THE STRUCTURE AND MECHANISM OF BACTERIAL DIHYDROOROTASE

A Dissertation

by

TAMIKO NEAL PORTER

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 2004

Major Subject: Chemistry

THE STRUCTURE AND MECHANISM OF BACTERIAL DIHYDROOROTASE

A Dissertation

by

TAMIKO NEAL PORTER

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

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Frank Raushel (Chair of Committee)	Gregory Reinhart (Member)
Victoria DeRose (Member)	Paul Lindahl (Member)
Emile Schweikert (Head of Department)	

December 2004

Major Subject: Chemistry

ABSTRACT

The Structure and Mechanism of Bacterial Dihydroorotase. (December 2004)
Tamiko Neal Porter, B.S., Michigan State University
Chair Advisory Committee: Dr. Frank Raushel

Dihydroorotase (DHO) is a zinc metallo-enzyme that functions in the pathway for the biosynthesis of pyrimidine nucleotides by catalyzing the reversible interconversion of carbamoyl aspartate and dihydroorotate. The X-ray crystal structure of the enzyme was obtained at a resolution of 1.7 Å. The pH-rate profiles for the hydrolysis of dihydroorotate or thio-dihydroorotate demonstrated that a single group of DHO must be unprotonated for maximal catalytic activity. The pH-rate profiles for the condensation of carbamoyl aspartate to dihydroorotate showed that a single group from the enzyme must be protonated for maximal catalytic activity. The native zinc ions within the active site of DHO were substituted with cobalt or CADmium by reconstitution of the apo-enzyme with divalent cations. The ionizations observed in the pH-rate profiles were dependent on the specific metal ion bound to the active site. Mutation of Asp-250 resulted in the loss of catalytic activity. These results are consistent with the formation of a hydroxide bridge between the two divalent cations that functions as the nucleophile during the hydrolysis of dihydroorotate. In addition, Asp-250 is postulated to shuttle the proton from the bridging hydroxide to the leaving group amide during dihydroorotate hydrolysis. The X-ray crystal structure of DHO showed that the side-chain carboxylate of dihydroorotate is electrostatically interacting with Arg-20, Asn-44 and His-254. Mutation of these residues resulted in the loss of catalytic activity, indicating that these residues are critical for substrate recognition. The thio-analog of dihydroorotate, (TDO) was found to be a substrate of DHO. A comprehensive chemical mechanism for DHO was proposed based on the experimental data presented in this dissertation. Armed with this understanding of the structure-function relationship of DHO, a rational approach was used to alter the substrate specificity of the enzyme. The R20/N44/H254 mutant of DHO was obtained and found to have increased activity on dihydrouracil compared to the wild-type enzyme. The sequence of the gene PA5541 from *Pseudomonas aeruginosa* has a glutamine at a position where most active DHO proteins have a histidine residue. Results from the characterization of PA5541 indicate that it is a functional DHO.

DEDICATION

To my parents, Fred and Catherine Neal, who’ve always been my biggest cheerleaders; to my husband, Christopher, who is my best friend and supporter; and to our wonderful sons, Nickolas and Donovan.

ACKNOWLEDGEMENTS

I thank my adviser, Dr. Frank Raushel, for his guidance and support throughout my graduate career and my committee members: Dr. Gregory Reinhart, Dr. Pual Lindahl and Dr. Victoria DeRose. I would also like to thank Dr. Paul Fitzpatrick. Also, many thanks go to all of my past and present labmates for their friendship and assistance through the years. Finally, it is very important to thank Dr. Hazel Holden and Jim Thoden of the University of Wisconsin at Madison for their X-ray crystallography work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

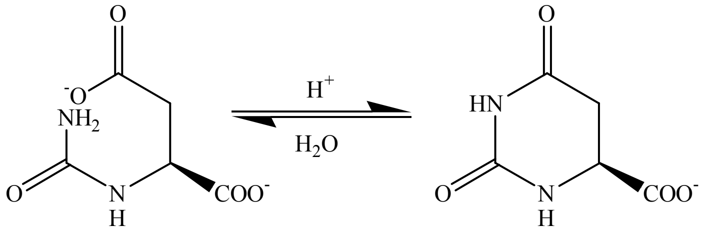
CHAPTER I

INTRODUCTION

In protein science the term “superfamily” has been invoked to describe a group of enzymes that have conserved tertiary structure but have varying catalytic activities due to divergent or convergent evolution. A specific example of an enzyme superfamily is the amidohydrolase group of metalloproteins. The enzymes of the amidohydrolase superfamily adopt the $(\beta/\alpha)_8$-barrel fold. The superfamily was first described by Holm and Sander based on structural alignments of three enzymes, adenosine deaminase, phosphotriesterase and urease (*1*). The superfamily was expanded based on amino acid sequence identity using database searches. With a few exceptions, members of the amidohydrolase superfamily conserve five active-site residues found at the C-terminal end of the β-strands. These residues are four histidines and an aspartic acid. Members of the superfamily bind either one or two metal ions per active-site. Since the original Holm and Sander description of the superfamily, the knowledge base has expanded as there are to date at least eighteen additional X-ray structures of amidohydrolase superfamily members. Analysis of the structure-function relationship of members of the superfamily will increase our knowledge of how Nature has evolved these enzymes. The more that is understood about how Nature has evolved protein function will better our ability to engineer proteins to capitalize on their usefulness for biotechnology and drug therapies.

This dissertation follows the style of *Biochemistry.*

Dihydroorotase (DHO) is a member of the amidohydrolase superfamily. DHO functions in the pyrimidine biosynthesis pathway by catalyzing the reversible cyclization of N-carbamoyl aspartate to form L-dihydroorotate and a molecule of water. The reaction is shown in Scheme 1.1. Extensive analyses of the primary structure of DHO clearly show that there are two classes of the enzyme (*2*). DHO proteins comprising Class I are found in higher organisms and are much larger than their Class II counterparts that are found in many bacteria and fungi. A Class I DHO is expressed in CAD, a multi-functional enzyme complex found in mammals, insects and molds, that consists of the first three enzymes of the pyrimidine biosynthesis pathway, carbamoyl phospate synthetase (CPS), aspartate transcarbamoylase (ATC) and DHO (*3, 4*). Monofunctional Class I DHO enzymes are also found in gram-positive bacterial strains. Class II enzymes consist of monofunctional proteins from gram-negative bacteria and yeast. Table 1.1 shows which class the known DHO sequences fall into. Class I proteins typically have subunit molecular weights of ~45 kDa compared to ~38 kDa for the Class II proteins.



Scheme 1.1

Within each class, the primary sequence identity is quite high (>40%), but poor sequence identity (<20%) is seen when one compares proteins between classes. An alignment of the human and *E. coli* proteins is shown in Figure 1.1. The two proteins

Table 1.1: DHO Classification[a]

Class I	g.i. number	Class II	g.i. number
Aquifex aeolicus	3914514	*Agrobacterium tumefaciens*	15887748
Archaeoglobus fulgidus	3914507	*Arabidopsis thaliana*	15235865
Bacillus anthracis	21401873	*Azotobacter vinelandii*	23103669
Bacillus caldolyticus	1172784	*Beta vulgaris*	17977977
Bacillus halodurans	10175158	*Bordetella pertussis*	33566830
Bacillus subtilis	2633923	*Bordetella bronchioseptica*	33603376
Bifidobacterium longum	23465368	*Buchnera aphidicola*	15616940
Caenorhabditidis elegans	17532699	*Burkholderia fungorum*	22988480
Chlorobium tepidum	21673868	*Campylobacter jejuni*	15791630
Clostridium acetobuytlicum	23821990	*Chloroflexus aurantiacus*	22971316
Corynebacterium glutamicum	19552823	*Emericella nidulans*	1870226
Deinococcus radiodurans	9087183	*Escherichia coli*	16129025
Desulfovibrio desulfuricans	23473834	*Helicobacter pylori*	15645206
Dictyostelium discoideum	400909	*Microbulbifer degradans*	23026307
Drosophila melanogaster	391583	*Neisseria meningitidis*	15676580
Enterococcus faecalis	22992871	*Nitrosomonas europaea*	22954697
Fusobacterium nucleatum	19703762	*Novosphingobium aromaticivoans*	23107613
Homo sapiens	33989959	*Oryza sativa*	14209529
Lactobacillus leichmannii	1346930	*Plasmodium falciparum*	23509919
Lactobacillus gasseri	23002207	*Plasmodium yoelii yoelii*	23479248
Leuconostoc mesenteroides	23024490	*Pseudomonas fluorescens*	23063090
Listeria innocua	21542203	*Pseudomonas syringae pv.*	23471210
Listeria monocytogenes	21542200	*Pseudomonas aeruginosa*	15598723
Mesocricetus auratus	131696	*Ralstonia solanacearum*	17545206
Methanococcus jannaschii	3122658	*Ralstonia metallidurans*	22978905
Mycobacterium bovis	31792575	*Rhodobacter sphaeroides*	22957385
Mycobacterium leprae	21542206	*Saccharomyces cerevisiae*	6323452
Mycobacterium tuberculosis	9087175	*Salmonella typhimurium*	16764519
Oceanobacillus iheyensis	23098944	*Salmonella enterica*	16760039
Oenococcus oeni	23037312	*Schizosaccharomyces pombe*	19115483
Pyrococcus abyssi	9087185	*Shewanella oneidensis*	24375192
Pyrococcus horikoshii	3914511	*Shigella flexneri*	24112473
Rattus rattus	2506025	*Sinorhizobium meliloti*	15964233
Squalus acanthias	3024509	*Synechocystis* sp.	16329543
Staphylococcus aureus	21204251	*Toxoplasma gondii*	21309874
Streptococcus mutans	24379638	*Ustilago maydis*	400913
Streptococcus pneumoniae	15901032	*Vibrio cholerae*	21542209
Streptococcus pyogenes	15674928	*Wigglesworthia brevipalpis*	24323648
Streptomyces coelicor	21219987	*Yersinia pestis*	16121855
Sulfolobus acidocaldaricus	34978387		

Table 1.1: continued

Class I	g.i. number	Class II	g.i. number
Thermatoga maritime	9087186		
Thermoanaerobacter tengcongensis	20807969		
Thermobifida fusca	23017985		
Thermus aquaticus	3122654		
Thermus thermophilus	46198734		
Trypanosoma cruzi	7494090		

[a] The sequences were all obtained from the database at NCBI (http://www.ncbi.nih.gov)

```
HUMAN    --TSQKLVRLPGLIDVHVHLREPGGTHKEDFASGTAAALAGGITMVCAM  TRPPIIDGPALALAQKLA
E. coli MTAPSQVLKIRRPD W L L D-----GDMLKTVVPYTSEI-YGRAIV   LAPPVTTVEAAVAYRQRI

HUMAN    EAGARCDFAL-FLGASSE------NAGTLGTVAGSAAGLKLYLNETFSELRL-DSVVQWMEHFETWPSH
E. coli LDAVPAGHDFTPLMTCYLTDSLDPNELERGFNEGVFTAAKLYPANATTNSSHGVTSIDAIMPVLERMEK

HUMAN    --LPIVAHAEQQTVAAVLMVAQLTQ--------------RSVHICHVARKEEILLIKAAKARGLPVTCE
E. coli IGMPLLVHGEVTHADIDIFDREARFIESVMEPLRQRLTALKVVFEHITTKDAADYVRDGN---ERLAAT

HUMAN    VAPHHLFLSHDDLERLGP-GKGEVRPELGSRQDVEALWEDMA---VIDCFASDHAPHTLEELPLKCGSR
E. coli ITPQHLMFNRNHMLVGGVRPHLYCLPILKRNIHQQALRELVASGFNRVFLGTDSAPHARHR---KESSC

HUMAN    PPPG-FPGLETMLLTAVSEGRLSLDDLLQRLHHNPRRIFHLPPQEDTYVEVDLEHEWTIPSHMPFSKAH
E. coli GCAGCFNAPTALGSYATVFEEMNALQHFEAFCSVNGPQFYGLPVNDTFIELVREEQQVAESIALTDDTL

HUMAN    WTPFEGQKVKGTVRRVVL
E. coli VPFLAGETVRWSVKK
```

Figure 1.1: Alignment of human and *E. coli* DHO sequences. The shaded residues are those that are conserved among an alignment of DHO sequences from over 30 organisms.

have a 17 % sequence identity. An alignment of DHO sequences from 39 bacterial sources is provided in Appendix I. Each class is represented in this master alignment. There are 17 conserved residues among DHO sequences and these are listed in Table 1.2. Generally, the conserved metal-binding HXH motif of the amidohydrolase superfamily is found in an DXHXHXR (residues 14-20 in *E. coli*) sequence. There are several bacterial organisms that contain sequences where the HXH motif is replaced by QXH (*2*). It is not known whether these enzymes are active. Among DHO proteins, there are also three additional conserved histidine residues (residues His-77, His-139 and His-254 in *E. coli*), the first two corresponding to the third and fourth conserved residues of the amidohydrolase superfamily. The conserved aspartate residue of the amidohydrolase superfamily is found in all DHO proteins. This aspartate (Asp-250) is found in a DXAPH motif.

In several organisms there are multiple DHO sequences. There are three situations where this is observed: organisms have an inactive DHO domain associated with a CAD-like multi-functional protein; organisms have an inactive DHO domain associated with ATC; or in addition to an active DHO with a HXH motif, the organism contains a DHO with a QXH motif. In *Saccharomyces cerevisiae* (yeast) and *Neurospora crassa,* a psuedo-CAD is found. While the enzyme has functional CPS and ATC domains, the DHO domain (encoded by ura2) is inactive (*5*). The sequence of the defective ura2 gene product contains several mutations of residues that are now known to be important for metal binding. The functional DHO protein in yeast is a Class I protein, is monofunctional and encoded by the gene ura4 (*6*). Shown in Figure 1.2 is a representation of the various forms of DHO.

Similar to DHO, there is more than one class of prokaryotic ATC proteins (*7*). In Class A, the catalytic subunit, pyrB, of ATC is associated with an active (pyrC) or

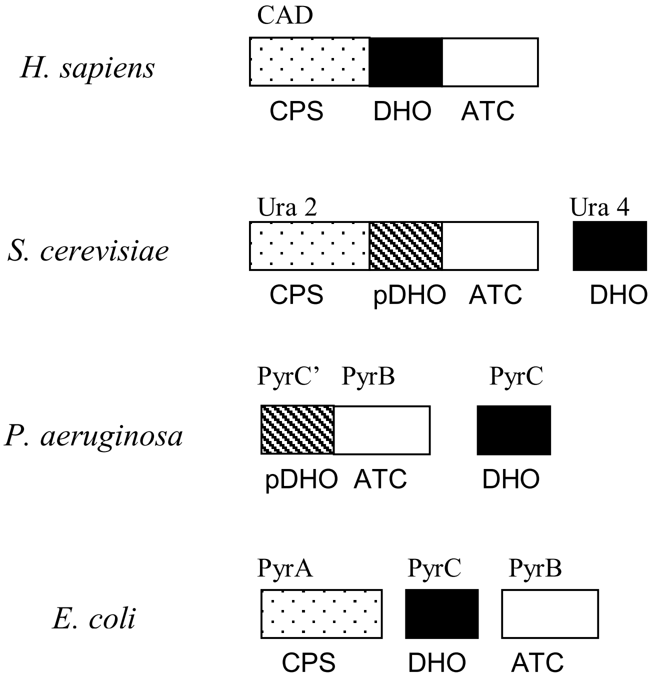


Figure 1.2: Organization of the genes for CPS, DHO and ATC in various organisms.

Table 1.2: Conserved Residues of DHO[a].

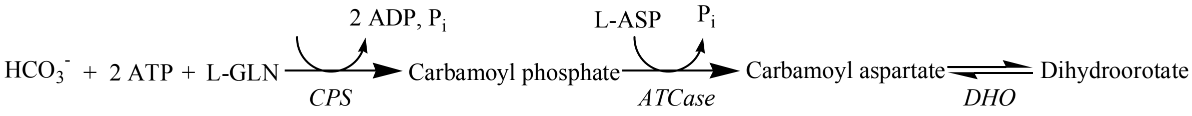
Residue	Probable Role
Asp-14	Hydrogen bonds to H17
His-16	Metal ligand
His-18	Metal ligand
Arg-20	Substrate-binding
Met-42	Unclear
Pro-43	Structural
Asn-44	Substrate-binding
Lys-102	Metal ligand
His-139	Metal ligand
His-177	Metal ligand
His-202	Unclear
Pro-223	Structural
Asp-250	Metal ligand and catalysis
Ala-252	Unclear
Pro-253	Structural
His-254	Substrate-binding
Lys-259	Unclear

[a] The numbers presented are those for the *E. coli* enzyme.

inactive (pyrC') regulatory domain (*8, 9*). The DHO domain, pyrC- or pyrC'-encoded, is believed to be required for assembly of the pyrB gene product into a stable dodecameric holoenzyme (*8, 10*). Class B ATC enzymes are associated with pyrIregulatory subunits. In this class, the holoenzyme is a dodecamer formed from two catalytic trimers and three dimers of pyrI-encoded regulatory subunits. The third class of ATC enzymes, Class C, consists of proteins that are trimers of only the catalytic domain (*11*).

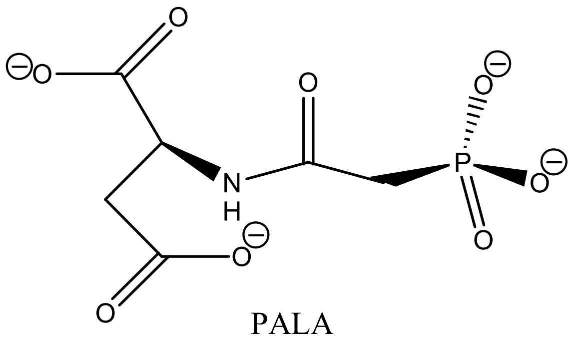
Pseudomonas aeruginosa and *Pseudomonas flourescens* contain three DHO-like sequences: an inactive DHO, a functional enzyme and an uncharacterized protein (*2, 12*). The inactive 45 kDa DHO protein of *P. aeruginosa* is associated with a 36 kDa ATC, which forms a dodecameric complex with a molecular weight of 486 kDa. The holoenzyme consists of six ATC catalytic chains (a dimer of trimers) and six pseudo-DHO (pDHO) chains (a trimer of dimers) (*12*). The ATC is encoded by pyrB while the gene pyrC' codes for pDHO. Vickrey *et al.* reported that plasmids containing only the gene for the ATC catalytic subunit did not complement *E. coli* strains deficient in ATC. The expressed protein was not soluble and was found in inclusion bodies. This suggested that pDHO is required for correct folding of the catalytic ATC domains. Similar ATC complexes are found in *P. putida, P. flourescens* and *Helicobacter pylori* (*8, 13*). In these systems, the inactive proteins are Class I proteins. In *Streptomyces griseus,* however, the ATC is associated with an active DHO (*9*). The functional DHO proteins of *P. aeruginosa* and *P. flourescens* are Class II proteins and are encoded by the gene pyrC.

In several organisms, including *P. aeruginosa, P. flourescens,* and *Xylella fastidiosa*, there is an uncharacterized protein that appears to be a Class I DHO that is 60 amino acids longer than sequences from organisms such as *E. coli*. These sequences all contain the unusual QXH motif discussed above but contain all of the perceived residue requirements for DHO activity. This is in contrast to URA2 of yeast or the previously described pDHO, which do not have the necessary conserved residues.



Scheme 1.2

The regions of the gene encoding CAD are organized in the following order: pyr1 (CPS), pyr3 (DHO) and pyr 2 (ATC). The gene from humans is very large, with over 6,679 base pairs (*4*). The trifunctional enzyme consists of 2,225 amino acids and is 243 kDa. The reactions catalyzed by this enzyme are shown in Scheme 1.2. The CPS domain of CAD is the rate-limiting step in the pyrimidine synthesis pathway (*3, 14*). CPS is allosterically activated by adenine triphosphate (ATP) or 5-phosphoribosyl-1-pyrophosphate (PRPP) and is feedback inhibited by uridine and cytidine nucleotides. It has been shown in *Drosophila melanogaster* that mutation in the CPS domain of CAD results in loss of feedback inhibition and thus an accumulation of large amounts of uridine triphosphate (UTP) (*15*). CAD is also regulated by Myc-dependent transcriptional changes, phosphorylation and caspase-mediated degradation. Myc is a family of transcriptional regulators involved in regulation of cell proliferation. Myc expression peaks in dividing cells and is suppressed as cells differentiate or exit the cell cycle. Myc has been shown to have oncogenic potential and is overexpressed in many human tumors (*16, 17*). In human and hamster CAD it was demonstrated that Myc binds to the E-box region of the CAD promoter increasing CAD transcription, which led to cell proliferation (*18*). When Myc was mutated, growth-factor stimulated increases of CAD promoter activity was blocked. Thus CAD expression is dependent on Myc and intact E-box sequences (*19, 20*). Myc-dependent increase in CAD expression is one mechanism that activity of this enzyme is increased in tumor cells. CAD is also regulated by reversible phosphorylation by mitogen-activated protein kinase (MAPK) in response to growth factors. When CPS is treated with MAPK, a loss of feedback inhibition by UTP and a gain of allosteric activation by PRPP is seen (*21*). CAD is also phosphorylated by cyclic AMP-dependent protein kinase (PKA)(*22*). Unlike what is found in mammalian CAD, there are no phosphroylation sites in ura2 (*23*). This suggests that CAD and ura2 regulation is different. Caspases, proteases that are activated during apoptosis, have been shown to target CAD (*24*). CAD is rapidly degraded in the early onset of apoptosis. Two caspase-3 sites have been located in CAD. Cleavage at these sites results in the loss of CPS activity and PRPP activation.



PALA

Scheme 1.3

An inhibitor of CAD is N-(phosphononacetyl)-L-aspartic acid (PALA). PALA, shown in Scheme 1.3, is a transition state analog inhibitor of ATC with a K_i of 27 nM (*25*). PALA was found to block the proliferation of mammalian cells in culture and thus is studied in clinical trials of various cancers alone or in combination with other compounds (*26, 27*). The effect of PALA on CAD is diminished as carbamoyl phosphate accumulates in the reaction mixture (*28*). The effect of PALA is also prevented if sufficient uridine is present (*27*).

Substrate channeling in CAD has been observed (*28-30*). For the hamster enzyme, experiments were conducted using ^{14}C-labeled HCO_3^- to initiate the reaction and allow for the determination of the concentration of the intermediates (*28, 29*). When dihydroorotate dehydrogenase was used to trap dihydroorotate as orotate, there was no carbamoyl phosphate accumulation in the assay mixture and carbamoyl aspartate was found at a concentration that was 12% of the value expected if the enzymatic activities were unlinked (*29*). The rate of dihydroorotate formation was 4 times the rate that would be attained from an unlinked system. Simply stated, the carbamoyl phosphate produced from the CPS reaction is channeled directly to the ATC active site where it is used as a substrate followed by the partial channeling of carbamoyl aspartate to the active site of DHO.

Early mechanistic studies of DHO were done on the hamster enzyme. The gene was first isolated from hamster by Kelly in 1986 (*31*). Unfortunately, the 46 kDa gene product did not have DHO activity. Later, other researchers found that there were additional residues at the C-terminal end required for activity (*32-34*). This discovery allowed for the successful subcloning of DHO into a cloning vector for expression in *E. coli*. An early study was the demonstration of inactivation of the enzyme by treatment with diethylpyrocarbonate (DEPC). This implicated one or more histidine residues in catalysis. The metal dependence of DHO was reported to be a single zinc ion. Using site-directed mutagenesis and $^{65}Zn^{2+}$-binding studies, Christopherson proposed that DHO had the same metal coordination geometry as carbonic anhydrase (CA). In CA, the single active site zinc ion adopts a tetrahedral geometry (Kannan, 1984). From the X-ray crystal structure of CA, the metal ligands have been determined to be three histidine residues and a hydroxide (*35*). The model proposed by Christopherson for DHO had three histidine residues assigned as metal ligands, His-16, His-18, and His-177 (numbers given are those that correspond to the *E. coli* enzyme) (*36*). A fourth active site histidine (His-139) played a role in catalysis. Mutation of the histidine residues to alanine, glycine or asparagine resulted in inactive enzymes. In CA, when the active-site metal is varied, the pK_a of the attacking hydroxide is altered (*37*). If a water molecule or a hydroxide were the ionizing group, one would expect that the acidity of this group would vary as the metal involved is altered. Christopherson did not find this when he substituted the native zinc for other divalent cations and determined the apparent pK_a values (*38*). The activity trend reported for the metal-substituted enzymes was $Co^{2+}>Zn^{2+}>Mn^{2+}>Cd^{2+}$. From the pH rate profile for the hydrolytic activity of the wild-type enzyme, he obtained an apparent kinetic pK_a on k_{cat} of 6.6 for the ionizable group involved in catalysis (*39*). Christopherson found that the apparent pK_a was relatively unchanged (6.6-6.9) when the active site metal was substituted. Also, in contradiction to what is found in CA, the visible absorption spectrum of the Co^{2+}-incorporated hamster DHO was consistent with penta-coordination (*38*). These findings indicated that CA may not be the appropriate model system for DHO.

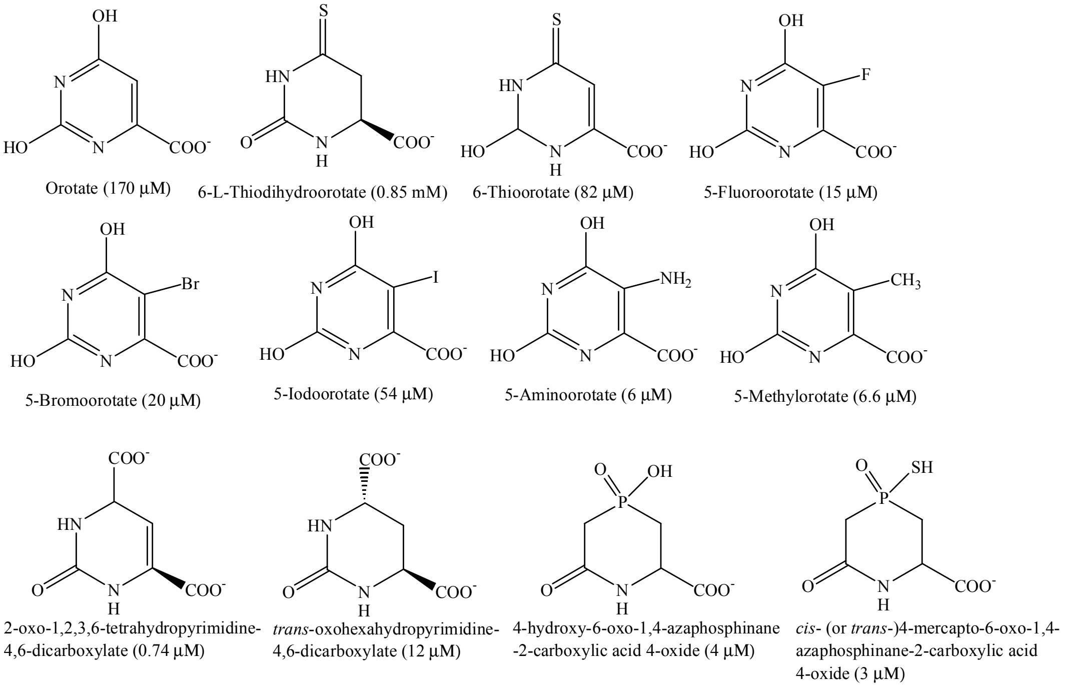
Several inhibitors of the mammalian enzyme have been described including cysteine and orotate analogs(*28, 36, 40-43*). Scheme 1.4 shows the most potent inhibitors with their K_i values for the hamster enzyme. These inhibitors were designed for possible anti-tumor or anti-malarial uses although none are used clinically. Two potent transition-state analogues of mammalian DHO have also been described, borocarbmoylethyl aspartate (K_i = 5 μM, pH 6) and *cis*-4-carboxy-6-(mercaptomethyl)-3,4,5,6-tetrahydropyrimidine-2(1H)-one (K_i = 140 nM, pH 7.4, 8.5) (*44, 45*). None of the inhibitors described for the hamster enzyme inhibited the bacterial enzyme (*36, 41*).

The explanation for this is unclear, although one would suppose that the difference is due to subtle perturbations in the active site or in how the substrate is recognized for binding.

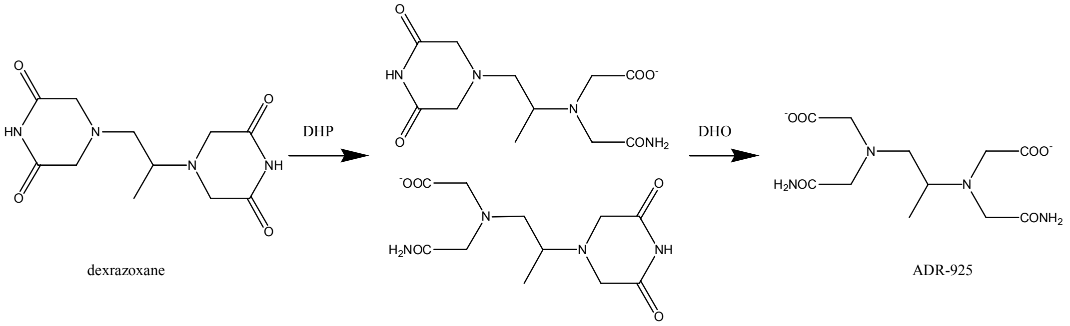
DHO has been implicated in the hydrolysis of the cardioprotective drug dexrazoxane [ICRF-187; (+)-1,2-bis(3,5-dioxopiperazin-1-yl)propane] (*46*). Cleaveage of dexrazoxane results in an active metal-binding compound ADR-925 [N,N'-[1S)-1-methyl-1,2-ethanediyl]bis[N-(2-amino-2-oxoethyl)glycine]. Scheme 1.5 illustrates how DHO, in conjunction with dihydropyrimidinase, cleaves dexrazoxane. ADR-925 reduces the cardiotoxicity of the anticancer drug doxrubicin by preventing oxygen radical formation by binding free iron or removing iron from the iron-doxrubicin complex (*47*). After dihydropyrimdinase acts on dexrazoxane to produce the two single ring containing metabolites B and C, DHO cleaves them to produce ADR-925 with a V_{max} that is 11- and 27- fold greater than that for the hydrolysis of dihydroorotate (*46*). The K_m values are 240- and 550- fold higher than that for dihydroorotate.

The DHO isolated from *E. coli* is homodimeric with a molecular weight of 76 kDa. It was also thought to bind a single molecule of zinc (*48*). It was reported by Collins that the cation could be substituted with cobalt to yield a more active enzyme (*48*). As found for the hamster enzyme, the Co^{2+}-substituted enzyme was more active than the native enzyme. Characterization of the enzyme was complicated by the fact that the enzyme was very unstable due to air oxidation (*49, 50*). Collins chemically modified the cysteines in the enzyme to conduct the experiments. A mass spectrometry and chemical derivatization experiment by Daniel proposed that two of the six cysteines, (C221 and C265), in the enzyme were responsible for the air oxidation (*51*).

Two weak noncompetitive inhibitors have been described for the bacterial enzyme, 4-chlorobenzenesulfonamide ($K_i = 200\ \mu M$) and 4-nitrobenzenesulfonamide



Scheme 1.4



Scheme 1.5

(K_i = 1.1 mM) (*52*). In the study where dexrazoxane was reported as a substrate for DHO, 4-chlorobenzesulfonamide was reported to not inhibit the mammalian enzyme (*46*). However, furosemide, a sulfonamide derivative with a carboxylate group, did inhibit the mammalian enzyme by 80 %.

In this dissertation a detailed study of the structure and mechanism of the *E. coli* DHO is presented. The X-ray crystal structure of the enzyme was determined. The structure allowed for the proposal of a novel mechanism for DHO. The mechanism of action and role of individual amino acids in binding and/or catalysis and the rate-limiting steps in substrate turnover were probed using mutagenesis and kinetic experiments. The roles of the metal ions in catalysis were studied by substituting the active site zinc for other divalent cations.

CHAPTER II

THE X-RAY CRYSTAL STRUCTURE OF DIHYDROOROTASE FROM *ESCHERICHIA COLI*

Dihydroorotase (DHO) catalyzes the reversible cyclization of N-carbamoyl-L-aspartate to form L-dihydroorotate in the *de novo* pyrimidine biosynthetic pathway. In bacteria and fungi, DHO is a monofunctional enzyme (*2, 49, 53*). In most eukaryotes, DHO is found in a multi-functional protein (CAD) that is composed of the first three enzymes of the de novo pyrimidine biosynthesis pathway (*3, 33*). CAD consists of carbamoyl phosphate synthetase (CPS), aspartate transcarbamoylase (ATC) and dihydroorotase (DHO). DHO from *E. coli* is a monofunctional, homodimeric zinc-containing metalloenzyme with a subunit molecular weight of 38,300 Daltons (*49*).

The amidohydrolase superfamily consists of metalloenzymes that all adopt a “TIM” barrel fold (*1*). Members of this superfamily generally catalyze hydrolysis reactions at a carbon or phosphorus center. An additional characteristic shared by most members of the amidohydrolase superfamily is that there are five conserved residues found at the C-terminus of β-strands 1, 5, 6, and 8. These residues include four histidines and an aspartic acid. The first two histidines are found in a metal-binding HxH motif. The members of the amidohydrolase superfamily are organized into two subgroups according to the number of metal ions that each enzyme coordinates per subunit. Phosphotriesterase (PTE) and urease (URE) contain two molecules of zinc and nickel respectively (*54, 55*). Most members of this subgroup contain a conserved carboxylated lysine residue that bridges the two metal ions found at the C-terminus of β-strand 4. A few enzymes, like the bacterial PTE homology protein, contain a glutamate residue at this position instead (*56*). The five conserved residues of the amidohydrolase superfamily are all metal ligands in this subgroup. A key representative of the subgroup of enzymes coordinating a single metal ion is adenosine deaminase (ADA) (*1, 55*). In adenosine deaminase, three of the histidines coordinate to the zinc ion, while the fourth conserved histidine serves as a general base in the active site.

Holm and Sander placed DHO in the adenosine deaminase subfamily based on previous assertions that DHO bound only a single zinc ion (*1*). Metal reconstitution experiments in the hamster and *E. coli* enzymes supported this belief (*36, 38, 48*). Also, the carboxylated lysine residue was not identified in primary sequence alignments. Here, the X-ray crystal structure of DHO is presented. Surprisingly, there was a binuclear metal center in the active site with a caboxylated lysine bridging the two zinc cations. Additionally, dihydroorotate was bound in one subunit of the crystallized dimer, while carbamoyl aspartate was bound in the other subunit. This discovery allowed for the proposal of a novel mechanism for the enzyme.

Materials and Methods

The gene for DHO, pyrC, was cloned from *E. coli* and ligated into the pBS$^+$ expression vector from Stratagene. The recombinant protein was expressed in *E. coli*. DHO was purified using a modified procedure adapted from Collins (*49*). The seleno-methionine derivative of DHO was also prepared. The seleno-methionine was incorporated using a method that inhibits the pathway used by cells to produce methionine (*57, 58*). This method requires that the cells are grown in minimal media. MALDI mass spectra of the native and derivative proteins were obtained to confirm the incorporation of the seleno-methionine. The mass spectroscopy was done by the

Laboratory for Biological Mass Spectrometry at Texas A & M University. The crystal structure was solved by our collaborators from the University of Wisconsin (*59*).

Results

Structure of DHO. The X-ray crystal structure of native DHO isolated from *E. coli* was obtained at a resolution of 1.7 Å. Contrary to expectations, each subunit of the dimer coordinated two ions of zinc. The overall dimensions for the dimer are 73 Å x 77 Å x 80 Å and it is shown in Figure 2.1. The two active sites are 25 Å apart. Each subunit of the dimer contains the conserved "TIM" barrel motif of the amidohydrolase superfamily. The eight β-strands of the "TIM" barrel are composed of residues Asp14 – His-18, Arg-38 – Val-41, Thr-73 – Leu-80, Phe-98 – Leu-130, Leu-136 – His-139, Lys-172 – Phe-175, Ala-195 – Ile-198 and Val-245 – Leu-247. In addition to a "TIM" barrel motif, each subunit of the dimer contains five extra β-strands, three α-helices and various type I, II and III turns. The dimensions of each subunit are 51 Å x 51 Å x 54 Å. The monomer is shown in Figure 2.2.

Subunit Interface. Three loop regions, His-144 – Arg-152, Arg-207 – Arg-227 and Asp-260 – Gly-264, define the interface between the two subunits. The surface lost by dimerization is approximately 2,300 $Å^2$ and the two subunits overlay with a root mean square deviation of 0.22 Å. The overlay of the α-carbons of each subunit is shown in Figure 2.3. The only movement seen is in loop 4, which is defined by residues Pro-

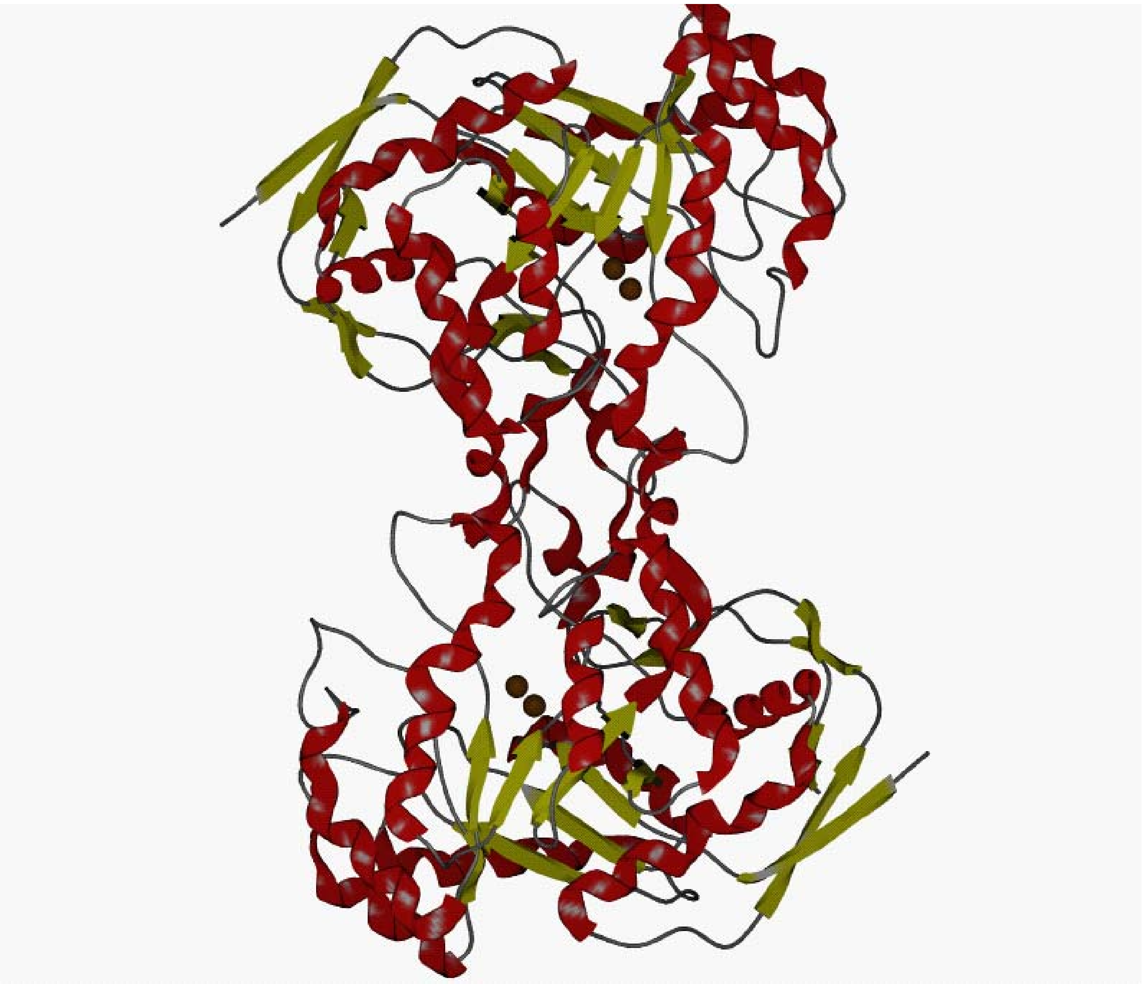


Figure 2.1: Ribbon representation showing the distribution of secondary elements in the DHO dimer (PDB 1J79). The strands are shown in yellow and the helices are in red. The overall dimensions of the dimer are 73 Å x 77 Å x 80 Å. The active sites are 25 Å apart.

Figure 2.2: Representation of the TIM-barrel of the DHO monomer (PDB 1J79). In addition to the core TIM barrel, there are five additional β-strands and three α-helices. The binuclear metal center and the ligands His-16, His-18, Lys-102, His-139, His-177 and Asp-250 are highlighted.

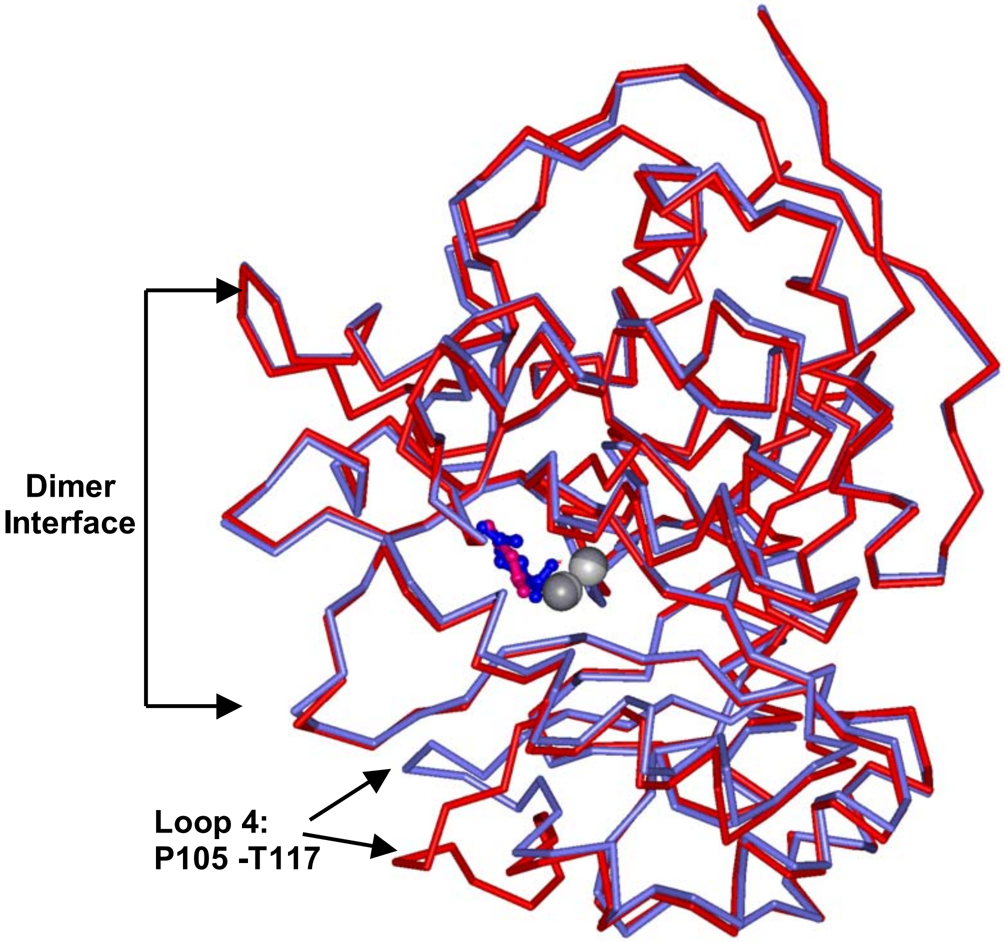


Figure 2.3: Overlay of the α-carbon backbone of the two subunits of DHO (PDB 1J79). Subunit I coordinating dihydroorotate is shown in blue while Subunit II with bound carbamoyl aspartate is shown in red/pink. The only movement between the two subunits appears to be in loop 4.

105 – Thr-117. The subunit interface is stabilized by two salt bridges between Asp-148 (subunit I) and Arg-207 (subunit II) and between Asp-151 (subunit I) and Arg-227 (subunit II). Hydrogen bonds are also formed between the side chain carboxamide group of Asn-208 (subunit II) and the backbone carbonyl oxygen of Ile-147 (subunit I) and the peptidic nitrogen of Ile-149 (subunit I).

Active Site of DHO. The crystallized dimer of DHO contained bound substrate and product. The overlay of the active sites of the two subunits is shown in Figure 2.4. In the active site of subunit I, dihydroorotate is bound as shown in Figure 2.5A. The location of the active site residues in the secondary structure is highlighted in Figure 2.6. The two metal ions are separated by 3.5 Å and are bridged by the carboxylated Lys-102 and a solvent molecule (believed to be a hydroxide ion). The more buried α-metal ion is coordinated to His-16, His-18, Lys-102, Asp-250 and the bridging solvent molecule. The α-metal ion is in a distorted trigonal bipyrimidal coordination sphere. The more solvent-exposed β-metal ion is distorted tetrahedrally and is coordinated by Lys-102, His-139, His-177 and the bridging solvent molecule. The carbonyl oxygen (O4) of dihydroorotate of the formed amide is coordinated to the β-metal ion at a distance of 2.9 Å and is 2.7 Å from the solvent molecule. N3 of dihydroorotate is 3.1 Å from the backbone carbonyl oxygen of Leu-222 and 2.9 Å from the bridging solvent molecule. O2 of dihydroorotate is coordinated to the backbone amide of Leu-222 with a distance of 2.7 Å, while N1 of dihydroorotate is 3.0 Å from the backbone carbonyl group of Ala-266. The carboxylate side chain of C6 appears to be stabilized by interactions with Arg-20, Asn-44 and His-254. Additionally, there are seven water molecules within 4.5 Å of the bound dihydroorotate. The active site of subunit II of DHO contained bound carbamoyl aspartate as shown in Figure 2.5B. The bridging active site water has been

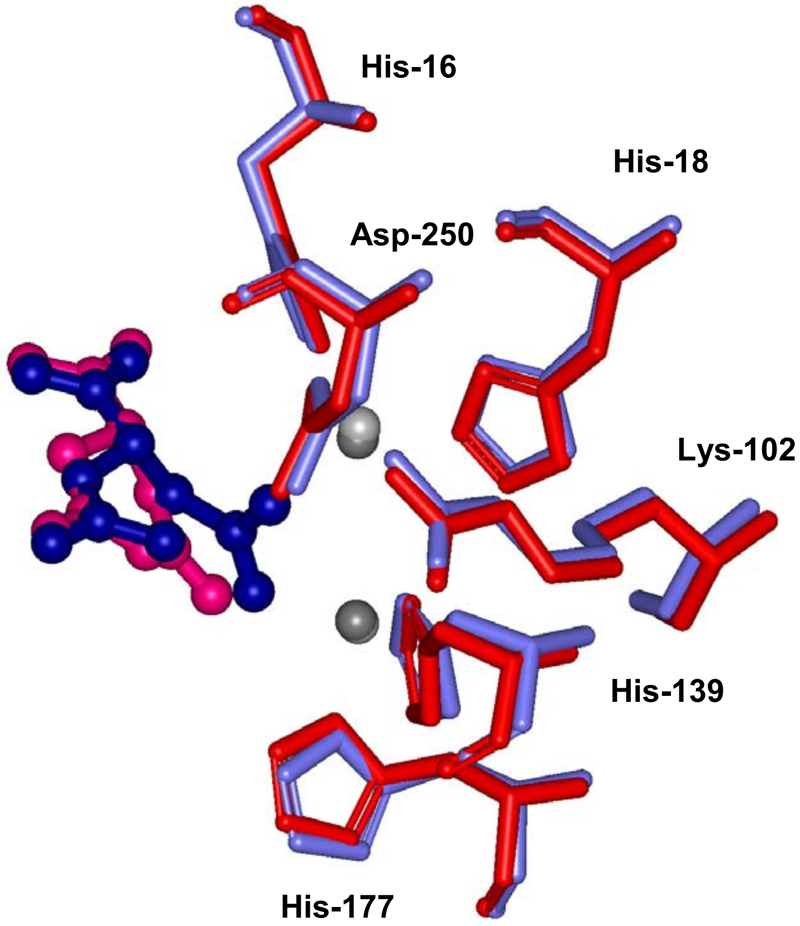


Figure 2.4: Overlay of the active sites of the two subunits of DHO (PDB 1J79). Subunit I coordinating dihydroorotate is shown in blue while Subunit II with bound carbamoyl aspartate is shown in red/pink.

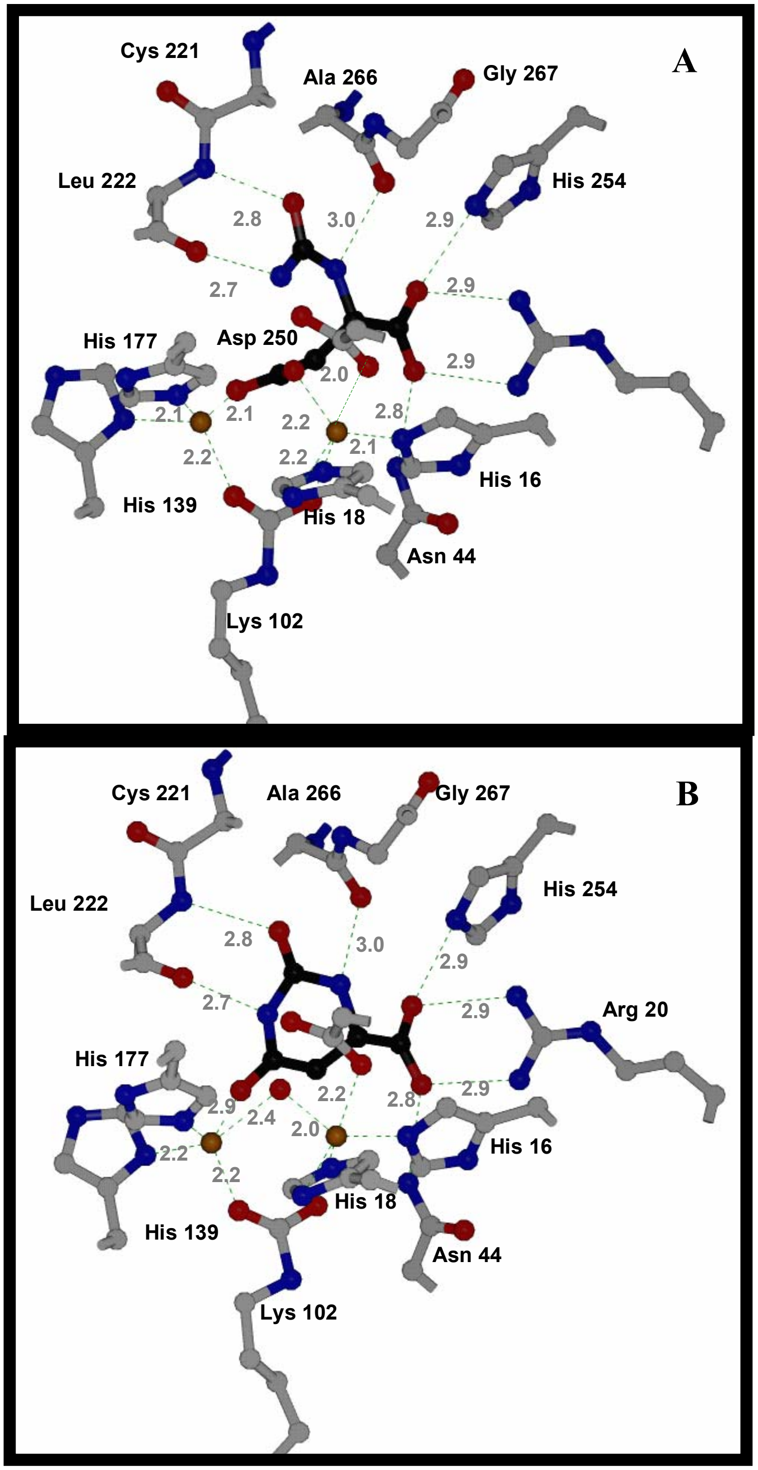
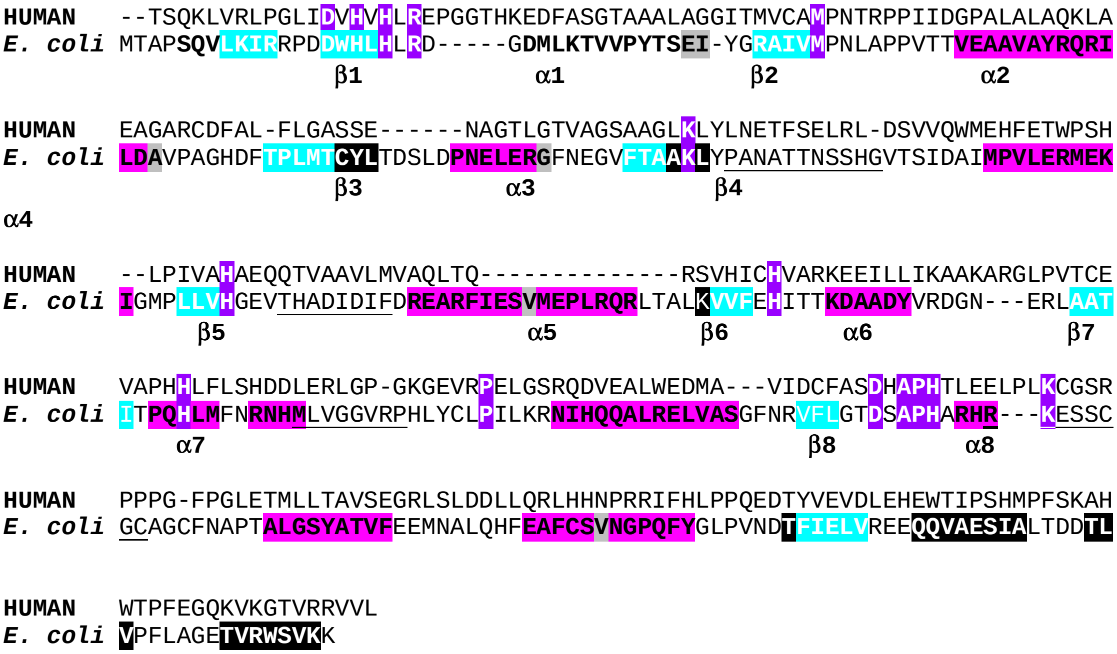


Figure 2.5: The active sites of DHO with bound carbamoyl aspartate (A) and dihydroorotate (B). The distances reported are in Å.



```
HUMAN   --TSQKLVRLPGLIDVHVHLREPGGTHKEDFASGTAAALAGGITMVCAMPNTRPPIIDGPALALAQKLA
E. coli MTAPSQVLKIRRPDDWHLHLRD-----GDMLKTVVPYTSEI-YGRAIVMPNLAPPVTTVEAAVAYRQRI
                       β1            α1             β2              α2

HUMAN   EAGARCDFAL-FLGASSE------NAGTLGTVAGSAAGLKLYLNETFSELRL-DSVVQWMEHFETWPSH
E. coli LDAVPAGHDFTPLMTCYLTDSLDPNELERGFNEGVFTAAKLYPANATTNSSHGVTSIDAIMPVLERMEK
                       β3          α3            β4
α4

HUMAN   --LPIVAHAEQQTVAAVLMVAQLTQ--------------RSVHICHVARKEEILLIKAAKARGLPVTCE
E. coli IGMPLLVHGEVTHADIDIFDREARFIESVMEPLRQRLTALKVVFEHITTKDAADYVRDGN---ERLAAT
              β5                    α5          β6        α6              β7

HUMAN   VAPHHLFLSHDDLERLGP-GKGEVRPELGSRQDVEALWEDMA---VIDCFASDHAPHTLEELPLKCGSR
E. coli ITPQHLMFNRNHMLVGGVRPHLYCLPILKRNIHQQALRELVASGFNRVFLGTDSAPHARHR---KESSC
            α7                                          β8         α8

HUMAN   PPPG-FPGLETMLLTAVSEGRLSLDDLLQRLHHNPRRIFHLPPQEDTYVEVDLEHEWTIPSHMPFSKAH
E. coli GCAGCFNAPTALGSYATVFEEMNALQHFEAFCSVNGPQFYGLPVNDTFIELVREEQQVAESIALTDDTL

HUMAN   WTPFEGQKVKGTVRRVVL
E. coli VPFLAGETVRWSVKK
```

Figure 2.6: Alignment of DHO sequences from human and E. coli. The conserved residues of DHO proteins are highlighted in purple. The residues composing the β-strands are highlighted in black, while those forming the α-helices are in grey. The secondary structure elements are numbered according to those that constitute the core TIM-barrel. The underlined residues are those comprising the subunit interface.

displaced by the side chain carboxylate group of carbamoyl aspartate. The backbone NH and CO groups of Leu-222, at 2.7 Å away, are within hydrogen bonding distance of the carbamoyl moiety of carbamoyl aspartate. N3 of the substrate is 3.0 Å from the carboxylate group of Asp-250. The α-carboxylate group of carbamoyl aspartate is stabilized by electrostatic interactions with Arg-20, Asn-44 and His-254. The two metal ions are separated by 3.7 Å.

Seleno-methionine DHO. The seleno-methionine derivative of DHO was successfully prepared. The MALDI mass spectra confirming the incorporation of the unnatural amino acid is shown in Figure 2.7. The mass of the native protein is 38,941 Daltons while the seleno-methionine derivative has a molecular weight of 39,408 Daltons. The mass difference between the two enzymes is 467 Da, which is consistent with the substitution of all 10 methionine residues with seleno-methionine.

Discussion

Comparison with Other Amidohydrolase Superfamily Members. The active sites of DHO, PTE and urease overlay very well as shown in Figure 2.8 (*59-61*). The root-mean-square deviation of DHO and urease is 1.24 Å (between the backbone of 56 residues of the 8 β-strands), while for PTE and DHO it is 2.9 Å (between 72 structurally equilavent α-carbons) (*59, 62*). The four histidine residues, the aspartic acid and the

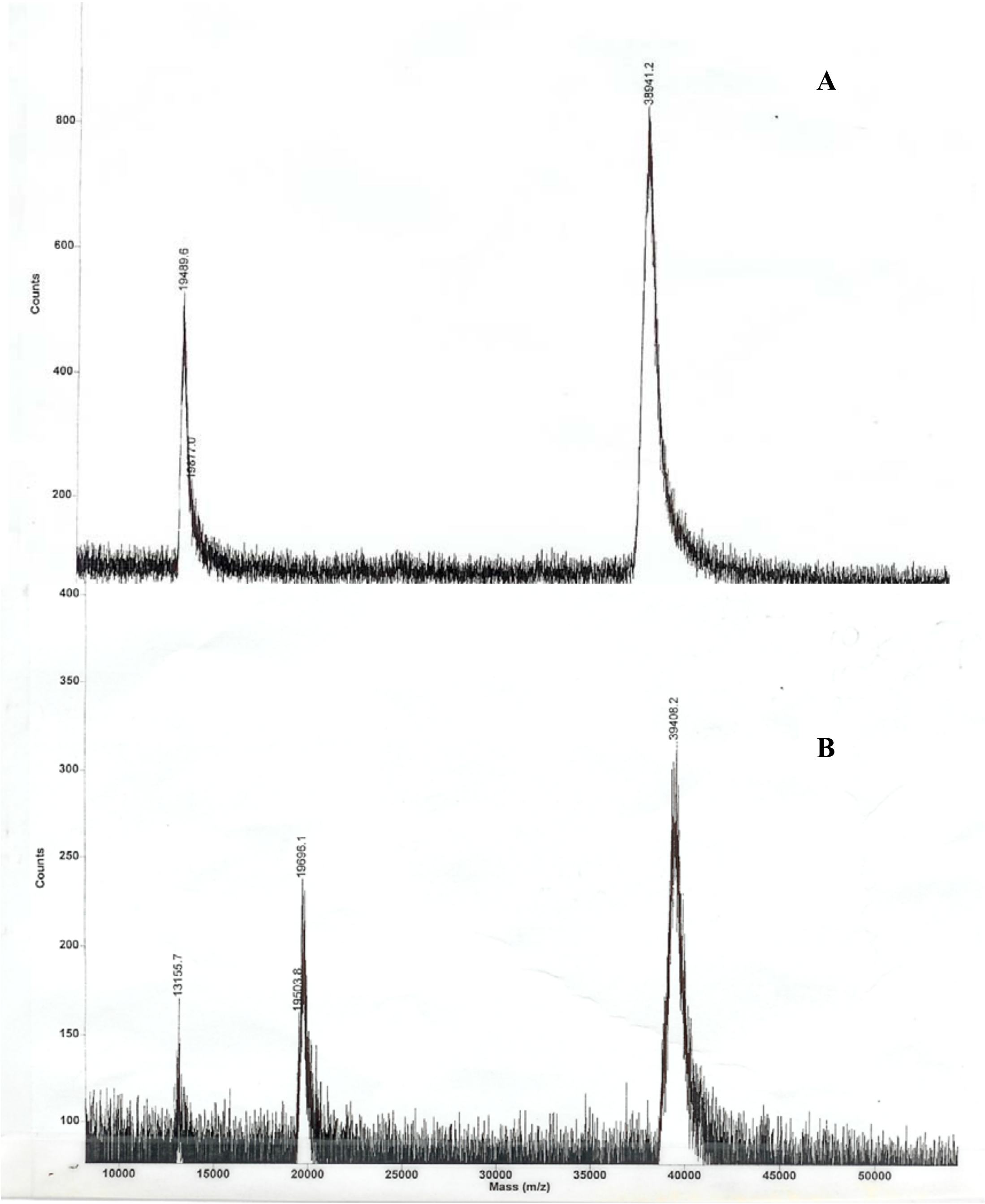


Figure 2.7: Mass spectra of native DHO (A) and the seleno-methionine derivative of DHO (B). The mass difference of 467 Daltons is representative of the incorporation of 10 seleno-methionine residues.

carboxylated lysine of all three enzymes align nicely. It is interesting that although the primary sequence identity is rather low (less than 20 %) for the three proteins, the active sites are so highly conserved.

Since the Holm and Sander paper there have been many additional X-ray crystal structures added to the amidohydrolase superfamily. Some of the structures fit nicely in the mold of the superfamily in that they have the requisite residues at the appropriate positions. Examples are the *E. coli* proteins isoaspartyl dipeptidase (IAD) and cytosine deaminase (CDA) (*63, 64*). Like DHO, IAD contains a binuclear zinc center (*63*). The structure has bound product, aspartate, in the active site. This enzyme is interesting as it is an octamer that is formed from a tetramer of dimers. Each subunit has two domains, the N-terminal region contains eight strands of β-sheet and the C-terminal region consisting of the TIM barrel. The active site of CDA is reminiscent of the active site of murine ADA as several substrate binding residues are conserved (*55*). The substrates of the two enzymes are bound in a similar orientation and position. CDA coordinates a single atom of iron, unlike the zinc-binding ADA. It is important to note that only the bacterial CDA proteins are members of the amidohydrolase superfamily. The fungal counterpart has a different fold that is similar to the one adopted by the bacterial cytidine deaminase (*65, 66*). The active sites of human renal dipeptidase and uronate isomerase differ from the other enzymes of the amidohydrolase superfamily (*67, 68*). In renal dipeptidase there is an HxD motif instead of the HxH from strand 1. Also, as seen in the PTE homology protein, there is a glutamate residue bridging the two metal ions in lieu of the carboxylated lysine at strand 4 (*56*). There also is the absence of an aspartate

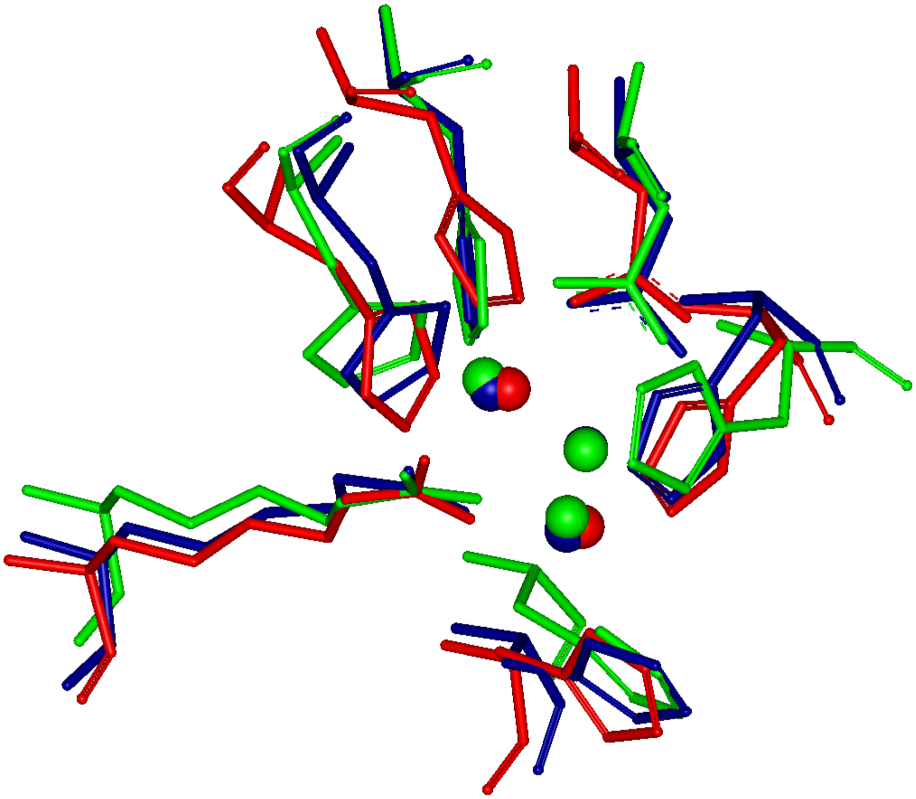
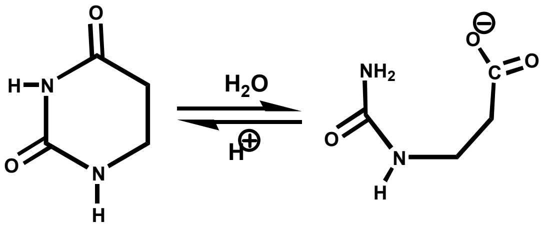


Figure 2.8: Overlay of the active sites of DHO (blue), PTE (green) and Urease (red). The structures were reported in Thoden et al. (*59*), Benning et al. (*60*) and Jabri et al. (*61*). The coordinates were obtained from the PDB (1J79, 1HZY, 2KAU).

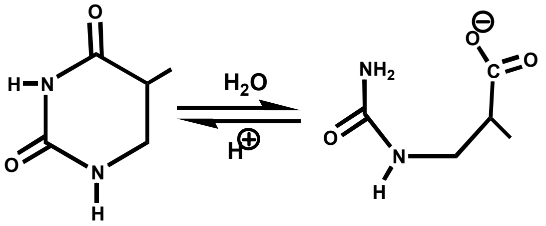
residue bound to the α-metal ion as seen in DHO, PTE and urease. Uronate isomerase (UAI) is an unusual member in that it does not catalyze a hydrolysis reaction. The structure of the *Thermotoga maritima* enzyme revealed that there is a tryptophan residue from strand 7. There are two domains of UAI (A and B) and only domain A contains the TIM barrel domain. The role of the helical domain B is unclear as the active site of the enzyme is found in domain A. The crystallographers saw electron density consistent with a metal ion. As they were unclear on the identity of the metal ion, they conservatively modeled a water molecule into the active site. Our unpublished data on the *E. coli* UAI indicates that the enzyme coordinates a single zinc ion. As more structures of amidohydrolase superfamily members are added, it will be interesting to see if these examples are outliers or more the norm than might be expected.

Cyclic Amidases. The enzymes of the amidohydrolase superfamily with the highest primary sequence similarity to DHO are other cyclic amidases catalyzing the hydrolysis of five or six-membered rings (*1, 69*). Scheme 2.2 shows the reactions of dihydropyrimidinase (DHP), hydantoinase and allantoinase. DHP functions in the second step of pyrimidine degradation. The enzyme catalyzes the reversible hydrolysis of 5,6-dihydrouracil (DHU) or 5,6-dihydrothymine (DHT) to form N-carbamoyl-β-alanine (NCBA) or N-cabamoyl-β-aminoisobutyrate respectively. DHP also hydrolyzes various 5'-mono-substituted hydantoins to 3-ureido acids (*70*). Hydantoinase is the microbial counterpart to DHP. Substrates for these enzymes are D- or L- isomers or racemic mixtures. The X-ray crystal structure for the mammalian DHP has not been

DIHYDROPYRIMIDINASE:

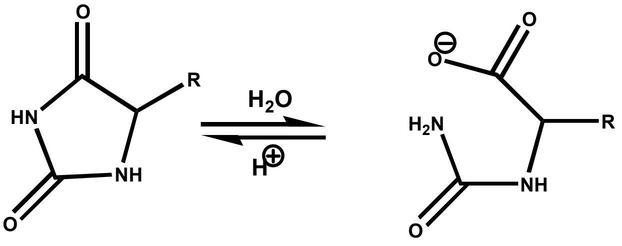


5,6-Dihydrouracil **N-carbamoyl-β-alanine**

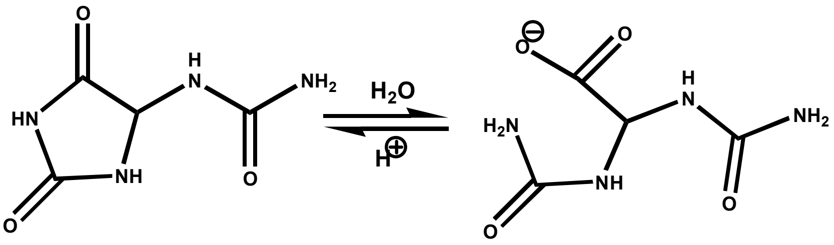


5,6-Dihydrothymine **N-carbamoyl-β-aminoisobutyrate**

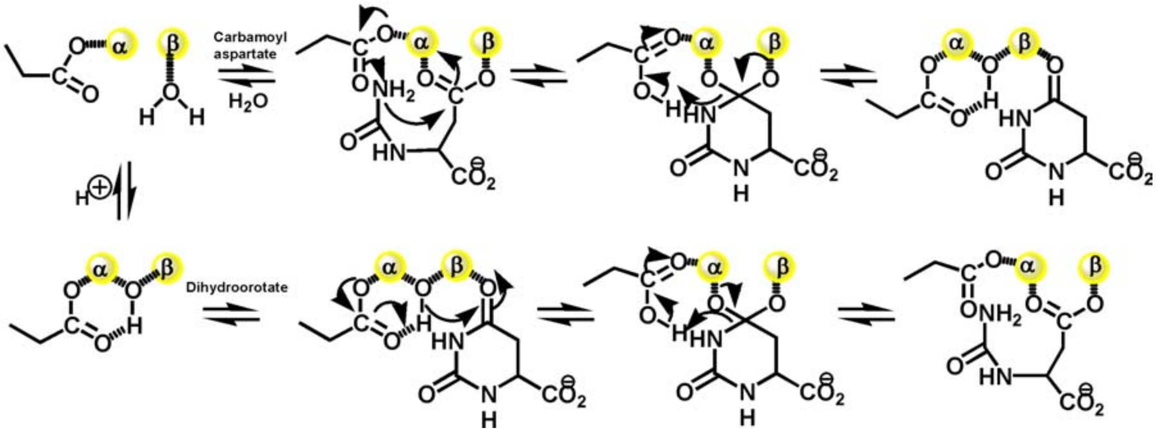
HYDANTOINASE:



ALLANTOINASE:



Scheme 2.2



Scheme 2.3

obtained yet, but as will be discussed later, researchers have developed a structural model of the enzyme. The structure of three hydantoinases has been determined, D-hydantoinase from *Thermus sp.*, L-hydantoinase from *Arthobacter aurescens* and D-hydantoinase from *Bacillus stearothermophilus* (*71-73*).

Proposed Catalytic Mechanism. The proposed mechanism suggested by the crystal structure of DHO is illustrated in Scheme 2.3. In the direction of dihydroorotate hydrolysis, Asp-250 initiates the reaction by activating the bridging hydroxide ion to attack at C-4 of the substrate. The β-metal ion polarizes the carbonyl group of the substrate making it more susceptible to attack. The ensuing tetrahedral intermediate collapses to form the product. In the biosynthesis of dihydroorotate, Asp-250 initiates the reaction by deprotonating the amide nitrogen of carbamoyl aspartate. The amide nitrogen attacks the carboxylate of carbamoyl aspartate to form the tetrahedral adduct. The intermediate collapses upon protonation of the amide nitrogen to produce dihydroorotate.

Comparison to Dihydropyrimidinase. DHP is another enzyme previously thought to have one active site metal ion, a zinc ion. Gojkonvic et al. were the first to present experimental data showing that DHP proteins, specifically those from *Saccharomyces kluyveri* and *Dictyostelium discoideum,* contained two ions of zinc per subunit (*74*). Additionally, Gojkonvic et al. modeled the human DHP sequence onto the DHO structure. The active site of DHP overlays with the DHO active site very well. Consequently, the mechanism proposed for DHP is highly reminiscent of that for DHO.

Subunit Interface. The observation of bound substrate and product in the crystallized dimer poses the question: Do the two subunits communicate with each other during catalysis? Or was the presence of both carbamoyl aspartate and dihydroorotate a fortunate coincidence? The topic of subunit communication has been studied before in the *E. coli* protein biotin carboxylase (*75*). Responsible for catalyzing the ATP-dependent carboxylation of the vitamin biotin, biotin carboxylase is a component of acetyl-CoA carboxylase. The multi-functional enzyme complex acetyl-CoA carboxylase is involved in fatty acid biosynthesis and is composed of biotin carboxylase, biotin carboxyl carrier protein and carboxyltransferase. As is the case with DHO, the enzyme is a homodimer with each subunit containing a complete active site.

Janiyani et al. used N-terminal tags to determine the significance of each subunit to the catalytic activity of biotin carboxylase (*75*). Hybrid molecules of the enzyme were formed in which one subunit of the dimer was wild-type and the other subunit a variant with active site mutations. To do this, the gene encoding for the wild-type (accC) or mutant enzyme was manipulated to incorporate a poly-histidine or FLAG tag.

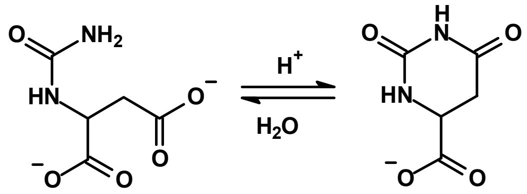
The genes were ligated into expression vectors. The plasmids were used to coexpress the differently tagged subunits. Three types of dimers could form: one with two His-tagged subunits (HH); a dimer that contained two FLAG-tagged subunits (FF); or one that was composed of a subunit with the His-tag and one with the FLAG-tag (HF). Affinity chromatography was used to purify the hybrids. The FF could be separated from the HH and HF dimers by running the sample through an immobilized nickel affinity column. Then the HF could be isolated by running an anti-FLAG affinity column. N-terminal protein sequencing results confirmed that the purified protein was indeed the HF hybrid as both tag sequences were observed. Four different HF biotin carboxylase hybrid dimers were used in kinetic analyses. In the hybrid dimers, one subunit was wild-type while the other contained one of four mutations to an active site residue (R292A, N290A, K238Q or E288K). The kinetic parameters of these hybrids were compared to the HF dimer in which both subunits were wild-type. Biotin carboxylase activity was greatly reduced in all mutant differentially tagged biotin carboxylases. The V_{max} of the R292A/WT HF hybrid was 2.54 min^{-1} compared to a value of 99.8 min^{-1} for the WT/WT HF hybrid. The values for V_{max} were 3.6, 1.06 and 0.35 min^{-1} for the N290A/WT, K238Q/WT and WT/E288K hybrid proteins respectively. These results suggest that two active subunits are required for maximum catalytic function.

Crystallization of Other DHO Proteins. Two papers have been published that report on the crystallization of DHO. Purcarea et al. crystallized the enzyme from the hyperthermophilic bacterium *Aquifex aeolicus* (*76*). The crystallization of the hamster enzyme was reported by Maher et al. (*77*). The X-ray crystal structure has not been published for either enzyme. The authors of the hamster enzyme paper report the formation of crystals containing a tetrameric form of DHO (*77*). The tetramer is believed to be linked through disulfide bonds. This was tested by reacting the dimer and tetramer with DTNB. The dimeric form of the enzyme gives an absorbance change at 412 nm that is consistent with one cysteine residue per monomer. No absorbance change is seen with the tetramer indicating that there are no free surface cysteines. Thus, the tetramer is formed from two disulfide linkages between two DHO dimers. The significance of this tetrameric species is unclear as the authors report that the activity is similar to that of the native form.

CHAPTER III

MECHANISM OF THE DIHYDROOROTASE REACTION

Dihydroorotase (DHO) functions in the biosynthesis of pyrimidine nucleotides by catalyzing the reversible cyclization of N-carbamoyl L-aspartate to L-dihydroorotate as shown in Scheme 3.1. An analysis of the amino acid sequences of DHO from multiple species reveals that there are two general classes of this enzyme (*2*). Members of Class I are found in higher organisms and are larger than their Class II counterparts that are found in many bacteria and fungi. Examples of the Class I form of DHO are exemplified by CAD, a multi-functional enzyme found in mammals, insects and molds. The CAD protein consists of the first three enzymes of the pyrimidine biosynthetic pathway - carbamoyl phosphate synthetase (CPS), aspartate transcarbamoylase (ATC) and DHO (*3, 4*). Monofunctional examples of Class I DHO are found in gram-positive bacterial strains including *Bacillus subtilis, Lactobacillus plantarum, Enterococcus faecalis, Clostridium acetobutylicum* and *Streptococcus aureus.* The Class II enzymes are all monofunctional proteins from gram-negative bacteria and yeast. Class I proteins have typical subunit molecular weights of ~45 kDa compared to ~38 kDa for the Class II proteins. Within each class of enzymes, the amino acid sequence identity is quite high (>40%), but poor sequence identity (<20%) is observed in a comparison of proteins between these two classes. For example, the sequence identity between the human (Class I) and *E. coli* (Class II) forms of DHO is only 17%. There are approximately 17 residues that are fully conserved within the DHO enzymes sequenced to date in both classes of enzyme.



Scheme 3.1

The DHO from hamster was initially reported to contain a single zinc ion that could be replaced by Co^{2+}, Mn^{2+}, or Cd^{2+} with retention of catalytic activity (*38*). However, the X-ray crystal structure of the *E. coli* enzyme demonstrated that the bacterial DHO contains two zinc ions per active site (*59*). The high resolution structure also confirmed the assignment of DHO as a member of the amidohydrolase superfamily with a $(\beta\alpha)_8$-barrel protein fold and whose binuclear metal center was identical to that previously observed for phosphotriesterase and urease (*1, 59*). The amidohydrolase superfamily of metalloenzymes functions primarily, but not exclusively, as catalysts for hydrolytic reactions at carbon and phosphorus centers. A representation of the binuclear metal center in DHO is presented in Figure 3.1. The six conserved residues found at the C-terminus of the central β-strands are His-16, His-18, Lys-102, His-139, His-177, and Asp-250. The two zinc ions are bridged by a carbamate functional group formed from the post-translational carboxylation of Lys-102 with CO_2 and a molecule from solvent that is most likely hydroxide.

The reversible reaction catalyzed by DHO is pH-dependent and the equilibrium constant of the overall reaction as written in Scheme 3.1 is 1.5×10^6 M^{-1} (*39*). At pH 6.2, the equilibrium between carbamoyl aspartate and dihydroorotate is unity. At lower pH values the formation of dihydroorotate is favored. Conversely, at higher pH values the formation of carbamoyl aspartate is dominant. DHO was first crystallized at a pH where the equilibrium constant for the interconversion between substrate and product

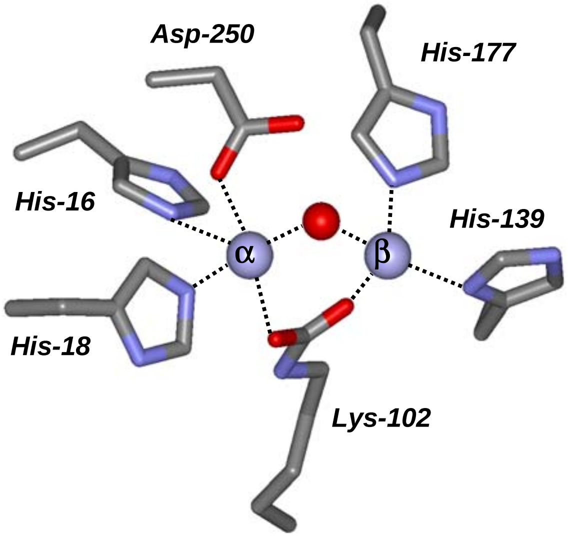


Figure 3.1: Representation of the binuclear metal center within the active site of DHO. The structure was obtained by Thoden et al. (*59*) and the coordinates obtained from the PDB (1J79).

was approximately one (*59*). This fortuitous event enabled the structural elucidation of the specific molecular interactions between substrate/product and the metal center in individual subunits to be unveiled within the same dimer of DHO in the crystalline state. In one subunit of the dimer, dihydroorotate was bound to the active site while in the adjacent subunit, carbamoyl aspartate was bound. Carbamoyl aspartate coordinates to the binuclear metal center via a bridging interaction through the side chain carboxylate of the substrate. In contrast, the amide carbonyl group of dihydroorotate interacts with the binuclear metal center by direct coordination to the β-metal ion. The bridging hydroxide is more closely associated with the α-metal ion. Electrostatic interactions with the backbone atoms of Leu-222, Ala-266 and Gly-267 stabilize the binding of the substrate and product in the active site of DHO. The most specific contacts between protein and substrate originate from the side chains of Arg-20, Asn-44 and His-254 via the formation of electrostatic interactions with the free α-carboxylate group of dihydroorotate and carbamoyl aspartate.

Here, we provide direct biochemical support for the chemical mechanism first suggested by the molecular contacts observed in the X-ray crystal structure of DHO. The role of Asp-250 in proton transfer reactions is assessed via mutation of the side chain carboxylate of this residue to functional groups that are unable to serve this function. Mutagenesis of specific residues was implemented to probe the functional requirement for the electrostatic interactions between the side chains of Arg-20, Asn-44 and His-254 with the α-carboxylate of dihydroorotate and carbamoyl aspartate. Metal-substituted variants of DHO were prepared to elucidate the roles of the metals in catalysis and to identify the role the bridging hydroxide plays in catalysis. The sulfur analog of dihydroorotate was analyzed as a substrate for DHO and used to probe the interactions between the amide bond that is formed and broken with the binuclear metal center of DHO.

Materials and Methods

Materials. N-Carbamoyl aspartate was purchased from Research Organics. Deuterium oxide was obtained from Cambridge Isotope Laboratories. Platinum pfx DNA polymerase was acquired from Invitrogen, while the remainder of the molecular biology products were from Promega or Stratagene. The buffers, substrates and other chemicals were acquired from Sigma-Aldrich. The Gene Technology Laboratory of Texas A&M University performed the DNA sequencing reactions and oligonucleotide synthesis. Thio-dihydroorotate (TDO) was synthesized using a protocol adapted from Christopherson (*42*).

Site-Directed Mutagenesis. The DHO used as the wild-type enzyme in these experiments was altered from the native *pyr*C gene isolated from *E. coli*. Four cysteine residues that are thought to cause sensitivity to oxidation were substituted for serine residues with no loss of activity or change in any other enzymatic property (*50, 51*). These residues are Cys-63, Cys-65, Cys-121 and Cys-179. The single point-mutations at any one of these sites did not consistently produce protein that was stable with maximum catalytic activity for an extended period of time. The quadruple mutant (C63S/C65S/C121S/C179S) was stable and had kinetic parameters consistent with previously reported values (*49, 50*). The DNA template for the mutagenesis experiments described in this paper utilized the altered sequence with the mutation of four cysteine residues to serine and designated as the QM protein.

The mutagenesis protocol was PCR-based and involved the use of four primers that produced two over-lapping fragments (*78*). The cloning vector in all cases was

pBS^+ from Stratagene. A typical reaction contained 2 ng of template DNA, 4 µM of each primer, 1X pfx buffer, 2.0 mM dNTP mix and 5 units of platinum pfx DNA polymerase. The PCR protocol was used as follows: 2 minutes at 95 °C, 25 cycles of 1 minute at 95 °C, 1 minute at 50 °C, and 4 minutes at 68 °C, followed by 1 cycle of 10 minutes at 68 °C. The PCR products were digested with *Eco*RI and *Hin*dIII for two hours at 37 °C, purified and then ligated with similarly digested pBS^+. XL1 Blue cells were transformed with the ligation reactions. Cells containing plasmid were selected on LB plates containing ampicillin. Individual colonies were used to inoculate 5 mL LB cultures, which were grown overnight at 37 °C. The plasmids were isolated from the overnight cultures using the Wizard Mini-Prep SV kit (Promega). The desired mutations were confirmed by sequencing of the isolated DNA.

Protein Purification. The wild type and mutant forms of DHO were all purified in the same manner. The *E. coli* strain, X7014a, that lacks a functional gene for dihydroorotase was obtained from the Yale *E. coli* Genetic Stock Center and transformed with the pBS^+ plasmids containing the gene for DHO. For each preparation, large cultures were incubated at 37 °C until they reached mid-log phase at which point IPTG was added to a concentration of 1.0 mM. After overnight incubation, the cells were collected by centrifugation and resuspended in 50 mM Tris-phosphate, pH 7.0, 100 µM $ZnCl_2$ and 5.0 mM carbamoyl aspartate. The cells were lysed by sonication and the nucleic acids precipitated by the addition of a 2.0% proteamine sulfate solution. After centrifugation, the cell extract was saturated to 60% in ammonium sulfate. After centrifugation, the pellet was resuspended in a minimal amount of buffer and then chromatographed with the aid of an AKTA Purifier (Pharmacia) using a Superdex-200 26/60 column. The buffer used for the gel filtration column was 50 mM bis-tris propane, pH 7.0, containing 5.0 mM carbamoyl aspartate. The flow rate was 1.0 mL/min with 1.5

mL fractions being collected. The fractions containing DHO activity were pooled and loaded onto a Resource-Q anion exchange column with the AKTA system. Buffer A was 20 mM bis-tris propane, pH 7.0, while Buffer B was the same buffer with 1.0 M NaCl. The protein was eluted from the column by the use of a salt gradient (0-30% buffer B in 30 column volumes). The flow rate was 4.0 mL/minute with 1.0 mL fractions being collected. Fractions were assayed for activity and those containing DHO were pooled. The purified enzyme was typically stored at –80 ºC in 20 mM bis-tris propane, pH 7.0, with 20% glycerol, 100 µM $ZnCl_2$ and 5.0 mM carbamoyl aspartate.

Preparation and Reconstitution of Apo-enzyme. Chelex 100 (BioRad) was used to remove contaminating metals from buffers. The buffers were degassed with argon prior to the addition of enzyme. The apo-enzyme was prepared by incubating DHO at a concentration of 1.5 mg/mL for 24 hours with 25 mM dipicolinate in 50 mM sodium acetate, pH 5.5, containing 50 mM sodium sulfate, 2.0 mM sodium hydrosulfite and 30% glycerol. The dipicolinate was removed by dialysis using three changes, 24 hours each, of 20 mM sodium acetate, pH 5.8, containing 2.0 mM sodium hydrosulfite and 20% glycerol. To reconstitute the apo-enzyme, the protein solution was made 100 mM in Hepes, pH 8.0, 2.0 mM sodium hydrosulfite and 100 mM potassium bicarbonate. Typical protein concentrations were 0.5-1.0 mg/mL. Excess divalent metal chloride salts at concentrations of 26-130 µM were added to the apo-enzyme for reconstitution of catalytic activity. Maximum recovery of activity required three days of incubation at 4 °C. Atomic absorption (AA) spectroscopy was used to measure the concentration of each cation in the protein samples. Prior to performing AA analysis, the protein solution was passed through a PD10 column (Pharamacia) equilibrated with metal-free buffer to remove excess metal ions.

Enzyme Assays. DHO activity was determined by a direct spectrophotometric assay at 230 nm (*13*). The assays were performed in a 96-well plate using a SPECTRAmax-340 (Molecular Devices) plate reader. Dihydroorotate absorbs at 230 nm with an extinction coefficient of 1.17 mM^{-1} cm^{-1} (*14*). TDO absorbs at 280 nm with an extinction coefficient of 17 $mM^{-1}cm^{-1}$. The assay volume was 250 μL which corresponds to a path length of 0.69 cm. The buffers for the pH-rate profiles were the potassium salts of MES from pH 5-6.75, HEPES from 6.75-8.5 and TABS from 8.5-9.5. The buffer concentration in each assay was 50 mM and the pH was varied in 0.25 unit increments. The pH was measured at the completion of the enzymatic reaction. For the variation of carbamoyl aspartate, the concentration range was 0.05-20 mM. A range of 0.025-1.75 mM was used for the variation of dihydroorotate and TDHO. For measurement of the solvent isotope effects, the same reaction conditions were used except that the buffers were prepared in D_2O. Enzyme dilutions were performed in 20 mM Hepes, pH 7.0, prepared with D_2O. The pD of each reaction was measured after completion of the assay by adding 0.4 to the pH electrode reading (*79*). The solvent deuterium isotope effects were determined over the pH range of 5.5-7.0 for the synthesis reaction and 8.0-9.8 for the hydrolysis reaction.

Data Analysis. The kinetic parameters, k_{cat} and k_{cat}/K_m, from the initial velocity experiments were determined from a fit of the data to equation 1, where v is the initial velocity, k_{cat} is the turnover number, E_t is the enzyme concentration, A is the substrate concentration and K_m is the Michaelis constant. For the pH-rate profiles, the effect of pH on k_{cat} and k_{cat}/K_m was determined for each substrate by a fit of the data to equations 2 and 3 using the computer programs of Cleland *(16).* Equation 2 was used to fit the data when the activity diminished at low pH whereas equation 3 was used when the activity was lost at high pH. In these equations, y is the value of k_{cat} or k_{cat}/K_m, c is the pH-independent value of y, H is the hydrogen ion concentration and K_a is the dissociation constant of the ionizable group.

$$v/E_t = k_{cat}A/(K_m + A) \tag{1}$$

$$\log y = \log[c/(1 + H/K_a)] \tag{2}$$

$$\log y = \log[c/(1 + K_a/H)] \tag{3}$$

Results

Kinetic Parameters. The k_{cat}, K_m and k_{cat}/K_m kinetic constants for the biosynthesis and hydrolysis of dihydroorotate by the zinc-substituted form of DHO are presented in Table 3.1. Consistent with previously reported data with the DHO from *E. coli*, the maximal rate of formation for the synthesis of dihydroorotate from carbamoyl aspartate is slightly faster than the rate of the hydrolysis reaction (*49*). However, the K_m value for dihydroorotate is about 10-fold lower than that of carbamoyl aspartate and thus k_{cat}/K_m for dihydroorotate is about an order of magnitude higher than it is for carbamoyl aspartate. Although the thio-substituted form of dihydroorotate has been reported to be a potent inhibitor of the eukaryotic DHO from hamster, this compound was found to be a reasonably good substrate for the bacterial enzyme (*42*). Shown in Figure 3.2 are the time courses for the changes in absorbance after the addition of DHO to a solution of TDO. There is a disappearance of the absorbance at 280 nm that is coupled with the

Table 3.1: Kinetic Parameters for Metal-Substituted Forms of Dihydroorotase[a].

Enzyme	Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
Zn/Zn-DHO	Carbamoyl Aspartate	160 ± 8	1.70 ± 0.2	1.0 (0.1) x 10^5
	Dihydroorotate	100 ± 1.6	0.080 ± 0.001	1.2 (0.1) x 10^6
	Thiodihydroorotate	4.4 ± 0.2	0.030 ± 0.001	1.5 (0.3) x 10^5
Co/Co-DHO	Carbamoyl Aspartate	25 ± 1.4	15 ± 0.2	1.6 (0.2) x 10^3
	Dihydrorotate	15 ± 1.4	0.70 ± 0.03	2.1 (0.2) x 10^4
Cd/Cd-DHO	Carbamoyl Aspartate	8.2 ± 0.2	4.0 ± 0.9	4.3 (0.1) x 10^3
	Dihydrorotate	1.9 ± 0.3	0.23 ± 0.06	8.3 (0.1) x 10^3
	Thiodihydroorotate	0.42 ± 0.01	0.009 ± 0.001	4.8 (0.9) x 10^4

[a] These data were collected at pH 5.8 with carbamoyl aspartate as the substrate and at pH 8.0 with dihydroorotate or thiodihydroorotate as the substrate.

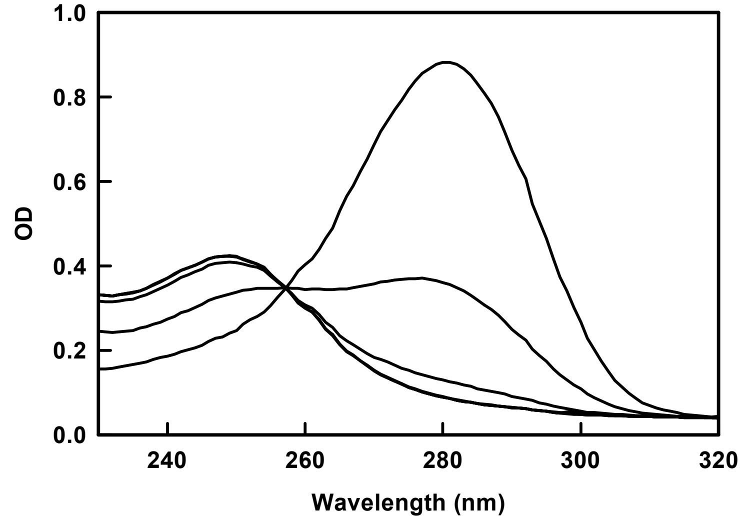
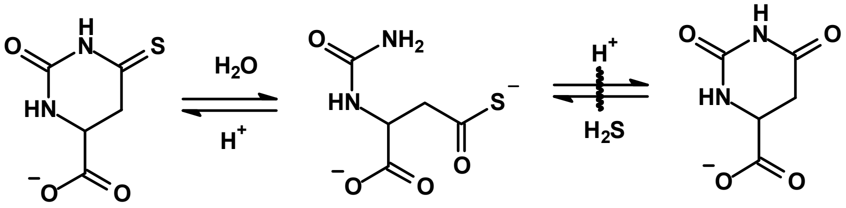


Figure 3.2: Time course for the hydrolysis of TDO at pH 8.0. TDO absorbs with a pH-maximum at 280 nm. After the addition of 0.2 μM DHO, spectra were recorded at 2, 4, and 12 minutes. The absorbance maximum at 280 nm diminishes with time and is replaced by an absorbance maximum at 250 nm that is consistent with the formation of the thio-acid analog of carbamoyl aspartic acid.

appearance of a new species that absorbs at ~250 nm. This observation is consistent with the formation of the thioacid analog of carbamoyl aspartate since the absorbance maximum of thioglycine is reported to be 247 nm (*80*). At pH 8.0, TDO is hydrolyzed at a rate that is 23-fold less than the rate of dihydroorotate hydrolysis and the kinetic constants are presented in Table 3.1. The equilibrium constant for the hydrolysis of TDO was determined by measuring the relative magnitude of the absorbances at 280 and 250 nm as a function of pH from 6.0 to 8.0. The data were fit to a modified form of equation 2 and the equilibrium constant was found to be 1.2 x 10^6 M^{-1}. During these experiments there was no indication for the formation of bisulfide. In addition, incubation of the enzyme with bisulfide (SH^-) and either carbamoyl aspartate or dihydroorotate did not generate a species that absorbs at either 280 nm or 250 nm at pH 6.0 or 8.0. Therefore, once the thioamide bond of TDO is hydrolyzed the reverse reaction does not form dihydroorotate with the liberation of bisulfide as illustrated in Scheme 3.2.



Scheme 3.2

Metal Substitution. The functional role of the two divalent metal ions within the binuclear metal center of DHO was probed via metal substitution experiments. The zinc ions of the native protein were removed by chelation with dipicolinate at pH 5.8 to make the inactive apo-enzyme. The binuclear metal center was reconstituted at pH 8.0 by the addition of 2-5 equivalents of Zn^{2+}, Co^{2+}, or Cd^{2+}. The metal substituted variants of DHO were catalytically active according to the following trend: $Zn^{2+} > Co^{2+} > Cd^{2+}$. The values for k_{cat}, K_m and k_{cat}/K_m of the reconstituted Zn^{2+}, Co^{2+} and Cd^{2+} enzymes are shown in Table 3.1. The Cd-substituted DHO was also utilized as a catalyst for the hydrolysis of TDO and the kinetic constants are presented in Table 3.1. With Cd-DHO the K_m for TDO is significantly lower than it is for the hydrolysis of dihydroorotate. In addition, the value of k_{cat}/K_m for the hydrolysis of TDO is 5-fold higher than it is for the hydrolysis of dihydroorotate.

pH-Rate Profiles. Ionizations within the active site of DHO that are critical for the maintenance of catalytic activity were identified by measuring the kinetic constants in both the forward and reverse directions as a function of pH. The pH-profiles showing the dependence of the kinetic parameters, k_{cat} and k_{cat}/K_m, were determined for all three substrates and the plots are presented in Figure 3.3. The log k_{cat} and log k_{cat}/K_m versus pH profiles (Figures 3.3A and 3.33B) with carbamoyl aspartate as the substrate shows that a single group must be *protonated* for optimal activity. The pH-rate profiles for the hydrolysis of either dihydroorotate or TDO indicates that a single group that must be *unprotonated* for activity is critical for catalytic activity (Figures 3.3C, 3.3D, 3.3E and 3.3F). The pH-rate profiles for the hydrolysis and synthesis of dihydroorotate was measure for the Co/Co-substituted enzyme and the results are also presented in Table 3.2. For the hydrolysis of dihydroorotate, the kinetic pK_a values observed for the effect of pH on k_{cat} and k_{cat}/K_m are shifted to higher values relative to the pK_a values determined with the Zn/Zn-substituted enzyme.

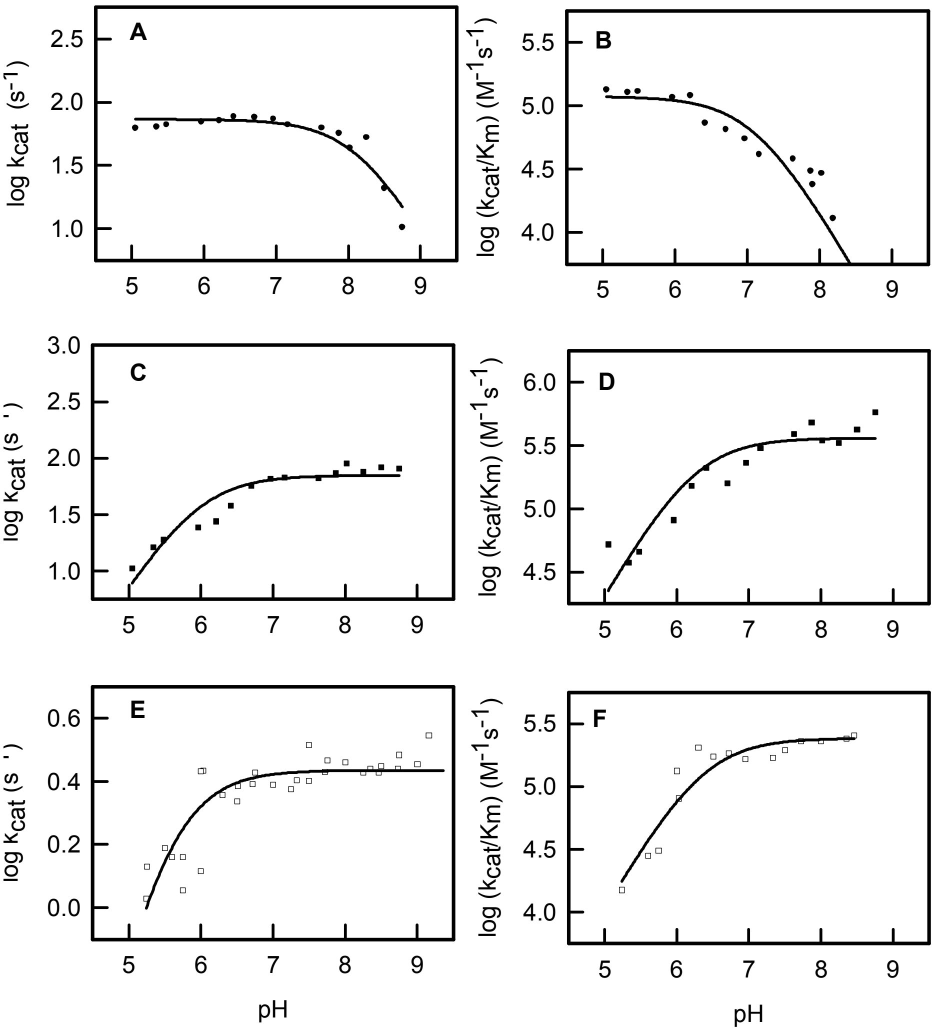


Figure 3.3: pH dependence of the reactions catalyzed by Zn/Zn-DHO with carbamoyl aspartate (A, B), dihydroorotate (C, D) or thiodihydroorotate (E, F) as the varied substrate.

Table 3.2: Kinetic pK_a Values from pH-Rate Profiles and Solvent Isotope Effects[a].

	pK			Solvent Isotope Effects	
Substrate	Solvent	k_{cat}	k_{cat}/K_m	$^{D}k_{cat}$	$^{D}k_{cat}/K_m$
Carbamoyl Aspartate	H_2O	8.2 ± 0.1	7.1 ± 0.1	1.9 ± 0.2	1.7 ± 0.1
	D_2O	> 8.5	7.1 ± 0.2		
	H_2O[b]	7.8 ± 0.2	7.0 ± 0.2		
Dihydroorotate	H_2O	6.0 ± 0.1	6.2 ± 0.2	2.5 ± 0.1	1.1 ± 0.1
	D_2O	5.8 ± 0.1	7.2 ± 0.1		
	H_2O[b]	6.9 ± 0.1	7.6 ± 0.1		
Thiodihydroorotate	H_2O	5.5 ± 0.2	6.4 ± 0.1	2.3 ± 0.1	1.6 ± 0.1
	D_2O	5.9 ± 0.1	6.4 ± 0.3		

[a] Zn-DHO was utilized for these experiments at 30 °C, except as noted. [b] Co-DHO was utilized for these experiments. The pK_a values were obtained by a fit of the data to equations 2 or 3.

Solvent Deuterium Isotope Effects. The pH-rate profiles for the Zn-substituted enzyme for the hydrolysis and synthesis of dihydroorotate were measured in the presence of D_2O and the kinetic pK_a values are presented in Table 3.2. In general, the kinetic pK_a values are shifted to slightly higher values relative to the values obtained for the same parameter measured in H_2O. Similar trends are observed for the hydrolysis of TDO. The pH-dependent equilibrium constant for the interconversion of carbamoyl aspartate and dihydroorotate precluded the accurate measurement of the kinetic constants for formation of dihydroorotate at pH values greater than 8.2 in D_2O. The solvent deuterium isotope effects were measured for all three substrates in the pH-independent region of the pH-rate profiles and the results are presented in Table 3.2. A solvent deuterium isotope effect of ~2.1 was obtained on k_{cat} for all three substrates. The largest effect was found on the k_{cat} for the hydrolysis of dihydroorotate and somewhat smaller effects were obtained for carbamoyl aspartate and TDO. The solvent isotope effects were slightly smaller on k_{cat}/K_m with values ranging of 1.1 to 1.7.

Site-Directed Mutagenesis. The role of specific amino acids in the binding and catalytic interconversion of substrates in DHO was addressed by site-directed mutagenesis. The X-ray structure of DHO shows that the α-carboxylate group of dihydroorotate is coordinated to the active site via multiple electrostatic interactions to Arg-20 Asn-44 and His-254 as illustrated in Figure 3.4. Mutation of the substrate-binding residues Arg-20, Asn-44 and His-254 had a significant affect on catalytic activity and the kinetic constants are presented in Table 3.3. The H254N variant had less than 1% of the activity possessed by the wild-type enzyme. A similar reduction in catalytic activity was obtained with the N44A mutant enzyme. The R20M mutant was inactive but the R20K variant retained significant activity, with a k_{cat} value of 15 s^{-1}.

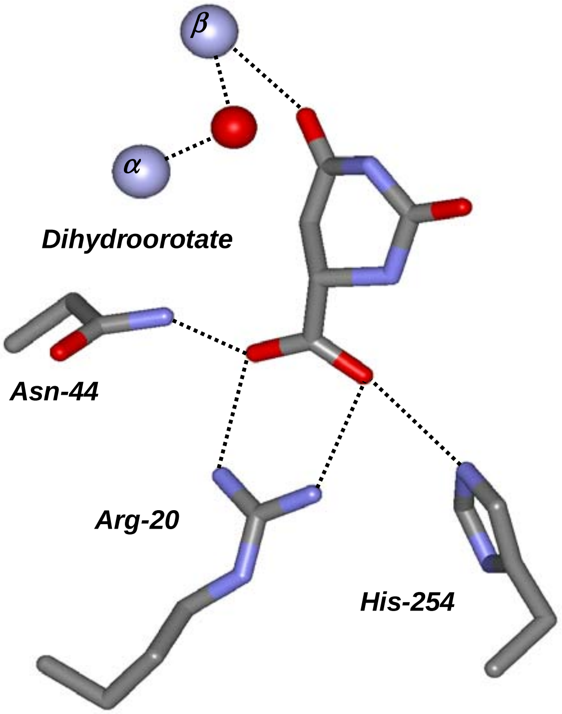


Figure 3.4: Representation of the electrostatic interactions between the α-carboxylate of dihydroorotate and the side chains of Arg-20, Asn-44, and His-254. The coordinates were taken from Thoden et al. (*59*) and obtained from the PDB (1J79).

Table 3.3: Kinetic Parameters for Mutants of Dihydroorotase[a].

DHO Variant	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
Wild-type	160 ± 8	1.7 ± 0.2	$1.5 (0.2) \times 10^5$
ASP-250A	< 0.01[b]	nd[b]	nd[b]
ASP-250E	12 ± 1	1.9 ± 0.3	$6.3 (0.3) \times 10^3$
ASP-250H	< 0.01[b]	nd[b]	nd[b]
ASP-250N	< 0.01[b]	nd[b]	nd[b]
ASP-250S	0.022 ± 0.002	0.51 ± 0.1	$4.3 (0.1) \times 10^1$
ARG-20Q	< 0.01[b]	nd[b]	nd[b]
ARG-20K	15 ± 1	0.9 ± 0.1	$1.7 (0.1) \times 10^4$
ARG-20M	< 0.01[b]	nd[b]	nd[b]
ASN-44A	< 0.01[b]	nd[b]	nd[b]
HIS-254N	< 0.01[b]	nd[b a]	nd[b]

[a] The kinetic constants reported here are for the cyclization of carbamoyl aspartate.

[b] Not determined because of the detection limit of the assay. These experiments were conducted at a carbamoyl aspartate concentration of 10 mM.

It has been postulated that the side chain carboxylate of Asp-250 shuttles the proton from the bridging hydroxide to the leaving group amide of carbamoyl aspartate during the interconversion of substrate and product (*59*). Five variations of the residue at position 250 of DHO were constructed and characterized using carbamoyl aspartate as the substrate. The conservative replacement in the D250E mutant had the highest activity with a k_{cat} of 6 s^{-1}, relative to a turnover number of 160 s^{-1} for the wild-type enzyme. The catalytic activity of the D250S mutant was diminished by a factor of 4000 in comparison with the wild type enzyme. The turnover numbers for D250A, D250H, and D250N were reduced by more than four orders of magnitude relative to the wild type enzyme. The kinetic constants are presented in Table 3.3. The pH-rate profile for the D250S mutant was measured using carbamoyl aspartate as the substrate. The kinetic pK_a for the group that must be protonated for activity was found to be 7.0 ± 0.1.

Discussion

The three-dimensional X-ray structure of DHO has provided a unique view of the active site of this enzyme and the manner in which dihydroorotate and carbamoyl aspartate are associated with the binuclear metal center (*5*). The trapping of the substrate and product within a single protein crystal has enabled a relatively rare glimpse of the molecular interactions of the two reactants in an enzymatic transformation immediately before and after the bond-making and bond-breaking events. This image of the active site has laid the groundwork for the proposed reaction mechanism and subsequent biochemical verification. The reversible hydrolysis of dihydroorotate likely requires three modes of catalysis by DHO: (*i*) the hydrolytic water molecule must be activated for nucleophilic attack, (*ii*) the amide bond of the substrate must be made more electrophilic by polarization of the carbonyl-oxygen bond, and (*iii*) the leaving group nitrogen must be protonated as the carbon-nitrogen bond is cleaved. The catalytic properties of the wild type enzyme and selected mutants of DHO are consistent with a chemical mechanism that incorporates all three forms of substrate activation.

Activation of Solvent Water. The X-ray structure of DHO in the presence of bound dihydroorotate reveals that there is a single solvent molecule associated with the binuclear metal center. The lone molecule from solvent is found bridging the two divalent cations as illustrated in Figure 3.1. The pH-rate profile for the enzymatic hydrolysis of dihydroorotate shows that a single group associated with the protein must be unprotonated for catalytic activity. The loss of catalytic functionality as the pH is lowered is consistent with the protonation of the hydroxide that is proposed to bridge the two divalent cations within the active site of DHO. In the direction of dihydroorotate hydrolysis, the kinetic pK_a for the Zn-substituted DHO is ~6.1 for k_{cat} and k_{cat}/K_m and is elevated to 6.9 and 7.6 for k_{cat} and k_{cat}/K_m, respectively, for the Co-substituted DHO. Therefore, the specific metal ion associated with the binuclear metal center dictates the pK_a for the bridging hydroxide. These results differ from what was found with the hamster enzyme as it was reported that the kinetic pK_a in the direction of dihydroorotate hydrolysis did not change significantly as the divalent cation was substituted with other metal ions (*38*).

The pH-rate profile is different in the direction for the synthesis of dihydroorotate from carbamoyl aspartate. Activity is lost as some group within the active site loses a proton. Microscopic reversibility dictates that this is the same group that must be unprotonated for catalytic activity during the hydrolysis reaction. Optimization of the biosynthetic reaction would require the protonation of the bridging hydroxide followed by dissociation of the resulting water molecule upon the binding of carbamoyl aspartate. This conclusion is supported by the X-ray structure of DHO in the presence of carbamoyl aspartate (*59*). In this enzyme-substrate complex the carboxylate group of carbamoyl aspartate was found bridging the two divalent cations. There was no other water molecule coordinated to either metal ion. The kinetic pK_a values measured for k_{cat}/K_m with the Zn-substituted DHO in the two directions are not the same and thus the apparent ionization constant for the bridging water/hydroxide must be influenced by the stickiness of one or both of the substrates (*18*).

Polarization of the Substrate. The activation of dihydroorotate as an electrophile is expected to be enhanced through polarization of the carbonyl group of the substrate. Complexation of dihydroorotate in the active site of DHO via a direct interaction of the carbonyl oxygen with the metal center would diminish the electron density at the carbon center and facilitate nucleophilic attack by the bridging hydroxide. This type of substrate activation is observed in the crystal structure of dihydroorotate bound to the Zn-substituted DHO. In this structure the carbonyl oxygen is 2.9 Å from the β-metal ion. The direct interaction of the metal center with the carbonyl oxygen of the amide bond to be cleaved is also supported by the kinetic properties of DHO upon substitution of sulfur for oxygen in the substrate using TDO. The more easily polarized sulfur is expected to complex better with the softer CADmium than with the harder zinc. In the CADmium-substituted enzyme the k_{cat}/K_m for TDO is 5-fold higher than with the Zn-substituted DHO, whereas with the Zn-substituted DHO, the k_{cat}/K_m for dihydroorotate is an order of magnitude higher than it is for TDO.

Protonation of the Leaving Group. During the hydrolysis of dihydroorotate the amide nitrogen must be protonated. Conversely, during the synthesis of dihydroorotate a proton must be abstracted from the attacking amide nitrogen. In the X-ray structure of DHO there are no acid/base groups that can facilitate this proton transfer except for Asp-250 from strand 8 that is also ligated to the α-metal ion. This residue has been proposed

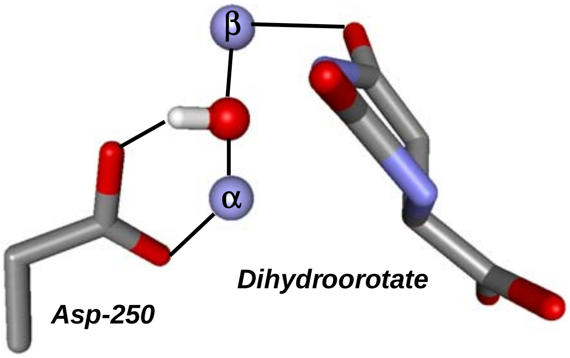
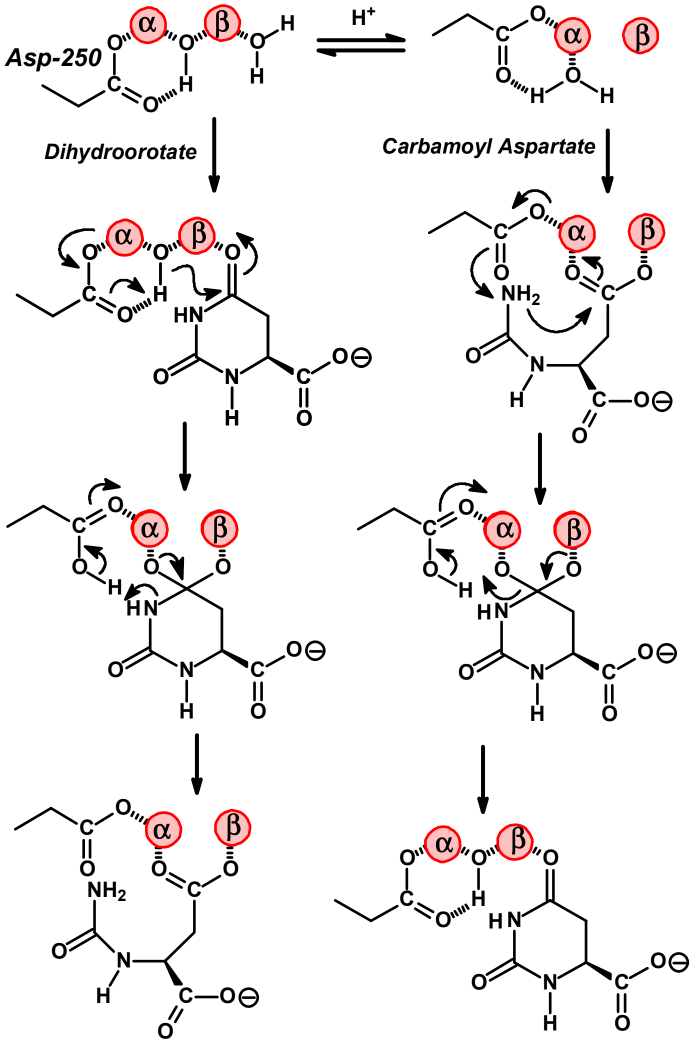


Figure 3.5: Relative orientation of Asp-250 and the hydroxide that bridges the two divalent cations within the active site of dihydroorotase (PDB 1J79). The proposed orientation of the hydrogen is shown for discussion purposes only.

as the group that shuttles the proton from the bridging hydroxide to and from the substrate and product during the course of the reaction (*5*). Shown in Figure 3.5 is a representation of the active site of DHO showing the relative positions of dihydroorotate, Asp-250, the two divalent cations, and the bridging hydroxide. In this structure the hydroxide that bridges the two divalent cations is poised to attack the carbonyl carbon of the bound dihydroorotate. If one positions the hydroxide bound hydrogen and the remaining lone pair attached to the oxygen in a tetrahedral arrangement with regard to the ligation to the two divalent cations, then the hydrogen is logically placed in a hydrogen bonding interaction with the carboxylate from Asp-250 and the lone pair is orientated directly toward the carbonyl carbon of dihydroorotate. The essential nature of Asp-250 for proton transfer reactions was addressed via site-directed mutagenesis. This residue was substituted with alanine, glutamate, histidine, asparagine and serine. Of these residues, the substitution by glutamate was the only perturbation that was tolerated at this position with any significant catalytic activity. In the hamster enzyme, mutation of the homologous aspartate residue to glutamate resulted in an enzyme that had a k_{cat} that was less than 10% of wild-type enzyme and had a 14-fold increase in K_m (*36*). The aspartate from strand 8 that resides in the active site of the related enzyme phosphotriesterase has also been demonstrated to be involved in proton transfer reactions with the bridging hydroxide (*20*). The solvent isotope effects measured for the formation and hydrolysis of dihydroorotate are modest. At this point it is not possible to determine if this reflects the effect of deuterium on the proton transfer during substrate attack or whether this is a reflection of the recharging of the binuclear metal center from water after each turnover of the substrate.

Substrate Binding Interactions. The exocyclic α-carboxylate group of dihydroorotate is bound to the protein via electrostatic interactions with the side chains of Arg-20, Asn-44, and His-254. Mutation of any one of these residues abolishes the ability to cyclize carbamoyl aspartate to dihydroorotate with the single exception of the replacement of Arg-20 with lysine. Loss of enzymatic activity was also seen when the homologous residues were mutated in the hamster enzyme (*36, 81*). It is clear from this result that the active site of DHO has a rather stringent requirement for substrate binding. In this regard we have attempted to hydrolyze dihydropyrimidine with the wild type DHO without success.

Mechanism of Action. The characterization of the catalytic properties of the wild type and mutated forms of DHO coupled with the high resolution X-ray crystal structure has provided sufficient insights to assemble a self consistent chemical mechanism for the interconversion of substrate and products within the active site of this enzyme. The reaction mechanism is presented in Scheme 3.3. For the hydrolysis of dihydroorotate the active form of the enzyme is one in which a hydroxide is bridging the two divalent cations within the enzyme active site. Dihydroorotate binds to the binuclear metal center where the carbonyl oxygen is ligated to the β-metal which polarizes the carbonyl group. This orientation is supported by the X-ray structure of the bound dihydroorotate and the relative kinetic properties of the thio-substituted dihydroorotate (*5*). Nucleophilic attack by the bridging hydroxide is facilitated by the transfer of the proton from the hydroxide to the carboxylate coordinated to the α-metal ion. A tetrahedral adduct is formed that bridges the two divalent cations. Collapse of the tetrahedral adduct occurs with protonation of the amide nitrogen and cleavage of the carbon-nitrogen bond. The resulting carbamoyl aspartate is coordinated to the binuclear metal center via the newly formed carboxylate group. This structure is supported by the X-ray structure of the bound carbamoyl aspartate (*5*). The product is released and the binuclear metal center is recharged with hydroxide via a process that has not been addressed by the experiments presented in this paper. In the reverse direction, the carbamoyl aspartate binds to the protonated form of the enzyme with the release of a bound water molecule. In this direction the reaction is initiated by the abstraction of a proton from the amide nitrogen by the carboxylate of Asp-250 concomitant with nucleophilic attack of the amide nitrogen on the carboxylate of carbamoyl aspartate that is bridging the binuclear metal center.



Scheme 3.3

CHAPTER IV

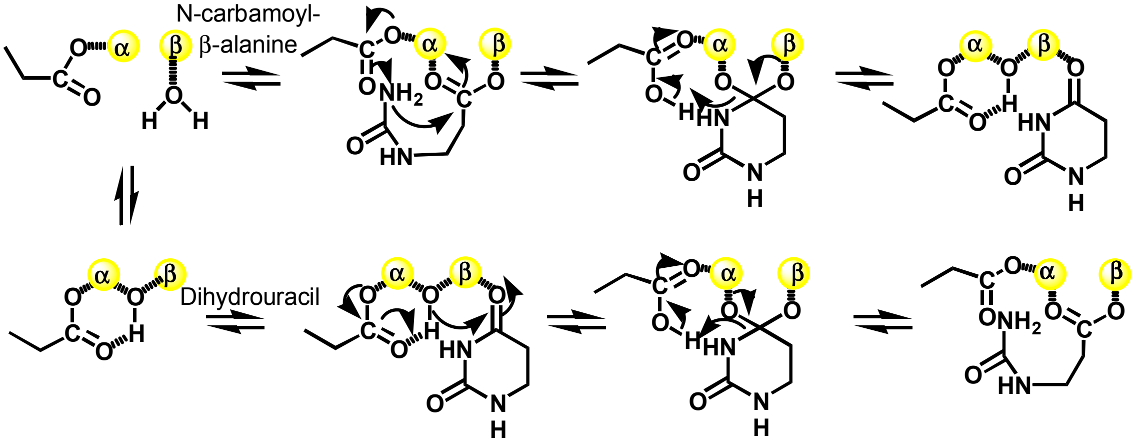
THE ISOLATION OF A PROBABLE DIHYDROOROTASE FROM *PSEUDOMONAS AERUGINOSA* AND THE EVOLUTION OF DIHYDROOROTASE INTO A DIHYDROPYRIMIDINASE

Involved in the de novo biosynthesis of pyrimidines, dihydroorotase (DHO), is an essential enzyme to the viability of microbial organisms. Unlike mammalian systems where there is a salvage pathway, there are no other pathways for the synthesis of uracil, cytosine and thymine nucleotides in pyrimidines. The full alignment of DHO proteins from 71 bacterial organisms is shown in Appendix I. There are six regions that are highly conserved. Due to the X-ray crystal structure of the *E. coli* enzyme, it is understood that these regions are important in delivering the active site residues (*59*). Four histidines, a lysine and an aspartic acid coordinate the binuclear zinc center of DHO in *E. coli*. Two of the histidines are found in an HXH motif. In some organisms including *Trichodesmium erythraeum*, *Nostoc punctiforme, Thermosynecoccus elongatus, Rhodopseudomonas palustris, Brucella suis, Rhodospirillum rubrum, Magnetospirillum magnetotacticum, Ralstonia metallidurans, Xanthomonas axonopodis, Xyllela fastidiosa, Pseudomonas fluorescens, Pseudomonas aeruginosa* and *Cytophaga hutchinsonii* this motif is replaced by a QXH motif, however, the four other active site residues are present. It is unclear whether these proteins are active because in some organisms there is only the one DHO sequence with the QXH motif. However, in *P. fluorescens* and *P. aeruginosa*, representatives of both sequences, QXH and HXH are present.

The bacterium *P. aeruginosa* is a pathogen that is a leading cause of human infections. Due to its resistance to antibiotics and disinfectants, *P. aeruginosa* infections are impossible to eradicate and cause heart failure and death (*82*). Due to its clinical significance, there is great interest in understanding the biochemistry of the organism. To that end, the genome of *P. aeruginosa* PA01 has recently been sequenced (*83*). As annotated by the institute for genomic research (TIGR) the *P. aeruginosa* has three DHO sequences: the inactive pyrC' (PA0401), a probable DHO (PA5541), and pyrC (PA3527) which encodes for the active enzyme. PA0401 has been shown to be associated with aspartate transcarbamoylase (ATC) (*12*). Without complexation with pyrC', ATC is inactive and insoluble. Thus, it appears that the inactive dihydroorotase assists in the folding of ATC. The probable DHO contains the QXH motif in lieu of the more typical HXH. It is known that the histidines of the HXH bind the α-metal ion in *E. coli*, or more buried zinc of DHO. Replacement of the first histidine with a glutamine could have an effect on metal-binding. PA5541 is a larger protein than the pyrC gene product with a molecular weight of 49,295 Da compared to 38,406 Da. Figure 4.1 shows the alignment between the proteins and highlights the shared residues. The sequence identity between PA5541 and PA3257 is 21 %. As previously mentioned, all the other residues that are known to be involved in binding the metal and substrate, as well as those involved in catalysis, are conserved between the two proteins. To further understand the function/structure relationship of DHO, the isolation and characterization of PA5541 was attempted. The goal is to determine whether it is an active DHO and if so, try to draw some conclusions on why the organism would have two active enzymes. That would answer why some organisms are viable even though they only have the QXH containing sequence.

```
PA3527 MSDRLT---------------LLR-----------------------------P---  W
PA5541 MQSLLIRNARMVNEGQVREGDLLVRHGRIERIAGCLENCGASREIDAAGRYLLPGMID -

PA3257  I L D-GAALANTVGDAART-----FGRAIV   LVPPVRNAAEADAYRQRILAARPAA
PA5541  V F EPGYPQKGSIASESRAAVAGGITSFMD   -TRPATLSLEALAEKKRLAAAHSVA

PA3527 SRFEPLMVLYLTDRTSTEEIRTAKASGFVHAA LYPAGATTNSDSGVTRIDNIFEALEAM
PA5541 NYGFHFGV----SRDNLDTV-AALDPREVAAV VF-MGAST----GDMLVDDLPTLERLF

PA3527 AEVGMPLLV GEVT-RAEVDVFDREKQFIDE-----HLR---------------RVVERF
PA5541 ASVPTLLLS CEDTPRIEANLARWRQRFGERIPAAAHPRIRDAEACYRSTALAVELAQRH

PA3527 PTLKVVFE ITTGDAAQFVREAP---ANVGATITAH LLYNRNHMLVGGIRPHFY-CL I
PA5541 GTLLHVL- LSTARELALFEDKPLCQKRITAEVCVH LLFDDSDYARLG---HLLKCN A

PA3527 LKRNTHQEALLDAAVSGNPKFFLGT S   ARHA EAA-----CGCAGCYSAYAAI-ELY
PA5541 IKSREDRDAL-RRALAGNRLDVIGT H   AWAE QQAYPQAPAGLPLVQHALPALLELV

PA3527 AEAFEQRNAL--DKLEGFASLHGPDFYGLPRN---TDRITLVREEWQAPAS-LPF-GDFD
PA5541 REGWLSLATLVAKTSHRVAELFAIADRGFLREGYWADLVLVSELEHPALASAMPLLSRCN

PA3527 VVPLR-------------AGETLRWK------------LLEAG
PA5541 WTPFRHRAFHHRIDTTIVSGQ-LAWHAGRLSDDCQGLPLRFSR
```

Figure 4.1: Alignment of *P. aeruginosa* DHO sequences, PA3527 and PA5541. The gray highlighted residues are the conserved residues of over 60 bacterial DHO proteins. The red highlighting is used to point out the His for Gln substitution in PA5541.



Scheme 4.1

Dihydropyrimidinase catalyzes the reversible hydrolysis of dihydrouracil or dihydrothymine to form N-carbamoyl-β-alanine or N-carbamoyl-β-aminoisobutyrate. Both dihydropyrimidinase and DHO are members of the amidohydrolase superfamily and conserve their overall structural TIM-barrel fold and active site residues. Like DHO, dihydropyrimidinase coordinates two molecules of zinc per active site (*74*). Due to the lack of an X-ray structure of the enzyme, dihydropyrimidinase was modeled onto the structure of DHO. The active site residues of the two enzymes overlayed very well. The model allowed for the proposal of the mechanism for dihydropyrmidinase, shown in Scheme 4.1, which is identical to that proposed for DHO (Sheme 2.3) in which an aspartate residue and a catalytic water mediate catalysis. The only difference between the structures of the substrates for the two enzymes, dihydrouracil and dihydroorotate, is that the latter has a side-chain carboxylate at C-6. The X-ray crystal structure of DHO provided details on how dihydroorotate is bound in the active site of the enzyme (*59*). As shown in Figure 2.5, each heteroatom of the substrate is coordinated by a group from the enzyme. The peptidic groups of Leu-222 and Ala-266 contribute electrostatic interactions with dihydroorotate. Arg-20, Asn-44 and His-254 are the only DHO residues with side-chain interactions with the bound substrate. The three residues appear to anchor dihydroorotate in the active site by forming hydrogen bonds with the carboxylate group at C-6. Gojkovic et al. aligned the primary sequences of several dihydropyrimidinases and the *E. coli* DHO (*74*). There is divergence, as expected, at the positions of Arg-20, Asn-44 and His-254 of DHO. At these positions, the dihydropyrimidinases conserve glutamine, aspartate and lysine residues, respectively. To further explore the evolution of the enzymes within the amidohydolase superfamily, the task of altering the substrate specificity of DHO is undertaken using a rational approach. The R20Q/N44D/H254K triple mutant of DHO was constructed and characterized to determine if dihydropyrimidinase functionality is gained.

Materials and Methods

Materials. The *P. aeruginosa* DNA was purchased from ATCC. Platinum pfx DNA polymerase was purchased from Invitrogen. All other molecular biology materials were purchased from Promega or Stratagene, while all chemicals were from Sigma. DNA sequencing and oligonucleotide synthesis was performed by the Gene Technologies Lab at Texas A & M University.

Amplification of PA5541 from P. aeruginosa. The sequence for the PA5541 gene (PA5541) was obtained from the TIGR website. The PA5541 gene was PCR-amplified using pfx polymerase. The gene-specific primers contained engineered restriction sites at the appropriate ends. At the 5'end of the gene, there was an NdeI site, while at the 3' end an EcoRI site was placed. The gene was digested and then ligated into similarly digested pet28 or pet30 vectors (Novagen). The pet28 construct produces the His-tagged protein.

Purification of His-tagged Protein. BL21 DE3 cells were transformed with the pet28 vector containing the PA5541 gene. Large cultures of the cells were incubated at 30 ºC and induced with IPTG upon reaching mid-log phase. The cells were collected by centrifugation after overnight incubation and suspended in 20 mM Tris, pH 7.9 containing 500 mM NaCl and 5 mM imidazole (Buffer A). The resuspended cells were lysed by sonication. After centrifugation, the cell extract was run on the AKTA Purifier (Pharmacia) using a 5 mL Hi-Trap Chelating Sepharose Column (Pharmacia). The column was loaded with Co^{2+} and equilibrated with buffer A. The column was washed with 2 column volumes of Buffer A followed by an additional wash of 5 column volumes of Buffer A with a 5 % step gradient of 20 mM Tris, pH 7.9 containing 500 mM NaCl and 250 mM imidazole (Buffer B). Elution of the protein was by a 5 – 100 % gradient of Buffer B in 5 column volumes.

Site-Directed Mutagenesis of DHO. The DHO used as the wild-type enzyme was the quadruple mutant (C63S/C65S/C121S/C179S). The four cysteine residues are believed to be responsible for the sensitivity to oxidation seen in the native enzyme(*50, 51*). The quadruple mutant was stable and had kinetic parameters consistent with previously reported values (*49, 50*). In the mutagenesis experiments, the DNA template is this altered gene. The protocol for mutagenesis was the overlap extension method previously described (*78*). The single mutations, R20Q, N44D and H254K were completed first using the QM-DHO (C65S/C68S/C121S/C179S) in the PBS^+ vector from Stratagene as wild type. The R20Q mutant was used as the template for the production of the double mutants R20Q/N44D and R20Q/H254K. The latter was used as the template for the triple mutant, R20Q/N44D/H254K. The mutant genes were digested with HindIII and EcoRI restriction enzymes and then ligated into similarly digested PBS^+ plasmid.

Preparation of Crude Extracts. The *E. coli* cell strains BL21 or the DHO$^-$ X7014a (Yale *E. coli* Genetic Stock Center) was transformed with the plasmid containing the DHO gene. Small 45 mL cultures were grown in LB at 37 ºC to mid-log phase and then induced with IPTG (final concentration of 1.0 mM). After incubation overnight, the cells were collected by centrifugation and then resuspended in 50 mM bis-tris propane, pH 7.0 containing 2 % (v/v) of Bugbuster (Novagen).

Purification of R20Q/N44D/H254K Mutant. Large cultures were grown and induced in the same manner as previously described. The cells were collected by centrifugation, resuspended in 50 mM bis-tris propane, pH 7.0 with 100 μM $ZnCl_2$ and 10 mM N-carbamoyl-β-alanine. The cells were lysed by sonication and the nucleic acids precipitated by addition of a 2.0% proteamine sulfate solution. After centrifugation, the cell extract was made 60% in ammonium sulfate. After centrifugation, the pellet was resuspended in a minimal amount of buffer and then run on the AKTA Purifier (Pharmacia) using a Superdex 200 26/60 column. The buffer used for the gel filtration column was 50 mM bis-tris propane, pH 7.0 containing 10 mM N-carbamoyl-β-alanine. The flow rate used was 1.0 mL/min with 1.5 ml fractions being collected. The fractions containing DHO activity were pooled and loaded unto the Resource-Q anion exchange column on the AKTA system. Buffer A was 20 mM bis-tris propane, pH 7.0, while Buffer B was the same buffer with 1M NaCl. The protein was eluted from the column by the use of a salt gradient (0 - 30% B in 30 column volumes). The flow rate used was 4.0 mL/min with 1.0 mL fractions being collected. Fractions were assayed for activity and those containing DHO were pooled. Purified enzyme was typically stored at –80 ºC in 20 mM bis-tris propane, pH 7.0 with 20% glycerol, 100 mM $ZnCl_2$ and 10 mM N-carbamoyl-β-alanine.

Kinetics. The direct spectrophotometric assay for DHO is measured at 230 nm for the native reaction and 225 nm for the dihydropyrimidinase activity (*74, 84*). To measure thiodihydroortate hydrolysis, 280 nm is used. The extinction coefficients are 1.17 $mM^{-1}cm^{-1}$ for dihydroorotate, 1.3 $mM^{-1}cm^{-1}$ for dihydrouracil and 17 $mM^{-1}cm^{-1}$ for thiodihydroorotate (*74, 85*). Using a SPECTAmax plate reader from Molecular Devices, the assays were performed in a 96-well plate. For the synthetic activities, the standard assay contained 10 mM carbamoyl aspartate or N-carbamoyl-β-alanine in 100 mM potassium phosphate, pH 6.0. The standard assays to monitor the hydrolysis reactions contained 1.0 mM dihydroorotate, 1.0 mM dihydrouracil or 60 μM thiodihydroorotate in 50 mM potassium phosphate, pH 8.0. The reaction mix was added to an aliquot of the enzyme to initiate the reaction.

For the substrate saturation curves, the substrate range was 0.24-10 mM for carbamoyl aspartate and 0.024-1.0 mM for dihydroorotate. When thiodihydroortate was the substrate, the concentration range was 4.4-53 μM. To determine the kinetic parameters, k_{cat} and k_{cat}/K_m, the initial velocities were fit to equation 1, where v is the initial velocity, k_{cat} is the turnover number, E_t is the enzyme concentration, A is the substrate concentration and K_m is the Michaelis constant.

$$v/E_t = k_{cat}A/(K_m + A) \qquad (1)$$

Results and Discussion

Isolation and Characterization of PA5541. The agarose gel showing the successful amplification of the PA5541 gene from *P. aeruginosa* is shown in Figure 4.3. The fragment obtained was consistent with the expected size of 1338 base pairs and sequencing confirmed that the desired gene was obtained. The resultant recombinant plasmids, pet28/PA5541 (produces His-tagged protein) and pet30/PA5541, exhibited overexpression in BL21 DE3 cells. The SDS-PAGE gel of the overexpression is shown in Figure 4.2. The His-tagged protein was purified and assayed for DHO activity. The kinetic parameters k_{cat}, K_m and k_{cat}/K_m for the synthesis and hydrolysis reactions are provided in Table 4.1. The values obtained for the activity of PA5541 are much lower than those for the *E. coli* DHO enzyme. The values obtained for k_{cat} for the synthesis and hydrolysis of dihydroorotate were 0.66 s^{-1} and 0.57 s^{-1} respectively. These values are 240-fold and 2000-fold less than the values for the *E. coli* enzyme. Shown in Figure 4.5 and Figure 4.6 are the substrate saturation curves for PA5541 with dihydroorotate or carbamoyl aspartate as the substrate. PA5541 also hydrolyzes thiodihydroorotate as illustrated in the UV scans shown in Figure 4.4. At pH 8.0, thiodihydroorotate absorbs at 280 nm. Upon the addition of PA5541, the absorbance at 280 nm is absent while a species absorbing at 250 nm is present. This species is the product, the thio-analog of carbamoyl aspartate. Thiodihydroorotate hydrolysis by PA5541 is slower than the cleavage of dihydroorotoate. This trend is also seen in the *E. coli* enzyme. However, the value of k_{cat} for the *E. coli* enzyme (4.4 s^{-1}), is more than 80 times greater than that of PA5541 (0.048 s^{-1}). The K_m values for each substrate obtained from PA5541 are much closer to those for the *E. coli* enzyme than the k_{cat} values. With carbamoyl aspartate, dihydroorotate and thiodihydroorotate as the substrate, the K_m values are 1.2, 0.5 and

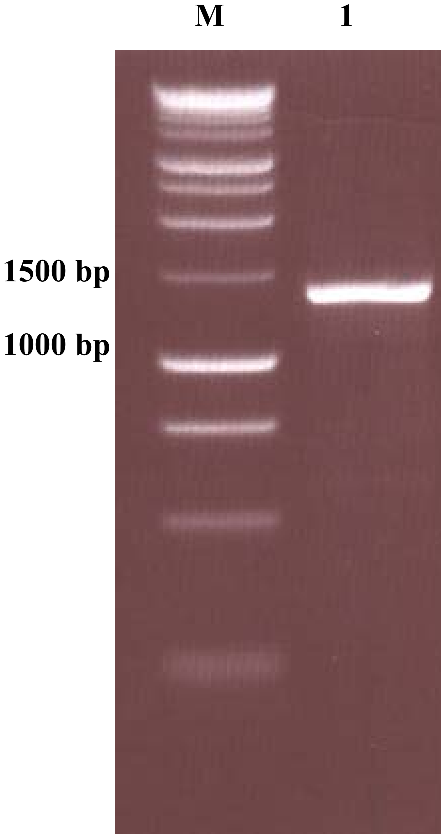


Figure 4.2: 1.0 % Agarose gel demonstrating the amplification of PA5541. The gene, PA 5541, is 1338 bp and is shown in Lane 1. The marker is the 1 kbp molecular weight ladder (Promega).

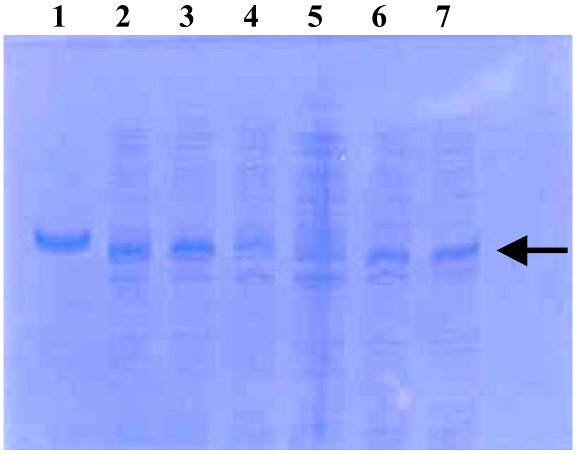


Figure 4.3: SDS-Page gel of the overexpression of PA5541. To prepare the samples for the gel, a small amount of culture was centrifuged, and the pellet was resuspended in sample buffer. In lane 1 is the marker protein, uronate isomerase, which has a MW of 53,000 Da. Lane 2-4 are the cell extracts of BL21/pet 28/PA5541, while lanes 5-7 show the extracts of BL21/pet 30/PA5541. Lanes 4 and 5 are from uninduced cultures while IPTG was added to the cultures that the remainder samples were made from. PA5541 has an expected MW of 49,295 Da. The His-tagged protein from BL21/pet28/PA5541 has a MW of 51,459 Da.

Table 4.1: Kinetic Parameters for PA5541[a].

Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
Carbamoyl Aspartate	0.66 ± 0.056	1.2 ± 0.33	5.7 (1.3) x 10^2
Dihydroorotate	0.57 ± 0.060	0.47 ± 0.11	1.2 (0.16) x 10^3
Thiodihydroorotate	0.048 ± 0.013	0.099 ± 0.037	4.8 (0.053) x 10^2

[a] The synthesis of dihydroorotate was measured at pH 6.0 while the hydrolysis reactions were measured at pH 8.0.

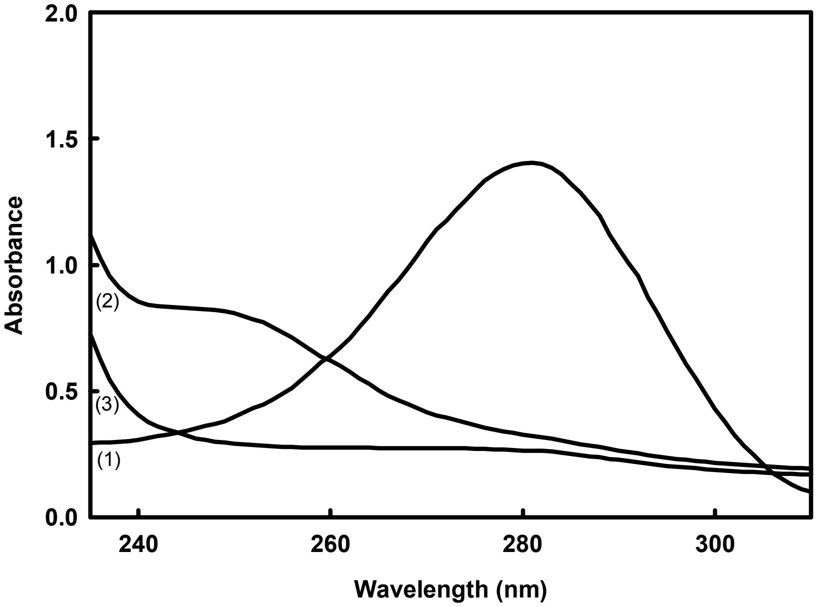


Figure 4.4: UV scan of the hydrolysis of thiodihydroorotate by PA5541. The absorbance versus wavelength was measured for: (1) 80 µM substrate, (2) the product formed from incubation of 80 µM substrate with 4.0 µM enzyme for 3 hours and (3) 4.0 µM enzyme only. Determined at pH 8.0, the scans clearly show the decrease in the absorbance of thiodihydroorotate at 280 nm and the appearance of product which absorbs at 250 nm.

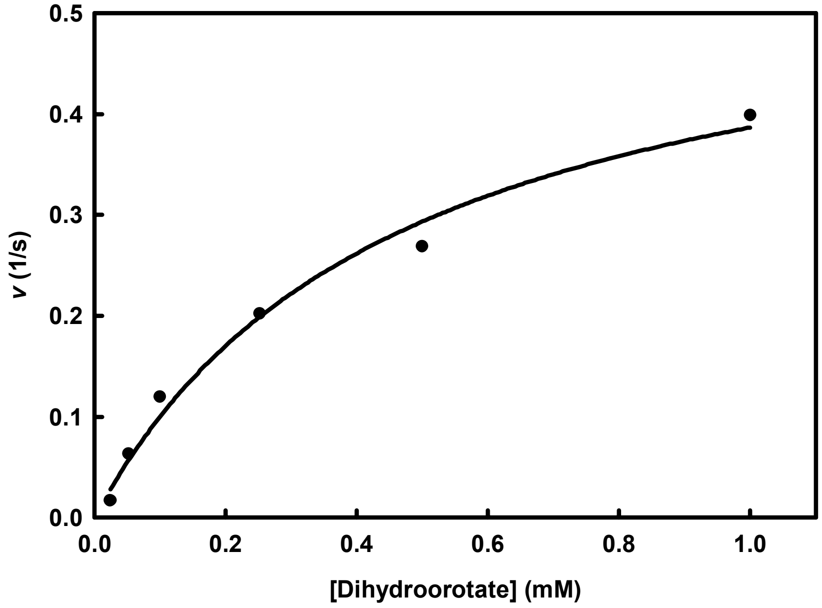


Figure 4.5: Substrate saturation curve for the hydrolysis of dihydroorotate by PA5541. The reaction was monitored at pH 8.0 with 0.02 μM enzyme in the assay.

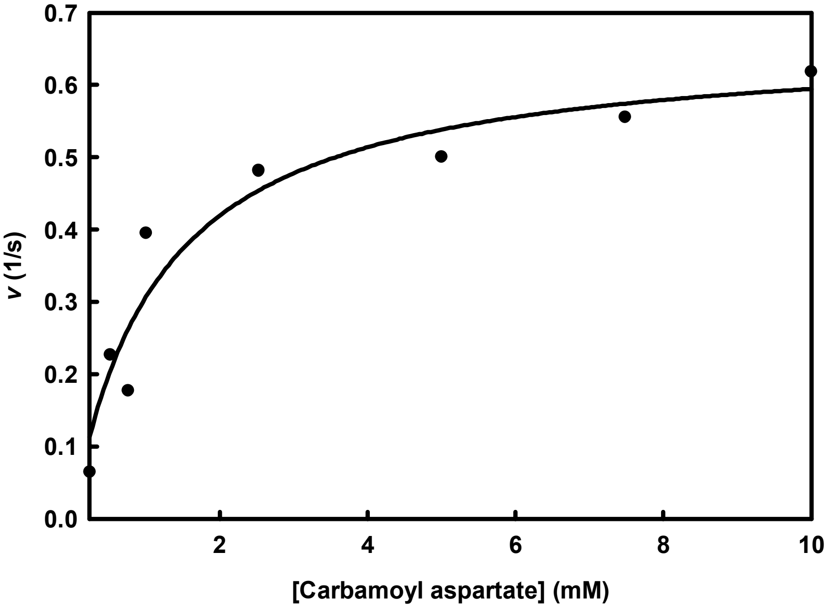


Figure 4.6: Substrate saturation curve for the synthesis of dihydroorotate by PA5541. The reaction was monitored at pH 6.0 with 0.02 µM enzyme in the assay.

0.099 mM respectively. These compare with values of 1.7, 0.08 and 0.03 mM for the *E. coli* enzyme.

His-tags, at the N-terminus in this case, can possibly affect protein folding and thus activity. This does not appear to be the case here as the activities obtained from crude extracts containing the tagged and untagged protein are within 5 % of each other. In the *E. coli* enzyme, when the metal is removed by the addition of a chelator, activity is lost. Both metal ions of the binuclear zinc center appear to be essential for activity. Thus, the low activity levels of PA5541 might be due to the lack of complete metal occupation of the active site. A decrease in the affinity for metal may be a consequence of the QXH for HXH substitution. It doesn't appear that any differences in activity between PA5541 and HXH containing enzymes are due to differences in substrate binding. The experiment to determine the metal concentration of PA5541 has yet to be completed but would be imperative in understanding why the protein has the level of activity seen.

Evolutionary Implications of PA5541. Comparison of PA5541 to the DHO of *P. aeruginosa* that has the HXH motif, PA3527, is also important in delineating the role of the protein in the organism. Though not characterized, PA3527 should have two metal ions per active site. All other active site residues are conserved between the proteins. Thus, comparison of the metal content of the two *P. aeruginosa* DHO enzymes is critical in understanding the role of PA5541.

It is unclear why there would be two functional DHO proteins in the same organism. The QXH motif of PA5541 is approximately 60 amino acids into the sequence, much later than the HXH motif appears in the sequence of PA3527 or the *E. coli* enzyme. This trend is also seen in the sequences of the other organisms that have the QXH motif. In fact, in many gram-positive bacteria and higher organisms with the

HXH motif, it appears later in the sequence and is a larger protein as in PA5541. These sequences form Class I of DHO proteins, while the enzymes of gram-negative bacteria and fungus form Class II. The possibility then arises that PA5541 evolved divergently from its PA3527 relative. At any rate, it would appear that PA5541 is not just an evolutionary artifact as the homologous protein in *Xyllela fastidiosa* and several other organisms is the only DHO sequence in the organism.

Evolution of DHO into a Dihydropyrmidinase. The PBS$^+$ plasmid containing the DHO triple mutant R20Q/N44D/H254K was successfully constructed as confirmed by sequencing. The protein was overexpressed in X7014a and BL21 as seen in the SDS-PAGE gel shown in Figure 4.4. Cell lysis was more efficient when BL21 cells were used as compared to X7014a, so only that strain was used. The values obtained from the crude assay of the triple mutant and wild type protein are in Table 4.1. The triple mutant is 5 times more active against N-carbamoyl-β-alanine than wild-type. For dihydrouracil, the mutant is about twice as active as the wild-type enzyme. The background rates obtained from the BL21 cell extracts are due to the contribution of the chromosomally derived DHO. Sensitivity is a problem in assays utilizing dihydroorotate and dihydrouracil because of the intrinsic complexity of monitoring a reaction in the 225-230 nm region. There are many other things that absorb in a crude assay. This is quite limiting as dihydroorotate and dihydrouracil absorb with extinction coefficients greater than 1. The amount of extract has to be maintained at a level that keeps the absorbance value of the reaction mix in a readable range for the spectrophotomer. Even after attempts to optimize the assay, the value for the rate of dihydrouracil hydrolysis by the triple mutant is only twice that of background. A method of minimizing the background would be to increase the quantity of the overexpressed protein in the cell. This could be accomplished by subcloning the gene into a different expression vector.

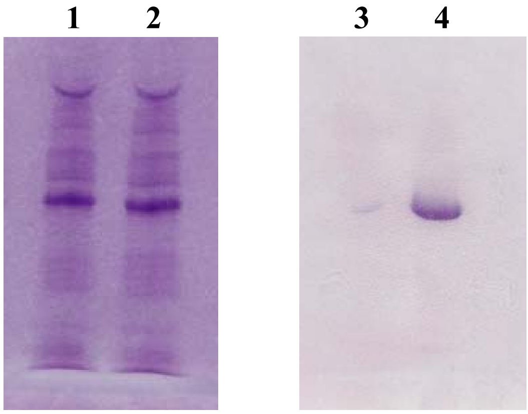


Figure 4.7: 12 % SDS-Page gel of the purification of the DHO mutant protein R20Q/N44D/H254Q. Lane 1 contains the cell extract of X7014a/pBS^{+}/WT DHO, while in lane 2 is extract containing the mutant construct. Lanes 3 and 4 show the purified WT and mutant proteins respectively. The MW of DHO is approximately 38,000 Da.

Table 4.2: Rates Obtained from Crude Assay of Cell Extracts[a]

	Cell strain/DHO (μmol/min/mg protein)		
Substrate	BL21	BL21/WT	BL21/ R20Q/N44D/H254K
Carbamoyl aspartate	0.21	24	0.84
Dihydroorotate	1.2	24	1.9
N-carbamoyl-β-alanine	0.035	0.085	0.43
Dihydrouracil	0.036	0.055	0.13

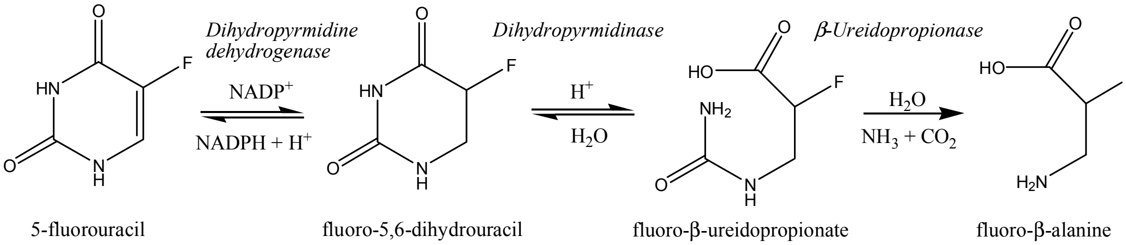
[a] The assays were performed at pH 6.0 when carbamoyl aspartate and N-carbamoyl-β-alanine were the substrates, and at pH 8.0 to monitor the hydrolysis of dihydroorotate and dihydrouracil.

PBS^+ is not the optimal vector for overexpression and other expression systems, such as the pet system (Novagen), are frequently referenced more in recent literature. Another solution may be to switch back to the DHO^- strain but use sonication to lyse the cells.

The triple mutant was purified to further characterize its activity. Unfortunately, the protein appears to be unstable. In a typical purification, the activity of the mutant would consistently decrease after each purification step. The protein was purified in the presence of 10 mM N-carbamoyl-β-alanine which appeared to stabilize the protein. No activity was seen in the purified protein prior to the addition of N-carbamoyl-β-alanine to the buffers. Although activity is seen, the numbers are much lower than the crude data would suggest. For N-carbamoyl-β-alanine turnover, the rate determined using 10 mM of substrate at pH 6.0 was 0.028 μmol/min/mg protein. This is only 1/15 of the activity seen in the crude assay. The instability could be due to the number of mutations in the protein from the native DHO.

The crude triple mutant data definitely suggest that it is possible to make DHO more promiscuous in its preference for substrate. A next step may be to backtrack and characterize the single point mutations, R20Q, N44D and H254K, or the double mutants to determine if fewer steps could be taken to reach the same goal. Fewer mutations may result in a more stable protein. Of particular interest would be to further develop the crude assay and apply it to screen potential substrate analogues or inhibitors. The combinatorial screen would be used to assay DHO and its mutants against compounds of interest.

Two approaches can be used to develop DHO variants for screening. A rational approach manipulates the knowledge of the structure-function relationship of DHO to alter substrate specificity. This method was utilized in the R20K/N44D/H254K mutant to gain dihydropyrimidinase activity. An additional way to produce variants with desired activities is through directed evolution techniques. Directed evolution involves using mutagenesis and/or recombination of one or more parent sequence to introduce genetic diversity (*86*). One method of directed evolution is to undergo several generations of single amino acid substitutions, selecting genes with the desired phenotype as the template for the next generation (*87, 88*). Another strategy is to use a gene shuffling technique (*89*). Here, structurally related genes are used in recombination events to produce a chimeric gene library for evolution of desired characteristics.



Scheme 4.2

Dihydropyrimidinase is a good target for directed evolution shuffling experiments due to its structural and mechanistic similarity to DHO. Also, dihydropyrimidinase has clinical significance. In addition to its role in the breakdown of dihydrouracil and dihydrothymine, dihydropyrimidinase is involved in the degradation of the chemotherapeutic agent 5-flourouracil. The catabolic pathway of 5-flourouracil is shown in Scheme 4.2. Deficiency in dihydropyrimidinase leads to severe 5-flourouracil toxicity which can cause death (*90*).

Dihydrouracil and 5-flourouracil are just two of the compounds that could be used in DHO directed evolution experiments. A series of dihydroorotate analogs with substituents at C-5 and C-6 could be synthesized or purchased and screened. A collaboration with Dr. Gary Sulikowski of Texas A & M University has already produced two compounds. Dihydroorotate analogs with a phenyl- or a methyl- group have been synthesized.

A success story of manipulating an enzyme to introduce a new reactivity is seen in the work reported by Schmidt et al. (*91*). In this work, proteins of the enolase superfamily were altered to introduce new activities. L-Ala-D/L-epimerase (AEE) from *E. coli* and muconate lactonizing enzyme II (MLE II) from *Pseudomonas* sp. were mutated to gain the functionality of o-succinylbenzoate synthase (OSBS). OSBS, AEE and MLE II are members of the same subgroup of the enolase superamily of TIM-barrel proteins. The enzymes also share similar active site geometries with OSBS. However, the three enzymes all perform different reactions and do not share substrates or products. Native AEE and MLE II do not catalyze the OSBS reaction. A rational approach was used to change AEE into an OSBS. The crystal structures of the two enzymes were superimposed. Asp-297 was targeted because in the superimposition, the residue was positioned where the succinyl moiety of o-succinylbenzoate was bound in the OSBS structure. The homologus residue to Asp-297 in OSBS is Gly-288. The D297G mutant indeed had OSBS activity. While WT-AEE had unmeasurable OSBS activity, a k_{cat} value of 0.0025 was obtained for the mutant. The value of k_{cat} for the wild-type OSBS was reported to be 24 s^{-1}. However, the mutation did result in loss of AEE activity, a decrease in k_{cat} to 0.043 s^{-1} from 10 s^{-1} for wild-type. Random mutagenesis was implemented to alter the specificity of MLE II. With a k_{cat} value of 1.5 s^{-1}, the E323G mutant was capable of catalyzing the OSBS reaction. In these experiments, the known structure-function relationships were manipulated to alter substrate specificity of structurally but mechanistically diverse enzymes. This work and the direction that the studies of DHO are leading utilizes the ability of the TIM-barrel fold to bring the active-site residues to an optimal position. This allows for the ability to change reactivities of enzymes with very few mutations.

CHAPTER V

SUMMARY AND CONCLUSIONS

Dihydroorotase (DHO) from *Escherichia coli* catalyzes the interconversion of N-carbamoyl aspartate to dihydroorotate with a loss of a molecule of water. The enzyme is a member of the amidohydrolase superfamily of metalloproteins. The X-ray crystal structure of DHO, obtained at 1.7 Å resolution, ascertained that the enzyme is indeed a member of the superfamily as it has the requisite core TIM-barrel domain (*1, 59*). In addition, the five conserved active site residues of the amidohydrolase superfamily are found at the C-terminus of the β-strands 1, 5, 6 and 8 as is seen in other members like phosphotriesterase, adenosine deaminase and urease. The conserved residues are four histidines and an aspartic acid. Contrary to previous assertions, the enzyme has a binuclear zinc center. This aligned the enzyme with phosphotriesterase and urease, proteins that form a subgroup of the amidohydrolase superfamily that coordinate two metal ions per subunit. Members of this subgroup generally conserve a carboxylated lysine residue that bridges the two metal ions. An exception is the bacterial PTE homology protein which has a bridging glutamate. Adenosine deaminase exists in a second subgroup where only a single catalytic metal ion is present. As new X-ray crystal structures have been solved in recent years, the general theme of the amidohydrolase superfamily has been expanded slightly. While most members catalyze a hydrolysis reaction at a carbon or phosphorus center, not all do. This is the case in uronate isomerase which catalyzes the isomerization of glucuronate to fructuronate.

The X-ray crystal structure of DHO provided was solved to a resolution of 1.7 Å. Views of the active sites with bound substrate or product were important in identifying the residues and/or groups involved in catalysis. In subunit I of the homodimer, carbamoyl aspartate was bound, while subunit II coordinated dihydroorotate. In subunit I, the carboxylate group of carbamoyl aspartate coordinates and bridges the two metal ions. Electrostatic interactions with the backbone atoms of Leu-222, Ala-266 and Gly-267 stabilize the substrate in the active site. Three side chains, from Arg-20, Asn-44 and His-254, form hydrogen bonds to the carboxylate group at C-6 of the substrate. In subunit II, the carbonyl oxygen, O4, of dihydroorotate is coordinated to the β or more solvent exposed metal ion.

The pH rate profile obtained for the hydrolysis of dihydroorotate by DHO shows that a single group must be unprotonated for catalytic activity. The group with a kinetic pK_a of 6.1 for k_{cat} and k_{cat}/K_m was assigned to the bridging hydroxide. Supporting data includes the increase in the pKa values when the active site zinc center is substituted for cobalt. In the direction of dihydroorotate synthesis, the pH rate profile shows that catalytic activity is lost as a single group, the hydroxide ion, on the enzyme is deprotonated. The synthetic reaction is dependent on the protonation of the bridging hydroxide and the subsequent dissociation of water. The kinetic pK_a values of the two reactions are not equal suggesting that the ionization constant is affected by substrate stickiness.

The sulfur analog of dihydroorotate, thiodihydroorotate (TDO), was found to be a substrate for the enzyme. At pH 8.0, the rate of TDO hydrolysis is 23-fold less than that of dihydroorotate hydrolysis. The TDO reaction appeared to be irreversible. In the zinc-substituted DHO, the value of k_{cat}/K_m for TDO was significantly lower than with CADmium-substituted enzyme. This was expected as both sulfur and CADmium are softer than the zinc. Zinc is better suited for harder ligands such as oxygen. This is seen in the higher value of k_{cat}/K_m for dihydroorotate than that of TDO in the zinc-substituted enzyme.

Site-directed mutagenesis experiments confirmed the roles of Asp-250, Arg-20, Asn-44 and His-254 in catalysis and substrate binding. Mutation of Asp-250 was generally not tolerated by the enzyme. The data and the X-ray crystal structure support the involvement of Asp-250 in proton transfer events in the catalytic mechanism of DHO. The mutagenesis of Arg-20, Asn-44 and His-254 to residues that cannot form hydrogen bonds with the carboxylate side chain of the substrate abolishes activity. Thus, interactions provided by the functional groups of these residues are required for substrate recognition and binding.

The X-ray crystal structure and the biochemical data provided insight into the catalytic mechanism of DHO. This is insight that could be applied to less understood members of the superfamily. For the hydrolysis of dihydroorotate, nucleophilic attack by the bridging hydroxide is facilitated by the transfer of the proton from the hydroxide to Asp-250. A tetrahedral intermediate is formed that bridges the two divalent cations and collapses with protonation of the amide nitrogen and cleavage of the carbon-nitrogen bond to form carbamoyl aspartate. In the reverse direction, the reaction is initiated by the abstraction of a proton from the amide nitrogen of the substrate by Asp-250. The deprotonated amide nitrogen nucleophilically attacks the carboxylate of carbamoyl aspartate forming a tetrahedral adduct that collapses to form the product.

The active site residues of DHO are generally conserved among active DHO proteins. An exception is seen in a few organisms where the first conserved histidine is replaced by a glutamine. The DHO sequence with this substitution from *Pseudomonas aeruginosa* was isolated. The protein is able to catalyze the reactions of DHO. Future work will need to determine the metal content of the protein. Additionally, the *P.*

aeruginosa DHO sequence that contains the conserved histidine residue needs to be isolated and characterized for comparison. A rational design method was implemented in altering the substrate specificity of DHO. A gain in dihydropyrimidinase activity was seen in the crude lysates of cells overexpressing the R20Q/N44K/H254K DHO mutant. Future directions include screening the single and double mutants at these positions for the ability to catalyze the interconversion of N-carbmoyl-β-alanine to dihydrouracil. Also, the optimization of the crude assay would allow for the development of a combinatorial screen. Libraries of compounds of clinical or biotechnological significance could be screened against libraries of DHO proteins. Directed evolution techniques would be an efficient way of producing enzyme libraries for the combinatorial screens.

REFERENCES

1. Holm, L., and Sander, C. (1997) An evolutionary treasure: unification of a broad set of amidohydrolases related to urease, *Proteins: Struct., Funct., Genet. 28*, 72-82.
2. Fields, C., Brichta, D., Shepherdson, M., Farinha, M. and O'Donovan, G. (1999) Phylogenetic analysis and classification of dihydroorotases: A complex history for a complex enzyme, *Path Pyr. 7*, 49-63.
3. Jones, M. E. (1980) Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis, *Annu. Rev. Biochem. 49*, 253-279.
4. Musmanno, L. A., Jamison, R. S., Barnett, R. S., Buford, E., and Davidson, J. N. (1992) Complete hamster CAD protein and the carbamylphosphate synthetase domain of CAD complement mammalian cell mutants defective in de novo pyrimidine biosynthesis, *Somat. Cell Mol. Genet. 18*, 309-318.
5. Souciet, J. L., Nagy, M., Le Gouar, M., Lacroute, F., and Potier, S. (1989) Organization of the yeast URA2 gene: identification of a defective dihydroorotase-like domain in the multifunctional carbamoylphosphate synthetase-aspartate transcarbamylase complex, *Gene 79*, 59-70.
6. Guyonvarch, A., Nguyen-Juilleret, M., Hubert, J. C., and Lacroute, F. (1988) Structure of the *Saccharomyces cerevisiae* URA4 gene encoding dihydroorotase, *Mol. Gen. Genet. 212*, 134-141.
7. Labedan B, X. Y., Naumoff DG, Glansdorff N. (2004) Using quaternary structures to assess the evolutionary history of proteins: the case of the aspartate carbamoyltransferase., *Mol. Biol. Evol.. 21*, 364-373.

8. Schurr, M. J., Vickrey, J. F., Kumar, A. P., Campbell, A. L., Cunin, R., Benjamin, R. C., Shanley, M. S., and O'Donovan, G. A. (1995) Aspartate transcarbamoylase genes of *Pseudomonas putida*: requirement for an inactive dihydroorotase for assembly into the dodecameric holoenzyme, *J. Bacteriol. 177*, 1751-1759.

9. Hughes, L. E., Hooshdaran, M. Z., and O'Donovan, G. A. (1999) Streptomyces aspartate transcarbamoylase is a dodecamer with dihydroorotase activity, *Curr. Microbiol. 39*, 175-179.

10. Van de Casteele, M., Chen, P., Roovers, M., Legrain, C., and Glansdorff, N. (1997) Structure and expression of a pyrimidine gene cluster from the extreme thermophile *Thermus* strain ZO5, *J. Bacteriol. 179*, 3470-3481.

11. Brabson JS, S. R. (1975) Purification and properties of Bacillus subtilis aspartate transcarbamylase, *J. Biol. Chem. 250*, 8664-8669.

12. Vickrey, J. F., Herve, G., and Evans, D. R. (2002) *Pseudomonas aeruginosa* aspartate transcarbamoylase. Characterization of its catalytic and regulatory properties, *J. Biol. Chem. 277*, 24490-24498.

13. Bergh, S. T., and Evans, D. R. (1993) Subunit structure of a class A aspartate transcarbamoylase from *Pseudomonas fluorescens*, *Proc. Natl. A*CAD. *Sci. U.S.A. 90*, 9818-9822.

14. Jones, M. E. (1980) The genes for and regulation of the enzyme activities of two multifunctional proteins required for the de novo pathway for UMP biosynthesis in mammals, *Mol. Biol. Biochem. Biophys. 32*, 165-182.

15. Simmons AJ, R. J., Piskur J, Davidson JN. (1999) A mutation that uncouples allosteric regulation of carbamyl phosphate synthetase in *Drosophila*, *J. Mol. Biol. 287*, 277-285.

16. Oster SK, H. C., Soucie EL, Penn LZ. (2002) The myc onocgene: marvelouslY complex, *Adv. Cancer Res. 84*, 81-154.

17. Henriksson M, L. B. (1996) Proteins of the Myc network: essential regulators of cell growth and differentiation, *Adv. Cancer Res. 68*, 109-182.

18. Miltenberger, R. J., Cortner, J., Farnham, P.J. (1993) An inhibitory Raf-1 mutant suppresses expression of a subset of v-raf-activa genes, *J. Biol. Chem. 268*, 15674-15680.

19. Miltenberger, R. J., Sukow, K. A., and Farnham, P. J. (1995) An E-box-mediated increase in CAD transcription at the G1/S-phase boundary is suppressed by inhibitory c-Myc mutants, *Mol. Cell Biol. 15*, 2527-2535.

20. Boyd, K. E., and Farnham, P. J. (1997) Myc versus USF: discrimination at the CAD gene is determined by core promoter elements, *Mol. Cell Biol. 17*, 2529-37.

21. Graves, L. M., Guy, H. I., Kozlowski, P., Huang, M., Lazarowski, E., Pope, R. M., Collins, M. A., Dahlstrand, E. N., Earp, H. S., 3rd, and Evans, D. R. (2000) Regulation of carbamoyl phosphate synthetase by MAP kinase, *Nature 403*, 328-332.

22. Carrey, E. A., and Hardie, D. G. (1988) Mapping of catalytic domains and phosphorylation sites in the multifunctional pyrimidine-biosynthetic protein CAD, *Eur. J. Biochem. 171*, 583-588.

23. Cervera, J., Bendala, E., Britton, H.G., Bueso, J., Nassif, Z., Lusty, C.J. and Rubio, V. (1996) Photoaffinity labeling with UMP of lysine 992 of carbamyl phosphate synthetase from *Escherichia coli* allows identification of the binding site for pyrimidine inhibitor, *Biochemistry 35*, 7247-7255.

24. Huang, M., Kozlowski, P., Collins, M., Wang, Y., Haystead, T. A., and Graves, L. M. (2002) Caspase-dependent cleavage of carbamoyl phosphate synthetase II during apoptosis, *Mol. Pharmacol. 61*, 569-577.

25. Jacobson, G. R., and Stark, G. R. (1975) Aspartate transcarbamylase of *Escherichia coli*. Mechanisms of inhibition and activation by dicarboxylic acids and other anions, *J. Biol. Chem. 250*, 6852-6860.

26. Huang, J. a. L., W. N. (2004) Aspartate Transcarbamylase (ATCase) of *Escherichia coli*: A new crystalline R-state bound to PALA, or to product analogues citrate and phosphate, *Biochemistry 43*, 6415-6421.

27. Swyryd, E. A., Seaver, S. S., and Stark, G. R. (1974) N-(Phosphonacetyl)-l-aspartate, a potent transition state analog inhibitor of aspartate transcarbamylase, blocks proliferation of mammalian cells in culture, *J. Biol. Chem. 249*, 6945-6950.

28. Christopherson, R. I., and Jones, M. E. (1980) The overall synthesis of L-5,6-dihydroorotate by multienzymatic protein pyr1-3 from hamster cells. Kinetic studies, substrate channeling, and the effects of inhibitors, *J. Biol. Chem. 255*, 11381-11395.

29. Mally, M. I., Grayson, D. R., and Evans, D. R. (1980) Catalytic synergy in the multifunctional protein that initiates pyrimidine biosynthesis in Syrian hamster cells, *J. Biol. Chem. 255*, 11372-11380.

30. Serre, V., Guy, H., Liu, X., Penverne, B., Herve, G., and Evans, D. (1998) Allosteric regulation and substrate channeling in multifunctional pyrimidine biosynthetic complexes: analysis of isolated domains and yeast-mammalian chimeric proteins, *J. Mol. Biol. 281*, 363-377.

31. Kelly, R. E., Mally, M. I., and Evans, D. R. (1986) The dihydroorotase domain of the multifunctional protein CAD. Subunit structure, zinc content, and kinetics, *J. Biol. Chem. 261*, 6073-6083.

32. Williams, N. K., Simpson, R. J., Moritz, R. L., Peide, Y., Crofts, L., Minasian, E., Leach, S. J., Wake, R. G., and Christopherson, R. I. (1990) Location of the dihydroorotase domain within trifunctional hamster dihydroorotate synthetase, *Gene 94*, 283-288.

33. Musmanno, L. A., Maley, J. A., and Davidson, J. N. (1991) Synthesis of the nonconserved dihydroorotase domain of the multifunctional hamster CAD protein in *Escherichia coli*, *Gene 99*, 211-216.

34. Zimmermann, B. H., and Evans, D. R. (1993) Cloning, overexpression, and characterization of the functional dihydroorotase domain of the mammalian multifunctional protein CAD, *Biochemistry 32*, 1519-1527.

35. Eriksson, A. E., Jones, T.A. and Liljas, A. (1988) Refined structure of human carbonic anhydrase II at 2.0 Å resolution., *Proteins: Struct., Funct., Genet. 4*, 274-282.

36. Williams, N. K., Manthey, M. K., Hambley, T. W., O'Donoghue, S. I., Keegan, M., Chapman, B. E., and Christopherson, R. I. (1995) Catalysis by hamster dihydroorotase: zinc binding, site-directed mutagenesis, and interaction with inhibitors, *Biochemistry 34*, 11344-11352.

37. Bauer, R., Limkilde, P., and Johansen, J.T. (1976) Low and high pH form of cadmium carbonic anhydrase determined by nuclear quadrupole interaction, *Biochemistry 15*, 334-342.

38. Huang, D. T., Thomas, M. A., and Christopherson, R. I. (1999) Divalent metal derivatives of the hamster dihydroorotase domain, *Biochemistry 38*, 9964-9970.

39. Christopherson, R. I., and Jones, M. E. (1980) The effects of pH and inhibitors upon the catalytic activity of the dihydroorotase of multienzymatic protein pyr1-3 from mouse Ehrlich ascites carcinoma, *J. Biol. Chem. 255*, 3358-3370.
40. Manthey, M. K., Huang, D. T., Bubb, W. A., and Christopherson, R. I. (1998) Synthesis and enzymic evaluation of 4-mercapto-6-oxo-1, 4-azaphosphinane-2-carboxylic acid 4-oxide as an inhibitor of mammalian dihydroorotase, *J. Med. Chem. 41*, 4550-4555.
41. Christopherson, R. I., Williams, N. K., Schoettle, S. L., Szabados, E., Hambley, T. W., and Manthey, M. K. (1995) Inhibitors of dihydroorotase, amidophosphoribosyltransferase and IMP cyclohydrolase as potential drugs, *Biochem. Soc. Trans. 23*, 888-893.
42. Christopherson, R. I., Schmalzl, K. J., Szabados, E., Goodridge, R. J., Harsanyi, M. C., Sant, M. E., Algar, E. M., Anderson, J. E., Armstrong, A., Sharma, S. C., and et al. (1989) Mercaptan and dicarboxylate inhibitors of hamster dihydroorotase, *Biochemistry 28*, 463-470.
43. Krungkrai, J., Krungkrai, S. R., and Phakanont, K. (1992) Antimalarial activity of orotate analogs that inhibit dihydroorotase and dihydroorotate dehydrogenase, *Biochem. Pharmacol. 43*, 1295-1301.
44. Adams, J. L., Meek, T. D., Mong, S. M., Johnson, R. K., and Metcalf, B. W. (1988) cis-4-Carboxy-6-(mercaptomethyl)-3,4,5,6-tetrahydropyrimidin-2(1 H)-one , a potent inhibitor of mammalian dihydroorotase, *J. Med. Chem. 31*, 1355-1359.
45. Kinder, D. H., Frank, S. K., and Ames, M. M. (1990) Analogues of carbamyl aspartate as inhibitors of dihydroorotase: preparation of boronic acid transition-

state analogues and a zinc chelator carbamylhomocysteine, *J. Med. Chem. 33*, 819-823.

46. Schroeder, P. E., Davidson, J. N., and Hasinoff, B. B. (2002) Dihydroorotase catalyzes the ring opening of the hydrolysis intermediates of the cardioprotective drug dexrazoxane (ICRF-187), *Drug. Metab. Dispos. 30*, 1431-1435.
47. Buss, J. L. a. H., B.B. (1993) The one-ring open hydrolysis product intermediates of the cardioprotective agent ICRF-187 (dexrazoxane) displace iron from iron-anthracycline complexes, *Agents Actions 40*, 86-95.
48. Brown, D. C., and Collins, K. D. (1991) Dihydroorotase from *Escherichia coli*. Substitution of Co(II) for the active site Zn(II), *J. Biol. Chem. 266*, 1597-1604.
49. Washabaugh, M. W., and Collins, K. D. (1984) Dihydroorotase from *Escherichia coli*. Purification and characterization, *J. Biol. Chem. 259*, 3293-3298.
50. Washabaugh, M. W., and Collins, K. D. (1986) Dihydroorotase from *Escherichia coli*. Sulfhydryl group-metal ion interactions, *J. Biol. Chem. 261*, 5920-5929.
51. Daniel, R., Caminade, E., Martel, A., Le Goffic, F., Canosa, D., Carrascal, M., and Abian, J. (1997) Mass spectrometric determination of the cleavage sites in *Escherichia coli* dihydroorotase induced by a cysteine-specific reagent, *J. Biol. Chem. 272*, 26934-26939.
52. Pradhan, T. K., and Sander, E. G. (1973) Noncompetitive inhibition by substituted sulfonamides of dihydroorotase from *Zymobacterium oroticum*, *Life Sci. 13*, 1747-1752.
53. Denis-Duphil, M. (1989) Pyrimidine biosynthesis in Saccharomyces cerevisiae: the ura2 cluster gene, its multifunctional enzyme product, and other structural or regulatory genes involved in de novo UMP synthesis, *Biochem. Cell Biol. 67*, 612-631.

54. 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.

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

56. 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.

57. Doublie, S. (1997) Preparation of selenomethionyl proteins for phase determination, *Methods Enzymol. 276*, 523-530.

58. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin, *J. Mol. Biol. 229*, 105-124.

59. 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.

60. 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.

61. Jabri, E., Carr, M.B., Hausinger, R.P. and Karplus, P.A. (1995) The crystal structure of urease from *Klebsiella aerogenes*, *Science 269*, 998-1004.

62. Hsu, C. S., Lai, W. L., Chang, W. W., Liaw, S. H., and Tsai, Y. C. (2002) Structural-based mutational analysis of D-aminoacylase from *Alcaligenes faecalis* DA1, *Protein Sci. 11*, 2545-2450.

63. 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.

64. 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.

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

66. Ko, T., Lin, J., Hu, C., Hsu, Y., Wang, A. and Liaw, S. (2003) Crystal structure of yeast cytosine deaminase, *J. Biol. Chem. 278*, 19111-19117.

67. 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.

68. Schwarzenbacher R, C. J., Brinen LS, Dai X, Deacon AM, Elsliger MA, Eshaghi S, Floyd R, Godzik A, Grittini C, Grzechnik SK, Guda C, Jaroszewski L, Karlak C, Klock HE, Koesema E, Kovarik JS, Kreusch A, Kuhn P, Lesley SA, McMullan D, McPhillips TM, Miller MA, Miller MD, Morse A, Moy K, Ouyang J, Robb A, Rodrigues K, Selby TL, Spraggon G, Stevens RC, van den Bedem H, Velasquez J, Vincent J, Wang X, West B, Wolf G, Hodgson KO, Wooley J, Wilson IA. (2003) Crystal structure of uronate isomerase (TM0064) from *Thermotoga maritima* at 2.85 Å resolution, *Proteins 53*, 142-145.

69. Kim, G. J., and Kim, H. S. (1998) Identification of the structural similarity in the functionally related amidohydrolases acting on the cyclic amide ring, *Biochem. J. 330 (Pt 1)*, 295-302.

70. Dudley, K. H., Butler, T. C. and Bius, D. L. (1974) The role of dihydropyrimidinase in the metabolism of some hydantoin and succinimide drugs., *Drug Metab. Dispos. 2*, 103-112.

71. Abendroth J, N. K., May O, Siemann M, Syldatk C and Schomburg D. (2002) Biochemistry 41, 8589-97. (2002) The structure of L-hydantoinase from *Arthobacter aurescens* leads to an understanding of dihydropyrimidinase substrate and enantio specificity, *Biochemistry 41*, 8589-8597.

72. 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.

73. Abendroth, J., Niefind, K. and Schomburg, D. (2002), *J. Mol. Biol. 320*, 143-156.

74. Gojkovic, Z., Rislund, L., Andersen, B., Sandrini, M. P., Cook, P. F., Schnackerz, K. D., and Piskur, J. (2003) Dihydropyrimidine amidohydrolases and dihydroorotases share the same origin and several enzymatic properties, *Nucleic Acids Res 31*, 1683-1692.

75. Janiyani, K., Bordelon, T., Waldrop, G. L., and Cronan, J. E., Jr. (2001) Function of *Escherichia coli* biotin carboxylase requires catalytic activity of both subunits of the homodimer, *J. Biol. Chem. 276*, 29864-29870.

76. Purcarea, C., Martin, P., Vickrey, J. F., Guy, H. I., Edwards, B. F., and Evans, D. R. (2002) Cloning, expression and preliminary X-ray analysis of the

dihydroorotase from the hyperthermophilic eubacterium *Aquifex aeolicus*, *Acta Crystallogr. D Biol. Crystallogr. 58*, 154-156.

77. Maher, M. J., Huang, D. T., Guss, J. M., Collyer, C. A., and Christopherson, R. I. (2003) Crystallization of hamster dihydroorotase: involvement of a disulfide-linked tetrameric form, *Acta Crystallogr. D Biol. Crystallogr. 59*, 381-384.

78. Ho, S. N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction, *Gene 77*, 51-59.

79. Schowen, K. B., and Schowen, R. L. (1982) Solvent isotope effects of enzyme systems, *Methods Enzymol. 87*, 551-606.

80. Weast, R. C., and Astle, M. J. (1985) *CRC Handbook of Data on Organic Compounds*, CRC Press, Boca Raton, FL.

81. Zimmermann, B. H., Kemling, N. M., and Evans, D. R. (1995) Function of conserved histidine residues in mammalian dihydroorotase, *Biochemistry 34*, 7038-7046.

82. Bodey, G. P., Bolivar, R., Fainstein, V., and Jadeja, L. (1983) Infections caused by *Pseudomonas aeruginosa*, *Rev. Infect. Dis. 5*, 279-313.

83. Stover, C. K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Spencer, D., Wong, G.K.S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V. (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen, *Nature 406*, 959-964.

84. Sander, E. G., Wright, L. D., and McCormick, D. B. (1965) Evidence for function of a metal ion in the activity of dihydroorotase from *Zymobacterium oroticum*, *J. Biol. Chem. 240*, 3628-3630.

85. Taylor, W. H., Taylor, M. L., Balch, W. E., and Gilchrist, P. S. (1976) Purification of properties of dihydroorotase, a zinc-containing metalloenzyme in Clostridium oroticum, *J Bacteriol 127*, 863-873.

86. Arnold, F. H., and Volkov, F.A. (1999) Directed evolution of biocatalysts, *Curr. Opin. Chem. Biol. 3*, 54-59.

87. Arnold, F. H. (1998) When blind is better: protein design by evolution, *Nat. Biotechnol. 16*, 617-618.

88. Arnold, F. H. (1996) Directed evolution: creating biocatalysts for the future, *Chem. Eng. Sci. 51*, 5091-5102.

89. Crameri, A., Raillard, S.A., Bermudez, E. and Stemmer, W.P.C. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution, *Nature 391*, 288-291.

90. van Kuilenburg, A. B. P., Meinsma, R., Zonnenberg, B. A., Zoetekouw, L., Baas, F., Matsuda, K., Tamaki, N., and van Gennip, A. H. (2003) Dihydropyrimidinase deficiency and severe 5-fluorouracil toxicity, *Clin. Cancer Res. 9*, 4363-4367.

91. Schmidt DM, M. E., Dojka M, Bermudez E, Ness JE, Govindarajan S, Babbitt PC, Minshull J, Gerlt JA. (2003) Evolutionary potential of (beta/alpha)$_8$-barrels: functional promiscuity produced by single substitutions in the enolase superfamily, *Biochemistry 42*, 8387-8393.

APPENDIX

MASTER ALIGNMENT OF DHO SEQUENCES

List of Organisms in Alignment

Trichodesmiumerythraeum
Nostoc punctiforme
Thermosynecoccuselongatus
Rhodopseudomonas palustris
Brucella suis
Rhodospirillum rubrum
Magnetospirillum magnetotacticum
Ralstonia metallidurans
Xanthomonas axonopodis
Xyllela fastidiosa
Pseudomonas flourescens
Cytophaga hutchinsonii
Saccharomyces cerevisiae
Schizosaccharomyces pombe
Ustilago maydis
Helicobacter pylori
Chloroflexus aurantiacus
Neisseria meningitidis
Beta vulgaris
Arabidopsis thaliana
Oryza sativa
Vibrio cholerae
Shewanella oneidensis
Microbulbifer degradans
Pseudomonas flourescens
Pseudomonas syringae
Pseudomonas aeruginosa
Azotobacterium vinelandii
Ralstonia solanacearum
Burkholderia fungorum
Agrobacterium tumefaciens
Sinorhizobium meliloti
Rhodobacter sphaeroides
Escherichia coli
Shigella flexneri
Salmonella typhimurium
Salmonella enterica
Yersinia pestis
Novosphingobium aromaticivoans
Buchnera aphidicola
Wigglesworthia brevipalpis
Plasmodium falciparum
Plasmodium yoelii
Toxoplasma gondii
Sulfolobus tokodaii
Streptococcus mutans
Geobacter
Clostridium acetobuytlicum
Thermoanaerobacter
Magnetococcus
Thermobifida fusca
Streptomyces coelicor
Corynebacterium glutamicum
Desulfovibrio desulfuricans
Bifidobacterium longum
Chlorobium tepidum
Thermus aquaticus
Bacillus caldolyticus
Enterococcus faecalis
Oceanobacillus iheyensis
Fusobacterium nucleatum
Lactobacillus leichmannii
Oenococcus oeni
Leuconostoc mesenteroides
Clostridium acetobuytlicum

Clustal W Protein Alignment

```
Trichodesmium      ---------- ---------- ---------- ------MAPI SS--LLIRRA
Nostoc             ---------- ---------- ---------- ------MSSP QI--LLIRRA
Thermosynecoccus   ---------- ---------- ---------- ---------- MA--IAIQNA
Rhodopseudomonas   ---------- ---------- ---------- ------MTGT FD--TILKGG
Brucella           ---------- ---------- ---------- ------MAAT FD--TILKGA
Rhodospirillum     ---------- ---------- ---------- -MERRDVAQS YD--LILKGG
Magnetospirillum   ---------- ---------- ---------S GSFAVFMAPV YEGWLYLKSG
Ralstonia          ---------- ---------- ---------- ----MSVAPT LHAATRNRGG
Xanthomonas        ---------- ---------- ---------- ---------- -MSRTVIVNA
Xyllela            MVSSTSAGTL PSVATGWPRP GTRQHQHTGQ SPFSSLYRTP PMNPTLIVNA
P. floruescens     ---------- ---------- ---------- ---------- -MSSVLIRNA
Cytophaga          ---------- ---------- ---------- ---------- -MGSIKISNA
Saccharomyces      ---------- ---------- ---------- ---------- ----------
Schizosacch.       ---------- ---------- ---------- ---------- ----------
Ustilago maydis    ---------- ---------- ---------- ---------- ----------
Helicobacter       ---------- ---------- ---------- ---------- ----------
Chloroflexus       ---------- ---------- ---------- ---------- ----------
Neisseria          ---------- ---------- ---------- ---------- ----------
Beta vulgaris      ---------- ---------- ---------- ---------- ----------
Arabidopsis        ---------- ---------- ---------- ---------- ----------
Oryza              ---------- ---------- ---------- ---------- ----------
Vibrio             ---------- ---------- ---------- ---------- ----------
Shewanella         ---------- ---------- ---------- ---------- ----------
Microbulbifer      ---------- ---------- ---------- ---------- ----------
P. flourescens     ---------- ---------- ---------- ---------- ----------
P. syringae        ---------- ---------- ---------- ---------- ----------
P. aeruginosa      ---------- ---------- ---------- ---------- ----------
Azotobacterium     ---------- ---------- ---------- ---------- ----------
Ralstonia          ---------- ---------- ---------- ---------- ----------
Burkholderia       ---------- ---------- ---------- ---------- ----------
A. tumefaciens     ---------- ---------- ---------- ---------- ----------
Sinorhizobium      ---------- ---------- ---------- ---------- ----------
Rhodobacter sp.    ---------- ---------- ---------- ---------- ----------
E. coli            ---------- ---------- ---------- ---------- ----------
Shigella           ---------- ---------- ---------- ---------- ----------
S. typhimurium     ---------- ---------- ---------- ---------- ----------
S. enterica        ---------- ---------- ---------- ---------- ----------
Yersinia           ---------- ---------- ---------- ---------- ----------
Novosphingobium    ---------- ---------- ---------- ---------- ----------
Buchnera           ---------- ---------- ---------- ---------- ----------
Wigglesworthia     ---------- ---------- ---------- ---------- ----------
Pl. falciparum     ---------- ---------- ---------- ---------- ----------
Pl. yoelii         ---------- ---------- ---------- ---------- ----------
Toxoplasma gondii  ---------- ---------- ---------- ---------- ----------
Sulfolobus         ---------- ---------- ---------- ---------- ----------
Streptococcus      ---------- ---------- ---------- ---------- ---MLLIKNG
Geobacter          ---------- ---------- ---------- ---------M N---LLIQGG
Clostridium        ---------- ---------- ---------- ---------M R---VLIKGG
Thermoanaerobacter ---------- ---------- ---------- ---------M R---IVIKNG
Magnetococcus      ---------- ---------- ---------- --------MM RTESLFIRGA
Thermobifida       ---------- ---------- ---------- --------MT TTTRYLIRNA
Streptomyces       ---------- ---------- ---------- ---------- -MSKTLIRGA
Corynebacterium    ---------- ---------- ---MVDSNTQ YPETGALAPA PADSLLISNV
Desulfovibrio      ---------- ---------- ---------- ------MSDH ISGTLFVSGA
Bifidobacterium    ---------- ---------- ---------- ---------- --MSITLRNI
Chlorobium         ---------- ---------- ---------- ---------- -MSTLFLN-A
Thermus aquaticus  ---------- ---------- ---------- ---------- ---MILIRNV
Bacillus           ---------- ---------- ---------- ---------- ---MNYLFKN
Enterococcus       ---------- ---------- ---------- ---------- ---MKTLIKN
Oceanobacillus     ---------- ---------- ---------- ---------- ---MKTILLN
Fusobacterium      ---------- ---------- ---------- ---------- ------MLLK
Lactobacillus      ---------- ---------- ---------- ---------- ---MATVIKN
Oenococcus         ---------- ---------- ---------- ---------- --MNSTLIKN
Clostridium        ---------- ---------- ---------- ---------- ---MNLLIKN
```

```
Trichodesmium      RILLPDG--- -----TFLIG DVQTQGR--- ---------- -----EIIQV
Nostoc             CIVLPNG--- -----EMMIG DVLTRDR--- ---------- -----QIVEV
Thermosynecoccus   VICTPEG--- -----ELRQQ QVRLEGD--- ---------- -----RIAEV
Rhodopseudomonas   TIVNQDG--- ------EGVG DIGIKDG--- ---------- -----RIAAL
Brucella           TIVNHDG--- ------IGQR DIGIRNG--- ---------- -----RIAAI
Rhodospirillum     TAVLPG---- -----RTMRA DIAVCGG--- ---------- -----KVVRV
Magnetospirillum   NSSLSGGIFP F-LGPRLLSS AAGEAGPTPF SVL------- -----RPFRF
Ralstonia          HSVTDGTVPG VGLPSARRHA DLLVIGGALV TPNGVERADV ACAAGRIVAL
Xanthomonas        RLVNEGK--- ------EFDA DLLIEGG--- ---------- -----RIAKI
Xyllela            RLVNEGR--- ------EFDS DLRIENG--- ---------- -----RIARI
P. flourescens     RLVNEGR--- ------EFDA DLLVSHG--- ---------- -----RIVKI
Cytophaga          NIVNEGQ--- ------IFQA DIWIEEG--- ---------- -----LIKQI
Saccharomyces      ---------- ---------- ---------- ---------- ----------
Schizosacch.       ---------- ---------- ---------- ---------- ----------
Ustilago maydis    ---------- ---------- ---------- ---------- ----------
Helicobacter       ---------- ---------- ---------- ---------- ----------
Chloroflexus       ---------- ---------- ---------- ---------- ----------
Neisseria          ---------- ---------- ---------- ---------- ----------
Beta vulgaris      ---------- ---------- ---------- ---------- ----------
Arabidopsis        ---------- ---------- ---------- -----MIKTL VSPYSGFGSQ
Oryza              ---------- ---------- ---------- ---------- ----------
Vibrio             ---------- ---------- ---------- ---------- ----------
Shewanella         ---------- ---------- ---------- ---------- ----------
Microbulbifer      ---------- ---------- ---------- ---------- ----------
P. flourescens     ---------- ---------- ---------- ---------- ----------
P. syringae        ---------- ---------- ---------- ---------- ----------
P. aeruginosa      ---------- ---------- ---------- ---------- ----------
Azotobacterium     ---------- ---------- ---------- ---------- -----MPPEA
Ralstonia          ---------- ---------- ---------- ---------- ----------
Burkholderia       ---------- ---------- ---------- ---------- ----------
A. tumefaciens     ---------- ---------- ---------- ---------- ----------
Sinorhizobium      ---------- ---------- ---------- ---------- ----------
Rhodobacter sp.    ---------- ---------- ---------- ---------- ----------
E. coli            ---------- ---------- ---------- ---------- ----------
Shigella           ---------- ---------- ---------- ---------- ----------
S. typhimurium     ---------- ---------- ---------- ---------- ----------
S. enterica        ---------- ---------- ---------- ---------- ----------
Yersinia           ---------- ---------- ---------- ---------- ----------
Novosphingobium    ---------- ---------- ---------- ---------- ----------
Buchnera           ---------- ---------- ---------- ---------- ----------
Wigglesworthia     ---------- ---------- ---------- ---------- ----------
Pl. falciparum     ---------- ---------- ---------- ---------- ----------
Pl. yoelii         ---------- ---------- ---------- ---------- ----------
Toxoplasma gondii  ---------- ---------- ---------- ---------- ----MAMTGE
Sulfolobus         ---------- ---------- ---------- ---------- ----------
Streptococcus      RVLDPKSGLD RVLDVLLDGK KVVKIA---- ---------- -----ENIES
Geobacter          RVIDPSQGID EVLDILVENG AVKELG---- ---------- -------KGV
Clostridium        HVVDPKTNTN GIMDILVEDG IITEIG---- ---------- -------KDI
Thermoanaerobacter IVIDG-FGGE EKADILINYG IIKGID---- ---------- -------KNI
Magnetococcus      RVVDPANERD EIADILIVDG VIREIG---- ---------- -------K-L
Thermobifida       RILG-----G EPTDLLLAHG RIDAI----- ---------- -------GPD
Streptomyces       KVLG-----G EPQDVLIDGT VVEAV----- ---------- -------GTN
Corynebacterium    LVYGE----G EPTNVFVKDG VIAAI----- ---------- -------GGT
Desulfovibrio      LLSG------ RTVDVTVSGG RIAAVS---- ---------- -----EHGAA
Bifidobacterium    KVWNTGEVID LVVPGTAAAY FADTSV---- ---------- -----VADG-
Chlorobium         RLLNPAENLD -TVGSIKIGD DGLIEA---- ---------- -----VATGG
Thermus aquaticus  RLVDAMGERG --PLDVLIGE -GRILS---- ---------- -----LEGG-
Bacillus           GRYMNEEGKI VATDLLVQDG KIAKVA---- ---------- -----ENIT-
Enterococcus       GQINTRKNMT TPAEIWIEDG RIKAIG---- ---------- -----TGFSE
Oceanobacillus     ANRLLPSNEL EKVEVLIEDN KIIKIA---- ---------- -----SQIP-
Fusobacterium      NCKILKNTKF EKVDILIRDN KIEKIS---- ---------- -----ENID-
Lactobacillus      GTVYQN-GRL IKADVLIEGK KIKAIG---- ---------- -----TDLD-
Oenococcus         AQILVGENSF LKRDLFIEKG KISAIG---- ---------- -----DGLS-
Clostridium        VNLIDESNNF FG-DIYIEKG VIKELG---- ---------- -----TELN-
```

```
Trichodesmium      APEISSSEAP EKIIDAEGLT LLPGVIDPQV HFREPGLEHK EDLFTASRAC
Nostoc             APEISQT-AV VTEIDAEGLT LLPGVIDPQV HFREPGLEHK EDLFTASCAC
Thermosynecoccus   AETVTVN-PS DTVIDATGLT LLPGVIDPQV HFREPGLEHK EDLFTASCAC
Rhodopseudomonas   G-SLGA-EKA GETIDCRGLH VLPGVIDTQV HFREPGLTHK EDLETGSRSA
Brucella           G-SLAG-NSA GEVIDCTGLH ILPGVVDSQV HFREPGLEHK EDLETGSLAA
Rhodospirillum     G-FLDA-GSA RQTIDCAGLH VLPGVIDTQV HFREPGLEHK EDLESGTRAA
Ralstonia          G-DLQCSWTA DATLDAKGLH VLPGVIDSQV HFREPGLTHK ETLEAGTRGA
Xanthomonas        DSKIVP-APG DTVVDAAGRW VLPGMIDDQV HFREPGLTHK GDIATESGAA
Xyllela            DTGLSA-QSK DIVIDAAGRR LLPGMIDDQV HFREPGLTHK GDIVSESAAA
P. flourescens     AGSIEG-ENA TREIDATGQW LLPGMIDDQV HFREPGAPAK GSVHTESRAA
Cytophaga          G-KITA-IA- DQTIDASGKY LFPGVIDDQV HFREPGLTHK ATIYTEAKAA
Saccharomyces      ---------- -----MVQEI DLGLTCDMHV HVREGAMCEL VTPKIRDGGV
Schizosacch.       ---------- -----MSLKI PG--LADMHV HLRQDNMLKA VVPTVAEGGV
Ustilago maydis    ---------- ----MSVQEI TIPAPGDFHV HLRQGKMSEL VTPHVAEGGV
Helicobacter       ---------- -------MEI TLFDPIDAHL HVRENALLKA VLGYSSEP-F
Chloroflexus       ---------- -------MTV TLFAPLDMHL HLREGAMLRL VAPFSAAQ-F
Neisseria          ---------- ------MQTL TIIRPDDMHL HLRDGDALKA VAPYTARQ-M
Beta vulgaris      ---------- -------MEL TLTRPDDWHL HLRDGDLLAA VAPHSARH-F
Arabidopsis        KLKFDRSSEK VKTRAVRMEL TITQPDDWHL HLRDGDLLHA VVPHSASN-F
Oryza              -------MAA SPSPPPLQEL TITRPDDWHL HLREGDVLAA VLPHSAMH-F
Vibrio             --MTT----- ---------L TITRPDDWHV HLRDGDVLAD TVRDISRY-N
Shewanella         --MKT----- ---------L TITRPDDWHI HLRDGAQLTD TVRDISRY-M
Microbulbifer      --MATGLNAN KQQELMSEKI TLTAPDDWHI HLRDGDALAY TVSDAARN-F
P. flourescens     ---------- -----MSDRL TLLRPDDWHI HLRDGAVLTN TVADVART-F
P. syringae        ---------- -----MSDRL TLLRPDDWHI HLRDGAVLPH TVADVART-F
P. aeruginosa      ---------- -----MSDRL TLLRPDDWHI HLRDGAALAN TVGDAART-F
Azotobacterium     FSGRMHGSLT FLEVSMPDRL ILLRPDDWHI HLRDGAALSR TVTDAART-F
Ralstonia          ---------- -----MIDTL TIVRPDDWHL HLRDGDALAD VVGDTARQ-F
Burkholderia       ---------M TASNASLDSI TLARPDDWHL HVRDGAMLAA VLPDTARQ-F
A. tumefaciens     ---------- ------MKSL TLRRPDDWHL HLRDGAMLEG VIGDTSRH-F
Sinorhizobium      ---------- ------MQDL VIRRPDDWHL HLRDGGMLRG VIADTSRH-F
Rhodobacter sp.    ---------- -----MTQSL TLRRPDDWHL HLRDGAMLEG VLPETTRH-F
E. coli            ---------- --MTAPSQVL KIRRPDDWHL HLRDGDMLKT VVPYTSEI-Y 37
Shigella           ---------- --MTAPSQVL KIRRPDDWHL HLRDGDMLKT VVPYTSEI-Y
S. typhimurium     ---------- --MTAPSQVL KIRRPDDWHV HLRDGDMLKT VVPYTSEI-Y
S. enterica        ---------- --MTAPSQVL KIRRPDDWHV HLRDGDMLKT VVPYTSEI-Y
Yersinia           ---------- --MTAQPQTL KIRRPDDWHI HLRDDEMLST VLPYTSEV-F
Novosphingobium    ---------- -----MTETL TIRRPDDWHV HLRDRDVLRG VVPYTARQ-F
Buchnera           ---------- --MSKFVKKI KIIKPDDWHV HLRDNEILNQ VIKYTGKF-Y
Wigglesworthia     ---------- -------MLL KIIKPDDFHI HLREKEILKK ILPYTSQF-F
Pl. falciparum     ---------- -----MKNYF YIPIADDMHC HLRQGDMLDF TVNSIRRGGC
Pl. yoelii         ---------- -----MENGF SIPLADDMHC HLRQDEMLKF TVPAIKKGGC
Toxoplasma gondii  EKGGNGASNS CCGQARAKPL VMPLISDMHT HLRQDEMAEF VTPMIKKGGC
Streptococcus      VD-------- AQVIDAEGLV VAPGLVDVHV HFREPGQTHK EDIHTGALAA
Geobacter          KAPSGT---- -ETIDASGLI VTPGLIDMHV HLRDPGHEYK EDIVSGTKAA
Clostridium        EISNG----- -DIIYAEGKL VLPGLVDAHC HLRDPGFEYK EDIETGTMSA
Thermoanaerobacter DVSDA----- -IVIDAEGKY VLPGFVDMHT HLRQPGFEEK ETIRTGTESA
Magnetococcus      DAPAHV---- -PVIEADGLI AAPGLVDMHV HLREPGYEYK ETIAGGTRAA
Thermobifida       LPADGA---- -EVVDAAGAI ALPGLVDLHT HLREPGREDA ETVATGSRAA
Streptomyces       LSAEGA---- -EVVEADGKV LLPGLVDLHT HLREPGREDS ETVLTGTRAA
Corynebacterium    HEAD------ -RTIDGNGGV LLPGFVDMHV HLREPGREDT ETIATGSAAA
Desulfovibrio      PAPQHA---- -ETVEAAGKI LFPSFIDCHV HLREPGFEYK EDIASGLAAA
Bifidobacterium    ---------- -ADIDATGLT VAPGFEDPHV HFRDPGQTYK ESMVSGCRAS
Chlorobium         ESIPAK--AE DNVIDLAGKV LAPGLFDMHC HFREPGQEYK ETLETGSAAA
Thermus aquaticus  --------EA KQVVDGTGCF LAPGFLDLHA HLREPGEEVK EDLFSGLLAA
Bacillus           -------ADN AEVIDVNGKL IAPGLVDVHV HLREPGGEHK ETIETGTLAA
Enterococcus       -------AEF DEVFDAKGQL ITPGLVDVHV HLREPGFTYK ETIEAGTRSA
Oceanobacillus     -------DKV ENYIDIEGKL LLPGLIDVHI HLREPGGEHK ETIKTGTMAA
Fusobacterium      -------ITD ENIIDIKNRF VTAGFIDVHV HWREPGFSKK ETVYTASRAA
Lactobacillus      -------AE- -KIIDAQGML VSPGLVDVHV HYRDPGQTYK EDIKTGSEAA
Oenococcus         -------NKA DFIVDAKGAL LLPGLIDVHV HFREPGFPEK ETIASGSKVA
Clostridium        --------KE CETLDGKGLV LMPAFIDTHA HFRDPGFEYK EDIESGSKAA
```

```
Trichodesmium      VKGGVTSFLE MPNTKPLTT- ----TQGALD DKLRRAEQKC VAN-------
Nostoc             AKGGVTSFLE MPNTRPLTT- ----TQQALD DKLERASQKS LVN-------
Thermosynecoccus   AKGGVTSFLE MPNTRPLTI- ----DQASLE NKLARAAAKC VVN-------
Rhodopseudomonas   VMGGVTAVFE MPNTNPLTI- ----TEETFT DKVKRAEHRM HCD-------
Brucella           VLGGVTAVFE MPNTKPLTT- ----SAEMLE DKIRRGRHRM HCD-------
Rhodospirillum     VLGGVTGVFE MPNTAPATD- ----SPEALE EKLTRAQGRT WCD-------
Magnetospirillum   AQAGVVAIFE MPNTKPSTT- ----TADAIL DKLARAKGRA WVD-------
Ralstonia          ALGGVTAVFE MPNTQPLTL- ----RAEDLR AKLDAAQGRA WCD-------
Xanthomonas        VAGGLTSFMD MPNTNPPTL- ----DAAALQ AKYDAAAGRA WAN-------
Xyllela            VAGGLTSFMD MPNTNPPTL- ----GAAALQ AKYDAARGRA WGN-------
P. flourescens     VAGGITSFMD MPNTHPATL- ----TLEALA DKKRRAAINS VAN-------
Cytophaga          VAGGVTSYME MPNTKPAAV- ----TQELLE QKYEIAAKTS LAN-------
Saccharomyces      SIAYI----- MPNLQPPITTLDRV- ----IEYKKT LQKLAP---K TTF--
Schizosacch.       SVAYV----- MPNLIPPITTVDAC- ----LQYKKE IEQLDS---K TTY--
Ustilago maydis    SLAYV----- MPNLVPPITSTQQA- ----MEYLER LAAVAP---Q TMF--
Helicobacter       SAAVI----- MPNLSKPLIDTPTT- ----LEYEEE ILNHSS---N FKP--
Chloroflexus       AGAVI----- MPNLV PPVDNADRL- ----ARYHEE IRAAVDP-YP FLP-
Neisseria          GRAVI----- MPNLK PPVVSVADA- ----LAYKAR IMAALPEGSA FEP-
Beta vulgaris      GRAIV----- MPNLR PPVTTTGAA- ----IAYRKS IMEVLPDDSD FNP-
Arabidopsis        KRAIV----- MPNLK PPVTSTAAA- ----IIYRKF IMKALPSESS FDP-
Oryza              GRAIV----- MPNLK PPVTTTARA- ----LEYREE ILRALPPGSN FVP-
Vibrio             GRALI----- MPNTV PPVTTTEMA- ----LAYRER IMAAQPQ-AH FEP-
Shewanella         GRAIV----- MPNLV PPAIDTETA- ----LAYYDR IKARVPAGSQ FEP-
Microbulbifer      RRAII----- MPNLV PPVLNAKQA- ----LDYKAR ILAHAPEDAD FTP-
P. flourescens     GRAII----- MPNLV PPVRNAAEA- ----DGYRQR ILAARPAGSR FEP-
P. syringae        GRAII----- MPNLV PPVRNAQQA- ----DAYRQR ILAARPATSR FEP-
P. aeruginosa      GRAIV----- MPNLV PPVRNAAEA- ----DAYRQR ILAARPAASR FEP-
Azotobacterium     GRAIV----- MPNLV PPVRNASEA- ----GAYRQR IQAARPADSR FEP-
Ralstonia          GRAII----- MPNLK PPVTTTAQA- ----RAYRER ILAALPAGTR FEP-
Burkholderia       GRAII----- MPNLK PPVTTTAMA- ----QAYRER IVTAIPEGAK FEP-
A. tumefaciens     ARAII----- MPNLV PPVVTTADA- ----SAYRER ILKAIPEGDR FEP-
Sinorhizobium      ARAII----- MPNLV PPVVTSADA- ----AAYRER ILAAIPAGDR FEP-
R. sphaeroides     ARAIV----- MPNLV PPVVTAAEA- ----QAYRAR ILRALPEGAR FEP-
E. coli            GRAIV----- MPNLA PPVTTVEAA- ----VAYRQR ILDAVPAGHD FTP- 75
Shigella           GRAIV----- MPNLA PPVTTVEAA- ----VAYRQR ILDAVPAGHD FTP-
S. typhimurium     GRAIV----- MPNLA SPITTVDAA- ----IAYRQR ILDAGPAGHD FTP-
S. enterica        GRAIV----- MPNLA SPITTVDAA- ----IAYRQR ILDAVPAGHD FTP-
Yersinia           ARAIV----- MPNLA QPITTVASA- ----IAYRER ILAAVPAGHK FTP-
Novosphingobium    ARAIV----- MPNLS PPMTDVAGV- ----AAYRDR ILAALPQGSA FTP-
Buchnera           KRAVI----- MPNLN SPITSCLKS- ----IAYRNR ILKSMHLNYK FKP-
Wigglesworthia     GRALV----- MPNLN QPIINSHFA- ----SVYKKE ICSFIPKLHK FNP-
Pl. falciparum     NRVLV----- MPNTH PIISTCSDA- ----QKYLYQ LKSRDD---D IEY-
Pl. yoelii         NRVLV----- MPNTT PIISSCEEA- ----KKYRDE LIKYDN---S IEY-
Toxoplasma gondii  RCVLV----- MPNTI PPVTTCAQA- ----AAYRER LVRIDP---N VDY-
Streptococcus      AAGGVTTVVM MANTTPTIS- ----DVATLT EVLESAAKEN -IH-------
Geobacter          AAGGFTSVAC MPNTKPVND- ----NKAVTS YIIAKAKAEG SVN-------
Clostridium        AMGGFTSIAC MPNTDPVCD- ----NKAVVK YIINKAKQDG YVN-------
Thermoanaerobacter AAGGFTTVAC MPNTNPPID- ----SEVVVE YVKAVAQREG VVK-------
Magnetococcus      AAGGVTSVAA MPNTKPVND- ----DPSVTG YMLDKARVAG FAN-------
Thermobifida       AMGGYTAVFA MANTDPVAD- ----TAGVVE QVWRLGRDAG YCD-------
Streptomyces       ASGGYTNVFA MANTFPVAD- ----TAGVVE QVWRLGQESG YCD-------
Corynebacterium    AKGGFTAVFT MANTTPVMD- ----QPVIAE SVWFKGQNIG LCD-------
Desulfovibrio      AHGGFGAVLP MANTSPVND- ----QGSVTE LMLERARKAW PHGP------
Bifidobacterium    ASGGYTNVLI MPNTLPALDG QTVSGPEATG AKEVLDAGFD NVIDFLQQYD
Chlorobium         VAGGFTGVAL MPNTRPVID- ----SPLGVA YIRHHSAGL- ----------
Thermus aquaticus  VRGGYTDLVS MPNTNPPVD- ----TPEAVR ALKEKAKALG ----------
Bacillus           AKGGFTTICA MPNTRPVPD- ----CREHME DLQNRIKEKA HVN-------
Enterococcus       ARGGFTTVCA MPNLNPVPD- ----TAEKLR QVYDIIRKDA VVK-------
Oceanobacillus     ARGGYTTVCA MPNTNPVPD- ----HPEALT SLLSKIADDA HIR-------
Fusobacterium      ARGGFTTVMT MPNLNPVPD- ----SVETLN KQLEIIKKDS VIR-------
Lactobacillus      ARGGFTTVGA MPNVTPVPN- ----TPELMK KMVEENKHKG VVH-------
Oenococcus         ARGGFTTVFA MPNLNPVID- ----NVEVFK QVQALNQQDG IIK-------
Clostridium        VRGGYTTVTL MPNTKPVCS- ----SKEILD YVVNKGKEVG LVD-------
```

```
Trichodesmium      ---------- -FGFFIGATA ENLPDLLTAN P---TPG--- ----------
Nostoc             ---------- -YGFFIGATA ENLPDLLLAK P---TPG--- ----------
Thermosynecoccus   ---------- -YGFFIGATK DNLAVLNTVH P---VCG--- ----------
Rhodopseudomonas   ---------- -FAFFIGGTR DNVEELPKLE RARGCCG--- ----------
Brucella           ---------- -FAFWVGGTR DNAKDVAELE RLPGAAG--- ----------
Rhodospirillum     ---------- -HAFFLGASA ANAAALGGWE RRDGCAG--- ----------
Magnetospirillum   ---------- -HAFFIGAAS DNVDHLAQWE RIPGCAG--- ----------
Ralstonia          ---------- -HAFYIGGSS VNAEHLEYLE SLPGCAG--- ----------
Xanthomonas        ---------- -YGFYMGASN DNLAHIQSLD PKT-APG--- ----------
Xyllela            ---------- -YGFYLGASN DNLAAIQALD PKT-SPG--- ----------
P. flourescens     ---------- -YGFHFGVSH DNLDTVAALN PSE-VAG--- ----------
Cytophaga          ---------- -YSFFMGTTN SNINELLKTN PAT-VCG--- ----------
Saccharomyces      ---------- -LMSFYLSKD LTPDLIHEAA QQHAIRG--- ----------
Shizosaccharomyces ---------- -LMSLYLSPE TTPEVIYEAA KK-GIRG--- ----------
Ustilago maydis    ---------- -VGTLYLSPD LTPAEIAEGA QN-GVRG--- ----------
Helicobacter       ---------- -LMSLYFNDG LTLEELQCAK EKG-VRF--- ----------
Chloroflexus       ---------- -LMTLFFRP- YDAQTLQALR EH--IFA--- ----------
Neisseria          ---------- -LMTLYLTDN ATPELVREAK AAGIV-A--- ----------
Beta vulgaris      ---------- -LMTLYLTDT TSPNEIKLAR KSEVVYA--- ----------
Arabidopsis        ---------- -LMTLYLTDK TLPEEIRLAR ESGVVYA--- ----------
Oryza              ---------- -LMTLYLTDN TSPEEIKLAK KSGVVFA--- ----------
Vibrio             ---------- -LMALYLTDN TSPEEIRKAK ASGKVVA--- ----------
Shewanella         ---------- -LMVLYLTDK TSPDEIRKAK ASGKVFA--- ----------
Microbulbifer      ---------- -LMVLYLTEK TSPADIAEAA AKG-IVA--- ----------
P. flourescens     ---------- -LMVLYLTDR TQPEEIREAK ASGFVHA--- ----------
P. syringae        ---------- -LMVLYLTDQ TTPDDIRTAK ASGFVYA--- ----------
P. aeruginosa      ---------- -LMVLYLTDR TSTEEIRTAK ASGFVHA--- ----------
Azotobacterium     ---------- -LMTLYLTDK TSPEDIRTAK AQGFVHA--- ----------
Ralstonia          ---------- -LMTLYLTDN TTPEEVRAAR ASGFVHG--- ----------
Burkholderia       ---------- -LMTLYLTDN TPPDEIRRAR ESGFVHG--- ----------
A. tumefaciens     ---------- -LMTLYLTED TVADDVEEGK KSGLITA--- ----------
Sinorhizobium      ---------- -LMTLYLTEG TDPGDVEAGF RSGLVKA--- ----------
Rhodobacter sp.    ---------- -LMVLYLTET TDPADVRAAA ASGLVTA--- ----------
E. coli            ---------- -LMTCYLTDS LDPNELERGF NEGVFTA--- ---------- 101
Shigella           ---------- -LMTCYLTDS LDPNELERGF NEAVFTA--- ----------
S. typhimurium     ---------- -LMTCYLTDS LDADELERGF HEGVFTA--- ----------
S. enterica        ---------- -LMTCYLTDS LDADELERGF HEGVFTA--- ----------
Yersinia           ---------- -LMTCYLTNS LDAKELTTGF EQGVFTA--- ----------
Novosphingobium    ---------- -LMTLYLTDS TDIEEVARGF AEGVFVA--- ----------
Buchnera           ---------- -LMTCYLTNS TSPKELEFGF SKKIFVA--- ----------
Wigglesworthia     ---------- -LMTLYLTEN CDKKMLVDGF LNKIFIA--- ----------
Pl. falciparum     ---------- -LMTLYLNKN TDENDILSNY YKCNLQG--- ----------
Pl. yoelii         ---------- -LMTLYLNKK TDENDILNNY KECNLQG--- ----------
Toxoplasma gondii  ---------- MMTLFLSPEV SADDLRQNAK MCHVTGFENY SRTDFGIYSG
Sulfolobus         ---------- -IGLIIDMPN TLPPVNTYER VIERIRE--- ----------
Streptococcus      ---------- VFPVGSITQG SKGELLSEMG EL-------- ----------
Clostridium        ---------- VYPIGAISKG QKGEELSEIG EL-------- ----------
Thermoanaerobacter ---------- VLPIGAMTKG MKGEEITEMA KLK------- ----------
Magnetococcus      ---------- LFPIGAVSKG LQGKEITEMG LLQ--AAG-- ----------
Thermobifida       ---------- VHPVGAVTVG L-- --- ----------
Streptomyces       ---------- VQPIGAVTVG LEGAKLAELG AMHESAAGVT VFSDDG----
Corynebacterium    ---------- VHPVGSITKG LEGKELTEFG MMARSEAKVR MFSDDG----
Desulfovibrio      ---------R VHPVGAATKG L--------- ---------- ----------
Bifidobacterium    TAHDVQLPVR YDLCVCASKD RAGHEASDVA DWLKYVPGFE DDA-------
Chlorobium         -------PID LEVIGAMTVE SRGEALAPYG KYASYS---- ----------
Thermus aquaticus  -------LAR LHPAAALTEK QEG------- ---------- ----------
Bacillus           ---------- VLPYGAITVR QAGSEMTDFE TL-------- ----------
Enterococcus       ---------- VLQYAPITEN LRSE------ ---------- ----------
Oceanobacillus     ---------- VLPYASITKS L--------- ---------- ----------
Fusobacterium      ---------- AIPYGAITKE EYGRELSDME AIASN----- ----------
Lactobacillus      ---------- IFQYGPITND ETTDIIPDYA AL-------- ----------
Oenococcus         ---------- IKQYAAISTG LTAN------ ---------- ----------
Clostridium        ---------- LYQTVSITKN LSGEEINHLR EFEGNPN--- ----------
```

```
Trichodesmium      --IKIFMGSM HGPLLVDTEE KLEPIFARGK RLIAVHAENQ ARIDE----R
Nostoc             --IKIFMGSM HGQLLVDGET TLETIFAKGD RLIAVHAEDQ ARINQ----R
Thermosynecoccus   --IKIFMGSM HGPLLVDEEP ILDRIFSEGK RLIAVHAEDQ GRIRA----R
Rhodopseudomonas   --VKVFIGSS TGSLLVEDDP SLKRILSVIQ RRAAFHAEDE YRLND----R
Brucella           --IKVFMGSS TGDLLVEDDD GVRSILRNTR RRAAFHSEDE FRLKE----R
Rhodospirillum     --VKIFMGSS TGSLLVDQDE AIARVLAGGF RRVAVHCEDE ARLKE----R
Magnetospirillum   --IKVFMGSS TGNLLVADDE ---------- ----TLGRVL AQGFRRVAVH
Ralstonia          --VKIFMGSS FGDLLADDDE ---------- ----VLCRIL AHGRRRMAVH
Xanthomonas        --IKVFMGAS TGNMLVDNPE ---------- ----TLDAIF RDAPTPIITH
Xyllela            --IKVFMGAS TGNMLVDNET ---------- ----TLEGIF RHAPTPIITH
P. flourescens     --VKVFMGAS TGNMLVDDPH ---------- ----TLERLF AEVPTILLAH
Cytophaga          --IKIFMGSS TGDMLVDNSA ---------- ----MLDEIF SQVKMLIAIH
Saccharomyces      --VKCYPAGV TTNSAAGVDP ------ND-F SAFYPIFKAM QEENLVLNLH
Schizosacch.       --VKSYPKGA TTNSESGVES ---------Y EPFYPTFAAM QETGMILNIH
Ustilago maydis    --VKSYPRGV TTNSDSGIED ---------Y ETYYPIFEEM QKHDMVLNLH
Helicobacter       --LKLYPKGM TTNAQNGTSD ------LLG- EKTLEVLENA QKLGFILCIH
Chloroflexus       --IKLYPEGV TTNSAGGVSD ------LTAI EPTLAMME-- -ELGIPLLVH
Neisseria          --FKLYPAGA TTNSDSGVTD ------LF-- -KLIPVLEEM AKQGILFLVH
Beta vulgaris      --VKLYPAGA TTNSQDGVTD ------LLG- -KCLPVLEEM AEQDMPLLVH
Arabidopsis        --VKLYPAGA TTNSQDGVTD ------LFG- -KCLPVLEEM VKQNMPLLVH
Oryza              --VKLYPSGA TTNSQDGVTD ------IFG- -KCLPVLEEM ARQEMPLLVH
Vibrio             --AKLYPAGA TTNSDSGVTS ------AK-- -NIYPVLQAM QEVGMLLLVH
Shewanella         --AKLYPAGA TTNSDSGVTD ------LK-- -NIYPALEAM QEVGMLFLVH
Microbulbifer      --CKLYPAGA TTNSDSGVTD ------IK-- -NCYDALAAM QEHNIKLLVH
P. flourescens     --AKLYPAGA TTNSDSGVTS ------ID-- -KIFPVLEAM AEAGMPLLIH
P. syringae v.     --AKLYPAGA TTNSDSGVTS ------ID-- -KIFPALEAM ADVGMLLLVH
P. aeruginosa      --AKLYPAGA TTNSDSGVTR ------ID-- -NIFEALEAM AEVGMPLLVH
Azotobacterium     --AKLYPAGA TTNSDAGVTR ------ID-- -NIFPILEAM AEAGLPLLVH
Ralstonia          --VKLYPAGA TTNSDAGVTD ------LR-- -RCAKTLEAM QDVGMPLLVH
Burkholderia       --VKLYPAGA TTNSDAGVTD ------IM-- -KCAKTLEVM QELGMPLLVH
A. tumefaciens     --VKLYPAGA TTNSHGGVRD ------FN-- -KAMPVLERM AKIGLPLCVH
Sinorhizobium      --VKLYPAGA TTNSSSGVRD ------ID-- -KAMPVLERM AEIGLPLCVH
Rhodobacter sp.    --VKLYPAGA TTNSASGVRD ------FD-- -RVRGVLETM AEIGLPLCVH
E. coli            --AKLYPANA TTNSSHGVTS ------ID-- -AIMPVLERM EKIGMPLLVH 140
Shigella           --AKLYPANA TTNSSHGVTS ------ID-- -AIMPVLERM EKIGMPLLVH
S. typhimurium     --AKLYPANA TTNSSHGVTS ------VD-- -AIMPVLERM EKLGIPLLVH
S. enterica        --AKLYPANA TTNSSHGVTS ------VD-- -AIMPVLERM EKLGMPLLVH
Yersinia           --AKLYPANA TTNSTHGVSD ------IP-- -AIYPLFEQM QKIGMPLLIH
Novosphingobium    --AKLYPAHA TTGSAHGVTD ------IR-- -NIYPVLEKM QEIGMPLLIH
Buchnera           --AKFYPNGC TTNSKTGIKK ------IS-- -DITPVLECM EKIGMPLLIH
Wigglesworthia     --AKMYISNT TTNSEKGIKN ------FE-- -NIFHILEIM QKIGMILSVH
Pl. falciparum     --VKIYPSNV TTNSSDGITS ------LEP- --YYKVFHAL EKLNKSIHIH
Pl. yoelii         --IKIYPSNV TTNSNDGVSN ------LEP- --YYKIFSTL EKINKSLHIH
Toxoplasma gondii  --IKSYPKGV TTNSDQGVES ------YEQ- --YYGIFEAM QELGLTLHLH
Sulfolobus         -VTKEIEKID TLPIAGYKIY PEDLDRTETK VLLE------ ---------H
Streptococcus      ---KAGAVGF SDDGIPLTSS KVVKEALDLA ---------- KKNDTFISLH
Geobacter          ---KESGCVA VSDDGHPVTN SELMRRALEY ---------A KGMGIMVISH
Clostridium        ---KFAGAVA ISDDGKPVKS SSLMKRALEY ---------S SMFDIAVISH
Thermoanaerobacter ---KAGVVAL SDDGFPIMSA GIMKRVMTYG ---------- KMYDLLMITH
Magnetococcus      ---KCVAFSD DGLPIMNSGL MRRALDYSRA FG-------- ----GLIIQH
Thermobifida       ---KGERLAE LGAMADSPAV RVFSDDGMCV SDAQLMRRAL EYVKAFDAQH
Streptomyces       ---KCVHDAV IMRRALEYVK AFN------- ---------- ----GVVAQH
Corynebacterium    ---KCVDDPQ VMRRALEYAK GMD------- ---------- ----VLIAQH
Desulfovibrio      ---KGEELAP MGELAAAGCV AFSNDGLPVG GAEMFRRCME YAADKIVIDH
Bifidobacterium    ---KTPAMLT HPITAISDDG SAVTPEILDQ VLENVKASD- ----LYLIEH
Chlorobium         --VKAVSDDG TAIQSSQIMR LAIEYAANFD ---------- ----LLLIQH
Thermus aquaticus  ---KTLTPAG LLQEAGAVLL TDDGRTNEDA GVLAAGLLMA APLGLPVAVH
Bacillus           ---KELGAFA FTDDGVGVQD ASMMLAAMKR ---------A AKLNMAVVAH
Enterococcus       ---KLVDQEA LIEEGAFAFT NDGVGVQTAG –TMYLAMKEA AKNNKALVAH
Oceanobacillus     ---KGEERTD IQSLIDAGAF AFTDDGVGIQ TADQMYQAMK ARHNTTIVAH
Fusobacterium      ---KAGVFAF TDDGRGVQSA NVMY EAMLM ---------G AKLNKAIVAH
Lactobacillus      ---KAFALSN DGHGVQTAQT MYLAMQKAKE ---------- --NNLIIATH
Oenococcus         ---KIDNIPA LSALGAIAFT NDGKGVQTAD –TMYQAMLAA KAAGKVLVAH
Clostridium        --VKAITDDG KGVSDSKIMM EAMKIAKENN ---------- ----WIVMSH
```

```
Trichodesmium      KKQFAGISDP AIHSQIQDNE AALLATKMAL KLSKKYERR- ---------L
Nostoc             RQEFANIHDP AVHSQIQDNQ AALLATQLVL KLSQKYHRR- ---------L
Thermosynecoccus   REQFAGITDV AIHSQIQDEI AALNATQLAV TLSRKYERR- ---------L
Rhodopseudomonas   KGERIEGD-P RSHPVWRDDV AALQATQRLV AIARETGKR- ---------I
Brucella           EGLRVEGD-P SSHPVWRDEI AALQCTQRLV RIARETGAR- ---------I
Rhodospirillum     KALVADGASV ALHPEWRDVE TALAATRRLL ALARQAGRK- ---------V
Magnetospirillum   KHVAEEGAHP RVHHLWRDEE TALMASKRLI ALAETARRR- ---------V
Ralstonia          RIIAESSGDV RQHPVWRDAR SALLATQRIV KLAGDTGRR- ---------L
Xanthomonas        KEKYGDALTP EMHPDIRSRQ ACLKSSQLAV SLARKHNTR- ---------L
Xyllela            HSRYGNTLSA EQHPDIRSRQ ACLKSSQLAV SLAKKNGTR- ---------L
P. flourescens     RERYGNQLPP DAHAQIRNAD SCFRSSSLAV DLAKRHGTR- ---------L
Cytophaga          KALYGDKLNA THHHLIRSEE ACYASSSKAV ALAKKHGTR- ---------L
Saccharomyces      GEKPSVHDGD KEPIHVLNAE EAFLP-ALKK LHNDFPNLK- ---------I
Schizosacch.       GEVPPSKDN- ----TVFTAE PKFLP-TLLD LHQRFPKLK- ---------I
Ustilago maydis    GELPSNADAG ---ICVLNAE EKFLT-HLFK MHGEFPKLK- ---------I
Helicobacter       AE------QT G---FCLDKE FLCHS-VLET FALSFPKLK- ---------I
Chloroflexus       GE------SH G---FVLDRE AEFLP-VYER WARTFPRLR- ---------I
Neisseria          GE------VT DPEIDIFDRE AAFIGRVMKP VLAQVPNLK- ---------V
Beta vulgaris      GE------VT DPDVDIFDRE KVFIESVLRP LIQKLPQLK- ---------V
Arabidopsis        GE------VT DPSIDVFDRE KIFIETVLQP LIQRLPQLK- ---------V
Oryza              GE------VT DQHVDTFDRE KVFIEKILAP LVQRLPQLK- ---------I
Vibrio             GE------VT THEVDIFDRE KTFLDTVLAP IVNDFPQLK- ---------I
Shewanella         GE------VT DSSIDIFDRE RVFIENILSK IVTDFPNLK- ---------I
Microbulbifer      GE------VT DADIDIFDRE ATFLSRTMEQ LVKDFPTLK- ---------I
P. flourescens     GE------VT RGDVDVFDRE KIFIDEHMRR VVERFPTLK- ---------V
P. syringae        GE------VT RGEIDVFDRE KVFIDEHLRR VVERFPSLK- ---------V
P. aeruginosa      GE------VT RAEVDVFDRE KQFIDEHLRR VVERFPTLK- ---------V
Azotobacterium     GE------VT HSEVDVFDRE KRFIDENLVR IIEHFPTLK- ---------V
Ralstonia          GE------VT DPTVDIFDRE AVFIDTVMQP LRRDFPALK- ---------V
Burkholderia       GE------VT DASIDLFDRE KVFIDRVMTP LRRAFPALK- ---------V
A. Tumefaciens     GE------VT TPDVDIFDRE KAFIDTVLEP LRQRLPELK- ---------V
Sinorhizobium      GE------VT TAEVDIFDRE AVFIETVLDP LRRRLPDLR- ---------I
Rhodobacter sp.    GE------VT DPAVDIFDRE AVFLDRVLDP IRRATPGLR- ---------V
E. coli            GE------VT HADIDIFDRE ARFIESVMEP LRQRLTALK- ---------V 174
Shigella           GE------VT HADIDIFDRE ARFIESVMEP LRQRLTALK- ---------V
S. typhimurium     GE------VT HADVDIFDRE ARFIDTVMEP LRQRLTALK- ---------V
S. enterica        GE------VT HADVDIFDRE ARFIDTVMEP LRQRLTALK- ---------V
Yersinia           GE------VT DAAVDIFDRE ARFIDQILEP IRQKFPELK- ---------I
Novosphingobium    GE------VT DSHVDIFDRE AVFIERTLTR LVADMPALR- ---------I
Buchnera           GE------EI NQNIDIYDRE AKFIEKTLDP LRKKFPKLK- ---------I
Wigglesworthia     GE------IS DEKTDIFDRE AKFIEKVMIP IRKNFPKLK- ---------I
Pl. falciparum     CE-------- EPNINPLYAE EKYLP-HIHD LAIKFPGLN- ---------I
Pl. yoelii         CE-------- EPNINPVYAE KEYLQ-HIHD ISIKFPHLK- ---------I
Toxoplasma gondii  GE-------- VPGVAPLDAE EAFIP-FFEQ IHSRFPSLK- ---------I
Sulfolobus         KSKKLKVLHP EIPLALKVPR KLR-NIWMEI AALHYVQGN- ---------V
Streptococcus      KIA-----KE HFKICGATGV AEYSMIARDV MIAYDRQAH- ---------I
Geobacter          GFV-----ST ELGLKGIPWA AEDAATARDV YLAEFTDSP- ---------L
Clostridium        GYW-----ST VMGLKGIPSA AEEIMVARDI ILSEYTKVP- ---------I
Thermoanaerobacter GII-----AT MLGLKGIPRE AEEVMLARNI ILAKATGAK- ---------L
Magnetococcus      GEV-----AT RLGLSGICNA AEDILVERDI RLVELTGGR- ---------Y
Thermobifida       GVV-----SD RLGLAGWPAV AEEAIIARDC LLAEHVGSR- ---------L
Streptomyces       GVV-----SA ELGLGGWPAV AEESVIARDV LLAEHVGSR- ---------V
Corynebacterium    GEN-----AA RLGLRGWPRV AEESIVVRDA IMARDYGNR- ---------V
Desulfovibrio      GVT-----SG RLGVKGQSVV AESVQVARDI LLAEYLGIP- ---------V
Bifidobacterium    GPV-----SR ELGVPGIPED TELKIVARDI EAARRTGVH- ---------V
Chlorobium         GAV-----SA MLGLKGIPEV AEPIMIARDL QLIAWLKKHK LNGAVAEPRY
Thermus aquaticus  GPL-----AD LLGLPGNPPE AEAARIARDL EVLRYALRR- ---SPATPRL
Bacillus           GKF-----SE KHGLNGIPSV CESVHIARDI LLAEAADCH- ---------Y
Enterococcus       GKK-----AE ELGLPGILSV TESSQIARDL LLAEATGVH- ---------Y
Oceanobacillus     GEV-----SE RLSLPGIPSL SESVQIARDV LLAEATGCH- ---------Y
Fusobacterium      GKR-----SA ELGIKGIPSI CESTQIVRDV LLAEAANCH- ---------Y
Lactobacillus      GVA-----AK KLDLPPVTEL AETTQIARDL LLAQKTGVH- ---------Y
Oenococcus         GLA-----SK NLGFPGAKKL SETSQLARDL MIAKATGAQ- ---------Y
Clostridium        ---------- ---------L AENMMTWRDI TLAKFVDCR- ---------L
```

```
Trichodesmium      HILHTSTGDE AELLR----- ---QDKPSWV TAEVTPQHLF LN-TSAYEKI
Nostoc             HILHMSTAEE ADLLR----- ---QEKPSWV TAEVTPQHLV LN-TSAYEKI
Thermosynecoccus   HILHLSTGIE VDFLR----- ---EHKRPWI TVEVTPQHLL LT-TEAYAKI
Rhodopseudomonas   HVLHVSTRQE MEFLR----- ---EHK-DVA SVEVTPHHLT LVAPDCYERL
Brucella           HVLHISTAEE IDFLK----- ---DYK-DVA TCEATPHHLT LS-ADDYKAL
Rhodospirillum     HVLHVTTAEE MALLQ----- ---DNR-DIA SVEVTPQHLT LAAPDCYEGL
Magnetospirillum   HVLHVTTAEE MEFLA----- ---GHK-DVA TVETTPQHLT LAAPECYERL
Ralstonia          HILHVSTAEE IAFLA----- ---DRK-DRV TVEVTPHHLT LEAPECYERL
Xanthomonas        HVLHISTADE LSLFEAGPLV DADGKLRKRI TAET-CIHFL RFDRSDYARL
Xyllela            HVLHISTADE LRLFEAGPLV DAAGKRRKQI TAET-CIHFL HFDRMDYARL
P. flourescens     HVLHLTTARE LALFEDKPLT ------QKRI TAEVCLHHLL FD-ERDYPNL
Cytophaga          HILHISTAKE LDLFTNTIPL E-----EKKI TAEACIHHLW FS-NEDYATK
Saccharomyces      ILEHCTSESA IKTIEDINKN VK-KATDVKV AATLTAHHLF LT-IDDWAG-
Schizosacch.       VLEHCTTADA VEAVK----- ----ACGESV AGTITAHHLY LT-QKDWQD-
Ustilago maydis    VLEHATTRKA VEAVK----- ----QCGDTV GCTITPHHLE LI-VDDWAG-
Helicobacter       IIEHLSDWRS IALIEKHDN- ---------L YATLTLHHIS MT-LDDLLGG
Chloroflexus       VMEHITTAAA LDLLDRYPN- ---------L FATVTLHHLL IT-LDDVAGG
Neisseria          VFEHITTAEA ARLVLEAG-D N--------V AATVTPQHLL LN-RNDLLVG
Beta vulgaris      VMEHITTADA VKFIESCNGG N--------V AATVTPQHLV LN-RNSLFQG
Arabidopsis        VMEHITTMDA VNFVESCKEG S--------V GATVTPQHLL LN-RNALFQG
Oryza              VMEHITTMDA VNFVESCKEG H--------V AATVTPQHLL LN-RNALFQG
Vibrio             VLEHITTADA VTFVQQAGDN ---------V AATITAHHLL FN-RNHMLVG
Shewanella         VLEHITTKDA VDFVTQASDN ---------V AATITAHHLL YN-RNHMLAG
Microbulbifer      VLEHITTEDA VKFVLKSGPN ---------V AATITAHHLL YN-RNHMLAG
P. flourescens     VFEHITTGDA VQFVNEASAN ---------V GATITAHHLL YN-RNHMLVG
P. syringae        VFEHITTGEA VQFVNEASAN ---------V AATITAHHLL YN-RNHMLVG
P. aeruginosa      VFEHITTGDA AQFVREAPAN ---------V GATITAHHLL YN-RNHMLVG
Azotobacterium     VFEHITTRDA VQFVETSSSN ---------V GATITAHHLL YN-RNHMLVG
Ralstonia          VFEHITTKHA AEYVRDAQGP ---------V GATITAHHLL YN-RNALFVG
Burkholderia       VFEHITTKDA VDYIREAGVA PE------VL GATITAHHLL YN-RNAIFQG
A. Tumefaciens     TMEHITTRDG VDYIKSSNAN ---------L AGSITTHHLI IN-RNAILVG
Sinorhizobium      TMEHVTTKDG VDYIREHAAN ---------L AGSITTHHLI IN-RNAILVG
Rhodobacter sp.    VLEHVTTREG LDYVRGGGPD ---------M AGTLTTHHLI IN-RNHILAG
E. coli            VFEHITTKDA ADYVRDGNER ---------L AATITPQHLM FN-RNHMLVG 214
Shigella           VFEHITTKDA ADYVRDGNER ---------L AATITPQHLM FN-RNHMLVG
S. typhimurium     VFEHITTKDA AQYVRDGNDY ---------L AATITPQHLM FN-RNDMLVG
S. enterica        VFEHITTKDA AQYVRDGSYN ---------L AATITPQHLM FN-RNDMLVG
Yersinia           VFEHITTKDA ADYVLAGNRF ---------L GATVTPQHLM FN-RNHMLVG
Novosphingobium    VFEHITTEEA AQFVEGAGDS ---------I AATITPQHLH IN-RNAMLVG
Buchnera           VLEHITTKES VEYIKNNDVN --------YL SATITPHHLM LN-RNDMFYG
Wigglesworthia     VFEHISTKIA VEYVISEDS- --------LL GATITPHHLM FN-YNDMLSN
Pl. falciparum     VLEHISSSES INVIKEFR-- --------NV AGSITPHHLY LT-IDDVVNM
Pl. yoelii         VLEHISTENM IEMVKKYP-- --------NV AGSITPHHLH LT-IDDVVDI
Toxoplasma gondii  VLEHVSTAAA IQAVKRMPA- --------NV AATITPHHLM LT-VDDVVKP
Sulfolobus         HVTHITNYET VKIAKELG-- ---------F STDITPHHLL VNGERDCITK
Streptococcus      HIQHLSKAES VKVVAFAQKL G------AQV TAEAAPQHFS KT-ENLLL--
Geobacter          HIAHVSTKGS LRIIRNAKAR G------VKV TCETA-PHYF SLT-DDAVRG
Clostridium        HIAHVSTELS VELIRNAKKR G------VKV TCETC-PHYF VLT-DEACKD
Thermoanaerobacter HIAHVSTKGS VELIRRAKEE G------VSI TAEVTPHHLT RT-DEAVYN-
Magnetococcus      HVAHISSAGA VASVAKAREK G------LRV SCEAAPHHLV LN-DTHVGN-
Thermobifida       HVCHVSTKGS VQIIRWAKSR G------CDV TAEVTPHHLL LT-DELAES-
Streptomyces       HICHLSTAGS VEIVRWAKSR G------IDV TAEVTPHHLL LT-DELVRS-
Corynebacterium    HICHASTEGT VELLRWAKSQ G------IPI TAEVTPHHLT LT-DERLET-
Desulfovibrio      HLAHISCRQS VELIAWAKQR G------VRV TAETC-PHYL LLDDLALEQ-
Bifidobacterium    HFQHVSTAIS FEAIRRAKAE G------LPI TCETA-PHYL ALSDEALLK-
Chlorobium         HVAHISTAES VALVRKAKAA G------LKV TCEVTPHHFT LT-EHDLSSS
Thermus aquaticus  HVQHLSTKRG LELVREAKRA G------LPV TAEATPHHLT LT-EEALRT-
Bacillus           HVCHVSTKGS VRVIRDAKRA G------IKV TAEVTPHHLV LC-EDDIPSA
Enterococcus       HVCHVSTKES VRVIRDAKKA G------IHV TAEVSPHHLI LI-DEDIPED
Oceanobacillus     HVCHVSTKES VRVIRDAKKA G------IHV TAEVSPHHLL LN-ETDIPDD
Fusobacterium      HVCHISAKES VRAVREGKKN G------IKV TCEVTPHHLL SC-DEDIKED
Lactobacillus      HICHVSTKTS VELVRLAKAR G------INV TCEVAPHHIL LT-DSDIPKD
Oenococcus         HMAHVSAAES VELIRIAKEH R------IDV TAEVSPHHLL LD-DSMITND
Clostridium        HMSHVSTKEA MKYIIEGKND G------VKV TCEITPHHLA LN------NK
```

```
Trichodesmium      G--------- ---------- ---------- ---TLAQMNP PLKSAGDNDI
Nostoc             G--------- ---------- ---------- ---TLAQMNP PLRSPHDNEV
Thermosynecoccus   G--------- ---------- ---------- ---SLAQMNP PLRTAVDNKK
Rhodopseudomonas   G--------- ---------- ---------- ---TKAQMNP PVRDAWHRDG
Brucella           G--------- ---------- ---------- ---NLIQMNP PVRDKRHREG
Rhodospirillum     G--------- ---------- ---------- ---TFAQMNP PIREARHREA
Magnetospirillum   G--------- ---------- ---------- ---TYAQMNP PIRGTRHRDA
Ralstonia          G--------- ---------- ---------- ---TLAQMNP PVRERCHRNA
Xanthomonas        G--------- ---------- ---------- ---NLIKCNP AIKDAEDRLA
Xyllela            G--------- ---------- ---------- ---NLIKCNP AIKEASDREA
P. flourescens     G--------- ---------- ---------- ---NLIKCNP AIKSRSDRDA
Cytophaga          G--------- ---------- ---------- ---NYIKWNP AVKTVADREA
Saccharomyces      ---------- ---------- ---------- --NPVNFCKP VAKLPNDKKA
Schizosacch.       ---------- ---------- ---------- --DPYCFCKP VAKTERDRRA
Ustilago maydis    ---------- ---------- ---------- --KPLNFCKP VAKYPDDRQA
Helicobacter       SL-------- ---------- ---------- --NPHCFCKP LIKTKKDQER
Chloroflexus       LL-------- ---------- ---------- --QPHLFCKP IAKRPTDRDA
Neisseria          GV-------- ---------- ---------- --RPHHFCLP VLKRETHRQA
Beta vulgaris      GL-------- ---------- ---------- --QPHNYCLP VLKREIHRQA
Arabidopsis        GL-------- ---------- ---------- --QPHNYCLP VLKREIHREA
Oryza              GL-------- ---------- ---------- --QPHNYCLP VLKRETHRQA
Vibrio             GI-------- ---------- ---------- --RPHFYCLP ILKRATHQHA
Shewanella         GI-------- ---------- ---------- --RPHFYCLP ILKRNTHQQA
Microbulbifer      GI-------- ---------- ---------- --RPHYYCLP ILKRSSHQQA
P. flourescens     GI-------- ---------- ---------- --RPHFYCLP ILKRNTHQEA
P. syringae        GI-------- ---------- ---------- --RPHFYCLP ILKRNTHQTA
P. aeruginosa      GI-------- ---------- ---------- --RPHFYCLP ILKRNTHQEA
Azotobacterium     GI-------- ---------- ---------- --RPHFYCLP ILKRRTHQEA
Ralstonia          GI-------- ---------- ---------- --RPHYYCLP VLKRETHRLA
Burkholderia       GI-------- ---------- ---------- --RPHYYCLP VLKRETHRVA
A. Tumefaciens     GI-------- ---------- ---------- --KPHYYCLP VAKREEHRLA
Sinorhizobium      GI-------- ---------- ---------- --KPHYYCLP VAKREAHRLA
Rhodobacter sp.    GI-------- ---------- ---------- --RPHYYCLP VAKRETHRLA
E. coli            GV-------- ---------- ---------- --RPHLYCLP ILKRNIHQQA 234
Shigella           GV-------- ---------- ---------- --RPHLYCLP ILKRNIHQQA
S. typhimurium     GI-------- ---------- ---------- --RPHLYCLP ILKRNIHQQA
S. enterica        GI-------- ---------- ---------- --RPHLYCLP ILKRNIHQQA
Yersinia           GI-------- ---------- ---------- --RPHLFCLP ILKRSTHQQA
Novosphingobium    GI-------- ---------- ---------- --RPHAFCLP VAKREKHRLA
Buchnera           GI-------- ---------- ---------- --QPYLYCLP ILKKNKHRMA
Wigglesworthia     KI-------- ---------- ---------- --KPHLYCFP ILKKKIDQFA
Pl. falciparum     DIYDHAIDNT Y--------- ------IEKY IKNTYHYCKP LPKLLEDKIA
Pl. yoelii         KNYDYSFNNT D--------- ------IEKY VKNVYNYCKP LPKTLDDKIA
Toxoplasma gondii  DAIKRAMETA AHRGLC---- ------CAEA VEKPHNFCKP LAKTKRDRDA
Sulfolobus         VN-------- ---------- ---------- ---------P PIRDYLTRLG
Streptococcus      -------EKG ---------- ---------- ---ANAKMNP PLRLEADRLA
Geobacter          ---------- ---------- ---------- -YDTNAKMNP PLRTADDVTA
Clostridium        ---------- ---------- ---------- -FNTLAKVNP PLRTRRDVEA
Thermoanaerobacter ---------- ---------- ---------- -YDTNTKVFP PLRTREDVEA
Magnetococcus      ---------- ---------- ---------- -YDTNAKMAP PLRNQRDVNA
Thermobifida       ---------- ---------- ---------- -YDPVYKVNP PLRTAEDVEA
Streptomyces       ---------- ---------- ---------- -YNPVYKVNP PLRTERDVMA
Corynebacterium    ---------- ---------- ---------- -YDAVNKVNP PLRESRDAEA
Desulfovibrio      ---------- ---------- ---------- -YSTAAKVNP PLRTPDDVAA
Bifidobacterium    ---------- ---------- ---------- -YGTLAKMNP PLRSEADRKA
Chlorobium         I--------- ---------- ---------- -EKGNFIMKP PLASVENRDA
Thermus aquaticus  ---------- ---------- ---------- -FDPLFKVAP PLRGEEDREA
Bacillus           D--------- ---------- ---------- ---PNFKMNP PLRGKEDHEA
Enterococcus       F--------- ---------- ---------- ---GFWKMNP PLRGREDREA
Oceanobacillus     N--------- ---------- ---------- ---ADWKMNP PLRSTEDQQA
Fusobacterium      N--------- ---------- ---------- ---GMWKMNP PLRSREDRNA
Lactobacillus      N--------- ---------- ---------- ---GYFKMNP PLRNKEDQAA
Oenococcus         N--------- ---------- ---------- ---SLMKMNP PLRSPDDRAA
Clostridium        I--------- ---------- ---------- ---SNYRVNP PIREEEDVNF
```

```
Trichodesmium      LWRALL--DG VIDFIATDHA PHTLAEKG-- ---------- ----------
Nostoc             LWQALR--DG VIDFIATDHA PHTLEEKA-- ---------- ----------
Thermosynecoccus   LWQGLL--DG VIDFIATDHA PHTLEEKA-- ---------- ----------
Rhodopseudomonas   IWHGLA--QG VVDVLGSDHA PHTLEEKA-- ---------- ----------
Brucella           VWKGID--QG IVDVLGSDHA PHTLEEKQ-- ---------- ----------
Rhodospirillum     LWRALE--QG VVDIIGSDHA PHTLKEKS-- ---------- ----------
Magnetospirillum   LWKAIA--DG TVDVLGSDHA PHTREEKD-- ---------- ----------
Ralstonia          LWQAIR--EG VGDVIGSDHA PHTPKEKS-- ---------- ----------
Xanthomonas        LIDALA--ED VIDVLATDHA PHTWEEKQ-- ---------- ----------
Xyllela            LIAALA--KD VIDVLATDHA PHTWEEKQ-- ---------- ----------
P. flourescens     LRQALL--SN RLDVIGSDHA PHTWAEKQ-- ---------- ----------
Cytophaga          IWQAVL--DN RIDVIATDHA PHTIEEKE-- ---------- ----------
Saccharomyces      LVKAAV-SGK PYFFFGSDSA PHPVQNK--- ---------- ----------
Schizosacch.       LIEAAT-SKN PKFFFGSDSA PHPRSSK--- ---------- ----------
Ustilago maydis    LRDVIR-QGH PRFFLGSDSA PHPLANKYPS AVTHGAPGTK ASDSGSDHLE
Helicobacter       LLSLAL-KAH PKISFGSDSA PHFISKK--- ---------- ----------
Chloroflexus       LLAAAL-AGH PKLMFGSDSA PHPIDRK--- ---------- ----------
Neisseria          LVAAVTGEKA HKFFLGTDSA PHAKSAK--- ---------- ----------
Beta vulgaris      LVSAVT-SGS KQYFLGTDSA PHERRRK--- ---------- ----------
Arabidopsis        IVKAVT-SGS KKFFLGTDSA PHERSRK--- ---------- ----------
Oryza              IVSAVT-SGS KQYFLGTDSA PHDKRRK--- ---------- ----------
Vibrio             LVAAAT-SGS KKFFLGTDSA PHAKGRK--- ---------- ----------
Shewanella         LLAAAA-SGN KKFFLGTDSA PHAKDRK--- ---------- ----------
Microbulbifer      LISAAI-SGN PKFFLGTDSA PHAKSKK--- ---------- ----------
P. flourescens     LLDAAT-SGS AKFFLGTDSA PHAQHAK--- ---------- ----------
P. syringae        LLDAAT-SGS GKFFLGTDSA PHAQHAK--- ---------- ----------
P. aeruginosa      LLDAAV-SGN PKFFLGTDSA PHARHAK--- ---------- ----------
Azotobacterium     LLDAAT-SGN TKFFLGTDSA PHARHAK--- ---------- ----------
Ralstonia          LVAAAT-SGH PRFFLGTDSA PHAKGLK--- ---------- ----------
Burkholderia       LVEAAT-SGN PRFFLGTDSA PHPKGLK--- ---------- ----------
A. Tumefaciens     LRAAAT-SGD ARFFLGTDSA PHVDPLK--- ---------- ----------
Sinorhizobium      LRQAAI-SGD VRFFLGTDSA PHVDPLK--- ---------- ----------
Rhodobacter sp.    LRAAAT-GGE ACFFLGTDSA PHVDAAK--- ---------- ----------
E. coli            LRELVA-SGF NRVFLGTDSA PHARHRK--- ---------- ---------- 260
Shigella           LRELVA-SGF NRVFLGTDSA PHARHRK--- ---------- ----------
S. typhimurium     LRELVA-SGF TRAFLGTDSA PHSRHRK--- ---------- ----------
S. enterica        LRDLVA-SGF TRAFLGTDSA PHSRHRK--- ---------- ----------
Yersinia           LRAAVA-SGS DRFFLGTDSA PHAKHRK--- ---------- ----------
Novosphingobium    LRKLAT-SGF SRVFLGTDTA PHAKHLK--- ---------- ----------
Buchnera           LRKAIS-NGD KHFFLGSDTA PHLHKNK--- ---------- ----------
Wigglesworthia     LHKAIS-GNC NRFFLGTDSA PHVLKNK--- ---------- ----------
Pl. falciparum     LQDVIK-DDF PRVFLGSDSA PHYKVMK--- ---------- ----------
Pl. yoelii         LCNIIK-EGN PKVFLGSDSA PHYQKFK--- ---------- ----------
Toxoplasma gondii  LRQVIQ-DGD THFFLGSDSA PHPRLAK--- ---------- ----------
Sulfolobus         LW-KAL---F EVDTVVSDHA PHSKEEKN-- ---------- ----------
Streptococcus      VIEGLK--SG VISVIATDHA PHHADEKN-- ---------- ----------
Geobacter          VKEALK--DG TIDAIATDHA PHHIDEK--- ---------- ----------
Clostridium        VIEGLK--DG TIDIIATDHA PHHADEK--- ---------- ----------
Thermoanaerobacter LIEGLK--DG TIDAIATDHA PHTKDDK--- ---------- ----------
Magnetococcus      IQEALA--RG VISVIATDHA PHEEDSK--- ---------- ----------
Thermobifida       LREGLA--DG TIDIVATDHA PHPSEAK--- ---------- ----------
Streptomyces       LREALA--DG TIDIVATDHA PHPHEDK--- ---------- ----------
Corynebacterium    LKKALL--DG TIDVVATDHA PHGSEDK--- ---------- ----------
Desulfovibrio      MRRAVA--DG TIDILVTDHA PHAAHEK--- ---------- ----------
Bifidobacterium    TIAAIA--DG TVDLLATDHA PHTLAEK--- ---------- ----------
Chlorobium         LIEGLR--DG TIDAIATDHA PHAKHEK--- ---------- ----------
Thermus aquaticus  LLEGLL--DG TLDAIATDHA PHTQAEK--- ---------- ----------
Bacillus           LIEGLL--DG TIDMIATDHA PHTAEEK--- ---------- ----------
Enterococcus       LIEGLL--DG TIDCIATDHA PHGLEEK--- ---------- ----------
Oceanobacillus     LFDGLM--DG TIDLIATDHA PHATEEK--- ---------- ----------
Fusobacterium      LIVGIL--DG TIDIIATDHA PHTMEEK--- ---------- ----------
Lactobacillus      LLVGLL--DG TIDLIATDHA PHAKSEK--- ---------- ----------
Oenococcus         LLAGLL--DG TIDIVATDHA PHTKEEK--- ---------- ----------
Clostridium        LIKAIK--MN YVDCIGTDHA PHSKEDK--- ---------- ----------
```

```
Trichodesmium      ---KGYPNTP SGMPGVETSL PLMLTQAIEG R--CSVAQVS NWMSTAVAKG
Nostoc             ---QEYPNSP SGMPGVETSL ALMLTAAMEG K--CTVSQVV NWMSKNVAVA
Thermosynecoccus   ---QPYPQSP SGMPGVETSL PLMLTQAMAG R--CTVPQVV RWMSTAVAAA
Rhodopseudomonas   ---KTYPASP SGMTGVQTLV PTMLDHVNAG R--LSLARFV DLTSAGPARL
Brucella           ---KSYPASP SGMTGVQTLV PIMLDHVNAG K--LSLERFV DLSSHGPNRI
Rhodospirillum     ---LPYPKSP SGMPGVQTLV PVMLDHVNTG K--LSLERFV DLTSAGPARV
Magnetospirillum   ---KPYPQSP SGMTGVQTLV PLMLDHVNQG R--LSLERLV DLTSAGPARI
Ralstonia          ---RPYPASP SGMTGVQTLV PVMLDHVHRG Q--LSLQRFV DLTSAGPARI
Xanthomonas        ---KPYAQAP SGLPLVQYAL VAALELVHEG K--LPITRIV QKFAHAPAQL
Xyllela            ---QPYAQAP SGLPLVQYAL VAALELVHAG R--LSVTEIV RKFAHAPAQL
P. flourescens     ---QAYEQAP SGLPLVQHAL PALLELVADG V--LPITTLV AKTSHRVADL
Cytophaga          ---LPYIDAP SGGPLVQHSL VAMLEKYHQG K--ISLERIA EKMSHNVAKL
Saccharomyces      ---ANYEGVC AGVYSQSFAI PYIAQVFE-E Q--NALENLK GFVSDFGISF
Schizosacch.       ---LKTP-PA AGVFTQPFAA SYLAEVFD-K E--GRLDALK DFACIFGRKF
Ustilago maydis    ATGVVSCGCA AAVYTSSILV PLWPTLLE-A F--GALDQLA NYVSINGRNF
Helicobacter       ----HSANIP AGIFSAPILL PALCELFE-K H--NALENLQ AFISDNAKTI
Chloroflexus       ----EAAFCA AGVFSAPVLL PMLVELFE-R H--NALDRLP DFVSGNARRI
Neisseria          ----ENACGC AGMFSAMTAI ELYAEVFE-K A--GALDKLE AFASKNGARF
Beta vulgaris      ----ECSCGC AGIYNSPVAL SLYAKVFE-E A--GALDKLE AFTSFNGPDF
Arabidopsis        ----ESSCGC AGIYSAPIAL SLYAKVFD-E A--GALDKLE AFTSFNGPDF
Oryza              ----ECSCGC AGIYSAPVAL SLYAKVFE-Q A--GALDKLE AFTSFNGPDF
Vibrio             ----EAACGC AGSYTAHAAL ELYAEVFE-K E--GKLENLE AFASFNGPDF
Shewanella         ----EAACGC AGSYTAHAAI ELYAEAFE-S V--NALDKLE AFASFNGPDF
Microbulbifer      ----EAACGC AGSYTAFAAL PLYAEAFE-E A--GALDKLE DFASHFGPDF
P. flourescens     ----EAACGC AGCYTAYAAI ELYAEAFE-Q R--NALDKLE AFASLNGPRF
P. syringae        ----ENACGC AGCYTAYAAI ELYAEAFE-Q R--NALDKLE GFASLHGPAF
P. aeruginosa      ----EAACGC AGCYSAYAAI ELYAEAFE-Q R--NALDKLE GFASLHGPDF
Azotobacterium     ----EAACGC AGCYTAYAAI ELYAEAFE-Q R--SALDRLE AFASHHGADF
Ralstonia          ----EHACGC AGCYTALHAM ELYAEAFE-D A--NALDKLE GFASLHGPDF
Burkholderia       ----EHACGC AGCYTALHAL ELYTEAFD-K A--GALDKLE GFASFFGADF
A. Tumefaciens     ----ECACGC AGIYTSINTM SCLAHVFE-D E--NALDKLE AFASLNGPAW
Sinorhizobium      ----ECACGC AGIYTSINTL SCLAHVFE-E E--GALDRLE AFTSLNGPAW
Rhodobacter sp.    ----ESGCGC AGCFTAPNTL SILAHVFE-E E--GALDRLE GFVSLHGPAF
E. coli            ----ESSCGC AGCFNAPTAL GSYATVFE-E M--NALQHFE AFCSVNGPQF 303
Shigella           ----ESSCGC AGCFNAPTAL GSYATVFE-E M--NALQHFE AFCSVNGPQF
S. typhimurium     ----ETSCGC AGCFNAPSAL GSYAAVFE-E M--NALAHFE AFCSLNGPQF
S. enterica        ----ETRCGC AGCFNAPSAL GSYAAVFE-E M--NALAHFE AFCSLNGPQF
Yersinia           ----ESSCGC AGVFNAPAAL PAYASVFE-E L--NALQHLE AFCALNGPRF
Novosphingobium    ----EAACGC AGIFNAPFAL ESYVTVFD-E E--GALDRFE AFASLNGPAF
Buchnera           ----INMLGC AGIFNAPSSL LSYVKVFE-E M--RALKYLQ SFCSENGPKF
Wigglesworthia     ----ESSMGF AGIFNSPIAL EMYATVFD-N L--NILNKLE NFCSINGANF
Pl. falciparum     ----RKPYYK PGIYTQPFLI NYVAHILN-K F--DALDKME NFTSKNASLF
Pl. yoelii         ----QEPHYK PGIFTQPFLL SYLAHIFN-K F--NSLDKIE NFACKNAAQF
Toxoplasma gondii  ----ESSPPA AGVFTQPLLL AYLVSIFA-E L--GCISKLR EFADGHAAAF
Sulfolobus         ---LNYDLCP PGIAAVSFTT PFIYSLVFKD L--LNIERAV NLLSKNPAKI
Streptococcus      --VADVTKAP SGMTGLETSL SLGLTYLVEA G-HLSLMDFL AKMTINPAQL
Geobacter          --DLEFNEAL NGIVGLETSL TLSLR-LVEE G-VLTLPVLV DKMACNPAKI
Clostridium        --NVEFNLAA NGMVGFETAL PLAITYLVKP G-HLTISQLV EKMCVNPSKL
Thermoanaerobacter --KVPYDMAP FGISGLETAF SVINTFLIQT G-IITMKALV NYMSMNPARI
Magnetococcus      --RVPFCQAA NGVVGLETLL PITLE-LVEA G-VLPLAKAL AAISCNPARL
Thermobifida       --ETEWNQAA MGMIGLETAL SVVQHTMVDT G-LLDWAGVA QRMSATPARI
Streptomyces       --DCEWAAAA MGMVGLETAL SVVQETMVDT G-LLDWAGVA DRMSFKPAKI
Corynebacterium    --CCEFENAK PGMLGLETSL SIIVDTFVAT G-LADWRFVA RVMSERPAEI
Desulfovibrio      --DTPLDEAP NGISGLDTAV ALTWR-LVQE G-LLTEADMV RLWCHAPGSL
Bifidobacterium    --ELGFLEAP NGIIGLECAY GVCHKVLVDG G-FISDERLI ELMSVGPAEL
Chlorobium         --ECPPDQAA FGIIGLETSL GLTITELVDK G-VITLSQAI ELLSTNPRRI
Thermus aquaticus  --EMDLLRAP FGIPSLEVAF PLLYTELHLK R-GFPLRRLV ELFTDGPRRV
Bacillus           --AQGIERAP FGITGFETAF PLLYTNLVKK G-IITLEQLI QFLTEKPADT
Enterococcus       --SQSFMKSP FGIVGSETAF QLIYTHFVET G-RFTLEQVI NWLAVKPAEI
Oceanobacillus     --AVGFKEAP FGIVGLETAF PLLYTHLVLK G-KMTLYQLV ERMTEKPAAT
Fusobacterium      --IRGIEKSS FGIVGSETAF AQLYTKFVKT D-IFSLEMLV KLMSENVAKI
Lactobacillus      --QGGMKNAA FGITGSETAF STLYTKFVKE EKVLSLEQLL ALLSDKPAKV
Oenococcus         --AQSILTAP NGVTGIETSF QLLYTHLVKP G-IMSLRQLL KAMNQRPADI
Clostridium        ------EKGA PGMIGIEQAF SICYTKLVKE N-HISLNKLS QLMSGNAAKL
```

```
Trichodesmium      YGILKKG--- ---------- ---------- ---------- ---------A
Nostoc             YGIPNKG--- ---------- ---------- ---------- ---------V
Thermosynecoccus   YEIPNKG--- ---------- ---------- ---------- ---------K
Rhodopseudomonas   FGIACKG--- ---------- ---------- ---------- ---------R
Brucella           FGMARKG--- ---------- ---------- ---------- ---------R
Rhodospirillum     YDIAGKG--- ---------- ---------- ---------- ---------R
Magnetospirillum   YNIAGKG--- ---------- ---------- ---------- ---------R
Ralstonia          FGVEGKG--- ---------- ---------- ---------- ---------R
Xanthomonas        FDVEERG--- ---------- ---------- ---------- ---------F
Xyllela            FDVIERG--- ---------- ---------- ---------- ---------F
P. flourescens     FAIPDRG--- ---------- ---------- ---------- ---------Y
Cytophaga          FHD------- ---------- ---------- ---------- ----------
Saccharomyces      YEVKDSE--- ---------- ---------- ---------- ---------V
Schizosacch.       YCIP------ ---------- ---------- ---------- ---------L
Ustilago maydis    YGYND-D--- ---------- ---------- ---------- ---------Q
Helicobacter       YGLEN----- ---------- ---------- ---------- ----------
Chloroflexus       HGLN------ ---------- ---------- ---------- ----------
Neisseria          YGIPE----- ---------- ---------- ---------- ----------
Beta vulgaris      YGLPR----- ---------- ---------- ---------- ----------
Arabidopsis        YGLPR----- ---------- ---------- ---------- ----------
Oryza              YGLPR----- ---------- ---------- ---------- ----------
Vibrio             YGLPR----- ---------- ---------- ---------- ----------
Shewanella         YNLPR----- ---------- ---------- ---------- ----------
Microbulbifer      YGLPR----- ---------- ---------- ---------- ----------
P. flourescens     YGLPA----- ---------- ---------- ---------- ----------
P. syringae        YGLPA----- ---------- ---------- ---------- ----------
P. aeruginosa      YGLPR----- ---------- ---------- ---------- ----------
Azotobacterium     YGIPR----- ---------- ---------- ---------- ----------
Ralstonia          YGLPR----- ---------- ---------- ---------- ----------
Burkholderia       YGLPR----- ---------- ---------- ---------- ----------
A. Tumefaciens     YGLAP----- ---------- ---------- ---------- ----------
Sinorhizobium      YGLPA----- ---------- ---------- ---------- ----------
Rhodobacter sp.    YRLPA----- ---------- ---------- ---------- ----------
E. coli            YGLPV----- ---------- ---------- ---------- ---------- 308
Shigella           YGLPV----- ---------- ---------- ---------- ----------
S. typhimurium     YGLPM----- ---------- ---------- ---------- ----------
S. enterica        YGLPV----- ---------- ---------- ---------- ----------
Yersinia           YGLPV----- ---------- ---------- ---------- ----------
Novosphingobium    YRMPV----- ---------- ---------- ---------- ----------
Buchnera           YNMPI----- ---------- ---------- ---------- ----------
Wigglesworthia     YNLPI----- ---------- ---------- ---------- ----------
Pl. falciparum     LNLAEK---- ---------- ---------- ---------- ----------
Pl. yoelii         LNLEPKS--- ---------- ---------- ---------- ---------I
Toxoplasma gondii  FGFEAAT--- ---------- ---------- ---------- ---------L
Sulfolobus         LNIPYG---- ---------- ---------- ---------- ---------E
Streptococcus      YDF------- ---------- ---------- ---------- ----DVGY--
Geobacter          LGI------- ---------- ---------- ---------- ----DRGT--
Clostridium        LGI------- ---------- ---------- ---------- ----NKGT--
Thermoanaerobacter LGI------- ---------- ---------- ---------- ----SNG---
Magnetococcus      LGM------- ---------- ---------- ---------- ----PRGT--
Thermobifida       GRL------- ---------- ---------- ---------- ----TDQGRP
Streptomyces       GQA------- ---------- ---------- ---------- ----TGHGRP
Corynebacterium    TRL------- ---------- ---------- ---------- ----PGQGRP
Desulfovibrio      FRL------- ---------- ---------- ---------- ----PVN--R
Bifidobacterium    MGHVPTDVAA LVEDYAEPAP TGDDPDGVIA GESDAQPQRE GVNRLLDLSR
Chlorobium         MG-------- ---------- ---------- ---------- -----LETIL
Thermus aquaticus  LG-------- ---------- ---------- ---------- -----LPPLH
Bacillus           FGLE------ ---------- ---------- ---------- -------AGR
Enterococcus       FGLN------ ---------- ---------- ---------- -------AGT
Oceanobacillus     FQLP------ ---------- ---------- ---------- -------YGK
Fusobacterium      FDLP------ ---------- ---------- ---------- -------YGK
Lactobacillus      FGIEN----- ---------- ---------- ---------- -------AGV
Oenococcus         FALKDV---- ---------- ---------- ---------- -------ARE
Clostridium        LNIN------ ---------- ---------- ---------- -------KGK
```

```
Trichodesmium      IAPGFDADLV LVDLNNY--- ---RPVLREE LMTKCRWSPF EGWSLTGWPV
Nostoc             IAPGYDADLV LVDLNTY--- ---RPVRREE LLTKCHWSPF EGWNLTGWAT
Thermosynecoccus   IAPGYDADLV LVDLHTY--- ---RPVRREE LLTKCGWSPF EGWSLTGWPV
Rhodopseudomonas   IAAGFDADFT VVDLKRS--- ---ETITNDW VASRAGWTPY DGVRVTGWPV
Brucella           IAVGYDADLT IVDMKRR--- ---ETITHEQ AGSKAGWTPY HGKTVTGWPV
Rhodospirillum     IAAGYDADFT VVDMNAE--- ---RTITNEW IASKCGWTPF HGLSVRGWPV
Magnetospirillum   IAVGYDADFT IVDMKAE--- ---RIITNDW IESRCGWTPF DGRKVTGWPV
Ralstonia          IALGYDADLT LVDLAAR--- ---RTIRNDW IASVSGWTPY DGIGVTGWPI
Xanthomonas        LREGYFADLV MIDNTP---- ---FTVKREQ VLSKCGWSPF EGTTFRSRIA
Xyllela            LREGYWADLV LVEDTP---- ---FTVKRQE ILSKCGWSPF EGTTFRSRIA
P. flourescens     LREGYWADLV LIQPEPKG-- ---VAVSRQP VLSQCGWTPF AQRSFRHRVS
Cytophaga          ---------- ---------- ---------- ---------- ----------
Saccharomyces      ASSDKAILFK KE-------- --QVIPQVIS DGKDISIIPF KAGDKLSWSV
Schizosacch.       DFKESNIVLK KES------- --FRVPESVA N----DLVPF HPNEVLQWHC
Ustilago maydis    HAKHGSIKLR KVRSRSCISP AAATVPAVYV HP-EFREVPD SDRSKVQVVP
Helicobacter       --LPSKKARL SK-------- -KPFMIPTHT LCLNEKIAIL RGGETLSWNL
Chloroflexus       --PPPRTVRL VE-------- -DPWQVPMR- ---YGDVVPF AAGQTLRWQV
Neisseria          --NTDTITLV KQ-------- -SQTVPASVP YGDG-ELVPM RAGGEIGWTV
Beta vulgaris      --NTSKIKLK KE-------- -PWKVLERIP FPSG-EIIPM FAGQMLDWKP
Arabidopsis        --NSSKITLK KS-------- -PWKVPDVFN FPFG-EIVPM FAGETLQWQP
Oryza              --NTSKIVLR KS-------- -AWKVPDTYS YSSG-EIVPM FTGNTLEWLP
Vibrio             --NQETVTLT KQ-------- -AWPVAESMP FGSD-IVVPI RAGENIEWTV
Shewanella         --NSDTITLV KK-------- -AWDIPASYP LGDT-NVVPI RAGEQIDWQV
Microbulbifer      --NTTKVTLT KQ-------- -AWTVPSNLP FGSD-TLVPV KAGETLNWTL
P. flourescens     --NTDRITLV RD-------- -EWTAPTSLP FGEL-TVIPL RAGEKLRWRL
P. syringae        --NQDTITLV RD-------- -EWTAPASLP FGEL-TVIPL RAGETLRWR-
P. aeruginosa      --NTDRITLV RE-------- -EWQAPASLP FGDF-DVVPL RAGETLRWKL
Azotobacterium     --NTDHITLI RE-------- -EWTAPACMT FGEH-SLVPL RAGEKLRWRL
Ralstonia          --NAGTLTLT RS-------- -QWQLPAEVP FGEQ-MLVPL RGGEMLRWKT
Burkholderia       --NEEKVTLR RE-------- -EWTLPAELP VGDT-PVVPL RGGESIGWRL
A. Tumefaciens     --NDETITLV KRE------- EAVAFPEKIE TGAG-PVTVF DPMFPLHWDV
Sinorhizobium      --NEETITLR KQE------- EPVSYPARIE TEAG-PVTVF DPMFPLHWAV
Rhodobacter sp.    --NEERITLR KG-------- AALALPGKIE TGAG-PVTVF DPGFPLLWHV
E. coli            --NDTFIELV RE-------- -EQQVAESIA LTDD-TLVPF LAGETVRWSV 346
Shigella           --NDTFIELV RE-------- -EQQVAESIA LTDD-TLVPF LAGETVRWSV
S. typhimurium     --NTGWVELV RD-------- -EQQIPGNIA LADD-SLVPF LAGETVRWSV
S. enterica        --NTGWVELV RD-------- -EQQIPENIA LADD-SLVPF LAGETVRWSV
Yersinia           --NDDVVELV RT-------- -PFLQPEEIP LGNE-SVIPF LAGQTLNWSV
Novosphingobium    --NEDRIVLE RA-------- -PIEVPEVID CNGT-AIVPF HAGETLGWRI
Buchnera           --NKETITII KKP------- CKIIKK--IN VGRN-VIIPF LSGEILNWSI
Wigglesworthia     --NKEHITFF KET------- FYPNKKQIIK INKNNYLSPI FFEKKLNWKI
Pl. falciparum     KKLAKYYICV EKHP------ ------FKLP -REYNGVVPF LAGKTLDYDI
Pl. yoelii         ENNQNSIIYI QKKN------ ------LTIP -NDYFGVVPF LANQTIDFTA
Toxoplasma gondii  EENLECAVIQ PTPQR----- --VPSIFKFP GQENEGVVPF MASLELPYTV
Sulfolobus         IRIGYVANFT IISKN----- -DWKYRTKFS KVTETPLDNF PLEAKVEFTI
Streptococcus      LAENGPADLV IFADK----- -EDRLVS-NH FASKAANSPF IGETLKGQVK
Geobacter          LKPGSVADIT VIDPK----- -ATWTVDADK LASKSKNSPF LGWEVKGAAA
Clostridium        LETGRSADIT IVDLN----- -EEFVVDVNK FKSKSKNSPF HGFKLNGSVY
Thermoanaerobacter IRVGATADIV IVNPH----- -EEYTVDKEK FKSKGKNTPY HGMKLKGVVE
Magnetococcus      LSLNAVGDVV LFDPQ----- -QSWTVDALA LHGSSKNTAF AGRQVKGRVK
Thermobifida       IAVGEPANLT LYDPT----- -ATRVIDPAT MVSKSRNTPF AGMTLPGRVL
Streptomyces       VSAGEPANLT LVDAA----- -YRGQVDPAG FASRSRNTPY EGRELPGRVT
Corynebacterium    IAEGEPANLA IVDPG----- -KTWTASGAD FASKAENTPF EGQEFSAKVT
Desulfovibrio      FTAGDPADFF LFDPA----- -HEWTVTPEA MHSKGKNTPF TGWKLTGKVT
Bifidobacterium    VDDADNVDLV VLNAA----- -EEWTVDPEQ FHSKARNTPF GGWQVTGRPL
Chlorobium         FRAGRKANLT IIDPD----- -CEWIVSESD FGSKSRNTPF MGRKLKGRAL
Thermus aquaticus  LEEGTEASLV LLDPK----- --ERPVDPQG FASKARFSPW TGWRLGGWPV
Bacillus           LKEGRTADIT IIDLE----- -QEEEIDPTT FLSKGKNTPF AGWKCQGWPV
Enterococcus       LTVGAPADVA VFDIT----- -QTCTIDKED FLSKGENTPF IGWKVKGETQ
Oceanobacillus     LEEDSVADIT VIDLE----- -KEETIHRET FYSKGKNTPF DNWKVKGIPV
Fusobacterium      LEENSFADIV VIDLE----- -KEITINPNN FLSKGKNTPY INEKINGIPV
Lactobacillus      LEPGKNADVA IFDIE----- -HKNEIKEAD FKSKGVNTPF TGQKVYGETV
Oenococcus         IAIGQVADLA LFDIR----- -HLHEIKANE FASKASNSPF IGWKVYGQTE
Clostridium        LQPGFLGDLV LIDLN----- -KKRIFKEED IVSRSKNTPF NGMEFYGDVV
```

```
Trichodesmium      VTIVGGEVAF NRGEFN-SEV RGRALIFSEI A--------- -
Nostoc             TTIVGGEIVY DKGQVN-TQV RGQALTFL-- ---------- -
Thermosynecoccus   YTFVNGEVVF SQGRVN-TSV RGRALKFGLG ---------- -
Rhodopseudomonas   GTFVRGAKVM WQGELT-TPS RGEPVRFLET LKP------- -
Brucella           GTFVRGIKVM WEAEIV-NPH KGEPVEFLEA LPRR------ -
Rhodospirillum     MTIIRGAIVM RDGALQ-GRP LGAPVRFHPL PPLTKESAVL R
Magnetospirillum   ATIVRGSTVM RDGQLL-GQA SGEPVRFQEC P--------- -
Ralstonia          HTVVRGHIVV RDEALVGTTP LGRPLRFLEA PPRVPAGV-- -
Xanthomonas        ATWVNGQQVW DGEQLV-GSA AGQRLTFDR- ---------- -
Xyllela            STWVNGNHVW DGNRLV-GVP NGQRLEFDR- ---------- -
P. flourescens     TTIVSGQIAW HDHRIH-DSC QGLPLRFMR- ---------- -
Cytophaga          ---------- ---------- ---------- ---------- -
Saccharomyces      RWEPR----- ---------- ---------- ---------- -
Schizosacch.       SWE------- ---------- ---------- ---------- -
Ustilago maydis    FWAVSAGIVR S--------- ---------- ---------- -
Helicobacter       QEIA------ ---------- ---------- ---------- -
Chloroflexus       VAE------- ---------- ---------- ---------- -
Neisseria          QY-------- ---------- ---------- ---------- -
Beta vulgaris      SF-------- ---------- ---------- ---------- -
Arabidopsis        LK-------- ---------- ---------- ---------- -
Oryza              SDQLEE---- ---------- ---------- ---------- -
Vibrio             K--------- ---------- ---------- ---------- -
Shewanella         E--------- ---------- ---------- ---------- -
Microbulbifer      TPAE------ ---------- ---------- ---------- -
P. flourescens     LEEHA----- ---------- ---------- ---------- -
P. syringae        LEEHA----- ---------- ---------- ---------- -
P. aeruginosa      LEAGA----- ---------- ---------- ---------- -
Azotobacterium     LEEKM----- ---------- ---------- ---------- -
Ralstonia          V--------- ---------- ---------- ---------- -
Burkholderia       V--------- ---------- ---------- ---------- -
A. Tumefaciens     N--------- ---------- ---------- ---------- -
Sinorhizobium      TQA------- ---------- ---------- ---------- -
Rhodobacter sp.    EP-------- ---------- ---------- ---------- -
E. coli            KQ-------- ---------- ---------- ---------- - 348
Shigella           KQ-------- ---------- ---------- ---------- -
S. typhimurium     KK-------- ---------- ---------- ---------- -
S. enterica        KK-------- ---------- ---------- ---------- -
Yersinia           KR-------- ---------- ---------- ---------- -
Novosphingobium    AAA------- ---------- ---------- ---------- -
Buchnera           ESD------- ---------- ---------- ---------- -
Wigglesworthia     KLKEY----- ---------- ---------- ---------- -
Pl. falciparum     HYVSKF---- ---------- ---------- ---------- -
Pl. yoelii         SYKYDVV--- ---------- ---------- ---------- -
Toxoplasma gondii  EVRRFSPSLL KECDEI---- ---------- ---------- -
Sulfolobus         VQGKIAFDGK NVLPIRGVNA FDKSSRYPV- ---------- -
Streptococcus      YTICSGEVIY QA-------- ---------- ---------- -
Geobacter          FTIVGGKVVY KR-------- ---------- ---------- -
Clostridium        YTLVNGNVVV REKVLL---- ---------- ---------- -
Thermoanaerobacter YTIVEGQIRY QKNKKFEKVE I--------- ---------- -
Magnetococcus      YTILAGRIVH QEA------- ---------- ---------- -
Thermobifida       ATFLRGVPTV LDGKIQ---- ---------- ---------- -
Streptomyces       HTWLRGKATL VDGKLT---- ---------- ---------- -
Corynebacterium    HTVLRGKVTC ADGVAQNA-- ---------- ---------- -
Desulfovibrio      SHWMGGHRIA ---------- ---------- ---------- -
Bifidobacterium    ATIIGSQLMF SRL------- ---------- ---------- -
Chlorobium         GIYHNSKLIM R--------- ---------- ---------- -
Thermus aquaticus  LTLVEGRIVH EALE------ ---------- ---------- -
Bacillus           MTIVGGKIAW QKESALV--- ---------- ---------- -
Enterococcus       MTLSTENLCG RRENKSMERL LILEDGTVFE GKAFGA---- -
Oceanobacillus     LTMVNGVVVY EEAKQHEEA- ---------- ---------- -
Fusobacterium      LTISNGKIAY IDKEEINL-- ---------- ---------- -
Lactobacillus      MTLVDGEVVY QRGTK----- ---------- ---------- -
Oenococcus         RTWVNGQQVY AKGDEK---- ---------- ---------- -
Clostridium        VTIKNGKIVY NGEF------ ---------- ---------- -
```

VITA

Tamiko Neal Porter

Education:

Texas A&M University
College Station, TX
Ph.D. Chemistry
December 2004

Michigan State University
East Lansing, MI
B.S. Biochemistry
May 1997

Permanent Address:

3209 Salzburg Ct.
College Station, TX 77845