FUNCTIONAL EXPLORATION AND CHARACTERIZATION OF THE DEAMINASES OF COG0402

A Dissertation

by

DANIEL STEPHEN HITCHCOCK

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

DOCTOR OF PHILOSOPHY

Chair of Committee, Frank M. Raushel
Committee Members, Tadhg Begley
Vishal Gohil
Paul Straight
Head of Department, Gregory Reinhart

May 2014

Major Subject: Biochemistry

ABSTRACT

High throughput sequencing technology and availability of this information has changed the way enzyme families can be studied. Sequence information from large public databases such as GenBank and UniProtKB can easily retrieved for the purpose of identifying unique enzymatic activities. The strategy adopted for this study is to identify characterized enzymes and the sequence features which give rise to their substrate specificity. Homologues of these enzymes are retrieved, and any active site variations can be readily identified. Cluster of Orthologous Groups (cog) 0402 is a family of enzymes which comprise a portion of the amidohydrolase superfamily. This group catalyzes a deamination reaction, releasing free ammonia and replacing it with a tautomerized oxygen. Cog0402 is most well known for guanine and cytosine deaminase, however other functions exist. One such function was that of *S*-adenosylhomocysteine deaminase, which was related to a large group of uncharacterized enzymes. These enzymes were predicted by us to deaminate 5'-modified adenosines. The enzymes were physically characterized these predictions were confirmed and a 5'-deoxyadenosine deaminase was discovered in addition to an 8-oxoadenine deaminase. During this study it was noted that background isoguanine deaminase activity was found at appreciable rates in *E. coli*. This activity was purified and identified using nanoLC-MS/MS and found to be caused by *E. coli* cytosine

deaminase. *E. coli* cytosine deaminase itself is found in a cluster of uncharacterized enzymes with a single amino acid difference in the active site. Representative enzymes were purified and a 5-methylcytosine deaminase was discovered. This enzyme is capable of rescuing thymine auxotrophs in the presence of 5-methylcytosine, and will confer sensitivity to 5-fluorocytosine. Finally, an enzyme distantly related to cytosine deaminase was purified and found to be a unique pterin deaminase. It was most efficient for oxidized pterin rings and would accept a variety of substituents on the C6 positions. Futhermore, it was thought to catalyze the first step of an undescribed pterin degradation pathway.

DEDICATION

To my Dad, who always inspired me to approach questions rationally and scientifically. To my Mom, who provided me with the unconditional love that only a mother could show. To my Brothers Bill and Tommy who gave me friendship. And to my Neice and Nephew, whose curiosity of the world around them reminds me why I became interested in science.

ACKNOWLEDGEMENTS

This work would not have been possible without the help of my advisor, my committee, my collaborators, my lab mates, and the various faculty members who used their time to help me. Dr. Raushel has provided me with all the tools guidance and encouragement I could use in a research lab. He provided me clear goals on what to do, however always was open to new ideas and gave me the creative freedom to explore projects which interested me the most. Both Dr. Manson and Dr. Dangott provided technical advice which made my projects possible. Dr. Gohil was very generous in using his contacts and assisting me with continuing my career after I finish at Texas A&M. My lab members were always friendly, provided me with advice and were particularly patient whenever my experiments were not cleaned up in a timely manner.

Also, a special acknowledgement goes to the Texas A&M Triathlon team. The group always provided me with friendship, a hobby outside of work, and helped me feel like an athlete again at times when when my days were spent sitting in lab. I gained confidence and physical health by racing which carried into my research, and my time in graduate school was enriched by the youthful energetic students on the team.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION: COG0402 OF THE AMIDOHYDROLASE SUPERFAMILY, NON-AMIDOHYDROLASE DEAMINASES AND THE AMIDOHYDROLASES OF HUMANS

The amidohydrolase superfamily (AHS) is an extraordinary set of enzymes, evolutionarily programmed to carry out various hydrolase activities (1,2). A $(\beta/\alpha)_8$ TIM barrel forms a scaffold for a metal center, poising a nucleophilic water for attack. While there are several variations on this catalytic theme, this dissertation focuses on research within a deaminase subgroup of the AHS known as cluster of orthologous groups 0402 (cog0402).

Cog0402 is a large group of enzymes, containing over 4000 members. The active site consists of a single divalent transition metal in its active site in a trigonal bipyramidyl coordination (**Figure 1.1**) (3-6). The TIM barrel scaffold can be used to easily pinpoint the location of certain active residues by referencing on which TIM barrel beta strand the residues reside. For example the defining features of cog0402 are a nearly universally conserved beta-strand 1 HxH motif, a strand 5 HxxE, a strand 6 His and a strand 8 Asp. The strand 5 Glu acts as a proton shuttle in the chemical mechanism while the remaining residues (and water) form the ligands to the metal center (7). This system catalyzes the deamination of an imine functional group, releasing free ammonia and substituting a tautomerized carbonyl group (Scheme 1). In all but one case, the imine functional group is found in an aromatic heterocyclic ring.

Because of the large size of this group, measures are taken to divide the cluster into smaller, manageable pieces. To accomplish this, we construct a sequence similarity network diagram with the software BLAST (8) and Cytoscape (9,10). In this diagram, the stringency can be changed to demonstrate how groups may cluster together (less stringent) or how they might separate (more stringent) based on phylogeny. Groups were originally defined at an E-value stringency of 10^{-70}, and the diagram is rendered at 10^{-50} to demonstrate clustering. The groups were arbitrarily numbered, and functions indicated (**Figure 1.2**).

The first known activities before cog0402 was defined were guanine (11-13) and cytosine deaminases (3,7,14). Both enzymes perform important functions in nucleotide metabolism. The sequences and structures of homologues have been solved(3,7) (guanine deaminase PDB: 2UZ9), showing the conserved structural nature of the active site (**Figure 1.3**). More importantly, this also displays the substrate binding of the imine functional group which will become important for function discovery. Closely related to cytosine deaminase was that of a creatinine deaminase (15). Additionally, enzymes closely related to cytosine deaminase and guanine deaminase were known to deaminate the atrazine intermediates *N*-isopropylammelide (*N*-isopropylammelide isopropylaminohydrolase, AtzC), melamine, and act as a chlorohydrolase for atrazine (16-18).

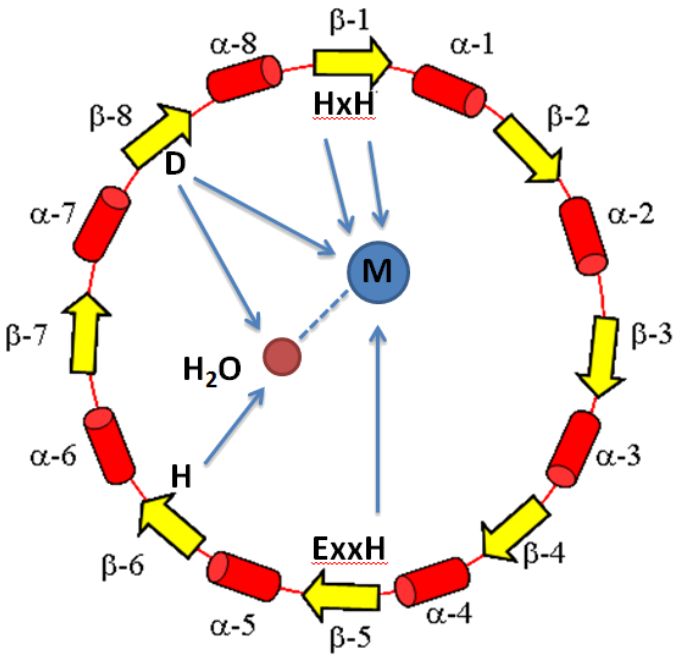
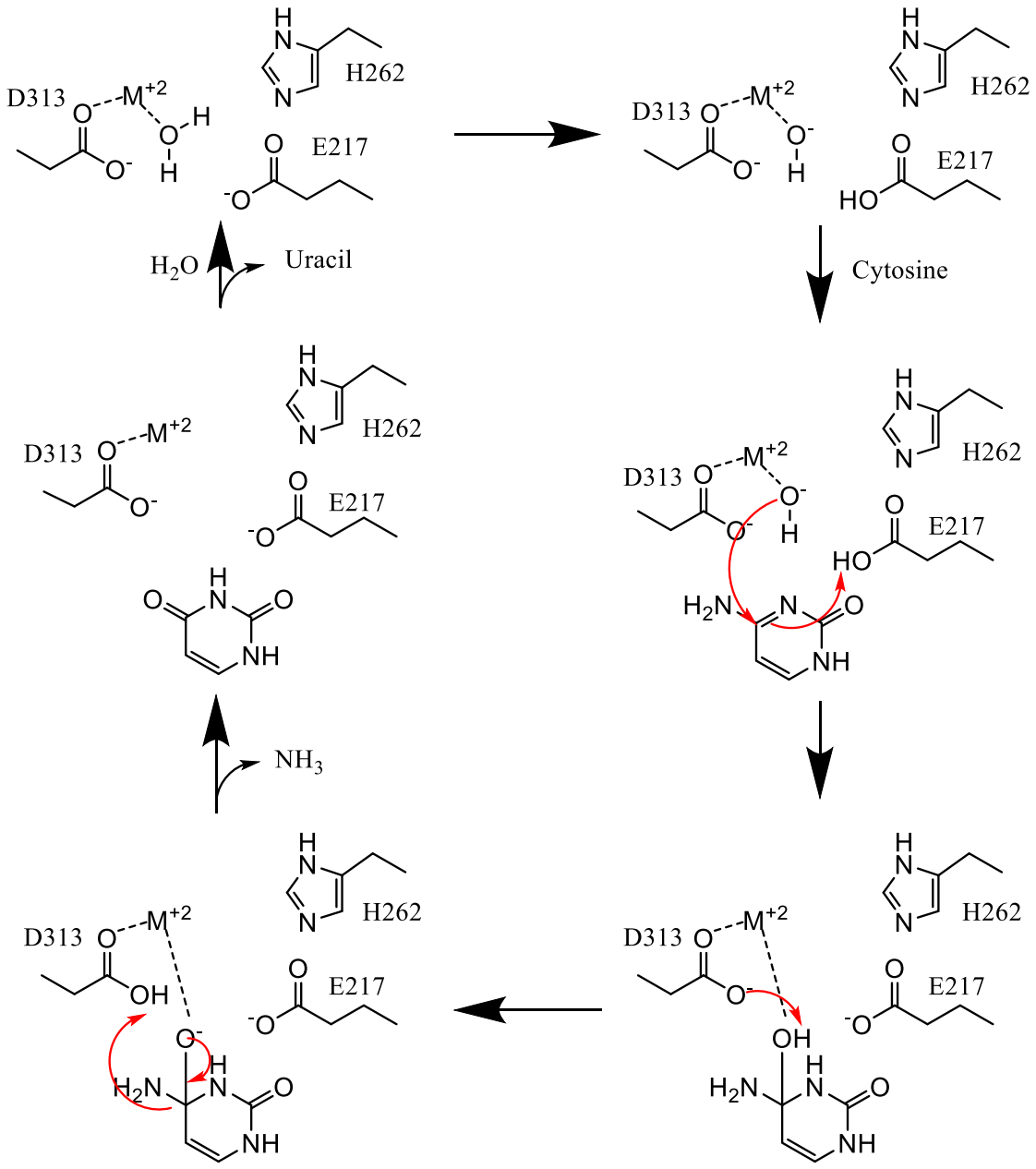


Figure 1.1: The active site of members of cog0402. Various active site residues are located on the N-termini of the beta strands which form the TIM barrel.



Scheme 1.1

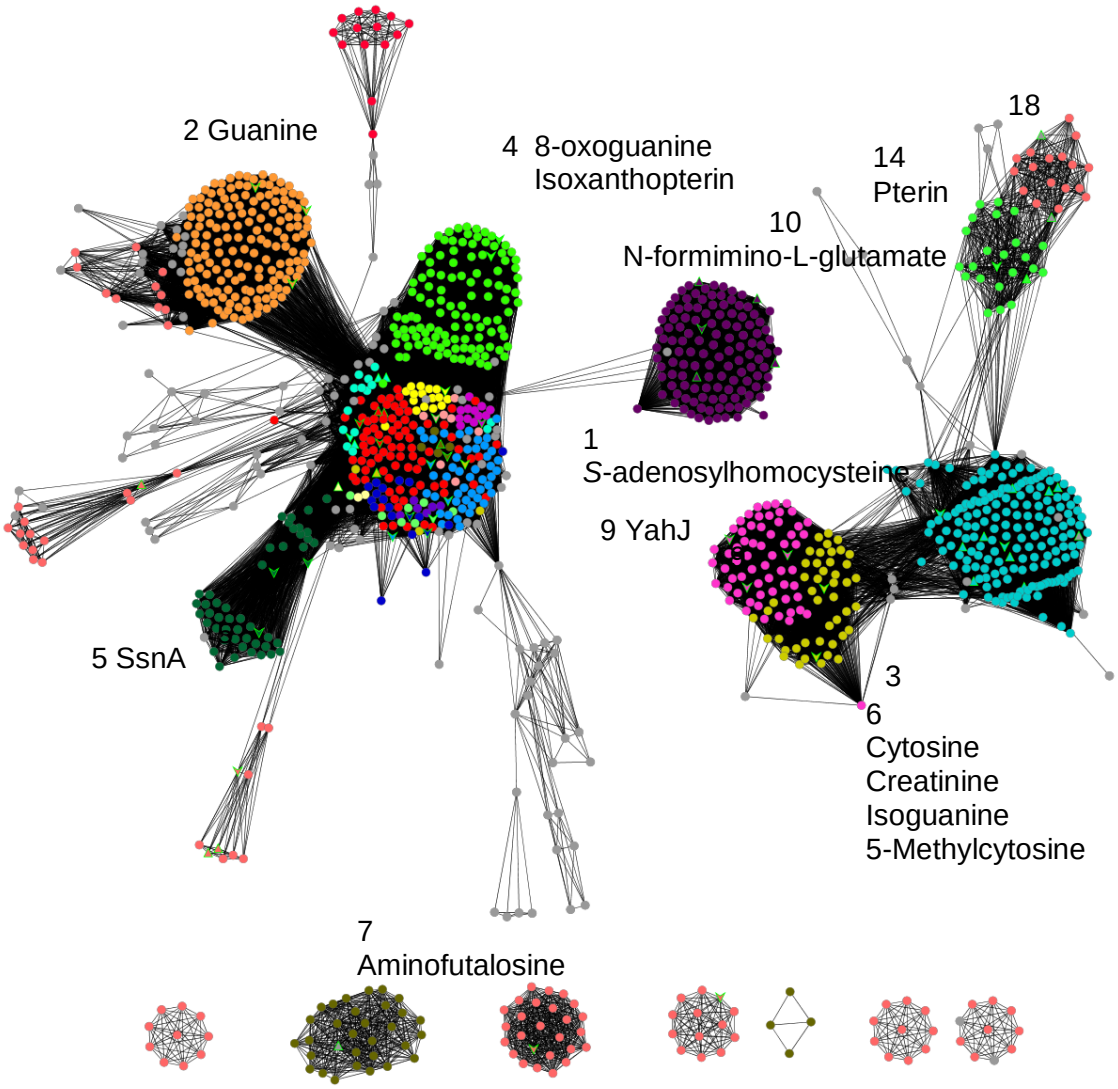


Figure 1.2: Cog0402 rendered at an E-value of 10^{-50}. Groups are arbitrarily numbered and functions are indicated.

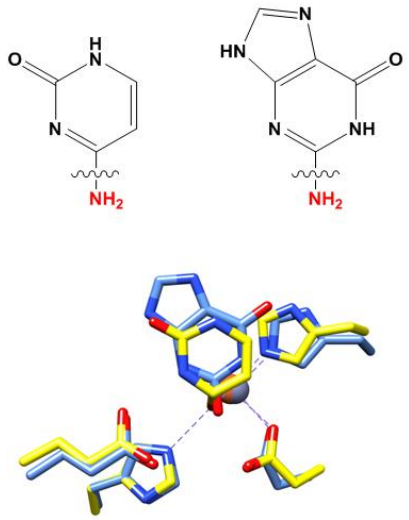


Figure 1.3: The active site overlay of cytosine and guanine deaminase. Due to the conserved nature of the active site, the imine functional group binds in an identical manner, aiding with function discovery.

Function discovery in the in cog0402 began with two enzymes, HutF and SAH deaminase. HutF was found in what is now labeled Group 10, and its genomic context suggested there to be a missing step of histidine metabolism (6). This was tested, and the substrate was found to be *N*-formimino-*L*-glutamate, forming ammonia in *N*-formyl-*L*-glutamate upon deimination. Next, SAH deaminase was discovered to be the function of a gene Tm0936 from *Thermatoga maritime* (19). This discovery was accomplished through genomic context, fragments bound in the active site structure and computational docking. Tm0936 was the first enzyme found to deaminate the byproduct of SAM mediated methylation and radical reactions, and was in fact promiscuous for 5'-deoxy-5'-methylthioadenosine.

The final two discoveries prior to the research in this manuscript were 8-oxoguanine deaminase (4) and isoxanthopterin deaminase (5). Guanine is often oxidized in DNA under oxidative stress (20), and can be removed by base excision repair (21). Once this takes place, free 8-oxogaunine can be deaminated to form uric acid, which in turn is recycled by the cell. A deaminase for this reaction was discovered active sequence differences compared to guanine deaminase as well as and library screening. The closely related isoxanthopterin deaminase was identified by active site variations from the newly discovered 8-oxoguanine deaminase.

Concurrently with this study, another function was established in menaquinone biosynthesis, aminofutalosine deaminase (22). Two members in cog0402 have this function, even though they are distantly related. One established the function of group 7, while the other only consisted of a small unlabeled group. Even with these numerous activities, still many genes remained unannotated and some groups in cog0402 had not been investigated to have promiscuous functions. The chapters in this study cover the newly discovered functions in cog0402, which will be briefly discussed in the following paragraphs.

The first discovery to be discussed (Chapter II) involved the dissection of Group 1, where the activities of 5'-deoxyadenosine deaminase and 8-oxoadenine were discovered, along with new isoforms of adenosine and guanine deaminase (23). Only two enzymes had been characterized from this group, deaminating SAM-mediated byproducts 5'-deoxy-5'-methylthioadenosine and *S*-adenosylhomocysteine. Chapter III

will discuss is the promiscuous activity of *E. coli* cytosine deaminase to deaminate 8-oxoadenine (isoguanine), an oxidized form of adenine (24). This was found during attempts to characterize group one proteins as background activity of *E. coli*. The structure was elucidated which provided insight into how isoguanine can effectively bind to the active site of cytosine deaminase. The subject of Chapter IV are the homologues of *E. coli* cytosine deaminase how they were investigated, and a new 5-methylcytosine deaminase was discovered. The active site differed at a single amino acid found directly after the strand 8 aspartate. This residue also matched one which had been mutated by other researchers to confer 5-fluorocytosine deaminase activity (25). Chapter V discussesthe discovery of a promiscuous pterin deaminase in group 14 (26). This enzyme shares a similar active site to *E. coli* cytosine deaminase. It is thought to initiate pterin ring degradation for photosynthetic organisms and plant symbiont and possibly forms part of a pathway for metabolizing oxidized pterin rings. Finally, Chapter VI will mention the attempts to functionally characterize *E. coli* genes YahJ and SsnA, two amidohydrolases of cog0402 with no known activity.

Several other deaminases exist as well, outside the realm of the amidohydrolase superfamily. Most non-amidohydrolase deaminases fall into what is known as the Cytidine deaminase-like superfamily. Members of this group includes cog0295 (cytidine deaminase), cog2131 (dCMP deaminase), cog0590 (tRNA adenosine deaminase, guanine deaminase, cytosine deaminase), cog0117 (RibG). Cog0717 (dCTP deaminase), while a deaminase, does not share the same active site machinery as the other families.

The APOBEC family also shares similarity with the cytidine deaminase superfamily, however this large group only has one catalytic function (deamination of DNA/RNA cytidine) even though there is incredible sequence diversity. APOBEC will not be discussed in the context of unknown functions for this reason.

Using the strategies discussed in the original research, it is possible to make a functional assessment of the uncharacterized members of these groups. It is likely that several unique functions exist in sequence homologues of the aforementioned deaminases. Presented is an overview of the groups with an assessment of these uncharacterized members.

Cog0295 houses Zn dependent deaminases which hydrolytically deaminate cytidine/2'-deoxycytidine or the antibiotic blasticidin-S. A sequence similarity network diagram was constructed by retrieving from NCBI the sequences annotated as "cog0295." The network diagram rendered at the E-values 10^{-30} are shown in **Figure 1.4**. The members share in common a zinc binding domain, although variations on this theme exist and are described below. There are three well known activities in cog0295, each which populate a separate subgroup in the network diagram. There are homodimeric CDA/2dCDA (HCdd), homotetrameric CDA/2dCDA(TCdd) and Blasticidin-S deaminases. An active site diagram of the 3 known functions is shown in **Figure 1.5**.

Homodimeric ccd forms a small cluster of 141 members in cog0295, forming Group 1. The group is mostly populated by gamma-proteobacteria, and includes representatives from Enterobacteria, Vibrionales, Pasteuralles and Alteromonadales.

Certainly the most studied cytidine deaminase from this group is that of *E. coli*, originally described in 1971 (27). Here *E. coli* cdd was purified by monitoring catalytic activity, and the enzyme was found to be inhibited by 3,4,5,6-tetrahydrouridine, which forms a transition state analogue. A purification in 1984 suggested that ccd formed a functional dimer *in vitro* (28), and this was confirmed in 1985 by crosslinking experiments (29). Furthermore, *E. coli* cdd was purified to homogeneity and found to efficiently deaminate cytidine, 2'-deoxycytidine and 5-methylcytidine (29). The structure was solved in 1994 (30), which revealed the nature of its active site.

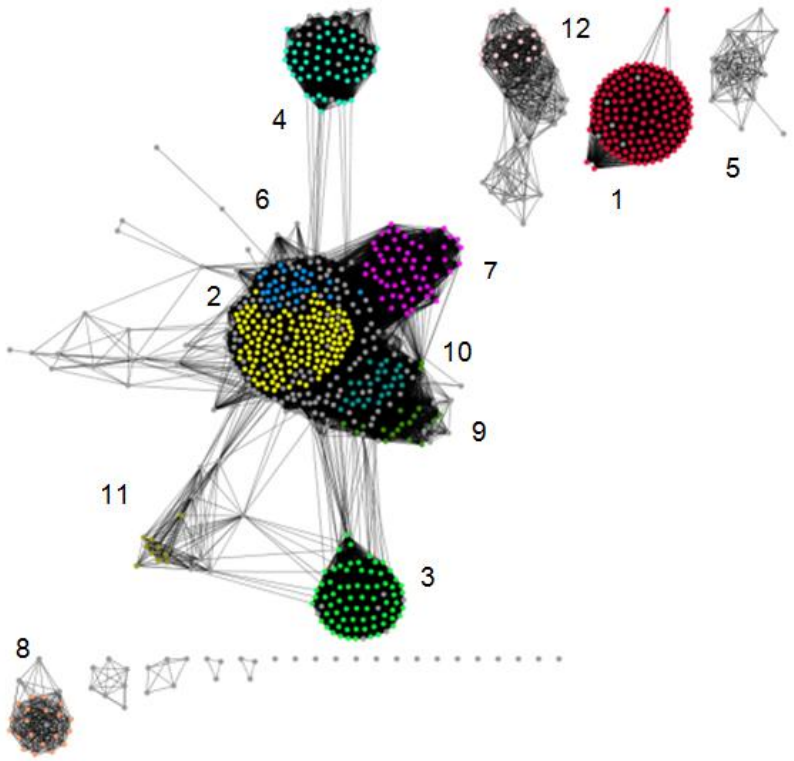


Figure 1.4: Network diagram of members of cog0295 rendered at a BLAST E-value of 10^{-30}.

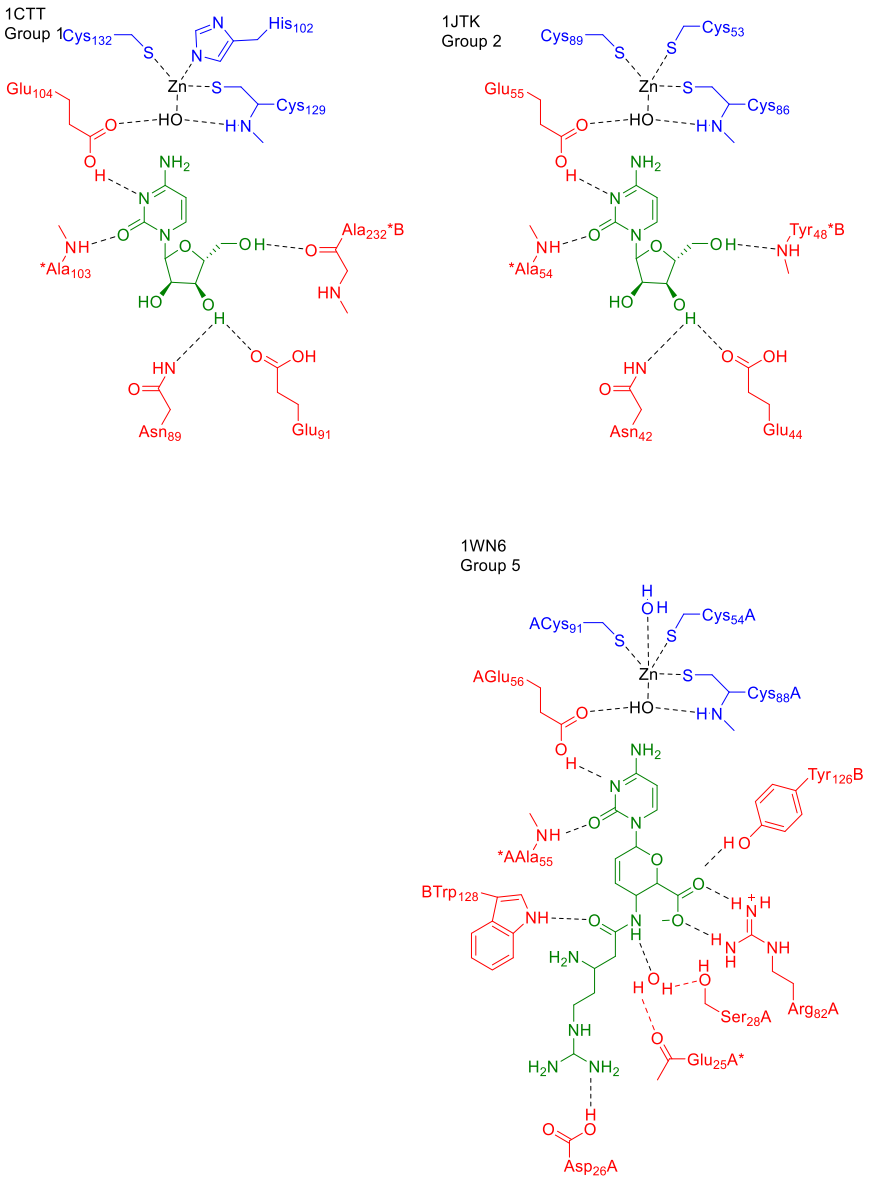


Figure 1.5: The active site schematics of various members of cog0295. Metal binding residues are in blue, the substrate in green and the ligand binding residues/catalytic machinery is in red.

A later structure was deposited in the PBD (PBD: 1CTT), where the enzyme was crystallized in the presence of a pro-inhibitor, zebularine, which is hydrated by the enzyme to form zebularine 3,4-hydrate (ZEB—H2O) (31). A mononuclear metal center is formed by Cys132, Cys139, and His102. A fourth ligand is formed as an activated water to complete the tetrahedral geometry. Cytidine is held in the active site by several interactions – O2 on the pyrimidine hydrogen bonds to the amide backbone on A103, the 3' hydroxyl binds to both Gln 89 and Glu 91, and the 5' hydroxyl binds across the dimer interface to the carbonyl backbone of Ala232. The mechanism is thought to occur through a hydrophilic attack using an activated hydroxyl ion bound to Zn^{2+} and Glu104, which acts as a proton shuttle (32). This was determined by solving the structure with various transition state inhibitors.

Homotetrameric CDAs comprise most of cog0295, and several organisms from various domains of life have a homologue. The characterized members form a central group which was labeled Group 2. Most representatives from this group are from Bacillales, and lactobacillales, however mouse (PDB: 1ZAB) (33), yeast (PDB: 1R5T) (34), and human (PDB: 1MQ0) (35) possess homologues. The first representative discovered was *Bacillus subtilis* in 1989 (36). Here the 14 kDa enzyme described was cloned, found to form a homotetramer, and is capable of deaminating cytidine and 2'-deoxycytidine. 5-Methylcytidine was not tested as a substrate. The structure was solved in 2002 (PDB: 1JTK) (37) and the active site as well as the tetrameric interface were described. The Zn^{2+} coordinating ligands differ only by one residue, where the corresponding ligand to

H102 in DCdd is a Cys. The binding of the substrate is similar to dCDA; O2 on the pyrimidine is hydrogen bonded to the amide backbone of Ala54, the 3'-hydroxyl bonds to Glu44 and Asn42, however the 5'-hydroxyl hydrogen bonds to the amide backbone of Tyr48 across the symmetric unit of the tetramer interface. The mechanism of TCDA is believed to occur in a similar manner to DCDA, and the additional charge of the 3xCys (vs Cys-Cys-His in DCDA) binding domain is thought to the neutralized by helix dipoles. A later study showed that Arg56, located on the active site helix, neutralized the charge of the Zn-coordinate cysteines and conferred tetramer stability (38). A functional outlier to this group is the yeast TCDA, which was found have mRNA editing activity of C→U (34), and shared the same substrates as the APOBEC-1 family. Although the structure was solved of the yeast enzyme, it remained unclear how mRNA could bind to the active site. Moreover, mouse TCDA was found to have mRNA binding capacity (33).

Blasticidin has been used as an antibiotic for rice since 1958 (39) and a deaminase for this compound was found in 1975 (40). The corresponding deaminase from *Aspergillus terreus* was found to be 193 amino acids and can confer Blasticidin S resistance as a selectable marker in eukaryotes (41). It forms a functional tetramer, like TCDAs (42) and was originally postulated to share a similar active site. However, its structure was solved in 2007 (43) which showed the free enzyme adopted a tetrahedral metal geometry, and shifted to trigonal bipyramidyl upon substrate binding. A schematic of the active site can be seen in **Figure 1.5**. Together, these enzymes populate group 5.

Groups 3, 4, 9, 10, 11 appear to be tetrameric cytidine deaminases. A member from each group was selected for a sequence alignment, and all contain similar residues to those found in Group 2. There are two minor differences. Groups 9 and 12 have a Phe at position 48 as opposed to a tyrosine. Since this is a backbone interaction, this substitution is expected to have little effect. Secondly, group 12 contains a histidine for the zinc binding, like the *E. coli* protein. However a lack of the functionally mysterious C-terminus (as in group 1) leads one to believe group 12 will be tetrameric.

Groups 6 and 8 both have slight variations and may contain new functions. Compared to group 2, both groups lack Asn42 (Cys in group 6, Ala in group 8) and Tyr48 (Gly in group 6, Ala in group 8), however group 6 possesses the other binding residues. Group 8 lacks Glu44, however it is replaced with an Asp at this position which may fill a similar role. Also, the Ala54 has been substituted with a Met; but given the backbone interaction utilized, this may have no effect. Given the intact active site and varied substrate binding residues, these groups may catalyze an undescribed function.

Cog0717 is comprised of a mix between dCTP deaminases, dUTP nucleotidohydrolase (dUTPase) and both a combination of these activities to form bifunctional enzymes. These enzymes provide dUMP for dTMP synthesis while minimizing the dUTP pool. The deaminases are unique in that no metal center is required for deamination, only magnesium which is found bound to the triphosphate substrate. Several structures have been solved to elucidate the catalytic mechanism for both of these functions, allowing functional predictions of enzymes to be made. A

sequence similarity diagram was constructed at a BLAST E-value of 10^{-30} and its groups were arbitrarily labeled 1-7. This diagram is rendered at 10^{-20} and 10^{-5} in **Figure 1.6**.

6.

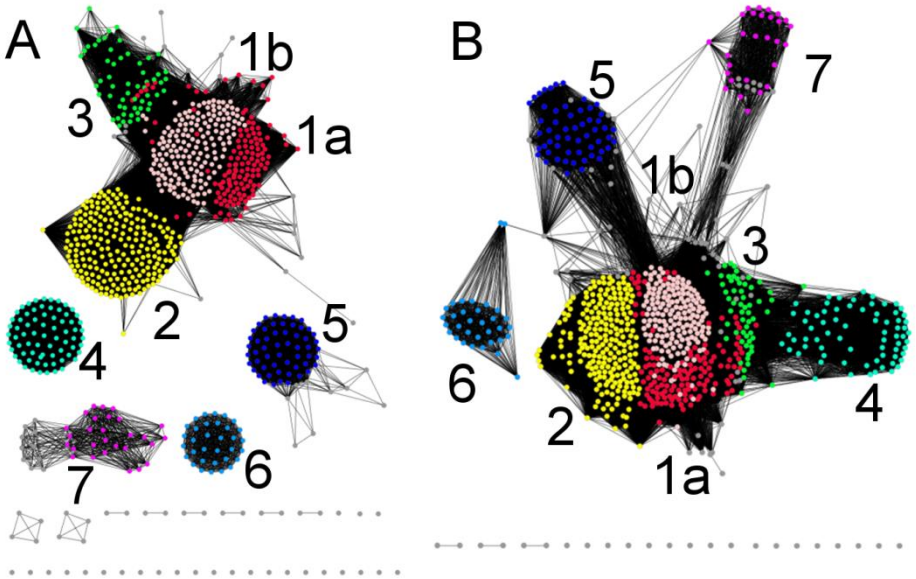


Figure 1.6: Network diagrams of cog0717.

In Group 1a, *E. coli* dCTP deaminase has been extensively studied and both its structure and mechanism have been deduced (44-46). The crystal structure revealed a homotrimer, forming in total 3 active sites, each occurring at the interface of two subunits. 10 sidechains are found to be responsible for substrate binding and catalysis (45), and specifically 3 residues are found to catalyze the reaction (46). Glu-138 deprotonates a catalytic water, which is stabilized by the polar hydrogen on Ser-111, which in turn is hydrogen bonded to R115. The resulting hydroxyl ion performs a

nucleophilic attack on the substrate, resulting in the release of ammonia. This enzyme is unable to perform dUTPase activity, which can be explained by the lack of a catalytic glutamine found in Group 7 *Mycobacterium tuberculousis* (PDB: 1SIX) (47) as well as *E. coli* dUTPase (PDB: 1SEH) (48). The designation of Group1a is defined as lacking this catalytic glutamine. Other group1 members which populate Group1b have been characterized as a bifunctional dCTP deaminase/dUTPase. Such examples include *Ethanocaldococcuss* (49) and *Methanocaldococcus jannaschii* (50,51). Deamination of dCTP is accomplished in an identical manner to the Group1a proteins.

Given these enzymes, it becomes possible to make substrate predictions for groups lacking functional annotations. Sequence alignments were constructed for each group, and conserved residues were compared with *E. coli* dCTP deaminase (S111, R115, E138, Q182, D128, S112, K178, Y171, R110, R126, PDB: 1XS1) as well as the catalytic residues for *M. tuberculosis* dUTPase (Q113, PDB: 1SIX). The active site of *E. coli* dCTP deaminase is presented in **Figure 1.7**, and the active residues of various groups in **Table 1.1**, with blue symbolizing metal binding residues, red for dUTPase residues and yellow for binding residues.

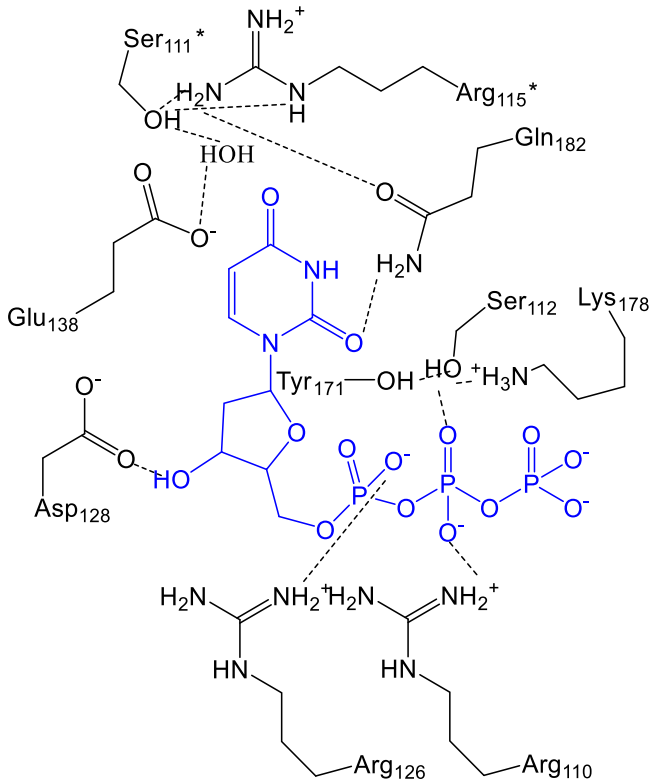


Figure 1.7: Active site of PDB: 1XS1.

Table 1.1. Active site residues of cog0717.

1A	**S111**	**R115**	**E138**	**Q182**	**D128**	**S112**	**K178**	**Y171**	**R110**	**R126**	**A**
1xs1											
1B	**S**	**R**	**E**	**Q**	**D**	**S**	**R/K**	**Y**	**R/K**	**Aro**	**Q**
2	**S**	**R**	**E**	**Q**	**E**	**T**	**K**	**Y**	**K**	**P**	**Q**
3	**S**	**R**	**E**	**N**	**gap**	**S**	**K**	**Y**	**R**	**gap**	**Q**
4	S	R	Ins	Q	D	S	N	Y	R	E	R
5	S	R	L	E	D	S	S	H	R	V	Q
6	N/T	A/E	Hypho	gap	Unc	E/D	gap	gap	R	Unc	T
7	S	K	Hypho	Unc	D	G	T/S	Unc	R	Unc	Q113 1six

Group 2 and Group 3 to our knowledge have not been experimentally verified, but reasonably can be assumed to have the function of a bifunctional dCTP deaminases/dUTPase. Group 2 contains two uncharacterized pdb entries (3KM3 and

4DHK). The catalytic machinery for both of these reactions remains intact. Substrate binding residues also are very similar. An Asn replaces Q182 for uracil binding in Group 3, but many of the phosphate binding residues are conserved compared to Group 1a and 1b. Finally, the catalytic Gln is present to perform the phosphohydrolase activity.

Group 4 also has no published experimental results, although a structure exists (PDB: 2R9Q). A sequence alignment predicts there to be a large insertion (>100AAs) centered around residues aligning to the catalytic Glu138. However the structure shows 370 AAs, whereas *E. coli* dCTP deaminase (PDB: 1XS1) is only 193 residues. A full length alignment of a duplicated 1XS1 sequence and 2R9Q reveal a 20% identity (vs. 23% for single 1XS1), suggesting Group 4 proteins may constitute a duplicated, fused version of dCTP deaminase. **Figure 1.8** shows a structural comparison between 2R9Q (red) and a 2 subunits of 1XS1 (light gray, dark gray), and the similarity is apparent. It is possible for Group 4 proteins to possess dCTP deaminase activity, but this would need to be experimentally verified.

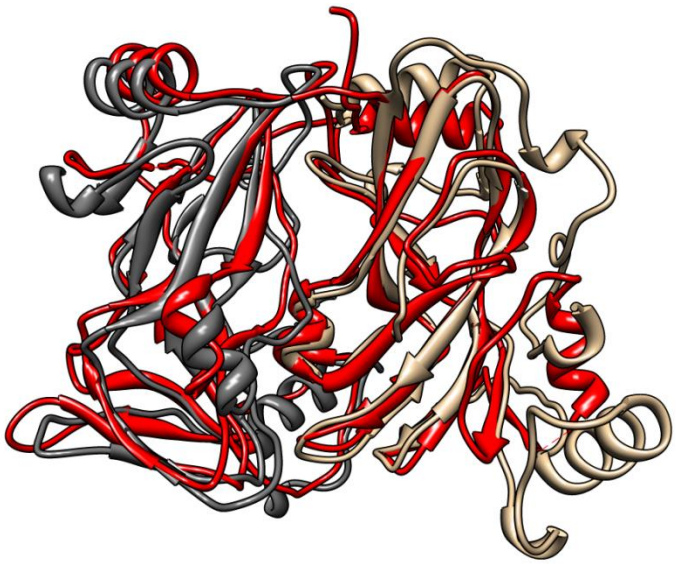


Figure 1.8: Structural comparison of a group 4 protein and a homodimer from group 1.

Group 5 shares many conserved residues with Group 1, but is unlikely to be capable of deaminase activity. The most striking feature is the lack of the catalytic Glu138, which, like Group 7, has been replaced by a hydrophobic residue. However, many substrate binding residues are intact, and most notably Group 5 shares with Group 7 a conserved Gln, which suggests these enzymes may be dUTPases. Group 6 proteins share few conserved residues with Group 1 or even Group 7, and their function is unknown.

dCMP deaminase of cog2131 is found in all domains of life, including viruses. These enzymes deaminate dCMP, forming the product dUMP which has several metabolic roles. The first activity was described by infecting *E. coli* with T2 bacteriophage, suggesting that the enzyme plays a role in viral proliferation (52). Since that discovery in 1960, other dCMPs have been studied including bacterial, mammalian

and viral. The primary structure of human dCMP deaminase was found in 1993 (53). Most members of cog2131 have an allosteric site, being activated by dCTP and inhibited by dTTP (54,55).

Three unique structures can be found in the PDB: *Streptococcus mutans* dCMP deaminase with inhibitors (PDB: 2HVW) (56) in Figure 1.9A, T4-bacteriophage (PDB: 1VQ2) (57) in Figure 1.9B and human dCMP deaminase (PDB: 2W4L). The active sites of these enzymes are common, and are accomplished through a conserved Zn^{2+} metal site. A tetrahedral Zn^{2+} ion is coordinated by a CXXC and HXE motif. The conserved glutamate does not bind to the Zn^{2+} metal ion but instead functions as a proton shuttle for the mechanism, binding to a nucleophilic water which forms the 4th ligand on the Zn^{2+} metal ion. The allosteric active site is displayed in **Figure 1.10**.

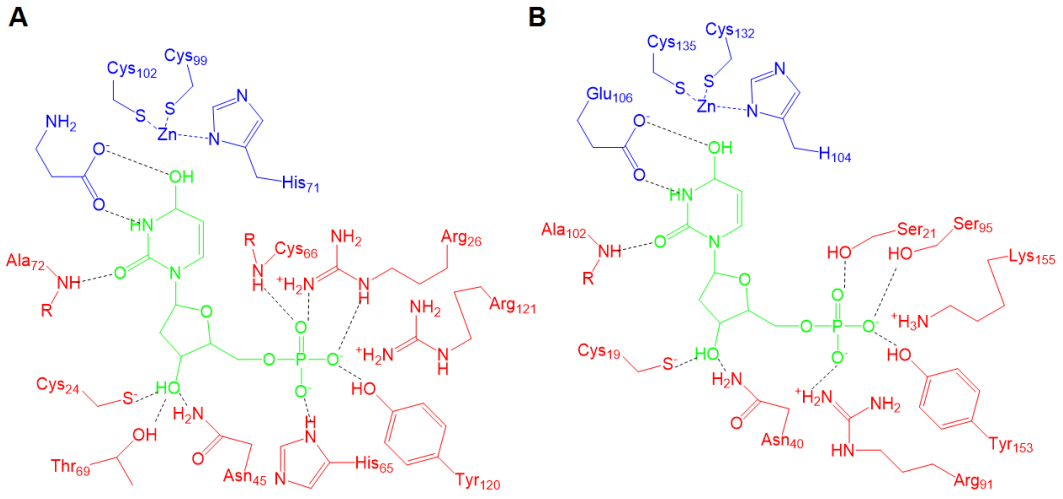


Figure 1.9: A) Active site of 2HVW B) Active site of 1VQ2.

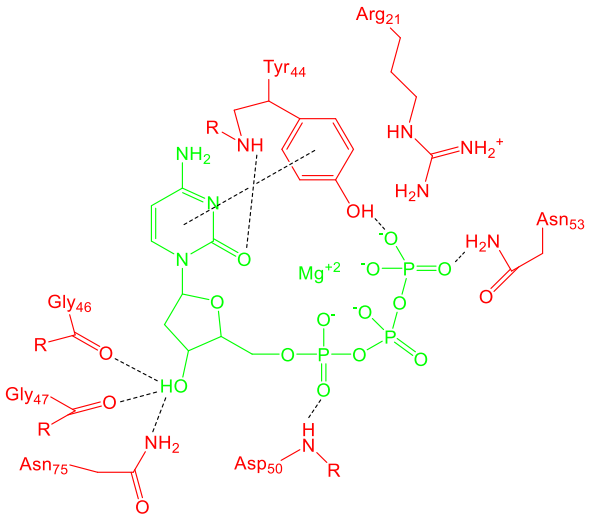


Figure 1.10: Allosteric site of 2HVW

A search in GenBank for "cog2131" retrieved 854 proteins. 99% identical redundancies were removed leaving only 440 sequences. At a stringency of 10^{-40}, the enzymes cleanly separated into 8 groups (**Figure 1.11**). The liganded active sites of 2HVW and 1VQ2 were used as a comparison for each group. **Table 1.2** shows the active site residues, and **Table 1.3** shows those ligands used for the allosteric site. The portion shown in yellow indicates mechanistic and metal binding residues.

Group 1 houses the structure of *Streptococcus mutans* (PDB: 2HVW). This group is mostly bacterial and is shown to cluster with groups 2, 3, and 6. The residues are almost universally conserved. Enzymes found in this group undoubtedly perform the same catalytic activity.

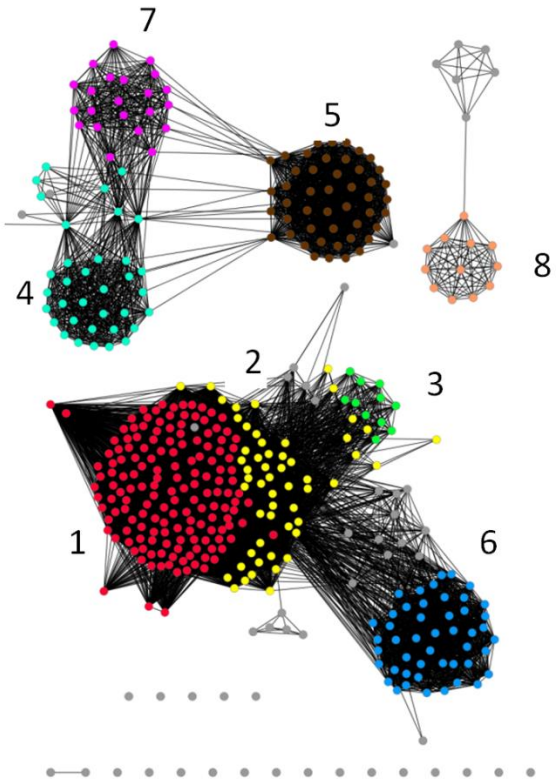


Figure 1.11: Cog2131.

Table 1.2. Active site residues of dCMP deaminase catalytic domain.

2HVW	H71	E73	C99	C102	C24	R26	N45	H65	C66	T69	A72	Y120	R121
1VQ2	H104	E106	C132	C135	C19	S21	N40	-H71	-C75	E102	A105	Y153	K155
2W4L	H	E	C	C	D	N	N	K	Y	V	A	K	Y
Group1	H	E	C	C	C	R	N	H	C	T	A	Y	R
Group2	H	E	C	C	C	R/A	N	Hdrph	C/S	A,G	A	Y/F	-
Group3	H	E	C	C	C	R	Unc	E	L	S	A	-	R
Group4	H	E	C	C	D	S,d	N	E,p,a	Y	T	A	F	D,e
Group5	H	E	C	C	D	N,h,s	N	E,D,t	Y	V,t,s	A,s	Y	E,D,unc
Group6	H	E	C	C	C	R	N	Y	V	-	A	Y	K/R
Group7	H	E	C	C	D	S,N	N	P	Y	V	A	E,k	S,e,q
Group8	H	E	C	C	C,d	S	N	-,unc	unc	unc	A	Y	Unc

Table 1.3. Amino acids aligning with allosteric domain residues.

2HVW	R21	Y44	G46	G47	D50	N53	N75
1VQ2	E	Y	G	S	G	N	N
2W4L	R	Y	G	M	G	-	N
Group1	R	Y	G	S,g	G,d,a,e	H,n	N
Group2	R	Y	G	Unc.	G,unc	H,n	N
Group3	R	Y	G	A	G	N	N
Group4	W	F	G	Y	G	-	N
Group5	R	Y	G	Hydrph	G	-	N
Group6	L/N	Y	G	T	G	N	N
Group7	F	F,Y	G	L	G	-	-
Group8	Unc	Unc	E/D	V	Unc	Unc	Unc

Group 2 shares many of the conserved residues as group 1, but seems to split into two subgroups. In one (with the equivalent to R26) we expect the substrate to be dCMP. However, a portion of this group lacks R26 (as well as C66 and Y120), which is crucial to binding the 5' phosphate group on dCMP. It is possible there may be a unique function, or perhaps a deoxycytidine deaminase hidden within this group.

Group 3 is populated by mostly clostridiales and is predicted to deaminated dCMP. Although not all residues are conserved, the arginines corresponding to R26 and R121 are.

Group 6 also clusters with group 1. This large group is made up of various bacteria like 1, 2, and 3. Much like group three, variations exist, however important residues are still conserved.

Groups 4, 5, and 7 cluster together and do not share many of the conserved active site contained in 2HVW and it is difficult to make an assessment without active site residues. They certainly possess the metal binding domain as well as the Asn corresponding to N45, but lack many others. Fortunately both groups 5 and 7 have

confirmed activities, even if the active site will remain unannotated. Group 5 houses bacterial enzymes as well as mammalian, including human dCMP deaminase and its unliganded structure (PDB: 2W4L). There is a confirmed activity in group 7, which was found to be a bifunctional dCMP/dCTP deaminase (58). This cluster is comprised primarily of chlorella virus. Group 4 (populated by gamma proteobacteria) has no confirmed activities.

Group 8 is unique and could potentially have a new or additional function. Compared to the rest of the enzymes (~130 AAs), group 8 averages over 500 AAs. The sequence corresponding to dCMP deaminase is found near the C-terminus, with the N-terminus residues showing only similarity to other members of the group. The remaining enzymes include mostly bacterial ones as well as the structure of T4-bacteriophage. Most of cog2123 seems to have the same dCMP function, however groups 2 and 8 may have new functions, and 4 remains untested. These enzymes are found in a wide variety of bacteria and should be experimentally verified.

Cog0590 houses several functions as well as uncharacterized groups. The known activities include tRNA adenosine deaminases, guanine deaminases and cytosine deaminases. Like most other non-amidohydrolase deaminases, a single Zn^{+2} ion is coordinated by 2 cysteine residues and a histidine. A glutamate (forming an EXH motif with the coordinating histidine) plays a role in the chemical mechanism. A search in GenBank revealed 2778 enzymes; after deleted those displaying 99% sequence identity,

1393 remained. The cog cleanly separates into 9 apparent groups when visualized at an E-value of 10^{-30}, shown in Figure **1.12**.

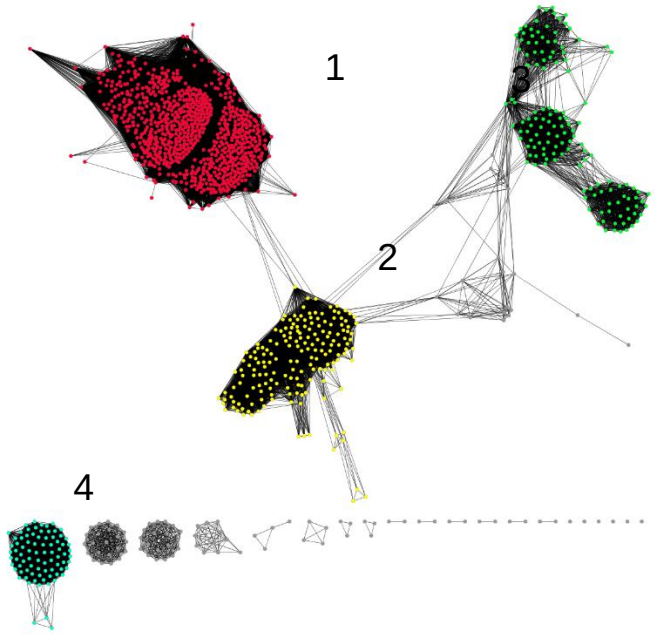


Figure 1.12: Cog0590 at 10^{-30}

The enzymes in group 1 are responsible for deaminating the wobble position on tRNA-34, postranscriptionally modifying the adenosine to an inosine. In prokaryotes, this gene is known as tadA and forms a functional homodimer (59), whereas in eukaryotes a functional heterodimer ADAT2-ADAT3 is formed (60). The biological mechanism for this reaction was elucidated, and the enzyme was found to have an exceptionally low K_m to manage the low concentrations of tRNA in the cell (61). On average there are 2500 acyl-tRNA molecules per *E. coli* cell (62). A liganded structure

was determined (PDB: 2B3J) (63) and the layout of the active site is presented in **Figure 1.13**. A sequence alignment was constructed, but the active site residues were not universally conserved across group one. The metal binding, and catalytic residues, Lys106, and the residue corresponding to Asp104 were conserved, with the exception of a group with an Asn instead of Asp104. Arg44 however show poor conservation, including a large pool of Thr. Asn similarly lacked conservation across the group. It is possible other functions exist in group 1.

Group 2 has one characterized activity, that of guanine deaminase (64). Inherently, this guanine deaminase is unrelated to the amidohydrolase guanine deaminase. There are two structures, one with a ligand of imidazole (PDB: 1WKQ). In order to determine how the liganded structure may appear, guanine was computationally modeled as an inhibitor and docked into the active site. This model is shown in **Figure 1.14**. A sequence alignment revealed that the metal binding, catalytic and those aligning to Phe26 and N42 are well conserved; however, Y156 and D114 appear in only half. If the modeled active site is accurate, there may be unique functions in group 2.

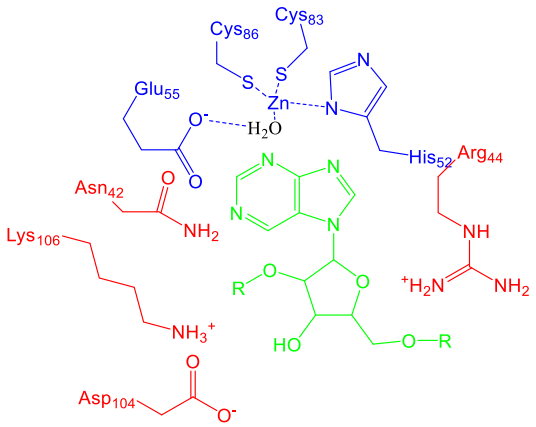


Figure 1.13: Active site shown in PBD: 2B3J

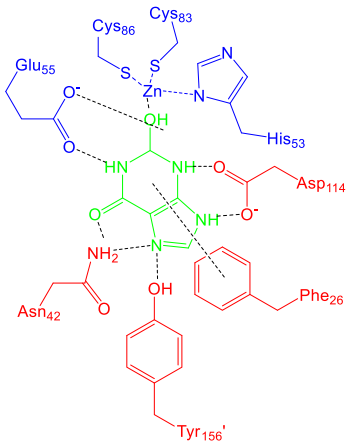


Figure 1.14: Active site of 1WKQ with modelled guanine-related inhibitor

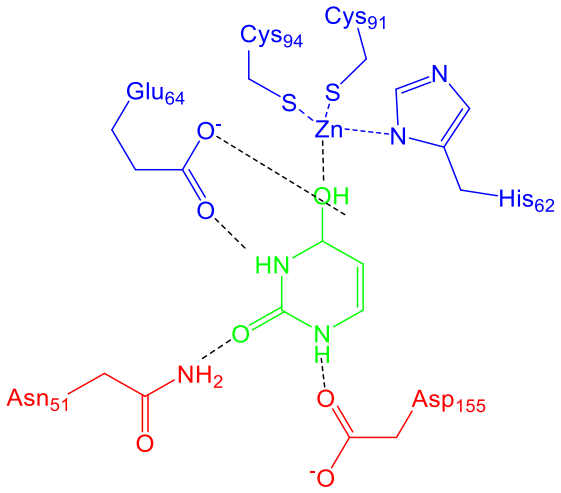


Figure 1.15: Active site of 1P6O

Group 3 contains no known functions or structures, nor do any predicted active site residues align with known functions with the exception of the metal binding domain. It is likely there are undescribed activities within this group.

Group 4 has been well characterized and has the function of cytosine deaminase. It was originally described as yeast and found to be unique compared to *E. coli* cytosine deaminase (65,66). Cytosine deaminase in addition has been studied as a potential treatment to cancer, and its liganded structure was solved to assist with directed evolution experiments (67). Shown in **Figure 1.15** is the active site of PDB: 1P6O. A sequence alignment compared to the remaining group 4 shows the conserved Asn and Asp. It is likely this group contains only cytosine deaminases.

No other functions are described in cog0590, meaning possibly additional enzymatic activities may be present in the smaller groups. Several of the larger groups certainly contain additional functions.

Cog0117 members are responsible for riboflavin biosynthesis and are known as RibD. These bifunctional enzymes catalyze the deamination of the pyrimidine and reduction of the sugar of 2,5-diamino-6-oxy-4-(5'-phosphoribosylamine)pyrimidine to form 5-amino-2,6-dioxy-4-(5'-phosphoribitylamine)pyrimidine (68,69). Compared to the other deaminase COGs mentioned, cog0117 is rather small and only seems for form 3 groups. Shown in the network diagram are 692 members labeled retrieved from GenBank as "cog0117", and with 99% identical sequences removed (**Figure 1.16**). The largest group contains several crystal structures, including that of *E. coli* RibD and the liganded *Bacillus subtilis* RibG (PDB: 4G3M) (70), which is used in this study for active site comparison.

The central cluster retains the active site residues shown in **Figure 1.17**, which the exception of arginines filling in the role of lysine. This would not be expected to affect substrate specificity. The two remaining groups share similar residues, but minor differences exist. The larger group uses a mixture of glutamate and aspartate to potentially bind to the sugar. The smaller group substitutes an arginine, much like part of the main group in this COG. Additionally, the smaller group is lacking a portion of the C-terminus, perhaps acting only as a monofunctional deaminase. It seems likely though that all members of cog0117 share the same substrate.

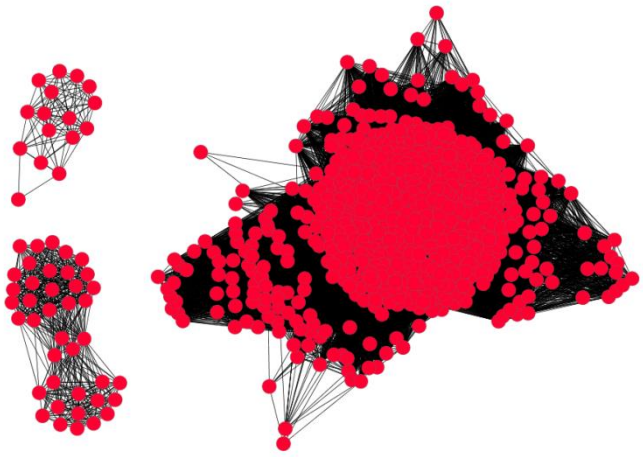


Figure 1.16: Cog0117 at 10^{-60}

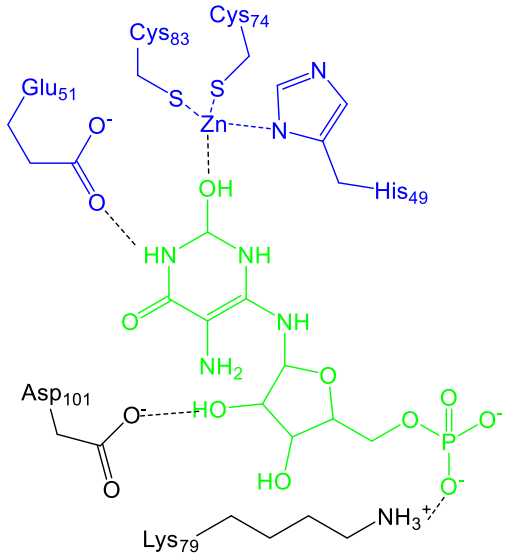


Figure 1.17: Active site of 4G3M

While the subject of this manuscript primarily discusses amidohydrolases from microbial organisms, it is important to remember the amidohydrolase spans across all domains of life, including *Homo sapiens.* Amidohydrolases in GenBank are known as

metallo-dependent hydrolases, and searching by name will reveal 97,671 sequences. The search results were filtered to those from Homo sapiens. Redundant sequences and fragments were deleted and 11 amidohydrolases were identified for humans by this method. Four more were identified by literature. It is possible there are additional sequences not represented.

The human genome contains two adenosine deaminases, both of which are characterized with crystal structures. Adenosine deaminase I is conserved in many organisms but was first detected in humans in kidney tissue (71). A recent structure was solved (PDB: 3IAR) however liganded structures of homologues (such as 91% identical bovine ADA) have been solved with ligands (PDB: 1KRM) (72). In humans, ADA-1 is associated with T-cell activation (73) and deficiency results in immunodeficiency (74). Adenosine build up causes apoptosis in lymphocytes (75), and inhibition of this enzyme with deoxycoformycin is a strategy for treating leukemia (76). ADA-2 is well characterized as well (PDB: 3LGG) (77) and was originally found to play an important role as an insect growth factor (78,79). It was discovered this was due to adenosine deaminase activity (80). In humans it is thought to induce differentiation of monocytes to macrophages in addition to aiding the proliferation of macrophages and CD4+ T cells (81).

Humans also possess three AMP deaminases (AMPD). AMPD-1 is expressed at high levels in skeletal muscle in rat (82), and in humans a deficiency leads to muscle cramps and fatigue (83). Isoform L (liver, AMPD-2) has AMPD activity (84). In

Arabidopsis this gene is known as Embryonic factor 1 and is one of the earliest expressed genes (85). The Arabidopsis gene is 51% identical to humans and its structure has been determined (PDB: 2A3L) (86). In humans this gene is not as well characterized other than its *in vitro* function, but it is known to have several different transcripts involving exon shuffling and varying termini (87). AMPD-3 is found in erthyrocytes and was cloned by cDNA human library (88). Deficiencies in this enzyme are entirely asymptomatic in humans (89).

Guanine deaminase, as previously mentioned, is found in humans (90) as well as several other organisms. This protein was primarily studied in mouse models, where its highest level of expression is in the small intestine (91). The mouse protein was expressed and purified to homogeneity (12). It is catabolic and forms an important part of the purine salvage pathway.

Humans also contain an interesting cluster of genes, including dihydropyrimidinase and 5 homologues which do not have described catalytic functions. Dihydropyrimidinase is a well characterized function involved in pyrimidine metabolism, and deficiencies in this gene can sometimes be asymptomatic or cause dihydropyrimidinuria. Several mutations have been found (92). The remaining predicted noncatalytic homologues have been phenotypically characterized, and renamed to collapsin response mediator proteins (CRMP) based on homology the chicken and rat genes (93,94). CRMP1 (PDB: 4B3Z) has been physiologically characterized to mediate signal transduction of axon guidance molecules (95) and in

mouse models deficiencies have shown to impair learning (96). CRMP2 is the most studied CRMP gene and plays a role in neuronal polarity. Its physiological function is characterized in a review (97) and its structure solved(PDB: 2GSE) (98). CRMP3 has been studied in mouse systems and found to be localized in portions of the brain involved with motor skills (99). Later this gene in mouse models was found to be important for neuronal plasticity, dendrite arborization and guide-posts navigation (100). Its structure was solved in 2013(PDB: 4BKN). CRMP4 was found to inhibit axon outgrowth, and inhibition of this protein may assist with neural regeneration after injury (101). Finally CRMP5, also known as CRMP-associate molecule (CRAM) is found to be imnportant in growth cone development (102). Additionally it regulates neural growth and plasticity in cerebellar purkinje cells (103). Its structure has been solved (PDB: 4B90) and its oligimerization with other CRMPs has been studied (104).

A gene for *N*-acetylglucosamine deacetylase is present in humans, however this gene remains uncharacterized to my knowledge. It can be found as gi:166233266 in GenBank.

Human renal dipeptidase is found in kidney tissue and had its sequenced deduced in 1990 (105). It is known to possess a high glycosylated C-terminus which is cleaved during maturation and is anchored to the membrane at Ser369 via glysocylphosphatidylinositol (106). Although its role is not quite understood, it is known to hydrolyze cysteinyl dipeptides and some beta-lactam antibiotics (107). Its crystal structure has been solved (PDB: 1ITQ) (108). Two other dipeptidases have been

characterized in mouse and are found as homologues in humans. Dipeptidase 2 cleaves leukotriene D4 (LTD4) like Dipeptidase 1, but not cystinyl-bis-glycine. Oppositely, Dipeptidase-3 cleaves cystinyl-bis-glycine but not LTD4 (109).

CHAPTER II

STRUCTURE-GUIDED DISCOVERY OF NOVEL ENZYME FUNCTIONS: FUNCTIONAL DIVERSITY IN GROUP 1 OF COG0402*

Introduction

The rate at which new genes are being sequenced greatly exceeds our ability to correctly annotate the functional properties of the corresponding proteins (110). Annotations based primarily on sequence identity to experimentally characterized proteins are often misleading because closely related sequences can have different functions, while highly divergent sequences can have identical functions (18,111). Unfortunately, our understanding of the principles that dictate the catalytic properties of enzymes, based on protein sequence alone, is often insufficient to correctly annotate proteins of unknown function. New methods must therefore be developed to define the sequence boundaries for a given catalytic activity and new approaches must be formulated to identify those proteins that are functionally distinct from their close sequence homologues. To address these problems, we have developed a comprehensive strategy for the functional annotation of newly sequenced genes using a combination of structural biology, bioinformatics, computational biology, and molecular enzymology (4,19,112). The power of -this multidisciplinary approach for

---

* This chapter is adapted with permission from **Structure-Guided Discovery of New Deaminase Enzymes** Daniel S. Hitchcock, Hao Fan, Jungwook Kim, Matthew Vetting, Brandan Hillerich, Ronald D. Seidel, Steven C. Almo, Brian K. Shoichet, Andrej Sali, and Frank M. Raushel *Journal of the American Chemical Society* **2013** *135* (37), 13927-13933. Copyright 2013 American Chemical Society.

discovering new reactions catalyzed by uncharacterized enzymes is being tested using the amidohydrolase superfamily (AHS) as a model system (2,19).

The AHS is an ensemble of evolutionarily related enzymes capable of hydrolyzing amide, amine, or ester functional groups at carbon and phosphorus centers (1,2). More than 24,000 unique protein sequences have been identified in this superfamily and they have been segregated into 24 clusters of orthologous groups (COG) (113). One of these clusters, cog0402, catalyzes the deamination of nucleic acid bases (2,19). Previously, we successfully predicted that Tm0936, an enzyme from *Thermotoga maritima,* would catalyze the deamination of *S*-adenosylhomocysteine (SAH) to *S*-inosylhomocysteine (SIH) (19). Here we significantly expand the scope of these efforts by addressing the functional and specificity boundaries for more than 1000 proteins homologous to Tm0936, resulting in the prediction and discovery of novel substrate profiles for neighboring enzyme subgroups. To do so, we have integrated a physical library screen with the computational docking of high-energy reaction intermediates to homology models of fifteen previously uncharacterized proteins. To identify enzymes most closely related to Tm0936, we retrieved all of the protein sequences that correlated with a BLAST E-value cutoff better than 10^{-36} (8). This procedure identified 1358 proteins that were further sorted into smaller subgroups through the construction of a sequence similarity network at a BLAST E-value cutoff of 10^{-100} (**Figure 2.1**) (9). The minimal sequence identity between any two

proteins in this network is 23% and twelve representative subgroups (sg-1a through sg-11) were arbitrarily defined, colored-coded, and numbered.

The three-dimensional structure of Tm0936 (from sg-8) was previously determined in the presence of the product SIH. The most salient structural features for substrate recognition include Glu-84, Arg-136, Arg-148 and His-173 (**Figure 2.2**). These residues form electrostatic interactions with the 2'- and 3'-hydroxyls from the ribose sugar, the α-carboxylate of the homocysteine moiety, and N3 of the purine ring (19). The catalytic machinery is composed of a zinc ion that is coordinated by three histidine residues (His-55, His-57, His-173) and an aspartate (Asp-279), while proton transfer reactions are facilitated by Glu-203 and His-228. The six residues required for metal binding and proton transfers are fully conserved in all 1358 proteins (**Figure 2.1**). However, only those proteins within sg-8 fully conserve the four residues that are utilized in the recognition of SIH; sg-1 through sg-7 lack one or both of the carboxylate-binding arginine residues, while sg-9, sg-10, and sg-11 lack the two adenine/ribose recognition residues, histidine and glutamate. All of these proteins are therefore anticipated to contain unique substrate profiles and to catalyze the deamination of unanticipated substrates.

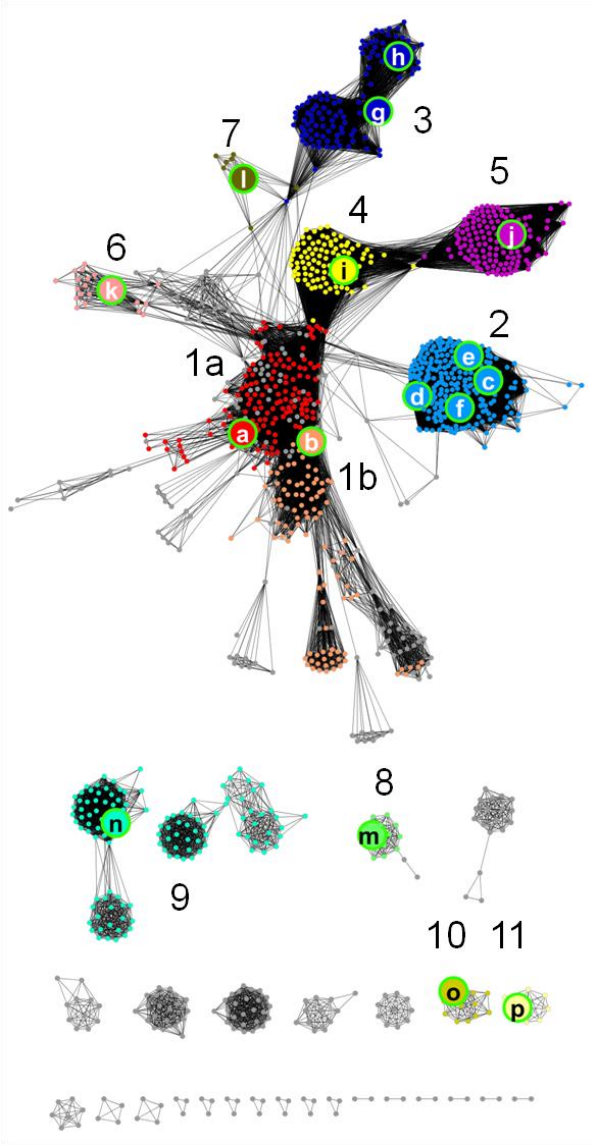


Figure 2.1: Sequence similarity network for proteins related to Tm0936 from *Thermotoga maritima*. A node (dot) represents an enzyme from a bacterial species and an edge (a connecting line) indicates that the two proteins are related by a BLAST E-value of 10^{-100} or better. Proteins sharing sequence similarity with Tm0936 cluster into apparent subgroups and 12 of these have been arbitrarily numbered and color-coded based on the network diagram. Subgroups are predicted to be functionally similar, and representatives from each subgroup, denoted by the letters **a-p**, were selected for purification and functional characterization.

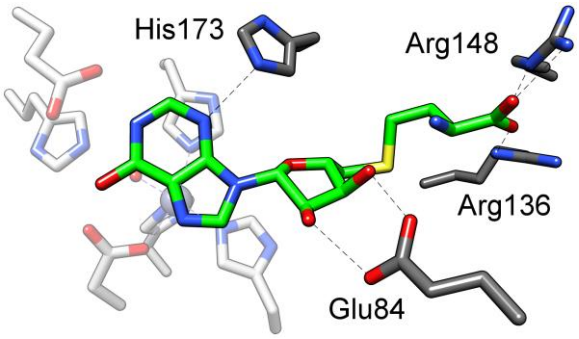


Figure 2.2: The active site of Tm0936. The crystal structure of Tm0936 (PDB id: 2PLM) in the presence of *S*-inosylhomocysteine (green) highlights the four residues (dark gray) that are important for substrate binding. Arg-148 and Arg-136 bind to the carboxylate moiety of SAH; His-173 and Glu-84 interact with N3 of the purine ring and the 2', 3' hydroxyls of the ribose moiety, respectively. Faded residues denote the six residues that bind the zinc or facilitate proton transfer reactions during the catalytic transformations. These residues are conserved in all of the proteins depicted in **Figure 2.1**.

Methods

Cloning and Purification of Mm2279, Spn3210, Moth1224, Avi5431, Tvn0515 Caur1903, and Cthe1199. The genes for Mm2279, Spn3210, Moth1224, Avi5431, and Tvn0515 were cloned into a pET-30a expression vector with C-terminal 6x His-tags. Caur1903 was cloned into pET-42a with an N-terminal GST tag. Cthe1199 was cloned into pET-30a without a His-tag. The *E. coli* BL-21 DE3 cells were transformed with the plasmids and plated onto LB-agarose containing 50 µg/mL kanamycin. 1 L cultures (LB broth, kanamycin 50 µg/mL) were inoculated from the resulting colonies and grown at 37 °C until the optical density at 600 nm reached 0.6. Protein expression was induced with 1 mM IPTG, and the cultures were shaken for 20 hours at 25 °C. The cultures were centrifuged at 7000 RPM for 15 minutes and the cell pellets were disrupted by

sonication in 35 mL of running buffer and PMSF. DNA was precipitated by protamine sulfate. Proteins with a 6x His-tag were loaded onto a 5 mL HisTrap column (GE HEALTHCARE, USA) with running buffer (20 mM HEPES, 250 mM NaCl, 250 mM NH_4SO_4, 20 mM imidazole, pH 7.5) and eluted with a linear gradient of elution buffer (20 mM HEPES, 250 mM NaCl, 250 mM NH_4SO_4, 500 mM imidazole, pH 7.5).

Caur1903 was loaded onto a 5 mL GSTrap column (GE HEALTHCARE, USA) with running buffer (20 mM HEPES, 10% [w/v] glycerol, 200 mM NaCl, 200 mM NH_4SO_4, pH 8.0) and eluted with a gradient of elution buffer (20 mM HEPES, 10% [w/v] glycerol, 200 mM NaCl, 200 mM NH_4SO_4, 10 mM reduced glutathione, pH 8.0). Cthe1199 was precipitated with 70% saturation NH_4SO_4 and the pellet resuspended in 4 mL of 50 mM HEPES, pH 7.5. The solution was loaded onto a HiLoad 26/60 Superdex 200 gel filtration column. Active fractions were loaded onto an anion exchange column with binding buffer (20 mM HEPES, pH 7.5), and eluted with a gradient of 20 mM HEPES, 1 M NaCl, pH 7.5. All proteins were concentrated with a 10 kDa molecular weight cutoff filter. Moth1224 and Mm2279 were almost entirely insoluble. Both proteins were verified by N-terminal sequencing.

Cloning and Purification of Cv1032, Pfl4080, Ef1223, Cpf1475, Dvu1825, and Bc1793. The genes for these proteins were PCR amplified and ligated into pSGX3, a derivative of pET26b (114), yielding a protein with Met-Ser-Leu followed by the PCR product and a C-terminal hexa-histidine tag. BL21 (DE3)-Condon 1 RIL (Invitrogen) cells were transformed with these plasmids and selenomethionine-labeled protein was

expressed in 3 L of HY media with 50 µg/mL of kanamycin and 35 µg/mL of chloramphenicol. Expression was induced with 0.4 mM IPTG at an OD 600 of 1.0 and grown for 21 hours whereupon the cells were pelleted by centrifugation. Cells were resuspended in 3x the volume with 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 25 mM imidizole, 0.1% (*v/v*) Tween20, lysed by sonication, and clarified by centrifugation at 4° C. The supernatant was applied to a 5-mL HisTrapHP column (GE HEALTHCARE, USA) pre-equilbrated with 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 25 mM imidizole, 10% (w/v) glycerol. The column was washed with 10 column volumes (CV) of 20 mM Tris–HCl pH 8.0, 500 mM NaCl, 10% (w/v) glycerol, and 40 mM imidazole, and subsequently step-eluted with 2 CV of same buffer with an imidazole concentration of 250 mM. Concentrated fractions were pooled and subsequently passed over a 120-mL Superdex 200 column (GE HEALTHCARE, USA) equilibrated with 10 mM HEPES pH 7.5, 150 mM NaCl, 10% (w/v) glycerol, and 5 mM DTT. Proteins exhibiting purity on SDS-GELS greater than 95% were pooled and concentrated to > 10 mg/mL using AMICON centrifugal filters. Samples were snap-cooled in liquid N_2, and stored at -80 °C.

Cloning and Purification of Pa3170 and Xcc2270. The genes for Pa3170 and Xcc2270 were amplified from their respective genomic DNA using primers that included a ligation independent cloning sequence and the N- and C-terminal sequences of each respective gene. In general, PCR was performed using KOD Hot Start DNA Polymerase (NOVAGEN). The amplified fragments for Pa3170, and Xcc2270 were cloned into the C-terminal TEV cleavable StrepII-6x-His-tag containing vector CHS30, a derivative of

pET30 by ligation-independent cloning (115). BL21 (DE3)-Condon 1 RIL (Invitrogen) cells were transformed with the resulting plasmids and protein was expressed by autoinduction in 2 L of ZYP-5052 media with 50 µg/mL of kanamycin and 35 µg/mL of chloramphenicol. Cells were grown at 37 °C to an OD of 2-4 and then reduced to 22 °C for a period of 12-16 hours whereupon the cells were pelleted by centrifugation. Cells were resuspended in 3x the volume with 50 mM Hepes, pH 7.5, 150 mM NaCl, 10% (*w/v*) glycerol, 20 mM imidazole, lysed by sonication, and clarified by centrifugation at 4 °C. The supernatant was applied to a 1-mL HisTrapHP column (GE HEALTHCARE) pre-equilbrated with 50 mM Hepes, pH 7.5, 150 mM NaCl, 25 mM imidizole, 10% (*w/v*) glycerol. The column was washed with 10 column volumes of 50 mM Hepes, pH 7.5, 150 mM NaCl, 10% (w/v) glycerol, 25 mM imidazole and subsequently step-eluted with the same buffer with an imidazole concentration of 250 mM directly onto a 120-mL Superdex 200 column (GE HEALTHCARE) equilibrated with 10 mM HEPES pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, and 5 mM DTT (protein storage buffer). Proteins exhibiting purity on SDS-GELS greater than 95% were pooled, concentrated to >10 mg/mL using AMICON centrifugal filters. Samples were snap-cooled in liquid N_2, and stored at -80 °C.

Substrate Characterization. Substrate screening was performed by monitoring the UV-vis spectral change from 235 to 320 nm. Potential substrates at a concentration of 100 µM were incubated overnight in a 96-well quartz plate with 1 µM enzyme in 50 mM HEPES, pH 7.5. Those showing a spectral change were subjected to more detailed

kinetic assays. The kinetic constants were determined in 20 mM HEPES buffer at pH 7.5 by a direct UV assay at 30 °C. Deamination of adenosine, 5'-dAdo, MTA, SAH and *S*-adenosylthiopropylamine (TPA) were monitored at 263 nm ($\Delta\varepsilon$ = 7900 M^{-1} cm^{-1}) or 274 nm ($\Delta\varepsilon$ = 2600 M^{-1} cm^{-1}), depending on the concentration of substrate. Non-adenosine compounds were as follows: guanine (245 nm, $\Delta\varepsilon$ = 5200 M^{-1} cm^{-1}), 2-hydroxyadenine ($\Delta\varepsilon$ = 294 nm, 6600 M^{-1} cm^{-1}), 8-oxoadenine ($\Delta\varepsilon$ = 272 nm, 3700 M^{-1} cm^{-1}). Values of k_{cat} and k_{cat}/K_m were determined by fitting the data to equation 2.1 using SigmaPlot 11, where v is the initial velocity, E_t is enzyme concentration, and A is the substrate concentration.

$$v/E_t = k_{cat} K_a / (A + K_a) \qquad \textbf{(Equation 2.1)}$$

Homology Modeling. Two structures (PDB code: 2PLM and 1P1M) have been solved for Tm0936 from *Thermotoga maritima* in the ligand-bound (holo) and ligand-free (apo) states, respectively. The holo structure of Tm0936 was used as a template to build homology models for 15 proteins from 11 subgroups. Because Moth1224 from sg-9 has a sequence identity similar to that of guanine deaminase (PDB code: 2UZ9) as it does to Tm0936, the structure of guanine deaminase was also used as template for Moth1224. For each target-template pair, the same procedure of homology modeling was applied. First, the sequence alignment was computed by MUSCLE (Multiple Sequence Comparison by Log-Expectation). Second, a total of 500 homology models

were generated with the standard “automodel” class in MODELLER (116), and the model with the best DOPE (117) score was selected to begin the model refinement. Third, side chains of active site residues that are within a distance of 5 Å from the bound ligand in the template structure, were optimized using the “side chain prediction” protocol in PLOP (118), resulting in one representative model. The bound ligand from the template structure was included in the second step for construction of the initial homology model, but was removed in the model refinement.

Virtual Screening. The high-energy intermediate (HEI) library (119,120) that contains 57,672 different intermediate forms of 6440 KEGG (Kyoto Encyclopedia of Genes and Genomes) (121,122) molecules was screened against the crystal structure of Tm0936 and each of the refined homology models using the docking program DOCK 3.6 (123). The computed poses were subjected to a distance cutoff to make sure that the O^- of the HEI portion of the molecule is found within 4 Å of the metal ions in the active site. Of the molecules that satisfied this constraint, the top 500 compounds ranked by DOCK score (consisting of van der Waals, Poisson–Boltzmann electrostatic, and ligand desolvation penalty terms) were inspected visually to ensure the compatibility of the pose with the amidohydrolase reaction mechanism. The details of the HEI docking library preparation, the molecular docking procedure, and the docking results analysis have been previously described (119,120,124,125).

Results

Target Selection and Structure Determination. Representative examples from each of the twelve major subgroups contained within the sequence similarity network of proteins (**Figure 2.1**) related to the initial protein target, Tm0936, were selected for interrogation (**Table 2.1**). Proteins from sg-1a (Cthe1199), sg-1b (Mm2279), sg-2 (Cv1032, Pfl4080, Xcc2270, and Pa3170), sg-3 (Ef1223 and Cpf1475), sg-4 (Bc1793), sg-5 (Spn13210), sg-6 (Dvu1825), sg-7 (Caur1903), sg-8 (Tm0936), sg-9 (Moth1224), sg-10 (Avi5431), and sg-11 (Tvn0515) were purified to homogeneity and the three-dimensional structures of three proteins were determined (**Table 2.2**). The structure of Cv1032 was determined in the presence of inosine (PDB id: 4F0S) and also with 5'-methylthioadenosine (MTA) in an unproductive complex (PDB id: 4F0R). The structures of Xcc2270 (PDB id: 4DZH) and Pa3170 (PDB id: 4DYK) were determined with zinc bound in the active site.

Predictions of Enzyme Function. Homology models were constructed for each of the purified proteins using the X-ray structure of Tm0936 (PDB id: 2PLM) as the initial structural template (126). A virtual library of 57,672 high-energy reaction intermediates (HEI) was screened against the homology models, as well as the X-ray structure of Tm0936 (112). For the thirteen targets from sg-1 through sg-8, analogues of adenosine (5'-deoxyadenosine, adenosine, 5'-methylthioadenosine, and SAH) were predicted as the most probable substrates (**Table 2.3**).

Table 2.1: Amidohydrolase selected from cog0402 for purification.

Subgroup	Organism	gi	Locus tag	Sequence identity to Tm0936
1a	*Clostridium thermocellum* ATCC 27405	125973714	Cthe_1199	42%
1b	*Methanosarcina mazei* Go1	21228381	MM_2279	40%
2	*Chromobacterium violaceum* ATCC 12472	34496487	CV_1032	32%
2	*Pseudomonas fluorescens* Pf0-1	77460301	Pfl01_4080	34%
2	*Xanthomonas campestris pv. campestris* str. ATCC 33913	21231708	XCC2270	34%
2	*Pseudomonas aeruginosa* PAO1	15598366	PA3170	33%
3	*Clostridium perfringens* ATCC 13124	110799822	CPF_1475	31%
3	*Enterococcus faecalis* V583	29375796	EF1223	28%
4	*Bacillus cereus* ATCC 14579	29895480	BC1793	36%
5	*Streptococcus pneumoniae* ATCC 700669	221232098	SPN23F_13210	33%
6	*Desulfovibrio vulgaris* str. Hildenborough	46580235	Dvu1825	36%
7	*Chloroflexus aurantiacus* J-10-fl	163847463	Caur_1903	32%
8	*Thermotoga maritima* MSB8	15643698	Tm0936	100%
9	*Moorella thermoacetica* ATCC 39073	83590072	Moth_1224	32%
10	*Agrobacterium vitis* S4	222106480	Avi_5431	31%
11	*Thermoplasma volcanium* GSS1	13541346	TVN0515	30%

The histidine residue that interacts with N3 from the purine ring of SAH in the structure of Tm0936 is conserved in all of the proteins except for the enzymes from sg-9 (Moth1224), sg-10 (Avi5431), and sg-11 (Tvn0515). In Moth1224 (sg-9), this histidine residue is replaced by an arginine (Arg-192), which has previously been observed in the active site of guanine deaminase (4). The sequence identity (33%) between Moth1224 and Tm0936 is similar to that between Moth1224 and human guanine deaminase (31%) (PDB id: 2UZ9). Therefore, a homology model was constructed for Moth1224 based on each one of these two templates separately. For the model based on Tm0936, the docking hits were predominantly adenosine analogues (**Table 2.3**). However, for the model based on guanine deaminase, the primary docking hits were analogues of

guanine. The highest ranking of these compounds included guanine (rank 10), 8-oxoguanine (rank 29), and thioguanine (rank 35).

In Avi5431 (sg-10) the sequence alignment predicts that a glutamate (Glu-189) replaces the histidine that interacts with N3 of the adenine moiety in the Tm0936 template. However, in the modeled active site of Avi5431, His-190 occupies the same physical position as His-173 in Tm0936, while the side chain of Lys-308 is located on the other side of the pocket, which has no available space for the binding of the ribose moiety of adenosine-like compounds. The primary docking hits for Avi5431 included small aromatic amines that are not attached to a ribose group. Among the best ranking compounds are mercarzole (rank 3), 6-hydroxymethyl 7,8-dihydropterin (rank 6), 8-oxoadenine (rank 21), melamine (rank 22), and guanine (rank 28).

Tvn0515 from sg-11 is unusual as proteins within this subgroup have a conserved glutamine in place of the histidine that binds to N3 of the adenine moiety in Tm0936. Originally, it was hypothesized that this substitution would enable the binding of 2-oxoadenosine in a manner similar to the way the homologous carbonyl group binds in cytosine, pterin, and guanine deaminases (26). However, virtual screening against the homology model of Tvn0515 suggested adenosine analogues as the most probable substrates; SAH was ranked higher than adenosine, MTA, and 5'-dAdo (**Table 2.3**).

Table 2.2. Crystallization data.

DATASET STATISTICS[a]	Cv_1032 (Zn-INOSINE)	Cv_1032 (Zn-MTA)	XCC2270 (ZN)	PA3170 (ZN)
Space Group	P41212	P41212	C2	P212121
Unit Cell (Å , °)	*a*=*b*=101.1 *c*=88.0	*a*=*b*=101.1 *c*=88.6	*a*=133.6 *b*=51.2 *c*=67.4 β=106.5	*a*=73.9 *b*=98.7 *c*=119.8
Resolution (Å)	50-1.85 (1.88-1.85)	50-1.8 (1.83-1.8)	50-1.53 (1.65-1.55)	50-2.0 (2.03-2.0)
Completeness (%)	98.1 (98.6)	99.2 (100.0)	99.6 (100.0)	99.6 (98.9)
Redundancy	11.4 (11.4)	6.1 (5.8)	3.7 (3.7)	5.2 (4.7)
Mean(I)/sd(I)	21.9 (4.3)	23.2 (2.3)	18.9 (2.8)	11.2 (2.7)
R_{sym}	0.140 (0.870)	0.102 (0.799)	0.068 (0.527)	0.068 (0.753)
STRUCTURE STATISTICS				
Resolution (Å)	40.3-1.85 (1.90-1.85)	50-1.80 (1.85-1.80)	50-1.53 (1.57-1.55)	50-2.0 (2.05-2.0)
Unique reflections	39008 (2648)	41034 (2892)	65346 (2581)	55862 (3609)
R_{cryst} (%)	15.7 (16.2)	16.3 (19.0)	20.8 (31.6)	17.6 (27.6)
R_{free} (%, 5% of data)	18.5 (19.2)	19.9 (22.2)	23.3 (36.8)	22.4 (35.4)
Contents of model (Native Range)[b]	1-439	1-439	1-449	1-444
Observed residues	A5-438	A5-440	A9-447	A6-442, B6-442
Waters	299	325	247	309
Atoms total	3674	3646	3673	7035
Average B-factor (Å^2)				
TLS groups (#)	0	0	2	8
Protein/Waters/Ligand/Zn	18.6/26.1/19.1/-	20.2/30.8/28.7/15.2	25.8/17.6/-/32.3	45.4/43.4/-/33.5
RMSD				
Bond lengths (Å) /Angles (°)	0.007/1.082	0.015/1.338	0.024/1.508	0.008/1.103
MOLPROBITY STATISTICS				
Ramachandran Favored / Outliers (%)	97.0 (0.23)	97.0 (0.23)	96.3 (0.23)	96.3 (0.3)
Rotamer Outliers (%)	0.0	0.29	1.2	2.0
Clashscore[d]	4.36 (98th pctl)	5.87 (94th pctl)	5.25 (92nd pctl)	7.99 (90th pctl)
Overall score[d]	1.39 (97th pctl)	1.49 (94th pctl)	1.57 (83rd pctl)	1.91 (80th pctl)
PDB ID	4F0S	4F0R	4DZH	4DYK

[a] Statistics in parenthesis are for the highest resolution bin
[b] numbering outside of listed native range are polyhistidine tags or cloning artifacts
[c] Scores are ranked according to structures of similar resolution as formulated in MOLPROBITY

Table 2.3 Enzyme kinetic constants at pH 7.5, 30 °C, and docking ranks.

Protein			Substrate k_{cat}/K_m (M^{-1} s^{-1})			
Label	**Subgroup**	**Locus Tag**	**5'-dAdo**	**Ado**	**MTA**	**SAH**
(a)	1a	Cthe1199	5.0×10^4 (6)	1.6×10^4 (5)	2.9×10^6 (9)	1.4×10^7 (1)
(b)	1b	Mm2279	1.2×10^7 (16)	2.2×10^5 (12)	1.7×10^7 (10)	1.3×10^7 (11)
(c)	2	Cv1032	3.4×10^6 (26)	6.6×10^6 (13)	1.2×10^6 (20)	<10 (80)
(d)	2	Pfl014080	8.9×10^6 (13)	4.6×10^5 (5)	1.3×10^7 (2)	<10 (316)
(e)	2	Xcc2270	4.7×10^6 (16)	3.4×10^6 (10)	6.8×10^6 (4)	<10 (60)
(f)	2	Pa3170	4.5×10^7 (23)	9.6×10^6 (9)	8.2×10^7 (19)	<10 (361)
(g)	3	Cpf1475	<10 (66)	<10 (50)	<10 (135)	<10 (212)
(h)	3	Ef1223	<10 (82)	<10 (44)	<10 (37)	<10 (373)
(i)	4	Bc1793	3.6×10^5 (12)	3.7×10^5 (6)	3.4×10^1 (32)	<10 (na)[a]
(j)	5	Spn13210	1.5×10^7 (34)	7.2×10^6 (23)	<10 (52)	<10 (334)
(k)	6	Dvu1825	1.2×10^7 (10)	2.6×10^5 (5)	6.2×10^2 (12)	<10 (na)
(l)	7	Caur1903	2.6×10^5 (6)	2.6×10^5 (3)	<10 (12)	<10 (14)
(m)	8	Tm0936	3.7×10^4 (6)	9.2×10^3 (3)	1.5×10^6 (2)	1.1×10^7 (1)
(n)	9	Moth1224	<10 (na)	<10 (na)	<10 (na)	<10 (na)
(o)	10	Avi5431	<10 (na)	<10 (na)	<10 (na)	<10 (na)
(p)	11	Tvn0515	<10 (175)	<10 (131)	<10 (106)	2.6×10^5 (28)

Physical Library Screen and Kinetic Assays. A physical library of more than 100 potential substrates was assembled and tested for catalytic activity using the 16 enzymes isolated for this investigation by monitoring the change in absorbance from 240-300 nm of the following compounds: The following compounds were tested for deamination with all purified enzymes: melamine, cytosine, thioguanine, guanine, adenine, 7-methylguanine, cytidine, 2'-deoxycytidine, pterin, biopterin, neopterin, guanosine, adenosine, 8-mercaptoguanine, 2,6-diaminopurine, xanthine, 8-oxoguanine, urate, 9-methylguanine, 6-chloropurine, 8-oxoadenine, isopentenyladenine, *N*6-methyladenine, 1-methyladenine, *cis*-zeatin, *trans*-zeatin, benzyladenine, 2-amino-6-benzylthiopurine, *N*-ethenopurine, *N*6-acyladenine, kinetin, *N*-dimethyladenine, *N*-butyladenine, isoguanine, 2-amino-6-methoxypurine, 6-methylthiopurine, 2-chloroadenine, 6-methoxyadenine, 6-mercaptopurine, 7-methyladenine, 2-dimethylaminoadenine, 3-iminisoindolinone, acycloguanosine, 2'-deoxyguanosine, xanthosine, 2'-deoxyadenosine, 3'-deoxyadenosine, 5'-deoxyadenosine, 2',5'-dideoxyadenosine, 5'-amino-5'-deoxyadenosine, 5'-chloro-5'-deoxyadenosine, 5'-deoxy-5'-methylthioadenosine, *S*-adenosylhomocysteine, *S*-adenosylthiopropylamine *N*6-methyl-2'-deoxyadenosine, isoguanine, 2'-deoxyisoguanine, cytosine-β-D-arabinofuranose, *N*-methylcytidine, 5-methyl-2'-deoxycytidine, 5-hydroxymethylcytidine, 2,4-diaminopyrimidine, 4,6-diaminopyrimidine, 4-amino-2,6-dihydroxypyrimidine, 2-amino-4,6-dihydroxypyrimidine, 5-methlcytosine, 5-hydroxymethylcytosine, 5-carboxycytosine, 5-aminocytosine, 5-fluorocytosine, 5-

chlorocytosine, 4-hydroxy-2,5,6-triaminopyrimidine, *N*-methylcytosine, 3-methylcytosine, 2,6-diaminopyrimidine, 2,4,6-triaminopyrimidine, uracil, 5-aminouracil, 5-formyluracil, 5-fluorouracil, 3-oxauracil, 5-hydroxymethyluracil, 6-hydroxyaminouracil, thiamine, 2-aminopyrimidine, 4-aminopyrimidine, toxopyrimidine, sulfamonomethoxine, 4-thiouracil, 2-thiouracil, thymine, 5-azacytosine, ammeline, ammelide, cyanuric acid, 2-amino-4-hydroxy-6-chlorotriazine, cyromazin, *N*-ethylammeline, desethylatrazine, 6-azacytosine, 2-aminopyridine, 2,6-diaminopyridine, 4,6-diamino-2-hydroxpyridine, 4-aminopyridine, 4-dimethylaminopyridine, creatinine, 5-amino-4-imidazolecarboxyamide, 5-hydantoin acetic acid, 3-amino-5-mercapto-1,2,4-triazole, 3-amino-5-carboxy-1,2,4-triazole, dihydrouracil, dihydrothymine, and dihydroorotate. The structures of the best substrates are provided in **Figure 2.3**. Compounds that exhibited catalytic activity with any of the 16 proteins were subjected to detailed kinetic assays using a direct spectrophotometric assay**.** Analogues of adenosine, including 5'-dAdo, MTA, and SAH, were deaminated by proteins from sg-1a, sg-1b, sg-2, sg-4 through sg-8, and sg-11. For most targets, the substrates with the highest activity had k_{cat}/K_m values that ranged from 10^5 - 10^7 M^{-1} s^{-1}, except for Cpf1475 and Ef1223 from sg-3 for which no substrates could be identified. The proteins from sg-1a, sg-1b, sg-8, and sg-11 preferentially deaminated SAH. MTA was deaminated by proteins from sg-1 and sg-8, but not from sg-11. The four proteins purified from sg-2 deaminated MTA, 5'-dAdo, and adenosine, but not SAH. The proteins purified from sg-4 through sg-7 preferentially deaminated adenosine and 5'-dAdo; MTA is either not a

substrate or it is deaminated 4-orders of magnitude less efficiently. For the proteins purified from sg-4, sg-5, and sg-7, adenosine and 5'-dAdo are equally good substrates, whereas for Dvu1825 from sg-6, 5'-dAdo is a substantially better substrate than is adenosine. The values of k_{cat}/K_m for adenosine, 5-dAdo, MTA, and SAH are presented in **Table 2.3** and the values of K_m and k_{cat} are provided in **Table 2.4**.

Moth1224 (sg-9) and Avi5431 (sg-10) were subjected to more extensive library screening because these proteins lacked many of the conserved residues found in Tm0936. Guanine was the only substrate identified for Moth1124 with values of k_{cat}/K_m and k_{cat} of 1.2 x 10^5 M^{-1} s^{-1} and 0.48 s^{-1}, respectively. Avi5431 deaminated two adenine related compounds, 8-oxoadenine and 2-oxoadenine, with values of k_{cat}/K_m of 1.8 x 10^5 and 2.3 x 10^4 M^{-1} s^{-1}, respectively. The values of k_{cat} for these two compounds are 3.4 and 13 s^{-1}. This enzyme did not deaminate adenine.

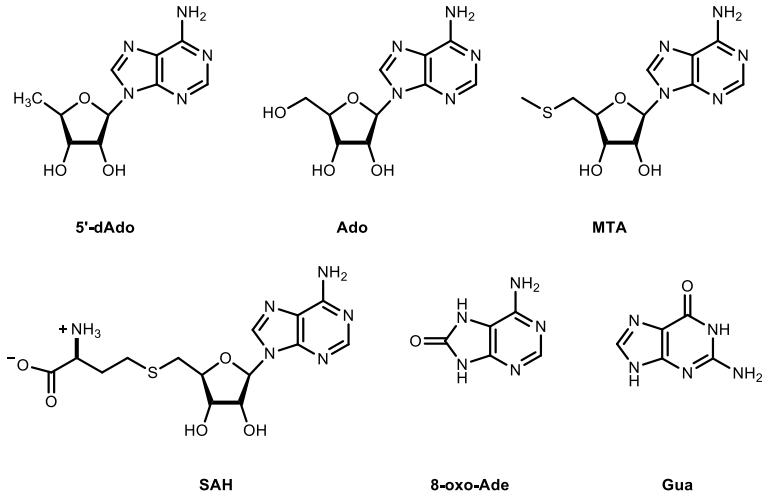


Figure 2.3: The structures of substrates for the enzymes related to Tm0936.

Table 2.4. Kinetic constants of adenosine deaminating enzymes.

Locus Tag	Subgroup	5dAdo			Ado			MTA			SAH			TPA		
		k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
Cthe1199	1a	3.9 0.1	78 ± 7	5.0 (0.3) x 10^4	2.6 ± 0.1	160 ± 20	1.6 (0.1) x 10^4	25 ± 2	9 ± 1	2.9 (0.6) x 10^6	25.6 ± 0.5	1.8 ± 0.2	1.4 (0.1) x 10^7	23 ± 1	4.4 ± 0.7	5.2 (0.5) x 10^6
MM2279	1b	27 ± 2	2.3 ± 0.4	1.2 (0.1) x 10^7	7.9 ± .7	36 ± 8	2.2 (0.3) x 10^5	16 ± 1	0.9 ± 0.3	1.7 (0.4) x 10^7	34 ± 3	2.7 ± 0.6	13 (2) x 10^7	16 ± 1	2.1 ± 0.4	7.6 (1) x 10^6
Cv1032	2	105 ± 3	31 ± 3	3.4 (0.3) x 10^6	107 ± 4	16 ± 3	6.6 (1) x 10^6	18.2 ± 0.3	16 ± 1	1.2 (0.1) x 10^6	ND	ND	<100	ND	ND	<100
Pf1014080	2	119 ± 3	13 ± 1	8.9 (0.6) x 10^6	40 ± 1	90 ± 10	4.6 (0.2) x 10^5	57 ± 2	4 ± 1	1.3 (0.1) x 10^7	ND	ND	<100	ND	ND	<100
XCC2270	2	67 ± 1	14 ± 1	4.7 (0.2) x 10^6	80 ± 2	24 ± 2	3.4 (0.1) x 10^6	74 ± 2	11 ± 1	6.8 (0.4) x 10^6	ND	ND	<100	ND	ND	<100
PA3170	2	180 ± 3	4.0 ± 0.3	4.5 (0.2) x 10^7	131 ± 9	14 ± 3	9.6 (1.3) x 10^6	155 ± 3	1.9 ± 0.2	8.2 (0.8) x 10^7	ND	ND	<100	ND	ND	<100
Bc1793	4	41 ± 1	130 ± 20	3.6 (0.3) x 10^5	46 ± 2	120 ± 10	3.7 (1) x 10^5	ND	ND	<100	ND	ND	<100	ND	ND	<100
SPN23F13210	5	850 ± 60	60 ± 10	15 (2) x 10^6	660 ± 40	90 ± 10	7.2 (0.4) x 10^6	ND	ND	<100	ND	ND	<100	ND	ND	<100
Dvu1825	6	38.7 ± 0.5	3.3 ± 0.2	1.2 (0.1) x 10^7	43 ± 3	170 ± 20	2.6 (0.2) x 10^5	ND	ND	620 ± 40	ND	ND	<100	ND	ND	<100
Caur1903	7	2.58 ± 0.05	11 ±1	2.3 (0.1) x 10^5	1.44 ± 0.04	5.6 ± 0.7	2.6 (0.3) x 10^5	ND	ND	<100	ND	ND	<100	ND	ND	<100
Tm0936	8	1.4 ± 0.1	39 ± 7	3.8 (0.5) x 10^4	2.2 ± 0.2	13 ± 4	1.7 (0.3) x 10^5	6.0 ± 0.3	4.0 ± 0.3	1.5 (0.1) x 10^6	7.1 ± 0.4	0.6 ± 0.2	1.1 (0.3) x 10^6	17 ± 1	9 ± 2	1.9 (0.2) x 10^6
TVN0515	11	ND	ND	<100	ND	ND	<100	ND	ND	<100	2.77 ± 0.05	11 ± 1	2.6 (0.2) x 10^5	ND	ND	<100

Discussion

Substrate Profiles. In this investigation we have successfully characterized fifteen proteins related in sequence to the first enzyme (Tm0936) found capable of deaminating SAH to SIH. Using a combination of homology modeling, virtual screening and physical library screening, six unique substrate profiles were established for twelve subgroups of proteins homologous to Tm0936. Tvn0515 (sg-11) deaminates only SAH, while enzymes from three subgroups (sg-1a, sg-1b, and sg-8) efficiently deaminate both SAH and MTA. Enzymes from sg-2 deaminate MTA, adenosine, and 5'-dAdo equally well, whereas proteins from sg-4, sg-5, and sg-7 cannot deaminate MTA but will deaminate adenosine and 5'-dAdo equally well. Enzymes from sg-6 deaminate 5'-dAdo about 100 times faster than adenosine. All of these substrates are linked to the recycling of reaction products from the utilization of *S*-adenosylmethionine in the cell. Not only was the correct family of substrates identified for these enzymes from full library screens, but much of the specificity among the adenosine analogues was also predicted, at least by relative rank. The substrate specificities for the 1000 proteins contained within sg-1 through sg-11 are provided in Appendix A.

Two of the subgroups that are closely related to Tm0936 deaminate quite different substrates. Moth1224 (sg-9) deaminates guanine while Avi5431 (sg-10) deaminates 8-oxoadenine. Avi5431 is the first enzyme ever reported to deaminate oxidatively damaged adenine. In addition, Dvu1825 (sg-6) is the first protein found capable of deaminating 5'-deoxyadenosine.

Deamination of SAH. In the crystal structure of Tm0936, the carboxylate of the homocysteine moiety of SAH interacts with two arginine residues, Arg-136 and Arg-148 (**Figure 2.2**). For Cthe1199 from sg-1a, the first arginine residue is conserved (Arg-147), while the second arginine is replaced with an aspartate (Asp-159). In the modeled active site of Cthe1199, these two residues interact with the carboxylate and amino groups of the homocysteine moiety of SAH, respectively (**Figure 2.4**). For Mm2279 (sg-1b), the two arginine residues are substituted with histidine (His-169) and aspartate (Asp-183), respectively. In the modeled active site of Mm2279, the amino group of the homocysteine moiety of SAH interacts with His-169 and Asp-183 (**Figure 2.5**). Sequence alignments for the remaining proteins in sg-1a and sg-1b suggest that catalytic activity with SAH is a consequence of an arginine (or lysine) residue aligned to Arg-136 in Tm0936, and a charged residue (Arg, Lys, Asp, Glu) aligned to Arg-148 in Tm0936. For proteins from sg-1b there is usually a tyrosine or histidine at residue position 136 and a glutamate or aspartate at residue position 148 from Tm0936.

This sequence-structure-activity relationship is not adequate to explain the SAH specificity for all of the proteins examined in this investigation. Bc1793 from sg-4 preserves the arginine (Arg-150) in the position aligned to Arg-136 in Tm0936, but SAH was not prioritized by docking against the homology model and not active in subsequent biochemical assays. The lack of SAH activity can be rationalized by the modeled active site of Bc1793, where an inserted loop from residue 126-129 prevents Arg-150 from interacting with potential substrates (**Figure 2.6**).

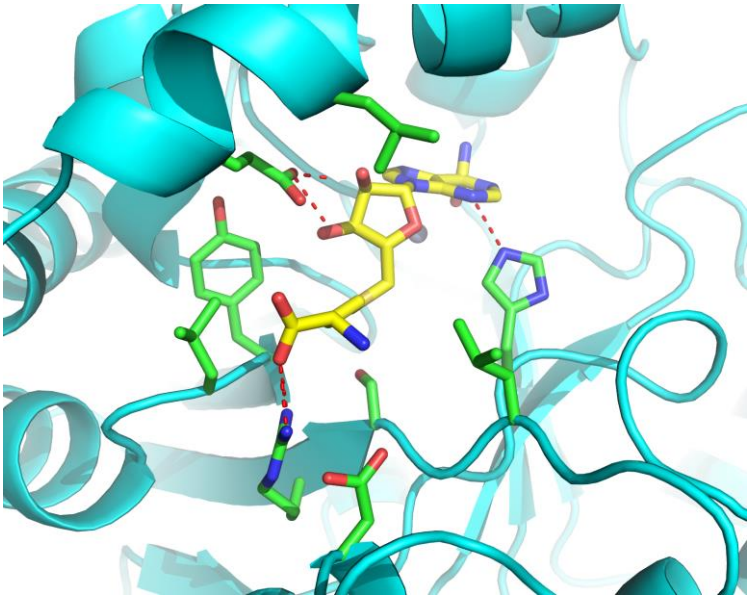


Figure 2.4: Modeled active site of Cthe_1199. In the modeled active sites of Cthe_1199 (cyan cartoon, green stick), the carboxylate and the amino groups of the homocysteine moiety of the docked SAH (yellow stick) interact with Arg-147 and Asp-159.

In Tvn0515 from sg-11, the two critical arginine residues from Tm0936 are replaced by the nonpolar residues Trp-140 and Pro-154. Surprisingly, SAH ranked high in docking to the homology model of Tvn0515 and this compound was efficiently deaminated. The carboxylate group from the homocysteine moiety of SAH apparently forms a hydrogen bond with the side chain from Lys-145 in the model of Tvn0515, while the amino group from the homocysteine moiety is exposed to the protein surface and forms a cation-π interaction with the side chain from Phe-85 (**Figure 2.7**). Uniquely, catalytic activity for Tvn0515 could not be detected with any compound other than SAH.

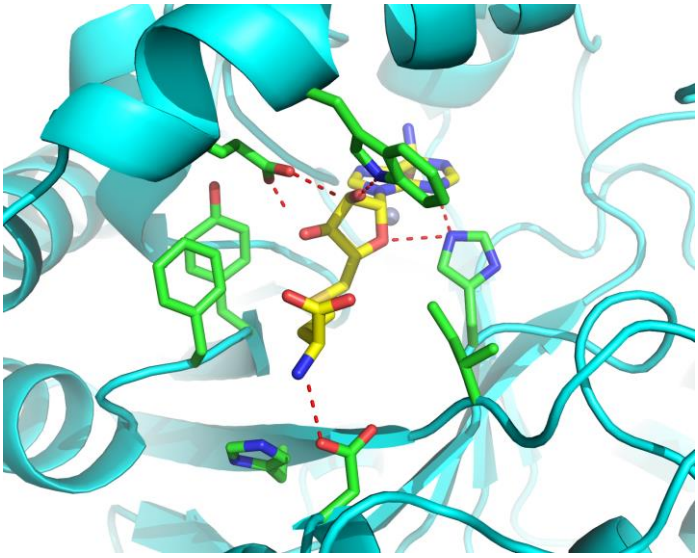


Figure 2.5: Modeled active site of MM_2279. In the modeled active sites of MM_2279 (cyan cartoon, green stick), the amino group of the homocysteine moiety of the docked SAH (yellow stick) interacts with His-169 and Asp-183.

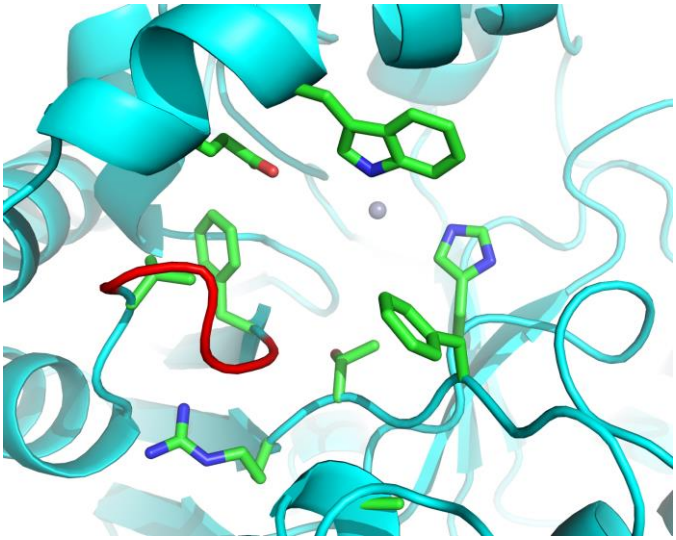


Figure 2.6: Modeled active site of BC1793. In the modeled active sites of BC1793 (cyan cartoon, green stick), the inserted loop (red cartoon) from residue positions 126-129 prevents the conserved Arg-150 from interacting with bound substrates.

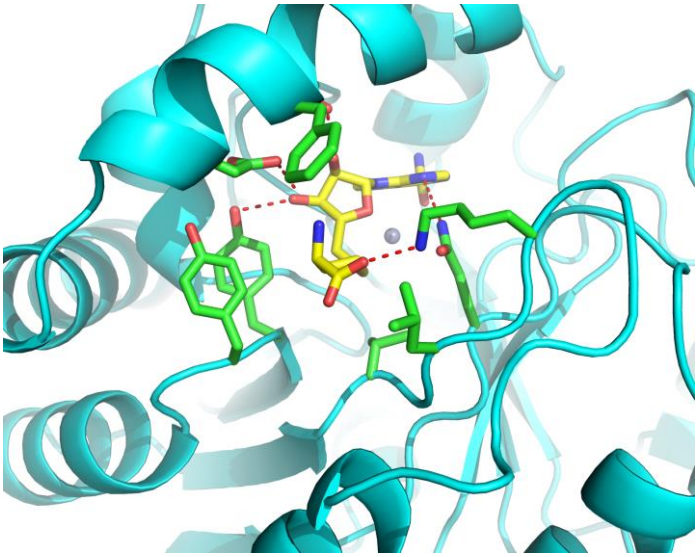


Figure 2.7: Modeled active site of TVN0515. In the modeled active sites of TVN0515 (cyan cartoon, green stick), the carboxylate group in the homocysteine moiety of the docked SAH (yellow stick) forms a hydrogen bond with the sidechain from Lys-145, while the amino group in the homocysteine moiety of the docked SAH was exposed to the protein surface and formed cation-π interaction with the sidechain from Phe-85.

Deamination of MTA. The four enzymes from sg-2 can deaminate MTA but not SAH. In the modeled active site of Pa3170, the methylthio group of the docked MTA resides in a hydrophobic pocket surrounded by Met-132, Tyr-133, Phe-134, Pro-155, Leu-157 and His-194. This model is consistent with the crystal structure of Pa3170 determined in the presence of 5'-methylthiocoformycin, an analogue of MTA (127). A superposition of the homology model for Cv1032 from sg-2 and the two x-ray structures of this protein are presented in **Figure 2.8A**. The modeled active site occupied by the HEI form of adenosine has an all-atom RMSD of 1.8 Å with respect to the crystal structure complexed with inosine, while the docking pose of adenosine is nearly identical to the crystal structure of inosine with a RMSD of 0.8 Å (**Figure 2.8B**).

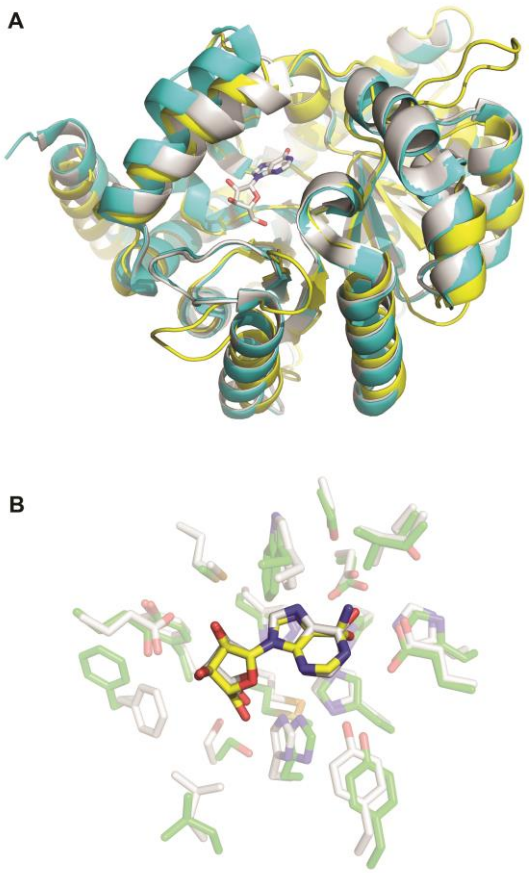


Figure 2.8: Comparison of the crystallographic and modeled structures of Cv1032. **(A)** The crystal structure of Cv1032 (PDB: 4F0S) in a catalytically productive state (white ribbon), the crystal structure of Cv1032 (PDB: 4F0R) in a catalytically unproductive state (cyan ribbon), and the homology model of Cv1032 based on the X-ray structure of Tm0936 (PDB: 2PLM) (yellow ribbon). The inosine bound in the active site of Cv1032 (PDB: 4F0S) is highlighted (white stick). (**B)** The crystal structure of inosine (white stick) in the active site of Cv1032 (transparent white stick), and the docking pose of adenosine (yellow stick) in the modeled active site (transparent green stick) composed of the same set of residues as the active site.

Deamination of Ado and 5'-dAdo. The enzymes from sg-4 through sg-7 catalyze the deamination of adenosine and 5'-deoxyadenosine, but not MTA and SAH. The

exclusion of MTA from the active sites of enzymes from these subgroups is likely the result of a hydrophobic amino acid residue found directly after β-strand 3 (**Figure 2.9**). In Tm0936, Val-139 is positioned near the C5'-carbon of the ribose moiety. This residue apparently permits Tm0936 and all proteins within sg-1a and sg-1b (Ile, Val) and sg-2 (Leu, Ile) to accommodate the thiomethyl substituent. In sg-4, sg-5, and sg-6 the residue equivalent to Val-139 is replaced with a bulky phenylalanine and in sg-7 this residue is replaced with methionine. These changes likely restrict the binding of only adenosine or 5'-dAdo within the active site. The lack of adenosine binding in the active site of Dvu1825 (sg-6) is less obvious, but it may be related to the residue corresponding to Gly-137 in Tm0936. In sg-4, sg-5, and sg-7 a conserved threonine is at this position. The crystal structure of Cv1032 (sg-2) suggests that a hydrogen bond is formed between this residue (Ser-153) and the 5'-hydroxyl of the bound inosine (PDB id: 4F0S). Dvu1825 contains an alanine (Ala-152) at this residue position and thus a hydrogen bond is not possible.

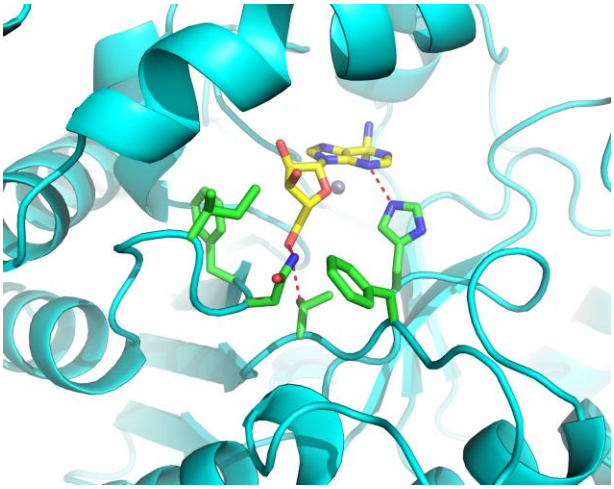


Figure 2.9: Interaction of Phe-153 in BC1793. In the modeled active sites of BC1793 (cyan cartoon, green stick), the 5'-hydroxyl group of the docked adenosine forms a hydrogen bond with the sidechain of the conserved Thr-151. The neighboring Phe-153 could prevent the binding of MTA

Convergent Evolution of Guanine Deaminase. Moth1224 from sg-9 was active as a deaminase only for guanine. The observed substrate specificity can be rationalized by the identity of the conserved residues located within the active site of this enzyme. Arg-192 and Gln-68 from Moth1224 align with residues Arg-213 and Gln-68 in human guanine deaminase (GuaD), which form critical hydrogen bonds to the substrate in the active site. The docking pose of guanine in the active site of Moth1224 (based on the GuaD structure) is presented in **Figure 2.10**. Although the protein is more similar to the entire group of enzymes related to Tm0936, the active site of Moth1224 bears the hallmark of the GuaD family. This example of convergent evolution underscores the importance of observing conservation of active site residues when predicting enzyme functions based on homology.

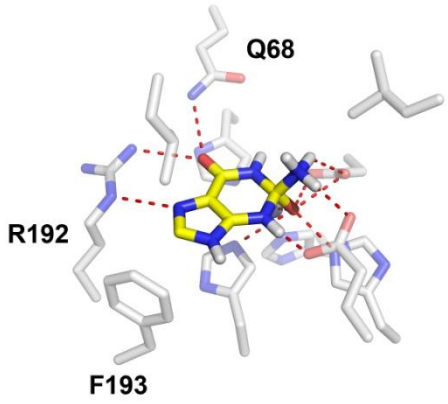


Figure 2.10: The binding pose of guanine in the homology model of Moth1224. The docking pose of guanine in a high-energy intermediate state (yellow stick) in the modeled active site of Moth1224 (transparent white stick) is presented.

8-Oxoadenine Deaminase. Avi5431 from sg-10 was able to deaminate only two substrates, 2-oxoadenine and 8-oxoadenine. To the best of our knowledge this is the first time that an enzyme has been identified that can deaminate these compounds. Both of these substrates are oxidatively damaged adenine within DNA (128,129). The deaminated products, 8-oxohypoxanthine and xanthine, can be further metabolized to uric acid (130). We had earlier identified the first enzyme capable of deaminating 8-oxoguanine (4). In the docking pose of 8-oxoadenine to the modeled Avi5431 active site, the N3 nitrogen and the C8 carbonyl oxygen from the purine ring interact with the side chains of His-190 and Lys-308, respectively (**Figure 2.11**).

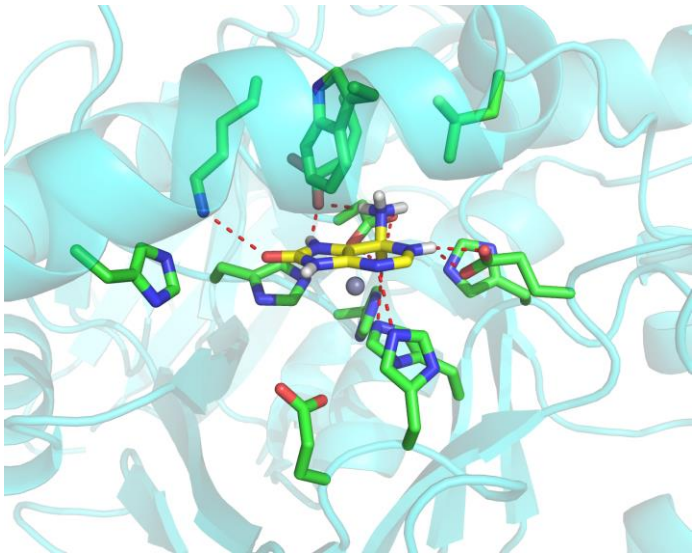


Figure 2.11: Modeled active site of Avi_5431. In the modeled active sites of Avi_5431 (transparent cyan cartoon, solid green stick), the N3 nitrogen and the C8 carbonyl oxygen from the purine ring of the docked 8-oxoadenine interact with His-190 and Lys-308.

Functional Annotation. Using a combination of bioinformatics, homology modeling, computational docking, structural biology, and molecular enzymology, we have interrogated the catalytic properties of a large cluster of proteins homologous to Tm0936, a protein from *Thermotoga maritima*, which was previously shown to deaminate *S*-adenosyl homocyteine to *S*-inosyl homocysteine. Subsets of enzymes were found to deaminate the catabolic metabolites of S-adenosyl methionine, including *S*-adenosyl homocyteine, methylthioadenosine, adenosine, and 5'-deoxy-adenosine. Not only were all of the actual substrates ranked within the top 100 of the metabolites in the virtual library, but the overall substrate specificities were also captured. We did not anticipate this level of agreement between prediction and experiment given the

well-known complexities of docking. Remarkably, these predictions were based on homology models of the targets, often templated at the limit of reliable primary sequence alignments. These results support the feasibility of prioritizing substrates and substrate-specificity from large library docking screens against homology models and presage large-scale functional annotation efforts that will significantly expand the range of metabolic diversity present in nature.

CHAPTER III

RESCUE OF THE ORPHAN ENZYME ISOGUANINE DEAMINASE*

Introduction

A major challenge for all aerobic organisms is the prevention and management of oxidative damage to DNA (131). In bacteria such as *Escherichia coli* there are specific repair enzymes for the removal of modified bases from damaged DNA (20,132-134). This class of repair enzymes includes MutM for excising 8-oxoguanine (8-oxoG), 8-oxoadenine (8-oxoA) and formamidopyrimidines (FAPY) from DNA (134-136). MutT catalyzes the hydrolysis of 2'-deoxy-8-oxoguanosine triphosphate to the monophosphate (137). The removal of mismatched A and isoguanine (2-oxoadenine) from DNA is catalyzed by MutY (138).

Isoguanine is mutagenic to *E. coli* (129,139). This base promotes A to C, G, and T transversions in addition to base substitutions and deletions (140). The formation of isoguanine in DNA occurs when 2'-deoxyATP is oxidized to 2'-deoxy-2-oxoadenosine triphosphate and then this modified base is incorporated into DNA by DNA polymerase III opposite guanine. To a lesser extent, adenine moieties in DNA can be oxidized directly to isoguanine (141).

---

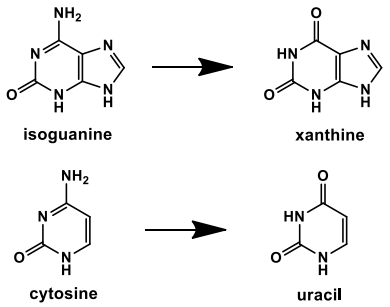
*Part of this chapter is adapted with permission from **Rescue of the Orphan Enzyme Isoguanine Deaminase** Daniel S. Hitchcock, Alexander A. Fedorov, Elena V. Fedorov, Lawrence J. Dangott, Steven C. Almo, and Frank M. Raushel *Biochemistry* **2011** *50* (25), 5555-5557. Copyright 2011 American Chemical Society.

A bacterial enzyme has recently been discovered that catalyzes the deamination of 8-oxoG to urate (4). This discovery suggests that other oxidized nucleotides may be metabolized in a similar manner. However, to the best of our knowledge, an enzyme that is able to catabolize isoguanine has not been identified and characterized. The recently discovered 8-oxoguanine deaminase (8-OGD) was found in cog0402 within the amidohydrolase superfamily (AHS). This superfamily also contains enzymes known to deaminate guanine, cytosine, *S*-adenosyl homocysteine (SAH) and adenosine (2,4-6,14,17,19).

Results and Discussion

A sequence similarity network for cog0402 is presented in **Figure 1.2** at an E-value cutoff of 10^{-70} (10). We postulated that there may be a subset of enzymes within cog0402 that is able to deaminate isoguanine to xanthine as shown in **Scheme 3.1**. The most likely candidate for this activity was originally predicted to be in a group of enzymes related to *S*-adenosylhomocysteine deaminase in Group 1 of **Figure 1.2,** prior to its characterization in Chapter II. Group 1 is the largest and most diverse group of enzymes within cog0402. SAH deaminase utilizes a conserved histidine residue (His-137) to hydrogen bond with N3 of the adenine moiety of SAH for substrate recognition (6). This particular histidine residue is fully conserved across all Group **1** enzymes of cog0402 except for three small subgroups, one of which contains a glutamine at this position. It was initially hypothesized that the carboxamide moiety of the glutamine side chain could be positioned in the active site of these enzymes to hydrogen bond

with the C2/N3 carbamoyl group of isoguanine, although this subgroup was ultimately found to specifically deaminate *S*-adenosylhomocysteine. From this small subgroup of enzymes, an uncharacterized protein from *Picrophilus torridus* was selected for purification and characterization. Since the DNA from this extremophilic archaeon was not commercially available, we purchased the codon-optimized gene (gi|48477797) from GenScript and then attempted to express the protein in an *E. coli* host.



Scheme 3.1

The target protein was largely insoluble when expressed from either a pET-28 or GenScript PGS-21a vector in *E. coli* BL21. However, an isoguanine deaminase activity was detected in cell extracts after centrifugation. The enzymatic activity was not thermostable and we were unable to isolate the protein using standard nickel affinity or GST columns, as would be expected for the recombinant protein with GST and polyhistidine tags. These results suggested that *E. coli* contained a native isoguanine deaminase. In *E. coli* there are two uncharacterized putative deaminases from cog0402 within the AHS. These proteins are YahJ (gi|16128309; Group 9) and SsnA (gi|33347706; Group 5), and the subject of Chapter VI. Both of these proteins were

purified to homogeneity, but no isoguanine deaminase activity could be detected with either enzyme. We therefore attempted to identify the specific enzyme responsible for the isoguanine deaminase activity in *E. coli* through classical purification methods.

E. coli BL21 (DE3) cells were grown in an LB medium to stationary phase and harvested. The isoguanine deaminase activity was determined at each step of the purification scheme by monitoring the decrease in absorbance at 300 nm using a $\Delta\varepsilon = -5.0 \times 10^3\ M^{-1}cm^{-1}$ for the conversion to xanthine. The cells were lysed by sonication and the DNA was removed through precipitation with protamine sulfate. Ammonium sulfate (40% - 50% of saturation) was used to fractionally precipitate the protein mixture. The pellet was redissolved in 50 mM HEPES, pH 7.7, and then loaded onto a HiLoad 26/60 Supderdex 200 gel filtration column. The active fractions from the gel filtration chromatography were pooled and loaded onto a ResourceQ anion exchange column. A 10% SDS-PAGE gel stained with Coomassie Blue revealed three bands, two of which correlated with the activity profile of the ResourceQ fractions with molecular weights of ~50 and ~70 kDa. The specific activity of isoguanine deaminase increased by approximately 1000-fold, relative to the initial cell lysate. The purification results are summarized in **Table 3.1**.

Table 3.1. Purification scheme for isoguanine deaminase.[a]

Step	volume (mL)	units (U)	protein (mg)	U/mg
lysate	75	8.6	490	0.018
protamine sulfate	147	6.2	130	0.048
$(NH_4)_2SO_4$	5	5.4	10	0.53
gel filtration	17	2.2	0.63	3.4
anion exchange	4.5	0.97	0.11	14

[a]A Unit (U) is defined as 1.0 µmol/min.

Table 3.2. Fragments from nanoLC Electrospray MSMS.

Position	Observed mass	Actual Mass
3-14	1317.26 (+2)	1315.25
15-30	1822.05 (+3)	1819.02
31-46	1647.20 (+2)	1645.18
95-103	1012.60 (+2)	1010.59
112-122	1323.68 (+2)	1321.67
193-201	1062.06 (+2)	1060.04
227-236	1059.00 (+2)	1056.96
244-262	2058.06 (+3)	2055.04
375-391	1817.46 (+3)	1814.43
405-427	2570.04 (+3)	2567.02

NanoLC electrospray MS/MS analysis was employed to identify the partially purified protein with the ability to catalyze the deamination of isoguanine. The gel from the SDS-PAGE separation was submitted to the Protein Chemistry Laboratory at Texas A&M University for trypsin digestion and analysis. The 70 kDa band was identified as catalase, and the 50 kDa band was identified as cytosine deaminase from *E. coli* (gi|16128322). The specific peptides that were identified in the mass spectrum are listed in **Table 3.2**.

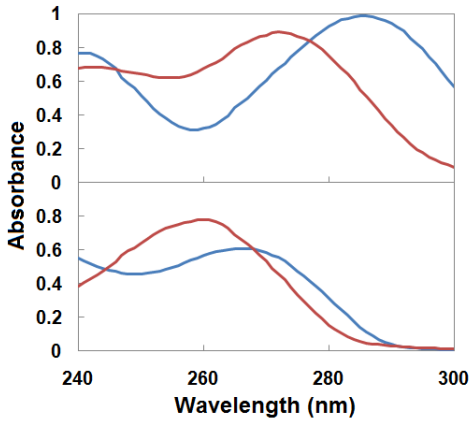


Figure 3.1: Absorption spectra of the reaction mixtures before and after the addition of purified cytosine deaminase to either isoguanine (top panel) or cytosine (bottom panel) at pH 7.7. Isoguanine and cytosine (spectra in blue) were incubated with 1.6 µM CDA for 1.5 hours in 45 mM HEPES, pH 7.7. After the reaction was complete, the UV spectra (shown in red) were collected. The red spectrum in the top panel is consistent with the formation of xanthine (λ_{max} = 272 nm) from isoguanine (λ_{max} = 286 nm). The red spectrum in the bottom panel is consistent with formation of uracil (λ_{max} = 260 nm) from cytosine (λ_{max} = 267 nm).

Cytosine deaminase (CDA) from *E. coli* was cloned and inserted into a pET30 expression vector using standard protocols. The CDA transformed cells were grown in

the presence of 90 µM dipyridyl supplemented with 1.0 mM Zn^{2+} to diminish the incorporation of iron in the active site (124). To ascertain whether CDA is able to catalyze the deamination of isoguanine, 1.6 µM of the purified protein was incubated with 100 µM cytosine or isoguanine and the UV spectra were obtained before and after addition of the enzyme. The spectra matched that of the two expected products, uracil and xanthine, as illustrated in **Figure 3.1**. Also, an activity profile from anion exchange showed the activities for cytosine and isoguanine deaminase coeluted. A structural comparison of cytosine and isoguanine is presented in **Scheme 3.1**.

The kinetic constants for the deamination of isoguanine and cytosine with the purified CDA were determined with a direct spectrophotometric assay at 294 and 255 nm using values for Δε of -6.6 x 10^3 M^{-1} cm^{-1} and +2.6 x 10^3 M^{-1} cm^{-1}, respectively. The kinetic constants for the deamination of isoguanine by CDA are 49 ± 2 s^{-1}, 72 ± 5 µM, and 6.7 (± 0.3) x 10^5 M^{-1} s^{-1} for the values of k_{cat}, K_m, and k_{cat}/K_m, respectively. Under identical reaction conditions the kinetic constants for the deamination of cytosine are 45 ± 4 s^{-1}, 302 ± 44 µM and 1.5 (± 0.1) x 10^5 M^{-1} s^{-1} for the values of k_{cat}, K_m, and k_{cat}/K_m, respectively, at pH 7.7. The values of k_{cat} are nearly identical for the two substrates but k_{cat}/K_m for the deamination of isoguanine is more than 4-fold greater than for the deamination of cytosine.

To further confirm that CDA is the only enzyme within *E. coli* that is capable of deaminating isoguanine we obtained a strain of this bacterium from the KEIO collection containing a knockout of the gene for cytosine deaminase (Δ*cod*A) (142). The Δ*codA E.*

coli cells were grown to stationary phase and lysed. The rate of deamination of isoguanine was measured by following the decrease in absorbance at 300 nm. Those cells lacking CDA had less than 1% of the isoguanine deaminase activity as the wild type strain.

The three-dimensional structure of cytosine deaminase (PDB code: 1K70) has previously been determined in the presence of an inhibitor that mimics the putative tetrahedral intermediate during the deamination of cytosine (3). In this structure, His-246 and Asp-313 are poised to serve as general acid/base groups to activate the metal-bound water molecule and the amino leaving group. In addition, Glu-217 is positioned to deliver a proton to N3 of the pyrimidine ring. The carbamoyl moiety at N1/C2 is recognized via hydrogen bonding interactions with the side chain of Gln-156. The crystal structure of CDA bound with isoguanine was determined and the molecular interactions with isoguanine are shown in **Figure 3.2** (PDB code: 3RN6). The orientation of isoguanine is nearly identical to that of the cytosine mimic in the previous structure. However, an additional interaction to the substrate is formed via a hydrogen bond between Asp-314 and N7 of the purine ring. The adenine deaminase from cog1816 of the amidohydrolase superfamily (PDB code: 3PAN) also possess an Asp-Asp motif at the end of β-strand 8 within the $(\beta/\alpha)_8$-barrel structure that forms a hydrogen bond with N7 of the purine ring.

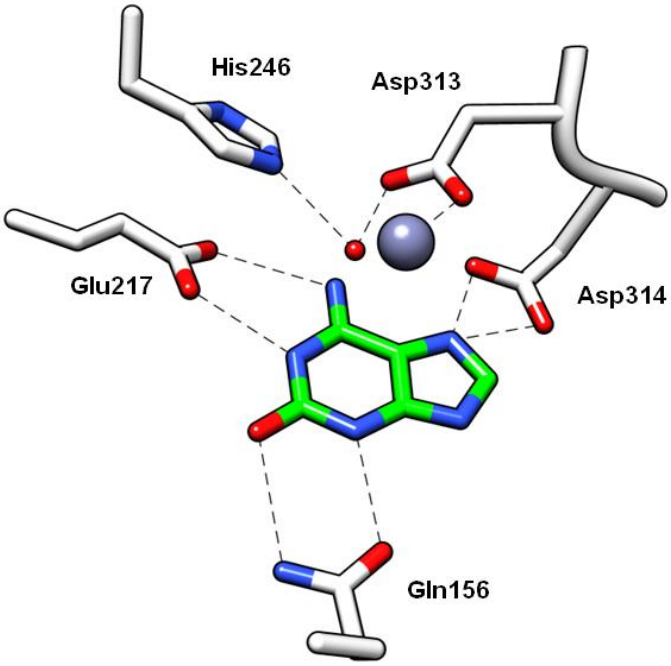


Figure 3.2: The structure of isoguanine (in green) bound in the active site of cytosine deaminase.

We were initially surprised to find that there were no reports of CDA from any source being able to deaminate isoguanine. However, a comprehensive literature search identified a single reference for the deamination of isoguanine by crude extracts of *E. coli* (143). The specific enzyme that was responsible for this transformation has not (until now) been identified in the 60 years since this initial discovery. We have now rescued the orphan isoguanine deaminase and have demonstrated that this enzyme also catalyzes the deamination of the structurally related base, cytosine. It is of interest to note that the value of k_{cat}/K_m for the deamination of isoguanine by CDA is greater than for the deamination of cytosine. Therefore, in *E. coli* the mutagenic base,

isoguanine, can be recycled via the formation of xanthine. It is likely that all of the bacterial cytosine deaminases have the ability to deaminate isoguanine.

CHAPTER IV

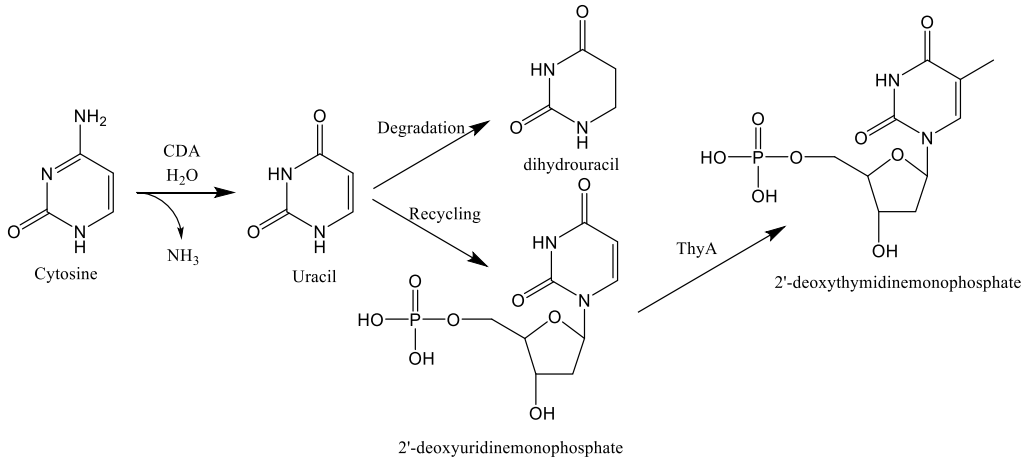
DISCOVERY OF A 5-METHYLCYTOSINE DEAMINASE

Introduction

Thymine, like other nucleobases, is essential for cell proliferation. Most organisms synthesize this metabolite through methylation of 2'-deoxyuracilmonophosphate (dUMP) by thymidylate synthase (ThyA) creating thymine in the form of 2'-deoxythymidinemonophosphate (dTMP) (**Scheme 4.1**) (144), where it is further phosphorylated to 2'-deoxythymidinetriphosphate (dTTP). Thymine starvation, however, differs from other nucleotide starvation. Not only is cell growth stalled, but cell death is induced (145). This phenomenon occurs in all organism, ranging from *E. coli* to cancer cells.

This strategy has been used in a very successful cancer drug, 5-fluorouracil (5FU) (146). The drug is received by cancer patients, where is phosphoribosylated to its active form, 2'-deoxy-5-fluorouridinemonophosphate (FdUMP). FdUMP is an irreversible mechanism based inhibitor of ThyA, and causes thymineless death. Rapidly dividing cancer cells are susceptible to this drug on account of rapid thymine depletion, yet healthy human cells are still killed in the process. One strategy being developed to reduce toxicity to healthy human cells is to utilize the relatively non-toxic 5-fluorocytosine (5FC) as a pro drug (147). This compound is not rapidly deaminated by any human genes products to form 5FU. Cancer cells could be specifically transfected with a 5FC deaminase, which in turn will form 5FU only in tumors.

5FC deaminase activity is minimal in any characterized enzyme, therefore 5FC deaminases have been engineered by directed evolution starting with *E. coli* cytosine deaminase (b0337) (25). It was found that a single active site mutation could confer 5FC deaminase activity. This engineered cytosine deaminase was capable of killing *E. coli* cells with 5FC, and since then more robust mutants mutants have been developed and effectively kill cancer cells by this strategy (148).



Scheme 4.1

Cytosine deaminase (CDA) itself is an integral enzyme in pyrimidine recycling and degradation (**Scheme 4.1**) (149,150). It is a member of the amidohydrolase superfamily, which use a TIM-barrel and a metal center to perform a nucleophilic attack from a coordinated water molecule (1,2). CDA is a member of cluster of orthologous groups 0402 (cog0402), along with guanine deaminase (90), *S*-adenosylhomocysteine deaminase (19) and 8-oxoguanine deaminase (4). These enzymes accomplish the

deamination reaction through a mononuclear metal center which is conserved across the entire cog. With b0337 having a well characterized structure and slight promiscuous activity for 5FC, it provided an excellent target for mutagenesis (3). The first round of mutants to confer 5FC deaminase activity involved modifications at the D314 position—most notably D314S, D314A, and D314G (**Figure 4.1**).

Upon closer inspection of the CDA family, it was found that this mutation already exists. Members were found to possess a Ser or Cys at the position corresponding to D314 in b0337. CDA homologues from Klebsielle pneuominae (KPN_00632) and Rhodobacter sphaeroides (RSP_0341) were cloned and purified and found to deaminate, in addition to cytosine, 5-methylcytosine (5MC) and 5FC. Furthermore, KPN_00632 was found to perform these activities in vivo in E. coli host cells, either killing them in the presence of 5FC or rescuing thymine auxotrophs in the presence of 5MC. The activity for 5MC deaminase could be physiologically relevant. Furthermore, these genes could provide new starting points on directed evolution to evolve more efficient 5FC deaminase for cancer gene therapy.

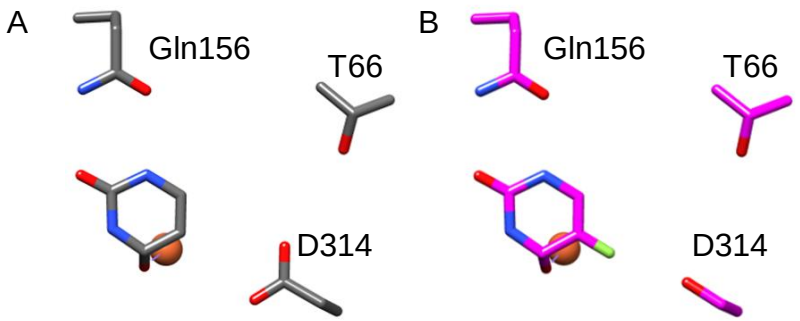


Figure 4.1: Active site of CDA. A) WT CDA bound to inhibitor. B) D314S mutant bound to 5-fluorouracil analogue,

Materials and Methods

Cloning and Purification of NCGl0075, KPN_00632, RSP_0341, and b0337. The genes for NCGl0075, KPN_00632, RSP_0341 were cloned into pET28 with a 6x C-terminal His tag using standard cloning practices. The plasmids were transformed into *E. coli* BL-21 DE3 cells and plated onto LB-agarose containing 50 µg/mL kanamycin. 1 L cultures (LB broth, kanamycin 50 µg/mL) were inoculated from the resulting colonies and grown at 37 °C until the optical density at 600 nm reached 0.6. Protein expression was induced with 1 mM IPTG, and the cultures were shaken for 20 hours at 25 °C. The cultures were centrifuged at 7000 RPM for 15 minutes and the cell pellets were disrupted by sonication in 35 mL of running buffer and PMSF. DNA was precipitated by protamine sulfate. Proteins with a 6x His-tag were loaded onto a 5 mL HisTrap column (GE HEAL/THCARE, USA) with running buffer (20 mM HEPES, 250 mM NaCl, 250 mM NH_4SO_4, 20 mM imidazole, pH 7.5) and eluted with a linear gradient of elution buffer (20 mM HEPES, 250 mM NaCl, 250 mM NH_4SO_4, 500 mM imidazole, pH 7.5).

The gene for b0337 was cloned into pBAD322C. *E. coli* (CGSC# 9145) (151) cells were transformed with the plasmid and plated onto LB-agarose containing 25 µg/mL chloramphenicol. A 1 L culture (LB broth, chloramphenicol 25 µg/mL) was inoculated from the resulting colonies and grown at 37 °C until the optical density at 600 nm reached 0.6. Protein expression was induced with 1 mM L-arabinose, and the culture was shaken for 20 hours at 25 °C. The culture were centrifuged at 7000 RPM for 15 minutes and the cell pellets was disrupted by sonication in 35 mL of running buffer and PMSF. DNA was precipitated by protamine sulfate. The protein were precipitated with 70% saturation NH_4SO_4 and the pellet resuspended in 4 mL of 50 mM HEPES, pH 7.5. The solution was loaded onto a HiLoad 26/60 Superdex 200 gel filtration column. Fractions were pooled and concentrated.

Cloning of KPN_00632, RSP_0341, and b0337-D314A for *in vivo* Assays. The genes for KPN_00632, RSP_0341 were cloned into pBAD322C using standard cloning practices. The b0337 D314S mutant was constructed using site directed mutagenesis by primer overlap extension from the pBAD322C-b0337 plasmid.

Activity Screen and Kinetic Constant Determination. 1 µM purified enzyme was incubated in 50 mM HEPES buffer pH 7.5 at 25 degrees C for 1 hr in a 96-well UV-Vis quartz plate. The spectrum was monitored from 240-350 nm and any spectral shifts were noted as deaminase activity. The library included cytosine, 5-methylcytosine, 5-hydroxymethylcytosine, 5-fluorocytosine, 5-aminocytosine, creatinine, isoguanine, cytosine-5-carboxylate, *N*-methylcytosine, and cytosine-5-formylcytosine.

Positive hits were subjected to kinetic assays. Initial reaction velocities were measured by a UV-Vis direct assay in a 96-well quartz plate for various enzyme/substrate combinations. Substrate concentration was varied while enzyme concentration was held constant. Product formation was monitored at the following wavelengths with experimentally derived molar extinction coefficients for the tested substrates: cytosine (255 nm, $\Delta\varepsilon$ = 2600 M^{-1} cm^{-1}), 5-methylcytosine (262 nm, $\Delta\varepsilon$ = 3700 M^{-1} cm^{-1}), creatinine (240 nm, $\Delta\varepsilon$ = 6100 M^{-1} cm^{-1}), isoguanine (294 nm, $\Delta\varepsilon$ = -6600 M^{-1} cm^{-1}), 5-fluorocytosine (262 nm, $\Delta\varepsilon$ = 3200 M^{-1} cm^{-1}), 5-aminocytosine (235 nm, $\Delta\varepsilon$ = -2000 M^{-1} cm^{-1}), 5-hydroxymethylcytosine (258 nm, $\Delta\varepsilon$ = 3700 M^{-1} cm^{-1}).

5-Fluorocytosine *in vivo* Deaminase Assay. *E. coli* pyrimidine auxotrophs (CGSC# 9145) cells were transformed with pBAD322C and the constructs containing the genes b0337, KPN_00326, and b0337-D314A, and grown on LB-agar plates with 25 µg/mL chloramphenicol. Single colonies were picked and overnight 5 mL cultures were grown in selection media plus 150 µM cytosine and 25 µg/mL chloramphenicol. Selection media consisted of: 1X M9 minimal salts, 1.96 g/L Yeast Synthetic Dropout Medium without uracil, 0.1% glycerol, 100 µM $CaCl_2$, 1 mM $MgSO_4$, and 1X trace elements. 5000x trace elements was created with 95 mL H_2O, 5 g citric acid, 5 g $ZnSO_4*7H_2O$, 4.75 g $FeSO_4*7H_2O$, 1g $Fe(NH_4)_2(SO_4)_2*6H_2O$, 250 mg $CuSO_4*H_2O$, 50 mg $MnSO_4*H_2O$, $Na_2MoO_4*2H_2O$. Yeast synthetic dropout with cytosine and glycerol was autoclaved separately from M9, $CaCl_2$ and $MgSO_4$ were sterile filtered and trace elements 5000x was autoclaved before all were combined. The culture was grown for

14 hrs. 25 mL cultures of selection media + 150 µM cytosine + 25 µg/mL + 100 µM MnSO4 + 100 µM zinc acetate + chloramphenicol were prepared in 100 mL flasks with the following experimental conditions: 1) no modification, 2) 100 µM arabinose (Induced), 3) 50 µM 5-FC + 100 µM arabinose, 4) 110 µM 5-FC + 100 µM arabinose, 5) 320 µM 5-FC + 100 µM arabinose, 6) 500 µM 5-FC + 100 µM arabinose, 7) 1 µM 5FU + 100 µM arabinose, 8) 5 µM 5FU + 100 µM arabinose. The cultures were inoculated with 25 µL of the overnight cultures and shaken at 200 rpm for 12 hrs at 37 degrees C. OD600 was measured every 2 hrs in a 96-well UV-Vis plate with 250 uL of the culture, pathlength of 0.63 cm.

Culture Activity Measurement. 5 mL of the active cultures was centrifuged at 6000 RPM for 10 minutes. The pellets were resuspended in Bugbuster 1x and 50 mM HEPES pH 7.5. The resuspended cultures were incubated at 25 degrees C on a while rotating for 30 minutes. The resulting suspension was centrifuged at 14000 RPM for 10 minutes. 5 µL of the resulting supernatant was combined with 200 µM of the appropriate substrate in a 250 µL reaction in a 96-well UV-Vis quartz plate, and reaction was monitored at the appropriate wavelength. If necessary, the supernatant was diluted in 50 mM HEPES pH 7.5 to allow measurement of initial velocity. All values were adjusted for extinction coefficient and cell density.

5-Methylcytosine *in vivo* Deaminase Assay. *E. coli* thymine auxotrophs (CGSC#: 4091) were transformed with pBAD322C and the constructs containing b0337, KPN_00326, and RSP_0341 and plated on to LB-Agar chloramphenicol (25 µg/mL)

plates supplemented with 200 µM thymine. Single colonies were picked and 5 mL cultures of LB + 100 µM thymine + 25 µg/mL were grown for 14 hours. 25 mL cultures of selection media + 100 µM MnSO4 + 100 µM zinc acetate + 25 µg/mL chloramphenicol were prepared in 100 mL flasks with the following experimental conditions: 1) no modification, 2) 500 µM 5FC, 3) 100 µM arabinose (Induced), 4) 100 µM thymine + 100 µM arabinose, 5) 100 µM 5MC + 100 µM arabinose, 6) 100 µM 5MC + 100 µM arabinose. The cultures were inoculated with 25 µL of the overnight cultures and shaken at 200 rpm for 12 hrs at 37 degrees C. OD600 was measured every 2 hrs in a 96-well UV-Vis plate with 250 µL of the culture, pathlength of 0.63 cm.

Results

Phylogeny of the Cytosine Deaminase Group. A BLAST search was conducted using the sequence of *E. coli* cytosine deaminase was submitted, and 1377 homologues were found. An All-by-All BLAST was conducted with the enzymes, and over 4 groups formed at an E-value stringency of 10^{-140}, which were arbitrarily labeled subgroups (sg) 1-4 (**Figure 4.2**). A sequence alignment was performed and one striking feature was the variation of a residue found directly after the strand 8 metal binding Asp. This residue (D314 in *E. coli* CDA) corresponds to that which was mutated to engineer a 5-fluorocytosine deaminase. Sg1 perserves this Asp, however sg2 posesses a Ser, sg3 a Cys, and sg4 a Ser. Four representatives were selected for cloning and purification: b0337, sg1, *E. coli* cytosine deaminase; sg2, KPN_00632, Klebsiella pneumonia; sg3, RSP_0341, Rhodobacter sphaeroides 2.4.1; sg4, NCGl0075, Corynebacterium

glutamicum ATCC 13032 creatinine deaminase. Despite their location in separate subgroups, sequence identity was high (**Table 4.1**). Additionally, the b0337 D314A mutant was constructed by primer overlap extension from the pBAD322C-b0337 plasmid.

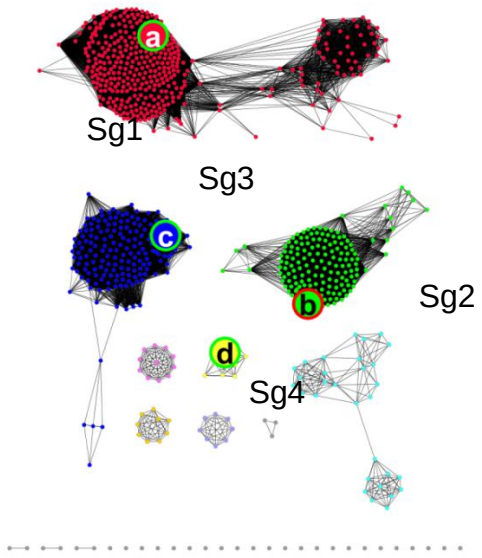


Figure 4.2: Sequence similarity network diagram of CDA homologues a BLAST E-value cutoff 10^{-140}. Proteins characterized in this study are: a) b0037, b) KPN_00632, c) RSP_0341, d)NCGl0057.

Table 4.1. Percent identity between protein sequences.

	RSP_0341	NCGl0075	b0037
NCGl0075	35%		
b0037	40%	42%	
KPN_00632	42%	48%	58%

Enzymatic Characterization. The targets were cloned into pBAD322C, expressed in *E. coli* and purified via gel filtration. B0337 contained 0.68 eq Zn and 0.18 eq Fe. KPN_00326 contained 0.33 eq Fe, 0.20 eq Zn, and 0.15 eq Mn. RSP_0341 contained 0.49 eq Fe, 0.35 eq of Zn, and 0.21 eq Mn. NCGl0075 contained 0.52 eq Fe, 0.19 eq Cu, 0.18 eq Mn, and 0.17 Zn.

B0337 has been previously characterized, however the kinetic constants were redetermined in the buffer conditions used in this experiment. The kinetic constants are summarized in **Table 4.2**. B0337 was found to deaminate only cytosine and isoguanine at appreciable rates. 5-methylcytosine deaminase activity was barely detectable (22 $M^{-1}s^{-1}$) and 5-fluorocytosine demainase activity somewhat higher (250 $M^{-1}s^{-1}$). KPN_00632 was found to deaminate 5-methylcytosine the most efficiently (the first recorded instance of this activity), followed by promiscuous activity for cytosine and isoguanine and 5-fluorocytosine. RSP_0341 was found to have a similar profile for as KPN_00632 except for it preferentially deaminates cytosine over methylcytosine. NCGl0075 was found to have creatinine as its best substrate. D314A activity.

5-Fluorocytosine Toxicity Assay. The empty vector as well as plasmid constructs containing b0337, D314A, and KPN_00632 transformed into *E. coli* CGSC #: 9145, and the constructs were tested for their ability to deaminate 5-fluorocytosine *in vivo*, which would cause cell death or slow growth from inhibition of ThyA (**Figure 4.3**). All constructs, including the empty vector, showed no growth in 1 and 10 µM 5-fluorouracil in 12 hrs. The empty vector showed slowed growth for 5-fluorocytosine,

whereas the effects for 5-fluorocytosine were more pronounced in other cultures. Only KPN_00632 showed no growth at 500 µM 5-fluorocytosine, suggesting it most potently produced 5-fluorouracil in this system. 5-fluorocytosine, 5-methylcytosine and cytosine deaminase activity were measured at 10 hrs for the induced culture and for the 50 µM 5FC culture (**Figure 4.4**).

5-Methylcytosine Deaminase Rescues Thymine Auxotrophs. The empty vector as well as plasmid constructs containing b0337, KPN_00632, and RSP_0341 were transformed into *E. coli* CGSC#: 4091 to test their ability to rescue the thymine autotroph by *in vivo* deamination of 5-methylcytosine (**Figure 4.5**). All strains were capable of growth when supplemented with 100 µM thymine. Neither the empty vector nor b0337 were capable of rescuing growth, even when supplemented with 400 µM 5-methylcytosine. Both KPN_00632 and RSP_0341 showed growth when the gene was induced in addition to 400 µM 5-methylcytosine, and to a lesser extent 100 µM 5-methylcytosine.

Table 4.2 Kinetic constants of the CDA homologues with various substrates.

k_{cat} (s^{-1}) K_m (μM) K_{cat}/K_m ($M^{-1}s^{-1}$)	Cytosine	5MC	Creatinine	Isoguanine	5FC	5AC	N-MCyt	HMC
b0337	33 (7) 400 (100) 8.4 (0.9) x 10^4	22 (1)	0.11 (.03) 830 (250) 1.3 (0.1) x 10^2	3.9 (0.1) 36 (4) 1.1 (0.1) x 10^5	250 (50)	8.0 (0.5) x 10^3	.0017 (.0001) 43 (8) 40. (5)	<10 M-1s-1
KPN_00632	13 (2) 440 (110) 2.9 (0.2) x 10^4	8.4 (1) 25 (7) 3.3 (0.6) x 10^5	6.8 (1.3)x 10^3	4.5 (0.5) 116 (26) 3.8 (0.5) x 10^4	6.5 (0.2) 56 (3) 1.15 (0.05)x 10^5	160 (10) 280 (30) 5.8 (0.3) x 10^5	<10 M-1s-1	2.2 (0.2) x 10^3
RSP_0341	30. (5) 210 (60) 1.4 (0.2) x 10^5	4.8 (.8) 245 (81) 2.0 (0.3) x 10^4	131 (5)	.035 (.003) 240 (30) 1.4 (0.1) x 10^2	25 ± 3 370 ± 70 6.8 (0.5) x 10^4	<10 M-1s-1	<10 M-1s-1	<10 M-1s-1
NCGl0075	0.35 (.05) 300 (90) 1.1 (0.2) x10^3	1.3 (0.2) 900 (200) 1.5 (0.1) x 10^3	32 (4) 510 (90) 6.3 (0.3) x 10^4	0.23 (0.04) 380 (90) 5.9 (0.3) x 10^2	0.055 (0.005) 270 (50) 2.0 (0.2) x 10^2	1.0 (0.1) x 10^3	<10 M-1s-1	<10 M-1s-1

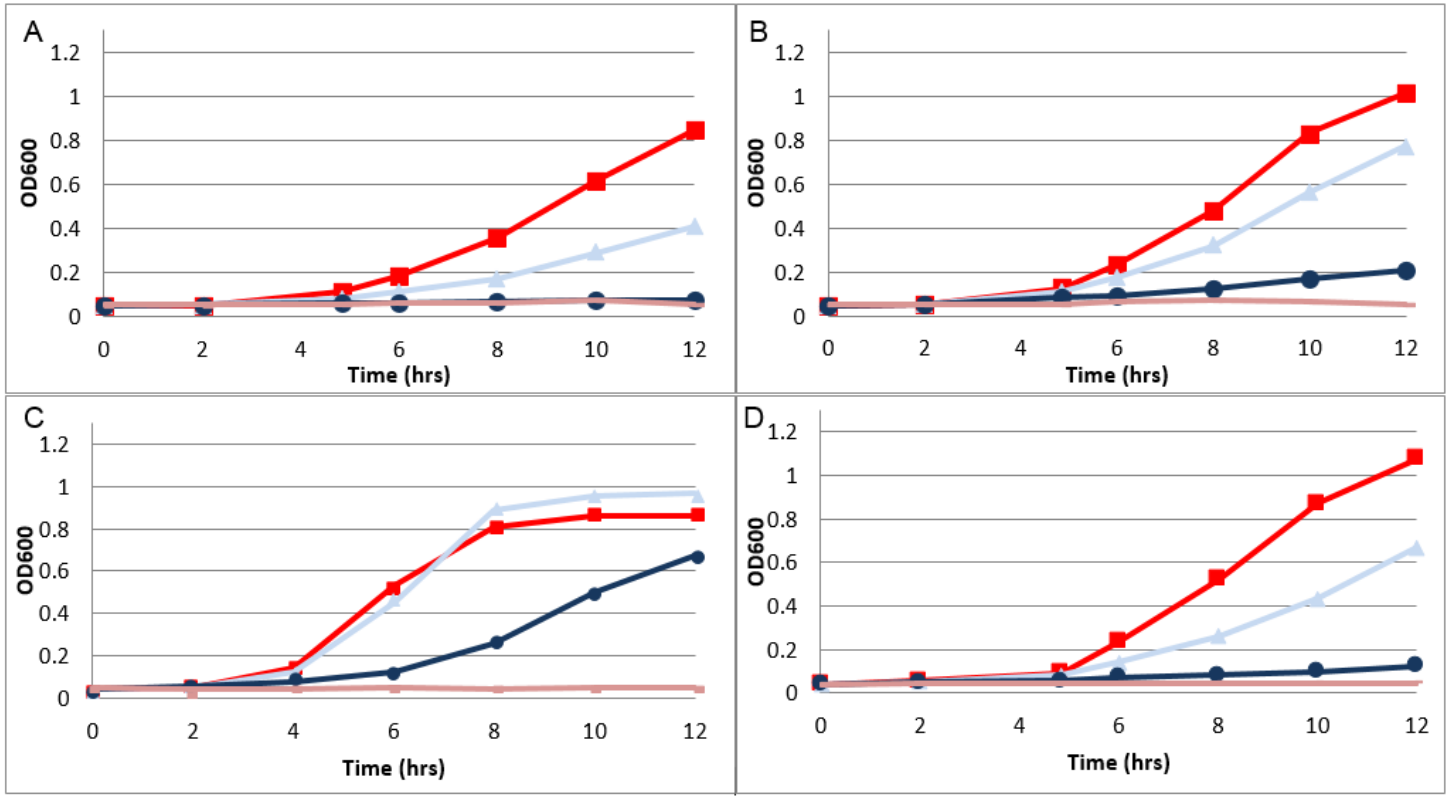


Figure 4.3: Toxicity of 5FC to E. coli with various cytosine deaminases. The experiment conditions are induced with 100 µM arabinose (IND) ( ), IND + 50 µM 5FC ( ), IND + 500 µM 5FC( ), and IND + 5 µM 5-FU ( ). A) KPN_00632 B) b0337, C) b0337 D314A, D) Empty pBAD322C

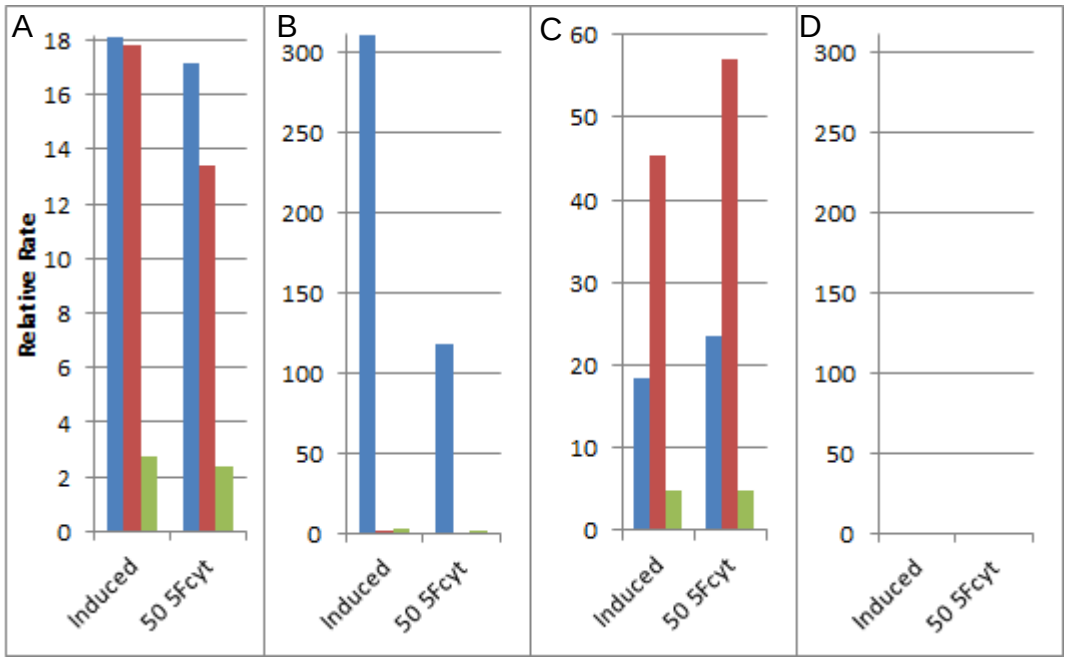


Figure 4.4. Activity for cytosine deaminase in cultures for 5-fluorocytosine toxicity test. Activity for cytosine (■), 5MC(■) and 5FC(■) was measured in the induced culture and induced culture with 50 µM 5FC added. A) KPN_00326, B) b0337, C) D314A, D)pBAD322C.

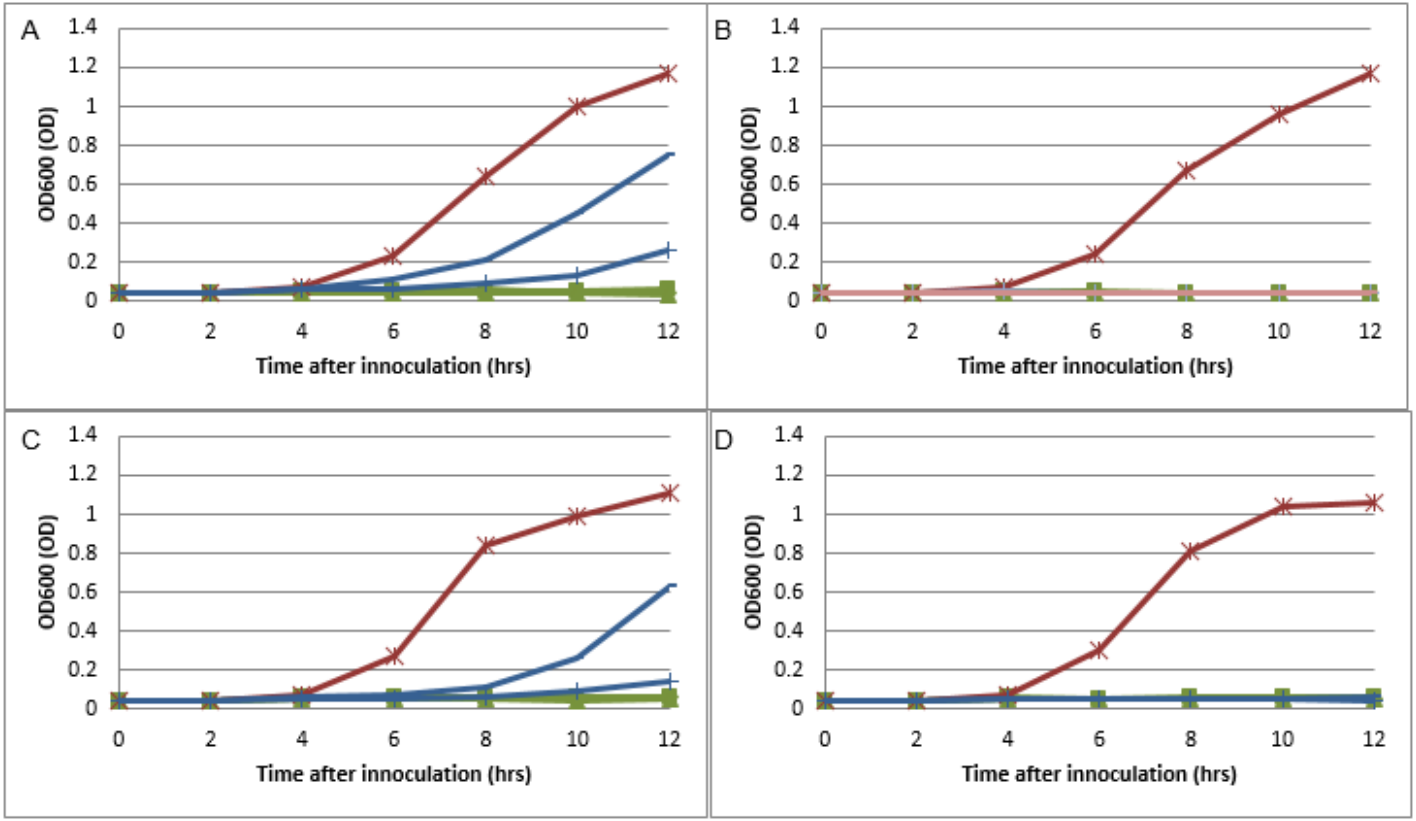


Figure 4.5: Growth of E. coli thymine auxotrophs supplemented with 5MC and various cytosine deaminases. E. coli expressing various deaminases had their growth monitored. Experimental conditions are: no addition ( ), 400 µM 5MC ( ), 100 µM Arabinose (IND) ( ), IND + 100 µM thymine ( ), IND + 100 µM 5MC ( ), IND + 400 µM 5MC( ). A) KPN_00632 B) b0337, C) RSP_0341, D) Empty pBAD322C

Discussion

Substrate Profiles and Structural Assessment. B0337 has several structures with various ligands and even mutants. Given the high sequence similarity between these homologues studied, it is clear a determining factor is the residue corresponding to Asp314. The D314S mutant is capable of binding to 5-fluorouracil, the CDA homologues in this study can be reasoned to have their various activities based on this evidence. The D314S mutant creates a larger pocket for a polar substitution atC5 on the pyrimidine substrate and the structure determined (PDB: 1rak) reflects this. Likewise, the D314G mutant similarly creates a larger pocket for lower steric energy. It remains unclear how b0337 is capable to binding isoguanine, but not the smaller 5-methylcytosine.

KPN_00632 possesses a Ser at this position. Presumably this allows for a bulky methyl group on the C5 position of the substrate. This serine is hydrogen bonded to an Asp corresponding to Asp317 in b0337. Likewise, RSP_0341 utilizes a Cys at this positions. Their substrate profiles are similar, however RSP_0341 is less efficient for 5-methylcytosine, was incapable of catalyzing the deamination of 5-aminocytosine, which is not known to be a naturally occurring compound.

NCGl0075 striking feature is its specificity for creatinine. This enzyme has previously been characterized as a creatinine deaminase. While the enzyme possesses a Ser at the corresponding 314 position, other sequence features may create a preference for smaller substrates. In all other homologues mentioned, the 314 residue

is bonded to Asp317, which in turn is hydrogen bonded to Thr66. This threonine is conserved except in the creatinine deaminase, where a Tyr occupies this position. Aside from potentially disrupting this hydrogen boding network and altering the backbone structure, Tyr could be capable of filling the active site and interfering with larger substrates such as 5-methylcytosine or isoguanine.

The Deamination of 5-Methylcytosine. While 5-methylcytidine deaminase activity has been observed by cytidine deaminase, deamination of the free pyrimidine has never been observed. The direct deamination of this compound results in thymine. Thymine is unique of the nucletodies in that starvation not only stalls cell growth, but actively induces cell death. Given the abililty of KPN_00632 and RSP_0341 to deaminate 5-methylcytosine at physiologically relevant rates, it was of importance to see if this reaction could rescue thymine auxotrophs and if this reaction could be physiologically relevant. BothRSP_0341 and KPN_00632 were capable of deaminating exogenous 5MC and supporting cell growth. B0337 has a very low k_{cat}/K_m for methylcytosine, and this is apparent in in vivo where cell growth does not occur with 5MC. Similarly, no growth is observed from the empty vector. In this case it is possible in KPN_00632 that this reaction is relevant.

Klebsiella pneumoniae additionally possesses a cytosine deaminase homologue which clusters with sg1 in the diagram. This cytosine deaminase would be expected to deaminate cytosine in a similar manner as b0337. KPN_00632 however in its operon contains codB which is a characteristic gene found in group 1 cytosine deaminases.

This gene encodes a cytosine membrane transport, suggesting that KPN_00632 could physiologically deaminated cytosine in addition to 5MC. It is unclear why there would be two deaminases.

The Deamination of 5-Fluorocytosine. The only known instance of the deamination of 5-fluorocytosine is from directed evolution experiments. No naturally occurring enzyme is found to have appreciable activity for this substrate, with the exception of low activity for b0337. The KPN_00632 is roughly 400 times faster than b0337. The test this deamination in vivo, a similar experiment was performed with pyrmidine auxotrophs (*ΔpyrF*) which are forced to use the pyrimidine salvage pathway and phosphoribosylate pyrimidines. If 5FU is formed as the deamination product of 5-fluorocytosine, it will be converted to 2'-deoxy-5-fluorouracil (d5FUMP) and stall growth or kill cells by inhibition of thymidylate synthase. The empty vector was found to not confer 5FC sensitivity, and slowed growth is presumably caused by incorporation of 5FC into DNA. However like b0337 and D314G, KPN_00632 slowed growth of *E. coli*, and at 500 µM 5FC did not allow any growth in 12 hrs. b0337 still slowed growth, and its activity was measured to account for expression levels. The higher expression of b0337 allowed for appreciable rates of 5FC deaminase *in vivo*, being expressed orders of magnitude higher than KPN_00632 (**Figure 4.4**).

Conclusion

Here is the first described described deamination of 5-methylcytosine, and additionally the first naturally occurring enzyme found which can catalyze the

deamination of 5FC. KPN_00632 as well as RSP_0341 may provide promising starting points for additional directed evolution experiments, owing to the already favorable ratios of 5FC/cytosine deamination. Additionally, the discovery of activity for 5-methylcytosine may provide an alternative route for organisms to procure thymine.

CHAPTER V

ASSIGNMENT OF PTERIN DEAMINASE ACTIVITY TO AN ENZYME OF UNKNOWN FUNCTION GUIDED BY HOMOLOGY MODELING AND DOCKING*

Introduction

With increasing availability of genomic sequences, a pressing challenge in biology is a reliable assignment of function to the proteins encoded by these genomes. Functional annotation of an uncharacterized protein can be conveniently accomplished by matching its sequence to that of a characterized protein (152,153). However, this strategy is often inaccurate and imprecise (154,155). Conservatively, over 50% of the sequences in the public databases have uncertain, unknown, or incorrectly annotated functions (156). Annotation of enzymes in functionally diverse superfamilies is particularly challenging (110).

A promising approach to functional assignment is the identification of the substrate by docking potential substrates against the binding site of an experimentally determined enzyme structure (111,157). This approach was used to predict the substrates of Tm0936 from *Thermotoga maritima* (120). High-energy intermediate forms of thousands of candidate metabolites were docked to the X-ray structure of Tm0936, and those highly ranked by docking energy score were tested experimentally,

---

*This chapter is adapted with permission from **Assignment of Pterin Deaminase Activity to an Enzyme of Unknown Function Guided by Homology Modeling and Docking** Hao Fan, Daniel S. Hitchcock, Ronald D. Seidel, II, Brandan Hillerich, Henry Lin, Steven C. Almo, Andrej Sali, Brian K. Shoichet, and Frank M. Raushel*Journal of the American Chemical Society* **2013** *135* (2), 795-803. Copyright 2013 American Chemical Society.

confirming a high deaminase activity against *S*-adenosylhomocysteine (SAH). This approach has been subsequently prosecuted against three other enzymes for activity determination (158-160). These studies suggest that structure-based docking might be a useful tool for enzyme function annotation, when an experimentally determined atomic structure of the enzyme is available.

Often, however, an enzyme of unknown function has no experimentally determined three-dimensional structure. In such situations, a three-dimensional model of the target sequence can be computed by homology modeling if a template structure of a related protein is known (161). Currently, the total fraction of protein sequences in a typical genome for which reliable homology models can be obtained varies from 20-75%, increasing the number of structurally characterized protein sequences by more than two orders of magnitude relative to the PDB (162). Therefore, homology models can, in principle, greatly extend the applicability of virtual screening for ligand discovery (126,163-167). For example, this strategy was successfully employed, in the prediction of function for Bc0371 from *Bacillus cereus* as an *N*-succinyl arginine/lysine racemase by docking dipeptides and *N*-succinyl amino acids to a homology model of Bc0371 based on the X-ray structure of the closest structural homologue, L-alanine-D/L-glutamate epimerase (168).

The accuracy of a homology model can be estimated from the target-template sequence identity. Generally, when the sequence identity exceeds 30%, a reliable alignment can be constructed, and the resulting models are often useful for virtual

screening. When the sequence identity decreases much below 30%, the target structure often deviates significantly from that of the template, resulting in large errors in side chain packing, loop conformations, core backbone conformations, alignment, and even fold assignment. Unfortunately, for many proteins from newly sequenced genomes, only distantly related template structures are available. Hence, it is of pressing interest to develop modeling and docking methods that account for backbone variation between homologues, so that homology models based on distant templates can be used for annotation of protein function.

The functionally diverse amidohydrolase superfamily (AHS) of enzymes provides an ideal test case for developing robust computational methods for function identification of uncharacterized enzymes. Enzymes within this superfamily possess a mononuclear or binuclear metal center embedded within a $(\beta/\alpha)_8$-barrel structural fold (1), and catalyze diverse reactions, including ester and amide hydrolysis, nucleic acid deamination, double bond hydration, carbohydrate isomerization, and decarboxylation (2). To date, more than 24,000 unique bacterial proteins have been annotated to this superfamily, segregated into 24 clusters of orthologous groups (COGs) (113,169).

One of these COGs, cog0402, contains approximately 1400 distinct proteins. At an E-value of 10^{-70}, cog0402 can be divided into 14 major groups of similar sequences (9,10)(**Figure 5.1**). Although 40% of the proteins in cog0402 can be reliably annotated as catalyzing the deamination of an aromatic base or a similar functional group (3,5,11,12,149,159,170), the rest remain uncharacterized. Here, we have attempted to

predict the substrate profile for enzymes of unknown function represented by Group 14 of cog0402, containing approximately 140 proteins. In particular, Arad3529 from *Agrobacterium radiobacter K84* was cloned, expressed, and purified to homogeneity. Substrates of Arad3529 were predicted by docking high-energy intermediate forms of candidate metabolites to the homology models constructed for Arad3529. High-ranking predictions were acquired and tested as candidate substrates by enzymology. The sequence identity between Arad3529 and its closest structurally characterized homologue, cytosine deaminase from *E. coli*, is only 26%, making the construction of the homology model and subsequent computational docking highly challenging.

Materials and Methods

General. All chemicals were obtained from Sigma-Aldrich unless otherwise specified. 7,8-Dihydro-L-biopterin was purchased from Santa Cruz Biotechnology. Formylpterin, pterin-7-carboxylate, hydroxymethylpterin, xanthopterin, 7,8-dihydrohydroxymethylpterin, and 7,8-dihydroneopterin were purchased from Schirks Laboratories.

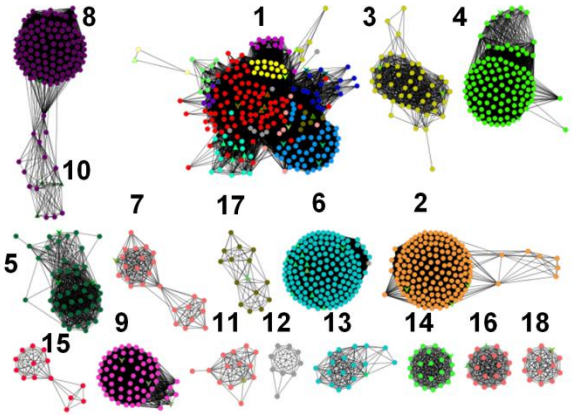


Figure 5.1. Sequence Similarity Network for cog0402. Protein sequences for cog0402 were retrieved from NCBI, subjected to an All-by-All BLAST to determine overall sequence similarity to all of the other proteins within this network(10). Each dot (node) represents an enzyme, and each connection between two nodes (an edge) represents those enzyme pairs that are more closely related than the arbitrary E-value cutoff (10^{-70}). Groups are arbitrarily numbered; those groups with experimentally determined substrate profiles are as follows: (1) *S*-adenosylhomocysteine/5'-deoxy-5'-methylthioadenosine deaminase; (2) guanine deaminase; (4) 8-oxoguanine/isoxanthopterin deaminase; (6) cytosine deaminase; and (8), N-formimino-L-glutamate deiminase.

Cloning, Expression, and Purification of Arad3529. The gene for Arad3529 was amplified from *Agrobacterium radiobacter* Strain K84 genomic DNA using 5'-TTAAGAAGGAGATATACCATGTCATACAGTTTCATGTCCCCGCC-3' as the forward primer and 5'-GATTGGAAGTAGAGGTTCTCTGCTGCGCCAATCACGGTGTCCAG-3' as the reverse primer. PCR was performed using KOD Hot Start DNA Polymerase (Novagen). The amplified fragment was cloned into the C-terminal TEV cleavable StrepII-6x-His-tag containing vector, CHS30, by ligation-independent cloning (171).

The Arad3529-CHS30 vector was transformed into BL21(DE3) *E. coli* containing the pRIL plasmid (Stratagene) and used to inoculate a 5 mL 2xYT culture containing 25 µg/mL kanamycin and 34 µg/mL chloramphenicol. The culture was allowed to grow overnight at 37°C in a shaking incubator. The overnight culture was used to inoculate 1 L of PASM-5052 auto-induction media (172) containing 150 mM 2-2-bipyridyl, 1.0 mM $ZnCl_2$, and 1.0 mM $MnCl_2$. The culture was placed in a LEX48 airlift fermenter and incubated at 37 °C for 5 hours and then at 22 °C overnight. The culture was harvested and pelleted by centrifugation.

Cells were resuspended in lysis buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, and 10% glycerol) and lysed by sonication. The lysate was clarified by centrifugation at 35,000 x g for 30 minutes. The clarified lysate was loaded onto a 5 mL Strep-Tactin column (IBA), washed with 5 column volumes of lysis buffer, and then eluted in StrepB buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, 10% glycerol, and 2.5 mM desthiobiotin). The eluent was loaded onto a 1.0 mL HisTrap FF column (GE Healthcare), washed with 10 column volumes of lysis buffer, and eluted in buffer containing 20 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole, and 10% glycerol. The purified sample was loaded onto a HiLoad S200 16/60 PR gel filtration column which was equilibrated with SECB buffer (20 mM HEPES pH7.5, 150 mM NaCl, 10% glycerol, and 5 mM DTT). Peak fractions were collected, analyzed by SDS-PAGE, snap frozen in liquid nitrogen, and stored at -80 °C. The purified protein was submitted

to ICP-MS for metal content analysis and found to contain 0.8 eq Mn, 0.1 eq Fe, 0.1 eq Ni, and 0.1 eq Zn.

Homology Modeling of Arad3529. The amino acid sequence of Arad3529 from *Agrobacterium radiobacter K84* (gi|222086854) was retrieved from the Structure Function Linked Database (SFLD) (173). Homology models of Arad3529 were generated in four steps. First, the primary sequence was submitted to the PSI-BLAST server at NCBI to search for suitable template structures (174). Cytosine deaminase (CDA) from *E. coli* (PDB id: 1K70) is the most closely related sequence of known structure. Cytosine deaminase is from group 6 of cog0402 and shares a 26% sequence identity to Arad3529 (**Figure 5.1**). In the PDB file, the enzyme is complexed with a mechanism-based inhibitor, 4-(S)-hydroxyl-3,4-dihydropyrimidine. Second, a sequence alignment between Arad3529 and cystosine deaminase was computed by MUSCLE (175) (Multiple Sequence Comparison by Log-Expectation). Third, 500 homology models were generated with the standard "automodel" class in MODELLER (116), and the model with the best DOPE (117) score was selected (Model-1) (**Figure 5.2**). Finally, two loop regions (residues 83-89, 174-186) in Model-1 were refined with the "loopmodel" class in MODELLER, resulting in Models-2, 3, and 4 (**Figure 5.3**). The "loopmodel" class in MODELLER includes several loop optimization methods, which all rely on scoring functions and optimization protocols adapted for loop modeling (176). Side chains in these two loops were optimized using the "side chain prediction" protocol in PLOP (118). The Fe^{2+} ion was included in all modeling steps. The cocrystallized inhibitor from

the template structure was included in the third step for the construction of the initial homology model, but was removed from the modeled active site in the last step.

Docking Screen Against Arad3529. A high-energy intermediate (HEI) library (119,177) that contains ~22,500 different stereoisomers of 4207 KEGG (Kyoto Encyclopedia of Genes and Genomes) molecules (121,122) was screened against the X-ray structure of the cytosine deaminase template (PDB id: 1K70) as a control, and each of the four homology models of Arad3529 using DOCK 3.5.54 (178). The computed poses were subjected to a distance cutoff to ensure that the O^- moiety of the HEI portion of the molecule is found within 4 Å of the metal ion in the active site. The top 500 compounds ranked by the docking score (the sum of van der Waals, Poisson–Boltzmann electrostatic, and ligand desolvation penalty terms) were inspected visually to ensure the compatibility of the pose with the amidohydrolase reaction mechanism. The details of the HEI docking library preparation, the molecular docking procedure, and the protocol for analyzing docking results have been previously described (120,124,125,179).

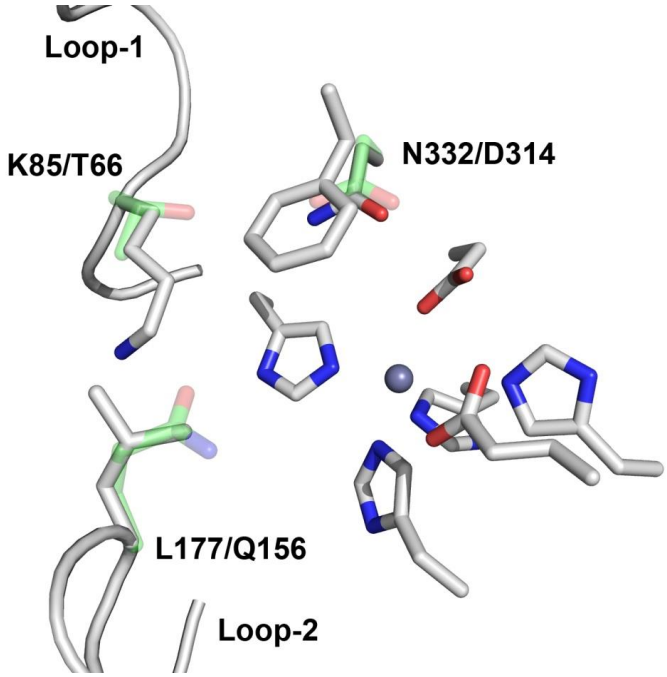


Figure 5.2. The active site of *E. coli* CDA and residues predicted to be important for Arad3529. Residues drawn in white are conserved across all members of cog0402. Thr-66, Gln-156, and Asp-314 in CDA, displayed in green, differed significantly in the sequence of Arad3529 and may be important for substrate binding. Thr-66 corresponds to Lys-85, Gln-156 corresponds to Leu-177, and Asp-314 corresponds to Asn-332. In the X-ray structure of CDA, Gln-156 forms hydrogen bonds with the bound inhibitor; Asp-304 is found within 4 Å of the CDA inhibitor but only forms steric interactions with it; Thr-66 is more distant from the active site. The conversion of a large lysine in Arad3529 suggests that the Lys-85 may reach into the active site and play a role in substrate binding.

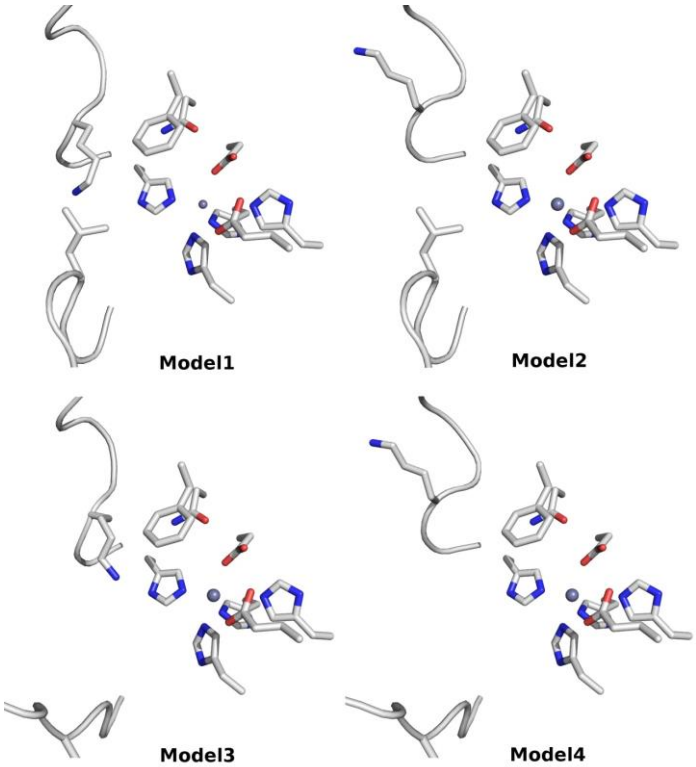


Figure 5.3. The active site conformations of 4 representative models of Arad3529 used in virtual screening. Model-1 is ranked the best by DOPE score among 500 homology models generated automatically by MODELLER. Model-2, Model-3 and Model-4 are different from Model1 in Loop-1, Loop-2, and both Loop-1 and Loop-2, respectively.

Chemoinformatic Analysis of the Docking Hit-Lists. The top 200 highest ranked compounds for each model were combined into ligand sets and used to calculate self-consistency expectation values (*E*-value) within each set using the Similarity Ensemble Approach (SEA) (180,181). Briefly, each ligand was broken into molecular fingerprints, here ChemAxon path fingerprints. The similarity between molecules is quantified by the number of bits they have in common divided by the total number of bits using the Tanimoto coefficient (*Tc*). The sum of all pairwise *Tc* values over a defined cutoff is calculated and compared to what is expected at random between two ligand sets of the

same size. The ratio of the calculated sum of *Tc* values over the sum of *Tc* values expected at random is divided by the standard deviation of the random similarity to give a *Z*-score and plotted against an extreme value distribution to yield an *E*-value. The self-similarity between the ligands of the same set provides a metric to how diverse or similar the ligands are based on the *E*-value. The more significant is the *E*-value (lower), the tighter the chemical space is in the ligand set with likely fewer numbers of distinct chemotypes. Each docking model's self-consistency *E*-value was then compared to the number of true substrates found from each ligand set.

Physical Library Screening. The initial screen of catalytic activity monitored changes in the UV–visible spectrum from 240 to 400 nm after the addition of 1.0 µM Arad3529 to a solution containing 100 µM of the target compound in 20 mM HEPES, pH 7.7 in a 96-well quartz plate using a Molecular Devices Spectramax 384 Plus spectrophotometer. The compounds that exhibited observable spectral change were subjected to more quantitative kinetic assays. On the basis of the docking ranks and compound availability, 11 compounds prioritized by docking were tested as potential substrates, including melamine, cytosine, thioguanine, guanine, adenine, 7-methylguanine, cytidine, 2'-deoxycytidine, pterin, biopterin, and neopterin. In addition, 104 compounds that were substrates of annotated amidohydrolases were tested as potential substrates: guanosine, adenosine, 8-mercaptoguanine, 2,6-diaminopurine, xanthine, 8-oxoguanine, urate, 9-methylguanine, 6-chloropurine, 8-oxoadenine, isopentenyladenine, *N*6-methyladenine, 1-methyladenine, *cis*-zeatin, *trans*-zeatin,

benzyladenine, 2-amino-6-benzylthiopurine, *N*-ethenopurine, *N*6-acyladenine, kinetin, *N*-dimethyladenine, *N*-butyladenine, isoguanine, 2-amino-6-methoxypurine, 6-methylthiopurine, 2-chloroadenine, 6-methoxyadenine, 6-mercaptopurine, 7-methyladenine, 2-dimethylaminoadenine, 3-iminisoindolinone, acycloguanosine, 2'-deoxyguanosine, xanthosine, 2'-deoxyadenosine, 3'-deoxyadenosine, 5'-deoxyadenosine, 2',5'-dideoxyadenosine, 5'-amino-5'-deoxyadenosine, 5'-chloro-5'-deoxyadenosine, 5'-deoxy-5'-methylthioadenosine, *S*-adenosylhomocysteine, *N*6-methyl-2'-deoxyadenosine, isoguanine, 2'-deoxyisoguanine, cytosine-β-D-arabinofuranose, *N*-methylcytidine, 5-methyl-2'-deoxycytidine, 5-hydroxymethylcytidine, 2,4-diaminopyrimidine, 4,6-diaminopyrimidine, 4-amino-2,6-dihydroxypyrimidine, 2-amino-4,6-dihydroxypyrimidine, 5-methlcytosine, 5-hydroxymethylcytosine, 5-carboxycytosine, 5-aminocytosine, 5-fluorocytosine, 5-chlorocytosine, 4-hydroxy-2,5,6-triaminopyrimidine, *N*-methylcytosine, 3-methylcytosine, 2,6-diaminopyrimidine, 2,4,6-triaminopyrimidine, uracil, 5-aminouracil, 5-formyluracil, 5-fluorouracil, 3-oxauracil, 5-hydroxymethyluracil, 6-hydroxyaminouracil, thiamine, 2-aminopyrimidine, 4-aminopyrimidine, toxopyrimidine, sulfamonomethoxine, 4-thiouracil, 2-thiouracil, thymine, 5-azacytosine, ammeline, ammelide, cyanuric acid, 2-amino-4-hydroxy-6-chlorotriazine, cyromazin, *N*-ethylammeline, desethylatrazine, 6-azacytosine, 2-aminopyridine, 2,6-diaminopyridine, 4,6-diamino-2-hydroxpyridine, 4-aminopyridine, 4-dimethylaminopyridine, creatinine, 5-amino-4-imidazolecarboxyamide, 5-hydantoin acetic acid, 3-amino-5-mercapto-1,2,4-

triazole, 3-amino-5-carboxy-1,2,4-triazole, dihydrouracil, dihydrothymine, and dihydroorotate.

After the pterins were identified as substrates, 11 compounds composed of a pteridine ring were added to the screening: 2,4-diamino-6-hydroxymethylpteridine, isoxanthopterin, pterin-6-carboxylate, sepiapterin, folate, formylpterin, pterin-7-carboxylate, hydroxymethylpterin, xanthopterin, 7,8-dihydrohydroxymethylpterin, and 7,8-dihydroneopterin.

Determination of Kinetic Constants. Once candidate substrates had been identified in parallel by docking and plate-based screening, quantitative k_{cat} and k_{cat}/K_m were determined in 20 mM HEPES buffer at pH 7.7. For each compound, the change in the extinction coefficient between the substrate and the deaminated product was determined experimentally by subtraction of the absorbance spectrum of the product from the spectrum of the substrate. The wavelengths (nm) and the differential extinction coefficients, Δε (M^{-1} cm^{-1}) for each of the compounds utilized in this investigation are as a follows: pterin-6-carboxylate (271 nm, 6211 M^{-1} cm^{-1}); pterin (255 and 314 nm, 7021 and 1877 M^{-1} cm^{-1}); biopterin (260 nm, 6527 M^{-1} cm^{-1}); D-neopterin (320 nm, 4431 M^{-1} cm^{-1}); isoxanthopterin (350 nm, 2223 M^{-1} cm^{-1}); sepiapterin (260 nm, 3137 M^{-1} cm^{-1}); folate (364 nm, 4312 M^{-1} cm^{-1}); formylpterin (316 nm, 2963 M^{-1} cm^{-1}), pterin-7-carboxylate (264 nm, 3706 M^{-1} cm^{-1}); hydroxymethylpterin (264 nm, 7356 M^{-1} cm^{-1}); xanthopterin (282 nm, 9704 M^{-1} cm^{-1}); 7,8-dihydrohydroxymethylpterin (536 nm, 3220 M^{-1} cm^{-1}); 7,8-dihydroneopterin (282 nm,

9424 M^{-1} cm^{-1}); and 7,8-dihydrobiopterin (262 nm, 3254 M^{-1} cm^{-1}). The values of k_{cat} and K_m were determined by fitting the parameters in **Equation 5.1** to the experimental data, using SigmaPlot 11, where v is the initial velocity, E_t is enzyme concentration, and A is the substrate concentration.

$$v/E_t = k_{cat} A / (A + K_a) \qquad \textbf{(Equation 5.1)}$$

Confirmation of Product Formation. 100 µM pterin-6-carboxylate was incubated with 150 nM Arad3529 in 50 mM NH4CO3H for 1 h at 25 °C. The reaction mixture was filtered with a 30,000 MW cutoff spin column, and the flow-through was collected. This flow-through and a sample of 100 µM pterin-6-carboxylate was submitted to ESI–mass spectrometry at the Texas A&M Laboratory for biological mass spectrometry. The reaction showed an *m*/*z* change from 206.04 to 207.02, consistent with the product 2,4-dihydroxypteridine-6-carboxylate from the deamination of pterin-6-carboxylate.

Results

Sequence Comparison of Arad3529 with Cytosine Deaminase from *E. coli*. Cytosine deaminase from *E. coli* is currently the closest structurally characterized protein to Arad3529. The four metal binding residues in cytosine deaminase (H61, H63, H214, and D313) align well with H80, H82, H231, and D331 from Arad3529, along with two additional residues (H246 and E217 aligned with H263 and E234 in Arad3529, respectively) that function as proton shuttles in cytosine deaminase. Because these

active site residues to perform a deamination were all intact in Arad3529, it is very likely also a deaminase. However, in Arad3529 there are three notable changes in the amino acid sequence, relative to that found in most other cytosine deaminases from cog0402. In Arad3529, Thr66, Gln156, and Asp314 from CDA are replaced by Lys85, Leu177, and Asn332, respectively (**Figure 5.2**). In CDA from *E. coli*, Thr66 located near the active site does not appear to hydrogen bond with cytosine. Conversely, Gln156 forms a pair of hydrogen bonds to the carbamoyl moiety of cytosine, while Asp314 is located immediately after the invariant Asp313 and within 4 Å of the bound inhibitor in the CDA structure. The substitutions to the active site residues thus suggest that Arad3529 deaminates a substrate different from cytosine.

Molecular Docking. In the initial homology model of Arad3529 (Model-1), which overall adopts the same backbone conformation as the template, changes are located in two unstructured loop regions that help define the active site (Loop-1, residues 83–89; Loop-2, residues 174–186). Therefore, Loop-1 and Loop-2 were remodeled separately, resulting in Model-2 and Model-3 (**Figure 5.3**). Finally, the Loop-1 conformation in Model-2 and the Loop-2 conformation in Model-3 were combined in Model-4. Thus, four homology models were obtained that are distinct from each other in their active site conformations.

Because homology models are constructed from template structures, the models often share some structural features with the templates, and thus recognize similar ligands. Here, however, small molecules that are highly ranked by the template

structure (PDB id: 1K70) are unlikely to be the true substrate of the remote homologue. Therefore, before targeting the homology models of Arad3529, the template structure 1K70 was docked against as a negative control. 57,672 high-energy intermediates (HEI) were docked into the template structure. As expected, cytosine, the natural substrate of the template, was ranked highly (10th out of 57 672), as were several other nucleosides (**Table 5.2**). These molecules were deprioritized for testing against Arad3529; we instead looked for molecules with high differential ranks between the models and the template docking screens.

The same HEI database was docked to each of the four homology models of Arad3529. From each model, putative substrates were selected from the top 500 scored molecules (**Table 5.1**). The compound sets selected using different homology models are overlapping but distinct. For Model-1, the docking hits were dominated by relatively small compounds with a 1,3,5-triazine skeleton. This pattern is also observed in the docking hits from Model-2. However, compounds composed of the purine or pteridine ring were also favored by Model-2. For Model-3, the 1,3,5-triazine skeleton completely disappeared, and the docking hits were dominated by compounds composed of the pteridine ring. Model-4 shared the pteridine pattern with Model-3, but also favored nucleosides having a ribose group.

Table 5.1. Virtual screening of the HEI database against four homology models of Arad3529*

Name	KEGG ID	Docking rank			
		Model-1	Model-2	Model-3	Model-4
melamine	C08737	10			
cyclopropylmelamine	C14147	13	18		
diethylatrazine	C06559	28	32		
Diisopropylatrazine	C06556	29	33		
Diisopropylhydroxyatrazine	C06557	31			
cytosine	C00380	52	45	185	126
2-amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine	C01300	108 (N1)	135 (N1)	10 (N1)	12 (N3)
thioguanine	C07648		11		
Guanine	C00242		20		
pterin	C00715		31 (N3)	27 (N1)	102 (N3) 127 (N1)
phenazopyridine	C07429		50		
biopterin	C06313		66 (N3)	14 (N1)	51 (N1) 71 (N3)
adenine	C00147		99		
neopterin	C05926			15 (N1)	143 (N1)
7-methylguanine	C02242				23
deoxycytidine	C00881				49
cytidine	C00475				75
5'-deoxy-5’-fluorocytidine	C16635				132
2-aminoadenosine	C00939				206

*The names of three molecules that were found in the control are written in italic. The 2-aminopteridine-4(3H)-one and the 2-aminopteridine-4(1H)-one tautomers were attacked by the hydroxide from the *re*-face and the *si*-face, and have the N-1 ring nitrogen and N-3 ring nitrogen protonated by Glu234, resulting in two intermediates named as N1 and N3, respectively.

Table 5.2. Virtual screening of the HEI database against the template structure 1K70

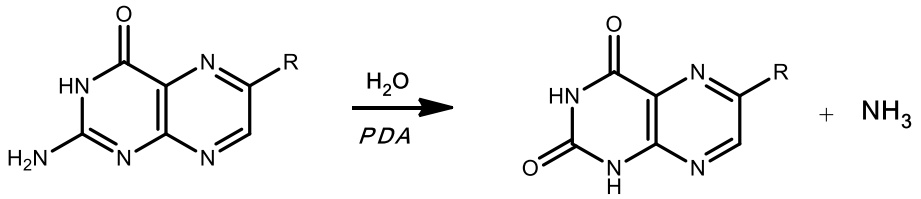
Name	KEGG ID	Rank
cytosine	C00380	10
creatinine	C00791	32
5-methylcytosine	C02376	48
adenine	C00147	50
N6-methyladenine	C08434	59
2'-deoxycytosine	C02026	73
guanine	C00242	77
toxopyrimidine	C01279	126

Chemoinformatic Analysis of the Docking Hit-Lists. On the basis of docking energies alone, it is difficult to prioritize one of these overall hit-lists over another, although the geometry of the interactions does provide guidance (below). One can, however, pose extra-thermodynamic criteria to judge among them. One such is to expect that in a well-behaved docking hit-list, the highly ranked molecules will resemble one another, and thus the hit-list will resemble itself by chemical similarity. We therefore compared the top-ranked 200 molecules for each of the four models of Arad3529 against themselves using the Similarity Ensemble Approach (SEA) (180,181). This was quantified by an expectation value (*E*-value) of the likelihood the self-similarity of each hit-list would occur at random. Of the four models, Model-3 and Model-4 had more self-consistent hit-lists (SEA *E*-value 1.5×10^{-100} and 4.9×10^{-98}, respectively) as compared to Model-1 and Model-2 (SEA *E*-value 2.7×10^{-89} and 1.2×10^{-85},

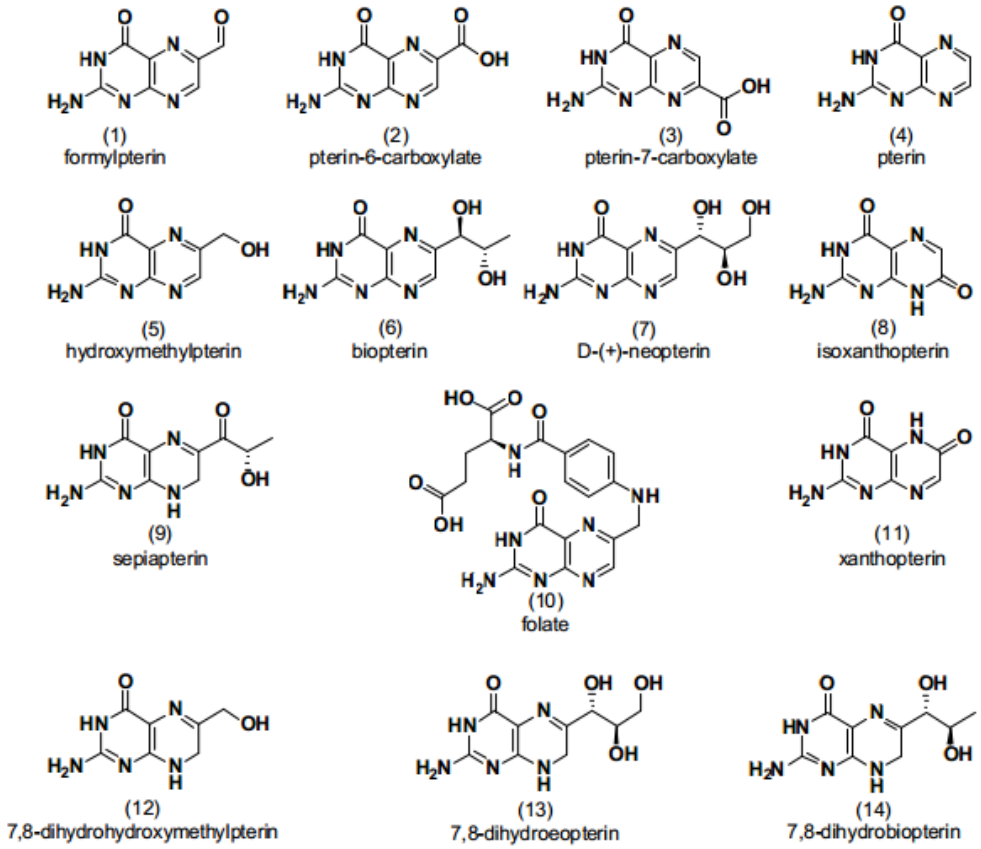
respectively). Comfortingly, the structures of what turned out to be the true pterin substrates of the enzyme were docked in catalytically competent geometries in Model-3 and Model-4. Naturally, this criterion has nothing to recommend it other than consistency, but it may merit further study as a metric to select multiple possible hit-lists for targets adopting different structures.

Substrate Profile for Arad3529. On the basis of commercial availability, eight compounds predicted to be potential substrates for Arad3529 were tested as enzyme substrates, as were three "negative control" nucleosides that also ranked well against the CDA template (**Table 5.1**). Arad3529 was found to deaminate three of the eight: pterin, biopterin, and neopterin with k_{cat}/K_m values of greater than 10^6 $M^{-1}s^{-1}$. We docked the ground states of three substrates, pterin, biopterin, and neopterin, to Model-3. The docking scores of these ground states (−19.29, −27.30, and −23.08, respectively) were substantially higher (worse) than the corresponding HEI states (−53.83, −58.45, and −58.26, respectively), consistent with the catalytic mechanism. The pterin deaminase (PDA) reaction is illustrated in **Scheme 5.1**. To investigate the mechanism further, the enzyme was screened with other substituted pterins and was found to deaminate several pterins (**Scheme 5.2**). Various substitutions on C6 of the pteridine ring are allowed (**Table 5.3**), and with the exception of sepiapterin, the fully aromatic pteridine ring is preferred. 7,8-Dihydroneopterin and 7,8-dihydrobiopterin were deaminated very slowly, and the deamination of dihydrofolate or

tetrahydrofolate could not be measured (<0.0008 s^{-1}). 2,4-Diamino-6-hydroxymethylpteridine was not deaminated.



Scheme 5.1

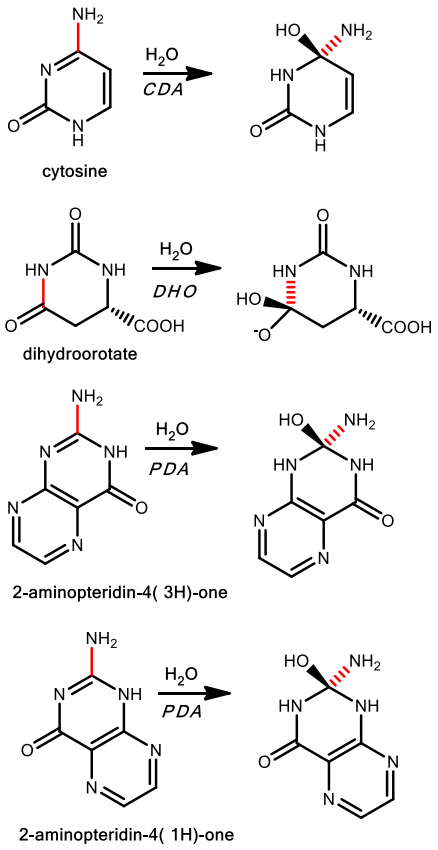


Scheme 5.2

Variations on Binding Orientation of Pterin. In all of the structurally characterized members of the amidohydrolase superfamily, the *re*-face of the carbonyl group of amide and ester substrates is presented to the attacking hydroxide; the same orientation was observed for the amidine or guanidine moieties of substrates in the deamination reactions. In docking screens against Arad3529, we observed two chiral intermediates of pterin (named as N1 and N3), that were generated by hydroxide *re*- and *si*-face attacking two pterin tautomers 2-aminopteridin-4(3*H*)-one and 2-aminopteridin-4(1*H*)-one, and by Glu234 protonating the N-1 and N-3 nitrogen in the pteridine ring, respectively (**Scheme 5.3**). In particular, N3 was selected by Model-2, N1 was selected by Model-3, and both intermediates were selected by Model-4 (**Table 5.1**). In Model-2, the docking pose of N3 had a stranded C-4 carbonyl group, which in N1 forms hydrogen bonds with Asn332 in Model-3 (**Figure 5.4**). This C-4 carbonyl group and the N-5 ring nitrogen in N1 also hydrogen bonded with Lys85 in Model-3. Furthermore, it is known that the 2-aminopteridin-4(1*H*)-one tautomer is less stable than the 2-aminopteridin-4(3*H*)-one tautomer (in *ab initio* and free energy perturbation calculations the ratio is 1:6 (182)). It is clear that the N1 orientation is the true binding motif, and that an important feature for recognition of a pterin deaminase is the DN dyad (D331, N332) found on β-strand 8 and the Lys85 found on β-strand 1.

Table 5.3. Catalytic constants for substrates of Arad3529

Substrate	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (μM)	Docking rank*
formylpterin	5.2 (0.5) x 10^6	64 ± 12	12 ± 2	
pterin-6-carboxylate	4.0 (0.2) x 10^6	110 ± 3	27 ± 2	
pterin-7-carboxylate	3.7 (0.2) x 10^6	48 ± 2	13 ± 1	
pterin	3.3 (0.3) x 10^6	131 ± 4	39 ± 4	27
hydroxymethylpterin	1.2 (0.1) x 10^6	28 ± 1	23 ± 3	
biopterin	1.0 (0.1) x 10^6	46 ± 4	47 ± 9	14
D-(+)-neopterin	3.1 (0.1) x 10^5	19 ± 1	61 ± 7	15
isoxanthopterin	2.8 (0.2) x 10^5	1.6 ± .03	5.7 ± .5	
sepiapterin	1.3 (0.4) x 10^5	2.9 ± .3	22 ± 9	
folate	1.3 (0.2) x 10^5	6.4 ± .5	50 ± 10	
xanthopterin	1.17 (.08) x 10^5	0.46 ± 0.01	40.0 ± 0.4	
7,8-dihydro-hydroxymethylpterin	3.3 (0.4) x 10^4	1.23 ± 0.09	37 ± 7	
7,8-dihydroneopterin	2.6 (0.3) x 10^2	0.036 ± 0.008	200 ± 100	
7,8-dihydrobiopterin	--	~0.01	--	



Scheme 5.3

Discussion

Three techniques were crucial for what turned out to be the correct prediction of pterin deaminase, activity for Arad3529. First, to leverage a distantly-related template, detailed modeling of two active-site loops was required to find models that were at once structurally sensible and catalytically competent, and that could discriminate new substrates from these precedential from the template. Second, the selectivity of models was enhanced by negative control docking screens against the

template structure; we sought substrates ranked highly by the models and poorly by the template. Finally, a close cycle of bioinformatics, biophysical modeling, and enzymology made this approach pragmatic. Moreover, identifying substrates for Arad3529 allows the functional annotation of about 140 previously uncharacterized amidohydrolase enzymes in Group 14 of cog0402. These proteins may be used by bacteria to salvage oxidized pterins, and form the pterin degradation pathway with neighboring genes.

Enzyme Specificity Recapitulated by Loop Modeling. The substrate specificity of enzymes comes from the corresponding compositions of binding-site residues. Remarkable changes in the binding-site residues between two enzyme sequences lead to differences in their binding-site structures. However, a homology model tends to inherit the backbone conformations of aligned regions from the template. Therefore, a homology model of the target enzyme often contains errors in the packing of side chains, and backbone conformations of structurally undefined regions (loops), especially when computed based a remote template. Accurate substrate prediction by docking against homology models of unknown enzymes requires precise positioning of the binding site residues that interact with the new substrates. In this study, we attempted to recapitulate binding-site features in the native structure of Arad3529, by remodeling two binding-site loops in the initial homology model. Because no target ligand is known, three new models that are structurally diverse were selected and docked against independently, followed by combining the screening results. True

substrates of the target enzyme Arad3529, pterins, were not captured by the initial homology model but these remodeled, demonstrating the necessity of the directed refinement of binding-site loops that contain residues different from the template.

Docking Against Template as a Negative Control. Given the definition of structure-based docking screens, the highly ranked compounds can be seen as fingerprint of the protein structure that was docked against. When a homology model of the target enzyme is computed based a distantly related template structure, on one side, the target enzyme very likely have different substrates from those of the template; on the other side, however, the homology model of the target even after refinement could still share some structural features with the template and thus prioritize similar compounds as the template structure in docking screens. In this study, we first generated the docking fingerprint of the template structure, and then compared to what the homology models of the target enzyme Arad3529 predicted. Several compounds, that were ranked highly by the template, were also prioritized by the homology models. However none of these compounds was confirmed as substrates of Arad3529 by enzymology testing. This result suggests that, for enzyme substrate prediction by homology model-based docking, it may be useful to account for compounds that are recognized by the template as negative control, especially when there is only a remote relationship between the target and the template sequence.

Additional Enzymes in Cog0402 that Catalyze Pterin Deamination. Figure 5.1 depicts 22 sequences in Group 14. To identify a more comprehensive list of proteins in

the current databases that share the same substrate profile as Arad3529, a BLAST search was conducted with the sequence of Arad3529. Of the top 200 hits, 165 proteins contained the characteristic β-strand 8 "DN" dyad and a total of 139 of these would populate Group 14 if included in the network diagram. The remaining 26 sequences cluster with the currently uncharacterized Group 13. Group 13 is closely related to cytosine deaminase and also possesses a mixture of enzymes that have either a "DN" or "NN" dyad. While the enzymatic function of Group 13 is unknown, all members of Group 14 are predicted to deaminate pterins. A list of these enzymes is found in Appendix B. Of course, this remains just a prediction, because even proteins with high sequence identity sometimes have different activities (18).

Biological Relevance of Substrates, Genomic Context in *Agrobacterium radiobacter*. Pterin rings form the backbone of the coenzymes folate and tetrahydrobiopterin. The 7,8-dihydro forms of hydroxymethylpterin and neopterin are intermediates in folate synthesis, whereas sepiapterin is an intermediate in the synthesis of biopterin. During the biosynthesis of these cofactors, the pterin substructure remains reduced in the 7,8-dihydro form. Fully oxidized pterins, formed by oxidation of 7,8-dihydropterins (183), are not known to be salvaged by reduction (184). Isoxanthopterin is formed by oxygenation of pterin by xanthine dehydrogenase (185), while pterin-6-carboxylate, formylpterin and pterin can be formed by photolysis of folate (186,187). It seems possible that Arad3529 and its orthologs are used by the bacteria to degrade oxidized pterins.

The gene for Arad3529 is situated in an apparent operon along with the genes TauC, GlcD and MdaB, which is flanked by TauA, TauB, LysR, and a protein from cog1402 (**Figure 5.4**). TauA, TauB and TauC resemble ATP-cassette binding transporters and LysR is a transcription factor. MdaB, GlcD and cog1402 are uncharacterized proteins resembling quinone oxidase, glycolate oxidase D, and creatininase, respectively. MdaB is predicted to contain an NAD(P)H binding domain, GlcD is predicted to be a flavin dependent redox enzyme, and the cog1402 protein will likely catalyze the hydrolysis of an amide bond. These enzymes may form a degradative pathway for catabolizing pterin rings. To our knowledge, the only known instance for the degradation of a pterin ring is by the soil bacteria *Alcaligenes faecalis* (188). In this bacterium, isoxanthopterin is deaminated to form 7-oxylumazine, which is further oxidized to 6,7-dioxylumazine (tetraoxypteridine). An isomerase then cleaves the C6/C7 bond, and reattaches forming xanthine-8-carboxylate (189). Xanthine-8-carboxylate is then decarboxylated, forming xanthine (190). Unfortunately, the recently sequenced genome of *A. faecalis* does not possess an Arad3529 homologue or any of its operon-related genes, and the protein that deaminates isoxanthopterin remains to be discovered.

Figure 5.4. Operon of Arad3529**.**

Certain caveats merit mentioning. Here, homology models of Arad3529 were predicted from a cytosine deaminase structure in complex with a mechanism-based inhibitor. This likely gave us an advantage over modeling from a ligand-free template. When such an apo- template is the only choice, enhanced sampling may be needed (166). Inevitably, this will lead to more models. To judge among these possible structures, and the docked molecules to which they lead, the criterion of a self-similar hit-list may be useful, as we found it to be. Admittedly, this metric remains largely untested, and its only theoretical virtue is that it quantifies an expectation that a well-behaved docking calculation should find putative ligands that resemble one another. From a docking standpoint, we have continued to use high-energy intermediate forms of substrates, rather than their ground states. This seems well-justified for amidohydrolases, where we have a good understanding of the structures that such intermediates adopt, and confidence that the chemical step is rate determining for the reactions the enzymes catalyze. For other families of enzymes, our ability to anticipate high-energy structures, and confidence that the chemical step is rate limiting, may be more limited. Finally, docking screens, in our hands, retain a human element as a final arbiter of which among the tens-to-hundreds of high-ranking metabolites to actually test experimentally. We have argued that human inspection integrates aspects of docking that are captured by the experience of the trained modeler, enzymologist, or medicinal chemist that are difficult to fully capture algorithmically (191).

Conclusion

Members of the amidohydrolase superfamily have been found in every sequenced genome; the functions of most of these proteins remain unknown. Here we focused on one amidohydrolase, Arad3529 from *Agrobacterium radiobacter K84,* which represents proteins from Group 14 in cog0402. Proteins in Group 14 were completely uncharacterized, with sequence similarity and knowledge of the mechanistically related chemistries within the superfamily wrongly suggesting annotations as chlorohydrolases or cytosine deaminases. A multidisciplinary approach was applied to the function annotation of Arad3529, integrating three-dimensional protein structure prediction by homology modeling, virtual metabolite screening by structure-based small molecule docking, and physical library screening by quantitative kinetic assays. The true substrates of Arad3529, a set of modified pterins, were prospectively predicted by homology model-based docking, and were efficiently deaminated by Arad3529 in quantitative kinetic assays. On the basis of the conservation of characteristic residues that interact with substrates in the docked structure, about 140 other previously unannotated amidohydrolase enzymes from different spieces may now be assigned as pterin deaminases. This approach may be useful in the discovery *in vitro* enzymatic and *in vivo* metabolical functions of unknown enzymes discovered in genome projects, especially for those targets with marginal sequence identities to template structures of know functions.

CHAPTER VI

ATTEMPTS TO ASCERTAIN THE ENZYMATIC ACTIVITY OF *E. coli* GENES SSNA AND YAHJ

Introduction

E. coli possesses four enzymes from cog0402. Two are well studied and characterized, and include cytosine/isoguanine deaminase (3,7,24,150) and guanine deaminase (11). The other two enzymes are SsnA and YahJ. SsnA is found in Group 5 (**Figure 1.2**), where it clusters with group 1. YahJ is found in Group 9 and is more closely related to Group 6, cytosine deaminase. While the functions of these genes remain unknown, there are several clues which may help identify the substrates of these enzymes.

Results and Discussion

Library Screens. A large library of compounds was screened to be deaminated by YahJ and SsnA using either a UV-Vis direct assay, or coupling release of ammonia to glutamate dehydrogenase, resulting in the oxidation of NADH (GLDH). These compounds tested and their structures are presented in Appendix C.

The SsnA Gene. SsnA is a gene encoding a 51 kDa protein. This enzyme possesses the catalytic machinery characteristic of all cog0402 members (**Figure 1.1**). Additionally it has been purified, containing 1 equivalent of manganese in its active site. Aside from its purification, much remains unknown about its function *in vivo*. A single *in vivo* study had been performed involving the overexpression and knockout of SsnA

(192). In this study it was determined that SsnA is regulated by RpoS and that expression is induced at early stationary phase. Disruption of the SsnA gene resulted in more colony forming units forming from cultures in stationary phase than wild type. Constitutive expression of SsnA in an expression vector slowed cell growth when compared to SsnA fragments being expressed. These results suggest that SsnA may play a role in the decline of cell viability during stationary phase.

SsnA Genomic Neighborhood. Analyzing the SsnA gene for genomic context reveals a set of genes which may be associated with SsnA. The genomes of 27 organisms containing a homologue of SsnA were compared and any homologues were noted. 56% of the organisms possessed a homologue of XdhA (b2866). In *E. coli*, deletion of this gene stops production of CO2 from adenine, suggesting it may be a xanthine dehydrogenase (193). The mutants also displayed an increased sensitivity to adenine, which was partially reversed by addition of guanosine. A homologue of b2867 is found in 48% of organisms, which is predicted to be the FAD binding subunit of b2866. A third protein is b2867, which is predicted to be the iron sulfur cluster binding subunit of this complex. It is found in 59% of organisms. 44% contain all three of these proteins adjacent to each other. B2870 has been studied and a structure has been determined (194). It shares homology with carbamoyltransferases, although it displays no activity with a variety of compounds, including amino acids. A homologue of this gene is found in 89% of the compared organisms. B2871 is found in 63% of organisms with SsnA, and is annotated as a PLP dependent diaminopropionate ammonia lyase

(195). Its structure has been determined (196). B2872 is found in 85% of organisms, annotated only as a peptidase. It shares 23% sequence identity to succinyl-diaminopimelate desuccinylase. B3880 and b2881 are found in 44% and 59% of the organisms with an SsnA homologue. They are annotated only has a dehydrogenase/oxidase and not much is known about them. 63% possess b2874, which shares 49% identity to a confirmed carbamate kinase (PDB: 1B7B). 56% possess a homologue of b2878, which bares only 28% identity to an uncharacterized gene labelled "xanthine dehydrogenase (PDB: 3ON5)." 81% of the organisms possess a homologue of b2873. This enzyme has been characterized as a D-phenylhydantoinase (197), however d-phenylhydantoin is not known to be naturally occurring. It is possible its physiological function has not been discovered. B2877, found in 56% of the genomes, is confirmed to catalyze molybdopterin cytidylyltransferase, forming the molybdopterin cytidine dinucleotide cofactor (198). This cofactor is essential for xanthine dehydrogenase, which may explain its proximity to the oxidoreductases in the genomic neighborhood. B2878 is found in 63% of the genomics discussed. It is an uncharacterized dehydrogenase with an iron sulfur binding domain.

SsnA Sequence and Structure. SsnA homologues populate group 5 in cog0402. This group is closely related to the group 1 proteins discussed in Chapter II and even clusters with this group at an E-value of 10^{-50}, however activity has never been found with any adenosine derivatives. Of course this is perfectly reasonable, considering the specific substrate requirements of certain group 1 proteins to deaminate only guanine

or 8-oxoadenine and their high similarity to other group 1 proteins. The hallmark of group 1 adenosine deaminases is a histidine found on beta strand 4 which binds to N3 of the purine ring and a glutamate found on beta strand 1 which binds to the 2' and 3' hydroxyls on the ribose moiety. In SsnA, this histidine is conserved however the residue corresponding to the glutamate exists as an aspartate. Shown in **Figure 6.1** is an active site alignment of the structure of *E. coli* SsnA (green) and of its closest liganded homologue, a group 1 methylthioadenosine deaminase (MTADA) from *Pseudomonas Aeruginosa Pao1* (127) (PDB: 4GBD)(cyan) and a transition state analogue (pink). The active sites align well, however the positioning of the aspartate suggests it may play a different role, if any, in substrate binding.

The surface also reveals some of the differences between SsnA and the group 1 proteins. **Figure 6.2** shows an active site cavity which is open to the solvent, unlike the Group 1 enzymes. The ribose binding pocket is much smaller and likely would not be able to fit a ribose. A new cavity forms near the C4 position on the adenine ring, possibly allowing extra space for substituents on that ring position. Of course it should be noted the structure could change substantially if bound to a ligand.

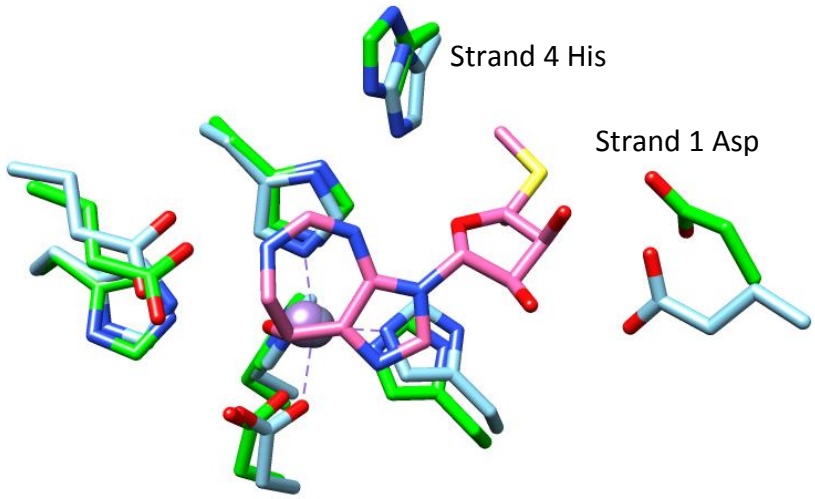


Figure 6.1: Active site of SsnA. The active site of SsnA (green) is aligned with the structure of MTADA (PDB: 4GBD) (blue) along with the ligand of MTADA (pink). Residues are labelled according to the characterized structure MTADA.

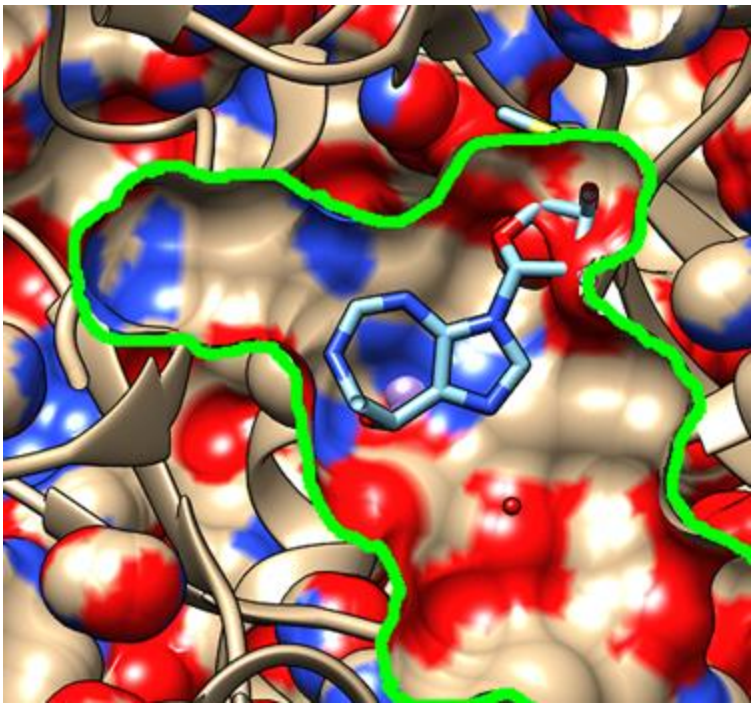


Figure 6.2: Surface of SsnA. The ligand from PDB: 4GBD is included in SsnA for comparison. The active site cavity is highlighted with a green line.

Functional Assessment of SsnA. Even with these features of SsnA known, a meaningful prediction remains difficult to establish. From its structural relationship with adenosine deaminases as well as the genomic context including dehydrogenases, carbamoyltransferases and the hydantoinase, and its expression at early stationary phase, it would seem this enzyme is degradative in nature and all predictions point to a heterocyclic imine functional group. The molybdenum cofactor produced by b2877 could provide an excellent starting point since little is known about its degradation.

While most common adenosine compounds have been tested, it is possible SsnA is an adenosine deaminase which is very specific on its function. This kind of activity is precedented by Tvn0515 , a specific group 1 SAH deaminase which would not deaminate any other tested adenosine compound. It is possible that the substrate is not known as a metabolite. Most predictions are based of compounds found in KEGG, which is continuingly being updated with newly discovered metabolites.

The YahJ Gene. While YahJ is included with SsnA in this chapter, it only superficially resembles this enzyme in that they are both unknown deaminases in cog0402. Being said, YahJ likely does not interact in any respect with SsnA. YahJ has not been studied as extensively as SsnA. YahJ encodes an amidohydrolase with a twin arginine translocation (TAT) peptide attached (199). This peptide assists in transporting the folded protein to the cell envelope or intermembrane space. This enzyme had been previously cloned and was found only to be soluble when using the 3rd methionine as a start codon. It would not fold properly inside the cell if the TAT

peptide were included on the N-terminus. There is no phenotype for the knockout, and genomic context too is limited. *E. coli* has in the operon a carbamate kinase, however other organisms which possess a YahJ homologue do not conserve this gene.

There is no structure, and the closest characterized homologues are *N*-isopropylammelide isopropylaminohydrolase (AtzC) (PDB: 2QT3) (29% identity) and cytosine deaminase (PDB: 1k70) (26% identity). The sequence of YahJ possesses some similar features to that of cytosine deaminase. Gln156 in *E. coli* CDA, which binds to the carbonyl of cytosine, is conserved in the members of YahJ. A slight variation occurs on Asp314, where the YahJ equivalent exists as a Ser. In *E. coli* cytosine deaminase, this mutation confers activity for C5 substituted cytosines such as 5-methyl and 5-fluorocytosine (PDB: 1RAK). Unfortunately none of these compounds are deaminated by YahJ. Lys65 in AtzC aligns with Thr66 in *E. coli* CDA, which is conserved in YahJ. A simple model of the potential active site of YahJ is present in **Figure 6.3**. On this assessment alone, it would seem likely that that the substrate of YahJ would be a 6-member ring. Because of the conservation of the *E. coli* CDA Gln156, the substrate would conceivably possess a carbonyl at position C2 to form a hydrogen bond. Also, there may be a polar substitute on C5 of this 6-member ring.

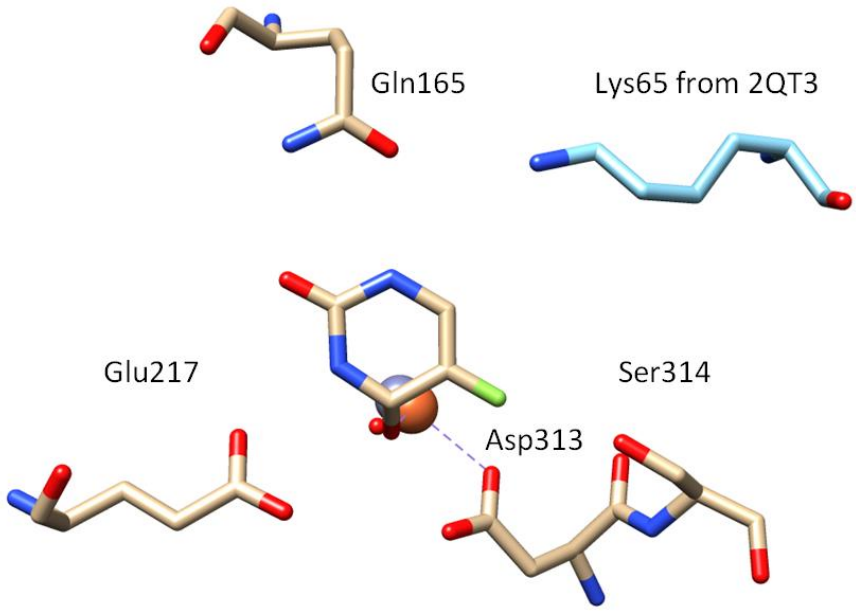


Figure 6.3: Proposed active site model of YahJ. White residues are from *E. coli* cytosine deaminase D314S mutant (PDB: 1RAK), bound to a transition state analogue of 5-fluorouracil. The blue residue is a Lys65 from AtzC (PDB: 2QT3) which aligns with Thr66 in *E. coli* CDA.

An interesting consideration is YahJ has only been purified from cell pellets, and never tested to be excreted into the media. It is possible, however unlikely, that YahJ may not be active until it is transported across the cell membrane. When solubility of the TAT construct was tested, it may have only been the misfolded enzyme inside the cell which was analyzed and not the active protein which was able to be exported. If this were the case, the full TAT construct could be expressed with a C-terminal His tag and purified out of the media in this manner.

CHAPTER VII

CONCLUSION

The age of sequencing has opened many new scientific possibilities and allowed new methods in examining the structure-function relationship of enzymes. In this dissertation we have explored the sequence universe of characterized enzymatic homologues and discovered new enzymatic function which would otherwise have remained unrepresented.

Chapter I, aside from being an introduction to chapters presented, is an exploration of the homologue universe of the known non-amidohydrolase family. A short review was presented over cog0295 (cytidine deaminase), cog2131 (dCMP deaminase), cog0590 (tRNA adenosine deaminase, guanine deaminase, cytosine deaminase), cog0117 (RibG), and cog0717 (dCTP deaminase). The sequences of members from these groups were retrieved from GenBank and isofunctional clusters of enzymes were identified based on known functions and conserved active site residues. New enzymatic activities are predicted to populate cogs 0295, 0717, 2313 and 0590. While the functional annotations here are only presented as unknown new function, it would be possible through analyses performed in the following chapters to provide more accurate descriptions.

Chapter II presents a similar strategy used in Chapter I to divide a large cluster of enzymes related to SAH/MTA deaminase (Tm0936) and ascertain the functional

boundaries as well as discovery enzymatic activities. About 1400 homologues of Tm0936 were retrieved and found to not possess the same active site residues. Most conspicuously absent were two arginines which bound to the homocysteine moiety on SAH. Furthermore, the sequence clustered, based on sequence similarity, to at least 12 different subgroups. A library of compounds was constructed based on modifications on the 5' end of the sugar moiety (based on the lack of arginines). A new function was discovered, that of 5'-deoxyadenosine deaminase. Also, new substrate profiles were discovered which contrast that of SAH deaminase. Enzymes were discovered which deaminate SAH/MTA/5dAdo, MTA/5dAdo/Ado, 5dAdo/Ado, specifically 5dAdo, and specifically SAH. Additional enzymes had wildly different active sites and were found to be 8-oxoadenine deaminases and guanine deaminase. These studies were performed concurrently with docking predictions to help predict, justify binding modes and synergistically improve function discovery and molecular simulation. This chapter was an adaptation of the published works from a 2013 JACS article (23).

Chapter III summarizes the work carried to discover the orphan enzyme isoguanine deaminase in *E. coli*. This project began while studying the enzymes in Chapter I. A homologue of TVN0515 from *Picrophillus torridus* was predicted to deaminate isoguanine. When expressed in *E. coli*, it was found to be entirely insoluble; however, background activity persisted for isoguanine deaminase. It was found that this background activity was from *E. coli* cell extract, and not from the exogenous gene. The activity for isoguanine activity was purified from *E. coli*, and the protein sequence

was identified by peptide mass fingerprinting by in the protein chemistry lab at Texas A&M by comparison to protein databases which have only become available from sequencing technology. The protein was identified as cytosine deaminase, and it was discovered that this enzyme was in fact more efficient for isoguanine than cytosine. The biological relevance for this reaction is not entirely established, but isoguanine is an oxidative adduct of adenine and this may provide a method of detoxification and recycling of the purine. This chapter was an adaptation of a 2011 Biochemistry publication (24).

Chapter IV summarizes unpublished results of the dissection of the bacterial cytosine deaminase family. Based on previous mutagenesis studies by a group attempting to develop 5-fluorocytosine deaminase activity, it was found that a single active site mutation (D314S) would confer this activity. By examining the cytosine deaminase family in the sequence universe, it was found that the family separated into at least 4 groups, with almost half the groups possessing an equivalent mutation in the active site. Members of each group were purified, resulting in the discovery of a novel 5-methylcytosine deaminase. This enzyme also had promiscuous 5-fluorocytosine deaminase activity. 5-methylcytosine deaminase activity was tested *in vivo*. E. coli thymine auxotrophs were transformed by plasmids containing WT CDA, and the new 5-methylcytosine deaminase (KPN00632) (as well as the empty vector). The cultures were grown in media lacking thymine, requiring deamination of 5-methylcytosine to form thymine to allow growth. The methylcytosine deaminase was capable of forming

thymine, while the empty vector and WT CDA were not. The deamination of 5-fluorocytosine was also measured in a similar experiment. The rate of 5-fluorocytosine deamination was measured by growth of cultures containing 5-fluorocytosine, which would produce 5-fluorouracil and inhibit the gene thyA. KPN00632 was found to be particularly effective at this reaction.

Chapter V discusses the discovery of a novel promiscuous pterin deaminase as well as the potential for an undescribed pterin degradation pathway. The gene product of Arad3529 was purified by collaboration with the EFI as part of a probe of cog0402. Arad3529 is found in group 14 and is distantly related to bacterial cytosine deaminase. Computational homology modelling provided models of the protein and allowed computational docking of the KEGG library. Concurrently, this enzyme was screened against a physical deamination library. Arad3529 was found to deaminate a variety of pterin compounds and would accept modifications on C6 of the pterin ring. Even oxidized folate was capable of interacting with the enzyme. Homologues of Arad3529 are found in primarily cyanobacteria and plant symbionts. It is likely that this enzyme is used to deaminate oxidized pterin molecules, which are formed by unintentional oxidation of reduced pterin rings, which are used in the production of cofactors. Furthermore, Arad3529 is found to be on an operon with several other enzymes which were suggested to be a part of a new pterin degradation pathway. These results are summarized in a 2012 JACS publication(26).

Chapter VI is a summary of two remaining projects which we have worked on, which still remain uncharacterized. These are the *E. coli* genes YahJ and SsnA, which are related to cytosine deaminase and the group 1 proteins respectively. No activity has been detected for either of these gene products. Knockout phenotypes have not produced any clues. SsnA appears to have a conserved genomic neighborhood among other organisms which contain an SsnA homologue. The active site of SsnA has been established in a structure, and the active site of YahJ can be approximated. However the substrate remains elusive and is thought to be either a purine or pyrimidine ring.

In conclusion, this research has attempted to utilize the availability of the sequence databases to its full extent. With the information available to us, and it being so easily extractable, pockets of new functional activity can be rapidly identified. In this dissertation, we have discovered a new pterin deaminase, a 5'-deoxadenosine deaminase, an 8-oxoadenine deaminase, an isoguanine deaminase, and a 5-methylcytosine deaminase by exploring uncharacterized enzymes closely related to enzymatically established homologues. We have postulated there to be a unique pterin degradation pathway and discovered an enzyme with 5-fluorocytosine deaminase activity which may be used in cancer gene therapy. The sequence universe will continue to grow, as will the potential to identify and discovery new enzymatic activities.

REFERENCES

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

2. Seibert, C.M. and Raushel, F.M. (2005) Structural and catalytic diversity within the Amidohydrolase Superfamily. *Biochemistry*, **44**, 6383-6391.

3. 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.

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

5. Hall, R.S., Agarwal, R., Hitchcock, D., Sauder, J.M., Burley, S.K., Swaminathan, S. and Raushel, F.M. (2010) Discovery and structure determination of the orphan enzyme isoxanthopterin deaminase. *Biochemistry*, **49**, 4374-4382.

6. Martí-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.

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

8. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403-410.

9. Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366-2382.

10. Atkinson, H.J., Morris, J.H., Ferrin, T.E. and Babbitt, P.C. (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS ONE*, **4**, e4345.

11. 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.

12. Yuan, G., Bin, J.C., McKay, D.J. and Snyder, F.F. (1999) Cloning and characterization of human guanine deaminase. *Journal of Biological Chemistry*, **274**, 8175-8180.

13. Hill, D.L. and Pittillo, R.F. (1973) Use of *Escherichia coli* mutants to evaluate purines, purine nucleosides, and analogues. *Antimicrobial Agents and Chemotherapy*, **4**, 125-132.

14. Danielsen, S., Kilstrup, M., Barilla, K., Jochimsen, B. and Neuhard, J. (1992) Characterization of the *Escherichia coli* codBA operon encoding cytosine permease and cytosine deaminase. *Mol. Microbiol.*, **6**, 1335-1344.

15. Bendt, A.K., Beckers, G., Silberbach, M., Wittmann, A. and Burkovski, A. (2004) Utilization of creatinine as an alternative nitrogen source in *Corynebacterium glutamicum*. *Arch. Microbiol.*, **181**, 443-450.

16. de Souza, M.L., Sadowsky, M.J. and Wackett, L.P. (1996) Atrazine chlorohydrolase from *Pseudomonas sp. strain ADP*: gene sequence, enzyme purification, and protein characterization. *J. Bacteriol.*, **178**, 4894-4900.

17. 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.

18. 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.

19. 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.

20. Michaels, M. and Miller, J. (1992) The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.*, **172**, 6321-6325.

21. van der Kemp, P.A., Thomas, D., Barbey, R., de Oliveira, R. and Boiteux, S. (1996) Cloning and expression in *Escherichia coli* of the OGG1 gene of *Saccharomyces cerevisiae*, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 5197-5202.

22. Goble, A.M., Toro, R., Li, X., Ornelas, A., Fan, H., Eswaramoorthy, S., Patskovsky, Y., Hillerich, B., Seidel, R., Sali, A. *et al.* (2013) Deamination of 6-Aminodeoxyfutalosine in menaquinone biosynthesis by distantly related enzymes. *Biochemistry*, **52**, 6525-6536.

23. Hitchcock, D.S., Fan, H., Kim, J., Vetting, M., Hillerich, B., Seidel, R.D., Almo, S.C., Shoichet, B.K., Sali, A. and Raushel, F.M. (2013) Structure-guided discovery of new eeaminase enzymes. *J. Am. Chem. Soc.*, **135**, 13927-13933.

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

25. 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 Engineering Design and Selection*, **17**, 625-633.

26. Fan, H., Hitchcock, D.S., Seidel, R.D., Hillerich, B., Lin, H., Almo, S.C., Sali, A., Shoichet, B.K. and Raushel, F.M. (2012) Assignment of pterin deaminase activity

to an enzyme of unknown function guided by homology modeling and docking. *J. Am. Chem. Soc.*, **135**, 795-803.

27. Cohen, R.M. and Wolfenden, R. (1971) Cytidine deaminase from *Escherichia coli* : purification, properties, and inhibition by the potential transition state analog 3,4,5,6-tetrahydrouridine. *Journal of Biological Chemistry*, **246**, 7561-7565.

28. Ashley, G.W. and Bartlett, P.A. (1984) Purification and properties of cytidine deaminase from escherichia coli. *Journal of Biological Chemistry*, **259**, 13615-13620.

29. Vita, A., Amici, A., Cacciamani, T., Lanciotti, M. and Magni, G. (1985) Cytidine deaminase from *Escherichia coli* B. Purification and enzymic and molecular properties. *Biochemistry*, **24**, 6020-6024.

30. Betts, L., Xiang, S., Short, S.A., Wolfenden, R. and Carter Jr, C.W. (1994) Cytidine eeaminase. The 2.3 Å crystal structure of an enzyme:transition-state analog complex. *J. Mol. Biol.*, **235**, 635-656.

31. Xiang, S., Short, S.A., Wolfenden, R. and Carter, C.W. (1995) Transition-state selectivity for a single hydroxyl group during catalysis by cytidine deaminase. *Biochemistry*, **34**, 4516-4523.

32. Xiang, S., Short, S.A., Wolfenden, R. and Carter, C.W. (1996) Cytidine deaminase complexed to 3-deazacytidine: A valence buffer in zinc enzyme catalysis. *Biochemistry*, **35**, 1335-1341.

33. Teh, A.-H., Kimura, M., Yamamoto, M., Tanaka, N., Yamaguchi, I. and Kumasaka, T. (2006) The 1.48 Å resolution crystal structure of the homotetrameric cytidine deaminase from mouse. *Biochemistry*, **45**, 7825-7833.

34. Xie, K., Sowden, M.P., Dance, G.S.C., Torelli, A.T., Smith, H.C. and Wedekind, J.E. (2004) The structure of a yeast RNA-editing deaminase provides insight into the fold and function of activation-induced deaminase and APOBEC-1. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 8114-8119.

35. Chung, S.J., Fromme, J.C. and Verdine, G.L. (2005) Structure of human cytidine deaminase bound to a potent inhibitor. *J. Med. Chem.*, **48**, 658-660.

36. Song, B.-H. and Neuhard, J. (1989) Chromosomal location, cloning and nucleotide sequence of the Bacillus subtilis cdd gene encoding cytidine/deoxycytidine deaminase. *Molecular and General Genetics MGG*, **216**, 462-468.

37. Johansson, E., Mejlhede, N., Neuhard, J. and Larsen, S. (2002) Crystal structure of the tetrameric cytidine deaminase from *Bacillus subtilis* at 2.0 Å resolution. *Biochemistry*, **41**, 2563-2570.

38. Johansson, E., Neuhard, J., WillemoÃ«s, M. and Larsen, S. (2004) Structural, kinetic, and mutational studies of the zinc ion environment in tetrameric cytidine deaminase. *Biochemistry*, **43**, 6020-6029.

39. Takeuchi, S., Hirayama, K., Ueda, K., Sakai, H. and Yonehara, H. (1958) Blasticidin S, a new antibiotic. *The Journal of Antibiotics*, **11**, 1-5.

40. Yamaguchi, I., Shibata, H., Seto, H. and Misato, T. (1975) Isolation and purification of blasticidin S deaminase from *Aspergillus terreus*. *The Journal of Antibiotics*, **28**, 7-14.

41. Kimura, M., Kamakura, T., Zhou Tao, Q., Kaneko, I. and Yamaguchi, I. (1994) Cloning of the blasticidin S deaminase gene (BSD) from *Aspergillus terreus* and its use as a selectable marker for *Schizosaccharomyces pombe* and *Pyricularia oryzae*. *Molecular and General Genetics MGG*, **242**, 121-129.

42. Kimura, M., Sekido, S., Isogai, Y. and Yamaguchi, I. (2000) Expression, purification, and characterization of blasticidin S deaminase (BSD) from *Aspergillus terreus*: the role of catalytic zinc in enzyme structure. *Journal of Biochemistry*, **127**, 955-963.

43. Kumasaka, T., Yamamoto, M., Furuichi, M., Nakasako, M., Teh, A.-H., Kimura, M., Yamaguchi, I. and Ueki, T. (2007) Crystal structures of blasticidin S deaminase (BSD): implications for dynamic properties of catalytic zinc. *Journal of Biological Chemistry*, **282**, 37103-37111.

44. Wang, L. and Weiss, B. (1992) dcd (dCTP deaminase) gene of *Escherichia coli*: mapping, cloning, sequencing, and identification as a locus of suppressors of lethal dut (dUTPase) mutations. *J. Bacteriol.*, **174**, 5647-5653.

45. Johansson, E., Fan, M., Bynck, J.H., Neuhard, J., Larsen, S., Sigurskjold, B.W., Christensen, U. and Willemoës, M. (2005) Structures of dCTP deaminase from Escherichia coli with bound substrate and product: reaction mechanism and

determinants of mono- and bifunctionality for a family of enzymes. *Journal of Biological Chemistry*, **280**, 3051-3059.

46. Thymark, M., Johansson, E., Larsen, S. and Willemoës, M. (2008) Mutational analysis of the nucleotide binding site of *Escherichia coli* dCTP deaminase. *Arch. Biochem. Biophys.*, **470**, 20-26.

47. Chan, S., Segelke, B., Lekin, T., Krupka, H., Cho, U.S., Kim, M.-y., So, M., Kim, C.-Y., Naranjo, C.M., Rogers, Y.C. *et al.* (2004) Crystal Structure of the *Mycobacterium tuberculosis* dUTPase: insights into the catalytic mechanism. *J. Mol. Biol.*, **341**, 503-517.

48. Barabás, O., Pongrácz, V., Kovári, J., Wilmanns, M. and Vértessy, B.G. (2004) Structural insights into the catalytic mechanism of phosphate ester hydrolysis by dUTPase. *Journal of Biological Chemistry*, **279**, 42907-42915.

49. Johansson, E., Björnberg, O., Nyman, P.O. and Larsen, S. (2003) Structure of the bifunctional dCTP deaminase-dUTPase from *Methanocaldococcus jannaschii* and its relation to other homotrimeric dUTPases. *Journal of Biological Chemistry*, **278**, 27916-27922.

50. Li, H., Xu, H., Graham, D.E. and White, R.H. (2003) The *Methanococcus jannaschii* dCTP deaminase is a bifunctional deaminase and diphosphatase. *Journal of Biological Chemistry*, **278**, 11100-11106.

51. Björnberg, O., Neuhard, J. and Nyman, P.O. (2003) A bifunctional dCTP deaminase-dUTP nucleotidohydrolase from the hyperthermophilic archaeon

Methanocaldococcus jannaschii. *Journal of Biological Chemistry*, **278**, 20667-20672.

52. Keck, K., Mahler, H.R. and Fraser, D. (1960) Synthesis of deoxycytidine-5′ -phosphate deaminase in *Escherichia coli* infected by T2 bacteriophage. *Arch. Biochem. Biophys.*, **86**, 85-88.

53. Weiner, K.X., Weiner, R.S., Maley, F. and Maley, G.F. (1993) Primary structure of human deoxycytidylate deaminase and overexpression of its functional protein in *Escherichia coli*. *Journal of Biological Chemistry*, **268**, 12983-12989.

54. Maley, F. and Maley, G.F. (1970) Mechanisms of enzyme modulation involving deoxycytidylate deaminase and thymidylate synthetase. *Adv. Enzyme Regul.*, **8**, 55-71.

55. Moore, J.T., Ciesla, J.M., Changchien, L.M., Maley, G.F. and Maley, F. (1994) Identification of a site necessary for allosteric regulation in T4-phage deoxycytidylate deaminase. *Biochemistry*, **33**, 2104-2112.

56. Hou, H.-F., Liang, Y.-H., Li, L.-F., Su, X.-D. and Dong, Y.-H. (2008) Crystal structures of S*treptococcus mutans* 2-deoxycytidylate deaminase and its complex with substrate analog and allosteric regulator dCTP·Mg2+. *J. Mol. Biol.*, **377**, 220-231.

57. Almog, R., Maley, F., Maley, G.F., MacColl, R. and Van Roey, P. (2004) Three-dimensional structure of the R115E mutant of T4-bacteriophage 2‘-deoxycytidylate deaminase. *Biochemistry*, **43**, 13715-13723.

58. Zhang, Y., Maley, F., Maley, G.F., Duncan, G., Dunigan, D.D. and Van Etten, J.L. (2007) Chloroviruses encode a bifunctional dCMP-dCTP deaminase that produces two key intermediates in dTTP formation. *J. Virol.*, **81**, 7662-7671.

59. Wolf, J., Gerber, A.P. and Keller, W. (2002) TadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *European Molecular Biology Organization*, **21**, 3841-3851.

60. Gerber, A.P. and Keller, W. (1999) An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science*, **286**, 1146-1149.

61. Auxilien, S., Crain, P.F., Trewyn, R.W. and Grosjean, H. (1996) Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. *J. Mol. Biol.*, **262**, 437-458.

62. Jakubowski, H. and Goldman, E. (1984) Quantities of individual aminoacyl-tRNA families and their turnover in *Escherichia coli*. *J. Bacteriol.*, **158**, 769-776.

63. Losey, H.C., Ruthenburg, A.J. and Verdine, G.L. (2006) Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nat. Struct. Mol. Biol.*, **13**, 153-159.

64. Liaw, S.-H., Chang, Y.-J., Lai, C.-T., Chang, H.-C. and Chang, G.-G. (2004) Crystal structure of *Bacillus subtilis* guanine deaminase: the first domain-swapped structure in the cytidine deaminase superfamily. *Journal of Biological Chemistry*, **279**, 35479-35485.

65. Erbs, P., Exinger, F. and Jund, R. (1997) Characterization of the *Saccharomyces cerevisiae* FCY1 gene encoding cytosine deaminase and its homologue FCA1 of *Candida albicans*. *Curr. Genet.*, **31**, 1-6.

66. 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.

67. Ireton, G.C., Black, M.E. and Stoddard, B.L. (2003) The 1.14 Å crystal structure of yeast cytosine deaminase: evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. *Structure*, **11**, 961-972.

68. Burrows, R.B. and Brown, G.M. (1978) Presence of *Escherichia coli* of a deaminase and a reductase involved in biosynthesis of riboflavin. *J. Bacteriol.*, **136**, 657-667.

69. Stenmark, P., Moche, M., Gurmu, D. and Nordlund, P. (2007) The crystal structure of the bifunctional deaminase/reductase RibD of the riboflavin biosynthetic pathway in *Escherichia coli*: implications for the reductive mechanism. *J. Mol. Biol.*, **373**, 48-64.

70. Chen, S.-C., Shen, C.-Y., Yen, T.-M., Yu, H.-C., Chang, T.-H., Lai, W.-L. and Liaw, S.-H. (2013) Evolution of vitamin B2 biosynthesis: eubacterial RibG and fungal Rib2 deaminases. *Acta Crystallographica Section D*, **69**, 227-236.

71. Schrader, W.P. and Stacy, A.R. (1977) Purification and subunit structure of adenosine deaminase from human kidney. *Journal of Biological Chemistry*, **252**, 6409-6415.

72. Kinoshita, T., Nishio, N., Nakanishi, I., Sato, A. and Fujii, T. (2003) Structure of bovine adenosine deaminase complexed with 6-hydroxy-1,6-dihydropurine riboside. *Acta Crystallographica Section D*, **59**, 299-303.

73. Kameoka, J., Tanaka, T., Nojima, Y., Schlossman, S. and Morimoto, C. (1993) Direct association of adenosine deaminase with a T cell activation antigen, CD26. *Science*, **261**, 466-469.

74. Martin, D.W. and Gelfand, E.W. (1981) Biochemistry of diseases of immunodevelopment. *Annu. Rev. Biochem.*, **50**, 845-877.

75. Minguet, S., Huber, M., Rosenkranz, L., Schamel, W.W.A., Reth, M. and Brummer, T. (2005) Adenosine and cAMP are potent inhibitors of the NF-κB pathway downstream of immunoreceptors. *Eur. J. Immunol.*, **35**, 31-41.

76. Dillman, R.O. (2004) Pentostatin (Nipent®) in the treatment of chronic lymphocyte leukemia and hairy cell leukemia. *Expert Review of Anticancer Therapy*, **4**, 27-36.

77. Zavialov, A.V., Yu, X., Spillmann, D., Lauvau, G. and Zavialov, A.V. (2010) Structural basis for the growth factor activity of human adenosine deaminase ADA2. *Journal of Biological Chemistry*, **285**, 12367-12377.

78. Matsushita, T., Fujii-Taira, I., Tanaka, Y., Homma, K.J. and Natori, S. (2000) Male-specific IDGF, a novel gene encoding a membrane-bound extracellular signaling molecule expressed exclusively in testis of *Drosophila melanogaster*. *Journal of Biological Chemistry*, **275**, 36934-36941.

79. Homma, K.-i., Matsushita, T. and Natori, S. (1996) Purification, characterization, and cDNA cloning of a novel growth factor from the conditioned medium of NIH-Sape-4, an embryonic cell Line of *Sarcophaga peregrina* (Flesh Fly). *Journal of Biological Chemistry*, **271**, 13770-13775.

80. Zurovec, M., Dolezal, T., Gazi, M., Pavlova, E. and Bryant, P.J. (2002) Adenosine deaminase-related growth factors stimulate cell proliferation in *Drosophila* by depleting extracellular adenosine. *Proceedings of the National Academy of Sciences*, **99**, 4403-4408.

81. Zavialov, A.V., Gracia, E., Glaichenhaus, N., Franco, R., Zavialov, A.V. and Lauvau, G. (2010) Human adenosine deaminase 2 induces differentiation of monocytes into macrophages and stimulates proliferation of T helper cells and macrophages. *J. Leukoc. Biol.*, **88**, 279-290.

82. Ogasawara, N., Goto, H., Yamada, Y. and Watanabe, T. (1978) Distribution of AMP-deaminase isozymes in rat tissues. *Eur. J. Biochem.*, **87**, 297-304.

83. Sabina, R.L., Swain, J.L., Olanow, C.W., Bradley, W.G., Fishbein, W.N., DiMauro, S. and Holmes, E.W. (1984) Myoadenylate deaminase deficiency. Functional and

metabolic abnormalities associated with disruption of the purine nucleotide cycle. *J Clin Invest.*, **73**, 720-730.

84. Bausch-Jurken, M.T., Mahnke-Zizelman, D.K., Morisaki, T. and Sabina, R.L. (1992) Molecular cloning of AMP deaminase isoform L. Sequence and bacterial expression of human AMPD2 cDNA. *Journal of Biological Chemistry*, **267**, 22407-22413.

85. Xu, J., Zhang, H.-Y., Xie, C.-H., Xue, H.-W., Dijkhuis, P. and Liu, C.-M. (2005) Embryonic Factor 1 encodes an AMP deaminase and is essential for the zygote to embryo transition in *Arabidopsis*. *The Plant Journal*, **42**, 743-758.

86. Han, B.W., Bingman, C.A., Mahnke, D.K., Bannen, R.M., Bednarek, S.Y., Sabina, R.L. and Phillips, G.N. (2006) Membrane association, mechanism of action, and structure of *Arabidopsis* Embryonic Factor 1 (FAC1). *Journal of Biological Chemistry*, **281**, 14939-14947.

87. Van den Bergh, F. and Sabina, R.L. (1995) Characterization of human AMP deaminase 2 (AMPD2) gene expression reveals alternative transcripts encoding variable *N*-terminal extensions of isoform L. *Biochem J.*, **132**, 401-410.

88. Yamada, Y., Goto, H. and Ogasawara, N. (1992) Cloning and nucleotide sequence of the cDNA encoding human erythrocyte-specific AMP deaminase. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, **1171**, 125-128.

89. Ogasawara, N., Goto, H., Yamada, Y., Nishigaki, I., Itoh, T. and Hasegawa, I. (1984) Complete deficiency of AMP deaminase in human erythrocytes. *Biochem. Biophys. Res. Commun.*, **122**, 1344-1349.

90. Gupta, N.K. and Glantz, M.D. (1985) Isolation and characterization of human liver guanine deaminase. *Arch. Biochem. Biophys.*, **236**, 266-276.

91. Mohamedali, K.A., Guicherit, O.M., Kellems, R.E. and Rudolph, F.B. (1993) The highest levels of purine catabolic enzymes in mice are present in the proximal small intestine. *Journal of Biological Chemistry*, **268**, 23728-23733.

92. Hamajima, N., Kouwaki, M., Vreken, P., Matsuda, K., Sumi, S., Imaeda, M., Ohba, S., Kidouchi, K., Nonaka, M., Sasaki, M. *et al.* (1998) Dihydropyrimidinase deficiency: structural organization, chromosomal localization, and mutation analysis of the human dihydropyrimidinase gene. *The American Journal of Human Genetics*, **63**, 717-726.

93. Hamajima, N., Matsuda, K., Sakata, S., Tamaki, N., Sasaki, M. and Nonaka, M. (1996) A novel gene family defined by human dihydropyrimidinase and three related proteins with differential tissue distribution. *Gene*, **180**, 157-163.

94. Wang, L.-H. and Strittmatter, S.M. (1996) A family of rat CRMP genes is differentially expressed in the nervous system. *The Journal of Neuroscience*, **16**, 6197-6207.

95. Yamashita, N., Uchida, Y., Ohshima, T., Hirai, S.-i., Nakamura, F., Taniguchi, M., Mikoshiba, K., Honnorat, J., Kolattukudy, P., Thomasset, N. *et al.* (2006)

Collapsin response mediator protein 1 mediates reelin signaling in cortical neuronal migration. *The Journal of Neuroscience*, **26**, 13357-13362.

96. Su, K.-Y., Chien, W.-L., Fu, W.-M., Yu, I.-S., Huang, H.-P., Huang, P.-H., Lin, S.-R., Shih, J.-Y., Lin, Y.-L., Hsueh, Y.-P. *et al.* (2007) Mice deficient in collapsin response mediator protein-1 exhibit impaired long-term potentiation and impaired spatial learning and memory. *The Journal of Neuroscience*, **27**, 2513-2524.

97. Arimura, N., Menager, C., Fukata, Y. and Kaibuchi, K. (2004) Role of CRMP-2 in neuronal polarity. *J. Neurobiol.*, **58**, 34-47.

98. Stenmark, P., Ogg, D., Flodin, S., Flores, A., Kotenyova, T., Nyman, T., Nordlund, P. and Kursula, P. (2007) The structure of human collapsin response mediator protein 2, a regulator of axonal growth. *J. Neurochem.*, **101**, 906-917.

99. Quach, T.T., Mosinger Jr, B., Ricard, D., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Stankoff, B., Honnorat, J., Belin, M.-F. and Kolattukudy, P. (2000) Collapsin response mediator protein-3/unc-33-like protein-4 gene: organization, chromosomal mapping and expression in the developing mouse brain. *Gene*, **242**, 175-182.

100. Quach, T.T., Massicotte, G., Belin, M.-F., Honnorat, J., Glasper, E.R., Devries, A.C., Jakeman, L.B., Baudry, M., Duchemin, A.-M. and Kolattukudy, P.E. (2008) CRMP3 is required for hippocampal CA1 dendritic organization and plasticity. *The FASEB Journal*, **22**, 401-409.

101. Alabed, Y.Z., Pool, M., Tone, S.O. and Fournier, A.E. (2007) Identification of CRMP4 as a convergent regulator of axon outgrowth inhibition. *The Journal of Neuroscience*, **27**, 1702-1711.

102. Hotta, A., Inatome, R., Yuasa-Kawada, J., Qin, Q., Yamamura, H. and Yanagi, S. (2005) Critical role of collapsin response mediator protein-associated molecule CRAM for filopodia and growth cone development in neurons. *Mol. Biol. Cell*, **16**, 32-39.

103. Yamashita, N., Mosinger, B., Roy, A., Miyazaki, M., Ugajin, K., Nakamura, F., Sasaki, Y., Yamaguchi, K., Kolattukudy, P. and Goshima, Y. (2011) CRMP5 (Collapsin Response Mediator Protein 5) regulates dendritic development and synaptic plasticity in the cerebellar purkinje cells. *The Journal of Neuroscience*, **31**, 1773-1779.

104. Ponnusamy, R. and Lohkamp, B. (2013) Insights into the oligomerization of CRMPs: crystal structure of human collapsin response mediator protein 5. *J. Neurochem.*, **125**, 855-868.

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

106. Adachi, H., Katayama, T., Inuzuka, C., Oikawa, S., Tsujimoto, M. and Nakazato, H. (1990) Identification of membrane anchoring site of human renal dipeptidase

and construction and expression of a cDNA for its secretory form. *Journal of Biological Chemistry*, **265**, 15341-15345.

107. Kim, H.S. and Campbell, B.J. (1982) β-lactamase activity of renal dipeptidase against N-formimidoyl-thienamycin. *Biochem. Biophys. Res. Commun.*, **108**, 1638-1642.

108. Nitanai, Y., Satow, Y., Adachi, H. and Tsujimoto, M. (2002) Crystal Structure of Human Renal Dipeptidase Involved in β-Lactam Hydrolysis. *J. Mol. Biol.*, **321**, 177-184.

109. Habib, G.M., Shi, Z.-Z., Cuevas, A.A. and Lieberman, M.W. (2003) Identification of two additional members of the membrane-bound dipeptidase family. *The FASEB Journal*. Published ahead of print May 8, 2003, doi: 10.1096/fj.02-0899fje.

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

111. Glasner, M.E., Fayazmanesh, N., Chiang, R.A., Sakai, A., Jacobson, M.P., Gerlt, J.A. and Babbitt, P.C. (2006) Evolution of structure and function in the O-succinylbenzoate synthase/N-acylamino acid racemase family of the enolase superfamily. *J. Mol. Biol.*, **360**, 228-250.

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

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

114. Sauder, M.J., Rutter, M.E., Bain, K., Rooney, I., Gheyi, T., Atwell, S., Thompson, D.A., Emtage, S. and Burley, S.K. (2008) High throughput protein production and crystallization at NYSGXRC. *Methods Mol. Biol.*, **426**, 561-575.

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

116. Sali, A. and Blundell, T.L. (1993) Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779-815.

117. Shen, M.Y. and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.*, **15**, 2507-2524.

118. Sherman, W., Day, T., Jacobson, M.P., Friesner, R.A. and Farid, R. (2006) Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.*, **49**, 534-553.

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

120. 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-U772.

121. Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27-30.

122. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354-D357.

123. Mysinger, M.M. and Shoichet, B.K. (2010) Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.*, **50**, 1561-1573.

124. Kamat, S.S., Fan, H., Sauder, J.M., Burley, S.K., Shoichet, B.K., Sali, A. and Raushel, F.M. (2011) Enzymatic deamination of the epigenetic base N-6-methyladenine. *J Am Chem Soc*, **133**, 2080-2083.

125. Goble, A.M., Fan, H., Sali, A. and Raushel, F.M. (2011) Discovery of a cytokinin deaminase. *ACS Chem. Biol.*, **6**, 1036-1040.

126. Fan, H., Irwin, J.J., Webb, B.M., Klebe, G., Shoichet, B.K. and Sali, A. (2009) Molecular docking screens using comparative models of proteins. *J. Chem. Inf. Model.*, **49**, 2512-2527.

127. Guan, R., Ho, M.-C., Fröhlich, R.F.G., Tyler, P.C., Almo, S.C. and Schramm, V.L. (2012) Methylthioadenosine deaminase in an alternative quorum sensing pathway in *Pseudomonas aeruginosa*. *Biochemistry*, **51**, 9094-9103.

128. Malins, D.C., Polissar, N.L., Ostrander, G.K. and Vinson, M.A. (2000) Single 8-oxo-guanine and 8-oxo-adenine lesions induce marked changes in the backbone structure of a 25-base DNA strand. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 12442-12445.

129. Kamiya, H. and Kasai, H. (1996) Effect of sequence contexts on misincorporation of nucleotides opposite 2-hydroxyadenine. *FEBS Letters*, **39**, 113-116.

130. Krenitsky, T.A., Neil, S.M., Elion, G.B. and Hitchings, G.H. (1972) A comparison of the specificities of xanthine oxidase and aldehyde oxidase. *Arch. Biochem. Biophys.*, **150**, 585-599.

131. David, S.S., O'Shea, V.L. and Kundu, S. (2007) Base-excision repair of oxidative DNA damage. *Nature*, **447**, 941-950.

132. Michaels, M., Cruz, C., Grollman, A. and Miller, J. (1992) Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 7022-7025.

133. Chetsanga, C.J. and Frenette, G.P. (1983) Excision of aflatoxin B1-imidazole ring opened guanine adducts from DNA by formamidopyrimidine-DNA glycosylase. *Carcinogenesis*, **4**, 997-1000.

134. Boiteux, S., Gajewski, E., Laval, J. and Dizdaroglu, M. (1992) Substrate specificity of the *Escherichia coli* Fpg protein formamidopyrimidine-DNA glycosylase: excision of purine lesions in DNA produced by ionizing radiation or photosensitization. *Biochemistry*, **31**, 106-110.

135. Cabrera, M., Nghiem, Y. and Miller, J. (1988) MutM, a second mutator locus in *Escherichia coli* that generates G.C----T.A transversions. *J. Bacteriol.*, **170**, 5405-5407.

136. Michaels, M., Pham, L., Cruz, C. and Miller, J. (1991) MutM, a protein that prevents G.C----T.A transversions, is formamidopyrimidine-DNA glycosylase. *Nucleic Acids Res.*, **19**, 3629-3632.

137. Maki, H. and Sekiguchi, M. (1992) MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature*, **355**, 273-275.

138. Hashiguchi, K., Zhang, Q.M., Sugiyama, H., Ikeda, S. and Yonei, S. (2002) Characterization of 2-hydroxyadenine DNA glycosylase activity of *Escherichia coli* MutY protein. *Int. J. Radiat. Biol.*, **78**, 585-592.

139. Kamiya, H. and Kasai, H. (1997) Mutations Induced by 2-hydroxyadenine on a shuttle vector during leading and lagging strand syntheses in mammalian cells. *Biochemistry*, **36**, 11125-11130.

140. Kamiya, H. and Kasai, H. (1997) Substitution and deletion mutations induced by 2-hydroxyadenine in *Escherichia coli*: effects of sequence contexts in leading and lagging strands. *Nucleic Acids Res.*, **25**, 304-310.

141. Kamiya, H. and Kasai, H. (1995) 2-Hydroxyadenine (isoguanine) as oxidative DNA damage: its formation and mutation inducibility. *Nucleic Acids Symposium Series*, 233-234.

142. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, **2**, 0008.

143. Friedman, S. and Gots, J. (1951) Deamination of isoguanine by *Escherichia coli*. *Arch. Biochem. Biophys.*, **1**, 227-229.

144. Hardy, L., Finer-Moore, J., Montfort, W., Jones, M., Santi, D. and Stroud, R. (1987) Atomic structure of thymidylate synthase: target for rational drug design. *Science*, **235**, 448-455.

145. Ahmad, S.I., Kirk, S.H. and Eisenstark, A. (1998) Thymine metabolism and thyeminless death in prokaryotes and eukaryotes. *Annu. Rev. Microbiol.*, **52**, 591-625.

146. Longley, D.B., Harkin, D.P. and Johnston, P.G. (2003) 5-Fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer*, **3**, 330-338.

147. Mullen, C.A., Kilstrup, M. and Blaese, R.M. (1992) Transfer of the bacterial gene for cytosine deaminase to mammalian cells confers lethal sensitivity to 5-fluorocytosine: a negative selection system. *Proceedings of the National Academy of Sciences*, **89**, 33-37.

148. Fuchita, M., Ardiani, A., Zhao, L., Serve, K., Stoddard, B.L. and Black, M.E. (2009) Bacterial cytosine deaminase mutants created by molecular engineering show

improved 5-fluorocytosine–mediated cell killing *in vitro* and *In vivo*. *Cancer Res.*, **69**, 4791-4799.

149. Danielsen, S., Kilstrup, M., Barilla, K., Jochimsen, B. and Neuhard, J. (1992) Characterization of the *Escherichia coli* CodBA operon encoding cytosine permease and cytosine deaminase. *Mol. Microbiol.*, **6**, 1335-1344.

150. Katsuragi, T., Sakai, T., Matsumoto, K.y. and Tonomura, K. (1986) Cytosine deaminase from *escherichia coli*-production, purification, and some characteristics. *Agric. Biol. Chem.*, **50**, 1721-1730.

151. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, **2**, 2006 0008.

152. Whisstock, J.C. and Lesk, A.M. (2003) Prediction of protein function from protein sequence and structure. *Q. Rev. Biophys.*, **36**, 307-340.

153. Gerlt, J.A. and Babbitt, P.C. (2000) Can sequence determine function? *Genome Biol*, **1**.

154. Brenner, S.E. (1999) Errors in genome annotation. *Trends Genet.*, **15**, 132-133.

155. Devos, D. and Valencia, A. (2001) Intrinsic errors in genome annotation. *Trends Genet.*, **17**, 429-431.

156. Gerlt, J.A., Allen, K.N., Almo, S.C., Armstrong, R.N., Babbitt, P.C., Cronan, J.E., Dunaway-Mariano, D., Imker, H.J., Jacobson, M.P., Minor, W. *et al.* (2011) The Enzyme Function Initiative. *Biochemistry*, **50**, 9950-9962.

157. Favia, A.D., Nobeli, I., Glaser, F. and Thornton, J.M. (2008) Molecular docking for substrate identification: the short-chain dehydrogenases/reductases. *J. Mol. Biol.*, **375**, 855-874.

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

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

160. Xiang, D.F., Kolb, P., Fedorov, A.A., Xu, C.F., Fedorov, E.V., Narindoshivili, T., Williams, H.J., Shoichet, B.K., Almo, S.C. and Raushel, F.M. (2012) Structure-based function discovery of an enzyme for the hydrolysis of phosphorylated sugar lactones. *Biochemistry*, **51**, 1762-1773.

161. Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 291-325.

162. Pieper, U., Eswar, N., Webb, B.M., Eramian, D., Kelly, L., Barkan, D.T., Carter, H., Mankoo, P., Karchin, R., Marti-Renom, M.A. *et al.* (2009) MODBASE, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.*, **37**, D347-D354.

163. Evers, A. and Klebe, G. (2004) Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. *J. Med. Chem.*, **47**, 5381-5392.

164. Kairys, V., Fernandes, M.X. and Gilson, M.K. (2006) Screening drug-like compounds by docking to homology models: a systematic study. *J. Chem. Inf. Model.*, **46**, 365-379.

165. Katritch, V., Byrd, C.M., Tseitin, V., Dai, D.C., Raush, E., Totrov, M., Abagyan, R., Jordan, R. and Hruby, D.E. (2007) Discovery of small molecule inhibitors of ubiquitin-like poxvirus proteinase I7L using homology modeling and covalent docking approaches. *J. Comput-Aided. Mol. Des.*, **21**, 549-558.

166. Carlsson, J., Coleman, R.G., Setola, V., Irwin, J.J., Fan, H., Schlessinger, A., Sali, A., Roth, B.L. and Shoichet, B.K. (2011) Ligand discovery from a dopamine D-3 receptor homology model and crystal structure. *Nature Chemical Biology*, **7**, 769-778.

167. Schlessinger, A., Geier, E., Fan, H., Irwin, J.J., Shoichet, B.K., Giacomini, K.M. and Sali, A. (2011) Structure-based discovery of prescription drugs that interact with

the norepinephrine transporter, NET. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 15810-15815.

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

169. Pieper, U., Chiang, R., Seffernick, J.J., Brown, S.D., Glasner, M.E., Kelly, L., Eswar, N., Sauder, J.M., Bonanno, J.B., Swaminathan, S. *et al.* (2009) Target selection and annotation for the structural genomics of the amidohydrolase and enolase superfamilies. *J. Struct. Funct. Genomics*, **10**, 107-125.

170. Marti-Arbona, R., Xu, C.F., 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.

171. Aslanidis, C. and Dejong, P.J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.*, **18**, 6069-6074.

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

173. Pegg, S.C.H., Brown, S.D., Ojha, S., Seffernick, J., Meng, E.C., Morris, J.H., Chang, P.J., Huang, C.C., Ferrin, T.E. and Babbitt, P.C. (2006) Leveraging enzyme

structure-function relationships for functional inference and experimental design: the structure-function linkage database. *Biochemistry*, **45**, 2545-2555.

174. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389-3402.

175. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792-1797.

176. Fiser, A., Do, R.K. and Sali, A. (2000) Modeling of loops in protein structures. *Protein Sci.*, **9**, 1753-1773.

177. Xiang, D.F., Kolb, P., Fedorov, A.A., Meier, M.M., Fedorov, L.V., Nguyen, T.T., Sterner, R., Almo, S.C., Shoichet, B.K. and Raushel, F.M. (2009) Functional annotation and three-dimensional structure of Dr0930 from *Deinococcus radiodurans*, a close relative of phosphotriesterase in the amidohydrolase superfamily. *Biochemistry*, **48**, 2237-2247.

178. Lorber, D.M. and Shoichet, B.K. (2005) Hierarchical docking of databases of multiple ligand conformations. *Curr. Top. Med. Chem.*, **5**, 739-749.

179. Kamat, S.S., Bagaria, A., Kumaran, D., Holmes-Hampton, G.P., Fan, H., Sali, A., Sauder, J.M., Burley, S.K., Lindahl, P.A., Swaminathan, S. *et al.* (2011) Catalytic mechanism and three-dimensional structure of adenine deaminase. *Biochemistry*, **50**, 1917-1927.

180. Keiser, M.J., Roth, B.L., Armbruster, B.N., Ernsberger, P., Irwin, J.J. and Shoichet, B.K. (2007) Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.*, **25**, 197-206.

181. Hert, J., Keiser, M.J., Irwin, J.J., Oprea, T.I. and Shoichet, B.K. (2008) Quantifying the relationships among drug classes. *J. Chem. Inf. Model.*, **48**, 755-765.

182. Schwalbe, C.H., Lowis, D.R. and Richards, W.G. (1993) Pterin 1H-3H tautomerism and Its possible relevance to the cinding of folate to dihydrofolate reductase. *J Chem Soc Chem Comm*, 1199-1200.

183. Pfleiderer, W. (1984) Chemistry of naturally occurring pterins. *Folates and Pterins*, **2**, 43-114.

184. Noiriel, A., Naponelli, V., Gregory, J.F. and Hanson, A.D. (2007) Pterin and folate salvage. Plants and *Escherichia coli* lack capacity to reduce oxidized Pterins. *Plant Physiol*, **143**, 1101-1109.

185. FORREST HS, G.E., MITCHELL HK. (1956) Conversion of 2-amino-4-hydroxypteridine to isoxanthopterin in *D. Melanogaster*. *Science*, **124**, 725-726.

186. Lowry, O.H., Bessey, O.A. and Crawford, E.J. (1949) Photolytic and enzymatic transformations of pteroylglutamic acid. *Journal of Biological Chemistry*, **180**, 389-398.

187. Off, M.K., Steindal, A.E., Porojnicu, A.C., Juzeniene, A., Vorobey, A., Johnsson, A. and Moan, J. (2005) Ultraviolet photodegradation of folic acid. *J Photoch Photobio B*, **80**, 47-55.

188. McNutt Jr., W.S. (1968) The metabolism of isoxanthopterin by *Alcaligenes faecalis*. *Journal of Biological Chemistry*, **238**, 1116-1121.

189. McNutt, W.S. and Damle, S.P. (1964) Tetraoxypteridine isomerase. *Journal of Biological Chemistry*, **239**, 4272-4279.

190. Dairman, W.M. and McNutt, W.S. (1964) The metabolism of xanthine-8-carboxylic acid by *Alcaligenes faecalis*. *Journal of Biological Chemistry*, **239**, 3407-3411.

191. Mysinger, M.M., Weiss, D.R., Ziarek, J.J., Gravel, S., Doak, A.K., Karpiak, J., Heveker, N., Shoichet, B.K. and Volkman, B.F. (2012) Structure-based ligand discovery for the protein-protein interface of chemokine receptor CXCR4. *Proc Natl Acad Sci U S A*, **109**, 5517-5522.

192. Yamada, M., Talukder, A.A. and Nitta, T. (1999) Characterization of the ssnA gene, which is involved in the decline of cell viability at the beginning of stationary phase in *Escherichia coli*. *J. Bacteriol.*, **181**, 1838-1846.

193. Xi, H., Schneider, B.L. and Reitzer, L. (2000) Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. *J. Bacteriol.*, **182**, 5332-5341.

194. Li, Y., Jin, Z., Yu, X., Allewell, N.M., Tuchman, M. and Shi, D. (2011) The ygeW encoded protein from *Escherichia coli* is a knotted ancestral catabolic transcarbamylase. *Proteins Struct. Funct. Bioinformat.*, **79**, 2327-2334.

195. Khan, F., Jala, V.R., Rao, N.A. and Savithri, H.S. (2003) Characterization of recombinant diaminopropionate ammonia-lyase from Escherichia coli and Salmonella typhimurium. *Biochem. Biophys. Res. Commun.*, **306**, 1083-1088.

196. Bisht, S., Rajaram, V., Bharath, S.R., Kalyani, J.N., Khan, F., Rao, A.N., Savithri, H.S. and Murthy, M.R.N. (2012) Crystal Structure of *Escherichia coli* diaminopropionate ammonia-lyase reveals mechanism of enzyme activation and catalysis. *Journal of Biological Chemistry*, **287**, 20369-20381.

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

198. Neumann, M., Mittelstädt, G., Seduk, F., Iobbi-Nivol, C. and Leimkühler, S. (2009) MocA Is a specific cytidylyltransferase involved in molybdopterin cytosine dinucleotide biosynthesis in *Escherichia coli*. *Journal of Biological Chemistry*, **284**, 21891-21898.

199. Tullman-Ercek, D., DeLisa, M.P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T. and Georgiou, G. (2007) Export pathway selectivity of *Escherichia coli* twin arginine translocation signal peptides. *Journal of Biological Chemistry*, **282**, 8309-8316.

APPENDIX A

FUNCTIONAL ANNOTATIONS OF GROUP 1 PROTEINS IN CHAPTER II

Organism	GI Number	Sg	Function
Acetonema longum DSM 6540	338814514	Sg1a	SAH/MTA Deaminase
Acidaminococcus fermentans DSM 20731	284048666	Sg1a	SAH/MTA Deaminase
Acidaminococcus intestini RyC-MR95	352684529	Sg1a	SAH/MTA Deaminase
Acidaminococcus sp. D21	227498515	Sg1a	SAH/MTA Deaminase
Acidithiobacillus caldus ATCC 51756	255021229	Sg2	MTA/5'-dAdo/Ado Deaminase
Acidithiobacillus ferrivorans SS3	344200120	Sg2	MTA/5'-dAdo/Ado Deaminase
Acidithiobacillus ferrooxidans ATCC 53993	198283616	Sg2	MTA/5'-dAdo/Ado Deaminase
Acidithiobacillus thiooxidans ATCC 19377	384084622	Sg2	MTA/5'-dAdo/Ado Deaminase
Aciduliprofundum boonei T469	254166651	Sg11	SAH Deaminase
Aciduliprofundum boonei T469	254167272	Sg11	SAH Deaminase
Acinetobacter lwoffii SH145	262376404	Sg10	8-Oxoadenine Deaminase
Acinetobacter sp. HA	389704937	Sg10	8-Oxoadenine Deaminase
Agrobacterium vitis S4	222106480	Sg10	8-Oxoadenine Deaminase
Akkermansia muciniphila ATCC BAA-835	187736175	Sg6	5'-dAdo Deaminase
Albugo laibachii Nc14	325189944	Sg2	MTA/5'-dAdo/Ado Deaminase
Albugo laibachii Nc14	325189945	Sg2	MTA/5'-dAdo/Ado Deaminase
Alcanivorax borkumensis SK2	110834613	Sg2	MTA/5'-dAdo/Ado Deaminase
Alcanivorax dieselolei B5	407696804	Sg2	MTA/5'-dAdo/Ado Deaminase
Alcanivorax hongdengensis A-11-3	408375135	Sg2	MTA/5'-dAdo/Ado Deaminase
Alcanivorax pacificus W11-5	407802061	Sg2	MTA/5'-dAdo/Ado Deaminase
Alcanivorax sp. DG881	254429653	Sg2	MTA/5'-dAdo/Ado Deaminase
Alkalilimnicola ehrlichii MLHE-1	114320072	Sg2	MTA/5'-dAdo/Ado Deaminase
Alkaliphilus metalliredigens QYMF	150392253	Sg9	Guanine Deaminase
Alkaliphilus oremlandii OhILAs	158319425	Sg9	Guanine Deaminase
Allochromatium vinosum DSM 180	288941706	Sg2	MTA/5'-dAdo/Ado Deaminase
Alteromonas sp. S89	372267296	Sg2	MTA/5'-dAdo/Ado Deaminase
Ammonifex degensii KC4	260892855	Sg1a	SAH/MTA Deaminase
Anaerococcus hydrogenalis ACS-025-V-Sch4	325846621	Sg1a	SAH/MTA Deaminase
Anaerococcus hydrogenalis DSM 7454	212696141	Sg1a	SAH/MTA Deaminase
Anaerococcus prevotii ACS-065-V-Col13	325479377	Sg1a	SAH/MTA Deaminase
Anaerococcus prevotii DSM 20548	257066470	Sg1a	SAH/MTA Deaminase
Anaerolinea thermophila UNI-1	320161507	Sg7	5'-dAdo/Ado Deaminase
Archaeoglobus fulgidus DSM 4304	11498602	Sg1b	SAH/MTA/5'-dAdo Deaminase
Archaeoglobus profundus DSM 5631	284161269	Sg1b	SAH/MTA/5'-dAdo Deaminase
Archaeoglobus veneficus SNP6	327401019	Sg1b	SAH/MTA/5'-dAdo Deaminase
Aromatoleum aromaticum EbN1	56475928	Sg2	MTA/5'-dAdo/Ado Deaminase
Azoarcus sp. BH72	119898878	Sg2	MTA/5'-dAdo/Ado Deaminase
Azoarcus sp. KH32C	358638550	Sg2	MTA/5'-dAdo/Ado Deaminase
Azotobacter vinelandii DJ	226943694	Sg2	MTA/5'-dAdo/Ado Deaminase
Bacillus alcalophilus ATCC 27647	402299138	Sg1a	SAH/MTA Deaminase
Bacillus alcalophilus ATCC 27647	402299011	Sg9	Guanine Deaminase
Bacillus anthracis str. A2012	65319185	Sg4	5'-dAdo/Ado Deaminase
Bacillus anthracis str. Ames	30261905	Sg4	5'-dAdo/Ado Deaminase
Bacillus cellulosilyticus DSM 2522	317128620	Sg1a	SAH/MTA Deaminase
Bacillus cereus 03BB102	225863770	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus 03BB108	196045020	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus 172560W	229178305	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus 95/8201	229121453	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus AH1134	206970647	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus AH1271	229029603	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus AH1273	229017192	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus AH187	217959387	Sg4	5'-dAdo/Ado Deaminase

Organism	GI Number	Sg	Function
Bacillus cereus AH603	229059557	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus AH621	229166770	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus AH676	229043659	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus AND1407	401100969	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus ATCC 10876	229189999	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus ATCC 10987	2462117	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus ATCC 10987	42736941	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus ATCC 14579	30019935	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus ATCC 4342	229155478	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus B4264	218233043	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BAG1X1-2	401644151	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BAG1X1-3	401637644	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BAG2X1-1	401649456	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BAG2X1-2	401647728	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BAG3X2-1	401102048	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BAG5X1-1	401135601	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BAG5X2-1	401140262	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BAG6X1-1	402435656	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BAG6X1-2	401144982	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BDRD-Cer4	229127222	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BDRD-ST196	229132732	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus BDRD-ST26	229138605	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus biovar anthracis str. CI	301053436	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus E33L	51977022	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus F65185	229069450	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus F837/76	376265752	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus FRI-35	402557855	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus G9241	47565590	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus G9842	218896835	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus H3081.97	206975034	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus HuA2-1	402455989	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus HuA2-4	401165478	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus HuA4-10	401170102	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus HuB1-1	402446765	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus ISP3191	401186307	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus m1293	229196109	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus m1550	229150105	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus MC67	401196778	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus MM3	229172558	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus NVH0597-99	196041244	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus Q1	222095526	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus R309803	229160865	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus Rock1-15	229109356	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus Rock1-3	229115336	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus Rock3-28	229102492	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus Rock3-42	229090873	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus Rock3-44	229084858	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus Rock4-18	229074831	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus Rock4-2	229079081	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VD014	401217546	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VD045	401227400	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VD048	401224952	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VD078	401231302	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VD102	401241928	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VD107	401249584	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VD115	401253694	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VD156	401273332	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VD166	401275063	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VD169	401285773	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VD200	401294813	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VDM022	401296447	Sg4	5'-dAdo/Ado Deaminase

Organism	GI Number	Sg	Function
Bacillus cereus VDM034	401303214	Sg4	5'-dAdo/Ado Deaminase
Bacillus cereus VDM062	401307506	Sg4	5'-dAdo/Ado Deaminase
Bacillus clausii KSM-K16	56963827	Sg1a	SAH/MTA Deaminase
Bacillus cytotoxicus NVH 391-98	152975255	Sg4	5'-dAdo/Ado Deaminase
Bacillus halodurans C-125	15614255	Sg1a	SAH/MTA Deaminase
Bacillus halodurans C-125	162416220	Sg1a	SAH/MTA Deaminase
Bacillus halodurans C-125	15613309	Sg9	Guanine Deaminase
Bacillus macauensis ZFHKF-1	392958953	Sg1a	SAH/MTA Deaminase
Bacillus megaterium DSM 319	295703084	Sg9	Guanine Deaminase
Bacillus megaterium QM B1551	294497714	Sg9	Guanine Deaminase
Bacillus megaterium WSH-002	384048466	Sg9	Guanine Deaminase
Bacillus methanolicus PB1	387928511	Sg9	Guanine Deaminase
Bacillus mycoides DSM 2048	229011199	Sg4	5'-dAdo/Ado Deaminase
Bacillus mycoides Rock1-4	229007884	Sg4	5'-dAdo/Ado Deaminase
Bacillus mycoides Rock3-17	228996976	Sg4	5'-dAdo/Ado Deaminase
Bacillus pseudofirmus OF4	288556132	Sg1a	SAH/MTA Deaminase
Bacillus pseudomycoides DSM 12442	228990902	Sg4	5'-dAdo/Ado Deaminase
Bacillus selenitireducens MLS10	297584388	Sg1a	SAH/MTA Deaminase
Bacillus selenitireducens MLS10	297582902	Sg9	Guanine Deaminase
Bacillus smithii 7_3_47FAA	365158272	Sg9	Guanine Deaminase
Bacillus sp. 1NLA3E	373858121	Sg9	Guanine Deaminase
Bacillus sp. 2_A_57_CT2	319653145	Sg9	Guanine Deaminase
Bacillus thuringiensis BMB171	296502493	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis HD-771	402561058	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis HD-789	401873394	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis IBL 200	228907597	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis IBL 4222	228900482	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis MC28	407704282	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar andalousiensis BGSC 4AW1	228936987	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar berliner ATCC 10792	228939014	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar chinensis CT-43	384185807	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar finitimus YBT-020	384179851	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar huazhongensis BGSC 4BD1	228920608	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar konkukian str. 97-27	49477409	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar kurstaki str. T03a001	228952282	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar pakistani str. T13001	228958183	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar pondicheriensis BGSC 4BA1	228926904	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar pulsiensis BGSC 4CC1	228914493	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar sotto str. T04001	228964893	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis serovar tochigiensis BGSC 4Y1	228984996	Sg4	5'-dAdo/Ado Deaminase
Bacillus thuringiensis str. Al Hakam	118477326	Sg4	5'-dAdo/Ado Deaminase
Bacillus weihenstephanensis KBAB4	163939704	Sg4	5'-dAdo/Ado Deaminase
Beggiatoa alba B18LD	386827021	Sg2	MTA/5'-dAdo/Ado Deaminase
Beggiatoa sp. PS	153873847	Sg2	MTA/5'-dAdo/Ado Deaminase
Bermanella marisrubri	94499387	Sg2	MTA/5'-dAdo/Ado Deaminase
beta proteobacterium KB13	254467827	Sg2	MTA/5'-dAdo/Ado Deaminase
Bilophila sp. 4_1_30	345889947	Sg9	Guanine Deaminase
Bilophila wadsworthia 3_1_6	317487162	Sg9	Guanine Deaminase
Blastocystis hominis	300121918	Sg2	MTA/5'-dAdo/Ado Deaminase
Brevibacillus brevis NBRC 100599	226310925	Sg1a	SAH/MTA Deaminase
Brevibacillus brevis NBRC 100599	226312096	Sg1a	SAH/MTA Deaminase
Brevibacillus laterosporus GI-9	372457910	Sg1a	SAH/MTA Deaminase
Brevibacillus laterosporus LMG 15441	339010906	Sg1a	SAH/MTA Deaminase
Brevibacillus sp. BC25	398816055	Sg1a	SAH/MTA Deaminase
Brevibacillus sp. BC25	398816901	Sg1a	SAH/MTA Deaminase
Brevibacillus sp. CF112	399051035	Sg1a	SAH/MTA Deaminase
Brevibacillus sp. CF112	399053627	Sg1a	SAH/MTA Deaminase
Caldalkalibacillus thermarum TA2.A1	335040522	Sg4	5'-dAdo/Ado Deaminase
Caldicellulosiruptor bescii DSM 6725	222528885	Sg1a	SAH/MTA Deaminase
Caldicellulosiruptor hydrothermalis 108	312128012	Sg1a	SAH/MTA Deaminase
Caldicellulosiruptor kristjanssonii 177R1B	312792882	Sg1a	SAH/MTA Deaminase

Organism	GI Number	Sg	Function
Caldicellulosiruptor kronotskyensis 2002	312622825	Sg1a	SAH/MTA Deaminase
Caldicellulosiruptor lactoaceticus 6A	344996935	Sg1a	SAH/MTA Deaminase
Caldicellulosiruptor obsidiansis OB47	302871479	Sg1a	SAH/MTA Deaminase
Caldicellulosiruptor owensensis OL	312134758	Sg1a	SAH/MTA Deaminase
Caldicellulosiruptor saccharolyticus DSM 8903	146296488	Sg1a	SAH/MTA Deaminase
Caldithrix abyssi DSM 13497	373458019	Sg7	5'-dAdo/Ado Deaminase
Caldithrix abyssi DSM 13497	373459944	Sg9	Guanine Deaminase
Caloramator australicus RC3	397905043	Sg9	Guanine Deaminase
Candidatus Accumulibacter phosphatis clade IIA str. UW-1	257092705	Sg2	MTA/5'-dAdo/Ado Deaminase
Candidatus Chloracidobacterium thermophilum	157273296	Sg9	Guanine Deaminase
Candidatus Chloracidobacterium thermophilum B	347755953	Sg9	Guanine Deaminase
Candidatus Desulforudis audaxviator MP104C	169830835	Sg1a	SAH/MTA Deaminase
Candidatus Micrarchaeum acidiphilum ARMAN-2	255513470	Sg11	SAH Deaminase
Carboxydothermus hydrogenoformans Z-2901	78042858	Sg1a	SAH/MTA Deaminase
Carboxydothermus hydrogenoformans Z-2901	78043497	Sg9	Guanine Deaminase
Celeribacter baekdonensis B30	407788291	Sg10	8-Oxoadenine Deaminase
Cellvibrio japonicus Ueda107	192359201	Sg2	MTA/5'-dAdo/Ado Deaminase
Cellvibrio sp. BR	388257211	Sg2	MTA/5'-dAdo/Ado Deaminase
Chloroflexus aggregans DSM 9485	219848329	Sg7	5'-dAdo/Ado Deaminase
Chloroflexus aurantiacus J-10-fl	163847463	Sg7	5'-dAdo/Ado Deaminase
Chondromyces apiculatus DSM 436	405354369	Sg9	Guanine Deaminase
Chromobacterium violaceum ATCC 12472	34496487	Sg2	MTA/5'-dAdo/Ado Deaminase
Clostridium cellulolyticum H10	220929580	Sg1a	SAH/MTA Deaminase
Clostridium clariflavum DSM 19732	374297278	Sg1a	SAH/MTA Deaminase
Clostridium papyrosolvens DSM 2782	326204458	Sg1a	SAH/MTA Deaminase
Clostridium sp. BNL1100	376261823	Sg1a	SAH/MTA Deaminase
Clostridium thermocellum ATCC 27405	125973714	Sg1a	SAH/MTA Deaminase
Congregibacter litoralis KT71	88705959	Sg2	MTA/5'-dAdo/Ado Deaminase
Corallococcus coralloides DSM 2259	383458147	Sg9	Guanine Deaminase
Coxiella burnetii CbuG_Q212	212212982	Sg2	MTA/5'-dAdo/Ado Deaminase
Coxiella burnetii Dugway 5J108-111	209364106	Sg2	MTA/5'-dAdo/Ado Deaminase
Coxiella burnetii 'MSU Goat Q177'	153207936	Sg2	MTA/5'-dAdo/Ado Deaminase
Cycloclasticus sp. P1	407716515	Sg2	MTA/5'-dAdo/Ado Deaminase
Dechloromonas aromatica RCB	71906868	Sg2	MTA/5'-dAdo/Ado Deaminase
Dechlorosoma suillum PS	372488586	Sg2	MTA/5'-dAdo/Ado Deaminase
Dehalobacter sp. DCA	410658783	Sg1a	SAH/MTA Deaminase
Desulfitobacterium dehalogenans ATCC 51507	392394730	Sg1a	SAH/MTA Deaminase
Desulfitobacterium hafniense DCB-2	219669780	Sg1a	SAH/MTA Deaminase
Desulfitobacterium hafniense DP7	361854316	Sg1a	SAH/MTA Deaminase
Desulfitobacterium hafniense Y51	89895347	Sg1a	SAH/MTA Deaminase
Desulfitobacterium metallireducens DSM 15288	354558628	Sg1a	SAH/MTA Deaminase
Desulfosporosinus acidiphilus SJ4	392426864	Sg1a	SAH/MTA Deaminase
Desulfosporosinus acidiphilus SJ4	392425377	Sg9	Guanine Deaminase
Desulfosporosinus meridiei DSM 13257	402574100	Sg1a	SAH/MTA Deaminase
Desulfosporosinus meridiei DSM 13257	402571851	Sg9	Guanine Deaminase
Desulfosporosinus orientis DSM 765	374997113	Sg1a	SAH/MTA Deaminase
Desulfosporosinus orientis DSM 765	374994062	Sg9	Guanine Deaminase
Desulfosporosinus sp. OT	345859710	Sg1a	SAH/MTA Deaminase
Desulfosporosinus sp. OT	345856704	Sg9	Guanine Deaminase
Desulfosporosinus youngiae DSM 17734	374582905	Sg1a	SAH/MTA Deaminase
Desulfosporosinus youngiae DSM 17734	374579958	Sg9	Guanine Deaminase
Desulfotomaculum acetoxidans DSM 771	258514475	Sg1a	SAH/MTA Deaminase
Desulfotomaculum gibsoniae DSM 7213	357038764	Sg1a	SAH/MTA Deaminase
Desulfotomaculum gibsoniae DSM 7213	357038847	Sg9	Guanine Deaminase
Desulfotomaculum kuznetsovii DSM 6115	333979669	Sg1a	SAH/MTA Deaminase
Desulfotomaculum nigrificans DSM 574	323702678	Sg1a	SAH/MTA Deaminase
Desulfotomaculum reducens MI-1	134299907	Sg1a	SAH/MTA Deaminase
Desulfotomaculum ruminis DSM 2154	334340360	Sg1a	SAH/MTA Deaminase
Desulfovibrio aespoeensis Aspo-2	317153206	Sg6	5'-dAdo Deaminase
Desulfovibrio africanus str. Walvis Bay	374301048	Sg6	5'-dAdo Deaminase
Desulfovibrio alaskensis G20	78356855	Sg6	5'-dAdo Deaminase

Organism	GI Number	Sg	Function
Desulfovibrio desulfuricans ND132	376295307	Sg6	5'-dAdo Deaminase
Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	220905193	Sg6	5'-dAdo Deaminase
Desulfovibrio fructosovorans JJ	303247207	Sg6	5'-dAdo Deaminase
Desulfovibrio magneticus RS-1	239907238	Sg6	5'-dAdo Deaminase
Desulfovibrio magneticus str. Maddingley MBC34	410462003	Sg6	5'-dAdo Deaminase
Desulfovibrio piger ATCC 29098	212702496	Sg6	5'-dAdo Deaminase
Desulfovibrio salexigens DSM 2638	242279072	Sg6	5'-dAdo Deaminase
Desulfovibrio sp. 3_1_syn3	303327872	Sg6	5'-dAdo Deaminase
Desulfovibrio sp. 6_1_46AFAA	345893536	Sg6	5'-dAdo Deaminase
Desulfovibrio sp. A2	347731243	Sg6	5'-dAdo Deaminase
Desulfovibrio sp. FW1012B	357632585	Sg6	5'-dAdo Deaminase
Desulfovibrio sp. U5L	386392840	Sg6	5'-dAdo Deaminase
Desulfovibrio vulgaris DP4	120602380	Sg6	5'-dAdo Deaminase
Desulfovibrio vulgaris str. Hildenborough	46580235	Sg6	5'-dAdo Deaminase
Desulfovibrio vulgaris str. 'Miyazaki F'	218885652	Sg6	5'-dAdo Deaminase
Dethiobacter alkaliphilus AHT 1	225181018	Sg1a	SAH/MTA Deaminase
Dialister microaerophilus UPII 345-E	313891988	Sg1a	SAH/MTA Deaminase
Dictyostelium discoideum AX4	66808699	Sg2	MTA/5'-dAdo/Ado Deaminase
Dictyostelium fasciculatum	328876856	Sg2	MTA/5'-dAdo/Ado Deaminase
Dictyostelium purpureum	330840449	Sg2	MTA/5'-dAdo/Ado Deaminase
Ectothiorhodospira sp. PHS-1	374622997	Sg2	MTA/5'-dAdo/Ado Deaminase
endosymbiont of Tevnia jerichonana (vent Tica)	345863456	Sg2	MTA/5'-dAdo/Ado Deaminase
Ferroglobus placidus DSM 10642	288931156	Sg1b	SAH/MTA/5'-dAdo Deaminase
Ferroplasma acidarmanus fer1	257076627	Sg11	SAH Deaminase
Fervidobacterium nodosum Rt17-B1	154249043	Sg8	SAH/MTA Deaminase
Fervidobacterium pennivorans DSM 9078	383786427	Sg8	SAH/MTA Deaminase
Filifactor alocis ATCC 35896	374307435	Sg1a	SAH/MTA Deaminase
Frankia sp. CN3	358458993	Sg10	8-Oxoadenine Deaminase
Frankia sp. EuI1c	312198398	Sg10	8-Oxoadenine Deaminase
Frateuria aurantia DSM 6220	383316999	Sg2	MTA/5'-dAdo/Ado Deaminase
gamma proteobacterium BDW918	386286357	Sg2	MTA/5'-dAdo/Ado Deaminase
gamma proteobacterium HIMB55	374619383	Sg2	MTA/5'-dAdo/Ado Deaminase
gamma proteobacterium HTCC2207	90416544	Sg2	MTA/5'-dAdo/Ado Deaminase
gamma proteobacterium HTCC5015	254447042	Sg2	MTA/5'-dAdo/Ado Deaminase
gamma proteobacterium IMCC1989	331005952	Sg2	MTA/5'-dAdo/Ado Deaminase
gamma proteobacterium IMCC3088	329894506	Sg2	MTA/5'-dAdo/Ado Deaminase
gamma proteobacterium NOR51-B	254281621	Sg2	MTA/5'-dAdo/Ado Deaminase
gamma proteobacterium NOR5-3	254515160	Sg2	MTA/5'-dAdo/Ado Deaminase
Glaciecola psychrophila 170	410612095	Sg10	8-Oxoadenine Deaminase
Hahella chejuensis KCTC 2396	83647052	Sg2	MTA/5'-dAdo/Ado Deaminase
Haladaptatus paucihalophilus DX253	322368161	Sg1b	SAH/MTA/5'-dAdo Deaminase
Halalkalicoccus jeotgali B3	300709789	Sg1b	SAH/MTA/5'-dAdo Deaminase
Halalkalicoccus jeotgali B3	300711133	Sg9	Guanine Deaminase
Haloarcula hispanica ATCC 33960	344210262	Sg9	Guanine Deaminase
Haloarcula marismortui ATCC 43049	55376787	Sg9	Guanine Deaminase
Halobacterium sp. DL1	354609692	Sg1b	SAH/MTA/5'-dAdo Deaminase
Halobacterium sp. DL1	354611609	Sg9	Guanine Deaminase
Halobacterium sp. NRC-1	15791066	Sg1b	SAH/MTA/5'-dAdo Deaminase
Halobacterium sp. NRC-1	15790271	Sg9	Guanine Deaminase
Halobiforma lacisalsi AJ5	383619876	Sg1b	SAH/MTA/5'-dAdo Deaminase
Halobiforma lacisalsi AJ5	383622348	Sg9	Guanine Deaminase
Halococcus hamelinensis 100A6	409723629	Sg1b	SAH/MTA/5'-dAdo Deaminase
Haloferax mediterranei ATCC 33500	389845656	Sg1b	SAH/MTA/5'-dAdo Deaminase
Haloferax mediterranei ATCC 33500	389848339	Sg9	Guanine Deaminase
Haloferax volcanii DS2	292654344	Sg1b	SAH/MTA/5'-dAdo Deaminase
Haloferax volcanii DS2	292657041	Sg9	Guanine Deaminase
Halogeometricum borinquense DSM 11551	313124800	Sg1b	SAH/MTA/5'-dAdo Deaminase
Halogeometricum borinquense DSM 11551	313125853	Sg9	Guanine Deaminase
Halogranum salarium B-1	399576134	Sg1b	SAH/MTA/5'-dAdo Deaminase

Organism	GI Number	Sg	Function
Halogranum salarium B-1	399578642	Sg9	Guanine Deaminase
Halomicrobium mukohataei DSM 12286	257386819	Sg1b	SAH/MTA/5'-dAdo Deaminase
Halomicrobium mukohataei DSM 12286	257388127	Sg9	Guanine Deaminase
halophilic archaeon DL31	345005326	Sg1b	SAH/MTA/5'-dAdo Deaminase
halophilic archaeon DL31	345006111	Sg1b	SAH/MTA/5'-dAdo Deaminase
halophilic archaeon DL31	345005185	Sg9	Guanine Deaminase
Halopiger xanaduensis SH-6	336252669	Sg1b	SAH/MTA/5'-dAdo Deaminase
Halopiger xanaduensis SH-6	336251757	Sg9	Guanine Deaminase
Haloquadratum walsbyi C23	385804851	Sg1b	SAH/MTA/5'-dAdo Deaminase
Haloquadratum walsbyi C23	385803442	Sg9	Guanine Deaminase
Haloquadratum walsbyi DSM 16790	110669287	Sg1b	SAH/MTA/5'-dAdo Deaminase
Haloquadratum walsbyi DSM 16790	110667994	Sg9	Guanine Deaminase
Halorhabdus tiamatea SARL4B	335438625	Sg1b	SAH/MTA/5'-dAdo Deaminase
Halorhabdus utahensis DSM 12940	257052590	Sg1b	SAH/MTA/5'-dAdo Deaminase
Halorhodospira halophila SL1	121997375	Sg2	MTA/5'-dAdo/Ado Deaminase
Halorubrum lacusprofundi ATCC 49239	222478500	Sg1b	SAH/MTA/5'-dAdo Deaminase
Halorubrum lacusprofundi ATCC 49239	222480611	Sg9	Guanine Deaminase
Haloterrigena turkmenica DSM 5511	284163440	Sg1b	SAH/MTA/5'-dAdo Deaminase
Haloterrigena turkmenica DSM 5511	284167168	Sg9	Guanine Deaminase
Heliobacterium modesticaldum Ice1	167630246	Sg1a	SAH/MTA Deaminase
Heliobacterium modesticaldum Ice1	167630730	Sg9	Guanine Deaminase
Herpetosiphon aurantiacus DSM 785	159901377	Sg7	5'-dAdo/Ado Deaminase
Hydrocarboniphaga effusa AP103	392954008	Sg2	MTA/5'-dAdo/Ado Deaminase
Ignavibacterium album JCM 16511	385810930	Sg9	Guanine Deaminase
Janthinobacterium sp. Marseille	152980942	Sg10	8-Oxoadenine Deaminase
Kangiella koreensis DSM 16069	256822116	Sg2	MTA/5'-dAdo/Ado Deaminase
Kyrpidia tusciae DSM 2912	295696295	Sg1a	SAH/MTA Deaminase
Kyrpidia tusciae DSM 2912	295696613	Sg9	Guanine Deaminase
Laribacter hongkongensis HLHK9	226939880	Sg2	MTA/5'-dAdo/Ado Deaminase
Lawsonia intracellularis PHE/MN1-00	94986921	Sg6	5'-dAdo Deaminase
Lawsonia intracellularis PHE/MN1-00	162416190	Sg6	5'-dAdo Deaminase
Limnobacter sp. MED105	149928285	Sg2	MTA/5'-dAdo/Ado Deaminase
Mahella australiensis 50-1 BON	332980925	Sg1a	SAH/MTA Deaminase
Marichromatium purpuratum 984	344342370	Sg2	MTA/5'-dAdo/Ado Deaminase
marine gamma proteobacterium HTCC2080	119505050	Sg2	MTA/5'-dAdo/Ado Deaminase
marine gamma proteobacterium HTCC2143	119476908	Sg2	MTA/5'-dAdo/Ado Deaminase
marine gamma proteobacterium HTCC2148	254481817	Sg2	MTA/5'-dAdo/Ado Deaminase
Marinobacter adhaerens HP15	385330969	Sg2	MTA/5'-dAdo/Ado Deaminase
Marinobacter algicola DG893	149374309	Sg2	MTA/5'-dAdo/Ado Deaminase
Marinobacter aquaeolei VT8	120555405	Sg2	MTA/5'-dAdo/Ado Deaminase
Marinobacter hydrocarbonoclasticus ATCC 49840	387814819	Sg2	MTA/5'-dAdo/Ado Deaminase
Marinobacter manganoxydans MnI7-9	358448853	Sg2	MTA/5'-dAdo/Ado Deaminase
Marinobacter sp. BSs20148	399544435	Sg2	MTA/5'-dAdo/Ado Deaminase
Marinobacter sp. ELB17	126665969	Sg2	MTA/5'-dAdo/Ado Deaminase
Marinomonas sp. MED121	87119283	Sg2	MTA/5'-dAdo/Ado Deaminase
Megamonas hypermegale ART12/1	291533823	Sg1a	SAH/MTA Deaminase
Megasphaera elsdenii DSM 20460	348026976	Sg1a	SAH/MTA Deaminase
Megasphaera genomosp. type_1 str. 28L	290968533	Sg1a	SAH/MTA Deaminase
Megasphaera sp. UPII 135-E	342218412	Sg1a	SAH/MTA Deaminase
Methanobacterium formicicum DSM 3637	408381208	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanobacterium sp. AL-21	325957773	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanobacterium sp. Maddingley MBC34	410720785	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanobacterium sp. SWAN-1	333986326	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanobrevibacter ruminantium M1	288559345	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanobrevibacter smithii ATCC 35061	148643319	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanobrevibacter smithii DSM 2375	222445560	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanocaldococcus fervens AG86	256811411	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanocaldococcus infernus ME	296109783	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanocaldococcus jannaschii DSM 2661	15669736	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanocaldococcus sp. FS406-22	289192226	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanocaldococcus vulcanius M7	261402768	Sg1b	SAH/MTA/5'-dAdo Deaminase

Organism	GI Number	Sg	Function
Methanococcoides burtonii DSM 6242	91773163	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanococcus aeolicus Nankai-3	150400987	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanococcus maripaludis C5	134045132	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanococcus maripaludis C6	159905566	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanococcus maripaludis C7	150402659	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanococcus maripaludis S2	45359054	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanococcus maripaludis X1	340624802	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanococcus vannielii SB	150399551	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanocorpusculum labreanum Z	124485874	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanoculleus bourgensis MS2	397780245	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanoculleus marisnigri JR1	126179037	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanofollis liminatans DSM 4140	395646173	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanohalobium evestigatum Z-7303	298674732	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanohalobium evestigatum Z-7303	298674749	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanohalophilus mahii DSM 5219	294495200	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanolinea tarda NOBI-1	355571893	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanolobus psychrophilus R15	410670209	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanoplanus limicola DSM 2279	374628529	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanoplanus petrolearius DSM 11571	307352986	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanoregula boonei 6A8	154150311	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanosalsum zhilinae DSM 4017	336476448	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanosarcina acetivorans C2A	20090140	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanosarcina acetivorans C2A	162416231	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanosarcina barkeri str. Fusaro	73670698	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanosarcina mazei Go1	21228381	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanosarcina mazei Go1	162416232	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanosphaera stadtmanae DSM 3091	84489028	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanosphaerula palustris E1-9c	219851668	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanospirillum hungatei JF-1	88603687	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanothermobacter marburgensis str. Marburg	304313866	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanothermobacter thermautotrophicus str. Delta H	15679502	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanothermococcus okinawensis IH1	336121909	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanothermus fervidus DSM 2088	312136929	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanotorris formicicus Mc-S-70	374635036	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methanotorris igneus Kol 5	333910056	Sg1b	SAH/MTA/5'-dAdo Deaminase
Methylobacillus flagellatus KT	91775930	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylobacter tundripaludum SV96	344943153	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylococcus capsulatus str. Bath	53804375	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylomicrobium album BG8	381151397	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylomicrobium alcaliphilum 20Z	357404855	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylomonas methanica MC09	333983431	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylophaga aminisulfidivorans MP	335043680	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylophaga sp. JAM1	387126837	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylophaga sp. JAM7	387129950	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylophaga thiooxidans DMS010	254492160	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylophilales bacterium HTCC2181	118594964	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylotenera mobilis JLW8	253996273	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylotenera versatilis 301	297538147	Sg2	MTA/5'-dAdo/Ado Deaminase
Methyloversatilis universalis FAM5	334131892	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylovorus glucosetrophus SIP3-4	253999389	Sg2	MTA/5'-dAdo/Ado Deaminase
Methylovorus sp. MP688	313201411	Sg2	MTA/5'-dAdo/Ado Deaminase
Moorella thermoacetica ATCC 39073	83590072	Sg9	Guanine Deaminase
Myxococcus fulvus HW-1	338536821	Sg9	Guanine Deaminase
Myxococcus xanthus DK 1622	108760686	Sg9	Guanine Deaminase
Natranaerobius thermophilus JW/NM-WN-LF	188586216	Sg1a	SAH/MTA Deaminase
Natranaerobius thermophilus JW/NM-WN-LF	188585774	Sg9	Guanine Deaminase
Natrialba magadii ATCC 43099	289581520	Sg1b	SAH/MTA/5'-dAdo Deaminase
Natrialba magadii ATCC 43099	289582777	Sg9	Guanine Deaminase
Natrinema sp. J7-2	397772383	Sg1b	SAH/MTA/5'-dAdo Deaminase
Natrinema sp. J7-2	397775256	Sg9	Guanine Deaminase
Natronomonas pharaonis DSM 2160	76801137	Sg1b	SAH/MTA/5'-dAdo Deaminase

Organism	GI Number	Sg	Function
Nitrococcus mobilis Nb-231	88810641	Sg2	MTA/5'-dAdo/Ado Deaminase
Nitrosococcus halophilus Nc4	292492450	Sg2	MTA/5'-dAdo/Ado Deaminase
Nitrosococcus oceani ATCC 19707	77165758	Sg2	MTA/5'-dAdo/Ado Deaminase
Nitrosococcus watsonii C-113	300114707	Sg2	MTA/5'-dAdo/Ado Deaminase
Nitrosomonas sp. AL212	325982084	Sg2	MTA/5'-dAdo/Ado Deaminase
Nitrosomonas sp. Is79A3	339484010	Sg2	MTA/5'-dAdo/Ado Deaminase
Nitrosospira multiformis ATCC 25196	82703305	Sg2	MTA/5'-dAdo/Ado Deaminase
Oscillochloris trichoides DG-6	309790136	Sg7	5'-dAdo/Ado Deaminase
Paenibacillus alvei DSM 29	402815300	Sg1a	SAH/MTA Deaminase
Paenibacillus curdlanolyticus YK9	304406901	Sg1a	SAH/MTA Deaminase
Paenibacillus curdlanolyticus YK9	304407468	Sg9	Guanine Deaminase
Paenibacillus dendritiformis C454	374606202	Sg1a	SAH/MTA Deaminase
Paenibacillus dendritiformis C454	374606828	Sg1a	SAH/MTA Deaminase
Paenibacillus elgii B69	357014314	Sg1a	SAH/MTA Deaminase
Paenibacillus lactis 154	354582691	Sg1a	SAH/MTA Deaminase
Paenibacillus mucilaginosus 3016	379720955	Sg1a	SAH/MTA Deaminase
Paenibacillus mucilaginosus K02	386721945	Sg9	Guanine Deaminase
Paenibacillus mucilaginosus KNP414	337747099	Sg1a	SAH/MTA Deaminase
Paenibacillus mucilaginosus KNP414	337745352	Sg9	Guanine Deaminase
Paenibacillus peoriae KCTC 3763	390453515	Sg1a	SAH/MTA Deaminase
Paenibacillus polymyxa E681	308069468	Sg1a	SAH/MTA Deaminase
Paenibacillus polymyxa SC2	310642521	Sg1a	SAH/MTA Deaminase
Paenibacillus popilliae ATCC 14706	410831044	Sg1a	SAH/MTA Deaminase
Paenibacillus sp. Aloe-11	375308996	Sg1a	SAH/MTA Deaminase
Paenibacillus sp. HGF5	329929741	Sg1a	SAH/MTA Deaminase
Paenibacillus sp. HGF7	334137813	Sg1a	SAH/MTA Deaminase
Paenibacillus sp. HGF7	334133556	Sg4	5'-dAdo/Ado Deaminase
Paenibacillus sp. JDR-2	251796461	Sg1a	SAH/MTA Deaminase
Paenibacillus sp. JDR-2	251797317	Sg9	Guanine Deaminase
Paenibacillus sp. oral taxon 786 str. D14	253577521	Sg1a	SAH/MTA Deaminase
Paenibacillus sp. Y412MC10	261406089	Sg1a	SAH/MTA Deaminase
Paenibacillus terrae HPL-003	374324456	Sg1a	SAH/MTA Deaminase
Paenibacillus terrae HPL-003	374324681	Sg1a	SAH/MTA Deaminase
Paenibacillus vortex V453	315646469	Sg1a	SAH/MTA Deaminase
Pelosinus fermentans DSM 17108	392960171	Sg1a	SAH/MTA Deaminase
Pelosinus fermentans DSM 17108	392960651	Sg1a	SAH/MTA Deaminase
Pelosinus fermentans JBW45	392525916	Sg1a	SAH/MTA Deaminase
Pelotomaculum thermopropionicum SI	147678063	Sg1a	SAH/MTA Deaminase
Peptoniphilus harei ACS-146-V-Sch2b	313888086	Sg1a	SAH/MTA Deaminase
Peptoniphilus lacrimalis 315-B	282883342	Sg1a	SAH/MTA Deaminase
Peptoniphilus rhinitidis 1-13	399924014	Sg1a	SAH/MTA Deaminase
Peptoniphilus sp. oral taxon 375 str. F0436	342215639	Sg1a	SAH/MTA Deaminase
Peptoniphilus sp. oral taxon 386 str. F0131	299144487	Sg1a	SAH/MTA Deaminase
Peptoniphilus sp. oral taxon 836 str. F0141	300814796	Sg1a	SAH/MTA Deaminase
Phytophthora infestans T30-4	301110462	Sg2	MTA/5'-dAdo/Ado Deaminase
Phytophthora infestans T30-4	301117868	Sg2	MTA/5'-dAdo/Ado Deaminase
Phytophthora sojae	348667430	Sg2	MTA/5'-dAdo/Ado Deaminase
Phytophthora sojae	348688636	Sg2	MTA/5'-dAdo/Ado Deaminase
Picrophilus torridus DSM 9790	48477797	Sg11	SAH Deaminase
Polysphondylium pallidum PN500	281202610	Sg2	MTA/5'-dAdo/Ado Deaminase
Porphyromonas asaccharolytica DSM 20707	332299943	Sg1b	SAH/MTA/5'-dAdo Deaminase
Porphyromonas asaccharolytica PR426713P-I	313887254	Sg1b	SAH/MTA/5'-dAdo Deaminase
Porphyromonas endodontalis ATCC 35406	229495478	Sg1b	SAH/MTA/5'-dAdo Deaminase
Porphyromonas sp. oral taxon 279 str. F0450	402846236	Sg1b	SAH/MTA/5'-dAdo Deaminase
Porphyromonas uenonis 60-3	228470332	Sg1b	SAH/MTA/5'-dAdo Deaminase
Pseudogulbenkiania ferrooxidans 2002	224824519	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudogulbenkiania sp. NH8B	347538759	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas aeruginosa 138244	334835211	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas aeruginosa ATCC 700888	404537409	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas aeruginosa M18	386057869	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas aeruginosa PA7	152983730	Sg2	MTA/5'-dAdo/Ado Deaminase

Organism	GI Number	Sg	Function
Pseudomonas aeruginosa PACS2	107102700	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas aeruginosa PAO1	15598366	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas aeruginosa UCBPP-PA14	116051159	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas avellanae BPIC 631	407990790	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas brassicacearum subsp. brassicacearum NFM421	330808285	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas chlororaphis O6	389680971	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas chlororaphis subsp. aureofaciens 30-84	397883504	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas entomophila L48	104780666	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas extremaustralis 14-3 substr. 14-3b	395651183	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas fluorescens A506	387892902	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas fluorescens F113	378949565	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas fluorescens Pf0-1	77460301	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas fluorescens Q2-87	397889241	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas fluorescens R124	404305413	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas fluorescens SBW25	229589162	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas fluorescens SS101	387997677	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas fluorescens WH6	312959696	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas fragi A22	402699915	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas fulva 12-X	333900206	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas fuscovaginae UPB0736	404401721	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas geniculata N1	408825073	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas mandelii JR-1	407365600	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas mendocina DLHK	400345472	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas mendocina NK-01	330502866	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas mendocina ymp	146306877	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas protegens Pf-5	70731666	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas pseudoalcaligenes CECT 5344	399520031	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas psychrotolerans L19	359781487	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas putida W619	170722310	Sg10	8-Oxoadenine Deaminase
Pseudomonas sp. Ag1	395798591	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. Chol1	409393717	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM102	398839969	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM17	399010547	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM18	399002584	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM21	398995407	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM24	398989436	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM25	398973358	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM30	398964531	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM33	398952126	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM41(2012)	398938561	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM49	398917336	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM50	398899111	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM55	398890629	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM60	398885429	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM67	398879366	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM74	398871471	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM78	398868653	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM79	398859092	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM80	398850871	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. GM84	398843691	Sg10	8-Oxoadenine Deaminase
Pseudomonas sp. HYS	409426090	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. M47T1	388545952	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. PAMC 25886	395499790	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. R81	408482975	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas sp. S9	374702041	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas stutzeri A1501	146282666	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas stutzeri ATCC 14405 = CCUG 16156	379064775	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas stutzeri ATCC 17588 = LMG 11199	339494279	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas stutzeri CCUG 29243	392421516	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas stutzeri DSM 10701	397687516	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas stutzeri DSM 4166	386020950	Sg2	MTA/5'-dAdo/Ado Deaminase

Organism	GI Number	Sg	Function
Pseudomonas stutzeri KOS6	409778759	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas stutzeri TS44	387967423	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas synxantha BG33R	388470050	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae Cit 7	330953526	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. aceris str. M302273	330972347	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. actinidiae str. M302091	330964795	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. aesculi str. NCPPB 3681	289623729	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. aptata str. DSM 50252	330980274	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. avellanae str. ISPaVe013	407996526	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. avellanae str. ISPaVe037	407996413	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. glycinea str. B076	320323354	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. lachrymans str. M301315	330986668	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. maculicola str. ES4326	330960548	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. mori str. 301020	330889349	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. morsprunorum str. M302280	330877043	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. phaseolicola 1448A	71733895	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. syringae 642	302187663	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. syringae B728a	66046878	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. tabaci str. ATCC 11528	257485851	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas syringae pv. tomato str. DC3000	28868949	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudomonas viridiflava UASWS0038	410092253	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudoxanthomonas spadix BD-a59	357417314	Sg2	MTA/5'-dAdo/Ado Deaminase
Pseudoxanthomonas suwonensis 11-1	319786813	Sg2	MTA/5'-dAdo/Ado Deaminase
Pyrococcus abyssi GE5	14520861	Sg1b	SAH/MTA/5'-dAdo Deaminase
Pyrococcus furiosus DSM 3638	18977910	Sg1b	SAH/MTA/5'-dAdo Deaminase
Pyrococcus horikoshii OT3	14591297	Sg1b	SAH/MTA/5'-dAdo Deaminase
Pyrococcus horikoshii OT3	161789036	Sg1b	SAH/MTA/5'-dAdo Deaminase
Pyrococcus sp. NA2	332157744	Sg1b	SAH/MTA/5'-dAdo Deaminase
Pyrococcus sp. ST04	389852822	Sg1b	SAH/MTA/5'-dAdo Deaminase
Pyrococcus yayanosii CH1	337284243	Sg1b	SAH/MTA/5'-dAdo Deaminase
Reinekea blandensis MED297	88800757	Sg2	MTA/5'-dAdo/Ado Deaminase
Rhodanobacter fulvus Jip2	389793007	Sg2	MTA/5'-dAdo/Ado Deaminase
Rhodanobacter sp. 115	389721634	Sg2	MTA/5'-dAdo/Ado Deaminase
Rhodanobacter sp. 116-2	389797714	Sg2	MTA/5'-dAdo/Ado Deaminase
Rhodanobacter spathiphylli B39	389774472	Sg2	MTA/5'-dAdo/Ado Deaminase
Rhodanobacter thiooxydans LCS2	389806384	Sg2	MTA/5'-dAdo/Ado Deaminase
Ricinus communis	255594903	Sg2	MTA/5'-dAdo/Ado Deaminase
Roseiflexus castenholzii DSM 13941	156744101	Sg7	5'-dAdo/Ado Deaminase
Roseiflexus sp. RS-1	148658401	Sg7	5'-dAdo/Ado Deaminase
Saccharophagus degradans 2-40	90021794	Sg2	MTA/5'-dAdo/Ado Deaminase
Salinisphaera shabanensis E1L3A	335420684	Sg2	MTA/5'-dAdo/Ado Deaminase
Selenomonas ruminantium subsp. lactilytica TAM6421	383754018	Sg1a	SAH/MTA Deaminase
Selenomonas sp. CM52	402833860	Sg1a	SAH/MTA Deaminase
Selenomonas sp. FOBRC6	401564592	Sg1a	SAH/MTA Deaminase
Selenomonas sp. FOBRC9	402301912	Sg1a	SAH/MTA Deaminase
Selenomonas sp. oral taxon 137 str. F0430	313895544	Sg1a	SAH/MTA Deaminase
Simiduia agarivorans SA1 = DSM 21679	410663080	Sg2	MTA/5'-dAdo/Ado Deaminase
Sporolactobacillus inulinus CASD	374710320	Sg9	Guanine Deaminase
Sporolactobacillus vineae DSM 21990 = SL153	404330030	Sg9	Guanine Deaminase
Stenotrophomonas maltophilia D457	386719260	Sg2	MTA/5'-dAdo/Ado Deaminase
Stenotrophomonas maltophilia JV3	344208162	Sg2	MTA/5'-dAdo/Ado Deaminase
Stenotrophomonas maltophilia K279a	190575196	Sg2	MTA/5'-dAdo/Ado Deaminase
Stenotrophomonas maltophilia R551-3	194366514	Sg2	MTA/5'-dAdo/Ado Deaminase
Stenotrophomonas sp. SKA14	254522291	Sg2	MTA/5'-dAdo/Ado Deaminase
Stigmatella aurantiaca DW4/3-1	115373684	Sg9	Guanine Deaminase
Stigmatella aurantiaca DW4/3-1	310823592	Sg9	Guanine Deaminase
Streptococcus anginosus 1_2_62CV	319939966	Sg5	5'-dAdo/Ado Deaminase
Streptococcus anginosus F0211	315222254	Sg5	5'-dAdo/Ado Deaminase
Streptococcus anginosus SK1138	400374374	Sg5	5'-dAdo/Ado Deaminase
Streptococcus anginosus SK52 = DSM 20563	335032486	Sg5	5'-dAdo/Ado Deaminase
Streptococcus anginosus subsp. whileyi CCUG 39159	383340131	Sg5	5'-dAdo/Ado Deaminase

Organism	GI Number	Sg	Function
Streptococcus cristatus ATCC 51100	322385066	Sg5	5'-dAdo/Ado Deaminase
Streptococcus gordonii str. Challis substr. CH1	157150044	Sg5	5'-dAdo/Ado Deaminase
Streptococcus infantis ATCC 700779	322388284	Sg5	5'-dAdo/Ado Deaminase
Streptococcus infantis SK1076	335029092	Sg5	5'-dAdo/Ado Deaminase
Streptococcus infantis SK970	343391918	Sg5	5'-dAdo/Ado Deaminase
Streptococcus infantis X	343400182	Sg5	5'-dAdo/Ado Deaminase
Streptococcus intermedius F0413	355363819	Sg5	5'-dAdo/Ado Deaminase
Streptococcus intermedius JTH08	392428100	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis ATCC 6249	306829721	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis B6	289167624	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis bv. 2 str. F0392	340771138	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis bv. 2 str. SK95	342833418	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis NCTC 12261	307701897	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis SK1073	339457332	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis SK1080	339455936	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis SK321	307706305	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis SK564	307708937	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis SK569	342837099	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis SK575	383350029	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis SK579	383353478	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis SK597	307705065	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis SK616	383344035	Sg5	5'-dAdo/Ado Deaminase
Streptococcus mitis SPAR10	395876851	Sg5	5'-dAdo/Ado Deaminase
Streptococcus oralis ATCC 35037	293365671	Sg5	5'-dAdo/Ado Deaminase
Streptococcus oralis SK10	383186488	Sg5	5'-dAdo/Ado Deaminase
Streptococcus oralis SK100	383184224	Sg5	5'-dAdo/Ado Deaminase
Streptococcus oralis SK1074	383347755	Sg5	5'-dAdo/Ado Deaminase
Streptococcus oralis SK255	334265937	Sg5	5'-dAdo/Ado Deaminase
Streptococcus oralis SK304	400367786	Sg5	5'-dAdo/Ado Deaminase
Streptococcus oralis SK313	343388443	Sg5	5'-dAdo/Ado Deaminase
Streptococcus oralis SK610	383183371	Sg5	5'-dAdo/Ado Deaminase
Streptococcus oralis Uo5	331266145	Sg5	5'-dAdo/Ado Deaminase
Streptococcus peroris ATCC 700780	322392322	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2061376	395594508	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2061617	395600874	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2070005	395573856	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2070035	395574791	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2070108	395579606	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2070335	395583699	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2070531	395587592	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2071004	395602303	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2080913	395607217	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2081074	395608263	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2082170	395613768	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2082239	395613286	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 2090008	395575076	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 459-5	381316034	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae 70585	225859123	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae ATCC 700669	221232098	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae CDC0288-04	183603522	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae CDC3059-06	168493257	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae CGSP14	182684315	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae D39	116515507	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae G54	194397682	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA05578	379637156	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA13224	379552008	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA13637	353812129	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA16531	353769117	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA17328	353827378	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA17570	332073663	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA40028	379574886	Sg5	5'-dAdo/Ado Deaminase

Organism	GI Number	Sg	Function
Streptococcus pneumoniae GA40410	379578678	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA41277	353838513	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA41301	332074936	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA41317	332200799	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA41437	353841597	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA43380	353849068	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA44452	353764072	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA47368	332201809	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA47461	379595656	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA47502	353748250	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA47522	379599992	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA47597	379599212	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA52306	353867138	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA54354	395889185	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA60080	395900240	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae GA60132	395909036	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae gamPNI0373	410476740	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae Hungary19A-6	169833999	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae INV104	387626615	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae JJA	225854804	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae MLV-016	183603825	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae Netherlands15B-37	353872354	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae P1031	225856991	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae R6	15458849	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae R6	161410744	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae SP11-BS70	148997592	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae SP14-BS69	149004511	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae SP18-BS74	149006995	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae SP19-BS75	149011588	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae SP23-BS72	149019308	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae SP3-BS71	148985696	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae SP6-BS73	148989473	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae SP9-BS68	148993087	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae SPNA45	405760708	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae TCH8431/19A	298502696	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pneumoniae TIGR4	111656887	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pseudopneumoniae IS7493	342163987	Sg5	5'-dAdo/Ado Deaminase
Streptococcus pseudopneumoniae SK674	383937960	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis ATCC 49296	315613384	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK1	327459442	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK1056	332358175	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK1057	327459045	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK1058	327490642	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK1059	332364739	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK1087	328945851	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK115	325689424	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK150	325693741	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK330	327468567	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK353	324990301	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK355	332364426	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK36	125717247	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK405	324992129	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK49	332363176	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK678	324994224	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis SK72	325688420	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sanguinis VMC66	323353448	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. 2_1_36FAA	262281915	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. AS14	401682181	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. BS35b	401684914	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. C300	322375454	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. F0441	410869346	Sg5	5'-dAdo/Ado Deaminase

Organism	GI Number	Sg	Function
Streptococcus sp. GMD1S	406587711	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. GMD2S	404468846	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. GMD6S	406576640	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. M143	270292519	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. oral taxon 056 str. F0418	339641238	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. oral taxon 058 str. F0407	358464559	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. oral taxon 071 str. 73H25AP	306825008	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. SK140	385260114	Sg5	5'-dAdo/Ado Deaminase
Streptococcus sp. SK643	385262097	Sg5	5'-dAdo/Ado Deaminase
Sulfobacillus acidophilus DSM 10332	379007998	Sg9	Guanine Deaminase
Sulfobacillus acidophilus TPY	339627639	Sg9	Guanine Deaminase
Sulfuricella denitrificans skB26	394989067	Sg2	MTA/5'-dAdo/Ado Deaminase
Sutterella parvirubra YIT 11816	378822009	Sg2	MTA/5'-dAdo/Ado Deaminase
Sutterella wadsworthensis 2_1_59BFAA	404657153	Sg2	MTA/5'-dAdo/Ado Deaminase
Sutterella wadsworthensis 3_1_45B	319942536	Sg2	MTA/5'-dAdo/Ado Deaminase
Symbiobacterium thermophilum IAM 14863	51892841	Sg1a	SAH/MTA Deaminase
Syntrophobotulus glycolicus DSM 8271	325290416	Sg1a	SAH/MTA Deaminase
Syntrophomonas wolfei subsp. wolfei str. Goettingen	114566319	Sg1a	SAH/MTA Deaminase
Syntrophothermus lipocalidus DSM 12680	297617458	Sg1a	SAH/MTA Deaminase
Tepidanaerobacter acetatoxydans Re1	332799863	Sg1a	SAH/MTA Deaminase
Teredinibacter turnerae T7901	254785507	Sg2	MTA/5'-dAdo/Ado Deaminase
Thauera sp. MZ1T	217970136	Sg2	MTA/5'-dAdo/Ado Deaminase
Thermacetogenium phaeum DSM 12270	410667372	Sg1a	SAH/MTA Deaminase
Thermaerobacter marianensis DSM 12885	317121933	Sg1a	SAH/MTA Deaminase
Thermincola potens JR	296133555	Sg1a	SAH/MTA Deaminase
Thermoanaerobacter ethanolicus CCSD1	256751722	Sg1a	SAH/MTA Deaminase
Thermoanaerobacter ethanolicus JW 200	326389977	Sg1a	SAH/MTA Deaminase
Thermoanaerobacter italicus Ab9	289578629	Sg1a	SAH/MTA Deaminase
Thermoanaerobacter mathranii subsp. mathranii str. A3	297544852	Sg1a	SAH/MTA Deaminase
Thermoanaerobacter pseudethanolicus ATCC 33223	167037289	Sg1a	SAH/MTA Deaminase
Thermoanaerobacter tengcongensis MB4	20808022	Sg1a	SAH/MTA Deaminase
Thermoanaerobacter wiegelii Rt8.B1	345017876	Sg1a	SAH/MTA Deaminase
Thermococcus barophilus MP	315231643	Sg1b	SAH/MTA/5'-dAdo Deaminase
Thermococcus gammatolerans EJ3	240102195	Sg1b	SAH/MTA/5'-dAdo Deaminase
Thermococcus kodakarensis KOD1	57641826	Sg1b	SAH/MTA/5'-dAdo Deaminase
Thermococcus litoralis DSM 5473	375084151	Sg1b	SAH/MTA/5'-dAdo Deaminase
Thermococcus onnurineus NA1	212224804	Sg1b	SAH/MTA/5'-dAdo Deaminase
Thermococcus sibiricus MM 739	242399953	Sg1b	SAH/MTA/5'-dAdo Deaminase
Thermococcus sp. 4557	341582507	Sg1b	SAH/MTA/5'-dAdo Deaminase
Thermococcus sp. AM4	223477357	Sg1b	SAH/MTA/5'-dAdo Deaminase
Thermococcus sp. CL1	390961208	Sg1b	SAH/MTA/5'-dAdo Deaminase
Thermococcus zilligii AN1	409096367	Sg1b	SAH/MTA/5'-dAdo Deaminase
Thermoplasma acidophilum	10640374	Sg11	SAH Deaminase
Thermoplasma acidophilum DSM 1728	16082586	Sg11	SAH Deaminase
Thermoplasma volcanium GSS1	13541346	Sg11	SAH Deaminase
Thermosinus carboxydivorans Nor1	121535792	Sg1a	SAH/MTA Deaminase
Thermosipho africanus H17ap60334	407514376	Sg8	SAH/MTA Deaminase
Thermosipho africanus TCF52B	217077947	Sg8	SAH/MTA Deaminase
Thermosipho melanesiensis BI429	150021441	Sg8	SAH/MTA Deaminase
Thermotoga lettingae TMO	157363801	Sg8	SAH/MTA Deaminase
Thermotoga maritima MSB8	15643698	Sg8	SAH/MTA Deaminase
Thermotoga naphthophila RKU-10	281413203	Sg8	SAH/MTA Deaminase
Thermotoga neapolitana DSM 4359	222100614	Sg8	SAH/MTA Deaminase
Thermotoga petrophila RKU-1	148270922	Sg8	SAH/MTA Deaminase
Thermotoga sp. EMP	403252676	Sg8	SAH/MTA Deaminase
Thermotoga sp. RQ2	170289617	Sg8	SAH/MTA Deaminase
Thermotoga thermarum DSM 5069	338730352	Sg8	SAH/MTA Deaminase
Thioalkalivibrio sp. K90mix	289208680	Sg2	MTA/5'-dAdo/Ado Deaminase
Thioalkalivibrio sulfidophilus HL-EbGr7	220934703	Sg2	MTA/5'-dAdo/Ado Deaminase
Thioalkalivibrio thiocyanoxidans ARh 4	350560955	Sg2	MTA/5'-dAdo/Ado Deaminase
Thiobacillus denitrificans ATCC 25259	74316965	Sg2	MTA/5'-dAdo/Ado Deaminase

Organism	GI Number	Sg	Function
Thiocapsa marina 5811	344339504	Sg2	MTA/5'-dAdo/Ado Deaminase
Thiocystis violascens DSM 198	390951578	Sg2	MTA/5'-dAdo/Ado Deaminase
Thiorhodococcus drewsii AZ1	345869285	Sg2	MTA/5'-dAdo/Ado Deaminase
Thiorhodospira sibirica ATCC 700588	350553685	Sg2	MTA/5'-dAdo/Ado Deaminase
Thiorhodovibrio sp. 970	381157376	Sg2	MTA/5'-dAdo/Ado Deaminase
Thiothrix nivea DSM 5205	386815608	Sg2	MTA/5'-dAdo/Ado Deaminase
uncultured bacterium	407007146	Sg10	8-Oxoadenine Deaminase
uncultured bacterium	406883814	Sg1b	SAH/MTA/5'-dAdo Deaminase
uncultured bacterium	406902473	Sg2	MTA/5'-dAdo/Ado Deaminase
uncultured bacterium	406915765	Sg2	MTA/5'-dAdo/Ado Deaminase
uncultured bacterium	406941479	Sg2	MTA/5'-dAdo/Ado Deaminase
uncultured bacterium	406943346	Sg2	MTA/5'-dAdo/Ado Deaminase
uncultured bacterium	406946171	Sg2	MTA/5'-dAdo/Ado Deaminase
uncultured bacterium	406982005	Sg9	Guanine Deaminase
uncultured haloarchaeon	148508286	Sg1b	SAH/MTA/5'-dAdo Deaminase
uncultured haloarchaeon	148508263	Sg9	Guanine Deaminase
Veillonella atypica ACS-049-V-Sch6	303231426	Sg1a	SAH/MTA Deaminase
Veillonella atypica ACS-134-V-Col7a	303228970	Sg1a	SAH/MTA Deaminase
Veillonella parvula ACS-068-V-Sch12	333977040	Sg1a	SAH/MTA Deaminase
Veillonella parvula ATCC 17745	282850234	Sg1a	SAH/MTA Deaminase
Veillonella parvula DSM 2008	269798007	Sg1a	SAH/MTA Deaminase
Veillonella sp. 3_1_44	294793747	Sg1a	SAH/MTA Deaminase
Veillonella sp. 6_1_27	294791887	Sg1a	SAH/MTA Deaminase
Veillonella sp. ACP1	401679851	Sg1a	SAH/MTA Deaminase
Veillonella sp. oral taxon 158 str. F0412	313894498	Sg1a	SAH/MTA Deaminase
Veillonella sp. oral taxon 780 str. F0422	342214436	Sg1a	SAH/MTA Deaminase
Xanthomonas albilineans GPE PC73	285018688	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas axonopodis pv. citri str. 306	21243112	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas axonopodis pv. citrumelo F1	346725272	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas axonopodis pv. malvacearum str. GSPB2388	410700374	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas axonopodis pv. punicae str. LMG 859	390992333	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas campestris pv. campestris str. ATCC 33913	21231708	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas campestris pv. campestris str. B100	188991303	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas campestris pv. musacearum NCPPB 4381	289671283	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas campestris pv. raphani 756C	384428174	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas campestris pv. vasculorum NCPPB 702	289665703	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas campestris pv. vesicatoria str. 85-10	78048131	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas citri pv. mangiferaeindicae LMG 941	381171484	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas fuscans subsp. aurantifolii str. ICPB 11122	294624431	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas gardneri ATCC 19865	325921158	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas oryzae pv. oryzae KACC 10331	58582326	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas oryzae pv. oryzae MAFF 311018	84624206	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas oryzae pv. oryzae PXO99A	188576191	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas oryzae pv. oryzicola BLS256	384419078	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas perforans 91-118	325926496	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas sacchari NCPPB 4393	380512828	Sg2	MTA/5'-dAdo/Ado Deaminase
Xanthomonas vesicatoria ATCC 35937	325918739	Sg2	MTA/5'-dAdo/Ado Deaminase
Xylella fastidiosa 9a5c	15839062	Sg2	MTA/5'-dAdo/Ado Deaminase
Xylella fastidiosa Dixon	71276240	Sg2	MTA/5'-dAdo/Ado Deaminase
Xylella fastidiosa M23	182682097	Sg2	MTA/5'-dAdo/Ado Deaminase
Xylella fastidiosa Temecula1	28199370	Sg2	MTA/5'-dAdo/Ado Deaminase
Yokenella regensburgei ATCC 43003	365847690	Sg10	8-Oxoadenine Deaminase

APPENDIX B

FUNCTIONAL ANNOTATIONS OF PTERIN DEAMINASES

Organism	Gi Number	Locus Tag
Agrobacterium radiobacter K84	222086854	Arad_3529
Agrobacterium tumefaciens 5A	358004348	AT5A_20871
Agrobacterium vitis S4	222149611	Avi_3553
Agrobacterium vitis S4	222106464	Avi_5410
alpha proteobacterium BAL199	163793175	BAL199_25339
Arthrospira maxima CS-328	209525534	AmaxDRAFT_2897
Arthrospira platensis NIES-39	291567086	NIES39_C04920
Arthrospira platensis str. Paraca	284051729	AplaP_010100009690
Arthrospira sp. PCC 8005	376001956	ARTHRO_1140019
Burkholderia multivorans CGD2M	221198870	BURMUCGD2M_6445
Burkholderia xenovorans LB400	91780827	Bxe_C0802
Chlorella variabilis	307110611	CHLNCDRAFT_34176
Comamonas testosteroni ATCC 11996	371450227	CTATCC11996_23416
Crocosphaera watsonii WH 0003	357261488	CWATWH0003_4509
Crocosphaera watsonii WH 8501	67924588	CwatDRAFT_1740
Cyanobium sp. PCC 7001	254432712	CPCC7001_2605
Cyanothece sp. ATCC 51142	172038506	cce_3593
Cyanothece sp. ATCC 51472	354554150	Cy51472DRAFT_2251
Cyanothece sp. CCY0110	126659334	CY0110_06099
Cyanothece sp. PCC 7424	218437242	PCC7424_0235
Cyanothece sp. PCC 7822	307151773	Cyan7822_1898
Cyanothece sp. PCC 8801	218246775	PCC8801_1952
Cyanothece sp. PCC 8802	257059817	Cyan8802_1979
Dinoroseobacter shibae DFL 12	159043930	Dshi_1381
Erwinia billingiae Eb661	300717286	EbC_27110
Gluconacetobacter oboediens 174Bp2	349687131	Gobo1_010100008018
Gluconacetobacter sp. SXCC-1	330992835	SXCC_02737
Hoeflea phototrophica DFL-43	163760226	HPDFL43_08189
Hydrogenophaga sp. PBC	388566070	Q5W_0868
Jannaschia sp. CCS1	89053965	Jann_1474
Ketogulonicigenium vulgare Y25	310816506	EIO_2064
Klebsiella oxytoca 10-5246	376395439	HMPREF9690_02815
Klebsiella pneumoniae 342	206580486	KPK_2297
Klebsiella pneumoniae KCTC 2242	386035174	KPN2242_13140
Klebsiella pneumoniae subsp. pneumoniae	378979178	KPHS_30190

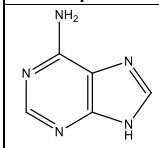
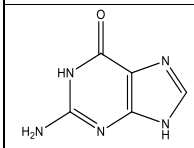
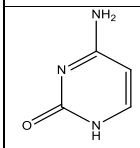
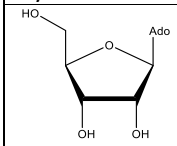
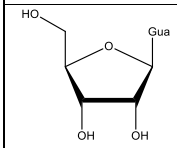
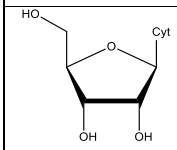
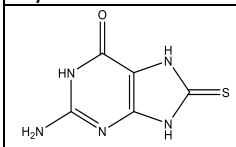
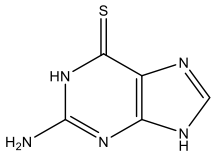
Organism	Gi Number	Locus Tag
HS11286		
Klebsiella pneumoniae subsp. pneumoniae MGH 78578	152970585	KPN_02036
Klebsiella pneumoniae subsp. pneumoniae NTUH-K2044	238895079	KP1_3111
Klebsiella sp. 1_1_55	290509179	HMPREF0485_00950
Klebsiella sp. 4_1_44FAA	365137953	HMPREF1024_00689
Klebsiella sp. MS 92-3	330015330	HMPREF9538_05868
Klebsiella variicola At-22	288935109	Kvar_2241
Labrenzia aggregata IAM 12614	118592963	SIAM614_02036
Labrenzia alexandrii DFL-11	254501071	ADFL11_1107
Lyngbya sp. PCC 8106	119485483	L8106_10086
Mesorhizobium alhagi CCNWXJ12-2	359791086	MAXJ12_16641
Mesorhizobium amorphae CCNWGS0123	357026772	MEA186_18497
Mesorhizobium australicum WSM2073	354569803	
Mesorhizobium ciceri biovar biserrulae WSM1271	319783022	Mesci_3325
Mesorhizobium loti MAFF303099	14022090	mll1290
Mesorhizobium loti MAFF303099	161621451	mll1290
Mesorhizobium opportunistum WSM2075	337268293	Mesop_3816
Methylobacterium chloromethanicum CM4	218530001	Mchl_2042
Methylobacterium extorquens AM1	240138303	MexAM1_META1p1649
Methylobacterium extorquens DM4	254560845	METDI2397
Methylobacterium extorquens DSM 13060	373564592	MetexDRAFT_4416
Methylobacterium extorquens PA1	163851150	Mext_1723
Methylobacterium populi BJ001	188580356	Mpop_1091
Microcoleus chthonoplastes PCC 7420	254410796	MC7420_274
Microcoleus vaginatus FGP-2	334120965	MicvaDRAFT_2986
Microcystis aeruginosa NIES-843	166363277	MAE_05360
Microcystis aeruginosa PCC 7806	159026135	IPF_4829
Microcystis aeruginosa PCC 7941	389764552	MICAD_400012
Microcystis aeruginosa PCC 9432	389677661	MICCA_3050004
Microcystis aeruginosa PCC 9443	389731332	MICAC_5760002
Microcystis aeruginosa PCC 9701	389881376	MICAK_3820002
Microcystis aeruginosa PCC 9717	389716015	MICAB_6620002
Microcystis aeruginosa PCC 9806	389790346	MICAE_2140020
Microcystis aeruginosa PCC 9807	389802797	MICAF_3630004
Microcystis aeruginosa PCC 9808	389824008	MICAG_3490012
Microcystis aeruginosa PCC 9809	389834001	MICAH_1010004

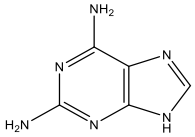
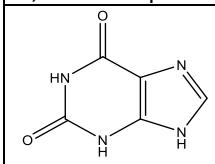
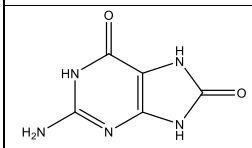
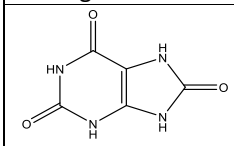
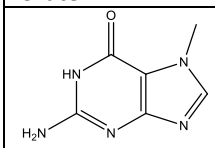
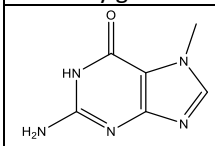
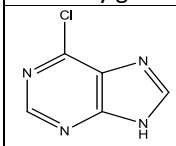
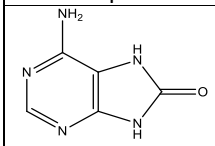
Organism	Gi Number	Locus Tag
Microcystis sp. T1-4	390441413	MICAI_4010001
Micromonas pusilla CCMP1545	303279224	MICPUCDRAFT_58167
Moorea producta 3L	332707052	LYNGBM3L_26730
Ochrobactrum anthropi ATCC 49188	153011192	Oant_3872
Octadecabacter antarcticus 307	254436998	OA307_1868
omamonas testosteroni KF-1	221070037	
Opitutaceae bacterium TAV1	390121479	ObacTA_020100014809
Opitutaceae bacterium TAV5	373851669	
Ostreococcus lucimarinus CCE9901	145349453	OSTLU_42867
Ostreococcus tauri	308806984	Ot08g02330
Pantoea sp. aB	304397339	PanABDRAFT_2477
Pantoea sp. At-9b	317046647	Pat9b_0413
Paulinella chromatophora	194476571	PCC_0086
Polymorphum gilvum SL003B-26A1	328542414	SL003B_0794
Prochlorococcus marinus str. MIT 9211	159903893	P9211_13521
Prochlorococcus marinus str. MIT 9301	126696828	P9301_14901
Prochlorococcus marinus str. MIT 9303	124023689	P9303_19891
Prochlorococcus marinus str. MIT 9312	78779784	PMT9312_1400
Prochlorococcus marinus str. MIT 9313	33862599	PMT0326
Prochlorococcus marinus str. NATL1A	124026429	NATL1_17241
Prochlorococcus marinus str. NATL2A	72382709	PMN2A_0870
Prochlorococcus marinus subsp. marinus str. CCMP1375	33240827	Pro1378
Pseudomonas sp. ADP	32455877	pADP-1_p093
Rahnella aquatilis CIP 78.65 = ATCC 33071	383189811	Rahaq2_1935
Rahnella sp. Y9602	322832607	Rahaq_1889
Rhizobium etli CFN 42	86281764	RHE_CH02043
Rhizobium etli CFN 42	162329640	RHE_CH02043
Rhizobium etli CIAT 652	190894434	RHECIAT_PC0000096
Rhizobium etli CNPAF512	327188280	RHECNPAF_9300154
Rhizobium leguminosarum bv. trifolii	1439551	
Rhizobium leguminosarum bv. trifolii WSM1325	241113027	Rleg_4666
Rhizobium leguminosarum bv. trifolii WSM2304	209549272	Rleg2_1675
Rhizobium leguminosarum bv. trifolii WSM597	392850220	Rleg9DRAFT_1550
Rhizobium leguminosarum bv. trifolii WU95	392515184	Rleg8DRAFT_0280
Rhizobium leguminosarum bv. viciae 3841	116255810	pRL110609

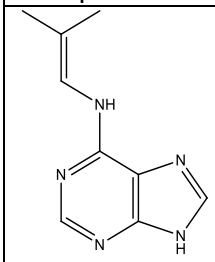
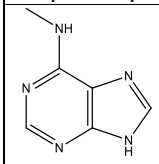
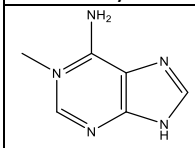
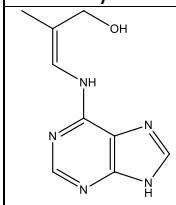
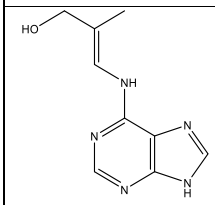
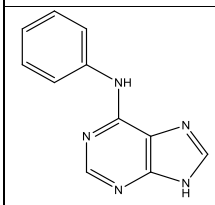
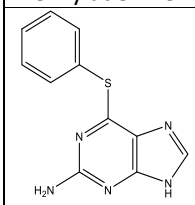
Organism	Gi Number	Locus Tag
Rhizobium leguminosarum bv. viciae 3841	116248977	pRL120308
Rhizobium sp. PDO1-076	375053111	PDO_3140
Roseobacter denitrificans OCh 114	110678617	RD1_1293
Ruegeria sp. TM1040	99078030	TM1040_3052
Serratia odorifera 4Rx13	270262662	SOD_c02830
Serratia sp. M24T3	383816172	SPM24T3_17475
Starkeya novella DSM 506	298292809	Snov_2842
Synechococcus elongatus PCC 6301	56750963	syc0954_d
Synechococcus elongatus PCC 7942	81299379	Synpcc7942_0568
Synechococcus sp. BL107	116072321	BL107_11056
Synechococcus sp. CB0101	318040418	SCB01_010100001875
Synechococcus sp. CB0205	317970568	SCB02_010100013621
Synechococcus sp. CC9311	113953404	sync_0740
Synechococcus sp. CC9605	78212398	Syncc9605_0854
Synechococcus sp. CC9902	78185105	Syncc9902_1538
Synechococcus sp. JA-2-3B'a(2-13)	86608541	CYB_1063
Synechococcus sp. JA-3-3Ab	86606230	CYA_1567
Synechococcus sp. PCC 7335	254421902	S7335_2052
Synechococcus sp. RS9916	116073349	
Synechococcus sp. RS9917	87125519	RS9917_02061
Synechococcus sp. WH 5701	87302978	WH5701_07386
Synechococcus sp. WH 5701	87303856	WH5701_16173
Synechococcus sp. WH 7803	148240087	SynWH7803_1751
Synechococcus sp. WH 7805	88809157	WH7805_05676
Synechococcus sp. WH 8016	352096259	Syn8016DRAFT_2486
Synechococcus sp. WH 8102	33866172	
Synechococcus sp. WH 8109	260436186	SH8109_1371
Synechocystis sp. PCC 6803	16330119	slr1237
Trichodesmium erythraeum IMS101	113477958	Tery_4570
Variovorax paradoxus EPS	319796039	Varpa_5413
Variovorax paradoxus EPS	319796483	Varpa_5860
Variovorax paradoxus S110	239817717	Vapar_4756
Variovorax paradoxus S110	239818099	Vapar_5141
Verminephrobacter eiseniae EF01-2	121610081	Veis_3139

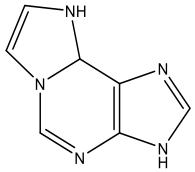
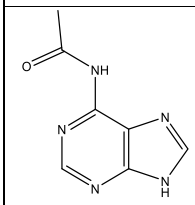
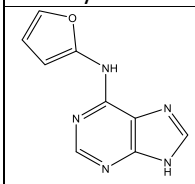
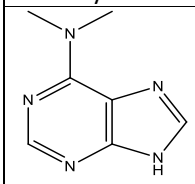
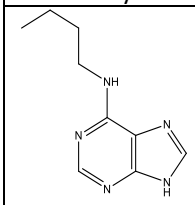
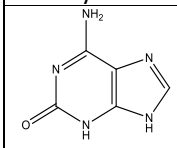
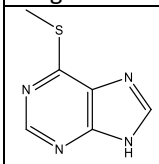
APPENDIX C

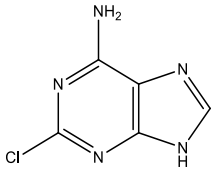
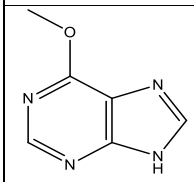
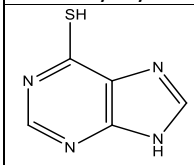
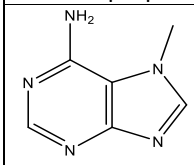
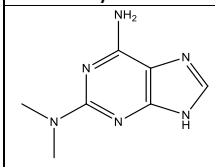
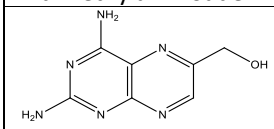
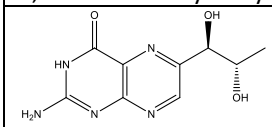
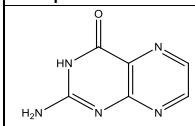
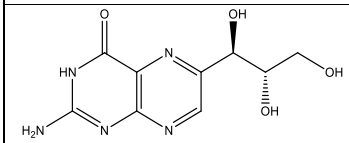
COMPOUNDS TESTED WITH YAHJ AND SSNA FOR DEAMINATION

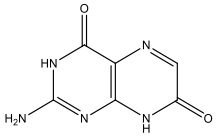
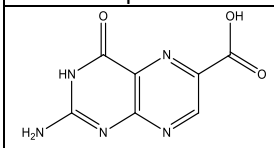
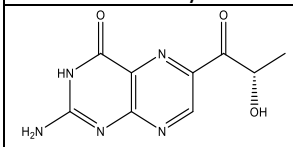
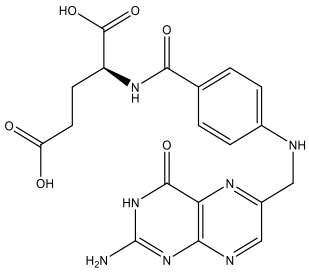
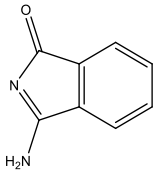
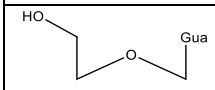
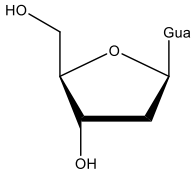
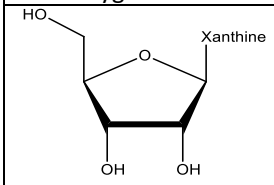
Compound Tested	SsnA	YahJ
 Adenine	No-UV	No-UV
 Guanine	No-UV	No-UV
 Cytosine	No-UV	No-UV
 Adenosine	No-UV	No-UV
 Guanosine	No-UV	No-UV
 Cytidine	Back Ground	No-UV
 8-mercaptoguanine	No-UV	No-UV
 Thioguanine	No-UV	No-UV

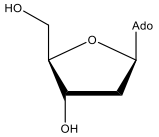
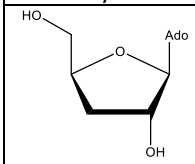
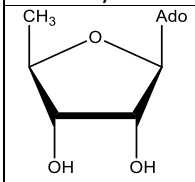
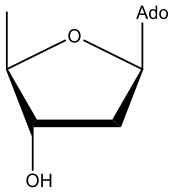
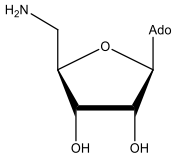
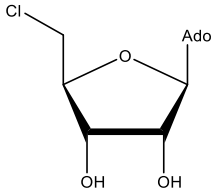
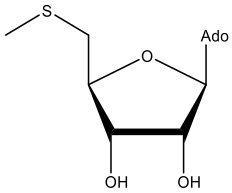
Compound Tested	SsnA	YahJ
 2,6-diaminopurine	No-UV	No-UV
 Xanthine	No-UV	No-UV
 8-oxoguanine	No-UV	No-UV
 Urate	No-UV	No-UV
 7-methylguanine	No-UV	False Positive
 9-methylguanine	No-UV	False Positive
 6-chloropurine	No-UV	No-UV
 8-oxoadenine	No-UV	No-UV

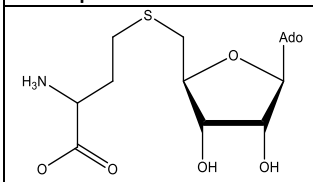
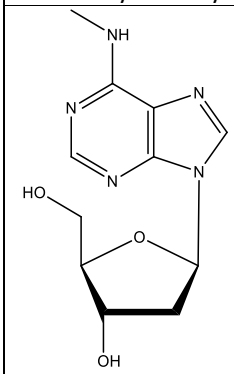
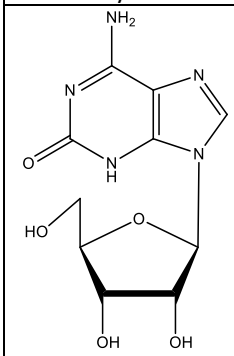
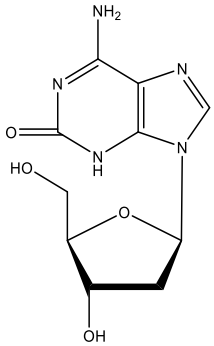
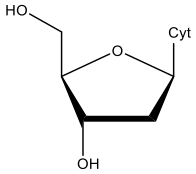
Compound Tested	SsnA	YahJ
 Isopentenyladenine	No-UV	No-UV
 N6-methyladenine	No-UV	No-UV
 1-methyladenine	No-UV	No-UV
 Cis-zeatin	No-UV	No-UV
 Trans-zeatin	No-UV	No-UV
 D2 Benzyladenine B7	No-UV	No-UV
 2-amino-6-benzylthiopurine	No-UV	No-UV

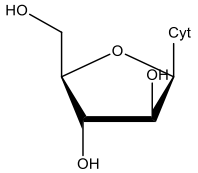
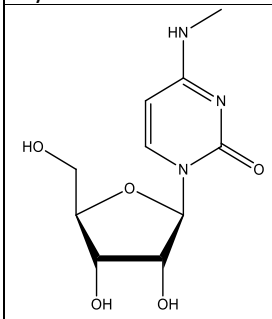
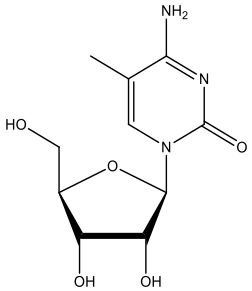
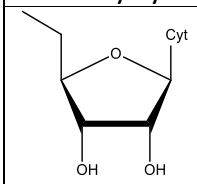
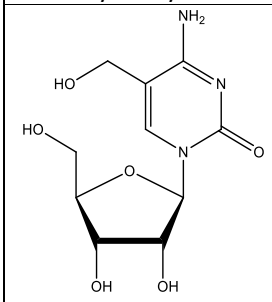
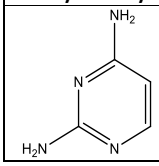
Compound Tested	SsnA	YahJ
 N-Ethenoadenine	No-UV	No-UV
 N6-acyladenine	No-UV	No-UV
 Kinetin Furfurly adenine	No-UV	No-UV
 N-dimethyladenine	No-UV	No-UV
 N-butyladenine	No-UV	No-UV
 Isoguanine	No-UV	No-UV
 6-methylthiopurine	No-UV	No-UV

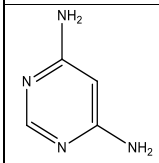
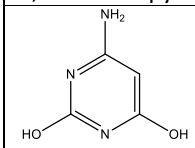
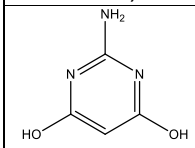
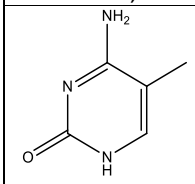
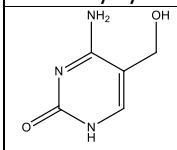
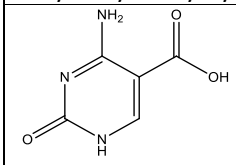
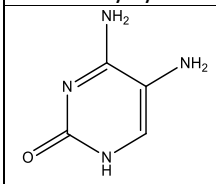
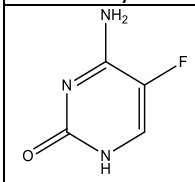
Compound Tested	SsnA	YahJ
 2-chloroadenine	No-UV	No-UV
 6-methyoxyadenine	No-UV	No-UV
 6-mercaptopurine	No-UV	No-UV
 7-methyladenine	No-UV	No-UV
 2-dimethylaminoadenin	No-UV	No-UV
 2,4-diamino-6-hydroxymethylpteridine	No-UV	No-UV
 Biopterin	No-UV	No-UV
 Pterin	No-UV	No-UV
 D-(+)-neopterin	No-UV	No-UV

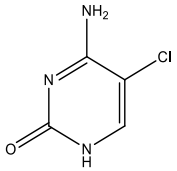
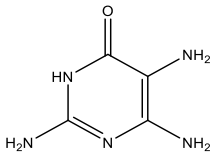
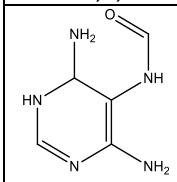
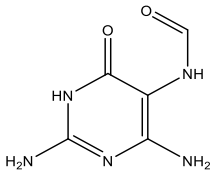
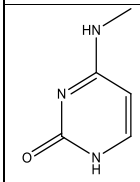
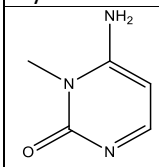
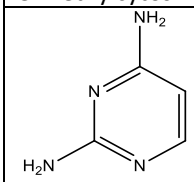
Compound Tested	SsnA	YahJ
 Isoxanthopterin	No-UV	No-UV
 Pterin-6-carboxylate	No-UV	No-UV
 Sepiapterin	No-UV	No-UV
 Folate	No-UV	No-UV
 3-iminiosoindolinone	No-UV	No-UV
 Acycloguanosine	No-UV	No-UV
 2’-deoxyguanosine	No-UV	No-UV
 Xanthosine	No-UV	No-UV

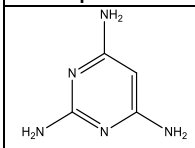
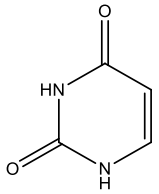
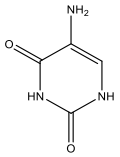
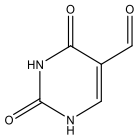
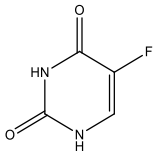
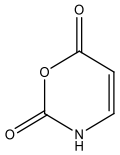
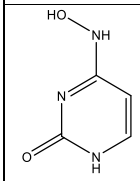
Compound Tested	SsnA	YahJ
 2’-deoxyadenosine	Back Ground	No-UV
 3’-deoxyadenosine	No-UV	No-UV
 5’-deoxyadenosine	No-UV	No-UV
 2’,5’-dideoxyadenosine	No-UV	No-UV
 5’-amino-5’-dAdo	No-UV	No-UV
 5’-chloro-5’-dAdo	No-UV	No-UV
 5’-deoxy-5’-methylthioaAdo	No-UV	No-UV

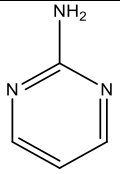
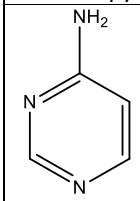
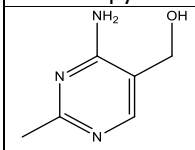
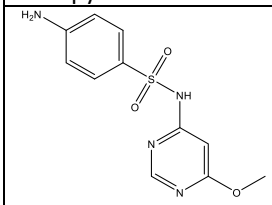
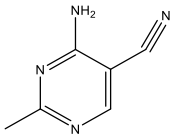
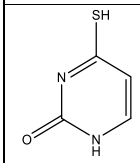
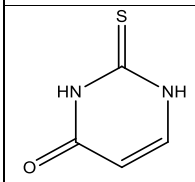
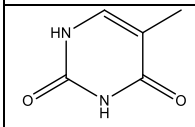
Compound Tested	SsnA	YahJ
 S-adenosylhomocysteine	No-UV	No-UV
 N6-methyl-2dAdo	No-UV	No-UV
 Isoguanosine	No-UV	No-UV
 2’-deoxyisoguanosine	No-UV	No-UV
 2’-deoxycytidine	Background	Background

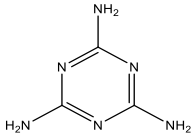
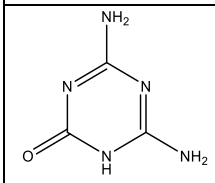
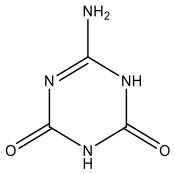
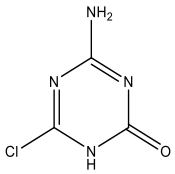
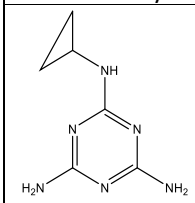
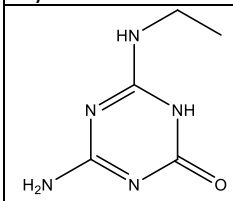
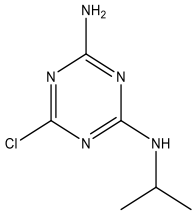
Compound Tested	SsnA	YahJ
 Cytosine-B-D-arabinofuranose	Background	No-UV
 n-methylcytidine	No-UV	No-UV
 5-methylcytidine	Back Ground	No-UV
 5-methyl-2dCyt	Back Ground	No-UV
 5-hydroxymethylcytidine	Back Ground	No-UV
	No-UV	No-UV

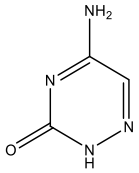
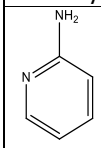
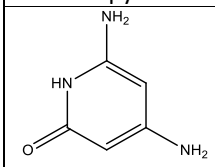
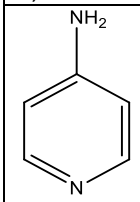
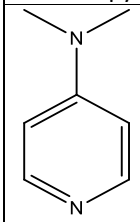
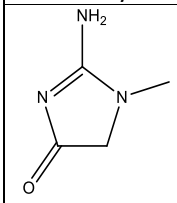
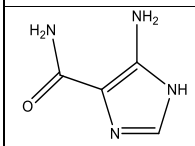
Compound Tested	SsnA	YahJ
2,4-diaminopyrimidine		
 4,6-diaminopyrimidine	No-UV	No-UV
 4-amino-2,6-dihydroxypyrimidine	No-UV	No-UV
 2-amino-4,6-dihydroxypyrimidine	No-UV	No-UV
 5-methylcytosine	No-UV	No-UV
 5-hydroxymethyl cytosine	No-UV	No-UV
 5-carboxy cytosine	No-UV	No-UV
 5-aminocytosine	No-UV	No-UV
 5-fluorocytosine	No-UV	No-UV

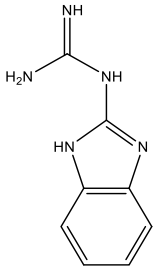
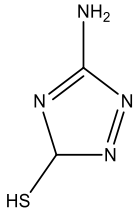
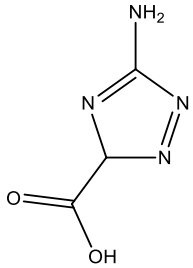
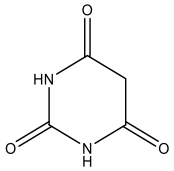
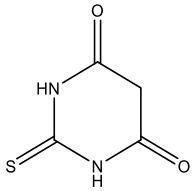
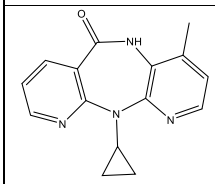
Compound Tested	SsnA	YahJ
 5-chlorocytosine	No-UV	No-UV
 4-OH-2,5,6-triaminopyrimidine	No-UV	No-UV
 FAPY Adenine	No-UV	No-UV
 FAPY Guanine	No-UV	No-UV
 n-methylcytosine Synthesized	No-UV	No-UV
 3-methylcytosine		
 2,6-diaminopyrimidine	No-UV	No-UV

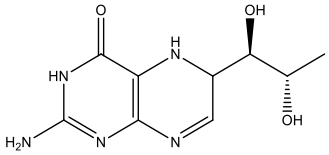
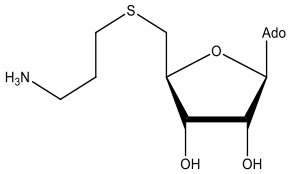
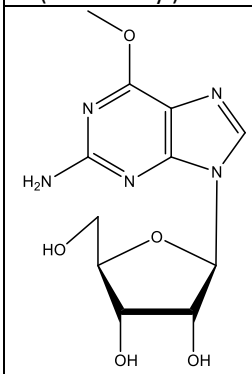
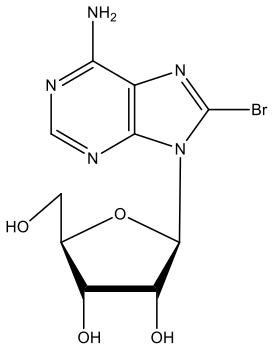
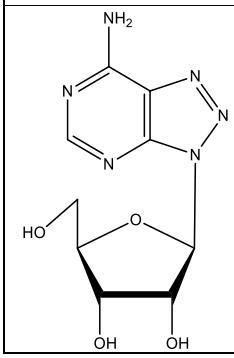
Compound Tested	SsnA	YahJ
 2,4,6-triaminopyrimidine	No-UV	No-UV
 Uracil	No-UV	No-UV
 5-aminouracil	No-UV	No-UV
 5-formyluracil	No-UV	No-UV
 5-fluorouracil A8	No-UV	No-UV
 3-oxauracil B8	No-UV	No-UV
 6-hydroxyaminouracil E4	No-UV	No-UV
Thiamin D8	No-UV	No-UV
Thiamin Pyrophosphate 20	No-UV	No-UV

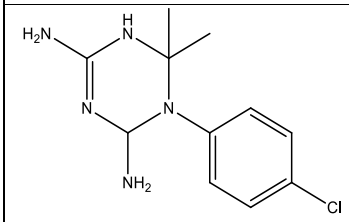
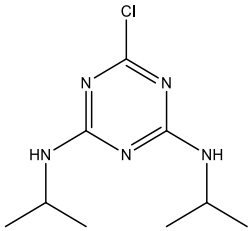
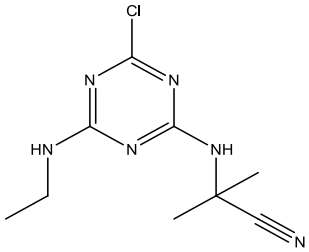
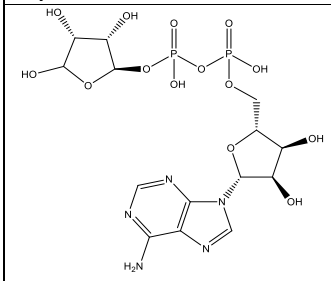
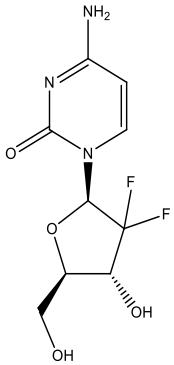
Compound Tested	SsnA	YahJ
 2-aminopyrimidine E8	No-UV	No-UV
 4-aminopyrimidine	No-UV	No-UV
 Toxopyrimidine	No-UV	No-UV
 Sulfamonomethoxine	No-UV	No-UV
 Thiamin Fragment	No-UV	No-UV
 4-thiouracil	No-UV	No-UV
 2-thiouracil	No-UV	No-UV
 Thymine	No-UV	No-UV

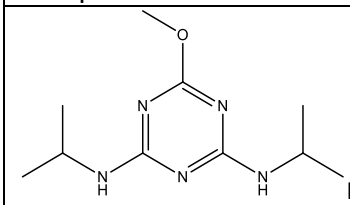
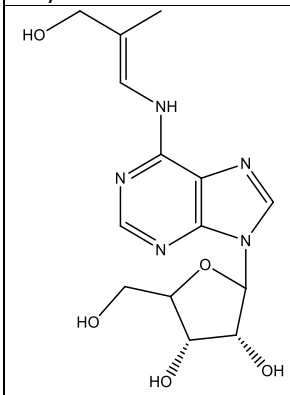
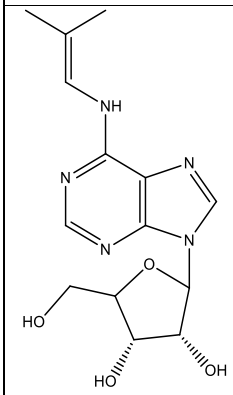
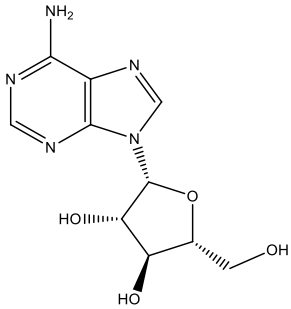
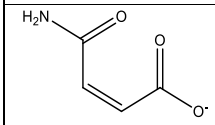
Compound Tested	SsnA	YahJ
 Melamine	No-UV	
 Ammeline	No-UV	
 Ammelide	No-UV	
 2-amino-4-hydroxy-6-chlorotriazine	No-UV	No-UV
 Cyromazin	No-UV	No-UV
 n-ethylammeline	No-UV	No-UV
 Desethylatrazine	No-UV	No-UV

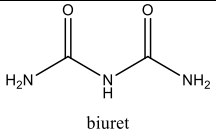
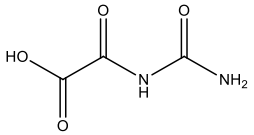
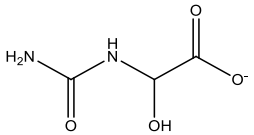
Compound Tested	SsnA	YahJ
 6-azacytosine	No-UV	No-UV
 2-aminopyridine	No-UV	No-UV
 4,6-diamino-2-hydroxypyridine	No-UV	No-UV
 4-aminopyridine	No-UV	No-UV
 4-dimethylaminopyridine	No-UV	No-UV
 Creatinine F9	No-UV	No-UV
 5-amino-4-imidazolecarboxyamide	No-UV	No-UV

Compound Tested	SsnA	YahJ
 2-guanidinobenzimidazole	No-UV	No-UV
 3-amino-5-mercapto-1,2,4-triazole	No-UV	No-UV
 3-amino-5-carboxy-1,2,4-triazole	No-UV	No-UV
 Barbiturate	No-UV	No-UV
 2-thiobarbiturate	No-UV	No-UV
 Nevirapine	No-UV	No-UV

Compound Tested	SsnA	YahJ
 7,8-dihydro-L-biopterin	No-UV	
 S-(5’-adenosyl)-3-thiopropylamine	No-UV	
 O6-methyl-2’-deoxyguanosine	No-UV	
Cyclic adenosine diphosphate-ribose	No-UV	
 8-bromoadenosine	No-UV	
	No-UV	

Compound Tested	SsnA	YahJ
8-azaadenosine		
 Cycloguanilhydrochloride	No-UV	
 Propazine	No-UV	
 Cyanazine	No-UV	
 Adenosine 5'-diphosphoribose sodium salt	No-UV	
 Gemcitabine	No-UV	

Compound Tested	SsnA	YahJ
 Prometon	No-UV	
Acyclovir	No-UV	
 Zeatin Riboside	No-UV	
 Isopentenyl Adenosine	No-UV	
 Adenosine beta-D arabinofuraose	No-UV	
 Maleamate		No-GLDH

Compound Tested	SsnA	YahJ
 Biuret		No-GLDH
 Oxalureate		No-GLDH
 Ureidoglycolate		No-GLDH