thr substrate recognition dyad was used to identify other proteins which may be performing a novel deaminase activity. Two of these enzymes were then tested for the deamination of a variety of pteridines, purines and pyrimidines. The high levels of activities obtained for the deamination of isoxanthopterin ($k_{cat}/K_m > 10^4$ M^{-1} s^{-1}) provide strong support that these enzymes function as isoxanthopterin deaminases *in vivo*. The moderate activity seen for pterin-6-carboxylate for both enzymes (~ 10^3 M^{-1} s^{-1}) may still be reasonably sufficient for biological relevance.

BLAST searches of protein sequences in the NCBI database were then used to identify the organisms containing quite similar genes. From this search, it was discovered that only a specialized group of 23 sequenced α, β and γ-proteobacterial and enterobacterial organisms share this gene as determined by the conservation of the Thr-Thr motif. The complete list of those genes and organisms are listed in Appendix B. The novel isoxanthopterin/pterin-6-carboxylate deaminases all share a sequence identity of greater than 41% and are predicted to participate in the deamination of isoxanthopterin within these organisms. Included with these 23 genes are 4 outliers which share a threonine-serine dyad at this position. These genes are: Mmwyl1_4101 (GI: 152998097) from *Marinomonas sp. MWYL1*; HMPREF0023_2011 (GI: 226953181) from *Acinetobacter sp. ATCC 27244*; PputW619_2587 (GI: 170721762) from *Pseudomonas putida W619*; and PP_3209 (GI: 26989928) from *Pseudomonas putida KT2440.* It appears that the Thr-Ser dyad could be an evolutionary intermediate between the Cys-Ser dyad of the 8-oxoguanine deaminases and the Thr-Thr dyad of the isoxanthopterin

deaminases. It is not yet clear whether this small difference will have significant implications for substrate specificity.

Both isoxanthopterin, and 2,4,7-trioxopteridine have been isolated from multiple biological sources such as: zebrafish embryos (*145*), the firebug *Pyrrhocoris apterus* (*146*), royal jelly and caviar (*147*) and in silkworm where isoxanthopterin activity was also measured (*142*). 6-Carboxypterin is a common product of folate and biopterin photo-oxidation, and 2,4-dioxo-pteridine-6-carboxylate production has been observed from *Pseudomonas fluorescens UK1* when grown on folate (*148*). Additonally, a purified extract of *Pseudomonas sp. Fo8* was found to deaminate pterin, 6-hydroxymethylpterin and pterin-6-carboxylate , while strain *Fo1* was found to yield less than 1% of that activity. (*149*).

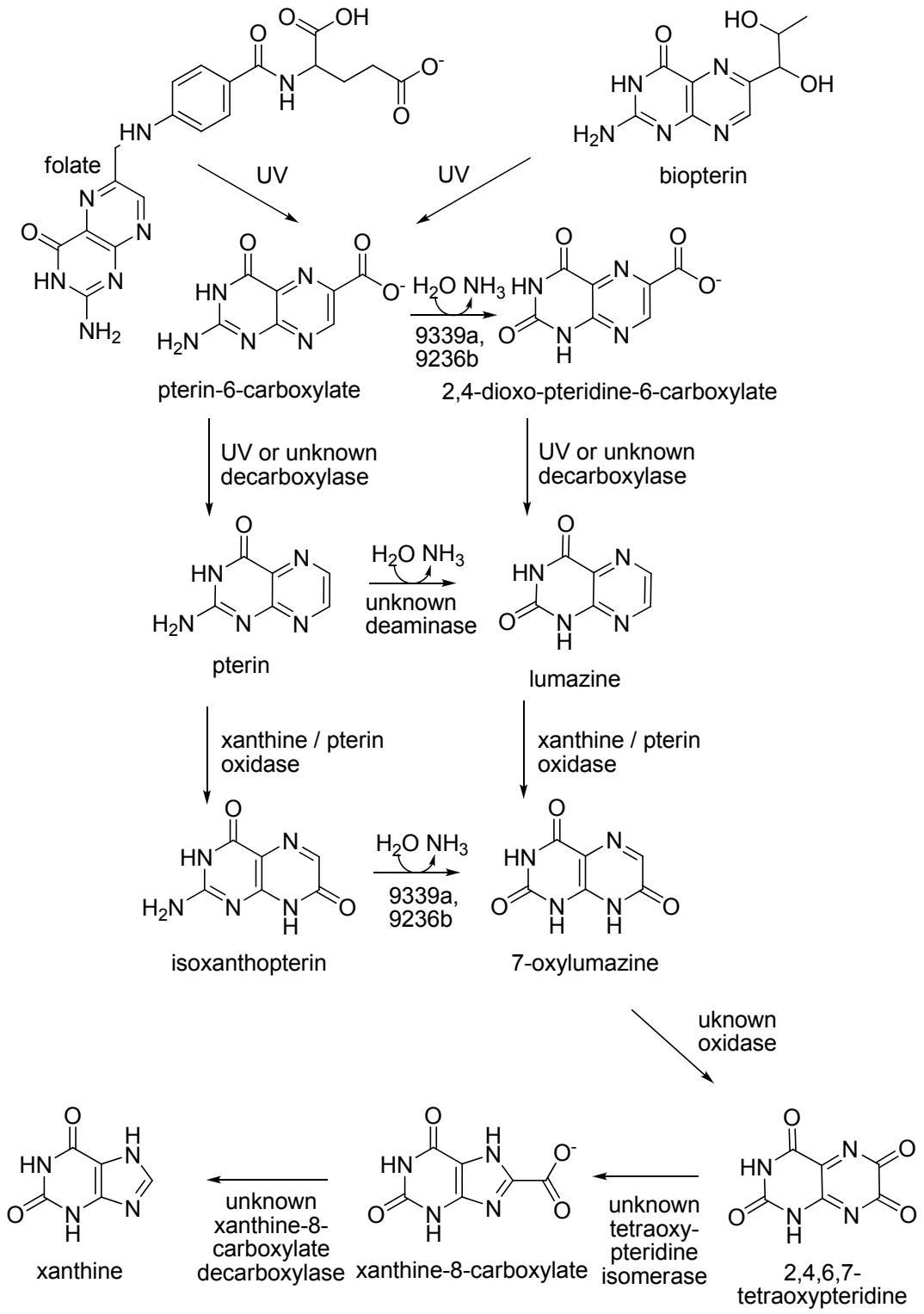
Noting the isoxanthopterin deaminase activity observed in extracts of *Alcaligenes faecalis* (*141*), a search of that genome was performed in pursuit of the specific deaminase which may be similar to 9339a and 9236b. Unfortunately, it appears that the genome of *A. faecalis* has not yet been fully sequenced. Of the genes sequenced for *A. faecalis*, there are none similar to the sequence of SGX-9339a or 9236b. Additional enzymes implicated in the degradation of isoxanthopterin by *A. faecalis*, are an unknown pterin-6-oxidase and a 2,4,6,7-tetraoxypteridine isomerase. These enzymes were found to catalyze the production of xanthine-8-carboxylate from isoxanthopterin (*141, 150*). An inducible xanthine-8-carboxylate decarboxylase of *A. faecalis* was found to complete

the degradation of pterines via entry into purine metabolism through the production of xanthine.

The proteins of unknown organism, NYSGXRC-9339a and 9236b are found to be most similar to the genes; Bxe_A2016 from *Burkholderia xenovorans LB400* and Bphyt_7136 from *Burkholderia phytofirmans PsJN.* They are the first proteins of known sequence found to possess significant deaminase activities against the pteridine compounds isoxanthopterin and pterin-6-carboxylate. These enzymes utilize a unique substrate binding motif that is capable of differentiating isoxanthopterin and pterin 6-carboxylate from 8-oxoguanine and guanine. While these substrate specificity profiles are somewhat different than those seen for the extracts *P. fluorescens UK1*, *Pseudomonas sp. Fo8* and *Fo1*, they are consistent the activity seen for *Alcaligenes faecalis*. Overall, these results demonstrate that the folate degradation pathways in bacteria such as *Burkholderia* and *Pseudomonas* are extremely diverse and that there are many new enzymes to discover in this remarkable pathway.

Based on the results of this study, a bacterial folate degradation pathway is proposed in Scheme 7.3. Pterin-containing molecules such as folate and biopterin, can easily decompose in the presence of solar radiation, forming pterin-6-carboxylate. Pterin-6-carboxylate may then become decarboxylated by some unknown decarboxylase or solar radiation. Another possible route would be the deamination of pterin-6-carboxylate by the enzymes homologous to 9339a or 9236b. The resulting pterin or 2,4-dioxopteridine is then oxidized at carbon 7 by xanthine oxidase, forming isoxanthopterin or 7-oxylumazine. 9339a and 9236b then deaminate isoxanthopterin, forming 7-oxylumazine. This product ultimately becomes oxidized to 2,4,6,7-tetraoxypteridine, by an unknown oxidase such as one of the oxidases seen in the genomic context figure shown previously. Tetraoxypteridine then undergoes isomerization to xanthine -8-carboxylate and is finally decarboxylated to xanthine. A look back at the genomic context surrounding 9339a and 9236b shows several probable decarboxylases which may be performing this role, or the initial decarboxylation of pterin-6-carboxylate.

Scheme 7.3: Bacterial Folate Degradation Pathway



CHAPTER VIII

SUMMARY AND CONCLUSIONS

The amidohydrolase superfamily is a remarkable set of evolutionarily and mechanistically related enzymes. Members of this superfamily catalyze a broad array of reactions including: hydrations, deaminations, deiminations, dehalogenations, decarboxylations, cyclizations and hydrolysis of heterocycles, peptides, amides, lactones and phosphtate esters. A result of the different enzymes sharing a common ancestral precursor could be the conservation of similar mechanisms for the catalysis of the various reactions.

Typically, amidohydrolases utilize one or two divalent transition metal cations for the activation of a water molecule and or substrate. Metals found in these enzymes may be Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, and Zn^{2+}. Activities have also been observed with enzymes prepared with added Cd^{2+}. Recent crystal structures have identified amidohydrolases which contain three or four active-site metals. The metal binding sites are named M_α, M_β, M_γ and M_δ. Variations in the metal content and metal binding ligands among structurally characterized AHS members, results in the identification of 16 distinct types of metal binding active sites. Of these different active sites, five are α/β-binuclear, 3 are β-mononuclear, four are α-mononuclear, three are α/β/γ trinuclear and one is α/β/γ/δ tetranuclear. The most common active site residues found in members of the amidohydrolase superfamily are: two histidines from the C-terminal end of β-strand 1 (HxH), a histidine from strand five, a histidine from strand 6 and an aspartate residue from strand 8.

The mechanistic details of α/β binuclear enzymes are still under debate, in spite of a large number of investigations. Studies have been performed on enzymes of this kind such as URE (*5, 61, 83, 86, 87, 151-156*), PTE (*46-51, 78, 80-82*), DHO (*31-33*) and IAD (*38-40*). Based on the analysis of those studies, a general mechanism for α/β binuclear enzymes was described in chapter I. The characterizations of the enzymes NagA and CDA presented here were directed toward a greater understanding of the mechanisms utilized by β-mononuclear and α-mononuclear amidohydrolase enzymes.

In NagA, it was uncertain if one or two metals were required for full activity or how the active site metal participates in catalysis because the NagA from *B. subtilis* contains two metals bound in the crystal structure 2VHL, while the crystal structure 1O12 from *T. maritima* had one metal in the active site. Additionally, in a group of NagAs including one from *E. coli*, it was noticed that the M_α forming HxH motif, was replaced by a QxN motif. It also was not known how the active site water becomes activated.

In chapter II of this dissertation, the catalytic mechanism of NagA from *E. coli* was addressed by the characterization of the catalytic properties of different metal-substituted forms of the enzyme. It was concluded that NagA requires one metal for catalysis, and that Mn^{2+}, Fe^{2+}, Ni^{2+}, Co^{2+}, Zn^{2+} and Cd^{2+} could all be used for catalysis at similar rates. Comparisons of the Zn and Cd forms of the enzyme in conjunction with the natural and thio-acetyl substrates demonstrated that the carbonyl oxygen interacts with the active site metal for transition state stabilization. pH-rate profiles with Zn and

Cd substituted forms of the NagA were used to demonstrate that the active site metal is also involved in the activation of the active site water.

N-acetyl-D-glucosamine-6-phosphate and N-acetyl-D-glucosamine-6-sulfate were compared through pH profiles to demonstrate that the doubly negative charge on the phosphate is extremely important for binding and catalysis. It was shown that NagA can utilize *N*-acetyl-D-galactosamine-6-phosphate equally as well as the canonical substrate, indicating that NagA may perform this function *in vivo*.*N*-formyl-D-glucosamine-6-phosphate and *N*-trifluoroacetyl-D-glucosamine-6-phosphate both acted as substrates. A 26-fold rate enhancement was seen for the trifluoroacetyl-substrate.

Mutants of the *E. coli* NagA were constructed to ascertain the roles of specific residues towards metal binding, substrate binding and catalysis. Substitution of the QxN motif for the HxH motif was performed to re-create the active site seen in the binuclear *B. subtilis* NagA. The double mutant was found to contain 50% more metal and a decrease in activity by about four orders of magnitude. Mutation of a Glu-131 from strand-three resulted in a loss of activity that was comparable to the loss in metal content, supporting its role in metal binding not catalysis. Significant losses in activity were seen for His-143 and His-251 mutants supporting their respective roles in transition state stabilization and substrate binding. There was no detectable activity found for Asp-273 mutants. It was concluded that this residue is primarily responsible for the activation of the hydrolytic water molecule.

The rates of substrate hydrolysis were compared in H_2O and D_2O, resulting in negligible solvent isotope effects. Solvent viscosity effects were also found to be

negligible. N-methylphosphonate-glucosamine-6-phosphate was found to be an excellent mimic of the tetrahedral intermediate, as supported by the low nanomolar level inhibition constant. These experiments, in conjunction with the large rate enhancement seen for the *N*-trifluoroacetyl-D-glucosamine-6-phosphate substrate, show that the chemical step involved in the formation or breakdown of the tetrahedral intermediate is rate limiting.

From these results, a catalytic mechanism was proposed that is similar to that of the binuclear enzymes. In this mechanism, catalysis is initiated by interactions of the carbonyl oxygen with the metal and His-143, resulting in polarization of the scissile bond. Asp-273 activates a metal bound water for hydrolytic attack at the carbonyl carbon of the substrate, resulting in the formation of a tetrahedral intermediate. The intermediate is stabilized through multiple interactions with Zn, His-143 and Asp-273. Proton transfer from Asp-273 to the leaving group amine initiates collapse of the tetrahedral intermediate.

The evolution and structure of NagA was addressed in chapter III. Sequence comparisons of *N*-acetyl-D-galactosamine-6-phosphate deacetylases (AgaA) and NagA indicate that all of the same residues required for binding and catalysis are present in both enzymes. This is supported by the wild-type level of activity seen in chapter II for NagA with the *N*-acetyl-D-galactosamine-6-phosphate substrate. Even though it contains all of the metal binding ligands seen in the α/β-binuclear active site of NagA from *B. subtilis*, metal chelation and reconstitution experiments were used to demonstrate that NagA from *T. maritima* utilizes a single divalent cation for catalysis, like the QxN

containing NagA enzymes. Sequence analysis indicated that all of the QxN containing enzymes as well as the mononuclear HxH containing *T. maritima* enzyme possess a histidine residue at the same position as His-143 of *E. coli* NagA. An X-ray crystal structure of NagA was obtained with a Zn^{2+} ion bound in the active site (PDB code: 2P50), this showed that the metal was bound to Glu-131, His-195 and His-216. Asp-273 was found to be interacting with a metal bound water. A structure of a D273N mutant with Zn^{2+} and the methylphosphonate inhibitor was also obtained (PDB code: 2P53). This structure elucidated the nature of the His-143 interactions with the transition state, showing that His-143 is aiding the β-metal in the polarization of the substrate and stabilization of the tetrahedral intermediate. This information provided a great deal of support for the establishment of the mechanism reported in chapter II.

Prior to this work, important mechanistic details of all α-mononuclear enzymes, particularly ADA and AMPD, were under debate. For instance, there was no consensus on which residues activate the hydrolytic water, or how the leaving group amnonia gets protonated. The mechanistic studies of CDA presented in chapter IV set out to shed light on the mechanism likely shared by all deaminases of the amidohydrolase superfamily.

CDA has been described as an Fe^{2+} dependent enzyme based on the greater level of activity seen for iron containing enzyme, compared to Zn^{2+} containing enzyme at pH 7.5. Metal chelation and reconstitution experiments showed that CDA was reactivated with 1.0 equivalent of Zn^{2+}. An activity assay was utilized with distinct extinction coefficients for each pH value measured. pH profiles of Zn^{2+} and Fe^{2+} containing enzymes were performed to elucidate the nature of the preference of CDA for Fe^{2+} over

Zn^{2+}. It was determined that the Fe^{2+} enzyme was able to effectively activate the hydrolytic water at low pH values, while at pH 7.5, the Zn^{2+} enzyme was already starting to lose activity.

Mutations performed at Glu-218, His-247 and Asp-314 all resulted in drastic losses of activity. This provides support for the vital roles that these enzymes play in catalysis. The E218A mutant was found to be able to catalyze the hydrolysis of 3-oxauracil, a compound which needs no protonation. This helps establish the role of Glu-218 in the protonation at N-3 of cytosine. Measurable levels of activity for the H247Q mutant were used in pH profiles. The resulting profiles dispayed no pH dependence on activity between pH 5 and 10. As a result, it was concluded that His-247 and all other catalytic residues need to be deprotonated for activity. The extremely low activity seen for the Asp-314 mutant supports its role in the final proton abstraction from hydrolytic water and its eventual protonation of the leaving group ammonia.

The sizable inverse isotope effects seen for CDA are indicative of a low-barrier H-bond occurring in the steps leading to the formation of the transition state. One such H-bond could be the bond formed between Glu-218 and N-3 of cytosine. Both groups have pK_a values close to four and should result in a compressed H-bond, but upon hydrolytic attack, the pK_a of the resulting uracil shifts to a much higher value, explaining why the H-bond in the structure appears to be of ordinary length. The two inhibitors served as excellent probes of the transition-state intermediate. Phosphonocytosine was a far better inhibitor than the phosphonouracil, resulting in an inhibition constant that is 20,000 times lower than the Michaelis constant of cytosine. This shows that transition

state stabilization is an essential component of the mechanism of CDA. Given the magnitude of the inverse solvent isotope effects, and the observation of the negliglible viscosity effects, it appears that the chemical step is rate determining.

Attempts at identifying a missing deaminase in the riboflavin pathway of archaea are described in chapter V of this dissertation. From the research that has been performed on the riboflavin biosynthetic genes in archaea, it has been noted that there is a deaminase that has yet to be identified. In bacteria and yeast, the deaminase is related to bacterial cytidine deaminase, which is not of the amidohydrolase superfamily. A cluster of genes containing many riboflavin biosynthetic proteins was identified in archaeal organisms such as *Methanococcus maripaludis, Methanosarcina mazei,* and *Sulfolobus solfataricus*. Within these clusters of proteins, there lies a hypothetical protein annotated as COG-SsnA or trzA-like. From the knowledge gained studying amidohydrolase enzymes, it was predicted that this enzyme contains all of the conserved active site residues found in deaminases.

For the production of the substrate for the deaminase, GTP cyclohydrolase II (RibA) from *E. coli* was expressed and purified. Mmp1348, a putative pyrimidine reductase from *Methanococcus maripaludis*, was cloned, expressed and purified. The substrate for the putative deaminase was prepared enzymatically with the two enzymes and verified through derivatization and UV/Vis characterization.

Three putative deaminases, MmarC7_0625 from *M. maripaludis,* Mm0823 from *M. mazei,* and Sso0398 from *S. solfataricus*, were cloned and expressed in BL21 cells, Rosetta cells, Rosetta-gami cells, and Rosetta pLysS cells, with and without $(His)_6$ tags,

and under a variety of conditions. Experimental conditions included various combinations of the following: low temperature expression at 15 ºC, High temperature expression at 45 ºC, variations from 0.1 to 1.0 mM IPTG, LB media with and without 500 mM NaCl and 1 mM glycine betaine. These conditions all failed to yield soluble or active protein.

The cell lysis pellets appeared to contain very pure protein. The purification of these inlusion bodies was attempted through partial denaturation with urea or guanidine HCl. Dialysis into less denaturing solutions resulted in immediate precipitation. For the thermophilic enzyme from *S. solfataricus* this dialysis was attempted at temperatures of up to 60 ºC, but the protein proved to be insoluble under all conditions.

Sso0398 was cloned into a pMAL-c4X maltose binding protein fusion vector. The resulting fusion protein proved to be soluble and was purified by amylose column. The protein fusion failed to yield any activity. The fusion was then cleaved using factor Xa protease. The resulting cleavage product precipitated, leaving only the maltose binding fusion partner in solution. This was verified by N-terminal sequencing.

It seems that without some form of subtantially different expression, the function of these deaminase-like proteins will remain elusive due to their extremely insoluble nature. Other remaining options for achieving soluble and active protein could be anaerobic expression, the incorporation of C-terminal fusion tags or the use of cell-free protein expression methods. It may also be possible that this protein needs some other subunit to be present for solubility and activity.

In chapter VI, a group of enzymes related to guanine deaminase were identified which appeared to be involved in a pathway with guanine deaminase and adenosine deaminase. While very similar in sequence to guanine deaminase, it was observed that these enzymes differed at two particular sites which are involved in substrate binding. Sequence alignments of these proteins were used to identify a key conserved Cys-Ser substrate binding dyad at the end of β-strand 4. One of these genes (9236e) was crystallized and assigned the PDB code: 3HPA. The crystal structure 3HPA showed that the protein had a Zn^{2+} bound in the active site and appeared to be very similar to guanine deaminase. A structural alignment of 3HPA with the human GDA crystal structure 2UZ9, showed that the Cys-232 may be participating in substrate recognition which is different than that of GDA. Another enzyme was identified within this group of Cys containing GDA-like enzymes is Pa0142, which shares a 46% sequence identity to 9236e.

Pa0142 and 9236e were tested for deaminase activity against a library of pteridines, purines and pyrimidines. These two proteins of unknown function were found to be active with isocytosine, guanine and 8-oxoguanine. 8-oxoguanine yielded 10-fold more activity for these two enzymes than any other compound, yielding k_{cat}/K_m values of up to 10^5 M^{-1} s^{-1}. The conservation of this cysteine group enabled the identification of 217 bacterial proteins which should possess 8-oxoguanine deaminase activity.

The final chapter of this dissertation describes the function discovery of two isoxanthopterin deaminases involved in the uncharacterized degradation pathway of folic acid. Thorough examination of unknown bacterial sequences related to guanine

deaminase produced a small group of 24 sequences which are highly similar to guanine deaminase and 8-oxoguanine deaminase, yet they contain another variation at the substrate recognition site identified in the previous chapter. This new group of enzymes contains a threonine at the very position where guanine deaminases contain a conserved arginine and 8-oxoguanine deaminases have a conserved cysteine. This group of genes is found among a cluster of genes which appear to be decarboxylases, oxidoreductases, dioxygenases, possible hydratases, other unknown deaminases, and xanthine/uracil permeases. The relevance of this genomic context became more apparent after the function of the deaminase was known.

Two of these proteins have been previously purified by the New York SGX Research Consortium. They are 9339a and 9236b. An X-ray crystal structure was also already known for 9339a and was assigned the PDB code: 2PAJ. It contained a single zinc atom in the active site and was missing the coordinates for several of the residues including the catalytic glutamate conserved among all deaminases. A structural alignment of 2PAJ with the human guanine deaminase structure 2UZ9 with bound xanthine showed that the threonine residue is positioned in the very spot as the arginine of guanine deaminase. It was not known what kind of activity this residue would impart, so the proteins 9339a and 9236b were tested for deaminase activity against a host of pteridines, purines and pyrimidines. These enzymes yielded maximum activities with isoxanthopterin that were 10 × greater than with any other compound. With k_{cat}/K_m on the order of 1.3×10^5 M^{-1} s^{-1} for 9339a and 2.4×10^4 M^{-1} s^{-1} for 9236b, it seems likely that isoxanthopterin is the natural substrate for these enzymes. The second best substrate

was pterin-6-carboxylate. It seems to be an odd substrate for activity, considering that the differences in size and charge between the two different pterin modifications. Considering the activities seen for these enzymes, it appears that the neighboring decarboxylases, dioxygenases, putative hydratases, and deaminases surrounding these isoxanthopterin deaminases fit nicely into the proposed pathway for folate degradation proposed in chapter VII. The two different activities make sense as well, because they eventually produce the same product. The genes near these enzymes would be of interest for further study, because the proteins of this pathway have never been identified. Genes related to 9339a and 9236b containing the threonine at this critical substrate recognition site are predicted to be isoxanthopterin deaminases.The conservation of this threonine was used to identify 24 genes which are predicted to deaminate isoxanthopterin. These genes are listed in Appendix B, and all share greater than 41% sequence identity to 9339a.

In conclusion, this research was successful in establishing detailed reaction mechanisms for the β-mononuclear and the α-mononuclear amidohydrolase enzymes which are consistent with experimental data. Residues implicated in substrate binding and catalysis were evaluated for individual contributions which may be used for future function discoveries. Knowledge of enzyme-substrate interactions for human guanine deaminase was used in the successful discovery of the related 8-oxoguanine deaminases and isoxanthopterin deaminases.

REFERENCES

1. Klug, D., Rabani, J., and Fridovich, I. (1972) A direct demonstration of the catalytic action of superoxide dismutase through the use of pulse radiolysis, *J Biol. Chem. 247*, 4839-4842.
2. Rotilio, G., Bray, R. C., and Fielden, E. M. (1972) A pulse radiolysis study of superoxide dismutase, *Biochim. Biophys. Acta 268*, 605-609.
3. Lewis, C. A., Jr., and Wolfenden, R. (2008) Uroporphyrinogen decarboxylation as a benchmark for the catalytic proficiency of enzymes, *Proc. Natl. Acad. Sci. U S A 105*, 17328-17333.
4. Lad, C., Williams, N. H., and Wolfenden, R. (2003) The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases, *Proc. Natl. Acad. Sci. U S A 100*, 5607-5610.
5. Lippard, S. J. (1995) At last--the crystal structure of urease, *Science 268*, 996-997.
6. Pauling, L. (1946) Molecular architecture and biological reactions, *Chemical & Engineering News*, 1375-1377.
7. 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.
8. Chiang, R. A., Sali, A., and Babbitt, P. C. (2008) Evolutionarily conserved substrate substructures for automated annotation of enzyme superfamilies, *PLoS Comput. Biol. 4*, e1000142.

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

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

11. Claren, J., Malisi, C., Hocker, B., and Sterner, R. (2009) Establishing wild-type levels of catalytic activity on natural and artificial (beta alpha)8-barrel protein scaffolds, *Proc. Natl. Acad. Sci. U S A 106*, 3704-3709.

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

13. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) A genomic perspective on protein families, *Science 278*, 631-637.

14. Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. (2003) The COG database: an updated version includes eukaryotes, *BMC Bioinformatics 4*, 41.

15. Matsui, H., Shimaoka, M., Kawasaki, H., Takenaka, Y., and Kurahashi, O. (2001) Adenine deaminase activity of the yicP gene product of *Escherichia coli*, *Biosci. Biotechnol. Biochem. 65*, 1112-1118.

16. Wang, Z., and Quiocho, F. A. (1998) Complexes of adenosine deaminase with two potent inhibitors: X-ray structures in four independent molecules at pH of maximum activity, *Biochemistry 37*, 8314-8324.

17. Wilson, D. K., Rudolph, F. B., and Quiocho, F. A. (1991) Atomic structure of adenosine deaminase complexed with a transition-state analog: understanding catalysis and immunodeficiency mutations, *Science 252*, 1278-1284.

18. Han, B. W., Bingman, C. A., Mahnke, D. K., Bannen, R. M., Bednarek, S. Y., Sabina, R. L., and Phillips, G. N., Jr. (2006) Membrane association, mechanism of action, and structure of Arabidopsis embryonic factor 1 (FAC1), *J. Biol. Chem. 281*, 14939-14947.

19. Ireton, G. C., Black, M. E., and Stoddard, B. L. (2001) Crystallization and preliminary X-ray analysis of bacterial cytosine deaminase, *Acta Crystallogr. D Biol. Crystallogr. 57*, 1643-1645.

20. Mahan, S. D., Ireton, G. C., Knoeber, C., Stoddard, B. L., and Black, M. E. (2004) Random mutagenesis and selection of *Escherichia coli* cytosine deaminase for cancer gene therapy, *Protein Eng. Des. Sel. 17*, 625-633.

21. Mahan, S. D., Ireton, G. C., Stoddard, B. L., and Black, M. E. (2004) Alanine-scanning mutagenesis reveals a cytosine deaminase mutant with altered substrate preference, *Biochemistry 43*, 8957-8964.

22. Porter, D. J., and Austin, E. A. (1993) Cytosine deaminase. The roles of divalent metal ions in catalysis, *J. Biol. Chem. 268*, 24005-24011.

23. 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.

24. Maynes, J. T., Yuan, R. G., and Snyder, F. F. (2000) Identification, expression, and characterization of *Escherichia coli* guanine deaminase, *J. Bacteriol. 182*, 4658-4660.

25. Marti-Arbona, R., and Raushel, F. M. (2006) Mechanistic characterization of N-formimino-L-glutamate iminohydrolase from *Pseudomonas aeruginosa*, *Biochemistry 45*, 14256-14262.

26. Seffernick, J. L., de Souza, M. L., Sadowsky, M. J., and Wackett, L. P. (2001) Melamine deaminase and atrazine chlorohydrolase: 98 percent identical but functionally different, *J. Bacteriol.183*, 2405-2410.

27. Shao, Z. Q., Seffens, W., Mulbry, W., and Behki, R. M. (1995) Cloning and expression of the s-triazine hydrolase gene (trzA) from *Rhodococcus corallinus* and development of *Rhodococcus* recombinant strains capable of dealkylating and dechlorinating the herbicide atrazine, *J. Bacteriol. 177*, 5748-5755.

28. Shapir, N., Pedersen, C., Gil, O., Strong, L., Seffernick, J., Sadowsky, M. J., and Wackett, L. P. (2006) TrzN from *Arthrobacter aurescens* TC1 Is a zinc amidohydrolase, *J. Bacteriol. 188*, 5859-5864.

29. Boundy-Mills, K. L., de Souza, M. L., Mandelbaum, R. T., Wackett, L. P., and Sadowsky, M. J. (1997) The atzB gene of *Pseudomonas* sp. strain ADP encodes the second enzyme of a novel atrazine degradation pathway, *Appl. Environ. Microbiol*. *63*, 916-923.

30. Sadowsky, M. J., Tong, Z., de Souza, M., and Wackett, L. P. (1998) AtzC is a new member of the amidohydrolase protein superfamily and is homologous to other atrazine-metabolizing enzymes, *J. Bacteriol. 180*, 152-158.

31. 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.

32. Anderson, M. A., Cleland, W. W., Huang, D. T., Chan, C., Shojaei, M., and Christopherson, R. I. (2006) 13C and 15N isotope effects for conversion of L-dihydroorotate to N-carbamyl-L-aspartate using dihydroorotase from hamster and *Bacillus caldolyticus*, *Biochemistry 45*, 7132-7139.

33. Porter, T. N., Li, Y., and Raushel, F. M. (2004) Mechanism of the dihydroorotase reaction, *Biochemistry 43*, 16285-16292.

34. Abendroth, J., Niefind, K., and Schomburg, D. (2002) X-ray structure of a dihydropyrimidinase from *Thermus* sp. at 1.3 A resolution, *J. Mol. Biol. 320*, 143-156.

35. 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.

36. 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.

37. Kim, K., Kim, M. I., Chung, J., Ahn, J. H., and Rhee, S. (2009) Crystal structure of metal-dependent allantoinase from *Escherichia coli*, *J. Mol. Biol. 387*, 1067-1074.

38. 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.

39. Marti-Arbona, R., Thoden, J. B., Holden, H. M., and Raushel, F. M. (2005) Functional significance of Glu-77 and Tyr-137 within the active site of isoaspartyl dipeptidase, *Bioorg. Chem. 33*, 448-458.

40. 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.

41. Kress, D., Alhapel, A., Pierik, A. J., and Essen, L. O. (2008) The crystal structure of enamidase: a bifunctional enzyme of the nicotinate catabolism, *J. Mol. Biol. 384*, 837-847.

42. Xiang, D. F., Xu, C., Kumaran, D., Brown, A. C., Sauder, J. M., Burley, S. K., Swaminathan, S., and Raushel, F. M. (2009) Functional annotation of two new carboxypeptidases from the amidohydrolase superfamily of enzymes, *Biochemistry 48*, 4567–4576.

43. 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.

44. Yu, Y., Liang, Y. H., Brostromer, E., Quan, J. M., Panjikar, S., Dong, Y. H., and Su, X. D. (2006) A catalytic mechanism revealed by the crystal structures of the imidazolonepropionase from *Bacillus subtilis*, *J. Biol. Chem. 281*, 36929-36936.

45. 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.

46. Xiang, D., Kolb, P., Fedorov, A., Meier, M., Federov, E., Nguyen, T., Sterner, R., Almo, S., Shoichet, B., and Raushel, F. (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.

47. Elias, M., Dupuy, J., Merone, L., Mandrich, L., Porzio, E., Moniot, S., Rochu, D., Lecomte, C., Rossi, M., Masson, P., Manco, G., and Chabriere, E. (2008) Structural basis for natural lactonase and promiscuous phosphotriesterase activities, *J. Mol. Biol. 379*, 1017-1028.

48. Afriat, L., Roodveldt, C., Manco, G., and Tawfik, D. S. (2006) The latent promiscuity of newly identified microbial lactonases is linked to a recently diverged phosphotriesterase, *Biochemistry 45*, 13677-13686.

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

50. Benning, M. M., Shim, H., Raushel, F. M., and Holden, H. M. (2001) High resolution X-ray structures of different metal-substituted forms of phosphotriesterase from *Pseudomonas diminuta*, *Biochemistry 40*, 2712-2722.

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

52. Omi, R., Goto, M., Nakagawa, N., Miyahara, I., and Hirotsu, K. (2004) Expression, purification and preliminary X-ray characterization of histidinol phosphate phosphatase, *Acta Crystallogr. D. Biol. Crystallogr. 60*, 574-576.

53. Peng, X., Masai, E., Katayama, Y., and Fukuda, M. (1999) Characterization of the meta-cleavage compound hydrolase gene involved in degradation of the lignin-related biphenyl structure by *Sphingomonas paucimobilis* SYK-6, *Appl. Environ. Microbiol. 65*, 2789-2793.

54. Peng, X., Masai, E., Kitayama, H., Harada, K., Katayama, Y., and Fukuda, M. (2002) Characterization of the 5-carboxyvanillate decarboxylase gene and its role in lignin-related biphenyl catabolism in *Sphingomonas paucimobilis* SYK-6, *Appl. Environ. Microbiol. 68*, 4407-4415.

55. Ishii, Y., Narimatsu, Y., Iwasaki, Y., Arai, N., Kino, K., and Kirimura, K. (2004) Reversible and nonoxidative gamma-resorcylic acid decarboxylase:

characterization and gene cloning of a novel enzyme catalyzing carboxylation of resorcinol, 1,3-dihydroxybenzene, from *Rhizobium radiobacter*, *Biochem. Biophys. Res. Commun. 324*, 611-620.

56. Martynowski, D., Eyobo, Y., Li, T., Yang, K., Liu, A., and Zhang, H. (2006) Crystal structure of alpha-amino-beta-carboxymuconate-epsilon-semialdehyde decarboxylase: insight into the active site and catalytic mechanism of a novel decarboxylation reaction, *Biochemistry 45*, 10412-10421.

57. Smiley, J. A., Kundracik, M., Landfried, D. A., Barnes, V. R., Sr., and Axhemi, A. A. (2005) Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from Neurospora crassa, *Biochim. Biophys. Acta 1723*, 256-264.

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

59. Callahan, B. P., Yuan, Y., and Wolfenden, R. (2005) The burden borne by urease, *J. Am. Chem. Soc. 127*, 10828-10829.

60. Estiu, G., and Merz, K. M., Jr. (2007) Competitive hydrolytic and elimination mechanisms in the urease catalyzed decomposition of urea, *J. Phys. Chem. B 111*, 10263-10274.

61. Marlier, J. F., and Cleland, W. W. (2006) Multiple isotope effect study of the hydrolysis of formamide by urease from jack bean (*Canavalia ensiformis*), *Biochemistry 45*, 9940-9948.

62. 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.

63. Williams, L., Nguyen, T., Li, Y., Porter, T. N., and Raushel, F. M. (2006) Uronate isomerase: a nonhydrolytic member of the amidohydrolase superfamily with an ambivalent requirement for a divalent metal ion, *Biochemistry 45*, 7453-7462.

64. Schwarzenbacher, R., Canaves, J. M., Brinen, L. S., Dai, X., Deacon, A. M., Elsliger, M. A., Eshaghi, S., Floyd, R., Godzik, A., Grittini, C., Grzechnik, S. K., Guda, C., Jaroszewski, L., Karlak, C., Klock, H. E., Koesema, E., Kovarik, J. S., Kreusch, A., Kuhn, P., Lesley, S. A., McMullan, D., McPhillips, T. M., Miller, M. A., Miller, M. D., Morse, A., Moy, K., Ouyang, J., Robb, A., Rodrigues, K., Selby, T. L., Spraggon, G., Stevens, R. C., van den Bedem, H., Velasquez, J., Vincent, J., Wang, X., West, B., Wolf, G., Hodgson, K. O., Wooley, J., and Wilson, I. A. (2003) Crystal structure of uronate isomerase (TM0064) from *Thermotoga maritima* at 2.85 A resolution, *Proteins 53*, 142-145.

65. Ferreira, F. M., Mendoza-Hernandez, G., Castaneda-Bueno, M., Aparicio, R., Fischer, H., Calcagno, M. L., and Oliva, G. (2006) Structural analysis of N-

acetylglucosamine-6-phosphate deacetylase apoenzyme from *Escherichia coli*, *J. Mol. Biol. 359*, 308-321.

66. 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.

67. 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.

68. Souza, J. M., Plumbridge, J. A., and Calcagno, M. L. (1997) N-acetylglucosamine-6-phosphate deacetylase from *Escherichia coli*: purification and molecular and kinetic characterization, *Arch. Biochem. Biophys. 340*, 338-346.

69. 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.

70. White, R. J., and Pasternak, C. A. (1967) The purification and properties of N-acetylglucosamine 6-phosphate deacetylase from *Escherichia coli*, *Biochem. J. 105*, 121-125.

71. White, R. J., and Pasternak, C. A. (1975) N-acetylglucosamine-6-phosphate deacetylase and glucosamine-6-phosphate deaminase from *Escherichia coli*, *Methods Enzymol. 41*, 497-502.

72. Cummings, J., Fedorov, A. A., Xu, C., Brown, S., Federov, E. V., 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*.

73. 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.

74. Yoshimune, K., Kanda, M., Wakayama, M., Kanda, S., Sato, A., Sakai, K., and Moriguchi, M. (2005) Role of arginine residues of D-aminoacylase from *Alcaligenes xylosoxydans* subsp. xylosoxydans A-6, *Protein Pept. Lett. 12*, 289-294.

75. Hara, H., Masai, E., Katayama, Y., and Fukuda, M. (2000) The 4-oxalomesaconate hydratase gene, involved in the protocatechuate 4,5-cleavage pathway, is essential to vanillate and syringate degradation in *Sphingomonas paucimobilis* SYK-6, *J. Bacteriol. 182*, 6950-6957.

76. Barnum, D. W. (1983) Hydrolysis of cations - formation-constants and standard free-energies of formation of hydroxy complexes, *Inorg. Chem. 22*, 2297-2305.

77. Del Vecchio, P., Elias, M., Merone, L., Graziano, G., Dupuy, J., Mandrich, L., Carullo, P., Fournier, B., Rochu, D., Rossi, M., Masson, P., Chabriere, E., and Manco, G. (2009) Structural determinants of the high thermal stability of SsoPox from the hyperthermophilic archaeon *Sulfolobus solfataricus*, *Extremophiles 13*, 461-470.

78. Buchbinder, J. L., Stephenson, R. C., Dresser, M. J., Pitera, J. W., Scanlan, T. S., and Fletterick, R. J. (1998) Biochemical characterization and crystallographic structure of an *Escherichia coli* protein from the phosphotriesterase gene family, *Biochemistry 37*, 10860.

79. 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.

80. Kim, J., Tsai, P. C., Chen, S. L., Himo, F., Almo, S. C., and Raushel, F. M. (2008) Structure of diethyl phosphate bound to the binuclear metal center of phosphotriesterase, *Biochemistry 47*, 9497-9504.

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

82. Jackson, C. J., Foo, J. L., Kim, H. K., Carr, P. D., Liu, J. W., Salem, G., and Ollis, D. L. (2008) In crystallo capture of a Michaelis complex and product-binding modes of a bacterial phosphotriesterase, *J. Mol. Biol. 375*, 1189-1196.

83. Pearson, M. A., Park, I. S., Schaller, R. A., Michel, L. O., Karplus, P. A., and Hausinger, R. P. (2000) Kinetic and structural characterization of urease active site variants, *Biochemistry 39*, 8575-8584.

84. Barrios, A. M., and Lippard, S. J. (2001) Decomposition of alkyl-substituted urea molecules at a hydroxide-bridged dinickel center, *Inorg. Chem. 40*, 1250-1255.

85. Estiu, G., Suarez, D., and Merz, K. M. (2006) Quantum mechanical and molecular dynamics simulations of ureases and Zn beta-lactamases, *J. Comput. Chem. 27*, 1240-1262.

86. Suarez, D., Diaz, N., and Merz, K. M., Jr. (2003) Ureases: quantum chemical calculations on cluster models, *J. Am. Chem. Soc. 125*, 15324-15337.

87. Marlier, J. F., Fogle, E. J., and Cleland, W. W. (2008) A heavy-atom isotope effect and kinetic investigation of the hydrolysis of semicarbazide by urease from jack bean (*Canavalia ensiformis*), *Biochemistry 47*, 11158-11163.

88. Merkler, D. J., and Schramm, V. L. (1993) Catalytic mechanism of yeast adenosine 5'-monophosphate deaminase. Zinc content, substrate specificity, pH studies, and solvent isotope effects, *Biochemistry 32*, 5792-5799.

89. Gleeson, M. P., Burton, N. A., and Hillier, I. H. (2003) The mechanism of adenosine deaminase catalysis studied by QM/MM calculations: the role of histidine 238 and the activity of the alanine 238 mutant, *Phys. Chem. Chem. Phys. 5*, 4272-4278.

90. Uehara, T., and Park, J. T. (2004) The N-acetyl-D-glucosamine kinase of *Escherichia coli* and its role in murein recycling, *J. Bacteriol. 186*, 7273-7279.

91. Uehara, T., Suefuji, K., Jaeger, T., Mayer, C., and Park, J. T. (2006) MurQ Etherase is required by *Escherichia coli* in order to metabolize anhydro-N-acetylmuramic acid obtained either from the environment or from its own cell wall, *J. Bacteriol. 188*, 1660-1662.

92. Nishiyama, T., Kawamura, Y., Kawamoto, K., Matsumura, H., Yamamoto, N., Ito, T., Ohyama, A., Katsuragi, T., and Sakai, T. (1985) Antineoplastic effects in rats of 5-fluorocytosine in combination with cytosine deaminase capsules, *Cancer Res. 45*, 1753-1761.

93. Morris, S. M. (1993) The genetic toxicology of 5-fluoropyrimidines and 5-chlorouracil, *Mutat. Res. 297*, 39-51.

94. Hayden, M. S., Linsley, P. S., Wallace, A. R., Marquardt, H., and Kerr, D. E. (1998) Cloning, overexpression, and purification of cytosine deaminase from *Saccharomyces cerevisiae*, *Protein Expr. Purif. 12*, 173-184.

95. Xiang, S., Short, S. A., Wolfenden, R., and Carter, C. W., Jr. (1997) The structure of the cytidine deaminase-product complex provides evidence for efficient proton transfer and ground-state destabilization, *Biochemistry 36*, 4768-4774.

96. Graupner, M., Xu, H., and White, R. H. (2002) The pyrimidine nucleotide reductase step in riboflavin and F(420) biosynthesis in archaea proceeds by the eukaryotic route to riboflavin, *J. Bacteriol. 184*, 1952-1957.

97. Uehara, T., Suefuji, K., Valbuena, N., Meehan, B., Donegan, M., and Park, J. T. (2005) Recycling of the anhydro-N-acetylmuramic acid derived from cell wall

murein involves a two-step conversion to N-acetylglucosamine-phosphate, *J. Bacteriol. 187*, 3643-3649.

98. 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.

99. Brouwer, A. C., and Kirsch, J. F. (1982) Investigation of diffusion-limited rates of chymotrypsin reactions by viscosity variation, *Biochemistry 21*, 1302-1307.

100. St Maurice, M., and Bearne, S. L. (2002) Kinetics and thermodynamics of mandelate racemase catalysis, *Biochemistry 41*, 4048-4058.

101. Pearson, R. G. (1966) Acids and bases, *Science 151*, 172-177.

102. Bond, M. D., Holmquist, B., and Vallee, B. L. (1986) Thioamide substrate probes of metal-substrate interactions in carboxypeptidase A catalysis, *J. Inorg. Biochem. 28*, 97-105.

103. Frick, L., and Wolfenden, R., (Eds.) (1989) *Design of Enzyme Inhibitors as Drugs*, Oxford University Press, Oxford.

104. Bartlett, P. A., and Marlowe, C. K. (1987) Possible role for water dissociation in the slow binding of phosphorus-containing transition-state-analogue inhibitors of thermolysin, *Biochemistry 26*, 8553-8561.

105. Holden, H. M., Tronrud, D. E., Monzingo, A. F., Weaver, L. H., and Matthews, B. W. (1987) Slow- and fast-binding inhibitors of thermolysin display different modes of binding: crystallographic analysis of extended phosphonamidate transition-state analogues, *Biochemistry 26*, 8542-8553.

106. Morgan, B. P., Holland, D. R., Matthews, B. W., and Bartlett, P. A. (1994) Structure-based design of an inhibitor of the zinc peptidase thermolysin, *J. Am. Chem. Soc. 116*, 3251-3260.

107. Radkiewicz, J. L., McAllister, M. A., Goldstein, E., and Houk, K. N. (1998) A theoretical investigation of phosphonamidates and sulfonamides as protease transition state isosteres, *J. Org. Chem. 63*, 1419-1428.

108. Morrison, J. F. (1969) Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors, *Biochim. Biophys. Acta. 185*, 269-286.

109. Sculley, M. J., Morrison, J. F., and Cleland, W. W. (1996) Slow-binding inhibition: the general case, *Biochim. Biophys. Acta. 1298*, 78-86.

110. Glaser, K. B., Li, J., Aakre, M. E., Morgan, D. W., Sheppard, G., Stewart, K. D., Pollock, J., Lee, P., O'Connor, C. Z., Anderson, S. N., Mussatto, D. J., Wegner, C. W., and Moses, H. L. (2002) Transforming growth factor beta mimetics: discovery of 7-[4-(4-cyanophenyl)phenoxy]-heptanohydroxamic acid, a biaryl hydroxamate inhibitor of histone deacetylase, *Mol. Cancer Ther. 1*, 759-768.

111. McClerren, A. L., Endsley, S., Bowman, J. L., Andersen, N. H., Guan, Z., Rudolph, J., and Raetz, C. R. (2005) A slow, tight-binding inhibitor of the zinc-dependent deacetylase LpxC of lipid A biosynthesis with antibiotic activity comparable to ciprofloxacin, *Biochemistry 44*, 16574-16583.

112. Jencks, W. P., and Westheimer, F. H. p*K*a Data Compiled by R. Williams. http://research.chem.psu.edu/brpgroup/p*K*a¥compilation. pdf (accessed 11/06/06), in part of *B. R. Peterson's web page at*

Penn State University (accessed 11/06/06). http://research.chem.psu. edu/brpgroup/grouphomepage.html (accessed 11/06/06).

113. Cleland, W. W. (1982) The use of pH studies to determine chemical mechanisms of enzyme-catalyzed reactions, *Methods Enzymol. 87*, 390-405.

114. Le Coq, J., An, H. J., Lebrilla, C., and Viola, R. E. (2006) Characterization of human aspartoacylase: The brain enzyme responsible for Canavan disease, *Biochemistry 45*, 5878-5884.

115. Uehara, T., and Park, J. T. (2003) Identification of MpaA, an amidase in *Escherichia coli* that hydrolyzes the gamma-D-glutamyl-meso-diaminopimelate bond in murein peptides, *J. Bacteriol. 185*, 679-682.

116. Uehara, T., and Park, J. T. (2002) Role of the murein precursor UDP-N-acetylmuramyl-L-Ala-gamma-D-Glu-meso-diaminopimelic acid-D-Ala-D-Ala in repression of beta-lactamase induction in cell division mutants, *J. Bacteriol. 184*, 4233-4239.

117. Brinkkotter, A., Kloss, H., Alpert, C., and Lengeler, J. W. (2000) Pathways for the utilization of N-acetyl-galactosamine and galactosamine in *Escherichia coli*, *Mol. Microbiol. 37*, 125-135.

118. Kobayashi, S., Nagayama, S., and Busujima, T. (1998) Lewis acid catalysts stable in water. Correlation between catalytic activity in water and hydrolysis constants and exchange rate constants for substitution of inner-sphere water ligands, *J. Am. Chem. Soc. 120*, 8287-8288.

119. Bartlett, P. A., Hunt, J. T., Adams, J. L., and Gehret, J. C. E. (1978) Phosphorus-containing purines and pyrimidines - new class of transition-state analogs, *Bioorg. Chem. 7*, 421-436.

120. Ganguly, S., and Kundu, K. K. (1994) Protonation/deprotonation energetics of uracil, thymine and cytosine in water from e.m.f/spectrophotometric measurements, *Can. J. Chem. 72*, 1120-1126.

121. Venkatasubban, K. S., and Schowen, R. L. (1984) The proton inventory technique, *CRC Crit. Rev. Biochem. 17*, 1-44.

122. Weiss, P. M., Cook, P. F., Hermes, J. D., and Cleland, W. W. (1987) Evidence from nitrogen-15 and solvent deuterium isotope effects on the chemical mechanism of adenosine deaminase, *Biochemistry 26*, 7378-7384.

123. Cleland, W. W. (1992) Low-barrier hydrogen bonds and low fractionation factor bases in enzymatic reactions, *Biochemistry 31*, 317-319.

124. Magalhaes, M. L., Argyrou, A., Cahill, S. M., and Blanchard, J. S. (2008) Kinetic and mechanistic analysis of the *Escherichia coli* ribD-encoded bifunctional deaminase-reductase involved in riboflavin biosynthesis, *Biochemistry 47*, 6499-6507.

125. Chatwell, L., Krojer, T., Fidler, A., Romisch, W., Eisenreich, W., Bacher, A., Huber, R., and Fischer, M. (2006) Biosynthesis of riboflavin: structure and properties of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-phosphate reductase of *Methanocaldococcus jannaschii*, *J. Mol. Biol. 359*, 1334-1351.

126. Grochowski, L. L., Xu, H., and White, R. H. (2009) An iron(II) dependent formamide hydrolase catalyzes the second step in the archaeal biosynthetic pathway to riboflavin and 7,8-didemethyl-8-hydroxy-5-deazariboflavin, *Biochemistry 48*, 4181-4188.

127. Hollander, I., and Brown, G. M. (1979) Biosynthesis of riboflavin: reductase and deaminase of *Ashbya gossypii*, *Biochem. Biophys. Res. Commun. 89*, 759-763.

128. Richter, G., Fischer, M., Krieger, C., Eberhardt, S., Luttgen, H., Gerstenschlager, I., and Bacher, A. (1997) Biosynthesis of riboflavin: characterization of the bifunctional deaminase-reductase of *Escherichia coli* and *Bacillus subtilis*, *J. Bacteriol. 179*, 2022-2028.

129. Bacher, A., and Lingens, F. (1970) Biosynthesis of riboflavin. Formation of 2,5-diamino-6-hydroxy-4-(1'-D-ribitylamino)pyrimidine in a riboflavin auxotroph, *J. Biol. Chem. 245*, 4647-4652.

130. Le Page, F., Gentil, A., and Sarasin, A. (1999) Repair and mutagenesis survey of 8-hydroxyguanine in bacteria and human cells, *Biochimie 81*, 147-153.

131. Mishra, P. C., Singh, A. K., and Suhai, S. (2005) Interaction of singlet oxygen and superoxide radical anion with guanine and formation of its mutagenic modification 8-oxoguanine, *Int. J. Quantum Chem. 102*, 282-301.

132. Druzhyna, N. M., Musiyenko, S. I., Wilson, G. L., and LeDoux, S. P. (2005) Cytokines induce nitric oxide-mediated mtDNA damage and apoptosis in oligodendrocytes. Protective role of targeting 8-oxoguanine glycosylase to mitochondria, *J. Biol. Chem. 280*, 21673-21679.

133. D'Errico, M., Parlanti, E., and Dogliotti, E. (2008) Mechanism of oxidative DNA damage repair and relevance to human pathology, *Mutat. Res. 659*, 4-14.

134. Redrejo-Rodriguez, M., Ishchenko, A. A., Saparbaev, M. K., Salas, M. L., and Salas, J. (2009) African swine fever virus AP endonuclease is a redox-sensitive enzyme that repairs alkylating and oxidative damage to DNA, *Virology 390*, 102-109.

135. Wentworth, P., Jr., Jones, L. H., Wentworth, A. D., Zhu, X., Larsen, N. A., Wilson, I. A., Xu, X., Goddard, W. A., 3rd, Janda, K. D., Eschenmoser, A., and Lerner, R. A. (2001) Antibody catalysis of the oxidation of water, *Science 293*, 1806-1811.

136. Shukla, P. K., and Mishra, P. C. (2007) H2O3 as a reactive oxygen species: formation of 8-oxoguanine from its reaction with guanine, *J. Phys. Chem. B 111*, 4603-4615.

137. Orsomando, G., Bozzo, G. G., de la Garza, R. D., Basset, G. J., Quinlivan, E. P., Naponelli, V., Rebeille, F., Ravanel, S., Gregory, J. F., 3rd, and Hanson, A. D. (2006) Evidence for folate-salvage reactions in plants, *Plant J. 46*, 426-435.

138. Blakey, R. L. (1969) The Biochemistry of Folic Acid and Related Pteridines, pp 58-59, Wiley, New York.

139. Dunlap, W. C., and Susic, M. (1986) Photochemical decomposition of pteridines and flavins in seawater exposed to surface solar radiation., *Mar. Chem. 19*, 99-107.

140. Lowry, O. H., Bessey, O. A., and Crawford, E. J. (1949) Pterine oxidase, *J. Biol. Chem. 180*, 399-410.

141. McNutt, W. S., Jr. (1963) The metabolism of isoxanthopterin by *Alcaligenes faecalis*, *J. Biol. Chem. 238*, 1116-1121.

142. Gyure, W. L. (1974) Catabolism of isoxanthopterin during the development of the silkworm, *Bombyx mori*., *Insect Biochem. 4*, 303-312.

143. Shaw, E., Baugh, C. M., and Krumdieck, C. L. (1966) The chemical degradation of folic acid. Photolysis of 2,4,7-trihydroxypteridine, *J. Biol. Chem. 241*, 379-382.

144. Jabri, E., Carr, M. B., Hausinger, R. P., and Karplus, P. A. (1995) The crystal structure of urease from *Klebsiella aerogenes*, *Science 268*, 998-1004.

145. Ziegler, I., McDonald, T., Hesslinger, C., Pelletier, I., and Boyle, P. (2000) Development of the pteridine pathway in the zebrafish, *Danio rerio*, *J. Biol. Chem. 275*, 18926-18932.

146. Bel, Y., Porcar, M., Socha, R., Nemec, V., and Ferre, J. (1997) Analysis of pteridines in *Pyrrhocoris apterus* (L.) (Heteroptera, Pyrrhocoridae) during development and in body-color mutants., *Arch. Insect Biochem. Physiol. 34*, 83-98.

147. Andondonskaja-Renz, B., and Zeitler, H. J. (1989) Determination of pteridines in royal jelly and caviar by reverse phase high-performance liquid chromatography., *J. Micronutr. Anal. 5*, 83-90.

148. Bacher, A., and Rappold, H. (1980) Bacterial degradation of folic acid, *Methods Enzymol. 66*, 652-656.

149. Rappold, H., and Bacher, A. (1974) Bacterial degradation of folic acid, *J. Gen. Microbiol. 85*, 283-290.

150. McNutt, W. S., and Damle, S. P. (1964) Tetraoxypteridine isomerase, *J. Biol. Chem. 239*, 4272-4279.

151. Jabri, E., Lee, M. H., Hausinger, R. P., and Karplus, P. A. (1992) Preliminary crystallographic studies of urease from jack bean and from *Klebsiella aerogenes*, *J. Mol. Biol. 227*, 934-937.

152. Moncrief, M. B., Hom, L. G., Jabri, E., Karplus, P. A., and Hausinger, R. P. (1995) Urease activity in the crystalline state, *Protein Sci. 4*, 2234-2236.

153. Park, I. S., Michel, L. O., Pearson, M. A., Jabri, E., Karplus, P. A., Wang, S., Dong, J., Scott, R. A., Koehler, B. P., Johnson, M. K., and Hausinger, R. P. (1996) Characterization of the mononickel metallocenter in H134A mutant urease, *J. Biol. Chem. 271*, 18632-18637.

154. Pearson, M. A., Michel, L. O., Hausinger, R. P., and Karplus, P. A. (1997) Structures of Cys319 variants and acetohydroxamate-inhibited *Klebsiella aerogenes* urease, *Biochemistry 36*, 8164-8172.

155. Pearson, M. A., Schaller, R. A., Michel, L. O., Karplus, P. A., and Hausinger, R. P. (1998) Chemical rescue of *Klebsiella aerogenes* urease variants lacking the carbamylated-lysine nickel ligand, *Biochemistry 37*, 6214-6220.

156. Yamaguchi, K., Cosper, N. J., Stalhandske, C., Scott, R. A., Pearson, M. A., Karplus, P. A., and Hausinger, R. P. (1999) Characterization of metal-substituted *Klebsiella aerogenes* urease, *J. Biol. Inorg. Chem. 4*, 468-477.

APPENDIX A

List of all genes predicted to possess 8-oxoguanine deaminase activity.

*8-oxoguanine deaminase activity was confirmed experimentally.

	Locus tag	GI	Organism
1*	PA0142	879447	Pseudomonas aeruginosa PA01
2	PaerPA_01000142	107099133	Pseudomonas aeruginosa PACS2
3	PA2G_04056	6715071	Pseudomonas aeruginosa 2192
4	PACG_03701	6720155	Pseudomonas aeruginosa C3719
5	PA14_01750	4383541	Pseudomonas aeruginosa UCBPP-PA14
6	PSPA7_0219	5356948	Pseudomonas aeruginosa PA7
7	Pmen_3705	5105700	Pseudomonas mendocina ymp
8	PSEEN0675	4088425	Pseudomonas entomophila L48
9	Pf101_0675	3713960	Pseudomonas fluorescens PfO-1
10	PFL_0724	3481313	Pseudomonas fluorescens Pf-5
11	PFLU0669	7819150	Pseudomonas fluorescens SBW25
12	PSPTOT1_3659	213969358	Pseudomonas syringae pv. tomato T1
13	atzB	1182385	Pseudomonas syringae pv. Tomato str. DC3000
14	Psyr_0669	3366147	Pseudomonas syringae pv. Syringae B728a
15	Psyrpo1_010100001930	237797581	Pseudomonas syringae pv. oryzae str. 1_6
16	PSPPH_4628	3555934	Pseudomonas syringae pv. phaseolicola 1448A
17	atzB	5096040	Pseudomonas stutzeri A1501
18	Arth_3425	4444155	Arthrobacter sp. FB24
19	AAur_3399	4639368	Arthrobacter aurescens TC1
20	Achl_3205	7294686	Arthrobacter chloophenolicus A6
21	ACIAD0761	2879221	Acinetobacter sp. ADP1
22	BlinB01003114	62423337	Brevibacterium linens BL2
23	VSAK1_24940	149188732	Vibrio shilonii AK1
24	V12B01_19391	84386830	Vibrio splendid 12B01
25	VSWAT3_19088	148978109	Vibrionales bacterium SWAT-3
26	P3TCK_01255	90413684	Photobacterium profundum 3TCK
27	PBPRA2267	3122820	Photobacterium profundum SS9
28	KT99_04812	163751045	Shewanella benthica KT99
29	PE36_19015	149908977	Moritella sp. PE36
30	SI859A1_03023	90420336	Aurantimonas manganoxydans SI85-

			9A1
31	Pcryo_1786	4033973	Psychrobacter cryohalolentis K5
32	Mnod_1111	7304423	Methylobacterium nodulans ORS 2060
33	MELB17_16823	126666105	Marinobacter sp. ELB17
34	Mext_1697	5832952	Methylobacterium extorquens PA1
35	Mchl_2016	7118716	Methylobacterium chloromethanicum CM4
36	BRADO1738	5118449	Bradyrhizobium sp. ORS278
37	Mchl_3248	7117587	Methylobacterium chloromethanicum CM4
38	Mpop_1121	6313798	Methylobacterium populi BJ001
39	M446_2144	6130066	Methylobacterium sp. 4-46
40	FP2506_15334	114705851	Fulvimarina pelagi HTCC2506
41	B116159	1048904	Brandyrhizobium japonicum USDA 110
42	MED121_14459	87118237	Marinomonas sp. MED121
43	MDG893_15717	149375925	Marinobacter algicola DG893
44	BBta 2050	5153053	Brandyrhizobium sp. BTAi1
45	Mrad2831_4071	6140129	Methylobacterium radiotolerans JCM 2831
46	HCH_01106	3841406	Hahella chejuensis KCTC 2396
47	Mext_3024	5835455	Methylobacterium extorquens PA1
48	CJA_0540	6415371	Cellvibrio japonicus Uedal07
49	Patl_2224	4172376	Pseudoalteromonas atlantica T6c
50	StermDRAFT_06890	229881741	Sebaldella termitidis ATCC 33386
51	Francci3_0845	3904327	Frankia sp. CcI3
52	Bphy_1293	6242785	Burkholderia phymatum STM815
53	Bxe_A1936	4003154	Burkholderia xenovorans LB400
54	Bphyt_2323	6281242	Burkholderia phytofirmans PsJN
55	BgramDRAFT_5200	170695243	Burkholderia graminis C4D1M
56	SADFL11_1395	7514273	Labrenzia alexandrii DFL-11
57	NOR53_2032	7635455	Gamma proteobacterium NOR5-3
58	Bcep18194_A5267	3750479	Burkholderia sp. 383
59	BuboB_010100016293	167586903	Burkholderia ubonensis Bu
60	BTH_I2073	3848116	Burkholderia thailandensis E264
61*	NYSGXRC-9236e	44264246	Env. Samp. of Sargasso sea
62	BDAG_01340	6703425	Burkholderia dolosa AUO158
63	BthaT_010100025531	167581531	Burkholderia thailandensis TXDOH
64	Bpse38_010100010869	167836982	Burkholderia thailandensis MSMB43
65	BH160DRAFT_3765	209518669	Burkholderia sp. H160

66	BCPG_02756	6709226	Burkholderia cenocepacia PC184
67	BCAL2029	6932197	Burkholderia cenocepacia J2315
68	Bcenmc03 1980	6123644	Burkholderia cenocepacia MC0-3
69	Bglu_1g13210	7908425	Burkholderia glumae BGR1
70	KoleDRAFT_1086	213014337	Thermotogales bacterium TBF 19.5.1
71	Bcen_6122	4097155	Burkholderia cenocepacia AU 1054
72	KT71_15846	88703340	Congregibacter litoralis KT71
73	atzB	3688614	Burkholderia pseudomallei 1710b
74	Bpse112_010100013174	167911441	Burkholderia pseudomallei 112
75	GBP346_A2491	7888169	Burkholderia pseudomallei MSHR346
76	Bpse17_030300030822	161374040	Burkholderia pseudomallei 1710a
78	BPSL2112	3092957	Burkholderia pseudomallei K96243
79	BURPS1106A_2422	4901051	Burkholderia pseudomallei 1106a
80	PputW619_3055	6112005	Pseudomonas putida W619
81	Bpse7_010100013763	167894812	Burkholderia pseudomallei 7894
82	BpseN_010100013161	167903203	Burkholderia pseudomallei NCTC 13177
83	Bamb_1943	4310181	Buckholderia ambifaria AMMD
84	Bpse110_020300007234	161355267	Burkholderia pseudomallei 1106b
85	pC2C203U28_p100	7636495	Clostridium botulinum
86	BamIOP4010DRAFT_5246	170702284	Burkholderia ambifaria IOP40-10
87	BamMEX5DRAFT_6744	171322619	Burkholderia ambifaria MEX-5
88	CSBG_01722	7864561	Clostridium sp. 7_2_43FAA
89	BamMC406_1870	6178204	Burkholderia ambifaria MC40-6
90	BURMUCGD1_1300	221215518	Burkholderia multivorans CGD1
91	Bcep1808_1882	4953605	Burkholderia vietnamiensis G4
92	PputGB1_3262	5871047	Pseudomonas putida GB-1
93	PP_2584	1045965	Pseudomonas putida KT2440
94	NT01EI_2443	7959175	Edwardsiella ictaluri 93-146
95	Pput_3132	5195102	Pseudomonas putida F1
96	Daci_1591	5747148	Delftia acidovorans SPH-1
97	PB2503_11584	84704671	Parvularcula bermudensis HTCC2503
98	BURMUCGD2M_1761	221198260	Burkholderia multivorans CGD2M
99	Noca_1626	4600005	Nocardioides sp. JS614
100	Bmul_1313	5764920	Burkholderia multivorans ATCC 17616
101	mma_0509	5352721	Janthinobacterium Sp. Marseille
102	Mpop_1247	6313306	Methylobacterium populi BJ001
103	CLL_A0628	6293100	Clostridium botulinum B str. Eklund

			17B
104	Mchl_2080	7115864	Methylobacterium chloromethanicum CM4
105	BoklC_010100010767	167570319	Burkholderia oklahomensis C6786
106	Pden_3916	4582467	Paracoccus denitrificans PD1222
107	Xaut_2741	5424708	Xanthobactor autotrophicus Py2
108	CLH_0589	6317741	Clostridium botulinum E3 str. Alaska E43
109	Rxyl_2838	4117689	Rubrobacter xylanophilus DSM 9941
110	Franean1_5674	5674001	Frankia sp. EAN1pec
111	Mrad2831_6206	6142345	Methylobacterium radiotolerans JCM2831
112	BACCAP_04381	154500708	Bacteroides capillosus ATCC 29799
113	Tmel_1728	5296651	Thermosipho melanesiensis BI429
114	Cbei_2067	5293279	Clostridium beijerinckii NCIMB 8052
115	FRAAL1474	4235777	Frankia alni ACN14a
116	h16_A1363	4249475	Ralstonia eutropha H16
117	Pmob_1736	5757974	Petrotoga mobilis SJ95
118	RALTA_A1283	6453012	Cupriavidus taiwanensis
119	RCCS2_14544	126735104	Roseobacter sp. CCS2
120	Mpe_A0782	4784170	Methylibium petroleiphilum PM1
121	CtesDRAFT_PD0564	221065227	Comamonas testosteroni KF-1
122	Lcho_1140	6163813	Leptothrix cholodnii SP-6
123	RB2150_16594	126727973	Rhodobacterales bacterium HTCC2150
124	Cbac1_010100016929	224388129	Clostridiales bacterium 1_7_47_FAA
125	R2601_20986	114762995	Roseovarius sp. HTCC2601
126	THA_25	7070815	Thermosipho africanus TCF52B
127	S7335_3628	7570418	Synechococcus sp. PCC 7335
128	HneapDRAFT_0499	225303291	Halothiobacillus neapolitanus c2
129	VDG1235_1698	7583649	Verrucomicrobiae bacterium DG1235
130	ANACOL_00024	167768708	Anaerotruncus colihominis DSM 17241
131	CLOSTASPAR_03878	225390128	Clostridium asparagiforme DSM 15981
132	CLOBOL_06600	160941704	Clostridium bolteae ATCC BAA-613
133	Reut_A1294	3611928	Ralstonia eutropha JMP134

134	Rpic_2259	6289356	Ralstonia pickettii 12J
135	VspiD_010100021095	171913717	Verrucomicrobi spinosum DSM 4136
136	OIHEL45_10143	163745939	Oceanibulbus indolifex HEL-45
137	MAB_2927c	5965434	Mycobacterium abscessus
138	Rpic12DDRAFT_0675	153888567	Ralstonia picketti 12D
139	Jann_1234	3933679	Jannaschia sp.CCS1
140	GobsDRAFT_06540	227405252	Geodermatophilus obscurus DSM 43160
141	atzB	7593298	Roseobacter sp. GAI101
142	SIAM614_20905	118591899	Stappia aggregata IAM 12614
143	TM1040_1906	4077403	Silicibacter sp. TM1040
144	BHWA1_02619	7667707	Brachyspira hyodysenteriae WA1
145	RB2654_08457	84686221	Rhodobacterales bacterium HTCC2654
146	BmurDRAFT_00390	227999304	Brachyspira murdochii DSM 12563
147	RR11_659	7624076	Ruegeria sp. R11
148	MSMEG_1297	4531257	Mycobacterium smegmatis str. MC2 155
149	OG2516_00245	89069700	Oceanicola granulosus HTCC2516
150	RRSL_01271	83748887	Ralstonia solanacearum UW551
151	RSMK01192	6954169	Ralstonia solanacearum Molk2
152	RB2083_1302	7579673	Rhodobacterales bacterium HTCC2083
153	SSCG_02005	6838461	Streptomyces clavuligerus
154	Magn03006515	46203056	Magnetospirillum magnetotacticum MS-1
155	RSK20926_22599	126739238	Roseobacter sp. SK209-2-6
156	RKLH11_2317	7627427	Rhodobacteraceae bacterium KLH11
157	atzB	4941115	Saccharopolyspora erythraea NRRL 2338
158	Dshi_2773	5713672	Dinoroseobacter shibae DFL 12
159	SnasDRAFT_44920	229865225	Stackebrandtia nassauensis DSM 44728
160	Bpse9_010100014316	167816318	Burkholderia pseudomallei 91
161	Gdia_3513	6976965	Gluconacetobacter diazotrophicus PA1 5
162	guaD	5789984	Gluconacetobacter diazotrophicus PA1 5
163	RGBS107_18053	163737460	Phaeobacter gallaeciensis BS107
164	OA238_3529	7523411	Octadecabacter antarcticus 238
165	RG210_13486	163740856	Phaeobacter gallaeciensis 2.10

166	RHOER0001_4250	229489548	Rhodococcus erythropolis SK121
167	EE36_11673	83944527	Sulfitobacter sp. EE-36
168	OA307_342	7603653	Octadecabacter antarcticus 307
169	SPO2957	3194390	Silicibacter pomeroyi DSS-3
170	RER_21130	7714318	Rhodococcus erythropolis PR4
171	atzB	1220960	Ralstonia solanacearum GMI1000
172	CaciDRAFT_38210	229248212	Catenulispora acidiphila DSM 44928
173	MED193_21601	86140035	Roseobacter sp. MED193
174	NAS141_16899	83955348	Sulfitobacter sp. NAS-14.1
175	SGR_1307	6209440	Streptomyces griseus subsp. Griseus NBRC 13350
176	atzB	4222059	Rhodococcus sp. RHA1
177	atzB	4196714	Roseobacter denitrificans OCh 114
178	RLO149_17318	163733433	Roseobacter litoralis Och 149
179	NamuDRAFT_10270	229222217	Nakamurella multipartita DSM 44233
180	Rmet_1186	4037987	Ralstonia metallidurans CH34
181	ROP_44280	7742160	Rhodococcus opacus B4
182	SSDG_03874	6842290	Streptomyces pristinaespiralis ATCC 25486
183	SCO6213	1101654	Streptomyces coelicolor A3 (2)
184	SAV_2017	1213459	Streptomyces avermitilis MA-4680
185	GbroDRAFT_06290	227400018	Gordonia bronchialis DSM 43247
186	SSEG_02091	6852719	Streptomyces sviceus ATCC 29083
187	Bpse14_010100013624	167739105	Burkholderia pseudomallei 14
188	RBY4I_2538	7619385	Rhodobacterales bacterium Y4I
189	SSBG_04646	6861992	Streptomyces sp. SPB74
190	OB2597_10014	84500269	Oceanicola batsensis HTCC2597
191	SrosDRAFT_80950	229857618	Streptosporangium roseum DSM 43021
192	Nfa52410	3110729	Nocardia farcinica IFM 10152
193	SSAG_01926	6859830	Streptomyces sp. Mg1
194	OM2255_10461	114769985	Rhodobacterales bacterium HTCC2255
195	MCAG_05318	238064352	Micromonospora sp. ATCC 39149
196	MMAR_2844	6227111	Mycobacterium marinum M
197	NdasDRAFT_4742	229209177	Nocardiopsis dassonvillei subsp. dassonvillei DSM 43111
198	MUL_2914	4553343	Mycobacterium ulcerans Agy99
199			
200	Krad_4281	5335917	Kineococcus radiotolerans SRA30216

201	BpseD_010100013846	167720093	Burkholderia pseudomallei DM98
202	Noca_0677	4599542	Nocardioides sp. JS614
203	BoklE_010100011296	167563133	Burkholderia oklahomensis EO147
204	R2601_07911	114767058	Roseovarius sp. HTCC2601
205	RHE_PC00089	3895310	Rhizobium etli CFN 42
206	pRL100336	116254776	Rhizobium leguminosarum bv. viciae 3841
207	HPDFL43_14672	163758676	Hoeflea phototrophica DFL-43
208	smzB	111183223	Herbaspirillum sp. B601
209	atzB	55416015	beta proteobacterium CDB21
210	Rleg2_4464	6977558	Rhizobium leguminosarum bv. Trifolii WSM2304
211	Magn03006970	23012454	Magnetospirillum magnetotacticum MS-1
212	Cbac1_010100004932	224385745	Clostridiales bacterium 1_7_47_FAA
213	CLOBOL_05339	160940449	Clostridium bolteae ATCC BAA-613
214	trzN	156257442	Arthrobacter sp. AD26
215	trzN	17432321	Nocardioides sp. C190
216	AtzB	117583158	Arthrobacter sp. AD25
227	AAur_pTC10087	119952514	Arthrobacter aurescens TC1

APPENDIX B

List of all genes predicted to possess isoxanthopterin deaminase activity.
*Isoxanthopterin deaminase activity was confirmed experimentally.

	Locus tag	GI	Organism
1*	NYSGSRC-9339a	145580545	Unknown environmental sample of the Sargasso Sea
2	Bxe_A2016	91783796	Burkholderia xenovorans LB400
3	BgramDRAFT_0187	170690211	Burkholderia graminis C4D1M
4	KPK_2302	206579968	Klebsiella pneumoniae 342
5	KP1_3107	238895075	Klebsiella pneumoniae NTUH-K2044
6	KPN_02032	152970581	Klebsiella pneumoniae subsp. pneumoniae MGH 78578
7	HMPREF0023_2011	226953181	Acinetobacter sp. ATCC 27244
8	ACIAD1467	50084641	Acinetobacter sp. ADP1
9	PputW619_2587	170721762	Pseudomonas putida W619
10	PP_3209	26989928	Pseudomonas putida KT2440
11	Bmul_3605	161520154	Burkholderia multivorans ATCC 17616
12	BURMUCGD1_3770	221209432	Burkholderia multivorans CGD1
13	H16_A3380	113869328	Ralstonia eutropha H16
14	Bcep18194_B0632	78061482	Burkholderia sp. 383
15	Reut_A3081	73542765	Ralstonia eutropha JMP134
16	BCAM2272	206564108	Burkholderia cenocepacia J2315
17	BBta_1262	148252823	Bradyrhizobium sp. BTAi1
18	BURMUCGD2M_4109	221200399	Burkholderia multivorans CGD2M
19	Pfl01_3434	77459656	Pseudomonas fluorescens Pf0-1
20	Bphyt_7136	187921762	Burkholderia phytofirmans PsJN
21	RALTA_A2835	194290915	Cupriavidus taiwanensis
22	SIAM614_03533	118593746	Stappia aggregata IAM 12614
23	Mmwyl1_4101	152998097	Marinomonas sp. MWYL1`
24*	NYSGXRC-9236b	44611670	Unknown environmental sample of the Sargasso Sea

VITA

Name: Richard Stuart Hall

Address: Department of Chemistry
Texas A&M University
P.O. Box 30012
College Station, TX 77842

Email Address: hallrich11@yahoo.com

Education: Texas A&M University
College Station, Texas
Ph.D., Chemistry
December 2009

The University of Arizona
Tucson, Arizona
B.S., Chemistry
May 2003