MECHANISTIC INSIGHTS INTO THE DIVERGED ENZYMES OF THE AMIDOHYDROLASE SUPERFAMILY

A Dissertation

by

TINH T. NGUYEN

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

DOCTOR OF PHILOSOPHY

December 2009

Major Subject: Chemistry

MECHANISTIC INSIGHTS INTO THE DIVERGED ENZYMES OF THE AMIDOHYDROLASE SUPERFAMILY

A Dissertation

by

TINH T. NGUYEN

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

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Frank M. Raushel
Committee Members,	Paul A. Lindahl
	Gregory D. Reinhart
	Coran M H. Watanabe
Head of Department,	David H. Russell

December 2009

Major Subject: Chemistry

ABSTRACT

Mechanistic Insights into the Diverged Enzymes of the Amidohydrolase Superfamily.

(December 2009)

Tinh T. Nguyen, B.S., Mary Washington College

Chair of Advisory Committee: Dr. Frank M. Raushel

The amidohydrolase superfamily is a functionally diverse set of enzymes that catalyzes predominantly hydrolysis reactions involving sugars, nucleic acids, amino acids, and organophosphate esters. A more divergent member of this superfamily, URI (uronate isomerase) from *Escherichia coli*, catalyzes the isomerization of D-glucuronate to D-fructuronate and D-galacturonate to D-tagaturonate. In *Bacillus halodurans*, two distinct operons were identified for the metabolism of D-glucuronate and D-galacturonate based on kinetics and genomic context. The canonical uronate isomerase is encoded by the gene Bh0705. A second URI in this organism, Bh0493, is the outlier of the group in terms of sequence similarity. Kinetic evidences indicate that Bh0705 is relatively specific for the isomerization of D-glucuronate, while Bh0493 is specific for the D-galacturonate pathway.

Bell-shaped pH-rate profiles were observed for the wild type URI from *Escherichia coli*. Primary isotope effects with [2-^{2}H]-D-glucuronate and solvent viscosity studies are consistent with product release as the rate limiting step. X-ray structure of Bh0493 was determined in the presence of D-glucuronate. A chemical

mechanism is proposed that utilizes a proton transfer from C-2 of D-glucuronate to C-1 that is initiated by the combined actions of Asp-355 and the C-5 hydroxyl of the substrate that is bound to the metal ion. The formation of the *cis*-enediol intermediate is further facilitated by the shuttling of the proton between the C-2 and C-1 oxygens by the conserved Tyr-50 and/or Arg-357.

Another divergent member of the AHS is the enzyme renal dipeptidase. Structural studies of the enzyme from *Streptomyces coelicolor* (Sco3058) demonstrate that the active site consists of a binuclear metal center. Bell-shaped pH-rate profiles are observed for both Zn^{2+} and Cd^{2+} enzymes. A chemical mechanism for renal dipeptidase is proposed based on structural analysis of the enzyme-inhibitor complex. The reaction is initiated by the polarization of the amide bond by the β-metal. Asp-320 activates the bridging hydroxide for nucleophilic attack at the peptide carbon center, forming a tetrahedral intermediate that is stabilized by the metal center and His-150. The protonated Asp-320 donates the proton to the α-amino group of the leaving group, causing the collapse of the tetrahedral intermediate and cleavage of the carbon-nitrogen bond.

DEDICATION

To the memory of Suhyung Park

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Frank Raushel, for his guidance and support throughout the course of this research. I also would like to thank my committee members: Dr. Paul Lindahl, Dr. Gregory Reinhart, and Dr. Coran Watanabe. I would like to extend my appreciation to Dr. David Barondeau for his generous time and effort in helping me with the crystallography of Sco3058.

Thanks to my family for their love and faith in me. Thanks to my best friends Thuy and Tuyet, who hold the key to my sanity. I also want to extend my gratitude to the past and present members of the Raushel group for their support and companionship. Thanks to the members of the Barondeau group for all of their help.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

The increasing number of the available genome sequences, together with advances in the computational methods that reveals the relationships between evolutionarily related genes, have generate a greater understanding into the evolution of proteins, their functions, and their associated biological processes (*1*). However, one of the difficult and ongoing tasks is the identification and assignment of molecular activity to functionally unknown enzymes encoded within completely sequenced genomes. Many sequenced genes in both prokaryotes and eukaryotes have functions that are uncertain, unknown, or incorrectly annotated (*2*-*6*). The task of protein function discovery is facilitated by the emerging number of high-resolution protein structures. A complete understanding of the interplay of structure and function is essential for decoding substrate specificity for these proteins of unknown function and for the development of catalysts for new reactions using the active sites of preexisting templates of a unique protein fold (*7*).

The practice of incorporating genomic context (sequence families, structures, and functions) into the study of enzyme reaction and mechanism is referred to as genomic enzymology (*8*). One of the driving forces for the development of genomic enzymology

---

This dissertation follows the style of *Biochemistry*.

was the recognition that enzymes that are evolutionary diverged can catalyze different overall reactions. Enzymes that are evolutionarily related (homologous enzymes) are described in several ways. A family refers to a group of homologous enzymes that catalyze the same reaction in terms of mechanism and substrate specificity (*8*). A superfamily is a group of enzymes that catalyze either the same chemical reaction with differing substrate specificities or different overall reactions that share a common mechanistic feature such as a partial reaction, intermediate, or transition state. This mechanistic attribute is facilitated by conserved active site residues that play similar roles in catalysis for all members of the superfamily (*8*). A suprafamily refers to a group of homologous enzymes that utilize different mechanistic elements to catalyze different overall reactions. Classification of enzymes and consequently functional annotations for members of mechanistically diverse superfamilies and functionally distinct suprafamilies are not possible from just sequence data, rather it entails the integration of functional and structural characterization (*8*). The number of structural types is much less than the set of reactions that needs to be deciphered due to the limited number of protein domain folds (*9-10*).

The number of unique protein folds discovered to date is estimated to be far less than the number of proteins encoded by the human genome (*11-13*). Consequently, protein folds must be reused by divergent evolution or independently formulated by convergent evolution to accommodate the number of functions represented by proteins found in living organisms. Based on the available evidence, it is possible to picture three distinct strategies that nature uses for the divergent evolution of enzyme function. Each

of these strategies involves the initial copying of the gene encoding the protein to be evolved so that the original enzyme activity is preserved by the organism (*8*).

The first divergent evolution strategy involves enzymes that catalyze reactions in a biosynthetic pathway (*14-15*). When a substrate for an enzyme is depleted, a new enzyme is evolved to supply the substrate from an available precursor using the template from an enzyme that uses the substrate. Because the original and the newly evolved enzyme must share the ability to bind the same substrate/product, this evolution strategy is restricted to retain binding specificity but not reaction mechanisms. Thus the old and new enzymes would be members of functionally distinct suprafamilies. In contrast to this approach is the hypothesis that the protein for divergent evolution come from a pool of enzymes whose mechanisms provide the needed partial reaction or strategy for stabilization of reaction intermediates or transition states (*16-17*). Evolution alters certain mechanistic characteristics and/or substrate specificity so that the new reaction is catalyzed with enhanced proficiency. Perhaps the selected enzyme to be evolved is promiscuous with regard to the desired new reaction even at a very low rate, and subsequent evolution then provides enhanced proficiency as the new metabolic pathway is optimized (*18-20*). This strategy of evolution creates enzymes that are members of mechanistically diverse superfamilies. The third scheme of divergent evolution suggests that the active site architecture is dominant instead of substrate specificity or chemical mechanism. For this approach, an active site is capable of utilizing shared functional groups in different mechanistic and metabolic contexts to catalyze an alternate reaction.

Thus, the old and the newly evolved enzymes are members of functionally distinct suprafamilies (*8*).

The most frequently found protein fold in nature is the $(\beta/\alpha)_8$-barrel. According to SCOP databases, there are currently 33 identified $(\beta/\alpha)_8$-barrel superfamilies including the amidohydrolase superfamily (*8*,*21*), the enolase superfamily (*8*,*22-24*), the thiyl radical superfamily (*8*,*25-26*), and the crotonase superfamily (*8,27-28*). The first crystal structure of an enzyme that posesses this structural fold was triose phosphate isomerase (TIM) from chicken muscle (*29*); hence the $(\beta/\alpha)_8$-barrel is also called the TIM-barrel fold. The canonical TIM-barrel consists of eight β-strands forming the beta barrel that is surrounded by eight α-helices. The active site of the TIM-barrel enzymes generally formed at the C-termini of the β-strands and the flexible loops followed the β-strands. The $(\beta/\alpha)_8$ motif of the triose phosphate isomerase structure is illustrated in **Figure 1.1**. The presence of this $(\beta/\alpha)_8$ structural fold in so many enzymes demonstrates its versatility in the divergence of enzyme function. Recent studies have shown that the $(\beta/\alpha)_8$-barrel enzyme fold originates from the duplication and fusion of ancestral $(\beta/\alpha)_4$-half-barrel, and artificial $(\beta/\alpha)_8$-barrel protein has been constructed experimentally by fusing two copies of the C-terminal half-barrel model protein (*30*).

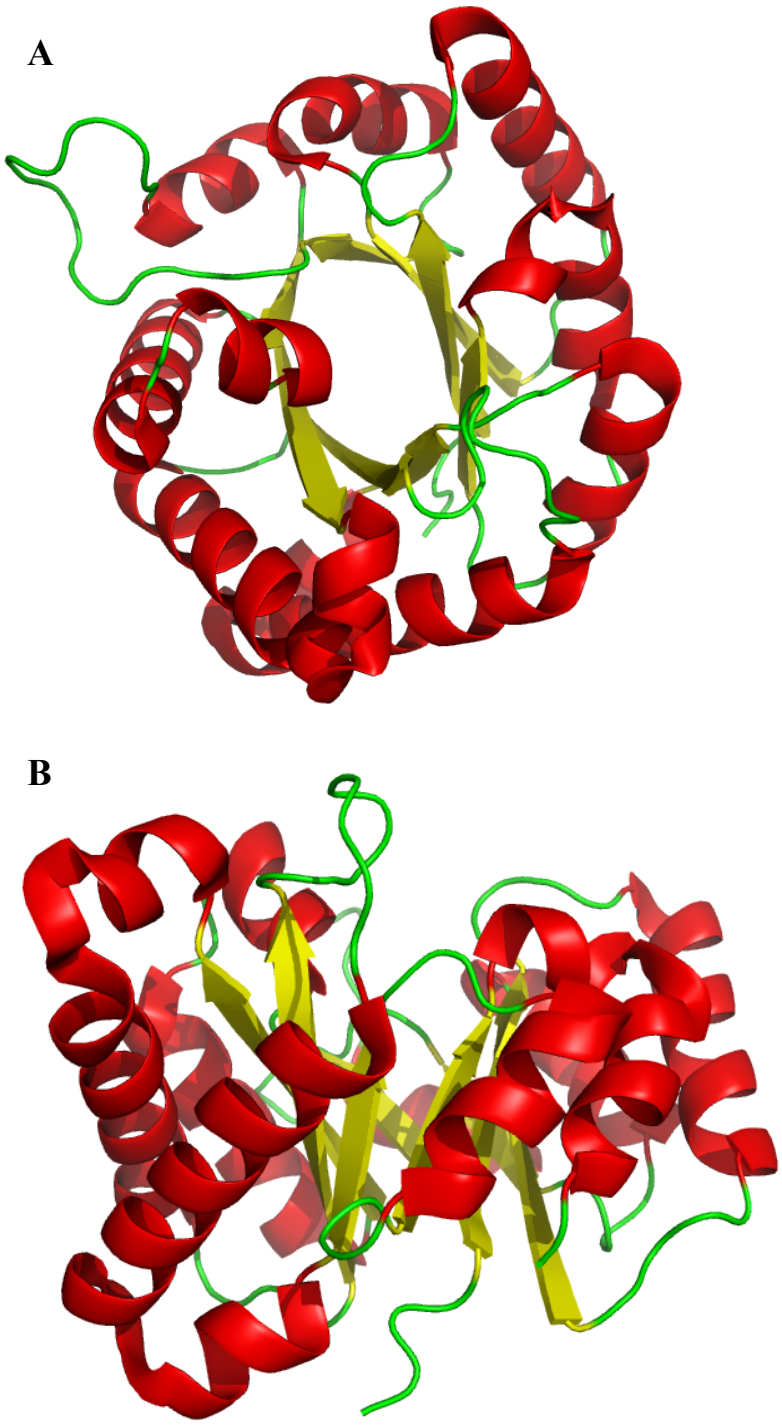


Figure 1.1: Crystal structure of triose phosphate isomerase from chicken muscle. **A)** Top view of TIM. **B)** Side view of TIM. (PDB: 1TIM).

One of the well-documented examples for the divergent evolution of protein architecture and function that can originate from a common ancestral precursor is the amidohydrolase superfamily (AHS) of enzymes. The AHS superfamily was first identified by Holm and Sander in 1997 based on the three-dimentional structural similarities of the protein folds and active sites of urease (URE), phosphotriesterase (PTE), and adenosine deaminase (ADA) (*21*). These enzymes, as mentioned earlier, all contain a $(\beta/\alpha)_8$ (TIM)-barrel structural domain. The most prominent structural landmark for this superfamily of enzymes is a mononuclear or binuclear metal center active site (*21,31*). The metals in a binuclear center are designated as α and β metal site. For the mononuclear enzymes, the metal is located at either the α or the β site. The divalent metal ions that have been found with these enzymes are zinc, cobalt, manganese, iron, and nickel. The metal center in the enzymes of the amidohydrolase superfamily has been shown to be essential for the expression of the overall catalytic activity (*31*).

X-ray crystal structures of several families of the amidohydrolase enzymes can be obtained from the Protein Data Bank (PDB). All of the amidohydrolase structures of the wild type and mutant enzymes (as of June 12, 2009) are shown in **Table 1.1**. From the 45 wild type structures, 29 have known catalytic function. These are urease (*32*) (URE, PDB code 2ubp), phosphotriesterase (*33*) (PTE, PDB code 1hzy), dihydroorotase (*34*) (DHO, PDB code 1j79), β-aspartyl dipeptidase (*35*) (IAD, PDB code 1onw), D and L-hydantoinases (*36-37*) (D-HYD, PDB code 1gkp and L-HYD, PDB code 1gkr), dihydropyrimidinase (*38*) (DHP, PDB code 1k1d), allantoinase (*39*) (ATase, PDB code

2e74), renal dipeptidase (*40*) (RDP, PDB code litq), *N*-acetyl glucosamine-6-phosphate deacetylase (*41-42*) (AGD, PDB code 1o12), uronate isomerase (*43-44*) (URI, PDB code 1j5s), D-amino acid deacetylase (*45*) (DAA, PDB code 1m7j), adenosine deaminase (*46*) (ADA, PDB code 1a4m), cytosine deaminase (*47*) (CDA, PDB code 1k6w), guanine deaminase (GDA, PDB code 2uz9), adenosine 5'-monophosphate deaminase (*48*) (AMPD, PDB code 2ael), 4-oxalomesaconate hydratase (PcmD, PDB code 2gwg), 2 pyrone 4,6-dicarboxylic acid hydratase (PcmC, PDB code 2qah), 2,6-dihydroxybenzoate decarboxylase (Rdc, PDB code 2dvt), amidozolonepropionase (*49*) (HutI, PDB code 2bb0), α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (*50*) (ACMSD, PDB code 2hbv), γ- and δ-lactonase (*4*) (Dr0930, PDB code 3fdk), phosphotriesterase-like-lactonase (*Sso*Pox, PDB code 2vc5) (*51*), two carboxypeptidases (*5-6*) (Cc0300 and Cc3125, PDB codes 2qs8 and 3be7, respectively), enamidase (*52*) (ENA, PDB code 2vun), N-isopropylammedile isopropylaminohydrolase (AtzC, PDB code 2qt3), a S-adenosylhomocysteine deaminase (7) (Tm0936, PDB code 1j6p), and histidinol-phosphate phosphatase from *T. thermophilus* (*53*) (HolPase, PDB code 2z4g). Sixteen of the reported structures remain with unknown biological function, these are the PTE homology protein (*54*) (PHP, PDB code 1bf6), three putative DNases from *Escherichia coli* (TatD, YjjV, and YcfH with PDB codes 1xwy, 1zzm, and 1yix, respectively), two of the three TatD-related protein from *Saccharomyces cerevisiae and Deinococcus radiodurans* (PDB codes 3e2v and 3gg7, respectively), and several uncharacterized amidohydrolase structures (PDB codes 3cjp, 2imr, 2i5g, 3h4u, 2paj, 2ics, 2p9b, 2r8c, and 3feg), solved by the efforts of the New York Structural Genomics Research

Consortium, and a TatD-related protein from *T. maritima* (PDB code 1j6o), solved by the efforts of the Joint Center for Structural Genomics.

Table 1.1: All of the structures belonging to the AHS that have been deposited in the Protein Data Bank (as of June 2009).

Deposition	PDB #	Enzyme	Organism	Active Site Ligands
9/10/08	3EGJ	NagA	*V. cholerae*	1 Ni (β)
4/21/09	3H4U	amidohydrolase	*Unidentified*	1 Zn (α)
11/30/08	3FEQ	eah89906	*Unidentified*	2 Zn
11/25/08	3FDK	unreleased hydrolase	*Unidentified*	Unknown
7/17/08	3DUG	arginine Carboxy-peptidase	*Unidentified*	2 Zn, Arginine
11/16/07	3BE7	arginine carboxy-peptidase	*Unidentified*	Arginine
9/10/07	2R8C	EAJ56179	*Unidentified*	2 Zn
5/21/07	2Q09	imidazolone-propionase	*Unidentified*	1 Fe, 3-(2,5-Dioxo-imidazolidin-4-yl)-propionic acid inhibitor
3/27/07	2PAJ	putative CDA / GuaD	*Unidentified*	1 Zn (α)
1/25/07	2OOF	imidazolone-propionase	*Unidentified*	1 Fe
8/20/01	1GKQ	D-hydantoinase	*Thermus sp.*	2 Zn
8/20/01	1GKP	D-hydantoinase	*Thermus sp.*	2 Zn
6/17/07	2Z4G	histidinol phosphate phosphatase	*T. thermophilus*	2 Fe, 1Zn
5/6/07	2Z00	DHO	*T. thermophilus*	2 Zn
5/3/07	2YZ5	histidinol phosphate phosphatase	*T. thermophilus*	2 Fe, 1Zn, Phosphate

Table 1.1: Continued.

4/26/07	2YXO	histidinol phosphate phosphatase	*T. thermophilus*	2 Fe, 1Zn, Sulfate
4/20/07	2PLM	SAH/TMA deaminase	*T. maritima*	1 Zn , inosylhomocysteine
8/11/07	2ANU	predicted phosphoesterase TM0559	*T. maritima*	4 Zn, 2Cl
4/12/03	1P1M	SAH/TMA deaminase	*T. maritima*	1 Ni, Methionine
10/15/02	1O12	NagA	*T. maritima*	1 Fe
7/9/02	1J6P	SAH/TMA deaminase	*T. maritima*	1 Ni
7/9/02	1J6O	TatD-related deoxy-ribonuclease	*T. maritima*	None
7/2/02	1J5S	uronate isomerase	*T. maritima*	None
9/18/07	2VC7	SsoPox lactonase, PTE	*S. solfataricus*	1 Fe, 1Co, (4S)-4-(decanoylamino)-5-hydroxy-3,4-dihydro-2H-thiophenium
9/18/07	2VC5	SsoPox lactonase, PTE	*S. solfataricus*	1 Fe, 1Co
6/15/07	2QAH	2-pyrone-4,6-dicarboxylic acid hydrolase	*S. paucimobilis*	None
6/3/08	3DC8	dihydro-pyrimidinase	*S. meliloti*	2 Zn
1/31/06	2FVM	dihydro-pyriminidase	*S. kluyveri*	2 Zn, N-carbamoyl-β-alanine
1/31/06	2FVK	dihydro-pyriminidase	*S. kluyveri*	2 Zn, Dihydrouracil
1/25/06	2FTY	dihydro-pyriminidase	*S. kluyveri*	2 Zn
8/6/08	3E2V	TatD like	*S. cerevisiae*	1 Mg
3/25/09	3GRI	DHO	*S. Aureus*	1 Zn (α)

Table 1.1: Continued.

5/12/06	2GZX	TatD deoxy-ribonuclease MW0446	*S. aureus*	2 Ni
8/1/06	2DVX	2,6-dihydroxy-benzoate decarboxylase	*Rhizobium sp.*	1 Zn, 2,3-dihydroxybenzaldehyde
8/1/06	2DVU	2,6-dihydroxy-benzoate decarboxylase	*Rhizobium sp.*	1 Zn, 2,6-dihydroxybenzoate
8/1/06	2DVT	2,6-dihydroxy-benzoate decarboxylase	*Rhizobium sp.*	1 Zn
11/25/08	3FDG	dipeptidase AC	*R. sphaeroides*	2 Mg
5/4/06	2GWG	4-oxalomesaconate hydratase (LigJ)	*R. palustris*	1 Zn
8/1/07	2QT3	N-isopropyl-ammelide isopropyl-amidohydrolase AtzC	*Pseudomonas sp adp*	1 Zn (α)
8/10/05	2AMX	ADA	*P. yeolii*	1 Co
8/8/07	2QVN	ADA	*P. vivax*	Guanosine-5'-monophosphate
4/10/07	2PGR	ADA	*P. vivax*	1 Zn, 2'-Deoxycoformycin
4/9/07	2PGF	ADA	*P. vivax*	1 Zn, Adenosine
12/19/05	2FFI	putative 2-pyrone-4,6-dicarboxylic acid hydrolase	*P. putida*	Phosphate
5/4/06	2GWN	DHO	*P. gingivalis*	2 Zn, β-mercaptoethnol, chloride, sulfate, glycerol, cacodylate ion
10/8/08	3ETK	amidohydro_3	*P. furiosus*	2 Mg

Table 1.1: Continued.

6/14/06	2HBX	α-amino-β-carboxymuconate-ε-semialdehyde-decarboxylase (ACMSD)	*P. flourescens*	1 Co
6/14/06	2HBV	α-amino-β-carboxymuconate-ε-semialdehyde-decarboxylase (ACMSD)	*P. flourescens*	1 Mg, 1 Zn
1/31/07	2OQL	PTE H254Q/H257F	*P. dimunita*	2 Zn, 2-[Bis-(2-hydroxy-ethyl)-amino]-2-hydroxymethyl-propane-1,3-diol
12/4/06	2O4Q	PTE G60A	*P. dimunita*	2 Zn, Dimethylarsenate
12/4/06	2O4M	PTE I106G/F132G/H257Y	*P. dimunita*	2 Zn, Dimethylarsenate, Acetate, Glycerol
5/9/00	1EZ2	PTE	*P. dimunita*	2 Zn, diisopropylmethyl phosphonate
5/9/00	1EYW	PTE	*P. dimunita*	2 Zn, triethylphosphate
10/23/07	3B40	putative peptidase K44Q, I62V, P64L, E69T, F73V, N79G, N195S, S203A, T215S, A216P, N314D	*P. aeruginosa*	None
8/24/06	2I5G	dipeptidase like T56A, K300E	*P. aeruginosa*	None
11/11/01	1KCX	dihydro-pyrimidinase related	*M. musculus*	None
1/31/98	1A4M	ADA	*M. musculus*	1 Zn, 6-hydroxy-1,6-dihydro purine nucleoside
1/31/98	1A4L	ADA	*M. musculus*	1 Zn, 2'-Deoxycoformycin

Table 1.1: Continued.

8/30/96	1UIP	ADA H238E	*M. musculus*	1 Zn, Purine riboside
8/30/96	1UIO	ADA H238A	*M. musculus*	1 Zn, 6-hydroxy-7,8-dihydro purine adenosine/ H238A
2/29/96	1FKX	ADA D296A	*M. musculus*	1 Zn, 6-hydroxy-1,6-dihydro purine adenosine
2/29/96	1FKW	ADA D295E	*M. musculus*	1 Zn, Purine riboside
12/2/94	2ADA	ADA	*M. musculus*	1 Zn, 6-hydroxy-7,8-dihydro purine adenosine
12/22/92	1ADD	ADA	*M. musculus*	1 Zn, 1-deaza-adenosine
11/29/05	2F6K	amidohydrolase II target Lp24	*L. plantarum*	1 Mn (α)
7/25/07	2QPX	putative metal dependent hydrolase	*L. casei atcc 334*	2 Zn
3/4/00	1EJW	urease at 298K	*K. aerogenes*	2 Ni
3/4/00	1EJX	urease at 100K	*K. aerogenes*	2 Ni
3/4/00	1EJV	urease H320Q	*K. aerogenes*	2 Ni
3/4/00	1EJU	urease H320N	*K. aerogenes*	2 Ni
3/4/00	1EJT	urease H219Q	*K. aerogenes*	2 Ni
3/4/00	1EJS	urease H219N	*K. aerogenes*	2 Ni
3/4/00	1EJR	urease D221A	*K. aerogenes*	2 Ni
2/5/00	1EF2	urease	*K. aerogenes*	2 Mn
2/17/98	1A5L	urease K217C	*K. aerogenes*	None
2/17/98	1A5O	urease K217C	*K. aerogenes*	2 Ni, Formate
2/17/98	1A5N	urease K217A	*K. aerogenes*	2 Ni, Formate
2/17/98	1A5M	urease K217A	*K. aerogenes*	None
2/17/98	1A5K	urease K217E	*K. aerogenes*	None
4/23/97	1FWJ	urease	*K. aerogenes*	2 Ni
4/23/97	1FWI	urease H134A	*K. aerogenes*	1 Ni (β)
4/23/97	1FWH	urease C319Y	*K. aerogenes*	2 Ni
4/23/97	1FWG	urease C319S	*K. aerogenes*	2 Ni
4/23/97	1FWF	urease C319D	*K. aerogenes*	2 Ni
4/23/97	1FWE	urease C319A	*K. aerogenes*	2 Ni, Acetohydroxamate

Table 1.1: Continued.

4/23/97	1FWD	urease C319A pH 9.4	*K. aerogenes*	2 Ni
4/23/97	1FWC	urease C319A pH 8.5	*K. aerogenes*	2 Ni
4/23/97	1FWB	urease C319A pH 6.5	*K. aerogenes*	2 Ni
4/23/97	1FWA	urease C319A pH 7.5	*K. aerogenes*	2 Ni, Carbonate
6/20/95	1KRC	urease H320A	*K. aerogenes*	2 Ni, Carbon Dioxide
6/20/95	1KRB	urease H219A	*K. aerogenes*	2 Ni
6/20/95	1KRA	urease	*K. aerogenes*	None
2/16/95	2KAU	urease	*K. aerogenes*	2 Ni
1/3/05	2BGN	dipeptidyl dipeptidase IV and ADA	*H. sapiens and B. taurus*	Tat N-terminal nonapeptide and Zn
N/A	1m7m	ADA	*H. sapiens*	Theoretical model
7/31/08	3E0L	designed ammelide deaminase	*H. sapiens*	1 Zn (α)
3/25/08	2VR2	dihydro-pyrimidinase	*H. sapiens*	2 Zn
1/24/08	2VM8	Crmp-2 D-HYD like	*H. sapiens*	None
4/26/07	2UZ9	GuaD	*H. sapiens*	1 Zn, Xanthine
4/26/06	2GSE	human dihydro-pyriminidase like	*H. sapiens*	None
2/3/02	1ITU	renal dipeptidase	*H. sapiens*	2 Zn, Cilastatin
2/2/02	1ITQ	renal dipeptidase	*H. sapiens*	2 Zn
11/1/00	1E9Z	urease	*H. pylori*	2 Ni
11/1/00	1E9Y	urease	*H. pylori*	2 Ni, Acetohydroxamate
4/29/03	1P6C	PTE H254G/H257W/L303T	*Flavobacterium sp.*	2 Zn, Diisopropyl methyl phosphonate, Diethyl-4-methylbenzyl phosphonate

Table 1.1: Continued.

4/29/03	1P6B	PTE H254G/H257W/ L303T	*Flavobacterium sp.*	3 Zn, Diethyl-4-methylbenzylphosphate, ethyl dihydrogenphosphate
9/13/06	2ICS	put. adenine deaminase	*E. faecalis*	1 Zn, Adenine
1/21/05	1YMY	NagA	*E. coli*	None
8/17/08	3E75	allantoinase	*E. coli*	2 Zn
8/17/08	3E74	allantoinase	*E. coli*	2 Fe
5/17/07	2Z2B	DHO deletion 107-116	*E. coli*	2 Zn
5/17/07	2Z2A	DHO T109G	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydropyrimidine-4-carboxylate
5/17/07	2Z29	DHO T109A	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydro-pyrimidine-4-carboxylate
5/17/07	2Z28	DHO T109V	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydropyrimidine-4-carboxylate
5/17/07	2Z27	DHO T109S	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydropyrimidine-4-carboxylate
5/17/07	2Z26	DHO T110A	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydropyrimidine-4-carboxylate
5/17/07	2Z25	DHO T110V	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydropyrimidine-4-carboxylate

Table 1.1: Continued.

5/17/07	2Z24	DHO T110S	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydropyrimidine-4-carboxylate
3/14/07	2P53	NagA D273N	*E. coli*	1 Zn, *N*-methylphosphonamidate-glucosamine-6-phosphate inhibitor
3/14/07	2P50	NagA	*E. coli*	1 Zn
2/28/07	2EG8	DHO	*E. coli*	2 Zn, 5-flouroorotate
2/28/07	2EG7	DHO	*E. coli*	2 Zn, 2-oxo-1,2,3,6-tetrahydropyrimidine-4,6-dicarboxylate
2/28/07	2EG6	DHO	*E. coli*	2 Zn
11/8/06	2E25	DHO T109S	*E. coli*	2 Zn, 5-Flouroorotate
8/18/05	2AQV	IAD Y137F	*E. coli*	2 Zn
8/18/05	2AQO	IAD E77Q	*E. coli*	2 Zn
6/14/05	1ZZM	TatD homolog	*E. coli*	1 Zn, 3,6,9,12,15,18-hexaoxaicosane-1,20-diol
2/4/05	1YRR	NagA	*E. coli*	Phosphate
1/13/05	1YIX	YcfH TatD homolog	*E. coli*	1 Zn
12/21/04	1YBQ	isoasparatyl dipeptidase D285N	*E. coli*	2 Zn, L-β-aspartylhistidine
11/2/04	1XWY	TatD deoxy-ribonuclease	*E. coli*	1 Zn
9/17/04	1XGE	DHO	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, 4(s)-2,6-dioxohexahydropyrimidine-4-carboxylate

Table 1.1: Continued.

10/31/03	1RAK	CDA D314S	*E. coli*	Fe, 5-fluoro-4-(S)-hydroxy-3,4-dihydropyrimidine
10/31/03	1RA5	CDA D314A	*E. coli*	1 Fe, 5-fluoro-4-(S)-hydroxy-3,4-dihydropyrimidine
10/31/03	1RA0	CDA D314G	*E. coli*	1 Fe, 5-fluoro-4-(S)-hydroxy-3,4-dihydropyrimidine
10/31/03	1R9Z	CDA D314S	*E. coli*	1 Fe
10/31/03	1R9Y	CDA D314A	*E. coli*	1 Fe
10/31/03	1R9X	CDA D314G	*E. coli*	1 Fe
6/15/03	1POK	isoaspartyl dipeptidase	*E. coli*	2 Zn, Asparagine
6/15/03	1POJ	isoaspartyl dipeptidase	*E. coli*	2 Zn, 2-{[[(1S-1-amino-2-carboxyethyl](dihydroxy)phosphoranyl]methyl}-4-methylpentanoic acid
6/15/03	1PO9	isoaspartyl dipeptidase	*E. coli*	2 Zn
5/14/03	1PB0	YCDX protein	*E. coli*	3 Zn
3/2/03	1ONX	isoaspartyl dipeptidase	*E. coli*	2 Zn, Aspartate
3/2/03	1ONW	isoaspartyl dipeptidase	*E. coli*	2 Zn
7/14/02	1M68	YCDX protein	*E. coli*	3 Zn, 1 Sulfate
7/12/02	1M65	YCDX protein	*E. coli*	1 Zn, 2 Na
10/17/01	1K70	CDA	*E. coli*	1 Fe, 4-hydroxy-3,4-dihydro-1H-pyrimidin-2-one inhibitor
10/17/01	1K6W	CDA	*E. coli*	1 Fe
5/16/01	1J79	DHO	*E. coli*	2 Zn, N-Carbamoyl-L-aspartate, dihydroorotate

Table 1.1: Continued.

5/27/98	1BF6	PTE homology protein	*E. coli*	2 Zn, Glycerol, (4S)-2-Methyl-2,4-pentanediol
5/27/08	2VUN	enamidase	*E. barkeri*	1 Fe, 1 Zn, 1Cl
2/27/09	3GG7	TatD like	*D. radiodurans*	1 Mn (α)
11/2/07	2ZC1	PTE	*D. radiodurans*	2 Co
10/4/06	2IMR	SAH/TMA deaminase DR_0824	*D. radiodurans*	1 Zn
1/25/06	2FTW	dihydro-pyriminidase	*D. discoideum*	2 Zn
3/17/09	3GNH	L-lysine/L-arginine carboxypeptidase	*C. vibrioides*	2 Zn, Methyl phosphonated L-Arginine
9/14/07	2RAG	Putative dipeptidase	*C. cresentus*	2 Zn, Cl
5/18/07	2Q01	Uronate isomerase	*C. cresentus*	None
9/6/06	2I9U	GuaD	*C. acetobutylium*	1 Fe, Guanine
3/13/08	3CJP	CAC3332	*C. Aceto-butylicum*	2 Zn
1/26/05	1YNY	D-hydantoinase	*Bacillus sp.*	2 Mn
2/28/09	3GGM	Bt9727_2919	*B. thuringiensis*	None
8/20/07	2Z7G	ADA	*B. taurus*	1 Zn (α), (2S,3R)-3-(6-amino-9H-purin-9-yl)nonan-2- ol
10/30/06	2E1W	ADA	*B. taurus*	1 Zn, 1-{(1R,2S)-2-Hydroxy-1-[2-(1-naphthyl)ethyl]propyl}-1H-imidazole-4-carboxamide

Table 1.1: Continued.

2/2/05	1WXZ	ADA	*B. taurus*	1 Zn, 1-((1R,2S)-1-{2-[2-(4-chlorophenyl)-1,3-benzoxazol- 7-yl]ethyl}-2-hydroxypropyl)-1H-imidazole- 4-carboxamide
2/2/05	1WXY	ADA	*B. taurus*	1 Zn, N-[4,5-Bis(4-hydroxyphenyl)-1,3-thiazol-2- yl]hexanamide
6/22/04	1W1I	ADA and dipeptidyl peptidase	*B. taurus*	1 Zn, Dipeptidyl peptidase (IV)
4/16/04	1VFL	ADA	*B. taurus*	1 Zn
12/14/03	1V7A	ADA	*B. taurus*	1 Zn, 1-{(1R,2S)-2-Hydroxy-1-[2-(2-naphthyloxy)ethyl]propyl}- 1H-imidazone-4-carboxamide
12/14/03	1V79	ADA	*B. taurus*	1 Zn, 1-{(1R,2S)-1-[2-(2,3,-Dichlorophenyl)ethyl]-2-hydroxypropyl}-1H-imidazole-4-carboxamide
10/5/03	1O5R	ADA	*B. taurus*	1 Zn, 1-[(1R)-3-(6-{[(Benzylamino)carbonyl]amino}- 1H-indol-1-yl)-1-(hydroxymethyl)propyl]-1H- imidazole-4-carboxamide
10/3/03	1UML	ADA	*B. taurus*	1 Zn, 1-((1R)-1-(hydroxymethyl)-3-{6-[(3-phenylpropanoyl)amino]- 1H-indol-1-yl}propyl)-1H-imidazole-4-carboxamide

Table 1.1: Continued.

9/8/03	1QXL	ADA	*B. taurus*	1 Zn, 1-((1R)-1-(hydroxymethyl)-3-(1-naphthyl)propyl)- 1H-imidazole-4-carboxamide
12/9/02	1NDZ	ADA	*B. taurus*	1 Zn, 1-((1R)-1-(hydroxymethyl)-3-(6-((3-(1-methyl- 1H-benzimidazol-2-yl)propanoyl)amino)-1H- indol-1-yl)propyl)-1H-IMIDAZOLE-4-carboxamide
12/9/02	1NDY	ADA	*B. taurus*	1 Zn, 1-((1R)-1-(hydroxymethyl)-3-(1-naphthyl)propyl)- 1H-imidazole-4-carboxamide
12/9/02	1NDW	ADA	*B. taurus*	1 Zn, 1-((1R)-1-(hydroxymethyl)-3-phenylpropyl)- 1H-imidazole-4-carboxamide
12/9/02	1NDV	ADA	*B. taurus*	1 Zn, N''-(4-(5-((1H-benzimidazol-2-ylamino)methyl)- 2-thienyl)-1,3-thiazol-2-yl)guanidine
1/10/02	1KRM	ADA	*B. taurus*	1 Zn, 6-hydroxyl-1,6-dihydropurine riboside
11/22/07	2VHL	NagA	*B. subtilis*	2 Fe, glucosamine-6-phosphate,
11/22/07	2VHL	NagA	*B. subtilis*	2 Fe, Glucosamine-6-phosphate product
2/19/06	2G3F	imidazolone-propionase	*B. subtilis*	1 Zn, substrate analogue imidazole-4-acetate
10/16/05	2BB0	imidazolone-propionase	*B. subtilis*	1 Zn, acetate
9/25/01	1K1D	D-hydantoinase	*B. stereo-thermophilus*	2 Zn

Table 1.1: Continued.

12/14/02	1NFG	D-hydantoinase	*B. pickettii*	2 Zn
1/14/04	1S3T	urease	*B. pasteurii*	2 Ni, Sulfate, Borate
4/9/01	1IE7	urease	*B. pasteurii*	2 Ni, Phosphate
2/25/99	4UBP	urease	*B. pasteurii*	2 Ni, Acetohydroxamate
12/16/98	3UBP	urease	*B. pasteurii*	2 Ni, Diamidophosphate
11/4/98	2UBP	urease	*B. pasteurii*	2 Ni, Sulfate
1/21/98	1UBP	urease	*B. pasteurii*	2 Ni, β-mercaptoethanol
3/24/07	2P9B	putative prolidase	*B. longum*	None
1/25/07	2OOD	GuaD	*B. japonicum*	1 Zn, Guanine
6/25/07	2QEE	BH0493	*B. halodurans*	1 Zn (α)
6/5/07	2Q6E	uronate isomerase	*B. halodurans*	1 Zn (α)
5/21/07	2Q08	uronate isomerase	*B. halodurans*	1 Zn (α)
4/24/07	2PNK	BH0493	*B. halodurans*	None
8/7/08	3E3H	PTE H254R, H275F	*B. dimunita*	2 Co, Diethyl 4-methylbenzyl-phosphonate
12/18/06	2OB3	PTE H257Y/L303T	*B. dimunita*	2 Zn
9/1/03	1QW7	PTE	*B. dimunita*	2 Co, Diethyl-4-methylbenzyl-phosphonate
6/26/01	1JGM	PTE	*B. dimunita*	2 Cd, Ethanol, Formate, 2-phenyl-ethanol
1/29/01	1I0D	PTE	*B. dimunita*	Zn, Cd, Ethanediol, Formate, 2-phenyl-ethanol
1/29/01	1I0B	PTE	*B. dimunita*	2 Mn, Ethanediol, Formate, 2-phenyl-ethanol
1/26/01	1HZY	PTE	*B. dimunita*	2 Zn, Ethanediol, Formate, 2-phenyl-ethanol
2/13/96	1DPM	PTE	*B. dimunita*	2 Zn, Diethyl-4-methylbenzyl-phosphonate
4/25/95	1PSC	PTE	*B. dimunita*	2 Cd, Diethyl-4-methylbenzyl-phosphonate

Table 1.1: Continued.

7/7/94	1PTA	PTE	*B. dimunita*	None
4/8/08	3CS2	PTE G60A	*B. diminuta*	2 Co, Dimethylarsenate (cacodylate)
2/20/08	3CAK	PTE	*B. diminuta*	2 Co, Ethylphosphate
7/31/08	3E0F	putative PTE	*B. adolescentis*	2 Fe, 1Zn, acetate, phosphate
2/10/08	3C86	OpdA PTE S92A	*A. tumefaciens*	1 Co, 1 Fe, Diethyl thiophosphate product, 1,2-ethanediol
8/23/07	2R1P	OpdA PTE S92A, K185R	*A. tumefaciens*	1 Co, 1 Fe , Diethyl thiophosphate product, 1,2-ethanediol
8/23/07	2R1O	OpdA PTE S92A	*A. tumefaciens*	1 Co, 1 Fe, Diethyl thiophosphate product, 1,2-ethanediol
8/23/07	2R1N	OpdA PTE S92A, N265D	*A. tumefaciens*	1 Co, 1 Fe, Diethyl 4-methoxyphenyl phosphate slow substrate,
8/23/07	2R1M	OpdA PTE S92A, N265D	*A. tumefaciens*	1 Co, 1 Fe, Diethyl phosphate product, 1,2-ethanediol
8/23/07	2R1L	OpdA PTE S92A, N265D	*A. tumefaciens*	1 Co, 1 Fe, Diethyl thiophosphate, 1,2-ethanediol
8/23/07	2R1K	OpdA PTE S92A, N265D	*A. tumefaciens*	1 Co, 1 Fe, Diethyl phosphate, 1,2-ethanediol
5/9/07	2PUZ	imidazolonepropionase	*A. tumefaciens*	1 Fe, N-formimino-L-glutamate product
1/5/07	2OGJ	DHO-like	*A. tumefaciens*	2 Zn, Imidazole
4/13/06	2GOK	imidazolonepropionase	*A. tumefaciens*	1 Fe
9/8/05	2D2J	PTE	*A. tumefaciens*	2 Co
9/8/05	2D2H	PTE	*A. tumefaciens*	2 Co, Trimethylphosphate

Table 1.1: Continued.

9/8/05	2D2G	PTE	*A. tumefaciens*	2 Co, Dimethylthiophosphate
6/25/05	2A3L	AMP deaminase	*A. thaliana*	1 Zn, Coformycin 5'-phosphate
7/30/07	2QS8	Xaa-Pro dipeptidase	*A. macleodii*	1 Mg (α) , Methionine
3/30/09	3GUW	TatD like	*A. Fulgidus*	2 Zn
11/20/03	1V51	D-aminoacylase	*A. faecalis*	2 Zn, Acetate
11/20/03	1V4Y	D-aminoacylase H220A	*A. faecalis*	1 Zn, AcetateLigand free/ Mutant
11/20/03	1RK6	D-aminoacylase	*A. faecalis*	1Zn, 1 Cd, Acetate
11/20/03	1RK5	D-aminoacylase D366A	*A. faecalis*	1 Zn, 1 Cu, Acetate
11/20/03	1RJR	D-aminoacylase D366A	*A. faecalis*	2 Zn, Acetate
11/20/03	1RJQ	D-aminoacylase D366A	*A. faecalis*	1 Zn, Acetate
11/20/03	1RJP	D-aminoacylase	*A. faecalis*	1 Zn, 1 Cu, Acetate
7/22/02	1M7J	D-aminoacylase	*A. faecalis*	2 Zn, Acetate
8/20/01	1GKR	L-hydantoinase	*A. aurescens*	2 Zn
5/20/08	3D6N	DHO	*A. Aeolicus*	1 Zn, Citrate
10/15/04	1XRT	DHO	*A. aeolicus*	2 Zn
10/14/04	1XRF	DHO	*A. aeolicus*	2 Zn, Sulfate

Based on the structurally characterized members of the AHS, there are several identified variations of the divalent metal centers, as depicted in **Figure 1.2**. The most common metal center is the binuclear metal center found in enzymes such as phosphotriesterase, dihydroorotase (DHO), isoaspartyl dipeptidase (IAD), urease, and hydantoinase (**Figure 1.2A**). In all of these proteins, the two metal ions are ligated to the protein via six amino acid side chains. According to the crystal structures, M_α, the more buried metal ion, is coordinated to the two histidines in the HxH motif at the end of β-strand 1 and an aspartate from β-strand 8. The more solvent exposed metal, M_β, is

ligated to the imidazole side chain of two histidines at the end of β-strands 5 and 6. The two metal ions are bridged by a solvent hydroxide as well as a carbamate functional group of a posttranslational modification of a carboxylated lysine from strand 4 (*55*). . One variation of this metal center subtype is the replacement of the bridging carboxylated lysine from β-strand 4 to a glutamate (**Figure 1.2A)**. This is seen in the structure of the PTE homology protein (PHP) from *E. coli*. Another variation of this metal center subtype is seen with the crystal structure of an organophosphorus hydrolase from *Deinococcus radiodurans* in which the β-metal ion is coordinated to a tyrosine from β-strand 3 in addition to the other four ligands (**Figure 1.2B)**. In the crystal structure of the enzyme *N*-acetyl glucosamine-6-phosphate deacetylase (AGD) from *Bacillus subtilis*, the binuclear metal center ligands are conserved except the bridging residue glutamate which derives from strand 3 instead of strand 4 (**Figure 1.2C**) (*42*). A more drastic variation to the binuclear center is seen with the enzyme renal dipeptidase (RDP) from humans (**Figure 1.2D**). According to the crystal structure, the M_α ligand motif from strand 1 is HxD instead of HxH, and the residue that bridges the two metals is a glutamate at the end of β-strand 3 instead of strand 4. Another significant difference in the structure of RDP is that the conserved aspartate from strand 8 is no longer a metal ligand to M_α, instead it interacts with the hydrolytic water molecule via hydrogen bonding (*40*).

Another subtype of metal center among the AHS enzymes is found in adenosine deaminase (ADA) and cytosine deaminase (CDA) (**Figure 1.2E**). These enzymes contain a single metal ion in the active site at the position that corresponds to the M_α

position of the binuclear center. According to the structures, the metal ion is coordinated to the two histidines from strand 1 as well as the conserved histidine from strand 5. The histidine from strand 6 is not a metal ligand but appears to play a role in catalysis. The residue that houses the bridging carboxylate from strand 4 is missing. In addition to the protein, the metal is also ligated to a solvent water molecule (*56-57*). One of the newest additions to this subtype is the protein α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (ACMSD) from *Pseudomonas fluorescens*, a member of the amidohydrolase that catalyzes a nonhydrolytic reaction. It has been suggested that this enzyme utilizes similar chemistry to that employed by ADA, CDA, and other M_{α} mononuclear enzymes of the AHS in its catalytic mechanism (*50,58*). One variation of this subtype metal center is seen in the structure of γ-resorcylate decarboxylase in which the HxH motif at the end of β-strand 1 has been replaced by an ExH motif (**Figure 1.2F)**. Another variation of the α-metal center is the replacement of the aspartate from β-strand 8 with a glutamate, as seen in the structure of 4-oxalomesaconate hydratase (PcmD) (**Figure 1.2G)**. A structure of a TatD-related DNase from *D. radiodurans* (PDB code 3GG7) indicates another metal center variation in which the histidine from strand 5 is present but no longer a metal ligand, instead a glutamate from strand 4 is now coordinating to the metal (**Figure 1.2H**). Another structure of an enzyme belonging to the TatD-related DNase family from *Saccharomyces cerevisiae* (PDB code 3e2v) contains a magnesium ion in which the metal is coordinated to an aspartate from strand 8, a histidine from strand 6, and three solvent water molecules. The histidine from strand 5 is conserved but it is not a metal ligand (**Figure 1.2I**).

The final variation of the α-metal center is represented by the uronate isomerase (URI) family. The crystal structures for three members of this family have been deposited in the Protein Data Bank (PDB codes 1j5s, 2q01, and 2q6e). From the structure of the *Thermotoga maritima* and *Caulobacter crescentus*, no metal ion is found in the active site. However, metal ligands for the M_α are fully present, suggesting that the proteins are capable of binding a divalent cation in the M_α site (**Figure 1.2J**). In addition, the crystal structure of a newly discovered second uronate isomerase from *Bacillus halodurans* (Bh0493) reveals the presence of one zinc ion in the active site. According to the structure, the zinc is coordinated to the two histidines in the HxH motif at the end of β-strand 1 and an evolutionary conserved aspartate from β-strand 8. Sequence and structural alignment indicate that the histidine from strand 5 is conserved among the URI from *E. coli*, *T. maritima*, and *C. crescentus*, but is no longer present in the Bh0493 enzyme (**Figure 1.2K**). Moreover, the histidine from strand 6 is also not conserved among the uronate isomerase family (*42-43*). A slight variation from this is seen with the structure of Dr0824 (PDB code 3imr), a hypothetical protein from *D. radiodurans* (**Figure 1.2L**). According to the structure, the enzyme contains a zinc ion in the M_α position where it is coordinating to the HxH from strand 1, a histidine from strand 5, and a water molecule. The histidine from strand 6 is present but does not take part in the metal coordination. The aspartate from strand 8 is also conserved but it too is not a metal ligand, instead it interacts with the water molecule that is coordinating to the metal via hydrogen bonding.

A variation of the mononuclear metal center is with the enzyme AGD from *T. maritima* (**Figure 1.2M**). This enzyme has all the metal ligands found in the PTE subgroup with the exception of the bridging carboxylate from strand 4. Instead, a glutamate from strand 3 is positioned as a potential bridging residue. Though the crystal structure only shows a single metal ion in the active site, the presence of all the conserved metal ligands suggests that the enzyme from *T. maritima* is capable of binding to two metal ions under the right conditions, but only a single metal ion is required for catalytic activity. In the crystal structure of AGD from *E. coli*, electron density corresponds to a single metal ion in the active site. In addition, the two histidines from strand 1 have been replaced with a glutamine and an asparagine. Unlike ADA and CDA, the metal ion in the active site of AGD from *T. maritima* and *E. coli* occupies the M_β site instead of the M_α site. From the structure, the metal ion is coordinated by the imidazole side chain of the two histidines from strand 5 and 6, a glutamate at the end of strand 3, and a solvent water molecule (**Figure 1.2N**) (*42*).

Another variation of a single divalent cation center in the M_β positon is seen with the enzyme D-amino acid deacetylase (DAA), as illustrated in **Figure 1.2O**. From the crystal structure of DAA, the histidines from strand 1 do not interact with the metal. The metal is coordinated by the imidazole side chain of the histidines from strand 5, the imidazole side chain of the histidine from strand 6, and an unusual cysteine residue from strand 2. The aspartate from strand 8 is positioned to act as an activator of the hydrolytic water. It has been shown that the second metal can be forced into the active site of DAA, but its presence is not necessary for enzymatic activity (*45,59*). Another variation of the

β-metal center subtype is seen in the structure of TatD (**Figure 1.2P**). According to the crystal structure, the metal is coordinated to a cysteine and a histidine from strand 6 as well as a glutamate in the ExD motif from strand 8.

Final variations of metal center in the AHS are with the structures of the histidinol-phosphate phosphatase family (PHP). According to the structure of PHP from *T. thermophilus* HB8 (**Figure 1.2Q**), the enzyme contains a trinuclear metal cluster in which there are two irons and one zinc (*53*). The distances between Fe_1-Fe_2, Fe_1-Zn, and Fe_2-Zn are 3.47, 4.70, and 5.84 Å, respectively. The two irons represent the binuclear center and their metal coordination is similar to the ones seen in the *B. subtilis* AGD subtype. The zinc, designated as M_γ, represents a novel subtype of metal center with distinct metal ligands (a histidine from loop 1, a histidine from strand 3, and a histidine from strand 8). Another structure belonging to the PHP family contains up to 4 zinc ions in the active site (**Figure 1.2R**). The metals in the tetranuclear metal cluster are designated as M_α, M_β, M_γ and M_δ sites. The M_α and M_β sites resemble the binuclear center in which the M_α is liganded to the HxH from strand 1, the aspartate from strand 8, and is bridged to the M_β by a glutamate from strand 3. The M_β, however, is coordinated to a histidine from strand 4 in addition to the glutamate from strand 3. The M_γ metal interacts with an aspartate from the loop after strand 1, a histidine from strand 2, a histidine from strand 8, and a water molecule. The metal M_δ of the tetranuclear cluster is bridged to the M_α by the aspartate from strand 8 as well s coordinating to an aspartate from strand 5. The structures of HolPase-family, therefore, represent extreme cases of structural diversity within the AHS.

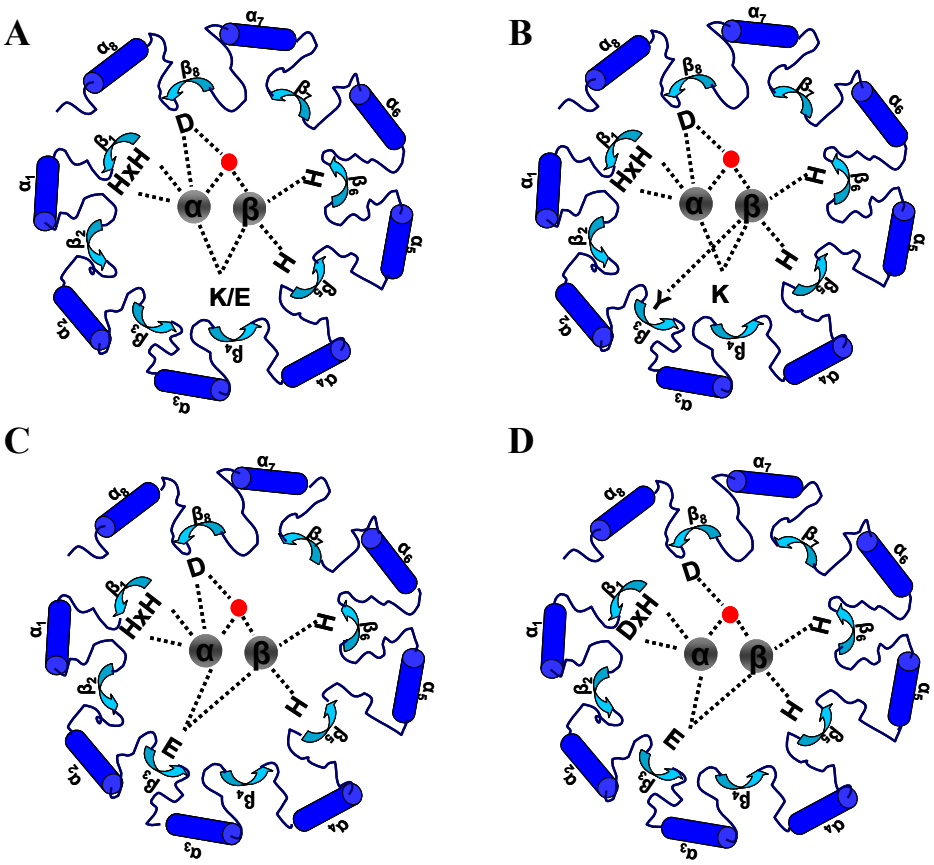


Figure 1.2: Variations of binuclear and mononuclear metal centers in the AHS. The metal ions are in grey and the water molecules are in red. Representations of the different variations are PTE from *P. diminuta* (A), PTE from *D. radiodurans* (B), AGD from *B. subtilus* (C), hRDP (D), CDA (E), rdc (F), PcmD (G), a TatD-related DNase from *D. radiodurans* (H), an unknown from *S. cerevisiae* (I), URI (J), URI from *B. halodurans* (Bh0493) (K), Dr0824 (L), AGD from *T. maritima* (M), AGD from *E. coli* (N), DAA (O), TatD (P), and HolPase-family (Q,R).

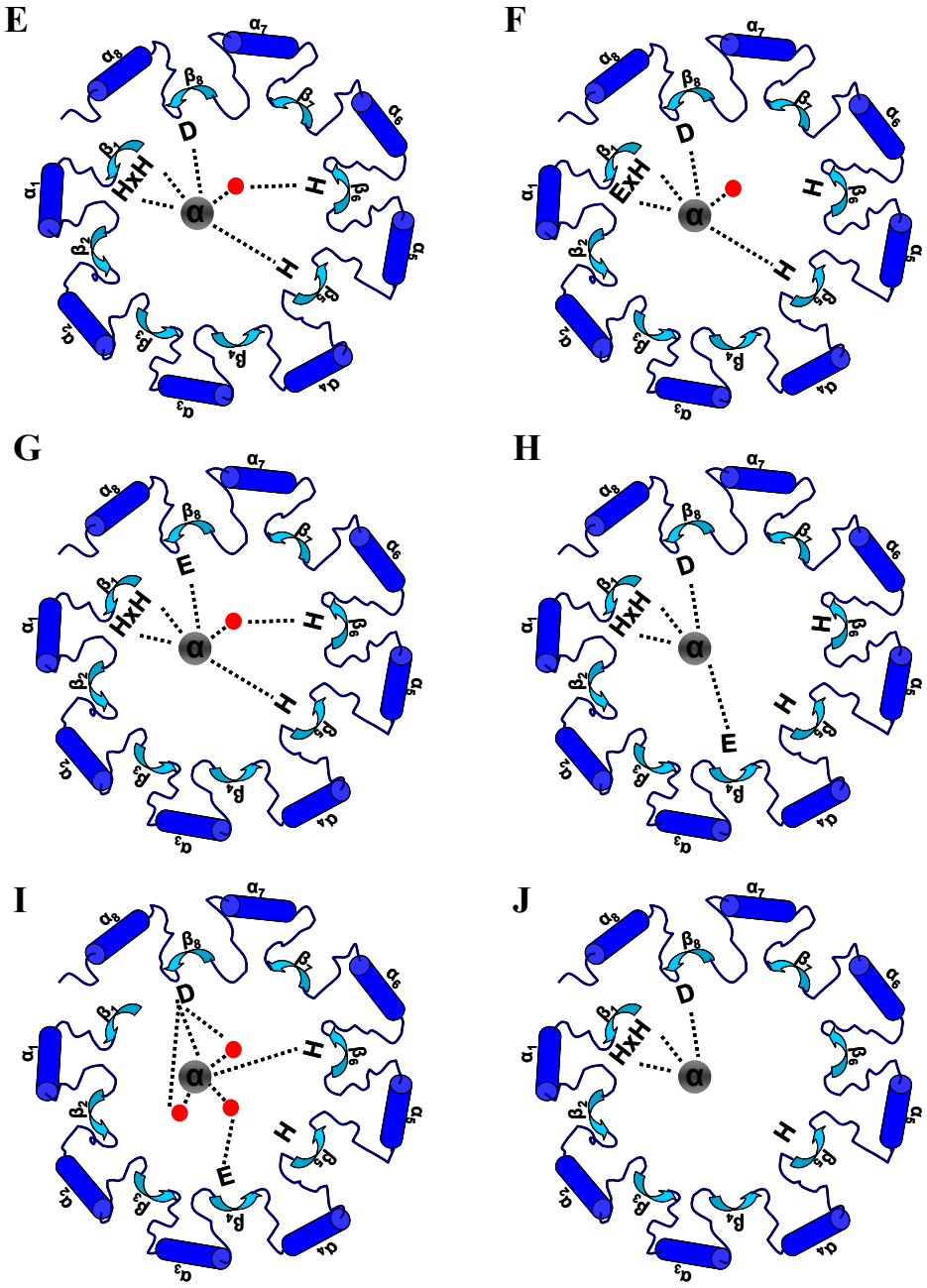


Figure 1.2: Continued.

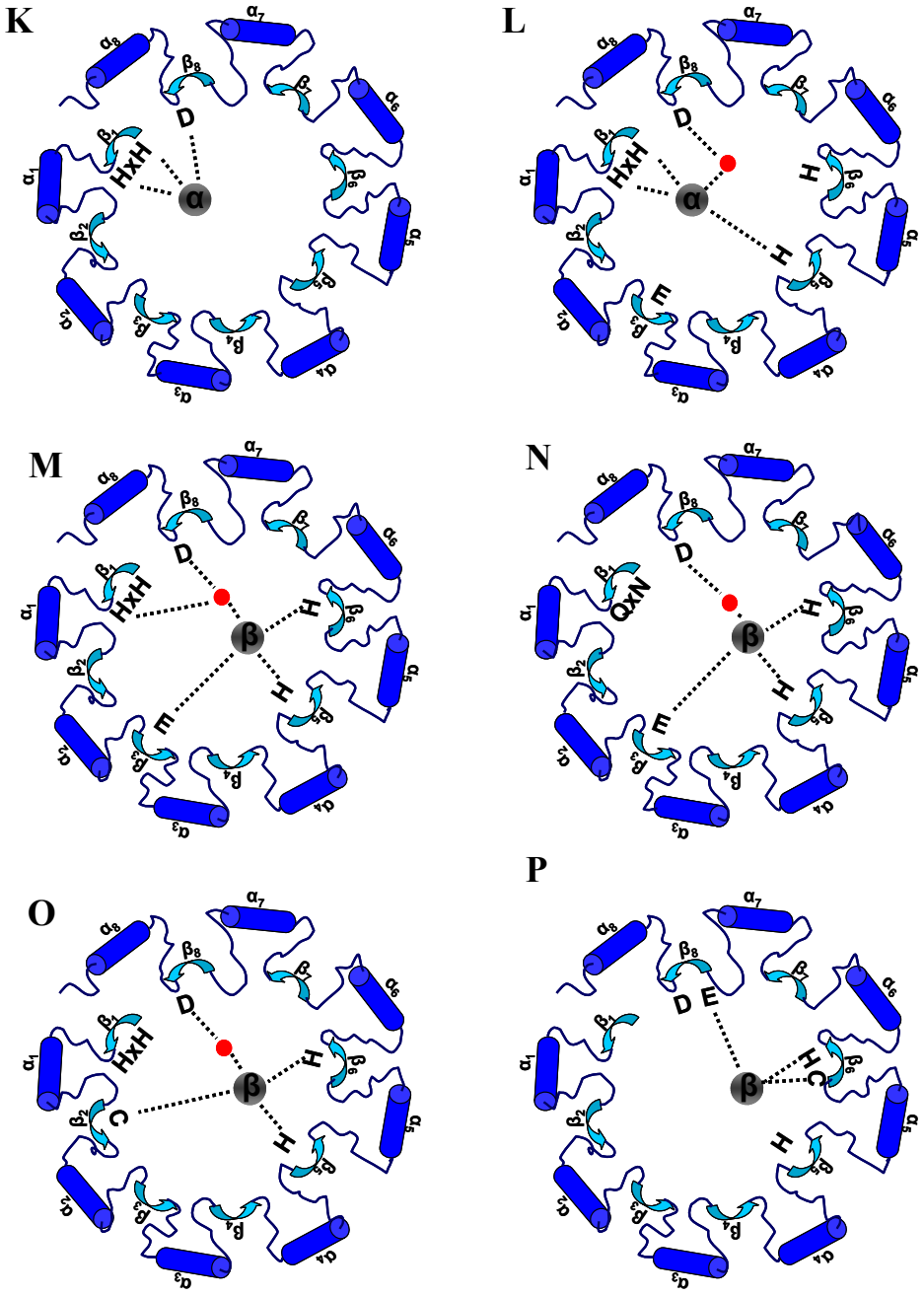


Figure 1.2: Continued.

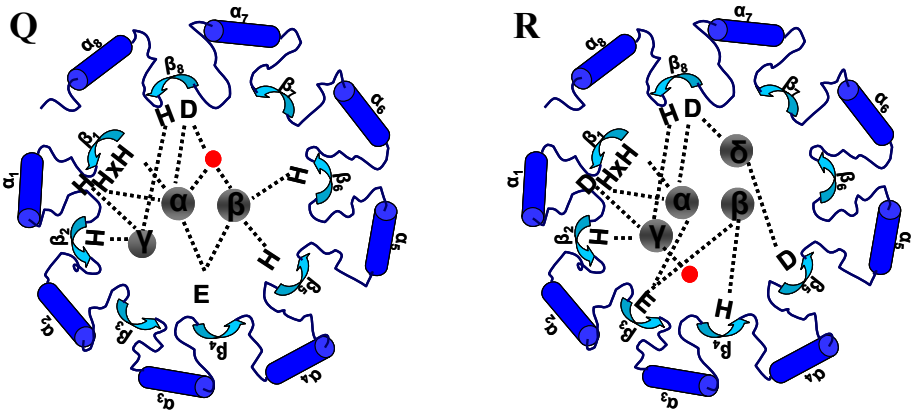
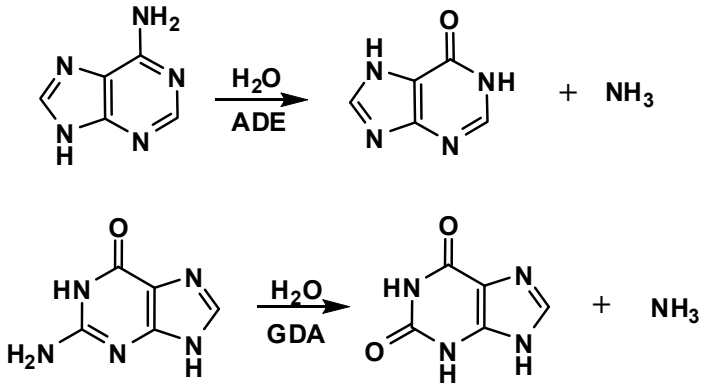


Figure 1.2: Continued.

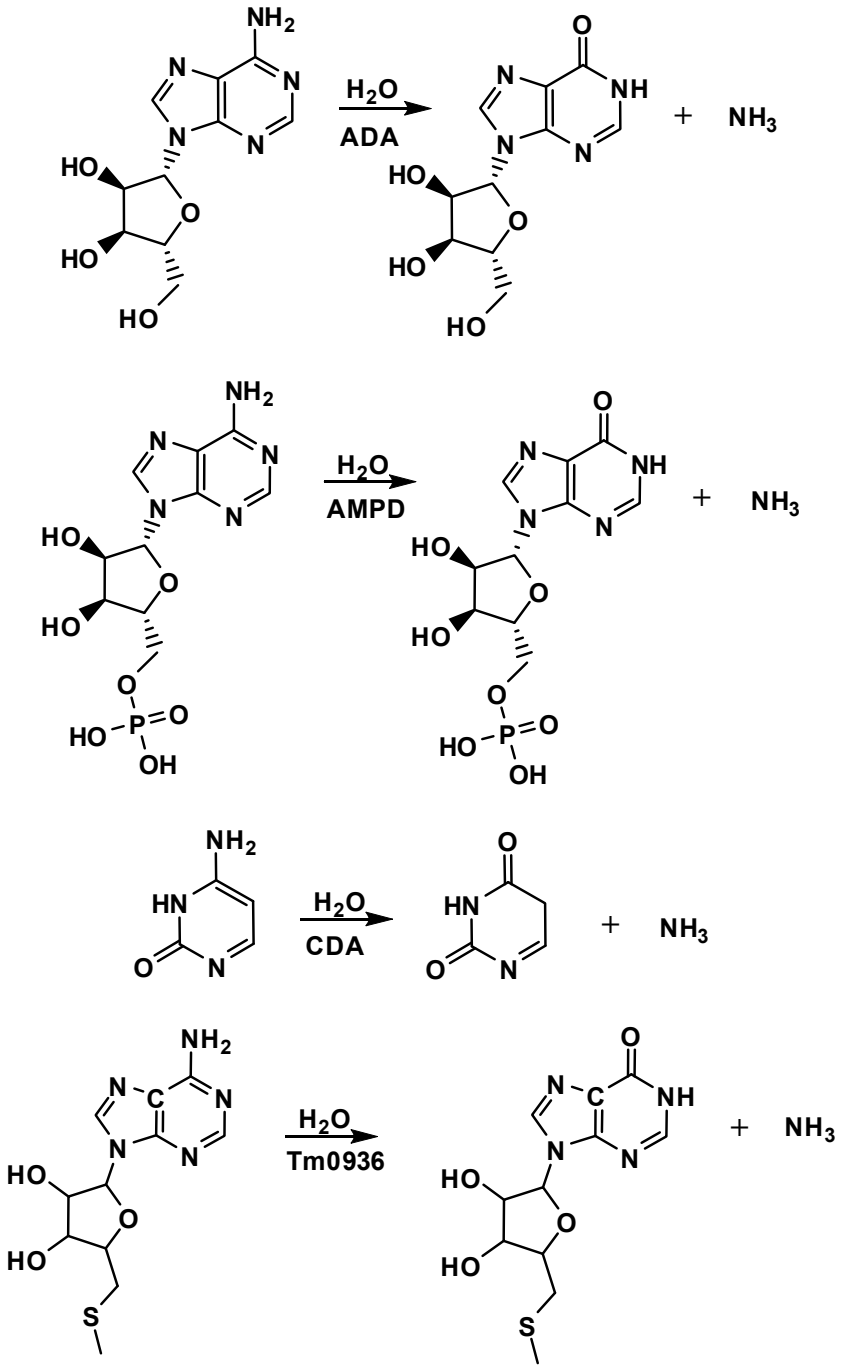
The amidohydrolase superfamily is a functionally diverse group of enzymes, and members of this superfamily have been found in every organism sequenced to date (*31*). These enzymes are found in the pathways of pyrimidine metabolism (CDA, DHO, DHP, and HYD) (*,36-38,47,60*), purine metabolism (ADA, ADE, AMPD, Tm0936, GDA, URE and ATase) (*7,32,39,46-48,61-62*), atrazine degradation (AtzA, AtzN, AtzB, AtzC, trzA, and TriA) (*63-67*), organophosphate degradation (PTE and adpB) (*68-69*), histidine metabolism (HolPase, HutI and HutF) (*3,49,53*), tryptophan catabolism (ACMSD) (*50*), sugars metabolism (AGD and URI) (*70,44*), amino acid metabolism (Cc3125, Cc0300, RDP, IAD, D-ANase, D-Amase, D-AAase, and D-AGase) (*5-6,40,71-74*), nicotinate catabolism (ENA) (*52*), phthalate catabolism (PcmC and PcmD) (*75*), hydroxybenzoates biodegradation (rdc) (*76*), lignin-related biphenyl structure degration (ligW and ligY) (*77-78*), and thymidine salvage (IDCase) (*79*). Since its

initial discovery, there are approximately forty families that have been functionally characterized and the number of sequences available on the Structure-Function Linkage Database (SFLD) now exceeds 10,000 (*10,80*). The reactions catalyzed by the enzymes of the AHS are predominantly hydrolytic reactions of the C-N amide bond of amino acids, nucleic acids, and sugars. These enzymes also catalyze reactions involing C-O and P-O ester bond cleavage. The mechanism of hydrolysis for this enzyme superfamily involves the activation of a water molecule by the metal center for nucleophilic attack. The metal ion also plays a role in the activation of the scissile bond of the substrate for bond cleavage and takes part in the stabilization of the tetrahedral intermediate. In addition to hydrolytic reactions, a few enzymes of this superfamily also catalyze nonhydrolytic reactions such as nonoxidative decarboxylation and isomerization reactions. Examples of reactions catalyzed by AHS enzymes (as of June 12, 2009) are catalogued in **Scheme 1.1**.

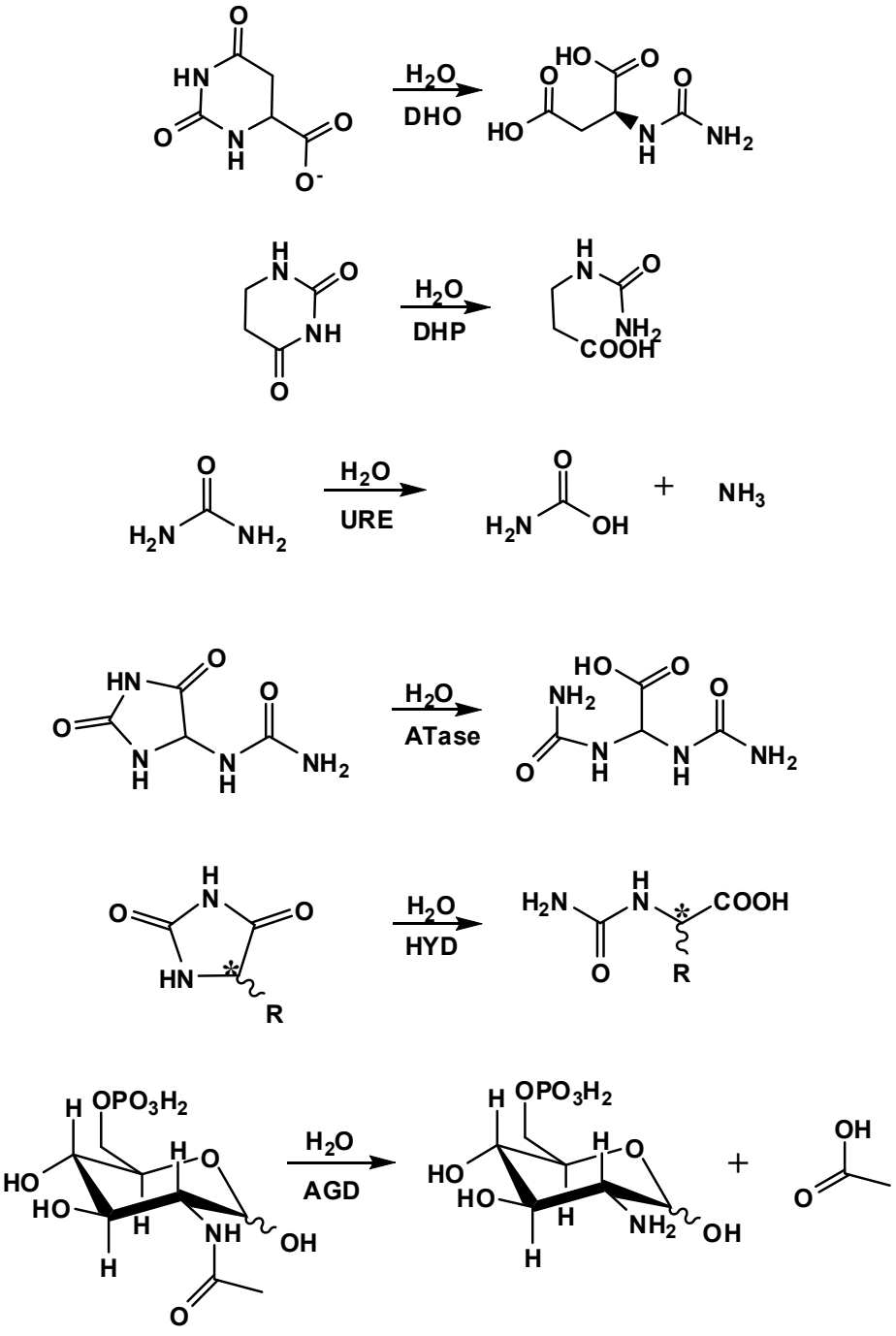
Scheme 1.1: Examples of reactions catalyze by enzymes of the AHS.



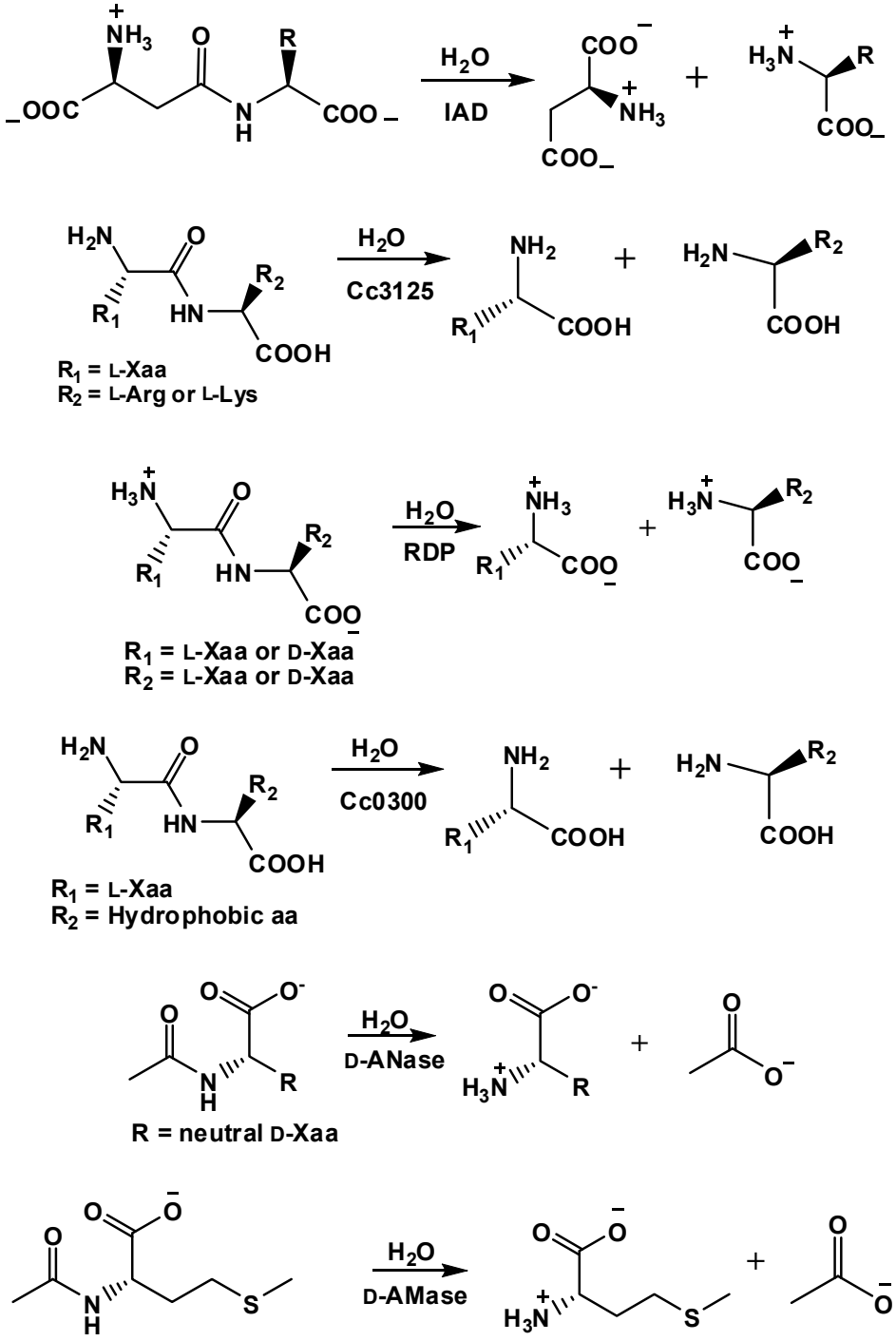
Scheme 1.1: Continued.



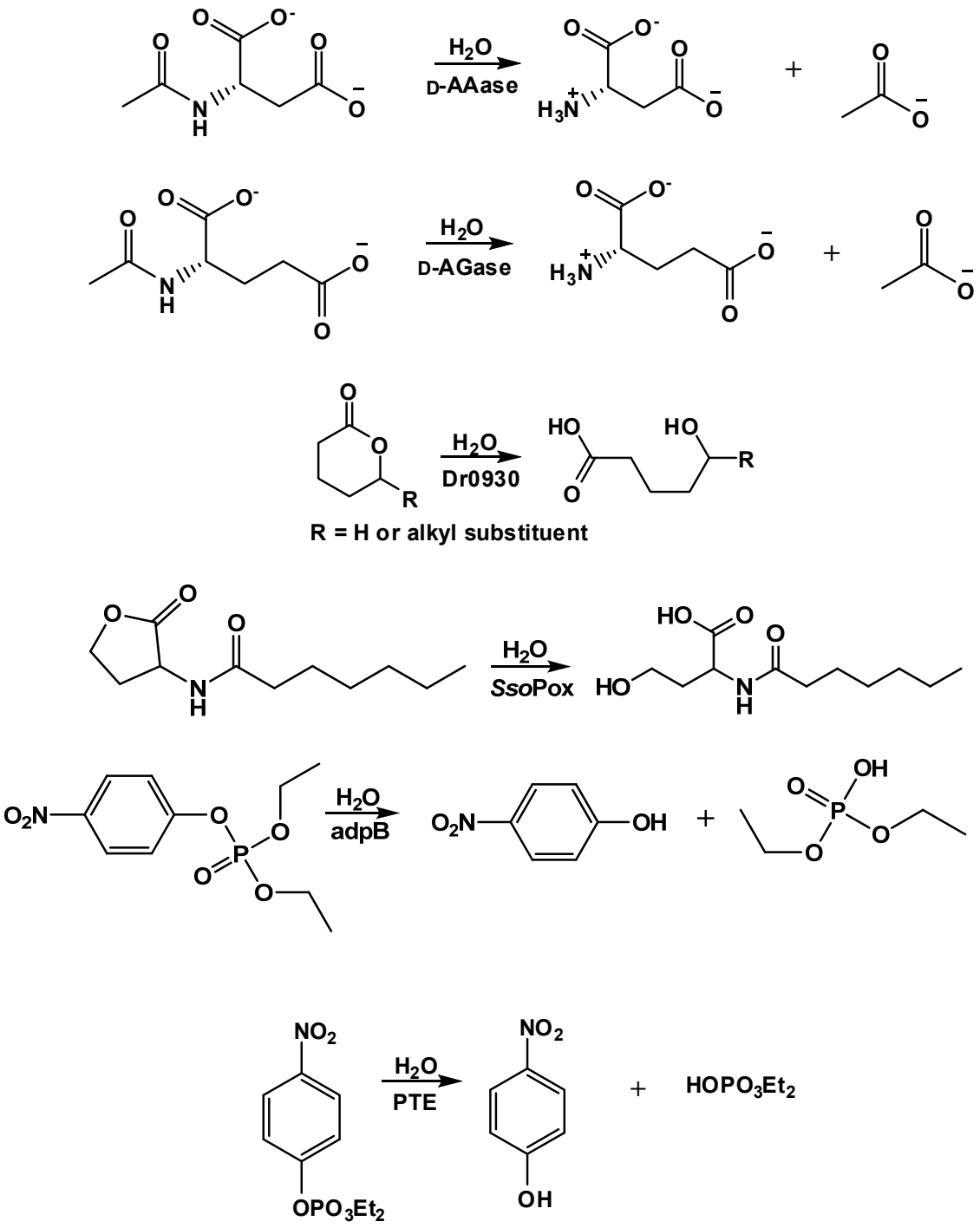
Scheme 1.1: Continued.



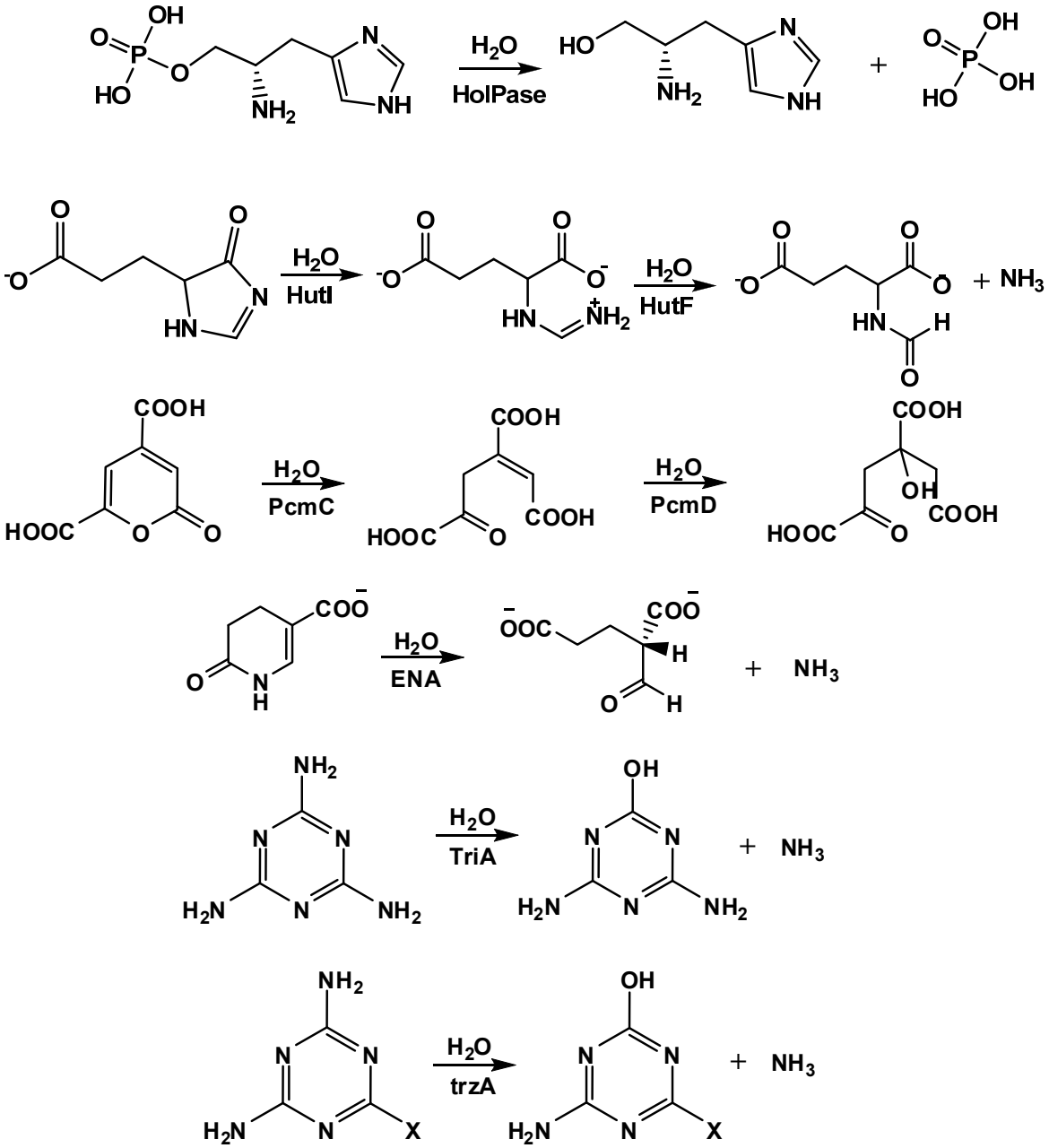
Scheme 1.1: Continued.



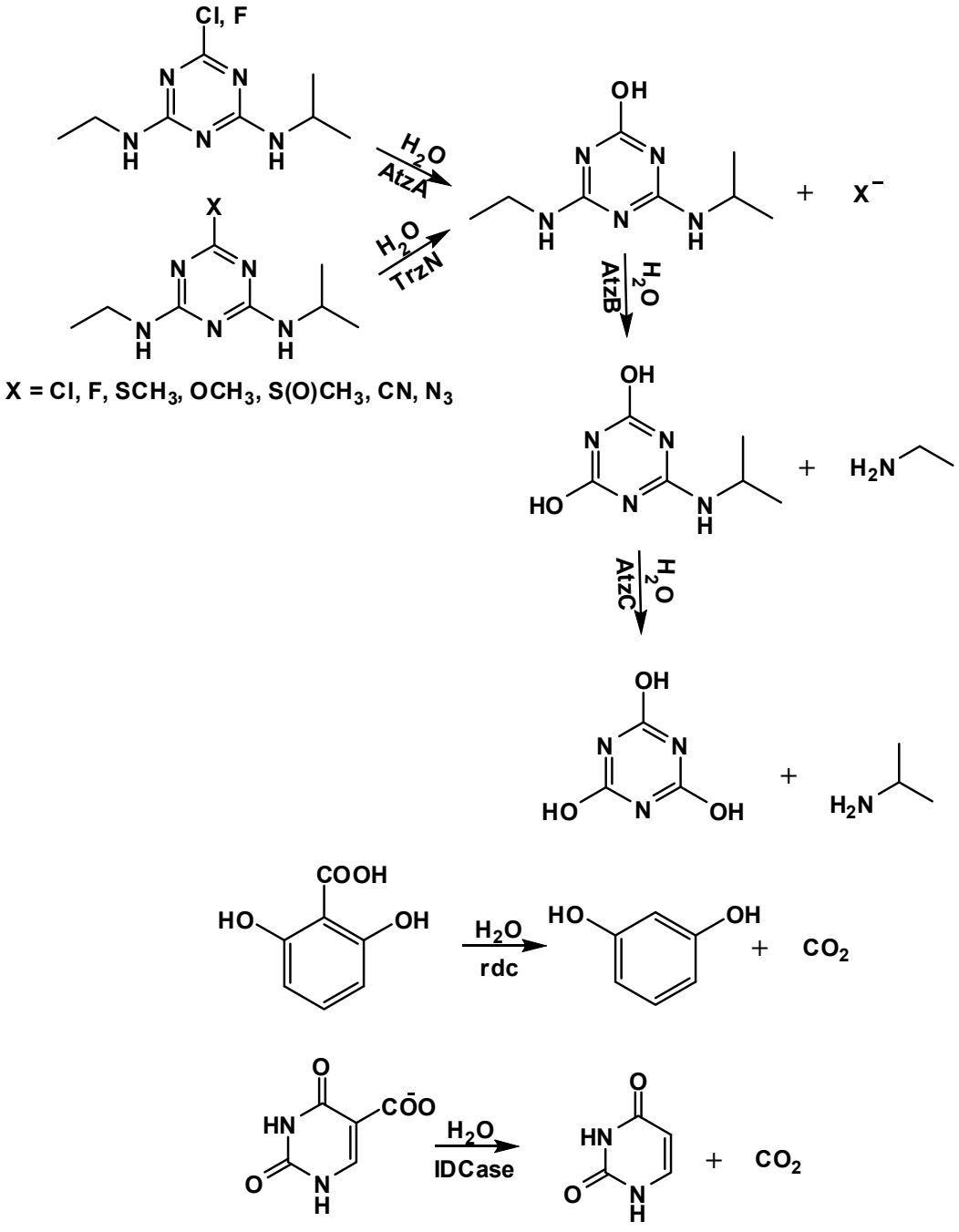
Scheme 1.1: Continued.



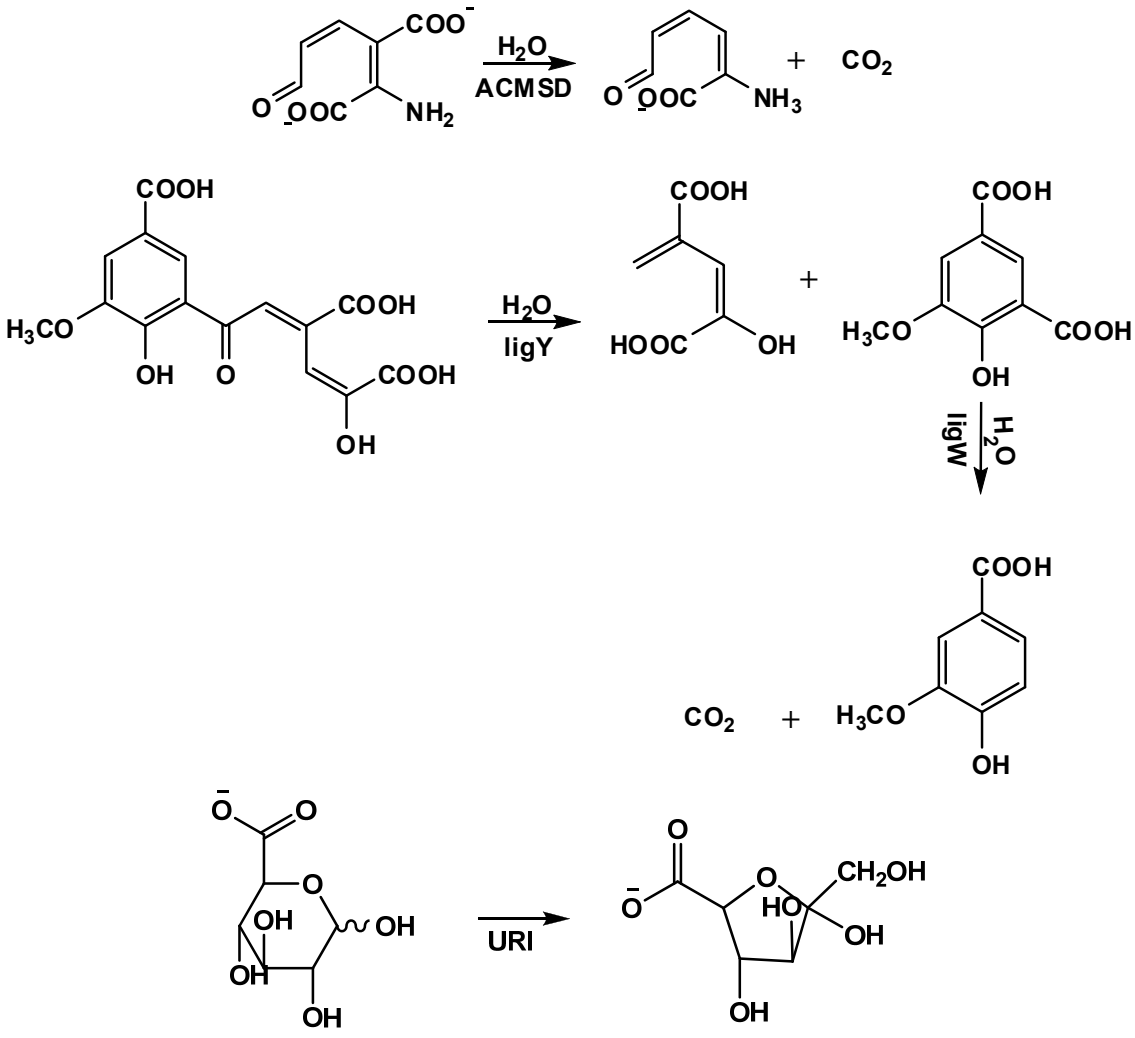
Scheme 1.1: Continued.



Scheme 1.1: Continued.



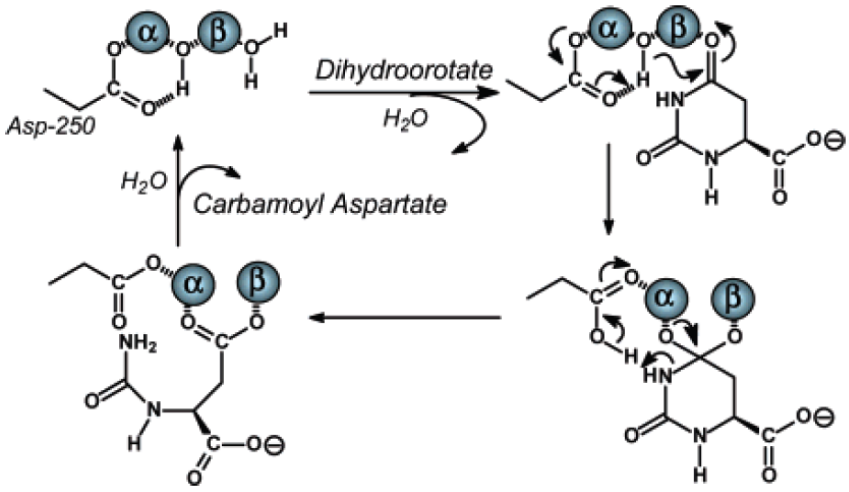
Scheme 1.1: Continued.



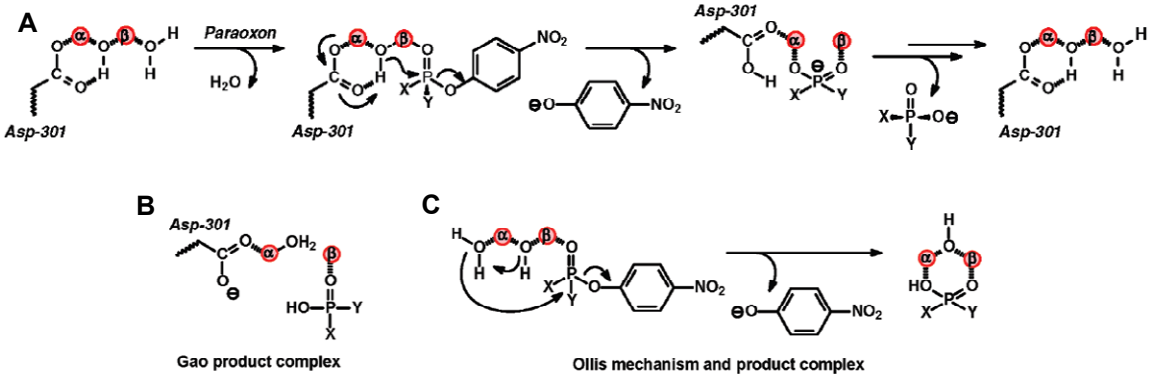
The detailed reaction for some of the enzymes of the AHS has been examined in depth. The best understood case is with the enzyme dihydroorotase from *E. coli* in which the crystal structure has been solved with substrate and product bound to separate monomeric subunits within the dimeric protein (*34*). The specific interactions within the enzyme active site before and immediately after catalysis are therefore manifested. Dihydroorotase catalyzes the reversible hydrolysis of dihydroorotate to carbamoyl aspartate. In the structure with dihydroorotate bound in the active site, the carbonyl group of the scissile amide bond is shown to be interacting with the M_β metal ion. This interaction helps polarize the amide bond and enhances the electrophilicity of the carbonyl carbon. The bridging hydroxide is hydrogen bonded to the aspartate from β-strand 8 and is positioned for the nucleophilic attack on the *re*-face of the carbonyl substrate. In the structure of DHO complexed with carbomoyl aspartate, the forming carboxylate group bridges the two metal ions in the active site. This indicates that the hydrolytic water is the bridging hydroxide. The conserved aspartate from strand 8 is positioned to protonate the amide nitrogen, suggesting its role as a proton shuttle between substrate and product (*34,60*). An illustration of the mechanism of DHO is shown in **Scheme 1.2** (*31*). Similar mechanisms have been proposed for isoaspartyl dipeptidase (*71*) and phosphotriesterase (*68*). Alternate mechanisms have also been proposed for PTE. In the PTE computational-based mechanism proposed by Wong and Gao, as the reaction advances from the enzyme-substrate complex to the enzyme-product complex, the diethyl phosphate product coordinates to the β-metal ion in a mononuclear manner. This coordination scheme is coupled with a lengthening of the

metal-metal distance from 3.6 to 5.3 Å (*81*). Another proposed mechanism for PTE has been postulated by Ollis et al. In this mechanism, Ollis and colleagues proposed that the bridging hydroxide acts as a base and abstracts a proton from a water molecule that is appararently loosely coordinated to the M_{α}. This mechanism is distinctly different from the previous two mechanisms where the bridging hydroxide is proposed as the nucleophile (*82*). The three PTE mechanisms are presented in **Scheme 1.3**.

Scheme 1.2: Proposed mechanism for the reversible hydrolysis of dihydroorotate by DHO (*31*).

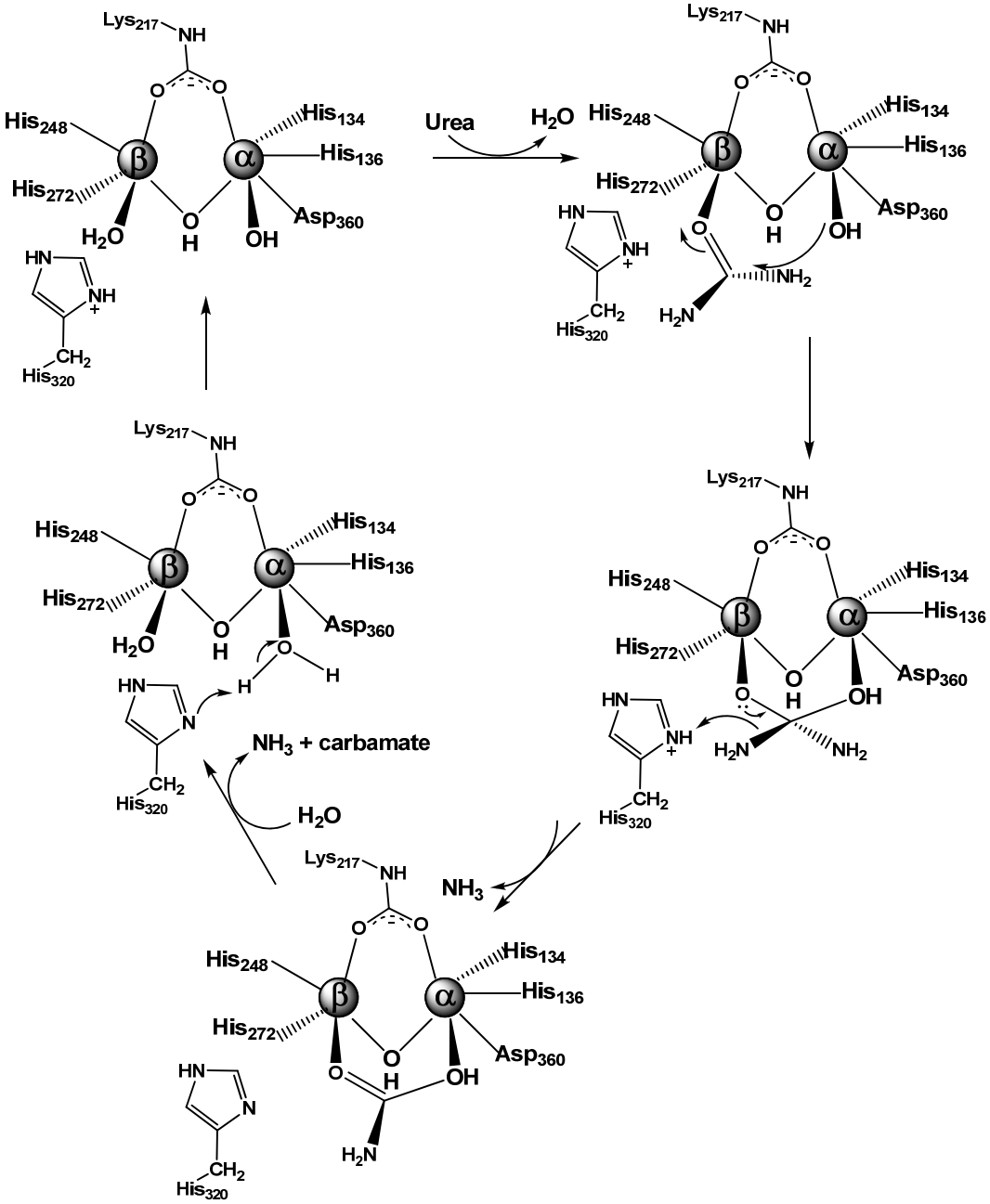


Scheme 1.3: The three mechanisms proposed for phosphotriesterase (*83*). A) A slightly modified version of the original mechanism proposed by Aubert et al. The bridging hydroxide nucleophilically attacks the phosphorus center of the substrate via a S_N2-like reaction. The bond to the phenol breaks and the resulting diethyl phosphate is bound within the active site as a bridging ligand between the two divalent cations. B) The Gao product complex where the diethyl phosphate is coordinated to the β-metal ion in a monodentate manner. C) The mechanism proposed by Ollis et al. The bridging hydroxide acts as a base that abstracts a proton from the nucleophilic M_α-bound water molecule. This deprotonation activates the attack on the phosphorus center of the substrate.

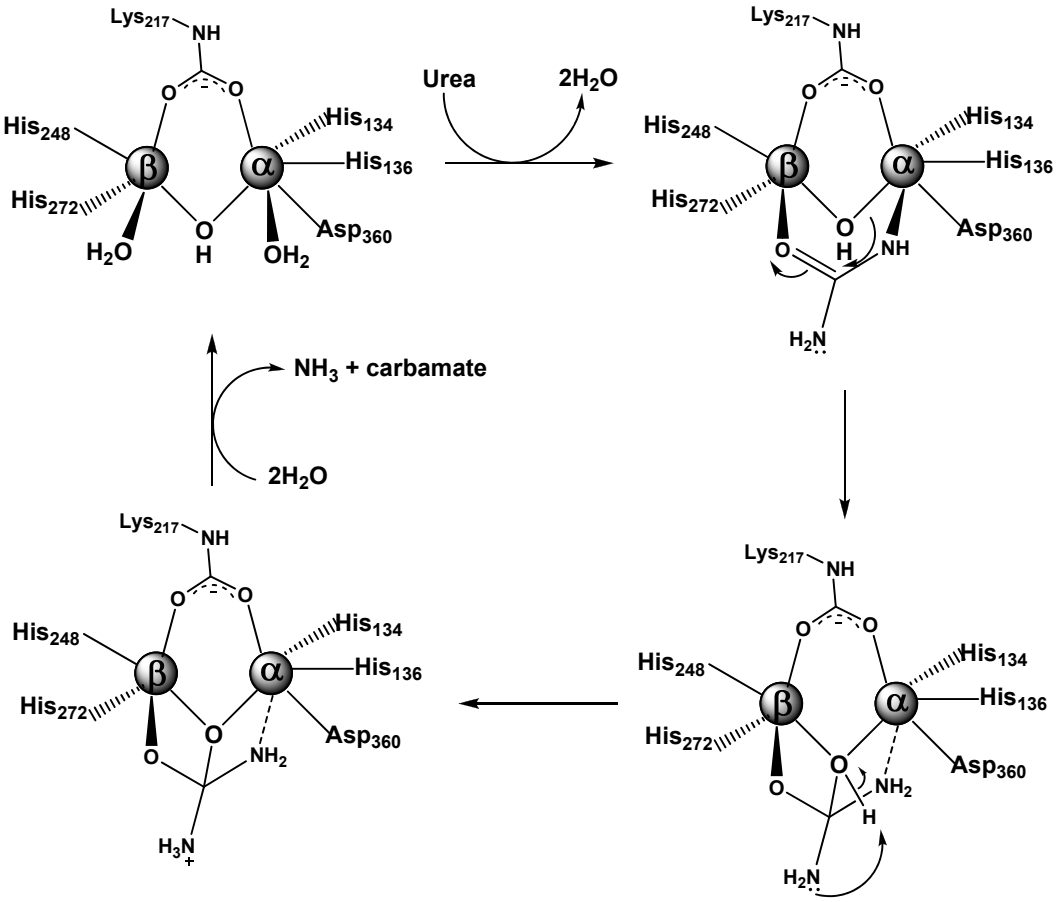


Another binuclear enzyme in which the mechanism has also been under great controversial debate is urease. There are three main reaction pathways that have been proposed for the enzyme-catalyzed decomposition of urea: the hydrolytic mechanism that involves the formation of a carbamate intermediate, the carbon dioxide mechanism in which urea is directly hydrolyzed to carbon dioxide and ammonia, and the elimination mechanism that involves the formation of a cyanate intermediate (*84*). The most commonly accepted mechanism, the carbamate mechanism, has been extracted from studies of bacterial urease from *K. aerogenes* and *B. pasteurii*. Two different variations of the carbamate mechanism have been proposed: the Karplus mechanism and the Benini mechanism. In the Karplus and Hausinger mechanism, urea binds to the active site and replaces a water molecule that is coordinating to M_β, while the water molecule (hydroxide) in the M_α site is retained. This Ni_α-bound OH acts as a nucleophile and attacks the carbonyl carbon of the urea molecule, which is further polarized by the Ni center. The reaction proceeds through a tetrahedral intermediate to form carbamic acid after the release of ammonia, which is assisted by proton transfer from the protonated His320 (*85*). A detailed illustration of this mechanism is shown in **Scheme 1.4**. An alternate mechanism for urease has been proposed by Benini et al, shown in **Scheme 1.5**. In this mechanism, urea replaces the two water molecules in the M_α and M_β sites and binds to the Ni active site center in a bidentate manner. The bridging hydroxide then acts as a nucleophile and attacks the urea carbonyl carbon atom, at the same time provides the proton that facilitates the release of ammonia (*32*).

Scheme 1.4: The mechanism of urease proposed by Karplus and Hausinger.

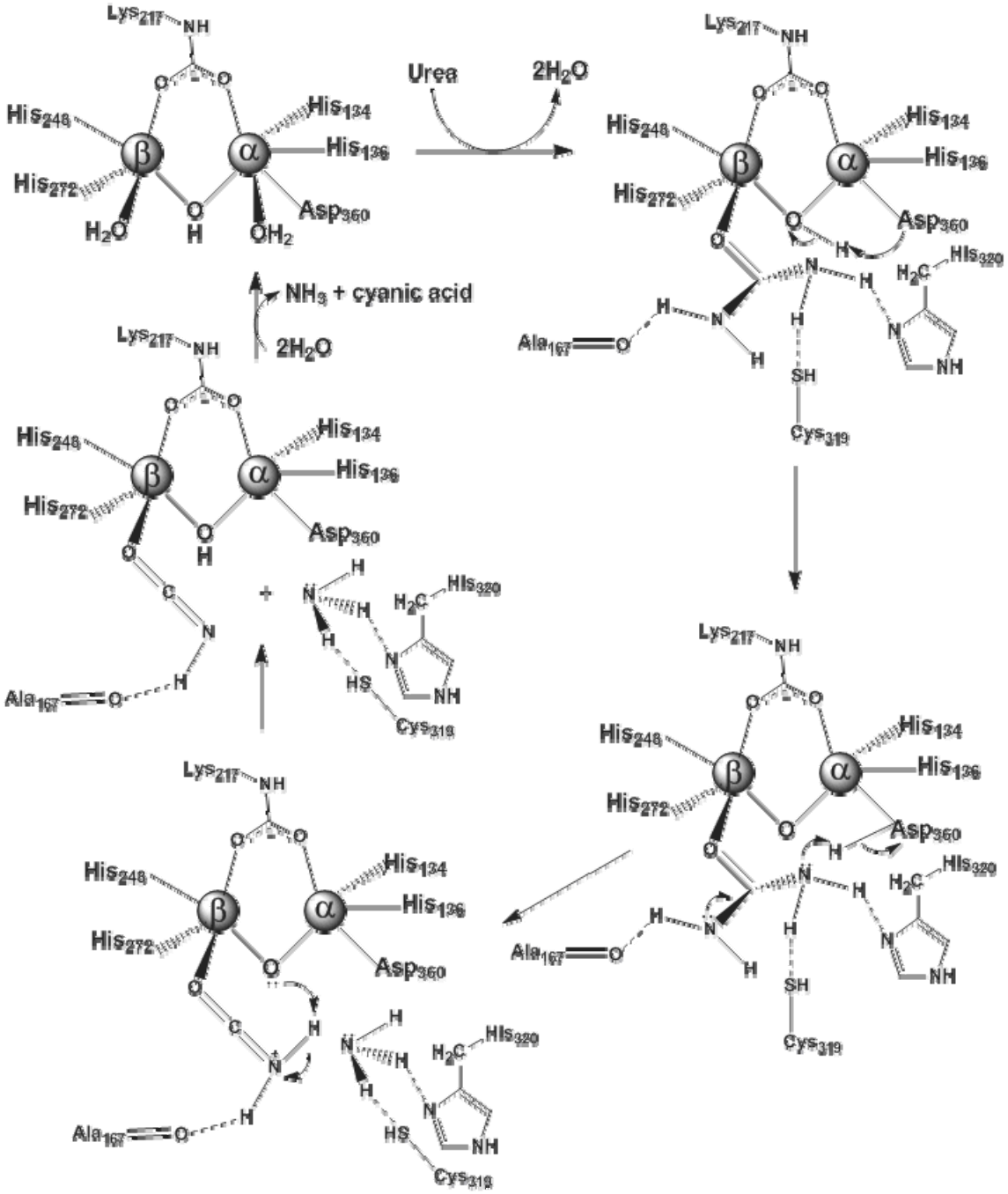


Scheme 1.5: The hydrolytic mechanism for urease proposed by Benini et al.



Computational methods have postulated the feasibility of a cyanic acid mechanism of urea decomposition, and this mechanism is also supported by kinetic studies of biomimetic complexes. The elimination mechanism is referred to as a "protein-assisted proton transfer" mechanism due to the major role of the protein in helping with the transfer of a proton between both NH_2 ends of urea. The first proposed protein-assisted proton transfer mechanism centers on Asp221 as the main residue relevant for catalysis. The second protein-assisted proton transfer, which can be described as "Asp-mediated proton transfer," involves the transfer of a proton from the bridging hydroxide to an NH_2 of urea. This process is facilitated by an active site Asp residue. The reactive nucleophilic di-Ni-coordinated oxygen is then capable of attacking the carbonyl carbon of urea to form a tetrahedral intermediate, as seen with the hydrolytic pathway. The result of this quantum mechanical analysis suggests that the active site of urease can undergo competitive elimination and hydrolytic mechanisms, leading to isocyanic acid or carbamic acid, respectively (*86*). A schematic example of a protein-assisted proton transfer mechanism is shown in **Scheme 1.6**.

Scheme 1.6: Schematic representation of a protein-assisted proton transfer mechanism in the elimination pathway of urea degradation.

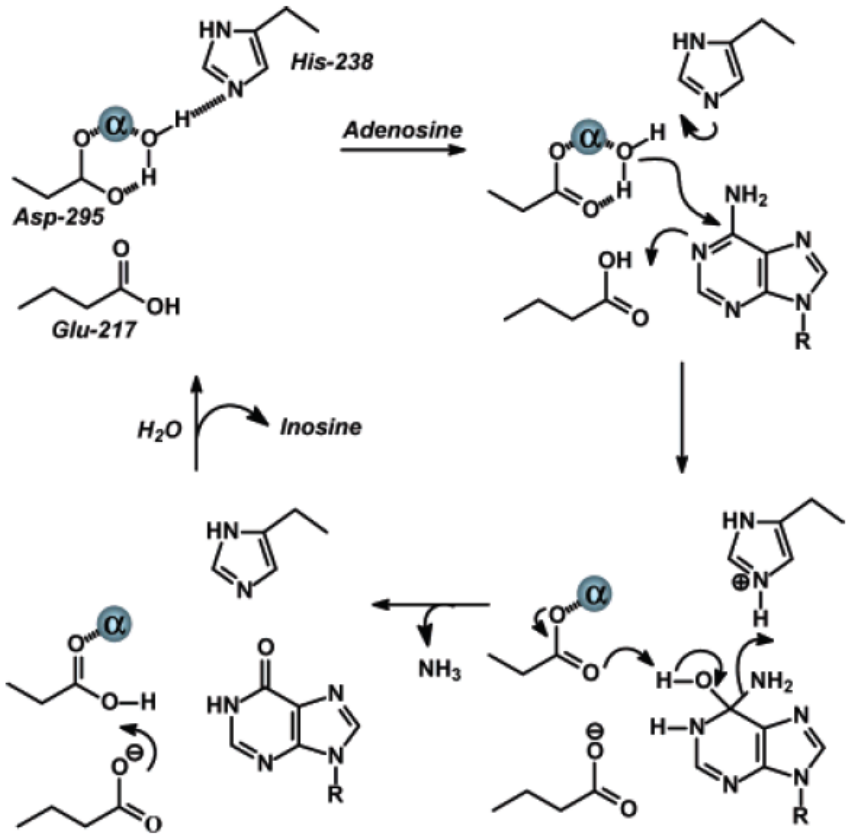


One other example of a binuclear metal center mechanism within the AHS is the mechanism proposed for the enzyme human renal dipeptidase (hRDP). The hRDP metal center is distinctly different from the binuclear centers described previously in that the metal ligands HxH from β-strand 1 is now HxD, and the conserved aspartate from strand 8 is no longer coordinating to the α-metal ion. Based on a crystal structure of the enzyme-cilastatin complex, Nitanai and colleagues suggested that the hydrolysis is initiated by a proton abstraction from the bridging water by the aspartate from strand 8. This deprotonation facilitates a nucleophilic attack on the carbonyl carbon of the amide bond by water, forming a tetrahedral adduct. The resulting oxyanion of the adduct is stabilized by the positive charge of the conserved His152 (*40*). This mechanism of intermediate stabilization is distinctive from other mechanisms proposed for enzymes of the AHS that have a binuclear metal center.

The reaction mechanisms for mononuclear enzymes of the AHS have been investigated to a great extent for adenosine deaminase (*46,56*) and *N*-acetyl-D-glucosamine-6-phosphate deacetylase (*70*). The reaction catalyzed by adenosine deaminase is an example of a nucleophilic aromatic substitution and the active site of ADA is a subtype of the mononuclear center where the metal ion is in the M_α position. In the mechanism proposed for ADA as outlined in **Scheme 1.7** (*31*), the histidine from strand 6 acts as a general base and abstracts the proton from the metal-bound water molecule. The resulting hydroxide attacks the aromatic ring forming a tetrahedral intermediate followed by a proton donation from a glutamate from strand 5 in a HxxE motif. Deprotonation of the carbinol intermediate by the conserved aspartate from

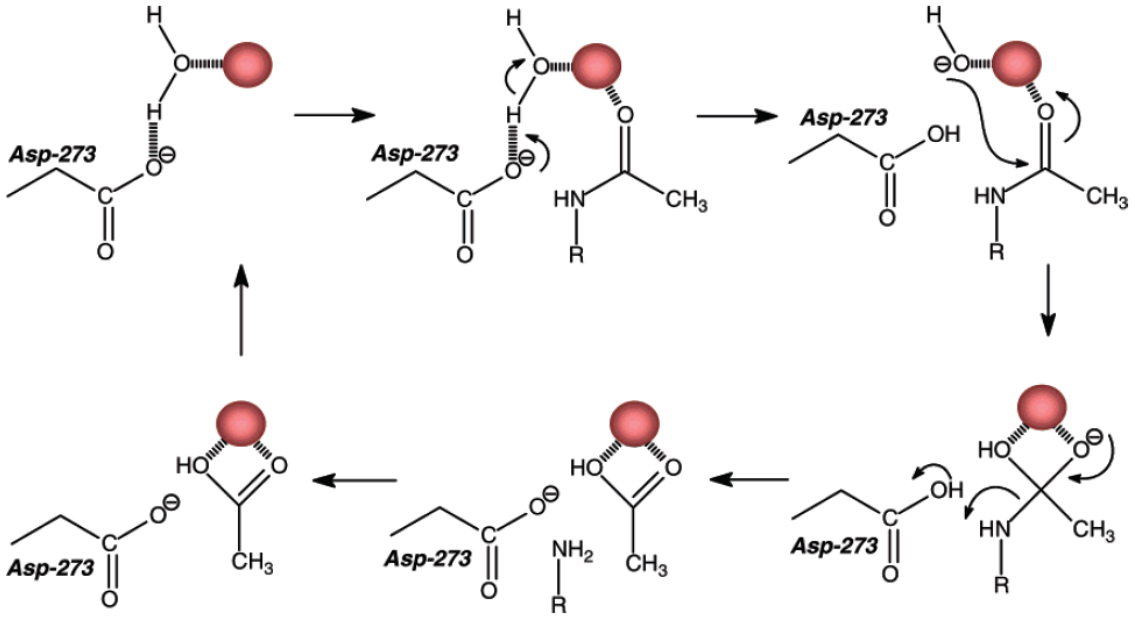
strand 8 and proton donation of the histidine from strand 6 assist the formation and release of the product (*46*).

Scheme 1.7: Proposed mechanism for the hydrolysis of adenine by ADA (*31*).



The structure and mechanism of AGD from *E. coli* was elucidated recently (*42*). AGD from *E. coli* catalyzes the deacetylation of *N*-acetyl-D-glucosamine-6-phosphate using a single metal ion positioned at the M_β site. The mechanism of AGD (**Scheme 1.8**), proposed by Hall et al (*70*), is initiated by the proton abstraction of the active site hydrolytic water (which is ligated to the metal ion) by the invariant aspartate from strand 8. The resulting hydroxide nucleophilicly attacks the carbonyl carbon of the substrate that is being polarized by the metal ion, forming a tetrahedral intermediate. The collapse of the tetrahedral intermediate is followed by a proton transfer from the aspartate of β-strand 8 to the amine leaving group of the product and ultimately the release of products. Apparently the mechanism utilized by NagA from *Bacillus subtilis* is different than that of the *E. coli* enzyme since it contains a binuclear metal center (*41*). These features highlight the significant diversity for the evolution of function within the amidohydrolase superfamily.

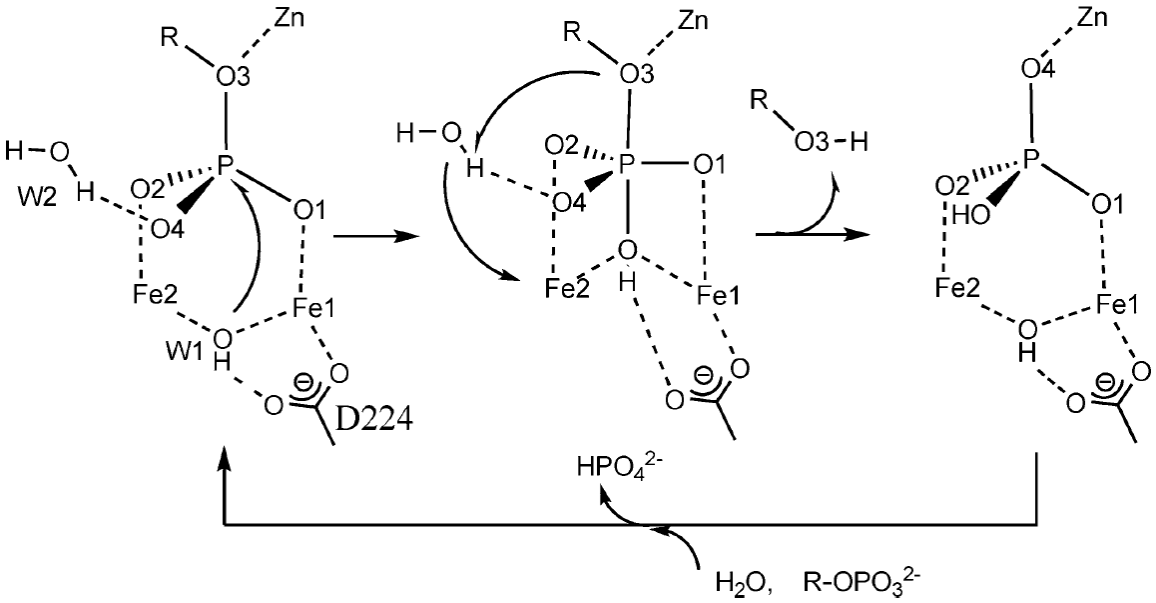
Scheme 1.8: Proposed mechanism for the hydrolysis of *N*-acetyl-glucosamine-6-phosphate deacetylase, an enzyme with a mononuclear metal center at the β-site (*70*).



The reaction mechanism for one of the most diverged members of the AHS, HolPase, has been proposed by Omi et al based on the crystal structure of the enzyme and an enzyme-substrate complex model (**Scheme 1.9**) (*53*). According to the mechanism, the substrate displaces a water molecule within the trinuclear cluster, and the substrate phosphate coordinates to the three metals. These interactions enhance the electrophilicity of the phosphorus atom. The bridging hydroxide acts as a nucleophile and makes an attack on the phosphorus center to form a trigonal pyramidal transition state, which is stabilized through the maintenance of the interactions of the phosphate and the metal center. This attack is facilitated by interactions between the hydroxide and

the two irons as well as between the hydroxide and Asp224. The geometrical constraint of the transition state makes the phosphorus atom move slightly toward the two irons. The P-O3 bond, which is being polarized by the Zn ion, loosens and the O3 gets protonated by a water molecule (W2 in **Scheme 1.9**). The histidinol is released from the active site, and another substrate enters to liberate the phosphate ion and start another round of catalysis.

Scheme 1.9: Reaction mechanism of histidinol-phosphate phosphatase from *T. thermophilus* (*53*). The formation and stabilization of a trigonal bipyramidal transition state is facilitated by interactions between the phosphate group of the substrate and the trinuclear metal center.

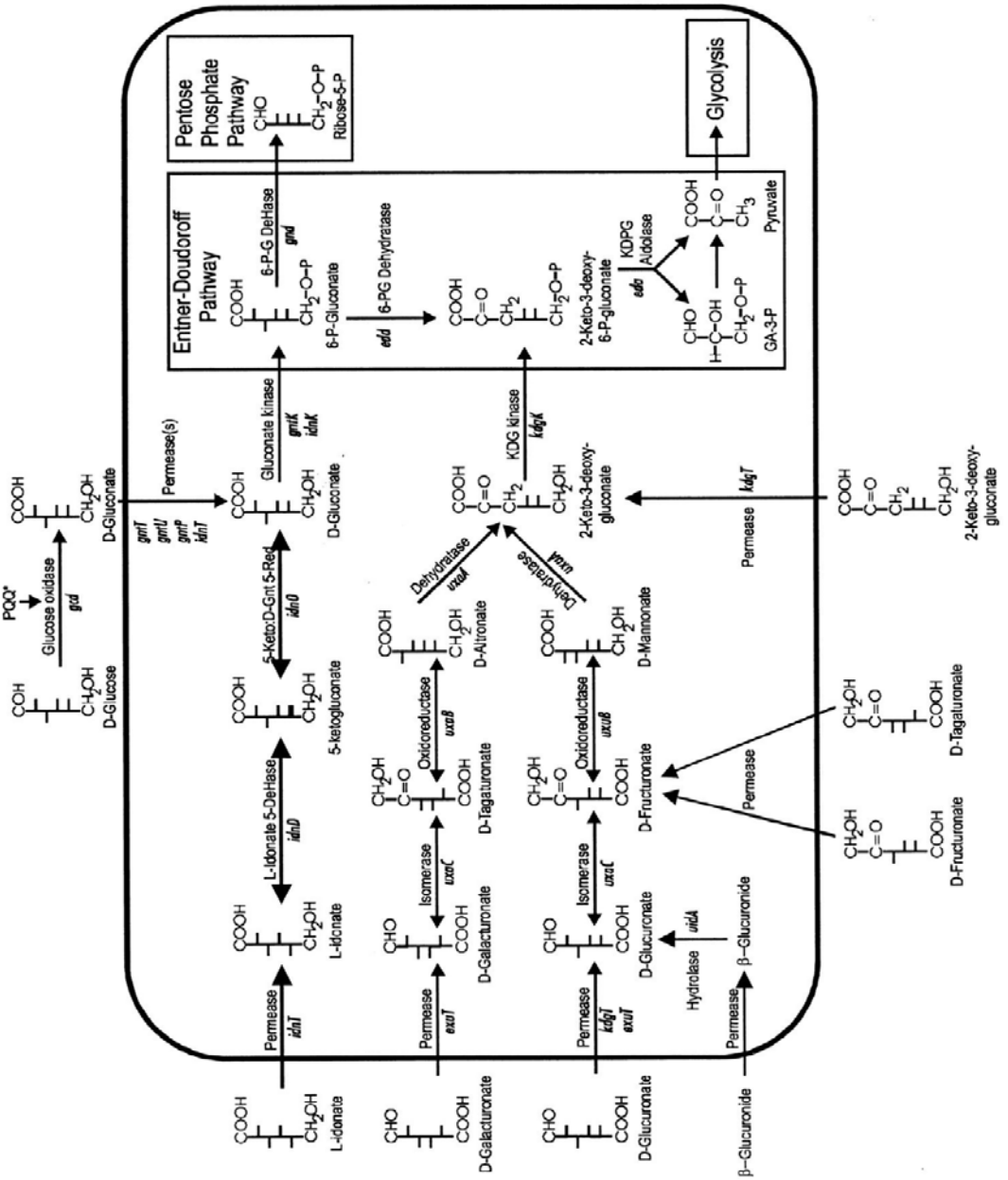


In this dissertation, the mechanisms of two members of the amidohydrolase superfamily were characterized and examined in detail using kinetics and X-ray crystallography. In the first part of the dissertation, the functional discovery of a second uronate isomerase (galacturonate isomerase) from *Bacillus halodurans* will be addressed. Unlike *E. coli* and *T. maritima*, the genomic context of the organism *B. halodurans* suggests that the pathways for the utilization of D-glucuronate and D-galacturonate are organized separately and involved separate enzymes for the isomerization of the respective hexauronate, the first reaction in the pathway. The X-ray crystal structure of galacturonate isomerase has also been solved, by the efforts of our collaborator Dr. S. Almo at the Albert Einstein College of Medicine, in the presence of substrate and intermediate mimics for the elucidation of the isomerization reaction mechanism utilized by enzymes of this family. The second part of the dissertation focuses on the mechanistic discussion of the enzyme uronate isomerase (URI) from *E. coli*. This enzyme is a unique member of the amidohydrolase superfamily due to its rare nonhydrolytic reactivity. Therefore characterization of this enzyme offers a striking example for the divergence of chemical reactivity of enzymes that share a common structural ancestor. The last part of this dissertation will be focused on the mechanistic details of a renal dipeptidase-like protein from *Streptomyces coelicolor*. The mechanism

for the hydrolysis of dipeptides by this enzyme will be presented, and kinetic and structural characterizations will be used for the corroboration of the proposed mechanism.

Uronate isomerase is considered an amidohydrolase enzyme based on sequence alignment. A crystal structure of the *T. maritima* URI has been obtained by the Joint Center for Structural Genomics (*43*). However, the structure contains a water molecule in place of the putative metal ion in the active site. The metal is predicted to be a zinc ion that is coordinating to the two histidines from β-strand 1 and the aspartate from strand 8. A sequence alignment of uronate isomerases from 36 different organisms reveals that uronate isomerases are all missing the histidine from strand 6. This enzyme represents the only well-characterized AH enzyme that does not contain a histidine at this location. Instead, it contains a conserved tryptophan at the end of strand seven. This enzyme is also unusual in that it is one of the enzymes from the AH superfamily that does not catalyze a hydrolysis reaction (*31*).

Uronate isomerase facilitates the conversion of D-glucuronate to D-fructuronate and of D-galacturonate to D-tagaturonate, the first step in the catabolism of hexuronate in *E. coli* (*87*). The parallel pathways for the break down of D-glucuronic and D-galacturonic acids converge with the formation of 2-keto-3-deoxy-gluconate. This intermediate is then phosphorylated and enters the Entner-Doudoroff pathway where it is further metabolized into glyceraldehyde-3-phosphate and pyruvate, two important metabolites of glycolysis. The pathways for the catabolism of hexuronate are outlined in **Scheme 1.10**. It was determined that the 72 *E. coli* strains isolated all possess the Entner-Doudoroff pathway and are capable of utilizing gluconate and glucuronate as energy sources (*87*). In humans, D-glucuronate is used for drug and toxin conjugation and is excreted in the urine and bile (*88*). D-glucuronate and D-galacturonate are found in high concentration in the intestinal tract and are believed to be substrates to support the growth of *E. coli* in the gut (*87*). Hence the activity of uronate isomerase in *E. coli* is essential for its survival in intestinal invironments.

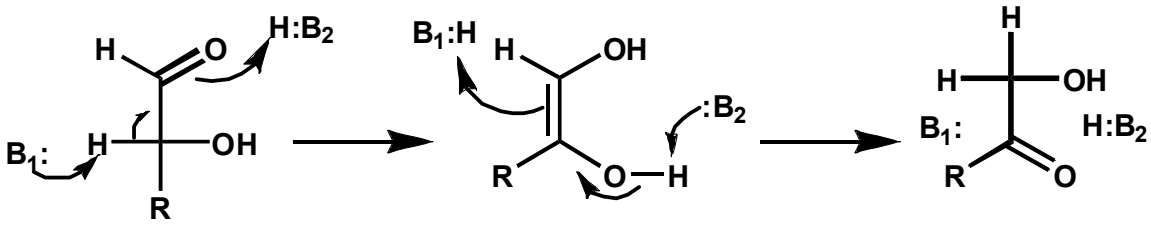


Scheme 1.10: Catabolic pathway of hexuronate in *E. coli* (87).

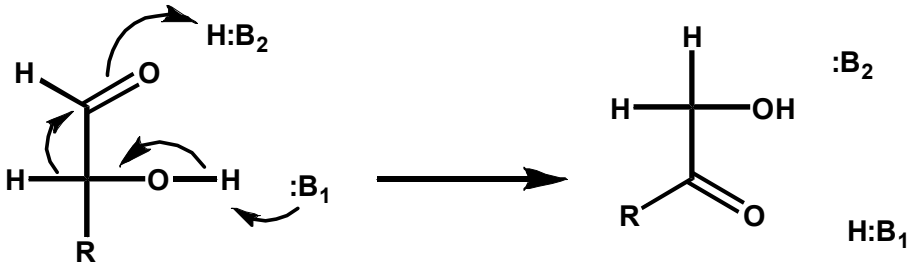
The reaction catalyzed by URI is that of the aldose-ketose isomerization reaction via a 1,2-hydrogen transfer. These enzymatic reactions generally proceed either through a proton transfer mechanism (**Scheme 1.11A**) such as phosphoglucose isomerase of the aldose-ketose isomerase family of enzymes (*89*) and triosephosphate isomerase (*90-94*) or a less common hydride transfer mechanism (**Scheme 1.11B**) as proposed for the metalloenzyme xylose isomerase (*95-97*). In the proton transfer mechanism of triosephosphate isomerase (TIM), Glu-165 abstracts the *pro-R* proton from C-1 of dihydroxyacetone phosphate (DHAP) and transfers it to the C-2 of glyceraldehyde 3-phosphate, while His-95 catalyzes the proton transfer between O1 and O2, as suggested by X-ray crystallography and mutagenesis studies (*90-94*). The mechanism of xylose isomerase involves two divalent cations that mediate the hydride transfer reaction. Evidence suggests that the enzyme binds the α form of the pyranose ring, and His-53 is involved in the protonation of O5 for ring opening. Following the ring-opening step, O2 is deprotonated by a metal-bound hydroxide ion accompanied by a shift in the position of the second metal (M2) towards the substrate. At a liganding distance to the substrate, M2 acts as an electron withdrawing force on the C1-O bond of substrate. This polarization results in a partial positive character on C-1 and thereby a more readily hydride transfer (*95-97*). Recently, the mechanism of URI from *E. coli* has been established as a proton transfer process based on the detection of substrate/solvent hydrogen exchange upon incubation with enzyme. The rate of deuterium incorporation at C2 of D-glucuronate was determined to be 0.018 s^{-1} compared to 196 s^{-1} for product formation (*98*).

Scheme 1.11: Aldo-keto isomerization of sugars. **A**) Proton transfer mechanism. **B**) Hydride transfer mechanism.

A



B



The isomerization of D-glucuronate to D-fructuronate requires a proton transfer from C2 of D-glucuronate to C1 at either the *re*- or *si*-face of D-fructuronate. The stereospecificity of this transfer was established by utilizing the enzymes D-mannonate dehydrogenase and D-mannonate dehydratase. The deuterium transferred from C2 of D-glucuronate is ultimately found at C6 of 2-keto-3-deoxy-D-gluconic acid. The detection of the deuterium at the *pro*-R position of 2-keto-3-deoxy-D-gluconic acid by NMR spectroscopy analyses confirmed that the proton is transferred to the *pro*-R position at

C1 of D-fructuronate (*98*). In chapter II of this dissertation, characterization of the mechanism of URI from *E. coli* will be discussed in detail using enzyme kinetics.

In the pursuit of protein function for unknown members of the AHS, one of the proteins in *B. halodurans* with an uncertain function is Bh0493. The sequence of Bh0493 is one of the most divergent protein sequences identified to date. A search of the DNA sequences available from NCBI reveals that the most closely related protein with a characterized catalytic function is Tm0064 from *Thermotoga maritima.* This protein is a uronate isomerase and a crystal structure of this enzyme has been deposited in the PDB (1J5S). However, the overall amino acid sequence identity between Bh0493 and Tm0064 is rather low. When these two proteins are compared to one another the sequence identity is less than 19%. Nevertheless, a sequence alignment between Bh0493 and other enzymes known to be uronate isomerases reveals that the residues suspected to be critical for the functioning of the isomerization reaction are conserved. These residues include the conserved HxH motif from β-strand 1, the aspartate from β-strand 8, and a WWF motif from β-strand 7. In addition, residues that are critical for functioning in the *E. coli* enzyme are conserved in the sequence of Bh0493 based on the sequence alignment of Bh0493 with other authentic uronate isomerases. In this part of the dissertation, cloning, expression, structure and functional characterization of Bh0493 will be reported. The X-ray structures of Bh0493 in the presence of substrate and intermediate mimics will aid in the elucidation of the isomerization mechanism employed by the uronate isomerase family of enzymes. The gene context within the genome of *B. halodurans* will also be examined to further obtain evidence for the

support of functional assignment of a gene encoding an enzyme in the catabolic pathways of hexuronate.

The third enzyme that will be characterized in this dissertation is a renal dipeptidase-like protein (Sco3058) from *Streptomyces coelicolor*. *Streptomyces* are recognized for their production of biologically active secondary metabolites and are responsible for almost two-thirds of all known natural antibiotics. The best genetically characterized strain is *Streptomyces coelicolor A3(2)*, which has become the preferred model organism in Streptomyces research (*99*).

The renal dipeptidase family of enzyme was first identified and characterized in the 1970s based on its physicochemical properties (*100*). In addition to a wide range of dipeptides, renal dipeptidase is also involved in the renal metabolism of glutathione and its conjugates and is the sole enzyme that is responsible for the hydrolysis of penem and carbapenem β-lactam antibiotics (*101-106*). The first crystal structure from the renal dipeptidase family is the enzyme from human (PDB: litq) (*40*). The quarternary structure of the hrDP is a homodimer with each polypeptide subunit consisting of 369 amino acid residues (42 kDa). The sequence of the hrDP shows a 75% identity to its mammalian homologs. From the crystal structure, the protein utilizes a HxD motif at strand 1 to bind to the M_{α} zinc ion instead of the conventional HxH. Moreover, the residue that bridges the two metals is a glutamate from β-strand 3 and not strand 4. The aspartate at the end of β-strand 8 is conserved, but it is no longer coordinating to the M_{α}. According to the X-ray structure, this aspartate residue appears to be interacting with the hydrolytic water via hydrogen bonding. This enzyme was also cocrystallized in the

presence of a known dipeptidyl inhibitor cilastatin (PDB entry code 1itu). The cilastatin-bound active site (**Figure 1.3**) suggested, as previously mentioned, that the hydrolytic water is activated by the conserved aspartate from β-strand 8. The deprotonated water then nucleophilically attacks the carbonyl carbon of the peptide bond to form a tedrahedral intermediate. It was also proposed that the conserved positively charged His-152 then acts as an oxyanion hole to stabilize the resulting negatively charged intermediate. This mechanism of hydrolysis is distintly different from the mechanism of characterized enzymes in the AHS in which the tetrahedral intermediate is stabilized by the solvent exposed M_β (*40*).

The first renal dipeptidase-like gene *acdp* from a prokaryote was cloned from *Acinebacter calcoaceticus* (*107*). This enzyme has a low sequence identity of 23% to the human renal dipeptidase, but the essential residues identified in the mammalian dipeptidases are all conserved. Unlike its mammalian homologs, the gene product of *acdp* could not hydrolyze the unsaturated dipeptide glycyldehydrophenylalanine; instead the protein showed substrate specificity towards various dipeptides with a preference for dipeptides contaiting a D-amino acid at the C-terminus (*107*). The protein Sco3058, on the other hand, contains a 45% and 22% sequence identities to the hrDP and *acdp*, respectively. Substrate specificity analysis of this protein indicates that the enzyme is promiscuous towards a wide range of dipeptides containing L- and D-amino acids with a preference for L-Xaa D-Xaa substrates. In this dissertation, kinetic and structural properties of the enzyme Sco3058 from *Streptomyces coelicolor* will be discussed in detail.

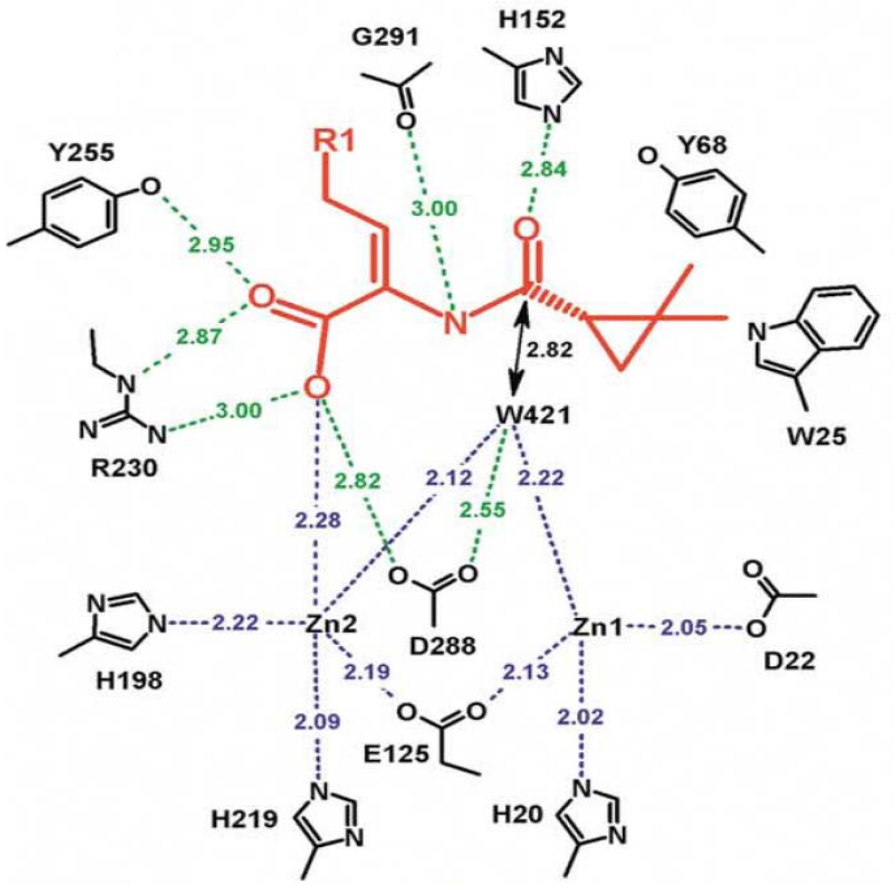


Figure 1.3: Active site of hRDP with bound cilastatin (in red) (*40*).

CHAPTER II

AT THE PERIPHERY OF THE AMIDOHYDROLASE SUPERFAMILY: BH0493 FROM *BACILLUS HADODURANS* CATALYZES THE ISOMERIZATION OF D-GALACTURONATE TO D-TAGATURONATE*

The identification and functional assignment of enzymes encoded within completely sequenced genomes is a difficult and demanding problem. According to some estimates, approximately 40% of newly sequenced genes have an *unknown*, *uncertain*, or *incorrect* functional assignment (*2*). The extent of the catalytic diversity inherent within metabolic and catabolic pathways will not be fully understood until the functional annotations of these newly sequenced genes have been addressed. The absence of a comprehensive understanding of the metabolic landscape has therefore motivated the development of new methodologies for the assignment of function for enzymes that catalyze reactions with ambiguous substrates. These emerging strategies include phylogenetic profiling 108, *109*), recognition of domain fusions and gene clusters (*87*,*110-113*), and other computational approaches (*88*). Most of the commonly used techniques require sequence and/or structural comparisons to well-characterized

*Reprinted with permission from "At the Periphery of the Amidohydrolase Superfamily: Bh0493 from Bacillus Halodurans Catalyzes the Isomerization of D-galacturonate to D-tagaturonate" by T. T. Nguyen, S. Brown, A. A. Fedorov, E. V. Fedorov, P. C. Babbitt, S. C. Almo, and F. M. Raushel, 2008. *Biochemistry*, 47, 1194-1206, Copyright 2008 by Journal of the American Chemical Society.

homologues to make high quality functional predictions. However, these attempts at functional characterization can be problematic when the sequence or structural similarity is remote (*2*).

Members of the amidohydrolase superfamily (AHS) have been identified in every organism sequenced to date (*20*, *31*). This enzyme superfamily catalyzes predominantly hydrolytic reactions where a water molecule is activated by one or two divalent metal ions embedded within the active site (*31*). All of the enzymes within the AHS adopt a $(\beta/\alpha)_8$-barrel structural fold with an active site that resides at the C-terminal end of the β-barrel (*21*, *31*). One of the most divergent members of the amidohydrolase superfamily is uronate isomerase (URI). This enzyme family exhibits few statistically significant sequence links to any other member of the AHS (**Figure 2.1**), and structure-based comparisons are required to link this family to other families in the superfamily.

The URI from *Escherichia coli* (UxaC) catalyzes the isomerization of D-glucuronate to D-fructuronate and of D-galacturonate to D-tagaturonate. These isomerization reactions represent the first step in the separate metabolic pathways for the utilization of D-glucuronate and D-galacturonate in bacteria. The isomerized products are subsequently reduced to D-mannonate and D-altronate by UxuB and UxaB, respectively, and then dehydrated by UxuA and UxaA to 2-keto-3-deoxy-D-gluconate. The entire metabolic pathway for the metabolism of these sugars in *E. coli* is summarized in **Scheme 2.1**.

We have begun to delineate the extent of the structural and functional diversity inherent within the amidohydrolase superfamily of enzymes. One of the most divergent

protein sequences identified to date from this investigation has emerged from the genome of the alkaliphilic bacterium *Bacillus halodurans* (gi:15613056). This protein (Bh0493) was tentatively assigned as a member of the AHS based upon a weak sequence similarity (<19%) with the structurally characterized uronate isomerase from *Thermotoga maritima* (Tm0064, gi: 15642839). Providing additional evidence for this assignment, the putative active site residues that originate from the ends of β-strands 1 and 8 are apparently conserved in the *B. halodurans* enzyme. In this chapter, we report the cloning, expression, structure and functional characterization of Bh0493. The X-ray structure of Bh0493, determined with a single zinc ion in the active site, confirms that this enzyme is a member of the AHS. This enzyme will catalyze the isomerization of D-galacturonate to D-tagaturonate in a metabolic pathway that has diverged from the isomerization of D-glucuronate. We have also functionally characterized Bh0705 from B. halodurans (gi: 15613268), the amino acid sequence that more closely resembles that of a canonical uronate isomerase. Crystal structures of Bh0493 in the presence substrate and intermediate mimics will aid in the elucidation of the isomerization mechanism utilized by this family of enzyme.

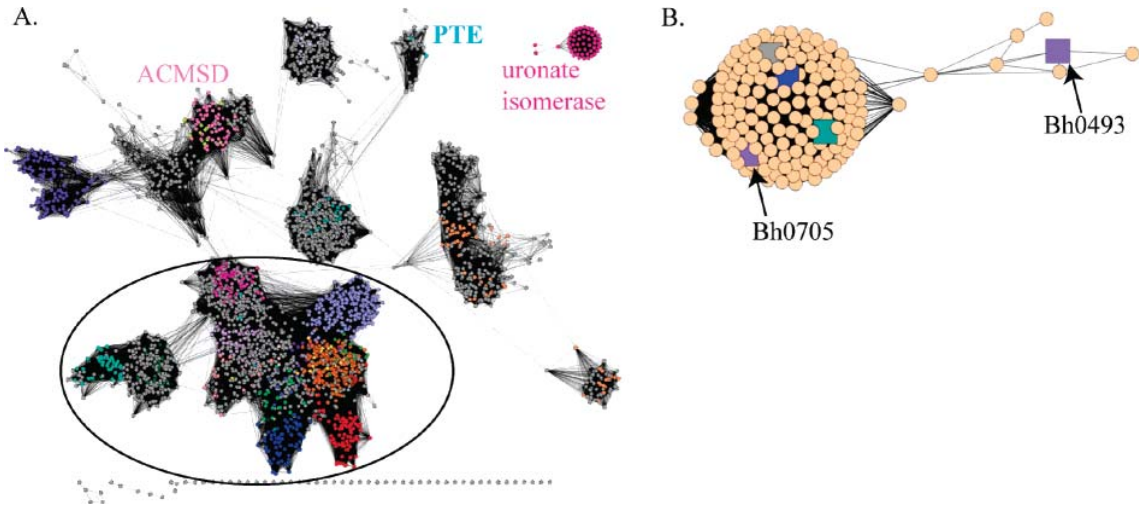
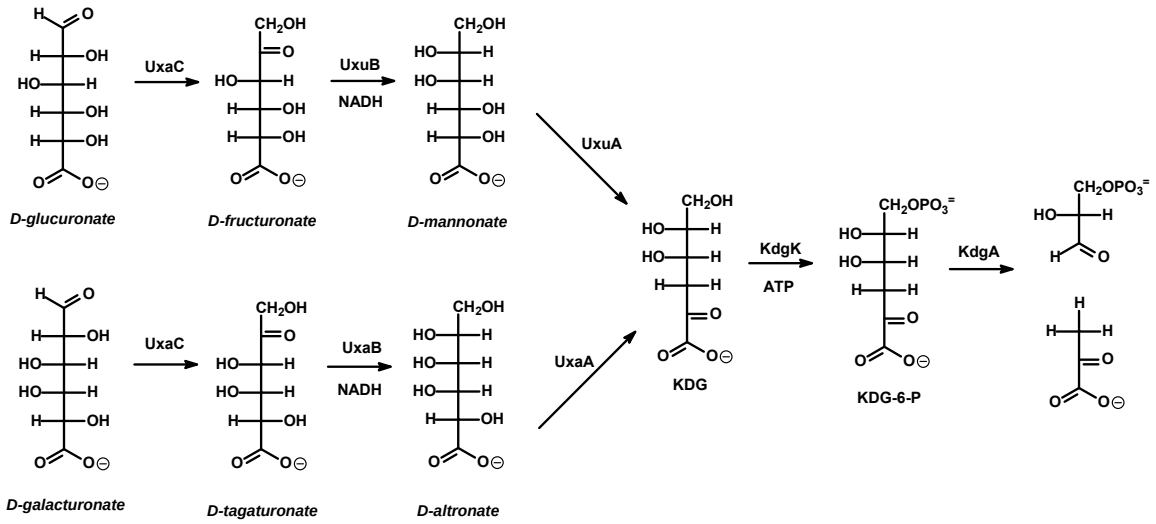


Figure 2.1: Network representation of the sequence relationships in the AHS. Each node in the network represents a single sequence in the uronate isomerase-like sequence set, and each edge represents the pairwise connection between two sequences with the most significant BLAST *E*-value (better than the cutoff) connecting the two sequences. Lengths of edges are not meaningful except that sequences in tightly clustered groups are relatively more similar to each other than sequences with few connections. (A) An 80% ID filtered set of AHS sequences from the SFLD showing connections with a BLAST *E*-value more significant than or equal to 10^{-4}. Sequences are colored by SFLD subgroup. Sequences with characterized functions originally identified as part of the AHS by Holm and Sander are contained within the circle (except for phosphotriesterase, (PTE)). Additional groups discussed in the text are indicated with labels. (B) An unfiltered set of uronate isomerase-like sequences at a BLAST *E*-value more significant than or equal to 10^{-10}. Sequences that have been experimentally characterized as uronate isomerase and/or have an X-ray crystal structure are colored as follows: gray, Tm0064; lavender, Bh0705 and Bh0493; blue Cc1490; and green, b3092.

Scheme 2.1: Metabolic pathways for the metabolism of D-glucuronate and D-galacturonate.



MATERIALS AND METHODS

Materials. D-Glucuronate and NADH were purchased from Sigma-Aldrich. D-galacturonate was obtained from Acros. The inhibitors, D-arabinaric acid (**1**) and the monohydroxamate derivative of D-arabinaric acid (**2**) were synthesized by Dr. Yingchun Li. The DNA sequencing reactions were conducted by the Gene Technologies Laboratory at Texas A&M University. Determination of the metal content of the purified proteins was done using ICP-MS. Gel filtration and anion exchange columns were purchased from GE HealthCare. Restriction enzymes and T4 DNA ligase were obtained from New England Biolabs. The pET30 vector was acquired from Novagen and the Platinum Pfx DNA polymerase was purchased from Invitrogen. The Wizard

Miniprep DNA purification kit was obtained from Promega and the genomic DNA of *B. halodurans* was acquired from ATCC (BAA-125D).

Cloning and Overexpression of Bh0705 and Bh0493. The gene *uxa*C (Bh0705, gi: 15613268) from *B. halodurans* was amplified by PCR with two primers, 5'-GCAG-GAGCATTAATGACGAATTTTCTTTCAGAAGACTTCCTCTTAATGAACGAAT-ACG-3'containing an *Ase*I restriction site and 5'-GCGCAAGCTTTCATCATAGTCG-AACAGTCTCCTTCATGTCG-3' containing a *Hin*dIII restriction site. The *uxa*C gene was inserted into a pET30a (+) vector that had been digested with *Ase*I and *Hin*dIII. The protein was expressed and purified from BL21(DE3) star cells (Novagen) that had been transformed with the pET30a (+) plasmid containing the *uxa*C gene from *B. halodurans*. The gene for Bh0493 (gi: 15613056) was cloned with two primers, 5'-GCAGGAGCCA-TATGTCCATAAACAGTAGGGAAGTGTTACGGG-3' and 5'-CGCGGAATTCTTA-TTACGTTTGCTGCTCCACCTTCACTGATGTGACG-3' containing the restriction sites for *Nde*I and *Eco*RI, respectively. The gene product was ligated into a pET30a (+) vector and expressed in BL21(DE3) star cells. The same protocol was utilized to purify Bh0705 and Bh0493. A single colony of the cells containing the gene of interest was inoculated and incubated overnight in 5 mL of LB medium containing 50 μg/mL kanamycin at 37 ºC. The overnight culture was used to inoculate 1 L of LB medium containing 50 μg/mL kanamycin. The cells were grown at 30 ºC and induced with 0.4 mM IPTG when the cell density at A_{600} reached ~0.4-0.6. After 24 hours, the cells were harvested at 4,000*g*. The cell pellet was resuspended in 5 mL of 50 mM HEPES, pH 8.0, (buffer A) for every gram of cell paste and disrupted by sonication. The nucleic

acids were removed by the dropwise addition of protamine sulfate to a final concentration of 2% w/v and subsequent centrifugation at 14,000*g*. The protein was precipitated between 50-70% saturation of ammonium sulfate, and the protein pellet was resuspended in a minimum amount of buffer A. The protein was loaded onto a pre-equilibrated Superdex 200 gel filtration column and eluted with buffer A. The fractions containing the protein of interest were pooled based on activity, loaded onto a Resource Q column, and eluted with a 0-30% gradient of 1 M NaCl in 20 mM HEPES, pH 8.0 (buffer B). The fractions containing the desired protein were pooled and the purity of the protein was determined by SDS-gel electrophoresis. The purified enzyme was sterile-filtered and stored at 4º C.

Cloning, Expression, and Purification of UxaB. The *uxa*B gene from *E. coli.* (gi: 49176119) was amplified by PCR with two primers, 5'-GCAGGAGCCATATGAAA-ACACTAAATCGTCGCGATTTTCCCGGTGC-3' and 5'-GCGCAAGCTTTTATTA-GCACAACGGACGTACAGCTTCGCGCATCCCTTTTTCG-3', containing *Nde*I and *Hin*dIII sites, respectively, and the resulting fragment was inserted into the *Nde*I and *Hin*dIII sites of the pET30 vector. The protein was expressed in the BL21(DE3) strain of *E. coli*. A single colony was used to inoculate a 5-mL overnight culture of LB medium containing 50 μg/mL kanamycin. The 5 mL overnight culture was inoculated into 1 L of LB containing the same concentration of kanamycin. The cells were grown at 30 ºC, induced with 0.4 mM IPTG at an A_{600} ~0.6, and incubated overnight. The cells were collected by centrifugation at 4,000*g*, and the cell pellet was resuspended in buffer

A. The cells were disrupted by sonication and centrifuged at 14,000*g*. The protein was soluble and the supernatant solution was used directly in the enzymatic assays.

Cloning, Expression, and Purification UxuB. The *uxu*B gene (gi: 26111644) from *E. coli* that encodes D-mannonate dehydrogenase (MDH) was cloned into a pET30 expression vector. The gene was transformed into BL21(DE3) and the cells grown in TB broth at 37 °C. D-Mannonate dehydrogenase was purified using the same methods previously described for the isolation of uronate isomerase, but no metal was added to the purification buffers. The cell extract was made 40-60% of saturation with ammonium sulfate, and the resuspended pellet was subjected to fractionation with the aid of a Superdex 200 gel filtration column. Fractions containing D-mannonate dehydrogenase were pooled, quick-frozen, and stored at -80 °C.

Sequence Analysis. The uronate isomerase sequence clusters were organized and created in collaboration with Patricia C. Babbitt at the School of Pharmacy, University of California, San Francisco, CA.

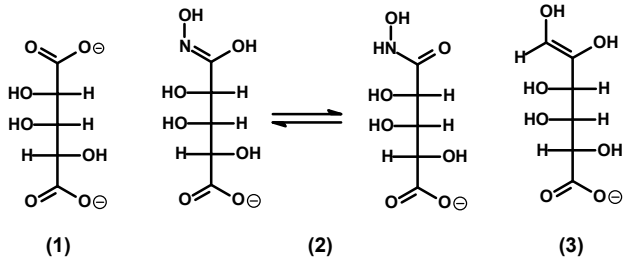
Crystallization of Bh0493. The structure of Bh0493 was solved in collaboration with the Steven C. Almo group at the Department of Biochemistry, Albert Einstein College of Medicine, at Bronx, New York. The PDB entry codes are 2q6e for the Se-Met-substituted enzyme and 2qo8 for the wild-type complexed with Zn.

Enzyme Activity and Substrate Specificity. The isomerase activities of Bh0705 and Bh0493 were measured with D-glucuronate and D-galacturonate. The enzymatic reactions with D-glucuronate and D-galacturonate were coupled to the reduction of D-fructuronate and D-tagaturonate with the appropriate dehydrogenase. These assays

monitor the oxidation of NADH to NAD^+ and the change in absorbance was measured spectrophotometrically at 340 nm. The standard assays contained 50 mM HEPES, pH 8.0, 0.2 mM NADH, ~ 2 μM of either D-mannonate dehydrogenase or D-tagaturonate dehydrogenase, and various substrate concentrations in a final volume of 250 μL.

Inhibition Studies. The inhibition of the reactions catalyzed by Bh0705 and Bh0493 was quantified using two compounds that have previously been shown to inhibit the isomerization of D-glucuronate by uronate isomerase from *E. coli*. D-Arabinaric acid (**1**) and (2S,3R,4S)-4-(hydroxycarbamoyl)-2,3,4-trihydroxybutanoate (**2**) were tested as inhibitors using standard assay conditions with D-glucuronate as the substrate and D-mannonate dehydrogenase as the coupling enzyme. The structures of these compounds are presented in **Scheme 2.2**.

Scheme 2.2: Structures of D-arabinarate (1), hydroxamate derivative (2) and *cis*-enediol intermediate.



Data Analysis. The kinetic parameters, k_{cat} and k_{cat}/K_m, for the enzyme Bh0705 were determined by fitting the initial velocity data to equation 2.1, where v is the initial velocity, E_t is the total enzyme concentration, k_{cat} is the turnover number, A is the substrate concentration, and K_a is the Michaelis constant. The double reciprocal plots for Bh0493 were nonlinear (concave down) and fitted to equation 2.2 (*114*). In this equation k_1 is the maximum velocity at low substrate concentration and the sum of k_1 and k_2 is the maximum velocity at high substrate concentration. The apparent Michaelis constants at low and high substrate concentrations are K_1 and K_2, respectively. Competitive inhibition patterns by the compounds that mimic the putative *cis*-enediol intermediate were fit to equation 2.3, where I is the inhibitor concentration and K_i is the slope inhibition constant.

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

$$v / E_t = (k_1[A] / (K_1 + [A])) + (k_2[A]) / (K_2 + [A])) \quad (2.2)$$

$$v / E_t = (k_{cat}[A]) / (K_a(1 + (I / K_i)) + [A]) \quad (2.3)$$

RESULTS

Bioinformatics. As of July 2009, approximately 190 uronate isomerase like sequences were identified in the NCBI nr database. All, except two archaeal sequences, are from bacteria, reflecting their role in bacterial glucuronic and galacturonic acid metabolism. As shown in **Figure 2.2**, these sequences cluster into three groups. Group 1 contains one of the two functionally characterized uronate isomerases from *B. halodurans* (Bh0705). Group 2 contains the structurally characterized uronate isomerase from *C. crescentus*. Group 3 contains the second functionally characterized uronate isomerase from *B. halodurans* (Bh0493).

Purification and Properties of Bh0705. The gene for the enzyme that encodes for *Bh0705* was expressed in BL21(DE3) cells. The protein Bh0705 was sparingly soluble after the cells were disrupted by sonication. Approximately 14 mg of purified protein was obtained from 14 grams of wet cell paste. The metal content of the purified protein was determined to contain 0.9 equivalents of Zn^{2+} per subunit using ICP-MS. SDS-PAGE analysis revealed the presence of a single band at approximately 55 kDa for Bh0705. This value is in agreement with the reported gene sequence.

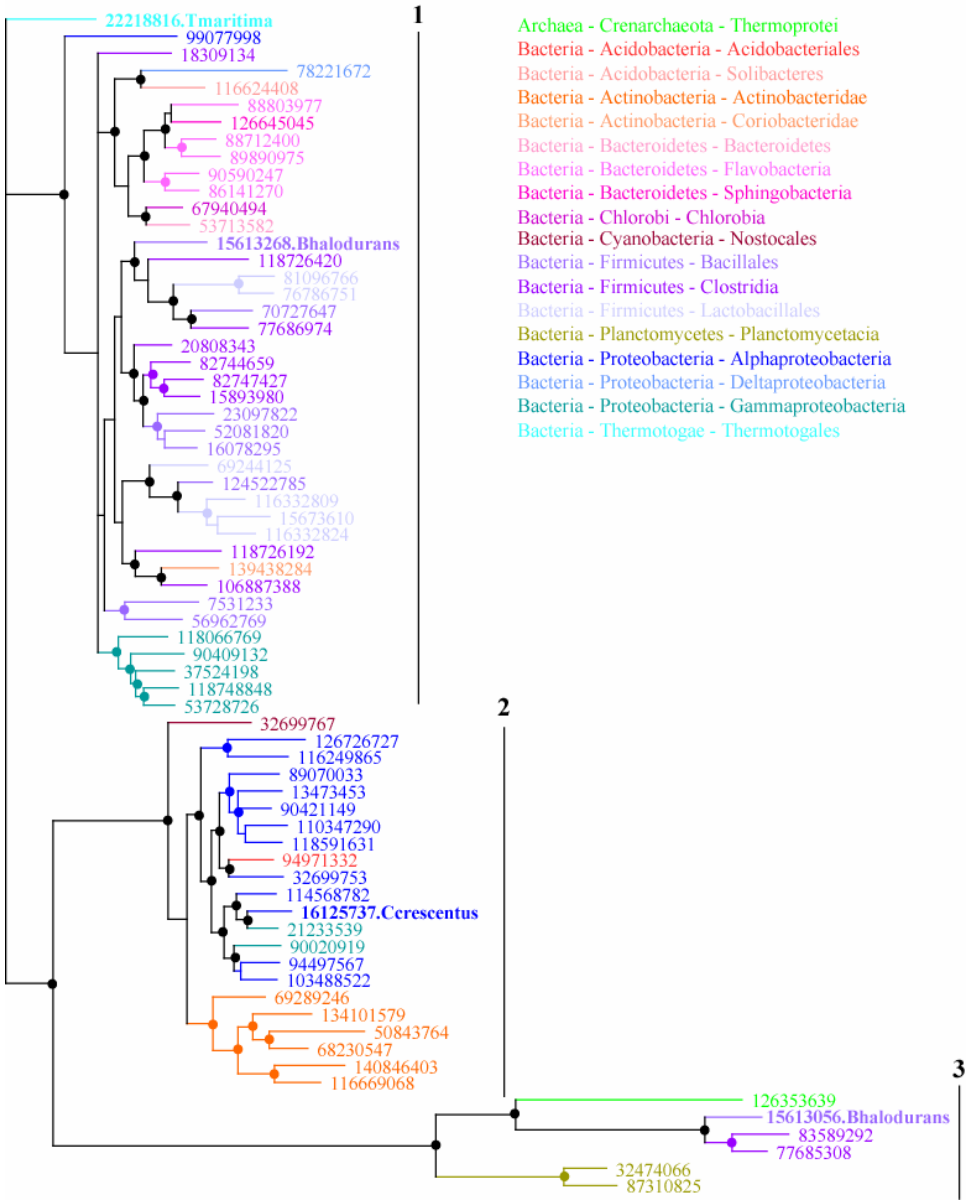


Figure 2.2: Bayesian phylogenetic tree of the proteins in the URI-like sequence set. A representative set of sequences was selected by filtered the complete URI sequence set to 70% identity. Sequences are listed according to their NCBI gi number. The crystallized URI sequences from *T. maritima*, *B. halodurans*, and *C. crescentus* are indicated with species abbreviations. Branch confidence values greater than 0.9 are indicated with circles. The three sequence clusters discussed in the text are indicated with numbers.

Characterization of Bh0705. The catalytic activity for Bh0705 was determined with D-glucuronate and D-galacturonate as substrates. This enzyme was found to isomerize either of these compounds to their respective ketoacid sugar products and the kinetic constants are presented in **Table 2.1**. The protein Bh0705 was found to be more specific for the isomerization of D-glucuronate compare to D-galacturonate since the ratio of k_{cat}/K_m for these two compounds is approximately 100. The two active site directed inhibitors, D-arabinaric acid (**1**) and the hydroxamate derivative (**2**), that were designed to mimic the proposed *cis*-endiol intermediate (**3**) were found to be potent competitive inhibitors for the isomerization of D-glucuronate. The inhibition constants are provided in **Table 2.2**.

Table 2.1: Kinetic parameters for Bh0493 and Bh0705 with D-glucuronate and D-galacturonate

enzyme	substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
Bh0493	D-glucuronate	5.2 ± 0.3	0.05 ± 0.01	(1.1 ± 0.3) x 10^5
		3.6 ± 0.3	15 ± 5	(2.5 ± 0.8) x 10^2
	D-galacturonate	2.4 ± 0.1	0.05 ± 0.01	(4.7 ± 0.8) x 10^4
		0.9 ± 0.1	12 ± 5	72 ± 30
Bh0705	D-glucuronate	9.7 ± 0.1	0.33 ± 0.01	(2.9 ± 0.10) x 10^4
	D-galacturonate	1.2 ± 0.01	3.3 ± 0.1	(3.6 ± 0.12) x 10^2

Kinetic constants were determined at 30 ºC, pH 8.0 from a fit of the data to equation 2.1 or 2.2.

Table 2.2: Inhibition constants with mimics of *cis*-enediol intermediate[a]

enzyme	inhibitor	K_i (μM)
Bh0493	D-arabinaric acid (**1**)	$(5.5 \pm 0.3) \times 10^{-2}$
	Monohydroxamate derivative (**2**)	2.1 ± 0.1
Bh0705	D-arabinaric acid (**1**)	0.40 ± 0.02
	Monohydroxamate derivative (**2**)	18 ± 2

[a]These constants were obtained at pH 8.0, 30 ºC from a fit of the data to equation 2.3.

Purification and Properties of Bh0493. The plasmid containing the gene encoding protein Bh0493 was transformed into BL21(DE3) cells, and the protein was expressed by induction with IPTG. The protein was soluble and significant amounts of protein were obtained and purified to homogeneity. Using ICP-MS, the protein was found to contain 0.5 equivalents of Zn^{2+} per subunit. SDS-PAGE indicated that the subunit molecular weight is approximately 49 kDa, which corresponds well to the calculated molecular weight of Bh0493 based on the DNA sequence.

Characterization of Bh0493. The enzymatic activity of Bh093 was tested using D-glucuronate and D-galacturonate as substrates. The enzyme catalyzes the isomerization of D-glucuronate to D-fructuronate and D-galacturonate to D-tagaturonate. The double-reciprocal plots for both substrates were biphasic (concave downward), and thus, the initial velocity data were fitted to eq 2.2. The double-reciprocal plots for the isomerization of D-glucuronate and D-galacturonate are shown in **Figure 2.3**. The kinetic constants at high and low concentrations of substrate are provided in **Table 2.1**.

The structural mimics of the proposed cis-enediol intermediate were found to be competitive inhibitors for the isomerization of D-glucuronate when the assays were conducted at low substrate concentrations. The kinetic constants are presented in **Table 2.2**.

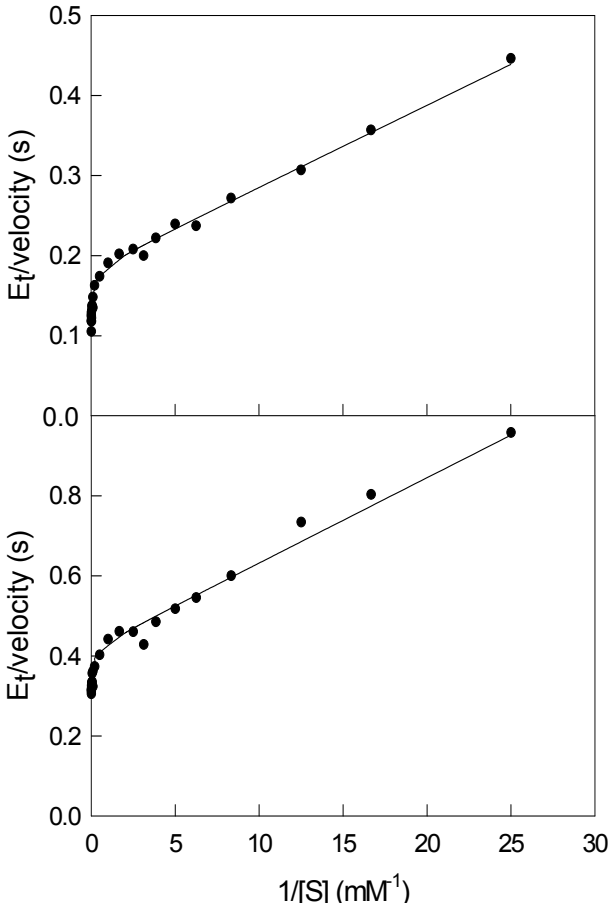


Figure 2.3: Double reciprocal plot for the isomerization of D-glucuronate (top) and D-galacturonate (bottom) by Bh0493. The initial velocities were fit to equation 2.2.

Structure of Bh0493. Two crystal forms of Bh0493 were solved and refined. A tetragonal crystal form was solved to a resolution of 2.4 Å and triclinic crystal form at a resolution of 2.0 Å. Both crystal forms contain similar Bh0493 trimers. There is one protein trimer in the asymmetric unit of the tetragonal form and there are four protein trimers in the asymmetric unit of the triclinic crystal form. Three molecules in every trimer are connected by non-crystallographic three-fold axis. Several ions were observed that sit on the local three-fold axes of every trimer. The Na^+ ion on the local three-fold axis in the tetragonal crystal form is coordinated by six water molecules, connected to protein side chains. This crystal form was crystallized from solutions containing NaCl. The triclinic Bh0493 crystal form had no NaCl in the crystallization solutions and the analogous metal position on the local three fold axis was occupied by Zn^{2+} as shown in **Figure 2.4**. The tetragonal crystal form additionally contains a Cl^- ion that sits on the local three-fold axis of the trimer and is surrounded by the side chains of Lys-278 from three adjacent protein molecules. All of these ions additionally stabilize the internal structure of the trimers.

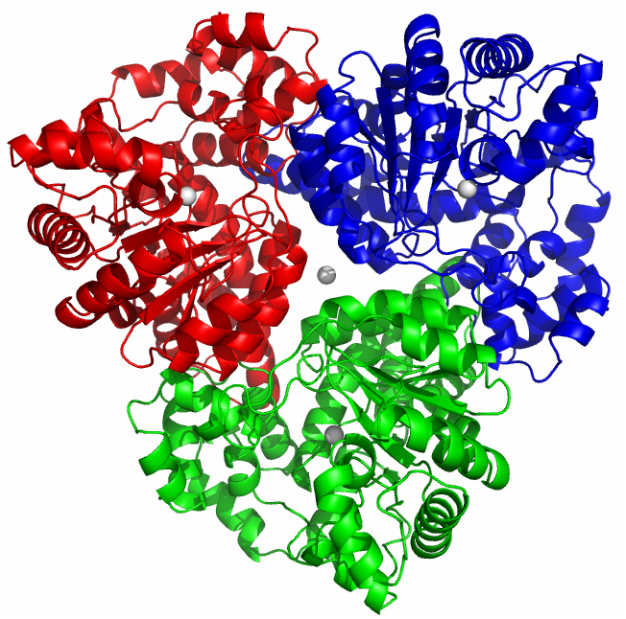


Figure 2.4: Ribbon representation of the trimeric form of Bh0493. Each color represents an individual subunit. The zinc ions are shown as grey spheres.

The structure of the Bh0493 monomer is composed of a $(\beta/\alpha)_7\beta$-barrel with N-terminal and C-terminal extensions of the chain from both sides of the barrel. The following chain segments are included in the eight β-strands of the barrel: β1 (residues 21-28), β2 (137-143), β3 (165-170), β4 (219-224), β5 (253-260), β6 (295-301), β7 (318-323) and β8 (349-356). The N-terminal extension of the chain includes α-helix 6-20. The C-terminal extension includes two long distorted α-helices from residues 360-386 and 393-413. These three long α-helices form a mutual hydrophobic core and stick out of the molecule. The long loop between strands β1 and β2 of the barrel contain 8 helices. The chain segment 40-122 of this loop contains 6 helices, sticks out of the molecule and can be considered as a separate domain B. The long interval between

strands β3 and β4 of the barrel contains 3 helices. The remaining intervals between the β-strands of the barrel contain one helix each.

The active site is located at the C-terminal end of the barrel in large domain A of the Bh0493 monomer and is open to external solvent. The three active sites from three adjacent monomers keep their positions approximately at the vertices of the trimer at a maximum distance from three-fold axis and from each other. The Zn^{2+} ion is bound in the active site of every monomer and is coordinated by His-26 and His-28 from the strand β1 and by Asp-355 from the strand β8 as illustrated in **Figure 2.5**. The fourth ligand for Zn is a water molecule. The inter-atomic distances between the Zn in the active site to these four ligands are 2.3, 2.5, and 2.3 and 2.0 Å, respectively. The other residues whose side chains may be important in substrate binding are: His-49 and Tyr-50 from the domain B, Arg-170 from the strand β3, Trp-325 from the loop after β7 and Arg-357 from the loop after β8.

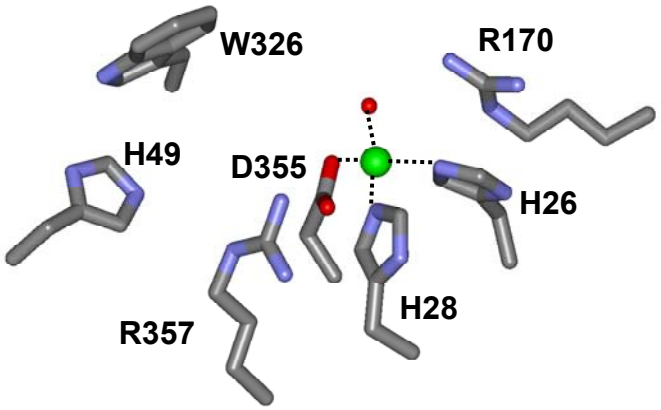


Figure 2.5: Close-up view of the active site region of Bh0493. The relative orientation of the zinc (green sphere), the four metal ligands and other conserved residues that are likely candidates for the binding of substrate and catalytic transformations are shown.

The Bh0493 trimer is mainly stabilized by interactions between the C-terminal helices 360-386 and 393-413 of one monomer and domain B (40-122) of the adjacent monomer. These interactions include main chain – main chain hydrogen bonds and side chain – main chain hydrogen bonds. Additional trimer stabilizing effects are interactions between the loops β5-β6, β6-β7 and β7-β8 of every monomer with the corresponding loops of neighboring monomer within the Bh0493 trimer.

DISCUSSION

Since Holm and Sander first identified a diverse set of amidohydrolases related to adenosine deaminase, phosphotriesterase, and urease in 1997 (*21*), increasingly sophisticated sequence search methods as well as the sequencing of additional genomes have revealed that the AHS contains far more diversity than previously thought, both in terms of function and in terms of sequence (**Figure 2.1A**). For example, α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (ACMSD) and related enzymes, recently identified as part of the AHS (*115*), catalyze nonoxidative decarboxylase reactions. Although these reactions likely share some mechanistic similarity with that of canonical AHS members, and not surprisingly, these sequences are quite different from those of the core superfamily members. As shown by the network representation of the AHS in **Figure 2.1A**, at an *E*-value cutoff where most of the amidohydrolase families identified by Holm and Sander are tightly clustered, ACMSD and related sequences form a distinct, although connected, cluster. The uronate isomerase-like sequences are even more divergent than those of ACMSD and the related decarboxylases. Even at the

permissive *E*-value cutoff used in **Figure 2.1A**, they do not connect to any other AHS sequences. Uronate isomerase is also an outlier of the superfamily in terms of function, being the only characterized superfamily member that catalyzes the interconversion of aldose and ketose functional groups (*98*).

Given currently available tools, enzymes can often be reliably assigned to a superfamily, especially when X-ray structures are available to augment sequence information. However, classifying superfamily sequences into appropriate families each associated with a specific overall reaction can be much more challenging, especially within superfamilies in which many divergent families catalyzing different overall reactions conserve active site residues common to all of the superfamily members. In such cases, even closely related enzymes may catalyze different overall reactions (*63*), and conversely, quite distantly related enzymes may catalyze the same overall reaction (*116*). The problem of functional inference in such superfamilies has been described in some detail for the muconate lactonizing enzyme subgroup of the enolase superfamily (*116*). This is also an issue in the AHS, where, for example, in silico docking approaches were required to identify the function of protein Tm0936 from *T. maritima* (PDB ID 1plm and 1j6p) as an adenosine and thiomethyl adenosine deaminase (*7*). By sequence similarity, this protein most resembles the large chlorohydrolase and cytosine deaminase subgroup within the AHS but shows no sequene similarity with adenosine deaminases.

Even within the canonical members of the URI family, substantial divergence is apparent, with the pairwise sequence identity among the top three sequences shown in the alignment given in **Figure 2.6** ranging from 35 to 52% and the sequence identities between these sequences and that from *C. crescentus* falling to between 22 and 27%. The new outlier uronate isomerase from *B. halodurans*, Bh0493, is even more distant, with a pairwise sequence identity to the other sequences in **Figure 2.6** ranging from 17 to 22%. As shown by the network representation in **Figure 2.1B**, it is only at an *E*-value threshold of 10^{-10} that the outlier uronate isomerases, including Bh0493, connect to the tightly clustered canonical uronate isomerases. At a slightly more stringent *E*-value cutoff of 10-12, the outlier uronate isomerases form a distinct group, losing their connection to the canonical uronate isomerases (not shown). However, because Bh0493 is so distant from even the canonical uronate isomerases, it is useful for identifying minimum sequence requirements for URI activity. The few active site residues conserved across Bh0493 and the canonical uronate isomerases (**Figure 2.6**) are likely candidates for residues required for substrate binding and catalysis. These residues include the conserved HxH motif from β-strand 1, the aspartate from β-strand 8, and a WWF motif from β-strand 7. We have demonstrated in unpublished experiments with

the uronate isomerase from *E. coli* that residues corresponding to His-26, His-28, His-49, Arg-170, Asp-355, and Arg-357 in Bh0493 are critical for the functioning of this enzyme. These residues are conserved in the sequence alignment of Bh0493 with other authentic uronate isomerases (**Figure 2.6**).

The identification of Bh0493 as a uronate isomerase is further supported by the gene context within the genome of *B. halodurans*. Presented in **Figure 2.7** is a schematic representation of the open reading frames in the vicinity of the gene that encodes for Bh0493. The protein Bh0492 is annotated as a tagaturonate oxidoreductase (UxaB) and Bh0490 is an altronate dehydratase (UxaA). The gene for Bh0492 is homologous to the *uxa*B gene in *E. coli* and the two proteins have an amino acid sequence identity of 38%. Likewise, Bh0490 has a sequence identity of 54% to altronate dehydratase from *E. coli* (UxaA). The co-location of these three genes suggests an operon for the metabolism of D-galacturonate to pyruvate and D-glyceraldehyde-3-phosphate (**Scheme 2.1**). This conclusion is further supported by the annotation of Bh0494 as an exopolygalacturonate lyase (PelX).

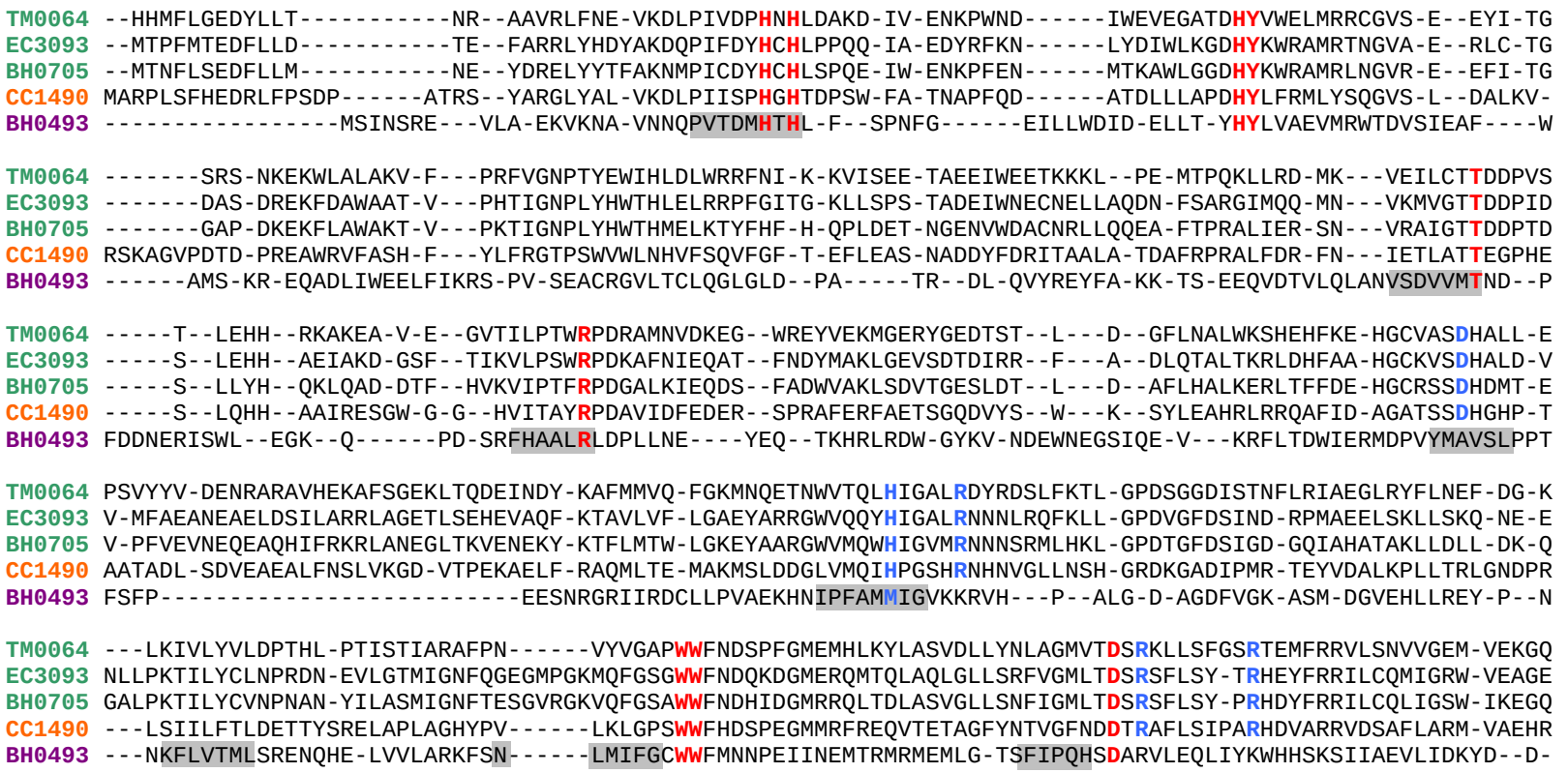


Figure 2.6: Protein sequence alignment of representative URI-like sequences from *T. maritima* (Tm0064), *E. coli* (b3092), *B. halodurans* (Bh0705 and Bh0493), and *C. crescentus* (Cc1490). The three sequence groups shown in **Figure 2.3** are indicated by the colors on the sequence headers (green=1; orange=2; and purple=3). Residues within the active site that may influence substrate binding and/or catalysis are colored. Red indicates residues conserved across the entire URI sequence set, and blue indicates positions that are completely conserved within a group but not across the entire sequence set. β-strands for Bh0493 are indicated with gray shading.

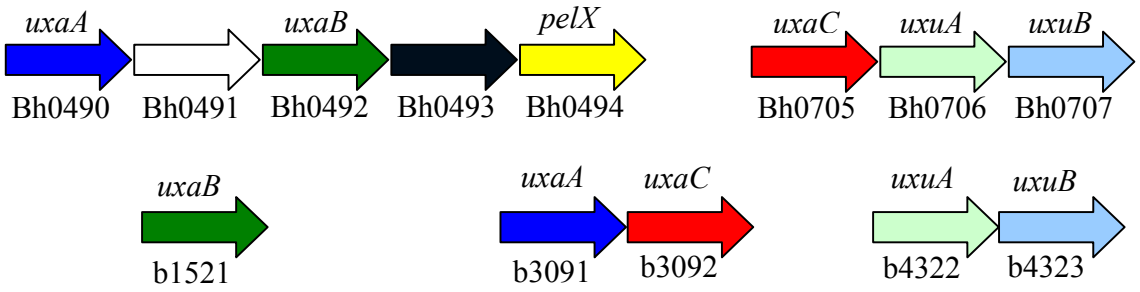


Figure 2.7: Chromosomal arrangement of the genes that code for enzymes involved in the metabolism of D-glucuronate and D-galacturonate. (top) *B. halodurans*. (bottom) *E. coli* K12. The relative lengths of each gene are not drawn to scale.

In nearly all microorganisms the isomerization of the two uronic acids, D-glucuronate and D-galacturonate, are catalyzed by the same enzyme. This is certainly true for *E. coli* and *T. maritima*. However, in the genome of *B. halodurans* there is another gene that is annotated as *uxa*C and apparently this enzyme also catalyzes the isomerization of uronic acids. The gene for this protein (Bh0705) is found adjacent to two other genes currently annotated as *uxu*B (Bh0707) and *uxu*A (Bh0706). The enzymes encoded by these two genes are expected to be D-mannonate oxidoreductase (UxuB) and D-mannonate dehydratase (UxuA). The amino acid sequence of Bh0705 is 53% identical with the prototypical uronate isomerase from *E. coli* (UxaC, b3092) while Bh0706 is 38% identical with the sequence for the authentic mannonate dehydratase from *E. coli* (UxuA, b4322). However, the protein Bh0707 has no apparent homology to the *uxu*B gene product from *E. coli* (b4323). Nevertheless, this enzyme does have a high sequence identity to the gene product of y*jm*F from *B. subtilis* and *uxu*B from *B. stearothermophilus*. Both of these proteins have been annotated as D-mannonate

oxidoreductases (*117*, *118*). Mannonate dehydrogenases generally belong to the long-chain alcohol dehydrogenase superfamily, but the protein sequence of *yjmF* is identified as a member of the short-chain alcohol dehydrogenase superfamily. The protein from *B. subtilis* was first recognized as a mannonate oxidoreductase because it was found in a cluster of genes that are part of the hexuronate catabolic pathway. The identification of Bh0707 as a mannonate dehydrogenase and Bh0706 as a mannonate dehydratase indicates that Bh0705 is in an operon for the metabolism for D-glucuronate. Therefore, it is apparent that in *B. halodurans* there are separate pathways for the metabolism of D-glucuronate and D-galacturonate. Bh0705 is in the operon for D-glucuronate whereas Bh0493 is in the operon for D-galacturonate.

The two putative uronate isomerase enzymes from *B. halodurans*, Bh0493 and Bh0705, were expressed, purified and characterized. It was found that the enzyme Bh0705 will catalyze the isomerization of both D-glucuronate and D-galacturonate. However, the enzyme is significantly more active with D-glucuronate than with D-galacturonate. The value of k_{cat}/K_m for the isomerization of D-glucuronate (2.9×10^4 M^{-1} s^{-1}) is approximately two orders of magnitude higher than it is for the isomerization of D-galacturonate (3.6×10^2 M^{-1} s^{-1}). This indicates that Bh0705, unlike the homologous enzyme from *E. coli*, is relatively specific for the isomerization of D-glucuronate. The corresponding enzyme from *E. coli* (b3092) has been shown to isomerize D-glucuronate with a k_{cat}/K_m of 4×10^5 M^{-1} s^{-1} and D-galacturonate with a k_{cat}/K_m of 2×10^5 M^{-1} s^{-1} (*98*). Therefore, the lone uronate isomerase from *E. coli* cannot discriminate between these two epimeric sugars.

The enzyme Bh0493 can catalyze the isomerization of D-glucuronate and D-galacturonate with a similar catalytic activity for either compound. The observed double-reciprocal plots for Bh0493 are nonlinear for either substrate with an apparent activation at high substrate concentrations. The dependence of the observed catalytic activity as a function of substrate concentration was fit to equation 2.2. At low substrate concentrations of D-glucuronate the value of k_{cat}/K_m is 1.1 x 10^5 M^{-1} s^{-1} whereas at low concentrations of D-galacturonate the value of k_{cat}/K_m is 4.7 x 10^4 M^{-1} s^{-1} and thus there is very little discrimination between these two acid sugars. At saturating concentrations of D-glucuronate the effective turnover number is ~8.8 s^{-1} (5.2 + 3.6 s^{-1}) and somewhat less with D-galacturonate (3.3 s^{-1}). The reason for the nonlinearity in the double reciprocal plots is unknown but it could represent a homotropic cooperativity between subunits. These results validate the hypothesis that Bh0705 catalyzes the isomerization reaction in the D-glucuronate pathway and Bh0493 is primarily responsible for the isomerization reaction that initiates the utilization of D-galacturonate, although this enzyme catalyzes the isomerization of both compounds.

Two compounds that mimic the proposed *cis*-enediol intermediate in the isomerization of D-glucuronate and D-galacturonate were tested as inhibitors of the reactions catalyzed by Bh0493 and Bh0705. D-Arabinaric acid can be considered as a mimic for either compound depending on how the dicarboxylic acid is oriented within the active site. This compound was found to be an excellent competitive inhibitor for the isomerization of D-glucuronate with either enzyme. With Bh0493, the lowest K_i value (at low substrate concentrations) was found to be 55 nM whereas with Bh0705 the

competitive inhibition constant was 400 nM. The competitive inhibition constants for the monohydroxamate derivative of D-arabinaric acid are considerably weaker for both enzymes relative to the effects of D-arabinaric acid. For Bh0493 and Bh705 the observed K_i values are 2 and 18 μM, respectively. The monohydroxamate derivative made for this investigation is an analogue of the *cis*-endiol of D-galacturonate rather than of D-glucuronate. This partially explains the weaker binding of the hydroxamate with Bh705 but it does not explain the weaker binding, relative to D-arabinaric acid, with Bh0493. In any event, the inhibition of these compounds is consistent with the formation of a *cis*-enediol-like reaction intermediate during the isomerization of the aldose substrate to the ketose product.

The first reported crystal structure in the uronate isomerase subfamily of the AHS was from *T. maritima*. The enzyme Bh0493, with an overall sequence identity of 19% to Tm0064, is structurally similar to this enzyme with an r.m.s.d. of 2.3 Å. Like Tm0064, the quaternary structure of Bh0493 is a homotrimer organized as a pinwheel (**Figure 2.4**). Members of the AHS are generally found in oligomerization states with an even number of subunits. For example, phosphotriesterase and dihydroorotase exist as homodimers (*34*, *119*). NagA from *E. coli* is tetrameric (*42*), whereas cytosine deaminase and isoapartyl dipeptidase adopt hexameric and octameric oligomerization states (*35*, *47*). Uronate isomerase is thus the first example for a trimeric quaternary arrangement of subunits within this superfamily of enzymes.

Structural comparisons of Bh0493 with Tm0064 indicate that tertiary fold of Bh0493 is also organized into two distinct domains (**Figure 2.8**). Domain A assumes

the amidohydrolase-like structural fold (TIM-barrel) and is the site for the catalytic transformations. Domain B is composed mainly of α-helices and loops and possesses a unique structural fold. The structural alignment of domain B for these two enzymes indicates a nearly identical fold with the exception of one helix being slightly longer in the *T. maritima* enzyme. Comparison of domain A also reveals a highly conserved structural fold in these two enzymes. However, one major difference between the two structures is an extra helix in Tm0064 of over 20 residues that occurs at the end of β-strand 4. The structural alignment also demonstrates that the histidine contributed from the end of β-strand 5 (highly conserved in nearly all of the enzymes in the AHS) is not conserved in Bh0493, suggesting that this histidine is not important for catalytic activity. Aside from the HxH motif at the end of β-strand 1 and the aspartate at β-strand 8, several other conserved residues are present in this alignment including His-49, Arg-170, Arg-357, and the WWF motif the end of β-strand 7. Although the details of the chemical mechanism for the enzymatic isomerization of D-glucuronate and D-galacturonate are uncertain, it is quite likely that these specific residues will be found to play the critical roles in the catalytic activity of the uronate isomerases. A sequence alignment including all uronate isomerase like sequences shows that the residues identified above are highly (in most cases, completely) conserved, supporting their importance for catalysis and/or substrate recognition (**Figure 2.6**). Additional positions within the active site appear to be conserved within one of the three uronate isomerase groups identified based on the phylogenetic tree, but not across the entire uronate isomerase sequence set (**Figure 2.6**), possibly indicating different strategies for substrate recognition within each group.

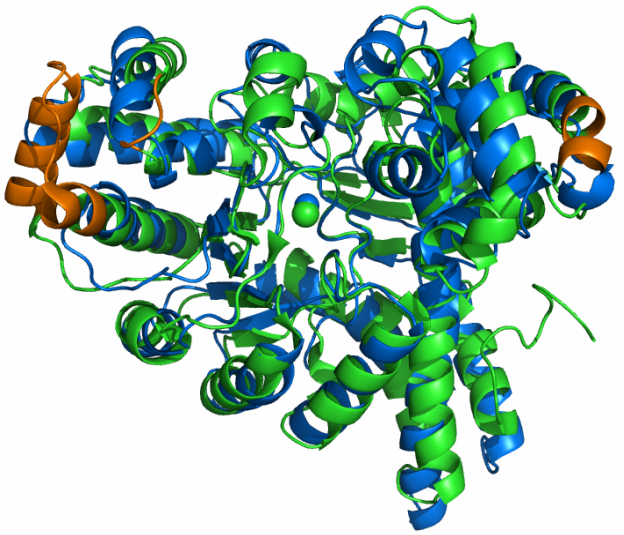


Figure 2.8: Overlay of the ribbon representation of the two structures, Tm0064 (green) and Bh0493 (blue). In orange are the non-overlapping regions between the two structures. The green sphere represents the putative metal ion in Tm0064 and the blue sphere represents the zinc ion in Bh0493.

B. halodurans may not be the only organism with separate enzymes for the isomerization of D-glucuronate and D-galacturonate. *Alkaliphilus metalliredigenes* QYMF, also an alkaliphilic bacterium like *B. halodurans*, has two members in the uronate isomerase sequence set, one of which (gi: 77686974) has 48% sequence identity to the prototypical uronate isomerase of *E. coli* and the other of which (gi: 77685308) has 57% identity to the outlier uronate isomerase Bh0493. The gene encoding protein 77686974 is located near homologues of *E. coli* mannonate dehydrogenase (UxuB, gi: 77686972) and mannonate dehydratase (UxuA, gi: 77686973), enzymes specific for the breakdown of D-glucuronate. Although the *A. metalliredigenes* genome has not been fully assembled, two fragments homologous to *E. coli* UxaB are found near 77685308,

the second putative UxaC (gi: 77685309 and 77685310). Full length homologs of *E. coli* UxaB and UxaA are also found in the genome (gi: 77686989 and 77686988, respectively). The homology of the two putative *A. metalliredigenes* UxaCs to the *B. halodurans* UxaCs, along with genomic context information, suggest that *A. metalliredigenes* has distinct isomerases for the metabolism of D-glucuronate and D-galacturonate.

Clostridium beijerincki NCIMB 8052 and *Saccharophagus degradans* 2-40 also contain multiple uronate isomerase like sequences, with genome context that suggests one may be specific for the isomerization of D-gluconate and the other for D-galacturonate. However, the putative D-galacturonate isomerases from these organisms do not cluster with those from *B. halodurans* and *A. metalliredigenes*, nor with each other. If they are, indeed, specific for D-galacturonate, this specificity appears to have evolved multiple times within the uronate isomerase like group.

Other proteins homologous to the outlier uronate isomerase Bh0493 include protein sequences from *Moorella thermoacetica* ATCC 39073 (gi: 83589292), *Blastopirellula marina DSM3645* (gi: 87310825), *Caldivirga maquilingensis* IC-167 (gi: 126353639) and *Rhodopirellula baltica SH 1* (gi: 32474066). These organisms are unusual because they do not appear to have a protein homologous to the prototypical uronate isomerase in *E. coli*. The *M. thermoacetica* homologue is located near genes homologous to *E. coli* UxuB and UxuA, indicating that it may be a genuine uronate isomerase. The remaining sequences, however, cannot be validated based on genomic

context. The function of the *C. maquilingensis* sequence is particularly open to question, as this is the only nonbacterial organism with a uronate isomerase like sequence.

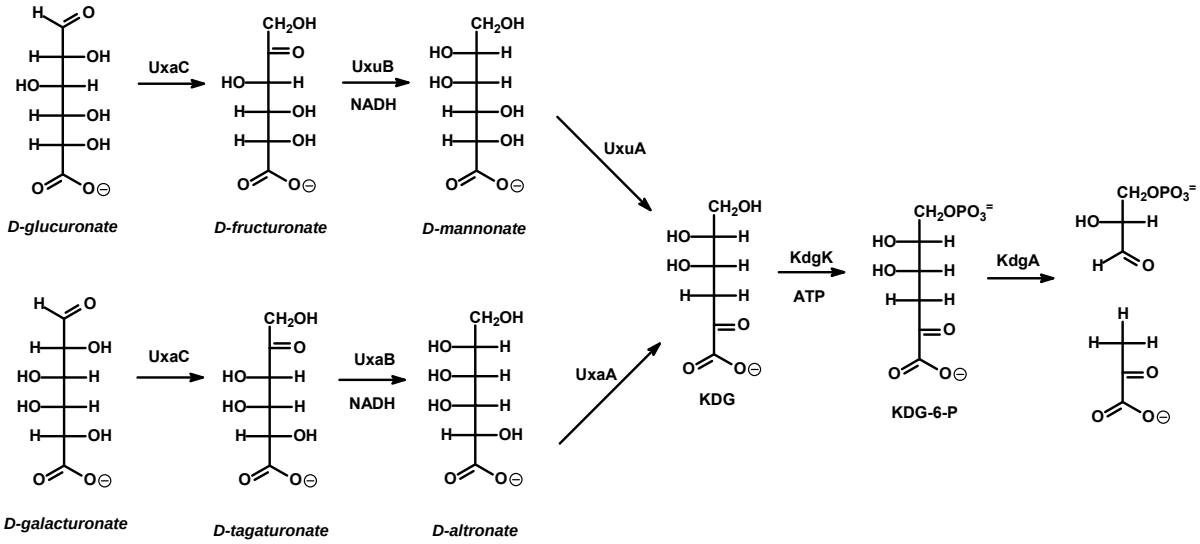
From the kinetic and structural studies, we have determined that the protein Bh0493 from *B. halodurans* is a uronate isomerase and a diverged member of the amidohydrolase superfamily. In *B. halodurans*, the gene encoding uronate isomerase had been identified previously as Bh0705 based on sequence identity. Initial velocity studies of Bh0705 and Bh0493 with the substrates D-glucuronate and D-galacturonate indicate that Bh0705 is highly specific for D-glucuronate while Bh0493 can utilize either substrate with similar efficiency. Bh0493 is located adjacent to genes that are annotated as tagaturonate oxidoreductase, altronate dehydratase, and exopolygalacturonate lyase (an operon for D-galacturonate metabolism), while the gene for Bh0705 is located near mannonate dehydratase and mannonate oxidoreductase (an operon for D-glucuronate metabolism). Sufficient information is not available to hypothesize as to why *B. halodurans* and some other organisms have more than one URI gene or whether this leads to a selective advantage for these organisms. Nor is the evolutionary path that connects this highly divergent family to the AHS clear. However, identification of this outlier Bh0493 sequence as a uronate isomerase extends the boundaries of the superfamily further into the "twilight" of functional and structural divergence than has been previously recognized.

CHAPTER III

THE MECHANISM OF THE REACTION CATALYZED BY URONATE ISOMERASE ILLUSTRATES HOW AN ISOMERASE MAY HAVE EVOLVED FROM A HYDROLASE WITHIN THE AMIDOHYDROLASE SUPERFAMILY

Uronate isomerase (URI[1]) catalyzes the first step in the pathway for the metabolism of D-glucuronate and D-galacturonate. In this transformation, D-glucuronate and D-galacturonate are initially isomerized into their corresponding keto products, D-fructuronate and D-tagaturonate, respectively (*120*). D-Fructuronate and D-tagaturonate are then reduced to D-mannonate and D-altronate, respectively, by mannonate and altronate dehydrogenase in the presence of NADH (*121*). The pathways converge through a dehydration reaction where mannonate dehydrase and altronate dehydrase convert mannonate and altronate to 2-keto-3-deoxy-D-gluconic acid (KDG). This product is then phosphorylated by the enzyme ketodeoxygluconic acid kinase with ATP to form 2-keto-3-deoxy-6-phosphogluconic acid (KDG-6-P). In the final step of this pathway, 2-keto-3-deoxy-6-phosphogluconic acid is cleaved by an aldolase to yield pyruvate and D-glyceraldehyde-3-phosphate, which enter the citric acid cycle and glycolysis. The entire pathway is summarized in **Scheme 3.1** (12*0,121*).

Scheme 3.1: Entire pathway for the metabolism of D-glucuronate and D-galacturonate.

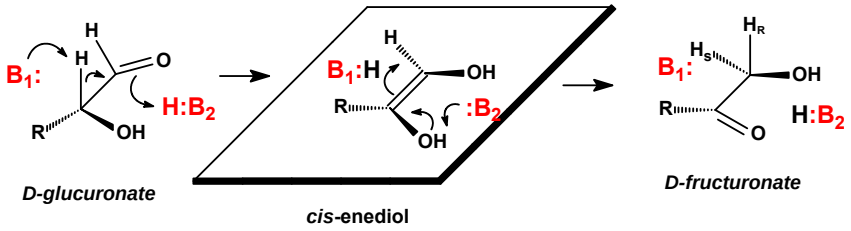


We have demonstrated that uronate isomerase is a member of the amidohydrolase superfamily of enzymes (*44*). The majority of the functionally characterized members of the amidohydrolase superfamily catalyze the hydrolysis of amide or ester bonds to carbon or phosphorus centers (*21,31*). Well characterized examples include dihydroorotase (*60*), urease (*32*) and phosphotriesterase (6*8*). Members of this superfamily also catalyze the deamination of many nucleotides including adenosine (*56*), cytosine (*47*), and guanine (*122*). The active sites of these enzymes generally consist of a mononuclear or binuclear metal center that is perched at the C-terminal end of the β-barrel in proteins within a $(\beta/\alpha)_8$ structural fold. The most highly conserved residues in the AHS include two histidines from β-strand 1, histidines after the ends of β-stands 5 and 6, and an aspartic acid from β-strand 8. Since URI

catalyzes an isomerization of an aldose sugar to the corresponding ketose product, this enzyme is one of the most divergent members of the amidohydrolase superfamily. The mechanistic details of this transformation are therefore of significant interest toward a greater understanding of how an active site that originally evolved to catalyze hydrolytic reactions has been re-forged to undergo an isomerization reaction.

We have previously demonstrated that the proton originally at C-2 of D-glucuronate is ultimately found at the *pro*-R position at C-1 of D-fructuronate and that this proton slowly exchanges with solvent (*98*). These results are consistent with a proton transfer mechanism with a *cis*-enediol intermediate. The general mechanism, shown in **Scheme 3.2**, indicates a requirement for at least two residues that participate in the transformation of D-glucuronate into D-fructuronate. A general base ($:B_1$) abstracts the proton from C-2 of D-glucuronate and a general acid ($H:B_2$) facilitates the transfer of a proton to the carbonyl oxygen at C-1 to produce the *cis*-enediol intermediate. In the subsequent step, the ketose product is generated by a proton transfer from the hydroxyl group at C-2 of the proposed intermediate and protonation of C-1 by $H:B_1$. For compounds such as D-glucuronate the enzymatic transformation is made more complicated by the fact that in solution the substrate exists almost entirely as a mixture of two anomeric cyclic hemiacetals.

Scheme 3.2: Proton transfer mechanism in a 1,2-hydrogen transfer reaction.



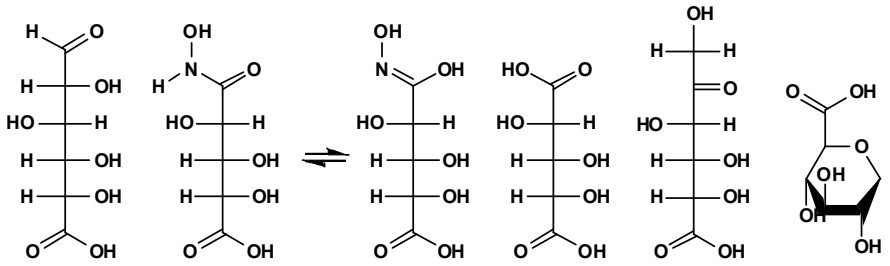
This chapter focuses on a determination of the chemical mechanism for the isomerization reaction catalyzed by URI from *E. coli*. The rate limiting steps have been interrogated by measuring the primary kinetic isotope effects with [2-^{2}H]-D-glucuronate and solvent isotope effects with D_2O for the wild type and mutant enzymes. The rate limitation imposed by product release has been examined using solvent viscosity effects. The identity of the residues involved in the proton transfer events has been probed by pH-rate profiles and characterization of the kinetic constants for mutant enzymes. These approaches have been augmented by the elucidation of the X-ray structure of a uronate isomerase from *Bacillus halodurans* (Bh0493) in the presence of D-glucuronate, D-fructuronate, and two mimics of the *cis*-enediol intermediate.

MATERIALS AND METHODS

Materials. D-Glucuronic acid (**I**), L-gulonic acid γ-lactone (**VI**), NADH, buffers, and all other chemicals were purchased from Sigma-Aldrich or Acros, unless otherwise stated. D-arabinaric acid (**III**) and the monohydroxamate derivative of this compound (**II**) were synthesized by Dr. Yingchun Li. [2-^{2}H]-D-glucuronate was also

synthesized by Dr. Yingchun Li. 2,6-Anhydro-L-gulonic acid (**V**) was synthesized by Dr. Chengfu Xu. The structures of these compounds are presented in **Scheme 3.3**. Oligonucleotide syntheses and DNA sequencing were performed by the Gene Technologies Lab of Texas A&M University. Metal analyses were done using inductively coupled plasma mass spectrometry (ICP-MS).

Scheme 3.3: Structures of D-glucuronate (I), hydroxamate derivative (II), D-arabinaric acid (III), D-fructuronate (IV) and 2,6-anhydro-L-gulonic acid (V).



Site-Directed Mutagenesis. Site-directed mutagenesis of URI was performed using the QuikChange mutagenesis kit from Stratagene. The following mutants were obtained by this method: H33N, H33A, H35N, H35A, H59N, H59A, Y60F, Y60A, R186K, R186M, D238N, H297N, R302K, R302M, H297A, W381F, W381A, D412N, D412A, R414K, and R414M. The mutations were confirmed by DNA sequencing of the modified plasmids.

Protein Expression and Purification of Uronate Isomerase. The *uxa*C gene encoding *E. coli* uronate isomerase (gi: 16130987) was amplified by PCR and inserted into pET28 without a histag. The protein was expressed in the *E. coli* strain BL21-(DE3), and single colonies were used to inoculate 5 mL of LB medium supplemented with 50 *í*g/mL kanamycin. The overnight culture was used to inoculate 1 L of TB medium supplemented with 50 μg/mL kanamycin. The large culture was grown at room temperature until an *A*600 of 0.4-0.6 was reached, and then 1.0 mM $ZnCl_2$ was added, followed by induction with 0.4 mM IPTG. The cells were incubated overnight and then centrifuged at 6000 rpm for 10 min at 4 °C. The cell pellet was resuspended in 50 mM HEPES, pH 8.0, and 500 μM $ZnCl_2$. The cells were lysed by sonication and the nucleic acids precipitated by dropwise addition of 2.0% (w/v) protamine sulfate solution. The solution was centrifuged at 12 000 rpm and the lysate fractioned between 50% and 80% saturation of ammonium sulfate. After centrifugation, the pellet was resuspended in a minimal amount of buffer (50 mM HEPES, pH 8.0), loaded onto a Superdex 200 gel filtration column, and eluted at a flow rate of 2.5 mL/min. Fractions containing uronate isomerase were identified by SDS-PAGE and activity assays. The pooled enzyme was loaded onto a Resource Q anion exchange column preequilibrated with buffer A (20 mM HEPES, pH8.0) and eluted with a 0-30% gradient of buffer B (20 mM HEPES, pH 8.0 and 1.0 M NaCl) at a flow rate of 4.5 mL/min. The fractions were assayed and pooled based on the enzymatic activity. The purified enzyme was stored in the refrigerator at 4 °C for up to 1 year without significant loss of activity. The enzymes contained up to 1 equivalent of zinc (depending on the mutant) as measured by ICP-MS.

Cloning and Purification of D-Mannonate Dehydrogenase. The *uxu*B gene (gi: 26111644) from *E. coli* that encodes D-mannonate dehydrogenase (MDH) was cloned into a pET30 expression vector. The gene was transformed into BL21(DE3) and the cells grown in TB broth at 37 °C. D-Mannonate dehydrogenase was purified using the same methods previously described for the isolation of uronate isomerase, but no metal was added to the purification buffers. The cell extract was made 40-60% of saturation with ammonium sulfate, and the resuspended pellet was subjected to fractionation with the aid of a Superdex 200 gel filtration column. Fractions containing D-mannonate dehydrogenase were pooled, quick-frozen, and stored at -80 °C.

Enzyme Assays. The conversion of D-glucuronate to D-fructuronate by URI was coupled to the reduction of D-fructuronate with NADH by mannonate dehydrogenase (MDH). The assays were monitored spectrophotometrically by following the decrease in absorbance at 340 nm. The standard assay conditions contained 50 mM HEPES (pH 8.0), varying concentrations of D-glucuronate, 0.2 mM NADH, excess MDH, and URI in a final volume of 250 μL. The pH-dependence of the kinetic parameters, k_{cat} and k_{cat}/K_m, were measured over the pH range of 5.3-10.3 at 0.20 pH intervals. The buffers used for the pH-rate profiles were MES, PIPES, HEPES, CHES, and CAPS. The pH values were recorded after the completion of the assays. The effects of solvent viscosity on the kinetic constants were determined at pH 8.0 using sucrose as the micro-viscogen at 25 ºC. The concentrations of sucrose were 0%, 10%, 14%, 20%, 24%, and 32% (w/w), and the corresponding relative viscosities were 1, 1.3, 1.5, 1.9, 2.2, and 3.2 (*15, 16*). The solvent isotope effects on the kinetic parameters for URI and two mutant

enzymes (D412N and R414M) were measured in 99% D_2O at a pD of 8.4. The primary deuterium kinetic isotope effects were obtained by direct comparison of the kinetic constants at pH 8.0 for [2-^{1}H]-D-glucuronate and [2-^{2}H]-D-glucuronate.

Data Analysis. The kinetic parameters, k_{cat} and k_{cat}/K_m, for uronate isomerase with D-glucuronate as the substrate were determined by fitting the initial velocity data to equation 3.1 where ν is the initial velocity, E_t is the total enzyme concentration, k_{cat} is the turnover number, [A] is the substrate concentration, and K_m is the Michaelis constant. The profiles for the variation of k_{cat} or k_{cat}/K_m with pH were fit to equation 3.3, where c is the pH-independent value of y, K_a and K_b are the dissociation constants of the ionizable groups and H is the proton concentration. The competitive inhibition patterns were fit to equation 3.2, where K_{is} is the slope inhibition constant and I is the concentration of the inhibitor.

$$v / E_t = (k_{cat}\,[A])/(K_a + [A]) \tag{3.1}$$

$$\nu / E_t = (k_{cat}[A])/(K_a(1 + (I/K_{is})) + [A]) \tag{3.2}$$

$$\log y = \log (c / (1 + (H / K_a) + (K_b / H))) \tag{3.3}$$

RESULTS

Requirement for Divalent Cation. The importance of a metal ion for the catalytic activity of uronate isomerase was investigated. The apo-enzyme was prepared and subsequently tested for enzymatic activity using D-glucuronate as the substrate. The wild-type URI from *E. coli* was found to contain 0.9 equivalents of zinc after purification. This protein (3 mL) at a concentration of 3.0 mg/mL was dialyzed against

1 L of dialysis buffer containing 20 mM dipicolinate in 50 mM MES, pH 6.0. The buffer was changed three times over the course of 48 h, after which the catalytic activity and metal content of the enzyme were determined. The chelator effectively removed more than 98% of the bound zinc as indicated by ICP-MS. The activity of the enzyme was assayed with mannonate dehydrogenase and NADH to detect the formation of D-fructuronate in the presence of 10 μM dipicolinate. The apo-enzyme exhibited less than 1% activity of the native enzyme with bound zinc.

*Inhibition by 2,6-Anhydro-L-Gulonic Acid (**V**).* Compound **V** was synthesized as a cyclic analogue mimic of the pyranose form of D-glucuronate. The inhibitory properties of **V** were determined with the wild-type uronate isomerase from *E. coli* and *B. halodurans* (Bh0493). This compound was found to be a competitive inhibitor for both enzymes. The data were fit to equation 3.2 and the values of K_{is} were determined to be 45 ± 4 and 24 ± 2 μM for the URI from *E. coli* and Bh0493, respectively.

pH-Rate Profiles. The kinetic constants for the conversion of D-glucuronate to D-fructuronate were obtained as a function of pH. The pH-rate profiles for the effects of pH on k_{cat} and k_{cat}/K_m are presented in **Figures 3.1A** and **3.1B**, respectively. The pH profiles are bell-shaped and are consistent with a single functional group that must be unprotonated for activity and another functional group that must be protonated for catalytic activity. From a fit of the data to equation 3.3 the kinetic pK_a values from the k_{cat}/K_m plot are 5.5 ± 0.1 and 9.5 ± 0.1, respectively. From the plot of k_{cat} vs. pH, kinetic pK_a values of 5.8 ± 0.1 and 10.2 ± 0.1 were obtained.

Site-Directed Mutants. Site-directed mutagenesis was utilized to identify the involvement of specific amino acids in metal binding, substrate recognition, and the catalytic mechanism of uronate isomerase. Conserved residues were chosen based on the location within the active site of Bh0493, a uronate isomerase found in *B. halodurans*. His-33 and His-35 were mutated to investigate the importance of metal binding and the potential role of the divalent cation on catalytic activity. Mutations at either of these two residues resulted in the dramatic loss of affinity for the divalent cation and a significant reduction in the catalytic activity. The diminution of catalytic activity for the mutants is more severe when these histidine residues are changed to alanine than to asparagine. The highly conserved histidine at the end of β-strand 5 (His-297) and the invariant aspartate at the end of β-strand 8 (Asp-412) were mutated to asparagine and alanine. For the mutations at His-297 there were substantial increases in the Michaelis constant. In contrast, with the mutation of Asp-412 the reduction in catalytic activity was more pronounced on k_{cat}. These four residues are broadly conserved in all members of the amidohydrolase superfamily.

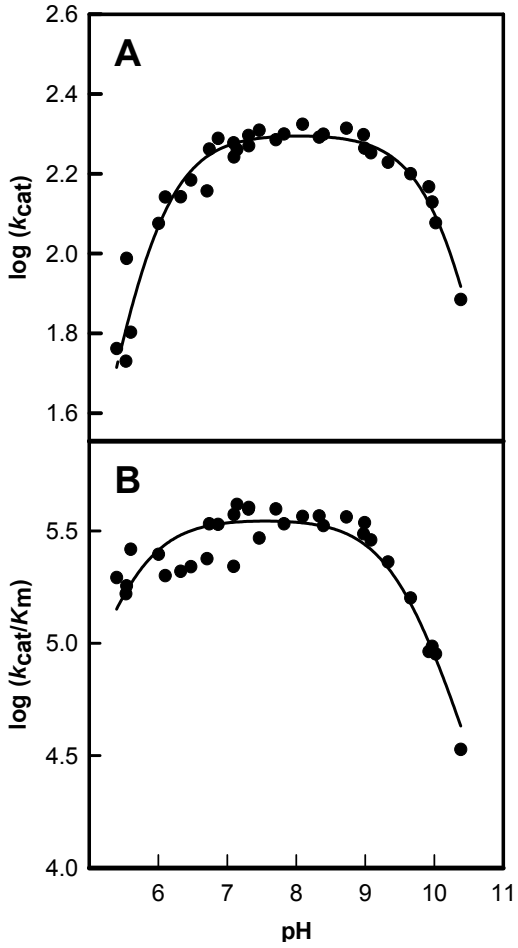


Figure 3.1: pH-rate profile for the wild type uronate isomerase containing 1 equivalent of zinc. The data were fit to equation 3.3. (**A**) Plot of log k_{cat} vs. pH. (**B**) Plot of log k_{cat}/K_m vs. pH.

Additional residues that are conserved among all of the known uronate isomerases were mutated as a probe of functional participation in binding and catalysis. These residues include Trp-381 from the conserved WWF motif after β-strand 7, His-59 and Tyr-60 after β-strand 1, and three conserved arginines (Arg-186, Arg-302, and Arg-414). Mutation of residues Trp-381, Arg-186, and Arg-302 resulted in increases in the value of K_m and small changes in k_{cat}, indicating that these residues most likely take part in substrate recognition and binding. For His-59 and Arg-414 there were relatively small changes in K_m but drastic reductions in the value of k_{cat}. The kinetic constants for the mutants constructed for this investigation are presented in **Table 3.1**.

Kinetic Isotope Effects. Primary deuterium kinetic isotope effects on the isomerization of D-glucuronate were obtained as a probe of the rate limiting steps in the overall reaction mechanism. For the solvent deuterium isotope effects, the kinetic parameters were obtained for the wild type enzyme in H_2O and D_2O with D-glucuronate as the substrate. The double reciprocal plots are presented in **Figure 3.2**. The measured solvent isotope effects for $^{D_2O}k_{cat}$ and $^{D_2O}(k_{cat}/K_m)$ are 1.22 ± 0.02 and 1.10 ± 0.09, respectively. Solvent isotope effects were also determined for four of the site directed mutants, H59N, Y60F, D412N and R414M. The solvent isotope effects for the H59N mutant were 1.7 ± 0.1 for $^{D_2O}k_{cat}$ and 1.4 ± 0.1 for $^{D_2O}(k_{cat}/K_m)$, while the effects for the Y60F mutant were 2.1 ± 0.1 and 1.2 ± 0.1 for $^{D_2O}k_{cat}$ and $^{D_2O}(k_{cat}/K_m)$, respectively. For the D412N mutant, the solvent isotope effects were determined to be 1.3 ± 0.1 and 1.5 ± 0.2 for $^{D_2O}k_{cat}$ and $^{D_2O}(k_{cat}/K_m)$, respectively. For the R414M mutant, the solvent isotope effects were 1.8 ± 0.1 for $^{D_2O}k_{cat}$ and 2.0 ± 0.3 for $^{D_2O}(k_{cat}/K_m)$.

Table 3.1: Kinetic parameters and metal content of mutants

Enzyme	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})	URI/Zn
WT	196 ± 6	0.50 ± 0.05	(4.0 ± 0.4) x 10^5	0.90
H33N	2.1 ± 0.1	(5.0 ± 0.4) x 10^{-2}	(4.7 ± 0.4) x 10^4	0.07
H33A	0.60 ± 0.01	0.20 ± 0.01	(3.0 ± 0.2) x 10^3	0.20
H35N	4.0 ± 0.2	9.4 ± 1.1	(4.3 ± 0.5) x 10^2	<0.05
H35A	0.70 ± 0.04	39 ± 5	18 ± 2	<0.05
H59N	15 ± 1	0.70 ± 0.04	(2.1 ± 0.1) x 10^4	0.96
H59A	0.60 ± 0.01	0.70 ± 0.07	(8.3 ± 0.1 x 10^2	0.95
Y60F	21.7 ± 0.1	0.16 ± 0.01	(1.4 ± 0.1) x 10^5	0.8
Y60A	13.9 ± 0.1	0.21 ± 0.01	(6.6 ± 0.3) x 10^4	0.8
R186K	54 ± 2	2.6 ± 0.2	(21 ± 2) x 10^3	0.94
R186M	4.7 ± 0.1	38 ± 3	(1.3 ± 0.1) x 10^2	0.91
D238N	60 ± 1	1.3 ± 0.1	(4.6 ±0.1) x 10^4	0.70
H297N	30 ± 2	56 ± 5	(5.0 ± 0.5) x 10^2	1.00
H297A	10 ± 1	(2.2 ± 0.3) x 10^2	43 ± 7	0.41
R302K	160 ± 4	2.5 ± 0.2	(6.3 ± 0.5) x 10^4	0.90
R302M	180 ± 9	(2.0 ± 0.3) x 10^2	(8.8 ± 1.3) x 10^2	0.99
W381F	16 ± 1	1.7 ± 0.1	(9.5 ± 0.4) x 10^3	0.90
W381A	250 ± 6	21± 2	(1.2 ± 0.1) x 10^4	0.69
D412N	0.60 ± 0.01	1.00 ± 0.04	(6.0 ± 0.3) x 10^2	0.36
D412A	(9.0 ± 0.3) x 10^{-3}	0.40 ± 0.05	21 ± 3	0.12
R414K	5.8 ± 0.1	0.82 ± 0.02	(7.1 ± 0.2) x 10^3	0.92
R414M	0.70 ± 0.01	1.4 ± 0.1	(5.4 ± 0.2) x 10^2	0.91
These data were obtained at 30 ºC, pH 8.0, with D-glucuronate as the substrate.				

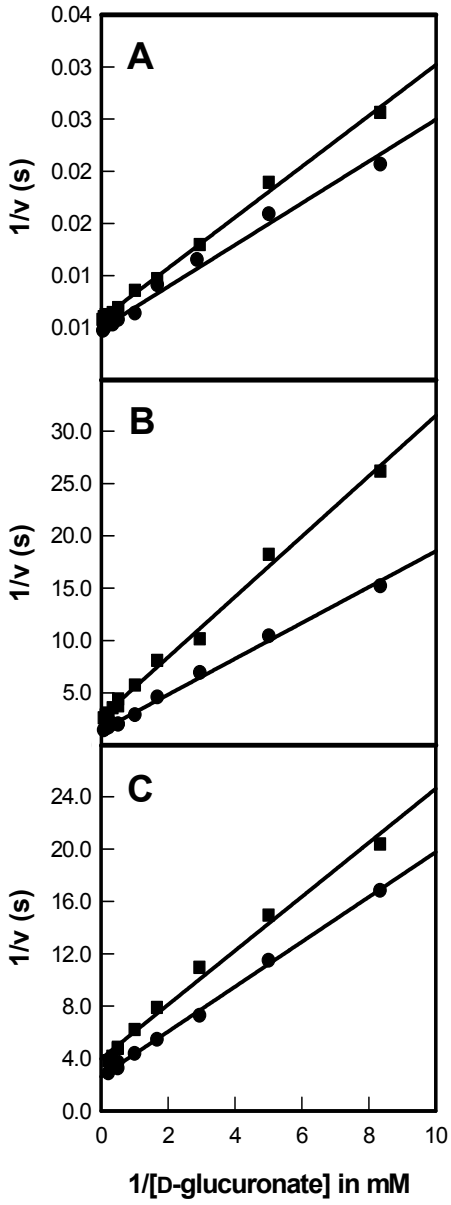


Figure 3.2: Double-reciprocal plots for the solvent isotope effect (H_2O *vs* D_2O) where $1/v$ (s) is plotted against 1/[D-glucuronate] (in mM^{-1}). The data sets in H_2O are represented by circles, and the squares correspond to the values obtained in D_2O. (**A**) represents the plot for the wild type enzyme, (**B**) for the R414M mutant, and (**C**) is the plot of the D412N mutant.

The primary deuterium isotope effects for abstraction of the proton from C-2 of D-glucuronate were determined for the wild type enzyme and two mutants, D412N and R414M. The double-reciprocal plots are shown in **Figure 3.3**. For the wild-type enzyme the primary deuterium isotope effects on k_{cat} and k_{cat}/K_m were determined to be 1.4 ± 0.04 and 1.2 ± 0.01, respectively. For the D412N mutant, the primary deuterium isotope effects were determined to be 2.0 ± 0.2 for $^{D}k_{cat}$ and 1.9 ± 0.3 for $^{D}(k_{cat}/K_m)$. For the R414M mutant, the isotope effect on k_{cat} was 3.2 ± 0.1 and the effect on k_{cat}/K_m was 3.5 ± 0.4.

Solvent Viscosity Effects. Alterations in solvent viscosity were utilized to probe the degree of rate limitation by the binding and dissociation of products and substrates on the kinetic constants of uronate isomerase (*123*). The effects of changes in solvent viscosity on k_{cat} and k_{cat}/K_m were made by the addition of sucrose (*124*). A plot of $^{o}k_{cat}/{}^{\eta}k_{cat}$ versus the relative solvent viscosity for the wild-type enzyme exhibits a slope of 0.72 ± 0.08. The slope for the effect of solvent viscosity on k_{cat}/K_m is 0.64 ± 0.05 for the wild-type enzyme. For the two mutant enzymes, D412N and R414M, the slopes for the effect of solvent viscosity on k_{cat} were found to be 0.07 ± 0.05 and 0.05 ± 0.01, respectively. For the effect on k_{cat}/K_m the slopes were 0.03 ± 0.05 and -0.17 ± 0.02, respectively, with the D412N and R414M mutant enzymes. The kinetic data are presented in **Figures 3.4A** and **3.4B**.

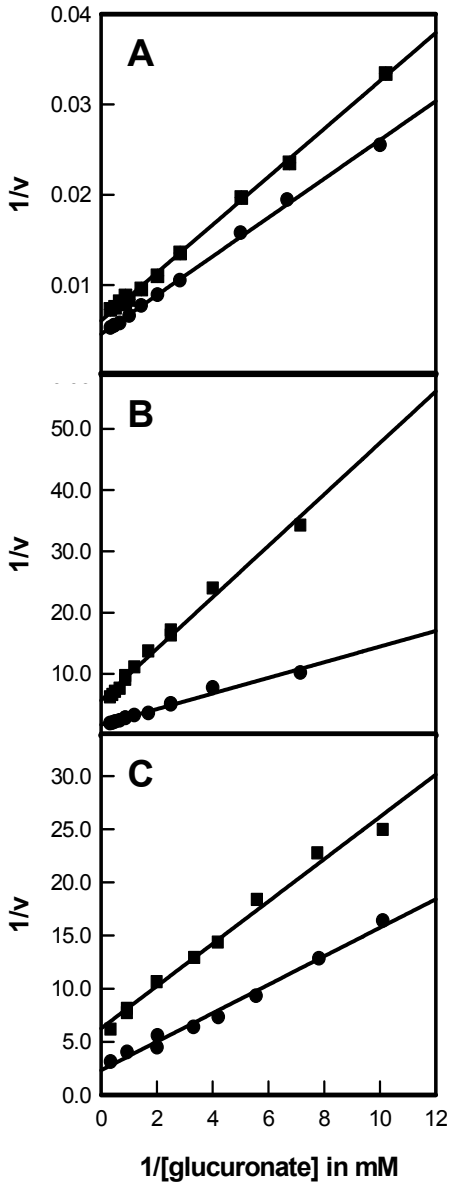


Figure 3.3: Primary isotope effects using protonated and deuterated D-glucuronate at the C-2 position are presented as double-reciprocal plots where (**A**) is the plot for the wild-type enzyme, (**B**) R414M mutant, and (**C**) D412N mutant. The data for the protonated substrate are represented as circles, and the values for the deuterated substrate are denoted as squares.

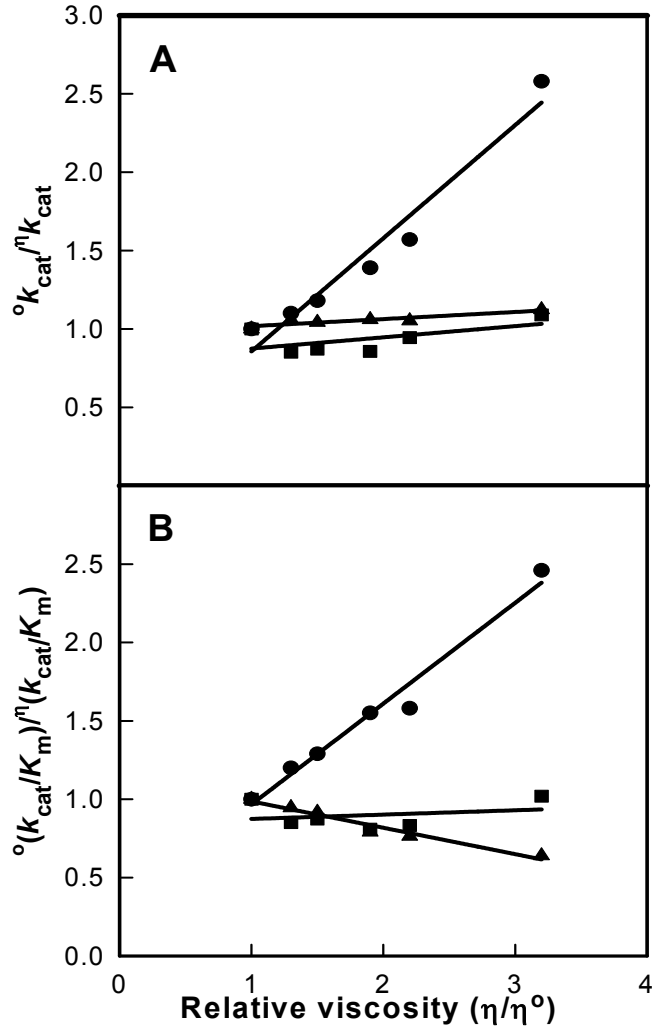


Figure 3.4: The effect of viscosity on the relative values of k_{cat} (**A**) and k_{cat}/K_m (**B**) using sucrose as the microviscogen. The circles indicate wild type enzyme, squares represent the data set for the D412N mutant, and triangles signify the values for the R414M mutant.

Structures of Inhibitor Complexes. The crystal structure of Bh0493 from *B. halodurans* was previously solved in the absence of bound substrates or inhibitors (PDB code: 2QO8, *44*). In that structure zinc is bound in the active site and coordinated to two histidines from β-strand 1 (H26 and H28) and the invariant aspartate found at the end of β-strand 8 (D355). The structure of Bh0493 was solved in the presence of the substrates D-glucuronate (**I**), D-fructuronate (**IV**), and two inhibitors that mimic the proposed *cis*-enediol(ate) intermediate, D-arabinarate (**III**) and its hydroxamate derivative (**II**). Portions of the electron density maps that show the D-glucuronate and D-fructuronate in the active site of Bh0493 are presented in **Figures 3.5A** and **3.5B**. These compounds are bound in the active site in a similar manner and are interacting with the same set of amino acid residues. Stereoscopic images of the active site complexes for compounds **I**, **II**, **III,** and **IV** are presented in **Figures 3.6A-D** and the distances between specific amino acid residues and the bound ligands are provided in **Figures 3.7A-D**.

In the complex with D-glucuronate, the substrate is bound in the open chain configuration. The terminal carboxylate is ion-paired with the guanidino group of Arg-170 and there is a monodentate coordination with the bound zinc. The zinc is also ligated with the hydroxyl group attached to C-5 of the substrate. The hydroxyl group at C-4 does not make any specific interactions with the protein, which is consistent with the observation that both D-glucuronate and D-galacturonate are substrates for this enzyme (*120*). The hydroxyl at C-3 interacts with both Arg-357 and His-49. The nearest residue to the hydroxyl at C-2 of D-glucuronate is Arg-357. The hydroxyl group from Tyr-50 hydrogen bonds with the carbonyl group at C-1 of the substrate. The closest residue to

the proton that is abstracted from C-2 of D-glucuronate is the side chain carboxylate of Asp-355. Similar interactions are found in the complexes with D-fructuronate, D-arabinarate and the hydroxamate derivative. The D-fructuronate complex was identified in a crystal that was grown after a long incubation and subsequent co-crystallization of Bh0493 with D-glucuronate.

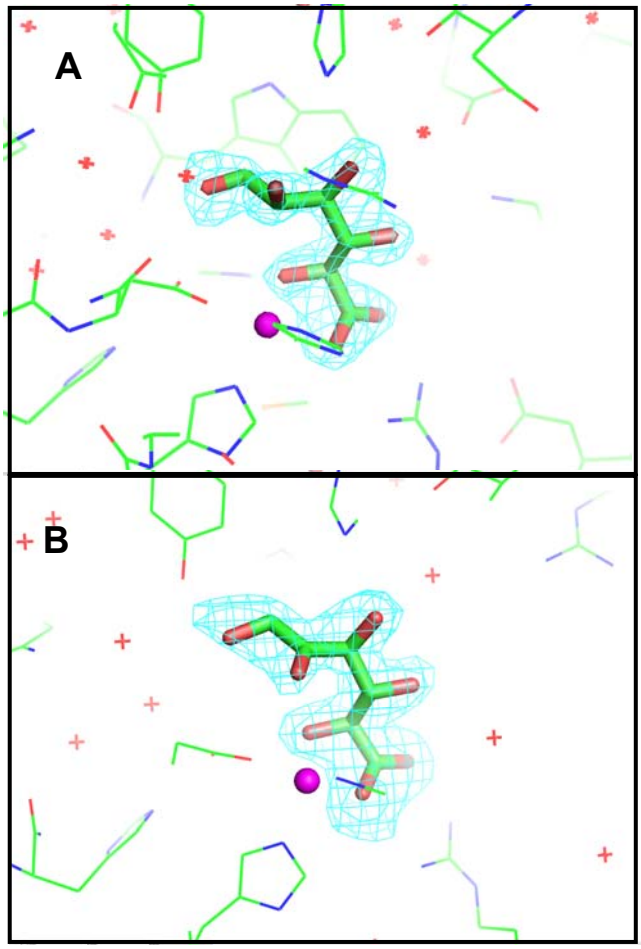


Figure 3.5: Electron density maps of (**A**) D-glucuronate and (**B**) D-fructuronate in the active site of Bh0493.

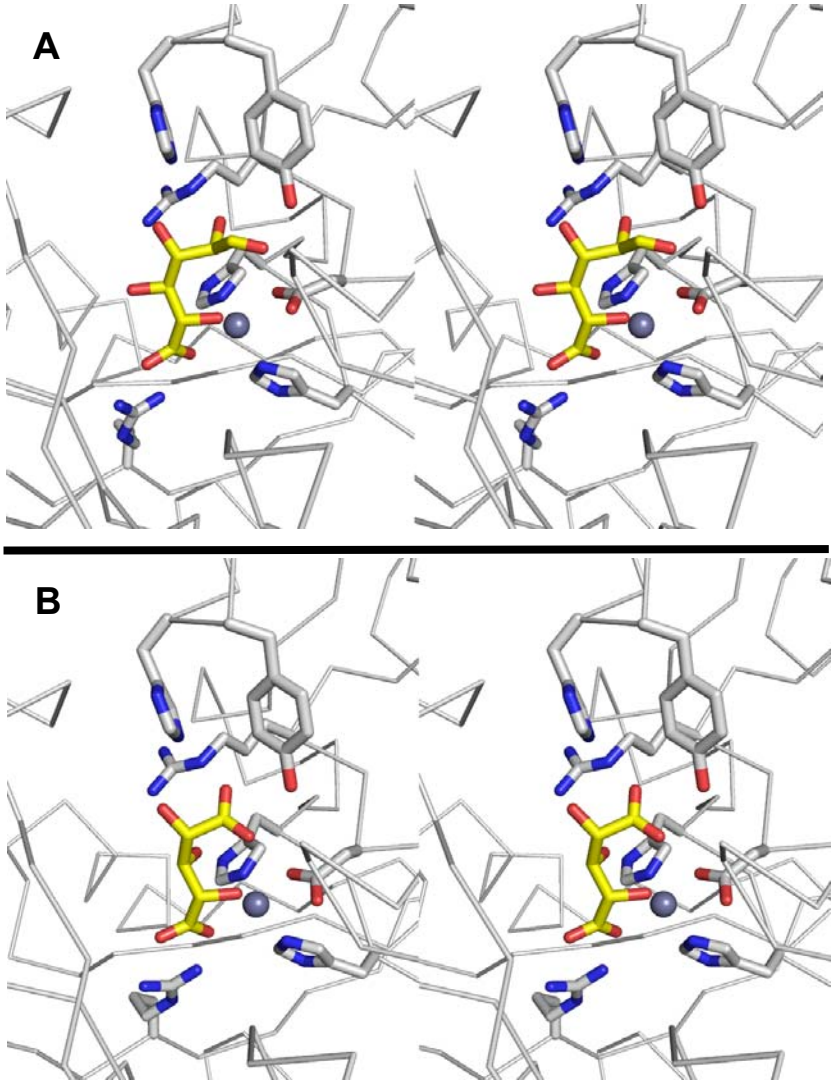


Figure 3.6: Stereoview images of Bh0493 in the presence of bound (**A**) D-glucuronate (**I**), (**B**) hydroxamate of arabinarate (**II**), (**C**) D-arabinarate (**III**), and (**D**) D-fructuronate (**IV**). This figure was created with PyMOL v0.99.

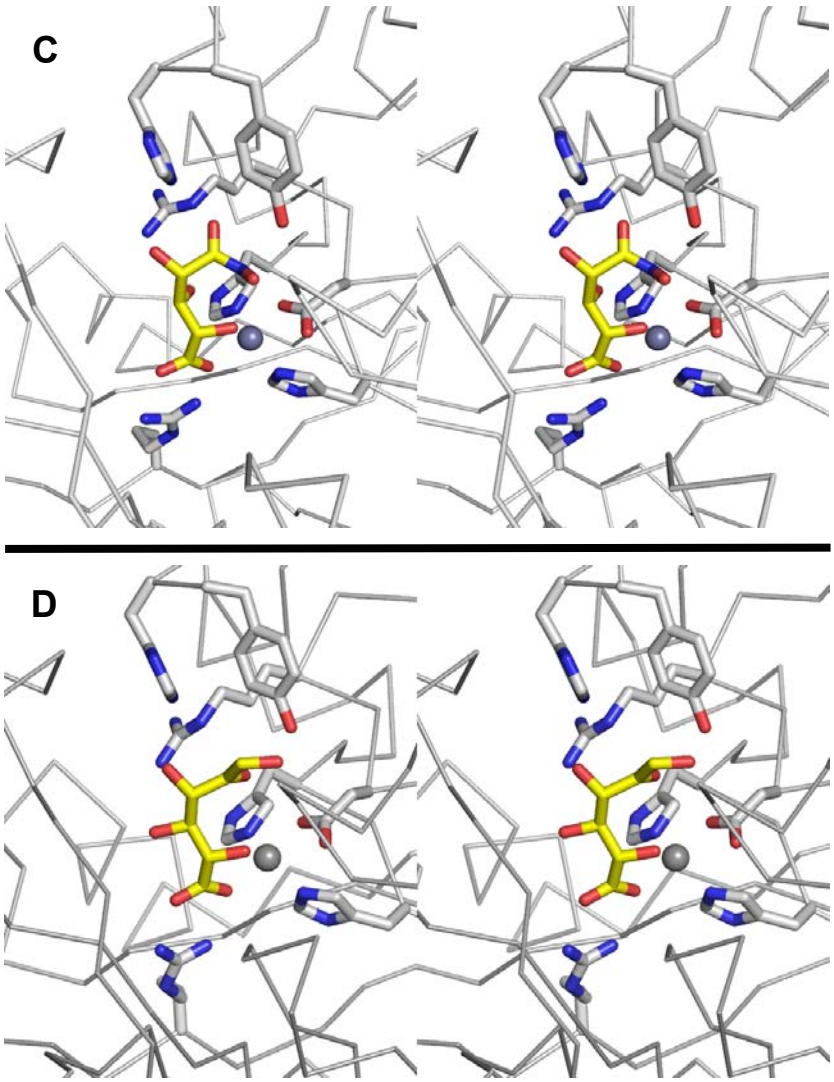


Figure 3.6: Continued.

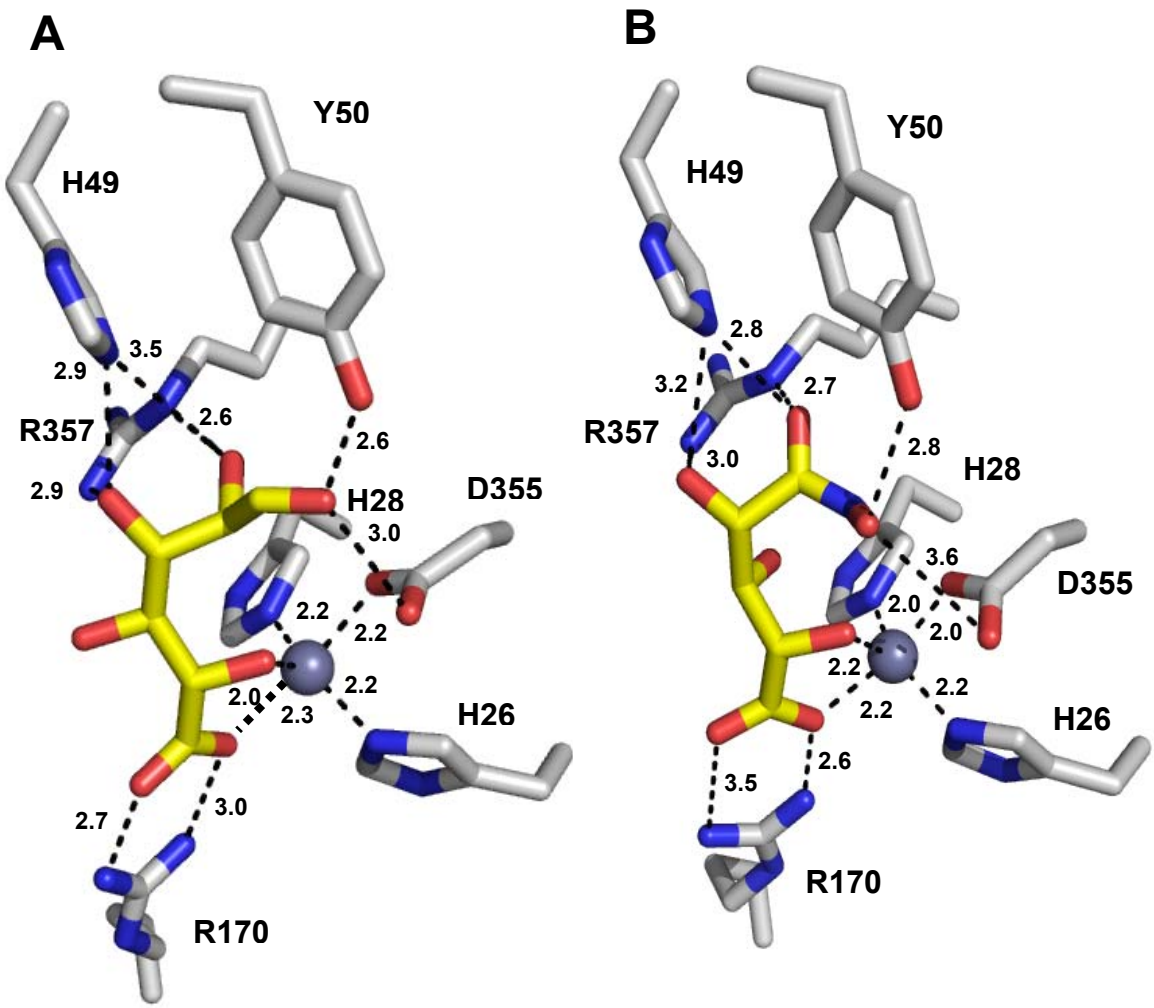


Figure 3.7: Active site of Bh0493 with (**A**) D-glucuronate, (**B**) hydroxamate of D-arabinarate, (**C**) D-arabinarate, and (**D**) D-fructuronate. Interactions between the enzyme and ligand with the distances are shown.

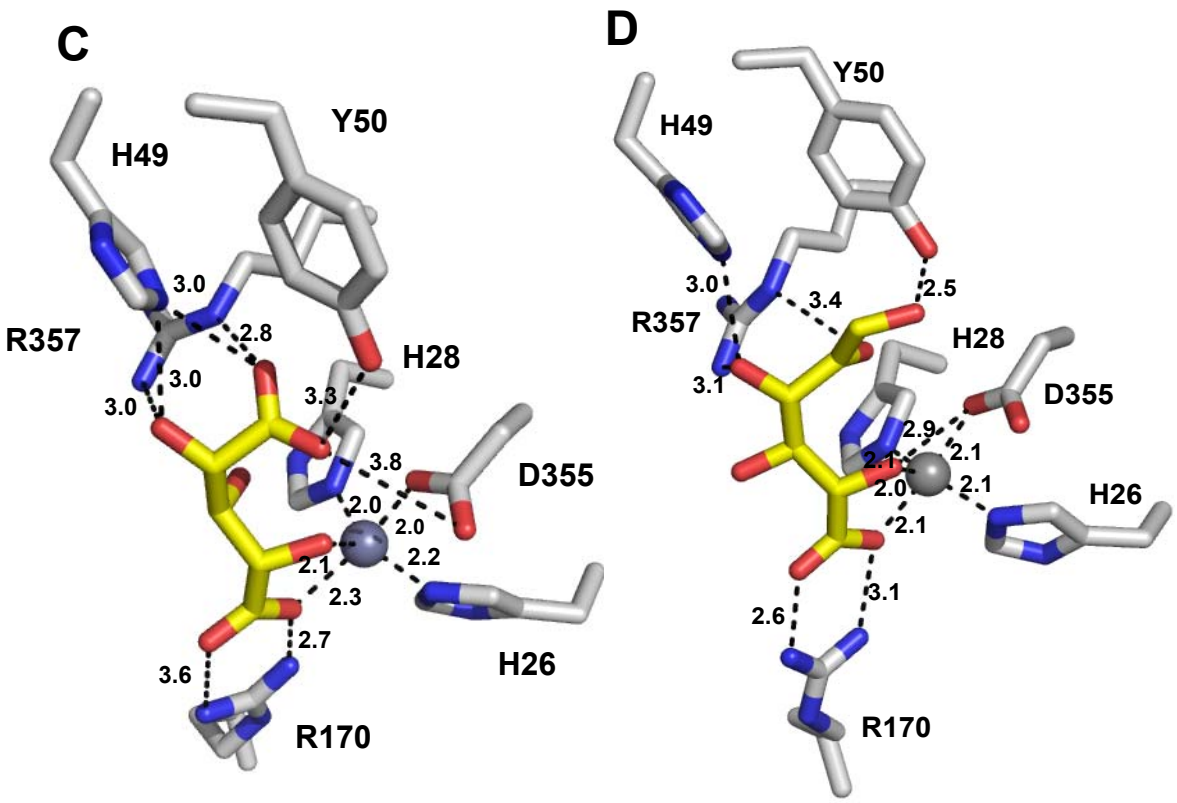


Figure 3.7: Continued.

DISCUSSION

The uronate isomerase from *E. c*oli has previously been shown to catalyze the isomerization of D-glucuronate and D-galacturonate to D-fructuronate and D-tagaturonate, respectively (*120*, *121*). With D-glucuronate the hydrogen at C-2 was shown to be transferred to the *pro*-R position at C-1 of the product (*98*). This hydrogen exchanges with solvent at a rate that is 4 orders of magnitude slower than the net interconversion of substrate and product. These results were interpreted to be consistent with a reaction mechanism that was initiated by proton abstraction at C-2, formation of a

cis-enediol(ate) intermediate and subsequent reprotonation at C-1 (*98*). This minimal reaction mechanism thus requires at least two different residues within the active site for these proton transfers. A general base (B_1:) is required for the removal of the proton at C-2 and delivery to C-1 and a general acid (B_2:H) is needed to shuttle a proton between the oxygens attached to C-1 and C-2 of the substrate/product pair. This transformation has been summarized in **Scheme 3.2**. A requirement for a minimum of two amino acids that must be in a specific state of protonation is experimentally supported by the measurement of the effects of pH on the magnitude of k_{cat} and k_{cat}/K_m for the conversion of D-glucuronate to D-fructuronate. The pH-rate profiles for both k_{cat} and k_{cat}/K_m are bell-shaped and indicate that one residue must be unprotonated and another protonated for catalytic activity. In these profiles the general base has a kinetic pK_a between 5.5 and 5.8 whereas the general acid has a kinetic pK_a between 9.5 and 10.2. Candidates for these residues were identified through the elucidation of the three-dimensional crystal structure of a uronate isomerase from *B. halodurans* (Bh0493) in the presence of D-glucuronate, D-fructuronate, and two mimics of the *cis*-enediol(ate) intermediate.

The crystal structure of Bh0493 determined with D-glucuronate has identified those residues in the active site that interact directly with the substrate. In this structure the zinc is ligated by three amino acids from the protein: the two conserved histidine residues from β-strand 1 and the aspartate from β-strand 8. The C-6 carboxylate group of the substrate is ligated to the zinc and also ion-paired with Arg-170. The hydroxyl group from C-5 is ligated to the zinc in the active site and hydrogen bonded to the aspartate from β-strand 8. At the other end of the substrate, Arg-357 hydrogen bonds

with the two hydroxyls from C-3 and C-2. The phenolic group of Tyr-50 forms a hydrogen bond with the carbonyl group at C1. The closest residue from the protein to the hydrogen at C-2 that must be abstracted during the chemical transformation is Asp-355 at 3.15 Å. In addition, His-49 is within hydrogen bonding distance to the hydroxyl at C-2 in the hydroxamate inhibitor (**II**) but not in the complex with the bound D-glucuronate (**I**) Thus, the most likely residues that are required for the isomerization reaction (in Bh0493) are His-49, Tyr-50, Asp-355 and Arg-357. In the *E. coli* enzyme these residues are equivalent to His-59, Tyr-60, Asp-412, and Arg-414, respectively. These residues, in addition to His-33 and His-35 (ligands to the zinc) Arg-186 (equivalent to R170 in Bh0493), His-297 (a conserved histidine at the end of β-stand 5 in most members of the amidohydrolase superfamily), Arg-302 and Trp-381, were mutated as probes of functional significance.

The mutation of specific residues within the active site of URI results in significant perturbations to the magnitude of the kinetic constants for substrate turnover. Changes to either of the two histidine residues that originate from the end of β-strand 1 weakens the binding of zinc to the active site and this results in a diminution of catalytic activity. This observation is consistent with the proposed role of zinc in the direct ligation of the substrate through the C-6 carboxylate and the hydroxyl from C-5. A drastic reduction in the affinity of the substrate occurs with the mutation of Arg-186 to methionine. In this case the K_m for the substrate increases by nearly two orders of magnitude and the value of k_{cat}/K_m is reduced by more than three orders of magnitude. This result is consistent with an ion-pair interaction between the C-6 carboxylate and the

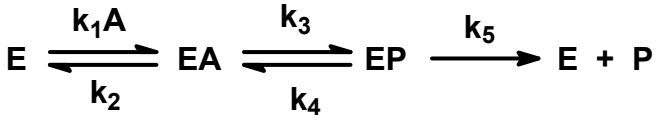
guanidino group of Arg-186 (R170 in Bh0493) that is observed in the X-ray structure of Bh0493. There is also a substantial increase in the K_m for D-glucuronate when His-297 is mutated to alanine or asparagine. This residue originates from the end of β-strand 5 and is highly conserved in nearly all members of the amidohydrolase superfamily but this residue is not conserved in Bh0493 and thus it is not easy to discern the effect on the structure of URI from *E. coli*.

The most dramatic reductions in k_{cat} occur with the mutation of Asp-412 to alanine or Arg-414 to methionine. Arg-414 is equivalent to Arg-357 from Bh0493, which is hydrogen bonded to the C-2 and C-3 hydroxyls of the bound D-glucuronate. Therefore, this residue is a suitable candidate for assisting in the movement of the proton from the hydroxyl at C-2 during the transformation to D-fructuronate. The other candidate for this process is the phenolic side chain from Tyr-60. This group is hydrogen bonded to the carbonyl oxygen at C-1 of D-glucuronate in the X-ray crystal structure. Mutation of Tyr-60 to phenylalanine results in reductions in k_{cat} and k_{cat}/K_m of about an order of magnitude. The only residue from the protein that is suitably positioned to function as the general base for the abstraction of the proton from C-2 and delivery to C-1 is Asp-412. Many members of the amidohydrolase superfamily that catalyze hydrolytic reactions have been shown to use this residue from the end of β-strand 8 to abstract a proton from the hydrolytic hydroxide or water (*60*, *68*, *70*, *71*).

Changes in solvent viscosity and isotopic substitution were used to address the degree of rate limitation on the bond breaking steps and product release (*123*, 1*24*, *126*, *127*). A simplified kinetic mechanism for the conversion of substrate to product for the

uronate isomerization reaction is presented in **Scheme 3.4**, where EA represents the enzyme-glucuronate complex and EP represents the enzyme-fructuronate complex. In this mechanism the expression for k_{cat} is given by $(k_3k_5)/(k_3 + k_4 + k_5)$. If one assumes that k_5 is inversely proportional to the relative solvent viscosity then the value of k_5, relative to the sum of k_3 and k_4 can be determined from a plot of the ratio of $^{o}(k_{cat})/^{\eta}(k_{cat})$ as a function of the relative solvent viscosity, η. The slope of this plot is equal to $(k_3 + k_4)/(k_3 + k_4 + k_5)$. For the wild-type enzyme the slope was found to be ~0.7 and thus the sum of k_3 and k_4 is greater than the product release step, k_5. This result is consistent with the release of the product as the rate limiting step for the wild type enzyme and the relatively small primary isotope effect for [2-^{2}H]-D-glucuronate. With the D412N and R412M mutants the slope of this plot, for changes in the relative value of k_{cat} as a function of solvent viscosity, is reduced substantially. This result is consistent with a significant reduction in the rate constants for the interconversion of the substrate/product pair to the point where k_5 is now greater than the sum of k_3 and k_4. This result is also consistent with the significant increase in the value of $^{D}k_{cat}$. With the D412N and R414M mutants, the primary isotope effects are 2.0 and 3.2, respectively, and thus with these two mutants the interconversion of the substrate/product pair is substantially rate limiting.

Scheme 3.4: A simplified kinetic mechanism for the conversion of substrate to product for the uronate isomerization reaction.



Mechanism of Action. Based upon the X-ray crystal structure of Bh0493 in the presence of the bound substrate, the catalytic properties of selected site-directed mutants and the stereochemical constraints for the conversion of D-glucuronate to D-fructuronate, a minimal chemical mechanism can be written for uronate isomerase. In the proposed mechanism for URI from *E. coli*, D-glucuronate is bound in the active site through electrostatic interactions to five highly conserved amino acid residues and the divalent cation. The carboxylate group at C-6 is coordinated to the divalent cation and Arg-186. The hydroxyl group at C-5 is also coordinated to the zinc. The hydroxyl groups at C-3 and C-2 interact with the side chain guanidino group of Arg-414 and the hydroxyl at C-2 is also apparently able to hydrogen bond to His-59. The carbonyl group at C-1 is hydrogen bonded to the phenolic oxygen of Tyr-60. The pH-rate profiles for URI are consistent with two amino acid residues that must be in a specific state of protonation for catalytic activity. The general base, with a kinetic pK_a of approximately 5.8, is consistent with Asp-412. This residue is conserved in all members of the amidohydrolase superfamily and for those enzymes that catalyze hydrolytic reactions it has been shown to initiate proton transfers from water/hydroxide to the leaving group

(*60*, *68*, *70*, *71*). The general acid, with a kinetic pK_a of approximately 10.2 from the pH-rate profiles, may be due to Tyr-60 but it is difficult to exclude a role for Arg-414.

In the simplest mechanism D-glucuronate binds in the open chain conformation in the active site and then Asp-412 abstracts the proton from C-2 as the carbonyl group of C-1 is protonated by Tyr-60 to form a *cis*-enediol intermediate. In the subsequent step, Tyr-60 abstracts the proton from the hydroxyl at C-2 as Asp-414 delivers a proton to C-1 with *pro*R stereochemistry. It should be noted that the mutation of Tyr-60 to alanine or phenylalanine diminishes k_{cat} by only a factor of 10. This reduction in rate is perhaps smaller than what may be expected for this role in catalysis. However, the lack of a primary deuterium isotope effect for the wild-type enzyme indicates that the chemical step is not rate-limiting. In addition, a water molecule may substitute for the phenolic group in the mutant enzymes. Additional mechanisms can be written that utilize a combination of Tyr-60, Arg-414, and His-59 to facilitate the proton movements between the oxygens at C-2 and C-1 within the *cis*-enediol intermediate and in the opening of the hemiacetal.

An alternative mechanism can be proposed in which Asp-412 abstracts the proton from the hydroxyl at C-5, which then abstracts a proton from C-2 to initiate the formation of the *cis*-enediol intermediate. The hydroxyl at C-5 is additionally activated through direct ligation to the bound zinc. These two variations are presented graphically in **Figure 3.8.** This latter mechanism is particularly attractive since it retains elements that are common to nearly all members of the amidohydrolase superfamily that have been interrogated mechanistically (*60*, *68*, *70*, *71*). The structural similarities in the

active site of uronate isomerase with those members of the amidohydrolase superfamily nicely illustrates the evolutionary link between those members of the AHS that catalyze hydrolytic reactions and those that catalyze 1,2-proton transfers. A structural alignment of the active sites of Bh0493/D-glucuronate and dihydroorotase (DHO)/dihydroorotate (PDB entry code 1j79) supports this proposition as illustrated in **Figure 3.9**. In this structural alignment one of the C-6 carboxylate oxygens from D-glucuronate is positioned in nearly the same place as the carboxylate oxygen of the bridging carbamate functional group in DHO. These oxygen atoms interact directly with the alpha-metal (M_α) in their respective structures. Moreover, the C-5 oxygen of D-glucuronate is positioned in the same way as the nucleophilic hydroxide in DHO and is oriented to favor proton abstraction by Asp-412 at a distance of 2.9 Å. The C-5 hydroxyl is 2.0 Å away from the hydrogen at C-2 of the bound substrate.

In addition to the crystal structures of Bh0493, the structures of uronate isomerase from *T. maritima* (Tm0064) and *C. crescentus* (Cc1490) are also available (PDB codes: 1j5s and 2q01, respectively). The three-dimensional structures of Bh0493, Tm0064 and Cc1490 are similar to each other despite the low sequence identities between Bh0493 and the other two proteins. All three proteins exist as a homotrimers. One significant difference in the three structures is that Bh0493 has a zinc ion in the active site, while the structure of Tm0064 contains electron density that corresponds to a water molecule that is in the same position as the zinc ion in the Bh0493 structure (*43*). The Cc1490 structure contains no metal ion or water molecules in the active site. Moreover, the histidine from β-strand 5 is conserved in Tm0064 and Cc1490 but not for

the Bh0493 enzyme. Despite the divergence in the amino acid sequences, the mechanism utilized by uronate isomerase family members is expected to be the same. Although the Tm0064 and Cc1490 structures do not contain a divalent cation, the metal ligand residues are present and therefore it is expected that they are capable of binding/utilizing a divalent cation for catalysis.

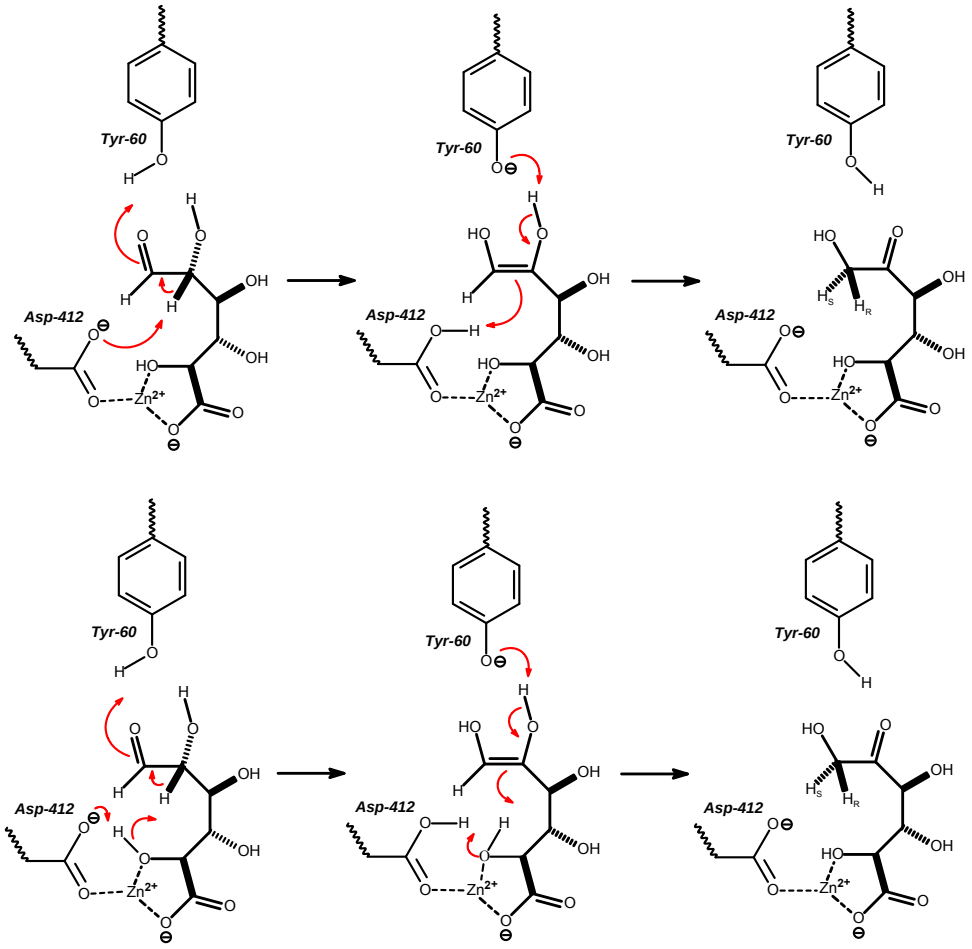


Figure 3.8: Proposed mechanisms for the isomerization of D-glucuronate by URI.

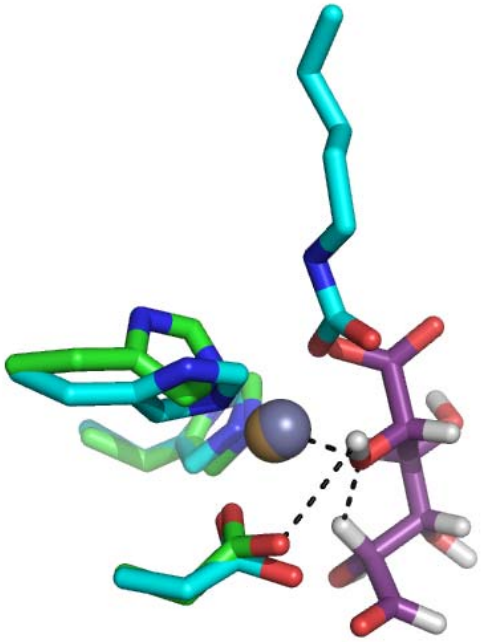


Figure 3.9: A structural alignment for portions of the active sites of Bh0493 (green) and DHO from *E. coli* (blue). The alpha metal (M_α) is shown as a yellow sphere for Bh0493 and in grey for DHO. The bound D-glucuronate for Bh0493, is shown in purple. The interactions between the C-5 oxygen of D-glucuronate and M_α, C-2 hydrogen and Asp-412 are indicated by dashed lines. The other ligands to the metal ion are shown as stick representations, including the HxH from β-strand 1 and aspartate from β-strand 8. The C-6 oxygen of D-glucuronate, aligns closely the the position occupied by one of the oxygens from the carboxyl group of the carboxylated lysine from strand 4 of DHO. The C-5 hydroxyl of D-glucuronate is orientated in the same position as the bridging hydroxide in the DHO structure.

CHAPTER IV

MECHANISTIC STUDIES OF SCO3058: A BACTERIAL RENAL DIPEPTIDASE FROM *STREPTOMYCES COELICOLOR*

The amidohydrolase superfamily (AHS) contains enzymes that catalyze primarily hydrolysis reactions of amide and ester bonds of sugars, amino acids, nucleic acids, and organophosphate ester substrates. A few rare families of the anidohydrolase superfamily catalyze nonhydrolytic reactions including isomerization and decarboxylation reaction (*21*,*31*). Structurally, all members of the AHS comprise of an $(\beta/\alpha)_8$-TIM barrel fold with a variation of a mononuclear or a binuclear metal center active site located at the C-terminal end of the beta strands. The metal center functions to activate the hydrolytic water as well as the scissile bond of the substrate for nucleophilic attack and to stabilize the resulting tetrahedral or trigonal bipyramidal transition state (*31*). Holm and Sander first discovered the amidohydrolase superfamily in 1997 based on the structural similarities among urease, adenosine deaminase, and phosphotriesterase (PTE) (*21*). Since then, more than 10,000 members of the AHS have been identified in prokaryotes and eukaryotes, and these enzymes have been shown to be involved in the catalysis of at least 40 different reactions.

One of the most common divalent cation centers within the AHS is the binuclear metal center found in enzymes such as urease, PTE, dihydroorotase, and isoaspartyl dipeptidase (*32-35*). In these proteins, the more buried metal, designated as M_α, is coordinated to the two histidines from the HxH motif at the end of β-strand 1 and an

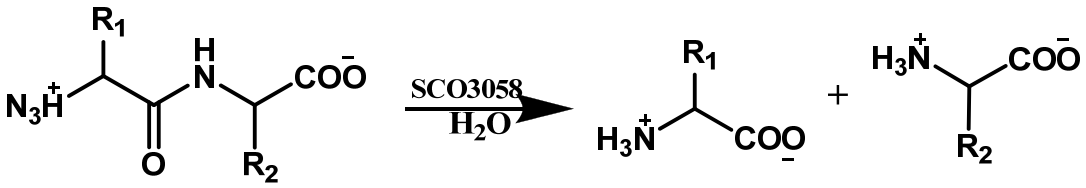
aspartate from strand 8. M_β, the more solvent exposed metal, is ligated to the two histidines from strands 5 and 6. The two metals are bridged by a solvent hydroxide as well as a carboxylated lysine or a glutamate that originates from strand 4.

Another distantly related binuclear member of the AHS is the renal dipeptidase from mammals. This family of enzymes was first identified and characterized in the 1970s based on its physicochemical properties (*100*). In addition to a wide range of dipeptides, renal dipeptidase is also involved in the renal metabolism of glutathione and its conjugates and is the sole enzyme that is responsible for the hydrolysis of penem and carbapenem β-lactam antibiotics (*101-106*). The crystal structure of the human renal dipeptidase was solved in the presence of the inhibitor cilastatin (PDB entry code: 1itu) (*40*). The quarternary structure of the hRDP is a homodimer with each polypeptide subunit consisting of 369 amino acid residues (42 kDa). The sequence of the hRDP shows a 75% identity to its homologs from rat and pig. From the crystal structure, the protein utilizes a HxD motif at β-strand 1 to bind to the M_α zinc ion instead of the more conventional HxH. Moreover, the residue that bridges the two metals is now a glutamate from β-strand 3 and not strand 4. The aspartate at the end of β-strand 8 is conserved but at a distance of 3.8 Å away, it is no longer coordinating to the M_α. The metal center of renal dipeptidases is illustrated in **Figure 4.1**. According to the X-ray structure, the aspartate residue is interacting with the hydrolytic water via a hydrogen bond. It was also suggested by Nitanai et al that the hydrolytic water is activated by the conserved aspartate from β-strand 8. The deprotonated water then nucleophilicly attacks the carbonyl carbon of the peptide bond to form a tetrahedral intermediate. It was

proposed that the positively charged His-152 acts as an oxyanion hole to stabilize the negatively charged intermediate (*40*). This mechanism of hydrolysis is distinctly different from the mechanism proposed for the enzymes in the AHS in which the tetrahedral intermediate is stabilized by the solvent exposed M_{β}.

The first renal dipeptidase-like gene *acdp* from a prokaryote was cloned from *Acinebacter calcoaceticus* (*107*). This enzyme has a low sequence identity of 23% to the human renal dipeptidase, but the essential residues identified in the mammalian dipeptidases are all conserved. Unlike its mammalian homologs, the gene product of *acdp* could not hydrolyze the unsaturated dipeptide glycyldehydrophenylalanine; instead the protein showed substrate specificity towards various dipeptides with a preference for dipeptides contaiting a D-amino acid at the C-terminus (*107*). In this chapter, the kinetic and structural properties of another bacterial renal dipeptidase-like enzyme from *Streptomyces coelicolor* are described. According to the sequence alignment shown in **Figure 4.2**, this protein is 45% and 22% identical in sequences to the hRDP and *acdp*, respectively. Substrate specificity analysis of Sco3058 against 55 dipeptide libraries indicated promiscuous dipeptidase activity with a preference for the hydrolysis of dipeptides with D-Leu, D-Met, D-Arg, and D-Lys at the N-terminus and either L- or D-Asp at the C-terminus. Overall, this enzyme is promiscuous towards a wide range of dipeptides containing L- and D-amino acids but the libraries with L-Xaa D-Xaa dipeptides were hydrolyzed the fastest. The best substrate tested was the dipeptide L-Arg-D-Asp, with kinetic constants of 68 s^{-1} for k_{cat} and (4.1 ± 0.5) x 10^5 for k_{cat}/K_m. The hydrolysis reaction catalyzed by Sco3058 is shown in **Scheme 4.1**.

Scheme 4.1: Hydrolysis reaction catalyzed by Sco3058.



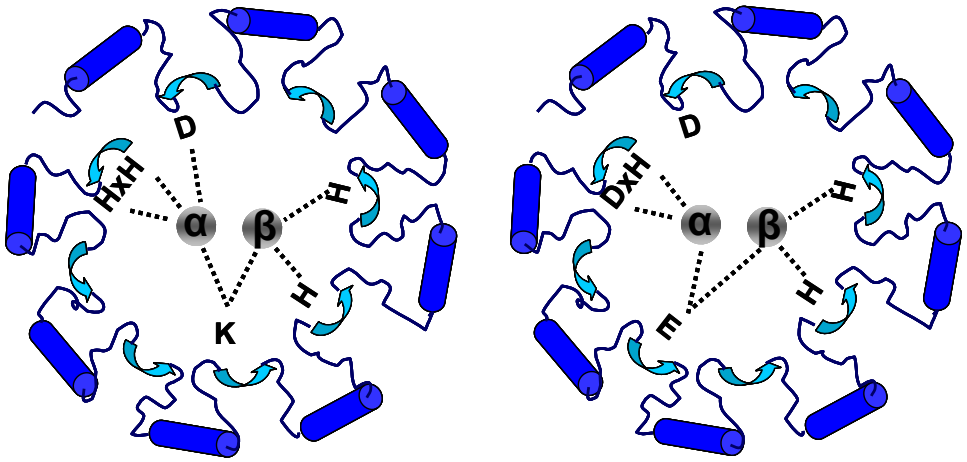


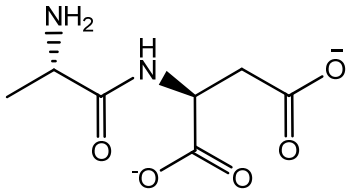
Figure 4.1: Variations of the binuclear metal center of the AHS. The metal center on the left is the most common binuclear center and is represented by enzymes such as DHO and PTE while the one on the right is unique for renal dipeptidase.

```
Sco3058  ----------------MTSLEKARELLREFPVVDGHNDLPWALREQVRYDLDARDIAADQ
hRDP     MWSGWWLWSLVAVCTADFFRDEAERIMRDSPVIDGHNDLPWQLLDMFNNRLQDERANLTT
acdp     ------------------------MKPSHIPVFDGHNDALTRL--WLSDHADPVHAFLHE

Sco3058  SAHLHTDLARLRSGGVGAQYWSVYVR-----------------SDLPG--AVTATLEQI
hRDP     LAGTHTNIPKLRAGFVGGQFWSVYTP-----------------CDTQNKDAVRRTLEQM
acdp     RLAGHLDLKRCQQAGFVGGMFAIFLPPFSYVQQHHSNKLFDQDATDFTQQQIEQICLEQL

Sco3058  DCVRRLIDRHPGELRAALTAADMEAARAEGRIASLMGAEGGHSIDNSLATLRALYALGVR
hRDP     DVVHRMCRMYPETFLYVTSSAGIRQAFREGKVASLIGVEGGHSIDSSLGVLRALYQLGMR
acdp     DIAHQLAQR-STDIQICTSVQDIQSCVNTQKLGIVLHMEGAEALQQNSDLLDVFYDRGLR

Sco3058  YMTLTHNDNNAWADS-----ATDEPGVGGLSAFGREVVREMNREGMLVDLSHVAATTMRD
hRDP     YLTLTHSCNTPWADNWLVDTGDSEPQSQGLSPFGQRVVKELNRLGVLIDLAHVSVATMKA
acdp     SIGPLWNRPSLFGHGLNAKFPHSPDTGAGLTQDGKAFIKRCADKKMVIDVSHMNEQAFWD

Sco3058  ALDTSTAPVIFSHSSSRAVCDHPRNIPDDVLERLSANGGMAMVTFVPKFVLQAAVDWTAE
hRDP     TLQLSRAPVIFSHSSAYSVCASRRNVPDDVLRLVKQTDSLVMVNFYNNYISCTNK-----
acdp     TVDLLQQPIVATHSNAHALCPQARNLTDQQLKAIRESKGIVGVNFDVAFLR---------

Sco3058  ADDNMRAHGFHHLDSSPEAMKVHAAFEERVPRPVATVSTVADHLDHMREVAGVDHLGIGG
hRDP     ---------------------------------ANLSQVADHLDHIKEVAGARAVGFGG
acdp     -------------------------SDGQRNANTSIDVILDHLDYLLDRLGEEHVAFGS

Sco3058  DYDGTPFTPDGLGDVSGYPNLIAELLDRGWSQSDLAKLTWKNAVRVLDAAEDVSRGLRAA
hRDP     DFDGVPRVPEGLEDVSKYPDLIAELLRRNWTEAEVKGALADNLLRVFEAVEQASN-LTQA
acdp     DFDGA-LIGTELEDVMGLHKLIHRMQQRAYSSELIEKLCFTNWINVLYRILGE-------

Sco3058  RGPSNATIEQLDGTAAEQPEG--------------------------
hRDP     PEEEPIPLDQLGGSCRTHYGYSSGASSLHRHWGLLLASLAPLVLCLSLL
acdp     -------------------------------------------------
```

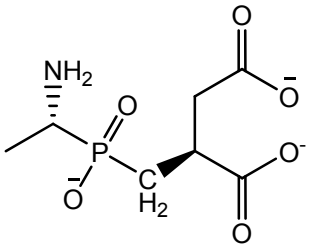
Figure 4.2: Sequence alignment of Sco3058, hRDP and acdp. Highlighted in yellow are the represented β-strands. Residues in red are the metal ligands which are also the conserved amino acids in the AHS.

MATERIALS AND METHODS

Materials. All chemicals were obtained from Sigma Aldrich, unless otherwise stated. The coupling enzymes glutamic-oxaloacetic transaminase and malate dehydrogenase were purchased from Sigma Aldrich and Calbiochem, respectively. The substrate L-Arg-L-Asp and the inhibitor L-Ala-L-Asp phosphinodipeptide were synthesized by Dr. Chengfu Xu. ICP standards were obtained from Inorganic Ventures Inc. Metal analyses were done using inductively couple plasma mass spectrometry (ICP-MS) from Perkin-Elmer. Oligonucleotide syntheses and DNA sequencing were performed by the Gene Technologies Lab of Texas A&M University.



L-ala-L-asp



L-ala-L-asp phosphinodipeptide

Cloning and Site-Directed Mutagenesis. The gene Sco3058 from *S. coelicolor* was cloned into pET 30 vector by Jennifer Cummings. Site-directed mutagenesis was performed on the gene at residues Asp-22, His-150, Arg-223, and Asp-320 using the Quik-Change kit from Stratagene. The wild type and the mutant enzymes were overexpressed in Rosetta 2 DE(3) competent cells purchased from Stratagene.

Protein Purification. The pET 30 vector containing the Sco3058 gene was transformed into Rosetta 2 DE(3) cells, and the cells containing the vector were selected on a kan^+ LB plate. A single colony was inoculated into 5-mL LB and after overnight incubation at 37 ºC, the cells were transferred into 1 liter of TB medium containing 50 ng/μL of kanamycin. Prior to induction, the cells were charged with 1 mM metal of interest (zinc or cadmium). After the optical density reached 0.6 at 600 nm, the expression of the protein was induced by the addition of 0.5 mM IPTG at room temperature and shaken overnight. The cells were centrifuged and resuspended in 50 mM Tris buffer at pH 7.5 and 100 μg/mL of phenylmethylsulphonyl fluoride (PMSF) serine protease inhibitor. The cells were lysed by sonication, and the soluble protein was isolated from the insoluble components by centrifugation. The nucleic acids were precipitated by the addition of 2% (w/v) protamine sulfate followed by centrifugation. The protein of interest was precipitated with 60% saturation of ammonium sulfate. After centrifugation, the pellet was resuspended in a minimum quantity of 50 mM Tris buffer, loaded onto a Superdex 200 gel filtration column, and eluted with the same buffer. The fractions containing the protein Sco3058 were identified by SDS-PAGE and activity assays, pooled, loaded onto a Resource Q anion exchange column, and eluted with a gradient of NaCl in 20 mM Tris buffer, pH 7.5. The active fractions of the purified enzyme were pooled, concentrated, and stored at -20 C.

Enzyme Assays. The substrate used for the determination of enzyme activity was L-Arg-L-Asp, unless otherwise stated. The hydrolysis of L-Arg-L-Asp was monitored by the coupling of the formation of L-aspartate to the oxidation of NADH in a system that involves glutamic-oxaloacetic transaminase and malate dehydrogenase. The change in the concentration of NADH was measured spectrophotometrically using a SPECTRAmax-340 plate reader (Molecular Devices Inc.) by following the decrease in absorbance at 340 nm. The standard assay conditions contained 100 mM Tris (pH 7.5), varying concentrations of L-Arg-L-Asp, 0.36 mM NADH, 7 units of glutamic-oxaloacetic transaminase, 1 unit of malate dehydrogenase, 3.7 mM α-ketoglutarate, 100 mM KCl and Sco3058 in a final volume of 250 μL at 30 ºC. The pH-dependence of the kinetic parameters, k_{cat} and k_{cat}/K_m, were measured over the pH range of 5.50 – 10.00 at 0.25 pH intervals for the Zn^{2+} and Cd^{2+} enzymes. The buffers used for the pH-rate profiles were MES, BIS-TRIS, TRIS and CABS. The pH values were recorded after the completion of the assays. The effects of solvent viscosity on the activity of Sco3058 enzyme were determined at pH 7.5 using sucrose as the microviscogen at 25 ºC. The concentrations of sucrose were 0%, 10%, 14%, 20%, 24%, and 32% (w/w), with corresponding η_{rel} values of 1.0, 1.3, 1.5, 1.9, 2.2, and 3.2. The solvent isotope effects on the kinetic parameters of the wild type Sco3058 were measured in 99% D_2O at a pD of 7.90.

Data Analysis. The kinetic parameters, k_{cat} and k_{cat}/K_m, for the renal dipeptidase Sco3058 with L-Arg-L-Asp as the substrate were determined by fitting the initial velocity data to equation 4.1 where ν is the initial velocity, E_t is the total enzyme concentration, k_{cat} is the turnover number, [A] is the substrate concentration, and K_m is the Michaelis constant. The profiles for the variation of k_{cat} and k_{cat}/K_m for the zinc enzyme with pH were fit to equation 4.2, where c is the pH-independent value of y, K_a and K_b are the dissociation constants of the ionizable groups and H is the proton concentration. For the cadmium enzyme, the profile for the variation of k_{cat} with pH was fit to equation 4.3, while the profile for the variation of k_{cat}/K_m with pH was fit to equation 4.2.

$$v / E_t = (k_{cat}\,[A])/(K_a + [A]) \tag{4.1}$$

$$\log y = \log (c / (1 + (H / K_a) + (K_b / H))) \tag{4.2}$$

$$\log y = \log (c / (1 + (K_b/H))) \tag{4.3}$$

Crystallization and Data Collection. A search for crystallization conditions of Sco3058 was conducted using the Wizard I crystallization matrices purchased from Emerald BioSystems and Crystal Screen HT formulations obtained from Hampton Research via hanging drop method of vapor diffusion at room temperature. For the Wizard I kit, the best crystals were observed with a condition that has 1.0 M citrate as the precipitant and 0.1 M imidazole buffer at pH 8.0. Subsequent optimization trials resulted in large single crystals with precipitant solutions of 0.7-0.9 M citrate, 0.1 M imizadole pH 8.0, and 12.0 mg/mL enzyme. With the Crystal Screen formulation, crystals were obtained for a condition that has 0.2 M magnesium chloride hexahydrate,

0.1 M Tris at pH 8.5, and 30% polyethylene glycol (PEG) 4000. Single crystals were obtained with subsequent optimizations from hanging drops with precipitant solutions of 18-24 % PEG-4000 containing 0.2 M $MgCl_2$·H20 buffered at pH 8.5 with 0.1 M Tris. The enzyme concentrations in the individual drops were typically around 5.5 – 6.0 mg/mL. Crystals were observed within 2 weeks. The crystal structure of the L-Ala-L-Asp phosphinodipeptide inhibitor was obtained from soaking the crystal in the presence of 10 mM inhibitor. These crystals also exibited diffraction consistent with the space group $P3_121$, with one molecule of Sco3058 per asymmetric unit and 62% solvent. Prior to data collection, the citrate-grown crystals were transferred to cryoprotectant solution composed of 80% of their mother liquors and 20% glycerol. The PEG-grown crystals were cryoprotected with 85-90% of the well solutions and 15-10% ethylene glycol. After a few seconds of incubation, the crystals were flashcooled and stored in liquid nitrogen. Data for the glycerol/citrate bound and phosphinodipeptide bound Sco3058 were collected to a maximum resolution of 1.7 and 1.9 Å, respectively, using a single-wavelength SSRL 7-1 beamline (Stanford Synchrotron Radiation Lightsource). For the citrate-bound data, intensities were integrated and scaled with DENZO and SCALEPACK programs (*128*). For the inhibitor-bound data, Imosflm was used to integrate the data (*129*). Scaling was done with Scala from the CCP4 suite (*130*,*131*). The data collection statistics are given in **Table 4.1**.

Structure Determination and Model Refinement. Both of the Sco3058 structures were solved by molecular replacement. The Sco3058-glycerol-citrate structure was solved using the program Amore (*131*,*132*), and the renal dipeptidase from human (PDB code 1itq) was used as a template. For the Sco3058-phosphinodipeptide structure, the program Phaser from CCP4 suite was utilized (*131*,*133*), and the template was taken from the Sco3058-citrate structure. For the Sco3058-citrate structure, manual model building and solvent building were done with Xfit and refinement with CNS (*134-136*). For the Sco3058-phosphinodipeptide structure, iterative cycles of refinement were done using the CCP4 suite programs Refmac and solvent building was done with Coot, both from the CCP4 suite (*131*,*137-138*). The two model complexes, Sco3058-glycerol-citrate and Sco3058-phosphinodipeptide, were refined at 1.7 and 1.9 A, respectively. The Zn^{2+} ions bound in the active sites were clearly visible in every molecule of every Sco3058 complex, and electron densities that correspond to the bound inhibitors were evident in the active site. Final crystallographic statistics for the two complexes are shown in **Table 4.1**.

Table 4.1: Crystallographic statistics for Sco3058-inhibitor complexes

	Sco3058-citrate	Sco3058-L-Ala-L-Asp phosphinodipeptide
Data collection		
Space group	P3121	P3121
# of mol. in asym. unit	1	1
Cell dimensions		
a, b, c (Å)	97.29, 97.29, 104.41	96.69, 96.69, 104.60
α, β, γ (º)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Resolution (Å)	1.7	1.9
# of unique reflections	62182	43310
Rmerge	.071	0.18
I/σI	29.7	18.5
Completeness (%)	98.3	97.0
Refinement		
Resolution (Å)	20.0-1.7	44.4-1.9
R_{cryst}	16.7	19.6
R_{free}	20.3	22.2
Protein atoms	3055	3015
Waters	504	356
Rmsd, bond lengths (Å)	0.018	0.019
Rmsd, bond angles (º)	2.4	1.5
Bound inhibitor	Citrate, glycerol	L-ala-L-asp phosphinodipeptide
Inhibitor atoms	19	15
Bound ions	2 Zn^{2+}	2 Zn^{2+}

RESULTS

pH-rate profiles. The kinetic constants for the hydrolysis of the dipeptide L-Arg-L-Asp by the zinc and cadmium enzymes were obtained as a function of pH. The zinc and cadmium enzymes were obtained by the addition of 1 mM $ZnCl_2$ or $CdCl_2$ to the cells before induction with ITPG. The metal content was measured with ICP-MS. For the zinc enzyme, 1.8 equivalents of Zn^{2+} were obtained per subunit. The cadmium enzyme contained 1.5 equivalent of Cd^{2+} per subunit. The pH-rate profiles for the effects of pH on k_{cat} and k_{cat}/K_m are presented in **Figures 4.3A** and **4.3B**, respectively. The pH profiles for the zinc enzyme are bell-shaped and are consistent with a single functional group that must be unprotonated for activity and another functional group that must be protonated for catalytic activity. From a fit of the data to equation 4.2, the kinetic pK_a values from the k_{cat}/K_m plot are 7.6 ± 0.4 for both the deprotonation and protonation groups for the zinc enzyme. Since the two pK_a's are less than 2 units apart, it is not possible to distinguish one from the other. The pK_a values for k_{cat}/K_m are 7.4 ± 0.1 and 9.8 ± 0.1 for the cadmium enzyme. From the plot of k_{cat} vs. pH, kinetic pK_a values of 6.2 ± 0.1 and 8.4 ± 0.1 were obtained for the zinc enzyme. The profile for the effect of pH on k_{cat} for the cadmium enzyme exhibits loss of activity with an ionization of a group that has a pK_a value 9.6 ± 0.1. This pK_a value is obtained by fitting the data to equation 4.3.

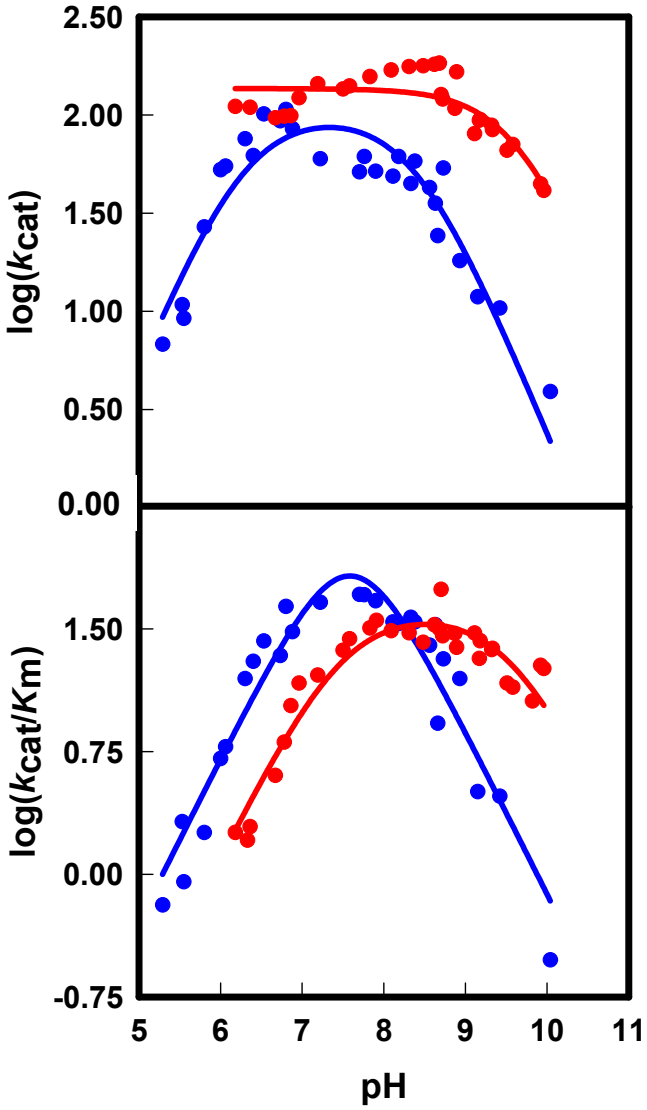


Figure 4.3: pH-rate profiles of the zinc (blue) and cadmium (red) Sco3058. (**A**) The effect of pH on k_{cat} and (**B**) the effect of pH on k_{cat}/K_m.

Site-directed Mutagenesis. Site-directed mutagenesis was utilized to identify residues that are critical for enzymatic activity. Several conserved residues including Asp-22, His-150, Arg-233, and Asp-320 were altered and the mutants were

characterized and kinetic results are presented in **Table 4.2**. When residue Asp-22 from the HxD motif at the end of β-strand 1 was changed to a histidine, no activity was detected even though it contained 2 equivalents of metal, suggesting a secondary role of this aspartate residue in addition to metal binding. His-150 was mutated to asparagine and alanine, and both of these mutants exhibited a significant decrease in k_{cat} (200 and 20 fold, respectively) and the K_m more than doubled. The conserved Arg-233 was changed to lysine and methionine. The lysine mutant was reduced in activity by 3 orders of magnitude relative to the wild type. No hydrolytic activity was detected with the methionine mutant. The conserved Asp-320 from β-strand 8 was modified to asparagine and alanine, and no activity was seen with either mutant.

Table 4.2: Kinetic parameters for the Sco3058 mutants

Enzyme	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
WT	132 ± 2	1.8 ± 0.1	(7.2 ± 0.4) x 10^4
D22H	na	na	<0.3
H150N	0.7 ± 0.01	4.9 ± 2.1	(1.4 ± 0.4) x 10^2
H150A	7.2 ± 0.4	4.3 ± 0.5	(1.7 ± 0.2) x 10^3
R223K	0.4 ± 0.04	8.8 ± 1.4	46 ± 9
R223M	na	na	<0.3
D320N	na	na	<0.3
D320A	na	na	<0.3
These data were obtained at 30 ºC at pH 7.5 with L-Arg-L-Asp substrate. *na = not applicable			

Solvent Isotope Effect and Solvent Viscosity Effect. Solvent isotope effects were utilized to determine the effect of proton transfers during turnover. The kinetic parameters were obtained for the wild type enzyme in H_2O and D_2O with L-Arg-L-Asp as the substrate. The double reciprocal plots are presented in **Figure 4.4**. The measured solvent isotope effects for $^{D_2O}k_{cat}$ and $^{D_2O}(k_{cat}/K_m)$ are 1.30 ± 0.02 and 1.10 ± 0.05, respectively. Alterations in solvent viscosity were utilized to probe the degree of rate limitation by the binding and dissociation of products and substrates on the kinetic constants of Sco3058 (*123*). The effects of changes in solvent viscosity on k_{cat} were made by the addition of sucrose (*124*). A plot of $^{o}k_{cat}$ / $^{\eta}k_{cat}$ versus the relative solvent viscosity for the wild-type enzyme exhibits a slope of 0.05 ± 0.03. The kinetic data are presented in **Figure 4.5**.

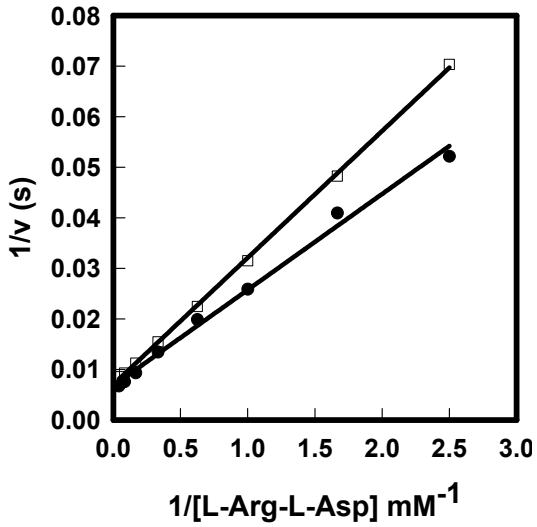


Figure 4.4: Double-reciprocal plot for the solvent isotope effect (H_2O *vs* D_2O) where $1/v$ (s) is plotted against 1/[L-Arg-L-Asp] (in mM^{-1}). The data sets in H_2O are represented by circles, and the squares correspond to the values obtained in D_2O.

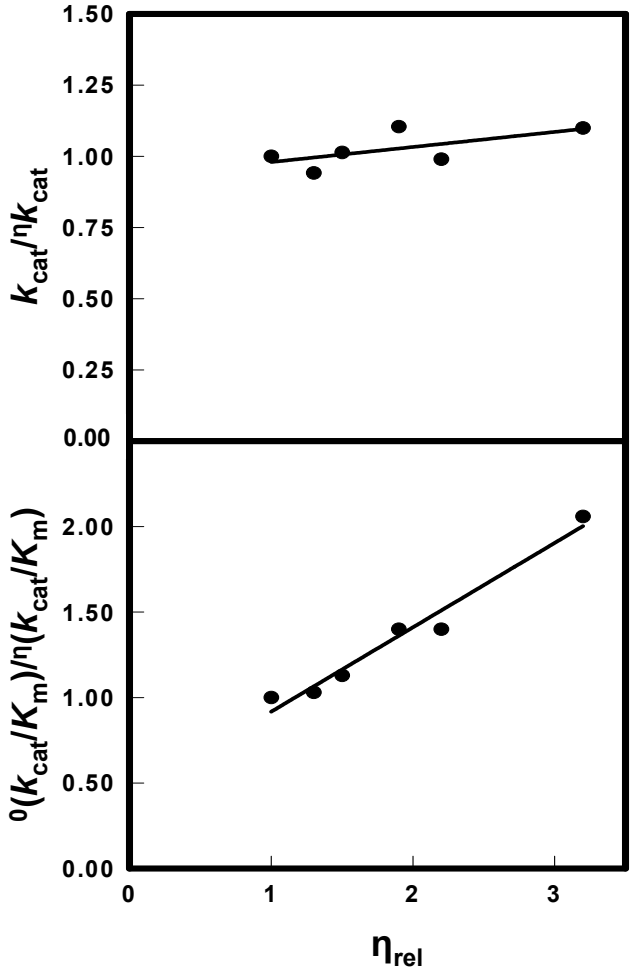


Figure 4.5: The effect of viscosity on the relative values of (**A**) k_{cat} and (**B**) k_{cat}/K_m for Sco3058 using sucrose as the microviscogen.

Structures of Sco3058-inhibitor Complexes. The crystal structures of Sco3058 complexed with glycerol/citrate and the inhibitor L-Ala-L-Asp phosphinodipeptide were solved to 1.7 and 1.9 Å, respectively. Both structures of Sco3058 complexes revealed that the zinc ion in the α site is coordinated to the HxD at the end of β-strand 1. The zinc ion in the β site is ligated to the two histidines from strands 5 and 6. The two metals are

bridged by the glutamate from β-strand 3 and a hydroxide/water molecule. The conserved aspartate from strand 8 (Asp-320) is 3.8 Å away from the M_α and therefore does not coordinate to the metal; instead it interacts with the bridging solvent molecule via hydrogen bonding at a distance of 3.0 Å. A ribbon diagram of the structure of Sco3058 is shown in **Figure 4.6**. The active sites of Sco3058 in the presence of glycerol/citrate and phosphinodipeptide inhibitor are shown in **Figure 4.7**.

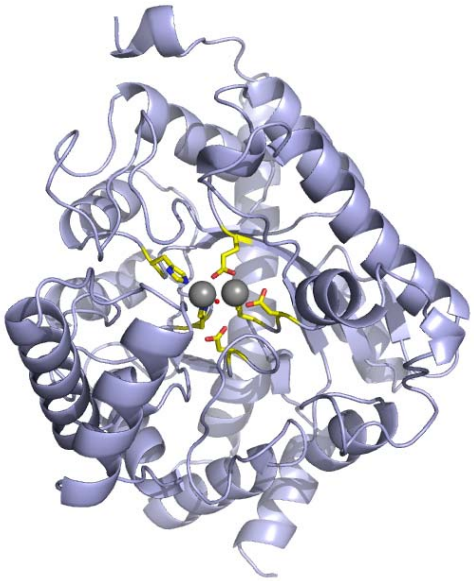


Figure 4.6: Ribbon diagram of the structure of Sco3058. The active site metals are shown in grey. The metal ligands are indicated in yellow.

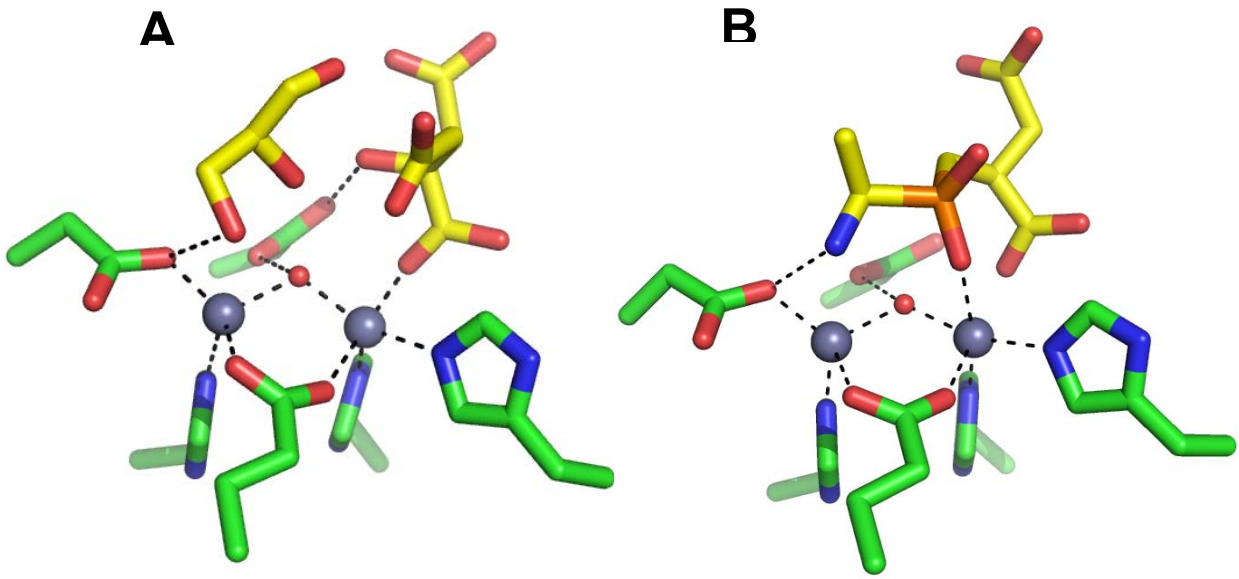


Figure 4.7: The active site of Sco3058 in the presence of bound (**A**) citrate and glycerol and (**B**) L-Ala-L-Asp-phosphinodipeptide.

In the glycerol/citrate bound structure, the α-metal is tetrahedral in coordination while the β-metal is in a distorted trigonal pyramidal geometry. Aside from the four ligands described above, the β-metal is coordinated to a carboxyl oxygen of citrate. This carboxyl group is also interacting with the conserved Arg-223. For the remaining two carboxylates of citrate, one is ion-paired to His-150 and the other one is interacting with Thr-324. The C3-OH of citrate is 2.81 Å from the conserved Asp-350. This hydroxyl from citrate is also interacting with the C2-OH of glycerol. In addition, the C1-OH of glycerol is within hydrogen bonding distance (2.97 Å) to Asp-22 from β-strand 1 and 2.64 Å away from the α-metal. A schematic drawing of the active site of Sco3058 in the presence of glycerol and citrate is illustrated in **Figure 4.8**.

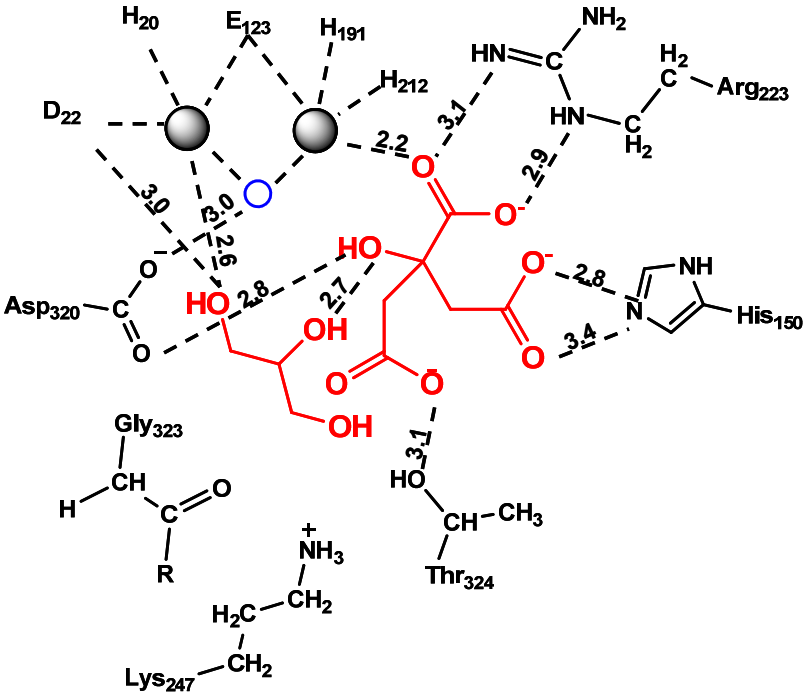


Figure 4.8: Schematic representation of the binding of citrate and glycerol in the active site of Sco3058. Glycerol and citrate are in red, the bridging solvent is in blue, and the residues on the protein are in black.

According to the Ala-Asp phosphinodipeptide structure, the electron density of the inhibitor revealed that the stereochemistry of the phosphinodipeptide corresponds to the compound L-Ala-L-Asp phosphinodipeptide. The F_o-F_c electron density map of the Sco3058-phosphinodipeptide structure is displayed in **Figure 4.9**. In the active site of the Sco3058-phosphinodipeptide complex, the β-zinc ion is coordinated by six ligands in a hexagonal fashion. Two of these interactions are from two different oxygens of the inhibitor, one from the phosphate and one from the C-terminal carboxylate. The rest are

interactions to Glu-123, His-191, His-212, and the bridging hydroxide. The α-metal is coordinated by four ligands in a distorted tetrahedral fashion to His-20, Asp-22, Glu-123, and the bridging hydroxide. The amino group of the inhibitor is approximately 2.8 Å away from the α metal. The C-terminal carboxylate of the inhibitor makes two interactions with the guanidino side chain of Arg-223. The residue His-150 is hydrogen bonded to the oxygens of the phosphate at a distances of 2.7 and 3.0 Å. The side chain carboxylate of the C-terminal aspartate interacts with Thr-324 as well as Lys-247. A schematic drawing of the active sites of Sco3058-glycerol-citrate and Sco3058-phosphinodipeptide complexes are illustrated in **Figure 4.10**.

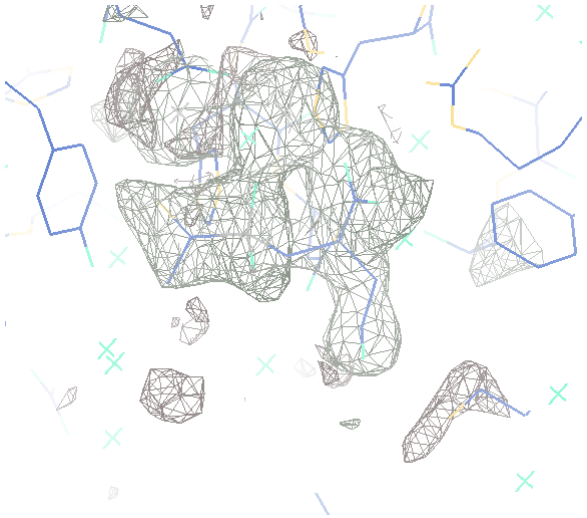


Figure 4.9: The F_o-F_c electron density map of the Sco3058-phosphinodipeptide superimposed on the structural model showing L-Ala-L-Asp phosphinodipeptide binding in the active site of Sco3058. The map in gray was calculated with phases from a model without the phosphinodipeptide, the two zinc ions, and the bridging water molecule.

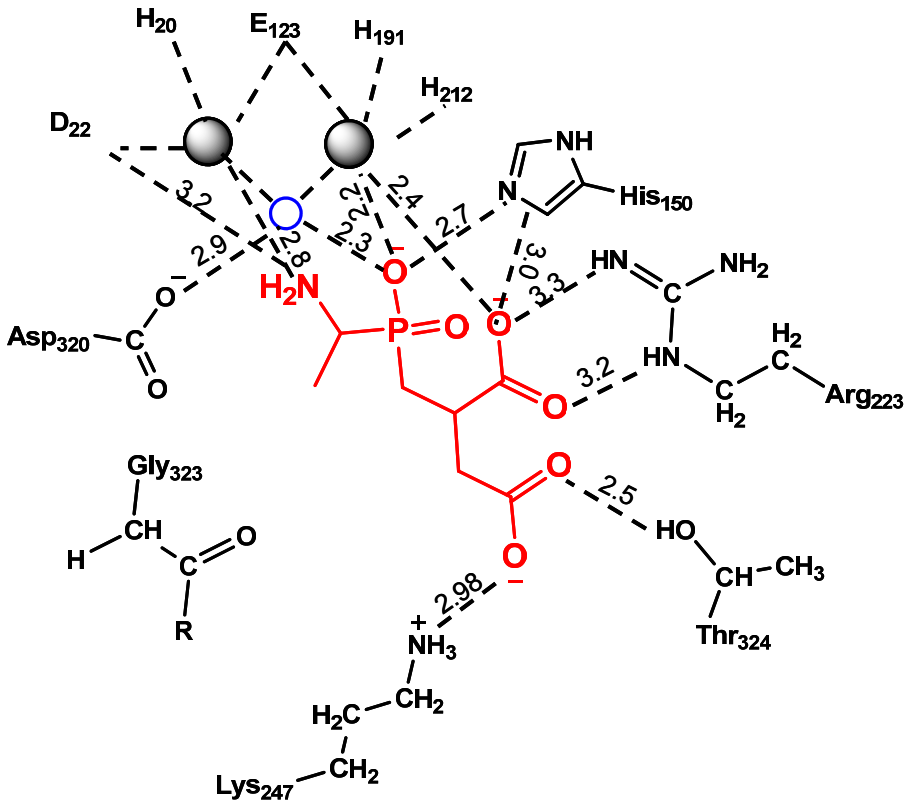


Figure 4.10: Schematic drawing of the active site of Sco3058 in the presence of L-Ala-L-Asp-phosphinodipeptide inhibitor. The inhibitor is colored red, the bridging solvent is in blue, and the residues on the enzyme are in black.

DISCUSSION

The enzyme renal dipeptidase from human has previously been crystallized in the absence and presence of the inhibitor cilastatin (*40*). This enzyme was identified as an amidohydrolase enzyme based on structural similarities to other enzymes of the amidohydrolase superfamily. According to the hRDP structure, the protein contains a TIM-barrel structural motif in which the active site of the enzyme enclosed a binuclear metal center of two zinc ions. The zinc ion in the α-site is ligated to HxD motif at the end of β-strand 1and is bridged to the β-metal via Glu-123 from strand 3 and a water molecule. In addition, the β-metal is coordinated to His-198 from strand 5 and H-219 from strand 6. Recently, a bacterial renal dipeptidase (Sco3058) have been identified and characterized. This enzyme has a 40% sequence identity to its human homolog. Although the hRDP is known to hydrolyze *S*(substitute)-L-cysteinyl-glycine adducts, the homologous protein from *Streptomyces coelicolor* (Sco3058) is promiscuous with its dipeptidase activity. However, this enzyme is also considered to be fairly specific in that it only works on dipeptides. In addition, the crystal structures of Sco3058-inhibitor complexes suggest that the N-terminus amino group of the substrate is interacting with the Asp-22 from β-strand 1 HxD motif. The elimination of hydrolytic activities from the D22N and D22A mutants provides further support for this hypothesis. Together, these results highlight the importance of the amino group at the N-terminus for substrate binding and catalysis.

The catalytic mechanism of Sco3058 was investigated using kinetic and crystallographic methodologies. The pH-rate profiles for the wild type enzyme were

investigated in order to enhance our understanding of the ionization state of the bridging solvent molecule in the resting state of the enzyme. The pH-rate profiles for either k_{cat} or k_{cat}/K_m show that one residue must be unprotonated and another protonated for catalytic activity for the zinc enzyme. The loss of activity at low pH is indicative of protonation of the bridging hydroxide that is utilized for substrate hydrolysis. If this were the case, then it is expected that the pK_a obtained from the pH-rate profiles will be a function of the active site divalent cation. In the log k_{cat} profile for the zinc enzyme, the pK_a obtained at low pH is 6.2. The pK_a for the cadmium enzyme, however, is not clear due to the limitation of the assays at low pH. In the log k_{cat}/K_m profile for the zinc enzyme, the pK_a obtained at low pH is 7.6 +/- 2. When zinc is replaced with cadmium, the kinetic pK_a is 7.4 +/- 0.1. With a wide error margin for the zinc enzyme, it is not possible for clearly state the effect of different metals on the pK_a of the ionizable group. It is expected, however, that the pK_a for the cadmium enzyme at low pH to be higher than for the zinc enzyme due to the ability of zinc to lower the pK_a of water relative to cadmium if the loss of activity at low pH is indeed due to protonation of the bridging hydroxide (*139*). The dependence of kinetic pK_a's on the identity of the active site metal ion has also been found with many enzymes in the AHS such as IAD and DHO (*60*,*71*). The functional group that ionizes at higher pK_a could possibly be the amino group from the substrate or a residue from the active site of Sco3058, possibly Lys-247.

The crystal structure of Sco3058 with bound citrate and glycerol was solved to 1.7 Å. The presence of citrate in the active site was due to the high concentration of citrate (1.0 M) in the crystallization conditions. Inhibition studies with citrate indicate

that citrate can competitively inhibit enzymatic activity with a K_i of 33 mM. The glycerol is presumed to originate from the crystal cryoprotectant solution that contains 20% glycerol and 80% of the mother liquor. A structural alignment of the Sco3058-glycerol-citrate and hRDP-cilastatin indicates that Arg-223 (Arg-230 in the hRDP) is interacting with the C-terminus carboxylate of the substrate. In addition, the inhibitor is also interacting with the β-metal at the C-terminus. The interactions at the amino N-terminus of the substrate are unclear in the hRDP-cilastatin structure; however it is suspected that the C1-OH of glycerol in the Sco3058-glycerol-citrate structure is occupying the amino group at the N-terminus of the substrate. Since the C1-OH of glycerol is in close proximity to Asp-22 and α-metal, it is supposed that Asp-22 and the α-metal are responsible for the main interactions at the N-terminus of the substrate. Based on the structural alignment, the C3-OH of citrate 2.8 Å away from Asp-320 and is occupying the same position as the nitrogen of the amide bond. The *pro*-S arm of citrate mimics the side chain of L-Aspartate at the C-terminus of the dipeptide Xaa-L-Asp substrate. The *pro*-R arm of citrate is representing the side chain of D-Aspartate in the Xaa-D-Asp dipeptide. The *pro*-R carboxylate of citrate is interacting with His-150 and represents the interaction between the enzyme and D-Aspartate. The *pro*-S carboxylate of citrate is within hydrogen bonding distance to Thr-324 and signifies the interaction between the enzyme and L-Aspartate. The obvious interactions at both side chains help explain the promiscuity of the enzyme in hydrolyzing substrates with D- and L- amino acid at the C-terminus. The interactions at the presumed N-terminus and C-terminus of the substrate are depicted in **Figures 4.11A** and **4.11B**, respectively.

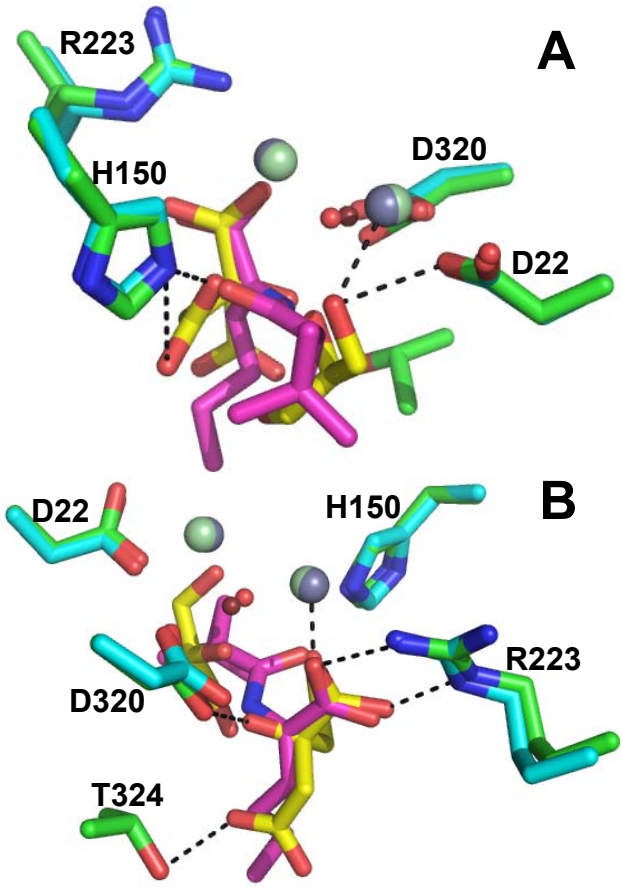


Figure 4.11: Structural alignment of the active site of hRDP (blue) and Sco3058 (green) in the presence of inhibitors. Cilastatin in the hrDP structure is in purple, while citrate and glycerol in the Sco3058 structure are in yellow. (**A**) Interactions at the N-terminus of the dipeptide substrate. (**B**) Interactions at the C-terminus of the dipeptide.

Another crystal structure of Sco3058 was solved to 1.9 Å in the presence of the dipeptide intermediate mimic L-Ala-L-Asp phosphinodipeptide. Structural alignment of the two Sco3058 complexes (glycerol/citrate and phosphinodipeptide) (**Figure 4.12**) supports the C-terminal carboxylate interactions with Arg-223 and the β-metal. The carboxylate side chain of aspartate interacts with Thr-324 and Lys-247 as expected for L-aspartate. The amino group at the N-terminus is in close proximity to the Asp-22 from β-strand 1 and also the α-metal. The oxygens from the phosphorus that is representing the electrophilic carbonyl center are interacting with His-150. One of these oxygens is also coordinating to the β-metal. One interesting thing about the structure is that the bridging hydroxide was never displaced; instead it interacts with one of the oxygens of the phosphorus center. This suggests that the inhitibor is possibly not binding in the correct orientation or that the metals are quickly charged with another solvent molecule soon after the nucleophilic attack on the substrate by the bridging hydroxide.

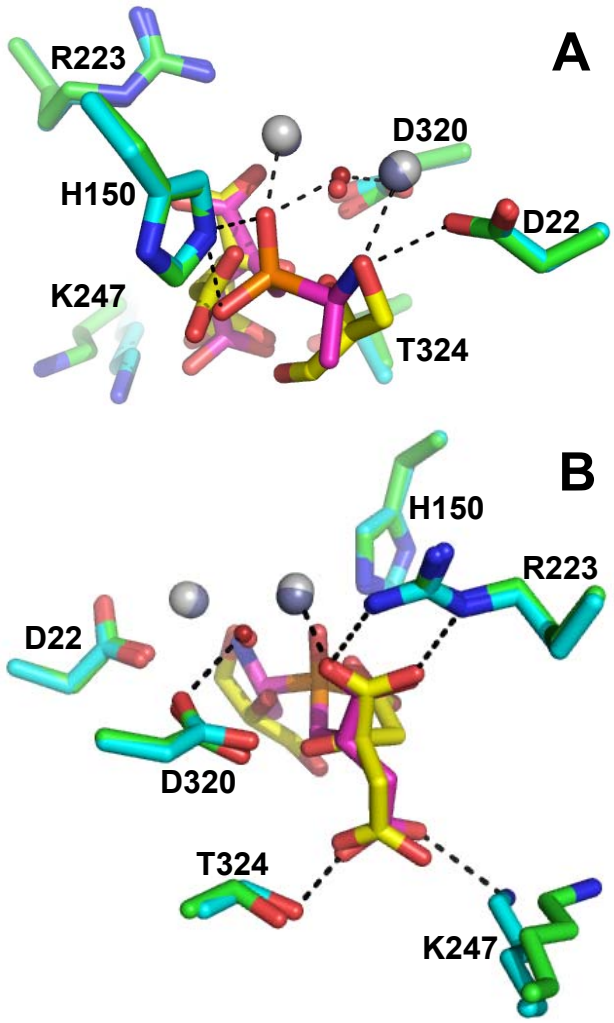


Figure 4.12: Structural alignment of the active sites of Sco3058 with different bound ligands. In green is the structure with citrate/glycerol and in blue is the structure with phosphinodipeptide inhibitor. (**A**) Interactions at the N-terminus of the dipeptide. (**B**) Interactions at the C-terminus of the dipeptide.

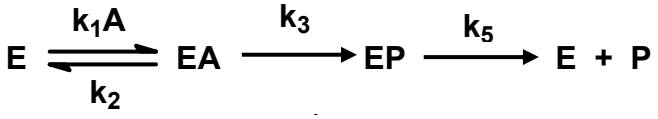
The importance of several conserved residues in the active site was investigated by mutagenesis. Mutation of the evolutionarily conserved aspartate from strand 8 (Asp-

320) resulted in a nonfunctional enzyme. No hydrolytic activity was detected with either the D320N or the D320A mutant. This indicates that Asp-320 may be responsible for abstraction of the proton from the bridging hydroxide to activate the nucleophilic attack. This is confirmed by crystallographic data in which Asp-320 is consistently within 3 Å of the solvent hydroxide. Another mutant that has no catalytic activity is the D22H mutant. The enzyme is capable of binding two metal ions but no hydrolysis could be detected. This suggests that Asp-22 plays an additional role in catalysis aside from ligating to the metal. Crystallographic data indicates that this residue takes part in the recognition of the N-terminal amino group and stabilization of the substrate. The structures of Sco3058 with bound inhibitors support the role of Arg-223 in the interaction with the C-terminal carboxylate. Mutation of this residue to lysine significantly reduced catalytic activity by three orders of magnitude while mutation to methionine abolished activity altogether. His-150 was suggested previously by Nitanai et al, as the key residue in stabilizing the tetrahedral intermediate. Mutation of this residue to asparagine and alanine reduced catalytic activity by two orders of magnitude.

Solvent isotope and solvent viscosity effects were used to address the degree of rate limitation on the bond breaking steps and product release (*123,124,126,127*). For the wild type enzyme, the relatively modest isotope effects of 1.30 for k_{cat} and 1.10 for k_{cat}/K_m indicate that proton transfer is not rate-limiting in the hydrolytic catalysis. For the solvent viscosity studies, in a simplified kinetic mechanism for the conversion of substrate to product for the hydrolytic reaction by Sco3058 presented in **Scheme 4.2**, k_1 and k_2 describe the substrate binding step, k_3 describes the chemical step, and k_5 depicts

the product dissociation step. The expression derived from this mechanism for k_{cat} is given by $(k_3k_5)/(k_3 + k_5)$ and the expression for k_{cat}/K_m is $(k_1k_3)/(k_2 + k_3)$. If one assumes that k_5 is inversely proportional to the relative solvent viscosity then the value of k_5, relative to k_3 can be determined from a plot of the ratio of ${}^0(k_{cat})/{}^\eta(k_{cat})$ and ${}^0(k_{cat}/K_m)/{}^\eta(k_{cat}/K_m)$ as a function of the relative solvent viscosity, η. The slope of these plots are equal to $(k_3 + k_5)/(k_3 + k_5)$ and $(k_2 + k_3)/(k_2 + k_3)$, respectively. When the rate-limiting step is dependent on something other than the chemical step such as substrate association/product dissociation ($k_3 >> k_2$), the slope equals a maximum value of 1. When the rate of the reaction is dependent entirely on the chemical step of the reaction ($k_3 << k_2$), the slope equals zero. For the wild type Sco3058 the slope of 0.5 for the effect of k_{cat}/K_m indicates that the substrate L-Arg-L-Asp is somewhat sticky ($k_3 > k_2$). The slope was found to be 0.05 for the effect on k_{cat}, suggesting that the overall rate for the hydrolysis of dipeptides by Sco3058 is not limited by release of products but possibly by the nucleophilic attack. Another possibility that could contribute to the overall rate of the reaction is conformational changes upon substrate binding and catalysis.

Scheme 4.2: Kinetic mechanism for the hydrolysis reaction catalyzed by Sco3058.



Mechanism of Action. Based on the crystal structures of Sco3058-inhibitor complexes and the kinetic characteristics of the mutants, mechanisms for the hydrolysis of dipeptides by Sco3058 can be proposed (**Figures 4.13** and **4.14**). In the proposed mechanism for Sco3058, initial binding of substrate is facilitated by Asp-22 at the N-terminus and by Arg-223 and M_{β} for the C-terminus for L-Xaa-L-Xaa dipeptides. In the next step, the conserved aspartate from strand 8 (Asp-320) abstracts a proton from the bridging solvent and activates the hydroxide for nucleophilic attack, forming a tetrahedral intermediate that is stabilized by the metal center. The collapse of the tetrahedral intermediate is facilitated by the proton transfer from the asp-320 to the leaving group of the amine product. In the case of L-Xaa-D-Xaa dipeptides, substrate recognition at the C-terminus is possibly due to the enzyme-substrate interaction at residue His-150. This mechanism, however, is not observed in the crystal structures of Sco3058 with bound inhibitors. It is possible that soaking of the crystals with the inhibitor solution does not allow the inhibitor to bind properly in the active site of the enzyme.

An alternative mechanism is proposed based on the crystal structures of Sco3058-inhibitor complexes. In the resting state of the enzyme, the α-metal is coordinated to a water molecule in addition to the briding hydroxide. Asp-320 initiates the reaction and indirectly activates the nucleophile by abstracting the proton from the bridging hydroxide. Simultaneously, the bridging solvent abstracts the proton on the water molecule that is coordinating to the α-metal. The resulting hydroxide acts as the nucleophile and attacks the carbonyl carbon of the amide bond to form a tetrahedral

intermediate, which is stabilized by its interaction with the M_β as well as His-150. A proton transfer from Asp-320 to the amine of the leaving group assists with the collapse of the tetrahedral intermediate. The products dissociate and the active site is ready for another round of catalysis with a charging of the α-metal center with another water molecule. One problem with this mechanism is that in the Sco3058 structure with no ligand bound in the active site (solved by Dr. Steven Almo and colleages), there is no electron density that would indicate the presence of a water molecule coordinating to the α-metal.

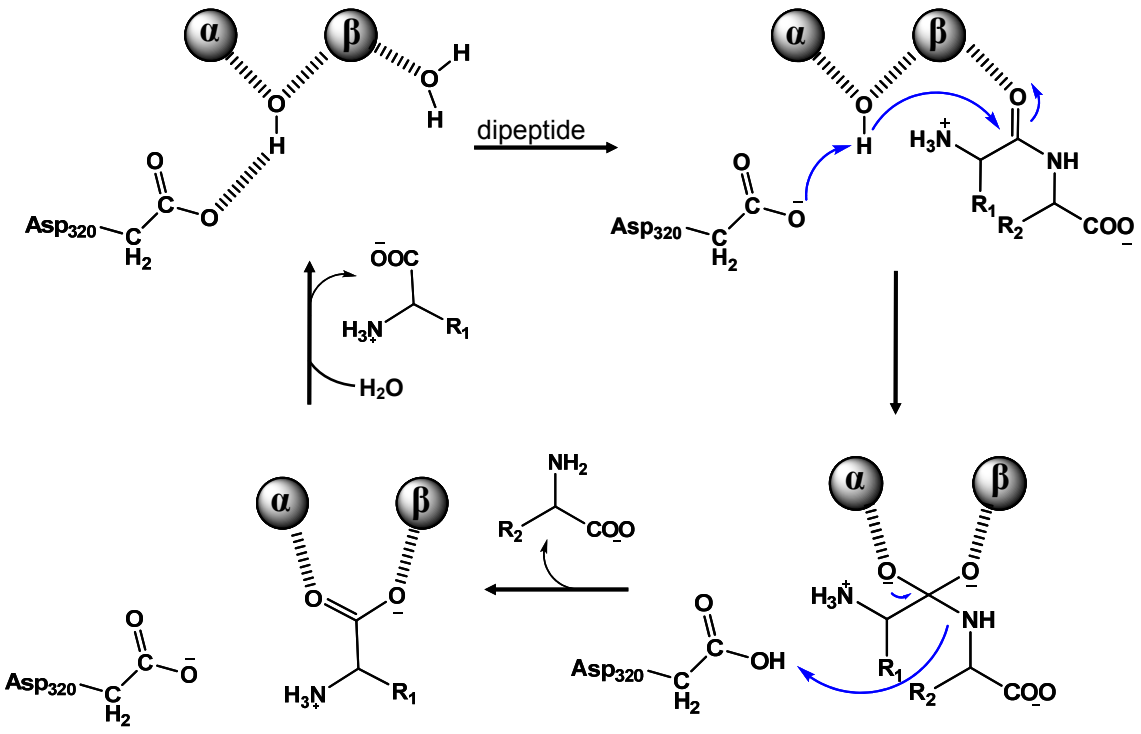


Figure 4.13: First proposed mechanism for the hydrolysis of dipeptides by Sco3058.

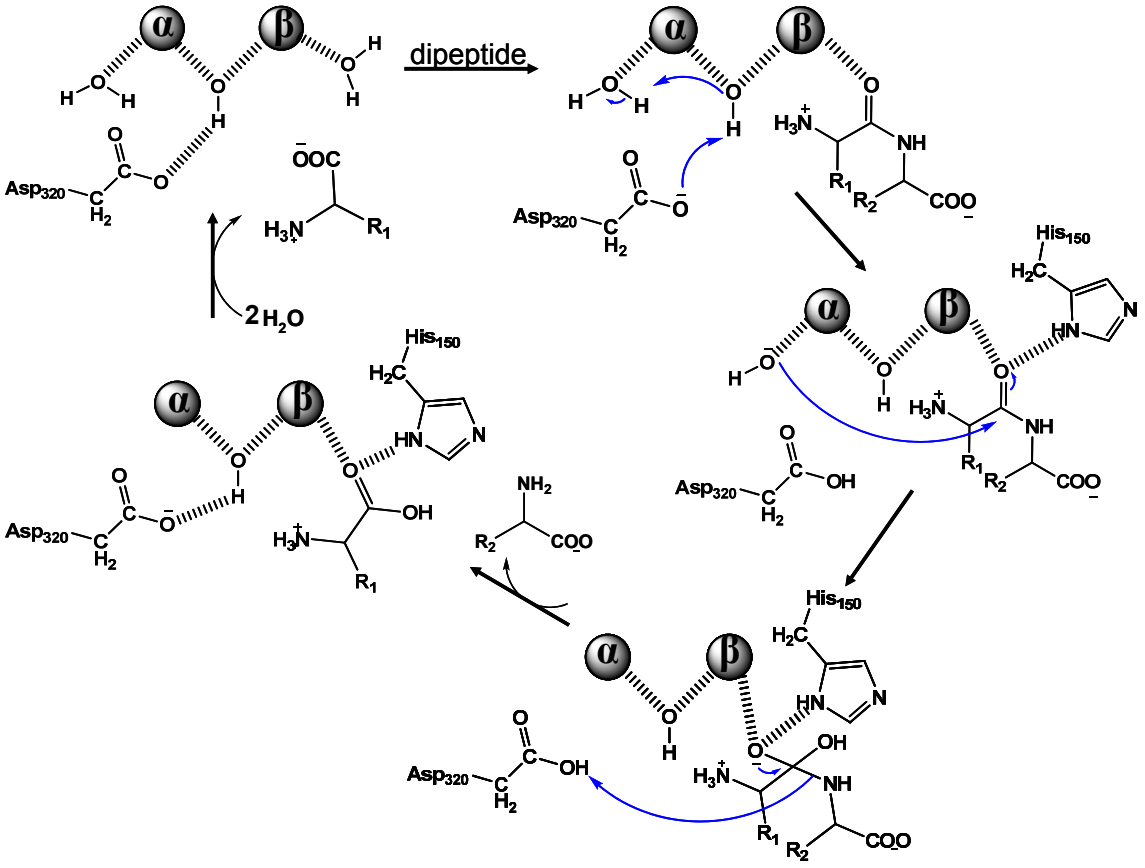


Figure 4.14: Alternative mechanism for the hydrolytic reaction catalyzed by Sco3058.

The second mechanism is further supported by comparing the active sites of NagA and Sco3058. NagA from *B. subtilis* is a binuclear metal center while the *E. coli* NagA is mononuclear at the β-site. The crystal structures of NagA from *E. coli* in the absence of bound ligand indicate a disordered loop after β-strand 3. In the presence of bound ligand, the loop becomes more ordered revealing a semi-conserved His-143 (*70*). In the crystal structure of D273N with bound phosphonamidate inhibitor, His-143 is believed to be interacting and stabilizing the intermediate. Structural alignment of the *E. coli* D273N NagA-inhibitor and Sco3058-inhibitor complexes (**Figure 4.15**) revealed that his-143 in NagA and His-150 in Sco3058 are superimposed onto each other. Structural analysis indicates that both residues originate from a loop after β-strand 3. The overlaying of the two inhibitors in the active sites indicates that the phosphinodipeptide inhibitor binds to the active site in a manner similar to that of the phosphonamidate. In the active site of Sco3058, however, the phosphate group of the phosphinodipeptide is shifted approximately 1.8 Å away from the metal center. This shift could be due to the flexibility of the P-C bond in the phosphinodipeptide, where the nitrogen of the amide is missing. In addition, the structure of NagA-inhibitor reveals that the oxygen of the phosphate group occupies the same space as the bridging solvent in the Sco3058 structure. These observations further support the idea of the latter proposed mechanism.

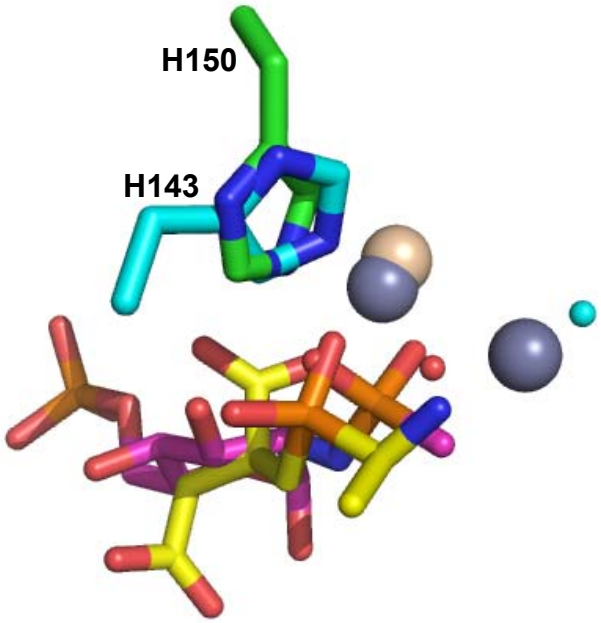


Figure 4.15: Structural alignment of the active sites of NagA and Sco3058 in the presence of inhibitors. His-150 of Sco3058 is presented in green and in blue is His-143 of NagA. The zinc ions in the Sco3058 structure are in grey, and the metal in the *E. coli* NagA (β-site) is in light brown. The blue circle presents a water molecule in the NagA structure. The phosphonamidate inhibitor of NagA is in purple while the phosphinodipeptide inhibitor of Sco3058 is in yellow.

CHAPTER V

SUMMARY AND CONCLUSIONS

The amidohydrolase superfamily contains enzymes that share a common ancestor but are evolved to catalyze a diverse set of hydrolytic and nonhydrolytic reactions of metabolites from various pathways. All enzymes of the AHS exhibits a TIM-barrel structural motif but the divergent evolution process is apparent in the variations of active site metal center. The most common active site variation found among the AHS is the binuclear center in the α and β positions in which the M_α is coordinated to the HxH motif from β-strand 1 and the aspartate from strand 8. The M_β is ligated to the two histidines from strands 5 and 6. The two metals are bridged by a carboxylated lysine from β-strand 4 and a solvent molecule (*21*,*31*). This type of binuclear active site is found with protein families such as PTE, DHO, and IAD (*33-35*). Deviations from this include the replacement of the carboxylated lysine to a glutamate from either strand 4 or 3 as seen with the PTE homology protein and *B. subtilis N*-acetylglucosamine 6-phosphate deacetylase, respectively (*41*,*54*). A more diverged binuclear center is observed for the structure of human renal dipeptidase in which the HxH from strand 1 is substituted with a HxD motif. In addition, the aspartate from strand 8 is further away from the M_α and no longer coordinates to the metal (*40*). For the mononuclear centers, the metal is either in the α- or the β-position. The metal ligands are generally similar to the ligands in the binuclear centers with a few exceptions. For example, M_α mononuclear centers generally have the HxH from strand 1, the histidine from strand 5 and the aspartate from

strand 8 as metal ligands. The histidine from strand 6 is present and may act as the general base as in the case with adenosine deaminse (*56*). The more common M_β metal center contains the two histidines from strand 5 and 6 as metal ligands. Additional residues from strands 2, 3, or 8 may also coordinate to the metal. The HxH motif from β-strand 1 may not be present as seen with the *E. coli* AGD (*42*). Another family of protein that is on the periphery of the AHS in terms of active site deviation is the histidinol phosphate phosphatase family (*53*). This family of protein contains enzymes that have trinuclear and tetranuclear active site metal clusters.

Despite the many variations in the active site metal centers and the diversity of substrates, the general hydrolytic mechanisms proposed for enzymes of the AHS are quite similar. The metal center has dual functionalities: activate the scissile bond of the substrate for bond cleavage and enhance nucleophilicity of the water molecule. The hydrolytic reaction is usually initiated by proton abstraction from the solvent molecule by either the aspartate from strand 8 or the histidine from strand 6. The nucleophilic attack on the substrate by the deprotonated water/hydroxide forms a tetrahedral or trigonal bipyramidal intermediate. The collapse of the intermediate is facilitated by a proton transfer from the enzyme to the product. The product(s) is released from the active site and the enzyme is charged with a solvent molecule for another round of catalysis (*31*).

In this dissertation, the kinetic mechanisms for two diverged protein families of the AHS were examined in detail. The first enzyme, uronate isomerase, catalyzes the isomerization of D-glucuronate to D-fructuronate as well as D-galacturonate to D-

tagaturonate, the first reaction in the metabolism of D-glucuronate and D-galacturonate, respectively. The products of the isomerization reaction then undergo a reduction reaction to D-mannonate and D-altronate by uxuB and uxaB, respectively. In the next step of the pathway, D-mannonate and D-altronate are dehydrated to 2-keto-3-deoxy-gluconate by the enzymes uxuA and uxaA, respectively.

The enzyme uronate isomerase is unique in that it catalyzes a nonhydrolytic reaction, an uncommon type of reaction for the AHS. Another interesting aspect of uronate isomerase is that in some organisms including *E. coli* and *T. maritima*, a single gene uxaC encodes for a uronate isomerase that works on both D-glucuronate and D-galacturonate. Recently, in the genome of *B. halodurans*, a gene encoding an unknown protein was discovered to have uronate isomerase activity. This gene sequence is less than 20% identical to the functionally characterized uronate isomerases. In addition, the gene uxaC encoding the canonical uronate isomerase in the organism *B. halodurans* has been identified (Bh0705) based on its high sequence identity to the *E. coli* URI. Both genes, Bh0493 and Bh0705, were cloned and their corresponding proteins were expressed and purified. The protein encoded by Bh0705 exibited activity with a preference for D-glucuronate while the protein encoded by the gene Bh0493 worked on both D-glucuronate and D-galacturonate. The functional assignment of Bh0493 and Bh0705 as both uronate isomerases were further supported by the gene context within the *B. halodurans* genome. The genes adjacent to Bh0705 are two genes currently annotated as uxuB and uxuA (mannonate dehydrogenase and mannonate dehydratase). Likewise, in the proximity of Bh0493 are two genes annotated as uxaB (Bh0492) and

uxaA (Bh0490). In addition, the gene Bh0494 (PelX) is assigned as an exopolygalacturonate lyase. The functional assignment of Bh0493 further extended the boundaries of the amidohydrolase superfamily.

The crystal structures of Bh0493 were solved in the absence and presence of ligands (D-glucuronate, D-fructuronate, and the intermediate mimics D-arabinaric acid and its hydroxamate derivative). Previously, two other uronate isomerase structures from *T. maritima* and *C. crescentus* have been deposited in the Protein Data Bank. Structural comparision of the three uronate isomerase structures indicated that the URI family operates with a mononuclear zinc center at the M_{α} position. Removal of the metal from the *E. coli* URI rendered the enzyme unactive, indicating that the metal is required for catalytic activity. Structures of Bh0493 in the presence of ligands showed that the metal coordinates to the two oxygens from C5 and C6 of the substrate and therefore anchored the substrate for catalysis. Structural data of Bh0493 in conjunction with kinetic results of the *E. coli* URI were utilized in the elucidation of the mechanism employed by enzymes of the uronate isomerase family. Previous studies by Williams et al revealed that URI utilizes a proton transfer mechanism in which a base (B_1**:**) abstracts a proton from C2 of the substrate to form a *cis*-enediol intermediate. The formation of product is facilitated by abstraction of the C2-OH proton by B2 and subsequent proton transfer from B1 to C1 of the product. NMR data also indicated that the proton at C2 is transferred to the *pro*-R position at C1. pH-rate profiles of the *E. coli* URI suggests two ionizable groups are essential for maximum catalytic activity, one with a pK_a around 6 and one with a pK_a of approximately 9.5. Site-directed mutagenesis and isotope effect

studies designated the group with a pK_a at lower pH to be from the carboxylate of the conserved aspartate from strand 8. The group that ionizes at higher pH is still unclear, and could possibly be assigned to a tyrosine or an arginine. Based on the observations of the ligand-bound Bh0493 structures and kinetic evidences of the *E. coli* URI, two mechanisms were proposed. In the first mechanism Asp-412 (*E. coli* number) acts as **:**B_1 and abstracts the proton from C2 of the sugar to form a *cis*-endiol intermediate. A tyrosine (Y60) or an arginine (Arg-414) behaves as H**:**B_2 and protonates C1 hydroxyl of the intermediate, follow by a proton transfer from **:**B_1 (Asp-412) to C1 to form the product. Alternatively, Asp-412 abstracts the proton from C5-OH of the substrate, which in turn abstracts the proton from C2 of itself to form the *cis*-enediol intermediate, following by product formation in the same manner as the first mechanism. The latter mechanism is unique in that it retains characteristics of the hydrolytic mechanism employed by the AHS enzymes. Structural alignment of Bh0493-glucuronate complex with DHO-dihydroorotate complex illustrated that one of the carboxylate oxygens of D-glucuronate that ligates to the zinc is aligned with one of the oxygens of the carboxylated lysine in the DHO structure. In addition, C5-OH of D-glucuronate is in the same position as the nucleophilic hydroxide and is oriented favorably for proton abstraction by Asp-412.

Renal dipeptidase was first identified in the 1970's based on its physicochemical properties. The enzyme is involved in the metabolism of glutathione and its conjugates, penem and carbapenem β-lactam antibiotics, as well as a wide range of dipeptides. Crystal structure of the hRDP identified this enzyme as a member of the AHS that

utilizes a binuclear metal center active site. The M_α of hRDP is coordinated to the HxD motif from strand 1 and the M_β is ligated to the two histidines from strand 5 and 6. The two metals are bridged by a glutamate from strand 4 and a solvent molecule. The aspartate from strand 8 is no longer coordinating to the M_α but is within hydrogen bonding distance to the bridging hydroxide.

One of the renal dipeptidase homologs from bacteria is Sco3058 from *Streptomyces coelicolor*. Chapter IV of this dissertation addressed the chemical mechanism of Sco3058 using kinetics and structural studies. pH-rate profiles of the zinc and cadmium enzymes both indicate the involvement of two functional groups for maximum catalytic activity, one with a pK_a of approximately 7 and the other around 9. Based on previous findings from other amidohydrolase enzymes, the group with a pK_a of 7 is attributed to the ionization of the nucleophile, the bridging hydroxide. The functional group that ionizes around the pH of 9 is possibly the N-terminal amine of the dipeptide. Site-directed mutagenesis of conserved residues indicates that the aspartate from the HxD motif plays a significant role in catalysis in addition to the metal binding function. The aspartate from strand 8 (Asp-320) is essential for catalytic activity and is proposed to act as a general base in the reaction mechanism. Arg-223 is also necessary for efficient catalysis. His-150 was proposed by Nitanai et al as the residue responsible for the stabilization of the intermediate based on the crystal structure of hRDP with bound cilastatin. Mutation of his-150 significantly compromised catalytic efficiency.

The crystal structure of Sco3058 was solved in the presence of glycerol/citrate and the inhibitor L-ala-L-asp phosphinodipeptide. Based on structural data and kinetic

results, two possible mechanisms were proposed the Sco3058. In the first mechanism, Asp-320 abstracts the proton from the bridging hydroxide, which in turn attacks the carbonyl carbon of the peptide bond, forming a tetrahedral intermediate that is stabilizing by the metal center. The proton abstraction from the hydroxide and the nucleophilic attack are made possible by polarization of the metal center. Proton transfer from Asp-320 to the leaving group amine facilitates the collapse of the intermediate and dissociation of products. In the second proposed mechanism, Asp-320 activates the bridging solvent to abstract a proton from the water molecule that is coordinating to the α-metal. The resulting hydroxide on the α-metal acts as the nucleophile and attacks the carbonyl carbon of the substrate amide, forming a tetrahedral intermediate. His-150 behaves as an oxyanion hole and stabilizes the negative charge on the oxygen of the intermediate. The tetrahedral intermediate collapses, assisting by the transfer of a proton from Asp-320 to the amine group of the leaving group. Products dissociate from the active site and the metal center is charged with another solvent molecule for another round of catalysis. The first mechanism is conventional to the proposed mechanisms for many of the amidohydrolase enzymes. One interesting observation is that when Sco3058 is structurally aligned to the AGD from *E. coli*, His-150 of Sco3058 is aligned to His-143 of AGD. The crystal structure of AGD with bound phosphonamidate inhibitor indicated that his-143 plays a role in stabilizing the transition state. Similarly, in the Sco3058-phosphinodipeptide inhibitor structure complex, His-150 interacts with both oxygens of the intermediate-mimic phosphate. This suggests that His-150 may be responsible for stabilizing the intermediate in the reaction mechanism.

REFERENCES

1. Orengo, C. A., and Thorton, J. M. (2005) Protein families and their evolution – A structural perspective, *Annu. Rev. Biochem. 74*, 867-900.
2. Gerlt, J. A., and Babbitt, P. C. (2000) Can sequence determine function? *GenomeBiology 5*, 1-10.
3. Marti-Arbona, R., Xu, C., Steele, S., Weeks, A., Kuty, G. F., Siebert, 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.
4. Xiang, D. F., Kolb, P., Federov, A. A., Meier, M. M., Federov, 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.
5. Xiang, D. F., Patskovsky, Y., Xu, C., Meyer, A. J., Sauder, J. M., Burley, S. K., Almo, S. C., and Raushel, F. M. (2009) Functional identification of incorrectly annotated prolidases from the amidohydrolase superfamily of enzymes, *Biochemistry 48*, 3730-3742.
6. Xiang, D. F., Xu, C., Kumaran, D., Brown, A. C., Sauder, J. M., Burley, S. K., Swamiathan, S. and Raushel, F. M. (2009) Functional annotation of two new

carboxypeptidases from the amidohydrolase superfamily of enzymes, *Biochemistry 48*, 4567-4576.

7. Hermann, J. C., Marti-Arbona, R., Federov, A. A., Federov, E., Almo, S. C., Shoichet, B. K., and Raushel, F. M. (2007) Structure-based activity prediction for an enzyme of unknown function, *Nature 448*, 775-779.
8. Gerlt, J. A., and Babbitt, P. C. (2001) Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies, *Annu. Rev. Biochem. 70*, 209-246.
9. Andreeva, A., and Murzin, A. G. (2006) Evolution of protein fold in the presence of functional constraints, *Curr. Opin. Struct. Biol. 16*, 399-408.
10. Pieper, U., Chiang, R., Seffernick, J. J., Brown, S. D., Glasner, M. E., Kelly, L., Eswar, N., Sauder, M. J., Bonanno, J. B., Swaminathan, S., Burley, S. K., Zheng, X., Chance, M. R., Almo, S. C., Gerlt, J. A., Raushel, F. M., Jacobson, M. P., Babbitt, P. C., and Sali, A. (2009) Target selection and annotation for the structural genomics of the amidohydrolase and enolase superfamilies, *J. Struct. Funct. Genomics 10*, 107-125.
11. Zhang, C., DeLisi, C. (1998) Estimating the number of protein folds, *J Mol Biol. 284*, 1301–1305.
12. Govindarajan, S., Recabarren, R., Goldstein, R. A. (1999) Estimating the total number of protein folds, *Proteins 35*,408–414

13. Wolf, Y. I., Grishin, N. V., Koonin, E. V. (2000) Estimating the number of protein folds and families from complete genome data, *J. Mol. Biol. 299*, 897-905.

14. Horowitz, N. H. (1945) On the evolution of biochemical syntheses, *Proc. Natl. Acad. Sci. USA 31*, 153-157.

15. Horowitz, N. H. (1965) The evolution of biochemical synthesis: retrospect and prospect in *Evolving genes and proteins*, ed. V Bryson, HJ Vogel, pp.15-23. New York: Academic.

16. Babbitt, P. C. and Gerlt, J. A., (1997) Understanding enzyme superfamilies: Chemistry as the fundamental determinant in the evolution of new catalytic activities, J. Biol. Chem. 272, 30591-30594.

17. Gerlt, J. A., and Babbitt, P. C. (1998) Mechanistically diverse enzyme superfamilies: The importance of chemistry in the evolution of catalysis, *Curr. Opin. Chem. Biol. 2*, 607-612.

18. O'Brien, P. J., and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities, *Chem. Biol. 6*, R91-105.

19. Palmer, D. R., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., Gerlt, J. A. (1999) Unexpected divergence of enzyme function and sequence: "*N*-acylamino acid racemase" is *o*-succinylbenzoate synthase, *Biochemistry 38*, 4252-4258.

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

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

22. Heidhart, D. J., Kenyon, G. L., Gerlt, J. A., Petsko, G. A. (1990) Mandelate racemase and muconate lactonizing enzyme are mechanistically distinct and structurally homologous, Nature 347, 692-694.

23. Babbitt, P. C., Mrachko, G. T., Hasson, M. S., Huisman, G. W., Kolter, R., Ringe, D., Petsko, G. A., Kenyon, G. L., Gerlt, J. A. (1995) A functionally diverse enzyme superfamily that abstracts the alpha protons of carboxylic acids, *Science 267*, 1159-1161.

24. Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., Gerlt, J. A. (1996) The enolase superfamily: A general strategy for enzyme-catalyzed abstraction of the alpha-protons of carboxylic acids, *Biochemistry 35*, 16489-16501.

25. Eklund, H., Eriksson, M., Uhlin, U., Nordlund, P., Logan, D. (1997) Ribonucleotide reductase - structural studies of a radical enzyme, *Biol. Chem. 378*, 821-825.

26. Eklund, H., and Fontecave, M. (1999) Glycyl radical enzymes: A conservative structural basis for radicals, *Struct. Fold Des. 7*, R257-262.

27. Modis, Y., Filppula, S. A., Novikov, D. K., Norledge, B., Hiltunen, J. K., Wierenga, R. K. (1998) The crystal structure of dienoyl-CoA isomerase at 1.5 Å resolution reveals the importance of aspartate and glutamate sidechains for catalysis, *Structure 6*, 957-970.

28. Benning, M. M., Haller, T., Gerlt, J. A., Holden, H. M. (2000) New reactions in the crotonase superfamily: Structure of methylmalonyl CoA decarboxylase from *Escherichia coli*, *Biochemistry 39*, 4630-4639.

29. Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., and Waley, S. G. (1975) Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution: Using amino acid sequence data, *Nature 255*, 609-614.

30. Hocker, B., Lochner, A., Seitz, T., Claren, J., and Sterner, R. (2009) High-resolution crystal structure of an artificial $(\beta/\alpha)_8$-barrel protein designed from identical half-barrels, *Biochemistry 48*, 1145-1147.

31. Siebert, C. M., and Raushel, F. M. (2005) Structural and catalytic diversity within the amidohydrolase superfamily, *Biochemistry 44*, 6383 – 6391.

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

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

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

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

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

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

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

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

40. 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.

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

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

43. Schwarzenbacter, R., Canaves, J. M., Brinen, L. S., Dai, X., Deacon, A. M., Elsliger, M. A., Eshaghi, S., Floyd, R., Godzik, A., Grittini, C., Grzechnik, S. K., Guda, C., Jaroszewski, L., Karlak, C., Klock, H. E., Koesema, E., Kovarik, J. S., Kreusch, A. Kuhn, P., Lesley, S. A., McMullan, D., McPhillips, T. M., Miller, M. A., Miller, M. D., Morse, A., Moy, K., Ouyang, J., Robb, A., Rodrigues, K., Selby, T. L., Spraggon, G., Stevens, R. C., van den Bedem, H., Velasquez, J., Vincent, J., Wang, X., West, B., Wolf, G., Hodgson, K. O., Wooley, J., and

Wilson, I. A. (2003) Crystal structure of uronate isomerase (TM0064) from *Thermotoga maritima* at 2.85 Å resolution, *Proteins 53*, 142-145.

44. Nguyen, T. T., Brown, S. Fedorov, A. A., Fedorov, E. V., Babbitt, P. C., Almo, S. C., and Raushel, F. M. (2007) At the periphery of the amidohydrolase superfamily: Bh0493 from *Bacillus halodurans* catalyzes the isomerization of D-galacturonate to D-tagaturonate, *Biochemistry 47*, 1194-1206.

45. Liaw, S. H., Chen, S. J., Ko, T. P., Hsu, C. S., Chen, C. J., Wang, A. H., and Tsai, Y. C. (2003) Crystal structure of D-aminoacylase from *Alcaligenes faecalis* Da1: A novel subset of amidohydrolases and insights into the enzyme mechanism, *J. Biol. Chem. 278*, 4957-4962.

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

47. 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.

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

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

50. Martynowski, D., Eyobo, Y., Li, T., Yang, K., Liu, A., and Zhang, H. (2006) Crystal structure of α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase: Insight into the active site and catalytic mechanism of a novel decarboxylase reaction, *Biochemistry 45*, 10412-10421.

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

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

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

54. Buchbinder, J. L., Stephenson, R. C., Dresser, M. J., Pitera, J. W., Scanlan, T. S., and Fletterick, R. J. (1998) Biochemical characterization and crystallographic structure of an *Escherichia coli* protein from the phosphotriesterase gene family, *Biochemistry 37*, 5096-5106.

55. Shim, H., and Raushel, F. M. (2000) Self-assembly of the binuclear metal center of phosphotriesterase, *Biochemistry 39*, 7357-7364.

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

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

58. Li, T., Walker, A. L., Iwaki, H., Hasegawa, Y., Liu, A. (2005) Kinetic and spectroscopic characterization of ACMSD from *Pseudomonas fluorescens* reveal a pentacoordinate mononuclear metallocofactor, *J. Am. Chem Soc. 127*, 12282-12290.

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

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

61. Hitchings, G. H., and Falco, E. A. (1944) The identification of guanine in extracts of *Girella nigricans*: The specificity of guanase, *Proc. Natl. Acad. Sci U.S.A. 30*, 294-297.

62. Ribard, C., Rochet, M., Labedan, B., Daignan-Fornier, B., Alzari, P., Scazzocchio, C., and Oestreicher, N. (2003) Sub-families of of α/β barrel enzymes: A new adenine deaminase family, *J. Mol. Biol. 334*,1117–1131.

63. 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*, 2405-2410.66)

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

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

66. 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.

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

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

69. Ohshiro, K., Kakuta, T., Nikaidou, N., Watanabe, T., and Uchiyama, T. (1999) Molecular cloning and nucleotide sequencing of organophosphorus insecticide

hydrolase gene from *Arthrobacter* sp. Strain B-5, *J. Biosci. and Bioeng. 87*, 531-534.

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

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

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

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

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

75. Eaton, R. W. (2001) Plasmid-encoded phthalate catabolic pathway in *Arthrobacter keyseri* 12B, *J. Bacteriol. 183*, 3689-3703.

76. Ishii, Y., Narimatsu, Y., Iwasaki, Y., Arai, N., Kino, K., and Kirimura, K. (2004) Reversible and nonoxidative γ-resorcylic acid decarboxylase: Characterization and gene cloning of a novel enzyme catalyzing carboxylation of resorcinol 1,3-dihydroxybenzene, from *Rhizobium radiobacter*, *Biochem. and Biophys. Res. Comm. 324*, 611-620.

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

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

79. Smiley, J. A., Kundracik, M., Landfried, D. A., Barnes, V. R. Sr., and Axhemi, A. A. (2004) Genes of the thymidine salvage pathway: Thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and *iso*-orotate decarboxylase from *Neurospora crassa*, *Biochimica et Biophysica Acta 1723*, 256-264.

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

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

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

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

84. Wang, J. H., and Tarr, D. A. (1955) On the mechanism of urease action, *J. Am. Chem. Soc. 77*, 6205-6206.

85. Karplus, P. A., and Pearson, M. A. (1997) 70 Years of crystalline urease: What have we learned? *Acc. Chem. Res. 30*, 330-337.

86. Estiu, G., and Merz, K. M. Jr. (2006) Catalyzed decomposition of urea. Molecular dynamics simulations of the binding of urea to urease, *Biochemistry 45*, 4429-4443.

87. Peekhaus, N., and Conway, T. (1998) What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*, *J. Bacteriol. 180*, 3495-3502.

88. Dutton, G. J. (1966) *Glucuronic acid: Free and combined*, New York: Academic Press.

89. Topper, Y. J. (1957) On the mechanism of action of phosphoglucose isomerase and phosphomannose isomerase, *J. Biol. Chem. 225*, 419-426.

90. Rieder, S. V., and Rose, I. A. (1959) The mechanism of the triosphosphate isomerase reaction, *J. Biol. Chem. 234*, 1007-1010.

91. Harris, T. K., Cole, R. N., Comer, F. I., and Mildvan, A. S. (1998) Proton transfer in the mechanism of triosephosphate isomerase, *Biochemistry 37*, 16828-16838.

92. Herlihy, J. M., Maister, S. G., Albery, W. J., and Knowles, J. R. (1976) Energetics of triosephosphate isomerase: the appearance of solvent tritium in substrate dihydroxyacetone phosphate and in product, *Biochemistry 15*, 5607-5612.

93. O'Donoghue, A. C., Amyes, T. L., and Richard, J. P. (2005) Hydrogen transfer catalyzed by triosephosphate isomerase. Products of isomerization of (R)-glyceraldehyde-3-phosphate in D_2O, *Biochemistry 44,* 2610-2621.

94. O'Donoghue, A. C., Amyes, T. L., and Richard, J. P. (2005) Hydrogen transfer catalyzed by triosephosphate isomerase. Products of isomerization of dihydroxyacetone phosphate in D_2O, *Biochemistry 44,* 2622-2631.

95. Fenn, T. D., Ringe, D., and Petsko, G. A. (2004) Xylose isomerase in substrate and inhibitor Michaelis states: Atomic resolution studies of a metal-mediated hydride shift, *Biochemistry 43*, 6464-6474.

96. Whitlow, M., Howard, A. J., Finzel, B. C., Poulos, T. L., Winborne, E., and Gilliland, G. L. (1991) A metal-mediated hydride shift mechanism for xylose isomerase based on the 1.6 Å *Streptomyces rubiginosus* structures with xylitol and D-xylose, *Proteins 9*, 153-173.

97. Allen, K. N., Lavie, A., Farber, G. K., Glasfeld, A., Petsko, G. A., and Ringe, D. (1994) Isotope exchange plus substrate and inhibition kinetics of D-xylose isomerase do not support a proton-transfer mechanism, *Biochemistry 33*, 1481-1487.

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

99. Borodina, I., Siebring, J., Zhang, J., Smith, C. P., van Keulen, G., Dijkhuizen, L., and Nielsen, J. (2008) Antibiotic overproduction in *Streptomyces coelicolor* A3(2) mediated by phosphofructokinase deletion, *J. Biol. Chem. 283*, 25186-25199.

100. Armstrong, D. J., Mukhopadhyay, S. K., and Campbell, B. J. (1974) Physicochemical characterization of renal dipeptidase, *Biochemistry 13*, 1745-1750.

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

102. Campbell, B. J., Lin, Y. C., Davis, R. V., and Ballew, E. (1966) The purification and properties of a particular renal dipeptidase, *Biochim. Biophys. Acta 118*, 371-386.

103. Kahan, F. M., Kropp, H., Sundelof, J. G., and Birnbaum, J. (1983) Thienamycin: Development of imipenem – cilastatin, *J. Antimicrob. Chemother. 12*, 1-35.

104. Kropp, H., Sundelof, J. G., Hajdu, R., and Kahan, F. M. (1982) Metabolism of thienamycin and related carbapenem antibiotics by the renal dipeptidase, dehydropeptidase, *Antimicrob. Agents Chemother. 22*, 62-70.

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

106. Adachi, H., Kubota, I., Okamura, N., Iwata, H., Tsujimoto, M., Nakazato, H. et al. (1989) Purification and characterization of human microsomal dipeptidase, *J. Biochem. (Tokyo) 105*, 957-961.

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

108. Eisen, J. A. (1998) Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis, *Genome Res. 8*, 163-167.

109. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D., and Yeates, T. O. (1999) Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles, *Proc. Natl. Acad. Sci. U.S.A. 96*, 4285-4288.

110. Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999) Detecting protein function and protein-protein interactions from genome sequences, *Science (Washington, DC) 285*, 751-753.

111. Chia, J., and Kolatkar, P. R. (2004) Implications for domain fusion protein-protein interactions based on structural information, *BMC Bioinf. 5*, 161, doi: 10.1186/1471-2105-5-161.

112. Enright, A. J., Iliopoulos, I., Kyrpides, N. C., and Ouzounis, C. A. (1999) Protein interaction maps for complete genomes based on gene fusion events, *Nature (London, U.K.) 402*, 86-90.

113. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D., and Maltsev, N. (1999) The use of gene clusters to infer functional coupling, *Proc. Natl. Acad. Sci. U.S.A. 96*, 2896-2901.

114. Cook, P. I., and Cleland, W. W. (2007) Initial velocity studies in the absence of added inhibitors, in *Enzyme kinetics and mechanisms*, pp 59-120, New York: Garland Science Publishing.

115. Liu, A., and Zhang, H. (2006) Transition metal-catalyzed nonoxidative decarboxylation reactions, *Biochemistry 45*, 10407-10411.

116. 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.

117. Rivolta, C., Soldo, B., Lazarevic, V. Joris, B., Mauel, C., and Karamata, D. (1998) A 35.7 kb DNA fragment from the *Bacillus subtilis* chromosome containing a putative 12.3 kb operon involved in hexuronate catabolism and a

perfectly symmetrical hypothetical catabolite-responsive element, *Microbiology (Reading, U.K.) 144*, 877-884.

118. Shulami, S., Gat, O., Sonenshein, A. L., and Shoham, Y. (1999) The glucuronic acid utilization gene cluster from *Bacillus stearothermophilus* T-6, *J. Bacteriol. 181*, 3695-3704.

119. Benning, M. M., Kuo, J. M., Raushel, F. M., and Holden, H. M. (1994) Three-dimensional structure of phosphotriesterase: An enzyme capable of detoxifying organophosphate nerve agents, *Biochemistry 33*, 15001-15007.

120. Ashwell, G., Wahba, A. J., Hickman, J. (1960) Uronic acid metabolism in bacteria. I. Purification and properties of uronic acid isomerases in *Escherichia coli*, *J. Biol. Chem. 235*, 1559-1565.

121. Wahba, A.J., Hickman, J. Ashwell, G. (1960) Uronic acid metabolism in bacteria. III. Purification and properties of D-altronic acid and D-mannonic acid dehydrogenases in *Escherichia coli*, *J. Biol. Chem. 235*, 1566-1570.

122. Roy, J. E., and Roy, K. L. (1967) The mechanism and specificity of guanine deaminase, *Can. J. Biochem. 45*, 1263-1269.

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

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

125. Hardy, L. W., and Kirsch, J. K. (1984) Diffusion-limited component of reactions catalyzed by *Bacillus cereus* β-lactamases I, *Biochemistry 23*, 1275-1282.

126. Leskovac, V. (2003) *Comprehensive enzyme kinetics*, New York: Kluwer Academic/Plenum Publishers.

127. Cleland, W. W. (1986) Enzyme kinetics as a tool for determination of enzyme mechanism, in *Investigation of rates and mechanisms of reactions*. (Beraasconi, C. F., ed.) 4th ed, vol. 6, part 1, pp 791-868, New York: John Wiley & Sons.

128. Otwinowski, Z., and Minor W. (1997) Processing of X-ray diffraction data collected in oscillation mode, in *Methods in enzymology* (Carter, C. W. J., Sweet, R. M., Abelson, J. N., and Simon, M. I., Eds.) pp 307-326, New York: Academic Press.

129. Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing folm and image plate data, *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* No. 26.

130. Weiss, M. S. (2001) Global indicators of X-ray quality, *J. Appl. Cryst. 34*, 130-135.

131. Collaborative Computational Project, Number 4, (1994) The CCP4 Suite: Programs for protein crystallography, *Acta. Cryst. D50*, 760-763.

132. Navaza, J. (1994) AMoRe: An automated package for molecular replacement, *Acta. Cryst. A50*, 157-163.

133. Read, R. J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood, *Acta. Cryst. D57*, 1373-1382.

134. Brunger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M.,

Simonson, T., and Warren, G. L. (1998) Crystallography & NMR System (CNS), A new software suite for macromolecular structure determination, *Acta Cryst. D54*, 905-921.

135. Brunger, A. T. (2007) Version 1.2 of the crystallography and NMR system, *Nature Protocols 2*, 2728-2733.

136. McRee, D. E. (1999) XtalView/Xfit – A versatile program for manipulating atomic coordinates and electron density, *J. Struct. Biol. 125*, 156-165.

137. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics, *Acta Crystallogr*. *D60*, 2126-2132.

138. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Cryst. D53*, 240-255.

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

APPENDIX

Uronate isomerase is an amidohydrolase enzyme containing a metal ion occupying the α-site of the active site center. Below is the sequence alignment of uronate isomerases from different organisms. Uronate isomerases are distinguished by the metal ligands HxH from β-strand 1, the HY motif from loop 1 after β-strand 1 that may participate in acid-base catalysis, a conserved R from β-strand 3 that anchors the C-6 carboxylate of the substrate, the conserved WWF motif at strand 7, and the DxR motif from strand 8 that are necessary for catalytic activity. The β-strands are highlighted in yellow, the conserved residues are indicated in red, and the semi-conserved residues are in green.

```
gi|94971332|Acid345_4306          ------------------------------------------------
gi|225874247|ACP_2685             ------------------------------------------------
gi|221639919|RSKD131_1820         ------------------------------------------------
gi|126462900|Rsph17029_2139       ------------------------------------------------
gi|146278446|Rsph170252412        ------------------------------------------------
gi|222149863|Avi_3892             ------------------------------------------------
gi|110680135|RD12934              ------------------------------------------------
gi|161620871|BCAN_B0804          ------------------------------------------------
gi|23500543|BRA0812               ------------------------------------------------
gi|163844934|BSUIS_B0804         ------------------------------------------------
gi|17988821|BMEll0476|gi|17988   ------------------------------------------------
gi|189022606|BAbS19_ll04000      ------------------------------------------------
gi|158424966|AZC3342             ------------------------------------------------
gi|13473453|mll4056              ------------------------------------------------
gi|16264678|SMb21354             ------------------------------------------------
gi|16125737|CC1490               ------------------------------------------------
gi|221234494|CCNA_01557          ------------------------------------------------
gi|167645720|Caul_1756           ------------------------------------------------
gi|197105913|PHZ_c2452           ------------------------------------------------
gi|84625827|XOO4170              ------------------------------------------------
gi|161898957|XOO4427             ------------------------------------------------
gi|188574617|PXO_03860           ------------------------------------------------
gi|21244968|XAC4251              ------------------------------------------------
gi|78049913|XCV4357              ------------------------------------------------
gi|66770505|XC4209               ------------------------------------------------
gi|21233539|XCC4117              ------------------------------------------------
gi|188993724|xccb100_4329        ------------------------------------------------
gi|114568782|Mmar100230          ----------------------------------------------MTPVTF 6
gi|148554803|Swit1887            ------------------------------------------------
gi|103488522|Sala3046           ------------------------------------------------
gi|90020587|Sde0940              ------------------------------------------------
gi|90020919|Sde1272              ------------------------------------------------
gi|116249865|RL0099             ------------------------------------------------
gi|209551594|Rleg2_4023          ------------------------------------------------
gi|86355751|RHECH00090           ------------------------------------------------
gi|190889727|RHECIAT_CH0000094   ------------------------------------------------
gi|119962615|AAur0527            ------------------------------------------------
gi|220911562|Achl_0787           ------------------------------------------------
gi|116669068|Arth0502            ------------------------------------------------
gi|163842140|RSal332093421|gi|   ------------------------------------------------
gi|158315003|Franean1_3197       ----------------------------------MAGKFAGAGNSAVS 14
gi|134101579|SACE5055           ----------------------------------MAATSGTRYDHEHA 14
gi|152964556|Krad_0587           ------------------------------------------------
gi|50955310|Lxx17350            ------------MTNPFYFGVIRGTQQQLKAAGYTQLLVDTEESDELED 37
gi|213691567|Blon_0669           ------------------------------------------------
gi|50843764|PPA2329              ------------------------------------------------
gi|161485962|CE2377              ------------------------------------------------
gi|145296483|cgR_2394            ------------------------------------------------
gi|62391323|cg2731               ------------------------------------------------
gi|161486711|NCgI2398            ------------------------------------------------
gi|126208494|APL_1020           ------------------------------------------------
gi|190150346|APP7_1077          ------------------------------------------------
gi|165976445|APJL_1038          ------------------------------------------------
gi|152977831|Asuc_0145          ------------------------------------------------
gi|161511000|MS0544             ------------------------------------------------
gi|148827203|CGSHiGG_02740      ------------------------------------------------
gi|68248599|NTHI0056            ------------------------------------------------
gi|113461738|HS_1602            ------------------------------------------------
gi|170718551|HSM_0409           ------------------------------------------------
gi|62181648|SC3078               ------------------------------------------------
gi|197249740|SeAg_B3309         ------------------------------------------------
gi|198243637|SeD_A3482          ------------------------------------------------
gi|207858396|SEN2980            ------------------------------------------------
gi|194734239|SeSA_A3319         ------------------------------------------------
gi|205354057|SG3031              ------------------------------------------------
gi|161616115|SPAB03917          ------------------------------------------------
gi|16766437|STM3137             ------------------------------------------------
```

```
gi|194448866|SeHA_C3383          ------------------------------------------------
gi|194446286|SNSL254_A3389       ------------------------------------------------
gi|16761915|STY3308              ------------------------------------------------
gi|56415079|SPA3005              ------------------------------------------------
gi|197364008|SSPA2803            ------------------------------------------------
gi|161506293|Sari04489          ------------------------------------------------
gi|37524198|plu0176             ------------------------------------------------
gi|28901561|VPA1706             ------------------------------------------------
gi|37677254|VV1594              ------------------------------------------------
gi|27367460|VV21070             ------------------------------------------------
gi|119943921|Ping0131           ------------------------------------------------
gi|152996793|Mmwyl12780         ------------------------------------------------
gi|117625400|APECO1_3327        ------------------------------------------------
gi|91212521|UTI89_C3530         ----------------------------------------MITLSA 6
gi|16130987|b3092                ------------------------------------------------
gi|161486116|c3850               ------------------------------------------------
gi|215488422|E2348C_3385        ------------------------------------------------
gi|170082634|ECDH10B_3268       ------------------------------------------------
gi|157155304|EcE24377A_3560     ------------------------------------------------
gi|218691395|ECED1_3759         ------------------------------------------------
gi|157162569|EcHS_A3275         ------------------------------------------------
gi|218701863|ECIAI39_3589       ------------------------------------------------
gi|170018657|EcolC_0608         ------------------------------------------------
gi|110643336|ECP_3183           ------------------------------------------------
gi|218560176|ECS88_3488         ------------------------------------------------
gi|170683343|EcSMS35_3384       ------------------------------------------------
gi|89109861|                     ------------------------------------------------
gi|30064435|S3339                ------------------------------------------------
gi|82545347|SBO2953             ------------------------------------------------
gi|56480255|SF3132              ------------------------------------------------
gi|110806974|SFV3133            ------------------------------------------------
gi|74313639|SSON3245            ------------------------------------------------
gi|187731951|SbBS512_E3528      ------------------------------------------------
gi|209397829|ECH74115_4407      ------------------------------------------------
gi|15803633|Z4445                ------------------------------------------------
gi|15833228|ECs3974             ------------------------------------------------
gi|218696796|EC55989_3506       ------------------------------------------------
gi|209920564|ECSE_3373          ------------------------------------------------
gi|218706716|ECUMN_3576         ------------------------------------------------
gi|218551598|EFER_4383          ------------------------------------------------
gi|206579202|KPK_0591           ------------------------------------------------
gi|152972036|KPN_03520          ------------------------------------------------
gi|146313181|Ent638_3546        ------------------------------------------------
gi|157148666|CKO_04494          ----------------------------------------MITLTA 6
gi|156935638|ESA_03503          ------------------------------------------------
gi|161484751|y3600               ------------------------------------------------
gi|161511340|YP2899             ------------------------------------------------
gi|108809201|YPA3210            ------------------------------------------------
gi|162419999|YPAngolaA1104      ------------------------------------------------
gi|145597678|YPDSF0366          ------------------------------------------------
gi|170022805|YPK_0554           ------------------------------------------------
gi|108810613|YPN0448            ------------------------------------------------
gi|16120908|YPO0579             ------------------------------------------------
gi|153949672|YpsIP317580491     ------------------------------------------------
gi|51597770|YPTB3478            ------------------------------------------------
gi|186896961|YPTS_3663          ------------------------------------------------
gi|123443892|YE3708             ------------------------------------------------
gi|157372554|Spro4321           ------------------------------------------------
gi|108810612|YPN0447            ------------------------------------------------
gi|50119594|ECA0645             ------------------------------------------------
gi|54303188|PBPRB1515           ------------------------------------------------
gi|60681862|BF1904G             ------------------------------------------------
gi|53713582|BF2293              ------------------------------------------------
gi|161511193|BT0823             ------------------------------------------------
gi|150005586|BVU3074            ------------------------------------------------
gi|212550592|CFPG_235           ------------------------------------------------
gi|150009769|BDI3184            MDPNHGYRCRGADRGESIRLVQGRRDLRRNGIQTLPQRSDSGKRLAIHYR 50
```

```
gi|146301998|Fjoh_4262          ------------------------------------------------
gi|192359962|CJA_2772           ------------------------------------------------
gi|182416100|Oter_4293          ------------------------------------------------
gi|116624408|acid5330           ------------------------------------------------
gi|225621141|BHWA1_02239        ------------------------------------------------
gi|18309134|CPE0152             ------------------------------------------------
gi|160880875|Cphy_2743          ------------------------------------------------
gi|189485070|TGRD_067           ------------------------------------------------
gi|56962769|ABC0995             ------------------------------------------------
gi|222529391|Athe_1402          ------------------------------------------------
gi|146296954|Csac1949           ------------------------------------------------
gi|150018894|Cbei_4082          ------------------------------------------------
gi|150019467|Cbei_4657          ------------------------------------------------
gi|20808343|TTE1939            ------------------------------------------------
gi|15893980|CAC0692             ------------------------------------------------
gi|150016705|Cbei_1832          ------------------------------------------------
gi|52081820|BL00708            ------------------------------------------------
gi|52787207|BLi03516            ------------------------------------------------
gi|157693739|BPUM2987           ------------------------------------------------
gi|154685671|RBAM012380         ------------------------------------------------
gi|16078295|BSU12300            ------------------------------------------------
gi|23097822|OB0367              ------------------------------------------------
gi|15613268|Bh0705              ------------------------------------------------
gi|138895421|GTNG_1765          ------------------------------------------------
gi|220929098|Ccel_1676          ------------------------------------------------
gi|81427739|LSA0129            ------------------------------------------------
gi|116332824|LVIS_0150          ------------------------------------------------
gi|116332809|LVIS_0135          ------------------------------------------------
gi|116512444|LAACR_1742        ------------------------------------------------
gi|125623707|llmg_0862           -----------------------------------------------
gi|15673610|L0019               ------------------------------------------------
gi|150389106|Amet1291           ------------------------------------------------
gi|70727647|SH2648              ------------------------------------------------
gi|25010729|gbs0674             ------------------------------------------------
gi|22536865|SAG0701             ------------------------------------------------
gi|76786751|SAK0827             ------------------------------------------------
gi|195977823|Sez_0689           ------------------------------------------------
gi|225868878|SZO_13120          ------------------------------------------------
gi|225870118|SEQ_0716           ------------------------------------------------
gi|94990680|MGAS10270_Spy1171   ------------------------------------------------
gi|222153335|SUB1203            ------------------------------------------------
gi|146318813|SSU051159|gi|1463  ------------------------------------------------
gi|146321021|SSU981174          ------------------------------------------------
gi|225859885|SP70585_2238       ------------------------------------------------
gi|157147928|CKO_03735          ------------------------------------------------
gi|220927672|Ccel_0214          ------------------------------------------------
gi|78221672|Gmet0451           ------------------------------------------------
gi|148269995|Tpet0860           ------------------------------------------------
gi|170288680|TRQ2_0883          ------------------------------------------------
gi|15642839|TM0064              ------------------------------------------------
gi|160902320|Pmob0854           ------------------------------------------------
gi|157364144|Tlet1289           ------------------------------------------------
gi|20808989|TTE2645             ------------------------------------------------
gi|159042183|Cmaq1622           ------------------------------------------------
gi|32474066|RB6095              ------------------------------------------------
gi|15613056|Bh0493              ------------------------------------------------

gi|94971332|Acid345_4306        ----------------MLIHPDR---LFPADPGTRTIAKRLYEHVRALPI 31
gi|225874247|ACP_2685           ----------------MLMHEDR---LFPADPTTRRIARSLYEQVRSLPI 31
gi|221639919|RSKD131_1820       -------------MG--LLQEDR---LFPVDEGTRALARSLYATVKDLPI 32
gi|126462900|Rsph17029_2139     -------------MG--LLQEDR---LFPVDEGTRALARSLYATVKDLPI 32
gi|146278446|Rsph170252412      -------------MG--LLHADR---LFPIEEGARALARALHGSVKDLPI 32
gi|222149863|Avi_3892           -----------------MLHPDR---LFPLDPTARTLARTLYDTVSDLPI 30
gi|110680135|RD12934            -------------MG--LTDPER---LFPIEPSARALARGLYASVCDLPI 32
gi|161620871|BCAN_B0804         -------------MA---LNPDR---LFSAEPGTREIARRLFASVEKLPI 31
gi|23500543|BRA0812             -------------MA---LNPDR---LFSAEPGTREIARRLFASVEKLPI 31
```

```
gi|163844934|BSUIS_B0804          ------------MA---LNPDR---LFSAEPGTREIARRLFASVEKLPI 31
gi|17988821|BMEll0476|gi|17988    ------------MA---LNPDR---LFSAEPGTREIARRLFASVEKLPI 31
gi|189022606|BAbS19_ll04000       -------------------------------------------------
gi|158424966|AZC3342              ------------MASPLIHPDR---LFPADPAVRAVARRLYEQVADLPI 34
gi|13473453|mll4056               ------------MVA-LTDPDL---LFPPEAHSRSLARDLYAGIKDLPI 33
gi|16264678|SMb21354              ------------MRG-LIDPDL---LFPAEERTRALARRLYAEVSGLPI 33
gi|16125737|CC1490                ------------MARPLSFHEDR---LFPSDPATRSYARGLYALVKDLPI 35
gi|221234494|CCNA_01557           ------------MARPLSFHEDR---LFPSDPATRSYARGLYALVKDLPI 35
gi|167645720|Caul_1756            ------------MVRPLVFHEDR---LFPADERTRGIARALYGQVKSLPI 35
gi|197105913|PHZ_c2452            ------------MTRPLRLHPDR---LFPADAGTRDIARRLYGSVKDLPI 35
gi|84625827|XOO4170               -----------MRSSVLSLHPDR---LLPADPGTRAIAGRLYAQIATLPI 36
gi|161898957|XOO4427              -----------MRSSVLSLHPDR---LLPADPGTRAIAGRLYAQIATLPI 36
gi|188574617|PXO_03860            -----------------MHPDR---LLPADPGTRAIAGCLYAQIATLPI 29
gi|21244968|XAC4251               -----------MRSSVLSLHPDR---LLPADPGTRAIARRLYAQVATLPI 36
gi|78049913|XCV4357               -----------MRSSVLSLHPDR---LLPADPGTRAIARRLYAQVATLPI 36
gi|66770505|XC4209                ------------MPTPLILHDDR---LLPADPATRAIARRLYAQTAALPI 35
gi|21233539|XCC4117               ------------MPTPLILHDDR---LLPADPATRAIARRLYAQTAALPI 35
gi|188993724|xccb100_4329         ------------MPTPLILHDDR---LLPADPATRAIARRLYAQTAALPI 35
gi|114568782|Mmar100230           ADQSANDRSLQSMPTPSFFHEDR---LFPAEAAVRGIARELHASVRDLPI 53
gi|148554803|Swit1887             -----------MTAAPLRLHPDR---LFPADPDTRVVARRLYAEVAGLPI 36
gi|103488522|Sala3046             ------------MPRPLYLSPDR---LFPSDPAQRDIARRLYKAVAGLPI 35
gi|90020587|Sde0940               ------------MSELILHPDR---LFPAEESVRGIARRLYAEVKDLPI 34
gi|90020919|Sde1272               ------------MTNPLNLHPDR---LFPAAPAARDIARRLYNDIKNLPI 35
gi|116249865|RL0099               -----------MDAGNGFLHPDR---LFPADPATRTVARDLYETVRNLPI 36
gi|209551594|Rleg2_4023           -----------MDVGNGFLHPDR---LFPADPATRTIARDLYETVRNLPI 36
gi|86355751|RHECH00090            -----------MDAGNGFLHPDR---LFPADPATRTIARDLYETVRNLPI 36
gi|190889727|RHECIAT_CH0000094    -----------MDAGNGFLHPDR---LFPADPATRTIARDLYETVRNLPI 36
gi|119962615|AAur0527             --------------MSLATHPDR---LLPADPGTRGIARELLQRVQDLPI 33
gi|220911562|Achl_0787            ------------MSQSIASHPDR---LLPADPGTRSIARSLLERVQDLPI 35
gi|116669068|Arth0502             ------------MVHSMAAHPDR---LLPAEPGVRDIARSLYNLVEGLPI 35
gi|163842140|RSal332093421|gi|    -------------------------------------------------
gi|158315003|Franean1_3197       GGVTVP-----GGGVPLRPHPDR---LFPVDLGVRALARRLYEAVRDLPI 56
gi|134101579|SACE5055             GPLATPRGAAPAGRVPLTPHPDR---LLPSEPSQRCIARRLYDSVRDLPL 61
gi|152964556|Krad_0587            ------------MRPLTLDDDR---LLPTEPGVRALAREVYGHVRDLPI 34
gi|50955310|Lxx17350              GTLQGLRRSFDGAILAASRLTDRRLTALAAEIPVVAVNRQTRG-VANVFI 86
gi|213691567|Blon_0669            -----------MVS---ALNEDR---LFPTDPEQRLIARRLYEAIKDRPI 33
gi|50843764|PPA2329               -----------MSTPTLTLSEDR---LLPRESSALAAAREIYRSTKGLPI 36
gi|161485962|CE2377               ------------MSTSHAAHPDR---LLPADPATRDIARRLLTYVEDLPI 35
gi|145296483|cgR_2394             ------------MTTSHASHPDR---LLPADPGTRDIARRLLAHVEDLPI 35
gi|62391323|cg2731                ------------MTTSHASHPDR---LLPADPGTRDIARRLLAHVEDLPI 35
gi|161486711|NCgI2398             ------------MTTSHASHPDR---LLPADPGTRDIARRLLAHVEDLPI 35
gi|126208494|APL_1020             -------------------------------------------------
gi|190150346|APP7_1077            -------------------------------------------------
gi|165976445|APJL_1038            -------------------------------------------------
gi|152977831|Asuc_0145            ---------------MKQFMDED---FLLSN---DVARTLYFDYAKDQPI 29
gi|161511000|MS0544               ---------------MKQFMDED---FLLSN---DVARTLYYDYAKDQPI 29
gi|148827203|CGSHiGG_02740        ---------------MKSFMDEN---FLLST---DTAKILYHDYAKNKPI 29
gi|68248599|NTHI0056              ---------------MKQFMDEN---FLLST---DTAKILYHDYAKNKPI 29
gi|113461738|HS_1602              -------------------MTED---FLLST---STAQKLYHDYAEEQPI 25
gi|170718551|HSM_0409             ---------------MMKFMTED---FLLST---STAQKLYHDYAEEQPI 29
gi|62181648|SC3078                ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|197249740|SeAg_B3309           ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|198243637|SeD_A3482            ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|207858396|SEN2980              ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|194734239|SeSA_A3319           ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|205354057|SG3031               ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|161616115|SPAB03917            ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|16766437|STM3137               ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|194448866|SeHA_C3383           ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|194446286|SNSL254_A3389        ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|16761915|STY3308               ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|56415079|SPA3005               ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|197364008|SSPA2803             ---------------MATFMTED---FLLKN---DIARTLYHKYAAPMPI 29
gi|161506293|Sari04489            ---------------MTTFMTKD---FLLKN---DIARTLYHKYAAPMPI 29
gi|37524198|plu0176               ---------------MKNFMCDD---FLLNN---ETARQLYHEHAADMPI 29
gi|28901561|VPA1706               ---------------MKNFLCED---FLLSN---ETARRLYHEHAFHQPI 29
gi|37677254|VV1594                ---------------MKNFLCED---FLLSN---ETARRLYHEHAFHQPI 29
```

```
gi|27367460|VV21070              --------------MKNFLCED---FLLSN---ETARRLYHEHACHQPI 29
gi|119943921|Ping0131            ----------MESSMKAFLCED---FLLNN---NVARRLYHDFAKDMPI 33
gi|152996793|Mmwyl12780          --------------MKNFMTED---FLLTT---ETAKRLYHEYAADQPI 29
gi|117625400|APECO1_3327         --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|91212521|UTI89_C3530          SHHKQASSLVERKTKMTPFMTED---FLLDT---EFARRLYHDYAKDQPI 50
gi|16130987|b3092                --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|161486116|c3850               --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|215488422|E2348C_3385         --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|170082634|ECDH10B_3268        --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|157155304|EcE24377A_3560      --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|218691395|ECED1_3759          --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|157162569|EcHS_A3275          --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|218701863|ECIAI39_3589        --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|170018657|EcolC_0608          --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|110643336|ECP_3183            --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|218560176|ECS88_3488          --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|170683343|EcSMS35_3384        --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|89109861|                     --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|30064435|S3339                --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|82545347|SBO2953              --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|56480255|SF3132               --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|110806974|SFV3133             --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|74313639|SSON3245             --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|187731951|SbBS512_E3528       --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|209397829|ECH74115_4407       --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|15803633|Z4445                --------------MTPFMTED---FLLDT---EFARXXYHDYAKDQPI 29
gi|15833228|ECs3974              --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|218696796|EC55989_3506        --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|209920564|ECSE_3373           --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|218706716|ECUMN_3576          --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|218551598|EFER_4383           --------------MTPFMTED---FLLDT---EFARRLYHEYAKDQPI 29
gi|206579202|KPK_0591            --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|152972036|KPN_03520           ------------------MTED---FLLDT---EFARRLYHDYAKDQPI 25
gi|146313181|Ent638_3546         --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|157148666|CKO_04494           SKHKQSDALVERKTIMTPFMTED---FLLDT---EFARRLYHDYAKDQPI 50
gi|156935638|ESA_03503           --------------MTPFMTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|161484751|y3600               --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|161511340|YP2899              --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|108809201|YPA3210             --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|162419999|YPAngolaA1104       --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|145597678|YPDSF0366           --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|170022805|YPK_0554            --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|108810613|YPN0448             --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|16120908|YPO0579              --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|153949672|YpsIP317580491      --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|51597770|YPTB3478             --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|186896961|YPTS_3663           --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|123443892|YE3708              --------------MAQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|157372554|Spro4321            --------------MATFLSED---FLLGN---EFARRLYHDYAADQPI 29
gi|108810612|YPN0447             --------------MSQFLTED---FLLDT---EFARRLYHDYAKDQPI 29
gi|50119594|ECA0645              --------------MPQFLSED---FLLDT---EFARRLYHEYAVDQPI 29
gi|54303188|PBPRB1515            -------------------------------------------------
gi|60681862|BF1904G              --------------MKNFMDKN---FLLQT---ETAQELYHNHAAKMPI 29
gi|53713582|BF2293               --------------MKNFMDKN---FLLQT---ETAQELYHNHAAKMPI 29
gi|161511193|BT0823              --------------MKNFMDEN---FLLQT---ETAQKLYHEHAAKMPI 29
gi|150005586|BVU3074             --------------MKNFMDEN---FLLQT---ETAQKLYHEHAAKMPI 29
gi|212550592|CFPG_235            --------------MKTFMDDN---FILQT---EIARQLYHENAKSLPI 29
gi|150009769|BDI3184             FMPQRLIYHSKVSPSMKNFLDQD---FLLQT---DTARELYHEHAAKMPI 94
gi|146301998|Fjoh_4262           -------------MSANTFIHDN---FLLEN---KYAEELYHNYSKNQPI 31
gi|192359962|CJA_2772            ---------------MSFIHDD---FLLDT---QQAKVLYHEYAKNMPI 28
gi|182416100|Oter_4293           --------------MRPFIHDD---FLLHT---DAARDLYHSFAKAEPI 29
gi|116624408|acid5330            ---------------MSFIHDD---FLLSS---KTARRFYHTFAEDQPI 28
gi|225621141|BHWA1_02239         --------------MKTFMDKD---FLLYN---ETAKTLFYNYACKCPI 29
gi|18309134|CPE0152              --------------MKKFMDKD---FLLEN---EVAKVLYHNYASKVPV 29
gi|160880875|Cphy_2743           --------------MKPFMDKD---FLLST---ETAKTLYHDYAAQVPI 29
gi|189485070|TGRD_067            --------------MKKFLDKD---FLLST---EAAKELFHKYAEKKPI 29
gi|56962769|ABC0995              ----------MISTGKAFIHEN---FMLQN---KTAETLYHTYAKTLPI 33
```

```
gi|222529391|Athe_1402            ---------------MKRFMDED---FLLNN---KTAKVLYEKYAKDMPI 29
gi|146296954|Csac1949             ---------------MKRFMDED---FLLNN---QTAKVLYEKYAKDMPI 29
gi|150018894|Cbei_4082            ---------------MKNFMDEN---FLLSN---QTAIDLYHNYAKNLPI 29
gi|150019467|Cbei_4657            ------------MMNLKKFMDEN---FLLSN---KTAMELYHNYAKNMPI 32
gi|20808343|TTE1939               ---------------MRKFMDED---FLLTN---ETAVKLYHQYAKDMSI 29
gi|15893980|CAC0692               ---------------MKNFMDEK---FMLST---KVAEDLYNDFAKDMPI 29
gi|150016705|Cbei_1832            ---------------MKKFMDEN---FLLSN---EVSEKLYHNYSEKMPI 29
gi|52081820|BL00708               ---------------MKPFLNDD---FLLTN---ETSKVLYHQYAKGMPI 29
gi|52787207|BLi03516              ---------------MKPFLNDD---FLLTN---ETSKVLYHQYAKGMPI 29
gi|157693739|BPUM2987             ---------------MKAFLNEQ---FLLNS---PTAEKLYHEFAKDLPI 29
gi|154685671|RBAM012380           ---------------MKAFMGDD---FLLNS---KTAVKLYREYAENMPI 29
gi|16078295|BSU12300              ---------------MEPFMGKN---FLLKN---ETAVSLYHNYAKDMPI 29
gi|23097822|OB0367                ---------------MKKFMDDN---FLLSN---DTAEELFHHYAKDMPI 29
gi|15613268|Bh0705                ---------------MTNFLSED---FLLMN---EYDRELYYTFAKNMPI 29
gi|138895421|GTNG_1765            ---------------MQPFIDDR---FLLQN---KHAEVLYHDYAKSLPI 29
gi|220929098|Ccel_1676            -------------MTVKTFINNN---FMLKS---KTAASLYHFNAASVPI 31
gi|81427739|LSA0129               ----------------MQLLDDD---FLLDN---DMAKTLYHDYAAQMPI 28
gi|116332824|LVIS_0150            ----------------MDLLDND---FLLGN---DMAKKLFHDYAQNMPI 28
gi|116332809|LVIS_0135            ----------------MTLLNED---FLLTN---EPAKRLYHEHAEKMPI 28
gi|116512444|LAACR_1742           ----------------MKFLSED---FLLNN---ENAKLLFHKHAEKMPI 28
gi|125623707|llmg_0862            ----------------MKFLSED---FLLNN---ESAKLLFHEHAEKMPI 28
gi|15673610|L0019                 ----------------MKFLSED---FLLTN---ESAKMLFHKHAEKMPI 28
gi|150389106|Amet1291             ----------MKGMIKISFINND---FMLEN---EIARHLYHTYAKSLPI 34
gi|70727647|SH2648                ----------------MAFINEH---FMLNN---ETGKHLYHDFAKDMPI 28
gi|25010729|gbs0674               ----------------MTFNTET---FMLKN---QAAIQLYEEVKR-QPI 27
gi|22536865|SAG0701               ----------------MAFNTET---FMLKN---QAAIQLYEEVKR-QPI 27
gi|76786751|SAK0827               ----------------MTFNTET---FMLKN---QAALQLYEEVKR-QPI 27
gi|195977823|Sez_0689             ----------------MAFNDDN---FMLKN---EAAKRLYQQIKD-QPI 27
gi|225868878|SZO_13120            ----------------MAFNDDN---FMLKN---EAAKRLYQQIKD-QPI 27
gi|225870118|SEQ_0716             ----------------MAFNDDN---FMLKN---EAAKRLYQQIKD-QPI 27
gi|94990680|MGAS10270_Spy1171     -------MEVITEEMNMTFNDAN---FMLKN---EAAKKLYQQIQD-QPI 36
gi|222153335|SUB1203              ----------------MTFNDKN---FMLKN---QAAKQLYTAVQD-QPI 27
gi|146318813|SSU051159|gi|1463    ----MLAVLLQLWRDEMSFNDKN---FMLKN---EPAKELYGKIAE-LPI 39
gi|146321021|SSU981174            ----MLAVLLQLWRDEMSFNDKN---FMLKN---EPAKELYGKIAE-LPI 39
gi|225859885|SP70585_2238         ----------------MSFNDKN---FMLKN---EPAKELYVKIAE-LPI 27
gi|157147928|CKO_03735            ----------------MSMINKS---FMISN---EPGQRLYSELAKTLPI 28
gi|220927672|Ccel_0214            -----------------MLSNN---FILSG---KTAVSLYERYAKDAPV 26
gi|78221672|Gmet0451              ----------------MTIRRN---FTASVR--ENALTLYERHAS-LPI 27
gi|148269995|Tpet0860             -----------------MFLGED---YLLTN----RAAVRLFNEVKDLPI 26
gi|170288680|TRQ2_0883            -----------------MFLGED---YLLTN----RAAVRLFNEVKDLPI 26
gi|15642839|TM0064                -----------------MFLGED---YLLTN----RAAVRLFNEVKDLPI 26
gi|160902320|Pmob0854             ----------------MSFLDEN---YLLQN----NTSKMLYNSIKDFPI 27
gi|157364144|Tlet1289             ----------------MAFLDER---YLLSS----KTAFDLYESVKNLPI 27
gi|20808989|TTE2645               --------------------------------------------------
gi|159042183|Cmaq1622            ------------------------------------MVSVQDVVEESPV 13
gi|32474066|RB6095                --------------MNAATMSDS---SSVAN----GPKSEIYEALTSIRL 29
gi|15613056|Bh0493                -----------------MSINSR----------EVLAEKVKNAVNNQPV 22


gi|94971332|Acid345_4306          VSPHGHTQAA-WFSRNESFPDPASLFVRP----DHYVFRMLYSQGVPLED 76
gi|225874247|ACP_2685             VSPHGHTQAA-WFAHNEPFPDPAKLFVQP----DHYIYRMLYSQGVTLED 76
gi|221639919|RSKD131_1820         ISPHGHTDPR-WFAENEPFPDPAQLFVTP----DHYVFRMLHSQGIPLEA 77
gi|126462900|Rsph17029_2139       ISPHGHTDPR-WFAENEPFPDPAQLFVTP----DHYVFRMLHSQGIPLEA 77
gi|146278446|Rsph170252412        VSPHGHTDPR-WFAENEAFPDPAQLFVTP----DHYVFRMLHSQGIPLEA 77
gi|222149863|Avi_3892             VSPHGHTDPR-WFAENEAFPDPAQLFVVP----DHYVFRMLYSQGIDLTA 75
gi|110680135|RD12934              ISPHGHCDPR-WFAENQRFENPAALFVIP----DHYVFRMLISQGVAPED 77
gi|161620871|BCAN_B0804           ISPHGHTEPI-WYARNEAFPDPASLFVKP----DHYITRMLYSQGHSLES 76
gi|23500543|BRA0812               ISPHGHTEPI-WYARNEAFPDPASLFVKP----DHYITRMLYSQGHSLES 76
gi|163844934|BSUIS_B0804          ISPHGHTEPI-WYARNEAFPDPASLFVKP----DHYITRMLYSQGHSLES 76
gi|17988821|BMEll0476|gi|17988    ISPHGHTEPI-WYARNEAFPDPASLFVKP----DHYITRMLYSQGHSLES 76
gi|189022606|BAbS19_ll04000       -------------------------------------------------- 
gi|158424966|AZC3342              VSPHGHTDPR-WYAENKNFPDPARLFVVP----DHYIFRMLYSQGIPLEQ 79
gi|13473453|mll4056               VSPHGHTDPR-WYALNEPFPDPAQLLIVP----DHYIFRMLFSQGVRLED 78
gi|16264678|SMb21354              VSPHGHTEPR-WYALDEAFPDPAQLLIVP----DHYVFRMLFSQGIRLEE 78
gi|16125737|CC1490                ISPHGHTDPS-WFATNAPFQDATDLLLAP----DHYLFRMLYSQGVSLDA 80
gi|221234494|CCNA_01557           ISPHGHTDPS-WFATNAPFQDATDLLLAP----DHYLFRMLYSQGVSLDA 80
gi|167645720|Caul_1756            ISPHGHTDPA-WFATNTPFEDATDLLLAP----DHYLFRMLYSQGISLDA 80
```

```
gi|197105913|PHZ_c2452              ISPHGHTDPE-WFAGDAPFPDATNLFLAP----DHYLYRMLYSQGVPLDR 80
gi|84625827|XOO4170                 ISPHGHTDPA-WFATNAPFADATELLLVP----DHYVFRMLYSQGIDLDA 81
gi|161898957|XOO4427                ISPHGHTDPA-WFATNAPFADATELLLVP----DHYVFRMLYSQGIDLDA 81
gi|188574617|PXO_03860              ISPHGHTDPA-WFATNAPFADATELLLVP----DHYVFRMLYSQGIDLDA 74
gi|21244968|XAC4251                 ISPHGHTDPA-WFATNAPFANATELLLVP----DHYVFRMLYSQGIDLDA 81
gi|78049913|XCV4357                 ISPHGHTDPA-WFATNAPFANATELLLVP----DHYVFRMLYSQGIDLDA 81
gi|66770505|XC4209                  ISPHGHTDPA-WFATDAPFANATELLLVP----DHYVFRMLYSQGIDLDQ 80
gi|21233539|XCC4117                 ISPHGHTDPA-WFATDAPFANATELLLVP----DHYVFRMLYSQGIDLDQ 80
gi|188993724|xccb100_4329           ISPHGHTDPA-WFATDAPFANATELLLVP----DHYVFRMLYSQGIDLDQ 80
gi|114568782|Mmar100230             ISPHGHTDPR-WFASNEAFENPTALLLAP----DHYVFRMLYSQGIRLAD 98
gi|148554803|Swit1887               VSPHGHTDPR-WFAEDAPFGDASSLLLQP----DHYVFRMLYSQGVPLEA 81
gi|103488522|Sala3046               VSPHGHTDPA-WFAGDAPFGNAAELLLHP----DHYVFRMLYSQGVSLDA 80
gi|90020587|Sde0940                 VSPHGHTDPS-WFADNQPFDSAADLLIKP----DHYVFRMLYSLGVPLES 79
gi|90020919|Sde1272                 VSPHGHTDPQ-WFAANQNFSNPAELLIKP----DHYVFRMLYSQGIRLES 80
gi|116249865|RL0099                 VSPHGHTEPS-WFADDKPFEDAASLLVIP----DHYLFRMLHSVGVTLDE 81
gi|209551594|Rleg2_4023             VSPHGHTEPS-WFADDKPFEDAASLLVIP----DHYLFRMLHSVGVALDE 81
gi|86355751|RHECH00090              VSPHGHTEPS-WFADDKPFEDAASLLVIP----DHYLFRMLHSVGVGLDE 81
gi|190889727|RHECIAT_CH0000094      VSPHGHTEPS-WFADDKPFEDAASLLVIP----DHYLFRMLHSVGVPLDA 81
gi|119962615|AAur0527               ISPHGHVDAA-VIEQNTPFPDPAALLVTP----DHYVTRLIHASGISMDK 78
gi|220911562|Achl_0787              ISPHGHVDAA-VIEHNTAFPDPAALLVSP----DHYVTRLIHASGVDQDL 80
gi|116669068|Arth0502               ISPHGHVDAA-VIEQNLPFPDPAALLVTP----DHYVTRLIHAGGVPMDQ 80
gi|163842140|RSal332093421|gi|      --------------------------------------------------
gi|158315003|Franean1_3197          LSPHGHVDPF-LLVHDEAFADPASLLITP----DHYVTRLLHASGVPLAE 101
gi|134101579|SACE5055               ISPHGHVEAR-LLADDDAFTDPASLLVTP----DHYVTRLLHAHGVPLAD 106
gi|152964556|Krad_0587              ISMHGHVEAQ-VLAEDEPFDDPARMLVVP----DHYVTRMLVSQGTRPED 79
gi|50955310|Lxx17350                DTPNGVEQAVGHLADDEPFPDPAALLITP----DHYVTRTLHSLGVPLRA 132
gi|213691567|Blon_0669              ISPHGHVPID-WFAEDKHFKNPTDLFITP----DHYVTRIMHGHSVPFSE 78
gi|50843764|PPA2329                 ISPHGHVPVS-WIADDMAFSDPTSLLITP----DHYVNRLLHANGVDLED 81
gi|161485962|CE2377                 ISPHGHLEAS-MFVKDEPFADPTTLLISP----DHYLTRVLHSAGVDLAD 80
gi|145296483|cgR_2394               ISPHGHLEAS-MFVKDEAFPDPTSLLISP----DHYLTRMMHSAGVDLAD 80
gi|62391323|cg2731                  ISPHGHLEAS-MFVKDEAFPDPTSLLISP----DHYLTRMMHSAGVDLAD 80
gi|161486711|NCgI2398               ISPHGHLEAS-MFVKDEAFPDPTSLLISP----DHYLTRMMHSAGVDLAD 80
gi|126208494|APL_1020               --------------------------------------------------
gi|190150346|APP7_1077              --------------------------------------------------
gi|165976445|APJL_1038              --------------------------------------------------
gi|152977831|Asuc_0145              FDYHCHLPPK-EIAENRLFKDLTEIWLA----GDHYKWRAMRSAGFDENV 74
gi|161511000|MS0544                 FDYHCHLPPK-EIAENRQFKDLTEIWLA----GDHYKWRAMRSAGVDENL 74
gi|148827203|CGSHiGG_02740          FDYHCHLNPR-EVAENRQFNDLAEIWLE----GDHYKWRALRTAGVPEEL 74
gi|68248599|NTHI0056                FDYHCHLNPR-EVAENRQFNDLAEIWLE----GDHYKWRALRTAGVPEEL 74
gi|113461738|HS_1602                FDYHCHLNPK-EIAENRQFNDLAEIWLE----GDHYKWRAMRSAGVEEHL 70
gi|170718551|HSM_0409               FDYHCHLNPK-EIAENRQFNDLAEIWLE----GDHYKWRAMRSAGVEEHL 74
gi|62181648|SC3078                  YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|197249740|SeAg_B3309             YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|198243637|SeD_A3482              YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|207858396|SEN2980                YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|194734239|SeSA_A3319             YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|205354057|SG3031                 YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|161616115|SPAB03917              YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|16766437|STM3137                 YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|194448866|SeHA_C3383             YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|194446286|SNSL254_A3389          YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|16761915|STY3308                 YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|56415079|SPA3005                 YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|197364008|SSPA2803               YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|161506293|Sari04489              YDFHCHLSPQ-EIADDRRFDNLGQIWLE----GDHYKWRALRSAGVDESL 74
gi|37524198|plu0176                 YDYHCHLNPR-EIAENRRFDNLGQIWLE----GDHYKWRAMRCAGIDESL 74
gi|28901561|VPA1706                 YDYHCHLNPA-EVTQNRQFDNLGQIWLE----GDHYKWRGMRSAGIEERL 74
gi|37677254|VV1594                  YDYHCHLNPA-EVAQNRQFDNLGQIWLE----GDHYKWRGMRSAGIEERL 74
gi|27367460|VV21070                 YDYHCHLNPA-EVAQNRQFDNLGQIWLE----GDHYKWRGMRSAGIEERL 74
gi|119943921|Ping0131               YDYHCHLNAQ-EVADNRQFDNLGQMWLE----GDHYKWRGMRTAGISESL 78
gi|152996793|Mmwyl12780             YDYHCHLSPQ-EIAENRRFTDLGEIWLE----GDHYKWRAMRTAGIEERL 74
gi|117625400|APECO1_3327            FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|91212521|UTI89_C3530             FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 95
gi|16130987|b3092                   FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|161486116|c3850                  FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|215488422|E2348C_3385            FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|170082634|ECDH10B_3268           FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
```

```
gi|157155304|EcE24377A_3560       FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|218691395|ECED1_3759           FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|157162569|EcHS_A3275           FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|218701863|ECIAI39_3589         FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|170018657|EcolC_0608           FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|110643336|ECP_3183             FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|218560176|ECS88_3488           FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|170683343|EcSMS35_3384         FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|89109861|                      FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|30064435|S3339                 FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|82545347|SBO2953               FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|56480255|SF3132                FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|110806974|SFV3133              FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|74313639|SSON3245              FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|187731951|SbBS512_E3528        FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|209397829|ECH74115_4407        FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|15803633|Z4445                 FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|15833228|ECs3974               FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|218696796|EC55989_3506         FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|209920564|ECSE_3373            FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|218706716|ECUMN_3576           FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|218551598|EFER_4383            FDYHCHLPPQ-QIAEDYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|206579202|KPK_0591             FDYHCHLPPQ-QVAENYRFKNLYDIWLK----GDHYKWRAMRTNGVPERL 74
gi|152972036|KPN_03520            FDYHCHLPPQ-QVAENYRFKNLYDIWLK----GDHYKWRAMRTNGVPERL 70
gi|146313181|Ent638_3546          FDYHCHLPPQ-QVAENYRFKNLYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|157148666|CKO_04494            FDYHCHLPPQ-QIAENYRFNNLYDIWLK----GDHYKWRAMRTNGVAERL 95
gi|156935638|ESA_03503            FDYHCHLPPQ-QIAENYRFKNLYDIWLK----GDHYKWRAMRTNGVPERL 74
gi|161484751|y3600                FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|161511340|YP2899               FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|108809201|YPA3210              FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|162419999|YPAngolaA1104        FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|145597678|YPDSF0366            FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|170022805|YPK_0554             FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|108810613|YPN0448              FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|16120908|YPO0579               FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|153949672|YpsIP317580491       FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|51597770|YPTB3478              FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|186896961|YPTS_3663            FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|123443892|YE3708               YDYHCHLPPE-QIAENTRFKNLYDIWLK----GDHYKWRAMRTNGVAERF 74
gi|157372554|Spro4321             FDYHCHLPPE-QIAENTRFKNLYEIWLK----GDHYKWRAMRTNGVAERL 74
gi|108810612|YPN0447              FDYHCHLPPE-QIAENYRFKNMYDIWLK----GDHYKWRAMRTNGVAERL 74
gi|50119594|ECA0645               FDYHCHLPPE-QIAENYRFKNLYDIWLK----GDHYKWRAMRTNGVPERL 74
gi|54303188|PBPRB1515             -------------------------------------MRTNGVNERF 10
gi|60681862|BF1904G               IDYHCHLNPQ-MVADDYRFKSLTEIWLG----GDHYKWRAMRSNGVDECF 74
gi|53713582|BF2293                IDYHCHLNPQ-MVADDYRFKSLTEIWLG----GDHYKWRAMRSNGVDECF 74
gi|161511193|BT0823               IDYHCHLIPQ-MVADDYKFKSLTEIWLG----GDHYKWRAMRTNGVDERY 74
gi|150005586|BVU3074              IDYHCHLIPQ-MVADDYQFKSLTEIWLG----GDHYKWRAMRTNGVDERY 74
gi|212550592|CFPG_235             IDYHCHLSPE-MIADDRQFDNLGQIWLE----GDHYKWRAMRTNGIDERY 74
gi|150009769|BDI3184              IDYHCHLDPR-QIAENHQFEDLTEIWLG----GDHYKWRAMRANGIPEEY 139
gi|146301998|Fjoh_4262            IDYHNHLNPQ-FIAEDKIFDNITNVWIN----GDHYKWRAMRTLGIDEQF 76
gi|192359962|CJA_2772             IDYHCHLPPE-QVGQNKQFRNLYEVWLA----GDHYKWRAMRSNGVDERY 73
gi|182416100|Oter_4293            FDYHCHLPQQ-QILENHQFADLAEIWLG----GDHYKWRAMRANGVKERF 74
gi|116624408|acid5330             LDYHCHLPPQ-DVAANRQFKDLFEIWLE----GDHYKWRAMRANGVPESH 73
gi|225621141|BHWA1_02239          FDYHCHLNPK-EIAENKKFKNITEIWLY----GDHYKWRMMRANGIDEKF 74
gi|18309134|CPE0152               FDYHCHLVPM-EIATDHKFKNLTEMWLY----HDHYKWRAMRSFGIDEEY 74
gi|160880875|Cphy_2743            IDYHCHINPE-EIAKDRSFDTITQVWLG----GDHYKWRQMRSNGIDEKY 74
gi|189485070|TGRD_067             LDYHCHINPE-EIAKDRQFENISQLWLG----SDHYKWRQMRSCGVEEKY 74
gi|56962769|ABC0995               IDYHCHVPPQ-EIAENRQFNNISEIWLH----GDHYKWRAMRAVGVEETF 78
gi|222529391|Athe_1402            VDFHCHLNPK-EIYENKTFKNITEVWLG----GDHYKWRLMRTNGIEEKY 74
gi|146296954|Csac1949             IDFHCHLNPK-EIYENKKFRNITEVWLG----GDHYKWRLMRANGIEEKY 74
gi|150018894|Cbei_4082            IDYHCHIDPK-EIYENKKFSNITEAWLY----GDHYKWRAMRSNGIDEKC 74
gi|150019467|Cbei_4657            IDYHCHIDPK-EIYENKKFNNITEAWLY----GDHYKWRAMRCNGIDEKY 77
gi|20808343|TTE1939               YDFHCHLSPK-EIYEDRRFKNITEVWLY----GDHYKWRLMRANGIDEKY 74
gi|15893980|CAC0692               IDYHCHISPQ-EICENKSFKNITEVWLY----GDHYKWRLMRSSGVDEKY 74
gi|150016705|Cbei_1832            IDYHCHINPK-EILEDKKFENITQVWLY----GDHYKWRQMRTLGIDEKY 74
gi|52081820|BL00708               IDYHCHLSPK-EIYENKTFKNLTEVWLY----GDHYKWRAMRANGISEEF 74
gi|52787207|BLi03516              IDYHCHLSPK-EIYENKTFKNLTEVWLY----GDHYKWRAMRANGISEEF 74
```

```
gi|157693739|BPUM2987            IDYHCHLSPK-DIYENKTFRNITEAWLY----GDHYKWRAMRANGIPETH 74
gi|154685671|RBAM012380          IDYHCHLSPK-EIYENKTFATITEAWLY----GDHYKWRIMRANGIEERC 74
gi|16078295|BSU12300             IDYHCHLSPK-EIYENKTFQNITEAWLY----GDHYKWRIMRANGIEETY 74
gi|23097822|OB0367               IDYHCHLSPK-EIYMNKRYSNITEVWLY----GDHYKWRAMRAAGVEESL 74
gi|15613268|Bh0705               CDYHCHLSPQ-EIWENKPFENMTKAWLG----GDHYKWRAMRLNGVREEF 74
gi|138895421|GTNG_1765           IDYHCHLSAK-EIAEDRRFHDMTELWLE----GDHYKWRAMRALGVEEKY 74
gi|220929098|Ccel_1676           FDYHCHLSPK-EILENRTFSNMTEIWLA----GDHYKWRLMRANGVDERC 76
gi|81427739|LSA0129              IDFHCHLNPS-EIYQNKNYTNITRIWLNEGTYGDHYKWRLMRANGVDEKY 77
gi|116332824|LVIS_0150           IDFHCHLNPE-EIYENKNYENITKIWLNEGTYGDHYKWRLMRANGVPEKL 77
gi|116332809|LVIS_0135           IDYHCHLEPK-DIYENKNYPNITRIWLNDGSLGDHYKWRLERANGVPEDL 77
gi|116512444|LAACR_1742          IDYHCHLEQE-EIYENKKYENLTQIWLG----GDHYKWRLLRANGIPEKL 73
gi|125623707|llmg_0862           IDYHCHLEPA-EIYENKKYENLTQIWLG----GDHYKWRLLRANGIPEKL 73
gi|15673610|L0019                IDYHCHLEPA-EIYENKKYENLTQIWLG----GDHYKWRLLRANGIPEKL 73
gi|150389106|Amet1291            FDYHCHLNTQ-HIAEDHEFADITELWLA----GDHYKWRAMRGNGVSEEK 79
gi|70727647|SH2648               YDYHCHLDPK-QISDNVACDNITDLWLS----GDHYKWRAMRAQGIEEQY 73
gi|25010729|gbs0674              FDYHCHLDPK-DIFEDRIFDNIVDLWLG----GDHYKWRLMRANGISEAE 72
gi|22536865|SAG0701              FDYHCHLDPK-DIFEDHIFDNIVDLWLG----GDHYKWRLMRANGISEAE 72
gi|76786751|SAK0827              FDYHCHLDPK-DIFEDRIFDNIVDLWLG----GDHYKWRLMRANGISEAE 72
gi|195977823|Sez_0689            FDYHCHLDPK-EIFEDKVYDNIVDLWLG----GDHYKWRLMRANGISEEE 72
gi|225868878|SZO_13120           FDYHCHLDPK-EIFEDKVYDNIVDLWLG----GDHYKWRLMRANGISEEE 72
gi|225870118|SEQ_0716            FDYHCHLDPK-EIFEDKVYDNIVDLWLG----GDHYKWRLMRANGISEEE 72
gi|94990680|MGAS10270_Spy1171    FDYHCHLDPK-EIFEDKVYDNIVDLWLG----GDHYKWRLMRANGISEEE 81
gi|222153335|SUB1203             FDYHCHLDPK-EIFEDKVFENIVDLWLG----GDHYKWRLMRANGISEEE 72
gi|146318813|SSU051159|gi|1463   YDFHCHLDPK-EIFEDKVYEDIVDLWLG----GDHYKWRLMRANGISEEE 84
gi|146321021|SSU981174           YDFHCHLDPK-EIFEDKVYEDIVDLWLG----GDHYKWRLMRANGISEEE 84
gi|225859885|SP70585_2238        YDFHCHLDPK-EIFEDKVYEDIVDLWLG----GDHYKWRLMRANGISEEE 72
gi|157147928|CKO_03735           IDYHCHLEAK-AIWENKPFADITQLWLE----GDHYKWRAMRANGIPERK 73
gi|220927672|Ccel_0214           YDYHCHLCAK-EIYEDEAFNDISSIWLG----YDHYKWRTMRFAGVPEEY 71
gi|78221672|Gmet0451             IDYHGHLPPA-RLASNAHFRDLSELWIE----GDHYKWRAMRIAGEREDL 72
gi|148269995|Tpet0860            VDPHNHLDAK-DIVENKPWSDIWEVEGA----TDHYVWELMRRCGISEEY 71
gi|170288680|TRQ2_0883           VDPHNHLDAK-DIVENKPWSDIWEVEGA----TDHYVWELMRRCGISEEY 71
gi|15642839|TM0064               VDPHNHLDAK-DIVENKPWNDIWEVEGA----TDHYVWELMRRCGVSEEY 71
gi|160902320|Pmob0854            LDAHNHGDVK-EIVENKGWDDIWQVEGA----TDHYVWESMRKRGVPEEK 72
gi|157364144|Tlet1289            VDAHNHGDVK-EIVENKGWNDIWEVEAA----TDHYVWELMRRRGVPEDK 72
gi|20808989|TTE2645              ------------------MKLYIAAWLA----EEGYDVFLAYLRGIETSF 28
gi|159042183|Cmaq1622            ADVHNHLNPR-----SLSPSGFQDVLLY------HYIVTELRSAGAPLGF 52
gi|32474066|RB6095               IDPHSHINPH-----SPASTTLADVLGY------HYYTELAHSAGMPKSQ 68
gi|15613056|Bh0493               TDMHTHL-F--SPNFGEILLWDIDELLTY-----HYLVAEVMRWTDVSIE 65

gi|94971332|Acid345_4306         LEI------GVHETANPRRVWRIFAENYHLFRGTPTRRWLDYAFSELFGM 120
gi|225874247|ACP_2685            LEI------GVEQIQNPRKVWRIFASHYYLFRGTPTRLWLDFAFETLFGL 120
gi|221639919|RSKD131_1820        MGVPRAD--GGPTETDGRKIWRLFASNFHLFRGTPSRLWLDHAFSEVFGV 125
gi|126462900|Rsph17029_2139      MGVPRVD--GGPTETDGRKIWRLFASNFHLFRGTPSRLWLDHAFSEVFGV 125
gi|146278446|Rsph170252412       MGVPRAD--GGPTETDGRTIWRLFAANYHLFRGTPSRLWLDHAFSEVFGL 125
gi|222149863|Avi_3892            LGVPRVD--GGMTETDGRKIWRLFAENFHLFRATPSRMWLDHAFEDVFGL 123
gi|110680135|RD12934             LGIPRVD--GGEVETDPRKIWRLFAQHFHLFRGTPSAMWLNHSFEHVFDL 125
gi|161620871|BCAN_B0804          LGIASRD--GRPSETDARKIWRLFATNWYLFRATPSRLWFEHAMETVFGI 124
gi|23500543|BRA0812              LGIASRD--GRPSETDARKIWRLFATNWYLFRATPSRLWFEHAMETVFGI 124
gi|163844934|BSUIS_B0804         LGIASRD--GRPSETDARKIWRLFATNWYLFRATPSRLWFEHAMETVFGI 124
gi|17988821|BMEll0476|gi|17988   LGIASRD--GRPSETDARKIWRLFATNWYLFRATPSRLWFEHAMETVFGI 124
gi|189022606|BAbS19_ll04000      --------------------------------------------------
gi|158424966|AZC3342             LGVPRRD--GGETEQDGRKIWRLFASHYHLFRGTPTRSWLDHAFATLFDI 127
gi|13473453|mll4056              LGVASLD--GAPVETDGRTIWRRFAEHYYLFRGTPTRLWFDHVLAHLFGI 126
gi|16264678|SMb21354             LGVPALD--GSPVETDGRAIWRRFCENYHLFRGTPTRLWFDYTLSELFGI 126
gi|16125737|CC1490               LKVRSKA--GV-PDTDPREAWRVFASHFYLFRGTPSWVWLNHVFSQVFGF 127
gi|221234494|CCNA_01557          LKVRSKA--GV-PDTDPREAWRVFASHFYLFRGTPSWVWLNHVFSQVFGF 127
gi|167645720|Caul_1756           LKVSSKS--GV-PGTDPREAWRLLAQNFHLFRGTPSWIWLNHVFSQVFDF 127
gi|197105913|PHZ_c2452           LGVPSKA--GP-SPADPREAWRMLASNMHLFRGTPSSMWLNHVFGEVFGF 127
gi|84625827|XOO4170              IGIPRAD--GMRAAVDPRAAWRVFAAHYTVLRGTPSALWLNHVFHDVFDL 129
gi|161898957|XOO4427             IGIPRAD--GMRAAVDPRAAWRVFAAHYTVLRGTPSALWLNHVFHDVFDL 129
gi|188574617|PXO_03860           IGIPRAD--GMRAAVDPRAAWRVFAAHYTVLRGTPSALWLNHVFHDVFDL 122
gi|21244968|XAC4251              LGIPRAD--GTRATVDPRAAWRVFAEHYTLLRGTPSALWLNHVFHDVFDL 129
gi|78049913|XCV4357              LGIPRAD--GTRAAVDPRAAWRVFAEHYTLLRGTPSALWLNHVFHDVFDL 129
gi|66770505|XC4209               LGIPHAD--GSRAPVDPREAWRVFASQFALLRGTPSALWLNHVFHQVFDL 128
gi|21233539|XCC4117              LGIPHAD--GSRAPVDPREAWRVFASQFALLRGTPSALWLNHVFHQVFDL 128
gi|188993724|xccb100_4329        LGIPHAD--GSRAPVDPREAWRVFASQFALLRGTPSALWLNHVFHQVFDL 128
```

```
gi|114568782|Mmar100230           LAIPDRN--GV-PATDPRAAWRLFAENFHLFRGTPSSIWLSHVLVKVFGV 145
gi|148554803|Swit1887             IGIG--------APVDPRAAWRVLADHYHLFRGTPSRLWLDWVFHSVFGL 123
gi|103488522|Sala3046             LGIGN-------ADADPRESWRLFAENYHLFRATPSRMWMDWVFAEVFGF 123
gi|90020587|Sde0940               LGVPTVD--GTEVETDGRKIWQTFAKHYTAFRGTPSRMWLDHVFKDVFGF 127
gi|90020919|Sde1272               IGLKNKE--GAPSETDARKIWQIFADHYYLFRGTPSQMWMDYVLHEVFGV 128
gi|116249865|RL0099               LGVPRLD--GK-PVASGRAIWRTFAAHYHLFRGTPSSLWVDHAMSAVLGC 128
gi|209551594|Rleg2_4023           LGVPRLD--GK-PVAEGRAIWRTLARHYHLFRGTPSSLWVDHAMSAVLGC 128
gi|86355751|RHECH00090            LGVPRLD--GN-PVAEGREIWRTFAKHYHLFRGTPSSLWVDHAMSAVLGC 128
gi|190889727|RHECIAT_CH0000094    LGVPRLD--GK-PVAAGRAIWQTFARHYHLFRGTPSSLWVDHAMSAVLGC 128
gi|119962615|AAur0527             LQPG-------NATP--REIWRHFCAAWPLFDGTASGYWLRTQFESVFGW 119
gi|220911562|Achl_0787            LRPQTGAKGDPNAVPDARTVWREFCKAWPLFEGTASGYWLRTQFQSVFGL 130
gi|116669068|Arth0502             LGLG-------ETAADSRQVWRNFCAAWPDFEGTASGYWIRQEFEHVFGL 123
gi|163842140|RSal332093421|gi|    --------------------------------------------------
gi|158315003|Franean1_3197        LGVGRGP--LPEE--QARAVWRTFCEHWAVFRGTPVRYWLESQLAEIFGV 147
gi|134101579|SACE5055             LGLRDSA--DDTDPAQPRDVWRTFCEHWHLFLGTASRGWLEAELHDVLGV 154
gi|152964556|Krad_0587            LGVPRTD--GGPVEQDGREIWRRFCAGWHLFRGTPSRYWLEHELHEVFGV 127
gi|50955310|Lxx17350              LGLAKNG---VPSPASGRAIWRTLCENWEAFLGTPVRFWFESEFSEVFGL 179
gi|213691567|Blon_0669            LGVGQKN----FTEEQSRNAFRLLGKYWYAYAGTPMRYWMEDSLSNVFGI 124
gi|50843764|PPA2329               LGVGRKT----MSEEDNRRAFRILCEHWRDFAGTAMRYWLVDQLIGIFGI 127
gi|161485962|CE2377               LRVGGTE------GKTPREAWRIFASNWDLYTGTATGYWVEQEFEHVFGI 124
gi|145296483|cgR_2394             LRVGGHE------GKSAREAWRIFMSHWDLYAGTATGYWVEQEFEHVFGI 124
gi|62391323|cg2731                LRVGGHE------GKSAREAWRIFMSHWDLYAGTATGYWVEQEFEHVFGI 124
gi|161486711|NCgI2398             LRVGGHE------GKSAREAWRIFMSHWDLYAGTATGYWVEQEFEHVFGI 124
gi|126208494|APL_1020             --------------------------------------------------
gi|190150346|APP7_1077            --------------------------------------------------
gi|165976445|APJL_1038            --------------------------------------------------
gi|152977831|Asuc_0145            ITG----------NASAYEKYQAWAKTVPLCIGNPIYHWTHLELRRPFGI 114
gi|161511000|MS0544               ITG----------NASNYEKYQAWAKTVPLCIGNPIYHWTHLELRRPFGI 114
gi|148827203|CGSHiGG_02740        ITG----------KATNYQKYLAWAKTVPLCIGNPIYHWTHLELRRPFGI 114
gi|68248599|NTHI0056              ITG----------KATNYQKYLAWAKTVPLCIGNPIYHWTHLELRRPFGI 114
gi|113461738|HS_1602              ITG----------SADKYSKYLAFANTVPKCIGNPIYHWTHLELRRPFGI 110
gi|170718551|HSM_0409             ITG----------SADKYSKYLAFANTVPKCIGNPIYHWTHLELRRPFGI 114
gi|62181648|SC3078                ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|197249740|SeAg_B3309           ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|198243637|SeD_A3482            ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|207858396|SEN2980              ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|194734239|SeSA_A3319           ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|205354057|SG3031               ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|161616115|SPAB03917            ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|16766437|STM3137               ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|194448866|SeHA_C3383           ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|194446286|SNSL254_A3389        ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|16761915|STY3308               ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|56415079|SPA3005               ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|197364008|SSPA2803             ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|161506293|Sari04489            ITGK---------ETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGI 115
gi|37524198|plu0176               ITGK---------QTRDYEKYQAWAETVPLTLGNPLYHWTHLELRRPFGI 115
gi|28901561|VPA1706               ITG----------DASDYDKYMAWAKTVPQTLGNPLYHWTHLELRRPFGI 114
gi|37677254|VV1594                ITG----------DASDYDKYMAWAKTVPQTLGNPLYHWTHLELRRPFGI 114
gi|27367460|VV21070               ITG----------DASDYDKYMAWAKTVPQTLGNPLYHWTHLELRRPFGI 114
gi|119943921|Ping0131             ITGK---------DTSDFEKYMAWAKTVPQTLGNPLYHWTHLELRRPFGI 119
gi|152996793|Mmwyl12780           VTG----------NASFKEKYQAWAKTVPQCIGNPIYHWTHLELRRPFGI 114
gi|117625400|APECO1_3327          CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|91212521|UTI89_C3530           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 135
gi|16130987|b3092                 CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|161486116|c3850                CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|215488422|E2348C_3385          CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|170082634|ECDH10B_3268         CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|157155304|EcE24377A_3560       CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|218691395|ECED1_3759           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|157162569|EcHS_A3275           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|218701863|ECIAI39_3589         CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|170018657|EcolC_0608           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|110643336|ECP_3183             CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|218560176|ECS88_3488           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|170683343|EcSMS35_3384         CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|89109861|                      CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
```

```
gi|30064435|S3339                CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|82545347|SBO2953              CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|56480255|SF3132               CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|110806974|SFV3133             CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|74313639|SSON3245             CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|187731951|SbBS512_E3528       CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|209397829|ECH74115_4407       CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|15803633|Z4445                CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|15833228|ECs3974              CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|218696796|EC55989_3506        CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|209920564|ECSE_3373           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|218706716|ECUMN_3576          CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|218551598|EFER_4383           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|206579202|KPK_0591            CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|152972036|KPN_03520           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 110
gi|146313181|Ent638_3546         CTG----------DATDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|157148666|CKO_04494           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 135
gi|156935638|ESA_03503           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|161484751|y3600               CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|161511340|YP2899              CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|108809201|YPA3210             CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|162419999|YPAngolaA1104       CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|145597678|YPDSF0366           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|170022805|YPK_0554            CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|108810613|YPN0448             CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|16120908|YPO0579              CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|153949672|YpsIP317580491      CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|51597770|YPTB3478             CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|186896961|YPTS_3663           CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|123443892|YE3708              CTG----------DASDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|157372554|Spro4321            CTG----------DAGDREKFDAWAATVPHTIGNPLYHWTHLELRRPFGI 114
gi|108810612|YPN0447             CTG----------DASDREKFDAWAATVPHTIGNPLYP------------ 102
gi|50119594|ECA0645              CTG----------DASDWEKFEAWAATVPHTIGNPLYHWTHLELRRPFGV 114
gi|54303188|PBPRB1515            CTG----------NASDKDKFTEWAKTVSQTIGNPLFHWTHLELRRPFGI 50
gi|60681862|BF1904G              CTGK---------ETSDWEKFEKWAETVPYTFRNPLYHWTHLELKTAFGI 115
gi|53713582|BF2293               CTGK---------ETSDWEKFEKWAETVPYTFRNPLYHWTHLELKTAFGI 115
gi|161511193|BT0823              CTGK---------DTTDWEKFEKWAETVPYTFRNPLYHWTHLELKTAFGI 115
gi|150005586|BVU3074             CTGK---------DTTDWEKFEKWAETVPYTFRNPLYHWTHLELKTAFGI 115
gi|212550592|CFPG_235            ITGE---------STSDFEKFEKWAETVPYTMRNPLYHWTHLELKTVFGI 115
gi|150009769|BDI3184             ITGD---------KPS-YEKFLKWAETMPYTMRNPLYHWTHLELSRIFGI 179
gi|146301998|Fjoh_4262           VTG----------NGSDKDKFLNWAKTVPYTMRNPLYHWTHLELARYFDI 116
gi|192359962|CJA_2772            CTG----------NASDWEKFEKWCETVPYTLRNPLYHWTHLELRKPFGI 113
gi|182416100|Oter_4293           CTG----------AATPREKFDAWVGAVPHTLRNPLYHWSHLELARYFGI 114
gi|116624408|acid5330            CTG----------NAPAYEKFMAWAKTVPATLRNPLYHWTHLELKRYFGI 113
gi|225621141|BHWA1_02239         ITG----------DASDYDKFIAWVKTVPNLIGNPLYHWSHLELQRYFDI 114
gi|18309134|CPE0152              ITG----------NASDYDKFYQFAKMMPYLIGNPIYHWSHLELKRFFGV 114
gi|160880875|Cphy_2743           ITG----------DASDIEKFEKWAETLQKAIGNPLYHWSHLELQRYFDY 114
gi|189485070|TGRD_067            ITG----------NASDSEKFQKWAETLEKAIGNPLYHWSHLELRRYFGY 114
gi|56962769|ABC0995              ITG----------DGDDKEKFLKWAETVPYTMGNPLYHWTHLELKRYFGI 118
gi|222529391|Athe_1402           ITG----------DADDYEKFLAWAKTIPMAIGNPIYHWTHLELKRYFGI 114
gi|146296954|Csac1949            ITG----------SADDYEKFLAWAKTIPMAIGNPIYHWTHLELKRYFGI 114
gi|150018894|Cbei_4082           ITG----------DGSDYDKFLAWSQTIPMAIGNPLYHWTHLELQRFFGI 114
gi|150019467|Cbei_4657           ITG----------DGSDYDKFLAWAKTIPMSIGNPLYHWTHLELQKFFGI 117
gi|20808343|TTE1939              ITG----------DADDYEKFVAYAKTIPMAIGNPVYHWTHLELQRYFGI 114
gi|15893980|CAC0692              ITG----------DSSDYEKFLAYVKAIETAIGNPLYHWSHLELQRYFGV 114
gi|150016705|Cbei_1832           ITG----------DGSDYEKFLAWAKTISFAIGNPLYHWTHLELKRFFGI 114
gi|52081820|BL00708              ITG----------DASDEEKFSAWARTVPMTIGNPLYHWTHLELRRFFGI 114
gi|52787207|BLi03516             ITG----------DASDEEKFSAWARTVPMTIGNPLYHWTHLELRRFFGI 114
gi|157693739|BPUM2987            VTG----------DASDYDKFLAWAKTVPMTIGNPLYHWTHLELRRYFEV 114
gi|154685671|RBAM01230           ITG----------NASDEEKFFAWAKTVPMAIGNPLYSWTHLELQRWFGI 114
gi|16078295|BSU12300             ITG----------DAPDEEKFMAWAKTVPMAIGNPLYNWTHLELQRFFGI 114
gi|23097822|OB0367               ITG----------DANDYDKFMAWAETVPKLIGNPLYNWTHLELQRYFGV 114
gi|15613268|Bh0705               ITG----------GAPDKEKFLAWAKTVPKTIGNPLYHWTHMELKTYFHF 114
gi|138895421|GTNG_1765           ITG----------SASPEEKFQAWAKTVPYCIGNPLYHWTHLELKRYFQV 114
gi|220929098|Ccel_1676           ITG----------DADPYEKFLKWAETVPKCIGNPLYHWTHLELKHYFDI 116
gi|81427739|LSA0129              ITG----------DGDDYLKFIEWAKTIENAYGNPLYEWTHLELRRFFHI 117
gi|116332824|LVIS_0150           ITG----------DGDDYDKFMAWAETIEKSMGNPLYEWTHLELKRFFHI 117
```

```
gi|116332809|LVIS_0135              ITG----------DGDEYDKFLAWAGTIEKAIGNPLYEWTHLELKRFFGI 117
gi|116512444|LAACR_1742             ITG----------DGDDYEKFLAFAKTLEKSLGNPIYEWTHLELKRFFHI 113
gi|125623707|llmg_0862              ITG----------DGDDYEKFLAFAKTLEKSLGNPIYEWTHLELKRFFHI 113
gi|15673610|L0019                   ITG----------DGEDYEKFLAFAKTLEKSFGNPIYEWTHLELKRFFHI 113
gi|150389106|Amet1291               ITG----------KASSEEKFQAWAETVESCVGNPLYHWTHLELKAYFGI 119
gi|70727647|SH2648                  ITG----------DAAPLDKFKKWTETLENSVGNPLYHWSQLELKMYFDI 113
gi|25010729|gbs0674                 ITG----------PASNLEKFKAFARTLERAYGNPVYHWSAMELKNVFGV 112
gi|22536865|SAG0701                 ITG----------PASNLEKFKAFARTLERAYGNPVYHWSAMELKNVFGV 112
gi|76786751|SAK0827                 ITG----------PASNLEKFKAFARTLERAYGNPVYHWSAMELKNVFGV 112
gi|195977823|Sez_0689               ITG----------SASKLDKFKAFARTLQRSYGNPVYHWSVMELKNVFGV 112
gi|225868878|SZO_13120              ITG----------SASKLDKFKAFARTLQRSYGNPVYHWSVMELKNVFGV 112
gi|225870118|SEQ_0716               ITG----------SASKLDKFKAFARTLQRSYGNPVYHWSVMELKNVFGV 112
gi|94990680|MGAS10270_Spy1171       ITG----------SASKLDKFKAFARTLERSYGNPVYHWSAMELKNVFGV 121
gi|222153335|SUB1203                ITG----------SASQLDKFKAFARTLQRSYGNPVYHWSAMELKNVFGI 112
gi|146318813|SSU051159|gi|1463      ITG----------SASKLDKFKAWARTLERAFGNPLYHWSHLELRQVFGV 124
gi|146321021|SSU981174              ITG----------SASKLDKFKAWARTLERAFGNPLYHWSHLELRQVFGV 124
gi|225859885|SP70585_2238           ITG----------FASKLDKFKAWARTLERAFGNPLYHWSHLELRQVFGI 112
gi|157147928|CKO_03735              ITG----------DASSEEKFEAWAQTVEASFGNPLYHWTHLELKYYFAI 113
gi|220927672|Ccel_0214              ITG----------NGDGRTKFRVWAKTCERLIGSPLYHWANMELKTYFGV 111
gi|78221672|Gmet0451                CTG----------SASPKERFDAWARTITRMASSPMQAWCQIELKRTFGI 112
gi|148269995|Tpet0860               ITG----------SRSNKEKWLALAKVFPRFVGNPTYEWIHLDLWRRFNI 111
gi|170288680|TRQ2_0883              ITG----------SRSNKEKWLALAKVFPRFVGNPTYEWIHLDLWRRFNI 111
gi|15642839|TM0064                  ITG----------SRSNKEKWLALAKVFPRFVGNPTYEWIHLDLWRRFNI 111
gi|160902320|Pmob0854               ITG----------NASNKEKWMALAKVFPGFVGNPTYEWIHLDLKRRFKI 112
gi|157364144|Tlet1289               ITG----------DATNYEKWIALAEVFPKFAGNPTYEWVHLDLKRRFGI 112
gi|20808989|TTE2645                 QIG----------FGRSLKRFLAYLKGIETNDSNPD---LPLHVS-VFSL 64
gi|159042183|Cmaq1622               G------------EVKGFNELKPIIPYFRRIRNTATHWGLMRILNDLYGL 90
gi|32474066|RB6095                  IEDK---------AIGPKELVERLVGGLQPLENTAQYSWLIEICRTFFGF 109
gi|15613056|Bh0493                  AF----W------AMS-KREQADLIWEELFIKRSPVSEACRGVLTCLQGL 103


gi|94971332|Acid345_4306            TQR----LSAETSDFYFDSISEKLCTPEFRPRALYEQFGIEVLATTDSPL 166
gi|225874247|ACP_2685               TES----LSAKTSDLYFDTISEKLQTPEFRPRALYERFHLEVLATTDSPL 166
gi|221639919|RSKD131_1820           NER----LSAETADAAYDQIADCLARPEFRPRALFERFGIEAIATTESAL 171
gi|126462900|Rsph17029_2139         NER----LSAETADAAYDQIADCLARPEFRPRALFERFGIEAIATTESAL 171
gi|146278446|Rsph170252412          ERR----LSAGTADASYDHIADCLSRPEFRPRALFERFGIEVIATTESAL 171
gi|222149863|Avi_3892               TTR----LSAQTADQTYDHIADCLTKPEFRPRALFERFNIEVIATTESPL 169
gi|110680135|RD12934                TTP----LSGETADACYDHIEAKLQTDAFRPRALFERFNVEVLATTEGAL 171
gi|161620871|BCAN_B0804             TER----LSQENADRIFDAIADQLTQPHMRPRALYDRFNIEAISTTDAAT 170
gi|23500543|BRA0812                 TER----LSQENADRIFDAIADQLTQPHMRPRALYDRFNIEAISTTDAAT 170
gi|163844934|BSUIS_B0804            TER----LSQENADRIFDAIADQLTQPHMRPRALYDRFNIEAISTTDAAT 170
gi|17988821|BMEll0476|gi|17988      TER----LSQENADRIFDAIADQLTQPHMRPRALYDRFNIEAISTTDAAT 170
gi|189022606|BAbS19_ll04000         --------------------------------------------------
gi|158424966|AZC3342                EER----LNAENADRLYDRIAERLETDACKPRALFQRFNIEAIATTESPL 173
gi|13473453|mll4056                 EEP----LDATTADRHYDTIATVLQWENFRPRALFERFNIEVIATTEGAL 172
gi|16264678|SMb21354                DEL----PSAASSDPLYDHVAECLTRPDYRPRALYERFNIEVISTTDSAL 172
gi|16125737|CC1490                  TEF----LEASNADDYFDRITAALATDAFRPRALFDRFNIETLATTEGPH 173
gi|221234494|CCNA_01557             TEF----LEASNADDYFDRITAALATDAFRPRALFDRFNIETLATTEGPH 173
gi|167645720|Caul_1756              TDF----LTADTADDYFNRINEALATEAFRPRALFDRFNIETLATTEGPQ 173
gi|197105913|PHZ_c2452              EIE----LSAETADLYFDRIGEALQSPAFRPRALFERFNIELLATTESPT 173
gi|84625827|XOO4170                 RLR----LDAGTADHYYDHITAALQTPDFLPRALFERFNIEVIATTESPL 175
gi|161898957|XOO4427                RLR----LDAGTADHYYDHITAALQTPDFLPRALFERFNIEVIATTESPL 175
gi|188574617|PXO_03860              RLR----LDAGTADHYYDHITAALQTPDFLPRALFERFNIEVIATTESPL 168
gi|21244968|XAC4251                 RIR----LDAGTADHYYDHITAALQTPAFLPRALFERFNIEVIATTESPL 175
gi|78049913|XCV4357                 RIR----LDAGTADHYYDHITAALQTPAFLPRALFERFNIEVIATTESPL 175
gi|66770505|XC4209                  RIR----LDAGSADHYYDHITAALQTPAFRPRALFERFNIEVIATTESPL 174
gi|21233539|XCC4117                 RIR----LDAGSADHYYDHITAALQTPAFRPRALFERFNIEVIATTESPL 174
gi|188993724|xccb100_4329           RIR----LDAGSADHYYDHITAALQTPAFRPRALFERFNIEVIATTESPL 174
gi|114568782|Mmar100230             EDR----LDGNTADAAYDTIDAALKRPDFRPRALYDRFNIELIATTESPV 191
gi|148554803|Swit1887               DVR----LAPETADLTFDSINEQLARPEFRPRALFDRFNIELLATTESPL 169
gi|103488522|Sala3046               DVQ----LSAETSDLYYDRITEALAIDAFRPRALFDRFGIEVIATTESPL 169
gi|90020587|Sde0940                 EKA----LSAATADDYYDGINAALQTPEFLPRALFDRFKIEWLATTESPL 173
gi|90020919|Sde1272                 THQ----LCNQTANTIYDSINEQLATDAFKPRALFERFNIELLATTESPL 174
gi|116249865|RL0099                 TEP----LTPDNADALYDHINAQLALPEFRPRALHQRFGIETIATTDGAL 174
gi|209551594|Rleg2_4023             TEP----LTSDNADALYDHINAQLALPEFRPRALHQRFGIETIATTEGAL 174
gi|86355751|RHECH00090              TEP----LTADNADALYDHINAQLARPEFRPRALHQRFGIETIATTEGAL 174
gi|190889727|RHECIAT_CH0000094      TEP----LTPNNADALYDHINAQLARPEFRPRALHQRFGIETIATTEGAL 174
```

```
gi|119962615|AAur0527             RGE----LSAEKADASFDSISAKLAEPGFRPRELFKDFNIEVLATTDDPL 165
gi|220911562|Achl_0787            REE----PDAGNPDASYDAISARLQEPGFRPRELFKDFNIEVLATTDDPL 176
gi|116669068|Arth0502             HEE----PTVENADRIFDAVSAKLSEPGFRPRQLFKEFNIEVLATTDDPL 169
gi|163842140|RSal332093421|gi|    ----------------------MLATPGFRPRALFDSFGIEVLATTDDPL 28
gi|158315003|Franean1_3197        TER----PGPDSADRIYDQVAERLADEAYRPRALFRRFDIEVLATTDDPC 193
gi|134101579|SACE5055             TVQ----PSADTADELFDELTVRLRKPEFRPRALYDSFAIAALATTDDPT 200
gi|152964556|Krad_0587            RVH----PSPETADEIYDQLAERIASPEFRPRALFDAFGVELLATTDAPT 173
gi|50955310|Lxx17350              TEH----PSAANADALYDQLAEMISSPAFRPRALFDRFRIAVLATTDDPA 225
gi|213691567|Blon_0669            NKP----LNEDTADSIYDELNELLASDDFTTRKLVKRFNIGFISTTDDPT 170
gi|50843764|PPA2329               TER----PSPENADRIYDTIAERIAQPDFLPRALMDSFDIAFIATTDDPC 173
gi|161485962|CE2377               NPDRLSA---ENADDIYDELSEILARPDFRPRALAEQFNLEILATTDDPL 171
gi|145296483|cgR_2394             NAERLNVGTPEHADAIFDELTDILAKPDFRPRALAEQFNLEVLATTDDPL 174
gi|62391323|cg2731                NAERLNVGTPEHADAIFDELTDILAKPDFRPRALAEQFNLEVLATTDDPL 174
gi|161486711|NCgI2398             NAERLNVGTPEHADAIFDELTDILAKPDFRPRALAEQFNLEVLATTDDPL 174
gi|126208494|APL_1020             --------------------------------------------------
gi|190150346|APP7_1077            --------------------------------------------------
gi|165976445|APJL_1038            --------------------------------------------------
gi|152977831|Asuc_0145            TDT---LFGPQSADKIWQECNELLQQPEFSARGIMRRMNVKFSGTTDDPI 161
gi|161511000|MS0544               TNT---LFNPQSADKIWQECNELLQQPEFSARGIMRQMNVKFSGTTDDPI 161
gi|148827203|CGSHiGG_02740        TNM---LFNPQNAEKIWHQCNEMLQQPEFSARGIMQKMNVKLVGTTDDPI 161
gi|68248599|NTHI0056              TNM---LFNPQNAEKIWHQCNEMLQQPEFSARGIMQKMNVKLVGTTDDPI 161
gi|113461738|HS_1602              TDT---IFSPETAEKIWHKGKELLQQPEFSARGIMKKMNVNLVGTTDDPI 157
gi|170718551|HSM_0409             TDT---IFSPETAEKIWHKGKELLQQPEFSARGIMKKMNVNLVGTTDDPI 161
gi|62181648|SC3078                TGT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|197249740|SeAg_B3309           TGT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|198243637|SeD_A3482            TGT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|207858396|SEN2980              TGT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|194734239|SeSA_A3319           TGT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|205354057|SG3031               TGT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|161616115|SPAB03917            TGT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|16766437|STM3137               TGT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|194448866|SeHA_C3383           TGT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|194446286|SNSL254_A3389        TGT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|16761915|STY3308               TGT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|56415079|SPA3005               TST---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|197364008|SSPA2803             TST---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|161506293|Sari04489            TDT---LFGPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPI 162
gi|37524198|plu0176               SGI---LFGPQTADEIWHQCNEKLAAPAFSARGIMQQMNVRMVGTTDDPV 162
gi|28901561|VPA1706               TNT---LFSPDTADQIWHQCNELLATPEFTARGIMQQMNVVMAGTTDDPI 161
gi|37677254|VV1594                TNA---LFSPDTADQIWHQCNELLATPEFTARGIMQQMNVVMAGTTDDPI 161
gi|27367460|VV21070               TNT---LFSPDTADQIWHQCNELLATPEFTARGIMQQMNVVMAGTTDDPI 161
gi|119943921|Ping0131             TGK---LFNPDTAEVIWHECNEMLATPEFSARGIMKKMNVVMAGTTDDPI 166
gi|152996793|Mmwyl12780           TDV---LFSPKTADAIWDQCNEMLQQPEFSARGIMQQMKVKMVGTTDDPA 161
gi|117625400|APECO1_3327          TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|91212521|UTI89_C3530           TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 182
gi|16130987|b3092                 TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|161486116|c3850                TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|215488422|E2348C_3385          TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|170082634|ECDH10B_3268         TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|157155304|EcE24377A_3560       TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|218691395|ECED1_3759           TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|157162569|EcHS_A3275           TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|218701863|ECIAI39_3589         TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|170018657|EcolC_0608           TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|110643336|ECP_3183             TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|218560176|ECS88_3488           TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|170683343|EcSMS35_3384         TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|89109861|                      TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|30064435|S3339                 TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|82545347|SBO2953               TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|56480255|SF3132                TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|110806974|SFV3133              TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|74313639|SSON3245              TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|187731951|SbBS512_E3528        TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|209397829|ECH74115_4407        TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|15803633|Z4445                 TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|15833228|ECs3974               TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
```

```
gi|218696796|EC55989_3506          TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|209920564|ECSE_3373             TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|218706716|ECUMN_3576            TGK---LLSPSTADEIWNECNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|218551598|EFER_4383             TGK---LLSPSTADEIWNQCNELLAQDNFSARGIMQQMNVKMVGTTDDPI 161
gi|206579202|KPK_0591              TGK---LLSPSTADEIWDQCNDLLAQDAFSARGIMKQMNVKMVGTTDDPI 161
gi|152972036|KPN_03520             TGK---LLSPSTADEIWDQCNDLLAQDAFSARGIMKQMNVKMVGTTDDPI 157
gi|146313181|Ent638_3546           TGK---LLSPTTADEIWNQCGDLLAQDNFSARGIMKQMNVKMVGTTDDPI 161
gi|157148666|CKO_04494             TGK---LLSPATADEIWDRCNELLAQDAFSARGIMQQMNVKMVGTTDDPV 182
gi|156935638|ESA_03503             TGK---LLSPKTADEIWNQCNDLLAQDAFSARGIMQQMNVKMVGTTDDPI 161
gi|161484751|y3600                 TGK---LLSPATSEEIWQRGNELLAQDPFSARGIMQQMNVKMVGTTDDPI 161
gi|161511340|YP2899                TGK---LLSPATSEEIWQRGNELLAQDPFSARGIMQQMNVKMVGTTDDPI 161
gi|108809201|YPA3210               TGK---LLSPATSEEIWQRGNELLAQDPFSARGIMQQMNVKMVGTTDDPI 161
gi|162419999|YPAngolaA1104         TGK---LLSPATSEEIWQRGNELLAQDPFSARGIMQQMNVKMVGTTDDPI 161
gi|145597678|YPDSF0366             TGK---LLSPATSEEIWQRGNELLAQDPFSARGIMQQMNVKMVGTTDDPI 161
gi|170022805|YPK_0554              TGK---LLSPATSEEIWQRGNELLAQDPFSARGIMQQMNVKMVGTTDDPI 161
gi|108810613|YPN0448               TGK---LLSPATSEEIWQRGNELLAQDPFSARGIMQQMNVKMVGTTDDPI 161
gi|16120908|YPO0579                TGK---LLSPATSEEIWQRGNELLAQDPFSARGIMQQMNVKMVGTTDDPI 161
gi|153949672|YpsIP317580491        TGK---LLSPATSEEIWQRGNELLAQDPFSARGIMQQMNVKMVGTTDDPI 161
gi|51597770|YPTB3478               TGK---LLSPATSEEIWQRGNELLAQDPFSARGIMQQMNVKMVGTTDDPI 161
gi|186896961|YPTS_3663             TGK---LLSPATSEEIWQRGNELLAQDPFSARGIMQQMNVKMVGTTDDPI 161
gi|123443892|YE3708                TGK---LLSPATADEIWQRGNELLAQDSFSARGIMQKMNVKMVGTTDDPI 161
gi|157372554|Spro4321              TGT---LLSPTTADDIWQRGNALLAQDDFRARGIMQQMNVKMVGTTDDPI 161
gi|108810612|YPN0447               --------------------------------------------------
gi|50119594|ECA0645                TGT---LLSPSTAKGIWDRCNAMLERDDFTARGIMQQMNVKMVGTTDDPI 161
gi|54303188|PBPRB1515              TDR---VLSLETADSIWNECNAMLETPEYSAQSLMKKMNVRMIGTTDDPI 97
gi|60681862|BF1904G                DK----VLNPKTAREIYDECNEKLSSQEYSARGMMRRYHVETVCTTDDPI 161
gi|53713582|BF2293                 DK----VLNPKTAREIYDECNEKLSSQEYSARGMMRRYHVETVCTTDDPI 161
gi|161511193|BT0823                DK----ILSPKTAREIYDECNEKLAQPEYSARGMMRRYHVEVVCTTDDPI 161
gi|150005586|BVU3074               NK----VLNPKTAREIFDECNEKLAKPEYSARGMMRRYHVETVCTTDDPV 161
gi|212550592|CFPG_235              NK----LLNPKSALEIYEHCTDLLRTPAFSTRNLMERYNVEVICTTDDPI 161
gi|150009769|BDI3184               HK----VLNPASAEEIYATCTDKLRTPEYRAQAIMKRMNVEVVCTTDDPI 225
gi|146301998|Fjoh_4262             YD----LLNEKSAEKIYTETTEKINSQAYSTQNLLKKVNAELVCTTEDPI 162
gi|192359962|CJA_2772              TDR---LLDSRSAKRTWDECNELLATPEFSARGLMTQANVKLVCTTDDPI 160
gi|182416100|Oter_4293             FD----LINQKSADKIWREANEKLAT--MRVHDILAANKVAVICTTDDPA 158
gi|116624408|acid5330              DE----LLNEQSAARVWEQANAVLATPELTAHRILEKFHVKAVCTTDDPT 159
gi|225621141|BHWA1_02239           HEV---INED-NADTIWEKANQKLQN--MTVKDILKKFKVHTIGTTDDPT 158
gi|18309134|CPE0152                EET---LSEK-TASTIWEKCNKVIENNNLTAKKLIEMANVVYIGTTDDPI 160
gi|160880875|Cphy_2743             YG----TLSAKTSEEVFKLCNEKLSQPEMSVRGLIKESNVDTICTTDDPI 160
gi|189485070|TGRD_067              EG----VLNGETAKEVWTICNEKLREKSMSARSMIKQSSVTLICTTDDPV 160
gi|56962769|ABC0995                DE----LLSSETAEAIWSATKEQLAAPERSVQGIIKESNVKVICTTDDPS 164
gi|222529391|Athe_1402             DD----VLNERSAPIIWEKTNKVL--KELGARDIILKSNVEIICTTDDPI 158
gi|146296954|Csac1949              DE----ILNEKSAPIIWEKANKVL--EELGARDIILKSNVEVICTTDDPV 158
gi|150018894|Cbei_4082             YE----PLDEDTAPEIWKRANELLNGEGFNVRDLIIKSNVETICTTDDPI 160
gi|150019467|Cbei_4657             YE----PLDEETAPAIWKKANELLSGEGFRARDLITKSKVETICTTDDPT 163
gi|20808343|TTE1939                YD----LLNEKTAKSIWERANEVISQEDFSARNILKKSNVKVVITTDDPV 160
gi|15893980|CAC0692                YE----VISEKNAPVIWEKANKVLN-DGLTVREIIKKSNVKAICTTDDPI 159
gi|150016705|Cbei_1832             DE----VLNEKSAPAIWEKVNKLLNSDDFTVRNLIKKSNVKVICTTDDPI 160
gi|52081820|BL00708                DD----RLDEKSAPHIWERVNEQLAGGGFGARDLIEKSNVETVVTTDDPT 160
gi|52787207|BLi03516               DD----RLDEKSAPHIWERVNEQLAGGGFGARDLIEKSNVETVVTTDDPT 160
gi|157693739|BPUM2987              QD----LLNEKNADTIWQKVNEKLQEEGFGARDFIMKSNVETVVTTDDPI 160
gi|154685671|RBAM012380            YD----VLNEKTAAAIWKKTNELLQGDGFGARDLILKSNVKVICTTDDPA 160
gi|16078295|BSU12300               YE----ILNEKSGSAIWKQTNKLLKGEGFGARDLIVKSNVKVVCTTDDPV 160
gi|23097822|OB0367                 DE----ILNKESGPSIWEKVNKCLAQDDFGVRELINQSNVQVVCTTDDPI 160
gi|15613268|Bh0705                 HQ----PLDETNGENVWDACNRLLQQEAFTPRALIERSNVRAIGTTDDPT 160
gi|138895421|GTNG_1765             GV----LLNEQTWKEVWDHCNELLQQEGYSARSFMIQSNVEWIGTTDDPL 160
gi|220929098|Ccel_1676             DE----LLVPESAPEIWKACNNKLKSPEFTTRGIIERFNVKALCTTDDPI 162
gi|81427739|LSA0129                DE----RLTAESAPRIWEKANQLLQTDDFKPRQLIKNSNVQVVCTTDDPA 163
gi|116332824|LVIS_0150             DE----VFSRKTAPEIWKKANALLQTEDFKPRNLIKNSNVKAVCTTDDPA 163
gi|116332809|LVIS_0135             DE----PFTTKNAPAIWEKANALLKTDDFKPRNLIKKSNVKVVCTTDDPA 163
gi|116512444|LAACR_1742            DK----LISSETAKEIWDEANQMLATDDFRPRALIKNSNVKVVCTTDAPV 159
gi|125623707|llmg_0862             DK----LISSETEKEIWDEANQMLATDDFRPRALIKNSNVKVVCTTDDPV 159
gi|15673610|L0019                  DK----LISSETAKEIWDEANQMLATDDFRPRALIKNSQVKVVCTTDDPV 159
gi|150389106|Amet1291              DE----LLKRSNWKSIYDRANKIIKEEGLTVRKLIKQSNVDFICTTDDPI 165
gi|70727647|SH2648                 ED----LLTSDNAEAIYHRANDYLKQHHTTTQSLITDSNVNLICTTDNPT 159
gi|25010729|gbs0674                NE----ILTESNAEEIYHRLNHFLKEHKISPRRLIADSKVMFIGTTDYPL 158
gi|22536865|SAG0701                NE----ILTESNAEEIYHRLNHFLKEHKISPRRLIADSKVMFIGTTDHPL 158
gi|76786751|SAK0827                NE----ILTESNAEEIYHRLNHFLKEHKINPRRLIADSKVMFIGTTDHPL 158
```

```
gi|195977823|Sez_0689              CE----LLTEDNAEEIYHRINAYLVEHQISPRKLIADSRVRFIGTTDHPL 158
gi|225868878|SZO_13120             CE----LLTEDNAEEIYHRINAYLVEHQISPRKLIADSLVRFIGTTDHPL 158
gi|225870118|SEQ_0716              CE----LLTEDNAEEIYHRINAYLVEHQISPRKLIADSRVRFIGTTDHPL 158
gi|94990680|MGAS10270_Spy1171      EE----LLTQDNAEEIYHRLNAYLVDHKISPRKLIADSKVTFIGTTDHPL 167
gi|222153335|SUB1203               EE----VLTEENAEEIYNRLNTYLLENKVSPRKLIADSKVTFIGTTDHPL 158
gi|146318813|SSU051159|gi|1463     EE----LLTEENAERLYHQLNAYLQEHQVSPRKLIADVRVTFIGTTDHPL 170
gi|146321021|SSU981174             EE----LLTEENAERLYHQLNAYLQEHQVSPRKLIADVRVTFIGTTDHPL 170
gi|225859885|SP70585_2238          EE----LLTEENAERLYHQLNTYLQEHQISPRKLIADARVAFIGTTDHPL 158
gi|157147928|CKO_03735             EE----TLNSKNWRTIMAQCNEQLRRDDFLPQALITRSNVEALCTTDGPL 159
gi|220927672|Ccel_0214             DE----ILKESNADRIYDHCNAKIKEEKLSPVKMIKASNVRLVCTTDDPV 157
gi|78221672|Gmet0451               TA----PLTPSSAECIWEEANSMLATENFRPRKLLARAGVEVLCTTDDPA 158
gi|148269995|Tpet0860              KK----VISEETAEEIWEETKKKLP--EMTPQKLLRDMKVEILCTTDDPV 155
gi|170288680|TRQ2_0883             KK----VISEETAEEIWEETKKKLP--EMTPQKLLRDMKVEILCTTDDPV 155
gi|15642839|TM0064                 KK----VISEETAEEIWEETKKKLP--EMTPQKLLRDMKVEILCTTDDPV 155
gi|160902320|Pmob0854              ED----TISKYTAEKIWQDTSLLLKSESMKPQRLLKEMNVEVMCTTNDPT 158
gi|157364144|Tlet1289              NE----TISRQTAETIWYRSKQILNEDKMKPQNILREMNVEIMCTTDDPT 158
gi|20808989|TTE2645                PKR----------D---------------------------------- 68
gi|159042183|Cmaq1622             RVSG---INEGNVNDVVKAIESVKGNEDRAVKVIRDSSRVRKSVLTLNPL 137
gi|32474066|RB6095                 EED---RLHLNNWEALYDRSESIMASAQWSET-VLDQSNVQAVFLTNDFD 155
gi|15613056|Bh0493                 GLD------PATRDLQVYREYFAKKTSEEQVDTVLQLANVSDVVMTNDPF 147

gi|94971332|Acid345_4306           DSLSDHKSLRAG-----DWKNRIIPTFRPDSVVDPDF-EGFASNVDDLGK 210
gi|225874247|ACP_2685              DSLADHQIIRDS-----GWPARILATFRPDSVVDPDF-TGFADNIATLGA 210
gi|221639919|RSKD131_1820          DDLRWHRMIRES-----GWQGRVVTAYRPDAVVDPEF-AGFAENVAQLGQ 215
gi|126462900|Rsph17029_2139        DDLRWHRMIRES-----GWQGRVVTAYRPDAVVDPEF-AGFAENVAQLGQ 215
gi|146278446|Rsph170252412         DDLRWHRMIRDS-----GWQGRVITAYRPDAVVDPEF-AGFAGNVAALGE 215
gi|222149863|Avi_3892              DELKWHETIRAS-----GWDGRVVTAYRPDPVVDPQF-EGFATNIERFGQ 213
gi|110680135|RD12934               DDLRWHREIRDS-----GWSGRVVTTYRPDAVVDPAF-EGFAENVAQLGA 215
gi|161620871|BCAN_B0804            DPLIYHDEVIAS-----GWHGRIIPAYRPDAAVDAGR-PDFASEVEKLVG 214
gi|23500543|BRA0812                DPLIYHDEVIAS-----GWHGRIIPAYRPDAAVDAGR-PDFASEVEKLVG 214
gi|163844934|BSUIS_B0804           DPLIYHDEVIAS-----GWHGRIIPAYRPDAAVDAGR-PDFASEVEKLVG 214
gi|17988821|BMEll0476|gi|17988     DPLIYHDEVIAS-----GWHGRIIPAYRPDAAVDAGR-PDFASEVEKLVG 214
gi|189022606|BAbS19_ll04000        -------------------------------------------------- 
gi|158424966|AZC3342               DELKWHQMIRDS-----GWNGRVVTAYRPDPVVDPDF-EGFRDNLTRFGE 217
gi|13473453|mll4056                DDLEWHKMIRDS-----GWEGRVVTAYRPDAVVDPDF-EGFSANLDRLGE 216
gi|16264678|SMb21354               DDLGWHAKILES-----GWKGRVVPAYRPDAVVDPDF-QGFPTNLDKLGD 216
gi|16125737|CC1490                 ESLQHHAAIRES-----GWGGHVITAYRPDAVIDFED-ERSPRAFERFAE 217
gi|221234494|CCNA_01557            ESLQHHAAIRES-----GWGGHVITAYRPDAVIDFED-ERSPRAFERFAE 217
gi|167645720|Caul_1756             DRLEHHRAIRES-----GWGGHVITAYRPDAVIDFED-ERFGRAFERFGE 217
gi|197105913|PHZ_c2452             DDLRHHKAMRAS-----GWKGRVVTAYRPDPVIDAEH-EDFRASLSRFGE 217
gi|84625827|XOO4170                DPLHHHAAIAAS-----GWQGRVVTAYRPDPVVDPEH-AQFAGALQQFGA 219
gi|161898957|XOO4427               DPLHHHAAIAAS-----GWQGRVVTAYRPDPVVDPEH-AQFAGALQQFGA 219
gi|188574617|PXO_03860             DPLHHHAAIAAS-----GWQGRVVTAYRPDPVVDPEH-EQFAGALQQFGA 212
gi|21244968|XAC4251                DRLQHHAAIAAS-----GWQGRVVTAYRPDPVVDPEH-EQFAGALQQFGA 219
gi|78049913|XCV4357                DRLQHHAAIAAS-----GWQGRVVTAYRPDPVVDPEH-EQFAGALQQFGA 219
gi|66770505|XC4209                 DTLEHHATIRDS-----GWTGRVLTAYRPDAVVDPEH-EQFASALQQFAA 218
gi|21233539|XCC4117                DTLEHHATIRDS-----GWTGRVLTAYRPDAVVDPEH-EQFASALQQFAA 218
gi|188993724|xccb100_4329          DTLEHHATIRDS-----GWSGRVLTAYRPDAVVDPEH-EQFASALQQFGA 218
gi|114568782|Mmar100230            DTLDHHHSIAAS-----DWSGRVITAYRPDCVIDADH-EDFHASLEAFGD 235
gi|148554803|Swit1887              DPLDHHAAIRAS-----GWTGRVITAFRPDPVVDPDF-EGFAANLDRLGA 213
gi|103488522|Sala3046              DSLDHHAVIRAANASG-EWGGRVITAYRPDPVVDPEF-EGFRDNLARFSN 217
gi|90020587|Sde0940                NDLRHHQKIRES-----GWKGNVITAYRPDPVVDPDF-EGFASNVAYLGK 217
gi|90020919|Sde1272                DNLAHHQTILDS-----GWTGRVITAFRPDPVVDPEF-EGFSQNIDKLAE 218
gi|116249865|RL0099                DPLAHHQKMAAD-----GWIGKVRTTYRPDSVTDPDA-VGFRDNLVKFGE 218
gi|209551594|Rleg2_4023            DPLAHHQKMAAD-----GWIGKVRTTYRPDSVTDPDA-IGFRDNLVKFGE 218
gi|86355751|RHECH00090             DPLVHHQKMAAD-----GWIGKVRTTYRPDSVTDPDA-VGFRDNLIKFGE 218
gi|190889727|RHECIAT_CH0000094     DSLAHHQKMAAD-----GWIGKVRTTYRPDSVTDPDS-VGFRDNLIKFGE 218
gi|119962615|AAur0527              DNLESHRALAEDP----SFHSRVLPTFRPDAYINIAH-PQWAANVERIIS 210
gi|220911562|Achl_0787             DSLSSHQAIAADP----SFHGRVLPTFRPDQYLNIAH-PQWQDNVERLIA 221
gi|116669068|Arth0502              DSLDSHAALAADP----AFAGRVLPTFRPDPYLNIAH-PEWRDNVERLIG 214
gi|163842140|RSal332093421|gi|     DSLSDHAALAADS----SFNGRVLPTFRPDAYLNAHA-PGFSDRVDRLIV 73
gi|158315003|Franean1_3197         SDLAAHRALTADP----DWPGRVIPTFRPDRYLEPTR-AGWDEAVRRLGE 238
gi|134101579|SACE5055              DDLSYHRALAEDP----SWSGRVVPTFRPDRCLDATG-SGWTTALDDLAA 245
gi|152964556|Krad_0587             SRLEHHAAIAAS-----GWSGTVVPTFRPDALVHLGR-PGWKRDVEELAE 217
gi|50955310|Lxx17350               DDLEAHARLAADP----GFTGRVVPTFRADRYMHPDE-PGRAGRLDRLAA 270
gi|213691567|Blon_0669             DDLVLHDKVRADA----NFPARLAPCFRPDRYLAVDR-VDWAQLCDQLGE 215
```

```
gi|50843764|PPA2329              DDLDGHARLAADE----TFTRRVAPTFRPDKYLEPAA-GGWTGLLARLSE 218
gi|161485962|CE2377              DDLADHKALAQDP----TFSPRVLPTFRPDAYTKMYN-PGWAEKTTRLID 216
gi|145296483|cgR_2394            DDLADHKALADDP----TFSPRVLPTFRPDAYTKMYN-AGWAEKTTKLID 219
gi|62391323|cg2731               DDLADHKALADDP----TFSPRVLPTFRPDAYTKMYN-AGWAEKTTKLID 219
gi|161486711|NCgI2398            DDLADHKALADDP----TFSPRVLPTFRPDAYTKMYN-AGWAEKTTKLID 219
gi|126208494|APL_1020            --------------------------------------------------
gi|190150346|APP7_1077           --------------------------------------------------
gi|165976445|APJL_1038           --------------------------------------------------
gi|152977831|Asuc_0145           DSLEYHKAIAEDNA----FDIEVAPSWRPDKAVKIEL-DQFNDYLEQLER 206
gi|161511000|MS0544              DSLEYHKAIAEDRD----FDIEVAPSWRPDKAVKIEL-PQFNDYIKQLEQ 206
gi|148827203|CGSHiGG_02740       DSLQYHQAIKNDES----FDIDVVPSWRPDKVFKIEL-PQFNDYLVQLSE 206
gi|68248599|NTHI0056             DSLQYHQAIKNDES----FDIDVVPSWRPDKVFKIEL-PQFNDYLVQLSE 206
gi|113461738|HS_1602             DSLEYHKAIAEDDT----FDVEVVPSFRPDRAFKIEL-PLFNDYIEQLGK 202
gi|170718551|HSM_0409            DSLEYHKAIAEDNT----FDVEVVPSFRPDRAFKIEL-PLFNDYIEQLGK 206
gi|62181648|SC3078               DSLEYHRQIAADDS----IDIEVAPSWRPDKVFKIEL-DGFVDYLRKLEA 207
gi|197249740|SeAg_B3309          DSLEYHRQIAADDS----IDIEVAPSWRPDKVFKIEL-DGFVDYLRKLEA 207
gi|198243637|SeD_A3482           DSLEYHRQIAADDS----IDIEVAPSWRPDKVFKIEL-DGFVDYLRKLEA 207
gi|207858396|SEN2980             DSLEYHRQIAADDS----IDIEVAPSWRPDKVFKIEL-DGFVDYLRKLEA 207
gi|194734239|SeSA_A3319          DSLEYHRQIAADDS----IDIEVAPSWRPDKVFKIEL-DGFVDYLRKLEA 207
gi|205354057|SG3031              DSLEYHRQIAADDS----IDIEVAPSWRPDKVFKIEL-DGFVDYLRKLEA 207
gi|161616115|SPAB03917           DSLEYHRQIAADDS----IDIEVAPSWRPDKVFKIEL-DGFVDYLRKLEA 207
gi|16766437|STM3137              DSLEYHRQIAADDS----IDIEVAPSWRPDKVFKIEL-DGFVDYLRKLEA 207
gi|194448866|SeHA_C3383          DSLEYHRQIAADDS----IDIEVAPSWRPDKVFKIEL-DGFVDYLGKLEA 207
gi|194446286|SNSL254_A3389       DSLEYHRQIAADDS----IDIEVAPSWRPDKVFKIEL-DGFVDYLGKLEA 207
gi|16761915|STY3308              DSLEYHHQIAADDS----IDIEVAPSWRPDKVFKIEL-DGFVDYLGKLEA 207
gi|56415079|SPA3005              DSLEYHRQIAADDS----INIEVAPSWRPDKVFKIEL-DGFVDYLGKLEA 207
gi|197364008|SSPA2803            DSLEYHRQIAADDS----INIEVAPSWRPDKVFKIEL-DGFVDYLGKLEA 207
gi|161506293|Sari04489           DSLAYHRQIAADDS----FGIEVAPSWRPDKAFKIEL-DGFVDYLGKLEA 207
gi|37524198|plu0176              DSLQYHRQIATDES----FDIEVLPSWRPDRVFKIEL-AGFIDYIKQLEA 207
gi|28901561|VPA1706              DSLEHHKAIAEDDT----FNVKVLPSWRPDKAFKIEL-DLFADYMHKLGE 206
gi|37677254|VV1594               DSLEHHKAIAEDDT----FNVKVLPSWRPDKAFKIEL-DVFADYMHKLGE 206
gi|27367460|VV21070              DSLKHHKTIAEDDT----FNVKVLPSWRPDKAFKIEL-DLFADYMHKLGE 206
gi|119943921|Ping0131            HSLEHHKAIAEDPA----CDIEVAPSWRPDRAFKIEI-EGFSEYMRLLGE 211
gi|152996793|Mmwyl12780          DSLEHHKFIANDNS----FDIAVTPSWRPDRAFKVDH-VGFKDYLNKLGK 206
gi|117625400|APECO1_3327         DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|91212521|UTI89_C3530          DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 227
gi|16130987|b3092                DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|161486116|c3850               DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|215488422|E2348C_3385         DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|170082634|ECDH10B_3268        DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|157155304|EcE24377A_3560      DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|218691395|ECED1_3759          DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|157162569|EcHS_A3275          DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|218701863|ECIAI39_3589        DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|170018657|EcolC_0608          DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|110643336|ECP_3183            DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|218560176|ECS88_3488          DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|170683343|EcSMS35_3384        DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|89109861|                     DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|30064435|S3339                DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|82545347|SBO2953              DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|56480255|SF3132               DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|110806974|SFV3133             DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|74313639|SSON3245             DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|187731951|SbBS512_E3528       DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|209397829|ECH74115_4407       DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|15803633|Z4445                DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|15833228|ECs3974              DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|218696796|EC55989_3506        DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|209920564|ECSE_3373           DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|218706716|ECUMN_3576          DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|218551598|EFER_4383           DSLEHHAEIAKDGS----FTIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|206579202|KPK_0591            DSLEHHAVVAKDTS----FDIKVLPSWRPDKAFNIEQ-ATFNDYMAKLGE 206
gi|152972036|KPN_03520           DSLEHHAAVAKDNS----FDIKVLPSWRPDKAFNIEL-ATFNDYMAKLGE 202
gi|146313181|Ent638_3546         DSLEHHAVVAKDSS----FDIKVLPSWRPDKAFNIEQ-ATFGDYMAKLAE 206
gi|157148666|CKO_04494           DSLEHHATVAKDST----FTVKVLPSWRPDKAFNIEQ-ATFGDYMTKLGE 227
gi|156935638|ESA_03503           DSLEHHATIAKDSA----FTVKVLPSWRPDKAFNIEQ-ATFADYMAKLAE 206
```

```
gi|161484751|y3600                DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|161511340|YP2899               DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|108809201|YPA3210              DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|162419999|YPAngolaA1104        DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|145597678|YPDSF0366            DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|170022805|YPK_0554             DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|108810613|YPN0448              DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|16120908|YPO0579               DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|153949672|YpsIP317580491       DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|51597770|YPTB3478              DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|186896961|YPTS_3663            DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|123443892|YE3708               DDLRHHKAIAADGS----FNIKVLPSWRPDKAFNIEA-AGFNDYMQRLEA 206
gi|157372554|Spro4321             DDLRHHKAIAADSS----FDIKVLPSWRPDKAFNIDA-PGFNDYLQKLEA 206
gi|108810612|YPN0447              --------------------------------------------------
gi|50119594|ECA0645               DDLRHHKAVAQDSS----FSIKVLPSWRPDKAFNIEL-ATFNDYMAKLGE 206
gi|54303188|PBPRB1515             DDLSAHKSIAGNKD----FNVKVLPSWRPDRSFNINA-SDYAHYLEQLGD 142
gi|60681862|BF1904G               DSLEYHIRTRES-G----FEIKMLPTWRPDKVMAVEVPSDFRTYIEKLSE 206
gi|53713582|BF2293                DSLEYHIRTRES-G----FEIKMLPTWRPDKVMAVEVPSDFRTYIEKLSE 206
gi|161511193|BT0823               DSLEYHIQTRES-G----FEIKMLPTWRPDKAMAVEVPADFRAYVEKLSA 206
gi|150005586|BVU3074              DSLEYHIKTRES-G----FEIKMLPTWRPDKAMAVEVPADFRAYMEKLSA 206
gi|212550592|CFPG_235             DSLEHHIKLKNE-G----FKIKVLPTWRPDKTMEVENPSDYRKYVEKLST 206
gi|150009769|BDI3184              DSLEFHQKIRSS-G----CHIRVYPAWRPDKVLAIDNFKALNDYLSKLEA 270
gi|146301998|Fjoh_4262            DSLEFHKKFANN-S----TGIKMSTAFRPDKAILIAN-DGYNAYLDTLGD 206
gi|192359962|CJA_2772             HDLAHHKTVAADKS----FKTAMLPTWRPDRAMMVEDAESYNKYLDRLAI 206
gi|182416100|Oter_4293            DSLEQHEKIKKL-G----IKTRVYPTFRPDKALNVGSPAAYNAWLEKLAG 203
gi|116624408|acid5330             DDLAAHQAIAAS-G----LATKVFPTFRPDKALHVHAPELFNPWVDRLQV 204
gi|225621141|BHWA1_02239          DNLEYHKLINEGKAQIGKIYTKVVPSFRPDKAINIEMPD-FSDYIKKLEN 207
gi|18309134|CPE0152               DDLRYHKQLKEDIN----FKCSVNPSFRPEKAMKIQN-EGFKEYIGKLAE 205
gi|160880875|Cphy_2743            DSLEWHKAIAKDST----FDVKVLPAWRPDKAMNLEKPD-YLDYLAKLES 205
gi|189485070|TGRD_067             DNLEWHKKIKEEES----FDVQVLPTWRPDRALNIEKP-GFKDYIARFSE 205
gi|56962769|ABC0995               DHLEAHQQIRKEGA----CQAAVYPAFRPDKALVASAPS-FVPYLETLGA 209
gi|222529391|Athe_1402            DTLEYHLKLKEDKD----FNVKVYPTFRPDKGVNIER-ETFIPWVKKLAE 203
gi|146296954|Csac1949             DTLEYHLKLKDDKN----FNVKVYPTFRPDKGVNIER-ETFIPWVEKLGE 203
gi|150018894|Cbei_4082            DTLEYHIKIKEDAS----FNVNVLPTLRPDKGIEINL-DGFVSWVKDLEK 205
gi|150019467|Cbei_4657            DTLEYHIKIKEDES----FNVNVLPTFRPDKGIEINL-DGFVSWVKKLEG 208
gi|20808343|TTE1939               DSLEYHIKLKEEKD----FDIKVLPAFRPDKGLNIEK-DDFLSWIKKLES 205
gi|15893980|CAC0692               DSLEYHLKLKEDTS----FNVKVLPAFRPDKALGINK-DGYTDWVSKLAK 204
gi|150016705|Cbei_1832            DSLEYHLAIKEDKS----FDVKVLPAFRPDKALGVNK-ETFEDWFNKLEE 205
gi|52081820|BL00708               DSLEYHVKLKDE-D----FNVSVLPGFRPDKAMEINQ-EGFAAWVRKLET 204
gi|52787207|BLi03516              DSLEYHVKLKDE-D----FNVSVLPGFRPDKAMEINQ-EGFAAWVRKLET 204
gi|157693739|BPUM2987             DSLQYHQKLREE-G----FSVQVLPGFRPDKALDIAN-DLFEKYVHELAE 204
gi|154685671|RBAM012380           DALTYHELLKES-D----FPVQVLPGFRPDKGLDISSP-GFADWVRSLES 204
gi|16078295|BSU12300              DSLEYHLLLKEDKD----FPVSVLPGFRPDKGLEINR-EGFPEWVQALED 205
gi|23097822|OB0367                DDLSFHQMLVEDDD----FSVKVLPGFRPDKAIEINQ-EGFIDYVDSLEK 205
gi|15613268|Bh0705                DSLLYHQKLQADDT----FHVKVIPTFRPDGALKIEQ-DSFADWVAKLSD 205
gi|138895421|GTNG_1765            DDLMDHQKIAQDSS----FSIKVVPSFRPDAVIEINRP-SFLDYVSKLGE 205
gi|220929098|Ccel_1676            DDLKSHIELEKDKN----FKVKVLPTFRPDKALQIEN-TGFADWIASLSK 207
gi|81427739|LSA0129               SDLHYHQLLKPEEAAN---GFKTLPAMRPDQLMQIDR-DGFGDYLKTLGK 209
gi|116332824|LVIS_0150            SDLHYHKEIKKDEAEN---GFRTLPAMRPDKLIQIDR-DGFAAYLKELGD 209
gi|116332809|LVIS_0135            SDLHYHQLLKQEEATN---GFKVLPAMRPDNALRINQ-GDYDAYLRQLSD 209
gi|116512444|LAACR_1742           SKLSYHKSLAKEEKD-----FKVLPALRPDHLISITD-GHFAEYLTELSA 203
gi|125623707|llmg_0862            SKLSYHKSLAKEEKD-----FKVLPALRPDHLISITD-GHFAEYLMELSA 203
gi|15673610|L0019                 STLSYHKLLAKEEKD-----FKVLPALRPDHLISITD-GHFADYLTELSA 203
gi|150389106|Amet1291             DSLEFHDEIAKDED----FKVKVVPTFRPDGAF-GIGEETFIDFIDKLGD 210
gi|70727647|SH2648                DDLNYHDAIKAKDG----FNTTVLPAFRPDDVF-KVGDPAFTDLLQKLEN 204
gi|25010729|gbs0674               DTLKWHKKLAADES----FKTVVAPTFRPDEAF-IEHRH-FVDFITKLGD 202
gi|22536865|SAG0701               DTLEWHKKLAADES----FKTVVAPTFRPDEAF-IEHRH-FVDFITKLGD 202
gi|76786751|SAK0827               DTLEWHKKLAADES----FKTVVAPTFRPDEAF-IEHRH-FVDFITKLGD 202
gi|195977823|Sez_0689             DDLAWHKRLTADDT----FETVVAPTFRPDEAF-IEHQR-FADFVARLAQ 202
gi|225868878|SZO_13120            DDLAWHKRLAADDT----FETVVAPTFRPDEAF-IEHQC-FADFVARLAQ 202
gi|225870118|SEQ_0716             DDLAWHKRLAADDT----FETVVAPTFRPDEAF-IEHQR-FADFVARLAQ 202
gi|94990680|MGAS10270_Spy1171     DDLEWHQQLADDST----FETVVAPTFRPDEAF-VEHRN-FAQFVERLGT 211
gi|222153335|SUB1203              DTLEWHQKLAEDKS----FETIVAPTFRPDEAF-IEHHN-FKGFLDKLSQ 202
gi|146318813|SSU051159|gi|1463    DSLEWHKRLAEDST----IDTVVAPTFRPDEAF-IEHRN-FNPFIKRLEE 214
gi|146321021|SSU981174            DSLEWHKRLAEDST----IDTVVAPTFRPDEAF-IEHRN-FNPFIKRLEE 214
gi|225859885|SP70585_2238         DNLEWHKRLAEDSP----IDTVVAPTFRPDEAF-IEHRN-FNLFIKRLEE 202
gi|157147928|CKO_03735            DDLEYHQRLAEKRD----FSTLVLPTFRPDELF-ETHPDRFLTFVSRLSQ 204
```

```
gi|220927672|Ccel_0214            DSLEYHLLLGKNSE----KGFDVFPTFRPDNVLKILN-DNFIEYIRKLEA 202
gi|78221672|Gmet0451              DQLEAHRQLVEDKE----VSFRVFPTFRPDSALCLHTRQNFHDWRSRLEA 204
gi|148269995|Tpet0860             STLEHHRKAKEVVEG-----VTILPTWRPDRAMNVDK-EGWKEYVEKMGE 199
gi|170288680|TRQ2_0883            STLEHHRKAKEVVEG-----VTILPTWRPDRAMNVDK-EGWKEYVEKMGE 199
gi|15642839|TM0064                STLEHHRKAKEAVEG-----VTILPTWRPDRAMNVDK-EGWREYVEKMGE 199
gi|160902320|Pmob0854             EDLSFHEKAKD-IEG-----IKILPTWRPDKAMNIEK-ENWKDFVEKLGE 201
gi|157364144|Tlet1289             SDLVYHKLAREKIPS-----IRILPTWRPDRTCKIDT-KDWKNYVEKLSM 202
gi|20808989|TTE2645               --------------------------------------------------
gi|159042183|Cmaq1622            EPLPEYDESLFTG----------ALRLDPLLPDIN----PQSINKLSK 171
gi|32474066|RB6095                DALEGFD-----------TETYIPCLRTDDLVFHLAKPEVRGRLAACTG 193
gi|15613056|Bh0493                DDNERISWLEGKQP-----DSRFHAALRLDPLLNEYEQTKHRLRDWGYKV 192


gi|94971332|Acid345_4306          QTNSDTATWDGYLEALRVSRDRFRELGCTATDHG-H-------PTARTAN 252
gi|225874247|ACP_2685             QTGEDTAHWSGYLKALRQARARFRSLGATATDHG-H-------PTAQTAN 252
gi|221639919|RSKD131_1820         LTGEDTATWEGYLAAHRKRRAYFKEFGATSSDHG-H-------ATARTED 257
gi|126462900|Rsph17029_2139       LTGEDTATWEGYLAAHRKRRAYFKEFGATSSDHG-H-------ATARTED 257
gi|146278446|Rsph170252412        LTGEDTATWQGYLAAHRKRRAYFKEFGATSTDHG-H-------ATARTED 257
gi|222149863|Avi_3892             LADVDATSWSGYLEAHRNRRAFFKSYGATSSDHG-H-------PSARTED 255
gi|110680135|RD12934              LTGEDTSTWQGYLNAHAARRAFFKSLGATASDHG-H-------PSARTAD 257
gi|161620871|BCAN_B0804           VAGTP-LTWQGYLDAHRNRREYFKRRGATSSDHG-H-------PTAQTAD 255
gi|23500543|BRA0812               VAGTP-LTWQGYLDAHRNRREYFKRRGATSSDHG-H-------PTAQTAD 255
gi|163844934|BSUIS_B0804          VAGTP-LTWQGYLDAHRNRREYFKRRGATSSDHG-H-------PTAQTAD 255
gi|17988821|BMEll0476|gi|17988    VAGTP-LTWQGYLDAHRNRREYFKRRGATSSDHG-H-------PTAQTAD 255
gi|189022606|BAbS19_ll04000       --------------------------------------------------
gi|158424966|AZC3342              LTGEDTFSWTGYLEAHRKRRAFFKSYGATSTDHG-H-------PSARTAD 259
gi|13473453|mll4056               ITGCDTGSWAGYLDAHRQRRAFFKSFGATSSDHG-H-------PTAETAN 258
gi|16264678|SMb21354              ITGADTGTWSGYLDAHRTRRAYFKDFGATSTDHG-H-------ATADTAN 258
gi|16125737|CC1490                TSGQDVYSWKSYLEAHRLRRQAFIDAGATSSDHG-H-------PTAATAD 259
gi|221234494|CCNA_01557           TSGQDVYSWKSYLEAHRLRRQAFIDAGATSSDHG-H-------PTAATAD 259
gi|167645720|Caul_1756            VSGQDVTSWKGYLEAHRVRRQAFIEAGATSSDHG-H-------PTAATAD 259
gi|197105913|PHZ_c2452            LTGQDVESWSGYLQAHRMRRAAFMEMGATSTDHG-H-------PTAATAD 259
gi|84625827|XOO4170               LTGEDVLTWNGYLRAHRQRRAFFAAHGATSTDHG-H-------PSAATAD 261
gi|161898957|XOO4427              LTGEDVLTWNGYLRAHRQRRAFFAAHGATSTDHG-H-------PSAATAD 261
gi|188574617|PXO_03860            LTGEDVLTWNGYLRAHRQRRAFFAAHGATSTDHG-H-------PSAATAD 254
gi|21244968|XAC4251               LTGEDVLTWDGYLRAHRQRRAFFAAHGATSTDHG-H-------PSAATAD 261
gi|78049913|XCV4357               LTGEDVLSWDGYLRAHRQRRAFFAAHGATSTDHG-H-------PSAATAD 261
gi|66770505|XC4209                LTGEDVMQWPGYLCAHRQRRAFFAAAGATSTDHG-H-------PSAATAD 260
gi|21233539|XCC4117               LTGEDVMQWPGYLCAHRQRRAFFAAAGATSTDHG-H-------PSAATAD 260
gi|188993724|xccb100_4329         LTGEDVMQWPGYLRAHRQRRAFFAAAGATSTDHG-H-------PSAATAD 260
gi|114568782|Mmar100230           ITGEDVSRWDGYLAAHRQRRVDFMAAGATSTDHG-H-------PTARTAD 277
gi|148554803|Swit1887             MTGEDVGSYAGYLAALRARRAAFAAAGATSTDHG-H-------PTAATAD 255
gi|103488522|Sala3046             LSGEDAFSYSGYLAAHRKRRAFFASMGATSTDHG-H-------PSAATAD 259
gi|90020587|Sde0940               ITGEDTSNWEGYLAALRNRREYFKAHGATSTDHG-H-------PTAVTAD 259
gi|90020919|Sde1272               LTGEDVSTFNGYLKALANRREYFKRYGATSTDHG-H-------PSALTAD 260
gi|116249865|RL0099               ITGTEVTRWDGLIEAHRRRRAYFRQFGATATDHGV--------PTAFTAD 260
gi|209551594|Rleg2_4023           ITGADVTRWDGLIEAHRRRRAYFRQFGATATDHGV--------PTAFTAD 260
gi|86355751|RHECH00090            ITGADVTRWDGLIEAHRRRRAYFRQFGATATDHGV--------PTAFTSD 260
gi|190889727|RHECIAT_CH0000094    ITGADVTRWEGLIEAHRRRRAYFRQFGATATDHGV--------PSALTAD 260
gi|119962615|AAur0527             EASDGGSGYRGYLTALENRRRYFVDHGAVSADHGV--------RTPATLK 252
gi|220911562|Achl_0787            AAGDGGTGYGAYITALENRRRYFVDNGAVSADHGV--------RTPRTLK 263
gi|116669068|Arth0502             VGGSGASGYAGYIGALESRRRHFVENGAVSADHGV--------RTPATLK 256
gi|163842140|RSal332093421|gi|    EAGEGAVGYRGYLVALANRRRYFIEHGAVSADHGV--------RIPLTVK 115
gi|158315003|Franean1_3197        AADADVGDYDGYVRALEARRRYFAEHGATSCDHS-H-------PDVRTDP 280
gi|134101579|SACE5055             ASGIDTGDYDGLIRALENRREHFRAHGATATDHSA--------PDARMAF 287
gi|152964556|Krad_0587            VSGEDTGDYAGFLAALRQRRAHFAAMGARATDHG-H-------LTADSAP 259
gi|50955310|Lxx17350              ASGTDTGSYAGLLAALRARRAAFAAAGGTATDTGVI--------DAGSEP 312
gi|213691567|Blon_0669            SAGVNTATYEGFVEAMRRRRLFFKQHGAVAADYGVDPSLDEVNWSGDTTR 265
gi|50843764|PPA2329               VSGCDATTLDGFTEAMEDRRAYFRQHGAVSSDH-SH-------RDLGTII 260
gi|161485962|CE2377               VAGDGKAGWEGYLQAMRNRRQYFIDHGATSADHGTH--------DTDTTP 258
gi|145296483|cgR_2394             TAGDGKAGWEGYLQAMRNRRQYFINHGATSADHGLH--------DTDTTP 261
gi|62391323|cg2731                TAGDGKAGWEGYLQAMRNRRQYFINHGATSADHGLH--------DTDTTP 261
gi|161486711|NCgI2398             TAGDGKAGWEGYLQAMRNRRQYFINHGATSADHGLH--------DTDTTP 261
gi|126208494|APL_1020             --------------------------------------------------
gi|190150346|APP7_1077            --------------------------------------------------
gi|165976445|APJL_1038            --------------------------------------------------
gi|152977831|Asuc_0145            VSDTDINSFDALKKALLKRLEHFDQHGCKSADHGMEI------VRFAPIP 250
```

```
gi|161511000|MS0544              VSDTEINGFDSLKKALSKRLDHFDKRGCKSADQGMEI------VRFAPVP 250
gi|148827203|CGSHiGG_02740       IADVDIYTFADLKKALLKRLEYFDAQGCKSADHGMEI------VRFSAIP 250
gi|68248599|NTHI0056             IADVDIYTFADLKKALLKRLEYFDAQGCKSADHGMEI------VRFSAIP 250
gi|113461738|HS_1602             VADIEINTFDKLKQALSKRIEHFDKYGCKSADHGMEI------VRFSPIP 246
gi|170718551|HSM_0409            VADIEINTFDKLKQALSKRIEHFDKYGCKSADHGMEI------VRFSPIP 250
gi|62181648|SC3078               AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|197249740|SeAg_B3309          AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|198243637|SeD_A3482           AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|207858396|SEN2980             AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|194734239|SeSA_A3319          AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|205354057|SG3031              AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|161616115|SPAB03917           AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|16766437|STM3137              AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|194448866|SeHA_C3383          AADVSITRFDDLRQALTRRLDHFAACGCLASDHGIET------LRFAPVP 251
gi|194446286|SNSL254_A3389       AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|16761915|STY3308              AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|56415079|SPA3005              AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|197364008|SSPA2803            AADVSITRFDDLRQALTRRLDHFAACGCRASDHGIET------LRFAPVP 251
gi|161506293|Sari04489           AADVNITRFDDLRRALTRRLDHFAACGCRAADHGIET------LRFAPVP 251
gi|37524198|plu0176              VADVSIVAFADVLNALERRLEHFAAHGCRAADHGIEN------LRYAPIP 251
gi|28901561|VPA1706              VADIDIRRFDDLLSALDKRLAHFDSHGCRAADHGIEI------VRYAPIP 250
gi|37677254|VV1594               VADIEIRRFDDLLSALDKRLAHFDAHGCRAADHGIEI------VRYAPIP 250
gi|27367460|VV21070              VADIEIRRFDDLLSALDKRLAHFDAHGCRAADHGIEI------VRYAPIP 250
gi|119943921|Ping0131            VTDIEIGRFSDLLSALEQRIEHFSAHGCRAADHGIEV------VRYATVP 255
gi|152996793|Mmwyl12780          AADINITRFSDLIAALAQRLEVFDAHGCRSADHGIEI------MRFANEP 250
gi|117625400|APECO1_3327         VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|91212521|UTI89_C3530          VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 270
gi|16130987|b3092                VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|161486116|c3850               VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|215488422|E2348C_3385         VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|170082634|ECDH10B_3268        VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|157155304|EcE24377A_3560      VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|218691395|ECED1_3759          VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|157162569|EcHS_A3275          VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|218701863|ECIAI39_3589        VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|170018657|EcolC_0608          VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|110643336|ECP_3183            VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|218560176|ECS88_3488          VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|170683343|EcSMS35_3384        VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|89109861|                     VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|30064435|S3339                VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|82545347|SBO2953              VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|56480255|SF3132               VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|110806974|SFV3133             VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|74313639|SSON3245             VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|187731951|SbBS512_E3528       VSDTDIRRFADLQTGLTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|209397829|ECH74115_4407       VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|15803633|Z4445                VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|15833228|ECs3974              VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|218696796|EC55989_3506        VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|209920564|ECSE_3373           VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|218706716|ECUMN_3576          VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|218551598|EFER_4383           VSDTDIRRFADLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|206579202|KPK_0591            VSDTDIRRFADLQSALTKRLDHFAAHGCKVSDHALDV------VLFAE-A 249
gi|152972036|KPN_03520           VSDTDIRRFADLQSALTKRLDHFAAHGCKVSDHALDV------VLFAE-A 245
gi|146313181|Ent638_3546         VSDTDIRRFTDLQSALTKRLDHFAAHGCKVSDHALDV------VLFAE-S 249
gi|157148666|CKO_04494           ASDTDIRRFTDLQAALTQRLDHFAAHGCKVSDHALDV------VMFAE-A 270
gi|156935638|ESA_03503           VSDTDIRRFSDLQTALTKRLDHFAAHGCKVSDHALDV------VLFAE-A 249
gi|161484751|y3600               AADTSISRFADLCVALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
gi|161511340|YP2899              AADTSISRFADLCVALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
gi|108809201|YPA3210             AADTSISRFADLCVALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
gi|162419999|YPAngolaA1104       AADTSISRFADLCVALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
gi|145597678|YPDSF0366           AADTSISRFADLCVALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
gi|170022805|YPK_0554            AADTSISRFADLCVALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
gi|108810613|YPN0448             AADTSISRFADLCVALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
gi|16120908|YPO0579              AADTSISRFADLCVALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
gi|153949672|YpsIP317580491      AADTSISRFADLCVALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
```

```
gi|51597770|YPTB3478              AADTSISRFADLCVALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
gi|186896961|YPTS_3663            AADTSISRFADLCVALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
gi|123443892|YE3708               AADTSISRFADLCTALNKRMDHFAAHGCKVSDHALDV------VVYGE-A 249
gi|157372554|Spro4321             AADINIGRFSALCDALKKRMDHFAAHGCKVADHALDV------VVYGE-A 249
gi|108810612|YPN0447              --------------------------------------------------
gi|50119594|ECA0645               VSDTDIRRFSDLQTALTKRLDHFAAHGCKVSDHALDV------VMFAE-A 249
gi|54303188|PBPRB1515             AADVYIIKYNQLCDALINRLDHFQSHGCCISDHALDT------VEYEE-S 185
gi|60681862|BF1904G               ISEITISDYNDMILALRKRHDYFAEQGCKLSDHGIEE------FYAED-Y 249
gi|53713582|BF2293                ISEITISDYNDMILALRKRHDYFAEQGCKLSDHGIEE------FYAED-Y 249
gi|161511193|BT0823               VSGVTISNFDDMIAALRKRHDFFAEQGCRLSDHGIEE------FYAED-Y 249
gi|150005586|BVU3074              VSGVTISSFDDMVAALRKRHDFFAEQGCKLSDHGIEE------FYAED-Y 249
gi|212550592|CFPG_235             VSGIAINCFEHLLEALKVRHDFFDSVGCKLSDHGLEM------FYAED-Y 249
gi|150009769|BDI3184              AADKTILTYKHLLEALQKRHDFFAAQGCRLSDHGLDT------FYAEP-Y 313
gi|146301998|Fjoh_4262            VSGVAINTFADLQAALRNRIEFFNANGCKLSDHGLDQ------IYFED-F 249
gi|192359962|CJA_2772             TADININTFDDLIKALQIRHDYFHENGCRLSDHGLET------VYAAD-Y 249
gi|182416100|Oter_4293            AAKTKIASFDDFLSALKKRHDDFHAIGGRLSDHGMEN------CYAEP-C 246
gi|116624408|acid5330             AADTHISTMAEFLDAIKKRHDFFHAMGGRLSDHGINH------AFSDF-P 247
gi|225621141|BHWA1_02239          ASKINIKDINSLTEALYNRIDYFKSLGCVSSDCSLSI------VPFNL-D 250
gi|18309134|CPE0152               VSNVEIKSFDDLKKALEIRLDYFYENGCMITDHSLER------VVFYK-F 248
gi|160880875|Cphy_2743            VSGVKIASFAGLMEALKVRLEFFNSMGCKVSDHALSY------VMYKP-A 248
gi|189485070|TGRD_067             VSGVNITSFKTMKEAFRKRLGFFISSGCRTTDHAPEY------IMYVP-S 248
gi|56962769|ABC0995               AAEVDISSLRSLLEALEKRATFFAEEGCVLSDHGLRT------LPFVD-T 252
gi|222529391|Athe_1402            VYGKKIESYDEFLDALKSRAEFFHSVGCRASDHAID-------DMVFADA 246
gi|146296954|Csac1949             VYGKKIESYDEFLDALKSRAEFFHSVGCRASDHAID-------NMVFAEA 246
gi|150018894|Cbei_4082            VSEVSINNYDEFLKALDSRIRFFHSVGCRIADHGIDG------VVVYADS 249
gi|150019467|Cbei_4657            ISEITISNYDEFLNAFDSRVRYFHSVGCRIADHGIDG------IVVYADA 252
gi|20808343|TTE1939               ASGIKITTYDDFLQALEKRIEFFHSVGCRISDHALD-------YVFYQKT 248
gi|15893980|CAC0692               VSKKNINSYDMFLEALNDRIEFFHSVGGRVSDHALD-------YVPYLEA 247
gi|150016705|Cbei_1832            VVGSKISNFDEYLEALKKRVEYFHEAGCRVSDNALD-------FVPVGNA 248
gi|52081820|BL00708               ASGMKIANYDDFLKALKNRIDFFHEAGGRISDHAIN-------QMMYTET 247
gi|52787207|BLi03516              ASGMKIANYDDFLKALKNRIDFFHEAGGRISDHAIN-------QMMYTET 247
gi|157693739|BPUM2987             ASAISIQSYQDFLNALRARIDFFHEHGCLISDHAIN-------EMTYEET 247
gi|154685671|RBAM012380           ASGMAVSSYQSYLDALESRVRFFHNAGGRVSDHALD-------QMVYAET 247
gi|16078295|BSU12300              AAAISITTYDEFLKALEKRVRFFHSAGGRVSDHAID-------TMVFAET 248
gi|23097822|OB0367                VTSHSTKTYDGFLRALKSRVDYFHQNGCSVADHALN-------TMMFTET 248
gi|15613268|Bh0705                VTGESLDTLDAFLHALKERLTFFDEHGCRSSDHDMT-------EVPFVEV 248
gi|138895421|GTNG_1765            VADLSVDDYGQLLQALENRVRYFHEAGCRMADHGLE-------SMPYAEC 248
gi|220929098|Ccel_1676            VTEMNISSFEDLILALQKRAEFFKSIGCMISDHSFGY------PDFTK-G 250
gi|81427739|LSA0129               VAGVEIHDFATIVTALTQRFEFFNKMGGRLSDHSLLT------YHFEE-A 252
gi|116332824|LVIS_0150            ISGVQISSFKDIIKALHQRFTFFNEMGGRLSDHSLLT------YHFVE-A 252
gi|116332809|LVIS_0135            VSGVTITDFDSLVAALAQRFDFFNSLGGHLSDHGLNT------YHFRK-A 252
gi|116512444|LAACR_1742           VSGIKIKDFKTMIKALEQRFEFFTSLGGRLSDHSLST------YTFAE-T 246
gi|125623707|llmg_0862            VSGIKIKDFKTMIKALEQRFEFFTSLGGRLSDHSLST------YTFAE-T 246
gi|15673610|L0019                 VSGIEIKGFKTMIKALEQRFEFFTSLGGRLSDHSLST------YTFAE-T 246
gi|150389106|Amet1291             LTSLPISSYQEMLVGIEARIDYFDKKGALISDHGLEK------LVYIE-S 253
gi|70727647|SH2648                LT-HPITTPSDFIEALYKRIQYFHDKGGRLADHGLEE------MHFEA-Y 246
gi|25010729|gbs0674               ITQKEITDFSTFIAAMEERIAYFAQNGCRASDISFTE------IVFEQ-T 245
gi|22536865|SAG0701               ITQKEITDFSTFIAAMEERIAYFAQNGCRASDISFTE------IVFEQ-T 245
gi|76786751|SAK0827               ITQKEITDFSTFIAAMEERIAYFAQNGCRASDISFTE------IVFEQ-T 245
gi|195977823|Sez_0689             ATGRTITDFKSFIAAMEERIAYFAENGCKASDISFTE------IVFEA-A 245
gi|225868878|SZO_13120            ATGRTITDFKSFIAAMEERIAYFAENGCKASDISFTE------IVFEA-A 245
gi|225870118|SEQ_0716             ATGRTITDFKSFIAAMEERIAYFAENGCKASDISFTE------IVFEA-A 245
gi|94990680|MGAS10270_Spy1171     STGNRITDFTTFMAAMEKRIAYFAEKGCKASDISFTE------IVFEK-A 254
gi|222153335|SUB1203              ATGKEMTEFKDFISAMEDRIAYFAENGCKASDISFTE------IVFEA-A 245
gi|146318813|SSU051159|gi|1463    VTGVAVRDFASFVEALGQRVSYFAQHGCRASDISITA------ITYEE-A 257
gi|146321021|SSU981174            VTGVAVRDFASFVEALGQRVSYFAQHGCRASDISFTA------ISYEE-A 257
gi|225859885|SP70585_2238         VTEVAVRDFASFVEALGQRVSYFAQHGCRASDISFTA------ITYEE-A 245
gi|157147928|CKO_03735            KTGKTITTLNEFLAALEARVDFFHSTGCRVSDHGPLA------VRFRP-L 247
gi|220927672|Ccel_0214            TTKLKIITYQDMLNVLKNRIEFFRQAGCLLSDHSLES------LVYLP-T 245
gi|78221672|Gmet0451              CSGIAIESLQTFMEALRVRYEFFHKLGCRVSDHGMDR------CYATP-C 247
gi|148269995|Tpet0860             RYGEDTSTLEGFLSALWKSHEHFKEHGCVASDHALLE------PSIYY-- 241
gi|170288680|TRQ2_0883            RYGEDTSTLEGFLSALWKSHEHFKEHGCVASDHALLE------PSVYY-- 241
gi|15642839|TM0064                RYGEDTSTLDGFLNALWKSHEHFKEHGCVASDHALLE------PSVYY-- 241
gi|160902320|Pmob0854             ITKENVERFEGFLNALYKTHKKFEKLGGVSSDHGILE------PISYP-- 243
gi|157364144|Tlet1289             STNTDISNLDDFLNALEKTHGYFNDLGCVCSDHALLN------PFLEP-- 244
gi|20808989|TTE2645               --------------------------------------------------
gi|159042183|Cmaq1622            TYGISISNPNDIDEALARLFKRFKGHVVAVTLNLQPD------ESFITLK 215
```

```
gi|32474066|RB6095                IELDG--SLESLRSSLKQRFEHFVSRGARACAISIPP------TFTPTPV 235
gi|15613056|Bh0493                NDEWNEGSIQEVKRFLTDWIERMDPV---YMAVSLPP------------- 226


gi|94971332|Acid345_4306          LSRAEAESLFPRVISGM-ANPDEKELFRGQMLTEMARMSLEDGLVMQIHA 301
gi|225874247|ACP_2685             LSEPEAAQLFITVLSGN-ATAAQRELFRAQMLTEMARMSVEDGLVMQIHP 301
gi|221639919|RSKD131_1820         LPQAEAAALFARALTGA-CTAEEADAFRGQMLTEMARMSLEDGLVLQIHP 306
gi|126462900|Rsph17029_2139       LPQAEAAALFARALTGA-CTAEEADAFRGQMLTEMARMSLEDGLVLQIHP 306
gi|146278446|Rsph170252412        LPQAEAAALLAKALKGT-CTPDEADAFRGHMLTEMARMSLDDGLVLQIHP 306
gi|222149863|Avi_3892             LPQDQAKALFDKALTGR-ITPDEADAFRGHMLTEMARMSLEDGLVLQIHP 304
gi|110680135|RD12934              LPQAEAAALFQKALSGQ-CSPEEQDVFRGHMLTEMARMSLEDGLVLQIHP 306
gi|161620871|BCAN_B0804           LSAGDASRLFDRVIKGN-ASTSDAEMFRAQMLTEMARMSIDDGLVMQIHP 304
gi|23500543|BRA0812               LSAGDASRLFDRVIKGN-ASTSDAEMFRAQMLTEMARMSIDDGLVMQIHP 304
gi|163844934|BSUIS_B0804          LSAGDASRLFDRVIKGN-ASTSDAEMFRAQMLTEMARMSIDDGLVMQIHP 304
gi|17988821|BMEll0476|gi|17988    LSAGDASRLFDRVIKGN-ASTSDAEMFRAQMLTEMARMSIDDGLVMQIHP 304
gi|189022606|BAbS19_ll04000       --------------------------------------------------
gi|158424966|AZC3342              LSLGEAEALFERVAGGQEVSAQDAELFRAQMLTEMARMSVDDGLVMQIHP 309
gi|13473453|mll4056               LSDAAAQELFNRIRRGS-QDERERKLFRAQMLTEMAKMSRDDGLVLQIHP 307
gi|16264678|SMb21354              LPQAEAAALFDKVRLGK-ANADERRLFRAQMLTEMAKMSLDDGLVMQIHP 307
gi|16125737|CC1490                LSDVEAEALFNSLVKGD-VTPEKAELFRAQMLTEMAKMSLDDGLVMQIHP 308
gi|221234494|CCNA_01557           LSDVEAEALFNSLVKGD-VTPEKAELFRAQMLTEMAKMSLDDGLVMQIHP 308
gi|167645720|Caul_1756            LSEVEAETLFLSLLKGE-VTPAKAELFRAQMLTEMAQMSLDDGLVMQIHP 308
gi|197105913|PHZ_c2452            LSLADAEALFRKLLSAD-FSPADAELFRAQMLTEMARMSLDDGLVMQIHP 308
gi|84625827|XOO4170               LSPAQAQRLFDTVVRGA-ATPEQAELFRAQVLTEMAAMSLDDGLVMQLHP 310
gi|161898957|XOO4427              LSPAQAQRLFDTVVRGA-ATPEQAELFRAQVLTEMAAMSLDDGLVMQLHP 310
gi|188574617|PXO_03860            LSPAQAQRLFDTVVRGA-ATPEQAELFRAQVLTEMAAMSLDDGLVMQLHP 303
gi|21244968|XAC4251               LSPAEAQRLFDTVVRGA-ATPEQAELFRAQVLTEMAAMSLDDGLVMQLHP 310
gi|78049913|XCV4357               LSPVQAQRLFDTVVRGE-ATPEQAELFRAQVLTEMAAMSLDDGLVMQLHP 310
gi|66770505|XC4209                LSPAEAQRLFDTVVRGQ-ATPAQAELFRAQVLTEMAAMSVDDGLVMQLHP 309
gi|21233539|XCC4117               LSPAEAQRLFDTVVRGQ-ATPAQAELFRAQVLTEMAAMSVDDGLVMQLHP 309
gi|188993724|xccb100_4329         LSPAEAQRLFDTVVRGQ-ATPAQAELFRAQVLTEMAAMSVDDGLVMQLHP 309
gi|114568782|Mmar100230           LDRGEAEALFDRIVSQK-QRDGDAELFRAQMLTEMAAMSLEDGLVMQIHP 326
gi|148554803|Swit1887             LDRGEAERLYAQVRGGG-ATAEQAELFRAHMLTVMAEMSIDDGLVMQIHP 304
gi|103488522|Sala3046             LSETQAEALFARVTGED-MSAADAELFRAHMLTVMAGMSLDDGLVMQIHP 308
gi|90020587|Sde0940               LSRPEIETLFAKALAGK-CSAAEAELFRGQILTEMAGMSLEDGLVMQIHP 308
gi|90020919|Sde1272               LSQSEKENLYLLARSGE-ISAQEAELFRGQMLTEMAAMSLDDGLVIQIHP 309
gi|116249865|RL0099               LPLTEKQALLDKALKGP-LSAEDAELFRGQMMTEMAGLSAEDGMVMQIHA 309
gi|209551594|Rleg2_4023           LPLAEKQALLDKALKGP-LSREEAELFRGQMMTEMAGLSAEDGMVMQIHA 309
gi|86355751|RHECH00090            LPLAEKQALLDKALKGP-LSAVDAELFRGQMMTEMAGLSAEDGMVMQIHA 309
gi|190889727|RHECIAT_CH0000094    LPLAEKQALLDRALNGP-LSAAEAELFRGQMMTEMAGLSAEDGMVMQIHA 309
gi|119962615|AAur0527             LEEADAEALFEKARSGQ-ATAHDRDVFEAHMMYQMARMSVEDGLVMTIHP 301
gi|220911562|Achl_0787            LSDSEAARLFDTARSGK-ATAQDREDFEAHMMYQMARMSVEDGLVMTIHP 312
gi|116669068|Arth0502             LDAADAERLFERARAGK-ASPEDREAFEAHMMYQMARMSVEDGLVMTIHP 305
gi|163842140|RSal332093421|gi|    LPFAEAERLFDSALRGE-LSALERDVFEGQMIYQMARMSVEDGLVMTLHP 164
gi|158315003|Franean1_3197        LPHAEAARIFRAALAGE-ASEAEATAFRRHMICESARMSCEDGLVMTLHP 329
gi|134101579|SACE5055             LDAVQAQRLFDAVRSGT-ASADDARLLSQQLLGEMARMSSEDGMVMALHT 336
gi|152964556|Krad_0587            LERADAERVYAAARAGA-VGPADASAFAAHMLFEMARMSRDDGLVMQIHP 308
gi|50955310|Lxx17350              LTKTAAERIHRSALAGA-LDPADAVAYRRNLLYRLAEMSAEDGLVMQLHP 361
gi|213691567|Blon_0669            LSDDVAIRLYTKARSGS-LTSDEARKLHAHLLNDQAKLAQDDGLVMTLHP 314
gi|50843764|PPA2329               LDHDRAASIFDASVAGR-ATVEEMALLRRHLFTDQARMASEDGLTMTVHP 309
gi|161485962|CE2377               LSHEDAQRILDKGLAGT-ATLAEMRAFEANTTYRFAEMCQDDGLVMTLHP 307
gi|145296483|cgR_2394             LSHEDAQKILDKGLAGT-ATLAEMRAFEANTTYRFAEMSQEDGLVMTIHP 310
gi|62391323|cg2731                LSHKDAQKILDKGLAGT-ATLAEMHAFEANTTYRFAEMSQEDGLVMTIHP 310
gi|161486711|NCgI2398             LSHKDAQKILDKGLAGT-ATLAEMHAFEANTTYRFAEMSQEDGLVMTIHP 310
gi|126208494|APL_1020             ----------------------MFCLITTAVLVWLASEYCKRNWVMQMHI 28
gi|190150346|APP7_1077            ----------------------MFCLITTAVLVWLASEYCKRNWVMQMHI 28
gi|165976445|APJL_1038            ----------------------MFCLITTAVLVWLASEYCKRNWVMQMHI 28
gi|152977831|Asuc_0145            DEKELDRILRLRRS-NQALTELQINQYSTALLVWLGTEYCKRNWAMQMHI 299
gi|161511000|MS0544               DEKELDRILQLRRN-EQPLTELQISQFSTALLVWLGAEYCKRNWVMQMHI 299
gi|148827203|CGSHiGG_02740        DESVLNSILQKRLQ-NQPLLEEEVAQFSTAILVWLASEYCKRHWVMQMHI 299
gi|68248599|NTHI0056              DESVLNSILQKRLQ-NQPLLEEEVAQFSTAILVWLASEYCKRHWVMQMHI 299
gi|113461738|HS_1602              DEKTLDQILQKRLN-NQPIEEEEIAQFSTALLVWLGTEYHKHHWVMQLHI 295
gi|170718551|HSM_0409             DEKTLDQILQKRLN-NQPIEEEEIAQFSTALLVWLGTEYHKHHWVMQLHI 299
gi|62181648|SC3078                DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|197249740|SeAg_B3309           DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|198243637|SeD_A3482            DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|207858396|SEN2980              DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
```

```
gi|194734239|SeSA_A3319          DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|205354057|SG3031              DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|161616115|SPAB03917           DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|16766437|STM3137              DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|194448866|SeHA_C3383          DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|194446286|SNSL254_A3389       DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|16761915|STY3308              DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|56415079|SPA3005              DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|197364008|SSPA2803            DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|161506293|Sari04489           DDAQLDAILGKRLA-GETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI 300
gi|37524198|plu0176              DETVLNGILQKRLA-GGILSELEIAQYTTAILVWLGRQYAARGWVMQMHI 300
gi|28901561|VPA1706              SEADLDALLARRLS-GEVLSELECAQFSTAVQVWLGKRYAQLGWVMQLHI 299
gi|37677254|VV1594               SEADLDALLARRLS-GEVLSELECAQFSTAVQVWLGKRYAQLGWVMQLHI 299
gi|27367460|VV21070              SEADLDVLLARRLN-GEVLSELECAQFSTAVQVWLGKRYAQLGWVMQLHI 299
gi|119943921|Ping0131            VESTLDSILTRRLS-QQTLTEQEIDQFTTAVQIWLGTQYAKRGWVMQLHI 304
gi|152996793|Mmwyl12780          SETDLDAILTKRIN-EQALTELEIAQFSSAVQAWLGKQYAKKGWVMQLHI 299
gi|117625400|APECO1_3327         NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|91212521|UTI89_C3530          NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 319
gi|16130987|b3092                NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|161486116|c3850               NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|215488422|E2348C_3385         NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|170082634|ECDH10B_3268        NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|157155304|EcE24377A_3560      NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|218691395|ECED1_3759          NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|157162569|EcHS_A3275          NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|218701863|ECIAI39_3589        NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|170018657|EcolC_0608          NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|110643336|ECP_3183            NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|218560176|ECS88_3488          NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|170683343|EcSMS35_3384        NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|89109861|                     NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|30064435|S3339                NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|82545347|SBO2953              NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|56480255|SF3132               NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|110806974|SFV3133             NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|74313639|SSON3245             NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|187731951|SbBS512_E3528       NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|209397829|ECH74115_4407       NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|15803633|Z4445                NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|15833228|ECs3974              NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|218696796|EC55989_3506        NEAELDSILARRLA-GEPLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|209920564|ECSE_3373           NEAELDSILARRLA-GEPLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|218706716|ECUMN_3576          NEAELDRILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|218551598|EFER_4383           NEAELDSILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|206579202|KPK_0591            TDAELDAILARRLA-GETLSAHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|152972036|KPN_03520           TDAELDDILARRLA-GETLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 294
gi|146313181|Ent638_3546         NEAELDSILARRLS-GESLSEHEVAQFKTAVLVFLGAEYARRGWVQQYHI 298
gi|157148666|CKO_04494           SDAELDSILARRLA-GETLSEHEVAQFKTAVLVWLGAEYARRGWVQQYHI 319
gi|156935638|ESA_03503           SESELNAILSRRLA-GETLSGHEIAQFKTAVLVWLGAEYARRGWVQQYHI 298
gi|161484751|y3600               DETTLDAILARRLA-GNQPSTEEIAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|161511340|YP2899              DETTLDAILARRLA-GNQPSTEEIAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|108809201|YPA3210             DETTLDAILARRLA-GNQPSTEEIAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|162419999|YPAngolaA1104       DETTLDAILARRLA-GNQPSTEEIAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|145597678|YPDSF0366           DETTLDAILARRLA-GNQPSTEEIAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|170022805|YPK_0554            DETTLDAILARRLA-GNQPSTEEIAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|108810613|YPN0448             DETTLDAILARRLA-GNQPSTEEIAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|16120908|YPO0579              DETTLDAILARRLA-GNQPSTEEIAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|153949672|YpsIP317580491      DETTLDAILARRLA-GNQPSTEEIAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|51597770|YPTB3478             DETTLDAILARRLA-GNQPSTEEIAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|186896961|YPTS_3663           DETTLDAILARRLA-GNQPSTEEIAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|123443892|YE3708              DEATLDGILARRLA-GNQPSTEETAQFKTAVLLFLSGEYHRREWVQQYHI 298
gi|157372554|Spro4321            DESTLDAILTARLS-GKLPTPEQSAQFKSAVLLFLAAEYQRREWVQQYHI 298
gi|108810612|YPN0447             --------------------------------------------------
gi|50119594|ECA0645              DDATLDKILARRLA-GETLSEHEVAQFKTGVLVWLGAEYARRGWVQQYHI 298
gi|54303188|PBPRB1515            TEGELDAILALRLQ-GKTVSMKQAAQFKTAVLILLGKEYARRGWVQQYHI 234
gi|60681862|BF1904G              TEGEIKTIFNKIYG-GSELTKEEVLKFKSAMLIVLGEMDWEKGWTQQFHY 298
gi|53713582|BF2293               TEGEIKTIFNKIYG-GSELTKEEVLKFKSAMLIVLGEMDWEKGWTQQFHY 298
```

```
gi|161511193|BT0823                TDAEIKAIFNKVYG-GAELTKEEILKFKSAMLVIFGEMDWEKGWTQQFHY 298
gi|150005586|BVU3074               TDAEINAIFNKVYG-GTELTKEEILKFKSAMLIVFGEMDWEKGWTQQFHY 298
gi|212550592|CFPG_235              TLSSINTIFNKVYDKGNKLNNDEIVRFKSCMLIELAKMDWEKNWAQQFHY 299
gi|150009769|BDI3184               TEQEIEVIFLKARM-GKSLTEQEVRKYRSALLYELAAMDARSGWVQQFHI 362
gi|146301998|Fjoh_4262             TESEISSIFKKKRE-NRIITPEEALKFQSAVLIFLSETYHEFGWVQQFHL 298
gi|192359962|CJA_2772              TDSEIKAIFLKIRA-HKQLDAVEIEKFQSAMMVEFALQDHAKGWVQQFHI 298
gi|182416100|Oter_4293             TATEAQAIFDAARA-GRAASVADQAKFASFMMLEFGRWDAKKGWTKQLHL 295
gi|116624408|acid5330              SEHEAARIFGRARC-GHAATSEEHGKFAAYMMLVFGRLDAEKGWTKQLHL 296
gi|225621141|BHWA1_02239           DEKNIDAIFKKAMN-KESLNFEDIEKYKTYILIKLIKKYKESNLVMQIHI 299
gi|18309134|CPE0152                SCEEINEIFIKALN-GENINNEEASKYSVALLISLGKMYAERNMVMQLHI 297
gi|160880875|Cphy_2743             TEEEIEAIFAKRLA-GGTISEMEELQFKTAFMVSVGREYNRLGWVMQLHY 297
gi|189485070|TGRD_067              SDKEVETIFDKRLS-GVALSKEEELKYKTAFMLFAGGEYSKNDLVMELHY 297
gi|56962769|ABC0995                TEKEAEAAFQKALN-QETLTPQEEEKYQTYVLLFLARLYNKLGWTMQFHL 301
gi|222529391|Athe_1402             SFDEVANIFKKALA-GEKLTEIEFAKYKTYTLRFLGKVYSSLGWAMQLHI 295
gi|146296954|Csac1949              SFDEVANIFKKALA-GEKLTEIEVAKYKTYTLRFLGKVYSSLGWAMQLHI 295
gi|150018894|Cbei_4082             SKEEAEAIFAKALE-GKSISIDEEKKYKTYTLRHVFKLYHELGWTMQLHI 298
gi|150019467|Cbei_4657             SKEDAAAIFARALN-GEVISLDEEKKYKTYTLTHVFKLYHELGWAMQLHI 301
gi|20808343|TTE1939                SKQEAEKVFKKVLT-EQHLTKEEIDSFKTYTMIFLGKKYAELNWVMQLHI 297
gi|15893980|CAC0692                SKEEVNTIFAKALK-GEKVSFEEETKFRTFTMKFLGKKYASLGWAMELHM 296
gi|150016705|Cbei_1832             SLEEVREIFANALK-NGQVSFEDENKLRVYILKFFGKLYHDLGWVMQFHM 297
gi|52081820|BL00708                DESEVRPIFAKVMN-GESASPEEECKFKSLTLHFLGTCYAEKGWAMQLHI 296
gi|52787207|BLi03516               DESEVRPIFAKVMN-GESASPEEECKFKSLTLHFLGTCYAEKGWAMQLHI 296
gi|157693739|BPUM2987              TQEEVETIFHKRMS-GYPLTEKEKIKFKTETFIMLGQAYCERGWAMQLHI 296
gi|154685671|RBAM012380            TEEEAARIFAAGLS-GEHVSFEDEKKFKTRTLQYLCGLYAELDWAMQFHI 296
gi|16078295|BSU12300               TKEEAGRIFSDRLQ-GTEVSCEDEKKFKTYTLQFLCGLYAELDWAMQFHI 297
gi|23097822|OB0367                 TKEKARNYYEKAMN-GQKLSPKEESDFKSFTLVFLGEQYADKGWVMQYHI 297
gi|15613268|Bh0705                 NEQEAQHIFRKRLA-NEGLTKVENEKYKTFLMTWLGKEYAARGWVMQWHI 297
gi|138895421|GTNG_1765             TLDEARVIFQKRKE-GFALSREEEEKYQTFTLCFLARLYYSLGWVMQLHI 297
gi|220929098|Ccel_1676             KLHEARSAFLKAMA-GEELEQGEIEAYKTQLMLALGRTYHRLGFAMQLHF 299
gi|81427739|LSA0129                TPAELDAIVAKG-INNAALSEHEINQYLTMLLEALMKLNTQFNWTMQFHI 301
gi|116332824|LVIS_0150             TDSELDAIVNKG-VNNEDLTPKELDQYITMLLENLMKLNKEFDWTMQFHV 301
gi|116332809|LVIS_0135             SKPELNQIIDKARQDNQNLSAADIDGYQTALLEALMRLNKQFGWTMQFHV 302
gi|116512444|LAACR_1742            DDVDLDSIIQRA-KEKQTLTDLEYNQYITALILEIMKLNKKFDWTMQLHV 295
gi|125623707|llmg_0862             KNVDLDSILQKA-KENQTLTDLEYNQYITALILEIMKLNKKFDWTMQLHV 295
gi|15673610|L0019                  KNVDLDSILQKA-KENQTLTDLEYNQYITALILEIMKLNKKFEWTMQLHV 295
gi|150389106|Amet1291              TDEEIENIFQKALG-KKEMTEEDYGKFITRLLIDLGKIYYHKDWIMQIHF 302
gi|70727647|SH2648                 TDQAIQDIFKKALN-HADISTYERFQFQSYMLNELSKAYYERGWVMQIHF 295
gi|25010729|gbs0674                DELELNDLFNKVCE-GYIPNQSEISKWQTAVFMELCRLYKKYGFVTQVHF 294
gi|22536865|SAG0701                DELELNDLFNKVCE-GYIPNQSEISKWQTAVFMELCRLYKKYGFVTQVHF 294
gi|76786751|SAK0827                DELELNDLFNKVCE-GYIPNQSEISKWQTAVFMELCRLYKKYGFVTQVHF 294
gi|195977823|Sez_0689              EPEQLDRLMTRVLE-GYQPQPLEIKQWQTAVFAELCRLYKHYGFVTQVHF 294
gi|225868878|SZO_13120             EPEQLDHLMTRVLE-GYQPQPLEIKQWQTAVFAELCRVYKHYGFVTQVHF 294
gi|225870118|SEQ_0716              EPEQLDRLMTRVLE-GYQPQPLEVKQWQTAVFAELCRLYKHYGFVTQVHF 294
gi|94990680|MGAS10270_Spy1171      TPEALNQIMAQVLA-GDTPSQLDINQWQTAVFADLCRLYKQYGFVTQVHF 303
gi|222153335|SUB1203               EESELNELLAKVKD-GYKPNALEVKQWQTAVFAELCQLYKKYGFVTQVHF 294
gi|146318813|SSU051159|gi|1463     TLEEM-----------------------------CRLYKKHGFVTQVHF 277
gi|146321021|SSU981174             TLEELDEILLARLA-GKEAGQSSINKWQTAIFRELCRLYKKHGFVTQVHF 306
gi|225859885|SP70585_2238          TLEELNDILLARMV-GKEVGQSGINKWQTAIFRELCRMYKKHGFVTQVHF 294
gi|157147928|CKO_03735             DAAGQAALFQRRLA-GDVLSEDELQAWDSLIFVALAKMYKQRDWAMQIHF 296
gi|220927672|Ccel_0214             DIVEAGSIFLKRLN-GEVITTHEAEKYKLYTLSFLARAYKEAGWAMQLHI 294
gi|78221672|Gmet0451               SDAEATTIFSRLLLSSSDIDADDLEKWRSWFMRQIAEWNHESGWTMMLHL 297
gi|148269995|Tpet0860              VDENRARAIHEKAFSGEKLTQDEINDYKAFMMMQFGRMNQETNWVTQLHI 291
gi|170288680|TRQ2_0883             VDENRARAVHEKAFSGEKLTQDEINDYKAFMMIQFGKMNQETNWVTQLHI 291
gi|15642839|TM0064                 VDENRARAVHEKAFSGEKLTQDEINDYKAFMMVQFGKMNQETNWVTQLHI 291
gi|160902320|Pmob0854              VERERVKEVYEKALSKKELTKQEISDFKSFMFYEFGTMNAESNWVMQLHI 293
gi|157364144|Tlet1289              VDKHKASQIFDNSMK-QSASHEDLMIFQSYMMYKFAEMNTKENWTMQLHI 293
gi|20808989|TTE2645                ---------------------------------------------------
gi|159042183|Cmaq1622             PSGSQIQPYLGTLTKGGVVNPSGRDLVASYLLRRILDNAGENGLVVQVML 265
gi|32474066|RB6095                 GDGRASTALNAVLQQGISADASHRDALARRVFWTLAELCDEFGLPFDLMI 285
gi|15613056|Bh0493                 -----------------TFSFPEESNRGRIIRDCLLPVAEKHNIPFAMMI 250


gi|94971332|Acid345_4306           GSERNHNRK---------------------------------------- 310
gi|225874247|ACP_2685              GSARNHNRK---------------------------------------- 310
gi|221639919|RSKD131_1820          GSFRNHSRP---------------------------------------- 315
gi|126462900|Rsph17029_2139        GSFRNHSRP---------------------------------------- 315
gi|146278446|Rsph170252412        GSFRNHSKP---------------------------------------- 315
```

```
gi|222149863|Avi_3892              GSYRNHSAG---------------------------------------- 313
gi|110680135|RD12934               GSRRNHSAQ---------------------------------------- 315
gi|161620871|BCAN_B0804            GSFRNHNPT---------------------------------------- 313
gi|23500543|BRA0812                GSFRNHNPT---------------------------------------- 313
gi|163844934|BSUIS_B0804           GSFRNHNPT---------------------------------------- 313
gi|17988821|BMEll0476|gi|17988     GSFRNHNPT---------------------------------------- 313
gi|189022606|BAbS19_ll04000        ------------------------------------------------
gi|158424966|AZC3342               GSVRNHNGG---------------------------------------- 318
gi|13473453|mll4056                GSWRNHSPS---------------------------------------- 316
gi|16264678|SMb21354               GSFRNHSPA---------------------------------------- 316
gi|16125737|CC1490                 GSHRNHNVG---------------------------------------- 317
gi|221234494|CCNA_01557            GSHRNHNVG---------------------------------------- 317
gi|167645720|Caul_1756             GSHRNHNTG---------------------------------------- 317
gi|197105913|PHZ_c2452             GSFRNHNAG---------------------------------------- 317
gi|84625827|XOO4170                GCFRNHNRP---------------------------------------- 319
gi|161898957|XOO4427               GCFRNHNRP---------------------------------------- 319
gi|188574617|PXO_03860             GCFRNHNRP---------------------------------------- 312
gi|21244968|XAC4251                GCFRNHNRQ---------------------------------------- 319
gi|78049913|XCV4357                GCFRNHNRQ---------------------------------------- 319
gi|66770505|XC4209                 GCFRNHNRQ---------------------------------------- 318
gi|21233539|XCC4117                GCFRNHNRQ---------------------------------------- 318
gi|188993724|xccb100_4329          GCFRNHNRA---------------------------------------- 318
gi|114568782|Mmar100230            GSFRNHNSP---------------------------------------- 335
gi|148554803|Swit1887              GSFRNHNAD---------------------------------------- 313
gi|103488522|Sala3046             GAFRNHNPW---------------------------------------- 317
gi|90020587|Sde0940                GSFRNHNPK---------------------------------------- 317
gi|90020919|Sde1272                GCYRNHNQQ---------------------------------------- 318
gi|116249865|RL0099                GSRRNTDSG---------------------------------------- 318
gi|209551594|Rleg2_4023            GSRRNTDSG---------------------------------------- 318
gi|86355751|RHECH00090             GSRRNTDSG---------------------------------------- 318
gi|190889727|RHECIAT_CH0000094     GSRRNTDSG---------------------------------------- 318
gi|119962615|AAur0527              GSYRNHHQP---------------------------------------- 310
gi|220911562|Achl_0787             GSFRNHHEP---------------------------------------- 321
gi|116669068|Arth0502              GSYRNHHGP---------------------------------------- 314
gi|163842140|RSal332093421|gi|     GSLRNH------------------------------------------ 170
gi|158315003|Franean1_3197         SVRRNHHTP---------------------------------------- 338
gi|134101579|SACE5055             GSRRNHHPD---------------------------------------- 345
gi|152964556|Krad_0587             GVLRDHDPA---------------------------------------- 317
gi|50955310|Lxx17350               GVHRNHHRP---------------------------------------- 370
gi|213691567|Blon_0669             GVMRNHYRK---------------------------------------- 323
gi|50843764|PPA2329                AVHRNHDTA---------------------------------------- 318
gi|161485962|CE2377                GVYRNHSAS---------------------------------------- 316
gi|145296483|cgR_2394              GVYRNHSAS---------------------------------------- 319
gi|62391323|cg2731                 GVYRNHSAS---------------------------------------- 319
gi|161486711|NCgI2398              GVYRNHSAS---------------------------------------- 319
gi|126208494|APL_1020              GALRNNNTR---------------------------------------- 37
gi|190150346|APP7_1077             GALRNNNTR---------------------------------------- 37
gi|165976445|APJL_1038             GALRNNNTR---------------------------------------- 37
gi|152977831|Asuc_0145             GALRNNNTR---------------------------------------- 308
gi|161511000|MS0544                GALRNNNTR---------------------------------------- 308
gi|148827203|CGSHiGG_02740         GAIRNNNSR---------------------------------------- 308
gi|68248599|NTHI0056               GAIRNNNSR---------------------------------------- 308
gi|113461738|HS_1602               GAIRNNNTR---------------------------------------- 304
gi|170718551|HSM_0409              GAIRNNNTR---------------------------------------- 308
gi|62181648|SC3078                 GAIRNNNTR---------------------------------------- 309
gi|197249740|SeAg_B3309            GAIRNNNTR---------------------------------------- 309
gi|198243637|SeD_A3482             GAIRNNNTR---------------------------------------- 309
gi|207858396|SEN2980               GAIRNNNTR---------------------------------------- 309
gi|194734239|SeSA_A3319            GAIRNNNTR---------------------------------------- 309
gi|205354057|SG3031                GAIRNNNTR---------------------------------------- 309
gi|161616115|SPAB03917             GAIRNNNTR---------------------------------------- 309
gi|16766437|STM3137                GAIRNNNTR---------------------------------------- 309
gi|194448866|SeHA_C3383            GAIRNNNTR---------------------------------------- 309
gi|194446286|SNSL254_A3389         GAIRNNNTR---------------------------------------- 309
gi|16761915|STY3308                GAIRNNNTR---------------------------------------- 309
gi|56415079|SPA3005                GAIRNNNTR---------------------------------------- 309
gi|197364008|SSPA2803              GAIRNNNTR---------------------------------------- 309
```

```
gi|161506293|Sari04489          GAIRNNNTR---------------------------------------- 309
gi|37524198|plu0176             GAIRNNNSR---------------------------------------- 309
gi|28901561|VPA1706             GAQRNNSTR---------------------------------------- 308
gi|37677254|VV1594              GAQRNNSTR---------------------------------------- 308
gi|27367460|VV21070             GAQRNNSTR---------------------------------------- 308
gi|119943921|Ping0131           GAQRNNNTR---------------------------------------- 313
gi|152996793|Mmwyl12780         GARRNNSSR---------------------------------------- 308
gi|117625400|APECO1_3327        GALRNNNLR---------------------------------------- 307
gi|91212521|UTI89_C3530         GALRNNNLR---------------------------------------- 328
gi|16130987|b3092               GALRNNNLR---------------------------------------- 307
gi|161486116|c3850              GALRNNNLR---------------------------------------- 307
gi|215488422|E2348C_3385        GALRNNNLR---------------------------------------- 307
gi|170082634|ECDH10B_3268       GALRNNNLR---------------------------------------- 307
gi|157155304|EcE24377A_3560     GALRNNNLR---------------------------------------- 307
gi|218691395|ECED1_3759         GALRNNNLR---------------------------------------- 307
gi|157162569|EcHS_A3275         GALRNNNLR---------------------------------------- 307
gi|218701863|ECIAI39_3589       GALRNNNLR---------------------------------------- 307
gi|170018657|EcolC_0608         GALRNNNLR---------------------------------------- 307
gi|110643336|ECP_3183           GALRNNNLR---------------------------------------- 307
gi|218560176|ECS88_3488         GALRNNNLR---------------------------------------- 307
gi|170683343|EcSMS35_3384       GALRNNNLR---------------------------------------- 307
gi|89109861|                    GALRNNNLR---------------------------------------- 307
gi|30064435|S3339               GALRNNNLR---------------------------------------- 307
gi|82545347|SBO2953             GALRNNNLR---------------------------------------- 307
gi|56480255|SF3132              GALRNNNLR---------------------------------------- 307
gi|110806974|SFV3133            GALRNNNLR---------------------------------------- 307
gi|74313639|SSON3245            GALRNNNLR---------------------------------------- 307
gi|187731951|SbBS512_E3528      GALRNNNLR---------------------------------------- 307
gi|209397829|ECH74115_4407      GALRNNNLR---------------------------------------- 307
gi|15803633|Z4445               GALRNNNLR---------------------------------------- 307
gi|15833228|ECs3974             GALRNNNLR---------------------------------------- 307
gi|218696796|EC55989_3506       GALRNNNLR---------------------------------------- 307
gi|209920564|ECSE_3373          GALRNNNLR---------------------------------------- 307
gi|218706716|ECUMN_3576         GALRNNNLR---------------------------------------- 307
gi|218551598|EFER_4383          GALRNNNLR---------------------------------------- 307
gi|206579202|KPK_0591           GALRNNNLR---------------------------------------- 307
gi|152972036|KPN_03520          GALRNNNLR---------------------------------------- 303
gi|146313181|Ent638_3546        GALRNNNQR---------------------------------------- 307
gi|157148666|CKO_04494          GALRNNNQR---------------------------------------- 328
gi|156935638|ESA_03503          GALRNNNQR---------------------------------------- 307
gi|161484751|y3600              GALRNNNSR---------------------------------------- 307
gi|161511340|YP2899             GALRNNNSR---------------------------------------- 307
gi|108809201|YPA3210            GALRNNNSR---------------------------------------- 307
gi|162419999|YPAngolaA1104      GALRNNNSR---------------------------------------- 307
gi|145597678|YPDSF0366          GALRNNNSR---------------------------------------- 307
gi|170022805|YPK_0554           GALRNNNSR---------------------------------------- 307
gi|108810613|YPN0448            GALRNNNSR---------------------------------------- 307
gi|16120908|YPO0579             GALRNNNSR---------------------------------------- 307
gi|153949672|YpsIP317580491     GALRNNNSR---------------------------------------- 307
gi|51597770|YPTB3478            GALRNNNSR---------------------------------------- 307
gi|186896961|YPTS_3663          GALRNNNSR---------------------------------------- 307
gi|123443892|YE3708             GALRNNNSR---------------------------------------- 307
gi|157372554|Spro4321           GALRNNNSR---------------------------------------- 307
gi|108810612|YPN0447            ------------------------------------------------
gi|50119594|ECA0645             GALRNNNLR---------------------------------------- 307
gi|54303188|PBPRB1515           GALRNNNQR---------------------------------------- 243
gi|60681862|BF1904G             GAIRNNNSR---------------------------------------- 307
gi|53713582|BF2293              GAIRNNNSR---------------------------------------- 307
gi|161511193|BT0823             GAIRNNNTK---------------------------------------- 307
gi|150005586|BVU3074            GAIRNNNTK---------------------------------------- 307
gi|212550592|CFPG_235           GALRNNNTK---------------------------------------- 308
gi|150009769|BDI3184            GANRNNNKR---------------------------------------- 371
gi|146301998|Fjoh_4262          GALRNNNAR---------------------------------------- 307
gi|192359962|CJA_2772           GAIRNNNPR---------------------------------------- 307
gi|182416100|Oter_4293          GALRNNNTR---------------------------------------- 304
gi|116624408|acid5330           GARRNNSTR---------------------------------------- 305
gi|225621141|BHWA1_02239        SAMRNNNEV---------------------------------------- 308
```

```
gi|18309134|CPE0152              GALRNNNTR---------------------------------------- 306
gi|160880875|Cphy_2743           GTKRDNNVF---------------------------------------- 306
gi|189485070|TGRD_067            GCKRNNNTL---------------------------------------- 306
gi|56962769|ABC0995              GALRNNNSR---------------------------------------- 310
gi|222529391|Athe_1402           NALRNNNTR---------------------------------------- 304
gi|146296954|Csac1949           NALRNNNTR---------------------------------------- 304
gi|150018894|Cbei_4082           AALRSNNTK---------------------------------------- 307
gi|150019467|Cbei_4657           AALRNNNNR---------------------------------------- 310
gi|20808343|TTE1939              GAMRNNNTK---------------------------------------- 306
gi|15893980|CAC0692              NAKRDNNTR---------------------------------------- 305
gi|150016705|Cbei_1832           NCVRDNNKV---------------------------------------- 306
gi|52081820|BL00708              NALRNNSSK---------------------------------------- 305
gi|52787207|BLi03516             NALRNNSSK---------------------------------------- 305
gi|157693739|BPUM2987            NALRNNNTK---------------------------------------- 305
gi|154685671|RBAM012380          NALRNTNTN---------------------------------------- 305
gi|16078295|BSU12300             NALRNTNTK---------------------------------------- 306
gi|23097822|OB0367               NALRNNNSR---------------------------------------- 306
gi|15613268|Bh0705               GVMRNNNSR---------------------------------------- 306
gi|138895421|GTNG_1765           GSIRNTNQK---------------------------------------- 306
gi|220929098|Ccel_1676           GVVRNTNTK---------------------------------------- 308
gi|81427739|LSA0129              NSDRDLNRP---------------------------------------- 310
gi|116332824|LVIS_0150           NSIRDLNRP---------------------------------------- 310
gi|116332809|LVIS_0135           NVLRNANGP---------------------------------------- 311
gi|116512444|LAACR_1742          NVNRSINGP---------------------------------------- 304
gi|125623707|llmg_0862           NVNRSINGP---------------------------------------- 304
gi|15673610|L0019                NVNRSINGP---------------------------------------- 304
gi|150389106|Amet1291            GAIRNNNSK---------------------------------------- 311
gi|70727647|SH2648               GAIRNNNTK---------------------------------------- 304
gi|25010729|gbs0674              GALRNNHST---------------------------------------- 303
gi|22536865|SAG0701              GALRNNHST---------------------------------------- 303
gi|76786751|SAK0827              GALRNNHST---------------------------------------- 303
gi|195977823|Sez_0689            GALRNNHSA---------------------------------------- 303
gi|225868878|SZO_13120           GALRNNHSA---------------------------------------- 303
gi|225870118|SEQ_0716            GALRNNHSA---------------------------------------- 303
gi|94990680|MGAS10270_Spy1171    GALRNNHSG---------------------------------------- 312
gi|222153335|SUB1203             GALRNNHSK---------------------------------------- 303
gi|146318813|SSU051159|gi|1463   GALRNNHTG---------------------------------------- 286
gi|146321021|SSU981174           GALRNNHTG---------------------------------------- 315
gi|225859885|SP70585_2238        GALRNNHTG---------------------------------------- 303
gi|157147928|CKO_03735           GAIRNNNTP---------------------------------------- 305
gi|220927672|Ccel_0214           GAIRSTNDN---------------------------------------- 303
gi|78221672|Gmet0451             GARRNNNTR---------------------------------------- 306
gi|148269995|Tpet0860            GALRDYRDS---------------------------------------- 300
gi|170288680|TRQ2_0883           GALRDYRDS---------------------------------------- 300
gi|15642839|TM0064               GALRDYRDS---------------------------------------- 300
gi|160902320|Pmob0854            GAIRDYRDK---------------------------------------- 302
gi|157364144|Tlet1289            GSYRDYRKK---------------------------------------- 302
gi|20808989|TTE2645              ------------------------------------------------
gi|159042183|Cmaq1622           G----AKRP---------------------------------------- 270
gi|32474066|RB6095               GVNRQVYPEGVFQGQDLYDSRVSLIQYRELFNSFPNVKFPVSVLASVTNQ 335
gi|15613056|Bh0493               GVKKRVHPA---------------------------------------- 268


gi|94971332|Acid345_4306         ------------------------------------------------
gi|225874247|ACP_2685            ------------------------------------------------
gi|221639919|RSKD131_1820        ------------------------------------------------
gi|126462900|Rsph17029_2139      ------------------------------------------------
gi|146278446|Rsph170252412       ------------------------------------------------
gi|222149863|Avi_3892            ------------------------------------------------
gi|110680135|RD12934             ------------------------------------------------
gi|161620871|BCAN_B0804          ------------------------------------------------
gi|23500543|BRA0812              ------------------------------------------------
gi|163844934|BSUIS_B0804         ------------------------------------------------
gi|17988821|BMEll0476|gi|17988   ------------------------------------------------
gi|189022606|BAbS19_ll04000      ------------------------------------------------
gi|158424966|AZC3342             ------------------------------------------------
gi|13473453|mll4056              ------------------------------------------------
```

```
gi|16264678|SMb21354              ------------------------------------------------
gi|16125737|CC1490                ------------------------------------------------
gi|221234494|CCNA_01557           ------------------------------------------------
gi|167645720|Caul_1756            ------------------------------------------------
gi|197105913|PHZ_c2452            ------------------------------------------------
gi|84625827|XOO4170               ------------------------------------------------
gi|161898957|XOO4427              ------------------------------------------------
gi|188574617|PXO_03860           ------------------------------------------------
gi|21244968|XAC4251               ------------------------------------------------
gi|78049913|XCV4357               ------------------------------------------------
gi|66770505|XC4209                ------------------------------------------------
gi|21233539|XCC4117               ------------------------------------------------
gi|188993724|xccb100_4329        ------------------------------------------------
gi|114568782|Mmar100230          ------------------------------------------------
gi|148554803|Swit1887            ------------------------------------------------
gi|103488522|Sala3046            ------------------------------------------------
gi|90020587|Sde0940              ------------------------------------------------
gi|90020919|Sde1272              ------------------------------------------------
gi|116249865|RL0099              ------------------------------------------------
gi|209551594|Rleg2_4023          ------------------------------------------------
gi|86355751|RHECH00090           ------------------------------------------------
gi|190889727|RHECIAT_CH0000094   ------------------------------------------------
gi|119962615|AAur0527            ------------------------------------------------
gi|220911562|Achl_0787           ------------------------------------------------
gi|116669068|Arth0502            ------------------------------------------------
gi|163842140|RSal332093421|gi|   ------------------------------------------------
gi|158315003|Franean1_3197       ------------------------------------------------
gi|134101579|SACE5055            ------------------------------------------------
gi|152964556|Krad_0587           ------------------------------------------------
gi|50955310|Lxx17350             ------------------------------------------------
gi|213691567|Blon_0669           ------------------------------------------------
gi|50843764|PPA2329              ------------------------------------------------
gi|161485962|CE2377              ------------------------------------------------
gi|145296483|cgR_2394            ------------------------------------------------
gi|62391323|cg2731               ------------------------------------------------
gi|161486711|NCgI2398            ------------------------------------------------
gi|126208494|APL_1020            ------------------------------------------------
gi|190150346|APP7_1077           ------------------------------------------------
gi|165976445|APJL_1038           ------------------------------------------------
gi|152977831|Asuc_0145           ------------------------------------------------
gi|161511000|MS0544              ------------------------------------------------
gi|148827203|CGSHiGG_02740       ------------------------------------------------
gi|68248599|NTHI0056             ------------------------------------------------
gi|113461738|HS_1602             ------------------------------------------------
gi|170718551|HSM_0409            ------------------------------------------------
gi|62181648|SC3078               ------------------------------------------------
gi|197249740|SeAg_B3309          ------------------------------------------------
gi|198243637|SeD_A3482           ------------------------------------------------
gi|207858396|SEN2980             ------------------------------------------------
gi|194734239|SeSA_A3319          ------------------------------------------------
gi|205354057|SG3031              ------------------------------------------------
gi|161616115|SPAB03917           ------------------------------------------------
gi|16766437|STM3137              ------------------------------------------------
gi|194448866|SeHA_C3383          ------------------------------------------------
gi|194446286|SNSL254_A3389       ------------------------------------------------
gi|16761915|STY3308              ------------------------------------------------
gi|56415079|SPA3005              ------------------------------------------------
gi|197364008|SSPA2803            ------------------------------------------------
gi|161506293|Sari04489           ------------------------------------------------
gi|37524198|plu0176              ------------------------------------------------
gi|28901561|VPA1706              ------------------------------------------------
gi|37677254|VV1594               ------------------------------------------------
gi|27367460|VV21070              ------------------------------------------------
gi|119943921|Ping0131            ------------------------------------------------
gi|152996793|Mmwyl12780          ------------------------------------------------
gi|117625400|APECO1_3327         ------------------------------------------------
gi|91212521|UTI89_C3530          ------------------------------------------------
```

```
gi|16130987|b3092                 ------------------------------------------------
gi|161486116|c3850                ------------------------------------------------
gi|215488422|E2348C_3385          ------------------------------------------------
gi|170082634|ECDH10B_3268         ------------------------------------------------
gi|157155304|EcE24377A_3560       ------------------------------------------------
gi|218691395|ECED1_3759           ------------------------------------------------
gi|157162569|EcHS_A3275           ------------------------------------------------
gi|218701863|ECIAI39_3589         ------------------------------------------------
gi|170018657|EcolC_0608           ------------------------------------------------
gi|110643336|ECP_3183             ------------------------------------------------
gi|218560176|ECS88_3488           ------------------------------------------------
gi|170683343|EcSMS35_3384         ------------------------------------------------
gi|89109861|                      ------------------------------------------------
gi|30064435|S3339                 ------------------------------------------------
gi|82545347|SBO2953               ------------------------------------------------
gi|56480255|SF3132                ------------------------------------------------
gi|110806974|SFV3133              ------------------------------------------------
gi|74313639|SSON3245              ------------------------------------------------
gi|187731951|SbBS512_E3528        ------------------------------------------------
gi|209397829|ECH74115_4407        ------------------------------------------------
gi|15803633|Z4445                 ------------------------------------------------
gi|15833228|ECs3974               ------------------------------------------------
gi|218696796|EC55989_3506         ------------------------------------------------
gi|209920564|ECSE_3373            ------------------------------------------------
gi|218706716|ECUMN_3576           ------------------------------------------------
gi|218551598|EFER_4383            ------------------------------------------------
gi|206579202|KPK_0591             ------------------------------------------------
gi|152972036|KPN_03520            ------------------------------------------------
gi|146313181|Ent638_3546          ------------------------------------------------
gi|157148666|CKO_04494            ------------------------------------------------
gi|156935638|ESA_03503            ------------------------------------------------
gi|161484751|y3600                ------------------------------------------------
gi|161511340|YP2899               ------------------------------------------------
gi|108809201|YPA3210              ------------------------------------------------
gi|162419999|YPAngolaA1104       ------------------------------------------------
gi|145597678|YPDSF0366            ------------------------------------------------
gi|170022805|YPK_0554             ------------------------------------------------
gi|108810613|YPN0448              ------------------------------------------------
gi|16120908|YPO0579               ------------------------------------------------
gi|153949672|YpsIP317580491       ------------------------------------------------
gi|51597770|YPTB3478              ------------------------------------------------
gi|186896961|YPTS_3663            ------------------------------------------------
gi|123443892|YE3708               ------------------------------------------------
gi|157372554|Spro4321            ------------------------------------------------
gi|108810612|YPN0447              ------------------------------------------------
gi|50119594|ECA0645               ------------------------------------------------
gi|54303188|PBPRB1515             ------------------------------------------------
gi|60681862|BF1904G               ------------------------------------------------
gi|53713582|BF2293                ------------------------------------------------
gi|161511193|BT0823               ------------------------------------------------
gi|150005586|BVU3074              ------------------------------------------------
gi|212550592|CFPG_235             ------------------------------------------------
gi|150009769|BDI3184              ------------------------------------------------
gi|146301998|Fjoh_4262           ------------------------------------------------
gi|192359962|CJA_2772            ------------------------------------------------
gi|182416100|Oter_4293           ------------------------------------------------
gi|116624408|acid5330            ------------------------------------------------
gi|225621141|BHWA1_02239         ------------------------------------------------
gi|18309134|CPE0152               ------------------------------------------------
gi|160880875|Cphy_2743           ------------------------------------------------
gi|189485070|TGRD_067            ------------------------------------------------
gi|56962769|ABC0995               ------------------------------------------------
gi|222529391|Athe_1402           ------------------------------------------------
gi|146296954|Csac1949            ------------------------------------------------
gi|150018894|Cbei_4082           ------------------------------------------------
gi|150019467|Cbei_4657           ------------------------------------------------
gi|20808343|TTE1939               ------------------------------------------------
```

```
gi|15893980|CAC0692              ------------------------------------------------
gi|150016705|Cbei_1832           ------------------------------------------------
gi|52081820|BL00708              ------------------------------------------------
gi|52787207|BLi03516             ------------------------------------------------
gi|157693739|BPUM2987            ------------------------------------------------
gi|154685671|RBAM012380          ------------------------------------------------
gi|16078295|BSU12300             ------------------------------------------------
gi|23097822|OB0367               ------------------------------------------------
gi|15613268|Bh0705              ------------------------------------------------
gi|138895421|GTNG_1765          ------------------------------------------------
gi|220929098|Ccel_1676          ------------------------------------------------
gi|81427739|LSA0129             ------------------------------------------------
gi|116332824|LVIS_0150          ------------------------------------------------
gi|116332809|LVIS_0135          ------------------------------------------------
gi|116512444|LAACR_1742         ------------------------------------------------
gi|125623707|llmg_0862           ------------------------------------------------
gi|15673610|L0019                ------------------------------------------------
gi|150389106|Amet1291           ------------------------------------------------
gi|70727647|SH2648               ------------------------------------------------
gi|25010729|gbs0674              ------------------------------------------------
gi|22536865|SAG0701              ------------------------------------------------
gi|76786751|SAK0827              ------------------------------------------------
gi|195977823|Sez_0689           ------------------------------------------------
gi|225868878|SZO_13120          ------------------------------------------------
gi|225870118|SEQ_0716           ------------------------------------------------
gi|94990680|MGAS10270_Spy1171   ------------------------------------------------
gi|222153335|SUB1203            ------------------------------------------------
gi|146318813|SSU051159|gi|1463  ------------------------------------------------
gi|146321021|SSU981174          ------------------------------------------------
gi|225859885|SP70585_2238       ------------------------------------------------
gi|157147928|CKO_03735          ------------------------------------------------
gi|220927672|Ccel_0214          ------------------------------------------------
gi|78221672|Gmet0451           ------------------------------------------------
gi|148269995|Tpet0860           ------------------------------------------------
gi|170288680|TRQ2_0883          ------------------------------------------------
gi|15642839|TM0064              ------------------------------------------------
gi|160902320|Pmob0854           ------------------------------------------------
gi|157364144|Tlet1289           ------------------------------------------------
gi|20808989|TTE2645             ------------------------------------------------
gi|159042183|Cmaq1622           ------------------------------------------------
gi|32474066|RB6095              ELVSYAWIFPNVITNGHWWYSNTPSFIARDAAARLEAVPQTKQIGYYSDA 385
gi|15613056|Bh0493               ------------------------------------------------


gi|94971332|Acid345_4306        --------------------LFEKYGRDVGADIPRAMN-YVDALGPLLEA 339
gi|225874247|ACP_2685            --------------------LYERYGRDVGADIPMPTN-YVDALRPMLDL 339
gi|221639919|RSKD131_1820        --------------------VLDRFGRDKGFDIPTRTD-YVRALRPLLSA 344
gi|126462900|Rsph17029_2139      --------------------VLDRFGRDKGFDIPTRTD-YVRALRPLLSA 344
gi|146278446|Rsph170252412       --------------------VMDRFGRDKGFDIPTRTD-YVRALRPLLSA 344
gi|222149863|Avi_3892            --------------------IMAKHGRDKGFDIPTRTD-YVRALKPLLDA 342
gi|110680135|RD12934             --------------------MFARFGRDKGFDIPGRTD-YVQDLRPLLDA 344
gi|161620871|BCAN_B0804          --------------------VFERFGLDKGADIPRQTG-FVDQLKPLLDA 342
gi|23500543|BRA0812              --------------------VFERFGLDKGADIPRQTG-FVDQLKPLLDA 342
gi|163844934|BSUIS_B0804         --------------------VFERFGLDKGADIPRQTG-FVDQLKPLLDG 342
gi|17988821|BMEll0476|gi|17988   --------------------VFERFGLDKGADIPRQTG-FVDQLKPLLDA 342
gi|189022606|BAbS19_ll04000      ------------------------------------------------
gi|158424966|AZC3342             --------------------ILARHGRDMGADIPSRTD-YVHALKPLLDR 347
gi|13473453|mll4056              --------------------VFQRFGRDKGFDIPTRTD-YVTALKPLLDC 345
gi|16264678|SMb21354             --------------------ILAKFGRDKGFDIPTRTD-YVTALKPLLDA 345
gi|16125737|CC1490               --------------------LLNSHGRDKGADIPMRTE-YVDALKPLLTR 346
gi|221234494|CCNA_01557          --------------------LLNSHGRDKGADIPMRTE-YVDALKPLLTR 346
gi|167645720|Caul_1756           --------------------LLASHGRDKGADIPMRTE-YVEALKPLLSK 346
gi|197105913|PHZ_c2452           --------------------LFARYGRDKGADIPTRTD-YVRALKPLLDR 346
gi|84625827|XOO4170              --------------------LFERYGRDKGADIPMRTD-YVHALKPLLDR 348
gi|161898957|XOO4427             --------------------LFERYGRDKGADIPMRTD-YVHALKPLLDR 348
gi|188574617|PXO_03860           --------------------LFERYGRDKGADIPMRTD-YVHALKPLLDR 341
gi|21244968|XAC4251              --------------------LFEQYGRDKGADIPMRTD-YVHALKPLLDR 348
```

```
gi|78049913|XCV4357               --------------------LFEQYGRDKGADIPMRTD-YVHALKPLLDR 348
gi|66770505|XC4209                --------------------LFERYGRDKGADIPMRTD-YVHALKPLLDR 347
gi|21233539|XCC4117               --------------------LFERYGRDKGADIPMRTD-YVHALKPLLDR 347
gi|188993724|xccb100_4329         --------------------LFERYGRDKGADIPMRTD-YVHALKPLLDR 347
gi|114568782|Mmar100230           --------------------LFASHGRDKGADIPTRTD-YVEALRPLLNR 364
gi|148554803|Swit1887             --------------------LFRRFGRDKGADIPTRTD-FVRALRPLLGR 342
gi|103488522|Sala3046             --------------------LFANHGRDKGADIPTATD-YVHALRPLLGR 346
gi|90020587|Sde0940               --------------------LFARYGRDKGADIPSQTE-YVNALKPLLDK 346
gi|90020919|Sde1272               --------------------IFAQFGRDKGADLPLATE-YVHNLKPLLDK 347
gi|116249865|RL0099               --------------------LFATRGPNMGADIPTSTD-WVGGLNALLSK 347
gi|209551594|Rleg2_4023           --------------------LFATRGPNMGADIPTRTD-WVGGLNALLSR 347
gi|86355751|RHECH00090            --------------------LFATRGPNMGADIPTRTD-WVGGLNALLSK 347
gi|190889727|RHECIAT_CH0000094    --------------------LFATRGPNMGADIPTRTD-WVGGLNALLSK 347
gi|119962615|AAur0527             --------------------TFEEFGADTGHDIPVATN-YTESIRPLLQD 339
gi|220911562|Achl_0787            --------------------TFNTYGADSGHDIPFAID-YTEAVRPLLQD 350
gi|116669068|Arth0502             --------------------TFEAFGADTGHDIPFAVN-YTEAIRPLLQD 343
gi|163842140|RSal332093421|gi|    --------------------------------------------------
gi|158315003|Franean1_3197       --------------------TLRRYGPDTGHDIPLRTE-FTQALQPLLDR 367
gi|134101579|SACE5055             --------------------TLRRFGPDTGHDIPLQAE-FTRALQPMLER 374
gi|152964556|Krad_0587            --------------------VHERFGSDKGFDIPIAVD-FVHGLRPLLSE 346
gi|50955310|Lxx17350              --------------------TFDQYGPDTGHDLHAVTA-FTEPLTPILRD 399
gi|213691567|Blon_0669            --------------------QLVNYGPDCGADCPMPAD-WAHWLRPMLNE 352
gi|50843764|PPA2329               --------------------AFHRFGADIGSDVPVTLE-VVDSLHPLLDK 347
gi|161485962|CE2377               --------------------AQKKFGNDIGADIPFQLE-YTRGLRPLLSD 345
gi|145296483|cgR_2394             --------------------AQKKFGADIGADIPFQME-FTNGLRPLLSD 348
gi|62391323|cg2731                --------------------AQKKFGADIGADIPFQME-FTNGLRPLLSD 348
gi|161486711|NCgI2398             --------------------AQKKFGADIGADIPFQME-FTNGLRPLLSD 348
gi|126208494|APL_1020             --------------------MFKLLGTDAGFDSIADLT-YASQLSQLLDA 66
gi|190150346|APP7_1077            --------------------MFKLLGTDADFDSIADLT-YASQLSQLLDA 66
gi|165976445|APJL_1038            --------------------MFKLLGTDAGFDSIADLT-YASQLSQLLYA 66
gi|152977831|Asuc_0145            --------------------MFKLLGADAGFDSIADRT-FAEPLSRLLDT 337
gi|161511000|MS0544               --------------------MFKLLGADSGFDSIADRT-FAEQLSRLLDA 337
gi|148827203|CGSHiGG_02740        --------------------MFALLGADSGFDSIGDRA-YAYPLSRLLDA 337
gi|68248599|NTHI0056              --------------------MFALLGADSGFDSIGDRT-YAYPLSRLLDA 337
gi|113461738|HS_1602              --------------------MFKLLGADAGFDSIGDRA-FAESLSRLLDS 333
gi|170718551|HSM_0409             --------------------MFKLLGADAGFDSIGDRA-FAESLSRLLDS 337
gi|62181648|SC3078                --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|197249740|SeAg_B3309           --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|198243637|SeD_A3482            --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|207858396|SEN2980              --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|194734239|SeSA_A3319           --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|205354057|SG3031               --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|161616115|SPAB03917            --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|16766437|STM3137               --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|194448866|SeHA_C3383           --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|194446286|SNSL254_A3389        --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|16761915|STY3308               --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|56415079|SPA3005               --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|197364008|SSPA2803             --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|161506293|Sari04489            --------------------MFRLLGPDTGFDSIGDNN-ISWALSRLLDS 338
gi|37524198|plu0176               --------------------MFRLLGADSGFDSIGDNN-IAWPLSRLLDS 338
gi|28901561|VPA1706               --------------------MFQLLGADAGFDSIGDRP-FAFELAHLLDE 337
gi|37677254|VV1594                --------------------MFQLLGADAGFDSIGDRP-FAFELAHLLDE 337
gi|27367460|VV21070               --------------------MFQLLGADAGFDSIGDRP-FAFELAHLLDE 337
gi|119943921|Ping0131             --------------------MFKLLGPDSGFDSIGDRP-FALQLGALLDA 342
gi|152996793|Mmwyl12780           --------------------MFKLIGGDSGFDSMDDRA-FAEPLSGFLNT 337
gi|117625400|APECO1_3327          --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|91212521|UTI89_C3530           --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 357
gi|16130987|b3092                 --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|161486116|c3850                --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|215488422|E2348C_3385          --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|170082634|ECDH10B_3268         --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|157155304|EcE24377A_3560       --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|218691395|ECED1_3759           --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|157162569|EcHS_A3275           --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|218701863|ECIAI39_3589         --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|170018657|EcolC_0608           --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
```

```
gi|110643336|ECP_3183            --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|218560176|ECS88_3488          --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|170683343|EcSMS35_3384        --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|89109861|                     --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|30064435|S3339                --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|82545347|SBO2953              --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|56480255|SF3132               --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|110806974|SFV3133             --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|74313639|SSON3245             --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|187731951|SbBS512_E3528       --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|209397829|ECH74115_4407       --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|15803633|Z4445                --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|15833228|ECs3974              --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|218696796|EC55989_3506        --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|209920564|ECSE_3373           --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|218706716|ECUMN_3576          --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|218551598|EFER_4383           --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|206579202|KPK_0591            --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 336
gi|152972036|KPN_03520           --------------------QFKLLGPDVGFDSINDRP-MAEELSKLLSK 332
gi|146313181|Ent638_3546         --------------------QFKLLGADVGFDSINDRP-LAEELSKLLSK 336
gi|157148666|CKO_04494           --------------------QFKLLGPDVGFDSINDRP-LAEELSKLLSK 357
gi|156935638|ESA_03503           --------------------QFRLLGPDVGFDSINDRP-LAEELSKLLSK 336
gi|161484751|y3600               --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|161511340|YP2899              --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|108809201|YPA3210             --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|162419999|YPAngolaA1104       --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|145597678|YPDSF0366           --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|170022805|YPK_0554            --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|108810613|YPN0448             --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|16120908|YPO0579              --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|153949672|YpsIP317580491      --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|51597770|YPTB3478             --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|186896961|YPTS_3663           --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|123443892|YE3708              --------------------MFNLVGPDIGFDSINDQP-LAQPLSRLLDA 336
gi|157372554|Spro4321            --------------------MLASVGPDIGFDSINDRP-LAEALSRLLDA 336
gi|108810612|YPN0447             ------------------------------------------------
gi|50119594|ECA0645              --------------------QFKLLGPDVGFDSINDRP-LAQELSRLLSK 336
gi|54303188|PBPRB1515            --------------------MLSILGPDTGFDSINDGL-IAAPLAKLLNA 272
gi|60681862|BF1904G              --------------------MFKQLGPDTGFDSIGEFA-TAKAMSKFLDR 336
gi|53713582|BF2293               --------------------MFKQLGPDTGFDSIGEFA-TAKAMSKFLDR 336
gi|161511193|BT0823              --------------------MFKLLGPDTGFDSIGEFT-TAKAMSKFLDR 336
gi|150005586|BVU3074             --------------------MFKLLGPDTGFDSIGEFT-TAKAMAKFLDR 336
gi|212550592|CFPG_235            --------------------MFMQIGADSGYDSIGDWN-TARSMSKFLNC 337
gi|150009769|BDI3184             --------------------MFKLLGPDTGFDAIDDQP-VAVSMNRFFSH 400
gi|146301998|Fjoh_4262           --------------------MHRILGPDTGWDSIGDYP-QAQKLSSFLNA 336
gi|192359962|CJA_2772            --------------------LFRTLGPDTGFDSIGDHN-YAKPLAKFLGR 336
gi|182416100|Oter_4293           --------------------LLATLGPDTGFDSIGDFP-QTRALSRYLDT 333
gi|116624408|acid5330            --------------------RFRELGADTGWDSIGDWP-QADALGTYLDR 334
gi|225621141|BHWA1_02239         --------------------MFKKLGADTGYDSVGDSN-IIEKLSFLLKT 337
gi|18309134|CPE0152              --------------------MFNRIGADAGFDSIDDGE-IAYSLSRILDE 335
gi|160880875|Cphy_2743           --------------------RFNQLGADTGFDCINTEG-SSAELANFLNA 335
gi|189485070|TGRD_067            --------------------VFDKLGPDTGHDCIDNFA-PGAQLSNFLDA 335
gi|56962769|ABC0995              --------------------MVEQVGPDSGFDSMADDR-FAESLNRFLNE 339
gi|222529391|Athe_1402           --------------------MFNILGPDTGYDSINDGH-IAFALVKFLDS 333
gi|146296954|Csac1949            --------------------MFNILGPDTGYDSINDGH-IALALVKFLDS 333
gi|150018894|Cbei_4082           --------------------MLEKIGVNTGFDSINDES-IAYPLSRLLDS 336
gi|150019467|Cbei_4657           --------------------MFEKIGPNTGFDSINDES-IAYPLSRLLDS 339
gi|20808343|TTE1939              --------------------MYRILGPDTGYDSIGDFP-IAYSLSRLLDS 335
gi|15893980|CAC0692              --------------------MYNKLGPDTGFDSVNDNG-VAGPLSRFLDS 334
gi|150016705|Cbei_1832           --------------------MFEKLGPDTGFDSLNDTE-VAIPLSRLLDA 335
gi|52081820|BL00708              --------------------MYRKLGPDTGYDAINDQD-IAKPLCSFLDS 334
gi|52787207|BLi03516             --------------------MYRKLGPDTGYDAINDQD-IAKPLCSFLDS 334
gi|157693739|BPUM2987            --------------------MFERLGPDTGYDAMNDED-IAKPLCRILDR 334
gi|154685671|RBAM012380          --------------------KFSSLGPDTGYDSINDER-IAKPLARLLDS 334
gi|16078295|BSU12300             --------------------MMKRLGPDTGYDSMNDEE-IAKPLYKLLNS 335
gi|23097822|OB0367               --------------------MYNNLGPDTGYDAMNDEV-ISKPLVNLLNE 335
gi|15613268|Bh0705               --------------------MLHKLGPDTGFDSIGDGQ-IAHATAKLLDL 335
```

```
gi|138895421|GTNG_1765            --------------------MFQQLGPNTGYDSINDFF-FAQPLNAFLNR 335
gi|220929098|Ccel_1676            --------------------MLKSTGVDAGFDSIGNGI-SADSLISLLDW 337
gi|81427739|LSA0129               --------------------MFAKIGPDTGYDAVGTQPDIVNHISALYTK 340
gi|116332824|LVIS_0150            --------------------MFDKLGADTGYDAVGTQPDIAGELQKLYTK 340
gi|116332809|LVIS_0135            --------------------MFRKLGADAGFDSMGSQPGIAGELMQLMSD 341
gi|116512444|LAACR_1742           --------------------ALRKIGENTGFDSMGTQANISEELTKLFSK 334
gi|125623707|llmg_0862            --------------------ALRKIGENTGFDSMGTQANISEELTKLFSK 334
gi|15673610|L0019                 --------------------ALSKIGENTGFDSMGTQANISEELTKLFSK 334
gi|150389106|Amet1291             --------------------MYELVGENTGFDSIVDQSNVAYALNRLLDA 341
gi|70727647|SH2648                --------------------MFEKVGKDAGFDSIRDQDNLAYHLNATLDM 334
gi|25010729|gbs0674               --------------------IFEKLGADVGVDSLGDQVALTVNMNRLLDS 333
gi|22536865|SAG0701               --------------------IFEKLGADVGVDSLGDQVALTVNMNRLLDS 333
gi|76786751|SAK0827               --------------------IFEKLGADVGVDSLGDQVALTVNMNSLLDS 333
gi|195977823|Sez_0689             --------------------IFNKLGADVGVDSLGDQAALAINMNRLLDH 333
gi|225868878|SZO_13120            --------------------IFNKLGADVGVDSLGDQAALAINMNRLLDH 333
gi|225870118|SEQ_0716             --------------------IFNKLGADVGVDSLGDQTGLAINMNRLLDH 333
gi|94990680|MGAS10270_Spy1171     --------------------IFKRLGADVGVDSLGDQTALTSNMNRLLDN 342
gi|222153335|SUB1203              --------------------LYQKLGADVGIDSLGDQTALTSNMNKLLDN 333
gi|146318813|SSU051159|gi|1463    --------------------LFEKLGADVGVDSIGDQTCLTVNLNRFLDN 316
gi|146321021|SSU981174            --------------------LFEKLGADVGVDSIGDQTCLTVNLNRFLDN 345
gi|225859885|SP70585_2238         --------------------LFEKLGADVGVDSIGDQTCLTGNLNRLLDN 333
gi|157147928|CKO_03735            --------------------MFEKIGINCGFDSIGDQTHLAESLNGLLNA 335
gi|220927672|Ccel_0214            --------------------MLKKAGVDSGFDIMNDFQ-IAEPLAKLLND 332
gi|78221672|Gmet0451              --------------------IFSSYGLDAGCDSIGDFP-QGESLVAFLDQ 335
gi|148269995|Tpet0860             --------------------LFKTLGPDSGGDISTNFLRIAEGLRYFLNE 330
gi|170288680|TRQ2_0883            --------------------LFKTLGPDSGGDISTNFLRIAEGLRYFLNE 330
gi|15642839|TM0064                --------------------LFKTLGPDSGGDISTNFLRIAEGLRYFLNE 330
gi|160902320|Pmob0854             --------------------LYNTLGPDTGGDISTSKIDLANGLRYFLNQ 332
gi|157364144|Tlet1289             --------------------LFDSLGPDSGGDISAGFVDIARGMKEFFNA 332
gi|20808989|TTE2645               --------------------------------------------------
gi|159042183|Cmaq1622             --------------------VPGASPPDYAVTLFNQQQ--LMQVMLLASE 298
gi|32474066|RB6095                YKLEFIWPKFDMYRKVLSGVLADQFVCDRGWSVERAVELGRKVLRDNVEE 435
gi|15613056|Bh0493                -------------------------LGDAGDFVGKAS--MDGVEHLLRE 290

gi|94971332|Acid345_4306          FGNEPRLTIILF----TLDESVYGR----------ELAPLAG----HYPC 371
gi|225874247|ACP_2685             VGNERGLTIILF----TLDETAYSR----------ELAPLAG----HYPC 371
gi|221639919|RSKD131_1820         VGGDPRLTVIVF----TLDETAYAR----------ELAPLAG----VYPA 376
gi|126462900|Rsph17029_2139       VGGDPRLTVIVF----TLDETAYAR----------ELAPLAG----VYPA 376
gi|146278446|Rsph170252412        VGGDPRLTVIVF----TLDETAYAR----------ELAPLAG----VYPA 376
gi|222149863|Avi_3892             VGMEKDLTVILF----TLDETSYSR----------ELAPLAG----AYPA 374
gi|110680135|RD12934              VGMEPALNIILF----TLDETTYSR----------ELAPLAG----AYPA 376
gi|161620871|BCAN_B0804           FGNDPRLTVILF----TLDETAYSR----------ELAPLAG----HYPA 374
gi|23500543|BRA0812               FGNDPRLTVILF----TLDETAYSR----------ELAPLAG----HYPA 374
gi|163844934|BSUIS_B0804          FGNDPRLTVILF----TLDETAYSR----------ELAPLAG----DYPA 374
gi|17988821|BMEll0476|gi|17988    FGNDPRLTVILF----TLDETALVANWRRSPATIQPMAPLAG----HYPA 384
gi|189022606|BAbS19_ll04000       ----------------------------------MAPLAG----HYPA 10
gi|158424966|AZC3342              FGNEKNLTIILF----TLDETSYSR----------ELAPLAG----HYPV 379
gi|13473453|mll4056               VGLERDLTVILF----TLDESSYAR----------ELAPLAG----VYPA 377
gi|16264678|SMb21354              VGLERDLTVILF----TLDETSYAR----------ELAPLAG----VYPA 377
gi|16125737|CC1490                LGNDPRLSIILF----TLDETTYSR----------ELAPLAG----HYPV 378
gi|221234494|CCNA_01557           LGNDPRLSIILF----TLDETTYSR----------ELAPLAG----HYPV 378
gi|167645720|Caul_1756            LGNDPRLSIILF----TLDETVYSR----------ELAPLAG----HYPV 378
gi|197105913|PHZ_c2452            FGNERGLSVIVF----TLDESSYAR----------ELAPLAG----HYPV 378
gi|84625827|XOO4170               YGNDPRLRLIVF----TLDETSYSR----------ELAPLAG----HYPS 380
gi|161898957|XOO4427              YGNDPRLRLIVF----TLDETSYSR----------ELAPLAG----HYPS 380
gi|188574617|PXO_03860            YGNDRRLRLIVF----TLDETSYSR----------ELAPLAG----HYPS 373
gi|21244968|XAC4251               HGNDPRLRLIVF----TLDETSYSR----------ELAPLAG----HYPS 380
gi|78049913|XCV4357               HGNDPRLRLIVF----TLDETSYSR----------ELAPLAG----HYPS 380
gi|66770505|XC4209                FGNDPRFRLIVF----TLDETSYSR----------ELAPLAG----HYPA 379
gi|21233539|XCC4117               FGNDPRFRLIVF----TLDETSYSR----------ELAPLAG----HYPA 379
gi|188993724|xccb100_4329         FGNDPRFRLIVF----TLDETSYSR----------ELAPLAG----HYPA 379
gi|114568782|Mmar100230           FGNEPDFGCIVF----TLDETSYAR----------ELAPLAG----HYPA 396
gi|148554803|Swit1887             FGNDPRLSLILF----TLDESAYAR----------ELAPLAG----HYPA 374
gi|103488522|Sala3046            YGNEADLTIILF----TLDETSYAR----------ELAPLAG----HYPA 378
gi|90020587|Sde0940               YGNEPNLTIILF----TLDETVYSR----------ELAPLAG----HYPA 378
gi|90020919|Sde1272               FGNNPDLTIILF----TLDETSYSR----------ELAPLAG----HYPA 379
```

```
gi|116249865|RL0099                YGHAPGLRVLLF----TLDETTYAR----------ELAPMVG----HWPC 379
gi|209551594|Rleg2_4023            YGHAPGLRVLLF----TLDETTYAR----------EMAPMVG----HWPC 379
gi|86355751|RHECH00090             YGHAPGLRILLF----TLDETTYAR----------ELAPMAG----HWPC 379
gi|190889727|RHECIAT_CH0000094     YGHAPGLRVLLF----TLDETTYAR----------ELAPMAG----HWPC 379
gi|119962615|AAur0527              FGTAKDFHLVLF----TLDETVFSR----------ELAPLAG----FYPS 371
gi|220911562|Achl_0787             FGAAKDFHLVLF----TMDETVFSR----------ELAPLAG----FYPS 382
gi|116669068|Arth0502              FGTAKDFHLVLF----TLDETVFSR----------ELAPLAG----FYPS 375
gi|163842140|RSal332093421|gi|     ------------------------------------------------
gi|158315003|Franean1_3197        HGTDPDFRLVLF----TLDETVFSR----------EIAPLAG----FYPA 399
gi|134101579|SACE5055             FGTNPRFQTVLF----TLDETVFSR----------ELAPLAG----FYPS 406
gi|152964556|Krad_0587             FGSDPAFRCVLF----TIDETTFSR----------ELAPLAG----AYPS 378
gi|50955310|Lxx17350               FGTNPTFRLVLF----TVDETAFSR----------EIAPLAG----FYPS 431
gi|213691567|Blon_0669             YGENPDFHLVAF----TMDETAYSR----------ELAPMAA----YYPA 384
gi|50843764|PPA2329                FG-NTDLKLVVF----TIDETLYSR----------EIAPLSG----WYRS 378
gi|161485962|CE2377                FGENKDFHFVMF----TIDETVFSR----------EVAPLAG----YYPA 377
gi|145296483|cgR_2394              FGENKDFHFVMF----TIDESVYSR----------EVAPLAG----YYPA 380
gi|62391323|cg2731                 FGENKDLISDSSFNRWLRTVSLGST----------QDADMAAA--SNLAA 386
gi|161486711|NCgI2398              FGENKDLISDSSFNRWLRTVSLGST----------QDADMAAA--SNLAA 386
gi|126208494|APL_1020              MDQTNQLPKTILYCLNPRDNEMIAS----------MIGNFQT--DGIAGK 104
gi|190150346|APP7_1077             MDQTNQLPKTILYCLNPRDNEMIAS----------MIENFQT--DGIAGK 104
gi|165976445|APJL_1038             MDQTNQLPKTILYCLNPRDNEMIAS----------MIGNFQT--GGIAGK 104
gi|152977831|Asuc_0145             MDQNDELPKTILYCLNPRDNEMIAT----------MIGNFQT--GGVAGK 375
gi|161511000|MS0544                MDQNNQLPKTILYCLNPRDNEMIAT----------MIGNFQT--GGIAGK 375
gi|148827203|CGSHiGG_02740         MDKENQLPKTILYCLNPRDNEMIAS----------MIGNFQG--DGIAGK 375
gi|68248599|NTHI0056               MDKENQLPKTILYCLNPRDNEMIAS----------MIGNFQG--DGIAGK 375
gi|113461738|HS_1602               MDQTDQLPKTILYCLNPRDNEMLGT----------MIGNFQT--GGIAGK 371
gi|170718551|HSM_0409              MDQTDQLPKTILYCLNPRDNEMLGT----------MIGNFQT--GGIAGK 375
gi|62181648|SC3078                 MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|197249740|SeAg_B3309            MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|198243637|SeD_A3482             MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|207858396|SEN2980               MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|194734239|SeSA_A3319            MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|205354057|SG3031                MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|161616115|SPAB03917             MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|16766437|STM3137                MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|194448866|SeHA_C3383            MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|194446286|SNSL254_A3389         MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|16761915|STY3308                MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|56415079|SPA3005                MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|197364008|SSPA2803              MDVTNELPKTILYCLNPRDNEVLAT----------MIGNFQG--PGIAGK 376
gi|161506293|Sari04489             MDVTNELPKTILYCLNPRDNEVLAT----------MTGNFQG--PGIAGK 376
gi|37524198|plu0176                MDVTDELPKTILYCLNPRDNEVIAT----------MASNFQG--GGIAGK 376
gi|28901561|VPA1706                MDQTNELPRTILYCLNPRDNEMMAT----------MIGNFQG--GGIAGK 375
gi|37677254|VV1594                 MDQTNELPRTILYCLNPRDNEMMAT----------MIGNFQG--GGIAGK 375
gi|27367460|VV21070                MDQTNELPRTILYCLNPRDNEMMAT----------MIGNFQG--GGIAGK 375
gi|119943921|Ping0131              MDITNELPKTILYCLNPRDNEMLAT----------MMGNYQG--GGIAGK 380
gi|152996793|Mmwyl12780            LDQSDELPKTILYCLNPMHNEMLAT----------MAGNFQG--GGVAGK 375
gi|117625400|APECO1_3327           QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|91212521|UTI89_C3530            QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 395
gi|16130987|b3092                  QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|161486116|c3850                 QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|215488422|E2348C_3385           QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|170082634|ECDH10B_3268          QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|157155304|EcE24377A_3560        QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|218691395|ECED1_3759            QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|157162569|EcHS_A3275            QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|218701863|ECIAI39_3589          QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|170018657|EcolC_0608            QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|110643336|ECP_3183              QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|218560176|ECS88_3488            QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|170683343|EcSMS35_3384          QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|89109861|                       QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|30064435|S3339                  QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|82545347|SBO2953                QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|56480255|SF3132                 QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|110806974|SFV3133               QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|74313639|SSON3245               QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
```

```
gi|187731951|SbBS512_E3528          QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|209397829|ECH74115_4407          QNEENLLPKTILYCLNPRDNEVLGT----------MSGNFQG--EGMPGK 374
gi|15803633|Z4445                   QNEENLLPKTILYCLNPRDNEVLGT----------MSGNFQG--EGMPGK 374
gi|15833228|ECs3974                 QNEENLLPKTILYCLNPRDNEVLGT----------MSGNFQG--EGMPGK 374
gi|218696796|EC55989_3506           QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|209920564|ECSE_3373              QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|218706716|ECUMN_3576             QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|218551598|EFER_4383              QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|206579202|KPK_0591               QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|152972036|KPN_03520              QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 370
gi|146313181|Ent638_3546            QNEQNLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|157148666|CKO_04494              QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 395
gi|156935638|ESA_03503              QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|161484751|y3600                  QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGEAGK 374
gi|161511340|YP2899                 QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGEAGK 374
gi|108809201|YPA3210                QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGEAGK 374
gi|162419999|YPAngolaA1104          QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGEAGK 374
gi|145597678|YPDSF0366              QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGEAGK 374
gi|170022805|YPK_0554               QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGEAGK 374
gi|108810613|YPN0448                QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGEAGK 374
gi|16120908|YPO0579                 QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGEAGK 374
gi|153949672|YpsIP317580491         QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGEAGK 374
gi|51597770|YPTB3478                QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGEAGK 374
gi|186896961|YPTS_3663              QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGEAGK 374
gi|123443892|YE3708                 QGLRNALPKTILYCLNPRDNEVIGT----------MVGNFQG--EGAAGK 374
gi|157372554|Spro4321               QARQGGLPKTILYCLNPRDNEVIGT----------MVGNFQG--EGTPGK 374
gi|108810612|YPN0447                ------------------------------------------------
gi|50119594|ECA0645                 QNEENLLPKTILYCLNPRDNEVLGT----------MIGNFQG--EGMPGK 374
gi|54303188|PBPRB1515               LDATDSLPKTILYCLNPRDNEVIAT----------MCGNFQG--GGIAGK 310
gi|60681862|BF1904G                 LNSKGKLTKTILYNLNPCANEVIAT----------MIGNFQD--GSIPGK 374
gi|53713582|BF2293                  LNSKGKLTKTILYNLNPCANEVIAT----------MIGNFQD--GSIPGK 374
gi|161511193|BT0823                 LNVNGKLTKTILYNLNPCANEVIAT----------MLGNFQD--GSIAGK 374
gi|150005586|BVU3074                LNTEGKLTKTILYNLNPCANEVIAT----------MLGNFQD--GTIPGK 374
gi|212550592|CFPG_235               LDKQNKLTRTILYNLNPNDNEMIAT----------MVGNFQD--GSIAGK 375
gi|150009769|BDI3184                LDQEGLLAKTIVYNLNPRDTELMVA----------NAYNFND--GSVPGK 438
gi|146301998|Fjoh_4262              LDSKDKLTKTIIYNLNPADNEVMAT----------MIGNFND--GSVRGK 374
gi|192359962|CJA_2772               LDDQNKLAKTILYNLNPRDNEMIGT----------MIGNFQD--GSAAGK 374
gi|182416100|Oter_4293              LDSTDELPRTVLYNLNPADNYAFAT----------MIGNFQD--GSVPGK 371
gi|116624408|acid5330               LDLENALPKTVIYNLNPADNYVIAT----------MIGNFQD--GVTAGK 372
gi|225621141|BHWA1_02239            ANNDGGLPKIIFYSLNHKDYYPLST----------LMGCFQE--GSIKGK 375
gi|18309134|CPE0152                 LDKEDKLPKTILYCLNPKDNEVLGT----------MIGNFQG--GNIAGK 373
gi|160880875|Cphy_2743              LNSSDELPKTIIYSLNPTDNAAIGT----------VLGCFQD--STAVGK 373
gi|189485070|TGRD_067               LDSAGQLPRTIIYSLNPNDNAVIGT----------ILGCFQD--SSALSK 373
gi|56962769|ABC0995                 LERTNELPKTILYTLNPIFNEVIAT----------TIGNFQG--GGIPGK 377
gi|222529391|Athe_1402              LEKENSLPKTILYSLNPKDNYVLAT----------IMGSFQG--GGIPGK 371
gi|146296954|Csac1949               LEKENSLPKTILYSLNPKDNYVLAT----------IMGSFQD--GSIPGK 371
gi|150018894|Cbei_4082              VDRENSLPKTILYTLNPKDNYVLGT----------MIGNFQG--DGIPGK 374
gi|150019467|Cbei_4657              VDKENSLPKTILYTLNPKDNYVLGA----------MIGNFQG--EGIPGK 377
gi|20808343|TTE1939                 LEIEGALPKTILYTLNPAANYVIAT----------MIGNFQD--GKIAGK 373
gi|15893980|CAC0692                 LEKEGSLPKTIIYSLNPNDNFVIGT----------LLGCFQG--TEAFGK 372
gi|150016705|Cbei_1832              LNSEKALPKTIIYSLNPNDNSTIGT----------LLGCFQG--DGIKGK 373
gi|52081820|BL00708                 LDRKDALPKTILYSLNPRDNVVISS----------LCGSFQD--GRIPGK 372
gi|52787207|BLi03516                LDRKDALPKTILYSLNPRDNVVISS----------LCGSFQD--GRIPGK 372
gi|157693739|BPUM2987               LEQEDALPNTILYSLNPRDNVVIST----------LAGSFQD--GKTPGK 372
gi|154685671|RBAM012380             AEKKRQLPKTILYSLNPNDNYIIAS----------MINSFQD--GKTPGK 372
gi|16078295|BSU12300                VEMKNQLPKTILYSLNPNDNYVIAS----------MINSFQD--GITPGK 373
gi|23097822|OB0367                  LEKKDRLPKTILYSLNPNDNPVIAS----------IIGSFQR--GGVPGK 373
gi|15613268|Bh0705                  LDKQGALPKTILYCVNPNANYILAS----------MIGNFTE--SGVRGK 373
gi|138895421|GTNG_1765              LERDGQLPKTIVYTLNPAYNYIVAS----------TIGNFPS--EGVKGK 373
gi|220929098|Ccel_1676              MNKTDELPRTVIYCLNDTDNAKIAS----------IIGCFQS--TMFPSK 375
gi|81427739|LSA0129                 MQQTEDVPKSIFYSLNSNDWLELAT----------MMGCFQG--GTTQ-K 377
gi|116332824|LVIS_0150              MRSSNDIPKSIFYSLNDNDWMQLAT----------MMGAFQE--GGQQ-R 377
gi|116332809|LVIS_0135              MAEENNIPKTIFFSTNPNDWLELAT----------GMQSFQG--GVTQ-K 378
gi|116512444|LAACR_1742             AAELEIIPKTILYSLNQNDWLELAT----------LIGCFQG--EGVQ-Q 371
gi|125623707|llmg_0862              AAELEIIPKTILYSLNQNDWLELAT----------LIGCFQG--EGVQ-Q 371
gi|15673610|L0019                   AAELETIPKTILYSLNQNDWLELVT----------MIGCFQG--EGVQ-Q 371
gi|150389106|Amet1291               MSRKNCLPKTVVYNLNPDYNHIVAS----------AVSNFQDNEEGIRGK 381
```

```
gi|70727647|SH2648                MEQEGHLPKTILYNLNPIYNDIVGS----------TIANFQT-EPGIKSK 373
gi|25010729|gbs0674               LVKKDSLPKMIWYNLNPAYNIAVAN----------TLANFQANELGVRSY 373
gi|22536865|SAG0701               LVKKDSLPKMIWYNLNPAYNIAVAN----------TLANFQANELGVRSY 373
gi|76786751|SAK0827               LVKKDSLPKMIWYNLNPAYNIAVAN----------TLANFQANELGVRSY 373
gi|195977823|Sez_0689             LVQRDSLPKMIWYNLNPSYNITVPN----------TLANFQANENGIAGY 373
gi|225868878|SZO_13120            LVQRDSLPKMIWYNLNPSYNITVAN----------TLANFQANENGIAGY 373
gi|225870118|SEQ_0716             LVQRDSLPKMIWYNLNPSYNIAVAN----------TLANFQANENGIAGY 373
gi|94990680|MGAS10270_Spy1171     LVQTDSLPKMIWYNLNPSYNIAVAN----------TLANFQANEAGVASY 382
gi|222153335|SUB1203              LVQKDALPKMIWYNLNPSYNIAVAN----------TLANFQANEEGVKSY 373
gi|146318813|SSU051159|gi|1463    LVKENALPKMIWYNLNPSYNIALAN----------TLANFQSNEEGVRSQ 356
gi|146321021|SSU981174            LVKENALPKMIWYNLNPSYNIALAN----------TLANFQSNEEGVRSQ 385
gi|225859885|SP70585_2238         LVKENALPKMIWYNLNPGYNIALAN----------TLANFQANEEGRRSQ 373
gi|157147928|CKO_03735            MAENNGLPKTILYNLNASYNDVVAS----------TIANFQSGEDGVKSP 375
gi|220927672|Ccel_0214            MDKMNCLPKTILYTLNSKDNLVLSS----------LPHCFTE--DGVPGK 370
gi|78221672|Gmet0451              LDTSERLPKTILFNSNPRDTLMFAT----------IAGSFWG--EGIRGK 373
gi|148269995|Tpet0860             FDGK--LKIVLYVLDPTHLPTVAT----------IARAFPN-------- 359
gi|170288680|TRQ2_0883            FDGK--LKIVLYVLDPTHLPTIAT----------IARAFPN-------- 359
gi|15642839|TM0064                FDGK--LKIVLYVLDPTHLPTIST----------IARAFPN-------- 359
gi|160902320|Pmob0854             FDEK--LKVVLYCLDPSLFPTVAT----------IARAFPN-------- 361
gi|157364144|Tlet1289             FDGK--VKIIIYCMDMTYLSVAAT----------IARAFEN-------- 361
gi|20808989|TTE2645               ------------------------------------------------
gi|159042183|Cmaq1622            YPNVKFDLIITEPLLNHAATVAAKN----------------------YPN 326
gi|32474066|RB6095                IFPLPKGATNSTSEEDSSDNADAPNAGVESLGALGVATAVAG----VTEV 481
gi|15613056|Bh0493                YPNNK-FLVTMLSRENQHELVVLAR-------------KFSN-------- 318

gi|94971332|Acid345_4306          LRLGPPWWFHDSPEGMA-----------------------------RF 390
gi|225874247|ACP_2685             LRLGPPWWFHDSPEGMM-----------------------------RF 390
gi|221639919|RSKD131_1820         LKLGPAWWFHDSPEGMR-----------------------------RF 395
gi|126462900|Rsph17029_2139       LKLGPAWWFHDSPEGMR-----------------------------RF 395
gi|146278446|Rsph170252412        LKLGPAWWFHDSPEGMR-----------------------------RF 395
gi|222149863|Avi_3892             LKLGPAWWFFDSPDGMR-----------------------------RF 393
gi|110680135|RD12934              LKLGPPWWFFDSYEGIK-----------------------------RF 395
gi|161620871|BCAN_B0804           LKLGPAWWFFDSPEGIL-----------------------------RY 393
gi|23500543|BRA0812               LKLGPAWWFFDSPEGIL-----------------------------RY 393
gi|163844934|BSUIS_B0804          LKLGPAWWFFDSPEGIL-----------------------------RY 393
gi|17988821|BMEll0476|gi|17988    LKLGPAWWFFDSPEGIL-----------------------------RY 403
gi|189022606|BAbS19_ll04000       LKLGPAWWFFDSPEGIL-----------------------------RY 29
gi|158424966|AZC3342              LRLGPAWWFFDAPEGMK-----------------------------RF 398
gi|13473453|mll4056               LKLGPAWWFHDSPEGMR-----------------------------RF 396
gi|16264678|SMb21354              LKLGPAWWFHDSAEGMR-----------------------------RF 396
gi|16125737|CC1490                LKLGPSWWFHDSPEGMM-----------------------------RF 397
gi|221234494|CCNA_01557           LKLGPSWWFHDSPEGMM-----------------------------RF 397
gi|167645720|Caul_1756            LKLGPSWWFHDSPEGMM-----------------------------RF 397
gi|197105913|PHZ_c2452            LKLGPAWWFHDSPEGMR-----------------------------RF 397
gi|84625827|XOO4170               LLLGPAWWFHDAPEGMW-----------------------------RF 399
gi|161898957|XOO4427              LLLGPAWWFHDAPEGMW-----------------------------RF 399
gi|188574617|PXO_03860            LLLGPAWWFHDAPEGMW-----------------------------RF 392
gi|21244968|XAC4251               LLLGPAWWFHDAPEGMW-----------------------------RF 399
gi|78049913|XCV4357               LLLGPAWWFHDAPEGMW-----------------------------RF 399
gi|66770505|XC4209                LLLGPAWWFHDAPEGMW-----------------------------RF 398
gi|21233539|XCC4117               LLLGPAWWFHDAPEGMW-----------------------------RF 398
gi|188993724|xccb100_4329         LLLGPAWWFHDAPEGMW-----------------------------RF 398
gi|114568782|Mmar100230           LKLGPAWWFHDSPEGMM-----------------------------RF 415
gi|148554803|Swit1887             LKLGPAWWFHDSPEGMR-----------------------------RF 393
gi|103488522|Sala3046             LKLGPAWWFHDSPEGMR-----------------------------RF 397
gi|90020587|Sde0940               LKLGPSWWFHDSPEGMR-----------------------------RF 397
gi|90020919|Sde1272               LKLGPAWWFHDSPEGMR-----------------------------RY 398
gi|116249865|RL0099               LMIGPPWWFHDSPLGIR-----------------------------RY 398
gi|209551594|Rleg2_4023           LMIGPPWWFHDSPLGIR-----------------------------RY 398
gi|86355751|RHECH00090            LMVGPPWWFHDSPLGIR-----------------------------RY 398
gi|190889727|RHECIAT_CH0000094    LMIGPPWWFHDSPLGIR-----------------------------RY 398
gi|119962615|AAur0527             VYIGAPWWFLDAPDAML-----------------------------RF 390
gi|220911562|Achl_0787            VFLGAPWWFLDAPDAML-----------------------------RF 401
gi|116669068|Arth0502             VYIGAPWWFLDAPDAML-----------------------------RF 394
gi|163842140|RSal332093421|gi|    MYLGAPWWFLDAPDAML-----------------------------RF 189
gi|158315003|Franean1_3197       VYVGAPWWFLDAPDSIR-----------------------------RF 418
```

```
gi|134101579|SACE5055          VYLGAPWWFLDAPRAMY----------------------------RF 425
gi|152964556|Krad_0587         VRLGAPWWFLDSPLAMR----------------------------RF 397
gi|50955310|Lxx17350           VYAGAPWWFLDTPAAIL----------------------------RY 450
gi|213691567|Blon_0669         LYIGAPWWFLDAPEPIL----------------------------RY 403
gi|50843764|PPA2329            LYIGVPWWFIDAPESVM----------------------------RF 397
gi|161485962|CE2377            AYVGAPWWFIDEIDAMN----------------------------RF 396
gi|145296483|cgR_2394          AYVGAPWWFIDEIDAMN----------------------------RF 399
gi|62391323|cg2731             NSKMARQNTRDILDAVSDGG----------------------VMLGRN 412
gi|161486711|NCgI2398          NSKMARQNTRDILDAVSDGG----------------------VMLGRN 412
gi|126208494|APL_1020          IQFGSGWWFNDQKDGME----------------------------RQ 123
gi|190150346|APP7_1077         IQFGSGWWFNDQKDGME----------------------------RQ 123
gi|165976445|APJL_1038         IQFGSGWWFNDQKDGME----------------------------RQ 123
gi|152977831|Asuc_0145         IQFGSGWWFNDQKDGME----------------------------RQ 394
gi|161511000|MS0544            IQFGSGWWFNDQKDGME----------------------------RQ 394
gi|148827203|CGSHiGG_02740     IQFGSGWWFNDQKDGME----------------------------RQ 394
gi|68248599|NTHI0056           IQFGSGWWFNDQKDGME----------------------------RQ 394
gi|113461738|HS_1602           IQFGSGWWFNDQKDGME----------------------------RQ 390
gi|170718551|HSM_0409          IQFGSGWWFNDQKDGME----------------------------RQ 394
gi|62181648|SC3078             VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|197249740|SeAg_B3309        VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|198243637|SeD_A3482         VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|207858396|SEN2980           VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|194734239|SeSA_A3319        VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|205354057|SG3031            VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|161616115|SPAB03917         VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|16766437|STM3137            VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|194448866|SeHA_C3383        VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|194446286|SNSL254_A3389     VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|16761915|STY3308            VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|56415079|SPA3005            VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|197364008|SSPA2803          VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|161506293|Sari04489         VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|37524198|plu0176            VQFGSGWWFNDQKDGML----------------------------RQ 395
gi|28901561|VPA1706            VQFGSGWWFNDQKDGMQ----------------------------RQ 394
gi|37677254|VV1594             VQFGSGWWFNDQKDGMQ----------------------------RQ 394
gi|27367460|VV21070            VQFGSGWWFNDQKDGMQ----------------------------RQ 394
gi|119943921|Ping0131          IQFGSGWWFNDQKDGMQ----------------------------RQ 399
gi|152996793|Mmwyl12780        VQFGSGWWFNDQLDGMQ----------------------------RQ 394
gi|117625400|APECO1_3327       MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|91212521|UTI89_C3530        MQFGSGWWFNDQKDGME----------------------------RQ 414
gi|16130987|b3092               MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|161486116|c3850              MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|215488422|E2348C_3385        MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|170082634|ECDH10B_3268       MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|157155304|EcE24377A_3560     MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|218691395|ECED1_3759         MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|157162569|EcHS_A3275         MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|218701863|ECIAI39_3589       MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|170018657|EcolC_0608         MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|110643336|ECP_3183           MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|218560176|ECS88_3488         MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|170683343|EcSMS35_3384       MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|89109861|                    MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|30064435|S3339               MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|82545347|SBO2953            MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|56480255|SF3132             MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|110806974|SFV3133           MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|74313639|SSON3245           MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|187731951|SbBS512_E3528     MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|209397829|ECH74115_4407     MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|15803633|Z4445               MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|15833228|ECs3974            MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|218696796|EC55989_3506      MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|209920564|ECSE_3373         MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|218706716|ECUMN_3576        MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|218551598|EFER_4383         MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|206579202|KPK_0591          MQFGSGWWFNDQKDGME----------------------------RQ 393
```

```
gi|152972036|KPN_03520          MQFGSGWWFNDQKDGME----------------------------RQ 389
gi|146313181|Ent638_3546        MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|157148666|CKO_04494          MQFGSGWWFNDQKDGME----------------------------RQ 414
gi|156935638|ESA_03503          MQFGSGWWFNDQKDGME----------------------------RQ 393
gi|161484751|y3600              MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|161511340|YP2899             MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|108809201|YPA3210            MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|162419999|YPAngolaA1104      MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|145597678|YPDSF0366          MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|170022805|YPK_0554           MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|108810613|YPN0448            MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|16120908|YPO0579             MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|153949672|YpsIP317580491     MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|51597770|YPTB3478            MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|186896961|YPTS_3663          MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|123443892|YE3708             MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|157372554|Spro4321           MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|108810612|YPN0447            --------------------------------------------------
gi|50119594|ECA0645             MQFGSGWWFNDQKDGMQ----------------------------RQ 393
gi|54303188|PBPRB1515           IQFGSGWWFNDQKDGMV----------------------------RQ 329
gi|60681862|BF1904G             IQFGSGWWFLDQKDGME----------------------------RQ 393
gi|53713582|BF2293              IQFGSGWWFLDQKDGME----------------------------RQ 393
gi|161511193|BT0823             IQFGSGWWFLDQKDGME----------------------------KQ 393
gi|150005586|BVU3074            IQFGSGWWFLDQKDGME----------------------------KQ 393
gi|212550592|CFPG_235           IQFGAGWWFLDQKMGME----------------------------AQ 394
gi|150009769|BDI3184            MQYGAAWWFLDQIKGME----------------------------DQ 457
gi|146301998|Fjoh_4262          VQFGSGWWFLDQKDGMT----------------------------KQ 393
gi|192359962|CJA_2772           IQFGSGWWFLDQMEGMT----------------------------RQ 393
gi|182416100|Oter_4293          MQFGSGWWFLDQKEAME----------------------------WQ 390
gi|116624408|acid5330           VQFGSGWWFLDQKEAMQ----------------------------WQ 391
gi|225621141|BHWA1_02239        MQLGSAWWFCDNRDGME----------------------------EQ 394
gi|18309134|CPE0152             IQFGSGWWFNDQKDGME----------------------------RQ 392
gi|160880875|Cphy_2743          IQQGSAWWFNDNKTGMI----------------------------DQ 392
gi|189485070|TGRD_067           IQQGAAWWFNDHKPGIT----------------------------EQ 392
gi|56962769|ABC0995             IQFGSGWWFNDTKDGME----------------------------KQ 396
gi|222529391|Athe_1402          MQLGAAWWFNDSKDGNI----------------------------QQ 390
gi|146296954|Csac1949           MQLGAAWWFNDSKDGNL----------------------------QQ 390
gi|150018894|Cbei_4082          MQFGAAWWFNDNKDGMI----------------------------EQ 393
gi|150019467|Cbei_4657          MQFGAAWWFNDNKDGMI----------------------------EQ 396
gi|20808343|TTE1939             MQFGAAWWFNDNKDGIK----------------------------EQ 392
gi|15893980|CAC0692             IQFGAAWWFNDHRDGMV----------------------------EQ 391
gi|150016705|Cbei_1832          IQFGSAWWFNDHKIGMQ----------------------------DQ 392
gi|52081820|BL00708             IQHGTAWWFNDTKDGML----------------------------EQ 391
gi|52787207|BLi03516            IQHGTAWWFNDTKDGML----------------------------EQ 391
gi|157693739|BPUM2987           MQHGTAWWFNDTKQGMT----------------------------EQ 391
gi|154685671|RBAM012380         IQFGTAWWFNDTKDGML----------------------------QQ 391
gi|16078295|BSU12300            IQFGTAWWFNDTKDGML----------------------------DQ 392
gi|23097822|OB0367              LQFGTAWWFNDTKSGMI----------------------------KQ 392
gi|15613268|Bh0705              VQFGSAWWFNDHIDGMR----------------------------RQ 392
gi|138895421|GTNG_1765          VQFGAAWWFNDHQDGII----------------------------RH 392
gi|220929098|Ccel_1676          LQLGAAWWFNDHRDGMV----------------------------NH 394
gi|81427739|LSA0129             LQLGAGWWFNDTAEGMT----------------------------QQ 396
gi|116332824|LVIS_0150          LQLGAGWWFNDTAEGMD----------------------------NQ 396
gi|116332809|LVIS_0135          MQLGCAWWFNDTREGMH----------------------------DQ 397
gi|116512444|LAACR_1742         LQLGAGWWFNDTAKGME----------------------------KQ 390
gi|125623707|llmg_0862          LQLGAGWWFNDTAKGME----------------------------KQ 390
gi|15673610|L0019               LQLGAGWWFNDTKKGME----------------------------NQ 390
gi|150389106|Amet1291           VQFGAGWWFNDTEQGML----------------------------RQ 400
gi|70727647|SH2648              VQHGAGWWFNDTKRGML----------------------------RQ 392
gi|25010729|gbs0674             LQFGAGWWFADTKLGMI----------------------------SQ 392
gi|22536865|SAG0701             LQFGAGWWFADTKLGMI----------------------------SQ 392
gi|76786751|SAK0827             LQFGAGWWFADTKLGMI----------------------------SQ 392
gi|195977823|Sez_0689           LQFGAGWWFADTKLGMI----------------------------SQ 392
gi|225868878|SZO_13120          LQFGAGWWFADTKLGMI----------------------------SQ 392
gi|225870118|SEQ_0716           LQFGAGWWFADTKLGMI----------------------------SQ 392
gi|94990680|MGAS10270_Spy1171   LQFGAGWWFADTKLGMI----------------------------SQ 401
gi|222153335|SUB1203            LQFGAGWWFADTKLGMI----------------------------SQ 392
```

```
gi|146318813|SSU051159|gi|1463     LQFGAGWWFNDTKLGMI------------------------------DQ 375
gi|146321021|SSU981174             LQFGAGWWFNDTKLGMI------------------------------DQ 404
gi|225859885|SP70585_2238          LQFGAGWWFNDTKLGMI------------------------------DQ 392
gi|157147928|CKO_03735             LQFGSGWWFNDTRRGMV------------------------------NQ 394
gi|220927672|Ccel_0214             VQFGAAWWFNDHKEGIT------------------------------AH 389
gi|78221672|Gmet0451               VQHGPPWWFLDQIYGIE------------------------------DH 392
gi|148269995|Tpet0860              VYVGAPWWFNDSPFGME------------------------------MH 378
gi|170288680|TRQ2_0883             VYVGAPWWFNDSPFGME------------------------------MH 378
gi|15642839|TM0064                 VYVGAPWWFNDSPFGME------------------------------MH 378
gi|160902320|Pmob0854              VSLGAPWWFNDSPFGME------------------------------IY 380
gi|157364144|Tlet1289              IFLGAPWWFNDSPFGMR------------------------------FQ 380
gi|20808989|TTE2645                ------------------------------------------------
gi|159042183|Cmaq1622             IYLNGYWWYSMYHDVIS------------------------------SY 345
gi|32474066|RB6095                 LDYGSTPSVEETQQVEIESDELESLDVDALPADEEVQELGVDDVELLDDS 531
gi|15613056|Bh0493                 LMIFGCWWFMNNPEIIN------------------------------E 336


gi|94971332|Acid345_4306           REHTTETAGFYNTA------GFNDDT---RAFLSIPGRHDMARRMDCAYL 431
gi|225874247|ACP_2685              REQVTETAGFYNTV------GFNDDT---RAFLSIPARHDVARRIDCAFL 431
gi|221639919|RSKD131_1820          REMTTETAGFYNTV------GFNDDT---RAFCSIPARHDVARRVDCAFL 436
gi|126462900|Rsph17029_2139        REMTTETAGFYNTV------GFNDDT---RAFCSIPARHDVARRVDCAFL 436
gi|146278446|Rsph170252412         REMTTETAGFMNTV------GFNDDT---RAFCSIPARHDVARRVDSAFL 436
gi|222149863|Avi_3892              RETTTETAGFYNTV------GFNDDT---RAFCSIPARHDVARRVDCAYL 434
gi|110680135|RD12934               RETTTETCGFYNTA------GFNDDT---RAFCSIPARHDVARRSDCAYL 436
gi|161620871|BCAN_B0804            RKLTTETAGFYNTV------GFNDDT---RAYLSIPARHDMARRVDCAYL 434
gi|23500543|BRA0812                RKLTTETAGFYNTV------GFNDDT---RAYLSIPARHDMARRVDCAYL 434
gi|163844934|BSUIS_B0804           RKLTTETAGFYNTV------GFNDDT---RAYLSIPARHDMARRVDCAYL 434
gi|17988821|BMEll0476|gi|17988     RKLTTETAGFYNTV------GFNDDT---RAYLSIPARHDMARRVDCAYL 444
gi|189022606|BAbS19_ll04000        RKLTTETAGFYNTV------GFNDDT---RAYLSIPARHDMARRVDCAYL 70
gi|158424966|AZC3342               RELTTETAGFYNTV------GFNDDT---RAFPSIPARHDVARRVDCAYL 439
gi|13473453|mll4056                REMTTETAGFYNTV------GFNDDT---RAFPSIPARHDVARRVDCAFL 437
gi|16264678|SMb21354               REMTTETAGFYNTV------GFNDDT---RAFPSIPARHDIARRVDCAFL 437
gi|16125737|CC1490                 REQVTETAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDSAFL 438
gi|221234494|CCNA_01557            REQVTETAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDSAFL 438
gi|167645720|Caul_1756             REQVTETAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDSAFL 438
gi|197105913|PHZ_c2452             REQTTETAGFYNTV------GFNDDT---RAFLSIPARHDVARRIDAGFL 438
gi|84625827|XOO4170                REQTLASAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDSAFL 440
gi|161898957|XOO4427               REQTLASAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDSAFL 440
gi|188574617|PXO_03860             REQTLASAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDSAFL 433
gi|21244968|XAC4251                REQTLASAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDSAFL 440
gi|78049913|XCV4357                REQTLASAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDSAFL 440
gi|66770505|XC4209                 REQTLASAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDSAFL 439
gi|21233539|XCC4117                REQTLASAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDSAFL 439
gi|188993724|xccb100_4329          REQTLASAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDSAFL 439
gi|114568782|Mmar100230            REQTTETAGFYNTV------GFNDDT---RAFLSIPARHDVARRVDCAFL 456
gi|148554803|Swit1887              RRQTTETAGFHNLV------GFNDDT---RAFLSIPARHDVARRIDCGYL 434
gi|103488522|Sala3046              RSQMTETAGFYNTV------GFNDDT---RAFLSIPARHDVARRIDCGFL 440
gi|90020587|Sde0940                REQVTETAGFYNTV------GFNDDT---RAFLSIPARHDVARRMDCGFL 438
gi|90020919|Sde1272                RQQITETAGFYNTV------GFNDDT---RAFLSIPARHDVARRMDCGFL 439
gi|116249865|RL0099                LDQVVETAGFANMA------GFNDDT---RALLSIPARHDVWRREVCRFL 439
gi|209551594|Rleg2_4023            LDQVVETAGFANMA------GFNDDT---RALLSIPARHDVWRREVCRFL 439
gi|86355751|RHECH00090             LDQVVETAGFANMA------GFNDDT---RALLSIPARHDVWRREVCRFL 439
gi|190889727|RHECIAT_CH0000094     LDQVVETAGFANMA------GFNDDT---RALLSIPARHDVWRREVCRFL 439
gi|119962615|AAur0527              RSAVTETAGFSRSS------GFIDDT---RAFCSIPARHDASRRIEASFL 431
gi|220911562|Achl_0787             RSAVTETAGFSRSS------GFIDDT---RAFCSIPARHDASRRIEASFL 442
gi|116669068|Arth0502              RAAITETAGFSRSS------GFIDDT---RAFCSIPARHDTSRRIESSFL 435
gi|163842140|RSal332093421|gi|     RSAVTETAGFSRSAGFSRSAGFVDDT---RTFCSIPARHDASRRVEASFL 236
gi|158315003|Franean1_3197         RRAVTETAGFTRTS------GFVDDT---RAFCSVPVRHDMSRRLDCGVL 459
gi|134101579|SACE5055              RDAVTETAGFYRNA------GFIDDT---RGFCSIPARHDTARRVDAAHL 466
gi|152964556|Krad_0587             REAVTETAGFYNTS------GFVDDT---RAFASIPARHDVSRRVDSGFL 438
gi|50955310|Lxx17350               RRAITDSAGVAKTS------GFIDDT---RAFCSIPTRHDMSRRVDAAFL 491
gi|213691567|Blon_0669             YEDVVPYAGFAKLS------GFIDDT---RALCSIPARHDMNRRLTARYI 444
gi|50843764|PPA2329                KHAVTEMAGFSRVS------GMIDDT---RAFCSIPARHDMSRRLDAAHL 438
gi|161485962|CE2377                RSLTTGTTGFSRYS------GFIDDT---RAYCSIPARHNTSRRVEANYL 437
gi|145296483|cgR_2394              RSATTGTTGFSRYS------GFIDDT---RAYCSIPARHNTSRRVEANYL 440
gi|62391323|cg2731                 GALVLGPVVGTLHIKFIAPLNKRVERVMYKTGLSEAA--------AAEQC 454
gi|161486711|NCgI2398              GALVLGPVVGTLHIKFIAPLNKRVERVMYKTGLSEAA--------AAEQC 454
```

```
gi|126208494|APL_1020              LQQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCEMI 163
gi|190150346|APP7_1077             LQQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCEMI 163
gi|165976445|APJL_1038             LQQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCEMI 163
gi|152977831|Asuc_0145             LQQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCEMV 434
gi|161511000|MS0544                LQQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCEMI 434
gi|148827203|CGSHiGG_02740         LQQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCEMI 434
gi|68248599|NTHI0056               LQQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCEMI 434
gi|113461738|HS_1602               LQQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCEMI 440
gi|170718551|HSM_0409              LQQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCEMI 434
gi|62181648|SC3078                 LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|197249740|SeAg_B3309            LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|198243637|SeD_A3482             LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|207858396|SEN2980               LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|194734239|SeSA_A3319            LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|205354057|SG3031                LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|161616115|SPAB03917             LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|16766437|STM3137                LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|194448866|SeHA_C3383            LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|194446286|SNSL254_A3389         LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|16761915|STY3308                LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|56415079|SPA3005                LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|197364008|SSPA2803              LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|161506293|Sari04489             LEQLSQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNLL 435
gi|37524198|plu0176                LEQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCNVL 435
gi|28901561|VPA1706                MEQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCDMV 434
gi|37677254|VV1594                 MEQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCDMV 434
gi|27367460|VV21070                MEQLSQLGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRILCDMV 434
gi|119943921|Ping0131              MQQLSQLGLLSKFV------GMLTDS---RSFLSY-TRHEYFRRILCNMV 439
gi|152996793|Mmwyl12780            LTSVAQMGLLSQFV------GMLTDS---RSFLSY-TRHEYFRRLLCEMI 434
gi|117625400|APECO1_3327           MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|91212521|UTI89_C3530            MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 454
gi|16130987|b3092                  MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|161486116|c3850                 MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|215488422|E2348C_3385           MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|170082634|ECDH10B_3268          MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|157155304|EcE24377A_3560        MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|218691395|ECED1_3759            MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|157162569|EcHS_A3275            MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|218701863|ECIAI39_3589          MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|170018657|EcolC_0608            MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|110643336|ECP_3183              MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|218560176|ECS88_3488            MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|170683343|EcSMS35_3384          MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|89109861|                       MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|30064435|S3339                  MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|82545347|SBO2953                MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|56480255|SF3132                 MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|110806974|SFV3133               MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|74313639|SSON3245               MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|187731951|SbBS512_E3528         MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|209397829|ECH74115_4407         MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|15803633|Z4445                  MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|15833228|ECs3974                MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|218696796|EC55989_3506          MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|209920564|ECSE_3373             MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|218706716|ECUMN_3576            MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|218551598|EFER_4383             MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|206579202|KPK_0591              MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|152972036|KPN_03520             MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 429
gi|146313181|Ent638_3546           MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|157148666|CKO_04494             MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 454
gi|156935638|ESA_03503             MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|161484751|y3600                 MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|161511340|YP2899                MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|108809201|YPA3210               MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|162419999|YPAngolaA1104         MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|145597678|YPDSF0366             MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
```

```
gi|170022805|YPK_0554              MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|108810613|YPN0448               MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|16120908|YPO0579                MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|153949672|YpsIP317580491        MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|51597770|YPTB3478               MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|186896961|YPTS_3663             MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|123443892|YE3708                MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|157372554|Spro4321              MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|108810612|YPN0447               --------------------------------------------------
gi|50119594|ECA0645                MTQLAQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRILCQMI 433
gi|54303188|PBPRB1515              MEQLSQLGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRLLCRLL 369
gi|60681862|BF1904G                LNALSLLGLLSRFV------GMLTDS---RSFLSY-PRHEYFRRTLCNLL 433
gi|53713582|BF2293                 LNALSLLGLLSRFV------GMLTDS---RSFLSY-PRHEYFRRTLCNLL 433
gi|161511193|BT0823                MNALSVLGLLSRFV------GMLTDS---RSFLSY-PRHEYFRRTLCNLV 433
gi|150005586|BVU3074               MNALSLLGLLSRFV------GMLTDS---RSFLSY-PRHEYFRRTLCNLL 433
gi|212550592|CFPG_235              MNSLSVLGLLSHFV------GMLTDS---RSFLSY-TRHEYFRRILCNLV 434
gi|150009769|BDI3184               LNALSSLGLLSRFV------GMLTDS---RSFLSY-PRHEYFRRILCNML 497
gi|146301998|Fjoh_4262             LNALSNMGLISCFV------GMLTDS---RSFLSF-PRHEYFRRILCNLL 433
gi|192359962|CJA_2772              IEALSQLGLLSRFV------GMLTDS---RSFLSY-SRHEYFRRILCGIF 433
gi|182416100|Oter_4293             MNALSNQGLLSRFV------GMLTDS---RSFLSY-TRHEYFRRTLCNLI 429
gi|116624408|acid5330              MNALSNCGMFSKFL------GMLTDS---RSFMSY-PRHEYFRRVLCDLF 431
gi|225621141|BHWA1_02239           IKILANTGSLALFV------GMLTDS---RSFLSY-SRHEYFRRILCNII 434
gi|18309134|CPE0152                MMALSQLGLISQFV------GMVTDS---RSFLSY-TRHEYFRRILCNYL 432
gi|160880875|Cphy_2743             MTSLANLGLLANFV------GMLTDS---RSFLSY-TRHEYFRRILCELI 432
gi|189485070|TGRD_067              LTSLANLGILSNFV------GMLTDS---RSFLSY-TRHEYFRRILCDLI 432
gi|56962769|ABC0995                MTDLANNGLLSLFV------GMLTDS---RSFLSY-TRHEYFRRILCNRI 436
gi|222529391|Athe_1402             MKDLANLGLLSRFV------GMVTDS---RSFLSY-ARHEYFRRLLCNLI 430
gi|146296954|Csac1949              MKDLANLGLLSRFV------GMVTDS---RSFLSY-ARHEYFRRLLCKLI 430
gi|150018894|Cbei_4082             MKALGNLGLLGRFV------GMLTDS---RSFLSY-TRHEYFRRIACNLI 433
gi|150019467|Cbei_4657             MKALANLGLLGRFV------GMLTDS---RSFLSY-TRHEYFRRIACNLI 436
gi|20808343|TTE1939                LKTLANVGLLGRFV------GMVTDS---RSFLSY-ARHEYFRRILCDLI 432
gi|15893980|CAC0692                METLANLGAFSTFI------GMLTDS---RSFLSY-TRHEYFRRILCDLI 431
gi|150016705|Cbei_1832             IRTLANLGMLSTFV------GMLTDS---RSFLSY-TRHEYFRRILCNVI 432
gi|52081820|BL00708                MKSLANIGLLSRFI------GMLTDS---RSFLSY-TRHEYFRRLLCDVI 431
gi|52787207|BLi03516               MKSLANIGLLSRFI------GMLTDS---RSFLSY-TRHEYFRRLLCDVI 431
gi|157693739|BPUM2987              MMTLSSIGLISRFI------GMLTDS---RSFLSY-TRHEYFRRLLCDII 431
gi|154685671|RBAM012380            MRALSNMGLFSRFI------GMLTDS---RSFLSY-PRHEYFRRLVCTLI 431
gi|16078295|BSU12300               MKALSNVGLFSRFI------GMLTDS---RSFLSY-TRHEYFRRIVCNLI 432
gi|23097822|OB0367                 MQTLADIGVFSQFI------GMLTDS---RSFLSY-PRHEYFRRLVCSLI 432
gi|15613268|Bh0705                 LTDLASVGLLSNFI------GMLTDS---RSFLSY-PRHDYFRRILCQLI 432
gi|138895421|GTNG_1765             LNDLANVGVFSTFV------GMLTDS---RSFLSY-VRHEYFRRIVCNLI 432
gi|220929098|Ccel_1676             MQTVANHGLLSGFI------GMLTDS---RSFLSY-TRHDYFRRILCDLI 434
gi|81427739|LSA0129                LQVFAQQSLLPHFV------GMLTDS---RSFLSY-PRHEYFRRVLCSFY 436
gi|116332824|LVIS_0150             LRIFAQESLLPNFV------GMLTDS---RSFLSY-PRHEYFRRVLCNFY 436
gi|116332809|LVIS_0135             LRIMAQQSLLPNFV------GMLTDS---RSFLSY-PRHEYFRRVLCDLI 437
gi|116512444|LAACR_1742            LEIFASQSLLPNFV------GMLTDS---RSFLSY-PRHEYFRRVLCNFV 430
gi|125623707|llmg_0862             LEIFASQSLLPNFV------GMLTDS---RSFLSY-PRHEYFRRVLCNFV 430
gi|15673610|L0019                  LEIFASQSLLSNFV------GMLTDS---RSFLSY-PRHEYFRRVLCNFV 430
gi|150389106|Amet1291              MSTLADHGLLMNFV------GMLTDS---RSFISY-PRHEYFRRILCNYV 440
gi|70727647|SH2648                 MSSLADQGLLMHFV------GMLTDS---RSFISY-SRHDYFRRILSSFI 432
gi|25010729|gbs0674                MNALAEQGMLANFI------GMLTDS---RSFLSY-QRHDYFRRILCTYL 432
gi|22536865|SAG0701                MNALAEQGMLANFI------GMLTDS---RSFLSY-QRHDYFRRILCTYL 432
gi|76786751|SAK0827                MNALAEQGMLANFI------GMLTDS---RSFLSY-QRHDYFRRILCTYL 432
gi|195977823|Sez_0689              MNALAEQGLLANFV------GMLTDS---RSFLSY-QRHDYFRRILSTYL 432
gi|225868878|SZO_13120             MNALAEQGLLANFV------GMLTDS---RSFLSY-QRHDYFRRILSTYL 432
gi|225870118|SEQ_0716              MNALAEQGLLANFV------GMLTDS---RSFLSY-QRHDYFRRILSTYL 432
gi|94990680|MGAS10270_Spy1171      MNALAEQGMLANFV------GMLTDS---RSFLSY-QRHDYFRRILSTYL 441
gi|222153335|SUB1203               MNALAEQGMLANFV------GMLTDS---RSFLSY-QRHDYFRRILCTYL 442
gi|146318813|SSU051159|gi|1463     MNAYAEQGMLANFV------GMLTDS---RSFLSY-QRHDYFRRILATYV 415
gi|146321021|SSU981174             MNAYAEQGMLANFV------GMLTDS---RSFLSY-QRHDYFRRILATYV 444
gi|225859885|SP70585_2238          MNAYAEQGMLANFV------GMLTDS---RSFLSY-QRHDYFRRILATYV 432
gi|157147928|CKO_03735             LNALADQGLLSNFI------GMLTDS---RSFVSY-TRHDYFRRILCDLI 434
gi|220927672|Ccel_0214             LKAIADQGMLAYFV------GMLTDS---RSFLSY-VRHDYFRRILCSFV 429
gi|78221672|Gmet0451               LNAQCAVGVLGTFV------GMASDS---RSFLST-ARHEYFRKAITLRL 432
gi|148269995|Tpet0860              LKYLASVDLLYNLA------GMVTDS---RKLLSFGSRTEMFRRVLSSVV 419
gi|170288680|TRQ2_0883             LKYLASVDLLYNLA------GMVTDS---RKLLSFGSRTEMFRRVLSNVV 419
gi|15642839|TM0064                 LKYLASVDLLYNLA------GMVTDS---RKLLSFGSRTEMFRRVLSNVV 419
```

```
gi|160902320|Pmob0854            LKYIATVDLLSDLA------GWVTDS---RKLISYGSRTEMFRRELSNVV 421
gi|157364144|Tlet1289            LQYIASVDLLSNFA------GMVTDS---RKLMSYGSRTEMFRRILCDVV 421
gi|20808989|TTE2645              --------------------------------------------------
gi|159042183|Cmaq1622           IRIRLQMLPYTKIG------GFFSDAY--VADWVYG-KVKLIKHVLALTL 386
gi|32474066|RB6095               VSSETSAEQLESLV---------DDEPLDVGNLLDD---ETPSEPITQLR 569
gi|15613056|Bh0493               MTRMRMEMLGTSFI------PQHSDA---RVLEQLIYKWHHSKSIIAEVL 377


gi|94971332|Acid345_4306         AKLVAEHRLAE---DEAYEVAHDLAYRLAKEAYRL--------------- 446
gi|225874247|ACP_2685            ARLVAEHRLSE---EEAAEVAHDLAYKLVKAAYRL--------------- 446
gi|221639919|RSKD131_1820        ATLVATGRLDA---EEAPEVAHDLAYRLAKEAYRL--------------- 451
gi|126462900|Rsph17029_2139      ATLVATGRLDA---EEAPEVAHDLAYRLAKEAYRL--------------- 451
gi|146278446|Rsph170252412       ATLVATGRLDE---DEAFEVAQDLAYRLAKQAYRL--------------- 451
gi|222149863|Avi_3892            AELVLTGRLEE---DEAQELAGDLAYGLAKKAYRL--------------- 449
gi|110680135|RD12934             ATLVRTGRLTE---AEALDLAYDLTYRLAKEAYRL--------------- 451
gi|161620871|BCAN_B0804          AGLVADHRLEE---DEAYEVAHDLAYRLAKQTYKL--------------- 449
gi|23500543|BRA0812              AGLVADHRLEE---DEAYEVAHDLAYRLAKQTYKL--------------- 449
gi|163844934|BSUIS_B0804         AGLVADHRLEE---DEAYEVAHDLAYRLAKQTYKL--------------- 449
gi|17988821|BMEll0476|gi|17988   AGLVADHRLEE---DEAYEVAHDLAYRLAKQTYKL--------------- 459
gi|189022606|BAbS19_ll04000      AGLVADHRLEE---DEAYEVAHDLAYRLAKQTYRL--------------- 85
gi|158424966|AZC3342             AELVTTHRLEE---DEAHEVAHDLAYRLAKKAYRL--------------- 454
gi|13473453|mll4056              ARLVAEHRLRE---DEAHELARELAHTLAKKAYRL--------------- 452
gi|16264678|SMb21354             ARLVAEHRLRE---DEAYELARDLAYGLAKEAYRL--------------- 452
gi|16125737|CC1490               ARMVAEHRMDL---VEAEELIVDLTYNLPKKAYKLDQRPDWARPA-TLRA 453
gi|221234494|CCNA_01557          ARMVAEHRMDL---VEAEELIVDLTYNLPKKAYKLDQRPDWARPA-TLRA 453
gi|167645720|Caul_1756           ARMVAEHRMDE---VEAAELIVDLTYTLPKRAYKLDQRPHWARPAKAARH 453
gi|197105913|PHZ_c2452           ARLVAEHRLSE---DDAHETAVDLAYNLPKRAYKLDE-P—ARTPAAA--- 453
gi|84625827|XOO4170              AKLVAEHRLEE---DEATEVAIDLAYRMPKRAYNL--------------- 455
gi|161898957|XOO4427             AKLVAEHRLEE---DEATEVAIDLAYRMPKRAYNL--------------- 455
gi|188574617|PXO_03860           AKLVAEHRLEE---DEATELAIDMAYRMPKRAYNL--------------- 448
gi|21244968|XAC4251              AKLVAEHRLEE---DEATEVAIDLAYRLPKQAYKLYKL------------ 455
gi|78049913|XCV4357              AKLVAEHRLEE---DEATEVAIDLAYRLPKQAYNL--------------- 455
gi|66770505|XC4209               AKLVAEHRLDE---DEAMEVAIDLAYRLPKQAYKL--------------- 454
gi|21233539|XCC4117              AKLVAEHRLDE---DEAMEVAIDLAYRLPKQAYKL--------------- 454
gi|188993724|xccb100_4329        AKLVAEHRLDE---DEAMEVAIDLAYRLPKQAYKL--------------- 454
gi|114568782|Mmar100230          ARLVAEHRLGL---DEAHETARDLAYNLAKRAYKLDAQN----------- 471
gi|148554803|Swit1887            AELVVEHQLDE---DEAADLAVDLTYRLVKSAYRL--------------- 449
gi|103488522|Sala3046            AQLVSEHRLEE---WEAAELAADLSYNLAKASYKL--------------- 453
gi|90020587|Sde0940              AKLVSEHRLQE---DEAREVAIDLTYNLVKQAYKLD-------------- 453
gi|90020919|Sde1272              AELVADHRLSE---NDAHELAVDLTYNLVKKAYKL--------------- 454
gi|116249865|RL0099              AQLAGEHRLSK---REAEIVAGELSYGNAKKAYKL--------------- 454
gi|209551594|Rleg2_4023          AGLAAEHRISK---KEAEIVAGELSYGNAKKAYKL--------------- 454
gi|86355751|RHECH00090           AGLAAEHRISK---KEAEIVARELSYDNAKKAYKL--------------- 454
gi|190889727|RHECIAT_CH0000094   AALAAEHRISK---RDAEIVAGELSYGNAKKAYKL--------------- 454
gi|119962615|AAur0527            ARLVAEHRISE---DRAHELIVDVVDSSPRRVFKL--------------- 446
gi|220911562|Achl_0787           ARLVAEHRVTE---ERAHELIVDIVDGSPRRVFKL--------------- 457
gi|116669068|Arth0502            ARLVAEHRISE---ERARELIVDIVDSSPRRVFKL--------------- 450
gi|163842140|RSal332093421|gi|   ARLVAEHRLAE---SRAAEIIVDLVDAASRRVFKL--------------- 251
gi|158315003|Franean1_3197       AELVAEHRLEE---DEAFETAVELVAVRPREAFRL--------------- 474
gi|134101579|SACE5055            AGLVAEHVLSE---DDAALVIHDLHDRQPRKAFRL--------------- 481
gi|152964556|Krad_0587           ATLVASGQLSL---EEAVETAVDLTTTVPRTAYAPRGA------------ 453
gi|50955310|Lxx17350             ASLVVSHQLGG---GGCFWGCWAVGG------------------------ 500
gi|213691567|Blon_0669           SGLVADHRLSY---EEGEQIAIRSVDGQPSDVFKL--------------- 459
gi|50843764|PPA2329              AELVVLGRLDL---DEAVEIAHRLVVEQPTQVFGL--------------- 453
gi|161485962|CE2377              ARLVAEHRISE---TRASEIIVDLIDASPRRVFKL--------------- 452
gi|145296483|cgR_2394            ARLVAEHRITE---SRAAEIIVDLIDASPRRVFKL--------------- 455
gi|62391323|cg2731               A-L--EDRLRE---EMAHALYQWNPGRDENYDLVINTGSMTYEQIVDLVV 504
gi|161486711|NCgI2398            A-L--EDRLRE---EMAHALYQWNPGRDENYDLVINTGSMTYEQIVDLVV 504
gi|126208494|APL_1020            GGWVVKGEAPN-DIELLGNMVKNICYHNAKSYFK---------------- 180
gi|190150346|APP7_1077           GGWVVKGEAPN-DIELLGNMVKNICYHNAKSYFK---------------- 180
gi|165976445|APJL_1038           GGWVVKGEAPN-DIELLGNMVKNICYHNAKSYFK---------------- 180
gi|152977831|Asuc_0145           GKWVVNGEAPN-DMNLLGNMVKNICYNNAKSYFK---------------- 451
gi|161511000|MS0544              GRWVVNGEAPN-DMNLLGNMVKNICFDNAKAYFK---------------- 451
gi|148827203|CGSHiGG_02740       GGWVEKGEAPN-DISLLGKMIEDICFNNAKNYFK---------------- 451
gi|68248599|NTHI0056             GGWVEKGEAPN-DISLLGKMIEDICFNNAKNYFK---------------- 451
gi|113461738|HS_1602             GGWVERGEAPN-DLNLLGKMVKDICYDNAKRYFK---------------- 447
gi|170718551|HSM_0409            GGWVERGEAPN-DLNLLGKMVKDICYDNAKRYFK---------------- 451
```

```
gi|62181648|SC3078                GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|197249740|SeAg_B3309           GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|198243637|SeD_A3482            GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|207858396|SEN2980              GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|194734239|SeSA_A3319           GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|205354057|SG3031               GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|161616115|SPAB03917            GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|16766437|STM3137               GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|194448866|SeHA_C3383           GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|194446286|SNSL254_A3389        GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|16761915|STY3308               GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|56415079|SPA3005               GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|197364008|SSPA2803             GQWAQDGEIPD-DEAMLSRMVQDICFNNAQRYFTIK---------- 452
gi|161506293|Sari04489            GQWAQDGEIPD-DEAMLSRMVQDICFNNAQHYFAIK---------- 452
gi|37524198|plu0176               GQWAEKGEIPN-HKAMLGQIVEDICFNNARRYFALPGENL------ 452
gi|28901561|VPA1706               GRWAENGEVPN-DLSLLGPMVEDICFGNAKRYFEERA---------- 451
gi|37677254|VV1594                GRWAENGEVPN-DLSLLGPMVEDICFGNAKRYFEERA---------- 451
gi|27367460|VV21070               GRWAENGEVPN-DLSLLGPMVEDICFGNAKRYFEERV---------- 451
gi|119943921|Ping0131             GQWVEDGEAPN-DIAMLGKLIQDVAFNNAKTYFKLPGEQ------ 456
gi|152996793|Mmwyl12780           GGWVERGEAPA-DMALLGEMVKGICAGNAKKYFGF---------- 451
gi|117625400|APECO1_3327          GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|91212521|UTI89_C3530           GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 471
gi|16130987|b3092                 GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|161486116|c3850                GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|215488422|E2348C_3385          GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|170082634|ECDH10B_3268         GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|157155304|EcE24377A_3560       GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|218691395|ECED1_3759           GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|157162569|EcHS_A3275           GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|218701863|ECIAI39_3589         GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|170018657|EcolC_0608           GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|110643336|ECP_3183             GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|218560176|ECS88_3488           GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|170683343|EcSMS35_3384         GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|89109861|                      GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|30064435|S3339                 GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|82545347|SBO2953               GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|56480255|SF3132                GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|110806974|SFV3133              GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|74313639|SSON3245              GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|187731951|SbBS512_E3528        GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|209397829|ECH74115_4407        GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|15803633|Z4445                 GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|15833228|ECs3974               GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|218696796|EC55989_3506         GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|209920564|ECSE_3373            GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|218706716|ECUMN_3576           GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|218551598|EFER_4383            GRWVEAGEAPA-DINLLGEMVKNICFNNARDYFAIELN------- 450
gi|206579202|KPK_0591             GRWVAAGEAPA-DIALLGEMVKNICFNNARDYFAIELN------- 450
gi|152972036|KPN_03520            GRWVAAGEAPA-DIALLGEMVKNICFNNARDYFAIELN------- 446
gi|146313181|Ent638_3546          GRWVAAGEAPA-DIQLLGEMVKNICFNNARDYFAIELN------- 450
gi|157148666|CKO_04494            GRWVNAGEAPA-DIQLLGEMVRNICFNNARDYFAIELN------- 471
gi|156935638|ESA_03503            GRWVEAGEAPA-DIDLLGEMVKNICFNNARDYFAIELN------- 450
gi|161484751|y3600                GRWVADGEAPA-DIALLGAMVKNICFDNAQQYFAIEL-------- 450
gi|161511340|YP2899               GRWVADGEAPA-DIALLGAMVKNICFDNAQQYFAIEL-------- 450
gi|108809201|YPA3210              GRWVADGEAPA-DIALLGAMVKNICFDNAQQYFAIEL-------- 450
gi|162419999|YPAngolaA1104        GRWVADGEAPA-DIALLGAMVKNICFDNAQQYFAIEL-------- 450
gi|145597678|YPDSF0366            GRWVADGEAPA-DIALLGAMVKNICFDNAQQYFAIEL-------- 450
gi|170022805|YPK_0554             GRWVADGEAPA-DIALLGAMVKNICFDNAQQYFAIEL-------- 450
gi|108810613|YPN0448              GRWVADGEAPA-DIALLGAMVKNICFDNAQQYFAIEL-------- 450
gi|16120908|YPO0579               GRWVADGEAPA-DIALLGAMVKNICFDNAQQYFAIEL-------- 450
gi|153949672|YpsIP317580491       GRWVADGEAPA-DIALLGAMVKNICFDNAQQYFAIEL-------- 450
gi|51597770|YPTB3478              GRWVADGEAPA-DIALLGAMVKNICFDNAQQYFAIEL-------- 450
gi|186896961|YPTS_3663            GRWVADGEAPA-DIALLGAMVKNICFDNAQQYFAIEL-------- 450
gi|123443892|YE3708               GRWVEDGEAPA-DIELLGSMVKNICFDNAKQYFAIEL-------- 450
gi|157372554|Spro4321             GGWVENGEAPA-DITLLGEMVKNICFDNAKNYFAIELA------- 450
gi|108810612|YPN0447              ---------------------------------------------
```

```
gi|50119594|ECA0645                GRWVEDGEAPA-DLPLLGEMVKNISFDNAKNYFAIEL------------ 450
gi|54303188|PBPRB1515              GQWVADGEVPH-DMTLLSKMVEDICFNNANNYFNLGLEK----------- 486
gi|60681862|BF1904G                GCDVENGEIPLSEMERVCQMVEDISYFNAKNFFHF--------------- 451
gi|53713582|BF2293                 GCDVENGEIPLSEMERVCQMVEDISYFNAKNFFHF--------------- 451
gi|161511193|BT0823                GRDVENGEIPVSEMDRVNQMIEDISYNNAKNFFKF--------------- 451
gi|150005586|BVU3074               GNDVENGEIPACEIERVNQMVEDICYNNAKKFFQF--------------- 451
gi|212550592|CFPG_235              GNDVEQGLLPVSELSFLGQLVENISYYNAKKYFGF--------------- 452
gi|150009769|BDI3184               GNEIEKGLLPASELSFIGQMVEDISYNNAKRYFDF--------------- 515
gi|146301998|Fjoh_4262             GDEIKRGELPN-DMEWIGKLVADISYNNAKEYFKF--------------- 450
gi|192359962|CJA_2772              GRDMVKGLVPD-DTHMVGKMIQDISFNNAKNYFPFVVPE----------- 450
gi|182416100|Oter_4293             GAEMERGEIPN-DRELVGPMVRRICFANAREYFRLELDPSFRG------- 447
gi|116624408|acid5330              GRDVENGELPD-NDELIGPVIRDICYGNAQRFLGLNV------------- 448
gi|225621141|BHWA1_02239           GEWADKGEVPN-DIKYLGSIIENICFNNSNIYFNN--------------- 451
gi|18309134|CPE0152                GNLVESGQYPY-EEEILGEIVQNICYKNSAKYFKR--------------- 449
gi|160880875|Cphy_2743             GNWVENGEYPN-DIEFLGQMVQDISYYNTKRYFGF--------------- 449
gi|189485070|TGRD_067              GSWVENGEYPY-DVKILSKIVSDISYDNAVRYFKFKL------------- 449
gi|56962769|ABC0995                GEWVERGEWPA-DEKWLGKVVEDISYYNAKRYFAFPK------------- 453
gi|222529391|Athe_1402             GEWVENGEYPY-DLETLGRIVQDICYYNAKEYFGF--------------- 448
gi|146296954|Csac1949              GEWVENGEYPY-DLEALSRIVQGVCYYNAKEYFGF--------------- 447
gi|150018894|Cbei_4082             GEWVENGEVPK-NDKLLKRIVQGICYSNAKEYFGFDEK------------ 450
gi|150019467|Cbei_4657             GEWVENGEVPG-NIKLLKTVVEGICYNNAKEYFNI--------------- 453
gi|20808343|TTE1939                GEWVENGEVPN-DIELLGKIVQDISFNNAKEYMGV--------------- 449
gi|15893980|CAC0692                GKWVENGEVPN-DMELLGRITKNICFNNANNYFEMGL------------- 448
gi|150016705|Cbei_1832             GEWVENGELPN-DIEHLGQIVADISYNNADKYFNM--------------- 449
gi|52081820|BL00708                GTWVENGEAPD-DIELLGRIVKGICYENAKHYFQFEVKDRLKA------- 448
gi|52787207|BLi03516               GTWVENGEAPD-DIELLGRIVKGICYENAKHYFQFEVKDRLKA------- 448
gi|157693739|BPUM2987              GDWVEKGEVPY-DLELLGEIVKGISYENAKQYFQFDRVKQLHHQSKIT-- 448
gi|154685671|RBAM012380            GGWAEQGEAPY-DMELLGKIVEGICYRNAEEYFRF--------------- 448
gi|16078295|BSU12300               GEWVENGEVPR-DMELLGSIVQGICYDNAKHYFQFQEEKANV-------- 449
gi|23097822|OB0367                 GEWVHNGEVPY-DLKSLGEIVQDISYYNAARYFDFGLLSDE--------- 449
gi|15613268|Bh0705                 GSWIKEGQLPP-DMERWGQIVQDICYNNVVDYFDMKETVRL--------- 449
gi|138895421|GTNG_1765             GSWIEKGEAPQ-DYAFLGKIVQDICYFNAKQYFDLS-------------- 449
gi|220929098|Ccel_1676             GEWVEHEQVPL-ELDMLGSMVKDICFNNSVRYFGLKLENEVQENKRCYNE 451
gi|81427739|LSA0129                GRLVEQGRVPN-DAAELGRVVQNIAYNNAHDYFDFF-------------- 453
gi|116332824|LVIS_0150             GKLAEQGRVPN-DEAALGQIVQDISYNNAYKFFGFFPDKTPDELFKTTAK 453
gi|116332809|LVIS_0135             GEWVVRGQVPD-DDAYLGKIVEDISYNNAHDYFGFFDQK----------- 454
gi|116512444|LAACR_1742            GQLIESGRIPD-DEALVGKMIEDISYNNVHDYFGFFKNEDSKQ------- 447
gi|125623707|llmg_0862             GQLIESGRIPD-DEALVGKMIEDISYNNVHDYFGFFKNEDSKQ------- 447
gi|15673610|L0019                  GQLIENGRIPD-NEALVGKMIEDISYNNVHDYFGFFKDEDSNQ------- 447
gi|150389106|Amet1291              GVQVENGKFPN-DEELLKRLIENICYYNAVNYFTKK-------------- 457
gi|70727647|SH2648                 GDLVEKGEIPN-DDQLLKRMIENICYNNAYNYFKLI-------------- 449
gi|25010729|gbs0674                GEWIEEGEVPE-DYQALGSMAKDIAYQNAVNYFKN--------------- 449
gi|22536865|SAG0701                GEWIEEGEVPE-DYQALGSMAKDIAYQNAVNYFKN--------------- 449
gi|76786751|SAK0827                GEWIEEGEVPE-DYQALGSMAKDIAYQNAVNYFKN--------------- 449
gi|195977823|Sez_0689              GEWIEEGEVPE-DYQVLGSMAKDIAYNNAIQYFS---------------- 449
gi|225868878|SZO_13120             GEWIEEGEVPE-DYQALGSMAKDIAYNNAIQYFS---------------- 449
gi|225870118|SEQ_0716              GEWIEEGEVPE-DYQALGKMAQDIAYNNAIQYFN---------------- 449
gi|94990680|MGAS10270_Spy1171      GEWIEEGEVPE-DYQAIGQMAKDIAYDNAVRYFN---------------- 458
gi|222153335|SUB1203               GEWIEEGEVPE-DYEALGHMAKDIAYHNAVNYFKHE-------------- 449
gi|146318813|SSU051159|gi|1463     GQWIVDEEVPE-DYNRLGQFVEAISYYNAKEFFEQ--------------- 432
gi|146321021|SSU981174             GQWIVDEEVPE-DYNRLGQFVEAISYYNAKEFFEQ--------------- 461
gi|225859885|SP70585_2238          GQWIVDEEVPE-DYDRLGQFVEAISYYNAKEFFEQ--------------- 449
gi|157147928|CKO_03735             GGWVERGEVPE-DDAILNNMIRGICVDNARRYFRFR-------------- 451
gi|220927672|Ccel_0214             GDLVEKGEFAD-DDAILGEIVEGICFKNIVRYLGL--------------- 446
gi|78221672|Gmet0451               ADDLESGLFNP-DQVDIDEFCRDLFYANAQSFFGWKVAHGNS-------- 449
gi|148269995|Tpet0860              GEMVERGQIPI---KEARELVKHVSYDGPKALFFG--------------- 434
gi|170288680|TRQ2_0883             GEMVEKGQIPI---KEARELVKHVSYDGPKSLFFR--------------- 434
gi|15642839|TM0064                 GEMVEKGQIPI---KEARELVKHVSYDGPKALFFG--------------- 434
gi|160902320|Pmob0854              GEMVEKGQIPI---REAFDLVRDISYFRPKRLFFEKI------------- 436
gi|157364144|Tlet1289              GEMVDRGQIPF---KEALELCNELCYFRPKELFFSRKR------------ 436
gi|20808989|TTE2645                ------------------------------------------------
gi|159042183|Cmaq1622             DDAVKAGYLSS---DLAIEIAEAMLYSNQLELYGIKA------------- 401
gi|32474066|RB6095                 GESSFSPDDDSLQLEADPNTGEFTLPVGDDDDDDDHVIITESDDSHSSSD 619
gi|15613056|Bh0493                 IDKYDDILQAGWEVTEEEIKRDVADLFSRNFWRFVGRNDHVTSVKVEQQT 394
```

VITA

Name: Tinh T. Nguyen

Address: Department of Chemistry
Texas A&M University
PO Box 30012
College Station, TX 77842-3012

Email Address: tnguyen@mail.chem.tamu.edu

Education: B.S., Chemistry, Mary Washington College, 2004
Ph.D., Chemistry, Texas A&M University, 2009