# STRUCTURAL AND FUNCTIONAL
# STUDIES OF PROLINE CATABOLIC ENZYMES

---

A Dissertation
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

---

In Partial Fulfillment
Of the Requirements for the Degree

Doctor of Philosophy

---

By

TOMMI ANNA WHITE

Dr. John J. Tanner, Dissertation Supervisor

MAY 2007

The undersigned, appointed by the dean of the Graduate School, have examined
the dissertation entitled

STRUCTURAL & FUNCTIONAL STUDIES OF BACTERIAL PROLINE
CATABOLIC ENZYMES

presented by Tommi Anna White,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

_____

Professor John J. Tanner

_____

Professor Lesa Beamer

_____

Professor Peter Tipton

_____

Professor Michael Henzl

_____

Professor Kent Gates

## DEDICATION

To the two most loving people, I dedicate the work contained in this

dissertation to my husband, Frank Barhorst, and my mother, Margaret White.

# TABLE OF CONTENTS

## 3.  MECHANISM-BASED INACTIVATION OF *THERMUS THERMOPHILUS* PROLINE DEHYDROGENASE BY *N*-PROPARGYLGLYCINE

## 4. PRELIMINARY STUDIES OF INTERACTIONS BETWEEN *T.THERMOPHLUS* PROLINE DEHYDROGENASE & $\Delta^1$-PYRROLINE-5-CARBOXYLATE DEHYDROGENASE

5.  **ATTEMPTED STRUCTURE DETERMINATION OF *BRADYRHIZOBIUM JAPONICUM* PUTA: INHERENT CRYSTALLOGRAPHIC PATHOLOGY**

# LIST OF FIGURES

4. CHAPTER 4

# LIST OF APPENDICES

# ABSTRACT

Catabolism is the oxidation of organic nutrients into simple end molecules to extract energy. For proline, the catabolic pathway involves two enzymes, L-proline dehydrogenase (PRODH, EC 1.5.99.8) and L-$\Delta^1$-pyrroline-5-carboxylate dehydrogenase (P5CDH, EC 1.5.1.12). Via the action of PRODH and P5CDH, proline can be utilized both as a carbon and nitrogen source. The conventional view of proline catabolism was that PRODH and P5CDH appear as separate enzymes in eukaryotes and as fused bifunctional enzymes (PutA) in bacteria. Analysis of genome sequence data, however, revealed a more complex situation for bacteria. The updated view is that PutAs are indeed restricted to bacteria, but monofunctional PRODHs and P5CDHs appear in both eukaryotes and bacteria. One of these newly discovered bacterial monofunctional PRODHs was chosen and characterized. This work resulted in kinetic and structural analysis of PRODH from *Thermus thermophilus*. *T. thermophilus* PRODH was also used for studying mechanism-based inactivation by *N*-propargylglycine. This work has also resulted in the first structure of a covalently modified PRODH as well as characterization of inactivation kinetics. Physical and functional interactions between monofunctional *T. thermophilus* PRODH and P5CDH utilizing coexpression have also been studied and preliminary results reported. Finally, the structure determination of *Bradyrhizobium japonicum* PutA from pseudomerohedrally twinned crystal is reported.

**INTRODUCTION**

## Proline - a unique amino acid

L-proline is unique among the 20 naturally occurring amino acids (FIGURE 1.1). It is the only one in which the side chain is bonded to both the backbone N atom and the $C_\alpha$ atom. Thus proline has a 5-membered ring that unites the side chain and main chain. Note also that proline is a secondary amine in contrast to the other amino acids, which have primary amino groups in the backbone.

Because of its unique structure, proline strongly influences local peptide conformation and protein folding energetics. For example, proline has low frequencies of occurrence in $\alpha$-helices and $\beta$-strands, and high frequency of occurrence in reverse turns. This is due, in part, to the fact that proline in a polypeptide chain does not have an N-H group that can donate $i$ to $i+4$ hydrogen bonds in $\alpha$-helices and interstrand hydrogen bonds in $\beta$-sheets. The 5-membered ring also imparts local rigidity to proline-containing polypeptides thereby decreasing the volume of available conformational space. Thus introduction of proline into proteins can increase thermal stability by lowering the conformational entropy of protein folding (1). Another unique feature of proline-containing polypeptides is that *cis* and *trans* peptide bond conformations are equally favorable for L-proline, whereas the *trans* conformation is favored for all other amino acids (2). Finally, L-proline and L-hydroxyproline (FIGURE 1.2) are largely responsible for the unique triple helical structure of collagen, which composes skin and bones.

Besides playing unique structural roles in proteins, L-proline protects organisms against various stresses. For example, plants accumulate proline under osmotic stress induced by drought, high salinity and freezing. In this context, proline is thought to serve as an osmolyte, which is a molecule that increases thermodynamic stability of folded proteins.

Proline also protects against oxidative stress. For example, it has been shown that proline and other pyrrolidine derivatives protect skin against UVA-induced photodamage by serving as effective quenchers of singlet oxygen, a reactive oxygen species (ROS) (3). Chen and Dickman found that proline prevents ROS-mediated apoptosis in fungus (4). Chen et al. subsequently showed that proline can protect yeast against oxidative stress (5). These studies suggest that proline may be a broad-based responder molecule during cellular stress and thus the enzymes that synthesize and degrade proline may be important in mediating the cellular stress response.

## The enzymes of proline biosynthesis and catabolism

Biosynthesis of proline from glutamate begins with the 2-step conversion of L-glutamate to L-glutamic semialdehyde via the intermediate L-glutamyl-5-phosphate (FIGURE 1.3). These two steps are catalyzed by the enzymes $\gamma$-glutamyl kinase ($\gamma$-GK, EC 2.7.2.11) and glutamate-5-semialdehyde dehydrogenase (GSAD, EC 1.2.1.41)(6). In eukaryotes, $\gamma$-GK and GSAD are fused into the bifunctional enzyme $\Delta^1$-pyrroline-5-carboxylate (P5C) synthetase (P5CS), whereas the two enzymes are separate molecules in bacteria. Proline is

then produced from the action of $\Delta^1$-pyrroline-5-carboxylate reductase (P5CR, EC 1.5.1.2) following nonenzymatic ring closure of L-glutamic semialdehyde to P5C (FIGURE 1.4).

Catabolism is the oxidation of organic nutrients into simple end molecules to extract energy. For proline, the catabolic pathway involves two enzymes, L-proline dehydrogenase (PRODH, EC 1.5.99.8) and L-$\Delta^1$-pyrroline-5-carboxylate dehydrogenase (P5CDH, EC 1.5.1.12) (SCHEME 1). PRODH is an FAD-dependent enzyme that catalyzes the oxidation of proline to P5C. As in proline biosynthesis, an equilibrium is established between P5C and its hydrolysis product L-glutamic semialdehyde. The latter molecule is the substrate for the second enzyme of proline catabolism, P5CDH, which catalyzes the oxidation of the semialdehyde to glutamate. P5CDH is an $NAD^+$-dependent enzyme having a catalytic Cys. As I elaborate in Chapter 2, PRODH and P5CDH are separate enzymes in eukaryotes and some bacteria, whereas the two enzymes are fused in other bacteria. The fused enzymes are known as Proline utilization A (PutA) and will be discussed in more detail below.

Via the action of PRODH and P5CDH, proline can be utilized both as a carbon and nitrogen source. In some organisms, like *Helicobacter* spp. and *Glossina morsitans*, it has been reported that L-proline is the preferred energy source (7,8). Electrons stored in the reduced $FADH_2$ of PRODH can be transferred to quinone-like molecules in the membrane to participate in the electron transport chain to generate adenosine triphosphate (ATP) for energy. The product of PRODH catalysis, P5C, can be further catabolized to generate

glutamate which can enter tri-carboxylic cycle as $\alpha$-ketoglutarate or P5C can be converted to ornithine for entry into the urea cycle. In eukaryotes, PRODH and P5CDH are mitochondrial enzymes and P5C can also be shuttled out of the mitochondria into the cytosol for conversion back to L-proline. This shuttling of reducing equivalents into and out of the mitochondria is known as a proline cycling and it has been shown using human cancer cell lines and yeast that proline cycling contributes to the redox status of the cell (9,10).

FIGURE 1.1.  The amino acid L-proline.
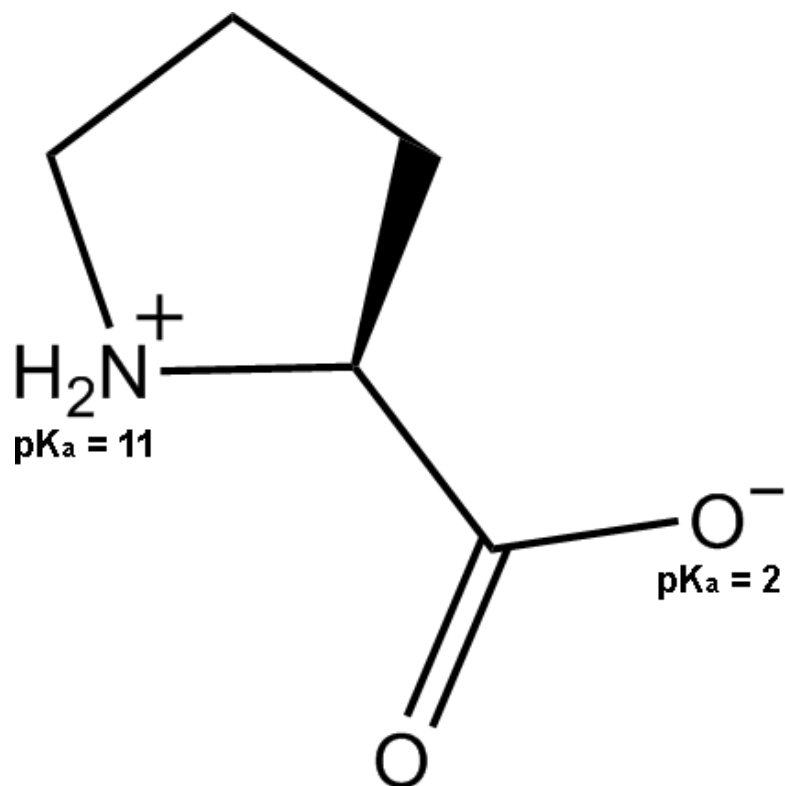


L-proline

FIGURE 1.2. L-proline and proline analogs. Reported substrates are in the top two rows, reported mechanism-based inactivators are in row three, and reported competitive inhibitors are in rows 4 and 5.



L-proline

3,4-dehydro-L-proline

L-thiazolidine-4-carboxylate

trans-4-hydroxy-L-proline

cis-4-hydroxy-L-proline

4-methylene-L-proline

N-propargylglycine

tetrahydro-2-furoic acid

L-Lactate

Acetate

L-mandelate

Glycine

L-azetidine-2-carboxylate
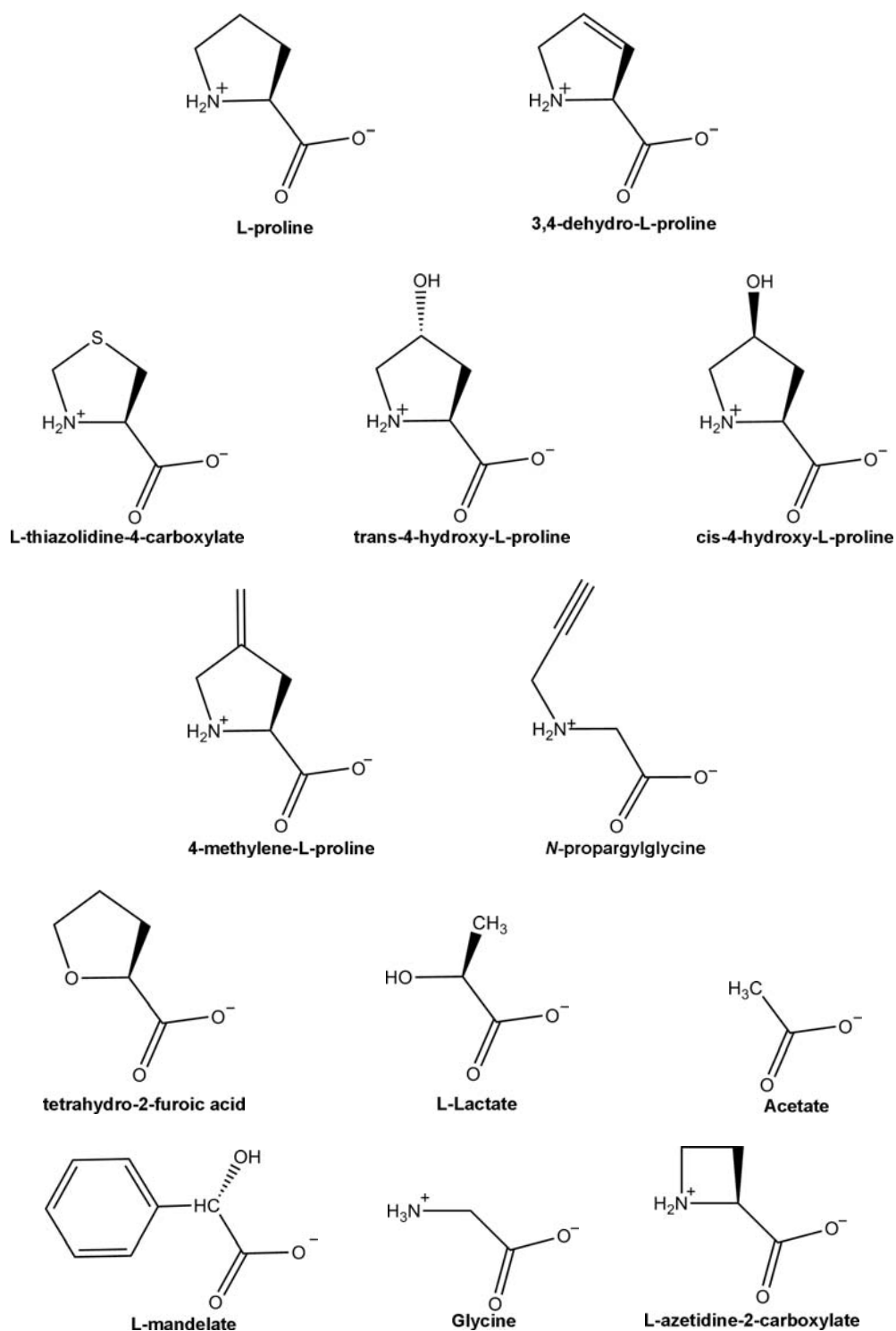
FIGURE 1.3. Proline Biosynthesis: L-glutamic semialdehyde formation from glutamate. $\gamma$-Glutamyl Kinase ($\gamma$-GK) converts glutamate to L-glutamyl-5-phosphate. Glutamic semi-aldehyde dehydrogenase (GSAD) catalyzes the next step of the reaction to form glutamic semialdehyde. In eukaryotes, these two enzymes are fused into the bifunctional enzyme $\Delta^1$-pyrroline-5-carboxylate synthetase (P5CS).
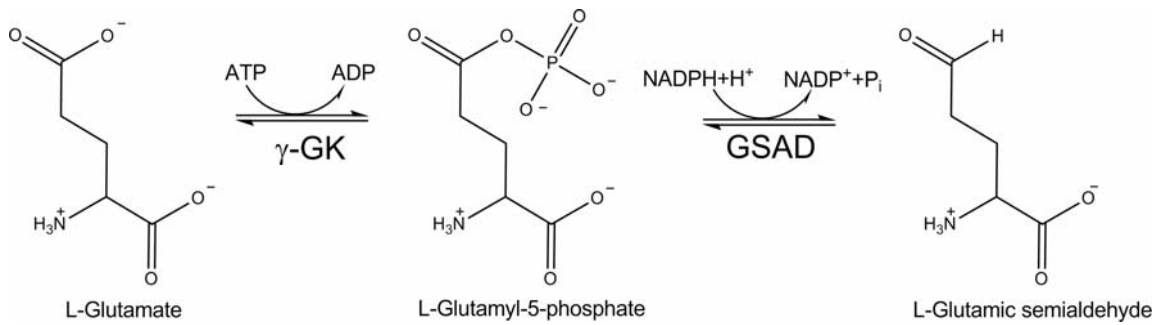
FIGURE 1.4. Proline Biosynthesis: Conversion of L-glutamic semialdehyde to L-proline. The final enzymatic step is catalyzed by $\Delta^1$-pyrroline-5-carboxylate reductase (P5CR).
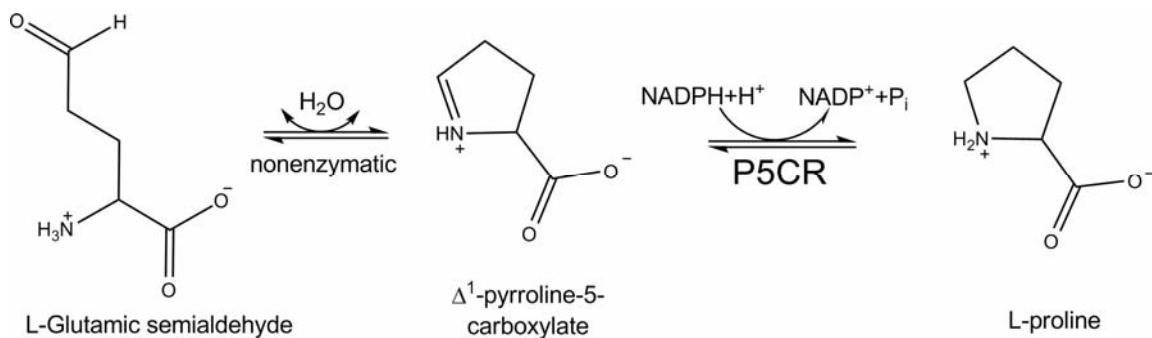
SCHEME 1.1. Reactions catalyzed by L-proline dehydrogenase (PRODH) and L-$\Delta^1$-pyrroline-5-carboxylate dehydrogenase (P5CDH).
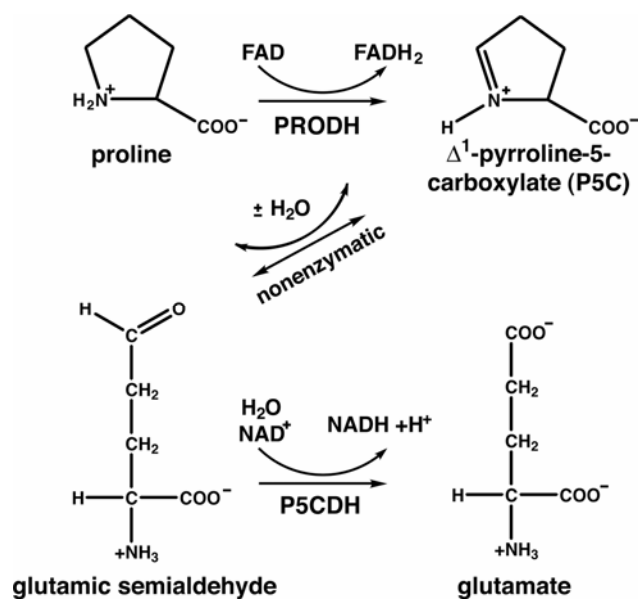
## Importance of Proline Catabolism in Health and Disease

PRODH and P5CDH in eukaryotes are peripherally-associated inner mitochondrial membrane enzymes. These enzymes are synthesized in the cytoplasm and transported into the mitochondria where the mitochodrial-localization sequence is cleaved (11). Humans have two isozymes of PRODH, which share about 50% amino acid sequence identity (12). PRODH1 is encoded on chromosome 19, is expressed almost exclusively in liver and kidney, and catalyzes oxidation of L-hydroxyproline. PRODH2 is encoded on chromosome 22q11, is expressed more widely than PRODH1 (brain, heart, pancreas, kidney, liver) and specifically oxidizes L-proline. PRODH2 is also known as proline oxidase or POX, since it, as well as the yeast homolog, exhibits reactivity with molecular oxygen to produce superoxide (11,13). Eukaryotic PRODHs have been studied in a few organisms using mitochondrial extracts or by enzyme overexpression, but there have been no reported characterizations of soluble, purified eukaryotic PRODH.

Inborn defects in PRODH and P5CDH cause the metabolic disorders hyperprolinaemia I & II, respectively. Hyperprolinaemia I is an autosomal recessive disorder and is characterized by increased plasma proline levels. Efron reported in 1965 that hyperprolinaemia I maps to mutations in chromosome 22q11.2, which implicated PRODH2 rather than PRODH1. Interestingly, mutations in 22q11 are associated with a variety of disorders, including schizophrenia susceptibility, DiGeorge syndrome and velocardiofacial syndrome.

The link between defects in PRODH2 and brain disorders has been

examined in a variety of studies. For example, the PRODH gene knockout strain of *Drosophila melanogaster* (*slgA* mutant) exhibits sluggish movement in response to light, no detectable proline oxidase activity and elevated proline levels (14). The authors speculated that the defect in PRODH affected neural glutamate pools and also suggested the possibility that proline itself may be a neurotransmitter.

A PRODH-mutant mouse has been informative in characterizing links between PRODH and neuropsychological disorders. In 1976, a mouse displaying hyperprolinaemia I was bred (15). Further characterization of these mice, known as ProRE mice, revealed a truncation of the C-terminal end of PRODH2, resulting in a non-functional enzyme (16). The crystal structure of the *E. coli* PutA PRODH domain would later show that the truncated PRODH expressed by ProRE mice lacked helix 8, which contributes several residues to the active site (17). These mice exhibit decreased sensorimotor gating and altered neurochemistry (16). Sensorimotor gating is a neural filtering process, which allows focus to be maintained on a stimulus and it is one of the few neuropsychological attributes that can be assessed similarly in humans and rodents. Based on this work, PRODH2 was speculated to be involved in various psychiatric and behavioral disorders, including schizophrenia.

The relationship between defects in PRODH2 and schizophrenia was further studied with genetic linkage analysis performed by Karayiorgou's group at The Rockefeller University. Analysis of DNA from patients diagnosed with schizophrenia revealed several single nucleotide polymorphisms (SNPs) within

exons that could potentially underlie disease susceptibility (18). I simulated the Leu441Pro missense mutation identified by the Rockefeller group by creating the homologous Leu432Pro mutation of the *E. coli* PRODH domain of PutA. This mutant had a 5-fold lower $k_{cat}$ and a severe decrease in thermostability compared to the native enzyme. This was the first demonstration that a schizophrenia-related mutation affects enzyme function (19). Valle's group subsequently determined the effect of several of the schizophrenia-associated SNPs on enzyme function using mitochondrial extracts. Four of the mutants, including Leu441Pro, showed greater than 30% decrease in activity compared to wildtype (20).

Human PRODH2 has also been implicated in apoptosis and cancer. Although it has been well documented that the p53 signaling cascade ends with damage to mitochondrial membranes, metabolic environment may play a role in expression of the cancer phenotype by alternative pathways or modulating known pathways. In a colorectal cancer cell line sensitive to p53, PRODH was shown by serial analysis of gene expression (SAGE) that the gene is upregulated (10-fold) prior to p53-mediated apoptosis. Only 14 other genes were upregulated to the same extent (21). This has been confirmed in various cancer cell lines including bladder, lung, ovarian and renal carcinomas, and colorectal cancer cell lines (13,22-25). By generation of reactive oxygen species, specifically superoxide radical, in a proline-dependent manner, various cancer cell lines undergo apoptosis (13). Proline-dependent PRODH apoptosis can be abrogated by overexpression of mitochondrial Mn-superoxide dismutase (24). This ROS

generation has been revealed to play roles in PRODH-mediated apoptosis, not only through intrinsic (mitochondrial) pathways, but also extrinsic pathways (26). TRAIL and DR5 are death receptors involved in extrinsic pathways and were reported to be activated by PRODH (26). Further mechanisms of regulation may involve the mTOR (mammalian target of rapamycin), which is similar to yeast regulation of proline catabolic enzymes (J. Phang, personal communication) (27).

## Bacterial Proline Catabolism

Early work on bacterial proline catabolic enzymes revealed that PRODH and P5CDH were combined into a single peripheral membrane associated polypeptide known as Proline utilization A (PutA). This early work focused on PutAs from *Escherichia coli* and *Salmonella typhimurium*, and the *E. coli* enzyme remains the best characterized PutA. In addition to having two enzymatic activities, PutAs from *E. coli* and S. *typhimurium* are autogenous transcriptional repressors. By binding to the 419 base-pair *put* intergenic region of DNA, PutAs repress divergent transcription of the genes *putA* and *putP* (encodes a sodium-proline symporter) (28). Thus, PutAs from *E. coli* and S. *typhimurium* are known as "trifunctional" PutA proteins. Upon reduction of the PutA flavin cofactor by L-proline, PutA gains affinity for the inner bacterial membrane, thereby relieving transcriptional repression (19). After oxidation of L-proline, the reduced cofactors of PRODH ($FADH_2$) and P5CDH (NADH) transfer reducing equivalents to membrane electron acceptors (quinone and menaquinone) for entry into the electron transport chain to obtain energy by chemiosmostic phosphorylation.

FIGURE 1.5. Domain composition of trifunctional and bifunctional Proline Utilization A proteins. Trifunctional (*E. coli*) PutA is composed of a N-terminal autogenous transcriptional repression domain (red), a proline dehydrogenase catalytic domain (yellow), a $\Delta^1$-pyrroline-5-carboxylate dehydrogenase catalytic domain (blue), and a C-terminal domain of unknown function (?). Bifunctional (*B.japonicum*) PutA have similar catalytic domains (yellow and blue), but lack the autogenous transcriptional repression and C-terminal domains. Corresponding protein structures for each domain, if solved, are shown.
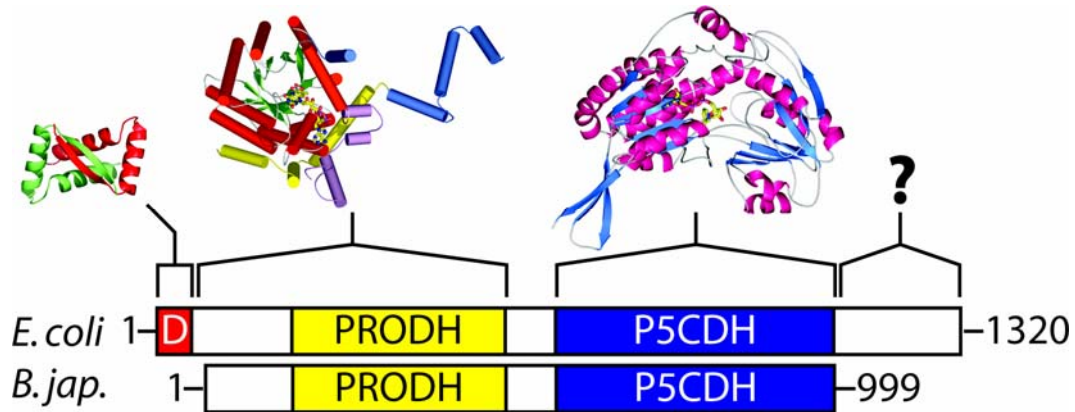
Not all PutAs serve as transcriptional repressors. One example of these "bifunctional" PutAs that has been studied in our lab is PutA from *Bradyrhizobium japonicum* (BjPutA). This protein has only 999 residues compared to over 1300 residues for trifunctional PutAs. Compared to the *E. coli* protein, BjPutA lacks the N-terminal DNA binding domain and a C-terminal 200 residue domain of unknown function. Characterization of BjPutA revealed that the protein is associated with the membrane independent of redox state, and there is no redox associated change in localization as with *E. coli* PutA.

Fusion of enzymes catalyzing sequential steps of a metabolic pathway, as in PutA, provides a kinetic advantage because the intermediate can be channeled between active sites (29,30). Indeed, Maloy's group reported kinetic data supporting substrate channeling for *Salmonella typhimurium* PutA (31). Becker's group has also obtained kinetic evidence for channeling in PutAs from *E. coli* and *B. japonicum* (D.F. Becker, personal communication).

Eisenberg and co-workers refer to such fused proteins as Rosetta Stone proteins because they decipher interactions between protein pairs (32). Thus, the Rosetta Stone hypothesis of protein evolution predicts that eukaryotic PRODH and P5CDH form physical and functional interactions. This hypothesis has not been tested for PRODH and P5CDH.

## Structural Information for Proline Catabolic Proteins

Our group has determined crystal structures of the PRODH and DNA-binding domains of *E. coli* PutA. The PRODH domain contains a unique $\beta_8\alpha_8$

barrel in which the final helix of the barrel topology lies atop the C-terminal face of the barrel rather than alongside the barrel. Structures of the PRODH domain complexed with inhibitors acetate, L-lactate, and L-tetrahydro-2-furoic acid (THFA) provided insights into substrate recognition and showed that the carboxyl group is the minimal functional unit recognized by the enzyme. Most recently, the structure of the PRODH domain reduced by dithionite provided a glimpse of local conformational changes caused by reduction of the flavin. The structure of the DNA-binding domain of *E. coli* PutA was also recently determined. This work showed that the DNA-binding domain is a ribbon-helix-helix domain. A structure of the ribbon-helix-helix domain complexed with DNA has also recently been determined (J.J. Tanner, personal communication). Structure determination for full-length PutAs has been problematic, although progress on this front will be described the last chapter of this dissertation.

## Summary of Research in This Dissertation

When I started my research, the conventional view of proline catabolism was that PRODH and P5CDH appear as separate enzymes in eukaryotes and as fused bifunctional enzymes (PutA) in bacteria. My analysis of genome sequence data, however, revealed a more complex situation for bacteria. The updated view is that PutAs are indeed restricted to bacteria, but monofunctional PRODHs and P5CDHs appear in both eukaryotes and bacteria. These bioinformatics studies are described in Chapter 2. I then proceeded to characterize one of the newly discovered bacterial monofunctional PRODHs. This work resulted in

kinetic and structural analysis of PRODH from *Thermus thermophilus*, which is described in Chapter 2. I continued this line of research by studying mechanism-based inactivation of *T. thermophilus* PRODH by *N*-propargylglycine. This work has resulted in the first structure of a covalently modified PRODH, as well as characterization of inactivation kinetics (Chapter 3). Chapter 4 describes initial studies that attempt to test the Rosetta Stone hypothesis using PRODH and P5CDH from *T. thermophilus*. For this work I cloned both genes into the pET-Duet1 vector and obtained preliminary data suggesting that the two enzymes co-purify. This work will be continued by others in the future. Finally, Chapter 5 describes attempts to solve the structure of the full-length PutA from *B. japonicum* using data collected from twinned crystals.

# REFERENCES

1. Agah, S., Larson, J. D., and Henzl, M. T. (2003) *Biochemistry* **42**, 10886-10895
2. Pauling, L., Corey, R. B., and Branson, H. R. (1951) *Proc Natl Acad Sci U S A* **37**, 205-211
3. Wondrak, G. T., Jacobson, M. K., and Jacobson, E. L. (2005) *J Pharmacol Exp Ther* **312**, 482-491
4. Chen, C., and Dickman, M. B. (2005) *Proc Natl Acad Sci U S A* **102**, 3459-3464
5. Chen, C., Wanduragala, S., Becker, D. F., and Dickman, M. B. (2006) *Appl Environ Microbiol* **72**, 4001-4006
6. Hu, C. A., Lin, W. W., and Valle, D. (1996) *J Biol Chem* **271**, 9795-9800
7. Hoek, J. B., Pearson, D. J., and Olembo, N. K. (1976) *Biochem J* **160**, 253-262
8. Nagata, K., Nagata, Y., Sato, T., Fujino, M. A., Nakajima, K., and Tamura, T. (2003) *Microbiology* **149**, 2023-2030
9. Phang, J. M., Yeh, G. C., and Hagedorn, C. H. (1981) *Life Sci* **28**, 53-58
10. Brandriss, M. C., and Magasanik, B. (1981) *J Bacteriol* **145**, 1359-1364
11. Wang, S. S., and Brandriss, M. C. (1987) *Mol Cell Biol* **7**, 4431-4440
12. Phang, J. M., Hu, C. A., and Valle, D. (2001) in *Metabolic and molecular basis of inherited disease* (Scriver, C. R., A.L., B., Sly, W., and Valle, D., eds), McGraw Hiss, New York
13. Donald, S. P., Sun, X. Y., Hu, C. A., Yu, J., Mei, J. M., Valle, D., and Phang, J. M. (2001) *Cancer Res* **61**, 1810-1815
14. Hayward, D., Delaney, S., Campbell, H., Ghysen, A., Benzer, S., Kasprzak, A., Cotsell, J., Young, I., and Gabor Miklos, G. (1993) *PNAS* **90**, 2979-2983
15. Blake, R. L., Hall, J. G., and Russell, E. S. (1976) *Biochem Genet* **14**, 739-757
16. Gogos, J., Santha, M., Takacs, Z., Beck, K., V L., Lucas, L., Nadler, J., and Karayiorgou, M. (1999) *Nature Genetics* **21**, 434-439
17. Lee, Y. H., Nadaraia, S., Gu, D., Becker, D. F., and Tanner, J. J. (2003) *Nature Structural Biology* **10**, 109-114
18. Liu, H., Heath, S., Sobin, C., Roos, J., BL, G., Blundell, M., Lenane, M., Robertson, B., Wijsman, E., Rapoport, J., Gogos, J., and Karayiorgou, M. (2002) *PNAS* **99**, 3717-3722
19. Zhang, M., White, T. A., Schuermann, J., Baban, B. A., Becker, D. F., and Tanner, J. J. (2004) *Biochemistry* **43**
20. Bender, U., S, A., G, S., A, W., and Valle, D. (2004) in *Proline Symposium*, Fort Deterich, MD
21. Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W., and Vogelstein, B. (1997) *Nature* **389**, 300-305
22. Rivera, A., and Maxwell, S. A. (2005) *J Biol Chem* **280**, 29346-29354
23. Maxwell, S. A., and Rivera, A. (2003) *J Biol Chem* **278**, 9784-9789

24. Liu, Y., Borchert, G. L., Donald, S. P., Surazynski, A., Hu, C. A., Weydert, C. J., Oberley, L. W., and Phang, J. M. (2005) *Carcinogenesis* **26**, 1335-1342

25. Hu, C. A., Donald, S. P., Yu, J., Lin, W. W., Liu, Z., Steel, G., Obie, C., Valle, D., and Phang, J. M. (2006) *Mol Cell Biochem*

26. Liu, Y., Borchert, G. L., Surazynski, A., Hu, C. A., and Phang, J. M. (2006) *Oncogene* **25**, 5640-5647

27. Kuruvilla, F. G., Shamji, A. F., and Schreiber, S. L. (2001) *Proc Natl Acad Sci U S A* **98**, 7283-7288

28. Wood, J. M. (1981) *J Bacteriol* **146**, 895-901

29. Miles, E. W., Rhee, S., and Davies, D. R. (1999) *J Biol Chem* **274**, 12193-12196

30. Huang, X., Holden, H. M., and Raushel, F. M. (2001) *Annu Rev Biochem* **70**, 149-180

31. Surber, M. W., and Maloy, S. (1998) *Arch Biochem Biophys* **354**, 281-287

32. Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999) *Science* **285**, 751-753

# CHAPTER 2

## Structure and Kinetics of Monofunctional Proline Dehydrogenase from *Thermus thermophilus*

Portions of this work are reproduced from the following:
- *Acta Crystallographica* Section F © 2005 Volume 61 pages 737-739 by copyright permission of the International Union of Crystallography
- *Journal of Biological Chemistry* © 2007 in press

## INTRODUCTION

Oxidation of amino acids is a central part of energy metabolism. The oxidative pathway for proline consists of two enzymatic steps and an intervening nonenzymatic equilibrium (Scheme 2.1, (1,2)). The first enzymatic step transforms proline to $\Delta^1$-pyrroline-5-carboxylate (P5C), which is nonenzymatically hydrolyzed to glutamic semialdehyde. The semialdehyde is oxidized in the second enzymatic step to glutamate.

This 4-electron transformation of proline is common to all organisms, but the enzymes of proline catabolism differ widely among the three kingdoms of life. Amino acid sequence analysis shows that bacteria and eukaryotes share a common set of proline catabolic enzymes called proline dehydrogenase (PRODH) and P5C dehydrogenase (P5CDH). Studies of the bacterial enzymes have shown that PRODH is an FAD-dependent enzyme with a $(\beta\alpha)_8$ barrel catalytic core (3,4) and that P5CDH is an $NAD^+$-dependent Rossmann fold enzyme featuring a nucleophilic Cys (5). These enzymes are unrelated in sequence and structure to hyperthermophilic archaeal proline catabolic enzymes, which appear in unique hetero-tetrameric and hetero-octameric complexes (6).

An intriguing aspect of proline catabolism in eukaryotes and bacteria is that PRODH and P5CDH are encoded on separate genes in some organisms, whereas the two genes are fused in other organisms. The traditional view has been that PRODH and P5CDH appear as separate enzymes in eukaryotes and fused bifunctional enzymes known as Proline utilization A (PutA) (7-15) in bacteria. Fusion of enzymes catalyzing sequential steps of a metabolic pathway provides a kinetic advantage because the intermediate can be channeled between active sites (16,17). Indeed, Maloy's group reported kinetic data supporting substrate channeling for *Salmonella typhimurium* PutA (18). Eisenberg and co-workers refer to fused proteins, such as PutA, as Rosetta Stone proteins because they decipher interactions between protein pairs (19). Thus, the Rosetta Stone hypothesis of protein evolution predicts that eukaryotic PRODH and P5CDH form physical and functional interactions.

As we elaborate in this paper, separate PRODH and P5CDH enzymes are not restricted to eukaryotes, but also appear in some bacteria. These bacterial monofunctional enzymes thus represent a new and unexplored group of enzymes. In addition, bacterial monofunctional PRODHs may serve as good models for understanding human PRODH, which has been difficult to purify for biochemical and biophysical study. Moreover, the bacterial enzymes are convenient systems for studying protein-protein interactions and intermolecular channeling. To begin exploring these questions, we previously reported the cloning, isolation and crystallization of the monofunctional PRODH from *Thermus*

*thermophilus* (TtPRODH) (20).  Here we report the crystal structure and kinetic characterization of this enzyme.


## MATERIALS AND METHODS

### Cloning of TtPRODH

The *T. thermophilus* PRODH gene was cloned from genomic DNA purchased from American Type Culture Collection and introduced into the plasmid pKA8H between BamH I and Nde I sites. The pKA8H vector codes for an N-terminal 8x-His affinity tag and a tobacco etch virus protease site.  Since the PRODH gene contains a BamH I site, digestion of the PCR product with BamH I was not possible.  Therefore, the staggered reannealing method (21) was used with the following three primers:

Forward: 5'-CCTTGATCATATGAACCTGGACCTGGCTTACCGTTC-3',

Reverse 1: 5'-GATCCCTAGCCGGAAACCAGGCTCCTCAGG-3'

Reverse 2: 5'-CCTAGCCGGAAACCAGGCTCCTCAGG-3'

Two separate PCR amplification experiments were performed using the forward primer in conjunction with each of the two reverse primers.   The two PCR products were purified, mixed in equimolar amounts, denatured at 96$^o$ C for 5 minutes, annealed by slow cooling and finally, digested with Nde I.   The resulting PCR product was ligated into pKA8H, which had been digested with BamH I and Nde I.  Sequencing confirmed that the gene had been successfully cloned into the vector.

### Native TtPRODH Expression and Purification

Unless otherwise stated, all chemicals were purchased from Fisher Scientific or Sigma Aldrich. PRODH was expressed in BL21(DE3)pLysS cells (Novagen) as follows. Small (10 mL) cultures were grown overnight in LB media and used to inoculate 1.5 L of LB media. Protein expression was induced with IPTG (0.5 mM) after the culture reached an optical density of $OD_{600}$ = 0.6. Cells were harvested 3 hours after induction, resuspended in 50 mM $NaH_2PO_4$, 300 mM NaCl, 10 mM imidazole, 5% glycerol, pH 8.0 and frozen. Cells were thawed at $4^o$ C and lysed by French press at 16,000 psi in the presence of five protease inhibitors (0.1 mM TPCK, 0.05 mM AEBSF, 0.1 $\mu$M Pepstatin, 0.01 mM Leupeptin, 5 $\mu$M E-64). The supernatant was collected after centrifugation at 15000 rpm for 30 minutes at $4^o$C, filtered through a 0.45 µm filter (Millipore) and applied to 5 mL of Ni-NTA Superflow resin (Qiagen). The column was washed in two steps with the loading buffer supplemented with 50 mM imidazole followed by 75 mM imidazole. PRODH was eluted with 250 mM imidazole and dialyzed overnight in the dark at 4 °C into 50 mM Tris HCl, 50 mM NaCl, 0.5mM EDTA, 0.5 mM DTT and 5% glycerol, pH 8.0 (FIGURE 2.1). Prior to dialysis, flavin adenine dinucleotide (FAD), at a concentration of 0.1 mM, was added to the dialysis bag containing the eluted PRODH. Excess FAD was removed using a desalting column (Biorad P100). The protein was concentrated to 13 mg/mL using centrifugal concentrators. Protein concentration was determined using the Bradford Method with bovine serum albumin as the standard.

The purified protein exhibited the intense yellow color that is characteristic of flavoenzymes. The enzyme displayed PRODH activity as measured by an

assay described previously (3). The molecular mass as determined by MALDI-TOF mass spectrometry at the University of Missouri-Proteomics Center was 37,968 ± 3 Da. This value was in good agreement with the theoretical mass predicted from the gene sequence (37,923 Da).

**Dynamic light scattering**

Gel-filtration chromatography suggested that the protein exhibited a high degree of aggregation. For example, protein injected onto a Superdex-200 column eluted entirely in the void volume. The correlation between monodispersity and crystallizability has been well established (22,23), therefore; the protein was exposed to various solvent conditions in an attempt to identify solution conditions that promoted monodispersity. The monodispersity of each protein solution tested was assessed with a Protein Solutions DynaPro 99 Molecular Sizing Instrument. The parameters varied included pH, ionic strength, addition of Pro and Pro analogues and addition of various detergents, including n-octyl β-D-glucopyranoside (BOG). The protein concentration was in the range 1 – 3 mg/mL.

Only BOG had a significant effect on protein aggregation. When BOG was added to a final concentration of 20 mM, high molecular weight species was dramatically reduced. The protein/BOG solution was shown by dynamic light scattering to be reasonably monodisperse by analyzing polydispersity (Cp/Rh = 37 %) with an apparent protein mass of 35 kDa. Based on these results, the purification procedure for both native and selenomethionine TtPRODH was modified by the addition of 20 mM BOG to the protein after the final dialysis step.

After overnight dialysis, excess detergent and FAD were removed using a desalting column (Biorad P100).

**Native Crystallization and Attempted Molecular Replacement**

All crystallization experiments were performed at 295 K, using the sitting drop method of vapor diffusion, forming the drops with equal volumes of the reservoir and protein solutions. Commercially available crystal screens (Hampton Research and Decode Genetics) were used to identify initial crystallization conditions. Several conditions in the screens yielded crystals of various size and quantity. The precipitating agent 2-methyl-2, 4-pentanediol (MPD) was present in many of the positive conditions (FIGURE 2.2). After several rounds of optimization, the best crystals were grown with a reservoir containing 50 mM $MgCl_2$, 100 mM imidazole pH 7.5, 35 % MPD (FIGURE 2.3). Since the mother liquor provided cryoprotection, the crystals were picked up with Hampton mounting loops and frozen directly in liquid nitrogen.

The space group is $P2_12_12_1$ with unit cell dimensions of a = 82.2 Å, b = 89.6 Å, and c = 94.3 Å. There are two protein molecules per asymmetric unit, with 46 % solvent content and Matthews coefficient of 2.3 $Å^3$ $Da^{-1}$ (24).

Molecular replacement calculations (high resolution limit = 4 Å) were performed with MOLREP (25) using the $\beta_8\alpha_8$ barrel of the PRODH domain of *E coli* PutA (3,4) as the search model (PDB code 1TIW). The top solution had R-factor = 0.57 and correlation coefficient = 0.23, which indicated that molecular replacement was not a suitable phasing method. Therefore, the structure of *T.*

*thermophilus* PRODH would be determined with multiwavelength anomalous dispersion phasing using a selenomethionyl derivative.

**Expression, Purification and Crystallization of Se-Met TtPRODH**

A 10 mL overnight culture grown in LB media was pelleted, and the LB medium was removed.  The cells were resuspended in 3 mL M9 media and diluted into 1.5 L of M9 media.  The culture was then grown to an optical density of 0.5 (OD = 600 nm).  Methionine production was inhibited for 30 minutes as described previously (26), followed by induction of protein expression by the addition of 0.5 mM IPTG.  After 12 hours of induction at 22 °C with a shaking rate of 200 rpm, cells were harvested, resuspended in 50 mM $NaH_2PO_4$ at pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM DTT and 5% glycerol, and frozen.

Se-Met TtPRODH was purified using procedures described for purification of TtPRODH(20), except for the following modifications.  As observed for native TtPRODH, the pellet obtained after centrifugation of lysed cells was bright yellow, indicating that a significant amount of Se-Met TtPRODH was bound to the cell debris in addition to Se-Met TtPRODH in the supernatant.  Therefore, the enzyme was extracted from the pellet by resuspending the pellet in 20 mL of 20 mM n-octyl β-D-glucopyranoside, 50 mM $NaH_2PO_4$, 300 mM NaCl, 10 mM imidazole, 1 mM DTT, 5% glycerol, pH 8.0, followed by centrifugation. The detergent extraction step was repeated 1-2 times until no additional TtPRODH was liberated.  The extracted protein was then purified using Ni-NTA chromatography as described previously (20).

Purified Se-Met TtPRODH was dialyzed overnight in the dark into 4 L of 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.5 mM EDTA, 5 mM DTT, 0.1 mM FAD and 5 % glycerol using 10000 MWCO dialysis tubing (Spectrum Spectra/Por). After dialysis, the enzyme was passed through a 12 mL Sephadex G-25 desalting column (GE Healthcare), to remove excess FAD, and then concentrated to 13 mg/mL using centrifugal concentrating devices (Amicon, Millipore). Molecular masses of Se-Met and native TtPRODH obtained from MALDI-TOF mass spectrometry were 38208 ± 6 Da and 37968 ± 3Da, respectively, which indicated 100 % incorporation of Se-Met into all 5 Met positions of the TtPRODH polypeptide chain.

Se-Met TtPRODH was crystallized using procedures described previously for native TtPRODH (20). Briefly, crystals were grown at room temperature in sitting drops by mixing equal volumes of the enzyme (2 $\mu$L) and reservoir (2 $\mu$L) solutions (FIGURE 2.4). The enzyme solution contained 2 - 3 mg/mL Se-Met TtPRODH and 20 mM of fresh n-octyl $\beta$-D thioglucopyranoside. The reservoir solution consisted of 100 mM imidazole pH = 7, 100 mM MgCl$_2$, 17% MPD and 5 mM DTT. The crystals typically appeared within one day after setup and grew to a maximum dimension of 0.2 mm. Crystals were prepared for cryogenic data collection by soaking in 100 mM imidazole pH = 7, 200 mM MgCl$_2$ and 25 % MPD, then picked up with Hampton mounting loops and plunged into liquid nitrogen.

The space group is $P2_12_12_1$ with unit cell dimensions of a = 82.1 Å, b = 89.6 Å, and c = 94.3 Å.  There are two protein molecules per asymmetric unit, with 46 % solvent content and Matthews coefficient of 2.3 $Å^3$ $Da^{-1}$ (24).

**X-ray Diffraction Data Collection, Phasing and Refinement**

Data were collected at Advanced Light Source beamline 4.2.2 using a NOIR-1 CCD detector.  The structure was solved by single-wavelength anomalous diffraction (SAD) phasing, using data collected at the energy corresponding to the experimentally determined maximum of f″ (FIGURE 2.5). The data set used for structure determination and refinement consisted of 180 frames with crystal-to-detector distance of 170 mm, oscillation range of 1°/frame and exposure time of 30 s/frame.  Integration and scaling were performed with d*TREK (27).  Data collection and processing statistics for the experimentally determined TtPRODH structure are listed in Table 2.1.  Data from other crystals, which were soaked in various inhibitors or substrates, are listed in Table 2.2

SOLVE (28) was used to identify a constellation of anomalous scattering centers, and the resulting SAD phases were improved with solvent flattening in RESOLVE (28).  The partial chain trace from RESOLVE was used to determine the non-crystallographic symmetry (NCS) transformation relating the two molecules in the asymmetric unit.  The RESOLVE phases were then improved with NCS averaging and solvent flattening in DM (29).  The DM phases were input to ARP/wARP for automated model building (30).  The model from ARP/wARP was improved with several rounds of model building in COOT, followed by refinement against the Se-Met peak data set with REFMAC5 (31).

Topology and parameter files for FAD were created using PRODRG (32) and the Libcheck module of CCP4i (33,34).

The final model includes residues 5-296 of TtPRODH chain A and residues 1-294 of TtPRODH chain B. In addition, 6 residues of the N-terminal affinity tag were built for chain B (residues -5 to 0). The C-terminal ends of the protein are disordered with residues 297-307 and 295-307 omitted in chains A and B, respectively. The model also includes 1 FAD cofactor per TtPRODH chain, 272 water molecules and 4 MPD molecules. See Table 2.1 for refinement statistics. Coordinates and structure factor amplitudes have been deposited in the Protein Data Bank (PDB, (35)) under accession number 2G37.

**PRODH Kinetic Characterization**

TtPRODH was expressed in *E. coli* and purified as described previously (20). The purified enzyme was dialyzed into buffer containing 50 mM Tris-HCl, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM DTT and 5% glycerol at pH 8.0 and stored at 4° C. PRODH activity was measured using the proline - dichlorophenolindophenol (DCPIP) oxidoreductase assay, as described previously for PutA proteins (11). One unit of PRODH activity is the quantity of enzyme that transfers electrons from 1 µmol of proline to DCPIP per minute at 25°C. Steady-state kinetic parameters for L-proline were obtained using the DCPIP assay with proline as the variable substrate in the range of 0.1-100 mM. Three trials were performed for each proline concentration. The parameters $K_m$ and $V_{max}$ were obtained by fitting the data to the Michaelis-Menten equation, using Origin software. Kinetic constants for an alternative substrate, 3,4-

dehydro-L-proline, were determined similarly, with the substrate concentration varied in the range 3-500 $\mu$M.  Inhibition by L-tetrahydro-2-furoic acid (THFA), L-lactic acid and L-mandelic acid were examined by steady-state inhibition kinetic measurements, using proline as the variable substrate.   Inhibition data were analyzed by the method of Dixon (36).

Thermostability was assessed by measuring PRODH activity as a function of incubation time at 90 °C.  For this study, the enzyme was incubated in a water bath at 90 °C, and aliquots were removed at various time points and stored at 4 °C.  After all of the aliquots were taken and cooled to 4 °C, activity assays were conducted at 25 °C in the presence of 25 mM proline.

**UV-visible Spectroscopy**

Potentiometric titrations of TtPRODH were recorded at 20 °C in 50 mM potassium phosphate buffer (pH 7.5) containing 50 mM NaCl and 5 % glycerol using a three - electrode single compartment spectroelectrochemical cell as previously described (14,37).   Measurements were made under a nitrogen atmosphere in a Belle Technology glove box. All potential values are reported versus the normal hydrogen electrode. Methyl viologen (20 mM) was used as a mediator dye, and pyocyanine (5 mM) and indigo disulfonate (3 mM) were used as indicator dyes. The UV-visible spectra in each experiment were recorded from 300-700 nm on a Cary 100 spectrophotometer. Corrected spectra of TtPRODH were obtained by subtracting the spectra of the dyes in the absence of protein measured under identical conditions. The reduction potential ($E_m$) and n values were calculated by linear regression of a Nernst plot of the data.

TtPRODH was titrated with proline under aerobic conditions in the proline concentration range 0 - 160 mM.  At each proline concentration, the absorbance spectrum was acquired after incubating the enzyme (56 mM) for 5 minutes with proline.   The absorbance values at $\lambda$ = 459 nm were analyzed as previously described (11,38) to obtain an apparent equilibrium constant for the formation of a reduced enzyme-P5C complex (eq. 1).

$$TtPRODH_{ox} + proline \leftrightarrow TtPRODH_{red}\text{-}P5C \quad (1)$$
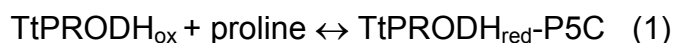
**Proline:$O_2$ Reactivity and Generation of Reactive Oxygen Species**

The rate of proline:$O_2$ activity was determined by monitoring the appearance of P5C as a function of proline concentration in the absence of an artificial electron acceptor.    P5C was detected as the yellow dihydroquinazolinium complex with *o*-aminobenzaldehyde (*o*-AB) by monitoring absorbance at $\lambda$ = 443 nm ($\varepsilon$ = 2900 M$^{-1}$cm$^{-1}$(39)).  The assay mixture included 50 mM potassium phosphate buffer pH=7.5, 20 $\mu$g of TtPRODH, 4 mM *o*-AB and 0.5 - 25 mM proline. The reaction was monitored for 15 minutes, and data from the 5-10 minute time period were used for rate calculations.   One unit of proline:$O_2$ activity is defined as the amount of TtPRODH that generates 1 $\mu$mol of P5C per minute at 25 °C.

Production of superoxide by TtPRODH was studied by measuring reduction of cytochrome c as described previously (40).    Reduction of cytochrome c was indicated by an increase in absorbance at $\lambda$ = 550 nm.  For

these assays, the L-proline concentration was 50 mM and the concentrations of cytochrome c and TtPRODH were 0.02 mg/mL each. Production of $H_2O_2$ by TtPRODH (0.02 mg/mL) was measured using the Amplex Red $H_2O_2$/peroxidase assay kit (Molecular Probes – Invitrogen) and a standard curve obtained from solutions of known $H_2O_2$ concentration.

## Bioinformatics Analysis of the Bacterial PutA/PRODH Family

Multiple sequence alignment calculations of PutA PRODH domains and bacterial monofunctional PRODHs were performed with ClustalW (41). The data set of sequences used for these calculations was obtained from the Protein Information Resource (PIR), using the Related Sequences tool with TtPRODH as the query sequence (see APPENDIX 2.2 for sequences). A total of 287 homologs of TtPRODH were identified.

## Tartrate Crystal Form

Another crystal form of TtPRODH was discovered when using 2 M sodium potassium tartrate, 100 mM Tris pH = 8.5, 0.5 % PEG 5000 MME, 4 % acetone (or 3% Isopropanol). These crystals grew as thin ovals with maximum dimension of 0.4mm. The space group was $P2_12_12_1$ and unit cell dimensions were a = 81.4 Å, b = 88.5 Å, c = 91.3 Å. Note that the cell dimensions are very similar to those of the MPD crystal form. These crystals diffracted to 2.3 Å and were reducible with proline.

## TtPRODH Mutagenesis

Site-directed mutagenesis to create the TtPRODH double mutation R202A/E207A was performed with Stratagene Site-directed Mutagenesis Kit. The following primers were used.

Forward:  5'-CCA GAC AAG GCC CTC ATT GAC GCC GCG TAC CTG-3'

Reverse:  5'-GGT CCA GGT ACG CGG CGT CAA TGA GGG CCT TG-3'

Mutations were confirmed with sequencing performed by the University of Missouri DNA core.

Crystals of TtPRODH R202A/E207A were obtained in Crystal Screen 27 and Wizard II 5.  Both conditions contained 0.1 M HEPES pH 7.5 an alcohol (20 %1.4-butanediol or 20 % 2-propanol), and a salt (NaCl (0.2 M) or sodium citrate tribasic dihydrate (0.2 M)).  These crystals exhibited the characteristic yellow color reminiscent of wildtype TtPRODH, but they diffracted sufficiently well to obtain the unit cell dimensions.  Those dimensions were similar to those of the MPD crystal form of native TtPRODH.

## RESULTS

### Structure Determination

TtPRODH contained 5 methionines that could be substituted with selenomethionine (Se-Met) for structure determination.  This corresponds to 1.5 % Met.  Note that the average occurrence of Met in proteins is 1.7 %(26).  To utilize the anomalous selenium (Se) signal for structure determination, it must be strong enough to detect (42).  This signal size decreases exponentially with protein molecular weight.  Utilizing the following equation, one can determine the

estimated Bijvoet diffraction ratio (signal size) based on the number of anomalous scatterers in a protein:

$$\text{rms}(\Delta F_{+/-h}) \,/\, \text{rms}|F| = (N_A/2N_T)^{1/2} \, (2f''_A/Z_{eff})$$

where $N_A$ is the number of anomalous scatters (5 for TtPRODH), $N_T$ is the total number of non hydrogen atoms in the molecule (327 for TtPRODH) and $Z_{eff}$ is the effective normal scattering at zero scattering angle (6.7 electrons for non-hydrogen atoms). Using a $f''$ of 4 electrons, the estimated Bijvoet diffraction ratio for TtPRODH is 0.104. This ratio is well above the expected level of signal utilizing the K-edge absorption of selenium.

Several Se-Met SAD data sets were collected but only one readily, lead to automated structure determination using the SOLVE/RESOLVE programs (TABLE 2.3). Although all the data sets had good completeness, redundancy and resolution, the one that led to automated structure determination was distinguished by superior $I/\sigma$ and $R_{merge}$ (see X3 in TABLE 2.3)

**Overall Fold**

The TtPRODH structure was determined at 2.0 Å resolution using Se-Met SAD phasing (TABLE 2.1). There are two protein chains in the asymmetric unit, labeled A and B, and 1 FAD cofactor bound to each protein (FIGURE 2.6). The structure reveals a distorted $(\beta\alpha)_8$ barrel fold. The first strand of the barrel ($\beta$1) is preceded by 4 $\alpha$-helices denoted $\alpha$A, $\alpha$B, $\alpha$C and $\alpha$0 (FIGURE 2.6, right side provides best view). The major distortion of the TtPRODH barrel from the classic

triosephosphate isomerase (TIM) barrel concerns the placement of $\alpha$8. Helix $\alpha$0 packs against $\beta$1 thus occupying the location reserved for $\alpha$8 in the classic TIM barrel. Consequently, helix $\alpha$8 of TtPRODH is located above the carboxyl terminal face of the $\beta$-barrel rather than alongside the $\beta$-barrel as in the classic TIM structure.

The FAD cofactor of TtPRODH is bound at the carboxyl ends of the strands of the barrel (FIGURE 2.6). The *re* face of the isoalloxazine packs tightly against strands 4-6, while the *si* face opens to the substrate binding pocket and is available for hydride transfer from the substrate proline.

Helices A-C and 8 form a small domain above the carboxyl terminal face of the $\beta$-barrel (FIGURE 2.6, right side provides best view). Helices A and B form a right angle, as do helices B and C. Helix 8 fits into the cleft formed by helices A-C. These four helices pack together to form a solvent-exposed hydrophobic patch (FIGURE 2.7; FIGURE 2.8). The patch has a pronounced hole in the middle (FIGURE 2.8). Several hydrophobic side chains form the rim of the hole and line its sides, while the hydroxyl groups of Thr287 and Tyr283 form the bottom of the hole (FIGURE 2.7; FIGURE 2.8). If the Rosetta Stone hypothesis is true for TtPRODH and TtP5CDH, then the function of the hydrophobic patch may be to serve as a docking interface for P5CDH. The patch presents a concave hydrophobic surface to solvent (FIGURE 2.8), and thus it is tempting to speculate that this patch mates with a complementary convex surface of P5CDH. Involvement of this patch in interaction with P5CDH also makes sense for substrate channeling because the patch is close to the PRODH active site.

The pair of molecules chosen for the asymmetric unit corresponds to the largest interface in the crystal lattice, which buries 992 $\text{Å}^2$ of surface area. The two molecules interact primarily through the packing of $\alpha 5$ (residues 200-218) of one chain against $\alpha 5$ and $\alpha 6$ (residues 228-242) of the other chain (FIGURE 2.6). The orientation of the two molecules is such that the $\alpha 5$ helices form an angle of 50°. The interface has 4 intermolecular ion pairs (Lys213-Glu207, Arg230-Asp200) and 4 intermolecular hydrogen bonds (His210-His210, Tyr238-Glu170). Several hydrophobic side chains are also buried in the interface: Phe198, Leu203, Leu209, Leu214.

The significance of this interface is unclear at this time. It is not the one typically found in dimeric TIM, nor is it similar to the tetramer interface of archaeal TIMs (43). The classic TIM dimer interface involves loops 1-4 (43) and buries about 1500 $\text{Å}^2$ of surface area, which is 50 % larger than the TtPRODH interface. Analysis of detergent-solublized TtPRODH by dynamic light scattering and gel filtration suggests the presence of apparent monomeric and dimeric species (FIGURES 2.9; FIGURE 2.10; TABLE 2.4). It is possible that the pair of proteins in the asymmetric unit represents the dimeric species observed in solution. However, it is also possible that this interface is an artifact of crystallization. An additional complicating factor in assessing the biological relevance of this interface is the possibility that TtPRODH interacts with TtP5CDH, as predicted by the Rosetta Stone hypothesis.

The TtPRODH structure is only the second PRODH structure solved to date, with the first being the PRODH domain of *E. coli* PutA (PutA86-669, PDB

code 1TIW (4,44)).  Residues 263 - 561 of PutA form a distorted TIM barrel that is similar to that of TtPRODH (FIGURE 2.11).  Within this catalytic core, the sequence identity between the two enzymes is 29 %, and the root-mean-square deviation for $C_a$ atoms is 1.7 Å for 243 aligned residues.  Also, the general location of the FAD at the carboxyl-terminal tips of the strands is similar in the two structures.  The major topological difference between the TtPRODH and the PutA86-669 barrels is an extra $\alpha$-helix that appears in PutA86-669 but is absent in TtPRODH.  Residues 437-449 of PutA form an $\alpha$-helix ($\alpha$5a in (FIGURE 2.11) that is replaced by a 5-residue loop in TtPRODH (residues 190-195).  As discussed in the next section, this topological difference has a dramatic effect on FAD conformation.

**FAD Conformation**

More than dozen residues interact with the FAD either through electrostatic or nonpolar interactions.  Electrostatic interactions with the FAD are shown in (FIGURE 2.12; FIGURE 2.13).  The interactions with the isoalloxazine in TtPRODH are also present in PutA86-669 structures (FIGURE 2.13).  Accordingly, the positioning of the isoalloxazine within the barrel is similar in TtPRODH, and PutA86-669 (FIGURE 2.12).  Surprisingly, the FAD conformations of the two enzymes are dramatically different (FIGURE 2.14; FIGURE 2.15).  In TtPRODH, the adenosine is under the dimethyl benzene ring of the isoalloxazine (FIGURE 2.14).  By contrast, in PutA86-669, the adenosine is below the pyrimidine ring of the isoalloxazine (FIGURE 2.15).  The difference

is dramatic.  When superimposed, the adenine groups of the two cofactors are separated by 13 Å (FIGURE 2.15).

Accordingly, the interactions involving the ribityl, pyrophosphate and adenosine groups are different in the two enzymes.  For example, in TtPRODH the ribityl 2'-OH is tucked under the pyrimidine ring of the isoalloxazine so that it forms hydrogen bonds with N1 of FAD (2.7 Å, (FIGURE 2.15) and the amine of Gly188 (FIGURE 2.12).  Gly188 is located on the loop between β5 and α5, which passes below the pyrimidine ring (FIGURE 2.12).  The 3'-OH of TtPRODH is also directed toward this loop, but does not form hydrogen bonds with the loop.   The 4'-OH is approximately trans to 3'-OH, and it forms a hydrogen bond to the pyrophosphate (3.0 Å, FIGURE 2.15).  In contrast, in PutA86-669 the 2'-OH and 3'-OH groups are rotated approximately 90° from the corresponding groups of TtPRODH (FIGURE 2.15).  Moreover, the 4'-OH of PutA86-669 is rotated 180° from that of TtPRODH and forms an intramolecular hydrogen bond with a ribose hydroxyl (FIGURE 2.15).

As noted above, the adenine bases of TtPRODH and PutA86-669 are separated by 13 Å.  This is due to different dihedral angle rotations of the pyrophosphate (FIGURE 2.15).  In TtPRODH, the adenine sits atop the N-terminal ends of α6 and α7 (FIGURE 2.6), right side provides best view).  In contrast, the adenine of PutA86-669 packs against α5a and forms an intimate stacking interaction with Trp438 (FIGURE 2.11).

Three factors contribute to the strikingly different FAD conformations in TtPRODH and PutA.  First, Asp228 of TtPRODH replaces Asn488 of PutA.

Although this seems like a conservative change, the structural ramifications are significant. Asn488 forms a hydrogen bond to the FAD pyrophosphate in PutA (FIGURE 2.16). Changing this residue to Asp not only eliminates the hydrogen bond but also creates electrostatic repulsion between the Asp228 carboxylate and the pyrophosphate. Apparently, this electrostatic clash is avoided in TtPRODH by rotation of the pyrophosphate dihedral angle so that the adenosine half of the pyrophosphate faces away from Asp288 (FIGURE 2.16). Note that Asp288 forms an ion pair with Lys187, which, in turn, forms an ion pair with the pyrophosphate (FIGURE 2.16). It is concluded that Asp288 helps set up the FAD conformation in TtPRODH by electrostatic repulsion with the pyrophosphate and electrostatic attraction to Lys187.

The other two factors contributing to the different FAD conformations are two helices that are present in PutA but absent in TtPRODH. As noted in the previous section, PutA $\alpha$5a is replaced by a short loop in TtPRODH (FIGURE 2.16). Thus, TtPRODH does not have a residue equivalent to PutA Trp438 for stacking against the adenine (FIGURE 2.16). In addition, PutA86-669 has extra secondary structural elements following $\alpha$8, owing to its longer polypeptide length. These extra elements are presumably important for linking the PRODH and P5CDH domains of PutA. One of these elements, a helix formed by PutA residues 565-570, is found to clash with the adenosine of TtPRODH when the two structures are superimposed (clash distances < 1 Å, (FIGURE 2.11). Thus, steric considerations prevent PutA86-669 from accommodating the FAD in the conformation observed in TtPRODH.

The aforementioned analysis compared conformations of oxidized cofactors from TtPRODH and PutA86-669. We recently reported the structure of dithionite-reduced PutA86-669 and showed that reduction of the FAD induces a 22° bend of the isoalloxazine and rotation of the 2'-OH by 90° so that it is tucked under the pyrimidine ring and forms a hydrogen bond to the FAD N1 (45). Thus, the conformations of the 2'-OH groups of oxidized TtPRODH and reduced PutA86-669 are identical (FIGURE 2.15). Electron density maps show that the isoalloxazine ring in TtPRODH is planar (FIGURE 2.14, right), as is the case for oxidized PutA86-669(4,44), so it is unlikely that TtPRODH was reduced by exposure to X-rays during data collection. We conclude that the TtPRODH structure presented here represents the conformation of the oxidized enzyme. Thus, in terms of the 2'-OH conformation, the oxidized cofactor of TtPRODH resembles the reduced cofactor of PutA86-669 (FIGURE 2.15). This result has implications for understanding the differences in membrane association of *E. coli* PutA and TtPRODH (see Discussion).

**Ligand-Free, Solvent-Exposed Active Site**

It is significant that the TtPRODH active site does not contain a bound proline analog because all PutA86-669 crystal structures solved to date have active site ligands bound, including acetate (PDB code 1TJ2), L-lactate (PDB code 1TJ0), THFA (PDB code 1TIW) and hyposulfite (PDB code 2FZM). Thus, the TtPRODH structure provides the first view of a ligand-free PRODH active site. Since the amino acid sequences of the proline binding pockets are identical in TtPRODH and *E. coli* PutA, comparison of ligand-free TtPRODH and inhibitor-

bound PutA86-669 possibly provides insights into conformational changes induced by the binding of proline. The following analysis refers to molecule A of the asymmetric unit but it also holds true for molecule B.

There are two major differences between TtPRODH and PutA86-669 in the region of the proline-binding pocket. First, $\alpha 8$ is shifted 3-4 Å away from the proline-binding pocket in TtPRODH (FIGURE 2.11; FIGURE 2.17). The consequences of this conformational difference are significant because $\alpha 8$ contributes two absolutely conserved Arg residues that participate directly in binding substrate. As shown in FIGURE 2.17, the two conserved Arg side chains of PutA86-669 (Arg555, Arg556) form ion pairs with the carboxyl group of THFA. Movement of $\alpha 8$ in TtPRODH pulls the analogous residues (Arg288, Arg289) out of the proline-binding pocket (FIGURE 2.17).

The second major difference between the proline binding pockets of TtPRODH and PutA86-669 involves a conserved ion pair that is present in PutA86-669 but absent in TtPRODH. As shown in FIGURE 2.17, Arg555 of PutA86-669 forms an ion pair with conserved Glu289. The analogous ion pair of TtPRODH (Arg288-Glu65) is not observed because Glu65 is flipped out of the active site and points into the solvent. This difference is quite large, with the carboxyl groups of TtPRODH Glu65 and PutA Glu289 separated by 10 Å. We note that Glu65 is free of crystal contacts in both molecules, thus the observed conformation of this residue is not likely an artifact of crystal packing.

Because of the conformational differences involving $\alpha 8$ and Glu65 (Glu289), the active site of TtPRODH is open and the FAD is solvent exposed,

sharply contrasting with the closed active site of PutA86-669/THFA, in which the FAD is buried (FIGURE 2.19). The solvent-accessible surface area (SASA) of the FAD in TtPRODH is 220 $Å^2$ for the A chain and 275 $Å^2$ for the B chain, based on analysis with the Ligand Protein Contacts server(46). For reference, the SASA of the isolated FAD is 963 $Å^2$. By contrast, the FAD in PutA86-669/THFA is buried, with SASA of only 25 $Å^2$. The THFA is also buried (0.0 $Å^2$ SASA). Exposure of the FAD in TtPRODH can be appreciated by viewing a space-filling model of the enzyme (FIGURE 2.18; FIGURE 2.19). It is evident from this view that almost the entire surface of the isoalloxazine *si* face is solvent exposed in TtPRODH. In particular, note that the hydride transfer acceptor atom of the flavin, N5, is open to solvent in TtPRODH. This atom is completely buried in PutA86-669/inhibitor complexes (FIGURE 2.19).

**MPD and Possible Electron Transfer Pathway**

There are two MPD molecules bound to each TtPRODH, and the binding sites are in identical locations in the two proteins. Both sites are located on bottom face of the barrel, which is the face opposite from the FAD-binding site. Sites 1 and 3 are located in the N-terminal ends of the β-strands, while sites 2 and 4 have different interactions due to crystal packing (FIGURE 2.20)

MPD sites 1 & 3 (FIGURE 2.21) are equivalent and are located in the center of the bottom face of the β-barrel among the loops that connect strands with helices. The MPD molecule binds in a hole formed by Tyr95, Phe129, Tyr221, Asn182, Arg273 and Arg131. Tyr95 and Arg171 form the bottom of the hole , and the other side chains form the sides of the hole. The MPD in site 1

forms nonpolar contacts primarily with Phe129 and Tyr221 and hydrogen bonds with Asn182.  The average B-factor for MPD in site 1 & 3 are 28 and 26 $\text{Å}^2$, respectively.

MPD site 2 is bound near on the periphery of the barrel near $\alpha 4$ and $\alpha 5$ in chain B. This ligand is interacting in a crystal contact.  This MPD molecule contacts side chains of Arg151, Glu152, Tyr179, and Pro178.  It forms hydrogen bonds with water and Gly30 of a symmetry-related molecule.  MPD site 2 has nonpolar interactions with Glu152 and Tyr 179.  The average B-factor for MPD in site 2 is 17.5 $\text{Å}^2$

MPD site 4 binds to chain A in the same location as MPD site 2 in chain B. Interestingly the crystal packing is different, and the B-factor of 40 $\text{Å}^2$ confirms that there are fewer interactions. There are the nonpolar interactions with Tyr 179, and MPD hydrogen bonds with water.

MPD has been shown to bind hydrophobic regions of proteins, reducing solvent-accessible areas, with implications in protein stability (47).  We hypothesize that the N-terminal ends of the $\beta$-strands associating with MPD sites 1 & 3 may be a potential conduit for transferring electrons from the reduced flavin to electron acceptors in the membrane.  The distance between MPD 1 and the flavin N5 is 12.5 Å.  Arg184, which hydrogen bonds to the isoalloxazine N5, is hypothesized to play a role in this electron transfer.  In inactivated TtPRODH (CHAPTER 3), Arg184 shows a movement of 0.5 Å upon flavin reduction. Dithionite-reduced EcPutA PRODH structures also show a movement of the homologous Arg431, which no longer forms a hydrogen bond with N5.  With

R184 movement, there are charged residues that could form an electron transfer pathway, with conservation of these residues among most bacteria.

The proposed electron transfer pathway is Arg184, Gln252, Glu250 and Arg273 (FIGURE 2.22). Among the proposed residues in the electron transfer pathway, Arg184 is completely conserved among bacteria. Gln252 and Glu250 are present in most all bacteria, excepting of those contained in branch 2a (*Helicobacter, Campylobacter, Corynebacterium* and *Bacteroides*). Arg273 is highly conserved, although this residue has more variation than previous residues in the pathway. Besides the branch 2a bacteria, Arg273 is also not present in *Oceanobacillus* (Cys), *Staphylococcus* (Thr), *Halobacterium* (Phe), and *Jannaschia* (Gln). Two PutA enzymes from branch 2a member, *Helicobacter sp.*, have been characterized and generate reactive oxygen species. Thus branch 2a PutAs may not have a need for an electron transport pathway (48).

Membrane association by SPR indicates that EcPutA prefers to bind neutral membranes (13). Perhaps MPD, which has been shown to bind hydrophobic regions, is binding to a potential location of membrane association on TtPRODH (FIGURE 2.22). Goals include the crystallization of TtPRODH in complex with quinone-like molecules, in order to gain insight into this association.

**Crystal Soaking Experiments**

Addition of L-proline to a solution of TtPRODH causes a change from bright yellow to colorless indicating flavin reduction. Optical spectroscopy confirm that the characteristic flavin spectra disappears upon reduction with L-

proline (FIGURE 2.23). Interestingly, addition of solid proline to the MPD crystal form did not produce a corresponding color change. DCPIP assays were conducted in the prescence of similar concentrations of MPD, MgCl$_2$ and imidizole produced no change in TtPRODH activity. It was thought that crystal packing interactions prevent closure of the active site, which is required for binding proline.

Other crystal forms were optimized, and mutagenesis was used to disrupt crystal packing (TABLE 2.2). The high-salt crystal form that could be reduced by proline (FIGURE 2.24; FIGURE 2.25). This crystal form had unit cell dimensions similar to those of the MPD form. Despite soaking with high concentrations of the isostructural analogs, L-THFA, and substrate, L-proline, the structure revealed in the same non-occupied, solvent-exposed active site. Mutagenesis to disrupt crystal packing between helices α5 was also attempted (FIGURE 2.11). Double mutant Arg202Ala / Glu207Ala was created, purified and crystals optimized. Although the crystallization conditions were different and crystals were reducible with proline, the unit cell dimensions were similar to those of the MPD form. This mutagenized TtPRODH crystal form was not pursued further.

**Absorbance Spectroscopy and Steady-State Kinetics**

We performed biochemical analyses of TtPRODH to understand how monofunctional PRODHs differ from PutA PRODH domains in terms of spectroscopic and steady-state kinetic properties. TtPRODH displayed a flavin absorption spectrum similar to that of previously characterized PutAs (49), with maxima at 381 nm and 452 nm (FIGURE 2.26). Potentiometric titration of

TtPRODH yielded an $E_m$ value of -75 mV for the bound FAD cofactor (FIGURE 2.26), which is similar to that previously reported for *E. coli* PutA ($E_m$ = -77 mV, pH 7.5,(50)). No significant stabilization of semiquinone species was observed during the titration (FIGURE 2.26). Titration of TtPRODH with proline under aerobic conditions yielded an apparent equilibrium constant for the formation of TtPRODH$_{red}$-P5C (eq. 1) of 4.5 mM$^{-1}$ proline, which is about two-fold lower than the value of 9.5 mM$^{-1}$ for *E. coli* PutA (38).

The kinetic parameters of TtPRODH using proline as the substrate were estimated to be $K_m$ = 27 mM, $V_{max}$ = 20.5 U/mg and $k_{cat}$ = 13 s$^{-1}$ (FIGURE 2.27)(TABLE 2.5). For reference, the corresponding values for PutAs from *E. coli* and *Helicobacter pylori* are listed in TABLE 2.5. Although, the $K_m$ parameter of TtPRODH is 3 - 5 times lower than those of the PutAs, the $k_{cat}$ value is comparable to the PutA values (TABLE 2.5). The kinetic parameters for TtPRODH more closely resemble those of PutA86-669 ($K_m$ = 60 mM, $k_{cat}$ = 17 s$^{-1}$) than those of the PutAs.

Interestingly, 3,4-dehydro-L-proline was more efficiently oxidized than proline, with $K_m$ = 4 mM, $V_{max}$ = 119 U/mg and $k_{cat}$ = 75 s$^{-1}$ (FIGURE 2.28) Analogous data for PutA are not available, although Wood reported, that in *E. coli* K12, PutA detoxifies 3,4-dehydro-L-proline by oxidation(51).

THFA and L-mandelic acid were identified as competitive inhibitors of TtPRODH, with $K_I$ values of 1.0 mM and 2.4 mM, respectively (FIGURE 2.29; FIGURE 2.30). THFA is also a competitive inhibitor of PutA86-669, *E. coli* PutA and *H. pylori* PutA ($K_I$ = 0.2 - 0.3 mM). Other proline analogs such as trans-4-

46

hydroxy-L-proline, cis-4-hydroxy-L-proline, L-azetidine-2-carboxylic acid and L-pipecolinic acid were neither inhibitors nor substrates of TtPRODH.

TtPRODH is highly thermostable, based on residual activity measurements, as expected for an enzyme from an extreme thermophile. TtPRODH exhibited over 85 % residual activity, after a 1-hour incubation at 90 °C, and over 60 % residual activity after 3 hours (FIGURE 2.31). The half-life estimated from these data was 257 +/- 6 min. For comparison, PutA86-669 exhibits a 50 % drop in activity after 2 hours at 45°C (44). The proline dehydrogenase activity of TtPRODH increased with increasing temperature, showing no maximum with current experimental capabilities (FIGURE 2.32).

**Proline:$O_2$ Reactivity and Generation of Reactive Oxygen Species**

The solvent-exposed active site of TtPRODH suggested the possibility that the reduced enzyme would exhibit higher reactivity with molecular oxygen than PutA from *E. coli*. A chromogenic assay based on *o*-AB was used to measure the rate of oxygen reactivity. In this assay, P5C produced by oxidation of proline forms a complex with *o*-AB, which is monitored by absorbance at 443 nm. Detection of the *o*-AB:P5C complex over time indicates reactivity of the reduced FAD with $O_2$ during catalytic turnover with proline.

TtPRODH exhibited significant proline:$O_2$ reactivity with kinetic parameters of $K_m$ = 1.3 mM, $V_{max}$ = 335 mU/mg, and $k_{cat}$ = 12.7 min$^{-1}$ (TABLE 2.5) (FIGURE 2.33). Data from the *o*-AB assay are typically expressed as the ratio of the specific activity from the DCPIP assay to the specific activity from the *o*-AB assay, which indicates the preference of utilizing DCPIP over $O_2$ as the electron

acceptor. For TtPRODH, this ratio is 61. For comparison, the corresponding ratios for PutAs from *Helicobacter pylori* and *E. coli* are 16 and > 2500, respectively (TABLE 2.5). This reactivity was not increased upon addition of *E. coli* vesicles containing electron acceptors in the membranes (FIGURE 2.34).

Generation of ROS ($O_2^-$, $H_2O_2$) by TtPRODH was also studied. The production of superoxide was examined by monitoring the reduction of cytochrome c during the proline oxidation catalytic cycle. Reduction of cytochrome c was observed, which is consistent with generation of superoxide. Addition of superoxide dismutase to the assay (30 mg) eliminated the observed effect, which further implicates superoxide as the product of the proline:$O_2$ reaction. Since superoxide decomposes in water to $H_2O_2$, the generation of $H_2O_2$ was also examined. $H_2O_2$ was produced in a proline-dependent manner (FIGURE 2.35).

**Conserved Sequence-Structure Motifs of the PRODH Family**

An intriguing aspect of proline catabolism is that PRODH and P5CDH are encoded on separate genes in some organisms, whereas the two genes are fused in other organisms (*putA*). The traditional view, which was developed prior to the whole genome sequence era, has been that PRODH and P5CDH appear as separate enzymes in eukaryotes and as fused bifunctional enzymes (PutA) in bacteria. Our analysis of genome sequence data, however, reveals a more complex situation for bacteria. The updated view is that PutAs are indeed restricted to bacteria, but monofunctional PRODHs and P5CDHs appear in both eukaryotes and bacteria.

The distribution of PutAs and monofunctional PRODHs in bacteria is depicted in the phylogenetic tree in FIGURE 2.36. Three main branches are evident. Branch 1 contains the best-characterized PutAs, including PutAs from *E. coli, Bradyrhizobium japonicum* and *S. typhimurium*. The organisms represented in branch 1 are primarily α-, β-, and γ-proteobacteria. The PutAs in branch 1 of our data set have 999 - 1361 residues and the pairwise sequence identities are 38 - 99% with an average of 49 %.

Branch 2 contains PutAs from Gram-negative cyanobacteria, δ- and ε-proteobacteria, and corynebacterium. The polypeptide length for branch 2 PutAs is 982-1294 and the pairwise sequence identity range is 23 - 73 % with an average identity of 38 %. Note that branch 2 is divided into two distinct groups, denoted 2A and 2B in FIGURE 2.36. PutAs from branch 2 have only recently been explored. For example, Krishnan and Becker showed that branch 2A PutAs from *Helicobacter pylori* and *Helicobacter hepaticus* appear to be unique among PutAs in that they exhibit oxygen reactivity and generate proline-dependent ROS (48).

Branch 3 consists entirely of monofunctional PRODHs. TtPRODH is the only enzyme of this branch to be purified and characterized. PRODHs of branch 3 have 279-333 amino acid residues, and the pairwise sequence identities for this group are 23 - 79 % with an average of 38 %. In some branch 3 organisms, the PRODH and P5CDH genes are very close together. For example, in *T. thermophilus,* only 15 bases separate the stop codon of the PRODH gene from the Met start codon of the P5CDH gene. In other organisms, the two genes are

quite far apart. For example, in *Staphylococcus aureus* subsp. aureus Mu50, the PRODH and P5CDH genes are separated by 800 kb. Interestingly, most of the organisms represented in branch 3 are Gram-positive bacteria. Counterexamples include *Thermus, Solibacter, Salinibacter*, and *Chlorobium*.

The availability of two PRODH structures (TtPRODH and PutA86-669) and many sequenced bacterial genomes allowed analysis of conserved sequence-structure motifs of the bacterial PutA/PRODH family. Nine conserved motifs were identified by mapping multiple sequence alignments onto the TtPRODH and PutA86-669 structures (TABLE 2.6). Each motif contains at least one residue that is identically conserved throughout the entire bacterial PutA/PRODH family (TABLE 2.6, bold letters) (FIGURE 2.37).

The TtPRODH and PutA86-669 structures show that the nine motifs cluster near the active site and that the identically conserved residues of the motifs have important roles in FAD binding and substrate recognition. Motifs 4-6 are primarily involved in FAD binding. For example, the conserved Gln of motif 4 forms a hydrogen bond to the FAD O2 (FIGURE 2.12, Gln163) while the Arg of motif 5 forms a hydrogen bond to the FAD N5 (FIGURE 2.12, Arg184). Also, the Lys and His of motifs 5 and 6 interact with the pyrophosphate (FIGURE 2.12, Lys187 and His227).

Motifs 1-3 and 7-9 are responsible for substrate recognition. For example, the Lys and Arg side chains of motifs 2 and 9 form ion pairs to the substrate carboxyl, as shown for the PutA86-669/THFA structure (FIGURE 2.17, PutA residues Lys329, Arg555, Arg556). The conserved Glu residues of motifs 1 and 9

play indirect roles in substrate recognition by providing stabilizing ion pairs to the two Arg residues of motif 9 when the substrate/product is bound (Arg555-Glu289 and Arg556-Glu559 in PutA, FIGURE 2.17). Size and shape complementarity are enforced by nonpolar contacts between the substrate and the Leu and Tyr side chains of motifs 7 – 9 (44). Finally, the conserved Asp of motif 3 (Asp133 of TtPRODH) plays a dual role in substrate recognition and FAD binding by forming stabilizing interactions with the Lys of motif 2 (substrate recognition, FIGURE 2.17), Tyr of motif 8 (substrate recognition, FIGURE 2.17) and the conserved Arg of motif 5 (FAD binding, FIGURE 2.17).

## DISCUSSION

### New Subfamily of PRODH

The genesis of this work was the realization, based on analysis of genome sequence data, that some bacteria lack putA genes and instead encode PRODH and P5CDH as separate monofunctional enzymes. This observation is significant because the traditional view of proline catabolism was that monofunctional enzymes are restricted to eukaryotes. Thus, bacterial monofunctional PRODHs represent a new subfamily of proline catabolic enzyme. Moreover, isolation of recombinant eukaryotic PRODHs in sufficient quantity and purity for biophysical study has been problematic (our unpublished results) and therefore the bacterial homologs are potentially attractive model systems for understanding human PRODH. We thus set out to characterize a bacterial monofunctional PRODH in order to establish paradigms for this new subfamily,

compare its structure and biochemical properties to those of PutAs, and set the stage for probing protein-protein interactions between monofunctional PRODH and P5CDH.

One major result of this work is that monofunctional PRODHs and PutAs share a common catalytic core consisting of a unique TIM barrel (FIGURE 2.11). The PutA/PRODH barrel is distinguished from other TIM barrels by placement of $\alpha8$ above the barrel. This distortion is functionally significant because $\alpha8$ contributes conserved motif 9 to the active site (TABLE 2.6). Observation of this structural distortion of the classic TIM barrel fold in enzymes from two branches of the bacterial PutA/PRODH family suggests that it is a defining structural signature of this family.

Interestingly, the TIM barrel catalytic core, FAD conformations, and proline binding pocket described here for the PutA/PRODH family bear no resemblance to those of PRODHs from the hyperthermophilic archaeon *Pyrococcus horikoshii* (PDH1 and PDH2) (FIGURE 2.38). For example, PDH1 from *P. horikoshii* is an $(\alpha\beta)_4$ hetero-octameric complex with the $\beta$-subunit binding an FAD cofactor and exhibiting PRODH activity(6). The FAD-binding domain of PDH1 consists of a Rossmann dinucleotide-binding fold similar to that of monomeric sarcosine oxidase, which is a member of the glutathione reductase family (6). As expected for a Rossmann fold protein, the FAD of PDH1 is highly extended, and the pyrophosphate interacts with a glycine-rich loop and associated conserved water molecule (52). Thus, the protein-FAD interactions in PDH1 are quite different from those described here. Furthermore, proline binds on the *re* face of the FAD

in PDH1, in contrast to PutA and TtPRODH which bind proline on the *si* face. Thus, the PutA/PRODH and PDH1 families represent two distinctly different solutions to the problem of catalyzing the oxidation of proline by a flavoenzyme.

**FAD Conformation**

Although the catalytic cores of TtPRODH and PutA86-669 are similar in overall fold, the FAD conformations are surprisingly different. This difference is attributed to an Asp/Asn sequence difference (FIGURE 2.16) and two helices present in PutA86-669 but absent in TtPRODH (FIGURE 2.11; FIGURE 2.16). Based on amino acid sequence alignments, the Asn-pyrophosphate hydrogen bond, $\alpha$5a, and the 560s helix are present in all PutAs and missing in all bacterial monofunctional enzymes. Interestingly, the stacking Trp of $\alpha$5a is present in most, but not all, PutAs. For example, branch 2A enzymes have Met or Leu in place of the stacking Trp, but these residues could also provide a nonpolar interaction with the FAD adenine. We therefore predict that all PutAs have the FAD conformation observed in PutA86-669 and that all bacterial monofunctional PRODHs have the FAD conformation observed in TtPRODH. It is concluded that the FAD conformation is the major structural difference between PutAs and bacterial monofunctional PRODHs.

The different FAD conformations in PutA and monofunctional PRODHs presumably reflect different structural and functional requirements. For example, we suggest that the difference in position of the ribityl 2'-OH reflects the different membrane association requirements of *E. coli* PutA and TtPRODH. *E. coli* PutA is membrane-associated only when the FAD is in the reduced state. In the

oxidized state, it remains in the cytoplasm and represses transcription of the putA and putP (encodes a proline transporter) genes (15,53,54). We recently showed that reduction of FAD in PutA86-669 triggers rupture of the hydrogen bond between the ribityl 2'-OH and Arg556, causing rotation of the 2'-OH so that it is tucked below, and hydrogen bonded to, the FAD N1 (45). We further showed that the 2'-OH-Arg556 hydrogen bond is a structural constraint that prevents oxidized PutA from binding the membrane(45). Interestingly, we find here that the 2'-OH of oxidized TtPRODH is tucked below the FAD N1, that is, locked in the membrane-binding position (FIGURE 2.14; FIGURE 2.15). This makes sense because TtPRODH does not have a repressor function and is presumably membrane-associated in both the oxidized and reduced states.

A larger question is how come the global FAD conformations of PutAs and monofunctional PRODHs differ so much, with the adenosine moieties of the two cofactors separated by 13 Å. It is possible that the particular FAD conformation found in PutA is necessary for coordinating the two catalytic functions of PutA, i.e., substrate channeling. Monofunctional PRODH and P5CDH may interact and exhibit intermolecular substrate channeling according to the Rosetta Stone hypothesis. We note that this occurs for tryptophan synthase, in which indole is channeled between separate $\alpha$ and $\beta$ subunits (16),(17). Presumably, the FAD conformation in TtPRODH is the one required for docking of TtPRODH with TtP5CDH. Observation of different FAD conformations in TtPRODH and PutA86-669 may indicate that the protein-protein interface between TtPRODH and TtP5CDH differs substantially from the PRODH:P5CDH domain interface in

PutA.  Clearly, structures of full-length PutAs, other monofunctional PRODHs, and a monofunctional PRODH:P5CDH complex would address how come PutAs and monofunctional PRODHs require such different FAD conformations.

**Insights into Protein Motions Associated with Substrate Binding**

Structural information on the ligand-free form of the PutA PRODH domain has been elusive due to the propensity of PutA86-669 to crystallize only in PEGs contaminated with either L-lactate or acetate (44), which are both competitive inhibitors.  These structures showed that the inhibitor is buried, which raised the question of how proline enters the active site and how P5C is released.  Clearly protein motion is required, but the nature of this motion is unknown.

TtPRODH crystallized in MPD, which allowed us to determine the first structure of a PRODH in the ligand-free state.  Although a structure of ligand-bound TtPRODH is not known, structural and biochemical data suggest that the active site of PutA86-669 complexed with THFA is a good model for the proline bound conformation of TtPRODH.  For example, all residues that contact THFA in PutA86-669 are identically conserved in TtPRODH, including Arg288, Arg289, Lys99, Tyr285, Leu254, Tyr275, Asp133 and Tyr190 (TtPRODH residue numbering).  Moreover, the steady state kinetics parameters for TtPRODH are similar to those of PutA86-669.  In particular, the $K_m$ for proline is 27 mM, which compares favorably with the value of 60 mM for PutA86-669.  Furthermore, THFA is a competitive inhibitor of TtPRODH with $K_I$ = 1 mM, which is similar to the value of $K_I$ = 0.2 mM for PutA86-669.  These data strongly suggest that TtPRODH and PutA86-669 bind the substrate proline similarly.   Thus,

comparison of ligand-free TtPRODH with PutA86-669/THFA provides clues about those parts of the TtPRODH active site move in response to substrate binding and product release.

Our analysis suggests that, prior to substrate binding, the enzyme adopts an open state in which α8 is shifted away from the FAD and the conserved ion-pair between the first Arg of motif 9 and Glu of motif 1 is broken (Arg289-Glu65). Upon binding substrate, α8 moves closer to the proline binding pocket to enable ion-pair formation between proline and the two Arg residues of motif 9, as well as formation of the ion-pair between motifs 1 and 9. The proline-binding pocket displays perfect charge balance when a carboxyl-containing molecule is bound in the active site (proline, P5C, THFA), (FIGURE 2.17). There are four Arg/Lys (motifs 2, 5, 9), two Glu (motifs 1, 9), one Asp (motif 3), and the substrate/inhibitor carboxyl in the pocket. In the absence of substrate/product, the binding pocket has one excess positive charge, necessitating movement to alleviate electrostatic repulsion. The TtPRODH structure suggests that the major conformational adjustment involves α8 and the Glu of motif 1.

Interestingly, there is a glycine residue in the loop connecting β8 and α8 (FIGURE 2.6, Gly279). This residue is identically conserved throughout the entire bacterial PutA/PRODH family except for branch 2A PutAs. In *E. coli* PutA, for example, the conserved glycine is Gly544. We suggest that this conserved Gly serves as a flexible hinge between β8 and α8, allowing α8 to shift in response to substrate binding.

**Relationships Between Bacterial and Human PRODH**

Humans have two isozymes of PRODH, which share about 50% amino acid sequence identity (55). PRODH1 is encoded on chromosome 19, is expressed almost exclusively in liver and kidney, and catalyzes oxidation of L-hydroxyproline. PRODH2 is encoded on chromosome 22q11, is expressed more widely than PRODH1 (brain, heart, pancreas, kidney, liver), and specifically oxidizes L-proline.

PRODH2 is part of the p53 signaling pathway with the PRODH2 gene identified as a p53 inducible gene (56). Up-regulation of PRODH2 and proline oxidation in lung, renal, and colon carcinoma cells has been shown to generate ROS and induce cell death by mitochondrial dependent processes (57-62). Perturbation of mitochondrial membranes by ROS causes the release of cytochrome c into the cytosol and subsequent activation of the intrinsic caspase pathway. Because the p53-apoptosis pathway involves ROS, PRODH2 appears to play a role in p53-mediated apoptosis by modulating the cellular redox environment. Indeed, antisense repression of PRODH2 prevents p53-induced apoptosis(62). Thus, PRODH2 is a pro-apoptotic protein that helps reduce carcinogenesis in humans by serving as a ROS generator. Accordingly, PRODH2 is often referred to as proline oxidase in the literature.

Bacterial monofunctional PRODHs are potentially attractive model systems for understanding the structure and biochemical function of human PRODHs. Although the sequence identity between TtPRODH and human PRODH1/2 is less than 20 %, conserved motifs 3-9 are clearly present in the sequences of both human PRODH1 and PRODH2. Considering residues within

10 Å of the active site, there is 45 % sequence conservation between human PRODH1/2 and TtPRODH. These data strongly suggest that the human enzymes have the $(\beta\alpha)_8$ catalytic core and active site structure common to TtPRODH and PutA86-669. On the other hand, human PRODH1 and PRODH2 have 536 and 600 residues, respectively, compared to only 307 for TtPRODH. Thus, the human enzymes clearly have additional structural elements not found in TtPRODH.

Interestingly, analysis of human PRODH1/2 amino acid sequences suggests that the conformation of FAD in human PRODH1/2 is similar to that of PutA rather than TtPRODH. Both human enzymes have Asn at the position equivalent to E. coli PutA Asn488. Furthermore, sequence alignments suggest that the human enzymes have the equivalent of PutA $\alpha$5a and that Leu replaces the stacking Trp. We note that Leu is also present at this position in some branch 2A PutAs. Thus, the FAD conformation observed in E. coli PutA is probably present in the human enzymes and the conformation observed in TtPRODH seems to be a unique signature of bacterial monofunctional enzymes. Production of proline-dependent superoxide is central to the role of PRODH2 in p53-mediated apoptosis (61,63). It is therefore highly significant that we observed reactivity of TtPRODH with molecular oxygen resulting in superoxide generation.

Reactivity of a FAD-dependent dehydrogenase with molecular oxygen implies that the reduced FAD is accessible to solvent. E. coli PutA is essentially nonreactive with molecular oxygen(48), suggesting that the FAD remains

sequestered even after the active site opens to release P5C. On the other hand, *H. pylori* PutA reacts strongly with molecular oxygen indicating that $O_2$ has access to the reduced cofactor (48). Krishnan and Becker have suggested that Asn291 of *H. pylori* PutA, which is replaced by the bulkier Tyr437 in *E. coli* PutA, may account for the differences in $O_2$ reactivity of these two PutAs (48). This particular Tyr is part of conserved motif 5 (TABLE 2.6, Tyr437 in *E. coli* PutA). It forms a water-mediated hydrogen bond to the O atom of the THFA ring (FIGURE 2.17) and is in position to protect the active site from bulk solvent by virtue of its location on the edge of the proline-binding site (see Tyr190 in FIGURE 2.18). Tyrosine, being larger than Asn, may afford more protection of the active site from solvent. Interestingly, human PRODH2 and TtPRODH - both ROS generators - have Tyr at this position (Tyr446 in PRODH2 and Tyr190 in TtPRODH, FIGURE 2.11). Thus, the Asn hypothesis, which may explain the difference in $O_2$ reactivity between *E. coli* and *H. pylori* PutAs, does not explain the $O_2$ reactivity of TtPRODH and human PRODH2. The TtPRODH structure shows that $\alpha 8$ is important for protecting the FAD from solvent and that there is sufficient flexibility in the $\beta 8$-$\alpha 8$ loop to allow movement of $\alpha 8$ away from the proline binding pocket resulting in exposure of the isoalloxazine. These results suggest that movement of $\alpha 8$ may contribute to $O_2$ reactivity of PRODHs.

**Obtaining the solvent-protected TtPRODH active site**

Crystal packing of TtPRODH in the $P2_12_12_1$ lattice must be highly favorable. Attempts to disrupt this crystal lattice were unsuccessful. Chapter 3 discusses results obtained with a mechanism-based inactivator of TtPRODH. It

was hypothesized that a covalently-attached inhibitor would induce TtPRODH to pack in a different crystal lattice. Unfortunately, no other crystal forms were discovered. However, the structure of the inactivated TtPRODH was solved using the MPD crystal form (see Chapter 3).

Strategies for obtaining a solvent-protected TtPRODH active site are important because it remains unresolved whether or not the accessibility of the TtPRODH active site results from crystal-packing. Perhaps lysine methylation would result in a different crystal form (64). This procedure methylates surface lysines turning primary amines into tertiary methylated amines which changes the surface potential. It is reported to be a "rescue strategy" for obtaining crystals when other methods are not fruitful. Another strategy would be relocating the His-tag to the C-terminal end of the protein. This strategy was attempted in Don Becker's lab, our collaborator. C-terminally his-tagged TtPRODH was reported precipitate after affinity chromatography (N. Krishnan, personal communication). A final strategy is homolog screening. For example, we cloned, expressed and purified *B. subtilis* PRODH. The enzyme was highly aggregated in solution which could be relieved by adding BOG, similar to TtPRODH. Unfortunately, no crystals were obtained with initial crystal screening. These screens were attempted again with the presence of 10 mM THFA, and again no crystals were obtained. There are plenty of other homologs to crystallize, although soluble aggregation may be a potential issue with bacterial PRODH, as it was seen in two of these enzymes already. Until a bacterial PRODH with a solvent-protected

active site is crystallized, questions about induced-fit substrate binding remain unanswered.

## REFERENCES

1.    Adams, E., and Frank, L. (1980) *Annu Rev Biochem* **49**, 1005-1061
2.    Phang, J. M. (1985) *Curr Top Cell Regul* **25**, 91-132
3.    Zhang, M., White, T. A., Schuermann, J., Baban, B. A., Becker, D. F., and Tanner, J. J. (2004) *Biochemistry* **43**
4.    Lee, Y. H., Nadaraia, S., Gu, D., Becker, D. F., and Tanner, J. J. (2003) *Nature Structural Biology* **10**, 109-114
5.    Inagaki, E., and Tahirov, T. H. (2004) *to be published*
6.    Tsuge, H., Kawakami, R., Sakuraba, H., Ago, H., Miyano, M., Aki, K., Katunuma, N., and Ohshima, T. (2005) *J Biol Chem* **280**, 31045-31049
7.    Menzel, R., and Roth, J. (1981) *J Biol Chem* **256**, 9762-9766
8.    Brown, E., and Wood, J. (1992) *Journal of Biological Chemistry* **267**, 13086-13092
9.    Muro-Pastor, A. M., and Maloy, S. (1995) *Journal of Biological Chemistry* **270**, 9819-9827
10.   Muro-Pastor, A. M., Ostrovsky, P., and Maloy, S. (1997) *J Bacteriol* **179**, 2788-2791
11.   Becker, D. F., and Thomas, E. A. (2001) *Biochemistry* **40**, 4714-4721
12.   Zhu, W., and Becker, D. F. (2003) *Biochemistry* **42**, 5469-5477
13.   Zhang, W., Zhou, Y., and Becker, D. F. (2004) *Biochemistry* **43**, 13165-13174
14.   Krishnan, N., and Becker, D. F. (2005) *Biochemistry* **44**, 9130-9139
15.   Larson, J. D., Jenkins, J. L., Schuermann, J. P., Zhou, Y., Becker, D. F., and Tanner, J. J. (2006) *Protein Sci* **15**, 2630-2641
16.   Miles, E. W., Rhee, S., and Davies, D. R. (1999) *J Biol Chem* **274**, 12193-12196
17.   Huang, X., Holden, H. M., and Raushel, F. M. (2001) *Annu Rev Biochem* **70**, 149-180
18.   Surber, M. W., and Maloy, S. (1998) *Arch Biochem Biophys* **354**, 281-287
19.   Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999) *Science* **285**, 751-753
20.   White, T. A., and Tanner, J. J. (2005) *Acta Crystallography* **F61**, 737-739
21.   Ailenberg, M., and Silverman, M. (1996) *Biochem Mol Biol Int* **39**, 771-779
22.   Ferre-D'Amare, A. R., and Burley, S. K. (1994) *Structure* **2**, 357-359
23.   D'Arcy, A., Banner, D. W., Janes, W., Winkler, F. K., Loetscher, H., Schonfeld, H. J., Zulauf, M., Gentz, R., and Lesslauer, W. (1993) *J Mol Biol* **229**, 555-557
24.   Matthews, B. W. (1968) *J Mol Biol* **33**, 491-497
25.   Vagin, A., and Teplyakov, A. (1997) *Journal of Applied Crystallography* **30**, 1022-1025
26.   Doublie, S. (1997) *Methods Enzymol* **276**, 523-530
27.   Pflugrath, J. (1999) *Acta Crystallography* **D55**, 1718-1725
28.   Terwilliger, T. C. (2003) *Methods Enzymol* **374**, 22-37
29.   Cowtan, K., and Main, P. (1998) *Acta Crystallogr D Biol Crystallogr* **54 ( Pt 4)**, 487-493

30. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2002) *Acta Crystallographica* **D58**, 968-975
31. Winn, M. D., Murshudov, G. N., and Papiz, M. Z. (2003) *Methods Enzymol* **374**, 300-321
32. van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W., and Vriend, G. (1996) *J Comput Aided Mol Des* **10**, 255-262
33. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) *Acta Crystallogr D Biol Crystallogr* **59**, 1131-1137
34. 4, C. C. P. N. (1994) *Acta Crystallography* **D50**, 760-763
35. Berman, H. M., Bhat, T. N., Bourne, P. E., Feng, Z., Gilliland, G., Weissig, H., and Westbrook, J. (2000) *Nat Struct Biol* **7 Suppl**, 957-959
36. Dixon, M. (1953) *Biochemistry Journal* **55**, 170-171
37. Stankovich, M. T. (1980) *Anal Biochem* **109**, 295-308
38. Brown, E., and Wood, J. (1993) *Journal of Biological Chemistry* **268**, 8972-8979
39. Mezl, V. A., and Knox, W. E. (1976) *Anal Biochem* **74**, 430-440
40. Tarpey, M. M., and Fridovich, I. (2001) *Circ Res* **89**, 224-236
41. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res* **22**, 4673-4680
42. Hendrickson, W., and Ogata, C. (1997) *Methods Enzymol* **276**, 494-523
43. Walden, H., Bell, G. S., Russell, R. J., Siebers, B., Hensel, R., and Taylor, G. L. (2001) *J Mol Biol* **306**, 745-757
44. Zhang, M. (2004) in *Chemistry*, pp. 36-54, University of Missouri, Columbia
45. Zhang, W., Zhang, M., Zhu, W., Zhou, Y., Wanduragala, S., Rewinkel, D., Tanner, J. J., and Becker, D. F. (2007) *Biochemistry* **46**, 483-491
46. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E., and Edelman, M. (1999) *Bioinformatics* **15**, 327-332
47. Anand, K., Pal, D., and Hilgenfeld, R. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 1722-1728
48. Krishnan, N., and Becker, D. F. (2006) *J Bacteriol* **188**, 1227-1235
49. Vinod, M. P., Bellur, P., and Becker, D. F. (2002) *Biochemistry* **41**, 6525-6532
50. Zhu, W., Gincherman, Y., Docherty, P., Spilling, C. D., and Becker, D. F. (2002) *Arch Biochem Biophys* **408**, 131-136
51. Wood, J. M. (1981) *J Bacteriol* **146**, 895-901
52. Bottoms, C. A., Smith, P. E., and Tanner, J. J. (2002) *Protein Sci* **11**, 2125-2137
53. Chen, C. C., and Wilson, T. H. (1986) *J Biol Chem* **261**, 2599-2604
54. Gu, D., Zhou, Y., Kallhoff, V., Baban, B., Tanner, J. J., and Becker, D. F. (2004) *J Biol Chem*
55. Phang, J. M., Hu, C. A., and Valle, D. (2001) in *Metabolic and molecular basis of inherited disease* (Scriver, C. R., A.L., B., Sly, W., and Valle, D., eds), McGraw Hiss, New York
56. Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W., and Vogelstein, B. (1997) *Nature* **389**, 300-305

57. Hu, C. A., Donald, S. P., Yu, J., Lin, W. W., Liu, Z., Steel, G., Obie, C., Valle, D., and Phang, J. M. (2006) *Mol Cell Biochem*

58. Rivera, A., and Maxwell, S. A. (2005) *J Biol Chem* **280**, 29346-29354

59. Pandhare, J., Cooper, S. K., and Phang, J. M. (2006) *J Biol Chem* **281**, 2044-2052

60. Maxwell, S. A., and Davis, G. E. (2000) *PNAS* **97**, 13009-13014

61. Liu, Y., Borchert, G. L., Donald, S. P., Surazynski, A., Hu, C. A., Weydert, C. J., Oberley, L. W., and Phang, J. M. (2005) *Carcinogenesis* **26**, 1335-1342

62. Maxwell, S. A., and Rivera, A. (2003) *J Biol Chem* **278**, 9784-9789

63. Donald, S. P., Sun, X. Y., Hu, C. A., Yu, J., Mei, J. M., Valle, D., and Phang, J. M. (2001) *Cancer Res* **61**, 1810-1815

64. Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I., and Grimes, J. M. (2006) *Structure* **14**, 1617-1622

SCHEME 2.1.  Proline Catabolism



proline     $\xrightarrow[\text{PRODH}]{\text{FAD} \quad \text{FADH}_2}$     $\Delta^1$-pyrroline-5-carboxylate (P5C)

$\pm\ H_2O$   nonenzymatic

glutamic semialdehyde    $\xrightarrow[\text{P5CDH}]{\text{H}_2\text{O} \quad \text{NAD}^+ \quad \text{NADH} + \text{H}^+}$    glutamate
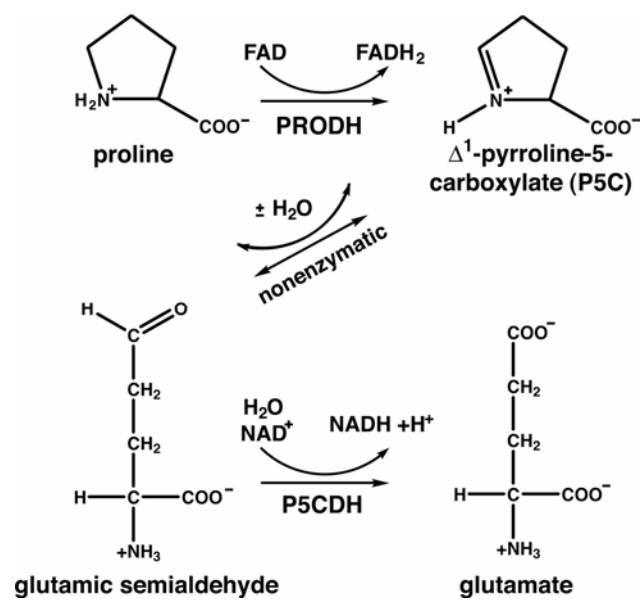
FIGURE 2.1. Ni-NTA Affinity Purification of TtPRODH. Lanes correspond to the following: 1) Marker (Biorad Precision Plus). 2) 14 μL pre-IPTG sample (1 mL of $OD_{600}$ = 0.6 cells were pelleted and resuspended in 1 mL $H_2O$). 3) 14 μL Post-IPTG sample (300 μL of induced cells were pelleted and resuspended in 1 mL $H_2O$). 4) 2 μL Cell free extract. 5) 2 μL pellet. 6) 2 μL flowthrough. 7) 14 μL wash 50 mM imidazole. 8) 14 μL wash 75 mM imidazole. 9) 14 μL elution 1. 10) 14 μL elution 2. 11) 14 μL elution 3 12) 14 μL elution 4.
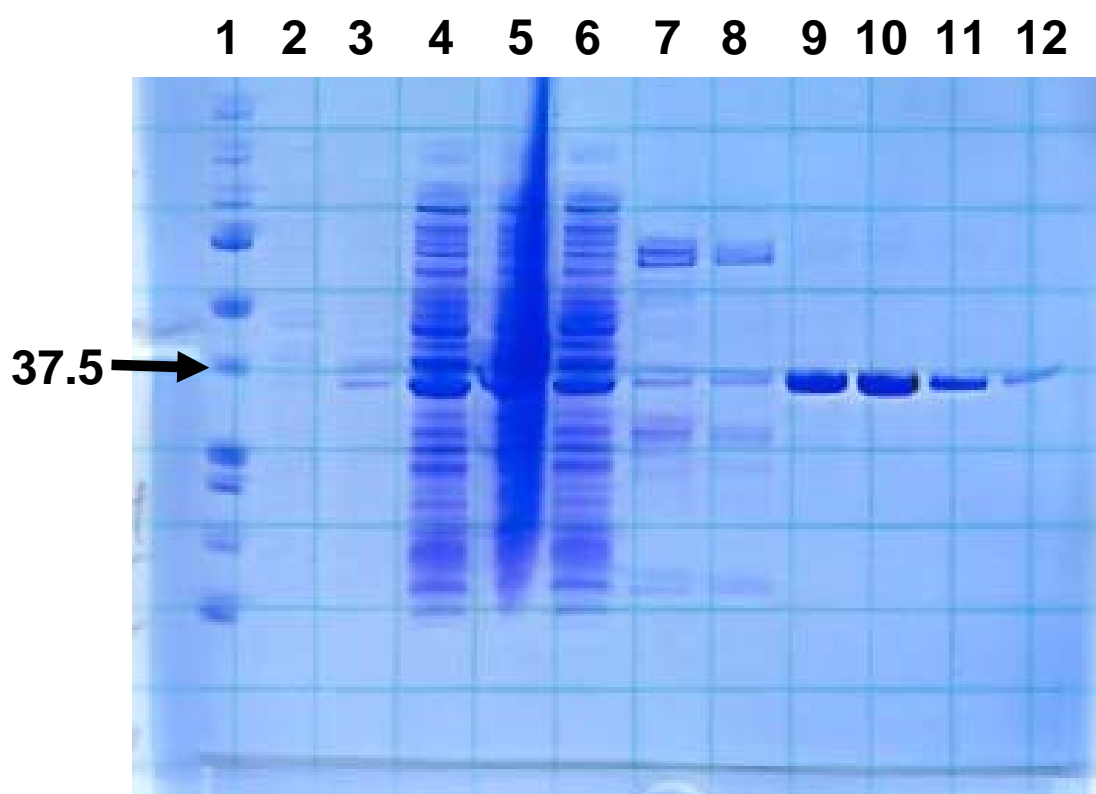
FIGURE 2.2. Initial crystals obtained from crystal screen condition Wizard 4 (35% 2-methyl-2,4-pentanediol, 100 mM Imidazole pH = 8, 200 mM MgCl$_2$).
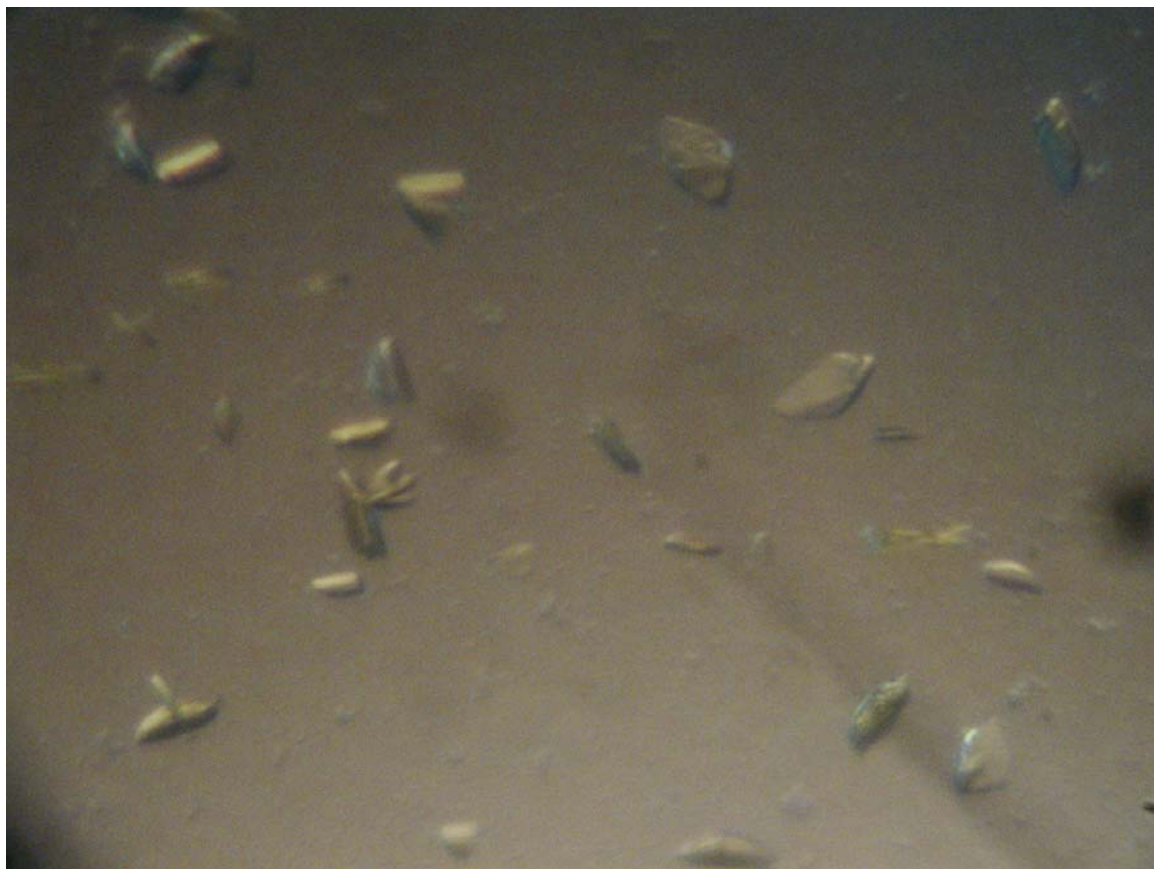
FIGURE 2.3.  Optimized TtPRODH MPD crystals which diffract to 2.3 Å (50 mM MgCl$_2$, 100 mM imidazole pH = 7.5, 35 % MPD).
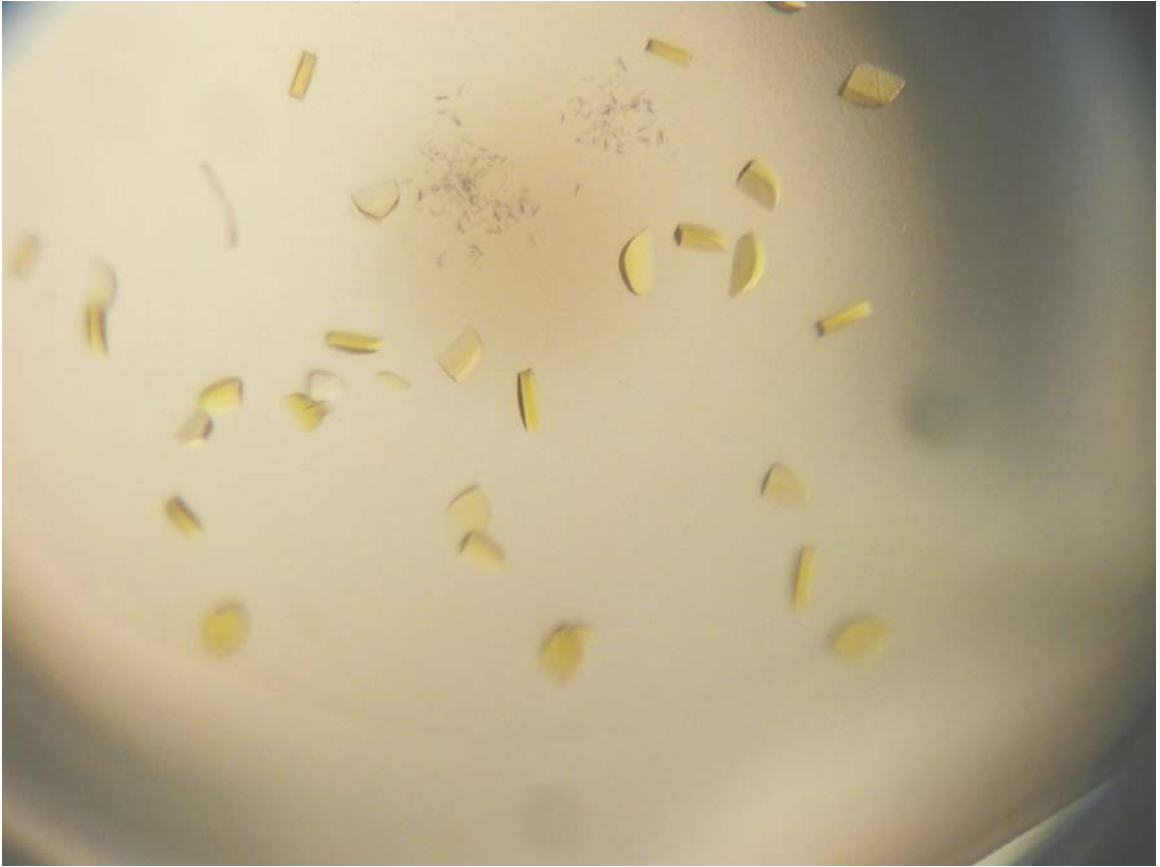
FIGURE 2.4. Selenomethionyl crystals used for structure determination which diffracted to 2 Å (100 mM imidazole pH = 7.0, 100mM $MgCl_2$, 17% MPD, 5 mM Dithiothreitol in reservoir and fresh 20 mM N-octyl β-D thioglucopyranoside was added to 3 mg/mL protein).

FIGURE 2.5. Experimental fluorescence spectra determined for selenomethionyl TtPRODH to utilize the anomalous signal at the peak wavelength of 12666 eV for singlewavelength anomalous diffraction.
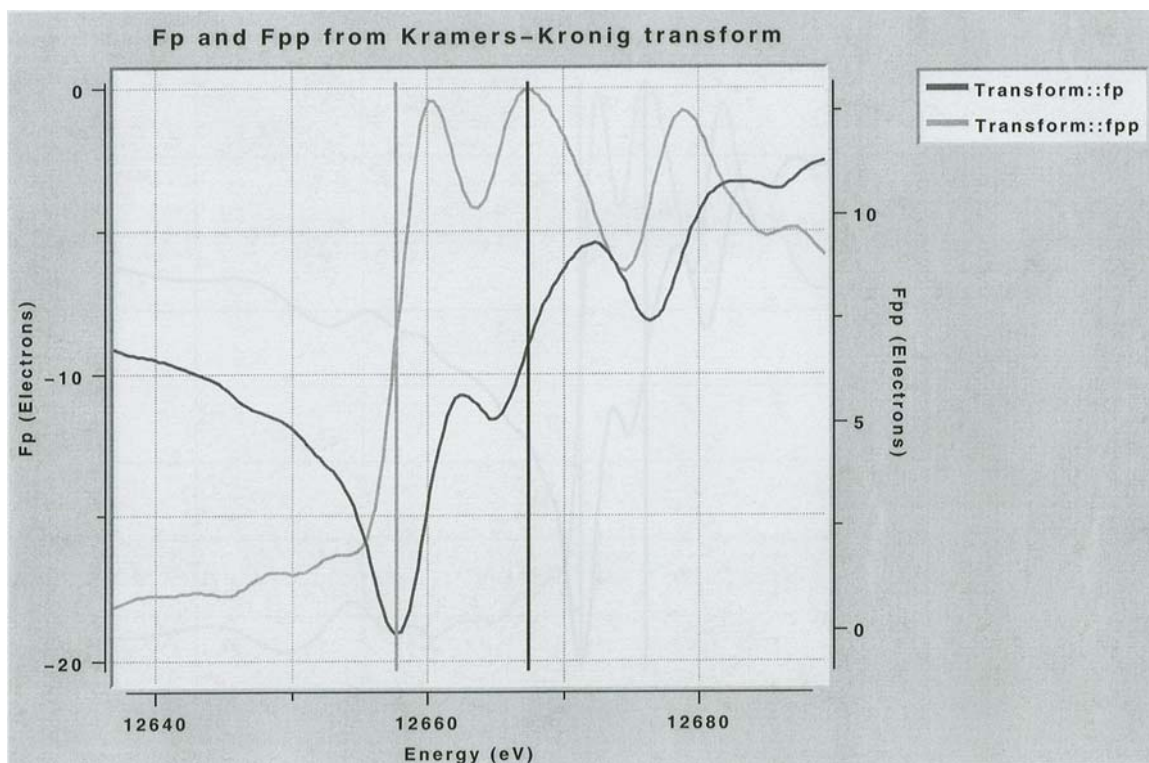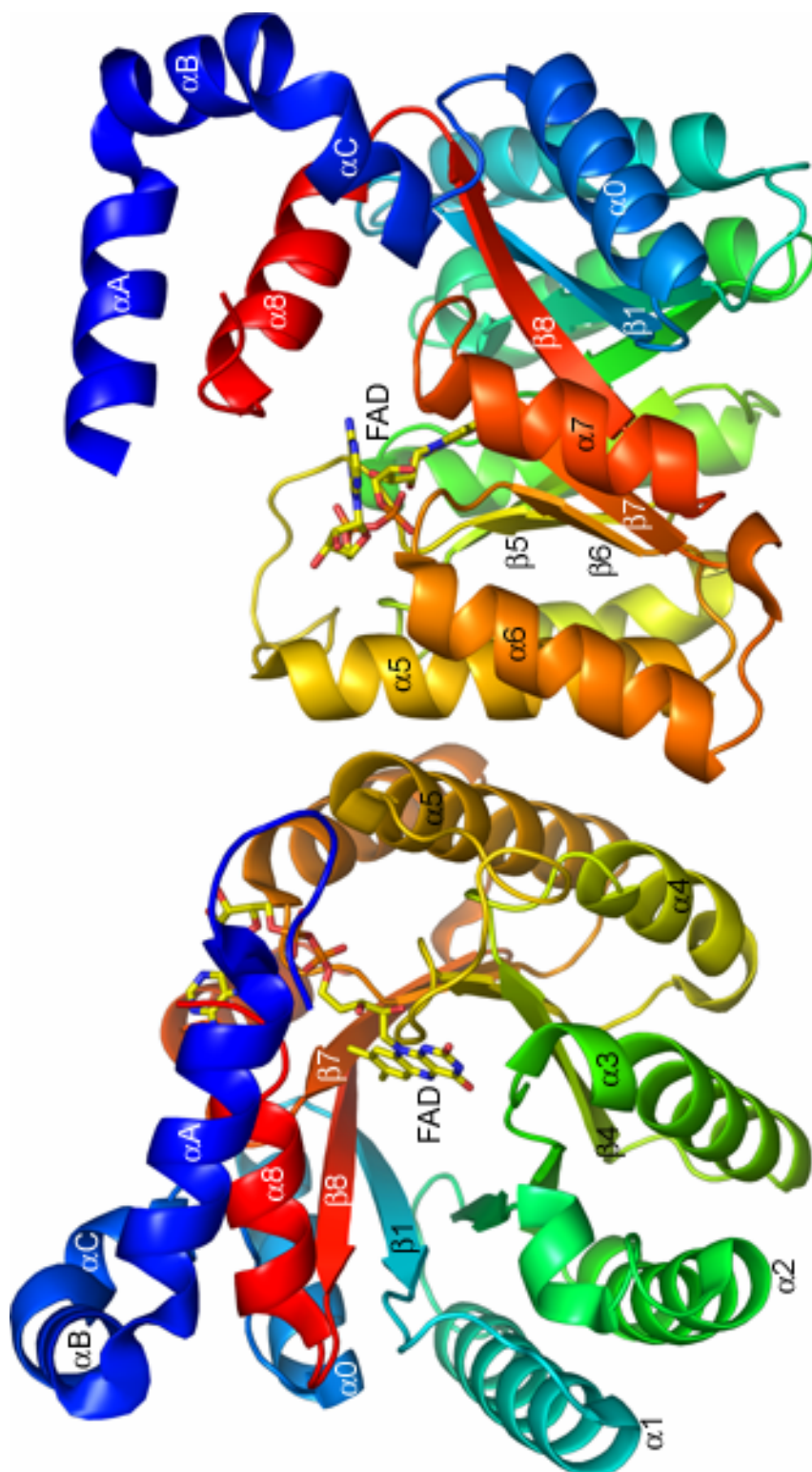
Figure 2.6. Overall structure of TtPRODH. Ribbon drawing of the two TtPRODH molecules in the asymmetric unit. The protein chains are colored in the rainbow scheme, with dark blue at the N-terminus and red at the C-terminus. Selected α-helices and β-strands are labeled. The FADs are drawn as stick models in yellow. The glycine hinge between β8 and α8 is noted (Gly279).

FIGURE 2.7. The solvent exposed hydrophobic patch formed by α-helices A, B, C and 8. The orientation is similar to that of the right hand protein in FIGURE 2.6. Hydrophobic side chains of the patch are colored green.
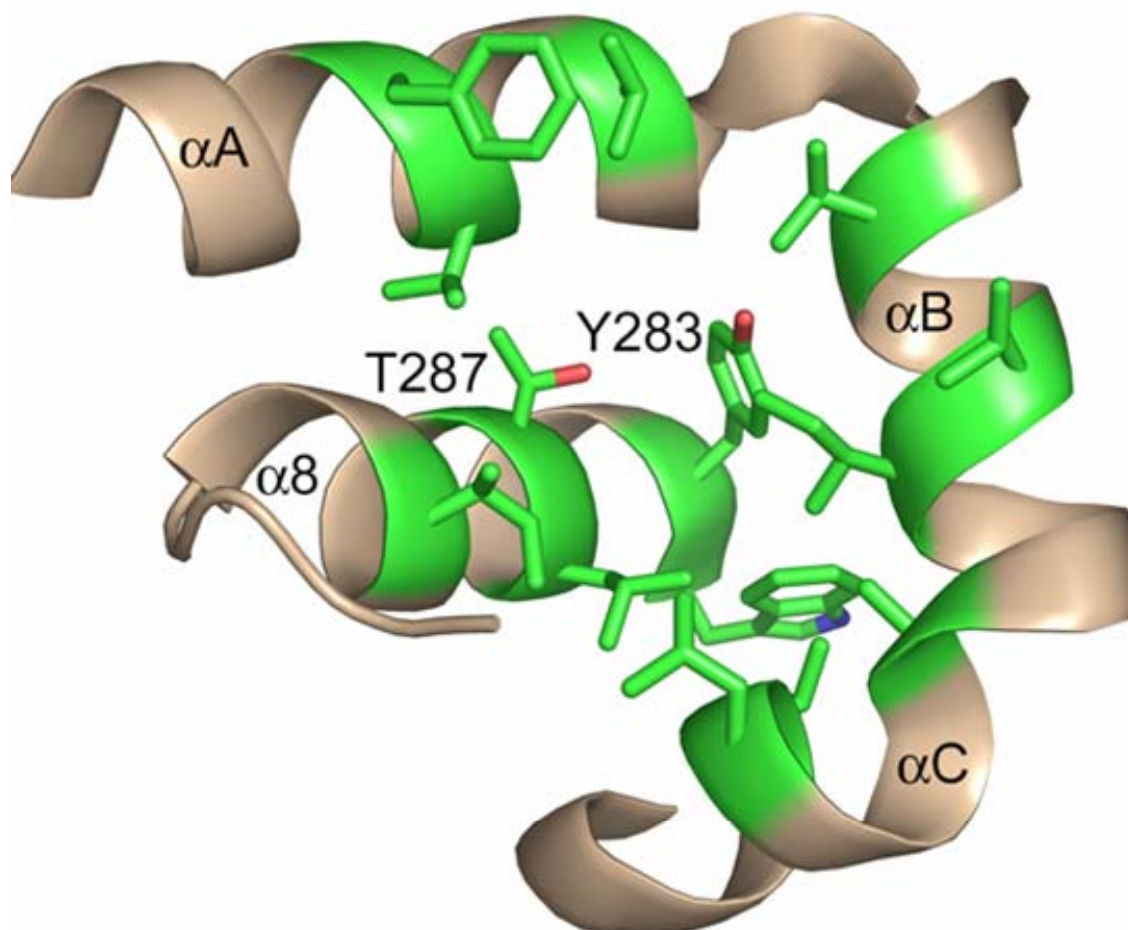
FIGURE 2.8. Surface representation of the hydrophobic patch.  The orientation is identical to that of FIGURE 2.7.
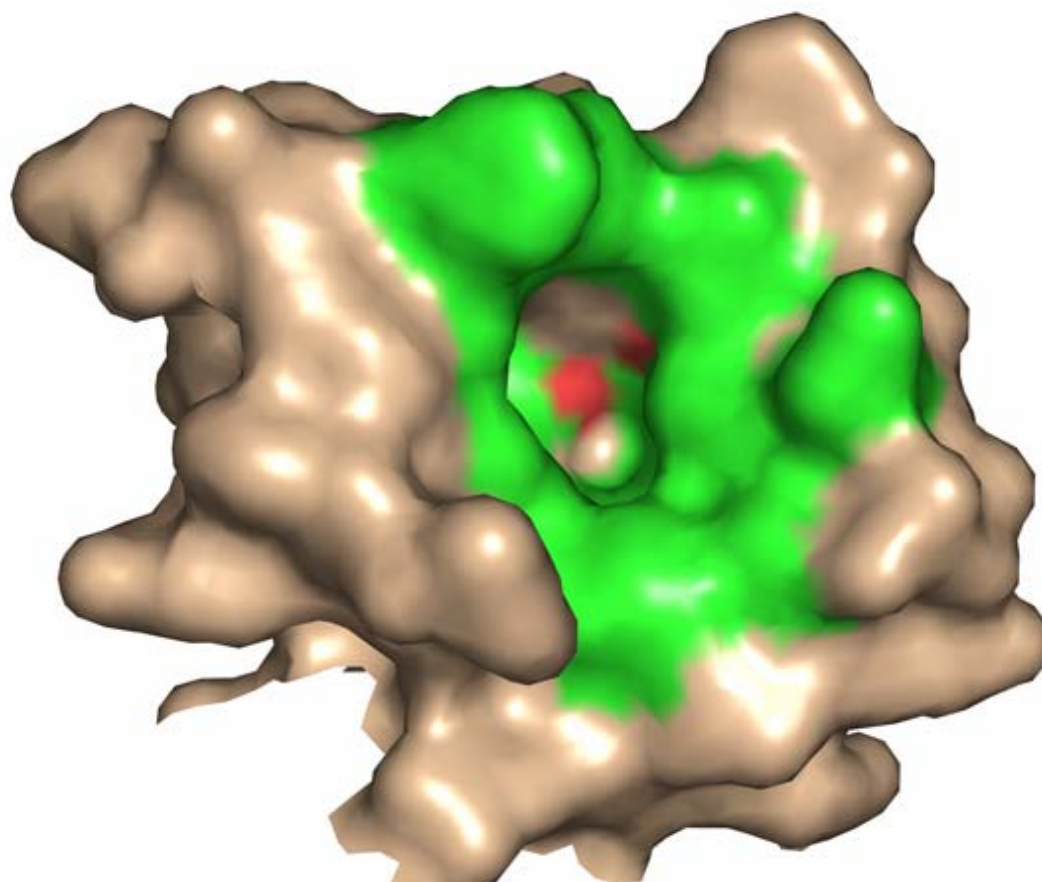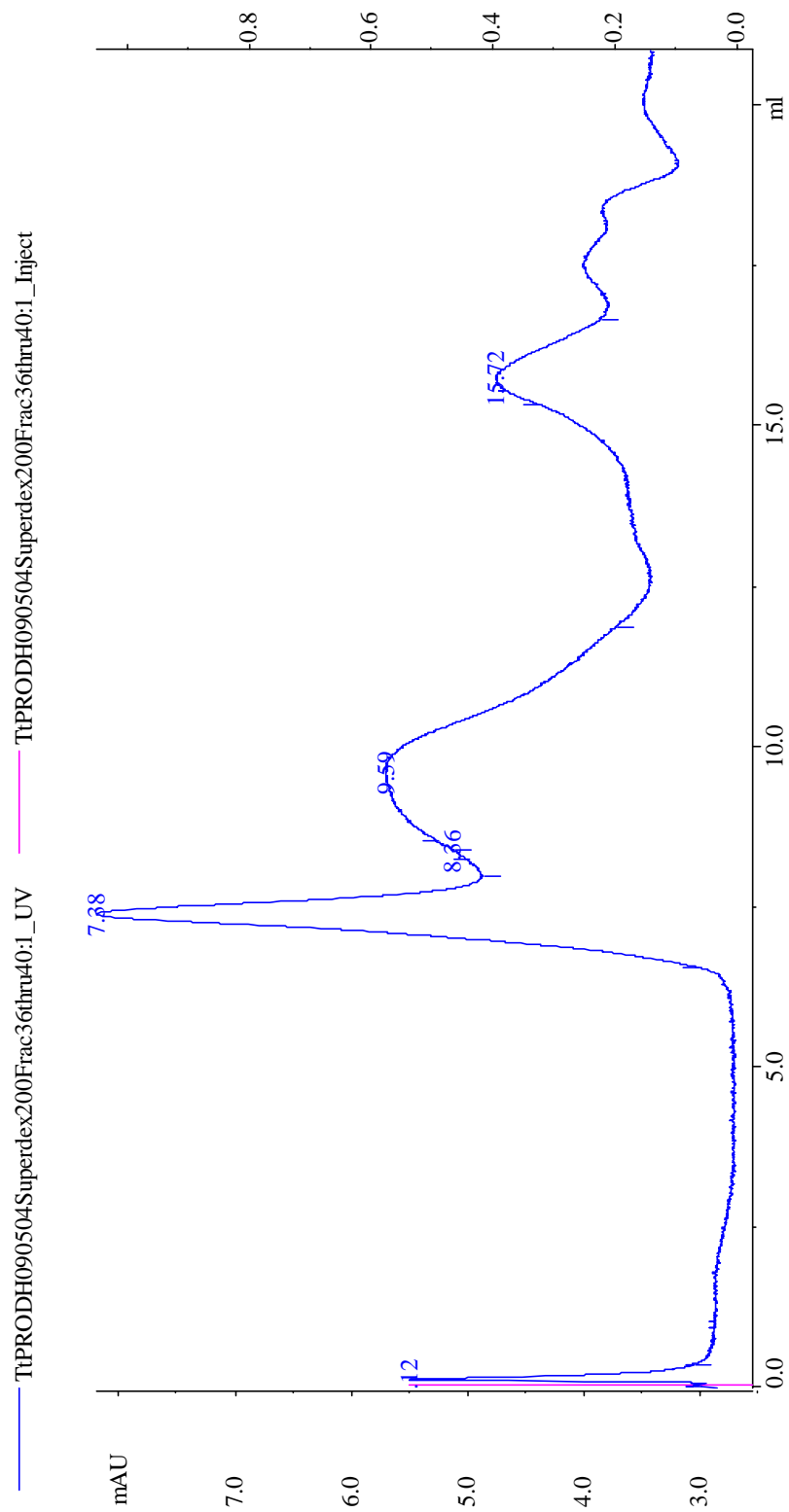
FIGURE 2.9. Gel filtration profile of soluble aggregated TtPRODH in 50 mM Tris pH 8, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM DTT, and 5 % glycerol.
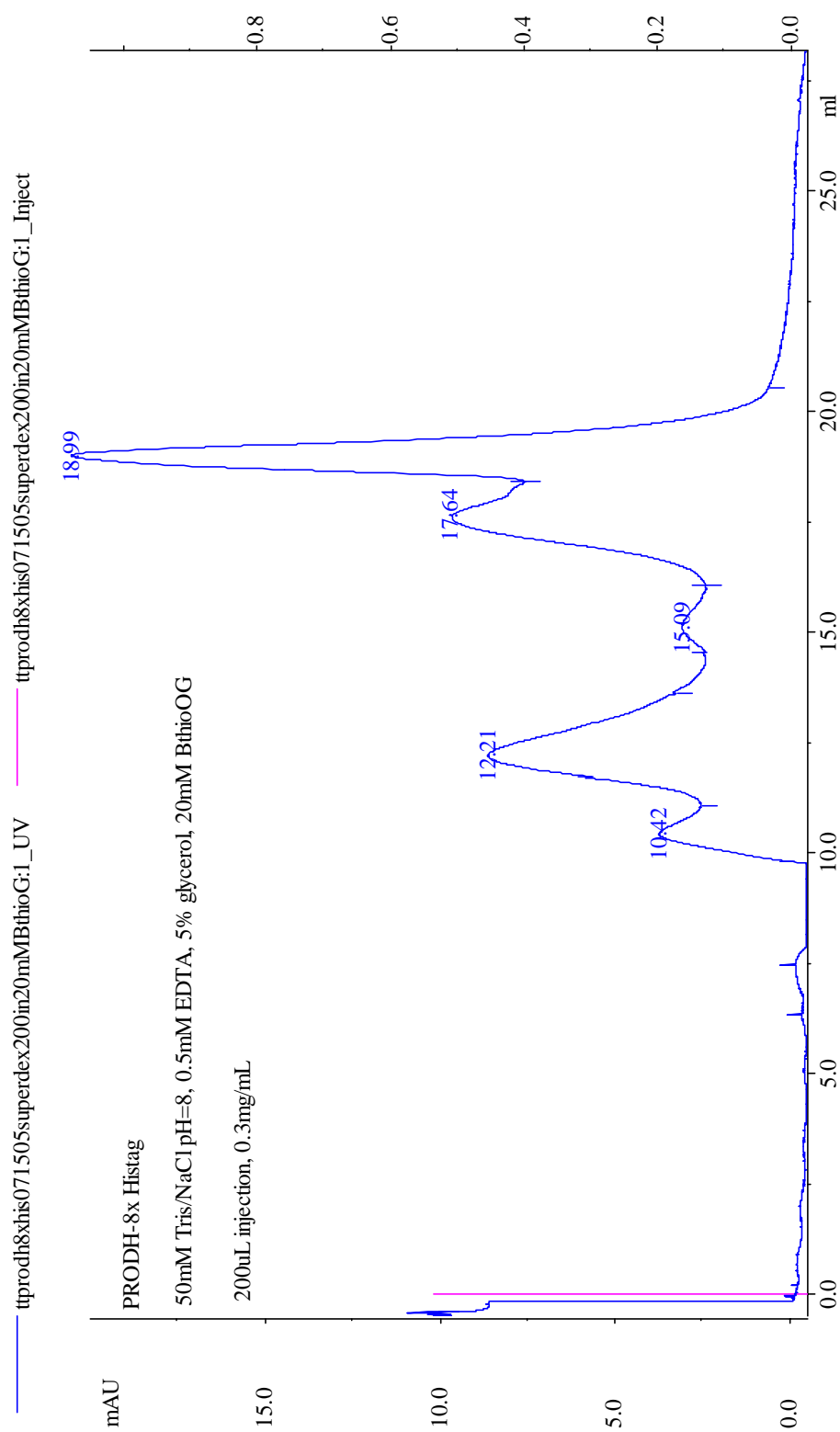
FIGURE 2.10. Gel filtration profile of TtPRODH upon addition of 20 mM n-octyl-β-glucopyanoside in 50 mM Tris pH 8, 50 mM NaCl, 0.5 mM DTT, 0.5 mM EDTA, and 5 % glycerol.

FIGURE 2.11. Superposition of TtPRODH (blue) and PutA86-669/THFA (white, PDB code 1TIW). The FAD cofactors of TtPRODH and PutA86-669 are colored yellow and green, respectively. Glu65 of TtPRODH is drawn in stick mode. Trp438 of PutA86-669, which stacks against the FAD adenine, is also drawn. The dashed lines indicate the Glu289-Arg555 ion pair in PutA86-669.
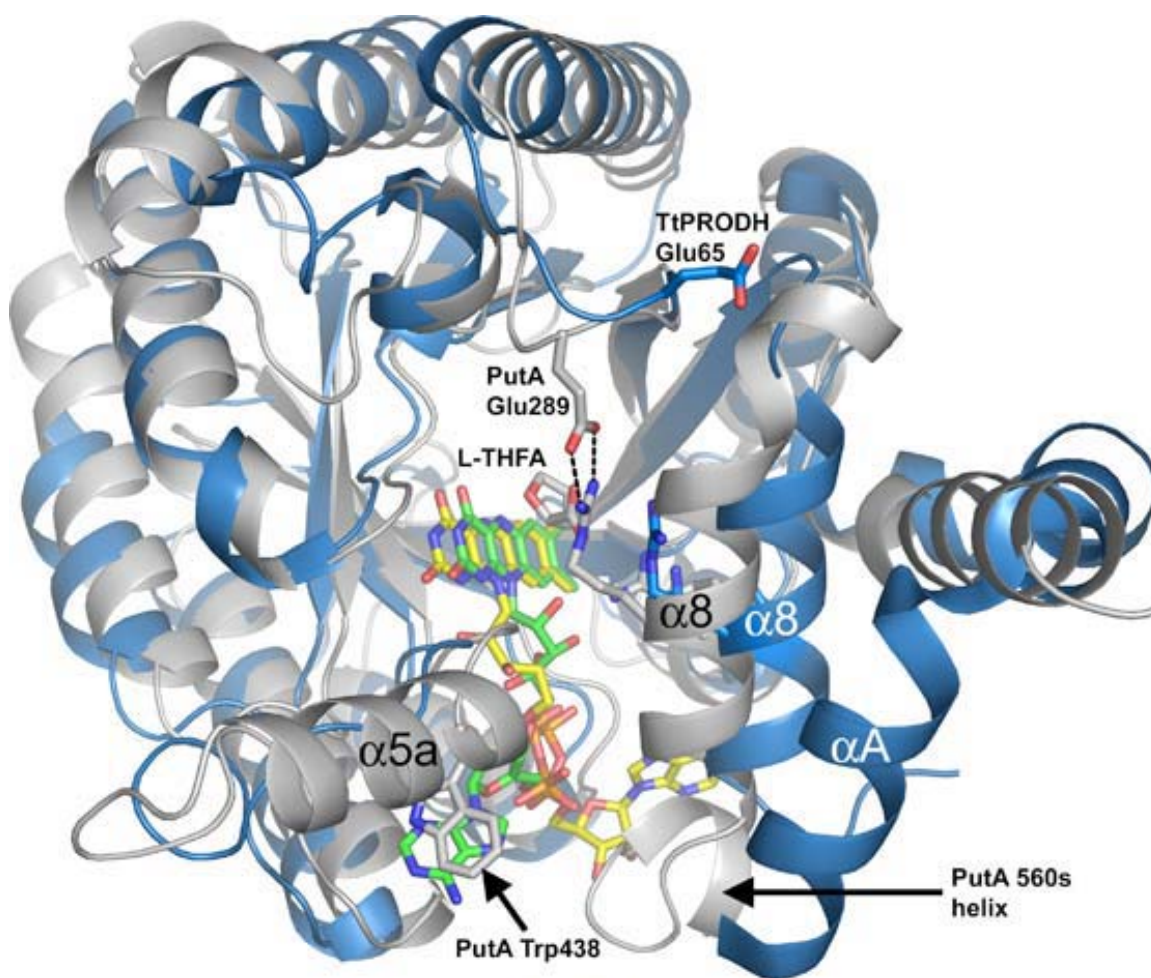
FIGURE 2.12. Stereographic drawing of protein-FAD interactions in TtPRODH. The dotted lines indicate hydrogen bonds and ion pairs. The FAD is colored yellow. Residues interacting with FAD are colored green. Strands β4 and β6 are indicated.
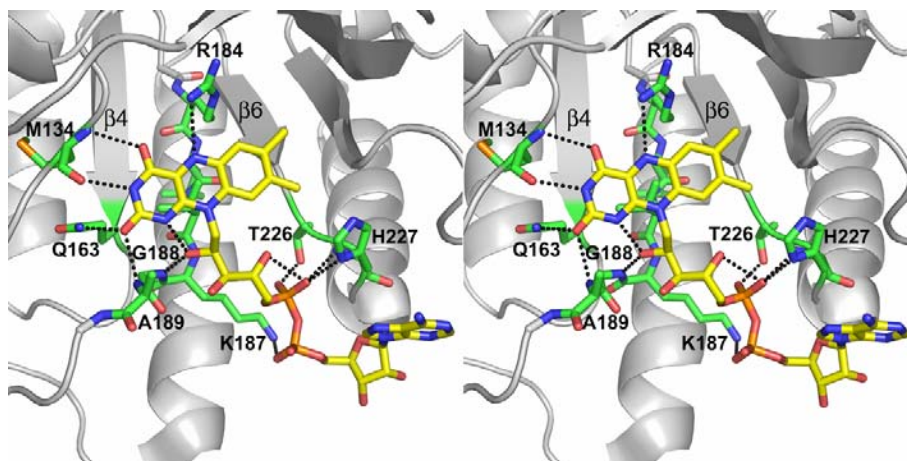
FIGURE 2.13.  Schematic of TtPRODH interactions with FAD.  Bacterial PutA/PRODH completely conserved interactions are boxed with the corresponding motif listed.
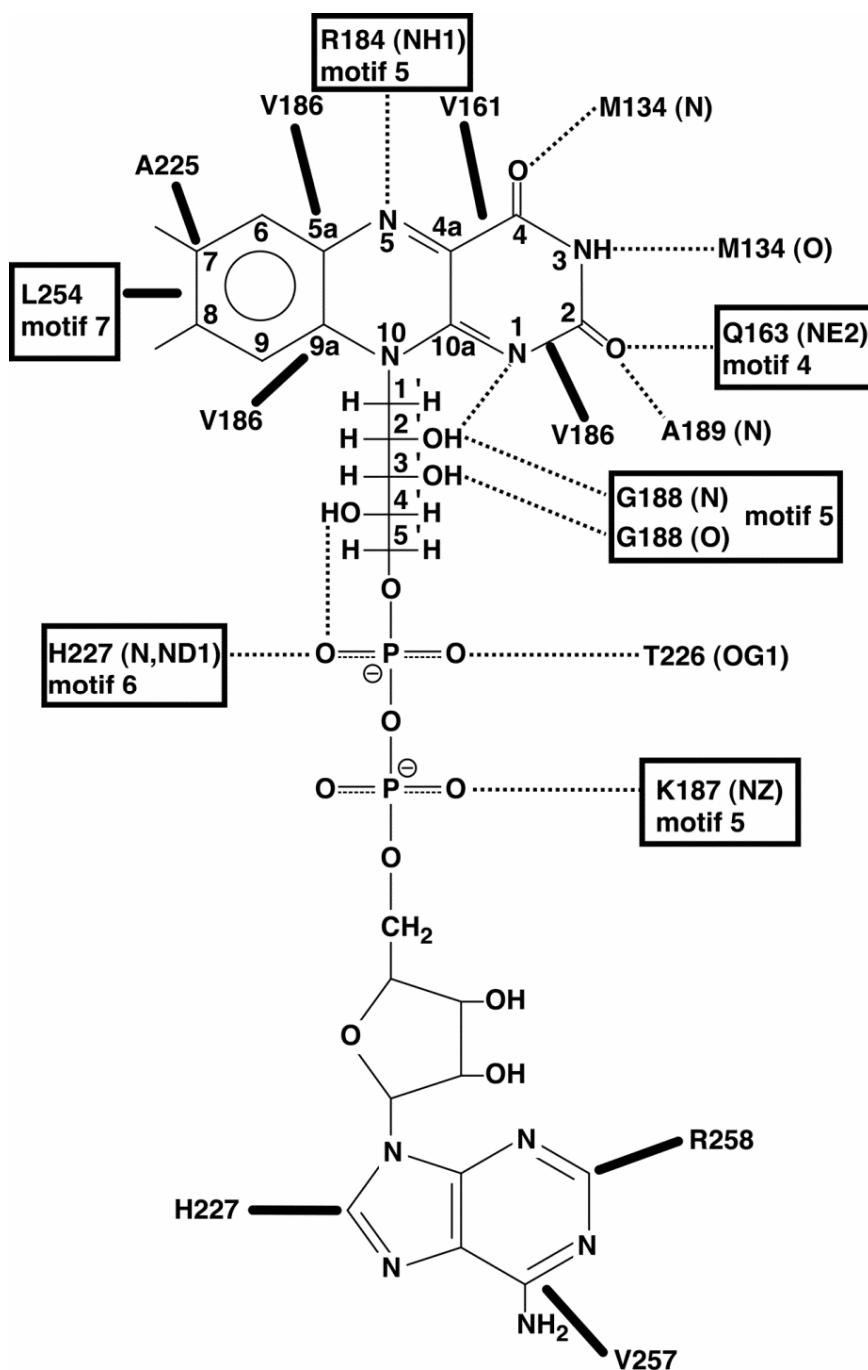
FIGURE 2.14. Two views of the FAD from TtPRODH covered by an experimental electron density map (1σ). The map was calculated using |F_obs| and experimental phases after density modification.
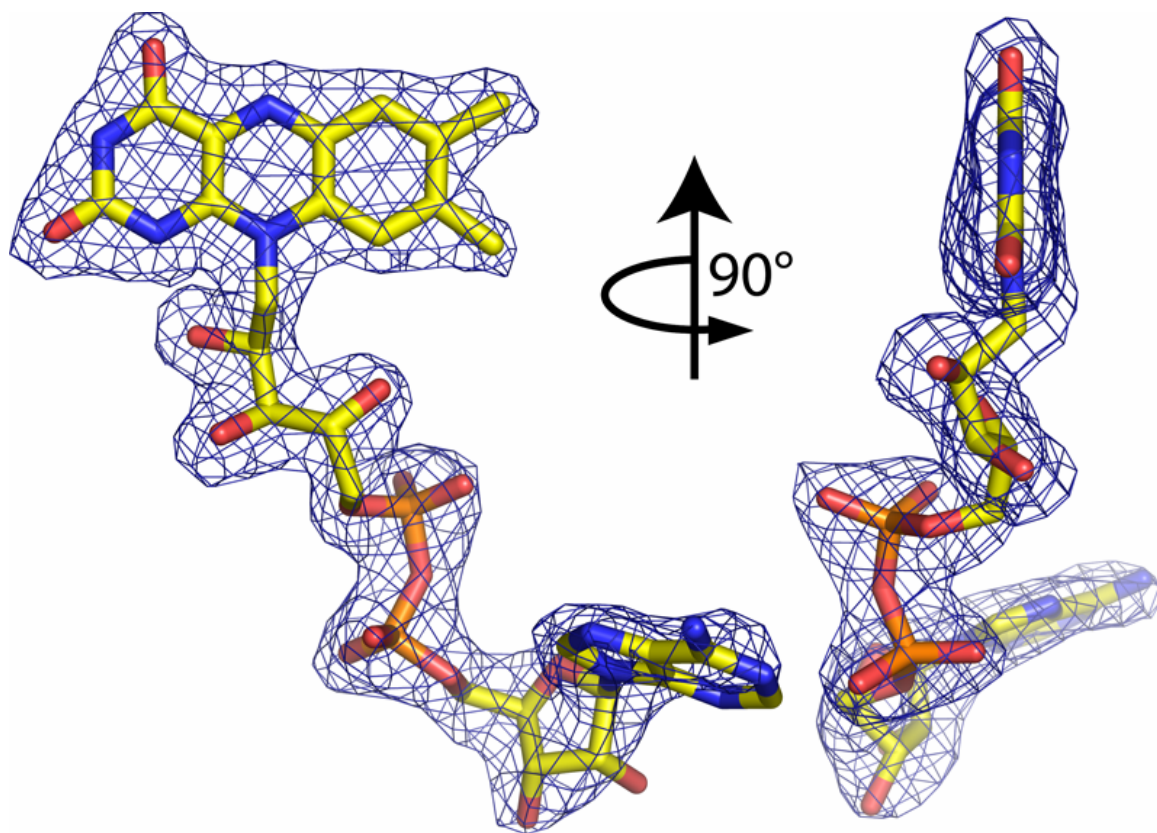
FIGURE 2.15. Superposition of the FAD cofactors from TtPRODH (yellow C atoms, orange P atoms), PutA86-669/THFA (green C atoms, cyan P atoms) and dithionite-reduced PutA86-669 (white C atoms, cyan P atoms). The 2'-OH and 4'-OH groups of TtPRODH are indicated. Note that the 2'-OH groups of TtPRODH and dithionite-reduced PutA86-669 superimpose nearly perfectly, whereas the 2'-OH group of PutA86-669/THFA points toward the viewer. The orange dotted lines denote hydrogen bonds in the TtPRODH cofactor. The black dotted lines denote hydrogen bonds in the PutA86-669 cofactors. The 3'-OH hydrogen bond to the ribose is present in both PutA86-669 structures but only one dotted line is drawn for clarity.
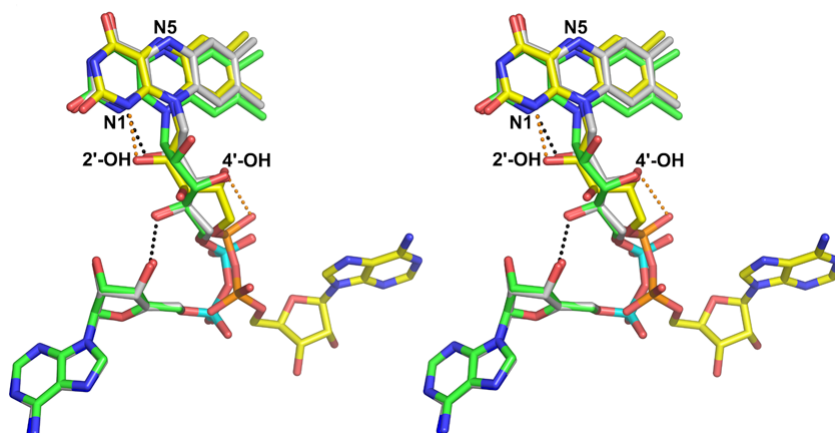
FIGURE 2.16.  Comparison of the active sites of TtPRODH (yellow) and PutA86-669/THFA (white).  This stereoview emphasizes differences between the two enzymes in interactions with the pyrophosphate and adenosine moieties.  Residues are labelled as TtPRODH/PutA86-669.  The orange and black dotted lines denote hydrogen bonds in TtPRODH and PutA86-669, respectively.
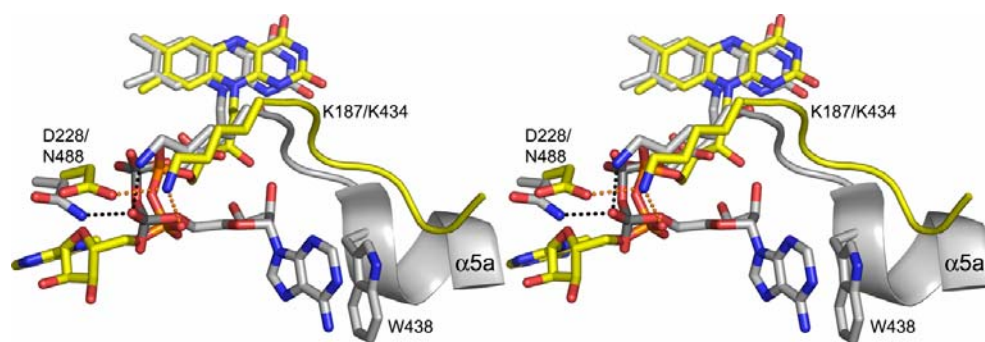
FIGURE 2.17.  Comparison of the active sites of TtPRODH (yellow) and PutA86-669/THFA (white). This stereoview shows differences between the two enzymes in the proline-binding pocket.   The inhibitor THFA is shown in green.  Residues are labelled as TtPRODH/PutA86-669.   The orange and black dotted lines denote hydrogen bonds in TtPRODH and PutA86-669, respectively.  For clarity, Y275/Y540, which form hydrogen bonds with D133/D370, are not labeled.
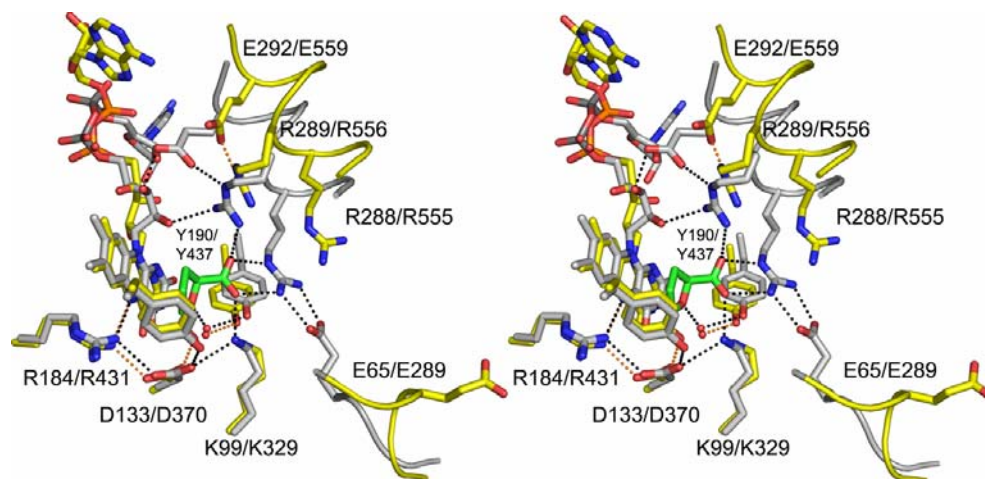
FIGURE 2.18. Space-filling representation of TtPRODH emphasizing solvent exposure of the FAD (yellow). Helix $\alpha8$ is colored magenta. Tyr190 is shown in green. The orientation is similar to that of FIGURE 2.11.
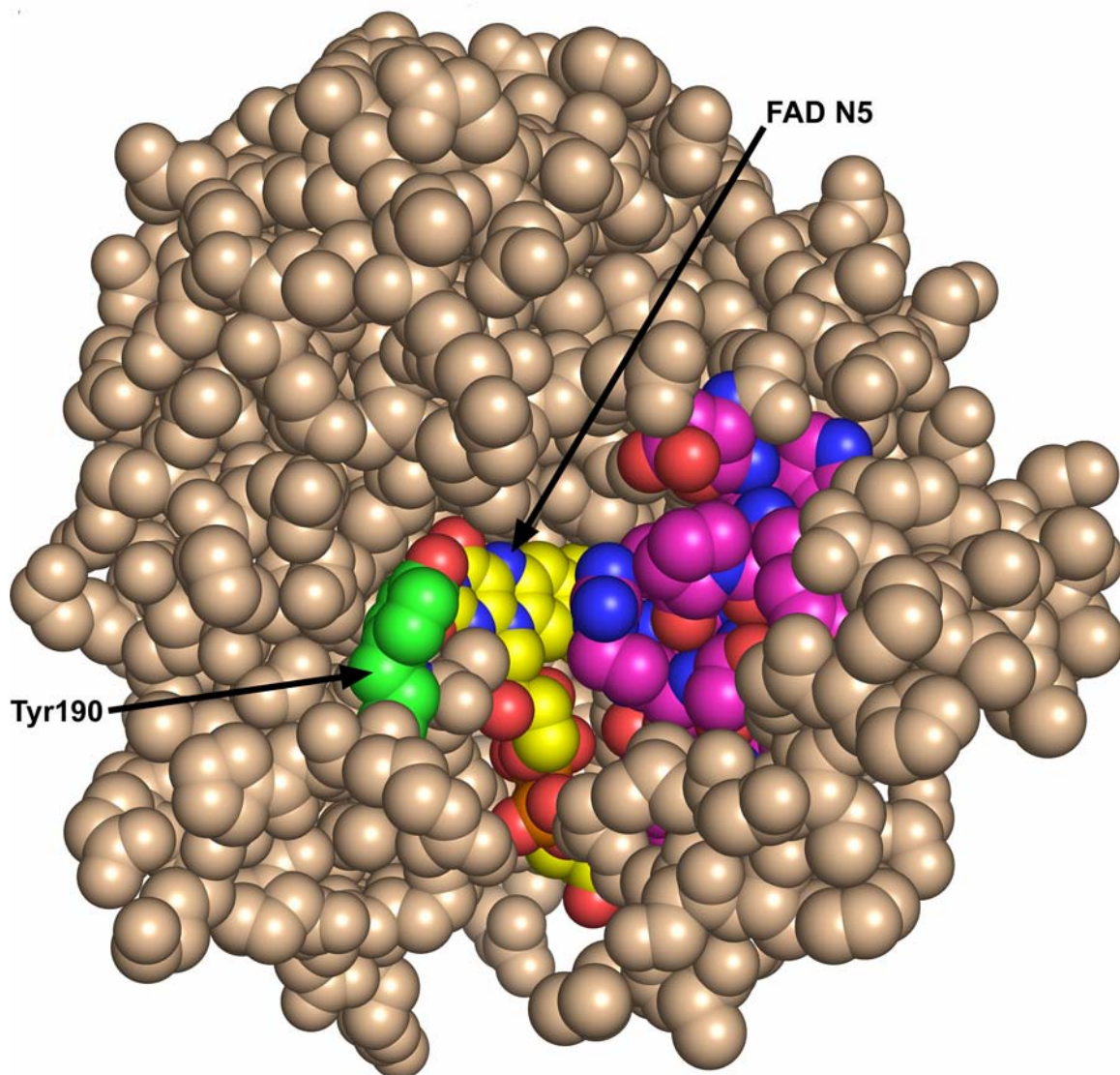
FIGURE 2.19. Comparison of active site solvent exposure in TtPRODH (top) and EcPutA (bottom) PRODH. These two structures were superimposed and then surface rendered with FAD in yellow CPK spheres.
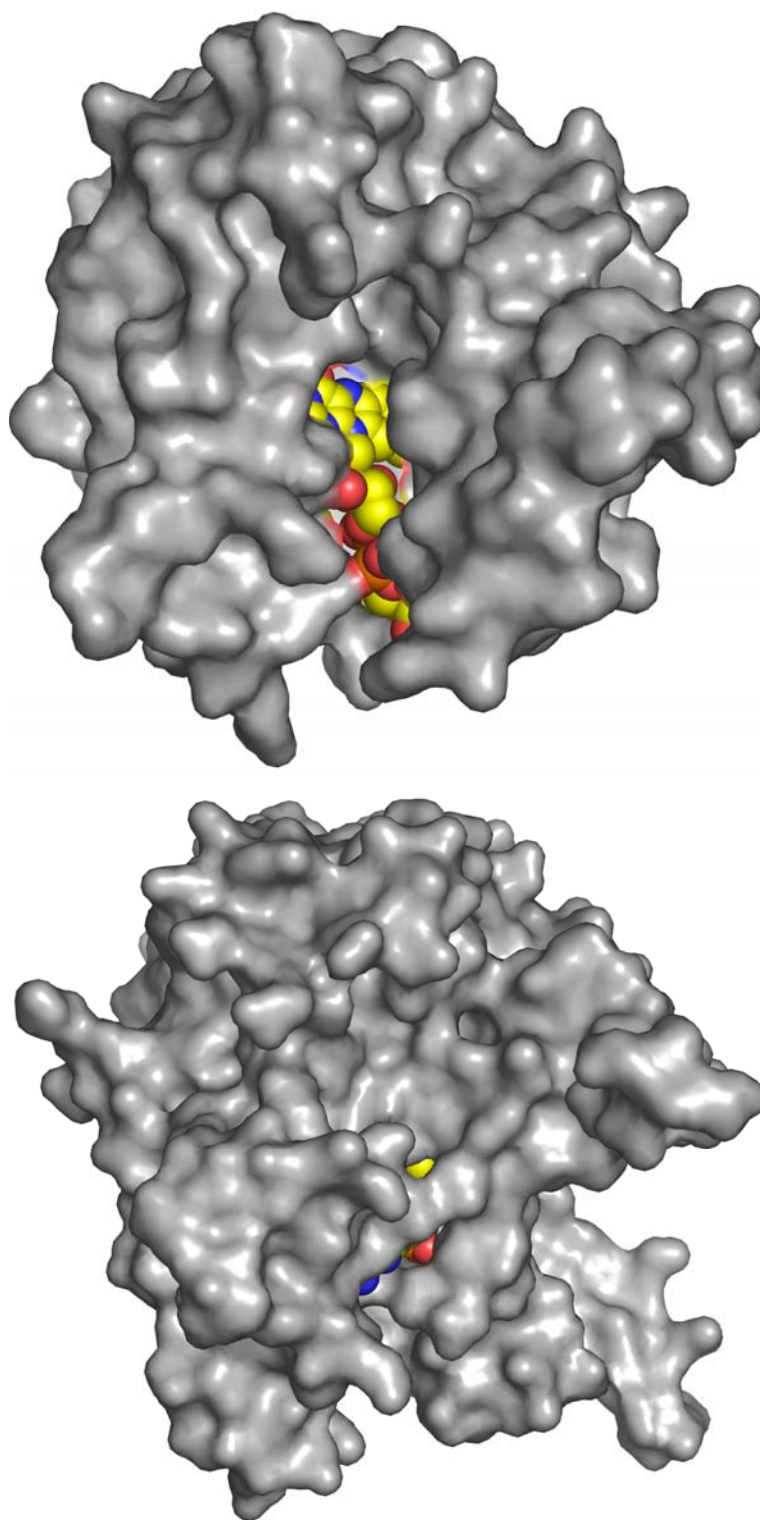
FIGURE 2.20. Locations of 2-methyl-2,4-pentanediol (MPD) in TtPRODH chain A while viewing the electrostatic surface potential. These MPD molecules occur in the same location in chain B. MPD 3 binds the N-terminal ends of the strands and is shown here on the right in yellow CPK sticks. MPD 4 is shown on the left in yellow CPK sticks.
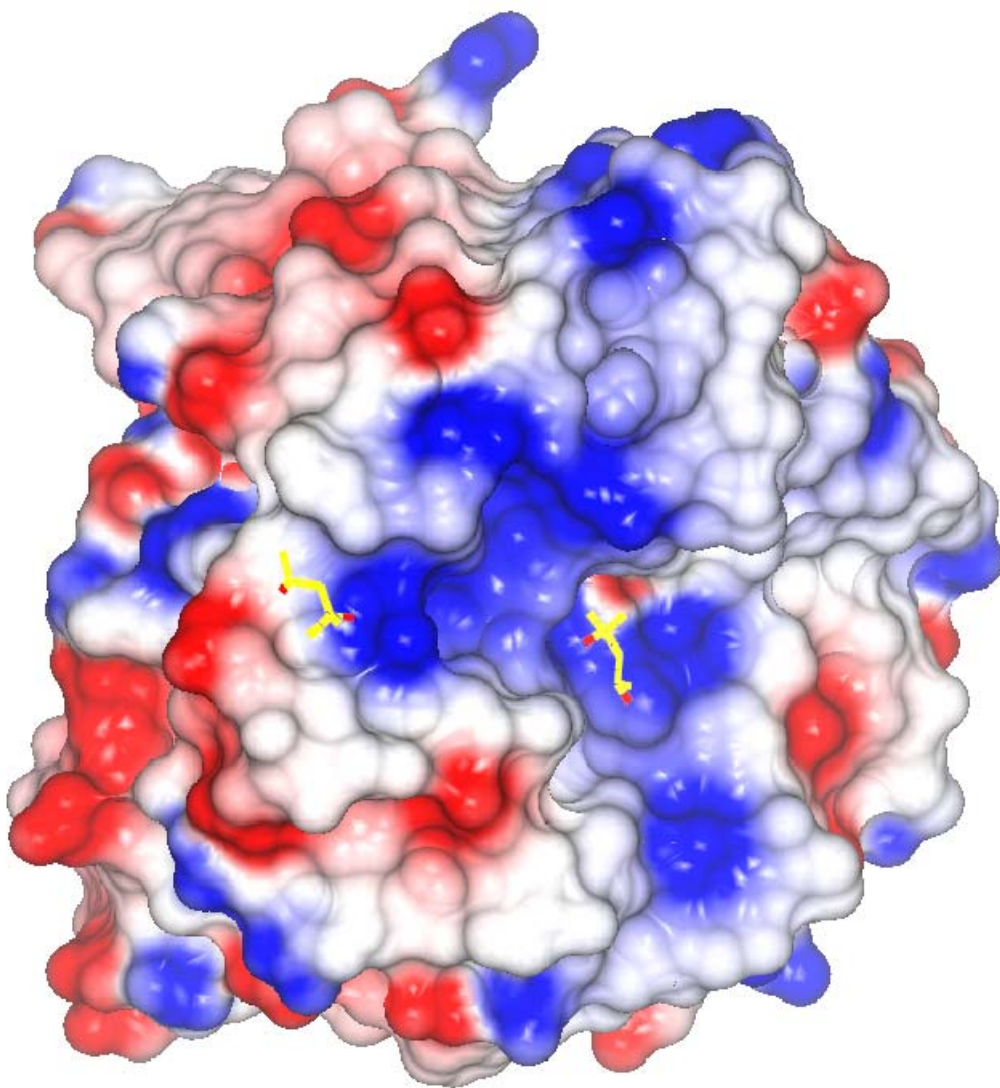
FIGURE 2.21.  TtPRODH electrostatic surface potential surrounding MPD site 1. Blue surface indicates areas of positive potential and red areas indicate areas of negative potential.  MPD in site 1 is shown in yellow.
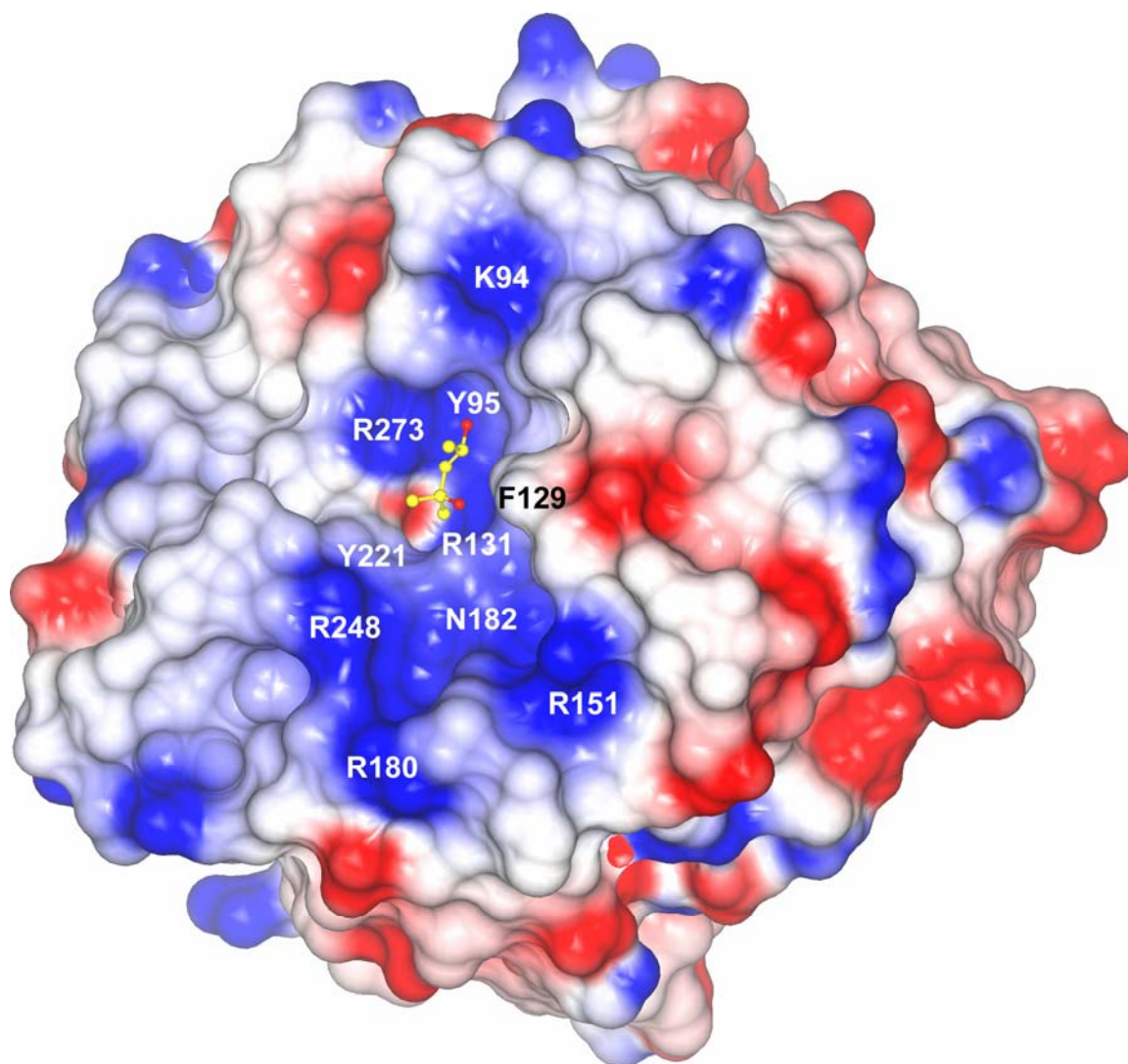
FIGURE 2.22. Proposed electron transfer pathway for TtPRODH catalyzed oxidation of proline. To transfer electrons from reduced $FADH_2$ to electron acceptors in the membrane, the enzyme could utilize charged residues such as conserved R184, Q252, E250 and R273. R184 is shown in light orange CPK sticks, Q252 is shown in light green CPK sticks, E250 is shown in teal CPK sticks and R273 is shown in salmon CPK sticks. FAD and MPD are shown in yellow and green CPK spheres, respectively. For clarity, some of the $(\beta\alpha)_8$ barrel has been removed.
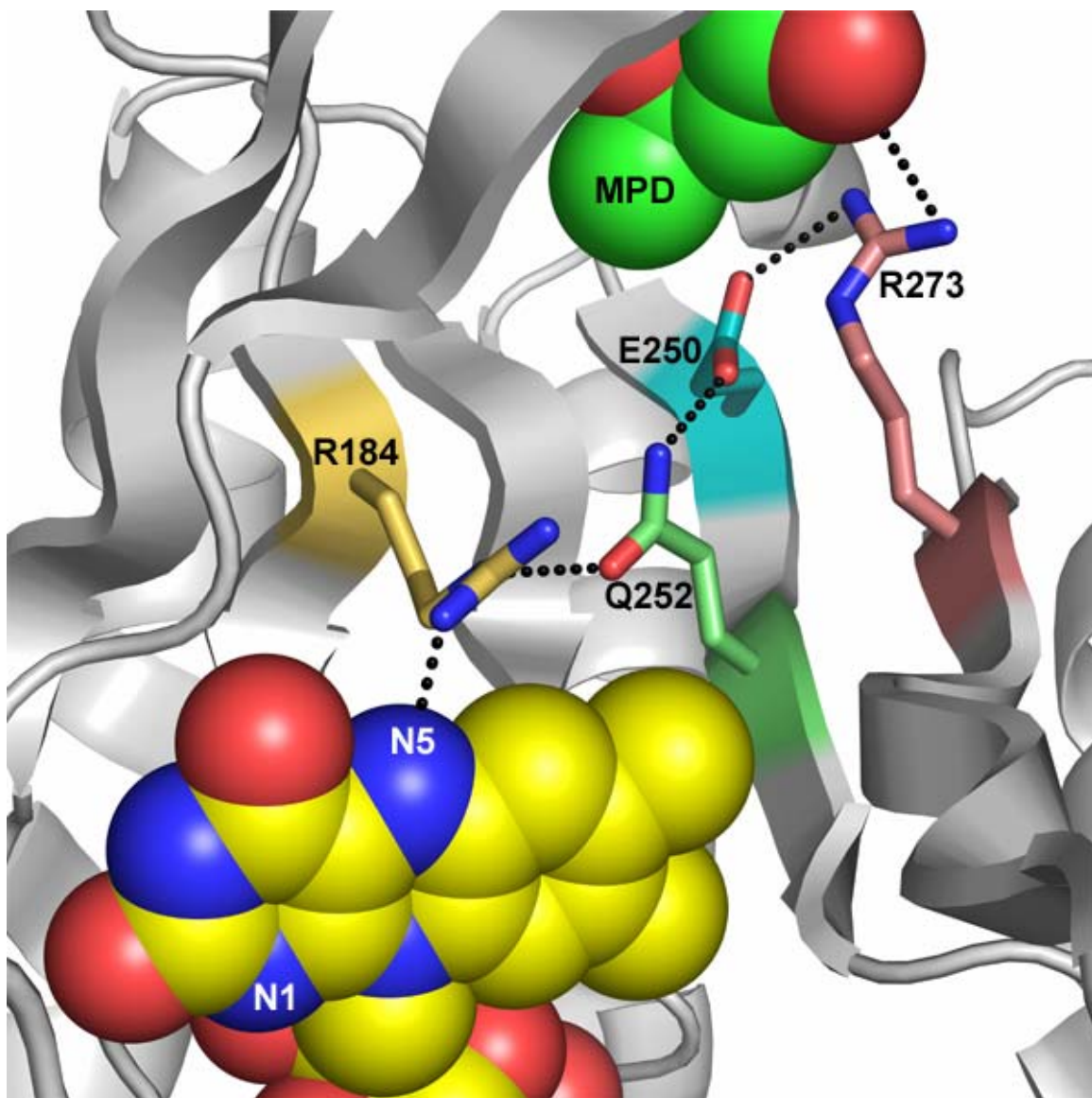
FIGURE 2.23. Proline reduction of TtPRODH. Red line is oxidized TtPRODH. Purple line is reduced with 500 mM L-proline after 1min. Maroon line is the same reduced TtPRODH sample after 16 hours.
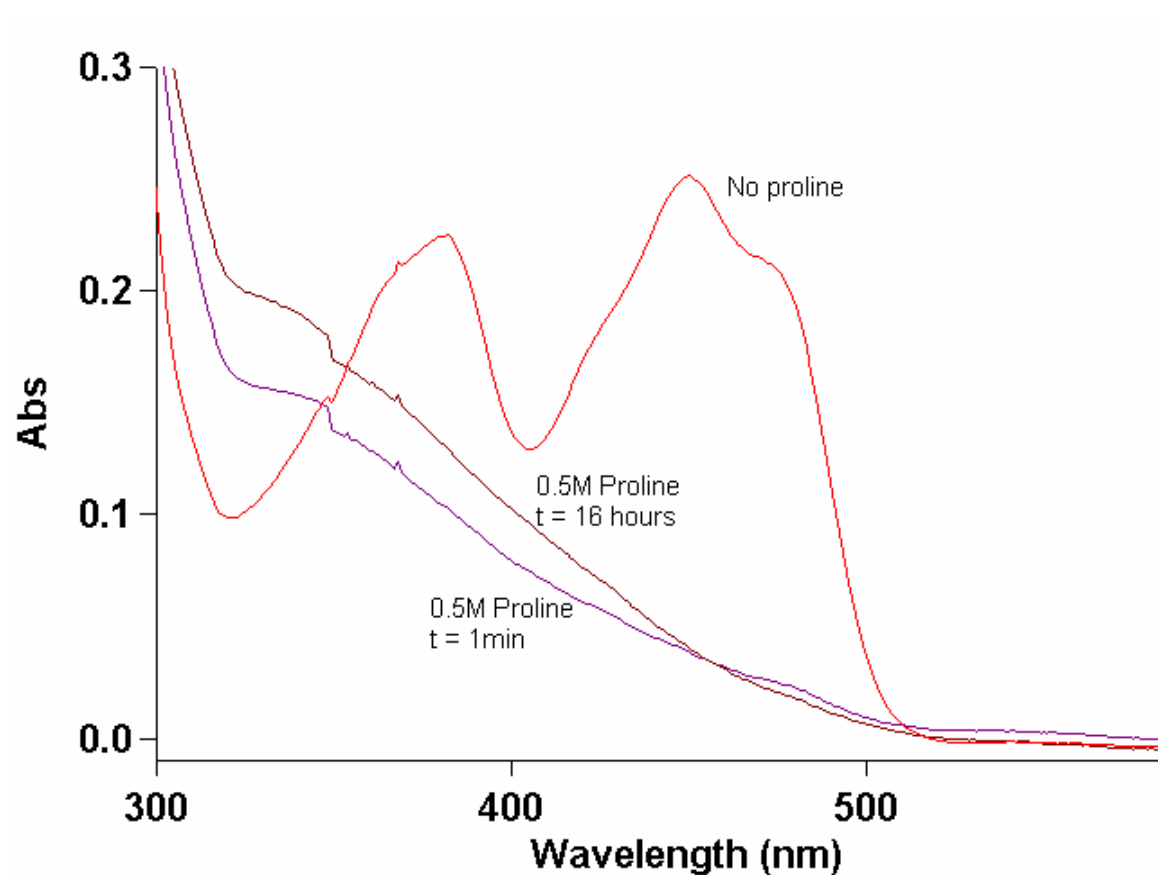
FIGURE 2.24.  High salt crystal form of TtPRODH (2 M Na K Tartrate, 100 mM Tris pH = 8.5, 0.5 % PEG 5000 MME, 4 % acetone (or 3% isopropanol) in reservoir and fresh 20 mM N-octyl β-D thioglucopyranoside was added to the 3 mg/mL protein).
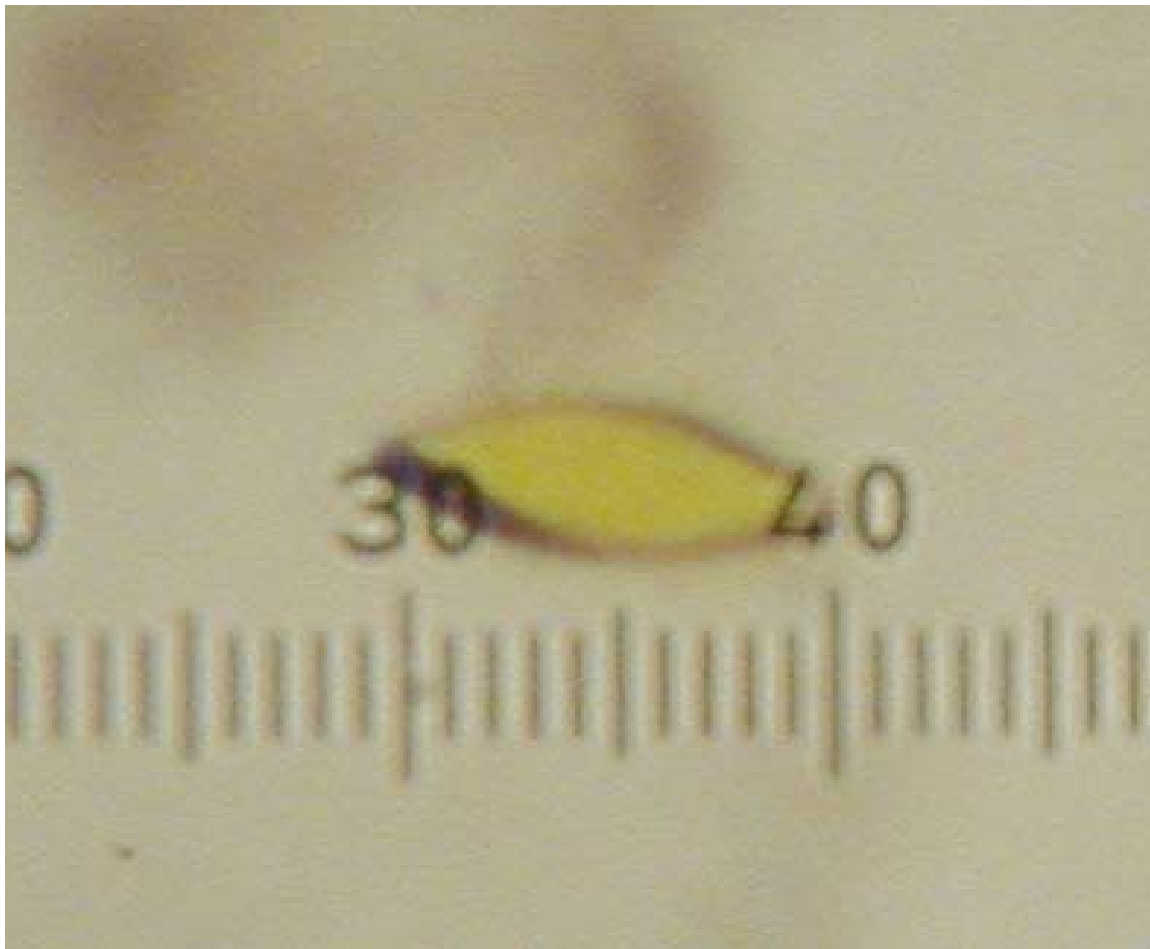
FIGURE 2.25. High salt crystal form of TtPRODH reduced with L-proline (*s*). (2 M Na K tartrate, 100 mM Tris pH = 8.5, 0.5 % PEG 5000 MME, 4 % acetone (or 3% isopropanol) in reservoir and fresh 20 mM N-octyl β-D thioglucopyranoside was added to the 3 mg/mL protein).
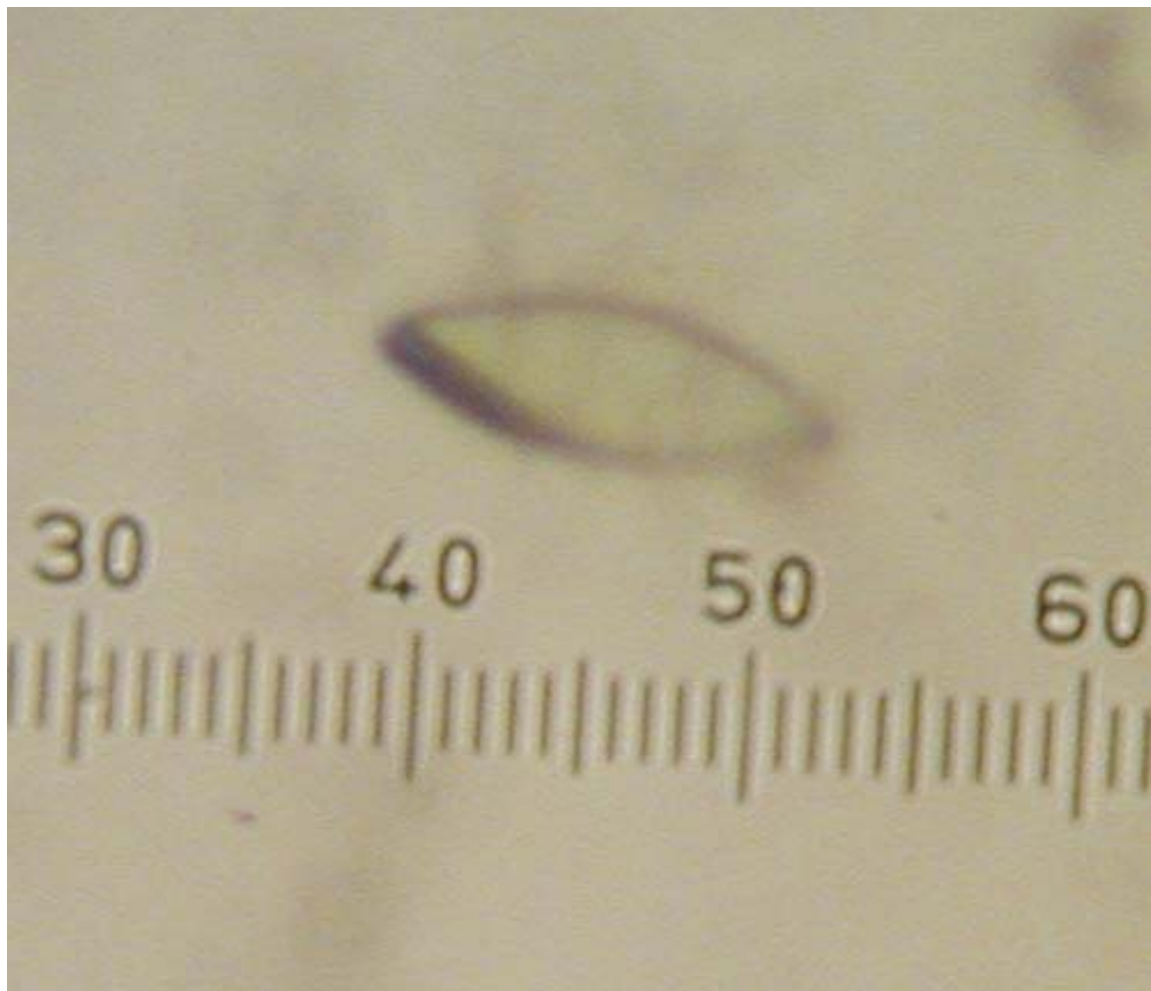
FIGURE 2.26. Potentiometric titration of TtPRODH (25 mM) at 20 °C (pH 7.5). Curves 1-11 correspond to fully oxidized, -0.041 V, -0.050 V, -0.063 V, -0.071 V, -0.08 V, -0.092 V, -0.098 V, -0.111 V, -0.118 V, and fully reduced, respectively. The inset is a Nernst plot of the potentiometric data, which yielded a midpoint potential of $E_m$ = -0.075 V with slope of 35 mV.
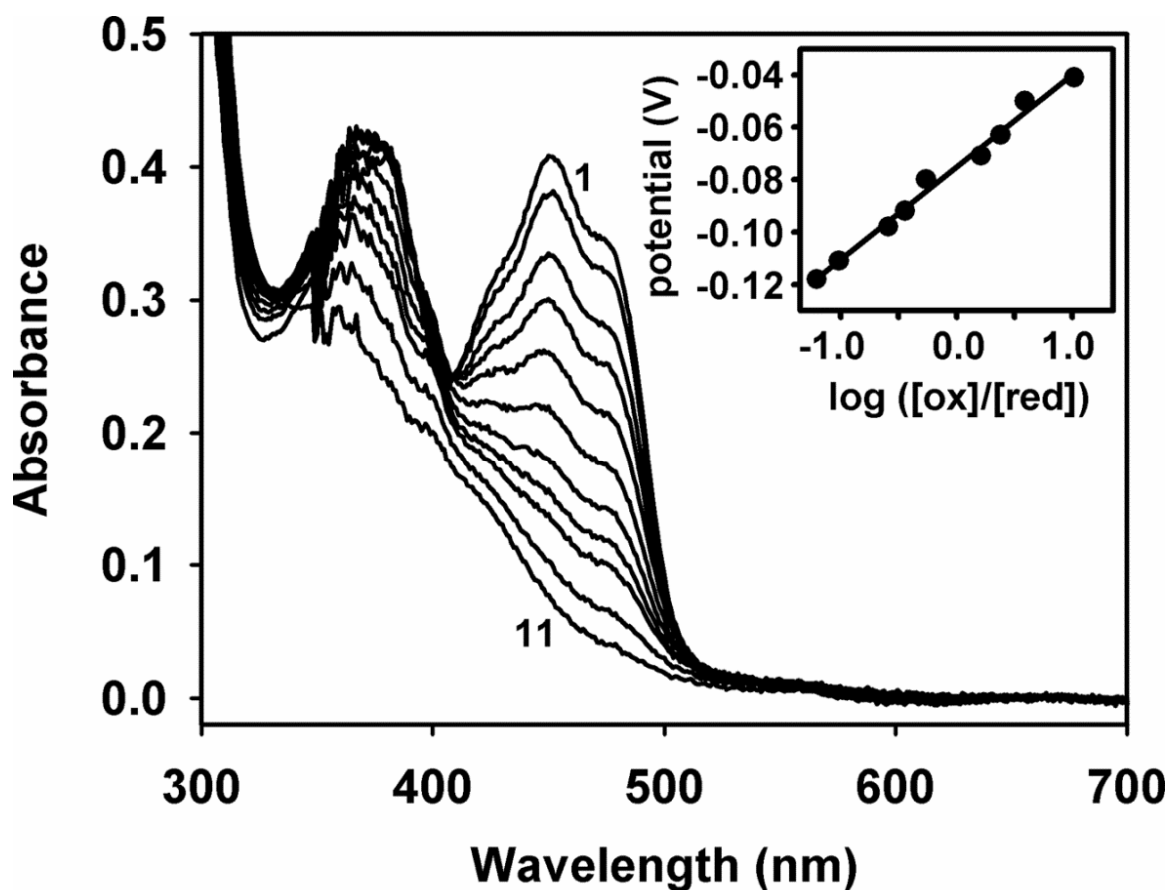
FIGURE 2.27. Steady-state kinetic parameter determination at 25$^o$C for TtPRODH catalysis of L-proline using DCPIP assay. Plot generated with Origin 7.0.



Data: Data1_SpActUmg
Model: Hyperbl
Equation:
y = P1*x/(P2 + x)
Weighting:
y          No weighting

Chi^2/DoF      = 0.15274
R^2      = 0.99652

| Vmax (U/mg) | 20.49529 | ±0.79222 |
| Km (M) | 0.02697 | ±0.00273 |
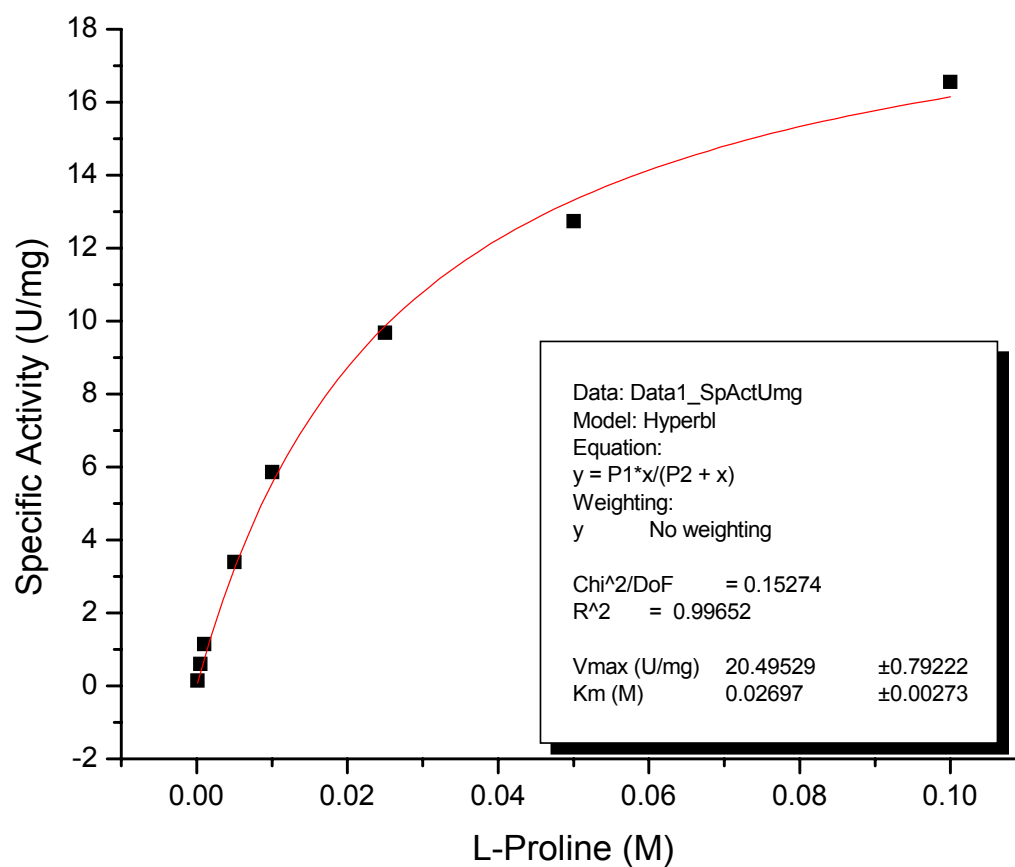
FIGURE 2.28. Steady-state kinetic parameter determination at 25$^{\circ}$C for TtPRODH catalysis of 3,4-dehydro-L-proline using DCPIP assay. Plot generated with Origin 7.0.
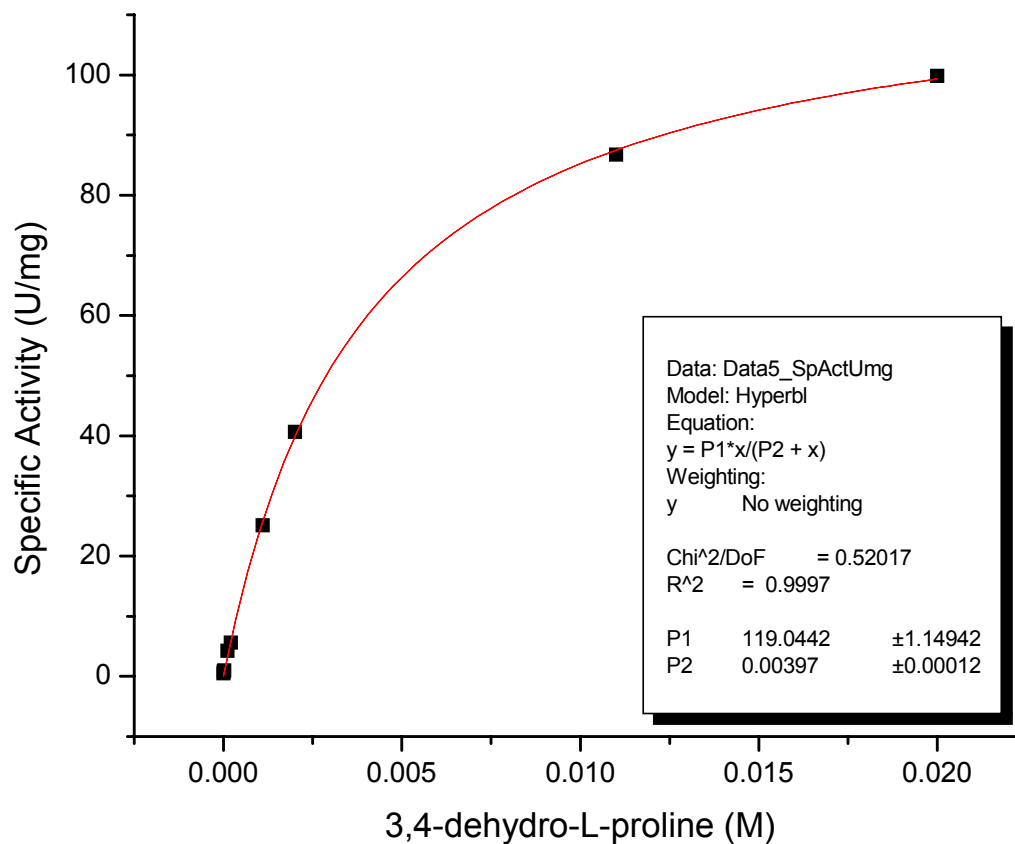
FIGURE 2.29. Determination of inhibition constant for L-Tetrahydrofuroic Acid. at 25$^o$C using DCPIP assay. Plot generated with Origin 7.0
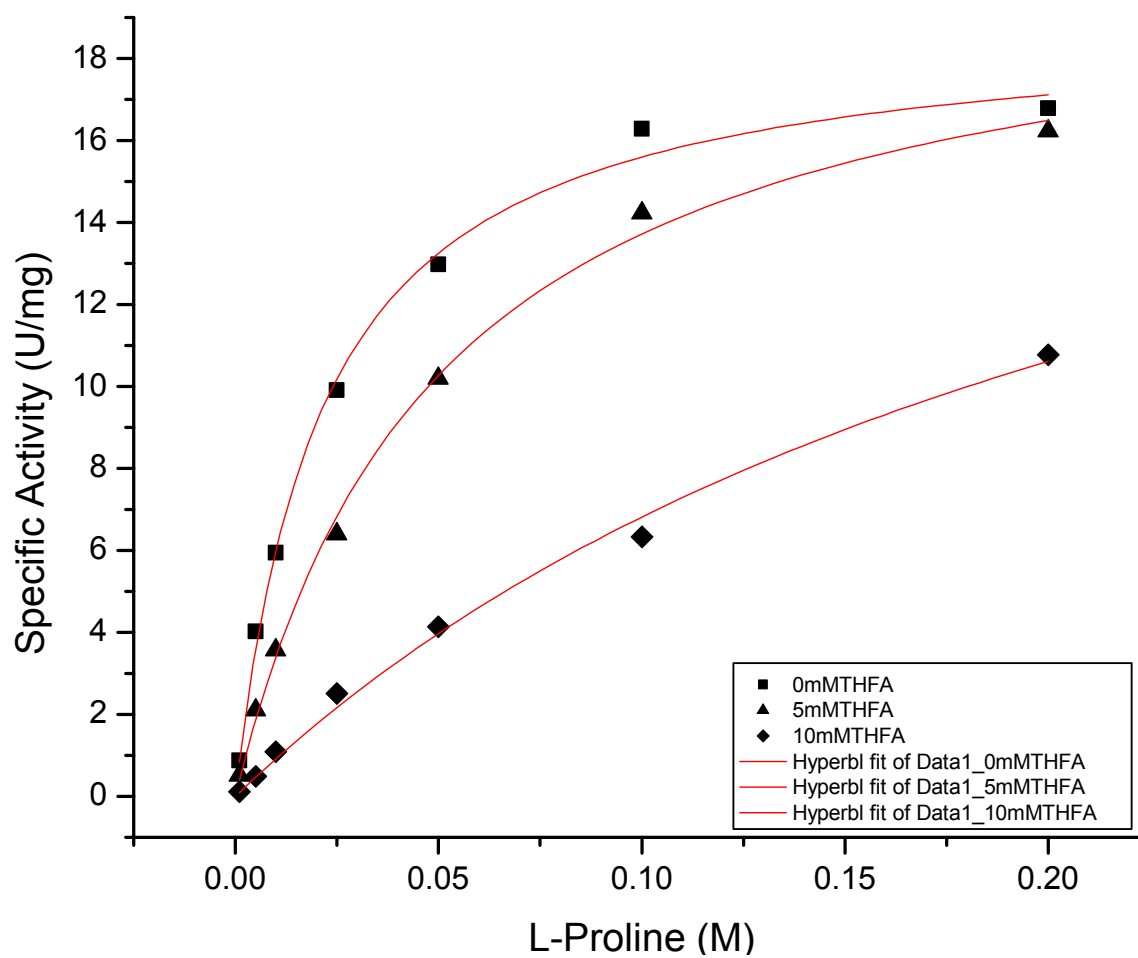
FIGURE 2.30. Determination of inhibition Constant of mandelate at 25$^{o}$C using DCPIP assay. Plot generated with Origin 7.0.
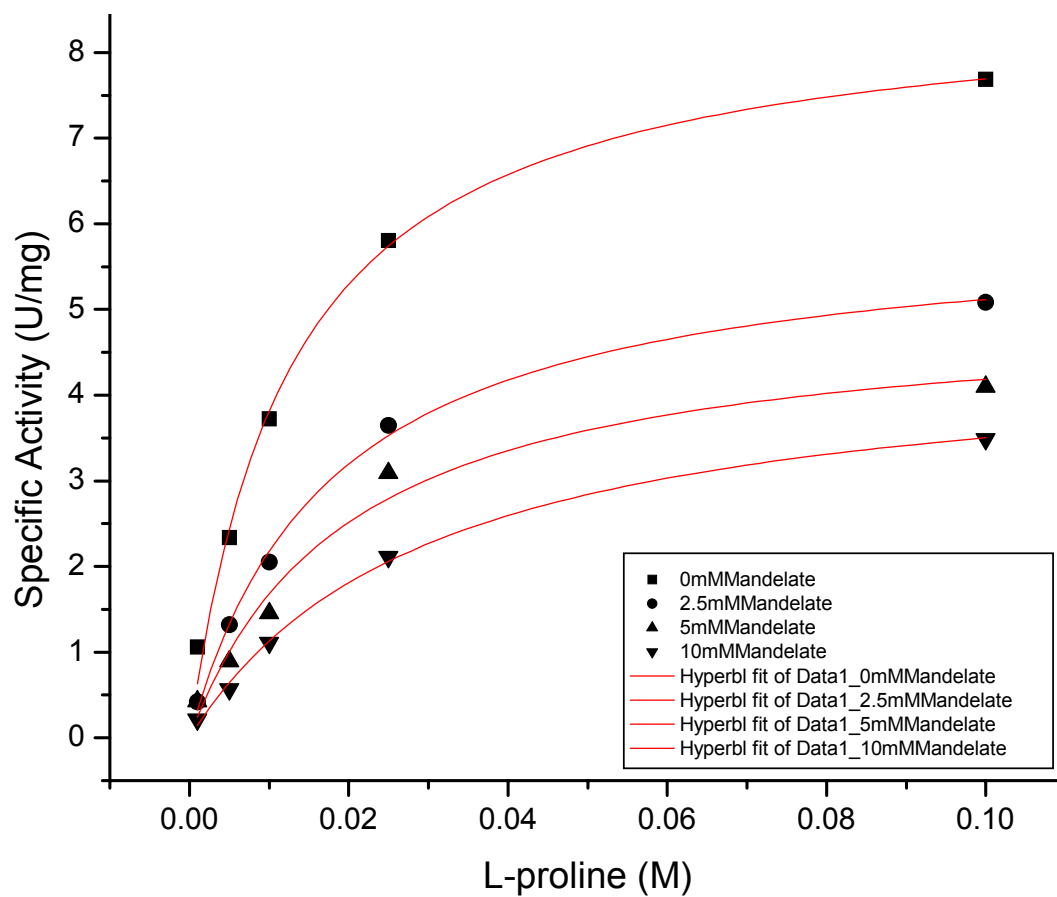
FIGURE 2.31.   Thermostability analysis of TtPRODH.   The percent activity remaining after incubation of the enzyme at 90 °C is plotted as a function of incubation time.
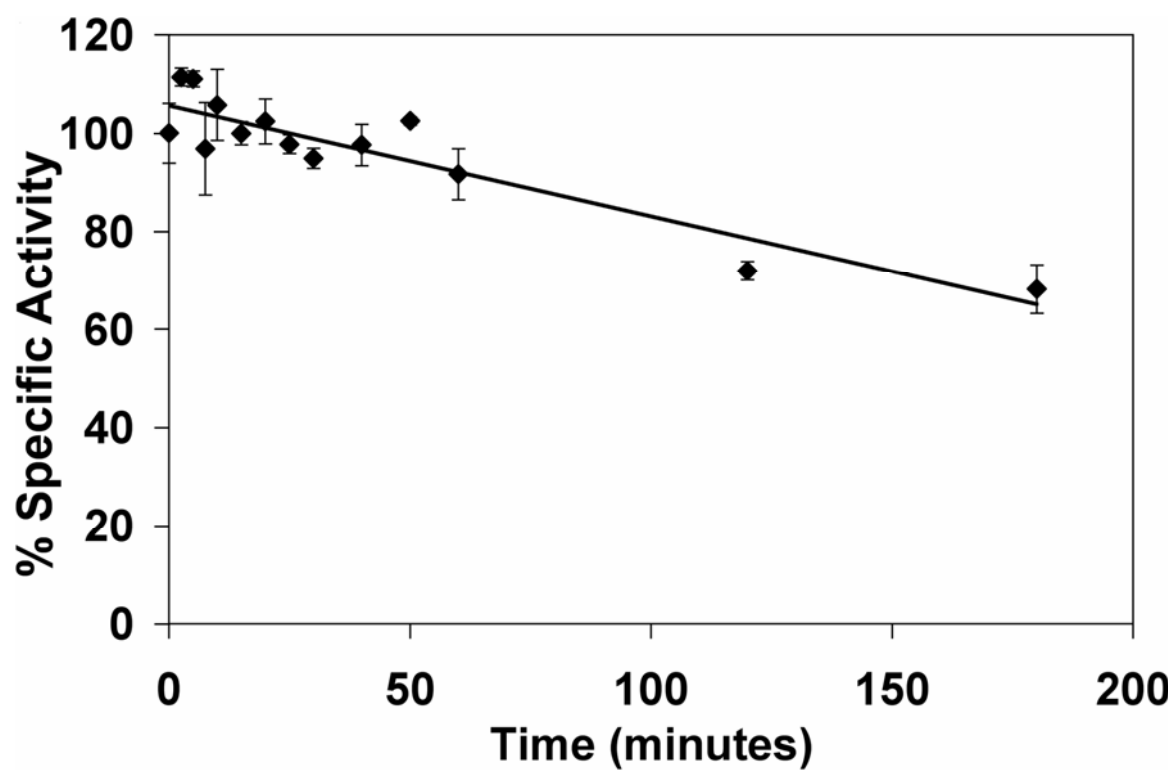
FIGURE 2.32.  TtPRODH specific activity as a function of temperature using the DCPIP assay.  Assay conditions were equilibrated to the temperature indicated before assaying for TtPRODH activity.
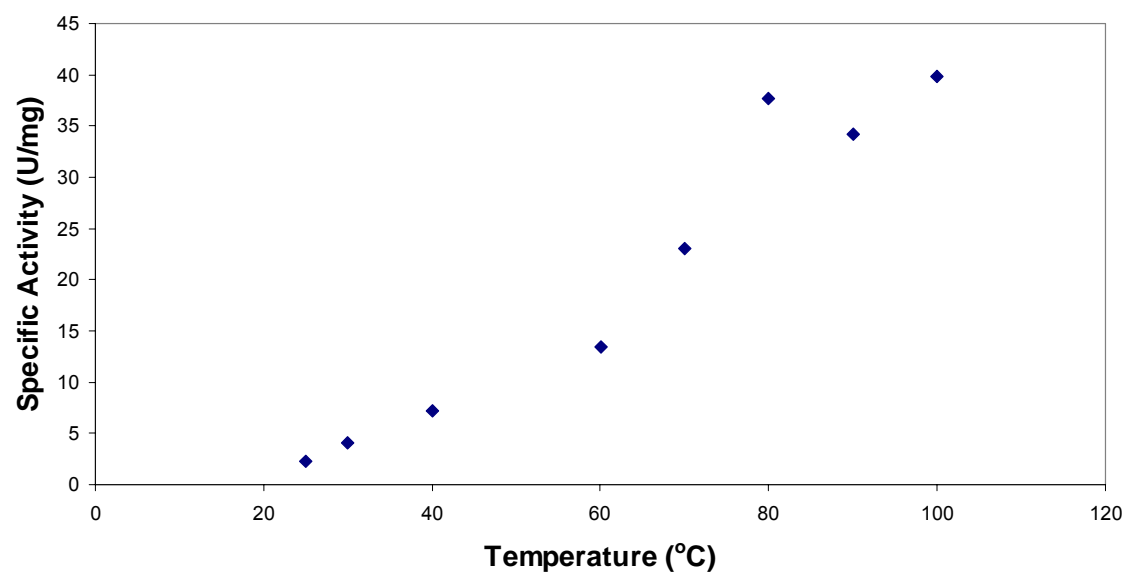
FIGURE 2.33. Steady-state kinetic parameter determination at 25$^o$C for TtPRODH oxygen reactivity utilizing *o*-aminobenzaldehyde. Plot generated with Origin 7.0.



Data: Data1_SpActUmg
Model: Hyperbl
Equation:
y = P1*x/(P2 + x)
Weighting:
y        No weighting

Chi^2/DoF     = 0.00004
R^2     = 0.99318

P1     0.3345 ±0.00411
P2     0.00131       ±0.00008
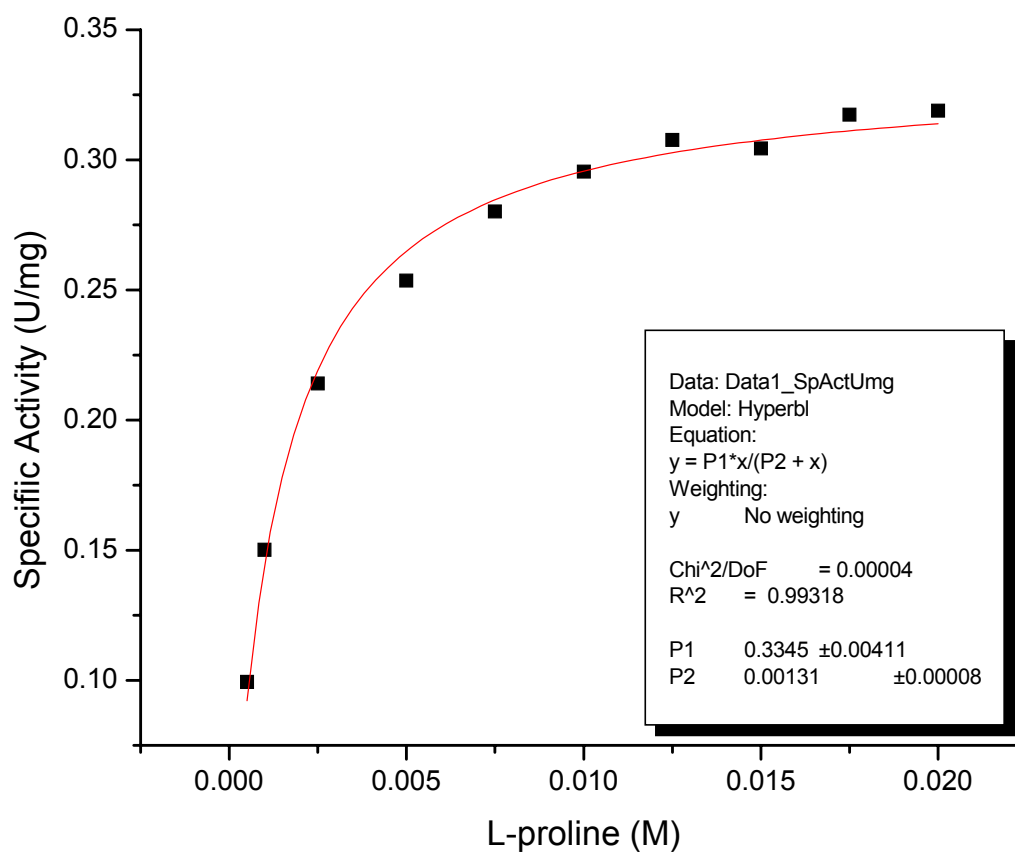
FIGURE 2.34.  Measuring P5C:*o*-aminobenzaldehyde complex formation at 443 nm.  The diamonds indicate complex formation in the absence of *E. coli* membrane vesicles.  The squares indicate complex formation in the presence of *E. coli* membrane vesicles.
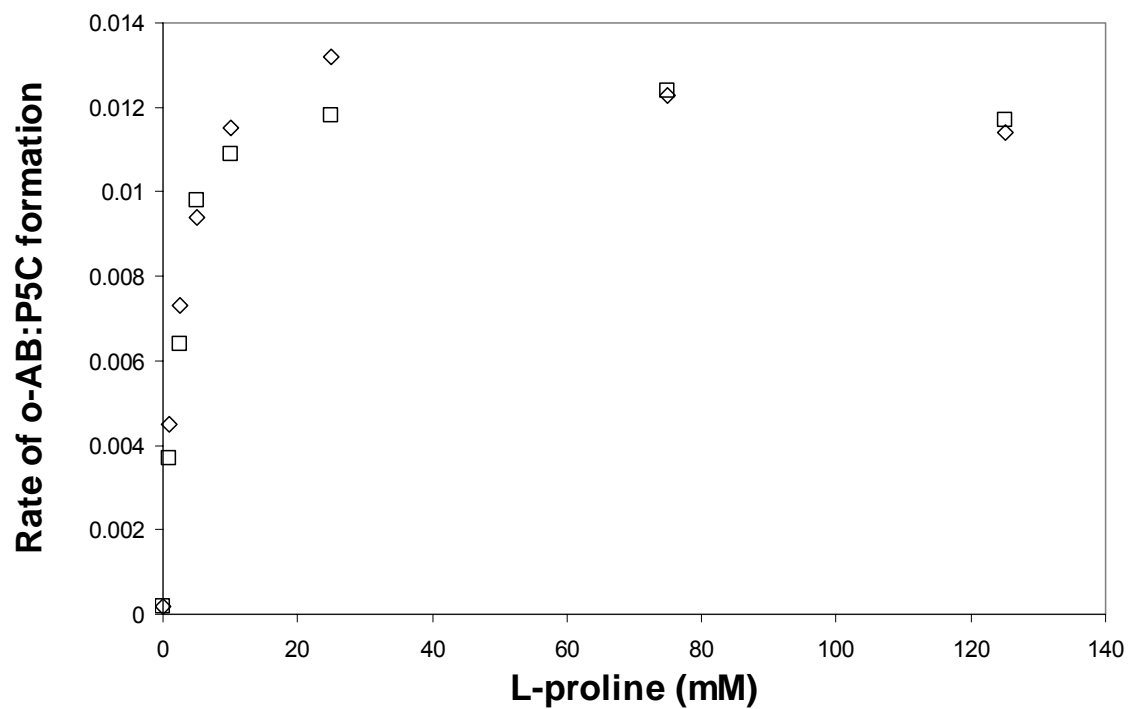
FIGURE 2.35. Generation of hydrogen peroxide by TtPRODH. Production of $H_2O_2$ was measured using the Amplex Red $H_2O_2$/peroxidase assay kit as described in the text.
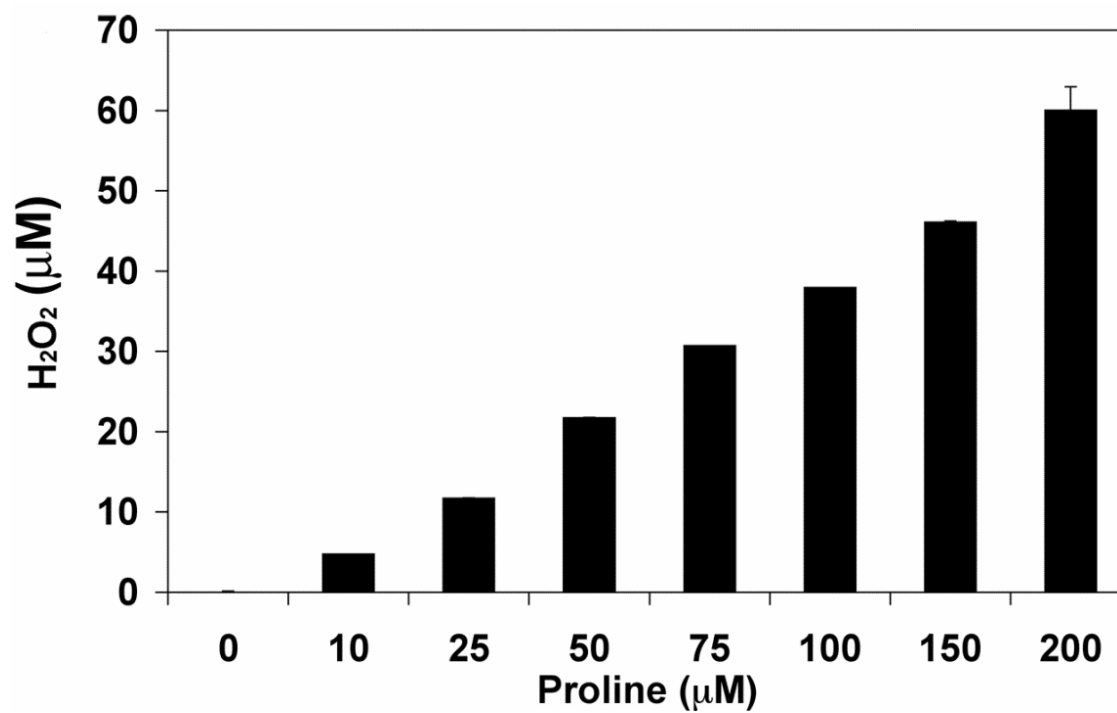
FIGURE 2.36. Unrooted phylogenetic tree representing the bacterial PutA/PRODH family generated with ClustalW.
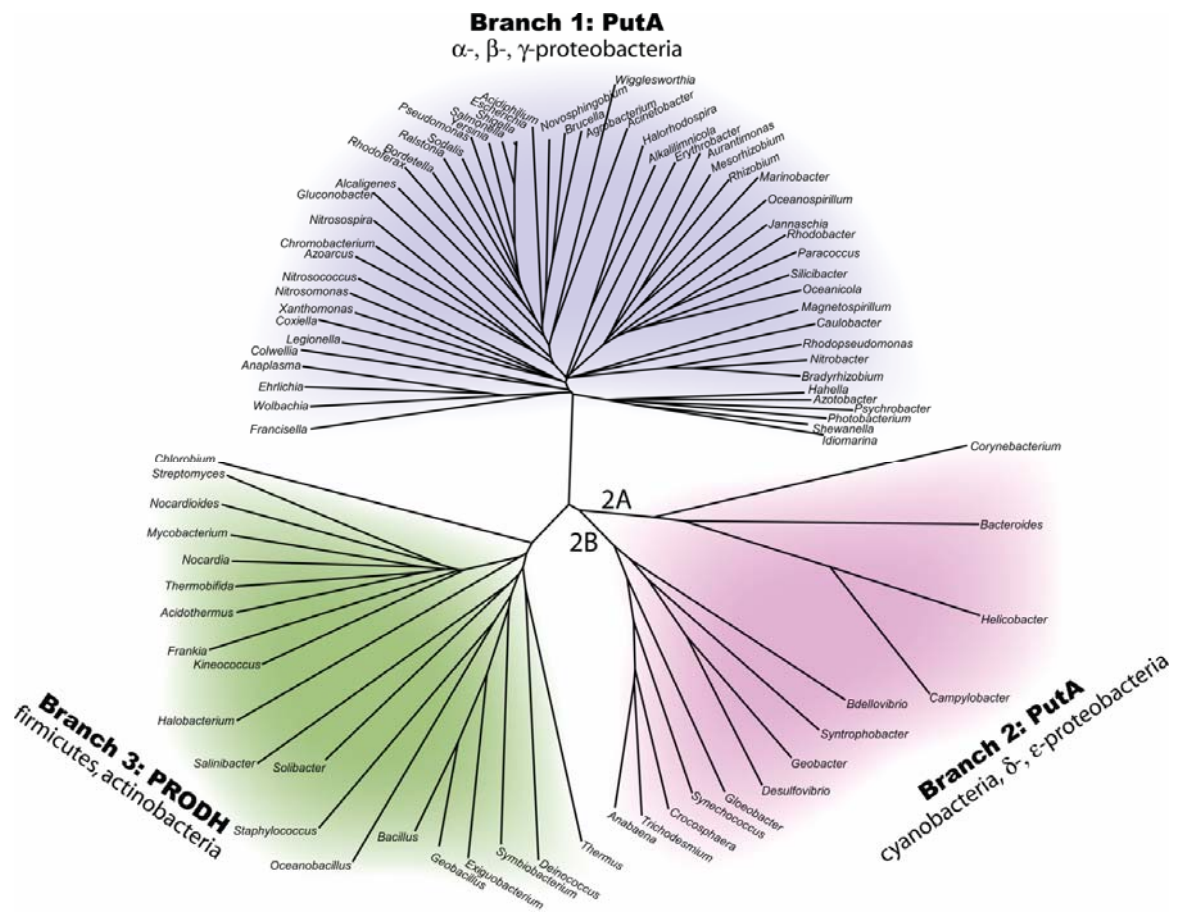
FIGURE 2.37. Conserved motif locations in TtPRODH. Each number corresponds to the conserved motif and each motifs conserved residues are highlighted in a different color.
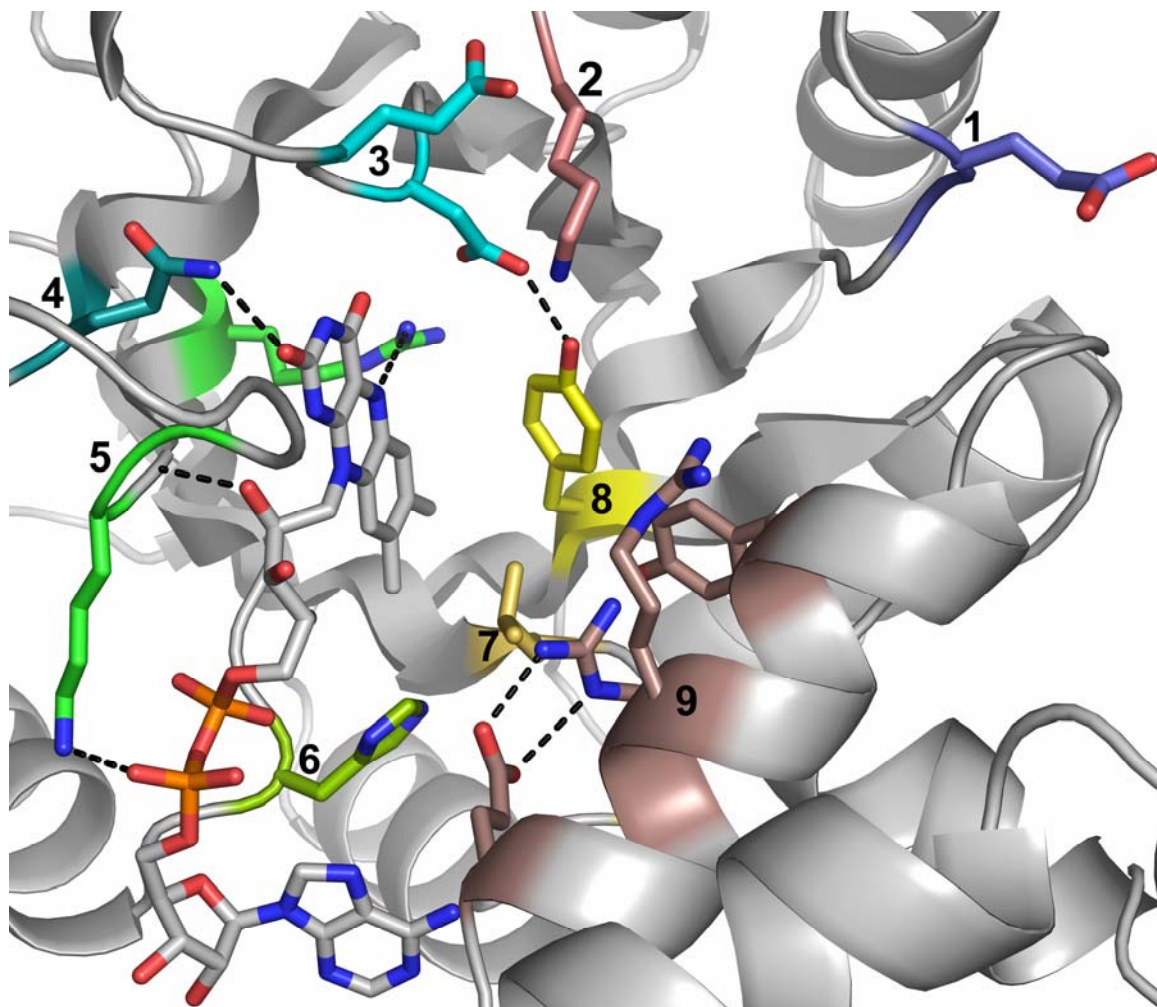
FIGURE 2.38. Archael L-proline dehydrogenase from *P. horikoshii*. FAD is shown in yellow CPK spheres, FMN is shown in green CPK spheres, and ATP is shown in magenta CPK spheres This heterooctameric complex $(\beta\alpha)_4$ has two domains. The $\beta$ domain (left) has proline dehydrogenase activity, contains FAD and has a monomeric sarcosine oxidase fold. The $\alpha$ domain has a unique fold with ATP-bound (right). FMN is bound at the interface of the two domains.
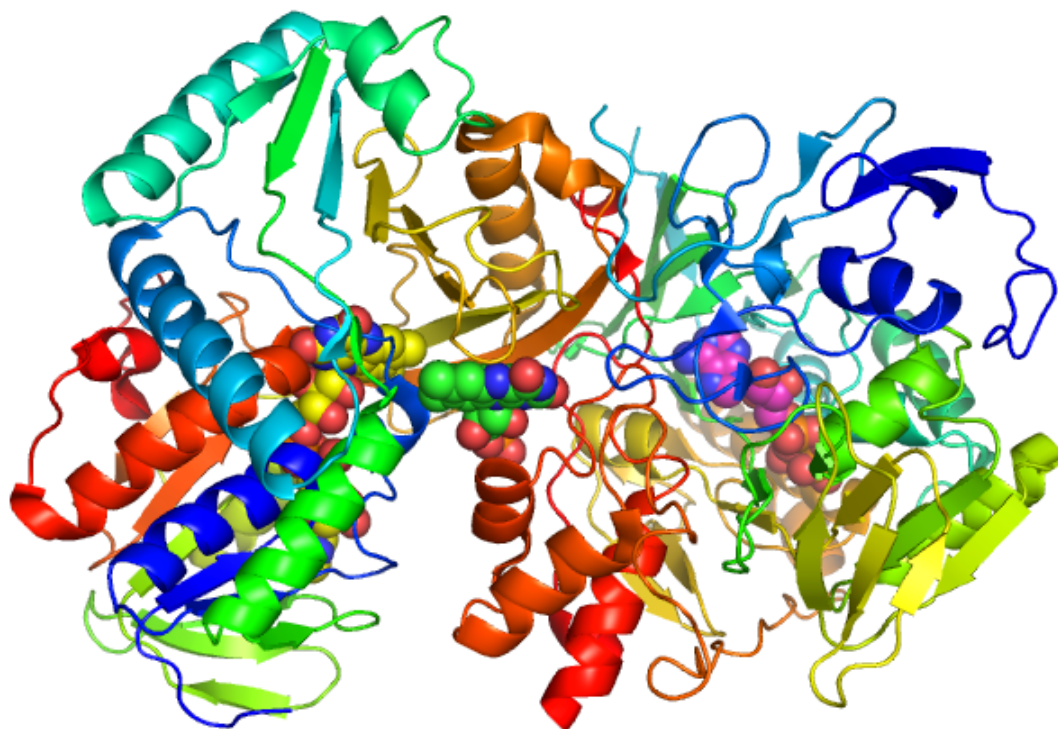
TABLE 2.1. Data Collection and Refinement Statistics[a]

| | |
|---|---|
| Wavelength (Å) | 0.97932 |
| Diffraction resolution (Å) | 47.16-2.00 |
| No. of observations | 292071 |
| No. of unique reflections | 46775 |
| Redundancy | 6.24 (3.12) |
| Completeness (%) | 98.0 (86.3) |
| $R_{merge}$ | 0.079 (0.399) |
| Average I/$\sigma$ | 13.6 (3.0) |
| Wilson B-factor (Å$^2$) | 16 |
| PDB accession code | 2G37 |
| No. of protein chains | 2 |
| No. of protein residues | 592 |
| No. of protein atoms | 5184 |
| No. of water molecules | 272 |
| $R_{cryst}$ | 0.190 (0.242) |
| $R_{free}$[b] | 0.231 (0.362) |
| RMSD[c] | |
|    Bond lengths (Å) | 0.014 |
|    Bond angles (deg.) | 1.5 |
| Ramachandran plot[d] | |
|    Favored (%) | 94.4 |
|    Allowed (%) | 5.4 |
|    Generously allowed (%) | 0.2 |
| Average B-factors (Å$^2$) | |
|    Protein | 16 |
|    Water | 16 |
|    FAD | 17 |

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]5% random $R_{free}$ test set.
[c]Compared to the Engh and Huber parameters (57).
[d]The Ramachandran plot was generated with PROCHECK (58).

**Table 2.2. Data Collection and Refinement Statistics of *Thermus thermophilus* Proline Dehydrogenase Structures**

| | I | II | III | IV | V |
|---|---|---|---|---|---|
| Name | Oxidized-Jan | THFA-March | Oxidized-Jun | THFA-Jun | Reduced-June |
| Soaks | nothing | THFA | Nothing | THFA | Proline |
| Precipitating Agent | 2-methyl-2,4-pentanediol | 2-methyl-2,4-pentanediol | Sodium Potassium Tartrate | Sodium Potassium Tartrate | Sodium Potassium Tartrate |
| Wavelength (Å) | 0.97932 | 0.9791 | 1.12718 | 1.12718 | 1.12718 |
| space group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| $a$ | 82.0669 | 82.0637 | 81.3705 | 81.917 | 81.5625 |
| $b$ | 89.5896 | 89.8929 | 88.5231 | 88.7067 | 88.423 |
| $c$ | 94.3157 | 94.79 | 91.2553 | 92.267 | 91.9451 |
| diffraction resolution (Å) | 47.16-2.00 | 39.42-1.80 | 45.63-2.30 | 40.93-2.20 | 45-97-2.30 |
| outer shell | 2.07 - 2.00 | 1.86 - 1.80 | 2.38 - 2.30 | 2.28 - 2.20 | 2.38 - 2.30 |
| no. of observations | 292071 | 397654 | 208978 | 120951 | 105038 |
| no. of unique reflections | 46775 | 64955 | 29833 | 34545 | 29153 |
| redundancy | 6.24 (3.12) | 6.12 (2.94) | 7.00 (6.89) | 3.5 (3.38) | 3.60 (3.65) |
| completeness (%) | 98.0 (86.3) | 99.0 (92.3) | 99.5 (98.4) | 99.1 (99.3) | 96.4 (94.7) |
| mean $I/\sigma_I$ | 13.6 (3.0) | 14 (2.9) | 14.5 (4.4) | 12.8 (3.2) | 10.2 (3.1) |
| $R_{merge}$ | 0.079 (0.399) | 0.074 (0.230) | 0.076 (0.410) | 0.045 (0.301) | 0.067 (0.370) |
| $R_{cryst}$ | 0.190 | 0.210 | 0.253 | 0.231 | 0.253 |
| $R_{free}$ | 0.231 | 0.244 | 0.307 | 0.280 | 0.312 |
| no. of protein residues | 592 | 594 | 540 | 554 | 537 |
| no. of protein atoms | 5184 | 4901 | 4384 | 4403 | 4465 |
| no. of water | 272 | 279 | 89 | 97 | 61 |
| RMSD bond lengths (Å) | 0.14 | 0.15 | 0.14 | 0.13 | 0.14 |
| RMSD bond angles (Å) | 1.5 | 1.7 | 1.6 | 1.6 | 1.7 |

Values for the outer resolution shell of data are given in parenthesis

Table 2.3. Data collection statistics for Se-Met SAD data sets.

| | thfa x4s1 | thfa x4s2 | thfa x1s1 | x3 |
|---|---|---|---|---|
| Resolution Limit | 2.2 | 2.2 | 2.7 | 2 |
| Exposure (sec) | 35 | 35 | 35 | 35 |
| Redundancy | 7.2 | 13.26 | 7.16 | 6.24 |
| Completeness | 99.8 | 99.9 | 100 | 98 |
| Average I/$\sigma$ [a] | 10.9 (21.4) | 11.8 (22.6) | 8.6 (15.0) | 13.6 (28.9) |
| $R_{merge}$ [a] | 0.111 (0.054) | 0.134 (0.068) | 0.166 (0.078) | 0.079 (0.043) |
| Se Sites found | - | - | - | 7/10 |
| SOLVE score | - | - | - | 0.24 |
| Solve FOM | - | - | - | 24.77 |
| Residues built | - | - | - | 73% |

[a] Parenthesis show values for low resolution bin

TABLE 2.4.  Dynamic Light Scattering results for TtPRODH in 20 mM n-octyl-β-D glucopyranoside and various controls.

**TtPRODH with 20 mM BOG**

|  | D (E-9cm^2/s) | R (nm) | MW (kDa) | Polyd (nm) | %Polyd | Polyd Index |
|---|---|---|---|---|---|---|
| Regu I Peak 1 | 1.44E+04 | 0.1341 | 0.02209 | 0.01935 | 14.4 | 0.02 |
| Regu I Peak 2 | 750.7 | 2.867 | 37.21 | 1.068 | 37.3 | 0.14 |
| Regu I Peak 3 | 0.3244 | 5853 | 3.99E+09 | 0 | 0 | 0 |

**Buffer**

|  | D (E-9cm^2/s) | R (nm) | MW (kDa) | Polyd (nm) | %Polyd | Polyd Index |
|---|---|---|---|---|---|---|
| Regu I Peak 1 | 2.38E+04 | 0.08361 | 0.007021 | 0.0187 | 22.4 | 0.05 |
| Regu I Peak 2 | 5.488 | 392.4 | 5.67E+06 | 149.6 | 38.1 | 0.15 |

**Buffer with 48 mM BOG**

|  | D (E-9cm^2/s) | R (nm) | MW (kDa) | Polyd (nm) | %Polyd | Polyd Index |
|---|---|---|---|---|---|---|
| Regu I Peak 1 | 2.84E+04 | 0.06697 | 0.004099 | 0 | 0 | 0 |
| Regu I Peak 2 | 879.2 | 2.341 | 22.76 | 0.6927 | 29.6 | 0.09 |
| Regu I Peak 3 | 46.78 | 42.68 | 2.61E+04 | 9.423 | 22.1 | 0.05 |
| Regu I Peak 4 | 0.3244 | 5853 | 3.99E+09 | 0 | 0 | 0 |

**BSA (1 mg/mL)**

|  | D (E-9cm^2/s) | R (nm) | MW (kDa) | Polyd (nm) | %Polyd | Polyd Index |
|---|---|---|---|---|---|---|
| Regu I Peak 1 | 2.86E+04 | 0.07343 | 0.005125 | 0 | 0 | 0 |
| Regu I Peak 2 | 601 | 3.553 | 62.63 | 0.4881 | 13.7 | 0.02 |
| Regu I Peak 3 | 0.3272 | 6418 | 4.99E+09 | 0 | 0 | 0 |

TABLE 2.5.  Kinetic parameters using L-proline as the substrate

| | TtPRODH | E. coli PutA[a] | H. pylori PutA[b] |
|---|---|---|---|
| proline:DCPIP assay | | | |
| $K_m$ (mM) | 27 | 100 | 146 |
| $V_{max}$ (U/mg of protein) | 20.5 | 5 | 3.6 |
| $k_{cat}$ (s$^{-1}$) | 13 | 12 | 8 |
| $k_{cat}/K_m$ (s$^{-1}$ M$^{-1}$) | 481 | 122 | 56 |
| $K_I$ for THFA (mM) | 1.0 | 0.2 | 0.35 |
| proline:$O_2$ assay | | | |
| $K_m$ (mM) | 1.3 | ND | 150 |
| $V_{max}$ (mU/mg of protein) | 335 | < 2 | 230 |
| $k_{cat}$ (min$^{-1}$) | 12.7 | <0.3 | 31 |
| DCPIP/$O_2$ activity ratio | 61 | > 2500 | 16 |

[a]Data obtained from Zhu et al. (42) and Krishnan and Becker (44).
[b]Data obtained from Krishnan and Becker (44).

Table 2.6.  Conserved Sequence Motifs of PutAs and Bacterial Monofunctional PRODHs[a]

| Motif | Branch 1[b] | Branch 2[c] | Branch 3[d] |
|---|---|---|---|
| 1 | s-y-D-m-L-**G**-**E** (288, 289) | v-d-l-l-**G**-**E** (146, 147) | l-d-X-l-**G**-**E** (64, 65) |
| 2 | i-S-i/v-**K**-l-S (329) | s-i/v-**K**-X-t/s (177) | S-i/v/l-**K**-X-s/t (99) |
| 3 | i-**D**-A-**E**-E (370, 372) | **D**-M-**E** (224, 226) | **D**-m-**E** (133, 135) |
| 4 | G-X-v-v-**Q**-a-y-q-k-R (404) | g-i-v-**Q**-A-Y-l-x-d (254) | G-X-v-l-**Q**-a-y-L (163) |
| 5 | **R**-L-v-**K**-**G**-A-Y-W-D-x-E-i-k (431, 434, 435) | **R**-l-V-**K**-**G**-a-y-w-d (285, 288, 289) | **R**-l-v-**K**-**G**-a-Y-X-E-p (184, 187, 188) |
| 6 | F-a-t-**H**-N (487) | i-a-s-**H**-N (342) | i-a-t-**H**-D (227) |
| 7 | E-f-Q-r-**L**-h-**G**-M-g-e (513, 515) | e-f-q-v-**L**-y-**G**-M-a (371, 373) | E-f-Q-m-**L**-y-**G**-i-r (254, 256) |
| 8 | i-**Y**-a-P-v-G (540) | l-**Y**-x-y-P-g (391) | v-**Y**-v-p-y-G (275) |
| 9 | L-L-a-**Y**-L-v-**R**-**R**-l-l-**E**-N-G (552, 555, 556, 559) | **Y**-L-v-**R**-**R**-l-l-**E**-N (406, 409, 410, 413) | **Y**-f-m-**R**-**R**-l-a-**E** (285, 288, 289, 292) |

[a]Letters in bold indicated amino acid residues that are identically conserved throughout the entire family of bacterial PutAs and PRODHs.  Upper case letters denote residues that are identically conserved within a branch of the family. Lower case letters indicate amino acid residues that are highly, but not identically, conserved within a branch.
[b]Numbers in parenthesis refer to *E. coli* PutA residue numbers.
[c]Numbers in parenthesis refer to *H. pylori* PutA residue numbers.
[d]Numbers in parenthesis refer to TtPRODH residue numbers.

APPENDIX 2.1. Purification protocol for pKA8H TtPRODH in BL21(DE3)pLysS

- Large scale purification (6 L)
  - 6 L LB media
  - 100 mg/mL Ampicillin
  - 34 mg/mL Chloramphenical
- Using 10 mL overnight cultures add to 1.5 L media
- Grow to $OD_{600}$ = 0.6
- Induce with 0.5 M IPTG for 3 hours.
- Harvest by centrifuging at 7000 rpm for 10 min at 4 $^{\circ}$C
- Resuspend harvested cell pellet in 50 mL:
  - 50 mM $NaH_2PO_4$, 300 mM NaCl, 10 mM Imidazole, 5% glycerol, pH = 8
- Freeze at -20 $^{\circ}$C overnight (or longer)
- Thaw
- Add protease inhibitors (AEBSF, TPCK, E-64, Leupeptin and Pepstatin)
- French Press at 16000 psi (2 passes)
- Centrifuge, collect supernatant, extract pellet with 20 mM BOG
- Filter with a 0.45 $\mu$m filter
- Affinity Chromatography
  - Wash with 3 column volumes of 50 mM Imidazole and another wash with 3 column volumes of 75 mM for Imidazole
  - Nickel-NTA resin and elute with Imidazole 300 mM
- Dialyze into following buffer overnight
  - 50mM Tris, 50 mM NaCl, 0.5mM EDTA, 0.5 mM DTT, 5% glycerol, pH = 8.0
  - Add 0.1 mM FAD to protein pre-dialysis
- Post dialysis - add 20 mM N-octyl β-D glucopyranoside
- Concentrate to 13 mg/mL
- Desalt to remove excess FAD and detergent

APPENDIX 2.2.  Sequences used to generate Unrooted Phylogenetic tree.

>Hahella Q2SL11 Q2SL11_HAHCH De
MFKASSVLSPDIQAVSLDSLWTAITQNYAVDEAAYLEELMALATPGTEELRAITQGATKL
IEDVRAQDDSVHMVDALLQEYSLDTQEGVLLMCLAEALMRIPDKATADALIRDKMSAAQW
DKHMGKSESTLVNASTWGLLLTGKVVKLDRNIDGTPANVLKRLINKCGEPVIRGAMNQAM
KIMGKQFVLGRDISEALQNGRKYRDKGYSYSFDMLGEAALTAEDAERYFQSYIKAIETVG
ADQYDTRPGASSISIKLSALHPRYEQAHQDRVLTEMYEKVLLLVRAARERNVSLTIDAEE
MDRLELSLRLFEKLYRSPDSRGWGEFGLVVQAYSKRALPVLCWLTALAAEQGDRIPVRLV
KGAYWDSEIKLCQQRGLPSYPVYTRKESTDVSYLACARFLFSEYARNHIYPQFATHNAHT
VASVQALAEKNDRPFEFQRLHGMGDALYNALLAKQKRTVRIYAPVGAHKDLLPYLVRRLL
ENGANSSFVHRLVDAETPIDSLVQHPVHELTRYSSLANHRIPLPAQIFGNRINSMGVNLF
VENQYSPLEKQLRAWDQHQWKACPVINGEHRTLSAPNPVLSPYDQSQSAGTVYWSQSNDV
DDALSAAYVGFEKWNNTPVEKRAQSLEKLADLMEAGMPELMALCCREAGKTLQDSIDEVR
EAVDFCRYYAEQARKHFSKPNVLPGPTGESNELYLEGKGVFVCISPWNFPLAIFIGQVTA
ALAAGNAVIAKPAEQTSLIAARAIDMMLEAGVAKEAIQFLPGDGAQLGPQLLSDNRVCGV
AFTGSTQTAHIINRSLAARDGAIATLIAETGGQNAMIVDSTALPEQVVKDVVQSAFASAG
QRCSALRVLFVQKDIADGILELLSGALKELKVDHPKHLSTDVGPVIDAEAQAGLLKHIEE
MKANSRWNAEAILPGDHNKGFFVTPSAFEIGSINELSKEHFGPILHIVRYAAEDLDKVID
SINDTGFGLTLGIHSRNETTAAYIEKRVKVGNTYINRNQIGAVVGVQPFGGHGLSGTGPK
AGGPHYLLRFANERTRTINTTAVGGNATLLSLGVEHI
>Shewanella Q3QL02 Q3QL02_9GAMM
MFKASEVLAGRYDNANLDELFSLISQNYIVDEEAYLKELIALVPSSDEEIARITSRAHDL
VAKVRQYEKKGLMVGIDAFLQQYSLETQEGIILMCLAEALLRIPDAETADALIADKLSGA
KWDEHMSKSDSVLVNASTWGLMLTGKIVQLDKNLDGTPSNLLSRLVNRLGEPVIRQAMYA
AMKIMGKQFVLGRTIEEGLKNAAEKRKLGYTHSYDMLGEAALTMKDADKYYRDYANAIQA
LGTAKFDESEAPRPTISIKLSALHPRYEVANEDRVMTELYATLIKLIEQARSLNVGIQID
AEEVDRLELSLKLFKKLYQSDAAKGWGLLGIVVQAYSKRALPVLMWLTRLAKEQGDEIPL
RLVKGAYWDSELKWAQQAGEAGYPLFTRKAATDVSYLACARYLLSEATRGVIYPQFASHN
AQTVAAITAMVGDRKFEFQRLHGMGQELYDTVLAEAAVPTVRIYAPIGAHKDLLPYLVRR
LLENGANTSFVHKLVDPKTPIESLVTHPLKTLQGYKTLANNKIVKPADIFGAERKNSKGL
NMNIISESEPFFAALEKFKDTQWSAGPLVNGETLSGEVRDVVSPYNTTLKVGQVAFANEA
TIEQAIAGADKAFASWCRTPVETRANALQKLADLLEENREELIALCTREAGKSIQDGIDE
VREAVDFCRYYAVQAKKMMSKPELLPGPTGELNELFLQGRGVFVCISPWNFPLAIFLGQV
AAALATGNTVIAKPAEQTCLIGFRAVQLAHEAGIPKDVLQFLPGTGAVVGAKLTSDERIG
GVCFTGSTTTAKVINRALAGRDGAIIPLIAETGGQNAMVVDSTSQPEQVVNDVVSSAFTS
AGQRCSALRVLYLQEDIAERVLDVLKGAMDELTLGNPGSVKTDVGPVIDAAAKANLNAHI
DHIKQVGRLIHQLSLPEGTENGHFVAPTAVEIDSIKVLTKENFGPILHVVRYKAAGLQKV
IDDINSTGFGLTLGIHSRNEGHALEVADKVNVGNVYINRNQIGAVVGVQPFGGQGLSGTG
PKAGGPHYLTRFVTEKTRTNNITAIGGNATLLSLGDAD
>Aurantimonas Q1YLD6 Q1YLD6_9RH
MTPETTAALRGTGDRDAVRQRYGEDEAIVVKDLAARISLSETDRAAISAAGADIVRRVRR
ETSPSMMESFLAEYGLSTEEGVGLMCLAEALLRVPDAETVDELIQDKIEPSDWGAHLGHS
SSSLVNASTWALMLTGRILDDDPSRPAAALRGLVRRVGEPVVRRAVAQSMRLLGRQFVLG
QTIEEGMRNARERERQGYTYSYDMLGEAARTEADALRYLKSYSDAIASIAKAAKGDVRSN
PGISVKLSALHPRYETTHRDTVMATLVPRALELARQAAKANIGFNIDAEEQDRLDLSLDI
IEAVLSDESLKGWDGFGVVVQAYGRRAVPVIEYLDTLARRLDRRIMVRLVKGAYWDTEIK
LAQEMGTHTFPVFTRKVNTDLSYMACAQMLLDRRDRIYPQFATHNAHTCAAILKMAGNDR
DSFEFQRLHGMGESLHDIVHRQNDTRCRIYAPVGAHRDLLAYLVRRLLENGANSSFVNQI
VNEAIAPENIARDPLSEVEALGEAIANPTIRTPAELFAPERRNSKGFRVNEPASITPLIE
ARQAFADAVWTARPMLADDLPVSGEGRAATSPADPGRRVGTVHEASVKDVHRALAAAEAG
FAAWSAVPVAERATTLRAIADLYEANIAELTVIATREAGKTLADGIAEIREAVDFLRYYA
DEAERLEVEEPGTPRGIFVCISPWNFPLAIFTGQIAAALAAGNAVLAKPAEQTPLIATRA
VELMREAGLPAAALQLLPGDGPNVGGLLTSDRRIAGVCFTGSTEVAQIIHKALAANAGPD
AVLIAETGGLNAMIVDSTALTEQAVRDILIASFQSAGQRCSALRMLYVQEDARDRLLHML

111

EGAMDALVVGDPWRIDTDVSPVIDADARDDIAGYIASRKAAGQALKTLPVPQDGLYAAPS
VVSVGGIDDLEREVFGPVLHVATFKAREIDGVVDAINARGYGLTFGLHTRIDDRVQQIVE
RIQVGNTYVNRNQIGAIVGSQPFGGEGLSGTGPKAGGPFYVARFRRTEPAEDLPAPAPPD
IPAATVAEAIARLDARNWAARGDRVAVLRRALSGKGGIIRRACSAVGAFDPGPFTLPGPT
GESNRLSLYPRGTVLCLGPTAEIALAQAVQALSAGCAVIIVAPGAAALAGPVIAAGAPLA
VFDGTLAPGDLEHLDGIDVVAAAGRSDWTWALRNALAARKGAIIPLETAIIGPSRYVVER
HLCIDTTAAGGNASLLAAAE
>Rhodobacter Q52711 Q52711_RHOC
MTDLSALGPKAKFAPEAEVLQALVAQAALPQPQLDRIAARGADLVARIRAEAKPSLMEHF
LAQYGLSTREGVALMCLAEAMLRVPDTATIDALIEDKIAPSDWGKHLGTAASSLVNASTW
ALMLTGKVLDDGAGGIAGTLRGAMRRLGEPVIRAAVGQAMREMGRQFVLGETIEKALERA
EKREAEGYTFSYDMLGEAALTAADAERYRLAYAQAITAIGKAATRGSIAANPGISIKLSA
LHPRYEVAQEARVMAELVPVVRDLARAAARAGIALHIDAEEQDRLALSLRVMAAVIADPE
TAGWEGFGAVVQAYGKRAGAAIDALAAMARAAGRRINIRLVKGAYWDAEMKRAQVEGHPG
FPLFTSKTGTDVAYICLAAKLFGLNDCIYPQFATHNAHTVAAVLEMAAGRPFEFQRLHGM
GARLHDIVLRETGGRCRIYAPVGAHRDLLAYLVRRLLENGANSSFVNQIVNESVPPAEVA
ACPFAALPTARAPRGLLAPADLFGAGRVNAQGFDLSDPEVLARIEAARDVTLPDAAPIVA
GPVSGTLRPVRNPATGAVVAQVTEADAATVALALDAAQVWSAPAATRAAVLCRAADLYEE
NFGPIFAALAQEAGKTLGDAVSELREAVDFLRYYAAEGAADTRPPRGAVVAISPWNFPLA
IFTGQVAAALMAGNAVLAKPAEQTPIIAALAVRLLHQAGVPETALQLLPGDGPTVGAALT
RDPRVAGVVFTGSTETAQIIARAMAAHLAPGTPLIAETGGLNAMVVDSTALPEQAVRDVV
ASAFRSAGQRCSALRCLYVQDDIAPHLIGMLKGAMEELVSGDPARLSTDVGPVIDAEAKA
GIETYLAANKARILHRSTAPEGGHFVAPALLQVGGIADLEREIFGPVLHLATFAAEDLPA
VIAAINARGYGLTFGLHSRIDARVETVAETIRAGNIYVNRNQIGAVVGSQPFGGEGLSGT
GPKAGGPLYLNRFYAPEPVVAVGGWTEAATPILPEARETQLDEIFLPGPTGELNRLTRHQ
RGPILCLGPGAEAASAQAAAVVALGGQAVQASGAVSPKALETLTPLAGVLWWGAAEMGRA
YAQALAVRPGPLVPLITAKPDLAHVAHERHLCVDTTAAGGNAALLAG
>Azotobacter Q4J680 Q4J680_AZOV
MFKASRVLQGNVLDTSAGEFFPFISANYCVDEAAWLNELLPLADPGEAGIAAIRERAGAL
IEAVRRRGNVEVTLDALLREYSLDTQEGLMLMCLAEALLRVPDPATAEALIRDKLSAAQW
AQHLGHSDNLLVNFAAWGLLMTGRLVSPESSDGRPKKVIGRLLQRSGEPVIRAALNQAMK
LMGNQFVLGRNIAEALRNARRARERGYGHSFDMLGEAALTEAAAERFLTDYRLAIEALGR
EPQVGGGPRPSISIKLSALHPRYEAAQRRRVLAELFASVRELAELARSLNVGITIDAEEA
DRLELSLELFEKLLCDPAIRGWGEFGLVVQAYSKRALPVLVWLTLLGRELDTRIPVRLVK
GAYWDSEIKQCQVQGLEGYPVFTRKEGTDTSYLACARYLLGEHCRGVLYPQFATHNAHTV
SCILAMAETHPPHDFEFQRLYGMGDALYDCLLERQQVQVRIYAPVGEHRELLPYLVRRLL
ENGANSSFVHRLVDPRIPVEALIGHPVEQLRRCSGLANPRIPLPVDIHGAGRKNSRGINL
NVRSQWEPFERALRVQLEHRWQAAPLVDGQVLAGEVHEVRSPQDLDRVVGTVWFAGAELV
DRAMARLAAAWPRWNATPMERRAAIFGRLADLLEEHRTELVALCILEAGKTIQDSLDEVR
EAVDFCRYYAQQARLTLKRTELRGPTGERNELFYEGRGLFVCISPWNFPLAIYLGQIVAA
LVAGNCVLAKPAEQTSLIAARALELLFAAGLPKEAIAFLPGDGPALGAACCADPHLAGVC
FTGSTETARLINRRLAWRDGPLVTLIAETGGQNVMIVDSTALPEQVVKDALHSAFTSAGQ
RCSALRVLYLQNEIAGPVIELLRGAMAELRVGLPQRRDTDVGPLIDTQARQALLEHVDLL
KGEERLLAEAGLAPALNGHFLAPLAFEIGGIGELRKEHFGPILHVVRYAAEDLEQVVAAI
NASGYGLTLGVHSRNEATARRIEALTRVGNLYVNRNQIGAVVGVQPFGGCGLSGTGPKAG
GPNYLLRFVHERSTAINTAAVGGNASLLSLTERE
>Chromobacterium Q7NXT8 Q7NXT8_
MQFESFAHQQSALRDAITAAYRRDERECVQALLPQAAMSPEQVASVQDLARRLVTEVRRE
RTRSSGVDALMHEFSLDSSEGIALMCLAEALLRIPDRETADKLIRDKISRGDWKAHLGNS
SSLFVNAAAWGLLVTGKLVSSHSANGLSAAMTRLIAKGGEPLIRKGVDMAMRMLGKQFVT
GETIEEALANGREREARGYRFSYDMLGEAAMTEADAQRYLKDYVTAIHAIGKESNGRGIY
DGPGISVKLSAIHPRYARLKHERMMTELLPRLKALFLLAKQYNIGLNIDAEEEADRLEISM
DLVEALANDADLNGFEGIGIVVQAYQKRCPFVIDFLIDLARRTGHRFMVRLVKGAYWDAE
IKRAQVDGLPGYPVYTRKVYTDVSYLACAKKLLAAQDAIYPQFATHNAYSLSAIYNLAAG
KDYEFQCLHGMGETLYDQVVGKDKLGKACRIYAPVGSHETLLAYLVRRLLENGANSSFVN
RIVDENVSIDELVTDPVAEAAGFSGMPHSKIPLPEALYGEGRRNSKGLDLSSEHVLATLQ
LGLQASEQQRWAAFPMLGDGDVTDGELQEVRNPADHSDVVGKVIEASAADVERALALSAG

```
IAAAWAATPVAERAASIRRMADLMEAHMPALMGLAVREAGKTLNNAIAEVREAVDFCRYY
AAQIVSEFDNASHRPLGPVVCISPWNFPLAIFIGEVVASLAAGNTVLAKPAEQTSLIAAY
AVRLLHEAGIPRAALQFLPGRGEVVGAALTGDARIQGVIFTGSTEVAQIINRTLAKRQDD
PVLVAETGGMNAMIVDSSALPEQVVTDVLSSAFDSAGQRCSALRVLYLQNDIADKVIAMI
KGAMDELTIGNPAKLTTDVGPVIDAEAQAGLLAHIARMKNSARAMHQTKLSAACEQGTFV
APTLFEIDNLSELKREVFGPVLHVLRYAASDLDKVVAEINATGYGLTHGIHSRIDETIAD
ICGKIKVGNIYVNRNIVGAVVGVQPFGGEGKSGTGPKAGGPYYLYRLTRAAWQPKLAAVP
AAADLSALDALAAAAKAQSLALDGAIAAARKESPLTHSVALPGPTGENNLLSFAGRGRIG
CVADDAQALAEQLAAAFAAGNRAVLADNELGRKFASALNGHVSLAADVLEADVDAVLYAG
AKAEEARRELAARDGALVPLILRGENGYNVHRLVVERALSVNTTAAGGNASLMSMTE
>Silicibacter Q5LP42 Q5LP42_SIL
MTHALRTHIDSQTYADQSALLDQLVAQAALSEVDRAAICAAAAGLVRDIRSSTAPGLMEV
FLAEYGLSTDEGIALMCLAEALLRVPDADTIDALIEDKIAPSDWGKHLGHSSSSLVNAST
WALMLTGKVLDERKASPVGALRGAIKRLGEPVIRTAVSRAMKEMGRQFVLGETIQSAMKR
ASGMEAKGYTYSYDMLGEAARTEADAARYHLSYSRAIAAIAEACTHGDIRANPGISVKLS
ALHPRYELAQDARVMDELVPRLRALALLAKAAGMGLNVDAEEADRLSLSLQVIEAVMAEP
ALKGWDGFGIVVQAYGPRAGLVIDTLYEMAQRHDRKLMVRLVKGAYWDTEVKRAQVEGVD
GFPVFTQKSATDVSYIANARKLLSMTDRIYPQFATHNAHTVAAILHMAKDGQPYEFQRLH
GMGETLHNLVLTANKTRCRIYAPVGAHRDLLAYLVRRLLENGANSSFVNQIVDENVPPEV
VAADPFETVKAPAKPLKRGPELFAPERPNSMGFDLGHQPTLDAIDAARDPFRSHRWQAGP
LLAEEASPEAEEAVVNPADPHDVPGSVAPASAADAELSLARAQPWQAPAAERAAVLNRAA
DLYEAHYGELFALLHREAGKTLPDAVAELREAVDFLRYYAANIPDAASAGIFTCISPWNF
PLAIFSGQIAAALATGNAVLAKPAPQTPLIAHRAVQLLHEAGVPRDALQLLPGGPAVGAA
LTSDPRVSGVAFTGSTATAQKIRAAMAANMRPGTPLIAETGGLNAMIVDSTALPEQAVQA
VIESAFQSAGQRCSALRCLYLQEDIAEDVLTMLKGAMDALNLGNPWALSTDSGPVIDEAA
RKIITDHVAGARAEGRVLKELALPPEGTFVAPTLIKVAGIGALEREIFGPVLHVATFRSN
ELDQVIAAINATGYGLTFGLQTRIDDRVQHVTEAVHAGNIYVNRNQIGAIVGSQPFGGEG
LSGTGPKAGGPNYMARFCAPDRQQAAADWAAPMATLPRATGTPAPVRVQSLPGPTGESNR
LSEMARPPLLCLGPTAATVATQAQAVEALGGTAIRATGALAAEALETVEGISGVIWWGDE
ETARAYELALSRRSGPILPLIPGLPDGARVLAERHVCVDTTAAGGNAALLGGMG
>Coxiella Q83DR6 Q83DR6_COXBU P
MTDTHLLFFEKAIAQNAIRPSLNKTYRMDETTCVNHLLKTIAFTPRLEAAVSRLAKELVT
AVREQESEKGGIEGFMMQYDLSTEEGILLMCLAEALLRVPDKETENLLIRDKLTSAEWNK
YVGASESSFVNFATWGLALSGKILKKEKDGQFKNVWRNLVRRSGEPVIRKAVREAMKLMS
EHFVLGRTIEEAVKRSQSAIKEGFRHSYDMLGEVARTQEDADRYYDSYHRAISVLGKSHP
TKSVHEAPGISVKLSALYPRYDFKKRELAVPFLIERVKELALHAKEQKIGMTIDAEEADR
LDISLDIFEALFTDEAFENWQGLGLAVQAYQKRAFYLIDWLIDLAQRQKRRIPVRLVKGA
YWDTEIKLAQMEGLSGYPVFTRKVNTDISYIACAQKMLNAQDAIYPQFATHNAYSVAAIL
NLMDHHYDNYEFEFQQLQGMGKALHHYIVTKLKLPCRVYAPVGYHEDLLPYLVRRLLENG
ANSSFVNRIADKTVPVDQLIESPVKKIEAFGDIPNPKIPLPKGIFKTRTNSSGIDLSNFA
ELMPLNEEIHHALEKEWEAAPFLQEIKNGKPVFDPTDNRRQIGVIELANESDVEKAIQAG
HSAFPNWDQKGISARATILRKMADLLEKHKAELMAVVVREGGRTLQNALSEVREATDFCR
YYAEQAEQHLSDKALPGYTGESNTLRMNGRGIILCISPWNFPIAIFTGQIAAALVTGNAV
IAKPSGQTPLTGALVTRLFHEAGVPKEILQLMPGSGKTVGQALIEDTKISGVIFTGSDAT
ARHIQKTLAARPGPIVPFVAETSGINAMIADSTALPEQLVNDVIVSAFDSAGQRCSALRI
LYIQEDIADDVIKMLKGAMAEIKMGDPLLLSTDVGPVIDANAQKTLQKHQALMQKEAKLI
YKVDLPRETDFGTTFVAPQAYELPNLGLITEEVFGPILHVIRYKRENLNKVIEEINGLYG
LTFGIQSRIDETVDYIQQRINAGNIYVNRNTVGAVVGVQPFGGSWLSGTGPKAGGPHYLP
RFCIESTLTINTTAAGGNASLMAMED
>Oceanicola Q2CBQ1 Q2CBQ1_9RHOB
MPRDVQAPALDEIAAATYAPEAETVAALAAAADLPEQARAEIGRRAAGLVTDIRKASRPG
LMEVFLAEYGLSTDEGIALMCLAEALLRVPDAETIDALIEDKIAPSDWGRHLGRSSSSLV
NASTWALMLTGRVLDDDRPAPFRALRGAVRRLGEPVIRTAVGRAMREMGRQFVLGEDIDR
AMARAAKMEAKGYTYSYDMLGEAARTEADAQRYLEDYARAITAISRAATHDDVAENPGIS
VKLSALHPRYEEPRRAQVLEELVPRLRRLCRLAAEANMGLNIDAEEEADRLLLSLDVIAET
LIDPALADWQGFGIVVQAYGQRAGAVIDWVADLAARQNRRIMVRLVKGAYWDTEIKRAQV
LGLDGFPVFTEKAATDVSYIANARKLLGLTDRLYPQFATHNAHTMAAVLHMAEGLPRAAW
```

```
EFQRLHGMGETLHNLLLEREGTRHRIYAPVGAHEDLLAYLVRRLLENGANSSFVNQIVDE
DVPAEVVAADPFAALGNPPPLRRGPDLFPDRGNAAGWDITHRPTLAALDAARAPFRRHAW
TPASSGQKYADATAATGAPLEVVNPADPDDRVGTVATASKEGAARAHAAARPWSAPAAER
AAALRRAAAAYEAHAPELFAILHREAGKSLLDAIGEVREAVDFLHYYAANATDAPGTRGT
WVCVSPWNFPLAIFTGQLTAALAMGNAVLAKPAEQTPLIAARAVELLHEAGVPAEALQLV
PGDGATGAALIADARTRGVAFTGSTDTALKIRASMAKHCAPGTPLIAETGGLNAMIVDST
ALPEHAVRDILASAFQSAGQRCSALRCLYVQEDIAERLTTMLTGAMDTLTLGDPWDIATD
VGPVIDTEARDGIEAYVEAARAEGRVLHEIAAPGRGTFIAPVLIRVRGIEDLPFEVFGPV
LHLATFAADEVGEVIERINARGYGLTFGLHTRIDDRVQEVVERIAVGNAYVNRNQIGAIV
GSQPFGGEGLSGTGPKAGGPDYLPRFAPAPAPAGGGAWDADDSPTRLARALAAAPAPEPV
ETIDLPGPTGESNRLTLLPRGPLLCLGPSAATARAQAEAVRALGGQAVEAEGRIDPAWLT
DLDGLAGVLWWGDADEARALEQALAKREGPIVALITGAPDRAHACHERHVCIDTTAAGGN
AQLLAEVAAR
>Magnetospirillum Q2W4B4 Q2W4B4
MIFTAPLPAPDPERQAIHRAAGTSEADLVSGLSAGIPLEDEARRRIVNRAVNLVDGARRN
RRTLGLDGLLNEYRLSTREGVVLMCLAEALLRIPDDHTVDLLIKDKIASADWDGHLGHSP
SVFVNASTWALVLGDRLLHLEEDGRAVLGRMAGRLGEAVVRRALRHAMGLMGRQFVLGRT
IAEALDNARAWEARGYRHSFDMLGEAARCEQAAQDYLRAYAGAIEALGRHAKGAGPIAGP
GLSVKLSALHPRFEMAQRQRVLGELVPRLRDLCHRARDAGIGLTIDAEEADRLDISLDVM
EAALADPALDGWDGFGMAVQAYQKRARPVIAWAGALAARRQQRLMIRLVKGAYWDGEVKR
AQERGLGGFPVFTTKEATDVSYLACAADLLARPDLFYPQFATHNAHTAAAVMEMTGGAGD
WEFQRLHGMGEALYAQLVPEFPCRTYAPVGSHQELLPYLVRRLLENGANSSFVSRLADEE
IPAHVVAADPLAALGRITPQLVAEPSALFGPSRRNSGGLDLSSPAVLAQLDLALAAVATP
ERSAPIVDGRERENQAAKPVLDPADHRRVVGEVVDASPADVEAALASARAAFPAWDDLGG
EARASILERAADRLEADRARFMALAIREAGKTIPDALSEVREAVDFLRFYAAEARARFSQ
PVRLPGPVGESNELMLGGRGVFACISPWNFPLAIFVGQVAAALAAGNAVVAKPAPQTPLM
AAAAVRLLHQAGVPPQALHLVPGGPAIGEALTVNPLVDAIAFTGSTATARHINRLRAAMD
GPLAPLIAETGGLNAMIVDSSALPEQVVADCLESAFRSAGQRCSALRVAFIQREAWTRIQ
PLLAGAMAELSLGDPALLSTDVGPVIDEASRRRLLAHGGRLRHAGRMIGQSACPPDCRVG
TFFAPMAHQLDNLDLLQSEVFGPILHVIPWEAGRLEQVLDCVAATSYGLTLGIHSRIDAT
IAQVIARARIGNIYVNRTMIGAVVGSQPFGGLGLSGTGAKAGGPNTLIRYGVERCLSVNT
AAAGGDVALMAGPQRHGTK
>Mesorhizobium Q3WRH9 Q3WRH9_9R
MPALDALRDEIRKHYLADEDVLLRRLIESADLSPARREAISAKAAELVRAVRAGSTRHLM
ESFLAEYGLSTSEGVALMCLAEALLRVPDAETMDDLIADKIAPHDWSAHSGESASIFVNA
STWALMLTGRVLDDEGNGIEATLRGMVRRLGEPVIRAAVAAAMREMGEQFVLGRTIAEAI
KRGRAMTAKGYTYSYDMLGEAARTEADARRYFQAYRNAIAALKGEAKSEDIRANPGISVK
LSAIHPRYEATQRERVMPVVTERLLLLAQDARAARMGLNIDAEEADRLDLSLDVIERVLA
DPSLEGWDGFGIVVQAYGRRAPFVIDWLYALAERLDRRIMVRLVKGAYWDTEIKRAQVMG
LSSYPVFTRKVNTDISYIACARKLLSLTDRIYPQFATHNAHTVAAILAMAENRNAFEFQR
LHGMGEALHDIIKRREGTRCRIYAPVGAHEDLLAYLVRRLLENGANSSFVHQIVDEEVSA
EEIARDPFALAEAAGSAPNPAIPLPALIFGRQRINAKGWDITDETALREIETARRPFQAP
HRWEGRPITRARGQGEARIIVNPARTSEIVGTVFDASPGQVSESVRLAVAAQPGWAATPV
TERAEILRRIADLYEAHAGEFFALCAREAGKTLPDGIAELREAVDFLRYYANEAAQAEEG
TGARGVIACISPWNFPLAIFTGQIAAALVTGNAVVAKPAEQTCLIATRAVALMHEAGVPE
DVLHLLPGDGPSVGAPLTADPRIAGVCFTGSTEVARAIERQLAKTAAPDAMLIAETGGLN
AMIVDSTALPEQAVRDILASAFQSAGQRCSALRILYIQKDVEARVTAMLRGAMQALTVGD
PWQVPTDVGPVIDDEALSGISGYCEEMERKGRLLARIERPEGGRFVAPAAFQVSGIEEME
REVFGPVLHVATFEAEDLGHVISAVNAKGYGLTFGLHTRIDARVQEVVDGIHAGNIYVNR
NQIGAVVGSQPFGGEGLSGTGPKAGGPHYLRRFRRGKAGPDLAPADAPAIAAEELVRAMP
DAQKSGWADRTDRIVALRRALRGRWADVLASAAALDLGPVDLPGPTGEANTLQLAPRGRV
LCLGADADALAAQVIQALAAGNAVLAVAKDAPAILQPLTRAGFPLAAFHGNVAPQTLETL
ALDLVAATGERQWLTAVRTALSAREGRIVPLVTELVYPAAYAHERSVCVDTTAAGGNASL
LAAA
>Psychrobacter Q1QBN4 Q1QBN4_PS
MNPVEQFTPQEPILFNPIDLLSPEYIAQSSTELHTRISPLYSVDEERWLTQLLPLAKPTD
AERDSAATQTRQLVEHVRNDGKAVKMVDSLLLEYSLDTQEGVLLMSLAEALIRVPDNYTA
```

```
DALIHDKMSVADWKKHLKNDNGFMVNASTWGMMMTGRVVSIDSDTTASGFLDRMTKKMGE
PVIRSAMQKAMRIMGHQFVLGETIEGANKNSQPYRNKGYTYSFDMLGEAAITHKDAEKYF
NDYLHAIKATASVKVKEGMPKPSVSIKLSALHPRYEATQEEQVMGLLRQRCLLLIEAARE
VNVDLSIDAEEADRLEISLKLFESLYRDNLTADWDGLGLVVQGYSKRAIAILSWLARLAT
EVGDRIPVRLVKGAYWDTEIKLAQQKGLSGYPVWTRKEGTDTAYLACARFLLSEHLRGLI
WPQFATHNAHTLASIMTMSSHRDFEFQRLHGMGDALYDHILQAYQIPVRIYAPVGAHKDL
LPYLVRRLLENGANSSFVHQLLDKSYPIDKLTVHPYDKLLTNATLHNPDIPLPLAIYGAR
RASFGPNIFVESQWIPFKAAIDTHIHKTWSATSVINGNAVNDNTVDGDTHQLDKQAICAP
WNHEVIAGEVVYANATIAQQAIAAAVAGQGEWQAVAATKRAAILRKIADLYEENYAELMA
LCQVEAGKTLQDSIDEIKEAVDFCRFYADEAERLDDVIYEFTDLAGKPSRQVYKARGTFI
CISPWNFPLAIYTGQVVAALAAGNTVVAKPAEQTSLIAHFGAQLMYQAGVPVEALQLVIG
AGDVGAALTAADNIAGVIFTGSTQTAQRINQSLNAHAQVSGELPVFIAETGGQNAMIMDS
TALPEQVVRDAVLSAFGSAGQRCSACRILCVQEEMADDLIELLQGNMAELIVGNPLYATT
DVGPVIDIDAKRGLEAHIERMRAEPTATILAQTPMSDSSEVSQDQATFVLPTAIEVKSID
VIGGEHFGPILHVLRYEARDLNKLIDAINATGFGLTLGIHSRIENTVEHIERRAFVGNTY
INRNQIGAVVNVQPFGGCGLSGTGPKAGGPHYVARLMSLTTEQLANEQVALIDNTAATST
QTIIA
>Rhodoferax Q21UN0 Q21UN0_RHOFD
MATVTLGIKVDETLRSIKDAATLQGRTSHWLIKQAVLQYVEGIERGHRPQVATAGATSS
VEAVEIDDADDLTLPASPALFPPQPFLDWAQNVLPQTEMRAAITAAWHRPEPECLPMLVQ
LAHVTDPVQRVAIAEVGTRLVDGLRSNKDSGGVEALVQEFSLSSQEGVALCMAEALLRI
PDNATRDALIRDKISHGDWHSHLGNSPSMFVNAAVWGLMLTGKLTATASEKSMASALTRM
IGKGGEPLIRQGVHRAMKLMGEQFVTGQTISEALANSRALEKKGFRYSYDMLGEAATTEL
DAERYVASYEQSIRAIGMASNGRGIFEGPGISVKLSALHTRYSRAQRDRVMGELLPRLTK
LALLARQYDIGINIDAEESDRLELSLDLLESLCFDPRLKGWNGIGFVVQAYLKRCPYVID
HVIDLARRSGRRLMIRLVKGAYWDAEIKRAQLDGLDGYPVYTRKVHTDVSYLACARKLLA
APDAVYPQFATHNAQTVASIYQMAGNNYYAGQYEFQCLHGMGEPLYEQVTGTVADGKLGR
PCRIYAPVGTHETLLAYLVRRLLENGSNSSFVNRIGDPKVRVEDLVADPVLEAQAIELEA
GQLGAPHPKIALPRQLFAALGAQSRLNSSGLNLANEQQLASLAAGLLRSTQGNYAAAATV
AGMAHVNPQTAAADGWQAVCNPADTRDQVGWVRPATVAEVELAVNRAARATQIWQVTPPQ
ERAACLKRAADLLEQRTQSILGLIVREAGKSLPNAISEIREAVDFLRYYAAQVEATFDNQ
TQRPLGVVLCISPWNFPLAIFAGQVAAALASGNCVLAKPAEQTPLVADLMVKILHEAGVP
MDAVQLVPGTGEVVGAALVANRQVAGVMFTGSTEVARLIAQTLSQRLSRQGRCIPLIAET
GGQNAMVVDSSALAEQVVGDVLSSAFDSAGQRCSALRLLCIQEDVAERVIGMIKDAMREW
VMGNPDRMHTDVGPVIDEDARAQIEQHIERMRSDGQPVTRMARDESGAQGHFVMPTLIEI
DRIERLQREVFGPVLHVLRYRRDDLDAVLDAINATGYGLTFGVHSRIDETIAQVTQKVQA
GNIYVNRNVIGAVVGVQPFGGMGLSGTGPKAGGPLYVYRLLQEDDAQSNPGLAALAASPA
SRVLVPGSFDRQLAAHPALQALQRLMASLQGPTLAAAQGALGSFDAAQGAAACEAYRACS
VLGQAFMLPGPTGESNRYQLLPRGAVWAVPQTALGLLHQLAAALASGNACWIETPASDSV
VARMLNTLPPEVLRFVQQRSFDQLRGEPHLSAMLFEGDGDALQALSPRVAQQPGAIVRIE
SLSPAQLAAGACYDLSALMHEQSISTNTAAAGGNAQLMTMD
>Bordetella Q2KZJ2 Q2KZJ2_BORA1
MASTTLGVKVDDALRDRLKAAAQKLQCTPHWLHKQALVAYLEKIERGHLPPEMAHLGSGE
DGAEEDAGGQAGATPPFYEFGQDVQPQSVLRAAITAAYRRPEPECVPLLLGQARMPHLEK
IHAMAAKLVQTLRAKRTGGGVEGLIQEFSLSSQEGVALMCLAEALLRIPDRATRDALIRD
KVSRGDWKSHMGGSQSLFVNAATWGLMITGKLVAVSSEQSLSKALTRLIGKGGEPLVRKG
VNMAMRMMGEQFVSGQTISEALANNRKMESRGFRYSYDMLGEAATTAQDAERYNAAYEQA
IHAIGKAAAGRGIYEGPGISIKLSALHPRYSRAQRERVMSELLPRVKKLAILARQYDIGL
NIDAEEADRLEISLDLLEALCFTPELEGWNGIGFVIQAYQKRAPFVIDYVIDLARRSGHR
IMVRLVKGAYWDSEIKRAQVDGLEGYPVYTRKLYTDVAYLACARKLLAVPEAVYPQFATH
NAYTLAAIYHLAGQNYYPGQYEFQCLHGMGEPLYDEVVGPLAQGKLNRPCRIYAPVGTHE
TLLAYLVRRLLENGANTSFVNLIGDDSIPVEQLVADPVEAAARIVPLGAPHDKIPLPREL
YGHARANSAGLDLSNEHRLGSLSAALLASAGMPWRAAPMLGESEFVWDESRAQDVLNPAD
QRDGVGRLIEADSHDVEAALRAAANTAPIWQATPVAERAQCLRRAAQLLEEQMQTLLGLI
VREAGKTLPNAIAEVREAVDFLRYYADQAEREFDNDTHRPLGSVLCISPWNFPLAIFTGQ
VAAALAAGNTVLAKPAEQTPLIAAQAVAILRAAGVPAGAVQLLPGKGETVGAQLVAHASV
RGVMFTGSTEVARLIARNLADRLDDRGHTIPLIAETGGQNAMVVDSSALSEQVVFDVLSS
```

```
AFDSAGQRCSALRVLCVQEDNADHVLTMLRGAMRELRMGNPDRLSTDVGPVIDAEARQNI
LRHIESMRAAGHEIVQIEGGAECRFGTFVPPTLIEINDIAELKREVFGPVLHVVRYARDD
LDRVIESINGTGYGLTFGVHTRIDETVARVTESVQAGNIYVNRNIVGAVVGVQPFGGEHL
SGTGPKAGGPLYLYRLLSMRPAGLPPGLEGQMPLTLTLPGPTGETNTYRVEPRGAVYCVA
ATAAGARAQWAAAQATGNLAWFADTPAARELLATLESAAREGQVALLEDAEVDAADFQAV
LFEGDGDALKALNQRIARREGPILAVHGLSPDELAAGASYVAERLLNERSCSVNTAAAGG
NASLMTIG
>Rhodopseudomonas Q6N9G9 Q6N9G9
MPSDPPLAEFTAAYAPDDAALAAELLTTATLTPDREAQIDAIATDLITAIRGSEHGLRGV
EAMLREFALSTKEGLALMVLAEALLRVPDAATADAFIEDKLGQGDFAHHRIKSDAVLVNA
SAWALGLSARLVHAGETPQGTLAALTRRIGAPAVRAATRQAMRLIGNHFVLGETIDAALA
RAQPYAREGSRYSYDMLGEGARTAADAERYYQSYADAITAIGRRAGNAALPARPGISVKL
SALHPRFEAISRDRVMRELTPRLLELAQLAKSHDLAFTVDAEEADRLELSLEVFAACFAD
PSLKGWDGYGLAVQAYQKRAATVIDHVAELARAFDRRMMLRLVKGAYWDTEIKRTQERGL
ADYPVFTRKAMTDLNYLHCARKLLALRPLLFPQFASHNALTVATILAEAGDGDGYEFQRL
HGMGEALYGRLLADHPQAVCRIYAPVGGHRDLLAYLVRRLLENGANSSFVAQAGDDSVPI
TELLARPATLIGHPENARNSAIPLPRDLYQPQRINSRGIEFGDRSALAALLGDIEGARRP
LPTVASTSPEQAAATVTAARKGFESWSRTSGDHRAAILQRAGDLLEQRRGELIALLQDEG
GKTLDDGVAEVREAVDYCRYYASEGRRLFGEPQALPGPTGERNTLALRGRGVFVAISPWN
FPLAIFLGQITAGLMAGNAVVAKPAEQTPVIAEVAVRLLHEAGVPPAALHLVQGDGRIGA
ALVEQRDIAGVVFTGSTEVARAINRTLAAKDGPIVPLIAETGGINAMIVDATALPEQVAD
DVIASAFRSAGQRCSALRLLCVQDDVADRVIAMIAGAARELTIGDPRDPATHVGPVIDAE
AKTRLDTHIATMKRQAQLHFAGTAPASGNFVAPHIFELNRASELTEEVFGPILHVVRYKA
AQFDDLLDDIATTGYALTLGVQSRIDDTIARVIARLPTGNVYVNRNIIGAVVGVQPFGGS
GLSGTGPKAGGPHYLPRFALEQTVSINTAAAGGNAALLTGSE
>Gluconobacter Q5FP43 Q5FP43_GL
MIIGLRPSVVSEAESMSTFVPSYPVRSSLRQAIAADMRRPEAECILPLIEQATLTETEQQ
NTFDVARHLTRTLRTQRRPGGVEALVQEFSLSSAEGVSLMCLAEALLRIPDAATRDALIR
DRIGTGDWLSHVGGKKSVFVNAASWGLMLTGKLTNDTEGLAAALFRLVGRGGQPLVRRA
LDIAMRMMGEQFVIGETIEEARKVSTEPEERGFKYSYDMLGEAAMTEADALRYRRDYERA
IDVIGQTARGANVYEKAGISIKLSALHPRYAFAQRERVLKELGQTLKDLVIQARRYDIGI
NIDAEESERLDLSLDLIEDLCHCPELDGWNGIGIVVQAYGRRAPKVLDYLIDLGRRSGHR
LMIRLVKGAYWDSEIKKAQVEGQTDFPVYTRKCYTDVSYIACAKKLLAARDVVFPQFATH
NARTLATIYTLAGLNFQIGDYEFQCLHGMGETLYNEVVGPQKLNRPCRVYAPVGSHETLL
AYLVRRLLENGANSSFVNQIGDESLPIETLIADPVALAKAVQPPGASHPAIALPKDLFVP
ERTNSRGLDLTDEHTVTALAEAVCASEKQTLEDLSCSGETQEIRNPANHLDRIGSVRFGT
EEDVRKAIDAAESEGAAWAALSADARSTKLDRAADLLEEHQSELMALLVREAGKSYANAL
SEVREAVDFLRYYAVQARETAQAGSSAPLGVVACISPWNFPLAIFLGQISVALAAGNTVV
AKPAEETPFIALRAVALLREAGVPENALRLVPGAGETGAALVADPRISGVMFTGSTAVAG
LIASTLTSRTGADGQPVPFVAETGGQNAMIVDSSALTEQVVADVLVSAFDSAGQRCSALR
VLCVQEDCADRVLTMLRGAVEELRVGNPAELHTDVGPVISAEAQSGIQTYIDASRAQNRA
VWSLPLPDTTSNGTFIAPTIIEIDSLADLKGEVFGPVLHVLRFEASGFEALINAINQSGY
GLTFGLHTRIESRMAHVTSRIEAGNLYVNRNMVGAVVGSQPFGGEKLSGTGPKAGGPLIL
RRLMSTAPAHTDWENQELPEPARLFLSWLMRTSFPLYQTIVDAMQHGMCGTTRELPGPVG
ETNLYQLLPRGAVLCVASDRETMLRAVGLALSGGNTAFVQGPNVASDWVSDLPDALALHI
RRTQGGRVAGCRTILASPEEKQMAEQTRAALSRSGMIVQLYMLDAKSPIRPEWVLQEKVV
STNTTAAGGNASLMTIG
>Alcaligenes Q6QHN6 Q6QHN6_ALCD
MQISQKHSTVMDIHSKFDPASRFRGFASGQPAASELRERITAACRTAEPVAVSALLAMAK
MEPEVARKANQLSMQIAQRLRERKNSAGRAGIVQGLLQEYALSSQEGVALMCLAEALLRI
PDMATRDALIRDKIARGHWQEHAGRSPSLFINAATWGLFLTGKLVSTYSESSLSTLLTRL
IAKGGEPLIRKGVDMAMRMMGEQFVTGETIKEALDHARGLEAQGFRYSYDMLGEAAMTAE
DADRYRQSYERAIHAIGEASNGRGVYEGPGISIKLSALHPRYSRAQYGRVMDELYPVVRS
LALLARRYDIGLNIDAEEADRLELSLDLLEKLCFEPALAGWNGIGLVIQAYQKRCLFVID
HVIELAQRSRHRLMIRLVKGAYWDSEIKRAQVEGQDGFPVYTRKVYTDVSYLACARKLLV
APAEIYPQFATHNAHTMAAIYNLAEPANYQPGQYEFQCLHGMGEPLYEQVVGSLADGKLG
RPCRIYAPVGTHETLLAYLVRRLLENGANTSFVNRIADPSIPLEALVEDPVAAVERMAQE
```
116

```
EGEAGRPHPCIPSPSALYGQARRNSHGLDLSNEQTLRGLEPILVDSARTGWTAEPLLGES
AEAEPTAWKPVLNPADHRDVVGKVRAASEQQIEAALRSATGFAPQWAATSAGDRAFVLER
AADLLQDQLPYLLGLLCREAGKSYANGVAEVREAIDFLRFYAAHARTDFDDPAHKPLGPV
VCISPWNFPLAIFVGQVSAALAAGNPVLAKPAEQTSLIAAEATRLMHTAGVPRAALQLIP
GRGSVVGARMVADARVQGVMFTGSTAVARQLQRTMAERLGPTGAPVPLIAETGGQNAMIV
DSSALVEQVVADVMASAFDSAGQRCSALRVLCVQDDVADRLLHMLRGAMAESNVGDPSSL
CTDVGPVIDLAARTTIESHIELMQSRGRRVNRAGRAAHEVLEGGTFVLPTMIELESISEL
EREVFGPVLHVVRYARQDLDTMLAQISATGYGLTMGIHSRIDETIGRVVSSAKVGNVYVN
RNMVGAVVGVQPFGGEGLSGTGPKAGGPLYMYRLLAESPHDVLVRAAERGLPGSVADGSV
TQASEPFEALLSWIETHRERPLTARCRLLKRTAGSGQMRVLVGPTGERNTYSLQPRVAAL
CLAEDEDSLLFQLAGALVVGCRAIWPSNTDTQALRATLPLDVQAEIALARDWAAPTVEFD
VALYQGSREGLLAASARLAERQGPITSLWRFPTDSGKVPLEALVVEQAVSTNTAAAGGNA
SLMTIS
>Paracoccus Q3P751 Q3P751_PARDE
MNRVNPTVRSDVFSAQAKFADEAELLSGLVAQAGLDDAQRAAITRRAADLVRRIRDEAKP
TMMEHFLAEYGLSTREGVALMCLAEAMLRVPDRMTIDALIEDKIAPSDWGKHLGEASSSL
VNASTWALMLTGKVLDDDQAGIAGTLRRAVRRLGEPVIRTAVGRAMKEMGRQFVLGQTIE
AALERAAKREAQGFTYSYDMLGEAAMTGADAARYDRAYSDAIAAIAKACTRGSVEDNPGI
SIKLSALHPRYEVAQESRVMTELVPVVLKLARQAKAAGMGMNIDAEEQDRLVLSLKVIEA
VLSDPSLAGWDGFGVVVQAYGKRAGQVIDWLYETATRLDRRIMVRLVKGAYWDTEIKRAQ
VEGFPGFPLFTSKVATDVSYIANARKLIGYADRIYPQFATHNAQTVAAILEMAGETRFEF
QRLHGMGERLHDIVLRDHNGRCRIYAPVGAHRDLLAYLVRRLLENGANSSFVHQIVDESV
SPEEVARDPFAALAEARPPLGLITPDALFGASRRNSTGFDLTDEETLARIYTARDVAIPD
AMPLTVSEPTGKMQDVLNPATGEKIARVMMVDADTAARAIADARIWDASAAERAAVLRRA
ADLYEDHYGPIFGILGREAGKSLADAVAELREAVDFLRYYAAEGEASPGAPRGIVGAISP
WNFPLAIFTGQIAAALMAGNAVIAKPAEPTPVIAAYAIGLLHQAGVPKAALQLLPGAGRV
VGTALSSDPRVSGLVFTGSTGTAQTIARTMAANLAPGTPLIAETGGLNAMVVDSTALPEQ
AVRDIVNSAFRSAGQRCSALRCLYVQEDVAPHLVEMIKGAMDELRLGDPWDLTTDVGPVI
DPGAQKEIADYIAANANRVLHKLAAPSRGWFIPPTMLKVSGIADLEREIFGPVLHVATFR
GDQLDQMIADINARGFGLTFGLHTRIDSRVQEVSDAIHVGNIYVNRNQIGAVVGSQPFGG
EGLSGTGPKAGGPRYVPRFFAPAAPSAVPGTWQGEGDEARLRGALADASPKKIDERLMPG
PTGELNRLTAHARPPVLCLGPGAEAVAAQVAAVQALGGQAVGADGALPPEALTRLPQLST
VLWWGDEAQGRAHAQALAAREGEIVQLITAMPDLAHIAHERHLCVDTTAAGGNAALLAG
>Sodalis Q2NTK7 Q2NTK7_SODGM Pr
MGTTTMGVKLDDEARERIKQAARQLDRTPHWLIKQAIFHYLDALEQGATPGLPLPAAEAD
DTLKVVSDEPRQPFIDLAEQILPQSITRAAVTAAWRRPETEAVPWLLEQARHPAPMAETI
QTLAGKLANQLRHQKRSGGRAGIVQDLLQEFSLSSQEGIALMCLAEALLRIPDKATRDAL
IRDKISNGNWQTHLGRSQSLFVNAATWGLLFTGKLVATHNEASLSRSLNRIIGKSGEPLI
RKGVDMAMRLMGEQFVTGEHIGEALANAHRLEQQGFRYSYDMLGEAALTEEDAQAYLLSY
QQAIHAIGKASSGRGIYEGPGISIKLSALHPRYCRAQYERVMAELYPRLLRLTLLARQYD
IGLNIDAEEADRLEISLDLLARLCFEPALAGWNGIGFVIQAYQKRCPQVIDELIDLAKRS
QRRLMIRLVKGAYWDSEIKRAQIEGLEDYPVFTRKVYTDLSYLACARKLLSVPHYIYPQF
ATHNAHTLAAIYHFAGQNYYPGQYEFQCLHGMGEPLYEQVVGKVAEGKLNRPCRIYAPVG
SHETLLAYLVRRLLENGANTSFVNRIADSAVPLEQLIADPVQEVTQLAVREGRAGLPHPK
IPLPRDLYGAQRRNSAGLDLANEHRLASLSAALLNVEHQTWQAAPVVVTEIGDGAVQKVL
NPADHRDVVGECRQASSEEVAHALAAAVHHGSLWSATPPADRAAVLKEAANRLETDMQPL
IGLLVREAGKSFSNAVAEVREAVDFLRYYACQIAQDFDNDNYRPLGPVVCISPWNFPLAI
FLGQIAAALAAGNTVLAKPAEQTPLIAARAVALMLEAGVPLGVLQLLPGAGETVGAALVA
DPRVRGVMFTGSTQVARLLQTTLAARLDPQGRPIPLIAETGGLNAMLVDSSALTEQVIID
VVTSAFDSAGQRCSALRLLCIQEDVAERTLRMLRGAMAEYTLGNPERLATDIGPVIDSSA
KAAIDSHITAMRERGYTVWQTPAAEAAGDIQGTFVPLTLIELDSIEALTQEVFGPVLHVV
RYQRRELDSIIDQINGSGYGLTLGLHTRIDETIQRVTERAHVGNCYINRNIVGAVVGVQP
FGGEGLSGTGPKAGGPLYLCLLSQRRDDALLPSLQALDAVHAPDFTRREILQQAHPALV
AWMETHQPTLVAQCRHLGEISQAGSVRLLTGPTGEQNSYRLLPREHILCLADQDPDLLLQ
LAAITSIGARALWAESPQSRRLFTTLPDSVRQRITLLADWTQEDVRLDAVLFHGDSDQLR
NLAQTLSKRPGPLITVQGNARGDGQIALERLLIERAISVNTAAAGGNASLMTIG
>Francisella Q2A415 Q2A415_FRAT
```

```
MNDLLNHSGEYPISKEIMNIYKYWLIDEKEAMTKLVEKAHMSSVQKAQVRERAYGLVEKV
RKNRLKKSGIDAFMIEYDLSSEEGVVLMCLAEALLRVPDKYTIDLLIKDKLTSAAWKNHV
GMEKHLFVNAATWSLMLTGKILKDPHRSYRKVFQNFLKKTSEPVIRQAMKQAMKIVGKQY
VLGETIEEALKVSEAKVARGYSYSYDMLGEAAMTDDAEYYYSQYLHAIHELAKYATNTE
IKKNPGISIKLSALHPRYEVAKHQRVHTELYPKLLKLTQLAKDYNVGMNIDAEETERLQI
SLELVERLAHEPSLEGFNGIGIVVQAYQKRAPYVLDYLANLAKKTNRRFMIRLVKGAYWD
AEIKHAQEQGLAGYPVFTRKYHTDVSYQACVKQLFENHQYIYPQFATHNAQTVAVVFELA
NGNRDFEFQCLHGMGDALYDNVVGKEGYEDIPCRIYAPVGGHKHLLAYLVRRLLENGANS
SFVNRIVDENLPIEELIEDPVKKAIDHGCGQHPNIPYPKDIVSPRLNSQGHNTNDFAVLA
DMYKQIEKYTHKNTYKAKPIVSGIDLDKTITESVINPNTNEVIGSVINADAKIAKRALKN
AQSAFEEWSNTPATKRADILEKFADLLEQDTNKFIAIAMIEAGKTLANAIDEVREAVDFC
RYYAAQARKEFNGPIELPALSDHLKQIEFTGRGAMVCISPWNFPLAIFLGQITAVLAAGN
TVVAKPAEQTPIIAYKAVKLLFKAGLPKNVLQFTPGDGATVGSALVKSPVCKGVIFTGST
EVAAIINQTLANKSSEIVPFIAETGGQNAMIVDSSSLPEQVTADVIRSAFDSAGQRCSAL
RILCLQEDIADNYIKMIVGAMKELKIGDSKYIDTDVGPVIDKEAADNLNAYIEEKKSQFK
LIYQSQPNQDTQKGTFVMPTAFEIDKISDLGREQFGPILHILRFKANQLDKLINDINATG
YGLTAGVHSRINEVMNYVKNNLKAGNIYVNRNIVGAVVGVQPFGGQGKSGTGPKAGGPYY
MHRLANEKLSGVGAVEEIYNPEKIAEYEKQTNKLIKDKYTITNIVAGEKAKKDKYTNLKA
ANGKTIGKKYVASVNTVDNAIEIASKEVEIWNHVNAEQRAATIEKFLELLEKERYLIASS
LVVESNISVEDAHIQIDKTIQQVAYYCLQAKKEFAHPQLLPGPTGEIDELSLKGRGVVVS
MCSSSDLLIRFVGQTAAALLAGNTVVAKPAYTGNLTAYNIVKLMLKAGIDSKVLHLVLSD
DEEITSALLFNSKVALVSFSGSVSAVKQVHQALVLRRGAIIPFVAESIAKNGKCTSLAIE
TASPLYLRRFVVEKTVSVDTTASGGNASLMSLEE
>Wigglesworthia Q8D2C0 Q8D2C0_W
MDTRILGIKLDNKLSHRIENISSRLNRTPHWIIKQAIFLHLDDLEKKSKSSEKLIENKNE
IQDYQDKINEKIIFKPFFEFAQQILPQSYLRFNTTNAWSTPEKEIVPKLLELSKCEEETE
KKIHELSKYLAKGLREKSKKYNREQMVQDLLREFPLSSKEGIALMCLAEALLRIPDVNTR
NSLIKDKIKNKDWKSHFGSNKGLFINSAVWGLCITNNLIKRHKEKNLSSYLNKVIQKIGE
PIVGKAINVAMKLMGKQFVVGENIQQALKNTNILEKKGFSYSYDMLGEAALTEEDAQKYM
TSYENAIHCIGKSSLGKGIYKGPGISIKLSAIHPRYSRNKYEKVMSELYPRLRSLTLLAR
HYDIGMNIDAEEANRLEISIDLLDKLCCEPSLFGWSGIGFVVQAYQKRCMYVIEELIDIA
KRTKRRLMVRLVKGAYWDSEIKNAQIEGIEDYPVFTRKAYTDISYLACAKKLLSESNHIY
PQFATHNAHTVSSIYHFAGENYYSGQYEFQCLHGMGEQLYEKVVGNAIDKKLDRPCRIYA
PVGTHKTLLAYLVRRILENGANNSFINRISDNNVSIEKLVLSPIKESISISKKENVEIGT
SHPNIPLPKNLYGNNRENSHGFNFSNENVLAKLSSDLLKNSDKFWKVTPIIHSKLDKGSI
SKIINPANKKDIIGECQSSTLKDVESALESSVKGIELWSSKEPLERSKILYDISNKIEKN
TNIFINLLVRESGKTFPNAVSEIREAVDFLRYYSNQIKPFNNKTHIPLGCVLCISPWNFP
LAIFLGQISAALASGNAVIAKPAEQTPIVAYEAIKLMIDSGIPSESLQFLPGSGKIIGNS
LSKDHRIHGVMFTGSTQVARLLQLNLSDRLSKGRPVPLIAETGGLNAMIVDSSALTEQVV
SDIIVSAFDSAGQRCSALRLLCIQEDVSEKTLNMIKGAIDSYTVGNPEYFSTDIGPVIDK
NAKINIDQHILNMKKSNHSVWQSEMSYDTKQDLGNFVLPTLIELDDIDQIKEEVFGPVLH
IVKFKYEEMNNIFEKINSSGYGLTIGIHSRIEENIEKVTTSTKVGNYYVNRNIVGAVVGV
QPFGGEGLSGTGPKAGGPLYLYRLLSDRKEKNILSSLEELDKNKKVDLTRCEKIQESCVS
LIEWISVNYPDLEKYCNYFISQSQSGSSRVLQGPTGEKNTYLLLPRERILCLSNDENDLL
IQLAAVTSIGGKAILSHNPITQKIFSILPQSVLKNVILIPNWKREDSMFDIVLFHGDELQ
LKEVLEILSKKRGPVINVHSHKNGDKKIFLEKLLIERTISLNTTAAGGNTTLLSMV
>Photobacterium Q6LFT6 Q6LFT6_P
MFNAADVLQPSFIERPLNEIWTLISPLYSVDESLWLEQLLPLAEPSDIERKHTTDKAAEL
IQRVRADKNAIQMIDALLLEYSLDTKEGILLMCLAEALMRIPDAATADALIRDKLSVADW
KSHLKNSDSLFVNASTWGLMLTGKVVTMDAKEDGKPSHVINRLVNKMSEPVIRQAMNQAM
KIMGHQFVLGRTISEAMKNGKSNRDKGFTYSFDMLGEAALTAQDAQKYFKDYIMAVESVG
RDEYASQKGSNQQSPDPTVSIKLSALHPRYDVANEARVLDEMYESVLTLLSRARELNVGI
TIDAEEADRLELSLKLFEKLYRSDAVKGWGRFGIVVQSYSKRALPVLAWLAALSKAQGDI
IPLRLVKGAYWDSELKLSQQSGFSDYPVFTRKESTDSAYLACARFLLSEHLRGVIYPQFA
SHNAHTISAIVAMTDHRDFEFQRLHGMGDALYNHAMDMYQANVRIYAPVGSHKDLLPYLV
RRLLENGANSSFVHRLVDANCPIESLTEHPVDTLRSRPSLNNSLIPLPPQIFGESRKNSA
SINIDIESEWTPFNAAIQAFSQHQWQAGPIVNGVTLDGEQHAITAPYNRSESVGQVAFSS
```

```
ASQVAQAIDVASEAYPAWSQVAASERGECLQRLADLLELHTGELVALCHREAGKTIQDSI
DEIREAVDFCRFYGEQASSELASAKSIKSFDGSTKQLTYQGRGVFACISPWNFPLAIFLG
QVSAALAAGNTVVAKPAEQTSLIAYRAIELMLEAGIPAGAIQLLPGTGAEVGATLTTDVR
IAGVAFTGSTETAQRINRSLVARDCEAVPFIAETGGQNAMIVDSTALPEQVVRDVVRSAF
ASAGQRCSALRVLFVQADIADRIITLIKGAMAELSVGRPELHSTDVGPVIDIAAKEKLLT
HIESLKNQSTLVAQTQLSDECKLGDFVAPTAFEINSIDILKNENFGPILHIVRFKANEID
QVVDQINHTGFGLTLGVHSRNERTYCRIEQRARVGNCYINRDQVGAVVGVQPFGGQGLSG
TGPKAGGPHYLYRFTKDVLSA
>Pseudomonas Q9R9T7 Q9R9T7_PSEP
MATTTLGVKLDDPTRERLKAAAQSIDRTPHWLIKQAIFNYLEKLEGGATLTELNGHASNP
ADDAGEVQADHSHQCFLEFAESILPQSVLRSAITAAYRRPEQEVVPMLLEQARLSAPLAD
ATNKLAASIAEKLRNQKSVGGRAGIVQGLLQEFSLSSQERRGVCLAEALLRIPDKGTRDA
LIRDKISTGNWQPHLGNSPSLFVNAATWGLLLTGKLVSTHNETGLTSSLTRIIGKSGEPM
IRKGVDMAMRLMGEQFVTGETIAEALANASRFEAKGFRYSYDMLGEAALTEHDAQKYLAS
YEQAIHSIGKASHGRGIYEGPGISIKLSALHPRYSRAQYERVMEELYPRLLSLTLLAKQY
DIGLNIDAEEADRLELSLDLLERLCFEPSLAGWNGIGFVIQAYQKRCPYLINYFFDLAKR
TPHRLIIRLLKGAYWDSEIKRAQVEGLEGYPVYTRKVYTDVSYVACARKLLAVPEAIYPQ
FATHNAHTLSAIYHIAGQNYYPGQYEFQCLHGMGEPLYEQVVGKIADGKLNRPCRVYAPV
GTHETLLAYLVRRLLENGANTSFVNRIADHSISIQELVADRWPASIAWVPRKGSIGLPHP
RIPLPRDLYGTERAKLAGIEMANEHRLGLLSCAMVATAHKQWEAAPLLACAARESAAAPV
LNPADHRNVVGHVQEATVAKFDNAIHCALNPAPIWQATPPAERAAILERTADLMEAEIHP
LMGLLIREAGKTFPNAIAELREAVDFLRYYAVQALNDFSNDAHRPLGPVVCISPWNFPLA
IFTGQVAAALAAGNPVLAKPAEQTPLIAAQAVRLLLEAGIPEGVLQLLPGRGETVGAGLV
GDERVKGVMFTGSTEVARLLQRNVAGRLDNQGRPIPLIAETGGQNAMIVDSSALTEQVVI
DVVSSAFDSAGQRCSALRVLCLQEDSADRVIEMLKGAMAESRLGCPDRLAVDIGPVIDAE
AKAGIEKHIQGMREKGRPVYQVAIADAAEIKRGTFVMPTLIELDSFDELKREIFGPVLHV
VRYNRRNLDQLIEQINNSGYGLTLGVHTRIDETIAKVVETATPATCRHRNIVGAVVGVQP
FGGEGLSGTGPKAGGPLYLYRLLSTRPADAIGRHFQQQDGEGTPDRTLHEQLVKPLHGLK
AWAENNQLADLAALCSQFASQSQSGIARLLPGPTGERNSYTILPREHVLCLADNETDLLA
QFAAVLAVGSSAVWVDGEPGKALRARLPRELQAKVKLVADWNKDEVAFDAVIHHGDSDQL
RGVCQQVAKRAGAIVGVHGLSSGDHQIALERLVIERAVSVNTAAAGGNASLMTIG
>Erythrobacter Q2N7F6 Q2N7F6_9S
MSKIAPLNRLHFRTAYRQEENACVAERLKQAAPVSARHEEAQALAVRLIEDARSRKASGI
DAFLHTYGLATEEGIALMCLAEALLRVPDHETADALIRDKLGDIDWGEHLGESSSTFVNA
ATFSLMLTGEVLERPEEHQRGMGKTLKRTVNRLGEPVIRKATLQAMRILGGQFVYGRTIG
EALKRAKPERAKGLTHSFDMLGEAAMTFEDAERYRRAYERAIERLAGETDGTIQSSPGIS
VKLSALYPKYDIFHQDAAVEALVPMLRDLAIKARDANMHFNIDAEEAERLEISLDIIERL
AADDALFTKADGTRWDGFGLAIQAYQKRAAPLCDWIVKLARRHDRRFMVRLVKGAYWDTE
IKLSQVGGYEDFPVFTRKVATDVSYLACAAKLLEAPDAIYPAFATHNAYTVGAVKALAGH
TEFEFQRLHGMGEELFEALAAQEGNRKTPVRIYAPVGVHKDLLAYLVRRLLENGANSSFV
NRMADAEVPAVELANDPVEELALLEPKRNPDIPLPADIFPGRRNSIGVDLADPLVRDPLL
QRLEALESRHWYAGPTFPSEDAGEEAPINAPQDLTHEVGTRRDSTEEEVRAAITQAEFIQ
PGWNALGGEKRAVLLEAAADLFEDHMDEFLSLCRREAGKTVLDGVLEVREAVDFLRYYAT
EARRQFSAPIMLPGPTGEENRLHLAGRGVFATISPWNFPLAIFIGPAAAAMAAGNTVIAK
PAEQTPLIAALAVKLCHEAGIPEEVFQLLPGAGDVGQMITGDPRIAGVAFTGSTETAQAI
NRSLAAREGPIATFIAETGGQNAMIVDSSALPEQVARDVVASAFQSAGQRCSAQRMLYIQ
DDVYDDMLAMIRGAFAALQIGDPTDFAIDVGPVIDPDAKAALERHVARRKKAGRTVWRRK
LPRGTANGCFVAPTIIELDSILDLKRENFGPILHVARYKADGLNRVIADINSTGYGLTLG
LHSRIEATRRYVEARARVGNFYVNRNQIGAVVESQPFGGEGLSGTGPKAGGPHYVARFAT
ERVVCIDTTAAGGNATLLAS
>Wolbachia Q73IQ2 Q73IQ2_WOLPM
MISSIQEPNELRKRLQGFYRTDEKSYIRYLVEKAELSADSKNRIYNIAKQVIEKIKHNKL
SIVDSFMQQYSLSNDEGIALMCLAESLLRIPDDYTIDEIIKDKIANQEWNKYLGHSSSLF
VNASTWSLMIGSSILRDNEGDSKFYYAISKLLKNLGEPIIRKAVKQAMSMLGNHFVVGET
IEEALRYAKLDDNSKFLYSFDMLGETAHTAEDAEEYFNSYMHSIKAIGESTEINDCFKSH
GVSIKLSALHPRYEFGQFDNIAEELRAKLLELCHEAKKYNISLCIDAEETERLEMSLILF
EQLRLDESLSEWEGLGLAVQAYQKRALSVLDFVEDVAIRSKHKIMVRLVKGAYWDSEIKR
```

```
TQELGLNDYSVFTRKSYTDVSYFVCAQKLLSKPNSFYPCFGTHNVYTFASIMELADKNHP
GFEFQRLHGMAKDLYDYAMSELATNVSCRVYAPVGEHSDLLPYLIRRLLENGANSSFVNQ
ISDPNVKIDELISDPLEKAINFNYEPHPGIPLPQDILGPERKNSLGMDISDSVIVSQFAD
DIKGFSEKKWQIGPIIDGKALFDNAEFIEVVNPAHLENVIGEVSSATSDQALNALEIAHS
AFAKWQNVSAEERAKCLEKAADLLEERMKELIYILIVEAGKILSDAIAEVREAIDFLRYY
AMIAKNELSDWKKLPGPTGEDNFIFFEGRGVFLCISPWNFPLAIFIGQVSAALAAGNAVL
AKPAEQTPIIAYEAVKILHEAGIPKNVLHLIPGDGGYLGKILVPDNRIAGVAFTGSTQTA
QIINKMLANRDGPIVPLIAETGGLNAMIVDSSALLEQVAMDALLSAFRSSGQRCSALRVL
FIQEDIAEKQIKMICGAAQELKIGDPIQLSTDIGPIIDKASIDMLTQHTQKMSEDEDSNL
LFKVPMDTNSHNGYFFPPYIYEIQKISQLKQEVFGPILHIIRFNKSQLNEVISDINSTGY
GLTFSLQSRIQSQIETISKKISVGNVYINRNQIGAAVGIQPFGGRGLSGTGPKAGGPHYL
QRFSTEKVVSVNTTAFGGNTTLMCLD
>Nitrosomonas Q3NDK3 Q3NDK3_9PR
MRIHPARAAMAHLYLCDETEAVNALIPRARLSAMERAATEALARDLVQRMRMQRWRGGLN
MFLHQYALSTEEGVALMCLAEALLRVPDAETQDRLIRDKIGSVHWEKHLGQSSSLFVNAS
TWALMLTGRVVRLGEGRSDWTNIFSRMVQRSGKPVIRQAMSAAMRIMGRQFVLGCTIEEA
LRRGTEAARRGYRFSFDMLGEAAVTREDAERYARVYAGAIAAITSHEARKEAQHGDIFAR
NSISVKLSALHPRLEYVKAERVLAEILPQLMPLCQAACKARIGFTIDAEEADRLDLQLDI
FEALSRTRELAGWEGLGLAVQTYQKRAPAVIDWLAAIAREHKRRIPVRLVKGAYWDSEIK
RAQELGIDGFPVFTRKASTDVSYLACARMLLADRSAFFPMFATHNAHTLAAVKVLAGNRS
DYEFQRLHGMGEELYDQAVGADHSGVQCRIYAPVGPHKDLLAYLVRRLLENGASTSFVNR
FADNKMPIASIIADPVDQIEAMSAKPHPHIPLPKNLFTDRKNSRGLLISDPAQARLLLNA
VAAAIEHGFEARPLIGGEPYPGEGAAVFDPALHSRQVGTVLNATPDMVTRAASTAAAAQG
AWDRLGGVRRAEILERAADLFERDMAVLIALCVREAGKTVSNALADVREAIDFLRYYAGI
ARADFAEPKVMPGPTGERNELSLHGRGVFAAISPWNFPVAIFTGQVASALAAGNSVLAKP
AERTSLAAYTAIQLMHEAGVPGEVLHFLPGSGRVIGTAMVRHPATTGVVFTGGTDTGHVI
NRMLAERAGPIVPFIAETGGLNALIADSTALPEQLVRDVLTSAFDSAGQRCSALRVLFLQ
EDIAERVISLIRGAMAELRIGDPMKLDTDIGPVIDGNSLAKLEAHAKRMSREAGFIAEAP
LSDETRDGFFFAPRAYEIDAISRIEREVFGPILHVIRFKATRLDKVCEAINATGYGLTLG
VHTRIQETVDFIRARVRIGNLYVNRNQIGAVVGVQPFGGEGLSGTGPKAGGSHYLHRFAV
ERTVCINTAATGGNTALLSLDEVACRGRDS
>Brucella Q578K6 Q578K6_BRUAB P
MTDNIPVSKVAVFQNFAPPIREQSALRQAITAAYRRPEAECVSALAEQATLPEETRQQIR
STARKLIEALRAKHKGTGVEGLVHEYSLSSQEGVALMCLAEALLRIPDMATRDALIRDKI
SNGDWKSHIGGGRSLFVNAATWGLVVTGKLTNTVNDRGLSAALTRLIARCGEPVIRRGVD
MAMRMMGEQFVTGETIDEALKRAKELEERGFRYSYDMLGEAATTAADAERYYKDYETAIH
AIGRASAGRGIYDGPGISIKLSALHPRYTRAQSERVMGELLPKVKALAAIAKSYNIGLNI
DAEEADRLELSLDLLQSLCEDPDLAGWDGIGFVVQAYGKRCPLVLDFIIDLARRTKRRVM
VRLVKGAYWDAEIKRAQVDGLEDFPVYTRKVHTDVSYIACARKLLAATDVIFPQFATHNA
QTLATIYHLAGPDFKTGKFEFQCLHGMGEPLYDEVVGPEKLGRPARIYAPVGPHETLLAY
LVRRLLENGANSSFVNRIGDKNVSVDELIADPVEVVRSMAVVGARHDQIALPENLYGARR
NSAGFDLSNEVTLAELSKTLKETAGRAWTAEPQVAGAKVKGVSRPVLNPGDRNDVVGTVT
EIAEEADVAKAMKAAQTATTSWSAVAPAERAACLERAADIMQRDMPALLGLVMREAGKSMP
NAIAEVREAIDFLRYYAEQTRRTLGVGHKALGPVVCISPWNFPLAIFTGQIAAALVAGNP
VLAKPAEETPLIAAEGVRILHEGGIPADALQLLPGDGRIGAALVAAPETCGVMFTGSTEV
ARLIQAQLASRLLPNGKPVPLIAETGGQNAMIVDSSALAEQVVFDVIASAFDSAGQRCSA
LRVLCLQEDVADRILTMLKGALRELSIGRTDQLKVDIGPVITDEAKNTIEKHIQAMRDLG
RKVEQLPLGPETQNGTFVAPTIIEIESLRDLKREVFGPVLHVVRYKRDDMESLIDDINST
GYGLTFGLHTRLDETIANVADRIRVGNIYINRNIIGAVVGVQPFGGRGLSGTGPKAGGPL
YLGRLVETAPIPPRHASVHTDAALKDFARWLGNRGMNDLAQAARDTGSASALGLELELPG
PVGERNLYALHPRGRVLLVPQTEIGLYRQLTAVLATGNTAVIDEACGLRAVLKDLPETVA
ARAIWTGDWQADAPFAGALIEGDSARIKEVNSRIAALPGPLVLTQAASPEDLAANQDAYC
LNWLLEEVSTSINTTAAGGNASLMAIG
>Escherichia coli P09546 PUTA_E
MGTTTMGVKLDDATRERIKSAATRIDRTPHWLIKQAIFSYLEQLENSDTLPELPALLSGA
ANESDEAPTPAEEPHQPFLDFAEQILPQSVSRAAITAAYRRPETEAVSMLLEQARLPQPV
AEQAHKLAYQLADKLRNQKNASGRAGMVQGLLQEFSLSSQEGVALMCLAEALLRIPDKAT
```

```
RDALIRDKISNGNWQSHIGRSPSLFVNAATWGLLFTGKLVSTHNEASLSRSLNRIIGKSG
EPLIRKGVDMAMRLMGEQFVTGETIAEALANARKLEEKGFRYSYDMLGEAALTAADAQAY
MVSYQQAIHAIGKASNGRGIYEGPGISIKLSALHPRYSRAQYDRVMEELYPRLKSLTLLA
RQYDIGINIDAEESDRLEISLDLLEKLCFEPELAGWNGIGFVIQAYQKRCPLVIDYLIDL
ATRSRRRLMIRLVKGAYWDSEIKRAQMDGLEGYPVYTRKVYTDVSYLACAKKLLAVPNLI
YPQFATHNAHTLAAIYQLAGQNYYPGQYEFQCLHGMGEPLYEQVTGKVADGKLNRPCRIY
APVGTHETLLAYLVRRLLENGANTSFVNRIADTSLPLDELVADPVTAVEKLAQQEGQTGL
PHPKIPLPRDLYGHGRDNSAGLDLANEHRLASLSSALLNSALQKWQALPMLEQPVAAGEM
SPVINPAEPKDIVGYVREATPREVEQALESAVNNAPIWFATPPAERAAILHRAAVLMESQ
MQQLIGILVREAGKTFSNAIAEVREAVDFLHYYAGQVRDDFANETHRPLGPVVCISPWNF
PLAIFTGQIAAALAAGNSVLAKPAEQTPLIAAQGIAILLEAGVPPGVVQLLPGRGETVGA
QLTGDDRVRGVMFTGSTEVATLLQRNIASRLDAQGRPIPLIAETGGMNAMIVDSSALTEQ
VVVDVLASAFDSAGQRCSALRVLCLQDEIADHTLKMLRGAMAECRMGNPGRLTTDIGPVI
DSEAKANIERHIQTMRSKGRPVFQAVRENSEDAREWQSGTFVAPTLIELDDFAELQKEVF
GPVLHVVRYNRNQLPELIEQINASGYGLTLGVHTRIDETIAQVTGSAHVGNLYVNRNMVG
AVVGVQPFGGEGLSGTGPKAGGPLYLYRLLANRPESALAVTLARQDAKYPVDAQLKAALT
QPLNALREWAANRPELQALCTQYGELAQAGTQRLLPGPTGERNTWTLLPRERVLCIADDE
QDALTQLAAVLAVGSQVLWPDDALHRQLVKALPSAVSERIQLAKAENITAQPFDAVIFHG
DSDQLRALCEAVAARDGTIVSVQGFARGESNILLERLYIERSLSVNTAAAGGNASLMTIG
>Nitrobacter Q3SN37 Q3SN37_NITW
MSPQPPPPFSVPYAPDDRTIAKQLLTGARLAPEQETRIDQTARRLIDAIRQNDDPLGGVE
DMLREFALSTKEGLALMVLAEALLRVPDAETADRFIEDRLGQGDFIHHETRSSAFLVNAS
AWALGLSARVIQPGETPQGTIGRLAKRLGVPAIRAATRQAMRLMGSHFVLGETIEAALER
THTHSAGCSRYSFDMLGEGARTAADADRYFDSYAAAIEAIGRGAGDNPLPDRPGISVKLS
ALHPRYEAVSRGRVMAELVPRAIDLARRARALDLAFTVDAEEADRLELSLDVVAAIAVDR
SLAGWSGFGLAVQAYQKRAGAVIDYIDDLAQALNRRMMVRLVKGAYWDTEIKRAQERGLE
GYPVFTRKAMTDLNYLACARKLLALRPRLFPQFATHNALTVATLLELTGGESGFEFQRLH
GMGEALYAALCKDRHAIACRIYAPVGSHRDLLAYLVRRLIENGANSSFVATASDPDVPPE
TLLRRPADIIGSADNAGHARIPLPRDLFQPERINSRGIEFGERAAVNRLVTDVATERVSA
VPVNYSGPNDVNRAVTTARDGFSTWSRAPATERAAALERAADLLEQRSARFIAWLQREGG
KTLDDCVSEVREAADFCRYYAAEGRKLFGDAQPLPGPTGERNTLRLRGRGVFAAISPWNF
PLAIFMGQVTAALMAGNAVVAKPAEQTPRIAVEAVRLLHEAGIPVKALHLIQGDGHIGAA
LVAHRDIAGVVFTGSTEVARSINRVLAGKDGPIVPLIAETGGINAMIADATALPEQVADD
VATSAFRSAGQRCSALRLLFVQEDVADRIIEMIVGAARELKIGDPSDPATHVGWVIDDDA
RKRLDAHIARMTREARVHFAGPAPSSGSFVAPHIFELDDARHLTEEVFGPILHIVRYSSS
ALDGVLEAIADSGYGLTLGIHSRIDDTVEAVVDRLQVGNVYVNRNMIGAVVGVQPFGGHG
LSGTGPKAGGPHYLTRFATEQTVTVNTAAAGGNAALMTGEG
>Caulobacter Q9AA07 Q9AA07_CAUC
MTDWDSLDAGKYRDEAAVIADLLAAKPLSSEDRAAVRAEAEALVRGARRSVRKQGVVESF
LQEFSLGTREGLALMCLAEALLRTPDDDTRDKLIAEKIGSADWASHLGGSDSLFVNASTW
GLMLTGKIVEPDETARNDMPGFIKKLAGRLGEPVIRAAVGQAIRIMGEQFVLGRTIEAAI
KRAAAEGDMCSFDMLGEGARTAADAARYEKAYADAIETVGKLSNGAGPEAGHGVSVKLSA
LCPRYEATHEDRVWEELYPRTLRLAKIAARHNLNFTIDAEEADRLALSLKLLDKLCREPE
LGDWTGLGLAVQAYQKRCGEVIARLKALSEETGRRLMVRLVKGAYWDSEIKRAQVAGRPD
YPVFTTKPATDLSYLVNAKALIEAAPHLYAQFATHNAHTLAAVVRMAKNTGVKIEHQRLH
GMGEALYKAADDLYDGITLRAYAPVGGHEDLLPYLVRRLLENGANTSFVHALLDERVPVE
KVVTDPIDTVEAHPDRHAKIPTPINVYGERRVNSAGLDLSVKADRERLSAAVAAQDGVTL
SAGPLVGGKVVAGGAPLPLIAPANDQKTVGVVSEAQSAQIDEAFKLARAAQPAWDRAGGV
ARAQVLRAMGDALEANIERLIALLSREAGKTLSDGIAEVREAVDFCRYYAMLAEDQFGEA
EILKGPVGETNSLRLAGRGVFVCISPWNFPLAIFTGQIAAALAAGNAVLAKPAEQTPLIA
FEAVKLYHAAGLDPRLLALLPGRGETVGAALTSHEDLDGVAFTGGTDTAWRINQTLAARQ
GPIVPFIAETGGLNGMFVDTTAQREQVIDDVIVSAFGSAGQRCSALRLLFLPHDTADHII
EGLKGAMDALVLGDPALAVTDVGPVIDAEAKDALDKHLVRLKSDAKVLHALAAPAGGTFF
APVLAEIPTADFLEREVFGPVLHVVRYKPENLEKVAGALAARRYGLTLGIHSRIESFAAD
VQRLVPAGNAYVNRSMTGAVVGVQPFGGEGLSGTGPKAGGPHALLRFAVERALSVNITAQ
GGDPALLNL
>Rhizobium Q98L65 Q98L65_RHILO
```

121

```
MTAGPLPDARSTGPAMPLDTIRQQIRANYLPDEDEAVKRLAEATGLSAADRDAISARAAD
LVRAVRGSSDPRLMEVFLSAYGLSTKEGVALMCLAEALLRVPDTETMDDLIADKIAPHDW
SAHSGGSSSIFVNASTWALMLTGRVLDEGEGGIEGTLRSMVRRLGEPVIRKAVAAAMREM
GEQFVLGRTIAEAVKRGRPMTQKGYLYSFDMLGEAARTEADALRYHKAYADAISSLDAGS
NGPDIRQNHGISVKLSALHPRYEVAQKEEMLPVMAERLLSLALAARHSRMGLNIDAEEAD
RLDLSLDVIERVLAEPELAGWNGFGVVVQAYGPRAAFVIDWLYALARKYDRTIMVRLVKG
AYWDTEIKRAQTLGLAGYPVFTRKANTDVSYMACAKKLLGMTDRIYPQFATHNAHTVAAI
LSMATNRDSFEFQRLHGMGEALHETVRQAEGTRCRIYAPVGAHSDLLAYLVRRLLENGAN
SSFVHQLTDEDVEPEDIARDPLETVESQGPAANPAIARPSQIFGAGRRNSRGFDITDTVT
LAAIDKARAAFAGPDRWHAKPITRAAGYGKPHPVVNPAKPDEVVGTVHEAAAKQVAIAVR
IAVEAQPAWAKRPVGERAAILNRAADLYEANAVEFFALATREAGKSLADGVAEVREAVDF
LRYYAAEAANAETGTQARGAIVCISPWNFPLAIFTGQIAAALVTGNSVIAKPAEQTPLIA
FRAVELLREAGVPEDIIQLLPGDGPSVGGPLTADPRIAGVCFTGSTEVAKLIEKQLAETA
APDAMLIAETGGLNAMIVDSTALPEQAVRDILASAFQSAGQRCSALRVLYVQKDVEKKML
DMLKGAMEALNIGDPWRISTDVGPVIDDEAQASIRDYCTRKGLEGRLIAKLEAPKDGRFV
APHVFRVKGIEEMEREVFGPVLHVASFDADEIDAVIAAINRKGYGLTFGLHTRIEGRVQH
FVDGIHAGNIYVNRNQIGAVVGSQPFGGEGLSGTGPKAGGPHYLRRFRKGPEAGTEVGEG
HKVTATELADNLPDPALGGWSTRPDRIAILRKHLRGKGAAAIGAAAAIDFGQVDLPGPTG
EANTLSLSPRGRVLCLGPDAETLLAQTIQALAAGNAVLAVAPGAPAALSALTGKGLPLAA
IDGRPDPVEARALRVDVVAFSGTPEAARIVRQVIADRTGPIVPLVSEVLNPAAYAHERAV
CVDTTAAGGNASLLAAA
>Halorhodospira Q2CQE2 Q2CQE2_E
MSAFVHPEPEQLLPAERRALAAAYRIDEPTRSRALLEEADCDPATRSRIQERARGLVHGM
IRAQRRQSSLTALLHEYDLSSEEGVALMCLAEALLRIPDAPTADQLIHDKLGTGGWSSHL
GRDRRLLVNAATLGLALTGRILDTRDAERWFGDRLHTAIARRGAPLIRRAVRRSMGLLGE
TFVLGRDIPEAQRRARKLEAKGYRYSYDMLGEAARTEADAEHFFQAYCRGIEHFGRSADP
DAPMDARAEVSVKLSALDPRFEPGQEERVQATVIPRLQALCRRAREAGIALCVDAEEAAR
IDLTLDVLEAVMADPELADWDGLGIAVQAYQKRAPEWIDWLAERAGHYRRRLRIRLVKGA
YWDTEIKDSQIQGLDDYPVFTRKAASDVCFLACARRMLRHPQQIYPQFATHNAHTVAAVM
ELADEQPFEFQRLHGMADDLYDQLVDARPGRGVPVRIYAPVGQHEALLPYLVRRLLENGA
NSSFVNRIHEGDVEELIADPVEHLRSRTTLRHPHLPLPSGIFGPERVNSRGIDFSNRQET
AALAAAMTTAAEPAREARPIINGQGATEADGTWAEVCSPTDTAQHVGRVLWAGHEHLEQA
LASAAAAWPRWAATPVDERARALERLADLYEAHTAELMTLCTLEGGKTLKDGIAEVREAV
DFCRYYAVQARRLMGEPTPLPGPTGETNALQLHGRGTYLCISPWNLPLAIFTGQITAALA
AGNAVIAKPAEQTPLIAHRAVELMHQAGIPGDVLHLLPGEGRRIGPPLVADRRIDGVAFT
GSVATAQQIHRTLAERDGPIVPLIAETGGLNALIVDSSALPEQAVVDVLRSAFFSAGQRC
SALRLLCIQEDIAEPFLAMLRGAMDALRVGDPRWLATDVGPVIDSDARARLEAHHEAMAA
AGRVVHRTPLGQAGERGHFVPPSLYRLDAIEDLQEEFFGPMLHYTTWRAGELDSVVERIN
AAGYGLTFGVHSRIDSHREMATRSIRAGNAYVNRDIVGAVVGSQPFGGEGLSGTGFKAGG
PNYLLRFVNERVVTENTAAAGGNASLFALGEDDEA
>Marinobacter Q36JQ3 Q36JQ3_MAR
MSLQQSVAPELNDIRQAIRANYLADEHEVIHRLIAEAQLSDETRKAISARAAELVRDVRN
SARPTIMEKFLAEYGLTTKEGVALMCLAEALLRVPDNLTIQDLIEDKITSGNWGAHVGKA
KSGMINSATLALLMTSNLLKDSERQSVGDTLRKLVKRLGEPVVRTVAGQAMKEMGRQFVL
GRTIEEAQDRGKSQEERGYTYSYDMLGEAARTDADAQRYYQAYSDAIDSISKRCKGDVRT
NPGISVKLSALLARYEYGHKERVMNELMPRALKLARKAAAANMGFNIDAEEQDRLDLSLD
VIEAILSDPELKDWQGFGVVVQAFGKRASQTLDWLYALSEKLDRRIMVRLVKGAYWDAEI
KRAQVMGLSDFPVFTRKACSDVAYLAGARKLLGMTDRIYPQFATHNAHSVSAVLELAKDL
SRDKFEFQRLHGMGESLHDQVLEDSGVPCRIYAPVGAHKDLLAYLVRRLLENGANSSFVN
QIVDTSITPEEIAKDPIDVVVGLGHNLSSKAIVHPSKIFGEQRRNSKGWDITDPVTVAEI
DEGRNRYKSHQWKGGPILAVDSVSDEVVEVRNPANPDDLVGHITYTSEADISSALGAAQE
GFKQWSAVPAEERAAMIRRVGDLYEENVHELFALTTREAGKSLLDAVAEIREAVDFAMFY
ANEGIRYKNDGEARGVMCCISPWNFPLAIFTGQILANLAAGNAVVAKPAEQTSLLAFRAV
ELMHQAGIPRAAIQLLPGTGATVGSGLTSDARVTGVCFTGSTATAQRINKAMTEHMEPDA
PLVAETGGLNAMIVDSTALPEQVVRDVLASSFQSAGQRCSALRMLYVQKDIADNLLEMLY
GAMEELGIGDPWQLSTDVGPVIDENARKKITDHCQKFEQQGKLLKKLNVPEKGLFVSPAV
LQVSGIEELEEEIFGPVLHVATFEAKDIDKVIDAVNAKGYGLTFGIHSRVDRRIEHIASR
```

```
IKVGNTYVNRNQIGAIVGSQPFGGEGLSGTGPKAGGPQYVRRFLRGEVVEKPAQSSDKVL
SADKAQKLIDKLAKVEVPEAEGRQALLVPFFGKVPAPLDEGYEDMPGPTGEQNHLSCHGR
GLVLCLGPDAESAVEQAGTALSQGNKVVVIAPGAEKALADAIKAGLPVVASDGMLDPDAL
SHLTGFEAVVSVAEKPLLKQYRMALSKRDGALLPVITEHKLDQRYVIERHLCIDTTAAGG
NASLIASAE
>Yersinia Q8ZF67 Q8ZF67_YERPE B
MANTTMGVKLDEATRDRIKSAAQRIDRTPHWLIKQAIFNYLEKLESNSELPELATTSSLS
LQDTEDAIPQLTENTHQPFLDFAEHVLPQSVTRAAITAAYRRPETEAIPMLLEQARLPAD
LAQATHKLAYSIAEKLRNQKSAHGRAGMVQGLLQEFSLSSQEGVALMCLAEALLRIPDKP
TRDALIRDKISNGNWHSHLGRSPSMFVNAATWGLLFTGRLVSTHNEAKLSGSLNRIIGKG
GEPLIRKGVDMAMRLMGEQFVTGETISEALANARKLEDKGFRYSYDMLGEAALTEADAQA
YLLSYQQAIHAIGKASNGRGIYEGPGISIKLSALHPRYSRAQYERVMDELYPRLLSLTLQ
ARQYDIGINIDAEEADRLEISLDLLERLCFEPQLAGWNGIGFVIQAYQKRCPSTIDAVID
MAQRSRRRLMIRLVKGAYWDSEIKRAQIDGLEGYPVYTRKVYTDVSYLACARKLLAVPNL
IYPQFATHNAHTLSAIYHLAGQNYYPGQYEFQCLHGMGEPLYEQVVGKVAEGKLNRPCRI
YAPVGTHETLLAYLVRRLLENGANTSFVNRIADATLPLDELVADPVSAVEAMAAAEGQLG
LPHPRIPLPRELFGKDRANSSGVDLANEHRLASLSSALLASASQVWRAEPVIDAEQDNGD
ALPVINPAEPADVVGYVREATEGEVSRALDAAARAGAIWFATPPAERAAILIRAAELMEN
QMQTLMGILVREAGKTFSNAIAEVREAVDFLHYYAGIVRDNFANDSHRPLGPVVCISPWN
FPLAIFTGQVAAALAAGNSVLAKPAEQTPLIAAQAVRILLDAGIPQGVLQLLPGRGDSVG
ALLVNDARVRAVMFTGSTEVATILQRSIAGRLDPQGRPTPLIAETGGLNAMIVDSSALTE
QVVTDVVASAFDSAGQRCSALRILCIQDDVAEHTLQMLRGAMAECRMGNPERLSTDIGPV
IDAEAKTGIERHIQAMRAKGRKVYQAARTNSLDEKEWQRGTFIKPTLIELDSFDELQKEV
FGPVLHVVRFQRQHLNELVDQINASGYGLTLGIHTRIDETIARVTEKAKVGNLYVNRNMV
GAVVGVQPFGGEGLSGTGPKAGGPLYLYRLLSSRPDDALANTLAHQDGEQQQNVAGREAL
LTAHRAFTQWATEQQHDSLATLCQRYASLAQGGTVRLLPGPTGERNTYALLPRERVLCLA
DTESDTLTQLAAVLATGSQVLWPENDVQKALLPQLPTEVQSRITLTHDWQTANITFDAVI
YHGDADQLRTLCEQVAQIDGPIVSVQGFARGETNILLERLLIEHSLSVNTAAAGGNASLM
TIG
>Shigella Q3Z3A0 Q3Z3A0_SHISS P
MGTTTMGVKLDDATRERIKSAATRIDRTPHWLIKQAIFSYLEQLENSDTLPELPALLSGA
ANESDEAPTPAEEPHQPFLDFAEQILPQSVSRAAITAAYRRPETEAVSMLLEQARLPQPV
AEQAHKLAYQLADKLRNQKNASGRAGMVQGLLQEFSLSSQEGVALMCLAEALLRIPDKAT
RDALIRDKISNGNWQSHIGRSPSLFVNAATWGLLFTGKLVSTHNEASLSRSLNRIIGKSG
EPLIRKGVDMAMRLMGEQFVTGETIAEALANARKLEEKGFRYSYDMLGEAALTAADAQAY
MVSYQQAIHAIGKASNGRGIYEGPGISIKLSALHPRYSRAQYDRVMEELYPRLKSLTLLA
RQYDIGINIDAEEADRLEISLDLLENLCFEPELAGWNGIGFVIQAYQKRCPLVIDYLVDL
ATRSRRRLMIRLVKGAYWDSEIKRAQMDGLEGYPVYTRKVYTDVSYLACAKKLLAVPNLI
YPQFATHNAHTLAAIYQLAGQNYYPGQYEFQCLHGMGEPLYEQVTGKVADGKLNRPCRIY
APVGTHETLLAYLVRRLLENGANTSFVNRIADTSLPLDELVADPVTAVEKLAQQEGQTGL
PHPKIPLPRDLYGHGRDNSAGLDLANEHRLASLSSALLNSALQKWQALPMLEQPVAAGEM
SPVINPAEPKDIVGFVREATPREVEQALESAVNNAPIWFATPPAERAAILHRAAVLMESQ
MQQLIGILVREAGKTFSNAIAEVREAVDFLHYYAGQVRDDFANETHRPLGPVVCISPWNF
PLAIFTGQIAAALAAGNSVLAKPAEQTPLIAAQGIAILLEAGVPPGVVQLLPGQGETVGA
QLTGDDRVRGVMFTGSTEVATLLQRNIASRLDAQGRPIPLIAETGGMNAMIVDSSALTEQ
VVVDVLASAFDSAGQRCSALRVLCLQDEIADHTLKMLRGAMAECRMGNPGRLTTDIGPVI
DSEAKANIERHIQTMRSKGRPVFQAVRENSEDAREWQSGTFVAPTLIELDDFAELHKEVF
GPVLHVVRYNRNQLPELIEQINASGYGLTLGVHTRIDETIAVVTGSAHVGNLYVNRNMVG
AVVGVQPFGGEGLSGTGPKAGGPLYLYRLLANRPESALAVTLARQDAEYPVDAQLKAALT
QPLNALREWAANRPELQALCTQYGELAQAGTQRLLPGPTGERNTWTLLPRERVLCIADDE
QDALTQLAAVLAVGSQVLWPDDALHRQLVKALPSAVSERIQLAKAENITAQPFDAVIFHG
DSDQLRALCEAVAARDGAIVSVQGFARGESNILLERLYIERSLSVNTAAAGGNASLMTIG
>Colwellia Q47VW3 Q47VW3_COLP3
MLFTGSLITDCPIRQKIREFYRIDENVAVDHILPAAEVNVSARSRAWERARKMVLKIRQD
QEGNGAIDSLLNEYSLSSEEGVVLMCLAEALLRVPDKHTQDELIRDKISQGQWSSHLGAS
DSLFVNASSWGLLLTGSMVNYADKRKKEQFGLLKKTVGRLGEPVIRKSMNYAMKVMGKQF
VMGETIKAATERAATKEQQGYVYSYDMLGEGARTMADANMYLKAYQDAINAIGEVAVASG
```

```
KNDPRKVPGISIKLSAIHPRYEFSHKERVMTEIVPKLKALCLQAKQYNIGLTVDAEESER
LDISLDIIEAVFSDSELGDWQGFGIALQAYQKRAIHVVDWLRDLTLRTERKMMVRLVKGA
YWDTEIKNAQKDGLNHFPVFTRKSSTDVSYHACANKLLEYRDTIYPQFATHNAYTAATIV
ELAGDDKAGFEFQCLHGMGDSLYDQIVKEESIQCRIYAPVGHHEDLLAYLVRRLLENGAN
SSFVNAIVDEDQPVESLLEDPVEKTQRLKAKYNNQIIMPIDLYRGEGEKGRDNSKGLDLT
DINEIIPLKAALDNWFVDHLLNKNEVPDGANAVMNPANRSEIIGFHQHASHGDMLLMIDK
AETAFSTWSKTPAVDRAALLCRIGDILERHIDELIAMCIKEAGKVAQDGIDEVREAVDFC
RYYAARAIELSSDERLEARGVVLCISPWNFPLAIFLGQVAAAIATGNTVLAKPAEQTGLI
ALRAIELMKSVGLPENVVQAVIARGSAVGNTIIPDSRIQTVMFTGSTETGTRISQTLSDR
GGDQVPLIAETGGQNCMIVDSTALPEQVVDDVISSGFQSAGQRCSALRVLFLQEDIADNV
ITMLQGALAELHIGNPAKLSTDIGPVIDQKALDALNAHAEYMKSHGKLLYQCEFSEDVVD
EDGHFFFAPRLYEIDDISVLKQEVFGPCVHIVRFKGNEIESVVDKINGTGFGLTMGIHTR
IEHRAINLAKLSRAGNIYINRNMIGAIVGVQPFGGRGLSGTGPKAGGPNYLTRLVKEKAT
PDERDFNFSANKTITLSGDAAENHQANHLMDKANWAEKEWRSTELNTRISCVRQLLAKIA
HVEIVDDLAEDLNHTLVLARSQLIDIEKRLKKPQQLPGPTGESNIIYLENRGNIICYADE
NVSFHFWVQSIVTSLATGNTVIAVVSDLFYQEALEFRDKFVATGAGKDVFQVAKQCHLEA
MLAHPALSGVVVDSGCEIKHYISEKLALRHGAILPVITSEYFDTMIQRLVTEKTISIDTT
ASGGNTSLMTLVEED
>Ralstonia Q46VY6 Q46VY6_RALEJ
MATTTLGVKLDDASRDRLKRVAQSIDRTPHWLIKQAIFTYLEQIERGHLPHEAMAGGAGE
ADGADGAEMSHADGAPQPFLEFAQSIQPQSVLRAAITSAYRRPETDCVPVLLEQARLPQQ
QAEAAIKMAKKLATALREQKVGTGREGLVQGLIQEFSLSSQEGVALMCLAEALLRIPDKP
TRDALIRDKISGANWQSHLGQSPSLFVNAATWGLLLTGKLVATHTESGLSKALTRIIGKG
GEPLIRKGVDMAMRLMGEQFVTGETISEALANARKYEAQGFRYSYDMLGEAAMTEEDAQR
YLASYEQAIHAIGQASRGRGIYEGPGISIKLSALHPRYSRAQHERVINELYERLKSLTLL
ARQYDIGININDAEEADRLEISLDLLERLCFEPELAGWNGIGFVVQGYQKRCPFVIDYLID
LARRSRHRLMIRLVKGAYWDSEIKRAQVEGLEGYPVYTRKVYTDVSYVACARKLLSVPDA
IYPQFATHNAHTLSAIYQIAGHSYYPGQYEFQCLHGMGEPLYDQVVGPIADGKFNRPCRI
YAPVGTHETLLAYLVRRLLENGANTSFVNRIADDTISLDELVADPVAVVETMHRDEGTLG
LPHPKIPSPRGLYGKSRANSAGIDLANEHRLASLSSALLAGTSERVVADPMLGTEVPRAA
DAITTPVLNPADHRDVVGQVTEASQADVDAALQAAANAAPIWQATPPDVRASALERAADL
MEAQMQSLMGIIMREAGKTFSNAIAEVREAVDFLRYYAAEVRRSFDNETHRPLGPVVCIS
PWNFPLAIFTGQVAAALAAGNTVLAKPAEQTPLIAAAAVRILREAGVPAGAVQLLPGRGE
TVGAALVGDARVKGVMFTGSTEVARILQRNIAGRLDAAGRPIPLIAETGGQNAMIVDSSA
LAEQVVGDVVNSAFDSAGQRCSALRVLCLQEDVADRVLAMLKGAMNELTMANPDRLSTDV
GPVIDEEARGNIVRHIDAMRAKGRRVHQADPNAPQGASCRHGTFVPPTLIELDSIDELKR
EVFGPVLHVVRFPRAALDTMIGQINGTNYGLTMGIHTRIDETIAQIVQHAEVGNLYVNRN
IVGAVVGVQPFGGEALSGTGPKAGGPLYLHRLLSVCPQDAVVRSVRHAEASGNPPAAPER
PVAAQALKDWAKTEMPEVAAACDRFAEASAAGISVTLRGPTGERNTYSLLPRHHVLCLAA
QEQDLAIQLAAVLAVGSQAVMAENPVSRGLFARLPKGVQSRVRIVSDWTSPEPAFDAVIH
HGDSDQLRTVCEQVAARPGPIIGVQGLAQGEPNIALERLLIERSLSVNTAAAGGNASLMT
IG
>Salmonella Q57QN1 Q57QN1_SALCH
MGTTTMGVKLDDATRERIKMAASRIDRTPHWLIKQAIFSYLDKLENSDTLPELPALFAGA
ANESEEPVAPQDEPHQPFLEFAEQILPQSVSRAAITAAWRRPETDAVSMLMEQARLSPPV
AEQAHKLAYQLAEKLRNQKSASGRAGMVQGLLQEFSLSSQEGVALMCLAEALLRIPDKAT
RDALIRDKISNGNWQSHIGRSPSLFVNAATWGLLFTGRLVSTHNEANLSRSLNRIIGKSG
EPLIRKGVDMAMRLMGEQFVTGETIAQALANARKLEEKGFRYSYDMLGEAALTAADAQAY
MVSYQQAIHAIGKASNGRGIYEGPGISIKLSALHPRYSRAQYDRVMEELYPRLKSLTLLA
RQYDIGLNIDAEEADRLELSLDLLEKLCFEPELAGWNGIGFVIQAYQKRCPLVIDYLVDL
ASRSRRRLMIRLVKGAYWDSEIKRAQMEGLEGYPVYTRKVYTDVSYLACAKKLLAVPNLI
YPQFATHNAHTLAAIYHLAGQNYYPGQYEFQCLHGMGEPLYEQVTGKVADGKLNRPCRIY
APVGTHETLLAYLVRRLLENGANTSFVNRIADATLPLDELVADPVEAVEKLAQQEGQAGI
PHPKIPLPRDLYGEGRINSAGLDLANEHRLASLSSALLSNAMQKWQAKPVLEQPVADGEM
TPVINPAEPKDIVGWGREATESEVEQALQNAVNQAPVWFATPPQERAAILQRAAVLMEDQ
MQQLIGLLVREAGKTFSNAIAEVREAVDFLHYYAGQVRDDFDNETHRPLGPVVCISPWNF
PLAIFTGQIAAALAAGNSVLAKPAEQTSLIAAQGIAILLEAGVPPGVVQLLPGRGETVGA
```

```
QLTADARVRGVMFTGSTEVATLLQRNIATRLDAQGRPIPLIAETGGMNAMIVDSSALTEQ
VVVDVLASAFDSAGQRCSALRVLCLQDDIAEHTLKMLRGAMAECRMGNPGRLTTDIGPVI
DSEAKANIERHIQTMRAKGRPVFQAARENNDDAQEWQTGTFVMPTLIELENFAELEKEVF
GPVLHVVRYNRNQLAELIEQINASGYGLTLGVHTRIDETIAQVTGSAHVGNLYVNRNMVG
AVVGVQPFGGEGLSGTGPKAGGPLYLYRLLAHRPPNALNTTLTRQDARYPVDAQLKTTLL
APLTALTQWAADRPALQTICRQFADLAQAGTQRLLPGPTGERNTWTLLPRERVLCLADDE
QDALTQLAAVLAVGSQALWSDDAFHRDLAKRLPAAVAARVQFAKAETLMAQPFDAVIFHG
DSDKLRTVCEAVAAREGAIVSVQGFARGESNILLERLYIERSLSVNTAAAGGNASLMTIG
>Xanthomonas Q4UPS6 Q4UPS6_XANC
MNAQPDTFASSHLARPLLAPELPAAPGALRAAITAAWLKDETEHVRELLEQARLPAAEQA
KVQALAADLVTRVRARAQDQGAIEAFMRQYDLGSEEGVLLMCVAEALLRIPDQDTADKLI
RDKLGEADWKKHVGGSDSVLVNASTWGLMLTGKLVQLNDLTRADVPGAFKRLIGRVGEPV
IRLAVRQAMKIMGHQFVMGRTIGEALARSRKGDNANYRYSFDMLGEGALTMKDAQRYLQA
YRDAIHAIGRSGSFVGTDVFAAPSISIKLSALYPRYEHAKRARVMAELVPGVLELAQLAK
SYGIGYTVDAEEADRLELSLDIIEATFSDPSLDGWEGYGLAVQAYQKRTPYTIDFLADLA
RRVGRRIPVRLVKGAYWDAEIKRAQIDGHPGYPVFTRKQNTDVSYLACARRMFAHSDALY
PMFATHNAQTIAAVRAISAGKTYEHQKLHGMGDDLYAEVIPADRLGLPCRVYAPVGSHED
LLPYLVRRLLENGANSSFVNRITDEDVAIEDLIRDPVEAVSSFASIPHPKIPLPADLLRS
QNQNRKNSMGANLANDNDLRQLADQLTAAIKPWKAAPLVPGAVITTPSEAVFNPADRRET
VGHWQPADPATVQKALASAAAAQPAWNRTPAASRATILEHAADLLEARMPEFMAICVKEA
GKTLPDAVAEVREAVDFLRYYAAQARAQFGAPERLPGPTGESNELQLHGRGVFVCISPWN
FPLAIFLGQVAAALAAGNTVIAKPAEQTNLVGFAAVKLLHEAGVPEAAVQFLPGDGATVG
AALTNDPRVAGVAFTGSTDTARIINRTLAARDAAIGVLIAETGGQNAFIADSSSLPEAVV
KDAISSAFISAGQRCSAARVLFVQDDIADKVMTMLAGAMGELKIGDPALLSTDVGPVIDA
DALKILEEHASRMDSEARLIGTTTLDSATAHGSFFAPRAYELKSLAQLQREIFGPVLHII
RWKADQLDSVIDQINATGYGLTLGVHSRIDETIDRITSRVAVGNVYVNRNQIGAVVGVQP
FGGQGLSGTGPKAGGPHYLLRFATEKVVTVNTTAAGGNASLLTLGD
>Jannaschia Q28LK1 Q28LK1_JANSC
MSDLSHFRQAIASAHLEDDALALDRLRSAHDIAPPVRKAASTRAARLVASIRAEDRGGLM
EVFLAEYGLSTAEGLALMRLAEALLRVPDDATVDALIEDKIVPAQWSGHRGKSKSGLVNA
STLALMLTGRVLSDEDGAGIAGTLKGAVKRLGEPVIRKATRRAMKEMGNQFVLGQTIAEA
MKRGRDRAAQGYTYSYDMLGEAAVTADEARIFYASYADSIAYLAKQATHDDIRQNPGISV
KLSALHPRYEEAQRDRVLSELVPRVLELAQAAKAARMGFNIDAEEADRLDLSLDVIEAVL
RDESLAGWDGFGVVVQAYGKRCAPVIDWLYALAEIHDRRIMVRLVKGAYWDTEIKRAQVA
GLPGFPVYTAKAATDVSYLCCARQLLDRTDRIYPQFATHNAHTVAATIEMAEDPLAFEFQ
RLHGMGEALHEQVRAANMTRCQIYAPVGAHRDLLAYLVRRLLENGANSSFVHKIMDEDTP
PEAVAEDPFLTLSHPATPVTLPGDLYAPERPNSHGIDLTDPASLAALDAARAPFREATWQ
AGCGMGETQEVRNPADPSDLVGYVTLSTAGDAQTAIETAQAWDAANRTDVLNRAADLYET
HAPELFALLAREAGKTLPDCEAELREAVDFLRYYAARIDDLTDPALGRVTCISPWNFPLA
IFTGQISAALAAGNAVLAKPAEATPLIAARAVALLHEAGVPKTALQLLPGEGGTVGAALT
SDPRIDGVAFTGSTATAQVIHRAMASNLAPSAPLIAETGGLNAMIVDSTALPEQAVKDVV
MSAFRSAGQRCSALRCLYVQKDIAPAFLKALYGAMDELSLGDAWPLANDIGPVIDARAQA
GIAAHIATARADGRILHEGTAPVLGTFIAPTAIRVTSIADMSEEIFGPVLHVATFKSSDL
DRVIDDINATGYGLTFGIHSRIDDRIADVTGRIMAGNLYVNRNQIGAIVGSQPFGGQGLS
GTGPKAGGPSYVPRFCRTPLGEGTPGTPADIEGAAHDLHVALPADRTLTTQDLPGPTGET
NRLTTHPRGRVLCLGPGSAAARDQASAATATGCAPVLAPDLPAAALTDLRAFDAVIYWGD
DAAARAYRKVLADRDGPILPLIMDADPKPRLILERHTCIDTTAAGGNAALLAAVA
>Idiomarina Q5QY28 Q5QY28_IDILO
MFKASEVLSKQHQAEDLKSLQQAITDNYIVDEDEYMRELLPLVPADDDTVSAVTERAAKL
VEQVREQADNGVVDAFLQEYSLDTKEGIILMCLAEALLRIPDAYTADALIQDKLSGGDWQ
KHMGQSASWLVNSGTWGLALTNSVTNPTGKAMETPRGAFRRLVRKLGKPIIRKATYTAMQ
IMGKQFVLGRTIEEALKESRDNRDKGYTHAYDMLGEAALTMKDADYYKQQYVNSIKTITK
EEFNNPDAPRPTISIKLSALHPRYEASNHERVLTELATTLTELVKLAKEADVGVTIDAEE
ADRHELSMELFEKVYRSGVCKGWPRFGLVVQAYSKRALPTLCWITALAKECGDEIPVRLV
KGAYWDNEIKWTQENGLLGYPVFTRKSHSDISYLACARYLLSDDTDGAIYPQFATHNAQT
FMAIQQMNETHQRRIEYQRLHGMGDSLYDTIMEQNPGMVVRIYAPVGPHKDLLPYLVRRL
LENGANSSFVHKLLDADTPVNDLVEHPMKTASGYEKYANSKIPLPSEMYGDRTNSLGLNM
```

```
NIHSQADDFIAAVQQYRDKQWQGGPIVDGKTIETDHHVSITSPQETSKQVGSIYWGDKKL
SEQALKSANAAYKSWRKVPTDERAECLEKFADLMEANRNELIALCSVEAGKGLQDGIDEV
REAVDFCRYYANQARELMGGPIQLPGPTGEDNELFVEGRGTFICISPWNFPLAIFVGQVA
AALVTGNSVIAKPAEQTGLIAYRAVQLALEAGIPGNVLHFMPGSGAEVGSYLTSQEDIGG
VCFTGSTYTAQAINRALAARTGPIVPLVAETGGQNAMMVDSTALPEQVVTDIVASAFQSA
GQRCSALRVLFVQDDVADRVLDLLRGAMEELQVGDPLLHETDVGPVIDGIAKTNLEQHIS
DIQQAGRLIARAPLPDYAMGGTFVAPTAIEIDSISQLVKENFGPILHVIRFKTSEIDEVI
ESINNTGFGLTFGIHSRNETFAHDVASRIDVGNVYINRNQIGAIVGVQPFGGRGMSGTGP
KAGGPHYLTRFITEKTRSDNITAVGGNATLLSLDD
>Oceanospirillum Q2BPW3 Q2BPW3_
MSITIEQSRHAIRNSYLADEAAVITGLINNSGLDETERQSISASAADLVRKVRSESSPSM
MEKFLAEYGLTTKEGVALMCLAEALLRVPDSLTIDALIEDKVVSGNWKGHLGQSASSLVN
SSTWALLITGKLLKDDEAKGIGSTVKGMIKRLGEPVVRTAVAQAMKELGRQFVLGRDINE
ATKRAKKLEKQGYSYSYDMLGEAARTDADAIRYHQAYSQAISELTPACTSNDIRTNPGIS
VKLSALHARYELGQKTRVMQELVDRTLTLALQAKDANMGFNIDAEEADRLDLSLDVIEAV
LADKRLADWDGFGIVVQAFGPRAAHVLDWLYATAQRLDRRIMVRLVKGAYWDAEIKRAQV
LGLDGFPVFTRKVNSDLSYMCCAEKLMQMTDRIYPQFATHNAHSAAAILFMAKKHNVDNF
EFQRLHGMGESLHDTVLKSNNTRCRIYAPVGAHRDLLAYLVRRLLENGANSSFVNQIVDT
RITPEEIAKDPFSQVLESQDQISNNYIPKPAEIFGDKRLNAKGWDITDHVTLEQLEPARE
AFQTTQWHAKPMIAGDATSSDTSEVRNPANPEDLVGMVTEANPADIESAIAAAELGFQRW
SKTAPSERAACLRRVADLFEANAPELFALASREAGKTLLDAVGEIREAVDFARYYANEAE
KHAESEARGVITCISPWNFPLAIFAGQVLAGLASGNTVIAKPADQTPLMAARAVALMHEA
GIPNDVIQLLPGSGINVGAPLTSDPRISAVCFTGSTLTAQRINKVMAENMAPDAPLVAET
GGMNAMIVDSTALPEQVVRDVLMSSFQSAGQRCSALRVLYVQEDIADNLLEMLYGAMDEL
RMGNPWLLSTDIGPVIDSNAQAKIEAHCEKFAAKGKLLKKLNTPEGGTFVAPTVLKVEGI
EEIEEEIFGPVLHVATFAAKDLDQVVDSINSKGYGLTFGLHTRVDNRVERIINQIKVGNI
YVNRNQIGAIVGSQPFGGEGLSGTGPKAGGPHYVKRFTRSAISQGTDAPSGQSVNTETLE
QLIQQLDLTTQGLKADPEQLQKVFAGIASELPKPVTEATEMPGPTGELNLLSETGRGVVL
CLGPDAESAVKQAATALFQGNAVVIIAQGIADLLKQCADAKLPVQGMVGQLTAEAITEAH
GFAAVSCNASHDVLKTIRQALAAREGALLPLITELDQPERFVMERHLCIDTTAAGGNASL
IATAG
>Alkalilimnicola Q34YB6 Q34YB6_
MSHFVHPQPDALLHPERAVLAEAYRADEAKVTSVMLERAALGETANRRIQAHARSLVMGM
ISAQKGEFGVDALLHEYDLSSEEGIVLMCLAEALLRVPDRYTADKLIHDKLTASHWETHL
GRDRPLFVNAATWGLLLTERIIDTDDRDRWLGGVLHHMVARAGEPVIRTAVRRAMGLLAD
TFVLGRDIDEALKRAHPNERKGYRYSYDMLGEAARTDADAQAYFEAYRNAIHRVGKAVDP
QARIRDRAGVSVKLSALDPRYEPGQEARVMGTVLPRLLQLCELARDYNIALCVDAEESWR
LDLSLDVIEAVLAEPGLADWEGFGLAIQAYQKRCYALVGWLEAQAARQNRALMVRLVKGA
YWDTEIKETQVQGLRDYPVFTRKAATDVAYLACARRLLSECPHLYPQFATHNAHTVAAVM
ELAGRQPYEFQRLHGMGEALYEQLLGRAGGPDIPCRIYAPVGSHEELLAYLVRRLLENGA
NSSFVNRIHEGDVEALVADPASSLRARESLRHPRIPLPRDLYGEQRLNSIGLDFSDRHEM
VWLAEAMEAASEHAWRAAPLIDGQVKAGAGEAVHCPADRERQVGRVVWSEHADVEHALEA
AVAGRETWANTRAETRARALEQIAELYEAHGAELMALCTREAGKGLKDGIAEVREAADFC
RYYAHQARRLFGQATVLPGPTGERNELRLHPRGTFLCISPWNFPLAIFTGQVTAALAAGN
TVVAKPAEQASLIAHQAVALMHRAGIPETALHLLPGDGGEIGPHLVADPRISGVAFTGGT
DTARRIQQGLAQREGPIVPLIAETGGLNAMVVDSSALPEQVVVDIIRSAFHSAGQRCSAL
RLLCVQEDIADDLLAMLKGAMDALTVGDPHWLATDVGPVIDPAARDGLLAHHERMVSAGR
LLHQAPLRPECERGSFVAPALYRLDRIDQLGREHFGPLLHWVTWRAGELDALVDRINGLG
YGLTLGVHSRVDETAAAVVNRARVGNAYVNRDMVGAVVGSQPFGGEGLSGTGFKAGGPHY
LLRFAAERVVTVNTAAAGGNASLFAMGEDD
>Acinetobacter Q6FBR9 Q6FBR9_AC
MSLEFGMNMMDNDPQMDTAKPHFGYISEFKEKTELEQHINDAWRRPEPETVETLLQAASI
SDELNHKIYELAFELANGLRERKTSSGKAGIVQGLLQEFSLSSQEGIALMCLAEALLRIP
DAATRDLLIRDKINQGNWKDHVGQSSLMFVNAAAWGLMLTGKLMETPKQGSLSGLLTSIL
ARSGRGVIRKAVDVAMRMMGEQFVTGETIEEALDNAKSHEHKGFRYSYDMLGEAALTTHD
AERYFNDYTQAIHAIGQASNGKGVYDPGISIKLSALHPRYQRSQIARVHDELYGKVFNL
ALLSKKYDIGLNIDAEEADRLEISLELLERLCFEPELADWKGIGFVIQAYQKRCFYVVDY
```

VVDLAKRSKKRLMIRLVKGAYWDSEIKKAQIDGMHDYPVFTRKVHTDLSYIACAKKLLAA
PEQIYPQFATHNAQSLATIYELADPGKYYAGQYEFQCLHGMGEPLYEQVVGSRSDKKLGI
PCRIYAPVGNHETLLAYLVRRLLENGANTSFVNRIADQNLKIEDLIQSPFDEIAINAQRE
GQAGLKHPAIPLPHNLYGSLRQNSKGFDLANDATLTALNETAQQLEHKVWESRPLLAQTV
DQIDVSPQALLNPAVHLDIVGYVQEATLEQVDVALESATQAQQAWANTEKNHRAATLKKA
ANLMESRMQELMVLLCREAGKTYANAISEVREAVDFLRYYATQIESLKPTAQIEPLGTVL
CISPWNFPLAIFTGQIAAALGAGNCVIAKPAEQTPLIAAQAVHILHEAGIPKSVLQLLPG
RGETVGAKLSSDQRIQGVMFTGSTEVAKILQKTVAKRLSPSGHPIPLIAETGGQNAMIVD
SSALTEQVVIDVLNSAFDSAGQRCSALRILCVQEDNSKTLVTMLKGAMQQLIMGNPFLLK
TDIGPVIDLEAKQTIDQHIQKMRSKGYPVYQLIFNHEAASQAQLSQGTFVTPTLIELPNL
DDLEREVFGPVLHVITYKHGEIKQLLQHINAKGYGLTMGLHTRIDETINTVVQHAEVGNL
YINRNIVGAVVGVQPFGGEGLSGTGPKAGGPLYLYRLMASCTEKAIQTPFGIPAENVAVP
LTQLSLFREFSAWIKKQYPAHQLPNVVPLNVGHAFDLQGPTGETNRYMMLPRKRILAIAS
NELELCHQMLAIFAVNSQVALLESNPSLQKFGQDLPTVVRQAIIEIRGVEHGDFDAVLHH
GSTEELQVLQTQIANRQGPIVGITHLQNSEQVIPLERLVIEHAISVNTAAAGGNASLMTM
SE
>Novosphingobium Q2GAH2 Q2GAH2_
MTQTAPFAAFAPRHSVPSDLRRAITAATRRAERECMVMLLPEATLPQATRASAQALARKL
VEALRAKPRGNGVEQLVQEYALSTHEGVALMCLAEALLRIPDNDTRDDLIRDKIAGGDWL
AHLGGDRSLFVNAATWGLVVTGKLASSVDDTGLGAALTRLIARAGEPVIRRGVDLAMRMM
GEQFVTGETIDEALKRARPLEDRGFRYSYDMLGEAAMTASDAARYHADYQAAIHAIGKAA
ARRGVYEGPGISIKLSALHPRYARAQQDRVFAELLPRLKALAYLALRYDIGLNIDAEEAD
RLELSLDLLEELAMDPELAGWNGLGFVVQAYGKRCPFVIDWIVDLARRSGRRIMVRLVKG
AYWDAEIKRAQVDGQSGFPVYTRKAHTDVAYIACARRLLAATDVVFPQFATHNAQTLATI
HEMAGPDFAVGRYEFQCLHGMGEPLYDEVVGKDKLDRPCRIYAPVGTHETLLAYLVRRLL
ENGANSSFVNRIANPDVPVDEIVADPVAQVAADADPGAPHPLIALPAALYPDRRNSSGLD
LADESTLAALTDRFQAISAHSRHAHPSADLPHVNPRPVLNPANHRDVVGHVSEAAPAAAG
MAASLAAASRWSATPVAARAMILERAADAMQAAMPELIALIVREAGKSVSNAIAEVREAV
DFLRYYASQAPAMSGSRPLGVALCISPWNFPLAIFTGQVAAALMAGNPVLAKPAEETPLI
ASEAVRILHEAGVPDDALVFLPGDGAIGAALVAAPEIAAVLFTGSTEVGRLIQRQLATRL
SAEGRPIPLIAETGGQNAMIVDSSALAEQVVADVIASAFDSAGQRCSALRVLCLQEDVAD
HVLAMLRGALAELTVGPTDRLSADIGPVIAAEAREAIEAHVARMEAAGCPVHRLPLGTAT
NEGTFVAPTIIELSAPDLLTREVFGPVLHVIRFRRADMDAMVERINAWGYGLTFGLHTRL
DETVARVTAKAHAGNIYVNRNVIGAVVGVQPFGGHGLSGTGPKAGGPLYLRRLVASAPAA
PALELTERELPGPVGERNIYALRPVGRVLVAAQSPERLAALTSRVERLGGTPIAADDGWQ
AQGPFARALVEGDAAFILSFQQAVAALDGPIAPVLASADDDTMLVAEVSLSINTTAAGGN
ASLMAMA
>Legionella Q5X4L5 Q5X4L5_LEGPA
MLEKQSIHLPEGLRAAINKAYRMDELSLITELSEQAALDPQQMMAIKTSATKLVQSVRSE
RKKSTGIDSFLTEYALSSDEGIALMCLAEALLRVPDNATIDNLIKDKLAGGDWGAHRGQS
ESFFVNATTWALMLTGKVLTPEKAENTLTKALLKLVNRSSEAVVRKAVDKAMRIMSKQFV
MGRTINEALARAKKKEDRGYRYSYDMLGEAALTSADAARYFEAYKEAIISIGEKADKHSD
VYRRPGISIKLSALHPRYSEFQYERVMAELPPKLLALSRLAKDYGIALTIDAEESERLDL
SLDVIEKVFTDESLQGWNGFGLAVQSYQKRAFYVLDWVAALARSKQRRIMVRLIKGAYWD
SEIKKTQMQGFSEYPVFTRKVFTDVSFQACAKKILTMTDAIYPQFATHNAYSVAMILNLV
GGYRDFEFQCLHGMGNELYEQIVPANCYGIPCRIYAPVGSHEDLLPYLVRRLLENGANSS
FVNRIVDDKAPISELVEDPVAKSRSLLDKINKNIPLPEDIFLPVRKNSKGFDFTNRLERA
LLQQELAKIESKEWQASPMIAGRKLSSDLLQTVMSPQQPAYAIGSVQQATLDDVEVALNQ
AKLAFESWSKKAVEERASCLNRFADLLQANMSELMVLTCREAGKTWSDGIAEVREAIDFC
RYYAKKAQELMSSPQRFNGYTGELNELSLHPRGTILCISPWNFPLAIFTGQVVAGLVTGN
CVIAKPAEQTPLIAAYAVKLMHQAGIPEGVIQLIPGAGETIGAALVADKRIKAVLFTGST
DTANLINRTLATRGGEIIPLIAETGGQNAMIVDSSALLEQVVVDAVTSAFGSAGQRCSAL
RVLYVQEEVYPRTVELLKGAMAELVVGDPQWLSTDVGPVIDKEALSILKNHVENMRKHHE
ILYQCTVDDEALSGYFMPPTAIAIDSISALEKEVFGPILHVIQFKRKDLDKVINQINQTG
YGLTLGIHSRINETVDYIRQRVHAGNCYVNRNMIGAVVGLQPFGGEGLSGTGPKAGGPNY
LIRLCHERTYTVDTTAAGGNASLMSIPEEG
>Azoarcus Q5P8F4 Q5P8F4_AZOSE B

```
MREAVQPCKNLVHPSHARRARSRAQVTAGCVRERQPQSPIALSGVLLAMKPEIPDPATAA
LPAAARSELRRRIDTAWRTPEPECVPPLIRAARIDAALQARIRAQARELVAGLRATRTHS
SGVDALMKEFSLSSQEGVALMCLAEALLRVPDSATADRLIRDKLAHRDWHAHLGHSPSLF
VNAATWGLLVTGRLVGTSSAQGLSSALTRMLARGGEPLIRRGMDLAMRLLGEQFVTGRNI
AEALERSAAARRRGYRYSFDMLGEAAMTAADAQRYLAGYEQAIHAIGKASAGRGVVDGDG
ISVKLSALHPRYVWSQRGRVLAELLPRLKSLCMLARHYDLGLNIDAEEADRLELSLDLLE
ALALDDELAGWQGLGFVVQAYQKRAPRVVDWIIELARRSGRRLMVRLVKGAYWDAEIKRA
QVEGLAGYPVFTRKLYTDVAYLACARQLLAARDVLYPQFATHNAHTLSAVFQLAGDDYAA
GDYEFQCLHGMGEPLYDQVVGDTGSRRRVRIYAPVGSHETLLAYLVRRLLENGANSSFVN
RIVDENVSIDALVADPVEEALPLAGAPHPRIPLPADLYGSERRNSAGHDLASEPVRQRIG
AALQLSRSVVRRAGPLLANGPASPASRDNVLPVRNPADHSDVVGMVVHADEIDVERALAA
ASGAAPGWAMRAAAARAACLERAAELIERDTDELVALAVREAGKSWANVLAEVREAVDFC
RYYAARVRDFDNASHAALGPVLCISPWNFPLAIFTGQIAAALAAGNPVLAKPAEQTPLIA
TAAVALFREAGVPAAVLQLLPGRGDIVGAALVADPRVRGVLFTGSVEVATRINRALARRG
GDVPLIAETGGQNAMIVDSTALAEQVVADAIASAFDSAGQRCSALRVLCLQDDVADAVLA
MLRGALDEMRLGDSSDVRNDIGPVIDADAQAALERHVAAMQASGARITRLALPQRCSRGT
FVAPTIIEVGDIRDVGDEHFGPILHVLRYRAEELDRLLDAINATGYGLTMGVHSRIDETI
DAVIAHARVGNLYVNRNMIGAVVGVQPFGGEGLSGTGPKAGGPLYVHRLLRRSPGPALEG
RDRLPDGEAFGCLVTWLDGAGRTLIDDAGRAFLHERIAACRATRLAGLRLALPGPTGEDN
SLRFVPRGLVAGVATTTVGRLHQLLAALASGNRIVFTDDAAHRALFDALPGPARKAVTLD
TDWPGHPFGALLLDGTAAEADAWRVRLAERDGPIVPLLQPEPDYDAARLVHERTVSINTA
AAGGNASLMALGA
>Ehrlichia Q3YS87 Q3YS87_EHRCJ
MGDIGMISALQAPNEIRKRMQMLYRTEENSYARYLTEKTEVSQDSKIRIYSVAKQIIEKI
RVDRNLGIIDAFMQEYGLSNEEGIALMCLAESLLRIPDDCTINDLIKDKIGNSMWSNHIG
SSSSMFVNAATWGLFIGGIVLKESNDSAKWFGTINNLLKTMGEPIIRKAIQQAMCTLGKH
FIKGRDITEALNNRKEGELYSFDILGEAAKTRKDAERYFSEYMQAVDSIGKSKGSDDTGR
LVDYDEISVKISALHSRYEFSQIDDVLDEIVDKLLQICRLAKEYNIRVCIDAEEEASRLEI
SLMILEKLRFESSLNGWEGLGLAVQAYQKRAFSVLDFVEDISVRSCHKMMVRLVKGAYWD
YEIKNSQELGLSSYPVFTRKVYTDVSYLACANKILSKPNTFYPCFATHNAYTLAAILEMA
NKDHPGFEFQRLHGMGASLYEYVTQELAANIKCRVYAPVGGYQDLLPYLIRRLLENGANS
SFINQLNDSNISLEQLIQDPLEKAKELEYLPHPNIPLPKDIFGPERLNSSGIDITDSVTL
ANFNDEMKNYQNCKFKASSIVNGEEFDGDFIEILSPSNSEDLVGEVLFASSTQALSALDI
AYSAFKDWSNVPVSTRASILEKAANLIEENKAKLIMLLIREGGKVISDAIAEIREAVDFL
RYYAVLGRQELEGSNRLPGPVGEDNYLYFRSRGVFVCISPWNFPLAIFIGPIAAALVTGN
TVIAKPAEQTSIIAYEAVKLLYDAGIPKGVLHLLLGDGKELGEVLLKNEKIGGVAFTGST
ETARIIINQSIAEKEGGIIPFIAETGGLNTMITDTSALVEQVTNDVITSAFKSAGQRCSAL
RVLFVQEEVADKQIEMICGAMEDLVIGDPMLLKTDIGPVIDKASQEMLIAHADRMSQEGK
LLCQVKLGEECQKGYFFAPCAYEIQNISQLQREVFGPILHIIRYKKGDLHKILSEINDTG
YGLTFAVQSRVQSNIDNIIDNINVGNVYVNRNQVGAVVGVQPFGGQGLSGTGPKAGGPYY
LHRFLTEKVVSINTTALGGNTSLMCLTDQ
>Agrobacterium Q44334 Q44334_9R
MADGASKADVNPQQTVNGIFQNFAPPVREQSPLRKAITAAYRRPEEECLAPLIDAATVTP
EQAAAIRTTATKLIEALRAKTKGTGVEGLVQEYSLSSHEGVALMCLAEALLRIPDTATRD
ALIRDKIARGDWKSHIGGGRSLFVNAATWGLVITGKLTSTVNDSGLSAALTKLIARAGEP
VIRRGVDMAMRMMGDQFVTGETIGEAIKRSKPLEEQGFQYSYDMLGEAATTAKDAERYYK
DYENAIHAIGKASAGRGIYGGPGISIKLSALHPRYARVQAERVMAELLPRVKSLMLLSKK
YDIGLNIDAEEADRLELSLDLLEELALDKDLAGWNGLGFVVQAYGRRCPFVLDYIIDLAQ
RAGRRIMVRLVKGAYWDAEIKRAQVDGLEDFPVFTRKVHTDVSYIACARKLLAARDVVFP
QFATHNAQSMATIYHLAGPDFKLGDYEFQCLHGMGEPLYSEVVGKKKLDRPCRFYAPVGT
HETLLAYLVRRLLENGANSSFVNRIADPAVPVASLLEDPVTVVKAYPVPGARHDRIAAPA
GLFGPERANSAGLDLSNETALAALDNVLKAGATTEWKAAAPHAGGKTRPVLNPGDHNDVV
GYVTEPTEADVEAAMQRAAASNWSSTPVEERAACLERAADRMQAEMPALLGLIMREAGKS
MPNAIAEVREAIDFLRYYAAEARKTFKANETPLGPVVCISPWNFPLAIFIGQVTAALVAG
NPVLAKPAEETPLIAAQGVRLLHEAGVPQDAVQLLPGDGKTGAALVGSALTAGVMFTGST
EVARLIQGQLAGRVLANGQPVPLIAETGGQNAMIVDSSALAEQVVADVIASAFDSAGQRC
SALRILCLQEDVADRTLTMLKGALHELRIGRTDSLSVDVGPVITAEAKGIIEKHVDSMRA
```

```
LGHRIEQISLAGETGKGTFVPPTIIEMKSLADLKKEVFGPVLHVIRFKRDNLDRLIDEIN
ATGYGLTFGLHTRLDDTIQHVLSRVAAGNLYVNRNIIGAVVGVQPFGGRGLSGTGPKAGG
PLYLGRVTQTAPKIDRVASQQDQAAVDLARWLDENGQSVAAEAARQAAALSGLGFETELA
GPVGERNVYALHPRGKVLLIPATEQGLYRQLAAALATGNSVVIDNASGLEKSIYGLPATV
TSRITWADSWEKSAPFAGALIEGDAERVVAINKKIPALPGPLVLVQAATTEALDRETQPY
NLDWLVEEVSVSVNTTAAGGNASLMSIG
>Nitrosospira Q2Y8S2 Q2Y8S2_NIT
MGKTAACPRFLRSAITHAYRHNETESVNGLLRQIDWPPCSRARAEELARKLVTSMRGKYS
HRGGVDALMHEFSLSSQEGVALMCLAEALLRIPDRETADRLIRDKISKGDWRVHVGHSPS
LFVNAAAWGLLISRKLVSTHREDGLSAALSGLMEKGGEPLIRKGMDLAMRMLGHQFVIGE
TIEEALERSREREMRGYRYSYDMLGEAAITEADAQRYWTSYASAIHAIGKSRGKGGIKSG
IYRGPGISVKLSALHPRYVRSQRERVMSELLPRLKSLLLLAKHYDMGFNMDAEEADRLDL
SLDLLEALVFDPDLAGWDGIGFVVQAYQKRCPFVVDWLIDLARRSGHRLMVRLVKGAYWD
GEIKRAQVEGLDGYPVYTRKIHTDICYMVCAQKLLRAAEAVFPQFATHNAHTLSAVYQIA
QDADIEDYEFQYLHGMGETLYDQVELLKEDRFVKPCRIYAPVGSYETLLAYLARRLLENG
ANSSFVNRILDDDIVVEELVADPVSQIEQEGVHPHPAIPLPGDLYGHKRRNSAGLDFSDE
QALAALSAELTEMERREWRARPVLAAGGNAGNAGSRVLNPADRNDSAGTVVEATEHDVED
AIAVAHAFVAEWQSVSHEARAAMLERAADALEAHRTELMGLAIREAGKSLPDAVAEVREA
VDFCRYYGQQIRDWPNMSPSSALGPVVCISPWNFPLAIFIGQVSAALAAGNPVLAKPAEQ
TPLIAAVAVRLLHAAGIPRAALQLLPGQGERVGEQLVKDVRVRGVIFTGSTEVAQLICRT
LALRAQKSEILFIAETGGQNAMIVDSSARIEQVVQDALTSAFDSAGQRCSALRVLCLQED
IAEQVLEVLRGAMGELTVGDPRELATDVGPVIDADAQHALLDHIEKMRAAGHTVFQAALP
SVCANGTFVAPTLIEIDHIDELEREVFGPVLHVVRFARERVDELVGQINATGYGLTHGIH
SRVDETIDFITKRVHAGNIYVNRNMIGAVVGVQPFGGEGKSGTGPKAGGPLYLHRLMPQT
PLLLADNGGLRVNSPPVELQVLAAWARQAGRDVLASLCEDYAMRTPFAFSIPLPGPTGES
NTLKFAPRRAVACIAVDEDALLEQMAAALATGNQIILADNPPLRALLDKLPSQVRNRLRI
EREWIHAPVSAVLYSGPEDEAYRLRNELAGREGALVAFITASGTDFPLYRLTAERVVSVN
TTAAGGNPGLMGLDFNGSSTH
>Acidiphilium Q2D8X7 Q2D8X7_ACI
MAVSTMGVKLDDQVRARLRALAEREGRTTHHLAKQLILAGLERLERGEALDAAVTAEDDA
PAAVPFLEFVQDVQPQTVLRAAITAAYRIPETDAVAGLIEGATLPAAAAKQARATATNLV
TALRARRNAGLVETLLQEYALSSHEGIALMCLAEALLRIPDAATRDALIRDKIGAGDWQS
HLGGNKSVFINAATWGLLLTGRLNATASEAGLSSALTRLLARGGEPVVRRGVNIAMRLMG
EQFVMGRTIGEALANARKREAKGFAYSYDMLGEAALTAADAARYYRDYEQAIHAIGAAAG
GKGIYRGPGISVKLSALHPRYCRAQRARVMEELAPRLKALALLARRYDIGINIDAEEADR
LDLSLDLLEGLCFEPGLAGWNGIGFVVQAYQKRARSVIEFLIDLARRSRHRLMVRLVKGA
YWDSEIKRAQLDGLEGFPVFTRKIHTDLSYLACARAMLAARTEIFPQFATHNALTLASIH
ALAGPAFEVGDYEFQCLHGMGEPLYEEVVGPQKLDRPCRIYAPVGTHETLLAYLVRRLLE
NGANTSFVNRLADRRVPVEALIEDPVAAARKIEPLGAPHPKIALPGAILGTRRNSAGFDL
TSEQRLASLGSALLASLEQDWRAAPPDADPAAPGRTVVNPANIADRVGTVIEPDAGAVGA
ALARAAAMAPIWQATPAAERAAILLRAADLLEHRMTALVGLIVREAGKSLPAAIGEVREA
VDFMRYYAGQLQDGFDNETHRPLGVVAAISPWNFPLAIFTGQVVAALAAGNVVLAKPAEE
TPLIAAQGVRIMHEAGVPADALQLLPGDGRTGAALVGDARVGGVVFTGSTAVARLIQATL
ARRLDADGMPVPLIAETGGQNAMLVDSSALLEQAVGDIVTSAFDSAGQRCSALRILLVQD
DIADPLLAMLRGAMEELAVGATDRLSSDVGPVITREARDNIRAHVETMEGRGHKVFSVAL
PPETEGGWFVAPTMIEIDHPSALEREVFGPVLHVVRYRREERNAMVAAVNALGYALTFGV
HSRIDETIAQAVGKAEAGNIYVNRNIVGAVVGVQPFGGHGLSGTGPKAGGPFYLRRLLAR
RPMPHGEGEPAPAFTKLVGWMREQGIETAPALRYAAASAFRDAELPGPVGEQNTLRLKPK
GLVLCRAASPEGMWRQVAACLATGNAPLVLCDGAEAEIARMPAEIRPAIAPADAVARAEF
SAVLFEGHRDALIALQMALAERDGPIVPVHALPAEDHSEWDYPLEFLLEEQSISVNTAAA
GGNASLMSIG
>Nitrosococcus Q3JAG9 Q3JAG9_NI
MVMISDIELPSLDSPRAAISAAYLADETVWVKSLLRETELAPEVNERVKRRASQWVQLVR
EKRRWHGGLDAFLHEYNLSSQEGVVLMCLAEALLRIPDDENADRLIHDKLLKGEWDRHLG
HSHSLFVNASTWGLMLTGRLMRLETPLIDDGKGVIKRLVKRGGEPLVRLALRQAMRIIGG
QFIMAPRIEQALTQCRENERYSFDMLGEVALTQARVEQYYSGYRHAIRVLGETDSGEHNR
EATGISVKLSALHPRYTFSQRRRVQAELVPRVLALAREACAADIGLTLDAEEADRLELML
```

```
DVFEAVFRDPSLRRWQGFGLALQAYQKRALPVLNYLRDLAQKEGRCIPVRLVKGAYWDTE
IKRAQEQGLAGYPVFTRKVNTDVSFLALARRLLSARDVFYPQFASHNAHTVAWVLETAGE
QQGFEFQRLYGMGELLYSALREQGISVPCRVYAPVGGYADLLPYLVRRLLENGANTSFVN
RIEDEEVPIEQIVADPSEYVRSLPSKSHPNIPLPLRLYGKVRRNSLGINLNDPTSLERLM
AELGRAMEKQRQALPLVSGEAGKGTVQVVRDPSDRRRVLGTVVEADQEAIAEALSEADAV
AAGWEATSVLGRAECLEQAADLFEERQVELMALCIREGGKTVADSLAEVREAVDACRYYA
AEARRLFAMPKMLPGPTGEHNELTLHGRGVFVCISPWNFPLAIFTGQVAAALVAGNTVIA
KPAGQTPLIAALVVQWFHEAGIPPRVLHFLPGRGSRVGQALVADYRISGVAFTGSTRTAA
VINQVLAERKGPIVPLIAETGGQNAMIVDSSALPEQVVVDVMTSAFNSAGQRCSALRVLF
LQEEVAEPILEMLIGAMEELRLGEPGRLDTDIGPLIDGEARARLETHCQRMNREARLLCR
LSLPEATQNGYYFAPRVYELENLAELTYEVFGPVLHVIRYSSKHLNKVIASINRTGYGLT
LGIHSRVDETVRYIQNRVHAGNIYVNRNMIGAVVGVQPFGGERLSGTGPKAGGPHYLLRF
ATERSTSINMAAVGGNTDLLSLGE
>Anaplasma Q5PAT8 Q5PAT8_ANAMM
MMISPLQSPDELRKRMQGLYNADEKSYVRYLTERTEVSQESKVRIYSLAKQIIEKVRANK
NTTIIDAFMQQYGLSTEEGLALMCLAESLLRIPDDCTIDDMIRDKIARTTWNKHIGRSTS
IFVNVSTLALSIGAHMLREVDDSRWYGILGNLLKDMGEPVIRKAALQAMQVLGKHYVCGR
TIEEAIARSYDAGHVCSFDVLGEAAKTRADADKYFAAYMGALETLGSNAQAVDDLGSRHG
ISVKLSSLHPRYEFGQADYVLKDISSKLLELCQVAKKYNVKLTVDAEEARRLELSLMILD
TVFSDSSLNGWEGLGFVIQAYQKRALAAIDFVEDIAIRANRKMIIRLVKGAYWDYEIRNA
QEMGLDGYPVFTRKVYTDVSYFACVQKLLSKPGTFYPVFGTHNAHSVSFILEMTDKDHPG
FEFQRLHGMAQDLYDYVTKEVAPNVCCRVYTPIGQHQELLPYLMRRLIENGANVSFVNMI
NDSDIPAECLCADPLEKAQSFEYAPHPYIPLPADMFPDGRVNSAGVNTSDSLTMLALSEE
VSAFNDTHWKAYPIIGGQDITEGELYEVFFPASLSTKVGEVLFATAEHAAQSIEAARSSF
YRWSNTPVGERAAILEHAADLLEKERGKFFSLLVREGGKVISDVIAEIREAVDFLRYYAM
LARDQLTDPIKLPGPAGEENYLYFESRGTFVCISPWNFPLAIFLGPIAAALVTGNTVIAK
PAEQTSLVAYEAVKLLYEAGVPTDVLHFVPGRGEVLGNALLSSANIAGVAFTGSTETANI
INQAIAGRGRDIIPLIAETGGINAMVVDSSALPEQVVEDVITSAFKSAGQRCSSLRVFL
QEEIADKTIEMLLGAVAELRLGNPMALSTDIGPVIDQQSFDMLTSYVEEMRKKKIKLLCK
ADISYLSGEEEGYFFPPHIFELESMSQLSREVFGPVLHVIRYKKSDLPSILDDINSTGYG
LTFAIQSRIQSSIDDITDRIGAGNVYVNRNQVGAVVGVQPFGGRGLSGTGPKAGGPHYLH
RFLTEKTVTVNTAALGGSVSLTCLEDE
>Bradyrhizobium Q89E26 gi|27355
MPNIPPPFTAPYAPDDAEIAARLLPASHLSPPQEARIHRTATRLIEAIRKRDDRLGGVEDMLREFALSTK
EGLALMVLAEALLRVPDARTADQFIEDKLGEGDFIHHETKSTAFLVNASAWALGLSARVIQPGETPDGTI
GRLVKRLGAPAVRTATRQAMRLMGNHFVLGETIEQALERGKPRSGQKTRYSFDMLGEGARTAADARRYFD
AYASAIETIGKAAGNHALPDRPGISVKLSALHPRFEAISRARVMVELVPQLLDLAQRAKAHDLNFTVDAE
EADRLELSLDVIAATLADPSLKGWDGFGLAIQAYQKRASAVIDYVDALARAHDRKLMVRLVKGAYWDTEI
KRAQERGLDGYPVFTRKAMTDLNYVACASKLLALRPRIFPQFATHNALTVATVLEMAEGSSGFEFQRLHG
MGEALYEQLAKDHADIAYRTYAPVGSHRDLLAYLVRRLLENGANSSFVAQAADYRVPVPALLQRPADAIV
RPQAAAHPRIPLPCDLFAPERRNSRGVEFGARTALDQLLTDVKAETGDLKPIADATPDQAHAAVAAARAG
FAGWSRTPAGIRAAALEQAAHLLESRSAHFIALLQREGGKTLDDALSELREAADFCRYYAAQGRKLFGSE
TAMPGPTGESNALTMRGRGVFVAISPWNFPLAIFLGQVTAALMAGNSVVAKPAEQTPRIAREAVALLHEA
GIPKSALYLVTGDGRIGAALTAHPDIAGVVFTGSTEVARSINRALAAKDGPIVPLIAETGGINAMIADAT
ALPEQVADDVVTSAFRSAGQRCSALRLLFVQEDVADRMIEMVAGAARELKIGDPSDVATHVGPVIDVEAK
QRLDAHIARMKTEARLHFAGPAPEGCFVAPHIFELTEAGQLTEEVFGPILHVVRYRPENLERVLRAIERT
GYGLTLGVHSRIDDSIEAIIDRVQVGNIYVNRNMIGAVVGVQPFGGNGLSGTGPKAGGPHYLARFATEQT
VTINTAAAGGNAALLAGEE
>Bacteroides Q8A340 Q8A340_BACT
MVQRPQDKKFLVKMLDESSQIRDRRILAKRIKTLLDQYGVPEFLNKRDSFLFKMYQAFGH
HFDFIAIPIIKKRLRMNTSQVIINEARPQLTKHLAIRAKEKIGQNVNLLGEVVLGNGEAD
HRYHHYLKALESPDINYISVKISGIYAQTHALNYEESFPELISRMSALYQKAIDFPYTDE
EGVRRSKFINLDMEEYKDTHFTLRLFKTVLSLPQFKNYSAGIVVQAYLPDAYDFQTELIE
FAKARVAEGGAPIKMRLVKGCNLEMETVISSLRGWPNPIRPSKEEVDANYLHLLERALMP
ENARVLHLGVASHNLFSIAYAYLLAQKYGTAEYMTFEMLEGMANHLWRAQSMLGNRVILY
TPVVKNEHFLNAVSYLVRRMDENTAPDNFLTHSFNLRPNTKEWDFLAKQFEDAYAMKDQL
SHVSPRTQNRNLPYTPVPPADVLKNEPDTDFDLPQNQEWVRSIFSKWKKDGTEQPEIIPL
```

```
QIGAETVVCESRYPYTDRCQDDEVCICEMSQADSAQVEKIIEIAEADPAGWRKTTLEERH
RIMYEAANRLADMRGDLIGCMCAVTGKTVIEGDVEVSEAVDYARFYTTAMKKFAALDDVE
MKPKGTILVISPWNFPCAIPVGGIVAGLAGGNTVILKPATVAAPVAWMFAKAFWDAGVPK
EALQVIITRREALKVLTTAPAIKHIILTGGTDTAQNIAKANPTTPLSAETGGKNVIILTA
SGDRDHAIMNIVTSAFGNAGQKCSACSLLLVERSVYEDENFRSKLKDAATSLKTGSVWNA
GNIVGPMITNKNDKLLQAFNLEPGESWLVPPRFIDRREYILAPTVKWGVKPESFSFRTEL
FGPLLSVACIENLEEGIRLVNGLDYGLTSGLQSLDEKEQKLWKNSVMAGNLYINRGITGA
IVNRQPFGGMKLSAFGGGIKAGGPNYCTCFLEITDKPDSRTDYRQSYAKAYQEEFSKPRD
VNRLYGEQNLFRYLPLKNMILRLFPKDTEEATMIAHAARICRTPLTISFGPTDDRSSRL
AGLGCTLRKESLEDFLKELPEYERVRTCSPDIPDVMYERAAETNKYIATAPPVKQGRIEL
IHYIKEQSIAFEYHRYGSISEVPPCE
>Corynebacterium Q8FUB9 Q8FUB9_
MGFLRCAGSPPPNPEILSAESSTDLWVNFRGFWEQELGVKAATVSGGASAPAPGTAGDRI
STCQYVGGHNNRDNWCGSHDGRVSARSVVGSGPRWPTNTARDVTNPGEQGASTMTATTDH
SMKLPIELATLTDRAVVKVRDWLEYAQQESVPNPTAERLAAILQDPNGLEFTVGFVDRVV
RTEDIHAAARALNDLGSIAPSTMSFIDRAQIQAGSLVGRALPHVVVPAARARIRQMVGHM
IVDARDKPFGRAVAELQGDGNRLNINLLGEAVLGEKEALKHLEDARRLLGRMDVDYVSIK
VSSVASQISLWGFDETVEAVVKRLTPLYKEAAAATGTPGGTKFINLDMEEYGDLRLTIEV
FKRLMSLPGVENLEAGIVLQAYLPDALASLQELAEFGAHRVDNGGAPIKIRLVKGANLPM
EHVHAEIAGWPVATMPSKQATDANYKRVLWWALRRENMTGLRLGVAGHNLFDIAFAHLLS
IERGVAEKVEFEMLQGMASDQARAASRDVGQLLLYVPAVRPQEFDAAISYLVRRLEENSA
SENFMSVIFDLEADNPAFRREESRFRASIEDLAGLIDAPVPGPNDTQDRGAQEAETLEAA
GERARELEAARAPDQLPPFANEPDTNPALTANQQWARAAIKRSAGEGWLEQQTAPLSVEE
EDIDALIEGTRAAAAEWAALPPMERARILYRTADILAARRGHLVSIAAAEVGKVVEQTDP
EISEAIDFARYYAHRALELEQVDNADFTPDRVAVITPPWNFPIAIPAGSTFAALAAGAGV
IHKPSKPSQHCSAAVVEALWEAGVPRRVLHCVYPANREAGRRLVSHPEVDRVILTGSSET
AAMFASWRPDLQINAETSGKNAMVITPAADRDLAVADLVHSAFGHAGQKCSAASLGILVG
TMYTSERFRSQLVDAASSLVVDWPTNMSATVGPLTEQPSDKLQHALTQLEPGERWLLEPK
PLDDTGRLWRPGIKEGVQPGSFFHLTEVFGPVLGLMKADDLEQAIEYQNAVEFGLTGGIQ
SLDIDEVRTWVDAVEVGNVYVNRGITGAIVERQSFGGWKKSSVGLGSKAGGPNYVMLMGR
WSDKPAKDAPVKSSPLINKWTHTLSPDDITWLEQANASDALAWTSEFGTPRDPSGLEAEA
NIFRYRPAKVVLRIADDAEPREVARAVLAARRAGAQLTAMVGPGVGGGVREVLADASTSV
ETIDDAVFITRLLRGDYDDGTGARVRVIGTVTPFQRERLAVRPEVAILDEPVTSSGRVEL
RYWLKEQAVSMTLHRFGNRSEAFHKLAADLKRPL
>Geobacter Q746X3 Q746X3_GEOSL
MLNSELNTKIVNRGKEFFGSISGEKPSLFNKGAWMGKAMDWSMQNEQFKIQMFRFVDVFP
SLTTSKLLTEHIREYFGNEQDMPAFMSTGAKVAGMLGSFGGAVLNKVLTSNIEEMARQFI
VGETTKEAVKNLEKLRKDGFAAVVDVLGEATLSEEEAEVYTNTYLELLEALKKEQGSWKG
LPGKGGDPGLDWGHAPKVNIAVKPTALFCLANPQDFEGSVVAILDRMRRIFKKVMELNGF
LCIDMESYRHKEIILEVFRRLKLEYRDYPHLGIVLQAYLKDNDKDLDDLLAWAKEHKVQI
SVRLVKGAYWDYETVKAKQNDWEVPVWTIKAESDAAYERQARKILENHQICHFACASHNI
RTISAVMEMARELNVPEDRYEFQVLYGMAEPVRKGILKVAGRIRLYAPYGNMVPGMGYLV
RRLLENTANESFLRQSFAEDAQIERLLEDPAVTVERERAARAAKPKKERKGLGGLPPFNN
EAMVDFTRADHRAAFPKHIAQVRTQLGKTYPLFINGKEVRTNDLIPTVNPNKPSEVLGQI
CQAGTTEVGDAIAAAKAAFPAWRDTDPRTRAEYLLKAAQAARKRLFELSAWQVLEIGKQW
DQAYADVTEAIDFLEYYAREMIRLGQPQRVGHAPGELNHYFYEPKGVAAVIAPWNFPLAI
SMGMASAAIVTGNCVVFKPSGITSIIGWHLVELFREAGLPEGVFNFTPGRGSVMGDYLVD
HPDISLIAFTGSMETGLRIIERAAKVHPGQANVKKIISEMGGKNAIIIDDDADLDEAVPH
VLYSAFGFQGQKCSACSRVIVLDAVYDKFIERLVSMAKATKVGPSEDPANYMGAVADDKA
MKSIKEYAEIGKREGHVLYESPVPAGEGYFVPMTIIGGIKPEHRIAQEEIFGPVLAVMRA
KDFDQAIEWANSTQFALTGGIFSRSPEHLAKARREFRVGNLYINRNNTGALVERQPFGGA
RMSGVGTKAGGPDYLLHFMDPRVVTENTMRRGFAPIEEDDDWVD
>Desulfovibrio Q725V6 Q725V6_DE
MDQQHLDGKVVERGKEFFRSISGEAPSIFNKGWWTGKVMDWAMQNEDFKVQLFRFVDVLP
YLNTSESLLRHIREYFATEDADIPPVLKWGAGKAGIGGALTAKLMGMTIRSNIEGMARQF
IIGDNSKEAVKGLAKLRKDGFTFTVDLLGEATVSEEESEAYAQGYHEVVDAIAREQEKWK
ALPGNGPVEGFDWGATPKVNVSIKPSALYSQAKPVDVEGSVRGILSRLVPIYRKVVAMGG
```

```
FLCIDMEQLKYKEMTLELFKRLRSDPEFRHYPHLSIVLQAYLRDTEKDLDDLLHWARSEK
LPIGIRLVKGAYWDYETVIAKQNGWEIPVWTDKPESDIAYEKLAHRILENSDIVYFACAS
HNVRTIAAVMETALALNVPEHRYEFQVLYGMAEPVRKGLKNVAGRVRLYCPYGELIPGMA
YLVRRLLENTANESFLRQSFAEGAALERLLENPQKTLHRLLAARPEPRAVEPGPGGLPPF
TNDAMIDFTVPDNRKAFVEALADVRSRFGQTVPLYIGGRDVTTADLIPTTNPAKPAEVVA
SICQAGRPEIDDAIAAAKKAALTWRDTSPADRAAYLRRAADICRKRIWELSAWQVVEVGK
QWDQAYHDVTEGIDFLEYYAREMLRLGAPRRMGRAPGEHNHLFYQPKGIAAVIAPWNFPF
AIAIGMASAAIVTGNPVIFKPSSISSRIGYNLAEVFREAGLPEGVFNYCPGRSSIMGDYL
VEHPDISLICFTGSMEVGLRIQEKAAKVQPGQRQCKRVIAEMGGKNATIIDDDADLDEAV
LQVLYSAFGFQGQKCSACSRVIVLDAIYDRFIERLVKAASSIHIGPSEDPSNYMGPVADA
TLQKNVSDYIRIAEEEGRVLLKRTDLPAEGCYVPLTIVGDIRPEHRIAQEEIFGPVLAVM
RAATFDEALSIANGTRFALTGAVFSRSPEHLDKARREFRVGNLYLNKGSTGALVERQPFG
GFAMSGVGSKTGGPDYLLQFMDPRVVTENTMRRGFTPIDEDDDWIV
>Syntrophobacter Q3MY03 Q3MY03_
MDSDLERRVQQTGLWLYELIEGESPSVFRKEFWTGKMLEWCMQNEAFKVEMFRFVDVFPY
LTRPESVARHVQEYFSRPGVNFPAVLQWGLRAVSPGSLTAKVIARSITHNLHNMARQFIV
GSNPSEALPNLERLRGQGMAFTIDLLGEAVVSEKEAEEYVSRYLELFDILDEAQRKWPAI
GGGAQQADWGHAPKVNVSIKASAMYSQMSARSFEDSVARSKEKLRPILRKALATGSFVNL
DMERHALKDLTLALFRSLMEEDEFRDYPHVGIVIQAYLKDSERDLEEILGWAKATGRHFT
IRLVKGAYWDSEVIWARQSEWPVPVFTSKPETDANFEKLADLIMENHQWVSLACASHNMR
SISYVMERARDLSVPAGRLEYQVLYGMGEPVRNALRKAGLPVRLYTPVGDMIQGMSYLVR
RLLENTANESFLRKSFFQGVSRELLLRNPMDVLAEERTAGPVPARDAPEYGDKGPFCNEP
CFDWTIPEHRAGFRDALDRVRATFPIKVPLTIGGSRFDTPVRLRSVNPNRAEEMVGDVAG
AGPLEADAAVEAAKAAFAAWRDTPPGERAEYLFKAAAAARRIRYDLAALQVYEVGKAWSE
ADADVCEAIDFLEYYGREMIRLSRPKRMGHAPGEISHLFYEPRGVAAVIAPWNFPMAIST
GMTSAALVTGNTVVYKPASQSPVVGSMVMNVFEEAGLPKGVLSFLPGPGAQIGDYLVHHP
DVAVIAFTGSKKVGLDIIAQANRDAERAGHVKTVVAEMGGKNAIVVDADADLDEALAQIV
HSAFGYQGQKCSACSRLIVLEEIYDKLVERLKAAAESIHLGPPEDPKNLMGAVIEAGARK
RIMEYIELGRKDGTVLVERTVPGNEGFFVPLTILADLPPDHRLAREEIFGPVLVVFKVKD
FARAIEIANDTEYALTGGVFSRSPANIDLARREFRTGNLYINRGCTGAVVERHPFGGFKL
SGIGSKAGGPDYLLQFMVPRNVVENTLRRGFAPADE
>Bdellovibrio Q6MNK1 Q6MNK1_BDE
MNDIQSQIVSRGEEILKRMESQSKASIFSKDFWYGSIMEWSMKNEKFKTNMFRFVDVLPS
INSGDEVARHLKEYFSEDGGTLPPVFNVGLGLGSLAPGLMAGAIKKNVMGMAKMFITGES
PDEALPVLKKARNKMTFTVDILGEATLSEKEAQDYSNKYMELVTWLAKDAEKWDEVPQI
DRDHEGALPKVNVSVKMTALYSQIKDAAWDESKKILKDRLRPVFRLGMEKGVFVNLDMEQ
YSVKHLTLEVFTELINEPEFKNYKFFGIVIQAYLRDSFEDVKSLTEFAQKRGTPFWVRLV
KGAYWDYETIEAEQRGWPVPVYTNKAESDANYELCAKYLLENIKFIRPAFASHNVRTLAA
CMLYAEKLNIPKEALEFQMLYGMAEPIKKTIVDMGYRMREYAPVGELIPGMAYLVRRLLE
NTSNESWLRGKFADNKSMAELLKDPAQGLTPTSPVIPKKPGKFYNEPLLDFAVKADREKM
LKALAEAKASLPVNVNIVINNKELQSGKIFDRVNPSQSDQIVGKIQMATTEQAEQAMQAA
QTAYKTWKNVPCEQRAALVDKLADIMTRDRFKLIATQVLEVGKPWAEADGDIGEAIDFCR
YYARHMRELQKPLRVGGLPGELSHYIYKSRGVTAVIAPWNFPLAILAGMVTAAAVAGNTV
VMKPAEQSTVVAWGLMKMIQEAGFPQGVINFLPGYGEEVGEYIVNHKYTTTIAFTGSKAV
GLHIMNRAAVVQPGQQHVKRCIIEMGGKNAVIIDNDADLDEAVDGVIYSAFGFSGQKCSA
ASRVIVLDEVYDRFVDRLVETAKSIEIHPAENPKAYMGPVVDKEAYDRILGTIAEAEKNH
KLLFKGSVPGGGFFAPPTIFGDVPGDAKLAQAEIFGPVVAVIRAKNLDQALDIANSTEYA
LTGGVFSRSPANINRVKEELEVGNLYVNRGITGAMVDRHPFGGFKMSGIGSKTGGPDYLK
QYMEPACVTENTLRRGFAPAEE
>Trichodesmium Q3HDN5 Q3HDN5_TR
MVEQISHTNYETKTQEIAKQLLKATQEKNRSFLAQLQNQMRWDDKLLDWAMASPGLRVQL
FRFIDCLPALRSKPEIAAHLQEYLTTQEVELPETLKKLLNFANPDSVPGQLAATTVAPAV
ETLAHKYIAGENVKQIIKTLEKLRKDKMCFTVDLLGEAVITETEAQLYLDRYLELMTQLS
QAANSWSSIPQIDEAEGEKLPRVQVSVKLTAFYSQFDPLDVKGSQERVSDHVRTLLRHAQ
KLGVAVHFDMEQYTYKDLTIDILKELLMEEEFRNRNDIGVTIQAYLRDSEKDLQGIIEWA
KIRGRPVTVRLVKGAYWDQETIKALQHDWPQPVFNDKPETDANFEKLTYMMLENHEYIYS
AIGSHNVRSQARAIAIAETLKVPRRCFEMQVLYGMGDKIATILAEKGYRVRVYCPYGDLL
```

```
PGMAYLIRRLLENTANSSFLKQSLENRPLEELLAAPSINGKTTIHDVVKPVFPNAADSDY
ANFQQRQQALNTIGQIRLELGKTYLPIINGEYTNTAQIVDSLNPSNSKEVVGKIGLISVE
QAEQAIQAAKAAFPSWKKTPVRERTRILRKAADLMEKRRHELAAWMVLEVGKPLNQGNAE
VSEAIDFCRYYADEMERLEQGYNYDVAGETDRYYYQPRGISLVISPWNFPLAIPTGMTVS
SLVTGNCTLLKPAAVSSVIASKISEILLEAGFPKGVFQFVPGNGSTVGDFMVKHPSVNSI
TFTGSMEVGCHIYAQAAVLQPRQKHLKRVIAEMGGKNAIIVDESADLDQAVAGVVYSAFG
YSGQKCSACSRVVVLATIYDSFVNRLVEATRSLNIGDAEKPSTQVGPVIDEKAQKRIQEY
IVKGKQEAEVAIEMSAPKDGYFVGPVIFKDVSPSATIAQEEIFGPVLAVMKADNFTQALE
IANGTNFALTGGLYSRTPSHIEQAKAEFEVGNLYINRGITGAIVSRQPFGGFKLSGVGSK
AGGPDYLLQFLEPRTITENIQRQGFAPIEGVDN
>Anabaena Q8YZD7 Q8YZD7_ANASP 1
MVLQVQTSTYEAKTQEIARQLLGATQENRSFFASLRDQMRWDDKLLAWAMSNPGLRVQLF
RFIDTLPALHSKAEIAAHLQEYLGDESVELPAALKGMLNFANPDSVPGQVAATTVSTAVE
TLAHKYISGENIKQVIKTVERLRKEKMAFTIDLLGEAVITETEAQSYLERYLELIAQLTE
ASKNWGTIAAIDEADGEQLAKVQVSVKLTAFYSQFDPLDAKGSEEKVSDRIRILLRRAKE
LGAAIHFDMEQYAYKDLTLNILQKILLEDEFRQRTDIGITIQAYLRDSEQDARNAIAWLK
QRGYPLTIRLVKGAYWDQETIKAAQKHWPQPVFNDKAATDANFEAITQLLLENHQYVYSA
IGSHNVRSQALAMAIAETLNVPRRRFEMQVLYGMGDKLAKALVDRGYRVRVYCPYGDLLP
GMAYLIRRLLENTANSSFLRQNLENRPVEELLAAPKVDLAQAKVHSPEAFPQGKHFVGAA
DTDYAEEEDRTKAARAFAVVRGELGKTYLPLINGEYVQTAEVIDSVNPSNFSEVIGKVGL
ISVEQAEQAMQAAKAAFPGWRRTSVKERAGILRRAGDLMEQRRAELSAWIVLEVGKPVKE
ADGEVSEAIDFCRYYADEMERLHQGINYDVDGETNRYIYQPRGIVVVISPWNFPLAIACG
MTVAALVTGNCTLLKPAETSSVITAKLTEILVEAGIPKGVFQYVPGKGSQVGAYLVSHPD
THLIAFTGSQEVGCRIYAEAATLKPQQRHMKRVIAEMGGKNAIIVDESADLDQAVVGVVQ
SAFGYSGQKCSACSRVVVVEAIYDAFIHRVVEATKSLNIGEAELPSTQVGPVIDANARDR
IREYIEKGKAESQVALELSAPNHGYFVGPVIFGEVPPHGTIAQQEIFGPVLAVIKAKDFA
QALAIANDTDYALTGGLYSRTPSHIQQAQEEFEVGNLYINRNITGAIVARQPFGGFKLSG
VGSKAGGPDYLLQFLEPRTITENIQRQGFAPIEGAE
>Crocosphaera Q4BWE2 Q4BWE2_CRO
MVVQLDSTTYEQRTQEIAKELIAQTREKRSLWSKLGDQMRLDDKLLDFAMANPGLRVQLF
HFIDTLPALQSNAEIAHHLQQYLGDESVELPSSLKGILNFTDYNSLPAKVAAETISKAVQ
TLAFKYISGETVPQVIKTVERLRKEKMGFTIDLLGEAVITESEAKAYLDSYLDLMEKLAT
ESKKWSNVAQIDTAGDENLSKVQVSVKLTAFYSQFDPIDPQGSKEKVCTLIRILLRRAQE
LGVAIHFDMEQYVYKDLTLSILRELLLEEEFRNRSDIGVTLQAYLRDSKQDLHDLISWAK
QRGTPITIRLVKGAYWDQETIKSEQNHWPQPVYNQKSATDVNYEEMTQLLLENHQYLYAA
IGSHNVRSQARAIAIAETLKVPSRAFEMQVLYGMGDQLAKALVKTGHRVRVYAPYGNLLP
GMAYLIRRLLENTANSSFLRQNLEERPIEDLIASPTLSGNTERIAQKESFVNVPDTDYSR
EVLRDKAEKALAKVKDSLGKTYLPLINGEYVQTDVIVESVNPSKSSEVVGQIGLISVSQA
EQALNAAREAFKDWKKTPATERARIIRKAGDLMEERRHELSAWICVEVGKVLQQADAEVS
EAIDFCRYYADEMERLDKGYNYDVAGETNRYHYQPRGIALVISPWNFPFAISTGMTVAAL
VTGNCTLLKPAETSTVIAAKIAEILVDAGIPKGVFQLVPGKGSKVGAYMVNHPDVHLIAF
TGSREVGCRIYADAAILQPGQKHLKRVIAEMGGKNAIIVDESSDLDQAVAGAVFSAFGYS
GQKCSAASRIIVLDPVYDAFLERFVEATKSLNVGATDQPSTQVGPVIDATAQKRILEYIE
IAKQESTLALQMEAPDNGFYVGPTIFGDVLPNHTIAQEEIFGPVVAVMRVKSFAEALGVA
NGTDYALTGGLYSRSPEHIEQAQKEFEVGNLYINRTITGAIVSRQPFGGFKLSGVGSKAG
GPDYLLQFLEPRHISENIQRQGFAPIEGNE
>Synechococcus Q2JNZ3 Q2JNZ3_SY
MVAQGAPTAYEAQTQALAQSLLASLWGEKRSLLAQWRDQLRWEDKLLAWAMENPHLRTQL
FRLIDVLPSLKSKAEVARHLQEYLSDPAVELPGMLKSLLNFADPDSLPGQLAATTLTAAV
EALARRYIAGETLQQVSKTLEALHRQGMAFTLDLLGEAVITEAEAQSYLDRYLQVMEHLA
KMVKTWPQPQPVPQVSVKLTAFYSQFDPIDPLGAKAKVGERIRILLRRAEALGVGIHFDM
EQYRYKSLTLAILQELLLLEPEFKSRTDVGLTLQAYLRDSYTDLQNLIAWAKERGSPVTVR
LVKGAYWDQETILAAQKHWPQPVYNSKSATDANFERMTQLLLEHHTHLHAAIASHNVRSQ
AKAIAIAQALEVPQSAFELQVLYGMADPLAKALVQKGQRVRVYCPYGELLPGMAYLIRRL
LENTANTSFLRQSVQGYTPEQLLAPPVLEEGSVFPAPSLSFAAAPDTDFSQPQQRDPFFQ
ALDEIRPQLGQRVQPWIAGEYRPTQQAMLSLNPADPQQVVAEIGLTEPEQVEQALAAAQA
SFPSWRRTPVAVRAGILRRVAEQIEAQRAELAAWMVLEVGKPIAQADGEISEAIDFCRYY
```

```
AQQMEALDRGRECNLPGETNHYHYQPRGVAVVIAPWNFPLSIPCGMTAAALVAGNCVLLK
PAEQSSLIAAKLAEFILKALAEAGIPEGVFQFLPGVGEAIGPLLVQDRRVHLIAFTGSQA
VGCQILAQAAQLQPGQHHIKRVIAEMGGKNAILIDESADLDQAVKGVVNSAFGYSGQKCS
ACSRAIVLQSIHDLFLGRLIEAVRSLNVGDPAHPSTQVGPVIDAEAQARLQQAIQSAKAD
ARLDLSLPVPEKGYFVGPTVFSQVDPQSELAQRELFGPVLAVISAPDFSSALQMANATPY
ALTGGLYSRTPSHIQQARRDFAVGNLYINREITGALVERQPFGGFKLSGVGSKAGGPDYL
LQFLEPRTVTENTQRQGFAPIEGS
>Gloeobacter Q7NGY2 Q7NGY2_GLOV
MELPPFQETRVLTVDPIESETQRIGQSLLAAPQQSRFSFFERSFWDEKLLEWAMADDELR
VQLFRFIDCLPALKTNREVASHLQEYLGSVKLPGPLAKALDFADPGSPAAAVATLGLRQG
ISQMARRFICGENLKTAVKTIEKLRSQNMAVSLDLLGEAVVSEAEAEVYQRRYLELLSDL
HAAQSRWPVIDQIDRADGQLLPRVHFALKLTSLYSQFDAIDPATTGRNVKARLRPILRKA
RELGAFVHIDMEQSQHKDLILQIFKEVMVEADFRDWIDTGICLQAYLRSAELDAANVIAW
SKERGYPVTVRLVKGAYWDGEVIRAAQQRWPVPVFTRKADTDAQYEKVLHLLMENHRSVH
IAVGSHNARTVALAIALARRNRVPRRAFEVQMLYGMADHLKAGVAQQGERLRVYAPYGEL
LPGMAYLIRRLLENTANTSFLRQSVRPGSDVVRLLAPPAASGEAAPSLMGSEFVNAPDRD
FAVAAEREKLAAGLAEVRTCFGQTYWPVVGGKARRGLPTLSVPNPAEPGTVLGALGLADA
ELADEAVKVAHTAWLSWSKTTPEARARLLGRLAELMEQQRSALTAWIVYETGKPWRESDA
DVSEAIDFCRYYGRQILKLTAQAQHRDVPGETNGYHYRSRGVAAVIAPWNFPLAILTGMT
TAALATGNTVIMKPAEQSSIVAAQLMRLCEQAGFAAGVVNYLPGRGEVVGAKLVAHPQVH
LIAFTGSLAVGQRILAEASMVRPGQRHLKRVIAELGGKNAIIVDSDADLDQAVVGVVQSA
FGYSGQKCSACSRLIVLEGIYDVFIERLVQATRSLIVGNPAQPATYTGPVIARDACERIR
GVIEDAKGRHRLVLEVDVSHLGEGYFIGPTIFADVDPQSSLAQEEIFGPVLAVIKARDFE
HAIEIANGTSFALTGGLFSRSPRHIAEARERFEVGNLYINRKITAALVDRQPFGGFRLSG
VGSKAGGPDYLLQFVEPVTVTENIQRQGFAPLPGEL
>Helicobacter Q9ZN12 Q9ZN12_HEL
MQKIIEDSLELAKKLQDSISNHLSEQEKAFHFKMQKLLNNPENKVMLIELMDRSFRCLDN
KARFEMIEHVLDKYKSREIFSSFEKWLLMGFLSFGKMLPDMSVPFFVNKIRSDTKAMVLD
QEESQLRERILKRKNEKIILNVNFIGEEVLGEEEASARFEKYSQALKSNYIQYISIKITT
IFSQINILDFEYSKKEIVKRLDALYALALEEEKKQGMPKFINLDMEEFRDLELTVESFME
SIAKFDLNAGIVLQAYIPDSYEYLKKLHAFSKERVLKGLKPIKIRFVKGANMESEETIAS
MKDWALPTFSSKQDTDSNYNKMLDFVLEGDNYKYIHIGAASHNIFEIAYVYTRIHALNDP
VVLEHFSFEMLEGMSLQASQELKEMHKLILYAPVCDEAHFNNAIAYLVRRLDENTSSDNF
MKAFFNLKVGTSEWKDQEQRFLNSLKGIATLDNTTHRTQDRNAKQTGHTTYPNHSFKNES
DTDFILKANREWAKKVRDKMHNAPILELYPEIDGRFEDPNLTPLEVFDRIHHKKIASVHL
ADKEAILKALEVAKSDKSRFSQKSFTEIHALLSQTAQLFRERRGDLIGISALEVGKTFAE
TDAEVSEAIDFLEFYPYSLRVLQEQNKKTKFTPKGVGVVIAPWNFPVGISVGTIAAPLAA
GNRVIYKPSSLSSVTGYKLCECFWDAGVPRDALIYLPSKGSDISEHLLKDESIKFAILTG
GEDTAYKMLKANPTLALSAETGGKNATIVSKMADRDQAIKNVIHSAFSNSGQKCSATSLL
VLEKEVYEDENFKKTLIDATLSLSVGDPFDFKNKIGALADKPNEKVIKAIDELKSYENYE
IPASFVDDNPYLMKPSIKYGTKKGDFTHQTELFTPILSVMKAKDLDEAIEIVNSTGYGLT
SALESLDEREWEYYLERIEAGNIYINKPTTGAIVLRQPFGGVKKSAVGFGRKVGIFNYIT
QFVNIHQEEEDENALKNPLSEALEGLTQKGYDEHTHELKRAIFMAKSYAYHYKHEFSQAK
DYVKIRGEDNLFSYTKVTSVGYRITEKDTLSDMLGVALACLISQIPLTISIENERANKDL
TFFLECLKALQANAPIVYESLQKFSEKLNAFNRVRYLKSDLDLLHEQASRLGMVLATAKP
CLNGRFELLYYHLERSVSISYHRYGNLGSRVLRQPTCHKSCCAEK
>Campylobacter Q4HED6 Q4HED6_CA
MIQKSLALAEELQGKIEANISNSEKEFHAKMQKLLNNPENKVMLIELLDRSFRCKDKRAS
FELIEHTLNKYGIADFFSAFEKFLLFSFLNFGKFAPNLSVPFFVKHLREDTKAMVLDANP
SVLEPHMRKRKDQDNITLNVNLIGEEVLGEAESKYRIQKYEEALKSSYITYISIKITTIF
SQINIIDFEYSKDEVVKRLDHLYALALEEEKKQGVSKFINLDMEEFRDLELTVAAFMESV
SKFDIKAGIVLQAYIPDSYEYLKKLFAFSKERVLKGMKPIKIRFVKGANMESEETIASQR
GWELPTFYKKIDTDSNYNKMLDFVLEGDNYKYINVGIASHNIFEIAYAYTRISEAGALDS
FTFEMLEGMSLQCSYELSKMHDLILYAPVCDEAHFNNAIAYLVRRLDENTSEDNFMRYFF
NLKVGDNNWNVQKELFLKSLEGIKTLDNSTHRTQDRNKAPNIVSAYESGMFNNESDTDFI
LPQNRAWAKVIRAKYENLENYDVMPVIGDLEFDKNTLNILEVKDKIKDRMIGKAYLAGEN
EIKKALKVAKNSNFKNTSFDEIHAILAKAANLMRERRGDLIGLAALEVGKTFLEIDPEVS
```

```
EAIDFTEFYPHSLKVLREQNKNTQFSPKGIGVTIAPWNFPIGISVGTIAAPLAAGNVVIY
KPSSLSTLTGYMLCKCFWDAGISKDALIFLPSKGSDISKYLLVDEAVKFSILTGGEDTAY
AMLKANPTLLLSAETGGKNATIVSKFADRDSAIKNIIHSAFSNSGQKCSATSLLVLEEEV
YEDEEFKKTLVDAAASMAVGNPFEFKNKLGTLADKPSAKVEKALNELAPYEEWALKPQFL
ENNPYLMTPGIKYGTKKGDFTHMNELFVPILSVMKAKDLKEAIEIVNSTGYGLTAGFESL
DEREWEYFHTHIEAGNIYINKPTTGAIVLRQPFGGVKKSAIGFGRKVGIYNYITQFLDIS
TSEADENLLNNEFVKTLEKIDLGTNEENKKALNNAILMAKSYAYHYRNEFSVSKDYVNIR
GEDNLFSYTKIKNLAFRVCENDNLQDILGVIIAANILNIPLSVSYEENDKISLVKELCKA
INLNLEITNESKAQFVSRLKNFERIRYHGKVSVEDEIYQEAAKEAKIIIREKPLLNGRFE
LLFYHNEKALSVSFHRYGNLGIRALNH
>Thermus Q72IB8 Q72IB8_THET2 Pr
MNLDLAYRSFVLGVAGHPQVERLIKHRAKGLVRRYVAGETLEEALKAAEALEREGVHAIL
DLLGEMVRTEEEARAFQRGLLELVWALAGKPWPKYISLKLTQLGLDLSEDLALALLREVL
REAEPRGVFVRLDMEDSPRVEATLRLYRALREEGFSQVGIVLQSYLYRTEKDLLDLLPYR
PNLRLVKGAYREPKEVAFPDKRLIDAEYLHLGKLALKEGLYVAFATHDPRIIAELKRYTE
AMGIPRSRFEFQFLYGVRPEEQRRLAREGYTVRAYVPYGRDWYPYLTRRIAERPENLLLV
LRSLVSG
>Geobacillus Q5KVJ1 Q5KVJ1_GEOK
MEQLMRDFFLFLSKNKTLTKWAKRYGLRFGASRFVAGETIEEAVRVIRQLNEKGLAVTVD
YLGEFVDNEQEANEMARHCLEAIEAISREKLNSQLSLKMTSMGLDISDELVMRNMRRILD
AAKQRGVFVTIDMEDYSRCQKTLDIFKTLKKEYDNVGTVLQAYLYRTEQDIEDLKPYRPN
LRLVKGAYKEPPEVAFPDKKDVDENFKKIIKMHLLNGNYTAVATHDDAIIEYTKQLVKEY
NIPNSQFEFQMLYGIRPERQVQLAREGYTMRVYVPYGTDWYGYFMRRLAERPANVAFVLK
GILRK
>Deinococcus Q9RW55 Q9RW55_DEIR
MIDQLYRKAVLTVAERPQVEQLARQKMWNLAERFVAGESIESAIQAVQALERDGIAGNLD
LLGEFIDSPAKCTEFADDVIKLIEAAHAAGIKPYVSIKLSSVGQGKDENGEDLGLTNARR
IIAKAKEYGGFICLDMEDHTRVDVTLEQFRTLVGEFGAEHVGTVLQSYLYRSLGDRASLD
DLRPNIRMVKGAYLEPATVAYPDKADVDQNYRRLVFQHLKAGNYTNVATHDERIIDDVKR
FVLAHGIGKDAFEFQMLYGIRRDLQKQLAAEGYRVRVYLPYGRDWYAYFSRRIAETPRNA
AFVVQGMLKG
>Bacillus Q81XI4 Q81XI4_BACAN P
MEQLMRNSFLFLSKNKALTKLAKKYGLRFGAGRFVAGETIELATAAIQALNKQGLCVTID
YLGEFVDNEAEANEMANQSIEAIRAIGREGLDSQLSLKMTSMGLDISDEIVMNNMRRILE
AAKENGVFVTIDMEDYTRCGKTIDIFKQLKSEYDNIGTVIQAYLYRTEKDIEDLNAYNPN
LRLVKGAYKEPEEVAFPDKKDVDDNYKKIIKMHLLNGNYTAIASHDEAIIEYTKKLAEEH
NIPRDQFEFQMLYGIRNERQLELVKEGYKMRVYVPYGNDWYGYFMRRLAERPANVAFVLK
GMVKK
>Exiguobacterium Q41B01 Q41B01_
MERVLRDGFIFLSQNKTLNGLAKRYGLKFGAGRFVAGDSLEASKRAIQDLNSKGLCVTMD
HLGEFISTVAEAEMMTQECIRAVEMIAEEKLDSQLSLKLTSLGLDISDELIRSNMERILT
RAKEVGVFVTIDMEDEPRCEKTIQLFEELRQSFDNIGTVIQSYLYRSEDDIKRVGRFETN
LRIVKGAYKEPATVAFPEKEDVDHNYIKLVKLQLSLGNYAAVATHDDAMIEQIIRYAKEN
GIGNDKFEFQMLYGIRVERQLELVKLGYKVRVYVPYGRDWYGYFMRRLAERPANVAFVLK
GMVKG
>Symbiobacterium Q67RR3 Q67RR3_
MSSLARKAILSLAGNRFVTRVMTRHGLKIGAGRFVAGVTLQDAVAVTRRLNDEGLAVTLD
LLGEGVTDRDGARAMAEGCAGILEAIGREKLDANLSVKLTQLGLAVDPEAAMENMVYLQD
IARQAGIFIRIDMESSDVTDVTLDIVRRLYARERNVGTVIQAYLYRSPKDLEELARLGMN
VRLVKGAYLEPPSVAYPKKADVDAAFRRLIAQHLSAGCYTAVATHDDAIIAFTESFVKEK
GIPRDQFEFQMLYGIRPERQRELARAGYRTRVYVPFGSDWYPYFVRRLAERPANVWFVLA
NLMKP
>Solibacter Q43US3 Q43US3_SOLUS
MVRAFFLFLSRQKQLRRWMETSPVAQRLSKRFVAGETLEQALAVSRKLNKQGITVTLDHL
GESVTSLAEAREARDVYRRTLDAIHAGDIQGNVSLKLTQFGLDLSYEECLANVEQLVRRA
AELGSFVRVDMESSEYVDRTLDLVRTLHARHGAVGVVIQSYLYRSKDDIAKLCAEKIRVR
LCKGAYLEPASLAFPQKSDVDRSYVELTQKLLTYGVYPAIATHDEAMIGQTKQFVSSKGI
```

```
GRECFEFQMLYGIRADLQRQLVADGYRLRLYVPFGKAWYPYYMRRLAERPANVLFILRNL
FRT
>Acidothermus Q2DYK6 Q2DYK6_ACI
MLRRLILEAADSQTLRRLIATAPPTRAVVHRFVAGTEVADGLAVARQLVADGLLVSLDRL
GEQVRDLSQARATAEAYRELAAAIEQAGLAAEVEISLKLSALGLGLAGNASVGSPEEPRR
VALDLAREICAAAQAAGTMVTFDMEDHTTTDDTLAIVAELRREFPSVGCVIQAYLRRSLA
DCRELAEAGARVRLCKGAYREPAEVAFTRRHEVDRNFVRCLRILMRGSGYPMVATHDPRL
IAIASVMATQAGRGKDTYEFQMLYGVRPDEQQRLVRRGERVRVYVPYGGQWYPYLMRRLA
ERPANVAFFLRALGSKK
>Kineococcus Q40WW1 Q40WW1_KINR
MFGQLLLGVAGNRGVRSLVTGSSLSRPVVARFVAGDDVDAATAAVRTLTGDGIAATLDRL
GEDVTEEAQADETVAGYRDLVERLAAEGLAAGNEISIKLSALGQGLGRSGPQRATERAHD
LAAHAREHGVDVTVDMEDHTRVDDTLTTVAALRADFPRTGCVLQAMLRRTEGDARDLAVA
GSRVRLVKGAYNEPPEVAYPAKADVDKAYVRCLRTLLDGGAYPMIATHDLRIVNLAEELL
RGRDAGTAEFQMLYGIRAPEQQRLARAGHVVRVYVPYGTDWYGYFSRRLAERPANLAFFA
RSLVAG
>Salinibacter Q2S542 Q2S542_SAL
MKLPFFLASRFVAGETLETSLPVVDDLNQDGLHVALDKLGEHVQNRSEAIAARDAYIDLV
RTMAERDDHGQRNRISIKLSMMGQLIDEDFCEENLRRLLEVAAEHDLFVRLDMEGSDLTQ
STLDLFEAVYPDFPNHVGPVLQAMLKRTDRDVDRMCELGVSVRLCKGAYAEPTSLAYQDM
SQIRERYLDYTERLLQHTDDSGIATHDDELIEATKSFVDRRGIDRDDFEFQMLYGLRRTT
QREMAADGYNMLVYVPYGTEWFPYFSRRLREKKENVWFVLRSLFQG
>Nocardioides Q3GZ84 Q3GZ84_9AC
MSLLRQPLLLLARSSLVKKVVSGMPVSAGIVRSYVPGETTESAVGATAELVDGGLRVTLD
YLGEDTTDAEQADATVAAYLDVLQELSARGLTRNAEVSVKLSAIGQFLPDAVGFGGGEKI
ALENARTICRAARNAGTAVTLDMEDHTTTDSTLSILRELRKDFPETGAVLQAALHRTESD
CRALAYEGSRVRLCKGAYLEPESVAFQDRLEISKSYVRCLKVLLAGEGYPMIATHDPRMI
EIASSLASRYGRAAGTYEYQMLYGIRPEEQRRLVASGETMRVYVPYGTEWYGYLMRRLAE
RPQNLSFFVKSLVSKK
>Oceanobacillus Q8CXJ6 Q8CXJ6_O
MANLTRDFFIGLSNNKLLNTNAKKYGFRLGAEKFVAGTNFDSIIGIIKDLNSEGISCTLD
NLGEFVTEKSEAIEARDDIIDMLYKIHEQRLDCHVSVKLTQLGLDIGDDFCINNMHAILK
VANRFEIFINIDMEKHIHYGKTLEILNELRKEYDNVGTVIQSYLYSAEDDLAALEDVRIR
LVKGAYKEDASIAYPSKEDIDRNFMELAKKRLLGNTFTSIATHDHNIIEELKSFVDEHNI
SRDIFEFQMLYGFRTEMHNELAQAGYHFCTYIPFGSDWFGYFMRRLAERPQNINLVLKDV
FYTKGNKLKKEPVIAGVAALSLLMYVNKKRKK
>Nocardia Q5YX89 Q5YX89_NOCFA P
MSGFTRLLRPALLAAAASPRVERTVTSLRATRALVDRFVAGTTEADALGTVEALLASGRW
ITVDHLGEHTTDPARARDTVAHYRRLLAALAELPAAGAPGTVRPLEVSVKLSALGQALPG
DGAAIALDHAREICTAASAAGIGVTVDAEDHTTTDATLAVVRELRADFPGVGTVLQAYLR
RTEGDCREFAGPGSRIRLCKGAYREPASVAFQRAAEVDDSYRRCLRILMSGRGYPMVASH
DPAMLAEADRLAAETGRGPDDFEYQMLFGIRDAEQRRLADAGHRMRVYVPYGDQWYGYFM
RRLAERPANLAFFLRAAASRDRA
>Streptomyces Q9Z560 Q9Z560_STR
MLGPVILAASRSDRMRRLVSAAPVTKPVVDRFIPGETVDQIVPIVRDLTDQGLELTMDVV
GEDITTPAQAEAARDAYLELIDRLKPLELGTRAEMSVKLSMFGQALDGGHELALANVRPV
VEAAAEIGTTVTLDAEDHTTLDSMFAIHEELRKDFPQTGCVIQAYLFRTEADARRLADSG
SRVRLVKGAYKEPAEVAYQQRHEIDKAYVRILRTLMEGEGYPMIGSHDPRLISIGQELAR
TAGRKLDEYEFQMLYGIRGDEHLRLAAEGHRMRVYTAYGTDWYGYFMRRLAEKPANLRFF
ARSMVSKG
>Chlorobium Q4AI77 Q4AI77_9CHLB
MNMFNKLIASTLPYMPQSLVWVFSKKYIAGKTLDSAIDISKELNKNGSMVTIDLLGEFIT
KLDEARSNKNEYLGIIDNVQKQKIDGNYSLKPTMFGLLLDEEACYRNIREIVKKAASYNN
FVRIDMEDSQCVDMEIKLFRRLKEEFPANVGLVLQAYLKRTYQDITDMLDLNSKQLPVNY
RLCKGIYVEPESIAYKKYEEINEHYLADLEYMFQQGIYPGIATHDFRLIEGAYKLIEKFN
VPKDKFEFQMLYGVTPGLRQSIIDKGFRMRVYVPFGEKWFAYSTRRLKENPRMAQDIIKG
IFIKG
```

>Burkholderia Q396J7 Q396J7_BUR
MRILNTMAARAIPLVPRSLIRKISRRYIAGETLCDARTRIRALHAAGFRTTVDVLGESAS
SSEQAESMTREYLNLIDALGAEREPTELSIKLTALGLHLDKDACMTRVAAILQSAAAHGI
NACIDMEDIRFTAKTLDAFSKFEADGYAIGIALQAYLKRTSDDLVPLLARKSSMRICKGI
YAEANEHLVDGASRDRAAINAHFVRHVSTALEAGSFVGIATHDAPLIDVLTGWLQRAQID
RSRFEFQMLLGVCEPMRDALRAQGFNVRIYVPYGQDWYGYSTRRIKENPRIAGYILAAMV
RADRMR
>Staphylococcus Q8CNU4 Q8CNU4_S
MSLFKDFFIALSNHTYLNKIAKKMGPQMGANRVVAGNTIHQLIETIQYLNDYNISVTVDS
LGEFVNTREESIKAKEEILEIIDAIYNNNVKAHMSVKISQLGSEFDLNLAYENMREILLK
ADKNGKMHINIDTEKYDSLSKIQHIIERLKGEFKNVGTVVQAYLYEADDIIDKYPELRLR
LVKGAYKEDASIAFQSKEEIDANYIRIIKKRLLNSKNFTSVATHDNEIINQVKQFMKENH
ISKDKMEFQMLYGFRTELSQKIANEGYFFTVYVPYGNDWFAYFMRRLAERPQNLSLAIKE
FTKPKILKKATLGIGIFATLLTSLILGIKRHKK
>Frankia Q2J961 Q2J961_FRASC Pr
MIQKLMPPVSRHPLTRALVTRTPLARGIIDRFVAGPSDTDAIAAVAGLADLGLLATVDLL
GEDITRAQDAADTLLAYLGLLDLADRAGIAAGLDVSVKLSALGQAVPRDGRKLSYENAAE
LCARAAAANATVTLDMEDHTTTDSTLDTLVELRRDFPFVGAVLQAQLLRTETDCRDLATA
GSRVRLCKGAYREPSAVAYQGRPAVAAAYARCLGILLTGSGYPMIATHDPALLNLAGRFA
EHVGRLPATFEYQLLYGVRPGEQRRLASTGATVRVYVPYGADCVPYLTRRLVERPTNLLL
AARALRSRA
>Mycobacterium Q73WR8 Q73WR8_MY
MAGLFAHTLRPAILAAGRRPGLRRAAEALPVTRRVVHRFIAGETIDSALDSVTALRDSGR
CVSVDHLGEDVSDADDADAAVRVYLELVDRLGCLGDGAAVRPLEVSLKLSALGQSLDRDG
EKIARENAWAICAAAQRAGVWVTVDAEDHTTTDSTLRIVRDLRREFDWLGVALQAYLRRT
LGDCEEFAAAGARVRLCKGAYDEPASLAYRDPAAVTDSYLRCLRVLMAGSGYPMVASHDP
AIIEAAPSLARESGRGTGEFEYQMLYGIRDDEQRRLAEAGHTVRVYVPFGTQWYGYFMRR
LAERPANLTFFLRALAARRH
>Thermobifida Q47SU5 Q47SU5_THE
MLRTLLLLAARSPRCRDFVTRSPLTRGVVRRFVAGDTLDATLPVLRQLTADRHVTVDYLG
EDITDAEQASRTVDAYTHLLTALGEHGLAPRAEVSVKLSALGLALPRDGAQLALDHARQI
CAAAEAVGTTVTVDMEDHTTTDATLDIVHQLRVDFPRTGAVLQAYLRRTEADCRDLAVPG
SRVRLCKGAYDEPATVAYRDRAEVDRSYVRCLRILMDGPGYPMVATHDPRLLRIAAFLAR
SAGRSADSYEYQMLYGVRTDLQAALADRGERVRVYVPYGEQWYGYFMRRLAERPANTAFF
ARALFAGGAVR
>Halobacterium Q9HNG0 Q9HNG0_HA
MIPPIANRFVAGETPAEAIEYAADLNDRGVTAILNLLGEHYHERPPADEDAAAYRRLVAD
VGDSDVDACVSVKPSQIGLDVGPEVFAENLETIAAAAADHDVFAWVDMEDHETTDATMDT
FESLARTHAGNMGLCVQANLKRTEQDLDRLADVPGKVRLVKGAYDEPDDVAYQDKPDVNE
AYRTHLEFMFREFDDGVAVGSHDPAMISYAAELHEEYGTDYEVQMLMGIREDAQTELAAT
GVPTYQYVPYGGKWFSYFYRRAMERKENLVFALRAVLGR

# CHAPTER 3

## Mechanism-Based Inactivation of *Thermus thermophilus*

## Proline Dehydrogenase by *N*-Propargylglycine

### <u>INTRODUCTION</u>

Mechanism-based enzyme inactivators contain latent functional groups that upon catalysis are transformed into a chemically reactive species (FIGURE 3.1). The reactive species typically becomes covalently bound to the enzyme before being released from the active site, thereby irreversibly inactivating the enzyme (1,2). These inactivators show time-dependent, first order and irreversible loss of enzyme activity, and inactivation is typically diminished in the presence of substrate (1). Mechanism-based inactivators are also known as suicide inhibitors, $k_{cat}$ inhibitors, suicide enzyme inactivators, and irreversible inhibitors. Mechanism-based inhibitors have proven useful in elucidating catalytic mechanisms for a variety of enzymes (1,2).

*N*-propargylglycine and related compounds have been reported to be mechanism-based inactivators of several enzymes, including flavoenzymes, since the late 1970s. For example, Walsh's group studied inactivation ofD- and L-amino acid oxidases, L-alanine transaminase, gamma-cystathioninase, and methionine gamma-lyase by various propargylglycines (3-7). Kraus and Belleau reported mechanism-based inactivation of the flavoenzyme monomeric sarcosine oxidase by *N*-propargylglycine (8). The proposed reaction mechanism included abstraction of the $\alpha$-proton from the propargyl group, and consecutive

nucleophilic attack by enzyme (8). These studies also established that proton abstraction was the rate-limiting step of the reaction because of the reduced inactivation rate observed with the deuterated inactivator $N$-($\alpha$,$\alpha$-didueteriopropargyl)glycine. Various substituents added to the propargyl frame have allowed mechanism-based enzyme inactivation of other enzymes. For example, N,N-dimethylglycine oxidase is irreversibly inhibited by $N$-methyl-$N$-propargylglycine (1). And several inactivators of the flavoenzyme monoamine oxidase have been developed around the propargyl moiety, including rasagiline (9,10).

Although reversible inhibition of proline dehydrogenases (PRODHs) by compounds such as THFA, L-lactate, and acetate is well documented, there is only one report of mechanism-based inactivation of PRODH. Tritsch, et al, reported that 4-methylene-L-proline (FIGURE 1.2) is a mechanism-based inactivator of PRODH activity in rat liver mitochondrial suspensions (11). It was shown that this compound is oxidized by PRODH with a $K_m$ of 10 mM. This inactivator showed pseudo-first order kinetics, and PRODH activity was protected upon addition of substrate, L-proline, or L-lactate. The reported Kitz and Wilson kinetic constants were $K_I$ = 5 mM and $k_{inact}$ = 0.117 min$^{-1}$ at 30 $^o$C. A proposed mechanism involves oxidation of the inactivator to $\Delta^1$-pyrroline-3-methylene-5-carboxylate and subsequent nucleophilic attack by the enzyme on the electrophilic methylene C atom, yielding the inactivation complex shown in FIGURE 3.2. The identity of the enzyme nucleophile is unknown.

We report here the first structure of a PRODH inactivation complex along

with characterization of inactivation kinetics and spectral changes caused by inactivation. The structure of *Thermus thermopilus* PRODH inactivated by *N*-propargylglycine shows that the ε-amino group of Lys99 is covalently attached to the FAD N5 atom via a 3-carbon linkage. To our knowledge, this is the first report of this type of an inactivation complex for a flavoenzyme.

## MATERIALS & METHODS

### Crystallization

*N*-propargylglycine was kindly provided by Prof. Christian Whitman of the University of Texas-Austin. TtPRODH was expressed and purified as previously reported (12). Purified TtPRODH (3 mg/mL) was incubated with a 300 molar excess of *N*-propargylglycine (1mg of protein per 1 mg of inhibitor) for 30 minutes. The inactivated enzyme was pipetted (2 μL) into sitting drop crystallization trays, mixed with an equal volume of reservoir solution, sealed with clear tape and placed in an incubator at room temperature. The optimal reservoir consisted of 100 mM imidazole pH = 7, 100 mM $MgCl_2$, 14% 2-methyl-2,4-pentanediol and 5 mM fresh dithiothreitol. Thin football shaped crystals formed overnight and had dimensions of 0.4 mm x 0.2 mm x 0.1 mm (FIGURE 3.3). The crystals were pale yellow compared to oxidized TtPRODH crystals, which are vibrant yellow (12). Loss of yellow color implicates of reduction of the FAD. These crystals were cryoprotected with reservoir buffer containing 25% 2-methyl-2,4-pentanediol and inhibitor, picked up with mounting loops, and plunged into $N_2$ (*1*).

**X-ray Diffraction Data Collection, Processing, and Refinement**

A 1.9 Å data set was collected at the Advanced Light Source Beamline 4.2.2 using a NOIR 1 detector. The data set used for structure determination consisted of 144 images with a crystal to detector distance of 130 mm, oscillation range of 1 degree per frame and exposure time of 20 s/frame. Data collection was performed at an X-ray wavelength of 1.0 Å at -180 $^{o}$C. Integration and scaling were performed with d*TREK (13)

The space group is $P2_12_12_1$, with unit cell dimensions a = 82.3 Å, b = 90.2 Å, c = 94.9 Å. We note that the structure of oxidized TtPRODH was determined from crystals also having space group $P2_12_12_1$ and similar unit cell lengths (a = 82.1 Å, b = 89.6 Å, c = 94.3 Å (14). Data processing statistics are listed in TABLE 3.1. The data set used for refinement is 99.3% complete and 5.8-fold redundant.

The structure of inactivated TtPRODH was determined using molecular replacement as implemented in MOLREP (15). The search model consisted of residues 40-281 of oxidized TtPRODH. FAD and active site residues including Lys99 were omitted from the search model. A clear solution was identified with 2 molecules in the asymmetric unit, having a correlation coefficient of 0.644 and R-factor of 0.372. The model from molecular replacement was extended and improved with several rounds of model building in COOT and restrained refinement with TLS in REFMAC5. Ten optimal TLS domains were determined with TLSMD (16). The test set used for R-free calculations was based on the set used previously for refinement of oxidized TtPRODH owing to the similarity in the

crystal forms.

Special attention was required to model the modified lysine and reduced FAD. The modified lysine 99 was created using CCP4's monomer library. Lysine was found in the library (LYS.pdb) by searching for amino acids. Using CCP4's monomer sketcher, a 3-carbon modification was drawn on lysine's $\epsilon$-amine. The 3-carbon modification contained a double-bond between the $\epsilon$-amine and the C1, and another double bond between C2 and C3. Monomer sketcher was then used to create a library description for this molecule, which was named LYX. FAD was found in the CCP4 monomer sketcher library (FAD.pdb), and the isoalloxazine planar restraints were changed by manually editing the cif file. One planar restraint group contained the atoms in the pyrimidine portion of the isoalloxazine. The other planar restraint group contained the atoms in the dimethylbenzene portion of the isoalloxazine. Atoms N5 and N10 of the isoalloxazine were not contained in the planar restraints. The planar restraints for the adenine portion of FAD were left unchanged. This molecule was named FDH, and a library description was created using the monomer sketcher.

**Kinetic and Spectroscopic Characterizations**

Inactivation parameters were determined using the Kitz and Wilson approach (17). The assay used for this study was the DCPIP-based assay described previously (14,18). Samples of TtPRODH (0.3 mg/mL) were incubated at 25$^{\circ}$C with 0.5 mM, 1.0 mM, 2.5 mM, and 5.0 mM *N*-propargylglycine. Aliquots were removed at various time points over 30 minutes and assayed immediately with a constant proline concentration of 50 mM to determine remaining activity.

The initial linear parts of the Kitz and Wilson plot were used in the replot analysis. Using Origin 7.0, the normalized data were plotted and fit globally sharing parameters A and $Y_o$ with the exponential decay equation

$$Y = Y_o + Ae^{-x/t}$$

The half-lives were determined by calculating t and multiplying t by 0.693. The half-lives were plotted versus inverse of the *N*-propargylglycine concentrations to make the Kitz and Wilson Replot. The data in the replot were fit linearly and inactivation parameters, $K_I$ and $k_{inact}$, were calculated from the equation of the line.

Absorption spectra of the bound FAD cofactor were acquired with a Cary 100 spectrophotometer. Spectral changes caused by inactivation were examined as follows. TtPRODH was added to a quartz cuvette at 0.2 mg/mL and spectra were acquired at 25 $^o$C in the wavelength range 300-600 nm. A quartz cuvette with buffer was used as the reference. Inhibitor was added to a final concentration of 500 $\mu$M, the solution was mixed and spectra were obtained every 30 s for 60 min.

An additional set of experiments was performed at 4 $^o$C for use in estimating inactivation constants. For this study, the decrease in absorbance at 452 nm was monitored with *N*-propargylglycine present at 5-10 mM and an enzyme concentration of 0.25 mg/mL.

**Mass Spectrometry**

TtPRODH was dialyzed overnight into 100 mM ammonium acetate, and then inactivator was added at a 300 molar excess. The inactivated enzyme was analyzed using Nanospray QqTOF mass spectrometry by the University of Missouri Proteomics Core.

## RESULTS

### Structure of TtPRODH inactivated by *N*-propargylglycine

Inactivation of the enzyme caused no noticeable change in the global structure of TtPRODH (FIGURE 3.4 and FIGURE 3.5). Two major changes, however, were observed in the active site. First, electron density maps clearly indicated that the FAD isoalloxazine was no longer planar, as in the oxidized enzyme, but rather it appeared highly bent with the ring system bowed toward the proline binding pocket. Second, it was apparent that Lys99 had been covalently modified at the ε-amino group. Electron density could be seen connecting the ε-amino of Lys99 with the N5 atom of the isoalloxazine, suggesting that the two were linked by a covalent connection. These features were present in both TtPRODH molecules in the asymmetric unit. A model consistent with these electron density features was built and refined.

The final refined structure shows that the FAD isoalloxazine is bent along the N5-N10 axis (FIGURES 3.6-3.9). This type of distortion from planarity is known as butterfly bending (19). The butterfly angle of inactivated TtPRODH is approximately 25°. The planar isoalloxazine conformation is typically associated with the oxidized state of the flavin, whereas highly nonplanar isoalloxazine

conformations indicate reduced states. The butterfly angle of 25° observed here is rather large for flavoenzymes, and it suggests that the FAD of inactivated TtPRODH is reduced.

Electron density between Lys99 and the FAD N5 was modeled as a 3-carbon covalent linkage that connects the ε-amino group of Lys99 with the FAD N5 (FIGURE 3.6). The model is consistent with mass spectral data. Native TtPRODH had a molecular mass of 37976 Da, which corresponds to the protein without FAD (FIGURE 3.10). The inactivated enzyme showed two mass spectral peaks, one at 37976 Da and another at 39797 Da (FIGURE 3.11), with the latter peak presumably representing the inactivated enzyme. The difference in mass between the apo and inactivated enzyme is 821 Da, which is approximately equal to the sum of the masses of reduced FAD (FADH⁻, 787 Da) and the modeled 3-carbon linker =CH-CH=CH- (39 Da).

Aside from the butterfly bend and covalent modification of N5, the conformation of the cofactor is identical to that of the oxidized enzyme. We do, however, observe that the cofactor shifts slightly in the active site toward Lys99 as shown in FIGURE 3.5.

Inactivation also causes conformational changes in active site residues other than Lys99 (FIGURE 3.12). Most notable is the movement of Arg184. In the oxidized enzyme, the guanidinium group of Arg184 donates a hydrogen bond to the flavin N5 (3.2 Å). This residue is universally conserved throughout the entire PRODH family as discussed in Chapter 2 and in White, et al. (14). The Arg-N5 hydrogen bond is also present in PutA PRODH domain structures

(20,21). Inactivation of TtPRODH causes the guanidinium to move away from the flavin, so that closest distance between the Arg guanidinium and FAD N5 is 3.9 Å. Rupture of this hydrogen bond is consistent with the transfer of hydride to the FAD so that N5 is protonated.

Other changes are also observed in the active site. The hydrogen-bonding pair Asp133-Tyr275, which forms the ceiling of the active site, moves 0.5 Å away from the active site. This movement is probably due to steric clash with the flavin's dimethylbenzene, which has shifted toward Lys99. Also, movements of up to 0.4 Å are observed in the loop between β–strand 5. This loop movement starts at Tyr190 and ends at Lys201. Similar movements are present in chain B, although they are not as large. Lastly, we see movement of helix 8 away from the active site. Helix 8 contains conserved Arg residues that orient carboxylate-containing ligands, such as the substrate proline, in the active site (21). Helix 8 containing completely conserved Arg288 and Arg289 has much weaker density than the rest of the barrel. As compared to the oxidized structure, this helix in the inactivated structure moves away slightly from the active site. The B-factors and deviation from the oxidized structure both increase proceeding along the helix, reaching B = 50 Å$^2$ and 0.5 Å deviation at Arg289. With the exception of Arg288, this trend is also present in chain A although the deviations are not as large comparing oxidized and inactivated enzymes. In chain A, Arg288 has moved out of the active site and forms hydrogen bonds with Thr287. In chain B, Arg288 is oriented toward the active site, as seen in oxidized TtPRODH.

A strong electron density feature that could not be assigned to the protein

or the modified cofactor was observed in the active site.   As discussed below, our proposed mechanism predicts that glycine is a byproduct of the inactivation reaction.   Thus, we considered the possibility that this electron density feature represents a bound glycine molecule.  Although the inhibition constant for glycine has not yet been determined, initial assays show that glycine is an inhibitor.  A glycine molecule was modeled into this density feature, but the average B-factor refined to the unacceptably high value of 74 $Å^2$ (FIGURE 3.13).   We note that the surrounding residues have B-factors in the range 40-50 $Å^2$.  Moreover, there are no hydrogen-bonding partners nearby, with the closest interactions occurring with Tyr190 at 4.8 Å, Y285 at 4.8 Å, isoalloxazine at 4.61 Å and R289 at 3.9 Å. Based on this information, it is likely that the occupancy of glycine in the active site is low or that this feature represents other bound solvent species such as water.  Given these results, this electron density feature was left unmodeled.

**Spectral changes caused by inactivation**

Oxidized TtPRODH showed a characteristic flavin absorption spectra with maxima at 452 nm and 381 nm (FIGURE 3.14).   The spectrum changed significantly upon addition of *N*-propargylglycine.   The peak at 452 nm disappears over time to baseline, which is similar to what was previously seen with electropotentiometric and proline titrations of TtPRODH.  The peak at 381 nm first decreases and then gradually increases throughout the duration of the experiment.  Note that the peak at 381 nm not only increases, but also shifts to a longer wavelength of 386 nm.

The absorbance values for $\lambda$ = 382 nm and 452 nm are plotted as functions

of time in FIGURE 3.15. We see that the rate of decrease of $A_{452}$ is much faster than the rate of appearance of $A_{382}$. This indicates that the two spectral features are monitoring different steps of the inactivation reaction. The consistent decrease of $A_{452}$ is very similar to the behavior observed during potentiometric and proline titrations of TtPRODH, which suggests that the cofactor is reduced during inactivation of the enzyme. In contrast, an increase in absorbance in the range $\lambda$=380-386 is not observed during potentiometric and proline titrations of TtPRODH. This suggests that the spectral feature in this region arises from a new flavin species that is formed during inactivation. This new species is presumably the covalently modified FAD described in the previous section.

The reductant sodium dithionite was added after 60 minutes of incubation of the enzyme with *N*-propargylglycine. As shown in FIGURE 3.16, the strong absorbance at 386 nm remains, and there is no further decrease in absorbance near 452 nm. These results suggest that the cofactor is in an irreversibly reduced state, which is consistent with the bent flavin and covalent modification observed by crystallography.

**Kinetics of inactivation**

Before describing our results, we present a brief summary of the kinetic framework used for analyzing mechanism-based inactivators because it differs somewhat from that used to analyze reversible inhibitors. Two kinetic constants describing rates of inactivation can be obtained for the mechanism-based inactivators (22). Referring to FIGURE 3.1, the inhibition constant, $K_I$, is the ratio of $k_{-1}/k_1$, when $k_2$ is rate-limiting. The second kinetic constant, $k_{inact}$, is more

complex and comprises rate constants $k_2$, $k_3$ and $k_4$. Step 2 ($k_2$) is most often the rate-limiting step, and assuming infinite concentration of inactivator the expression for $k_{inact}$ is

$$k_{inact} = (k_2 * k_4) / (k_3 + k_4)$$

The rate constant of inactivation is sometimes incorrectly referred to as $k_2$. The rate constant $k_3$ describes the escape of the inactivator from active site before covalent modification. $k_4$ describes conversion of the inactivating species to the final inactivated enzyme. $k_3$ and $k_4$ are related by the partition ratio which describes the released product per inactivation event and is defined by $k_3 / k_4$.

To determine these kinetic constants, $K_I$ and $k_{inact}$, one ideally utilizes a continuous assay. Enzyme and various concentrations of inactivator are incubated together. Aliquots are removed at various time points and assayed with substrate. The log of the percent of remaining enzyme activity is plotted versus time, a so-called Kitz and Wilson Plot (FIGURE 3.17) (17). Half-lives of inactivation are determined from the Kitz and Wilson plot for each value of the inactivator concentration, and these half-lives are plotted versus the inverse of inactivator concentration. The resulting plot is known as the Kitz and Wilson Replot and the curve, in theory, should be linear as shown in FIGURE 3.18.

The kinetics of inactivation for TtPRODH was studied using a dye-coupled continuous activity assay. The Kitz and Wilson plot is shown in FIGURE 3.19. Interestingly, the initial parts of the curves are linear but the inactivation rate

flattens out at longer time, indicating a departure from pseudo first-order kinetics. This phenomenon, called non-pseudo first-order kinetics, can occur for various reasons, including high partition ratio ($k_3/k_4$ or escape vs. turnover) and generation of a secondary metabolite that binds to the active site (23). In the present case, the second explanation may be reasonable because, as discussed below, $N$-propargylglycine is oxidized by TtPRODH and subsequently hydrolyzed to two metabolites, propynal and glycine. Moreover, electron density maps suggested that glycine may be bound in the active site, albeit weakly (FIGURE 3.13).

The inactivation constants estimated with these data are $K_I$ = 838 $\mu$M and $k_{inact}$ = 0.3 min$^{-1}$ (FIGURE 3.19). These data agree with the previously determined kinetic parameters determined with 4-methylene-L-proline using rat liver PRODH at 30$^{\circ}$C (11). Spectroscopic data were also used to estimate inactivation constants. Kitz and Wilson plot and replot were also obtained when using $A_{452}$ to follow the reaction (FIGURE 3.20). These data were collected at 4 $^{\circ}$C in order to slow the inactivation rate. This analysis resulted in estimated inactivation constants of $K_I$ = 206 mM and $k_{inact}$ = 6.6 min$^{-1}$.


## DISCUSSION

### Inactivation of monoamine oxidase by propargylglycine compounds

Maycock and colleagues performed pioneering studies of inactivation of monoamine oxidase (MAO) by propargyl-based compounds. This work has inspired our thinking about the mechanism of inactivation of TtPRODH by $N$-

propargylglycine, we briefly summarize their results in this section.

MAO is a tightly bound outer mitochondrial membrane enzyme that oxidizes biogenic amines to their corresponding imines (24). Two isoforms, A and B, have been characterized. The different isoforms are localized together in some organs like liver, although only MAO A is found in placenta, and only MAO B is found in platelets (25). MAO plays roles in metabolism of neurotransmitters, like dopamine and serotonin, and is a target for treatment of depression, dementia, and early stages of Parkinson's disease. MAO inhibitors are currently used for treatments of various neurological disorders.

Maycock and Abeles studied inactivation of MAO by propargyl-containing compounds such as dimethylpropargylamine (1,26). Upon flavoenzyme catalysis of a mechanism-based inactivator, the flavin itself becomes modified (FIGURE 3.21). Based on the chemical and spectral properties of the isolated flavin-adduct (27), Maycock determined that flavin N5 was modified and proposed three likely reaction mechanisms (27). The proposed reaction mechanisms included carbanion formation, radical pair complex, or Michael addition of flavin nucleophile to the acetylenic moiety (FIGURE 3.22).

In the carbanion-based mechanism, the enzyme catalyzes the removal of the $\alpha$-proton from the inactivator to yield carbanion (1,27) (FIGURE 3.22B), which can subsequently react with the oxidized flavin N5 to form a modified flavin (1,27). In this proposed mechanism, a reduced inactivator reacts with an oxidized flavin (27). The most likely scenario for this reaction mechanism to occur is using $\alpha$-amino or $\alpha$-hydroxy acids where an activating group is adjacent to the site of

proton removal (27).

The radical pair-based mechanism involves reaction of a partially reduced flavin and partially oxidized inactivator (27) (FIGURE 3.22C). The enzyme could catalyze the removal of the $\alpha$-proton to the flavin or could catalyze electron transfer from an initially formed substrate carbanion to the oxidized flavin. Radical pair collapse and subsequent protonation leads to formation of the adduct (27).

Complete oxidation of the inactivator generates reduced flavin and the highly reactive species dimethylpropargylimine (27) (FIGURE 3.22D). Upon Michael addition, a reduced flavin would be reacting with an oxidized inactivator. The reduced flavin has been shown to act as a nucleophile, utilizing both the N5 and C4a positions to perform this reaction (27). Williams studied free dihydrolumiflavin and its anionic N1 form on reactions with various carbonyl-containing pyruvate derivatives and found that indeed, reduced flavin could act as a nucleophile occurring by the proposed carbanion mechanism (28). Confirmation of a nucleophilic enzyme-bound flavin was demonstrated in UDP-galactopyranose mutase (29). Addition of the reduced flavin N5 across the $\gamma$-carbon acetylene of dimethylpropargylimine by Michael addition would give rise to the monoamine oxidase flavin adduct that Maycock isolated (27) (FIGURE 3.21D).

Crystal structures have been determined for MAO inactivated by propargyl-containing compounds have been determined (PDBIDs 2C64, 2C65, 2C66 1S2Q, 1S2Y, 1S3B, 1S3E) (9,10,30)). These studies are important because they

definitively established the structure of the flavin adduct of the inactivated enzyme. Significantly, the structures for the enzyme inactivated by rasagiline reveals that the flavin is modified at the isoalloxazine N5 position, as in TtPRODH (FIGURE 3.23).

**Proposed mechanism of inactivation of TtPRODH by *N*-propargylglycine**

The structural, spectroscopic and mass spectral data presented here shows that the cofactor of TtPRODH is reduced and that the flavin N5 is covalently attached to the $\varepsilon$-amino group of Lys99 by a 3-carbon linkage. The mechanism proposed in FIGURES 3.24-3.26 is consistent with these data and previous studies of MAO inactivation by propargyl compounds.

In the first step of the proposed mechanism, *N*-propargylglycine is oxidized to *N*-propargyliminoglycine with concomitant reduction of FAD. In step 2, *N*-propargyliminoglycine is nonenzymatically hydrolyzed of to propynal and glycine. Note that these two steps are analogous to oxidation of the substrate proline to glutamic semialdehyde. The third and fourth steps are (3) formation of a Schiff base between Lys99 and the propynal aldehyde group and (4) Michael addition of the reduced flavin to the acetylene group.

Considering the reaction of propynal with Lys99, the general reaction for typical Schiff base formation involves attack of a lone pair from the $\varepsilon$-amino of lysine at an aldehyde or ketone group (FIGURE 3.27). A carbinolamine intermediate is formed, and finally the imine is produced with the exclusion of water. Lys99 must be in the neutral ionization state ($-NH_2$) to form a Schiff base. The $pK_a$ of Lys is typically 10-11 (depending on reference used), we expect that

Lys99 to be positively charged at the pH used in these studies (pH=8). The structure of oxidized, uninhibited TtPRODH shows that Lys99 forms hydrogen bonds with water molecules in the active site. Structures of the *E. coli* PutA PRODH domain complexed with reversible inhibitors show that Lys99 interacts with the carboxyl groups of Asp370 (3.1 Å) and the inhibitor (2.7 Å). Based on the inhibited EcPutA PRODH, it appears that a general base is required for deprotonation of Lys99 during inactivation by *N*-propargylglycine. The identity of this base is unknown, but possible candidates are Asp133 or Glu135.

It is proposed that the cofactor is in a reduced state, based on the bleaching of the crystal, decrease of absorbance at 425 nm, and large butterfly bending of the isoalloxazine. In other flavoenzymes, isoalloxazine bending is associated with reduction (19,31). This bending is due to a change in hybridization of N5 from $sp^2$ to $sp^3$ upon reduction. The isoalloxazine is planar in structures of oxidized TtPRODH and the PutA PRODH domain. The recent structure of the E. coli PutA PRODH domain reduced by dithionite shows that reduction of the flavin induces a butterfly angle of $22^o$ (31), which is similar to the isoalloxazine of inactivated TtPRODH (FIGURE 3.28). Comparing the isoalloxazines of inactivated TtPRODH and reduced PutA PRODH reveals a similar butterfly conformation, but the ring system of reduced PutA is more symmetric than that of TtPRODH due to twisting of the isoalloxazine in inactivated TtPRODH (FIGURE 3.28 and FIGURE 3.29). The asymmetry in inactivated TtPRODH presumably results from strain due to covalent linkage to Lys99.

Following Maycock's work, we propose that the reduced FAD attacks the

acetylene group of the hydrolyzed inhibitor. Previous reports of nucleophilic flavins suggest that the flavin is most likely in the anionic form (FADH⁻), with a negative charge on N1. Negative charge on N1 enhances nucleophilicity of N5 by increasing the electron density in the isoalloxazine (29). Soltero-Higgin measured the p$K_a$ of anionic flavin using fluorescence, and the value of 6.7 indicates that a majority of the flavin is in the anionic form at physiological pH (29). In addition, the ribityl 2'-OH of TtPRODH is positioned to donate a hydrogen bond to the N1, which may further stabilize the anionic reduced FAD.

As described in the previous section, Maycock envisioned three possible mechanisms by which propynal reacts with the FAD: carbanion formation, radical pair complex, or Michael addition of nucleophile across the acetylenic moiety (27). A radical mechanism is the least likely scenario for TtPRODH inactivation because no stabilization of flavin semiquinone was observed under potentiometric titrations (14). Furthermore, sulfite reactivity, an indicator of radical formation, was not observed (14). The carbanion mechanism is more likely, since this reaction is favored for $\alpha$-amino acids due to the activating group next to the site of proton abstraction (1). The activating group is normally a carbonyl or carboxylate, but in the case of proline and *N*-propargylglycine, the carboxylate is 3-4 atoms away. Most carbanions need stabilization from an electronegative group, such as amino acid residue like glutamate and aspartate. In the active site, Asp133 is close to the site of proton abstraction. Based on the available data, we favor Michael addition of the anionic flavin N5 to propynal $\gamma$-carbon. Nucleophilic flavins have been investigated in free solution and

proposed for flavoenzymes (27,32). Soltero-Higgin, et al. reported convincingly that UDP-galactopyranose mutase utilizes a nucleophilic FADH$^-$ (29). Also, the imine formed by the Schiff base reaction is highly reactive and a good Michael acceptor.

**Comparison to structures of other inactivated flavoenzymes**

Mechanism-based inactivation has been observed for numerous flavoenzymes (33). These include monoamine oxidase, monomeric sarcosine oxidase, dimethylglycine oxidase, lactate dehydrogenase and cytochrome c (1). Inactivators that have been used include *N*-propargylglycine, dimethylpropargylamine amine, vinylglycine, 2-hydroxy-3-butynoate and rasagiline. Interestingly, X-ray crystallographic data is sparse for these mechanism-based inactivated covalently-modified flavins. Binda, et al. determined structures of inactivated monoamine oxidase A and B with various propargyl inactivators, such as rasagiline(9,10). The flavin is modified at the N5 atom as in *N*-propargylglycine inactivated TtPRODH (FIGURE 3.23). The structure reported here is different from rasagiline-inactivated MAO in that the FAD is covalently connected to a protein side chain. To our knowledge, this is the first report of such a modified flavoenzyme.

**Potential applications of mechanism-based inactivators of PRODH**

*Stress response*

Organisms accumulate proline in response to various stresses, including oxidative stress. Since PRODH degrades proline, specific mechanism-based inactivators of PRODH could enhance protection by proline. Plants, for example,

accumulate proline in response to extreme temperatures, salinity, drought, and heavy metals such as Cd, Cu, Zn (34-38). Proline concentrations as high as 1 M have been measured in plant tissues subjected to osmotic stress due to high salinity (34). Yeast and fungi also utilize proline accumulation upon exposure to stress (39).

The basis for protection is that proline can serve as an osmolyte, radical scavenger, electron sink, macromolecule stabilizer and component of cell walls (34). As an osmolyte, proline has been shown to prevent freezing by stabilizing membranes (40). Proline binds to head groups of membrane phospholipids, replacing water and preventing cell shrinkage upon freezing (41). As an antioxidant, proline scavenges reactive oxygen species like hydroxyl radical (42) It has been proposed that the hydroxyl radical either removes a proton from proline on the C5 to form water or can react to form proline-nitroxyl radical (43,44). Proline has also been shown to quench singlet oxygen ($^1O_2$) by forming an electron charge-transfer complex (45).

*Pesticide Development*

Proline is a major fuel for the flight muscles in *Glossinia morsitans* (46). *Glossinia* is a major vector for transmission of parasitic African trypanosomes, which cause trypanosomiasis (African Sleeping Sickness) and Chagas disease. Upon uptake by *Glossinia,* trypanosomes utilize L-proline as the respiratory substrate (47). If proline is the major fuel for Glossinia during flight and is important for trypanosome respiration, PRODHs from these organisms are attractive targets for pesticide and anti-trypanosome development. In fact, the

only other reported mechanism-based inactivator of PRODH, 4-methylene-L-proline, and was lethal to *Glossinia* within 12 hours at a concentration of 250 μg, although flight was not impaired (11).

*Manipulation of proline catabolism in eukaryotic cells*

Specific mechanism-based inactivators of human PRODH may be useful reagents for modulating PRODH activity in eukaryotic cells. One potential application is the study of the function of PRODH in cancer and apoptotsis. As discussed in the introduction to this thesis, PRODH (specifically PRODH2) is a pro-apoptotic protein that helps reduce carcinogenesis in humans by serving as a superoxide generator. The PRODH2 gene is induced by the cell proliferation regulator p53. Many cancers are due to mutations in p53, or inactivation of p53-mediated apoptosis. More than half of colorectal cancer cell lines show inactivation of p53. Overexpression of PRODH (also called proline oxidase or POX in this context) in colorectal cancer cell lines and various carcinomas induces p53-mediated apoptosis (48-51). The mechanism by which this happens is currently being investigated, although there is evidence it happens by both intrinsic and extrinsic pathways (52). Further research shows that PRODH induces apoptosis in a proline-dependent manner by superoxide generation, even without utilizing p53-mediated pathways (48). This is currently being investigated as a potential cancer treatment, although it is unclear whether proline-dependent superoxide generation is specific to cancerous cells. Mechanism-based inhibitors of human PRODH offer an alternative to gene inactivation and RNAi for studying the connection between proline metabolism

and apoptosis in the setting of cancer cells.

*Biophysical and structural studies of reduced PRODH and PutA*

Mechanism-based inactivators of PRODH lock the enzyme in the reduced state under aerobic conditions.  Thus, these inactivators could prove useful for biophysical and structural studies of PRODHs and PutAs because the reduced enzyme can be studied outside of an anaerobic chamber.   Potential uses include crystallization, NMR and ligand binding studies.   This should be especially useful for PutAs, which are thought to exhibit large conformational changes upon reduction of the FAD.

# REFERENCES

1.  Maycock, A. L. (1980) *Methods Enzymol* **66**, 294-302
2.  Bugg, T. D. (2001) *Nat Prod Rep* **18**, 465-493
3.  Marcotte, P., and Walsh, C. (1976) *Biochemistry* **15**, 3070-3076
4.  Soper, T. S., Manning, J. M., Marcotte, P. A., and Walsh, C. T. (1977) *J Biol Chem* **252**, 1571-1575
5.  Abeles, R. H., and Walsh, C. T. (1973) *J Am Chem Soc* **95**, 6124-6125
6.  Johnston, M., Jankowski, D., Marcotte, P., Tanaka, H., Esaki, N., Soda, K., and Walsh, C. (1979) *Biochemistry* **18**, 4690-4701
7.  Johnston, M., Marcotte, P., Donovan, J., and Walsh, C. (1979) *Biochemistry* **18**, 1729-1738
8.  Kraus, J.-L., and Belleau, B. (1975) *Canadian Journal of Chemistry* **53**
9.  Binda, C., Hubalek, F., Li, M., Herzig, Y., Sterling, J., Edmondson, D. E., and Mattevi, A. (2004) *J Med Chem* **47**, 1767-1774
10. Binda, C., Hubalek, F., Li, M., Herzig, Y., Sterling, J., Edmondson, D. E., and Mattevi, A. (2005) *J Med Chem* **48**, 8148-8154
11. Tritsch, D., Mawlawi, H., and Biellmann, J. F. (1993) *Biochim Biophys Acta* **1202**, 77-81
12. White, T. A., and Tanner, J. J. (2005) *Acta Crystallography* **F61**, 737-739
13. Pflugrath, J. (1999) *Acta Crystallography* **D55**, 1718-1725
14. White, T. A., Krishnan, N., Becker, D. F., and Tanner, J. J. (2007) *J Biol Chem*
15. Vagin, A., and Teplyakov, A. (1997) *Journal of Applied Crystallography* **30**, 1022-1025
16. Painter, J., and Merritt, E. A. (2006) *Acta Crystallogr D Biol Crystallogr* **62**, 439-450
17. Kitz, R., and Wilson, I. B. (1962) *J Biol Chem* **237**, 3245-3249
18. Abrahamson, J., Baker, L., Stephenson, J., and Wood, J. (1983) *European Journal of Biochemistry* **134**, 77-82
19. Lennon, B. W., Williams, C. H., Jr., and Ludwig, M. L. (1999) *Protein Sci* **8**, 2366-2379
20. Lee, Y. H., Nadaraia, S., Gu, D., Becker, D. F., and Tanner, J. J. (2003) *Nature Structural Biology* **10**, 109-114
21. Zhang, M., White, T. A., Schuermann, J., Baban, B. A., Becker, D. F., and Tanner, J. J. (2004) *Biochemistry* **43**
22. Silverman, R. B. (1988) *J Enzyme Inhib* **2**, 73-90
23. Silverman, R. B. (1995) *Methods Enzymol* **249**, 240-283
24. Silverman, R. B. (2002) *The Organic Chemistry of Enzyme-Catalyzed Reactions*, Revised Ed., Academic Press
25. Muller, F. (2000) *Chemistry and Biochemistry of Flavoenzymes* (CRC, Ed.), CRC Press
26. Abeles, R. H., and Maycock, A. L. (1976) *Accounts of Chemical Research* **9**, 313-319
27. Maycock, A. L., Abeles, R. H., Salach, J. I., and Singer, T. P. (1976) *Biochemistry* **15**, 114-125

28. Williams, R. F., and Bruice, T. C. (1976) *J Am Chem Soc* **98**, 7752-7767
29. Soltero-Higgin, M., Carlson, E. E., Gruber, T. D., and Kiessling, L. L. (2004) *Nat Struct Mol Biol* **11**, 539-543
30. Binda, C., Li, M., Hubalek, F., Restelli, N., Edmondson, D. E., and Mattevi, A. (2003) *Proc Natl Acad Sci U S A* **100**, 9750-9755
31. Zhang, W., Zhang, M., Zhu, W., Zhou, Y., Wanduragala, S., Rewinkel, D., Tanner, J. J., and Becker, D. F. (2007) *Biochemistry* **46**, 483-491
32. Ghisla, S., Hartmann, U., Hemmerich, P., and Muller, F. (1973) *Justus Liebegs Ann. Chem*, 1338
33. Abeles, R. H., and Tashjian, A. H., Jr. (1974) *Biochem Pharmacol* **23**, 2205-2207
34. Sharma, S. S., and Dietz, K. J. (2006) *J Exp Bot* **57**, 711-726
35. Aspinall, D., and Paleg, L. G. (1981) in *The physiology and biochemistry of drought resistance in plants*, pp. 205-240, Australia
36. Ashraf, M., and Harris, P. J. C. (2003) *Plant Science* **166**, 3-16
37. Naidu, B. P., Paleg, L. G., Aspinall, D., Jennings, A. C., and Jones, G. P. (1991) *Phytochemistry* **30**, 407-409
38. Schat, H., Sharma, S. S., and Vooijs, R. (1997) *Physiologia Plantarum* **101**, 477-482
39. Chen, C., and Dickman, M. B. (2005) *Proc Natl Acad Sci U S A* **102**, 3459-3464
40. Yancey, P. H. (2005) *J Exp Biol* **208**, 2819-2830
41. Rudolph, A. S., and Crowe, J. H. (1985) *Cryobiology* **22**, 367-377
42. Smirnoff, N., and Cumbes, Q. (1989) *Phytochemistry* **29**, 1057-1060
43. Rutgsi, S., Joshi, A., Moss, H., and Riesz, P. (1977) *International Journal of Radiation Biology and Related Studies of Physics, Chemistry and Medicine* **31**, 415-440
44. Floyd, R. A., and Nagy, Z. S. (1984) *Biochemica et Biophysica Acta* **790**, 94-97
45. Alia, Mohanty, P., and Matysik, J. (2001) *Amino Acids* **21**, 195-200
46. Bursell, E. (1963) *Bull World Health Organ* **28**, 703-709
47. Vickerman, K., Tetley, L., Hendry, K. A. K., and Turner, M. R. (1988) *Biology of the Cell* **64**, 109-119
48. Donald, S. P., Sun, X. Y., Hu, C. A., Yu, J., Mei, J. M., Valle, D., and Phang, J. M. (2001) *Cancer Res* **61**, 1810-1815
49. Maxwell, S. A., and Rivera, A. (2003) *J Biol Chem* **278**, 9784-9789
50. Rivera, A., and Maxwell, S. A. (2005) *J Biol Chem* **280**, 29346-29354
51. Maxwell, S. A., and Davis, G. E. (2000) *PNAS* **97**, 13009-13014
52. Liu, Y., Borchert, G. L., Surazynski, A., Hu, C. A., and Phang, J. M. (2006) *Oncogene* **25**, 5640-5647

FIGURE 3.1. Kinetic scheme for mechanism-based enzyme inactivation.

$$E + I \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E{*}I \xrightarrow{k_2} E{*}I' \xrightarrow{k_4} E{-}I''$$

$$\downarrow k_3$$

$$E + P$$
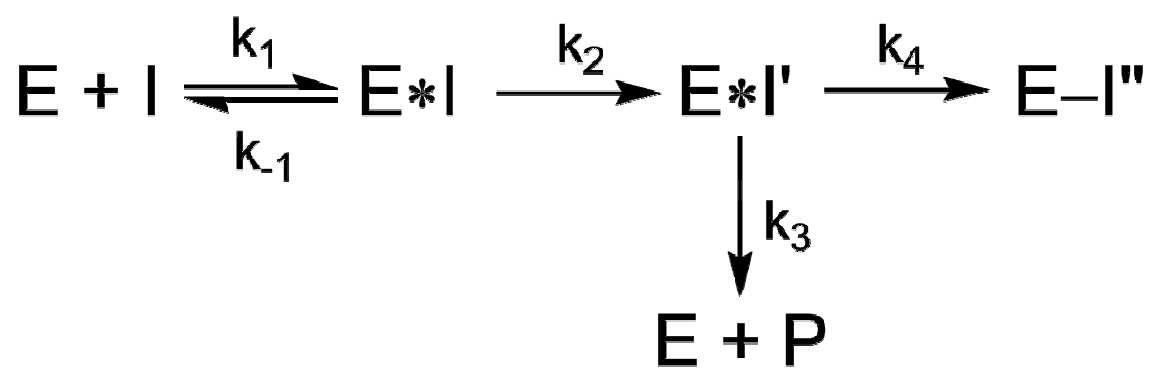
FIGURE 3.2. Proposed action of PRODH mechanism-based inactivator, 4-methylene-L-proline (adapted from Tritsch, 1993).
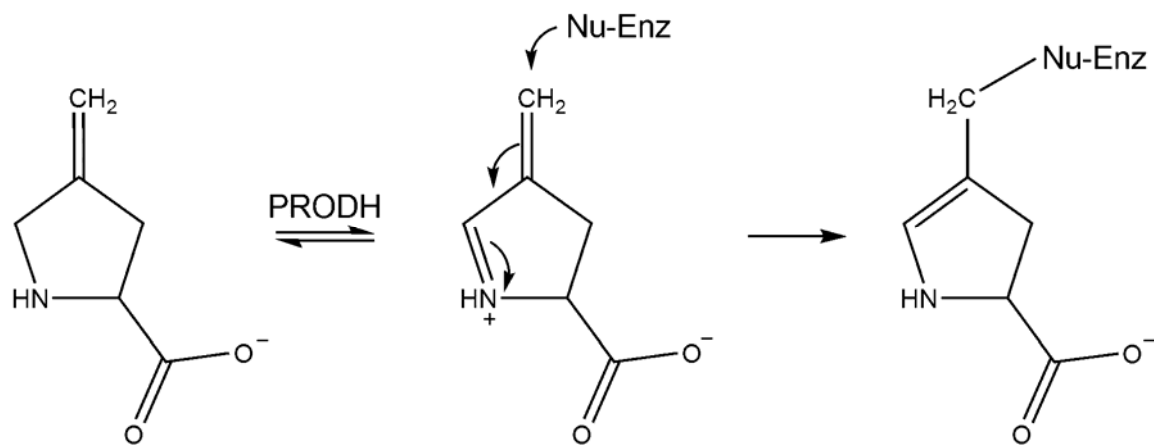
FIGURE 3.3. Crystals of *Thermus thermophilus* proline dehydrogenase inactivated by *N*-propargylglycine.

FIGURE 3.4. *Thermus thermophilus* proline dehydrogenase monomer inactivated with *N*-propargylglycine. The $(\beta\alpha)_8$ barrel is shown in cartoon rainbow coloring, covalently-modified flavin adenine dinucleotide is shown in white CPK sticks.



Lys99

FIGURE 3.5. Superposition of oxidized TtPRODH (white) and *N*-propargylglycine inactivated TtPRODH (rainbow). Oxidized FAD is shown in white sticks and inactivated TtPRODH flavocyanine shown in green sticks.
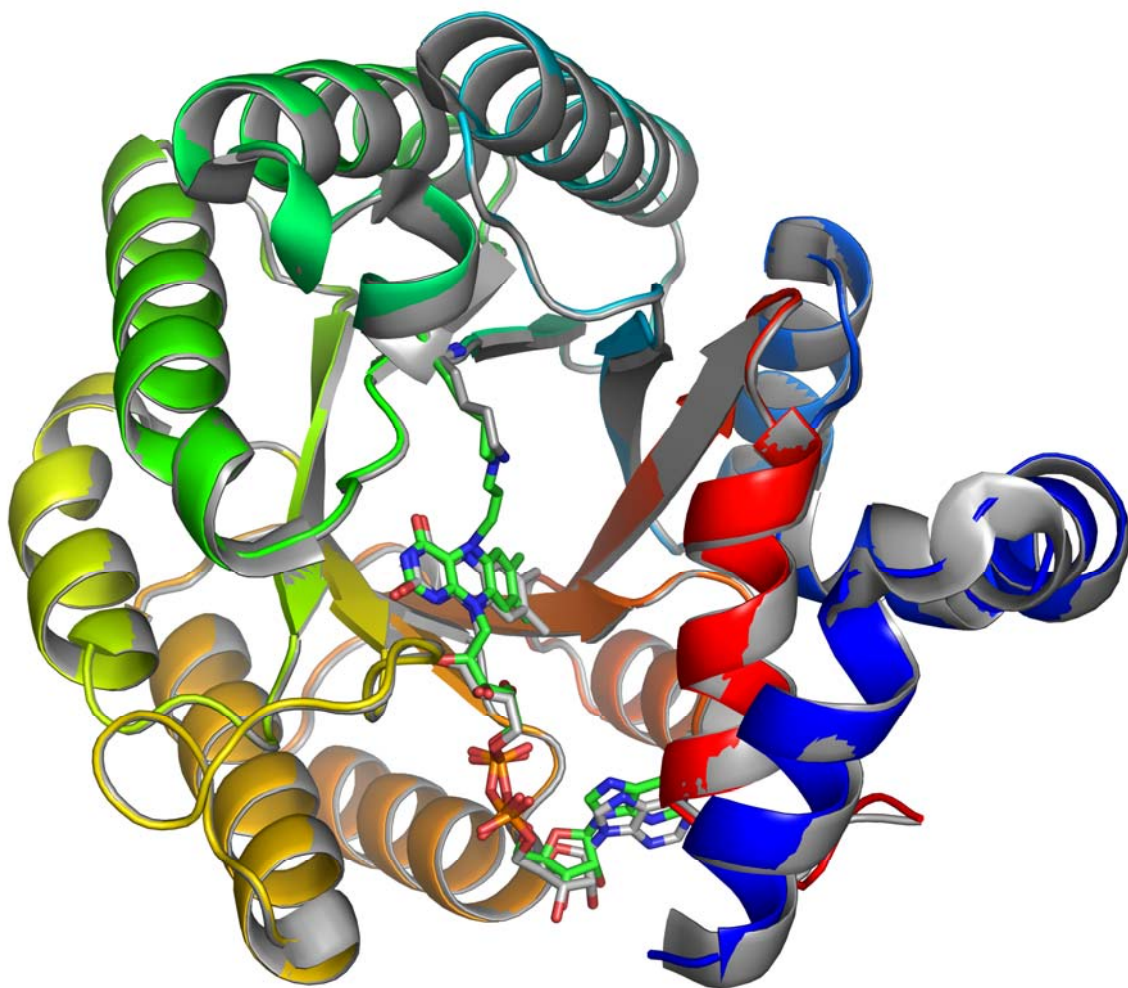
FIGURE 3.6. Two views of the modified flavin of inactivated TtPRODH (chain A) covered by an electron density map (2.5 σ). The map is a simulated annealing $F_o - F_c$ omit map using. The modified lysine is shown in cyan sticks and the flavin shown in yellow sticks.
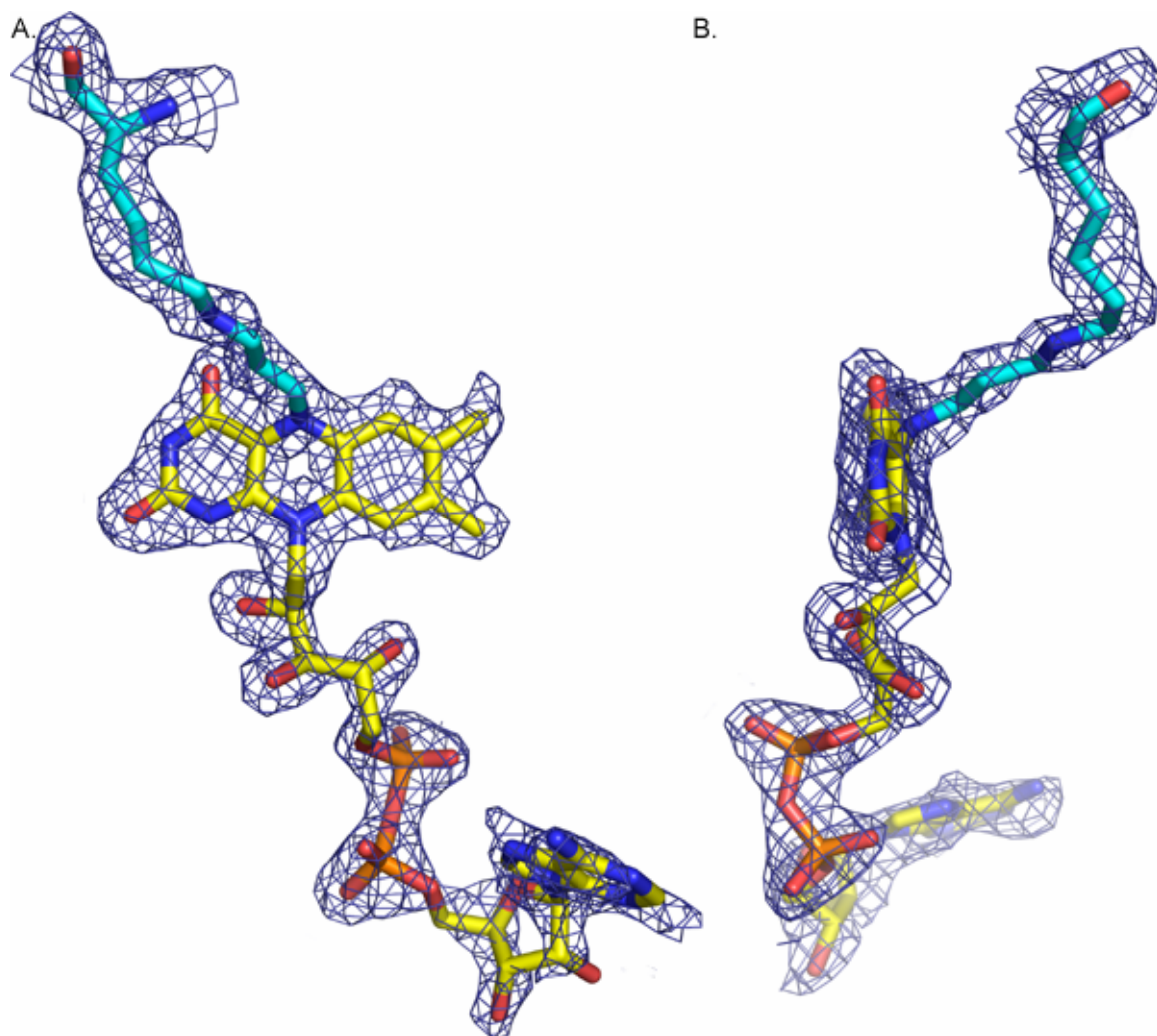
FIGURE 3.7.  Side view of the *N*-propargylglycine-inactivated flavin of TtPRODH. FAD is shown in yellow, Lys99 in green.
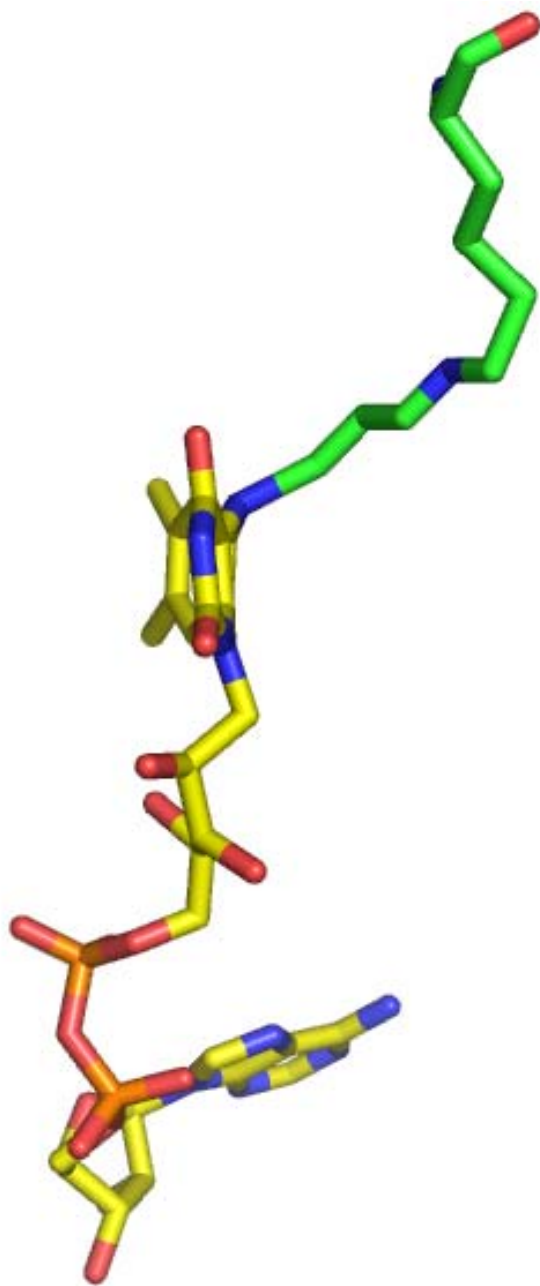
FIGURE 3.8. Views of the FAD isoalloxazine from TtPRODHs. A) isoalloxazine of *N*-propargylglycine inactivated TtPRODH B) isoalloxazine of inactivated TtPRODH with electron density from a simulated annealing $F_o - F_c$ omit map (2.5 $\sigma$ ). The isoalloxazine is oriented so that the plane of the dimethyl benzene ring is perpendicular to the page, which allows one to see the twisting of the isoalloxazine. C) Same as in (B) but the isoalloxazine is oriented so that the plane of the pyrimidine ring is perpendicular to the page. D) Superposition of isoalloxazines from oxidized (magenta) and inactivated (yellow) TtPRODHs.

**A.**



**B.**



**C.**



**D.**
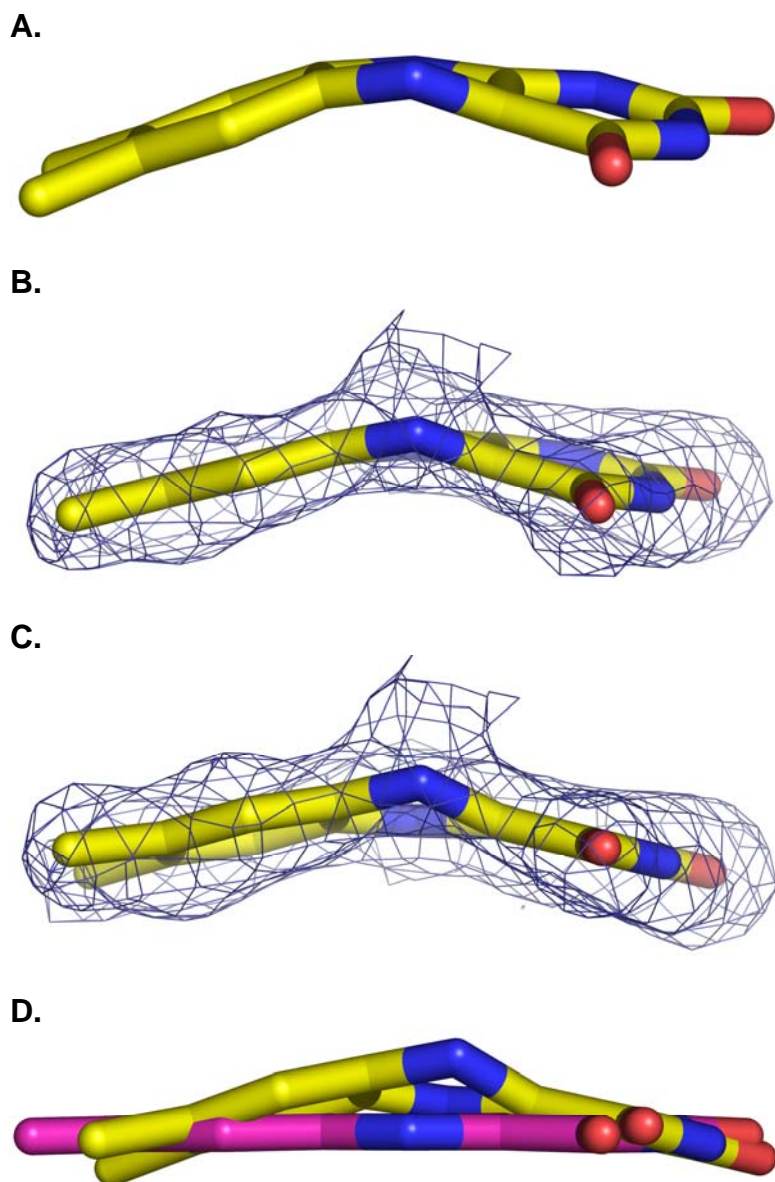
FIGURE 3.9. Four views of *N*-propargylglycine covalently-modified FAD of TtPRODH.



A.    B.    C.    D.
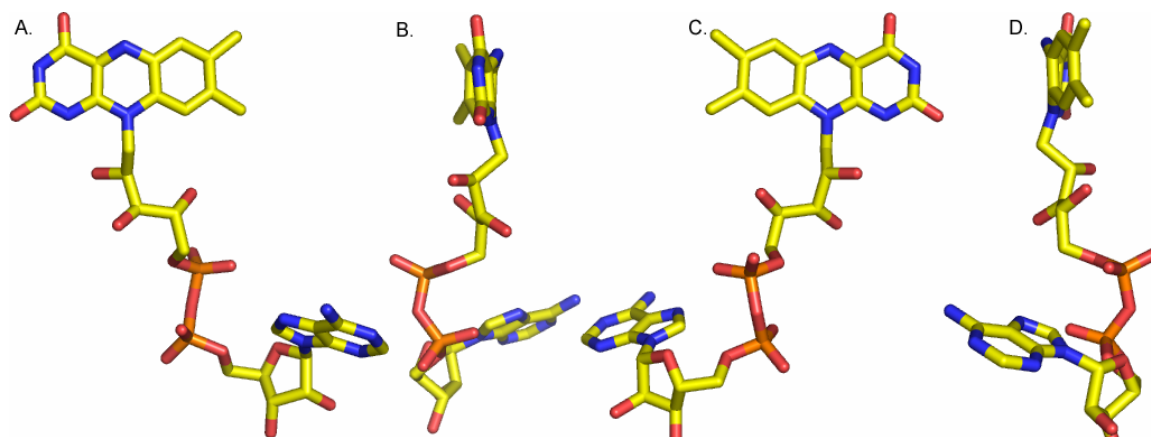
FIGURE 3.10. Electrospray Ionization – time of flight mass spectrometry of TtPRODH. The peak at 37976 corresponds to the molecular weight of the apo enzyme.
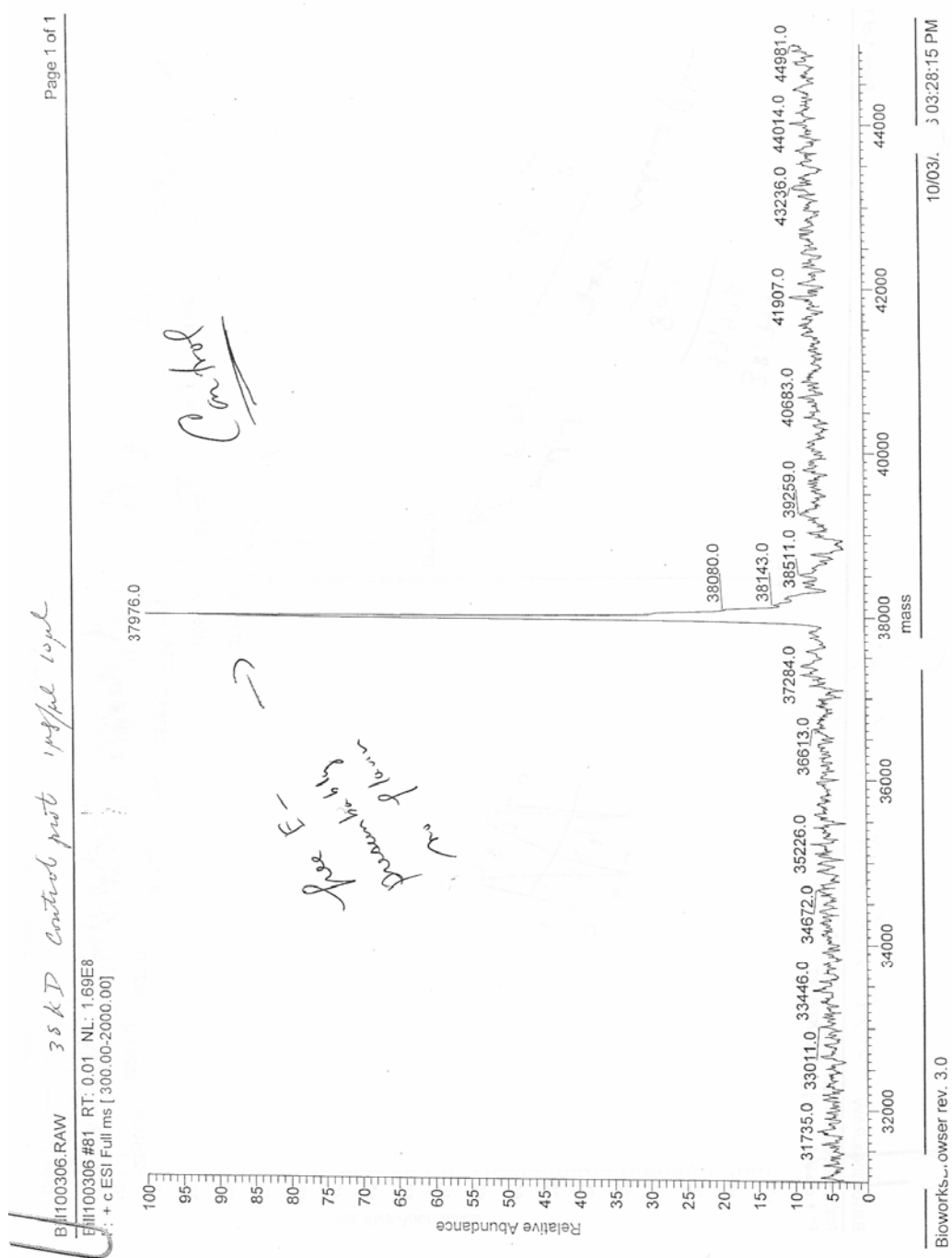


171

FIGURE 3.11. Electrospray ionization – time of flight mass spectrometry of *N*-propargylglycine-inactivated TtPRODH. The peak at 37976 Da corresponds to the apo enzyme and the peak at 38797 Da corresponds to the inactivated enzyme.



172

FIGURE 3.12. Active site changes beyond the flavocyanine adduct. Arg184 is shown in orange, Asp133 is shown in green and Tyr275 shown in cyan. Hydrogen bonds are shown by black dotted lines. Flavocyanine shown in white CPK sticks. Strands are shown in light pink cartoons and helices are shown in light cyan cartoons
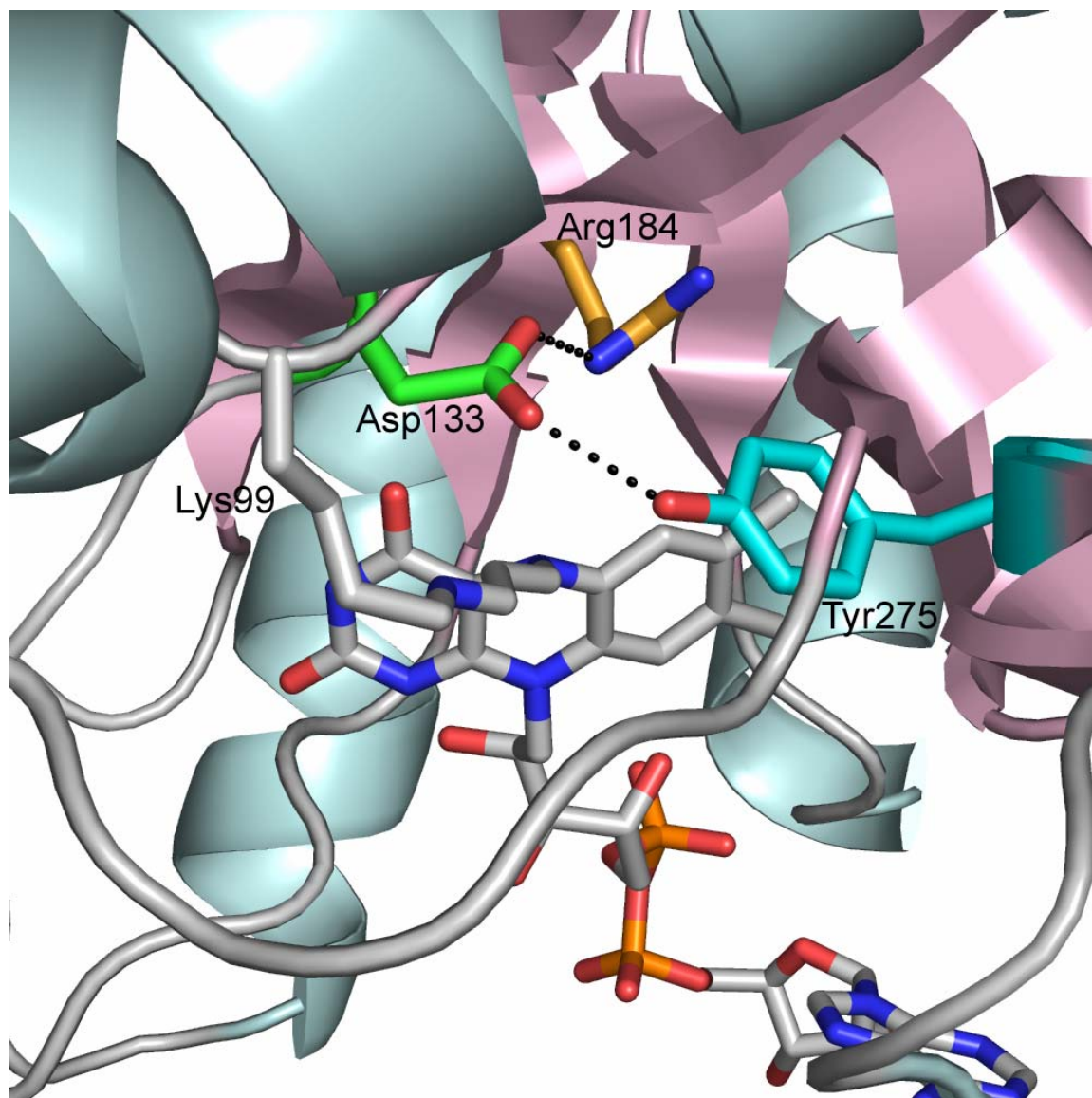
FIGURE 3.13. Residual active site electron density. The map is a simulated annealing $F_o - F_c$ omit map (2.5 σ) in which the modeled glycine molecule was omitted prior to simulated annealing. Note that the density is quite weak and suggests low occupancy for glycine. Glycine was omitted from the final model.
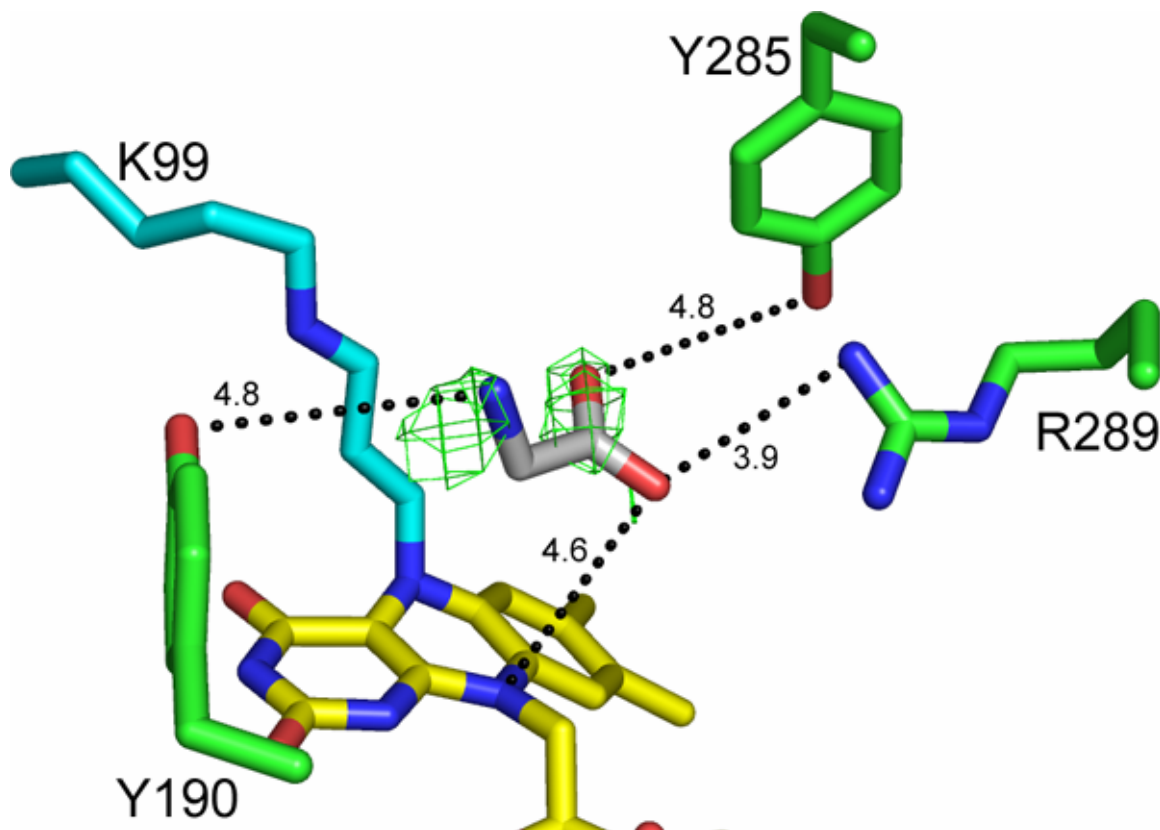
FIGURE 3.14. Spectral changes caused by inactivation of TtPRODH by *N*-propargylglycine.  The flavin spectrum of oxidized TtPRODH is shown in red  and displays characteristic peaks at 452 and 382 nm.  The other curves correspond to various time points up to 60 minutes.   Note that the peak at 452 nm disappears and a new peak centered at 386 nm appears as time advances.
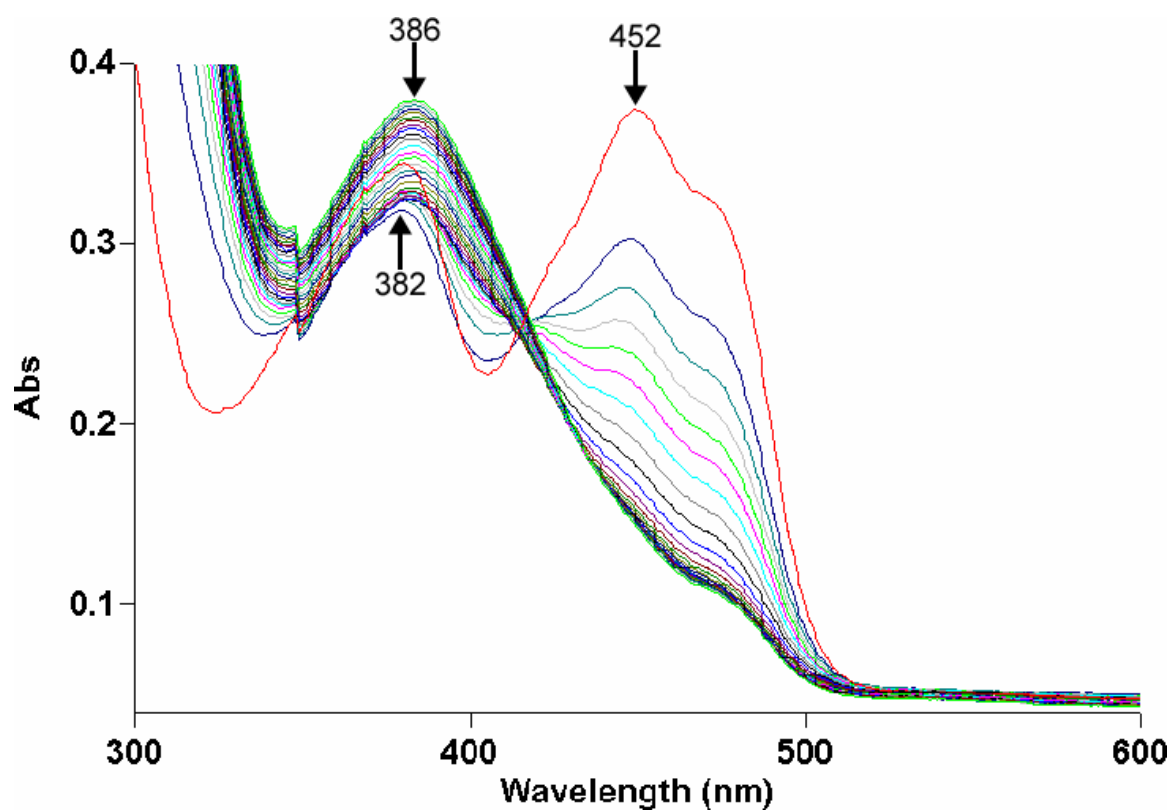
FIGURE 3.15.   Time course of inactivation of TtPRODH by *N*-propargylglycine monitoring 452 nm, 382 nm and enzyme activity.
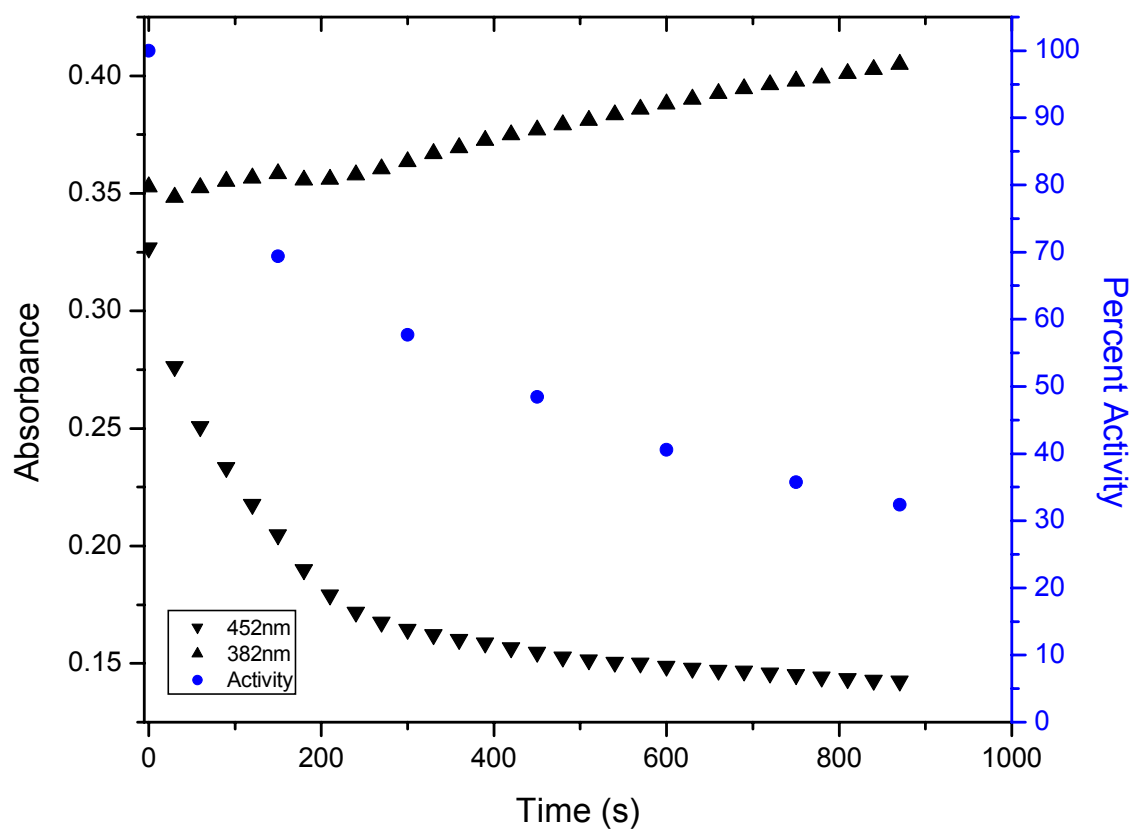
FIGURE 3.16. Effect of adding sodium dithionite to a sample of TtPRODH that had been inactivated by *N*-propargylglycine. Spectra are labeled as: oxidized enzyme (Eox), dithionite reduced enzyme (Ered), *N*-propargylglycine inactivated (EI''), and dithionite-reduced inactivated (EI''red).
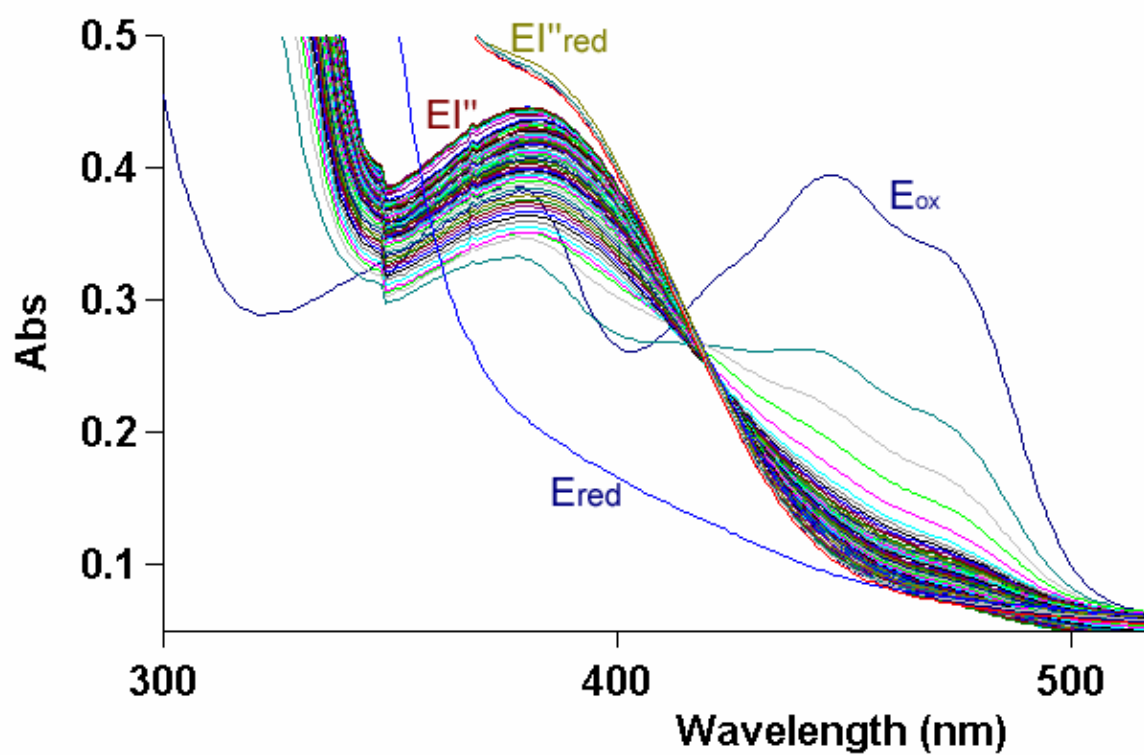
FIGURE 3.17. Theoretical Kitz and Wilson plot (adapted from Silverman, 2002).
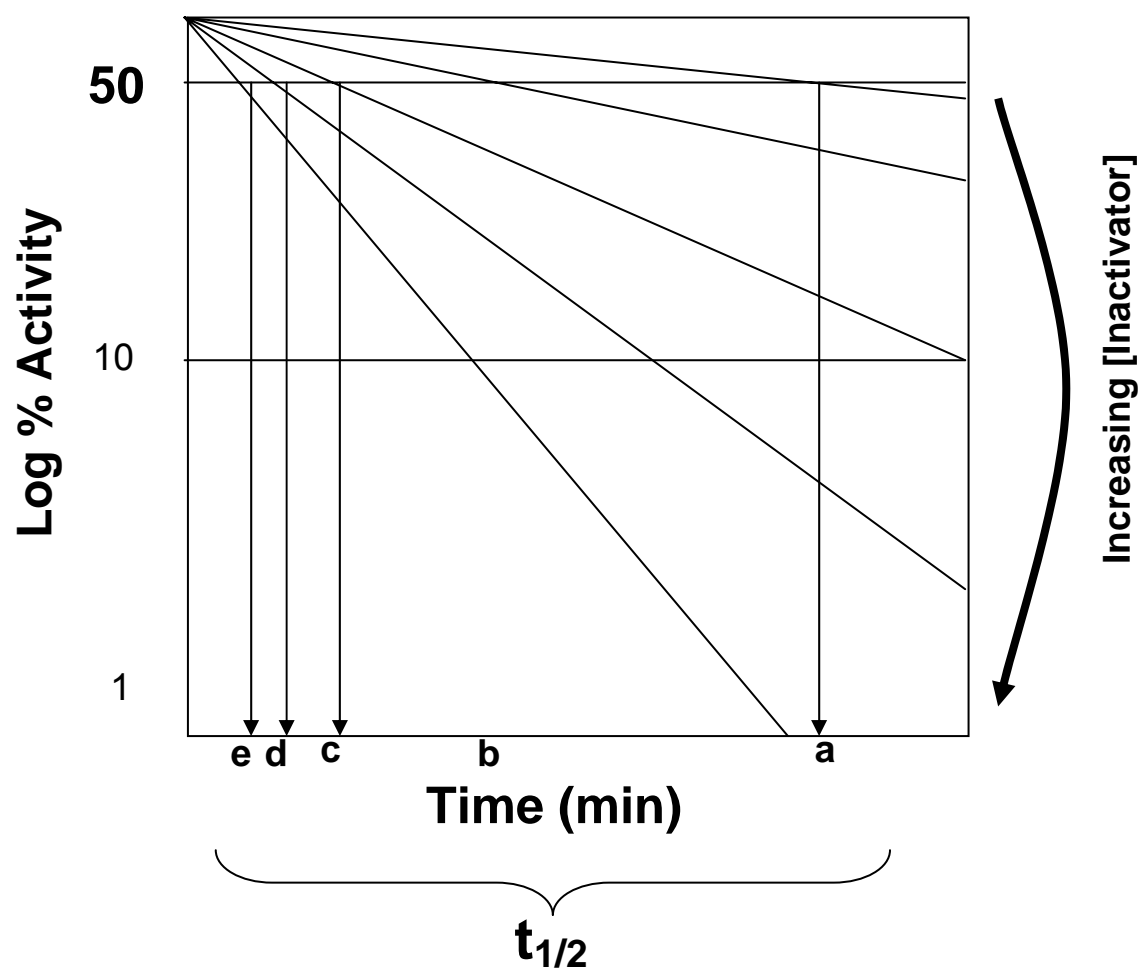
FIGURE 3.18. Theoretical Kitz and Wilson replot (adapted from Silverman, 2002).
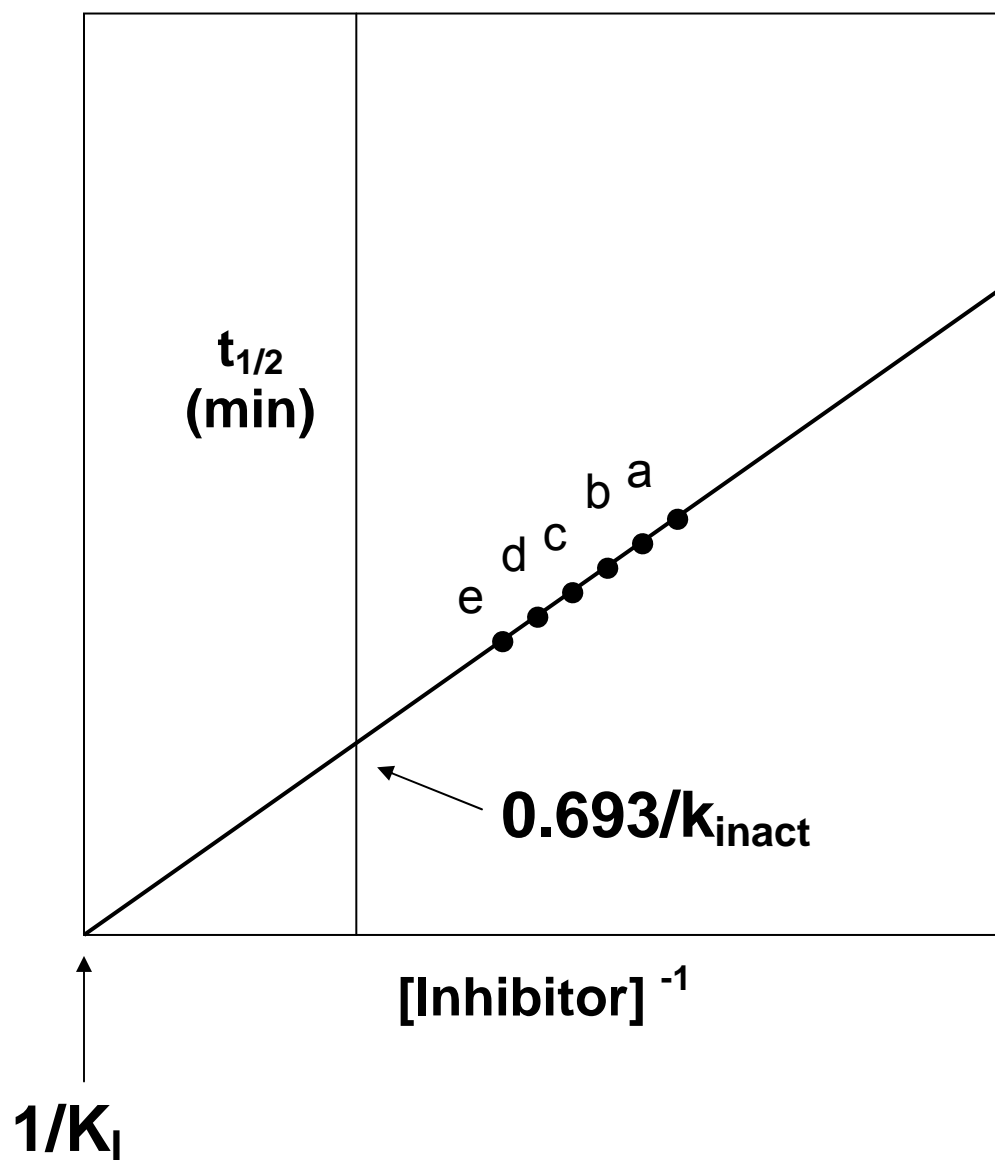
FIGURE 3.19. Kitz and Wilson plot (top) and replot (bottom) for TtPRODH inactivation by *N*-propargylglyine at 25 $^{\circ}$C. The legend lists *N*-propargylglyine concentrations.
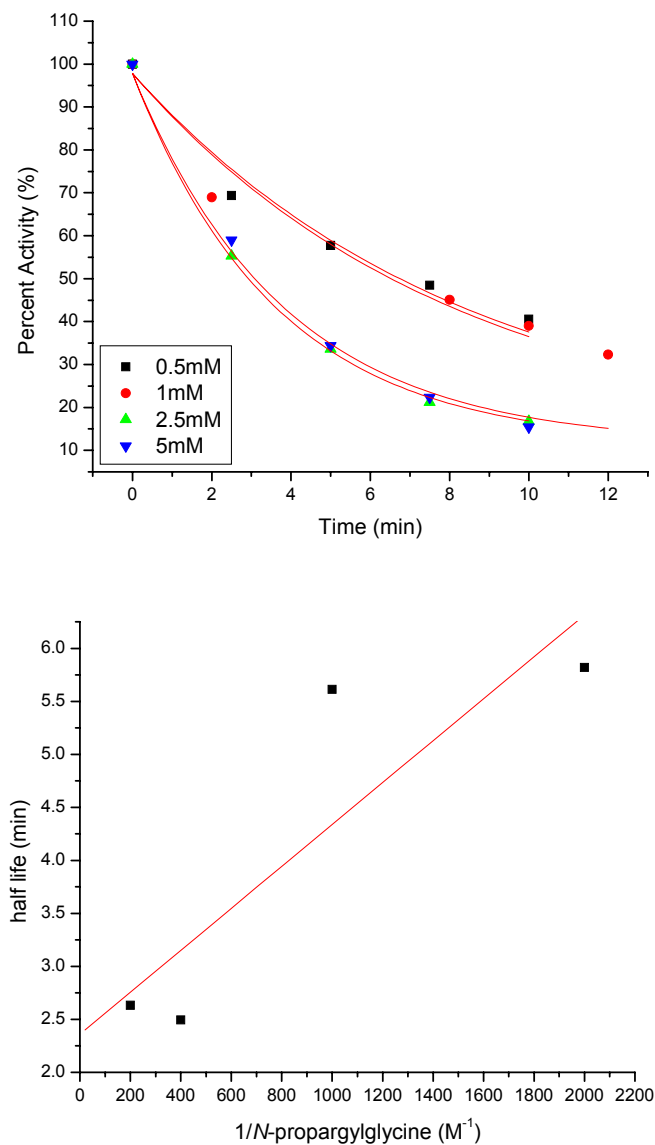
FIGURE 3.20.   Kitz and Wilson plot (top) and replot (bottom) of TtPRODH inactivation with *N*-propargylglycine at 4 °C.
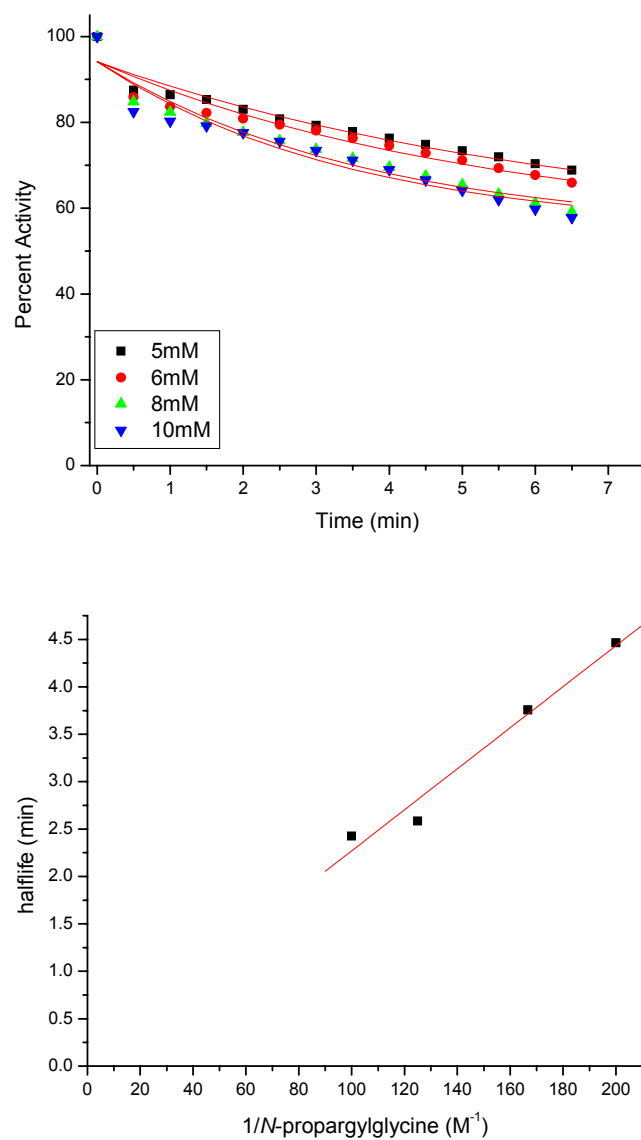
FIGURE 3.21.  Possible flavin adducts formed by mechanism-based inactivation. The sites of modification include N5 (A), C4a (B), both N5 and C4a (C).   (D)  The modified flavin observed by Maycock and colleagues. (E) The modified flavin observed in TtPRODH inactivated by *N*-propargylglycine.
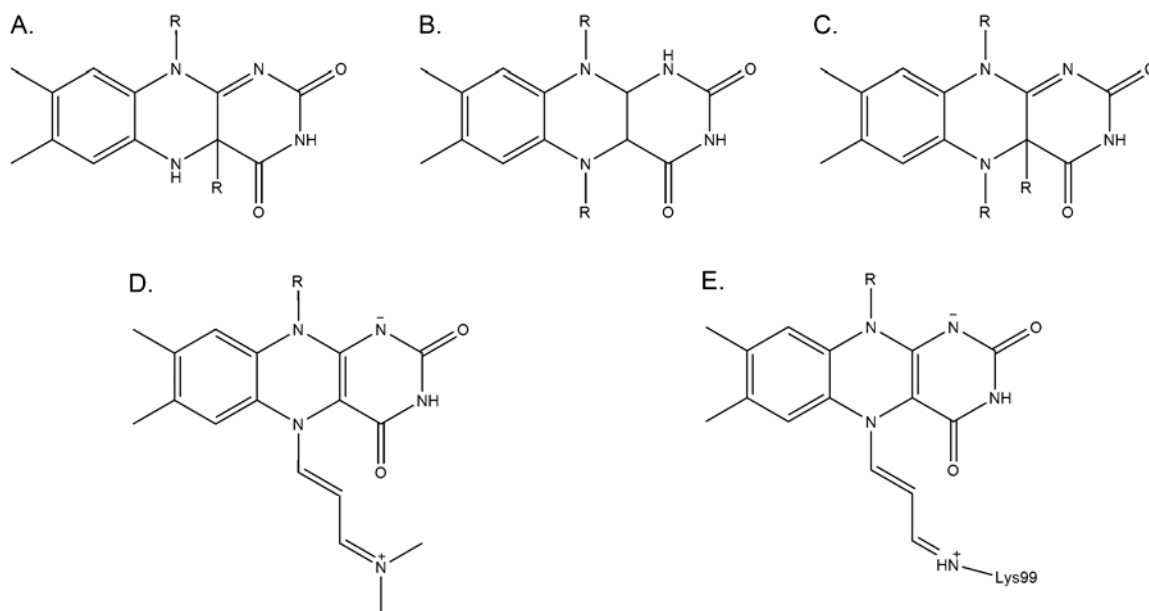
FIGURE 3.22. Maycock's proposed reaction mechanisms for flavin adduction by mechanism-based inactivators with an acetylenic group. A) Theoretical acetylenic mechanism based inactivator. B) Carbanion mechanism. C) Radical pair complex. D) Michael addition (note the isoalloxazine has truncated in these drawings).
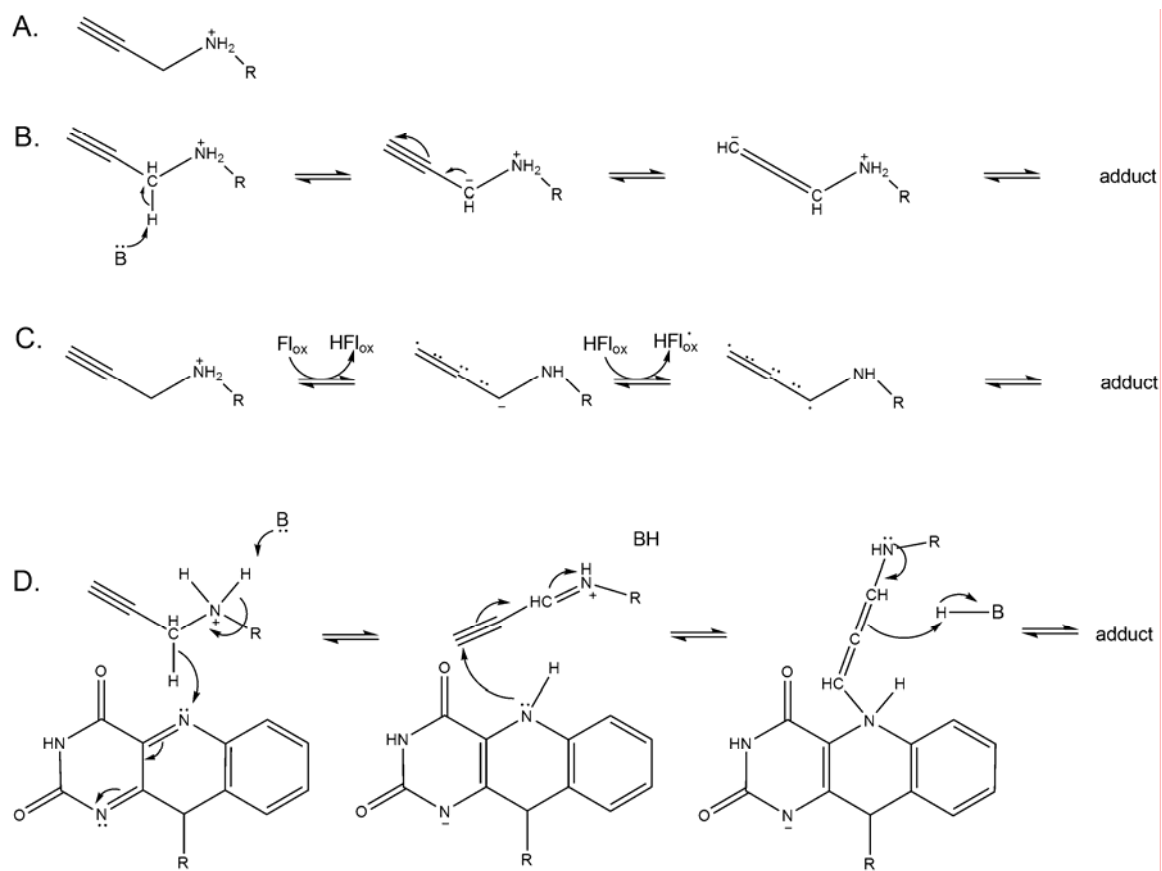
FIGURE 3.23. Structure of monoamine oxidase inactivated with rasagiline. A) Monoamine oxidase B shown in rainbow cartoon with black arrow head indicating the flavocyanine adduct (PDBID 1S2Q). B) Monoamine oxidase B flavocyanine adduct with FAD in yellow and rasagiline adduct in cyan. C) Monoamine oxidase A shown in rainbow cartoon with black arrow head indicating the flavocyanine adduct (PDBID 2BK4). D) Monoamine oxidase A flavocyanine adduct with FAD in yellow and rasagiline adduct in cyan.
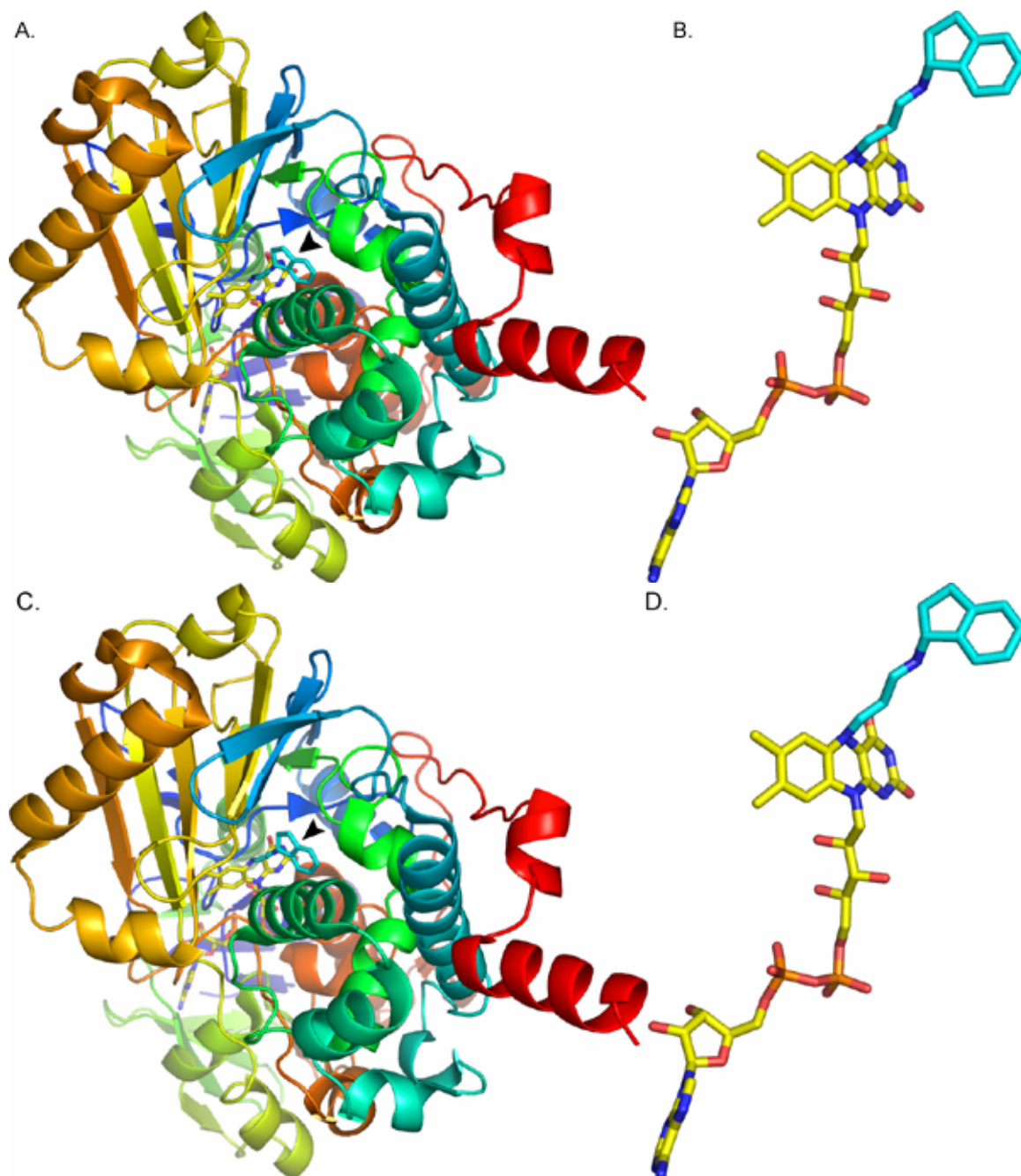
FIGURE 3.24. Steps 1 and 2 of the proposed mechanism of inactivation of TtPRODH by *N*-propargylglycine.
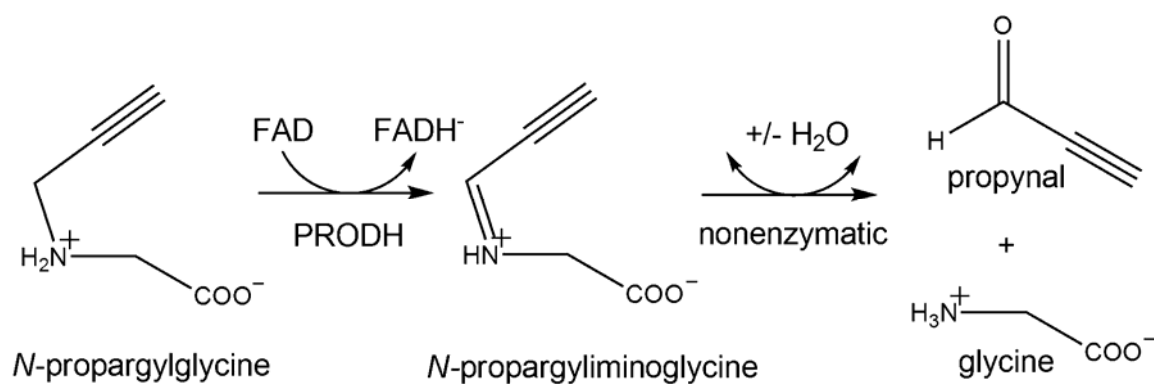
FIGURE 3.25. Proposed Schiff base formation with propynal and K99. There is a strategically placed active site $H_2O$ to form the carbinolamine intermediate.
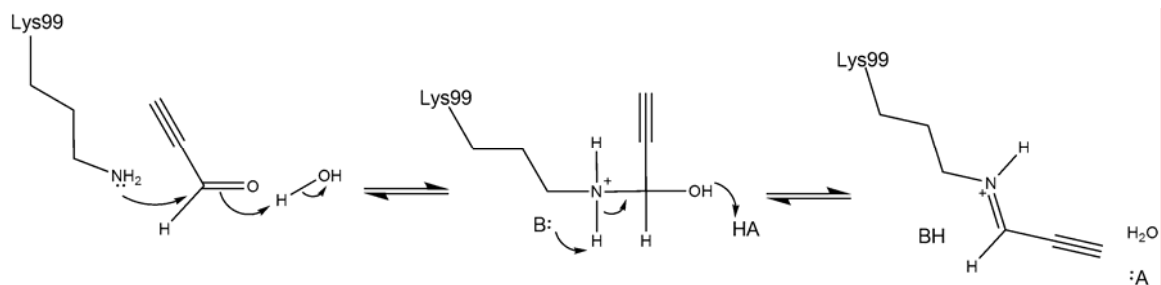
FIGURE 3.26. Last step of the proposed mechanism of inactivation of TtPRODH by *N*-propargylglycine. The reaction is a Michael addition of reduced flavin to propynal-modified lysine 99.
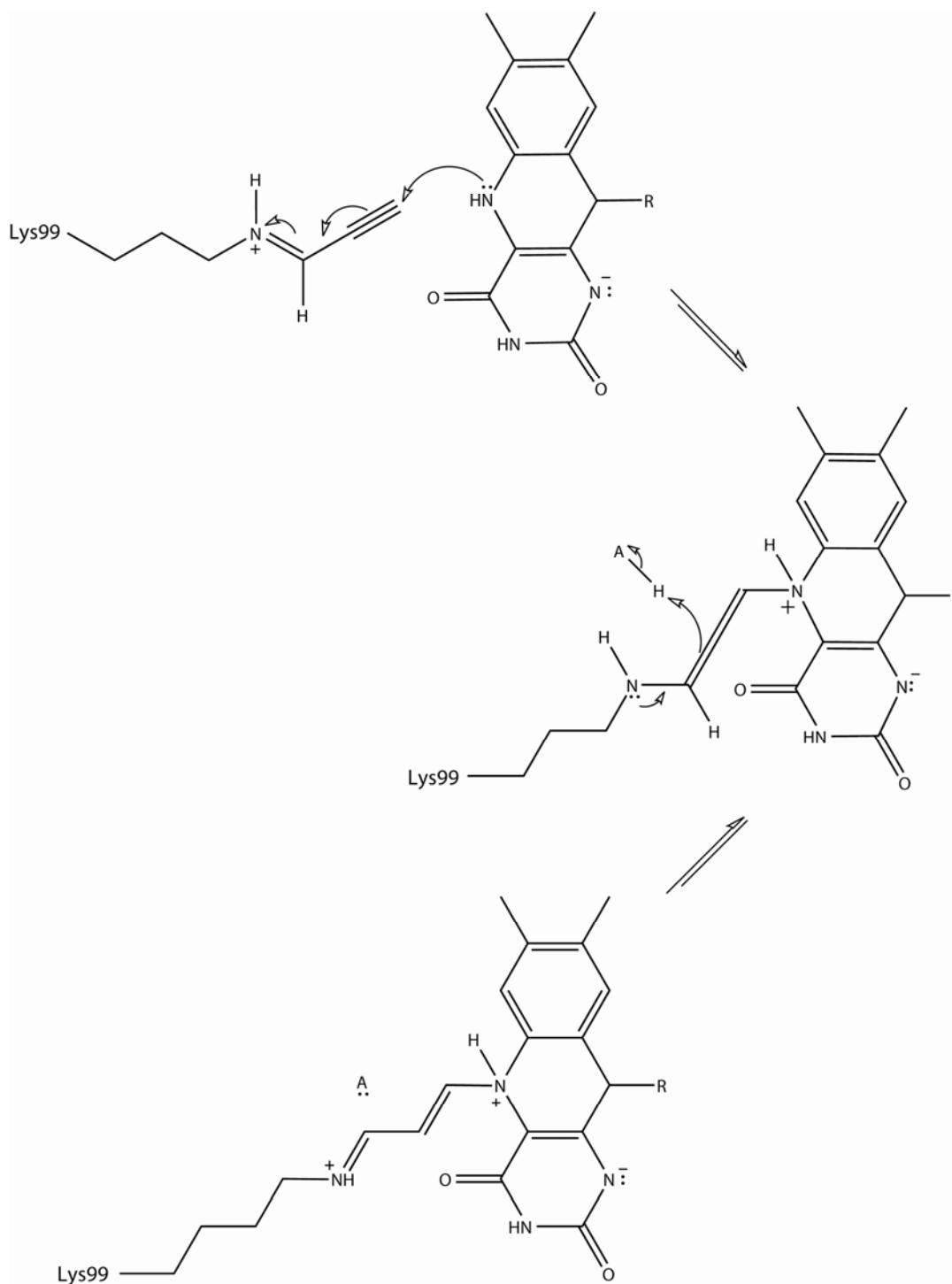
FIGURE 3.27. Generic Schiff base reaction between an amine and an aldehyde or ketone to form the carbinolamine intermediate and finally the imine and water.
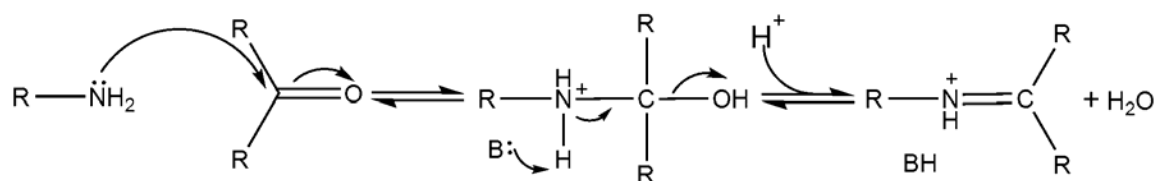
FIGURE 3.28. Comparison of isoalloxazines from inactivated TtPRODH (yellow) and dithionite-reduced *E. coli* PutA PRODH domain (green).
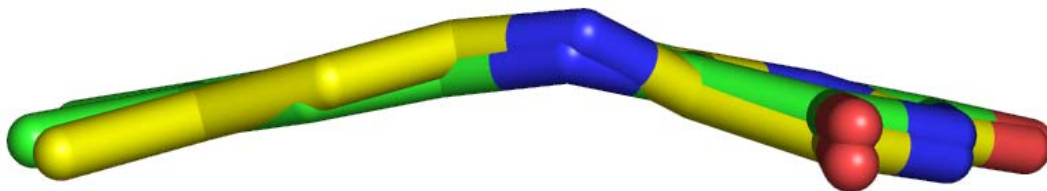
FIGURE 3.29. Comparison of FAD from inactivated TtPRODH (yellow) and dithionite-reduced *E. coli* PutA PRODH domain (green).
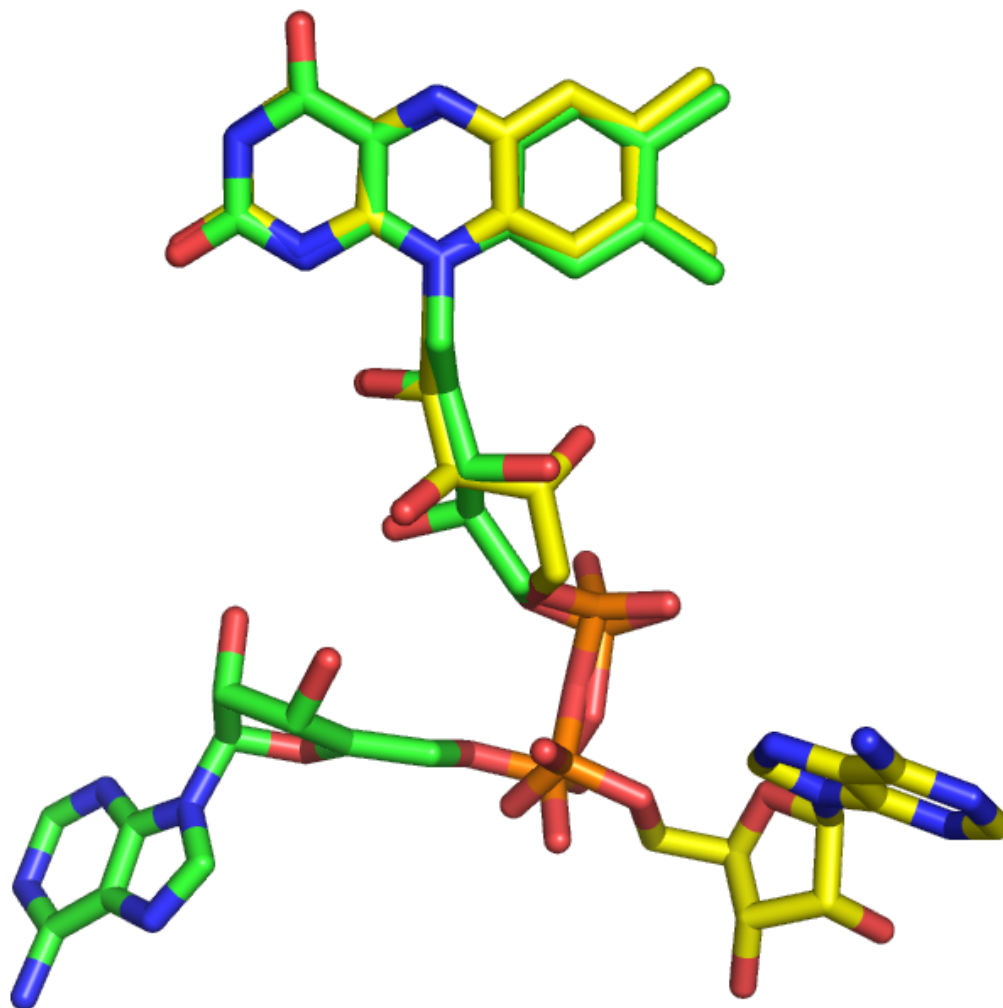
TABLE 3.1.  Data Collection and Refinement Statistics for TtPRODH inactivated by *N*-propargylglycine[a]

| | |
|---|---|
| Wavelength (Å) | 1.00 |
| Diffraction resolution (Å) | 1.9 |
| No. of observations | 325777 |
| No. of unique reflections | 55959 |
| Redundancy | 5.82 (5.78) |
| Completeness (%) | 99.3 (100) |
| $R_{merge}$ | 0.059 (0.356) |
| Average I/ | 12.5 (3.3) |
| Wilson B-factor ($Å^2$) | 33 |
| No. of protein chains | 2 |
| No. of protein residues | 589 |
| No. of protein atoms | 5221 |
| No. of water molecules | 462 |
| $R_{cryst}$ | 0.193 (0.238) |
| $R_{free}$[b] | 0.220 (0.299) |
| $RMSD^{c}$ | |
| Bond lengths (Å) | 0.012 |
| Bond angles (deg.) | 1.296 |
| Ramachandran plot[d] | |
| Favored (%) | 95 |
| Allowed (%) | 5 |
| Average B-factors ($Å^2$) | |
| Protein | 33.3 |
| FAD | 36.4 |
| Modified Lysine | 36.6 |
| MPD | 50.0 |
| Water | 38.1 |

[a]Values for the outer resolution shell of data are given in parenthesis.

[b]5% random $R_{free}$ test set.

[c]Compared to the Engh and Huber force field (*51*).

[d]The Ramachandran plot was generated with PROCHECK

# CHAPTER 4

## Preliminary Studies of Interactions Between *T. thermophilus* Proline Dehydrogenase and $\Delta^1$-Pyrroline-5-Carboxylate Dehydrogenase

### INTRODUCTION

Channeling is the direct transfer of intermediate between the active sites of enzymes that catalyze sequential reactions (1). This can occur in multifunctional enzymes, tightly-associated multienzyme complexes or transiently associating enzymes. Channeling of intermediates of many metabolic pathways has been reported, including enzymes involved in nucleotide biosynthesis, amino acid metabolism, lipid metabolism, glycolysis, TCA cycle, DNA replication, RNA synthesis and protein biosynthesis (2). Benefits of channeling intermediates include decreased transit time, prevention of intermediate entry into competing metabolic pathways, and protection of reactive, toxic or labile intermediate (1).

Tryptophan synthase (TrpS) is perhaps the best characterized system for intermediate channeling (3). TrpS catalyzes the last two steps of tryptophan biosythesis (3). In bacteria and higher plants, the enzymes form a tightly associated heterotetramer $(\alpha\beta)_2$ enzyme complex, whereas in mold and yeast, TrpS is a bifunctional enzyme with fused $\alpha$ and $\beta$ domains (2). The $\alpha$-subunit catalyzes breakdown of indole 3-glycerol phosphate to form indole and D-glyceraldehyde 3-phosphate. The $\beta$-subunit utilizes cofactor PLP, the indole formed from the $\alpha$-subunit, and serine to catalyze the formation of tryptophan (2). The intermediate indole is channeled from the $\alpha$-subunit to the $\beta$-subunit. The

crystal structure of *S.typhimurium* TrpS revealed a 25 Å tunnel leading from the α-subunit to the β-subunit (3) which can hold up to 4 indole molecules (2). Kinetic studies confirmed channeling for TrpS. Steady state kinetics failed to trap the indole intermediate indicating that it was not released from the protein. And introducing a bulky aromatic residue in the proposed channel by mutagenesis made catalysis less efficient (3).

Yeast TrpS is an example of what Eisenberg and colleagues refer to as a Rosetta Stone protein because it deciphers protein-protein interactions (4). For TrpS, the Rosetta Stone hypothesis predicts that monofunctional α and β enzymes interact, which, of course, is true. The Rosetta Stone hypothesis makes a similar prediction about monofunctional proline catabolic enzymes proline dehydrogenase (PRODH) and $\Delta^1$-pyrroline-5-carboxylate dehydrogenase (P5CDH) based on the observation that these enzymes are fused in some organisms into the bifunctional enzyme Proline utilization A (PutA).

Other considerations also argue in favor of interactions between monofunctional PRODH and P5CDH as well as channeling between the two enzyme active sites. PRODH and P5CDH catalyze the oxidation of L-proline to L-glutamate via the intermediates pyrroline-5-carboxylate (P5C) and glutamic semialdehyde (GSA). Evidence is mounting for the toxicity of the proline catabolic intermediate P5C/GSA. Yeast deficient in P5CDH (Δ*put2*) have been reported to generate reactive oxygen species (ROS) upon proline treatment (5,6). P5C (1 mM) treatment of mutant *A.thaliana* having hypersensitivity to proline results in visible damage within 9 hours and death within 3 days, whereas

treatment of hypersensitive mutant plants with L-proline (100 mM) weakly affected the whole plant after 3 days (6). Upon treatment with exogenous proline, P5CDH-deficient *A.thaliana* display sensitivity to proline resulting in apoptosis, callous deposition, and ROS generation (7).

Besides toxicity issues, P5C and GSA are reactive intermediates. The aldehyde moiety of glutamic semialdehyde is especially prone to nucleophilic attack. Other competing metabolic pathways share P5C and GSA as intermediates. For example, GSA enters the urea cycle by ornithine aminotransferase, and P5C can be converted back to proline by P5C reductase. The toxicity, reactivity, and metabolic utility of P5C and GSA make them ideal candidates for channeling.

Another argument in favor of interactions between PRODH and P5CDH is that kinetic data suggest that channeling occurs in PutAs. Maloy's group reported data showing that *S.typhimurium* PutA channels the intermediate P5C/GSA (8). Upon examination of substrate preference by measuring NADH formation as a function of P5C concentration, they found that PutA had a greater preference (14-fold) for P5C from PRODH catalysis rather than exogenously supplied P5C. When measuring NADH formation as a function of time, P5CDH oxidized endogenous P5C more efficiently (8-fold) then compared to exogenous P5C. Also, they found that, in the absence of $NAD^+$, PRODH was active with concomitant release of P5C and that exogenous P5C can be oxidized by P5CDH. Based on these data, Maloy and coworkers termed this mechanism of intermediate transport a "leaky" channel (8). More recently, kinetic data in

support of channeling of has been obtained by Prof. Donald Becker's group for PutAs from *B. japonicum* and *E. coli* (unpublished data).

Based on the Rosetta stone hypothesis, the reactivity of P5C/GSA, and the observation of channeling in PutAs, we have begun to examine the possibility that monofunctional PRODH and P5CDH from *Thermus thermophilus* interact and engage in intermolecular substrate channeling.

## MATERIALS AND METHODS

### Cloning

The genes encoding PRODH and P5CDH from *T. thermophilus* HB27 were cloned from genomic DNA by the University of Missouri DNA core into plasmid pKA8H. The resulting plasmids encode proteins having N-terminal 8x His tags that are cleavable with TEV protease. TtPRODH was cloned as previously described (9) and TtP5CDH was cloned into pKA8H using the following primers:

Forward 5' – CCTTGATCATATGACGGTGGAACCTTTCCGGAACG – 3'

Reverse 5' – TTTGGATCCCTAGAAGCGCTCGGCCACCGCC – 3'

Correct sequences were confirmed using T7 forward and reverse primers.

Coexpression vector pET-Duet1 containing two multi-cloning sites (MCS1, MCS2) was purchased from Novagen. Various pET-Duet1 constructs containing one or both genes were created using PCR primers listed in Table 4.1. The genes were subcloned from pKA8H. First, two plasmids corresponding to pET-Duet1 with either the TtPRODH or TtP5CDH gene in MCS2 were created. NdeI

and EcoRV restriction sites were used for these two constructs.  Presence of the TtPRODH and TtP5CDH genes in MCS2 were confirmed with sequencing.  Next the P5CDH gene with DNA encoding a TEV-cleavable 8x-histag was subcloned into MCS1 of the pET-Duet1 vector containing the TtPRODH gene in MCS2. XbaI and EcoRI restriction sites were used.  Presence of theTtP5CDH gene in MCS1 was confirmed by sequencing.

**Expression and Purification**

TtPRODH and TtP5CDH were separately expressed from the pKA8H vector using *E. coli* strain, BL21(DE3)pLysS.  An 10 mL overnight culture was used to inoculate 1.5 L of LB broth.  When the $OD_{600}$ = 0.6, protein was induced using 0.5 mM IPTG for 3 hours at 37 $^{o}$C.  Cells were harvested and frozen in 10 mL 50 mM $NaH_2PO_4$ pH 8, 300 mM NaCl, 10 mM imidazole and 5 % glycerol. Upon thawing, protease inhibitors were added (AEBSF, TPCK, E64, Pepstatin and Leupeptin) and cells were lysed by French press at 16,000 psi.  TtPRODH is pure after Ni-NTA affinity chromatography (Qiagen), whereas TtP5CDH requires an extra step of anion exchange using a HiTrap Q column. Both enzymes were separately dialyzed into 50 mM Tris pH 8, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM DTT, and 5 % glycerol.

Coexpression of TtPRODH and TtP5CDH using pET-Duet1 resulted in production of TtP5CDH with cleavable 8x His tag and TtPRODH without a His tag.  From freshly transformed BL21(DE3)pLysS cells containing the pET-Duet1 vector (clones 173181, 173182, and 173192), an overnight culture was used to inoculate 700 mL of LB broth.  When $OD_{600}$ = 0.6, protein was expressed using

0.5 mM IPTG for 3 hours at 37 $^{\circ}$C.  Cells were harvested and frozen in 10 mL 50

mM NaH$_2$PO$_4$ pH 8, 300 mM NaCl, 10 mM imidazole and 5 % glycerol.  Upon

thawing, protease inhibitors were added (AEBSF, TPCK, E64, Pepstatin and

Leupeptin) and cells were lysed by French press at 16,000 psi.  Using 2.5 mL Ni-

NTA affinity resin, gravity flow purification was used to isolate His tagged

TtP5CDH and potential partner TtPRODH by elution with 250 mM imidazole.

FAD (50 $\mu$M) was added to the eluted protein and it was dialyzed into 4L of 50

mM Tris pH 8, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM DTT, and 5 % glycerol.

Total protein concentration was determined using the Comassie Plus reagent

(Pierce).  The purity was analyzed by SDS-PAGE and activity assays.

**TtPRODH and TtP5CDH activity**

TtPRODH activity was analyzed using the DCPIP assay with proline (30

mM) as the substrate (10).  TtP5CDH was assayed by monitoring NADH

formation at 340 nm using 0.2 mM NAD$^+$ and 0.5 mM P5C in 20 mM MOPS pH

7.5.

**Assessment of Intermediate P5C/GSA Channeling**

Channeling was assessed by appearance of the intermediate P5C.  P5C

was detected as the yellow dihydroquinazolinium complex with *o*-

aminobenzaldehyde (*o*-AB) by monitoring absorbance at $\lambda$ = 443 nm ($\varepsilon$ = 2900

M$^{-1}$cm$^{-1}$, (11)).  The assay mixture included 50 mM potassium phosphate buffer

pH=7.5, 200 ug of *o*-AB, 300 mM proline and 0.2 mM NAD$^+$.  For some of the

assays, membranes from *E. coli* (250 $\mu$g) or *T. thermophilus* (140 $\mu$g) were

added as electron acceptors.  Channeling assessment was also performed by

monitoring NADH formation upon addition of 300 mM proline as the substrate and 0.2 mM NAD$^+$. These assays were likewise performed in the presence and absence of membranes from *T. thermophilus* (140 $\mu$g) or *E. coli* (250 $\mu$g).

**Gel Filtration**

Gel filtration was used to assess association between TtPRODH and TtP5CDH. A Superdex 200 column was used with a buffer of 50 mM Tris pH 8, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM DTT, and 5 % glycerol.

The separately expressed and coexpressed enzymes were analyzed. For the separately expressed enzymes, many variations of mixing and incubating the two enzymes were attempted and gel filtration conducted. The enzymes were mixed at an apparent 1:1 molar ratio and 300 $\mu$L of 0.2 mg/mL protein solution was injected on the column. In a separate experiment the separately expressed enzymes were mixed in a 1:1 molar ratio and incubated at 75 $^o$C for 60 min before gel filtration. Similar studies were performed for the proteins after the His tags were removed.

For the enzymes coexpressed from pET-Duet1, gel filtration was performed after Ni-NTA affinity purification and a subsequent dialysis step. After determining total protein concentration using Comassie Plus (Pierce), 300 $\mu$L of 0.25 mg/mL protein solution was injected on the gel filtration column. A standard consisting of $\beta$-amylase (200,000 Da) was analyzed using the same buffer as the samples. The sample volume was 300 $\mu$L and the protein concentration was 0.25 mg/mL.

## RESULTS

**Analysis of separately expressed TtPRODH and TtP5CDH**

The gel filtration profile of TtPRODH (38 kDa) showed a broad peak with $V_e$ = 10.49 (FIGURE 4.2, green). Upon addition of 20 mM n-octyl β-D-glucopyranoside this peak narrowed and shifted to $V_e$ = 12.2 mL (FIGURE 4.2, brown) which indicates a species larger than bovine serum albumin (66 kDa, data not shown). This result suggested that TtPRODH forms an apparent homodimer in the presence of detergent. TtP5CDH (59.5 kDa) displayed a fairly sharp elution peak at $V_e$ = 10.0 mL (FIGURE 4.2, blue), which is larger than the elution volume of β-amylase (200 kDa, data not shown). This result is consistent with TtP5CDH forming an apparent homotetramer under these conditions (TAW unpublished result).

The enzymes were mixed at a molar ratio of 1:1 and then analyzed by gel filtration. The chromatogram had a large peak at $V_e$ = 10 mL (FIGURE 4.3, blue) and a significant peak at the void volume of 7 mL. There was also a shoulder peak near $V_e$ = 8.0-8.5 mL, which was not present in the chromatograms of the individual enzymes. Heating the mixed proteins at 75 °C resulted in almost complete disappearance of the $V_e$ = 10 mL peak and enhancement of the $V_e$ = 8.0-8.5 mL shoulder (FIGURE 4.3, green). The effect of removing the His tags from both enzymes (without heating) was investigated. As with heating, removal of the His tags resulted decreased the $V_e$ = 10 mL peak and enhanced the $V_e$ = 8.0-8.5 mL shoulder (FIGURE 4.3, red). His tag removal also resulted in a large peak at the void volume. The profile obtained after His tag removal and heating

(FIGURE 4.3, brown) was similar to that obtained by His tag removal alone (compare brown and red curves in FIGURE 4.3).

It appeared that His tag removal caused the disappearance of the $V_e$ = 10 mL peak associated with TtP5CDH and appearance of a new peak at $V_e$ =8.0-8.5 that was not observed when either enzyme was analyzed separately. We followed up on this observation by examining protein mixtures in which only one of the His tags was removed by proteolysis (FIGURE 4.4). The result for a 1:1 mixture of His-TtPRODH and untagged TtP5CDH is shown in blue in FIGURE 4.4. The peak at $V_e$ = 10 is present and there is large void volume peak. Heating the sample produced a similar chromatogram except that the peak at $V_e$ = 10 mL was decreased (FIGURE 4.4, red). Experiments corresponding to mixtures of untagged TtPRODH and His-TtP5CDH are shown in green and brown in FIGURE 4.4. In the absence of heating, this mixture resulted in a gel filtration profile with a prominent peak at $V_e$ = 8.3 mL and a peak at the void volume (FIGURE 4.4, green). The heated sample also resulted in a similar chromatogram but the peak at $V_e$ = 10 mL was larger (FIGURE 4.4, brown). These results show that the peak at $V_e$ = 8.3 is associated with removal of the His tag from PRODH (FIGURE 4.4, green and brown) and that removal of the tag from P5CDH results in a large peak at the void volume (FIGURE 4.4, blue and red).

Another set of gel filtration studies was performed using enzymes that were expressed separately, but the cells from two expression experiments were combined, lysed in the same vessel and purified using Ni-NTA chromatography.

Gel filtration profiles are shown in FIGURE 4.5. The resulting profile is very similar to that obtained when the proteins were purified separately (compare blue curves of FIGURE 4.5 and FIGURE 4.3). Heating produced a broad peak at $V_e$ = 8.4 mL (FIGURE 4.5 green). Removal of the His tags had minimal effects, which was unexpected (FIGURE 4.5 red). In particular, the peak at $V_e$ = 8.3 was absent in this chromatogram. Removing the His tags and heating decreased the $V_e$ = 10 mL peak and increased the void volume peak, but did not cause appearance of the $V_e$ = 8.3 mL peak, which was unexpected.

These studies were inconclusive, due perhaps to problems involving self-association of TtPRODH. We note that addition of detergent was necessary to obtain crystals of TtPRODH, as described previously (9).

**Coexpression of TtPRODH and TtP5CDH**

Coexpression of the two enzymes was pursued using a pET-Duet1 construct with the P5CDH gene in MCS1 and the PRODH gene in MCS2. P5CDH is expressed with a His tag while PRODH does not have an affinity tag. SDS-PAGE analysis of the protein sample after elution from a gravity flow Ni-NTA affinity column showed two major bands having the molecular weights expected for TtP5CDH and TtPRODH (FIGURE 4.6). This result suggests that TtPRODH copurified with TtP5CDH.

There was also a substantial amount of apparent TtP5CDH in the wash fraction (FIGURE 4.6). We note that in our experience the enzyme encoded by MCS1 is always expressed at a much higher level than the enzyme encoded by MCS2. The cause is unknown.

**Enzymatic activities of coexpressed TtPRODH and TtP5CDH**

The eluted protein sample displayed both PRODH and P5CDH activities. Utilizing the DCPIP assay(10), PRODH activity was detected at a level 10-fold above background when using 30 mM L-proline and 100 μL of eluted and dialyzed enzyme sample (FIGURE 4.7). P5CDH activity was also detected by measuring increasing production of NADH at 340 nm upon addition of neutralized P5C (0.5 mM), NAD$^+$ (0.2 mM) and increasing enzymes concentration (FIGURE 4.8).

The presence of P5C released into solution was tested using the *o*-AB assay. Using a proline concentration of 300 mM proline as substrate, the P5C-*o*AB complex was not detected under any condition tested (FIGURE 4.10; FIGURE 4.11; FIGURE 4.12; FIGURE 4.13). Several assays were performed: without NAD$^+$ and membranes (FIGURE 4.10), no NAD$^+$ and 140 ug of T. *thermophilus* membranes (FIGURE 4.11), 0.2 mM NAD$^+$ and no membranes (FIGURE 4.12), and 0.2 mM NAD$^+$ and 140 ug T. *thermophilus* membranes (FIGURE 4.13).

The sample was also assayed for P5CDH activity by monitoring NADH production at 340 nm using 300 mM proline and 0.2 mM NAD$^+$ (0.2 mM). No activity was detected. The assay was repeated in the presence of membranes isolated from *T. thermophilus* and *E. coli* which could act as the electron acceptor for the reduced flavin. An increase at 340 nm was observed but the same increase was observed in the absence of enzyme indicating that production of NADH was likely due to activity of enzymes associated with the membranes,

such as NAD$^+$ dehydrogenase.  We note that Maloy's group used membranes from a NAD$^+$ dehydrogenase-deficient (*ndh$^-$*) *E. coli* strain MWC215 (8).

**Gel filtration studies of coexpressed TtPRODH and TtP5CDH**

Gel filtration analysis was also conducted on protein sample purified from coexpression of the two enzymes (FIGURE 4.9).  Major peaks were observed at $V_e$ = 9.7 mL, 13.9 mL and 17.6 mL.  The $V_e$ = 13.9 peak represents a contaminant that appears on SDS-PAGE near 20 kDa and the peak at $V_e$ = 17.6 results from excess FAD added during purification.  The peak at $V_e$ = 9.7 mL has a pronounced shoulder.  Recall that the profile for separately expressed TtP5CDH has a sharp peak at $V_e$ = 10.0 mL (FIGURE 4.2 blue) and that the profile for TtPRODH has a broad peak at $V_e$ = 10.5 mL (FIGURE 4.2 green).  It is tempting to speculate that the peak at $V_e$ = 9.7 mL in the chromatogram for the coexpressed enzymes represents a species not present in the individual protein samples.

**DISCUSSION**

Purification of TtPRODH and TtP5CDH separately and mixing the two proteins did not yield definitive identification of a functional complex.  Highly aggregated species were observed in gel filtration with elution volumes similar to the void volume.  This is logical since the N-terminally His tagged TtPRODH displays significant self-association and aggregation in the absence of detergent.

Coexpression of the two enzymes yielded promising preliminary results.  SDS-PAGE and PRODH activity assays suggested that TtPRODH copurifies with

TtP5CDH.   And gel filtration chromatography revealed a new peak that was not present in profiles of the individually purified proteins.

The P5C trapping experiments were less encouraging.  Based on previous results with PutA, we expected that P5C would be produced when the enzyme pair is incubated with proline and membranes in the absence of $NAD^+$.  P5C, however, was not detected in such an assay for the coexpressed enzymes. P5CDH activity was not detected when proline was used as the substrate, although the method for detecting P5CDH activity is not optimal.  Perhaps NADH was consumed by NAD dehydrogenase isolated with membranes.  This may imply lack of a functional channel, but a better method of determining NADH formation is necessary.  Measuring the product, glutamate, instead may provide a better analysis of P5CDH activity.

Future studies will focus on performing a larger scale purification of the coexpressed enzymes for analysis with mass spectrometry, analytical ultracentrifugation, and additional enzyme assays.

# REFERENCES

1.  Miles, E. W., Rhee, S., and Davies, D. R. (1999) *J Biol Chem* 274, 12193-12196
2.  Huang, X., Holden, H. M., and Raushel, F. M. (2001) *Annu Rev Biochem* 70, 149-180
3.  Schlichting, I., Yang, X. J., Miles, E. W., Kim, A. Y., and Anderson, K. S. (1994) *J Biol Chem* 269, 26591-26593
4.  Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999) *Science* 285, 751-753
5.  Deuschle, K., Funck, D., Hellmann, H., Daschner, K., Binder, S., and Frommer, W. B. (2001) *Plant J* 27, 345-356
6.  Hellmann, H., Funck, D., Rentsch, D., and Frommer, W. B. (2000) *Plant Physiol* 123, 779-789
7.  Deuschle, K., Funck, D., Forlani, G., Stransky, H., Biehl, A., Leister, D., van der Graaff, E., Kunze, R., and Frommer, W. B. (2004) *Plant Cell* 16, 3413-3425
8.  Surber, M. W., and Maloy, S. (1998) *Arch Biochem Biophys* 354, 281-287
9.  White, T. A., and Tanner, J. J. (2005) *Acta Crystallography* F61, 737-739
10. Abrahamson, J., Baker, L., Stephenson, J., and Wood, J. (1983) *European Journal of Biochemistry* 134, 77-82
11. Mezl, V. A., and Knox, W. E. (1976) *Anal Biochem* 74, 430-440

FIGURE 4.1.  Schematic diagram of proline catabolic enzymes from *E. coli, B. japonicum*, and *T. thermophilus*.
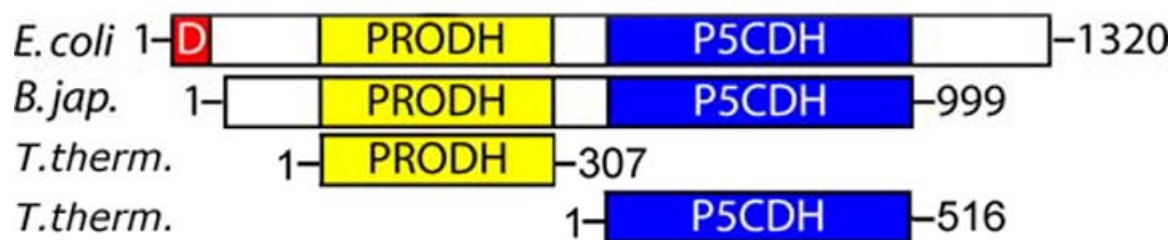
FIGURE 4.2. Gel filtration profiles for TtPRODH and TtP5CDH. The enzymes were expressed separately from pKA8H and purified by Ni-NTA affinity chromatography. Profiles for TtPRODH are shown in green (no BOG) and brown (20 mM BOG). The profile for TtP5CDH is shown in blue. The void volume is approximately 7 mL.
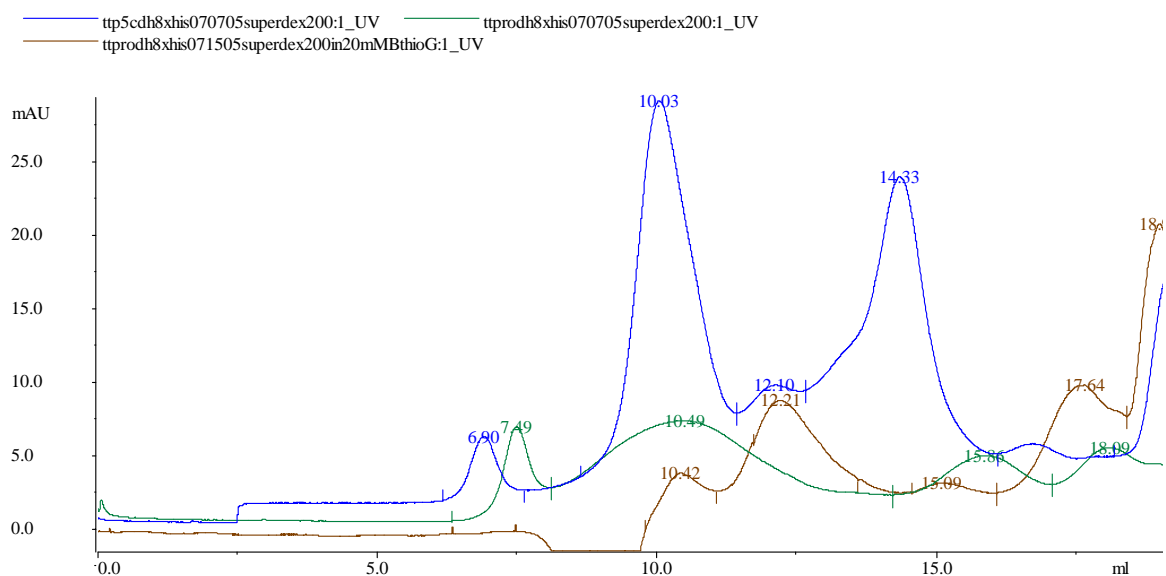
FIGURE 4.3. Gel filtration profiles for mixtures of TtPRODH and TtP5CDH. Each protein was separately expressed, lysed and purified. The purified proteins were mixed in a 1:1 molar ratio. The profile for the mixture with no further manipulation is shown in blue. The green curve corresponds to heating the mixture at 75 $^{\circ}$C for 60 min. The red profile corresponds to the mixture after cleavage of the His tags with TEVP. The brown curve results from His tag cleavage and heating at 75 $^{\circ}$C for 60 min. Void volume is approximately 7 mL.
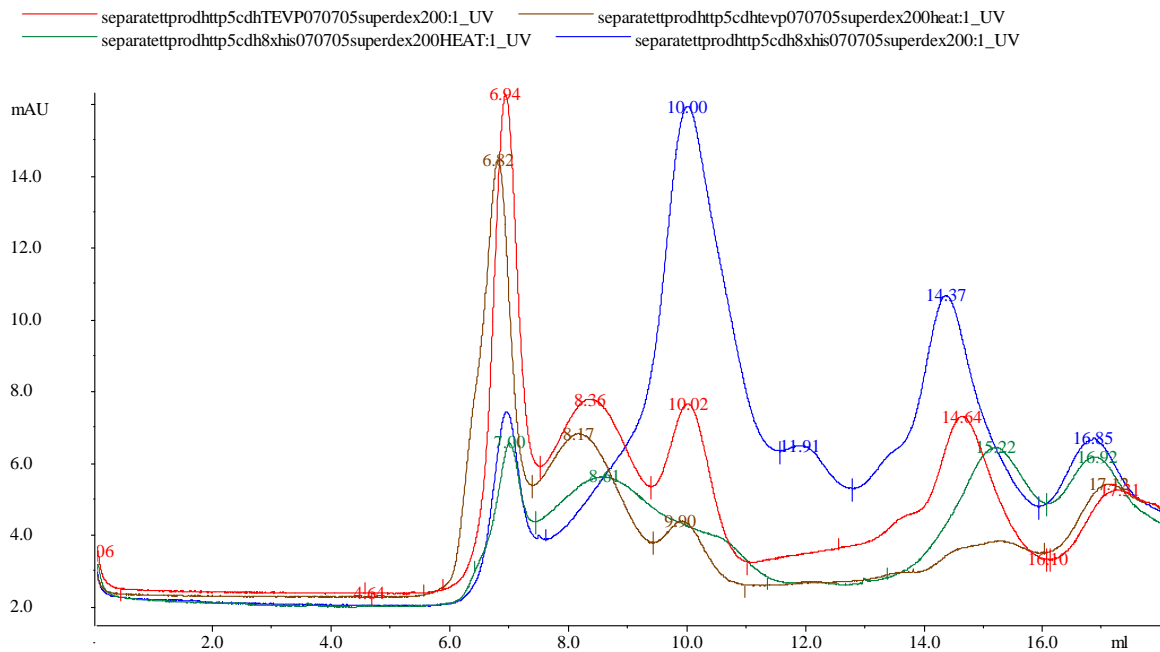
FIGURE 4.4. Gel filtration profiles for mixtures of TtPRODH and TtP5CDH in which the His tag of one protein had been cleaved. Each protein was separately expressed, lysed and purified. The proteins were mixed in 1:1 molar ratios. The blue curve represents a mixture of His-TtPRODH and cleaved TtP5CDH. The red curve represents a mixture of His-TtPRODH and cleaved TtP5CDH incubated together at 75 $^\circ$C for 60 min. The green curve represents a mixture of His-TtP5CDH and cleaved TtPRODH. The brown curve represents a mixture of His-TtP5CDH and cleaved TtPRODH incubated together at 75 $^\circ$C for 60 min. Void volume is approximately 7 mL.
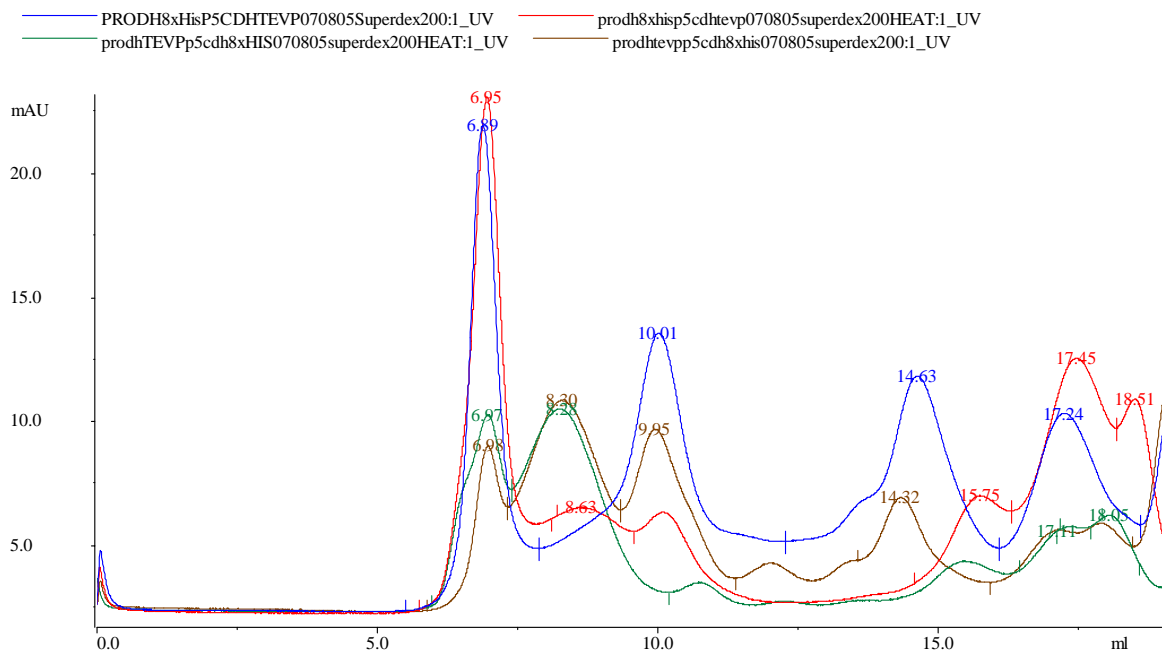
FIGURE 4.5. Gel filtration profiles for separately expressed, but copurified samples of TtPRODH and TtP5CDH. The His-tagged enzymes were expressed separately, combined as whole cells, lysed in the same vessel and purified with Ni-affinity chromatography. The blue curve represents the purified protein sample with no further manipulation. The green curve represents the sample incubated at 75 $^o$C for 60 min. The red curve represents the sample after cleavage with TEVP. The brown curve represents the sample after cleavage with TEVP and heating at 75 $^o$C for 60 min. Void volume is approximately 7 mL.

FIGURE 4.6. SDS-PAGE analysis of coexpressed His-TtP5CDH and untagged TtPRODH. The sample was purified using gravity-flow Ni-NTA affinity chromatography.
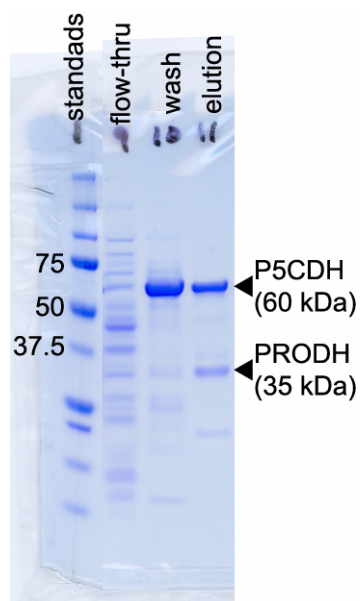
FIGURE 4.7. PRODH activity assay of coexpressed His-TtP5CDH and TtPRODH after Ni-NTA chromatography. The slope from a control assay performed without enzyme -0.0026.



Calculated Slope-0.0214538

DCPIP ASSAY:
8x-Histag-TEV-TtP5CDH coexpressed
with TtPRODH in pET-Duet1
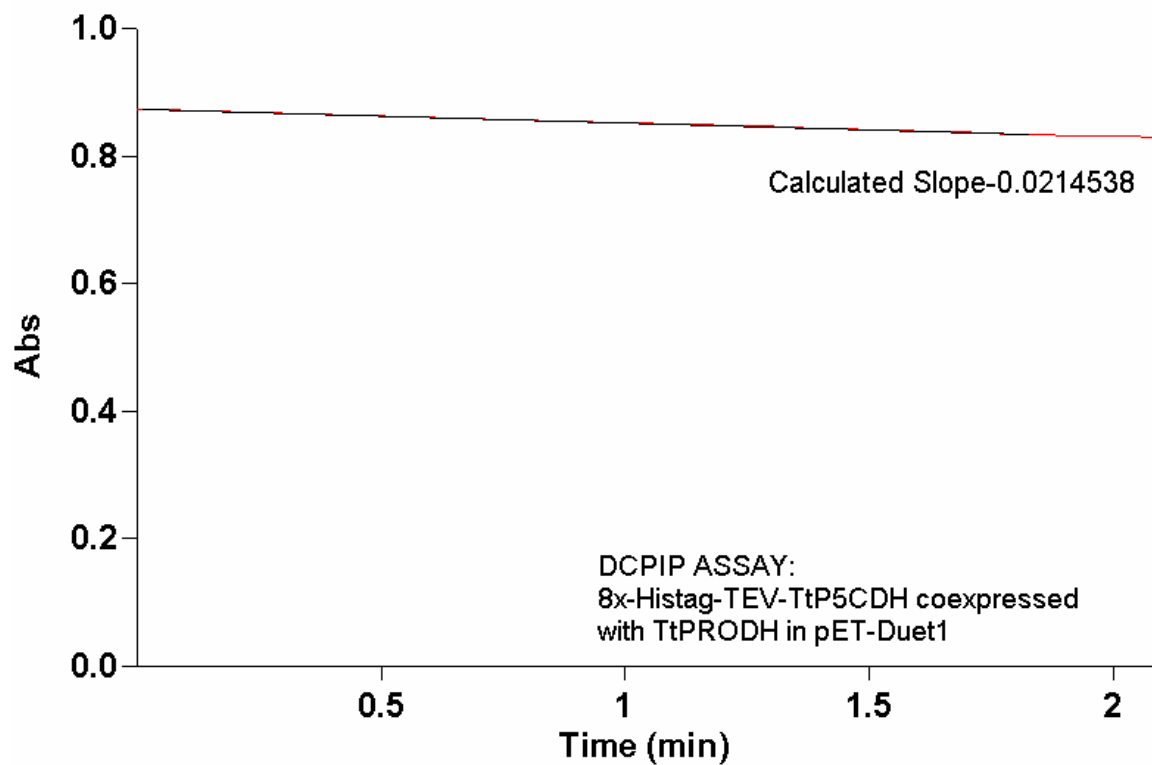
FIGURE 4.8. TtP5CDH activity assay of coexpressed His-TtP5CDH and TtPRODH after Ni-NTA chromatography using exogenous P5C as the substrate. NADH was monitored as absorbance at 340 nm. The panels correspond to different amounts of enzyme added. The assay contained P5C (0.5 mM, pH 7), and NAD$^+$ (0.2 mM) in MOPS (20 mM pH 7.5).
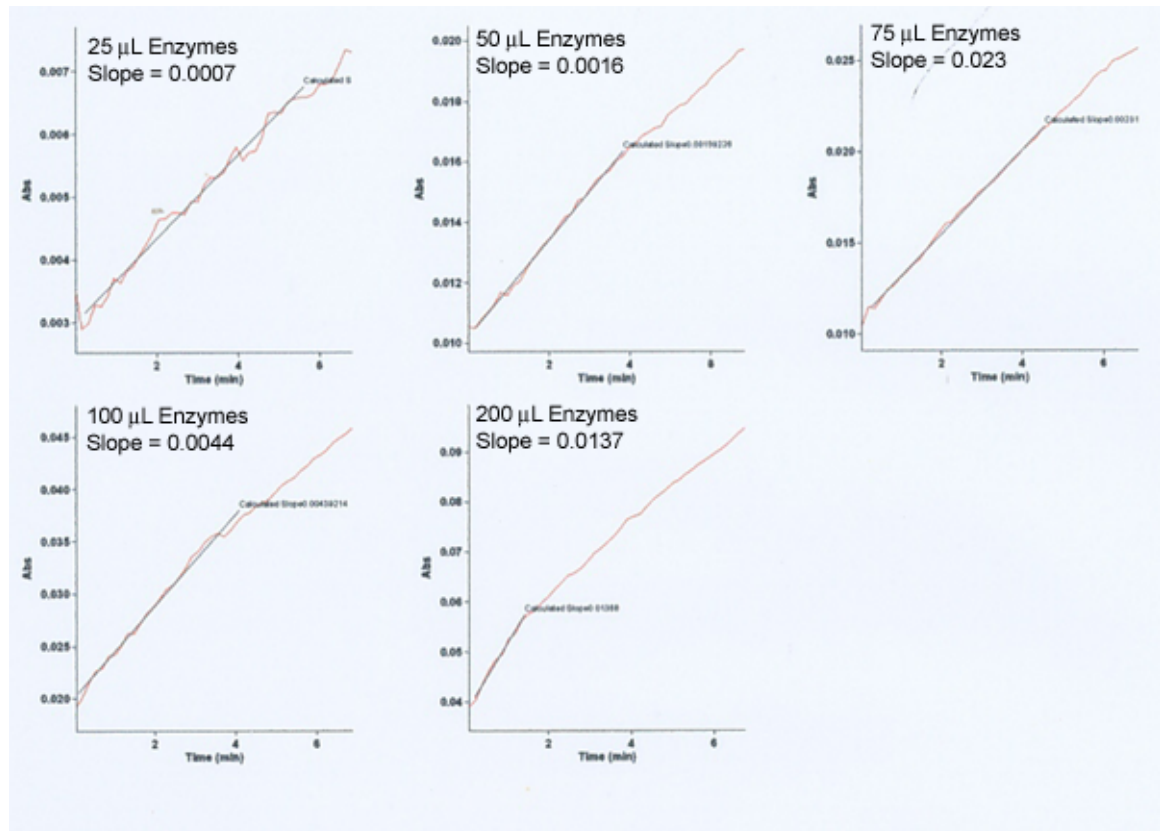
FIGURE 4.9. Gel filtration chromatogram of coexpressed His-TtP5CDH and TtPRODH after Ni-NTA chromatography. A) Chromatogram for the enzyme sample. Note the peak at 17.6 mL is most likely due to excess flavin added during dialysis. B) β-amylase standard (200 kDa) run on the same day.

**A.**



**B.**
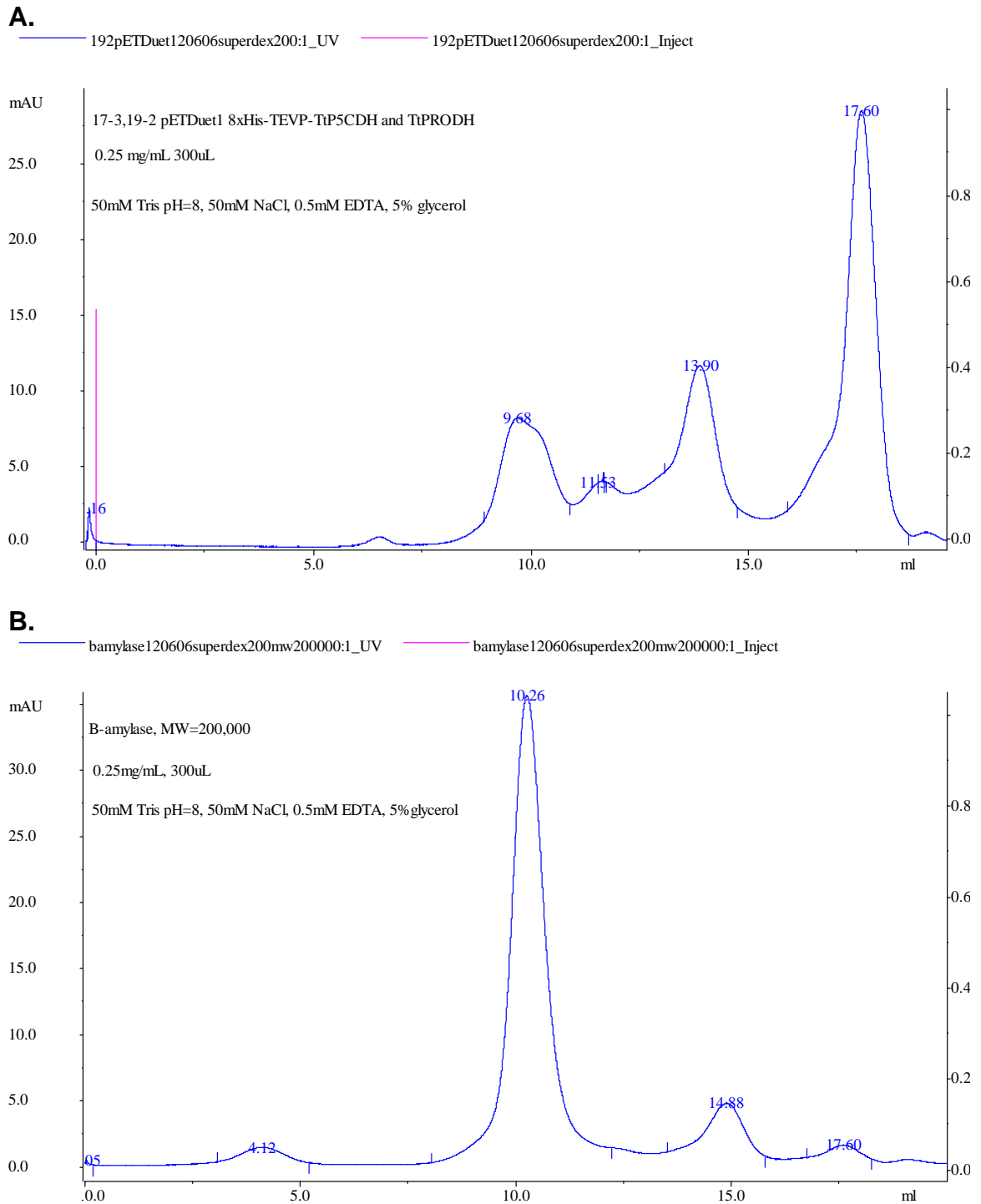
FIGURE 4.10. Detection of P5C using the *o*-AB assay with coexpressed His-TtP5CDH and TtPRODH after Ni-NTA chromatography. Increasing volumes of the purified enzymes were assayed for P5C:o-AB complex formation. No NAD$^+$ or *T. thermophilus* membranes containing electron acceptors were added to these experiments. Purified 8x-histagged-TEV-TtPRODH (20 $\mu$g) assayed with proline (30 mM) was used as a positive control (last graph).
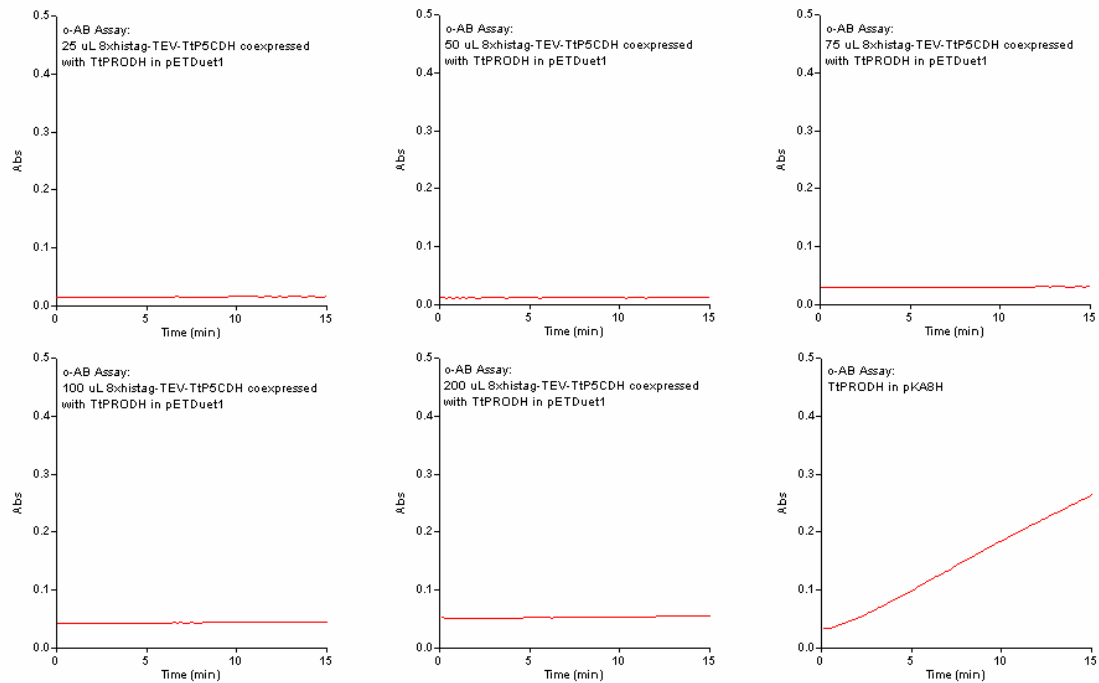
FIGURE 4.11. Detection of P5C using the *o*-AB assay with coexpressed His-TtP5CDH and TtPRODH after Ni-NTA chromatography. Increasing volumes of the purified enzymes were assayed for P5C:*o*-AB complex formation. Proline (300 mM) and *T. thermophilus* membranes (140 $\mu$g) containing electron acceptors were added. No NAD$^+$ was added to these experiments.
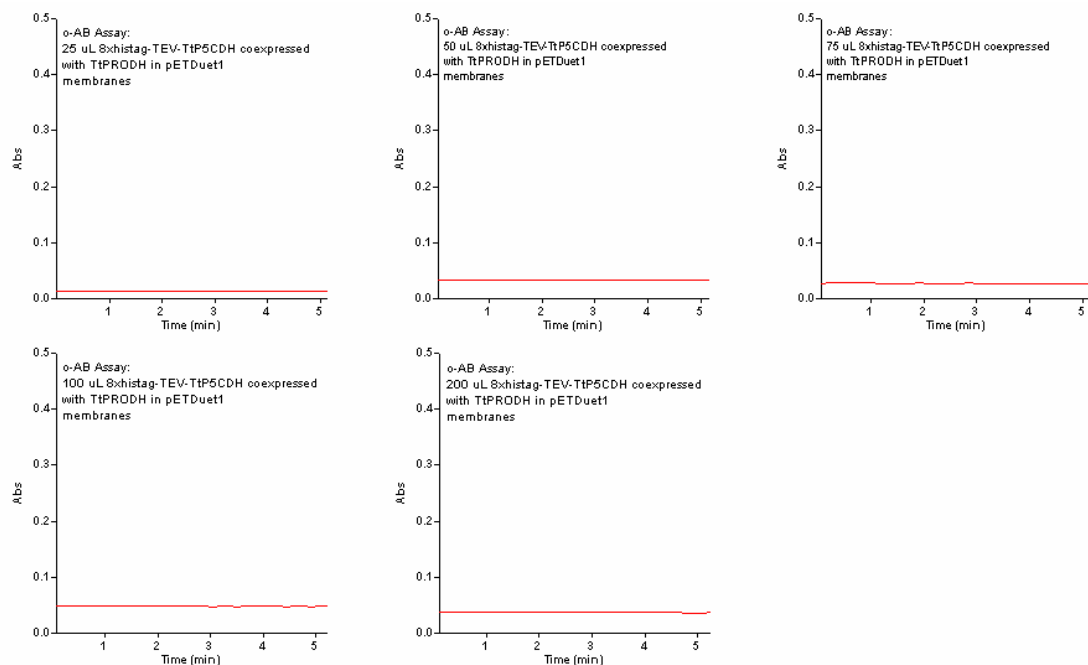
FIGURE 4.12. Detection of P5C using the *o*-AB assay with coexpressed His-TtP5CDH and TtPRODH after Ni-NTA chromatography. Increasing volumes of the purified enzymes were assayed for P5C:o-AB complex formation. Proline (300 mM) and $NAD^+$ (0.2 mM) were added to these experiments. No *T. thermophilus* membranes containing electron acceptors were added. Purified 8x-histagged-TEV-TtPRODH (20 $\mu$g) assayed with proline (30 mM) was used as a positive control (last graph).
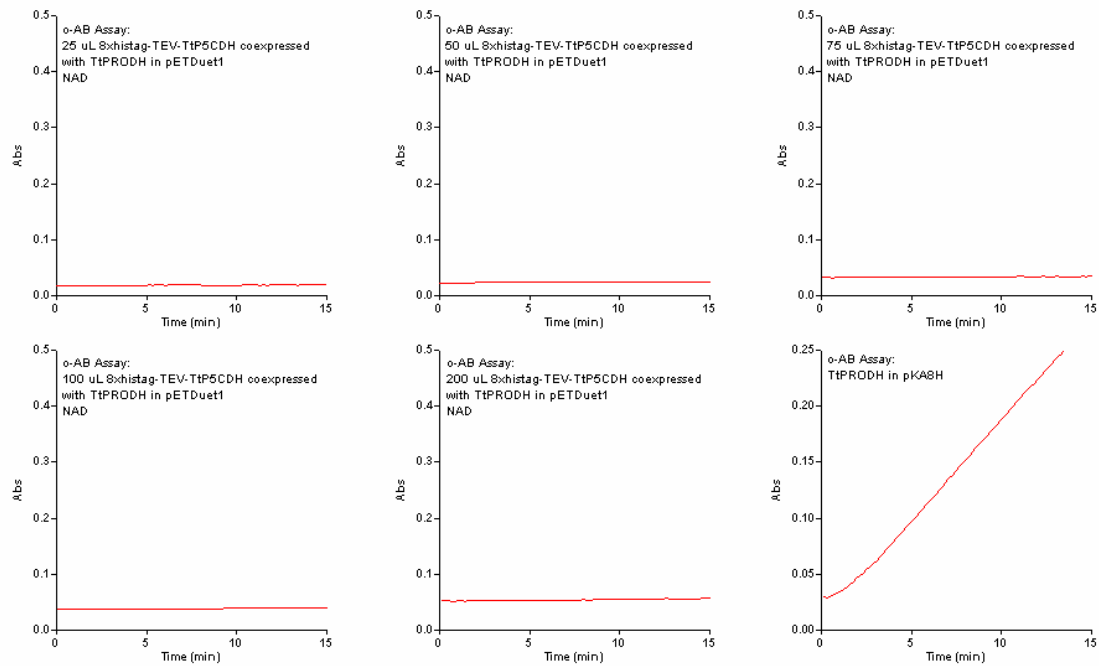
FIGURE 4.13. Detection of P5C using the *o*-AB assay with coexpressed His-TtP5CDH and TtPRODH after Ni-NTA chromatography.  Increasing volumes of the purified enzymes were assayed for P5C:*o*-AB complex formation.  Both NAD$^+$ (0.2 mM) and *T. thermophilus* membranes (140 $\mu$g) containing electron acceptors were added to these experiments.  Purified 8x-histagged-TEV-TtPRODH (20 $\mu$g) assayed with proline (30 mM) was used as a positive control (last graph).
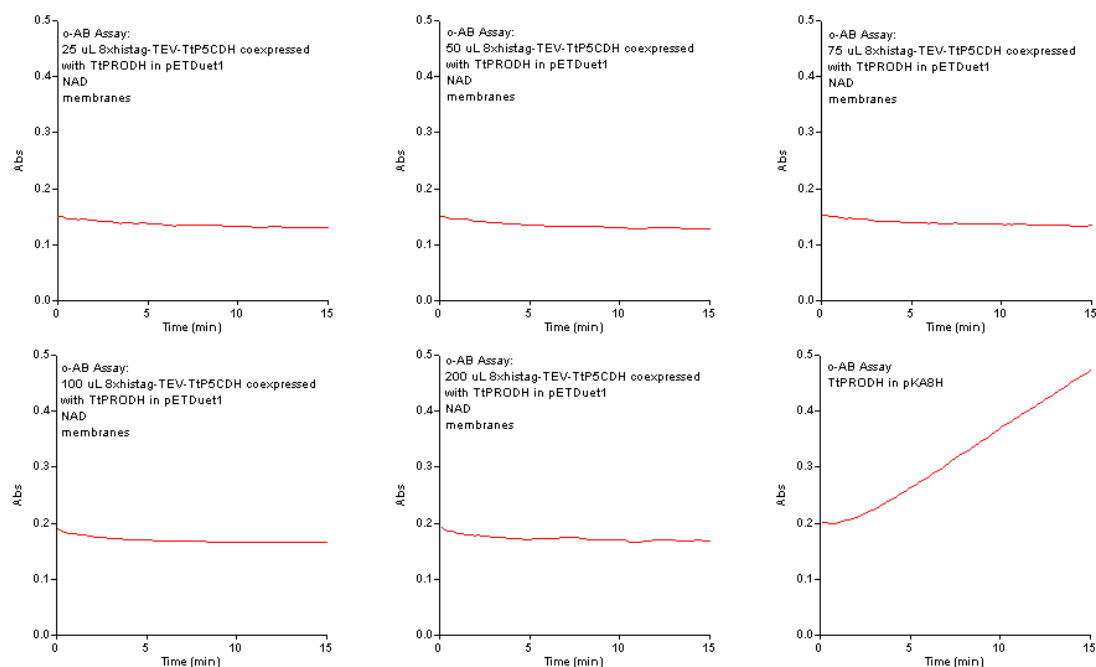
TABLE 4.1. PCR primers used for subcloning TtP5CDH and TtPRODH into pET-Duet1.

| TtPRODH – MCS 1 | |
|---|---|
| Forward | 5' – TAATACGACTCACTATAGGG - 3' |
| Reverse | 5' - TGCAGTGAATTCGGATCCCTAGCCGGAAACCAGGCT - 3' |

| TtP5CDH – MCS 1 | |
|---|---|
| Forward | 5' – TAATACGACTCACTATAGGG - 3' |
| Reverse | 5' - TGCAGTGAATTCGGATCCCTAGCCGGAAACCAGGCT - 3' |

| TtPRODH – MCS 2 | |
|---|---|
| Forward | 5' - TAATACGACTCACTATAGGG - 3' |
| Reverse | 5' - TGTTTGACAGCTTATCATCG - 3' |

| TtP5CDH – MCS 2 | |
|---|---|
| Forward | 5' - TAATACGACTCACTATAGGG - 3' |
| Reverse | 5' - TGTTTGACAGCTTATCATCG - 3' |

# CHAPTER 5

## Attempted Structure Determination of *Bradyrhizobium japonicum* PutA: Inherent Crystallographic Pathology

### INTRODUCTION

Research described in this chapter continues work started by J.P. Schuermann on structure determination of *Bradyrhizobium japonicum* PutA (BjPutA). Initial structure determination was performed by JPS using crystals with hexagonal bipyramidal shape and apparent space group $P6_222$. Models were built using MAD and SAD phases from a selenomethionyl derivative (JPS, dissertation). Electron density was quite strong for residues 362-989 (?) and a nearly complete model in this part of the protein could be built. Density in parts of the PRODH domain was much weaker, particularly for the FAD isoalloxazine and strands near the isoalloxazine. These best model had R=0.26 $R_{free}$ =0.36 for data to 2.6 Å resolution. This model included 943 residues and 7142 atoms. Although the R-values is acceptable for a large protein at moderate resolution, the high value of $R_{free}$ and the large difference between $R_{free}$ and R indicated potential problems with the data and/or model.

A second crystal having a diamond shape was obtained in the same crystallization conditions as the hexagonal bipyramids. These crystals had apparent space group $P4_32_12$ and the diffraction was weaker than that of the hexagonal form. Refinements against data collected from the tetragonal form also produced unacceptable statistics.

Data from both crystal forms displayed unusual intensity statistics

indicating inherent pathologies such as twinning, anisotropy and pseudosymmetry. This chapter discusses my attempts to identify these pathologies and to account for them during refinement. In the rest of this section, I discuss these various pathologies and methods of diagnosing them. I apply these techniques to the hexagonal and tetragonal crystal forms in the remainder of this chapter.

**Crystal Pathology: Twinning, Pseudosymmetry, and Anisotropy**

Hemihedral twinning in protein crystals occurs when two crystal domains are oriented differently (1). These different domain orientations can be described by whether or not they overlap in all dimensions (2). These different orientations are related by a transformation that does not correspond to the crystal's point group symmetry, but is related to the crystal lattice (FIGURE 5.1) (1,2).

Merohedral twinning refers to crystals having a higher lattice symmetry than point group (class) symmetry (1)  Merohedral twin domains overlap in all three dimensions (2). The transformation that describes the symmetry element contributing to twinning is known as the twinning operator.  This is classic twinning, and is possible in higher symmetry spacegroups of trigonal, tetragonal, hexagonal or cubic  (1). In the case of hemihedry, each observed intensity ($I_{obs}$) is the weighted sum of the intensities of the two reflections, $h_1$ and $h_2$. These reflections are not related by crystallographic symmetry, but by a twinning operation (TABLE 5.1) (2).  The twin domains may contribute equally to the observed diffraction intensities (perfect twinning) or unequally (partial twinning), and is defined by the twin fraction, $\alpha$.  For perfectly twinned crystals, the twin

fraction is 0.5 and for partial twining the twin fraction ranges from $0 < \alpha < 0.5$.
The following equations describe the relationships between the twin-related
reflections.

$$I_{obs} (h_1) = (1 - \alpha)/I(h_1) + \alpha\, I(h_2)$$

$$I_{obs} (h_2) = \alpha\, I(h_1) + (1 - \alpha)/\, I(h_2)$$

Pseudomerohedral twinning occurs when, by chance, the crystal metric
corresponds to a higher symmetry system (1). The twin domains of pseudo-
merohedrally twinned crystals overlap approximately, but not exactly, in 3-
dimensions (2). Although pseudomerohedry is commonly unreported, this can
occur in centered-orthorhombic, tetragonal, and hexagonal (3,4). Some
pseudomerohedral twins may be recognized at higher resolution since the spot
profiles start to split (1). Most pseudomerohedral twins have identical
characteristics to merohedral twins.

Non-merohedral twinning occurs when only a subset of reflections overlap
(in less than 3-dimensions), which can be seen upon inspection of the diffraction
image (1,2). Inspecting diffraction intensities from twinned crystals can be useful
for diagnosing non-merohedral twinning.

Besides twinning, another potential pathology is the occurrence of
pseudosymmetry due to a rotation. Rotational pseudosymmetry (RPS) occurs
when the rotational component of the non-crystallographic symmetry operator is
approximately equal to the twinning operator in a potentially twinned crystal.

Lebedev, 2006 discusses cases of RPS, effects on these cases on the perfect twinning test, and identification (3). When a crystal is twinned, it is common that RPS exists resulting in ambiguous twinning test results.

Translational non-crystallographic symmetry (TNCS) occurs if multiple molecules occur in the asymmetric unit and are related by a parallel translation by a vector close to a fraction of one unit cell dimension. This pathology will cause some classes of reflections will have abnormally weak intensities, shifting the overall intensity distribution toward a centrosymmetric character. TNCS can occur for both twinned and non-twinned crystals. Pseudocentering is a special type of translational non-crystallographic symmetry (1).

Anisotropy is another pathology which may affect refinement (5). Anisotropic correction may be necessary if a crystal diffracts less in one dimension of reciprocal space than in the others. Anisotropically-diffracting crystals have reflections with unusual intensity distributions, similar to what is seen with translational non-crystallographic symmetry.

With the many potential crystal pathologies that can be present, it is necessary for the crystallographer to be aware of potential problems that can arise. Detection of these pathologies by careful inspection of intensities during data collection and later during refinement will facilitate speed structure determination. Methods for detecting twinning and the above pathologies are be discussed below.

**Detection of Crystal Pathology**

Howells first introduced cumulative intensity distribution N(z) plots to

differentiate between centrosymmetric and non-centrosymmetric crystals (1,6). The variable z is the local average intensity calculated in small resolution bins and $N(z)$ is the fraction of reflections that have intensities lower than the average. For non-centrosymmetric crystals, the fraction of reflections with lower than average intensity is less than for centrosymmetric crystals. For a twinned crystal, the fraction of the non-centrosymmetric reflections with lower than average intensities are even less. A plot can be constructed and plotting z versus $N(z)$ results in two lines for centric and acentric reflections, known as a cumulative intensity distribution. Plotting intensity statistics for normal crystals results in exponential curves, although with a twinned crystal, a sigmoidal shape is observed for acentric reflections. Although qualitative, the cumulative intensity distribution can provide insights into unusual intensities, which may be indicative of inherent crystal pathology. For an example of a plot of cumulative intensity distributions for non-centrosymmetric reflections see FIGURE 5.2.

Normalized within resolution shells, observed intensities can be sampled in one-dimension and assigned the variable Z. The perfect twin test plots resolution versus the ratio of the average of the squared intensities over the average intensity squared, $< I^2 > / < I >^2$, also known as the second moment of Z, or $Z^2$ (FIGURE 5.3). To perform these tests, higher resolution data should be removed where $R_{standard}$, $< \sigma (F) > / < F >$, increases. $R_{standard}$ can be obtained from the program SFCHECK (7). As a function of resolution, these values should fall around 2 for an untwinned crystal. With a perfect merohedrally twinned crystal, the value of the perfect twin test will be around 1.5 (TABLE 5.2, FIGURE

5.3).

Estimation of the twin fraction is necessary for properly dealing with intensities from twinned crystals. There are multiple tests for estimating twin fraction. The most popular tests are the Britton plot, Murray Rust Plot, Yeates S(H) plot (2), Rees N(z) plot, and more recently the L-function (1,8-12). Dauter (2003) provides a good overview of each of these methods championing the Yeates S(H) plot to determine twin fraction(1). Calculation of the ratio of difference to sum of twin-related diffraction intensities ($I_1$ and $I_2$) allows one to accurately determine twinning fraction, this term is named H and is represented by the following equation.

$$H = |I_1 - I_2| / |I_1 + I_2|$$

Plotting H of non-centrosymmetric reflections versus the dependence of the cumulative distribution of H, known as S(H), leads to straight lines with a slope of $1 / (1 - 2\alpha)$. Twin fraction can be estimated from the slope of this line, which can be performed with SFCHECK, only if the program detects twinning and the spacegroup allows merohedral twinning.

Various methods of dealing with hemihedral twinning have been described. One approach is to detwin data. This results in a loss of reflections resulting in loss of completeness. This can be done with CCP4i DETWIN or CNS partial/perfect detwin scripts (13,14). Another method is to take twinning into account during refinement was described by Herbst-Irmer & Sheldrick and

implemented in SHELXL and CNS(15). During refinement, one enters the twinning operator and twin fraction.

Translational NCS or pseudocentering has the opposite effect on the perfect twin test displaying values much greater than 2, although this deviation from 2 will decrease with increasing symmetry. Although the perfect twin test may indicate twinning, it doesn't mean that the twinning will be merohedral.

For crystals displaying anisotropy and pseudocentering (TNCS), Padilla and Yeates introduced the Local Intensity function, or L-function to compare intensities that may vary systematically within crystals exhibiting these abnormalities (1,12). The L-function is based on the idea of the H-function, which was discussed earlier with merohedral twinning (1,12). The H-function which identifies twin fraction $\alpha$, subtracts twin-related intensities and divides this by their sum (2). Instead of using twin-related intensities, the L-function uses reflections in reciprocal space that are proximally located (1,12).

$$L = \{I(h_1) - I(h_2)\} / \{I(h_1) + I(h_2)\}$$

Using the local intensity function, anisotropic crystals that had unusual intensity statistics indicating twinning now display values that indicated these crystals were not twinned (12). Other crystals that were highly merohedrally twinned and pseudocentered displaying normal intensity statistics, utilizing L-function analysis displayed the twinned values which were indicative of the crystal's pathology. TABLE 5.1 shows expected values for normal and twinned crystals.

Examination of the self-rotation function as calculated by GLRF can help identify extra symmetry elements that are not described by the point symmetry (16). This test identifies symmetry that is related by rotation around a certain axis. Also defined as rotational pseudosymmetry, extra peaks arise in the calculated rotation functions. From these peaks, it can be determined if there are extra elements of symmetry not described by the point group.

Translational pseudosymmetry can be identified by viewing a native Patterson map at 4 Å resolution. Off-origin peaks with peak heights of 15-20% of the origin peak height indicate translational non-crystallographic symmetry. Translational non-crystallographic symmetry (TNCS) can also be detected in the perfect twin test (3). A perfect twin test having a value above much greater than 2 at moderate resolution is indicative of translational non-crystallographic symmetry. It is essential that a moderate resolution cutoff for the data is utilized to examine TNCS due to increased error of high resolution data which twinning tests do not take into account (3). Another test that can be utilized for detection of pseudocentering is the parity test (17). Systematically weak reflections are indicative of pseudocentering, a specific type of TNCS, and can be determined by calculating the average intensity ratios for even and odd reflections. A crystal displaying no pseudocentering will result in a parity test ratio of one since the intensities will be equivalent. A pseudocentered crystal will have systematically weak absences, resulting in a ratio less than one. For example, a crystal with C-face pseudocentering will have the H+K reflections odd/even less than 0.5. The weak class of reflections indicates which face is pseudocentered. Different high

resolution cutoffs should be employed to detect pseudocentering, which is more likely to show up at lower resolution cutoffs. Weak average intensity ratios can be examined with DATAMAN using the parity test (17).

Regular refinement programs such as Refmac and CNS can scale mild or moderate anisotropy. When it becomes strong or severe, other strategies for handling anisotropy are needed. Anisotropy affects R-factor in two ways. Considerable quantities of poorly measured weak reflections ($F/\sigma < 3$) increases the R-factor in anisotropic diffracting crystals. The features of electron density maps are also degraded with anisotropic diffraction. This is because refinement programs scale up the weaker diffraction and scale down the stronger diffraction. Strong, et al. suggest using an ellipsoidal cutoff to remove these weak, poorly measured reflections and has designed a server to complete these calculations (5).

## MATERIALS AND METHODS

### Crystals and Soaks

Crystals used for these experiments were grown by J.P. Schuermann or D. Karr. Mutant BjPutA A310V was characterized as described previously (18). Most of the work described here is focused on crystals of native BjPutA. Some work on the A310V mutant is also described. This mutant was characterized previously by Becker's group (18).

For radical scavenging experiments, crystals were cryoprotected with the final solution containing 250 mM ascorbic acid, looped and plunged into liquid $N_2$.

For reduction of crystals, crystals were cryoprotected with the final solution containing 25 mM L-proline and 25 mM sodium dithionite.  When the yellow crystals were bleached upon reduction, they were plunged into $N_2$ (*l*).

**Data Collection and Processing**

Several data sets were collected at APS and ALS (APPENDIX 5.1).  I was responsible for data sets collected in Jan. 2005 and beyond.  Jon Schuermann collected the other data sets.  Although many data sets were collected and analyzed, I will focus only on a few representative ones in this thesis.  Unless otherwise indicated, the representative data sets discussed here are the d8x5 data set collected from a hexagonal crystal reduced with 25 mM dithionite and 25 mM L-proline (ALS 4.2.2, September 2006), and the d3_04_2 data collected from a tetragonal crystal (ALS 4.2.2, May 2006).  Indexing logs for these data sets are provided in APPENDIX 5.2 for d8x5 and APPENDIX 5.3 for d3_04_02.  Data processing statistics are summarized in APPENDIX 5.4 for d8x5 and APPENDIX 5.5 for d3_04_02.  Resolution cutoffs were determined by analyzing the high resolution bin values for $R_{merge}$ (less than 0.4) and I/$\sigma$ unaveraged (greater than 2).

**Molecular Replacement**

Molecular replacement was conducted using with CCP4 MOLREP when the spacegroup was known and PHASER when exploring alternate choices of spacegroup (19,20).  Models used for were initially built by J.P. Schuermann and J.J Tanner  For model 104nofad.pdb 943 residues and 7142 atoms and was built using Se-Met SAD data that had been locally scaled by RESOLVE.  The

refinement R-factors are R = 0.2968 and $R_{free}$ = 0.3859. From this model, another model (d8x5_1_001_refmac12.pdb) was generated from multiple refinements alternating between CNS simulated annealing and Refmac5 maximum likelihood refinements utilizing TLS using the d8x5 data processed with MOSFLM in $P6_222$ (14,21,22). This model contained chain breaks where density was lacking, especially in loops, but consisted of residues 31-992. This model has 764 residues and 5040 atoms.

**Refinement**

Refinement (without twinning taken into account) was performed with CNS simulated annealing and Refmac5 maximum-likelihood restrained refinement utilizing TLS (14,22). Refinement with merohedral twinning was performed with CNS (14). COOT was used for viewing maps and model building (23).

**Crystal Pathology Detection**

Several tests and programs were used to detect crystal pathologies. Inspection of cumulative intensity distributions were viewed using the CCP4i import module log graphs output (13,24). Perfect twin tests were generated using CCP4 SFCHECK utilizing all data (shown) and data truncated to 3.5 Å (not shown). L-function tests were performed using moderate resolution data (8-3.5 Å) using the web server at http://nihserver.mbi.ucla.edu/pystats/ (12). Parity tests were conducted with DATAMAN using various values of the high resolution cut-off in the range 3 Å - 11Å (17). Native Patterson maps were generated using the CCP4i Generate Patterson maps module using data to 4 Å resolution (13). Self rotation function maps were generated using CCP4i MOLREP using data to

4 Å resolution (19). Anisotropy was detected and resolution cutoffs made using the web server at [http://www.doe-mbi.ucla.edu/~sawaya/anisoscale/](http://www.doe-mbi.ucla.edu/~sawaya/anisoscale/) (5).

**Merohedral Twinning**

Merohedral twinning refinements for the hexagonal crystal form were conducted using CNS least-squares twinning refinements in all possible space groups and using various twin fractions (TABLE 5.2) (14). For the tetragonal form, the twin fraction was obtained using SFCHECK and twin operator used was provided by CNS.

## RESULTS

**BjPutA Mutant A310V**

BjPutA Mutant A310V was utilized for crystallographic experiments since this mutant can be reduced fully with the substrate L-proline. Hexagonal bipyramids were grown in similar conditions as wildtype (D.Karr, personal communication). Data sets collected at ALS beamline 4.2.2 had high resolution limit of only 3 Å. A minimal model having ~63% of the residues modeled (residues 362-989) was used for refinement in REFMAC5 using space group $P6_2 22$. Typical statistics from refinement were R=0.43-0.5 and $R_{free} > 0.5$. These crystals were not pursued further due to the weak resolution and poor refinement statistics.

**Crystal Reduction, Radical Scavenging and Merging Initial Data**

On ALS beamline 8.3.1, it became apparent that the portion of the BjPutA hexagonal bipyramid crystal illuminated by the X-ray beam was being bleached

(JP Schuermann, personal communication). Radical scavenging experiments using crystals soaked in ascorbate (250 mM) and merging the initial $5^o$ of data from several crystals (TABLE 5.3) were attempted to prevent possible conformational changes upon this potential reduction. Data was collected and processed assuming $P6_222$ spacegroup. Molecular replacement using MOLREP was conducted (19) with both the P5CDH domain and the entire 104nofad.pdb structure (TABLE 5.4). Refinements were conducted using CNS rigid body and simulated annealing refinement (14). The best R-factor was obtained with initial $5^o$ merged data set had value of 0.3908 with an $R_{free}$ of 0.494, which was obtained by refinement of 104nofad.pdb against ALS beamline 4.2.2 May 2006 datasets 1, 2, 5, 6 and 7 (see TABLE 5.3 and TABLE 5.4 for the actual dataset names). The best $R_{free}$ that was obtained with any single data set had a value of 0.4875 with an R-factor of 0.3935. This was obtained by refinement of 104nofad.pdb against the single dataset ALS beamline 4.2.2 May 2006 dataset b12_05.

**Proline and Dithionite Reduction of BjPutA**

To lock-in conformational changes that may occur upon BjPutA reduction in the X-ray beam, the hexagonal bipyramidal crystals were reduced. Before looping, these crystals were soaked 25 mM each dithionite and L-proline until bleached indicating that these crystals were reduced (FIGURE 5.4 A-C shows diamond form). The dataset d8x5 was collected at ALS beamline 4.2.2 in September 2006. Data were processed to 2.6 Å with MOSFLM with in an apparent spacegroup of $P6_222$ (21). MOLREP found one molecule in the

asymmetric unit using the P5CDH domain (residues 546-992) from 104nofad.pdb as the search model.  The R-factor was 0.4571 and $R_{free}$ was 0.4894.  Using both programs iteratively, after 24 rounds of refinement in CNS utilizing simulated annealing then REFMAC5 with TLS (12 rounds of each), a model (d8x5_1_001_refmac12.pdb) consisting of residues 31-992 (5040 atoms) was constructed which was missing disordered loops throughout the PRODH domain and between the PRODH domain and P5CDH domain.  The R-factor was 0.3529 and $R_{free}$ was 0.4204 after REFMAC5 refinement using two TLS domains consisting of residues 31-535 and 542-992.

**Oxidized Diamonds**

Data from diamond crystals were collected assuming spacegroup of $P4_32_12$.  These crystals diffracted to 2.6 Å. Molecular replacement using MOLREP provided a suitable solution with two molecules in the asymmetric unit using 104nofad.pdb as the search model with an R of 0.506 and a correlation coefficient of 0.225 (19).  Restrained REFMAC5 refinement (no TLS) resulted in an R-factor of 0.4176 and $R_{free}$ of 0.489 (22).

**Spacegroup investigation for the Hexagonal Bipyramidal form**

Alternative choices for the spacegroup were investigated for the hexagonal bipyramids.  PHASER was utilized to investigate alternate choices of spacegroup using d8x5_1_001_refmac12.pdb (20).  The spacegroups with the highest log-likelihood gain were $P6_222$, $P6_2$, $P3_221$, and $P32$.  CNS simulated annealing refinements of the molecular replacement solutions obtained with PHASER are listed in TABLE 5.5 (14,20).  The best solution was obtained with

P3$_2$ (best R-factor of 0.3652) and P6$_2$ (best R$_{free}$ of 0.4165). This result is also reflected in those obtained using REFMAC5 refinement calculations post-simulated annealing in CNS, with P3$_2$ (best R-factor of 0.3399) and P6$_2$ (best R$_{free}$ of 0.3906) (not shown), giving the best results.

**Analysis of Pathology Indicators**

*Cumulative Intensity Distribution*

Upon analysis of the cumulative intensity distributions, it was observed that the data from the hexagonal bipyramids have acentric reflections with centric character (FIGURE 5.5). This indicates TNCS or pseudocentering. The diamond data sets had acentric reflections which had sigmoidal curvature, indicating twinning (FIGURE 5.6).

*Perfect Twin Test*

A valute of 2 is expected for nontwinned data and 1.5 indicates perfect twinning. Hexagonal bipyramids had a value much, greater than 2 (FIGURE 5.7B, last panel). This value much greater than two indicates TNCS or pseudocentering. The perfect twin test for the diamonds had a value lower than 2, indicating twinning (FIGURE 5.8B, last panel).

*Partial Twin test*

The partial twin test will only be reported in SFCHECK if the spacegroup allows and cumulative intensity distributions suggests twinning. For the diamonds, the partial twin test indicated a twin fraction of 0.363 (FIGURE 5.8A and FIGURE 5.8B, middle right panel.)

*L-function*

This test is more reliable at detecting twinning when it may be masked by other pathologies. The L-function for the hexagonal bipyramids did not track along the theoretical twinned or the theoretical untwinned line (FIGURE 5.9). Although not specifically mentioned in the algorithm documentation, this indicates TNCS and pseudocentering. The diamond data track between the theoretical twinned and the untwinned curv4es, indicatingf twinning (FIGURE 5.10).

*Native Patterson*

Translational pseudosymmetry can be detected by the native Patterson map calculated to 4 Å. Native Patterson maps calculated for both hexagonal bipyramids (FIGURE 5.11) and diamonds (FIGURE 5.12) showed no large, off origin peaks with heights greater than 10% of the origin.

*Parity Tests*

Pursuing the idea of pseudocentering for the hexagonal bipyramids further, the parity test is an indicator of pseudocentering (17). Parity tests were conducted on the d8x5 data using many different resolution cutoffs. The parity test ratios obtained never indicated pseudocentering, with all ratios at or above 0.9 (not shown).

*Self Rotation Functions*

Rotation pseudosymmetry can be detected in the self rotation function. No extra symmetry elements were detectable in the $\chi$ = 180$^o$ for hexagonal bipyramids (FIGURE 5.13) or diamonds (FIGURE 5.14).

*Anisotropy Detection*

Anisotropic data can cause refinements to stall (5). Data from the

hexagonal bipyramids have strong anisotropy. Diffraction drops off quickly in the a* and b* dimensions (FIGURE 5.15). The diamond crystal data are much less anisotropic, displaying mild anisotropy (FIGURE 5.16). Anisotropic corrections for data can be obtained through the web-based server. Refinements in CNS against the anisotropic-corrected data did help R-factor and $R_{free}$, although the split between the two was not reduced.

**Merohedral Twinning Refinements**

PHASER determined molecular replacement solutions in $P6_2$, $P3_221$, and $P32$ using d8x5_1_001_refmac12.pdb for the hexagonal bipyramids. Entering various twin fractions and twin operators (TABLE 5.6), least-squares merohedral twinning refinements were conducted using CNS (14). The statistical indicators revealed $R_{free}$ greater than or equal to 0.40. Using the diamond data and 104nofad.pdb as the search model, $P4_3$ was determined by PHASER to be the best spacegroup and found 4 molecules in the asymmetric unit. Merohedral twinning refinements were conducted in this spacegroup with operator *h, -k, -l* utilizing a twinning fraction of 0.4. Since these crystals indicated twinning it was anticipated that this would provide the drop in R-factor and $R_{free}$ expected. Unfortunately, the $R_{free}$ did not drop below 0.42.


**DISCUSSION**

**Exploration of BjPutA oxidation and reduction**

*B.japonicum PutA A310V Mutant*

BjPutA has a thermodynamic barrier to become fully reduced with L-

proline due to an extremely negative reduction potential of -132 mV (18). A mutant was found that could be fully reduced with L-proline BjPutA A310V. Based on homologous structures, the proposed location of this alanine to valine mutation occurs behind the C4 and C4a portion of the isoalloxazine (FIGURE 5.17). Most other PutA proteins have valine in this position. This mutant BjPutA was utilized for crystallization trials. Unfortunately this mutated BjPutA did not resolve the issues with statistical indicators, R and $R_{free}$, with value of these staying above 0.4.

*Radical Scavenging and Merging Wedges of Initially-Collected Data*

During data collection, it was observed that the distinct yellow color of the flavin-containing BjPutA crystals was fading. This change of color is similar to what happens upon reduction of FAD with L-proline or sodium dithionite (18). It was thought that perhaps the intense X-ray beam at the synchrotron was reducing the FAD cofactor. Examples reported in the literature of reduction of enzymes by X-rays include the photolyase from *A.nidulans* and horseradish peroxidase. These enzymes contain redox sensitive sites of deazaflavin and haeme (25,26).

Synchrotron radiation-induced damage to various proteins has been reported previously (27). There are two types of radiation damage, primary and secondary (28). Ionization of an atom through photo-electric absorption or Compton scattering comprises primary radiation damage. Secondary damage is caused by formation of secondary electrons, such as free-radicals, which cause further ionization events (28). Also, photoreactive molecules may have

superexcited states which result in relaxation with damage to the enzyme their contained in or reaction with molecular oxygen to create reactive oxygen species (26,29). Dose-rate effects of radiation damage have been studied and Leiros, et al. proposes that at high dose-rates, the damage may be related to steady state free-radicals generated upon X-ray exposure (30). The total free-radical population is composed of the formation, recombination and diffusion of free-radicals and at the highest dose rates recombination may not be fast enough leading to free-radical diffusion to reduce sensitive sites (30).

It has been reported that soaking crystals with free-radical scavengers reduces the ability of radiation to reduce sensitive sites (28). Examples of scavengers are ascorbate, 1, 4-benzoquinone, 2, 2, 6, 6,-tetramethyl-4-piperidone (TEMP), and dithiothreitol (FIGURE 5.18). The effectiveness of these scavengers was investigated and varies upon the type of system utilized, but ascorbate and quinone were effective on two reduction-sensitive models (disulfide-containing protein and lipoic acids) (28) (FIGURE 5.18).

Using hexagonal bipyramidal crystals of BjPutA soaked in 250 mM ascorbate, data were collected for $P6_2 22$ spacegroup. The best data sets from ALS 4.2.2 May 2006 were identified by low merging statistics and high resolution (TABLE 5.3). To catch the oxidized enzyme before changes from radiation damage and flavin reduction occurred, merging of the first $5^o$ of data from many crystals was attempted. TABLE 5.3 displays the data collection statistics for the merged datasets. TABLE 5.4 contains the R and $R_{free}$ values after CNS simulated annealing refinement with 104nofad.pdb. The drop in R and $R_{free}$ was

not large as anticipated with the values staying in the 0.40s.

*Proline & Dithionite soaks*

Since it was suspected that BjPutA changed conformation during data collection, the crystals were reduced with proline and sodium dithionite prior to data collection. If these crystals were no longer able to undergo conformational changes in the beam, it was suspected that data collection statistics and resulting maps would be better thus allowing better models to be built ultimately reflected in the statistical indicators, R and $R_{free}$. Once reduced, these crystals were looped and plunged into N2 (*l*) to preserve this reduced state.

Data collection on proline/dithionite reduced crystals seemed to result in better resolution and better density maps than similar oxidized crystals. Comparing maps provided insights into the quality of structure determination. These maps were compared by viewing the unmodeled flavin density and the β-strands behind the flavin isoalloxazine which seemed blurred in some maps. In the case of the proline/dithionite reduced hexagonal bipyramids, the flavin density looked better with the exception of the isoalloxazine ring for which density was weak  The blurring of the density for the strands was still evident although not as much perhaps due to higher resolution data. Unfortunately, the resulting R/$R_{free}$ statistics did not improve greatly as anticipated, with the R factor at 0.36 and $R_{free}$ at .42 using REFMAC 5 refinement with TLS.

**Spacegroup Investigation for the Hexagonal Bipyramidal Form**

Initial spacegroup determination was $P6_222$, but upon R/$R_{free}$ lagging with further model building, assignment of space group was revisited (APPENDIX 5.1

& 5.2). PHASER was utilized to test all spacegroups during molecular replacement (20). Hexagonal and trigonal spacegroups were used in refinements to see if drops in R/$R_{free}$ occurred (TABLE 5.5). One caveat to consider when making these comparisons is to include the number of atoms used in the refinement (indicated by mol/ASU in TABLE 5.5). The more atoms in the model, the more likely that artificial decreases in R-factor are seen. Another consideration is the completeness of the data utilized. The R-factor and $R_{free}$ can be artificially decreased by using less complete data, which can occur from attempting to decrease point group symmetry when only enough data was collected for the higher symmetry point group,. Data utilized in CNS refinements is indicated by percent of total reflections included in the refinement (indicated by % reflections in TABLE 5.5). R/$R_{free}$ statistics were only decreasing slightly, most likely to the increasing number of atoms in the asymmetric unit. To no avail, this investigation did not shed light on an alternative space group.

**Examination of Intensity Distributions, Anisotropy and Twinning**

*Hexagonal Bipyramids*

After much crystal manipulation, literature review of various pathological cases proved fruitful. Examining at the various pathology indicators described above, provided new insights about BjPutA hexagonal bipyramidal crystals.

Cumulative intensity distributions N(z) plots created with CCP4 did not show the signature sigmoidal curves of a twinned crystal (FIGURE 5.5) (13). Although, on closer inspection it revealed that the acentric reflections had a very centric character, with the observed acentric reflections tracking just above of the

centric theoretical reflections. According to Dauter, this is indicative of a crystal with translational non-crystallographic symmetry, whether it is twinned or untwinned (1). Using SFCHECK to analyze the data, the second moment of Z displayed a much higher value than 2 (FIGURE 5.7B, last panel) (7). Although Dauter would argue that for proper analysis only lower resolution data (less than 3.5 Å) should be utilized for the perfect twin test, cutting off data at 3.5 Å still gave an average second moment of Z higher than 2 (not shown). These two tests are indicating that perhaps we have translational non-crystallographic symmetry.

To investigate further this idea of translational non-crystallographic symmetry, a native Patterson map with data truncated to 4 Å was generated using CCP4i (FIGURE 5.11A) (13). As stated earlier, if TNCS is present, a peak in the map with 15-20 % of the origin peak height will appear. Peaks with this height were never encountered upon calculation of the native Patterson (FIGURE 5.11 A & 5.12 B), although some smaller peaks were present in the native Patterson however. SFCHECK confirms the lack of TNCS in the output report stating that there is no translational pseudosymmetry detected (FIGURE 5.7A).

The L-function, which has been shown to be a robust way to analyze intensities, displayed an unusual feature not described by the Padilla and Yeates paper (12). Analyzing our moderate resolution data and with the L-function server showed our data tracking below the line for untwinned data (FIGURE 5.9). This result most likely indicates pseudocentering.

The possibility that the crystal is pseudocentered was investigated. From

$P6_222$, symmetry allows for C-centered orthorhombic to be a potential space group. The indexing of d8x5 from ALS May 2006 is shown in APPENDIX 5.2. This shows a least squares residual of 0.094 for selection of a centered orthorhombic space group as compared to 0.156 for P622. Although perfect twin tests, the L-test and cumulative intensity distributions indicated pseudocentering, parity tests did not indicate pseudocentering (not shown). Refinements were conducted against 2.6 Å C222 d*TREK-processed data (APPENDIX 4). With only 83.8% of the data used for CNS simulated-annealing refinements, statistical indicators did not drop significantly, with R-factor = 0.36 and $R_{free}$ = 0.42.

Rotational pseudosymmetry was examined using a resolution cut-off of 4 Å using MOLREP (19). The $P6_222$ d8x5 ALS September data $\chi = 180$ section of the self rotation function shows a 6-fold symmetry (FIGURE 5.13A). When processed in P3 (FIGURE 5.13B, the $\chi = 180$ still shows a 6-fold symmetry, although the self rotation function looks a little different but the features are not strong. These maps indicate no rotational pseudosymmetry.

Anisotropy, which can cause refinement to stall, was also investigated. Indeed d8x5 showed strong anisotropy as indicated by the Diffraction Anisotropy Server (5). FIGURE 5.15A and FIGURE 5.15B displays the results from analysis of our data for anisotropy. FIGURE 5.15A shows much higher diffraction in the c* direction and diffraction in the a* and b* directions. Ellipsoidal truncation of the data to 2.7 Å was applied (FIGURE 5.15B) and used in refinement. Although we did see better statistics, truncation did not significantly improve our refinement values.

Twinning was investigated utilizing CNS least-squares twin simulated annealing refinement. Various space groups with merohedral twin operators and various twin fractions were used. Using $P6_2$, $P3_221$, $P3_2$ space groups, refinement taking twinning into account dropped the R-factor to 0.30-0.32 depending on space group and twin fraction, but the $R_{free}$ never dropped below 0..40 (TABLE 5.6).

*Diamonds*

Diamond-shaped crystals would sometime occur in the same crystallization drops as the hexagonal bipyramidal crystals. The best diffraction obtained with these crystals was 2.7 Å. Originally, the Laue group was thought to be 4/mmm and data were collected accordingly at the Advanced Light Source, beamline 4.2.2. Indexing of these crystals is shown in APPENDIX 5.3. Further investigation of systematic absences with dtcell showed the most likely spacegroup was $P4_32_12$ (24). Molecular replacement provided a suitable solution, although upon refinement, the $R/R_{free}$ still remained high.

After looking at cumulative intensity distributions, the observed acentric reflections showed a sigmoidal curvature, indicating twinning (FIGURE 5.6). The perfect twinning test $< I^2 > / < I >^2$ was ~1.7 with this being displayed in the SFCHECK output (FIGURE 5.8B last panel). Utilizing the L-function analysis to analyze local intensities, the d3_04_2 data tracked between the theoretically untwinned and perfectly twinned curves (FIGURE 5.10) indicating twinning. Peak heights with 15-20% of the origin peak height were not detected in 4 Å native Patterson map (FIGURE 5.12). Pseudo translations were not detected by

SFCHECK (FIGURE 5.8A). Self-rotation functions of data processed in P4 indicated a four-fold symmetry at $\chi$ = 180$^o$ (FIGURE 5.14), although no other strong rotational pseudosymmetry was seen. Anisotropy was checked and these crystals had mild anisotropy (FIGURE 5.16A). Slight ellipsoidal truncation was not performed since refinement programs can handle mild anisotropy (FIGURE 5.16B).

Pursuing the idea of twinning further, least-squares merohedral twinning refinements were conducted. The Yeates S(H) plot and the partial twinning test estimated a partial twin fraction of 0.36-0.4 (2) (FIGURE 5.8A & 9B middle right panel). Upon reindexing the data choosing P4 with unit cell dimensions of a=138.17 Å c= 267.3 Å, PHASER was utilized to find a proper molecular replacement solution, with the best log-likelihood gain occurring in P4$_3$ (20). Although twinning was suspected, least-squares twinning refinement utilizing the P4 merohedral twinning operator and calculated twin fraction 0.4 resulted in 81% data used in refinement, with a R = 0.27 and R$_{free}$ = 0.40.

**Pseudomerohedral twinning**

Interestingly, Parsons discusses pseudomerohedrally twinned crystals that emulate tetragonal symmetry which are actually orthorhombic ($\beta$~90$^o$) (4). These are related by a two-fold rotation around the longer dimension's (usually c*) axis or around the diagonal along the a*-b* axis.

Parsons also discusses a similar case for pseudomerohedrally twinned monoclinic crystals that emulate hexagonal crystals ($\beta$~120$^o$), although there are more than two twin domains that occur (4). A three-domain twinning case was

recently described for the aclacinomycin oxidoreductases where they utilized a three-domain twin operator to describe the symmetry (31).

Studies of BjPutA and the d3_04_2 datasets are currently being investigated for the occurrence of pseudomerohedral twinning. Since the diamond crystals may be emulating a tetragonal spacegroup, it is suspected that the twin operator will be a two-fold around the c* axis described by $-k, h, l$ or a diagonal in the a*-b* plane $k, h, -l$. Current difficulties include finding datasets to merge together to obtain a complete dataset to perform refinements. Only two programs are currently available to handle refinements of data with pseudomerohedral twinning; SHELX and PHENIX (15,32).

A new program from the PHENIX program suite (PHENIX.xtriage) was also used to analyze data collected from the diamond-shaped crystals (apparent tetragonal form). This analysis was performed on a d3_04 data set that had been processed in $P2_12_12_1$. The data set was 90 % complete to 2.8 Å. The analysis indicated pseudomerohedral twinning with twin operator $- k, - h, - l$. Thus, refinement assuming pseudomerohedral twinning was investigated.

Using a search model created from d8x5_1_001_refmac12.pdb (04.pdb obtained from JJ Tanner), a suitable molecular replacement solution was identified using MOLREP with 4 molecules in the asymmetric unit. PHENIX.refine was used for refinement since it is the only program currently available that properly treats pseudomerohedral twinning at moderate resolution. The least squares twinning refinements previously described by Herbst-Irmer and Sheldrick (15) has been implemented in PHENIX.refine. (We note that SHELX

also treats pseudomerohedral twinning but it is not recommended at moderate resolution.)

Non-crystallographic symmetry restraints were utilized for refinements with or without twinning.   The untwinned refinement resulted in R = 0.4093 and Rfree = 0.4491.  The analogous refinement taking twinning into account resulted in significantly lower R-factors of R = 0.3089 and Rfree = 0.3554.  After 7 rounds of PHENIX least squares twin refinement the current model has 844 residues and R = 0.2691 and Rfree = 0.3132.  These twinning refinements have revealed previously unobserved density for the N-terminal domain which in previously modeled structures had been incorrectly modeled or not modeled at all.

**Strategies for Structure Completion**

The current crystal forms of BjPutA have an inherent pathology which could be described by pseudomerohedral twinning and current refinements have lowered both R-factor and $R_{free}$ substantially with anticipation of structure completion. Other strategies that are currently being pursued are finding a new crystal form of BjPutA and homolog screening.

One way other crystal forms could be obtained is by cleaving off the TEV-protease cleavable 8x-Histidine affinity tag.  Removal of the histidine tag may allow the enzyme to pack differently in the crystal lattices resulting in a new crystal form.  Another strategy is lysine methylation, which has been shown to be an effective strategy for obtaining crystals (33). Lysine methylation targets the primary amine of surface lysines and converts the primary amine to a tertiary amine by methylation.   Lysine methylation is a better alternative to surface

mutagenesis since the modifications only happen on the surface of a folded protein. The result is a change in the overall surface charge, and disruption of salt bridges formed on the surface. As discussed in Chapter 3, *N*-propargylglycine has been shown to inhibit PutA (D Srivastava and JD Larson, unpublished results). Inactivation of BjPutA may be an effective way to obtain the reduced crystal form without reoxidation and may provide a new crystal form.

Homolog screening is yet another method for obtaining a bifunctional PutA structure. *Legionella pneumophila* and *Pseudomonas aeruginosa* are two bifunctional PutA enzymes that are cloned and show promise with expression of monodisperse, soluble protein. Moreover, small crystals of LpPutA have been obtained (JD Larson and D Srivastava, unpublished results). LpPutA also has a TEV-protease cleavable 6x-Histidine tag and cleavage could aid in crystallization trials. There are more PutA homologs to be examined, some of which include *Shewanella* sp., *Wolbachia* sp., *Xanthomonas* sp., and *Rhodopseudomonas* sp.

Improvements in detecting crystal pathology during data collection and refinement should aid in structure determination. Outlined in this chapter are the various pathologies that have been detected in the literature, methods for their remedy and application to pathological crystals of BjPutA.

## REFERENCES

1. Dauter, Z. (2003) *Acta Crystallogr D Biol Crystallogr* **59**, 2004-2016
2. Yeates, T. O. (1997) *Methods Enzymol* **276**, 344-358
3. Lebedev, A. A., Vagin, A. A., and Murshudov, G. N. (2006) *Acta Crystallogr D Biol Crystallogr* **62**, 83-95
4. Parsons, S. (2003) *Acta Crystallogr D Biol Crystallogr* **59**, 1995-2003
5. Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006) *Proc Natl Acad Sci U S A* **103**, 8060-8065
6. Howells, E., DC, P., and D, R. (1950) *Acta Crystallography D* **3**, 210-214
7. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) *Acta Crystallogr D Biol Crystallogr* **55**, 191-205
8. Britton, D. (1972) *Acta Crystallography A* **28**
9. Murray Rust, P. (1973) *Acta Crystallogr B* **29**, 2559-2566
10. Yeates, T. O. (1988) *Acta Crystallography A* **44**, 142-144
11. Rees, D. C. (1980) *Acta Crystallographica A* **36**, 578-581
12. Padilla, J. E., and Yeates, T. O. (2003) *Acta Crystallogr D Biol Crystallogr* **59**, 1124-1130
13. 4, C. C. P. N. (1994) *Acta Crystallography* **D50**, 760-763
14. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54**, 905-921
15. Herbst-Irmer, R., and Sheldrick, G. M. (2002) *Acta Crystallogr B* **58**, 477-481
16. Tong, L., and Rossmann, M. G. (1997) *Methods Enzymol* **276**, 594-611
17. Kleywegt, G. J., and Jones, T. A. (1996) *Acta Crystallogr D Biol Crystallogr* **52**, 826-828
18. Krishnan, N., and Becker, D. F. (2005) *Biochemistry* **44**, 9130-9139
19. Vagin, A., and Teplyakov, A. (1997) *Journal of Applied Crystallography* **30**, 1022-1025
20. Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 432-438
21. Leslie, A. G. (2006) *Acta Crystallogr D Biol Crystallogr* **62**, 48-57
22. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr D Biol Crystallogr* **53**, 240-255
23. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
24. Pflugrath, J. (1999) *Acta Crystallography* **D55**, 1718-1725
25. Berglund, G. I., Carlsson, G. H., Smith, A. T., Szoke, H., Henriksen, A., and Hajdu, J. (2002) *Nature* **417**, 463-468
26. Kort, R., Komori, H., Adachi, S., Miki, K., and Eker, A. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 1205-1213
27. Banumathi, S., Zwart, P. H., Ramagopal, U. A., Dauter, M., and Dauter, Z. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 1085-1093

28.    Southworth-Davies, R. J., and Garman, E. F. (2007) *J Synchrotron Radiat* **14**, 73-83

29.    Wondrak, G. T., Jacobson, M. K., and Jacobson, E. L. (2005) *J Pharmacol Exp Ther* **312**, 482-491

30.    Leiros, H. K., Timmins, J., Ravelli, R. B., and McSweeney, S. M. (2006) *Acta Crystallogr D Biol Crystallogr* **62**, 125-132

31.    Sultana, A., Alexeev, I., Kursula, I., Mantsala, P., Niemi, J., and Schneider, G. (2007) *Acta Crystallogr D Biol Crystallogr* **63**, 149-159

32.    Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 1948-1954

33.    Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I., and Grimes, J. M. (2006) *Structure* **14**, 1617-1622

FIGURE 5.1. An example of a twinned diffraction pattern from a hexagonal crystal in the P6 space group. The two upper figures represent different orientations of diffraction patterns from a P6 spacegroup, but the patterns are oriented within the hexagonal P622 symmetry. They are related by a two-fold rotation around the a* axis. In the upper figures, the selected reflections have varying intensities when in P6. In a perfect twin (bottom), the two upper orientations superimpose, the diffraction pattern aquires a P622 symmery and the intensities average together. Taken from Dauter "Twinned Crystals and anomalous phasing" 2003 (1).
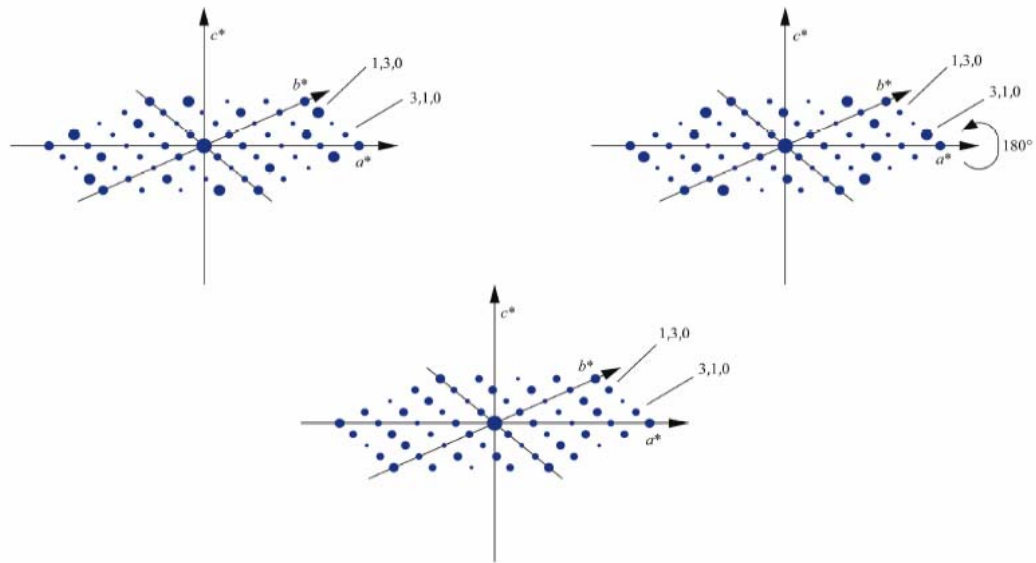
FIGURE 5.2. Example of expected cumulative intensity distribution for an untwinned crystal (upper exponential curve) and purely twinned crystal (sigmoidal curve). Dotted lines represent pathological crystals from gpD (blue dots) and IL-β (red dots). Taken from Dauter "Twinned Crystals and anomalous phasing" 2003(1).
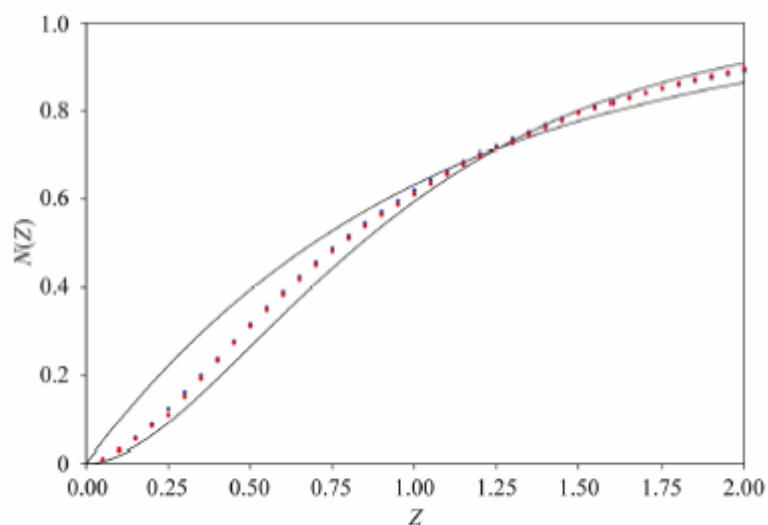
FIGURE 5.3.  Example of the Perfect Twin Test and sample data (taken from Dauter, 2003).  The green curve indicates a untwinned data and the blue and red curves indicate twinning (1).
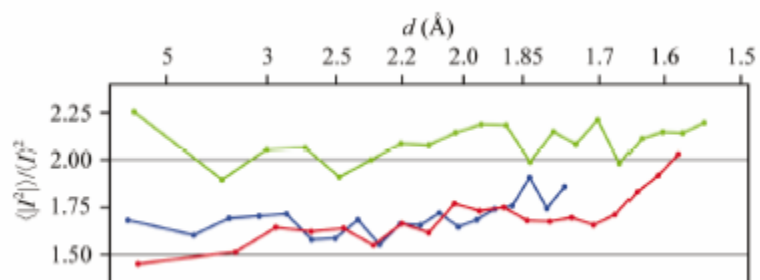
FIGURE 5.4.  Redution of BjPutA with 25 mM sodium dithonite.  A) BjPutA fully oxidized.  B) BjPutA after addition of sodium dithionite.  C)  Fully reduced BjPutA.
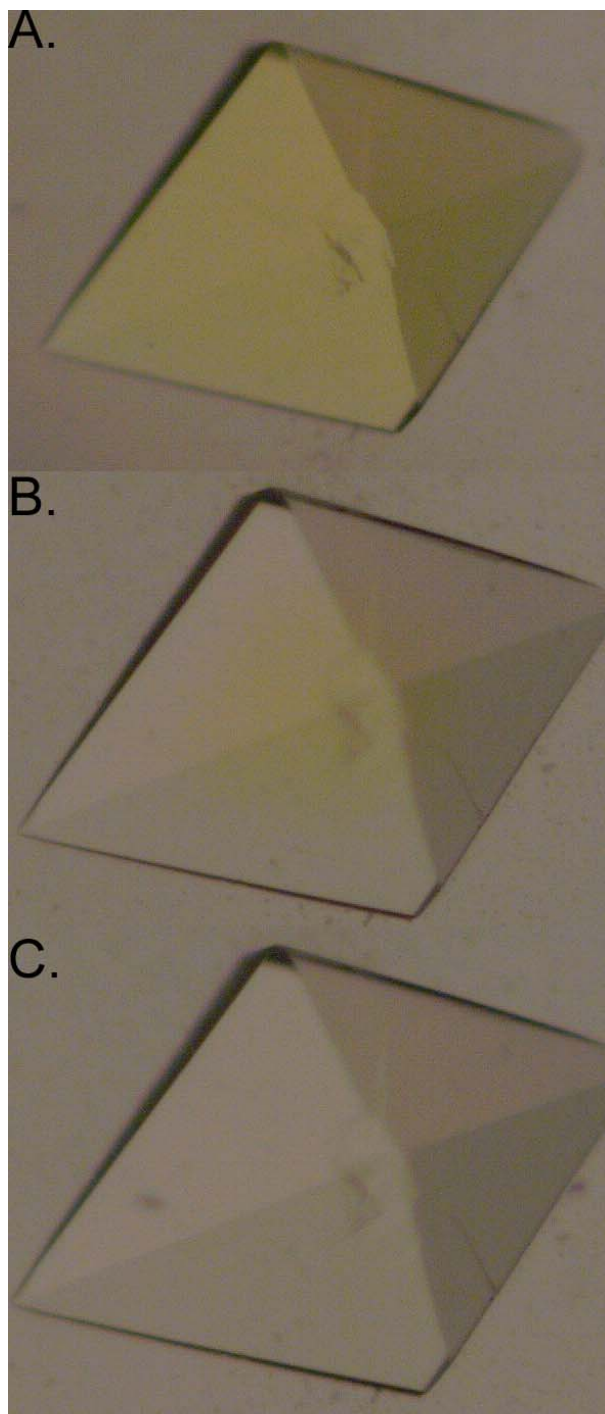
FIGURE 5.5. Cumulative intensity distribution created upon import of scaled data into CCP4 using MOSFLM-processed ALS May 2006 d8x5 to 2.6 Å. Green and red lines track normal theoretical intensity distributions for centric and acentric reflections, respectively. Black and blue lines track our experimentally determined intensity distributions for centric and acentric reflections, respectively.
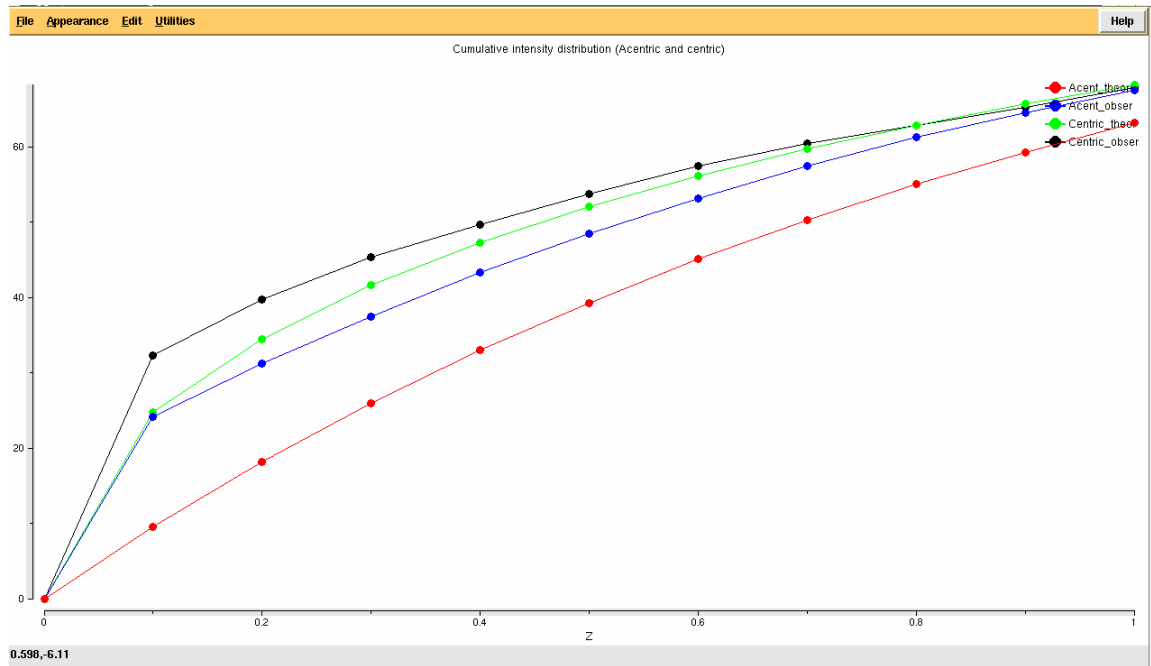
FIGURE 5.6.  Cumulative intensity distribution created upon import of scaled data into CCP4.  Using d*TREK-processed ALS May 2006 d3_04_2 to 2.8 Å. Green and red lines track normal theoretical intensity distributions for centric and acentric reflections, respectively.  Black and blue lines track our experimentally determined intensity distributions for centric and acentric reflections, respectively.
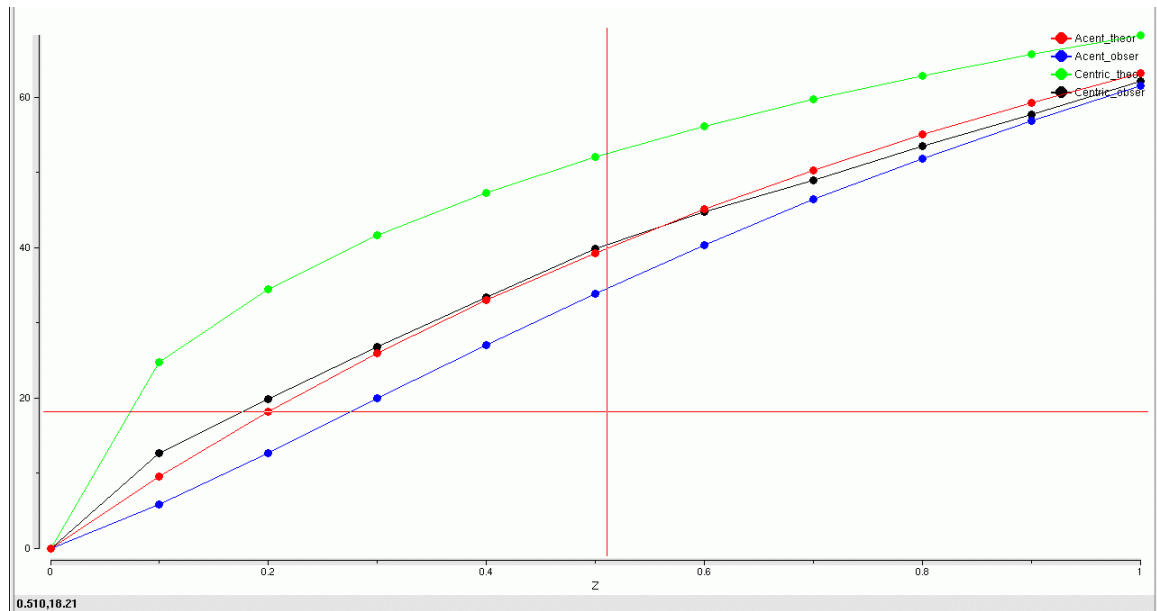
FIGURE 5.7A. SFCHECK output for ALS September 2006 d8x5 MOSFLM-processed data to 2.6 Å showing the data statistics, pseudotranslations, Wilson Plot, completeness and Rstandard, optical resolution, coordinate errors and the perfect twinning test.



"Judge not, that you be not judged" Bible, Matthew, ch.7

# Structure Factor Check
## XXXX

Title:
Date:
PDB code:   XXXX

### Crystal

Cell parameters:
  a: 145.25 Å     b: 145.25 Å     c: 185.42 Å
  α: 90.00°      β: 90.00°       γ: 120.00°

Space group: P 62 2 2

### Structure Factors

Input
Nominal resolution range:           43.50 − 2.65 Å
Reflections in file:                          33744
Unique reflections  above 0:               33744
                    above 1 σ:              33077
                    above 3 σ:              21143

SFCHECK
Nominal resolution range:           43.50 − 2.65 Å
    (max from input data, min from author)
Used reflections:                            33744
Completeness:                               98.8 %
R_stand(F) = <σ(F)>/<F> :              0.052
Anisotropic distribution of Structure Factors
    ratio of eigen values:  1.0000 1.0000 0.4855
B_overall (by Patterson):              45.5 Å²
Optical resolution:                       2.02 Å
Expected opt. resol. for complete data set: 2.02 Å
Estimated minimal error:               0.052 Å
Pseudo-translation is not detected

256

FIGURE 5.7B. Continued SFCHECK output for ALS September 2006 d8x5 MOSFLM-processed data to 2.6 Å showing the data statistics, pseudotranslations, Wilson Plot, completeness and $R_{standard}$, optical resolution, coordinate errors and the perfect twinning test.
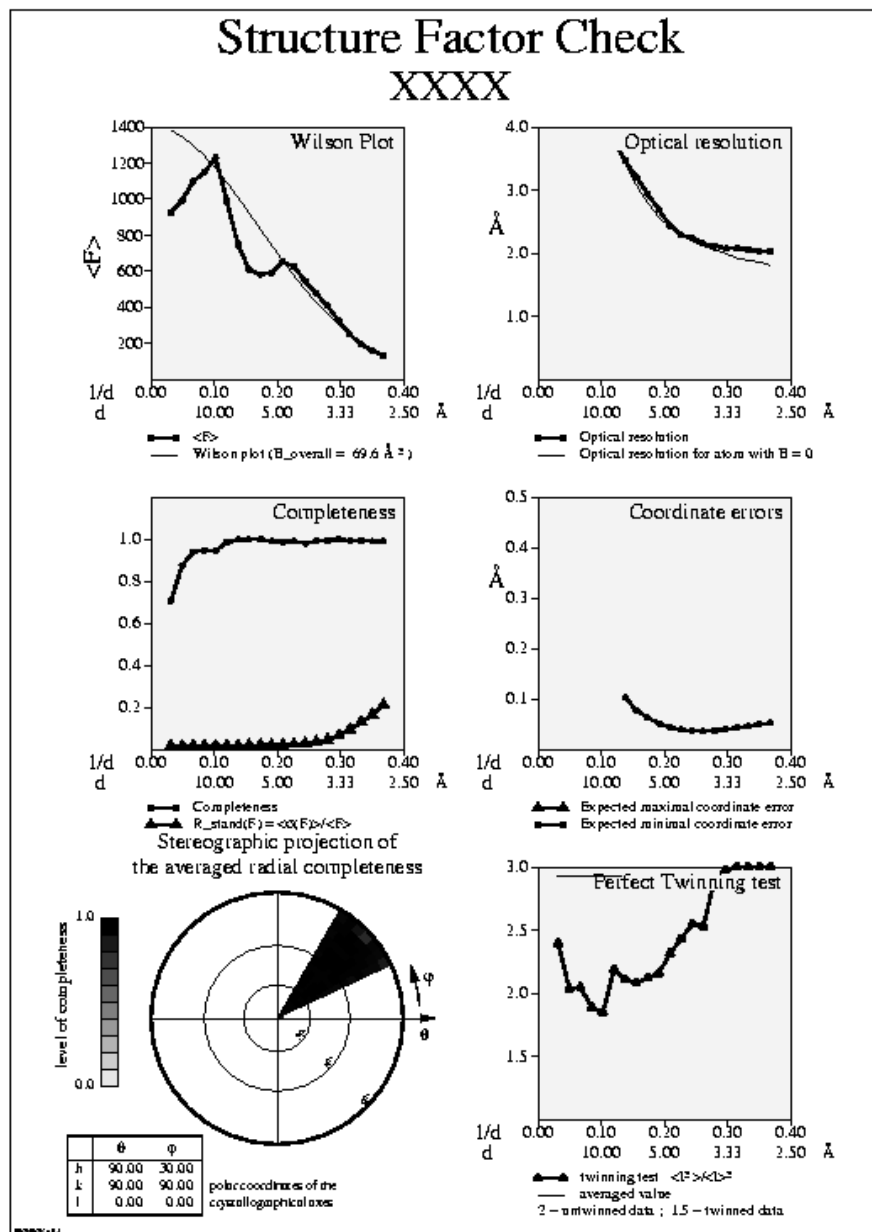
FIGURE 5.8A.  Output from SFCHECK utilizing d*TREK-processed ALS May 2006 d3_04_2 data.

FIGURE 5.8B. Output from SFCHECK utilizing d*TREK-processed ALS May 2006 d3_04_2 data.
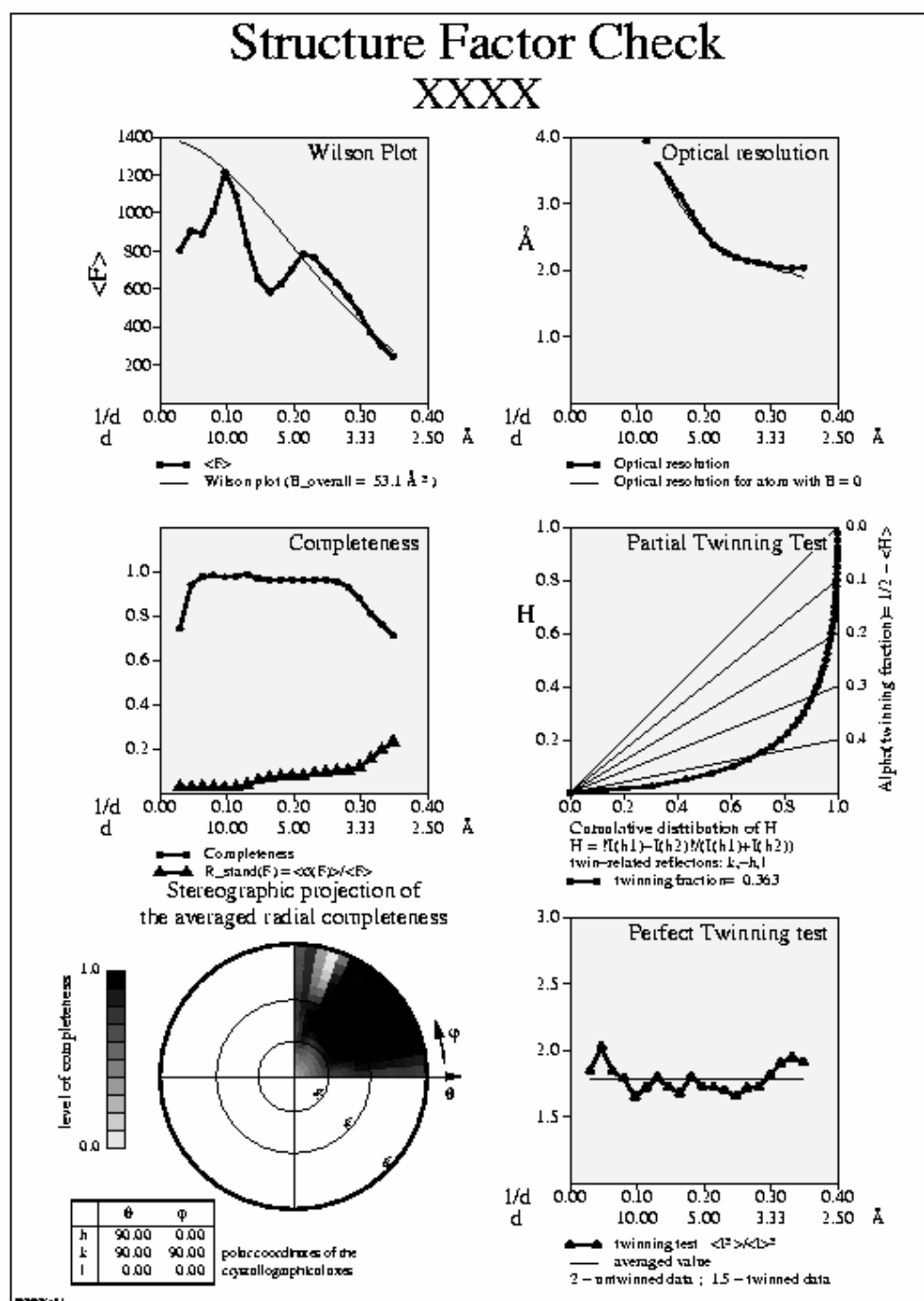
FIGURE 5.9.  L-function output of ALS May 2006 d8x5 truncated to 8 - 3.5 Å shown in blue.  The red line displays results of L-function on untwinned crystals and upper red curve is the result for perfectly twinned crystals.



Δ,  Created by DATAMAN at Mon Dec 11 12:06:13 2006 for an UNKNOWN user *** XPS_GRAF 1998 12 16/3.1.2 *** (C) GJ

File /usr/httpd/html/Services/pystats/upload/twin9242.ps

X = |L|

Y = N(|L|) acentrics

DATAMAN Local Intensity Statistics plot
Cumulative N(|L|) vs. |L| (acentrics)
Dataset M1 File /usr/httpd/html/Services/pystats/upload/9242_file.x
Comment Read from /usr/httpd/html/Services/pystats/upload/9242_file.x
<|L|> = 0.545 Untwinned = 0.500 Perfectly twinned = 0.375
<L^2> = 0.389 Untwinned = 0.333 Perfectly twinned = 0.200
See: JE Padilla & TO Yeates, Acta Cryst D59, 1124 (2003).
Red line = theoretical untwinned
Red curve = theoretical perfectly twinned
Blue curve + points = observed

FIGURE 5.10.  Using the L-function to analyze proximal reflection intensities of ALS May 2006 d3_04_2 data truncated from 8 - 3.5 Å shown in blue.  The linear red line displays results of L-function on untwinned crystals and the upper red curve is the result for perfectly twinned crystals.

Δ.  Created by DATAMAN at Wed Feb 28 18:45:40 2007 for an UNKNOWN user *** XPS_GRAF 1998 12 16/3.1.2  *** (C) GJ
File /usr/httpd/html/Services/pystats/upload/twin1479.ps

X = |L|

Y = N(|L|) acentrics

DATAMAN Local Intensity Statistics plot
Cumulative N(|L|) vs. |L| (acentrics)
Dataset M1 File /usr/httpd/html/Services/pystats/upload/1479_file.x
Comment Read from /usr/httpd/html/Services/pystats/upload/1479_file.x
<|L|> = 0.432 Untwinned = 0.500 Perfectly twinned = 0.375
<L^2> = 0.257 Untwinned = 0.333 Perfectly twinned = 0.200
See: JE Padilla & TO Yeates, Acta Cryst D59, 1124 (2003).
Red line = theoretical untwinned
Red curve = theoretical perfectly twinned
Blue curve + points = observed

FIGURE 5.11 A.  Native Patterson results from the ALS May 2006 d8x5 dataset truncated to 4 Å resolution.  These native pattersons were calculated for data processed in P6$_2$22.  Shown here are the x vs w and the u vs w Patterson maps.
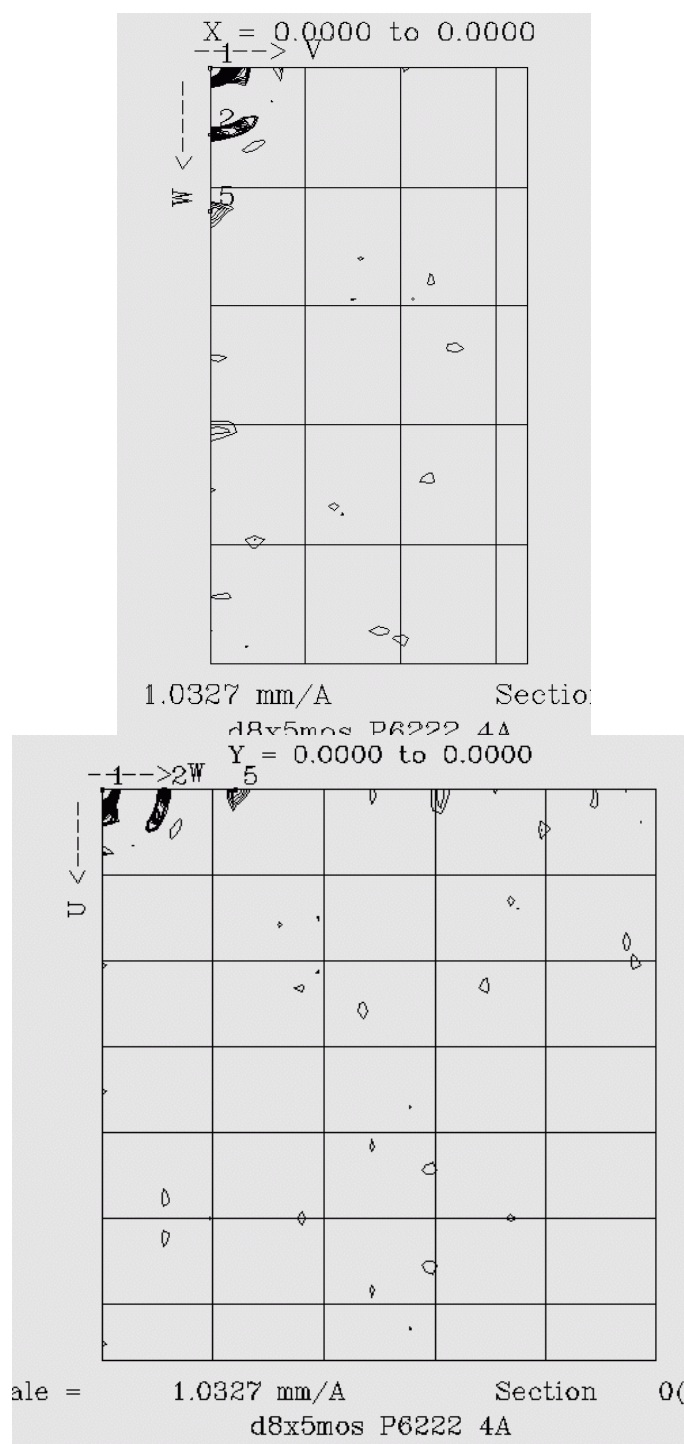
FIGURE 5.11B.  Native Patterson results from the ALS May 2006 d8x5 dataset truncated to 4 Å resolution.  These native pattersons were calculated for data processed in P6$_2$22.  Shown here are two u vs v Patterson at Z = 0 and Z = 0.3.
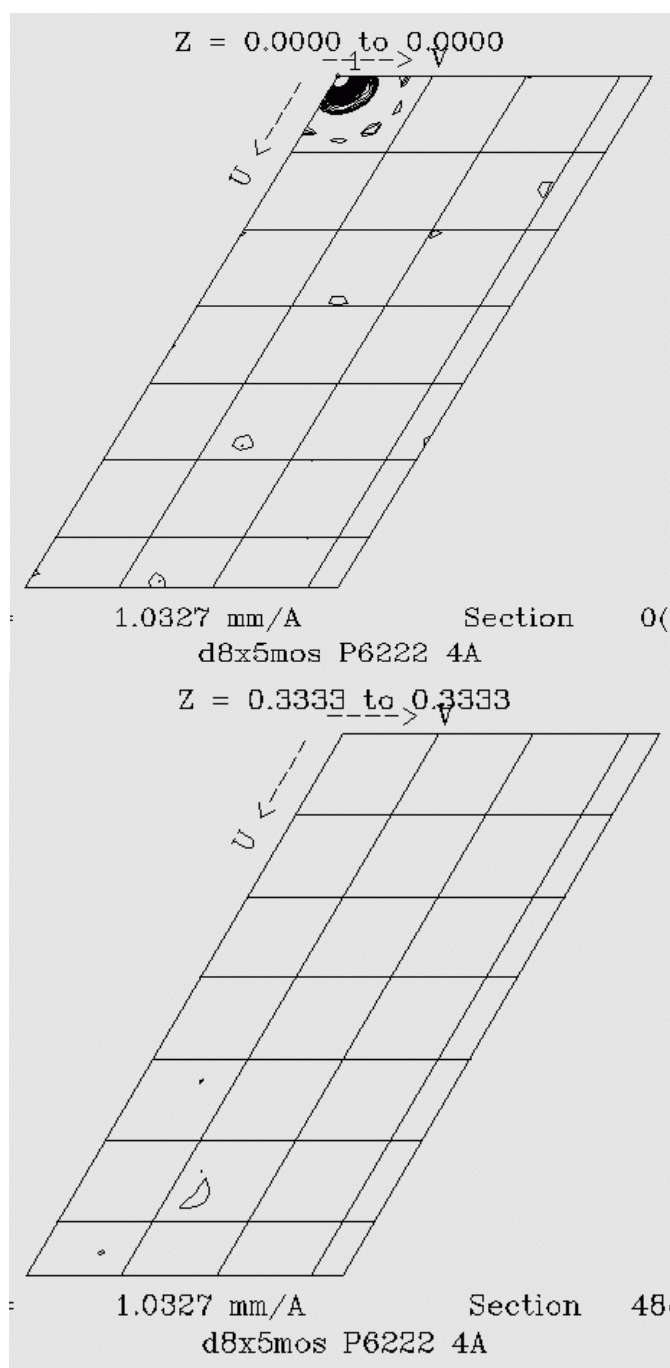
FIGURE 5.12.  Native Patterson calculated to 4 Å using ALS May 2006 d3_04_2
d*TREK processed data.
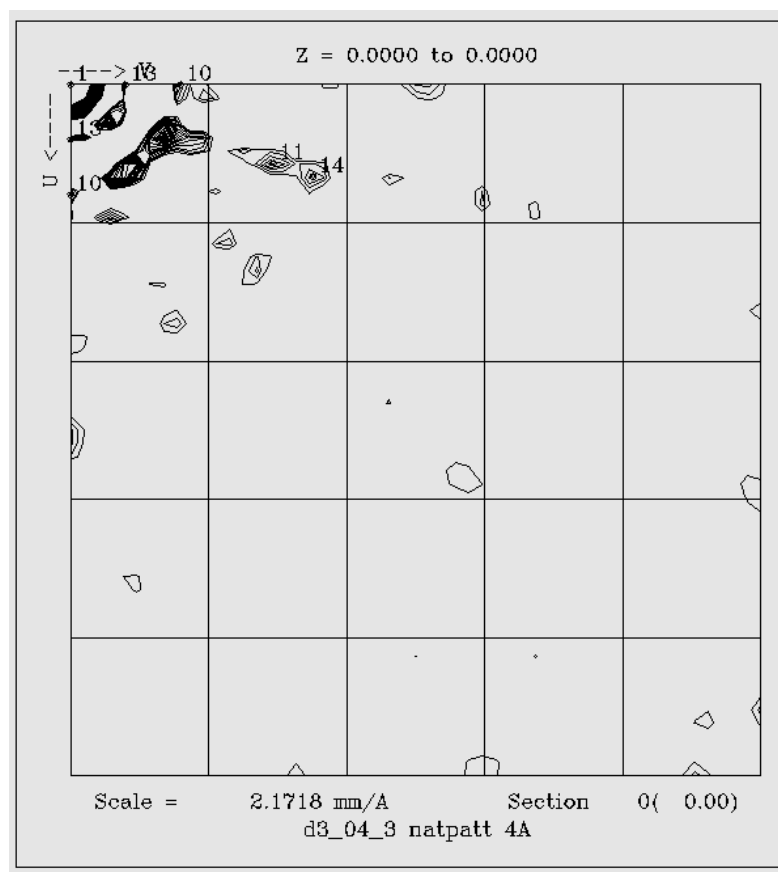
FIGURE 5.13A.  Self Rotation function of P6$_2$22 MOSFLM processed d8x5 ALS September 2006 data
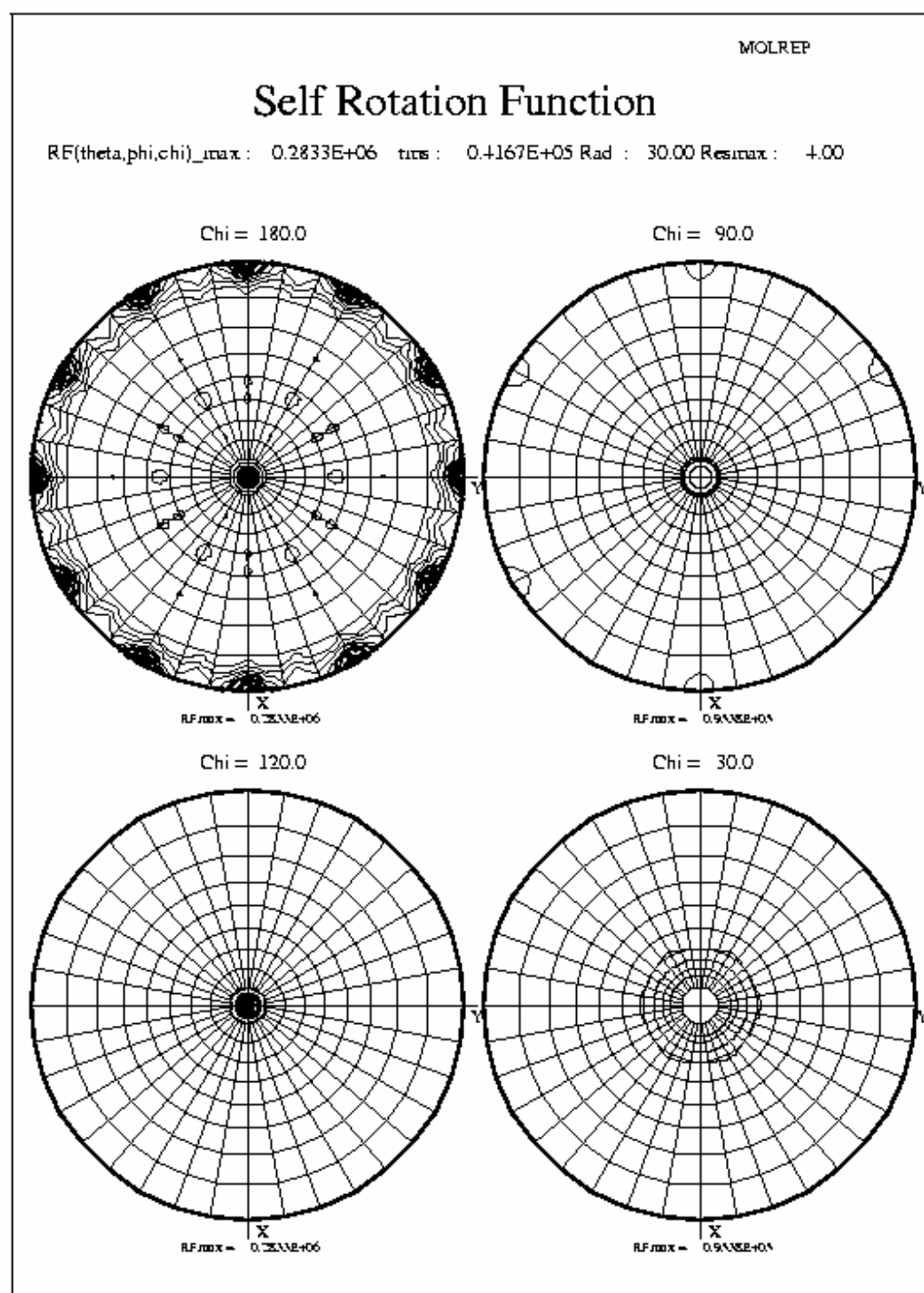
FIGURE 5.13B. Self Rotation function of P3 MOSFLM processed d8x5 ALS September 2006 data
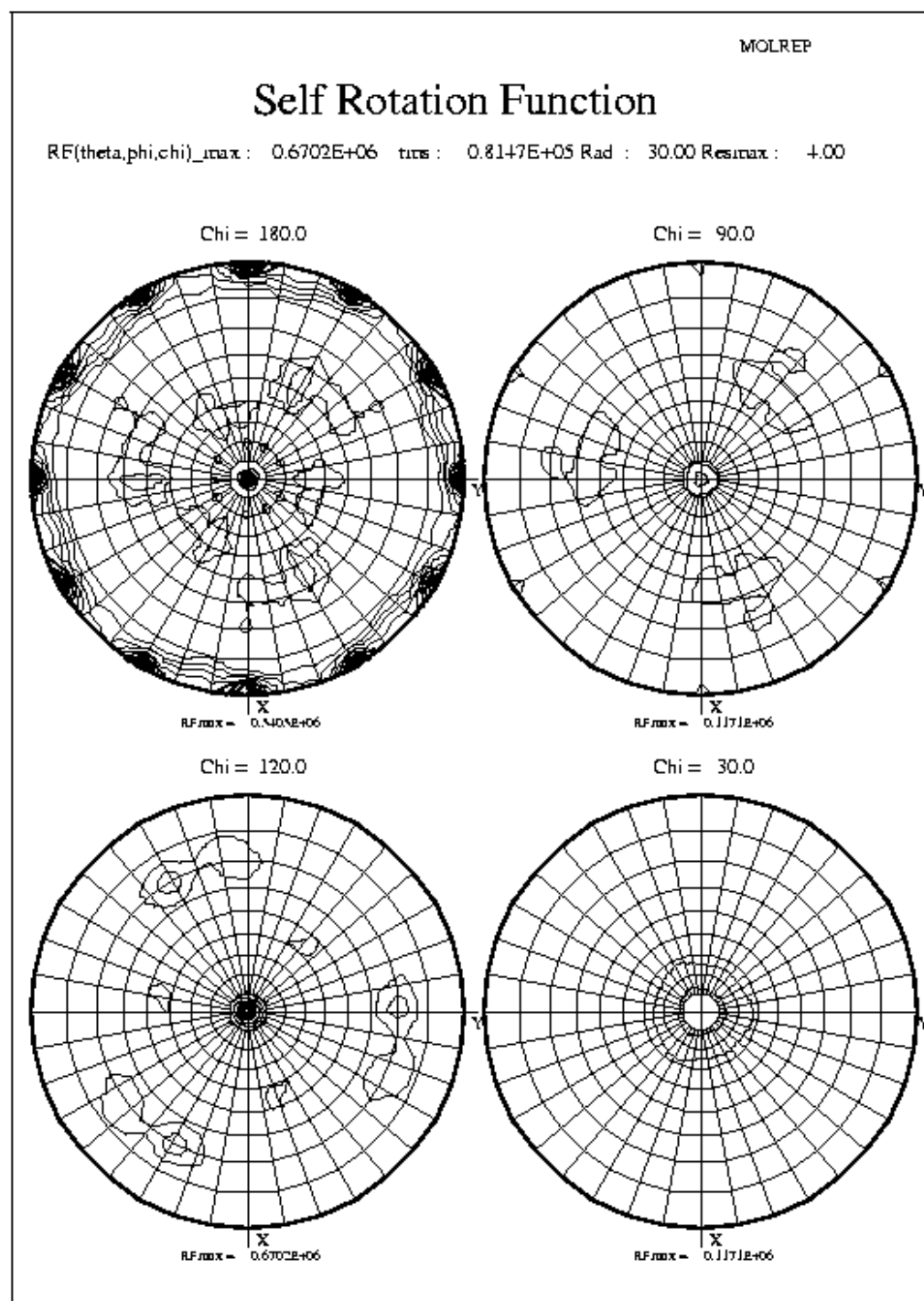
FIGURE 5.14.  Self rotation function calculated for P4 d*TREK processed
d3_04_2 ALS May 2006
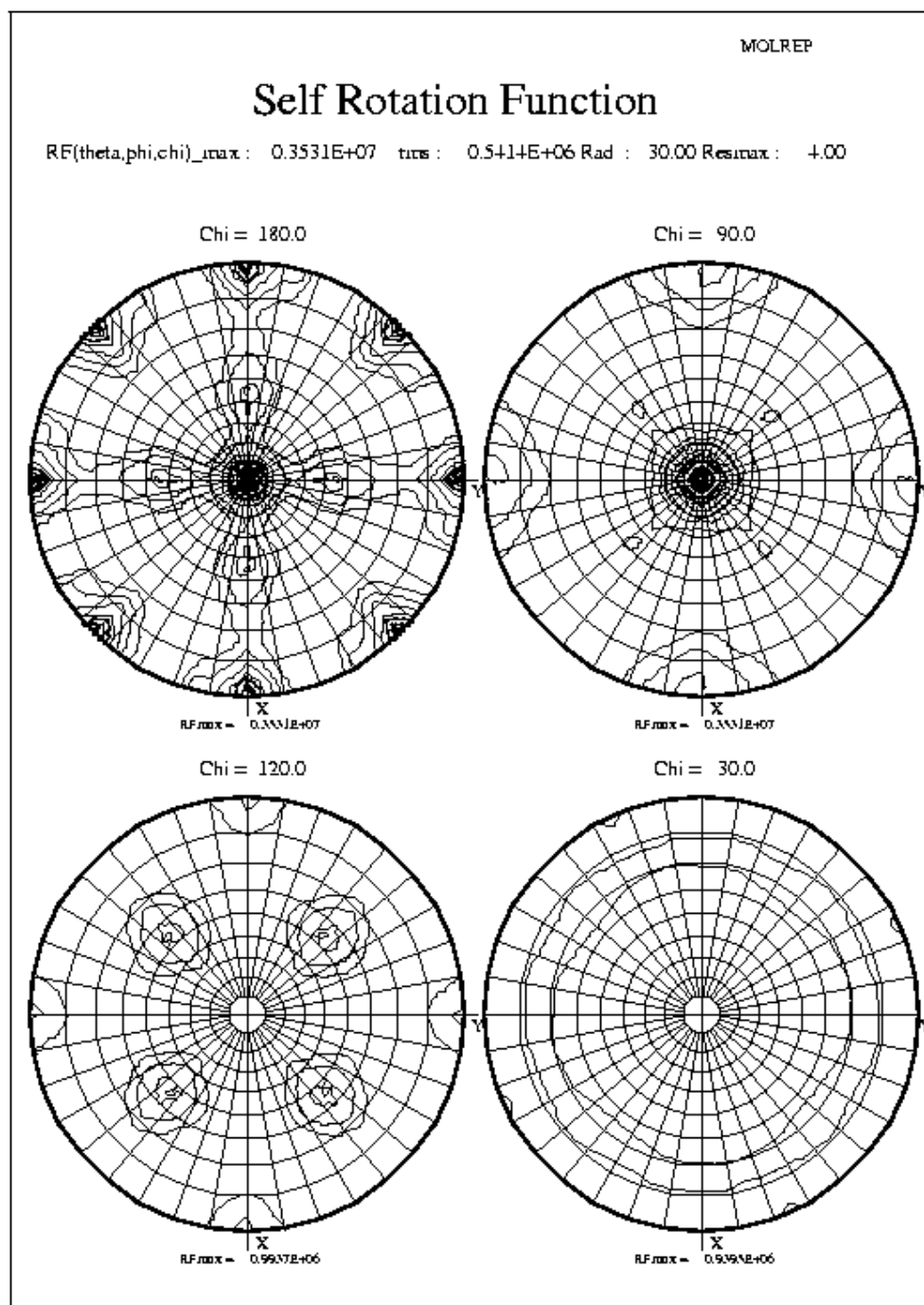
FIGURE 5.15A.  Results from the Diffraction Anisotropy Server using d8x5 from ALS May 2006.

# FIGURE 5.15B. Results from the Diffraction Anisotropy Server using d8x5 from ALS May 2006. Ellipsoidal Data Truncation to remove weakly diffracting spots.

35707 reflections were in the initial data set. 1744 were discarded because they fell outside the specified ellipsoid with dimensions 1/2.7, 1/2.7, 1/2.6 $\text{Å}^{-1}$ along a*,b*,c*, respectively. These discarded reflections had an average F/sigma of 3.67.

33963 reflections remain after ellipsoidal truncation. Anisotropic scale factors were then applied to remove anisotropy from the data set. Lastly, an isotropic B of -12.58 $\text{Å}^2$ was applied to restore the magnitude of the high resolution reflections diminished by anisotropic scaling. The following pseudo precession images illustrate the individual steps.
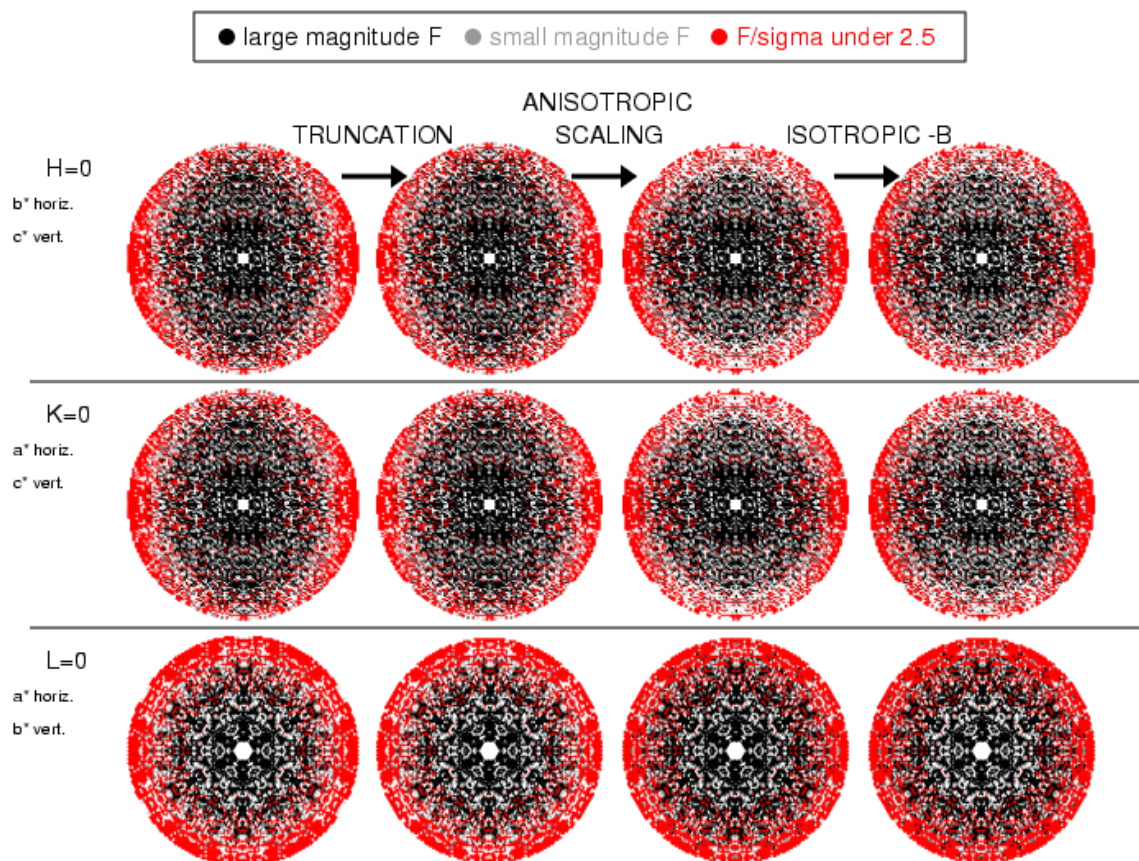
FIGURE 5.16A. Results from the Anisotropy Diffraction server for ALS May 2006 d*TREK-processed d3_04_2 data.

Your data has MILD anisotropy
based on the spread in values of the
three principle components= 24.59 A**2

The principle components are the exponential
scale factors used to correct for anisotropy.
They may be regarded as B factors applied to
the three principle directions of the data set.
Larger |values| indicate stronger anisotropy.

In decreasing order the 3 components are:
9.21    6.17   -15.38 A**2

a* direction
b* direction
c* direction

F/sigma

100 A**2
75 A**2
50 A**2
25 A**2
0 A**2

3.0

8.5  6.0  4.9  4.3  3.8  3.5  3.2  3.0  2.8  2.7
resolution

The recommended resolution limits along a*,b*,c* are
2.7 Ang    2.7 Ang    2.7 Ang
These are the resolutions at which F/sigma
drops below an arbitrary cutoff of 3.0

# FIGURE 5.16B. Ellipsoidal truncation results from ALS May 2006 d*TREK-processed d3_04_2 data.

171043 reflections were in the initial data set. 0 were discarded because they fell outside the specified ellipsoid with dimensions 1/2.7, 1/2.7, 1/2.7 $\text{Å}^{-1}$ along a*,b*,c*, respectively. These discarded reflections had an average F/sigma of 0.00.

171043 reflections remain after ellipsoidal truncation. Anisotropic scale factors were then applied to remove anisotropy from the data set. Lastly, an isotropic B of -7.68 $\text{Å}^2$ was applied to restore the magnitude of the high resolution reflections diminished by anisotropic scaling. The following pseudo precession images illustrate the individual steps.
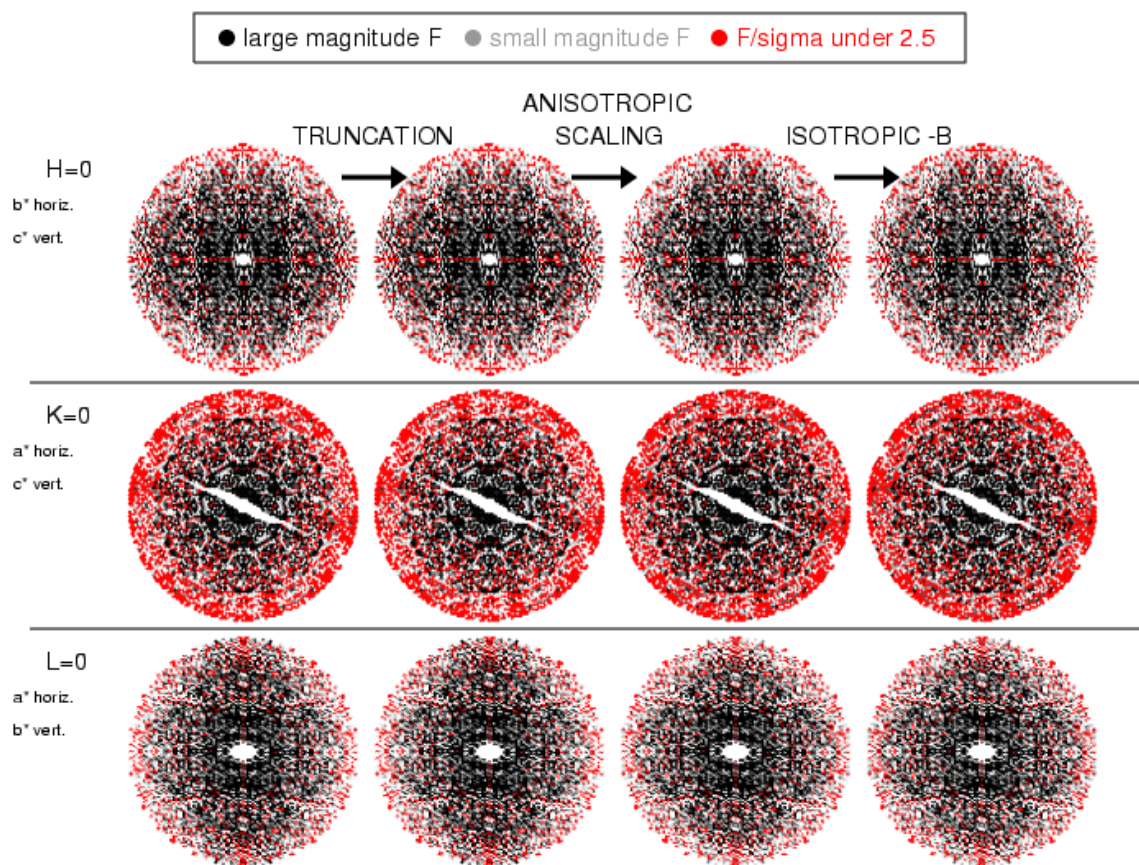
FIGURE 5.17.  Location of the BjPutA A310V mutation (pink).  A)  EcPutA
PRODH domain   B) TtPRODH
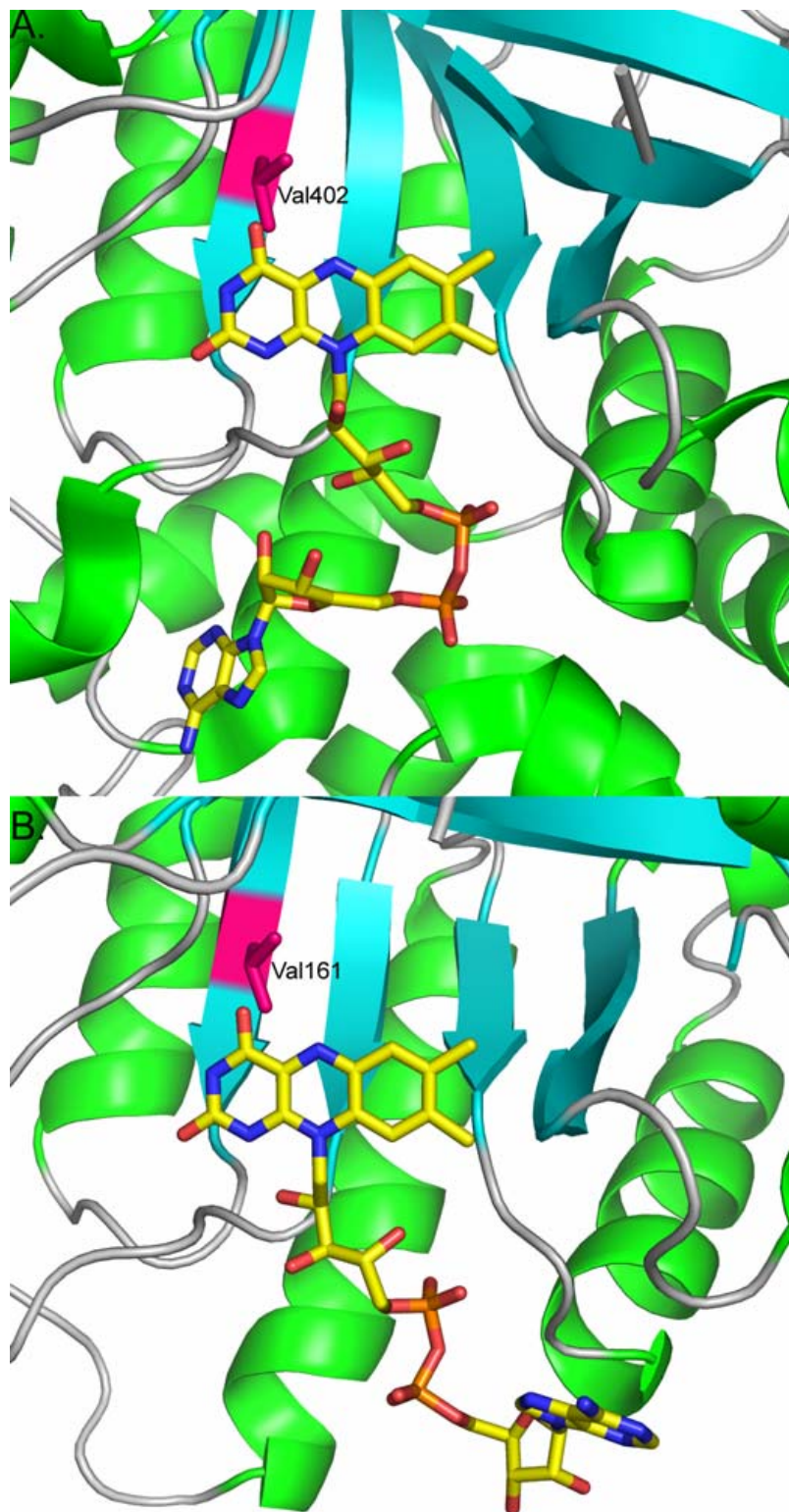
FIGURE 5.18. Radical scavengers A) oxidized ascorbate B) reduced ascorbate and C) oxidized quinone D) reduced quinone.
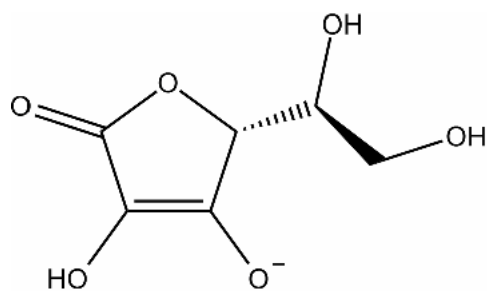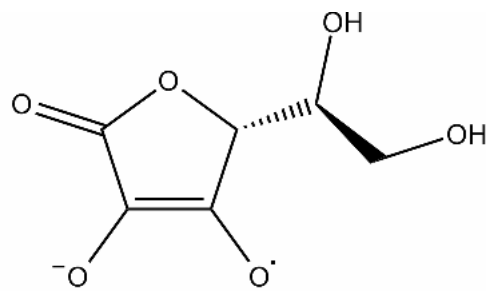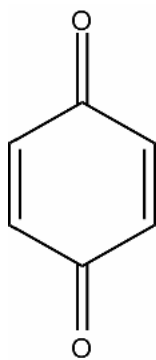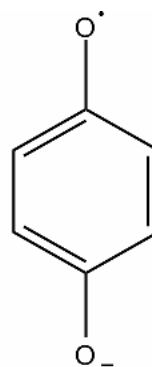
TABLE 5.1. Expected intensity distributions values for normal and twinned crystals (taken from Dauter, 2003).(1)

Cumulative intensity distributions
  Non-centrosymmetric reflections
    Untwinned          $_1N(z) = 1 - \exp(-z)$
    50% twinned        $_1N(z, 0.5) = 1 - (1 + 2z)\exp(-2z)$
    $\alpha$ twinned   $_1N(z, \alpha) = (\alpha[\exp(-z/\alpha) - 1] - (1 - \alpha)\{\exp[-z/(1 - \alpha)] - 1\})/(1 - 2\alpha)$

  Centrosymmetric reflections
    Untwinned          $_{-1}N(z) = \mathrm{erf}(z/2)^{1/2}$
    50% twinned        $_{-1}N(z, 0.5) = 1 - \exp(-z)$
Cumulative $S(H)$ distributions
  Non-centrosymmetric reflections    $S(H) = H/(1 - 2\alpha)$
  Centrosymmetric reflections        $S(H) = \cos^{-1}[H/(2\alpha - 1)/\pi]$

($b$) Moments of $I$ and $H$.

| | $\langle I^2\rangle/\langle I\rangle^2$ | $\langle F\rangle^2/\langle I\rangle$ (Wilson ratio) | $\langle H\rangle$ | $\langle H^2\rangle$ |
|---|---|---|---|---|
| Non-centrosymmetric | | | $\frac{1}{2} - \alpha$ | $(1 - 2\alpha)^2/3$ |
| Untwinned | 2.0 | 0.785 ($= \pi/4$) | 0.5 | 0.333 |
| 50% twinned | 1.5 | 0.885 | 0.0 | 0.0 |
| Centrosymmetric | | | $2(1 - 2\alpha)/\pi$ | $(1 - 2\alpha)^2/2$ |
| Untwinned | 3.0 | 0.637 ($= 2/\pi$) | 0.637 | 0.5 |
| 50% twinned | 2.0 | 0.785 | 0.0 | 0.0 |

($c$) Moments of $L$ and cumulative $N(|L|)$ distributions.

| | $\langle|L|\rangle$ | $\langle L^2\rangle$ | $N(L)$ | $N(|L|)$ |
|---|---|---|---|---|
| Non-centrosymmetric | | | | |
| Untwinned | 1/2 | 1/3 | $(L + 1)/2$ | $|L|$ |
| Twinned | 3/8 | 1/5 | $(L + 1)^2(2 - L)/4$ | $|L|(3 - L^2)/2$ |
| Centrosymmetric untwinned | $2/\pi$ | 1/2 | $\arccos(-L)/\pi$ | $(2/\pi)\arcsin(|L|)$ |

TABLE 5.2. Merohedral Twinning operations for various point groups.

| True Point Group | Twin Operation | *hkl* related to |
|---|---|---|
| 3 | 2 along a,b | *h, -h-k, -l* |
|  | 2 along a*,b* | *h+k, -k, -l* |
|  | 2 along c | *-h, -k, l* |
| 4 | 2 along a,b,a*,b* | *h, -k, -l* |
| 6 | 2 along a,, a*,b* | *h, -h-k, -l* |
| 321 | 2 along a*,b*,c | *-h, -k, l* |
| 312 | 2 along a,b,c | *-h, -k, l* |
| 23 | 4 along a,b,c | *k, -h, l* |

TABLE 5.3.  Data collection statistics processed with d*TREK sets used in ascorbate soaks and statistics of merging of initial data from hexagonal bipyramids.  The data set labels correspond to the following:

1) b12_05 – 2.7 Å
2) b2_04 – 2.55 Å
3) b7_05 – 2.6 Å
4) b7_03 – 2.6 Å
5) b7_02 – 2.6 Å
6) b2_03 – 2.9 Å
7) b3_04 – 2.8Å
8) b7_03 – 5-2.5Å

**7 data sets**

|  | $R_{merge}$ | $R_{merge}$ **low** | $R_{merge}$ **high** | % Complete | Multiplicity |
|---|---|---|---|---|---|
| 1234567 | 0.106 | 0.073 | 0.372 | 91.9 | 3.82 |
| 1235678 | 0.105 | 0.072 | 0.072 | 91.9 | 3.83 |

**6 data sets**

|  | $R_{merge}$ | $R_{merge}$ **low** | $R_{merge}$ **high** | % Complete | Multiplicity |
|---|---|---|---|---|---|
| 123456 | 0.100 | 0.068 | 0.367 | 90.0 | 3.43 |
| 123567 | 0.101 | 0.063 | 0.373 | 90.8 | 3.28 |
| 123568 | 0.100 | 0.068 | 0.371 | 90.0 | 3.44 |

**5 data sets**

|  | $R_{merge}$ | $R_{merge}$ **low** | $R_{merge}$ **high** | % Complete | Multiplicity |
|---|---|---|---|---|---|
| 12356 | 0.095 | 0.059 | 0.373 | 88.8 | 2.88 |
| 12456 | 0.096 | 0.065 | 0.348 | 85.6 | 2.96 |
| 12567 | 0.098 | 0.062 | 0.357 | 86.8 | 2.80 |
| 12568 | 0.095 | 0.065 | 0.350 | 85.6 | 2.97 |
| 12578 | 0.101 | 0.101 | 0.354 | 82.6 | 3.13 |

**4 data sets**

|  | $R_{merge}$ | $R_{merge}$ **low** | $R_{merge}$ **high** | % Complete | Multiplicity |
|---|---|---|---|---|---|
| 1256 | 0.089 | 0.053 | 0.350 | 83.9 | 2.39 |
| 1257 | 0.096 | 0.096 | 0.353 | 80.5 | 2.54 |
| 1258 | 0.093 | 0.093 | 0.348 | 77.1 | 2.80 |

TABLE 5.4. CNS simulated annealing refinements of 104nofad.pdb P5CDH (542-992) or PutA (2-992) against the first 5$^o$ of various data sets. The data set labels correspond to the following:
1) b12_05 – 2.7 Å
2) b2_04 – 2.55 Å
3) b7_05 – 2.6 Å
4) b7_03 – 2.6 Å
5) b7_02 – 2.6 Å
6) b2_03 – 2.9 Å
7) b3_04 – 2.8Å
8) b7_03 – 5-2.5Å

| P5CDH refine.pdb | R | $R_{free}$ | PutA refine.pdb | R | $R_{free}$ |
|---|---|---|---|---|---|
| dm1234567 | 0.4693 | 0.5174 | dm123456b | 0.3999 | 0.5012 |
| dm123456 | 0.4662 | 0.513 | dm1234567b | 0.3974 | 0.4998 |
| dm1235678 | 0.4688 | 0.5215 | dm12356b | 0.3963 | 0.4934 |
| dm123567 | 0.464 | 0.5227 | dm123567b | 0.3957 | 0.4879 |
| dm123568 | 0.4667 | 0.5139 | dm1235678b | 0.3991 | 0.4883 |
| dm12356 | 0.4661 | 0.5158 | dm123568b | 0.3944 | 0.4951 |
| dm12456 | 0.4693 | 0.5206 | dm12456b | 0.4055 | 0.5096 |
| dm12567 | 0.4651 | 0.5232 | dm12567b | 0.3908 | 0.494 |
| dm12568 | 0.4657 | 0.5109 | dm12568b | 0.3989 | 0.5182 |
| dm12578 | 0.4661 | 0.5304 | dm12578b | 0.4003 | 0.5154 |
| dmb12_05 | 0.4651 | 0.516 | dmb12_05b | 0.3935 | 0.4875 |
| dmb2_04 | 0.4992 | 0.5418 | dmb2_04b | 0.4239 | 0.4967 |
| dmb7_02 | 0.4943 | 0.531 | dmb7_02b | 0.4272 | 0.5164 |
| dmb7_03 | 0.4867 | 0.5165 | dmb7_03b | 0.4221 | 0.4942 |
| dmb7_05 | 0.4891 | 0.5295 | dmb7_05b | 0.426 | 0.5125 |

TABLE 5.5. Spacegroup investigation. Using model d8x5_1_001_refmac12.pdb, CNS simulated annealing refinements after 3 rounds of rigid body refinements using data set d8x5 processed with MOSFLM. Percent of reflections utilized for this refinement is indicated by % reflections and number of molecules in the asymmetric unit are listed as mol/ASU.

| Spacegroup | R | $R_{free}$ | % reflections | mol/ASU |
|---|---|---|---|---|
| $P6_222$ | 0.3748 | 0.4446 | 93.6 | 1 |
| $P6_2$ | 0.3665 | 0.4165 | 99.0 | 2 |
| $P3_221$ | 0.3695 | 0.4233 | 99.0 | 2 |
| $P3_2$ | 0.3652 | 0.4232 | 97.6 | 4 |

TABLE 5.6.   Utilizing CNS least squares merohedral twinning refinements for various space groups and twinning fractions with ALS May 2006 d8x5 data. Note, $P3_2$ has three twinning operators, the results for only one is shown in this table.  The other two $P3_2$ twinning operators are the same operators used for $P6_2$ and $P3_221$, *h, -h-k, -l* and *-h, -k, l,* respectively.

| Spacegroup | twin operator | twin fraction | R | $R_{free}$ |
|---|---|---|---|---|
| $P6_2$ | *h, -h-k, -l* | 0.00 | 0.3445 | 0.4337 |
| | | 0.25 | 0.3128 | 0.4228 |
| | | 0.35 | 0.3039 | 0.4134 |
| | | 0.45 | 0.2987 | 0.4061 |
| | | 0.50 | 0.2979 | 0.4060 |
| $P3_221$ | *-h, -k, l* | 0.00 | 0.3482 | 0.4384 |
| | | 0.25 | 0.3164 | 0.4203 |
| | | 0.35 | 0.3061 | 0.4199 |
| | | 0.45 | 0.2996 | 0.4147 |
| | | 0.50 | 0.2960 | 0.4127 |
| $P3_2$ | *h+k, -k, -l* | 0.00 | 0.3360 | 0.4371 |
| | | 0.25 | 0.3023 | 0.4181 |
| | | 0.35 | 0.2919 | 0.4142 |
| | | 0.45 | 0.2859 | 0.4114 |
| | | 0.50 | 0.2859 | 0.4065 |

APPENDIX 5.1.  Data Set Summary of BjPutA

Advanced Photon Source 19-BM
- MARCH 17-19th, 2004  (BjPutA native and HA soaks)

Advanced Photon Source 19ID
- NOVEMBER 14-16th, 2003 (BjPutA I90, w1-47)
- JUNE 15-16th, 2003
- AUGUST 11-12, 2004 (BjPutA SMProaps1, SMProAPS6)

Advanced Light Source
- DECEMBER 7-8th, 2004 (BjPutA SeMet BH4, Pro+dithio)
- JANUARY 22-23rd, 2005 (BjPutA SeMet)
- AUGUST 27-28th, 2005 (A310V Pro)
- OCTOBER 11-12th, 2006 (A310V crosslink, A310V THFA & A310V oxidized)
- MAY 12-13th 2006 (ascorbate hexagonal P6222, ascorbate diamond P422)
- SEPTEMBER 11-12th (hexagonal P6222 oxidized and dithionite/proline reduced)

## APPENDIX 5.2.  Indexing from ALS May 2006 d8x5 utilizing d*TREK.

```
Least-squares fit of reduced primitive cell to 44 lattice characters
sorted on decreasing (highest to lowest) symmetry.
Only solutions with residuals <=  55.____0 are listed.
=======================================================================
 Soln  LeastSq Spgrp Cent    Bravais type        a         b         c
 num residual  num* type    Cell volume       alpha      beta     gamma
=======================================================================
   1   47.215   195   P          cubic   159.503   159.503   159.503
                                4057935    90.000    90.000    90.000

   2   49.668   197   I          cubic   166.702   166.702   166.702
                                4632563    90.000    90.000    90.000

   4   37.499   146   R rhomb/hexagonal  158.662   158.662   649.292
                               14155210    90.000    90.000   120.000

   5    0.156   143   P  trig/hexagonal  145.024   145.024   185.181
                                3372911    90.000    90.000   120.000

   6   35.936    79   I      tetragonal  186.601   186.601   117.497
                                4091247    90.000    90.000    90.000

   7   41.941    75   P      tetragonal  144.968   144.968   185.181
                                3891714    90.000    90.000    90.000

   8   34.273    23   I    orthorhombic  144.907   167.466   415.096
                               10073163    90.000    90.000    90.000

   9    0.094    21   C    orthorhombic  144.907   251.398   185.181
                                6746040    90.000    90.000    90.000

  10   18.868    22   F    orthorhombic   91.719   242.443   358.825
                                7979092    90.000    90.000    90.000

  11   41.941    16   P    orthorhombic  144.907   145.029   185.181
                                3891712    90.000    90.000    90.000

  12    0.047     5   C      monoclinic  144.964   251.496   235.290
                                6751327    90.000   128.091    90.000

  13    0.123     3   P      monoclinic  144.907   185.181   145.029
                                3373019    90.000   119.920    90.000

  14    0.000     1   P       triclinic  144.907   145.029   185.181
                                3373017    89.964    89.939    60.080


=======================================================================
*Suggested spacegroup number until systematic absences are examined.

To view least-squares fits to other lattices,
enter a new residual between 55 and 100 at the following prompt.
```

## APPENDIX 5.3.  Indexing from ALS May 2006 d3_04_2 utilizing d*TREK.

```
Least-squares fit of reduced primitive cell to 44 lattice characters
sorted on decreasing (highest to lowest) symmetry.
Only solutions with residuals <=   3.0 are listed.
=======================================================================
 Soln LeastSq Spgrp Cent   Bravais type        a        b        c
 num residual  num* type   Cell volume      alpha     beta    gamma
=======================================================================
   7    1.875    75    P    tetragonal    138.601  138.601  266.580
                            5121070        90.000   90.000   90.000

   9    1.870    21    C    orthorhombic  195.762  196.261  266.580
                            10242108       90.000   90.000   90.000

  11    1.859    16    P    orthorhombic  138.135  139.065  266.580
                            5120955        90.000   90.000   90.000

  12    0.513     5    C    monoclinic    196.261  195.762  266.580
                            10240716       90.000   90.945   90.000

  13    0.368     3    P    monoclinic    139.065  138.135  266.580
                            5120088        90.000   91.055   90.000

  14    0.000     1    P    triclinic     138.135  139.065  266.580
                            5120011        88.945   89.720   89.854


=======================================================================
*Suggested spacegroup number until systematic absences are examined.

To view least-squares fits to other lattices,
enter a new residual between 15 and 100 at the following prompt.
```

# APPENDIX 5.4. Data processing statistics using various spacegroups for ALS May 2006 d8x5 processed with MOSFLM.

```
P6222
========================================================================

Summary data for Project: ProDithio Crystal: d8x5_1 Dataset: d8x5_1

                                      Overall   OuterShell

  Low resolution limit                 47.56      2.74
  High resolution limit                 2.60      2.60

  Rmerge                               0.076     0.698
  Rmeas (within I+/I-)                 0.080     0.733
  Rmeas (all I+ & I-)                  0.080     0.733
  Fractional partial bias             -0.024    -0.040
  Total number of observations        375060     53088
  Total number unique                  36104      5187
  Mean(I)/sd(I)                         20.0       2.3
  Completeness                          99.9     100.0
  Multiplicity                          10.4      10.2


========================================================================


P62
========================================================================

Summary data for Project: ProDithio Crystal: d8x5_1 Dataset: d8x5_1

                                      Overall   OuterShell

  Low resolution limit                 47.56      2.74
  High resolution limit                 2.60      2.60

  Rmerge                               0.074     0.671
  Rmeas (within I+/I-)                 0.081     0.741
  Rmeas (all I+ & I-)                  0.081     0.741
  Fractional partial bias             -0.021    -0.043
  Total number of observations        375101     53093
  Total number unique                  67882      9940
  Mean(I)/sd(I)                         14.5       1.6
  Completeness                          99.8     100.0
  Multiplicity                           5.5       5.3


========================================================================
```

APPENDIX 5.4.  Data processing statistics for ALS May 2006 d8x5 processed with MOSFLM in various spacegroups (continued).

```
P32
========================================================================

Summary data for Project: ProDithio Crystal: d8x5_1 Dataset: d8x5_1

                                          Overall   OuterShell

  Low resolution limit                      47.56      2.74
  High resolution limit                      2.60      2.60

  Rmerge                                     0.067     0.581
  Rmeas (within I+/I-)                       0.080     0.693
  Rmeas (all I+ & I-)                        0.080     0.693
  Fractional partial bias                   -0.020    -0.039
  Total number of observations             375179     53098
  Total number unique                      132057     19381
  Mean(I)/sd(I)                              10.4       1.2
  Completeness                               98.1      98.2
  Multiplicity                                2.8       2.7


========================================================================


P3221
========================================================================

Summary data for Project: ProDithio Crystal: d8x5_1 Dataset: d8x5_1

                                          Overall   OuterShell

  Low resolution limit                      47.56      2.74
  High resolution limit                      2.60      2.60

  Rmerge                                     0.073     0.668
  Rmeas (within I+/I-)                       0.081     0.738
  Rmeas (all I+ & I-)                        0.081     0.738
  Fractional partial bias                   -0.021    -0.049
  Total number of observations             375096     53093
  Total number unique                       69849     10143
  Mean(I)/sd(I)                              14.3       1.6
  Completeness                               99.8     100.0
  Multiplicity                                5.4       5.2


========================================================================
```

284

## APPENDIX 5.4.  C222 d*TREK processing of ALS September 2006 d8x5 data.

```
Rmerge vs Resolution
-------------------------------------------------------------------------------

  Resolution   Average   Num     Num    I/sig  I/sig  Rducd  Model Rmerge Rmerge
     range      counts   rejs   mults   unavg   avg   ChiSq  Eadd*  shell  cumul
-------------------------------------------------------------------------------

 57.03 - 5.38   13870    409    10241   13.9   26.9   0.64   0.06  0.028  0.028
  5.38 - 4.27    8327    292    10279    8.8   17.9   0.64   0.08  0.041  0.033
  4.27 - 3.73    5625    359    10280    6.1   12.6   0.74   0.11  0.057  0.038
  3.73 - 3.39    3226    329    10311    4.4    8.9   0.84   0.15  0.081  0.042
  3.39 - 3.15    1440    287    10384    2.7    5.5   0.97   0.21  0.129  0.046
  3.15 - 2.96     778    234    10497    2.0    4.1   1.03   0.26  0.172  0.049
  2.96 - 2.82     422    292    10507    1.6    3.0   1.13   0.32  0.246  0.052
  2.82 - 2.69     284    240    10524    1.3    2.3   1.21   0.36  0.315  0.055
  2.69 - 2.59     185    372    10584    1.1    1.8   1.35   0.42  0.425  0.057
  2.59 - 2.50     150    418    10517    1.0    1.5   1.41   0.48  0.512  0.059
-------------------------------------------------------------------------------

 57.03 - 2.50    3451   3232   104124    4.3    8.5   0.99   0.10  0.059  0.059

I/sig unavg is the mean I/sig for the unaveraged reflections in the input file.
I/sig avg   is the mean I/sig for the unique reflections in the output file.
 * When EMul == 1.41


Summary of data collection statistics
------------------------------------------------------------

 Spacegroup                   C222
 Unit cell dimensions         144.87   251.36   185.01
                               90.00    90.00    90.00
 Resolution range             57.03 - 2.50   (2.59 - 2.50)
 Total number of reflections  427543
 Number of unique reflections 113474
 Average redundancy           3.77           (3.65)
 % completeness               97.6           (98.7)
 Rmerge                       0.059          (0.512)
 Reduced ChiSquared           0.99           (1.41)
 Output <I/sigI>              8.5            (1.5)
------------------------------------------------------------

  Note: Values in () are for the last resolution shell
```

## APPENDIX 5.5.  Processing statistics of ALS May 2006 d3_04_2 data in various spacegroups utilizing d*TREK.

```
Summary of data collection statistics
--------------------------------------------------------------
 Spacegroup                    P43
 Unit cell dimensions          138.03   138.03   267.04
                               90.00   90.00   90.00
 Resolution range              46.01 - 2.80   (2.90 - 2.80)
 Total number of reflections   209861
 Number of unique reflections  107065
 Average redundancy            1.96            (1.55)
 % completeness                87.6            (70.3)
 Rmerge                        0.092           (0.241)
 Reduced ChiSquared            0.97            (0.96)
 Output <I/sigI>               6.7             (2.4)
--------------------------------------------------------------
  Note: Values in () are for the last resolution shell.


Rmerge vs Resolution
-----------------------------------------------------------------------------
 Resolution   Average   Num    Num   I/sig  I/sig  Rducd  Model Rmerge Rmerge
    range     counts    rejs  mults  unavg   avg   ChiSq  Eadd*  shell  cumul
-----------------------------------------------------------------------------
 46.01 - 6.03   1234    597   11293    8.7   14.7   0.89   0.09  0.058  0.058
  6.03 - 4.79    705    233   10288    4.7    6.9   1.01   0.17  0.099  0.070
  4.79 - 4.18    965    219    9809    4.5    6.3   1.03   0.18  0.101  0.078
  4.18 - 3.80    752    167    9539    4.1    5.5   1.06   0.20  0.108  0.083
  3.80 - 3.53    569    223    8552    4.2    5.4   1.02   0.20  0.107  0.085
  3.53 - 3.32    419     65    7113    4.2    5.3   0.93   0.20  0.107  0.086
  3.32 - 3.15    275     35    6040    3.4    4.3   0.92   0.22  0.131  0.087
  3.15 - 3.02    192     14    5525    2.7    3.4   0.97   0.27  0.171  0.089
  3.02 - 2.90    135     18    5049    2.2    2.7   0.98   0.32  0.213  0.090
  2.90 - 2.80    108     15    4739    1.9    2.4   0.96   0.34  0.241  0.092
-----------------------------------------------------------------------------
 46.01 - 2.80    649   1586   77947    4.6    6.7   0.97   0.17  0.092  0.092

I/sig unavg is the mean I/sig for the unaveraged reflections in the input file.
I/sig avg   is the mean I/sig for the unique reflections in the output file.
 * When EMul == 2.18
```

APPENDIX 5.5 (continued). Processing statistics of ALS May 2006 d3_04_2
data in various spacegroups utilizing d*TREK.

```
Summary of data collection statistics
-------------------------------------------------------------
 Spacegroup                      P422
 Unit cell dimensions            138.17   138.17   267.30
                                  90.00    90.00    90.00
 Resolution range                46.06 - 2.80    (2.90 - 2.80)
 Total number of reflections     210980
 Number of unique reflections    59993
 Average redundancy              3.52           (2.70)
 % completeness                  93.0           (78.0)
 Rmerge                          0.111          (0.296)
 Reduced ChiSquared              0.92           (1.01)
 Output <I/sigI>                 8.0            (2.9)
-------------------------------------------------------------
  Note: Values in () are for the last resolution shell.
```

```
Rmerge vs Resolution
-------------------------------------------------------------------------------
 Resolution   Average   Num     Num    I/sig  I/sig  Rducd  Model Rmerge Rmerge
    range      counts   rejs   mults   unavg   avg   ChiSq  Eadd*  shell  cumul
-------------------------------------------------------------------------------
 46.06 - 6.03   1220     523    6736     7.2   16.4   0.80   0.10  0.070  0.070
  6.03 - 4.79    707     283    6482     4.5    8.9   0.90   0.17  0.108  0.082
  4.79 - 4.18    964     207    6393     4.3    8.0   0.97   0.18  0.116  0.091
  4.18 - 3.80    749     184    6378     3.8    6.7   0.96   0.21  0.128  0.097
  3.80 - 3.53    564     232    6111     3.8    6.4   0.93   0.20  0.129  0.100
  3.53 - 3.32    416      63    5713     3.8    6.2   0.91   0.20  0.135  0.103
  3.32 - 3.15    270      32    5185     3.1    4.9   0.92   0.24  0.169  0.105
  3.15 - 3.02    189      32    4815     2.7    4.3   1.05   0.25  0.211  0.107
  3.02 - 2.90    133      24    4508     2.2    3.4   1.01   0.30  0.259  0.109
  2.90 - 2.80    106      16    4260     1.9    2.9   1.01   0.32  0.296  0.111
-------------------------------------------------------------------------------
 46.06 - 2.80    644    1596   56581     4.2    8.0   0.92   0.17  0.111  0.111
```

```
I/sig unavg is the mean I/sig for the unaveraged reflections in the input file.
I/sig avg   is the mean I/sig for the unique reflections in the output file.
 * When EMul == 2.50
```

**VITA**


Tommi Anna White is the daughter of Calvin and Margaret White and grew up in St. Peters, Missouri. She attended University of Missouri-Columbia as an undergraduate and graduated with a Bachelors of Science in Biochemistry in December of 2000. After employment conducting research at both Pharmacia and at University of Missouri-Columbia, she was accepted into the University of Missouri-Columbia Ph.D. program in Biochemistry. Tommi joined Jack Tanner's group to pursue crystallographic studies of proline catabolic enzymes. She completed her Ph.D. in Biochemistry in April 2007. In July 2007, Tommi will start her post-doctoral fellowship at the National Cancer Institute under Jacqueline Milne, studying pyruvate dehydrogenase complex using cryo-electron microscopy and electron tomography.