STRUCTURAL DIVERSITY OF PROLINE CATABOLIC ENZYMES REVEALED BY SMALL ANGLE X-RAY SCATTERING, X-RAY CRYSTALLOGRAPHY AND LIGHT SCATTERING

A Dissertation

Presented to the Faculty of Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

By

RANJAN KUMAR SINGH

Dr. John J. Tanner, Dissertation Supervisor

DECEMBER 2012

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

STRUCTURAL DIVERSITY OF PROLINE CATABOLIC ENZYMES REVEALED BY SMALL ANGLE X-RAY SCATTERING, X-RAY CRYSTALLOGRAPHY AND LIGHT SCATTERING

presented by Ranjan Kumar Singh,

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 J. Beamer

Professor Michael Henzl

Professor Kent S. Gates

Dedication

I would like to dedicate my thesis

To my parents,

And all of friends,

Without whom none of my success would be possible

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude towards my PhD advisor Prof. John J. Tanner for his patience, motivation, enthusiasm and immense involvement in each and every aspect of my PhD; right from its humble beginnings to the compilation of my research work culminating into this thesis in its present form. I would specially like to thank him for his perseverance in guiding me through difficult times during my research and encouraging me to overcome the relevant scientific problems successfully. His drive for excellence and thoughtful understanding of the various aspects of scientific research work was infectious and catapulted me to achieve greater heights in my research endeavors. During my entire Ph.D. program, I was fortunate to have countless discussions regarding my research work, during which he has provided me with thoughtful insights that have helped me substantially in accomplishing my doctoral work

I would also take this opportunity to thank Professor Donald. F. Becker and his group for the numerous meaningful and relevant discussions on our project during various Skype-meetings. My grateful acknowledgements are also extended to Professor *Michael Henzl* for allowing me to submit a plethora of samples for analysis using Analytical Ultracentrifugation and ITC. I would also gratefully appreciate Professor Lesa Beamer for letting me work in one of her projects, for being very supportive and nice and a big thanks to Professor Kent S. Gates for helping me ease through my first year organic chemistry course work with flying colors. Finally, I extend my sincere thanks to each of these accomplished scientists to have graced my committee with their esteemed presence.

I would like to thank members of Tanner lab for keeping a congenial environment in the lab and giving various inputs in my research work during the weekly lab meetings. Special thanks to Travis Pemberton for being cool and easy going. My sincere thanks also goes to Min Luo for helping me driving one of my projects towards completion. Thanks to John Larson, Dale Karr and Dhiraj Srivastava for teaching me most of the basic techniques used in the Tanner lab and to Ritcha Mehra-Chaudhary for letting me borrow her lab supplies and for being nice. I would also like to thank Jay nix for helping me in remote data collection.

I would like to extend my sincere gratitude towards my fiancée Sudha for being lovely. My sincere thanks also goes out to my friend Shatadru Chakravarty for being understanding and supportive during my PhD. I would also like to thank my childhood friend Sunil, who used to visit my parents frequently during my prolonged absence and providing them with much needed moral support and comfort.

No words of gratitude can suffice the hardships and immense sacrifice that my parents have endured during my entire educational endeavor. I owe to them my each and every success and it was their zeal towards providing me a better form of education that has today culminated into a successful doctoral stint. A special mention to my beloved mother for being a pillar of strength through difficult and trying times and for being a never ending source of inspiration.

Finally I would like to thank to all the staff members of Chemistry Department and University of Missouri for giving prompt responses of any query I ever had

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPENDIX

ABSTRACT

The proline catabolic enzymes catalyze the 4-electron oxidation of proline to glutamate. The reaction involves two enzymes, proline dehydrogenase (PRODH) and Δ^1-pyrroline -5-carboxylate dehydrogenase (P5CDH). Some bacterial organisms have both of these enzymes fused together, and the fused bifunctional enzymes are called Proline utilization A (PutA). In addition to these bifunctional enzymes, some PutAs are trifunctional, because they moonlight as transcription repressors of their own gene. Our lab recently reported that the quaternary structure of the bifunctional PutA from B. *japonicam* (BjPutA) is a ring-shaped tetramer. However, the structural organization of PutAs from other organisms is still unknown. In particular, there are no structures available for moonlighting trifunctional PutAs. We therefore utilized small angle X-ray scattering (SAXS) to obtain the overall shape of a trifunctional PutA from *Escherichia coli* (EcPutA). In addition, rigid body modeling of full-length PutA has been done with the help of SAXS data and crystal structures of DNA-binding and PRODH domains of EcPutA, and BjPutA crystal structure. Unique structural features of PutA have also been explored through multiple sequence alignments and homology modeling using the webservers like ClustalW, Espript, Phyre, and Swiss Model. The results obtained from sequence alignment study led us to work on finding the diversity in oligomeric states of PutAs. Finally, the structural basis of HPII disease that is related to disorder in human P5CDH was determined through X-ray crystallographic studies.

CHAPTER 1

Introduction

1.1. Proline catabolism

The interest in proline catabolism biochemistry lies in proline's critical role in bioenergetics, cellular redox control, and cancer. The proline catabolic pathway involves the oxidation of proline to glutamate (1-4). There are two enzymes in the pathway: proline dehydrogenase (PRODH) and Δ^1-pyrroline -5-carboxylate dehydrogenase (P5CDH) (4). Along with the two catalytic steps, the process also involves a non-enzymatic hydrolysis step. These enzymes are widespread in the bacterial and eukaryotic kingdoms of life. Proline oxidation to glutamate in archaea involves a different pair of enzymes, which will not be discussed here.

Figure 1.1 shows reactions involved in proline catabolism. The first step is the PRODH-catalyzed oxidation of proline to an intermediate Δ^1-pyrroline -5- carboxylate (P5C). Flavin adenine dinucleotide (FAD) is the cofactor involved in the PRODH catalyzed reaction. In second step, this intermediate is non-enzymatically hydrolyzed to glutamate semialdehyde (GSA). Finally, in the third step, P5CDH catalyzes the oxidation of glutamate semialdehyde to glutamate. Nicotinamide adenine dinucleotide (NAD^+) is the cofactor involved in the P5CDH-catalyzed reaction. Overall proline catabolism is a four electron oxidation process which is common to any organism.

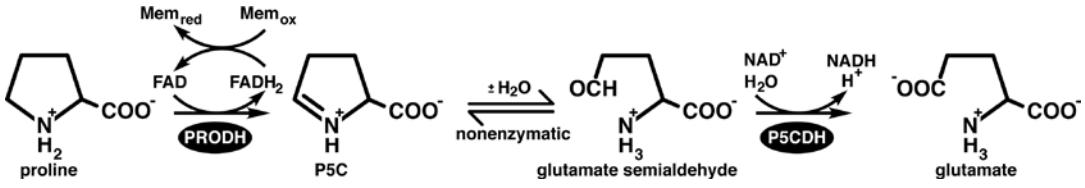


Figure 1.1 The proline catabolic pathway. This figure was taken from Singh *et al*. JBC 2011 (5).

1.2. Biological significance

PRODH and P5CDH in eukaryotes are mitochondrial enzymes. PRODH is associated with the inner mitochondrial membrane, whereas P5CDH is in the mitochondrial matrix. Similar to any other metabolic process, proline catabolism also has disorder associated with it. Any unusual genetic mutation that changes protein structure may affect the catalytic efficiency of the enzyme. The diseases related with proline catabolism are called hyperprolinemia disorders. They are inherited inborn autosomal recessive disorders and are rare disease. Hyperprolinemia is further categorized in two types, defined by mutations in the genes encoding PRODH and P5CDH. Below is the categorical explanation for the two inherited forms of hyperprolinemia.

1.2.1. Hyperprolinemia type I (HPI)

HPI is caused by mutation in the *PRODH* gene. The HPI patients can have proline levels as high as 10 times compared to the normal concentration. People with this condition may not show any significant symptoms (6). In contrast, some patients diagnosed with HPI show severe phenotypes, including mental retardation and increased frequency of seizures (7-9). Though in past, scientists have suggested that the neurological disturbance observed in HPI patients could be coincidental and may not reflect the extent of hyperprolinemia (10). There are various point mutations in *PRODH* which has been reported earlier, that might be responsible for HPI. Some of them could have a severe effect while others may have a milder effect. Missense mutations that are thought to be disease-causing include are V427M, L441P, and R453C (7, 11).

1.2.2. Hyperprolinemia type II (HPII)

HPII is caused by a deficiency of P5CDH. Some of the clinical phenotype includes 10-15-fold increase in plasma proline concentration, accumulation of P5C, and increased excretion of urinary proline. Similar to HPI, patients with HPII may not show any symptoms. However, HPII is usually characterized by mental retardation and convulsions. The missense mutation S352L and the frameshift mutation G521fs(+1) are considered to be pathological mutations (7, 12, 13). Part of this thesis focuses on S352L.

1.3. Diversity in proline catabolic enzymes

One interesting feature of proline catabolism is that PRODH and P5CDH are combined into a single polypeptide in some organisms (Figure 1.2). The combined enzymes are known as proline utilization A (PutA) and were discovered by Roth's group in the late 1970s during their studies of proline utilization in *Salmonella typhimurium* (14). Analysis of genome sequence data suggests that PutAs are limited to Gram-negative bacteria (Figure 1.2, branches 1, 2), whereas PRODH and P5CDH are separate enzymes encoded by distinct genes in Gram-positive bacteria (branch 3B) (15). In eukaryotes, PRODH and P5CDH are also separate enzymes and are localized to mitochondria (branch 3A). Human PRODH is a p53-induced tumor suppressor protein localized to the inner mitochondrial membrane and is often referred to as POX to emphasize its role as a superoxide-generating oxidase (2, 4, 16-22). Human P5CDH (ALDH4 (23)) is also induced by p53 (24) and is located in the mitochondrial matrix. ALDH4 has been characterized biochemically, including elucidation of the oligomeric

state in solution (dimer) and kinetic mechanism (25, 26).

The PutA part of the PutA/PRODH/P5CDH family tree has two branches (15, 27). Branch 1 primarily consists of PutAs from alpha-, beta-, and gamma- proteobacteria. Branch 2 includes PutAs from delta- and epsilon-proteobacteria as well as cyanobacteria. The PutAs in branch 1 have chain lengths from 999 to almost 1400 residues, and the pairwise sequence identities are greater than 38 %. The polypeptide length for branch 2 PutAs ranges from around 980 to almost 1300 residues and the pairwise sequence identity range can be as low as 23 %. Thus, branch 2 PutAs appear to be a more diverse group than branch 1 PutAs. Between branches 1 and 2, the pairwise sequence identities are typically less than 30 %. Nevertheless, the residues in the PRODH and P5CDH active sites are highly conserved, indicating that the three-dimensional structures of the catalytic domains are conserved by PutAs. Whether the three-dimensional arrangement of the catalytic and other domains is likewise conserved remains to be determined. PutAs are further classified as bifunctional or trifunctional. Bifunctional PutAs exhibit only PRODH and P5CDH catalytic activities, have polypeptide chain lengths in the range of ~980 residues to over 1300 residues, and are found in both PutA branches. Bifunctional PutAs from *Bradyrhizobium japonicum* (BjPutA, (28-30) and *Helicobacter species* (31-33) have been studied. Trifunctional PutAs constitute a subset of branch 1 PutAs and are distinguished by the presence of a DNA-binding domain (a ribbon-helix-helix domain) in the first ~50 residues of the polypeptide chain. The polypeptide chain length of trifunctional PutAs are in the range of ~1270-1361. In addition to functioning as dual PRODH/P5CDH enzymes, trifunctional PutAs have a third function of repressing transcription of the put regulon, which contains the genes

encoding PutA and the proline transporter PutP, when proline levels are low (24-27). High levels of proline in the bacterium's environment cause PutA to disengage from the put control region thus activating transcription of *putA* and *putP* genes. Thus, trifunctional PutAs are remarkable proteins that link transcription and metabolism in response to an environmental cue (proline level). Trifunctional PutAs from *S. typhimurium* (25, 26, 28-31) and *Escherichia coli* (EcPutA) (24, 32-46) have been studied. PutA from *E. coli* is the most studied trifunctional PutA and is considered to be the archetypal trifunctional PutA.

Three-dimensional structural studies have contributed to our understanding of proline catabolic proteins. Crystal structures have been solved for the monofunctional PRODH and P5CDH enzymes from *Thermus thermophilus* TtPRODH (27, 34, 35), TtP5CDH (36, 37), the DNA-binding domains of two trifunctional PutAs (38, 39), a PRODH domain construct of EcPutA (EcPutA86-630, (40-43), and full-length BjPutA (44). The structure of TtPRODH and TtP5CDH structures shows the basic features of these two domains. The structure of trifunctonal and bifunctional PutAs are discussed later. Here are some brief details involving the structure of monfunctional PRODH and P5CDH from *Thermus thermophiles.*

1.4. Monofunctional proline catabolic enzymes

1.4.1. Monofunctional PRODH

The crystal structure of the monofunctional PRODH from *Thermus thermophilus* (TtPRODH) has been determined. TtPRODH consists of 327 amino acid residues and has FAD as the cofactor. The crystal structure shows that TtPRODH adopts a distorted $(\beta\alpha)_8$

barrel (Figure 1.3B). It consists of 8 alternating α-helices and parallel β-strands. This kind of fold was first observed in the metabolic enzyme triosephosphate isomerase and is therefore called the TIM barrel (Figure 1.3A)(45, 46). In addition to the TIM barrel fold, the TtPRODH structure also contains a 3 α-helix bundle (αA, αB, αC) that precedes the first helix of the barrel (β0).

The deviation from the normal TIM barrel comes as a result of the position of α8 helix. Figure 1.4 shows the pictorial representation of a TIM barrel fold (pdb code 1TIM) and the distorted TIM barrel fold of TtPRODH. In clearly shows that the original place of α8 helix in classical TIM barrel is taken by α0 helix in the TtPRODH structure. As a result, the α8 helix is located above the carboxy terminal face of the β-barrel (45).

FAD is the redox cofactor of PRODH and is involved in oxidation of proline to P5C (47). FAD is bound to the C-terminal ends of the strands the barrel (Figure 1.3). It abstracts a hydride ion from proline resulting in formation of reduced FAD ($FADH_2$) and P5C. The reduced flavin subsequently transfers two electrons to the electron transport chain thus regenerating the catalyst (Figure 1.1).

The solution state of TtPRODH is not well studied to date. The substrate recognition by TtPRODH could be explained with the inhibitor bound TtPRODH structure. The well –known inhibitor for PRODH is L-tetrahydrofuroic acid (L-THFA). L-THFA mimics the substrate proline very closely. The structure of the L-THFA bound TtPRODH is yet to be reported. Fortunately, the inhibitor bound isolated PRODH domain of *E.coli* PutA is known, which signifies the substrate recognition site of the PRODH domain (41). A summary of substrate recognition is discussed in a later section of the introduction. In addition, a later section also discusses the basic differences between monofunctional

PRODH and the PRODH domain within PutA.

1.4.2. Monofunctional P5CDH

The crystal structure of P5CDH from *Thermus thermophilus* (27) (TtP5CDH) has also been determined. TtP5CDH consists of 516 amino acid residues with NAD^+ as the cofactor. P5CDH belongs to the aldehyde dehydrogenase (ALDH) family. The overall structure of TtP5CDH suggests a similar fold as the other ALDHs. The protomer basically contains three different domains. 1) NAD^+-binding domain, 2) catalytic domain, 3) and oligomerization domain (Figure 1.6).

The NAD^+-binding domain has a non-classical Rossmann fold with a 5-stranded parallel β-sheet, which contrasts the 6-stranded sheet in the classic fold. Figure 1.5 shows the difference between the classical and non-clasical Rossmann fold. The non-classical Rossmann fold domain along with some other parts of the protein constitutes NAD^+-binding domain of TtP5CDH.

The catalytic domain contains the active site cysteine, which acts as nucleophile for the incoming substrate GSA. The catalytic domains consist of an α/β fold. The active site is formed between the NAD^+-binding domain and the catalytic domain.

Two TtP5CDH protomers associate to form a domain swapped dimer. The dimerization domain contains an antiparallel β-strand (also called the β-flap) which interacts with the catalytic domain of the other subunit. This kind of dimerization domain is widespread in proteins of ALDH superfamily.

1.5. Proline utilization A (PutA)

1.5.1. Bifunctional PutA

PutA from *Bradyrhizobium japonicum* (BjPutA) is one of the simplest bifunctional PutA whose crystal structure is known(44). At 999 residues in length, it is the shortest branch 1 PutA known and is thus considered to be a minimalist PutA. Our lab recently reported the crystal structure of BjPutA (PDB code 3haz), which is the first structure of a full-length PutA. As discussed above, PRODH and P5CDH enzymes are one of the domains in PutAs. Though, this is not simply a fusion of PRODH and P5CDH domain, but it also has some other domains not found in individual monofunctional PRODH and P5CDH. The protomer comprises seven domains: arm, alpha, PRODH barrel, linker, NAD^+-binding, P5CDH catalytic, and oligomerization domain (Figure 1.7). In addition, it also have one conserved C-terminal motif found in all branch 1 PutAs.In addition BjPutA have a short linker connecting the two domains.

Similar to TtPRODH, BjPutA PRODH domain also adopts a distorted $(\beta\alpha)_8$ barrel. Despite the similarities in overall fold and amino acid sequence (28 % identity), there is an important difference between monofunctional PRODH and PutA. The PutA PRODH barrel has an extra helix (alpha5a) inserted between beta5 and alpha5 . This insertion in helix causes a different FAD conformation in BjPutA.

The P5CDH domain of BjPutA is very similar to that of TtP5CDH (38 % sequence identity). Like TtP5CDH, the P5CDH domain of BjPutA also has a non-classical Rossmann fold.

The oligomerization domain in BjPutA is similar to that of TtP5CDH, as here too a C-terminal β-flap motif is involved in domain-swapped dimerization. The β-flap of one

protomer interacts with the catalytic domain of the other protomer to form the dimer interface. BjPutA is not just a simple domain swapped dimer but is a dimer-of-dimers tetramer. The second dimer interface is mostly formed by the N-terminal arms and the alpha domains. The BjPutA crystal structure showed many special features in the protein that distinguished it from just being a normal fusion protein. For example, it is able to show substrate channeling phenomenon from PRODH active site to P5CDH active site (44). The PRODH and P5CDH active sites are separated by 41 Å and connected by a large irregularly shaped internal cavity (silver surface in Figure 1.7). This is the proposed cavity for channeling. The channeling mechanism is actively studied by the group of our collaborator, Prof. Donald F. Becker of the University of Nebraska, Lincoln. Apart from the aforementioned features, there are many more unique features of BjPutA not found in monfunctional PRODH and P5CDH that is in the scope of the second chapter.

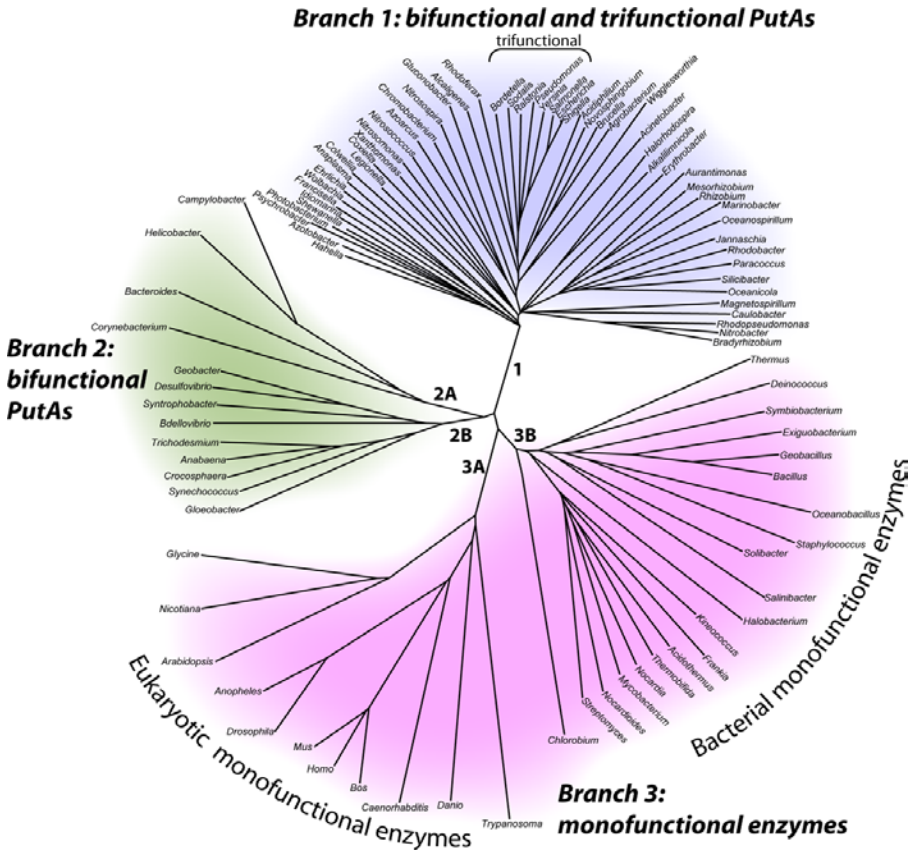


Figure 1.2. Phylogenetic tree representing the organization of proline catabolic enzymes in bacteria and eukaryotes. PutAs are found in branches 1 and 2. Monofunctional PRODH and P5CDH enzymes are found in branch 3. A cluster of trifunctional PutAs is indicated. This figure was taken from Singh et al. FBS 2012 (48).

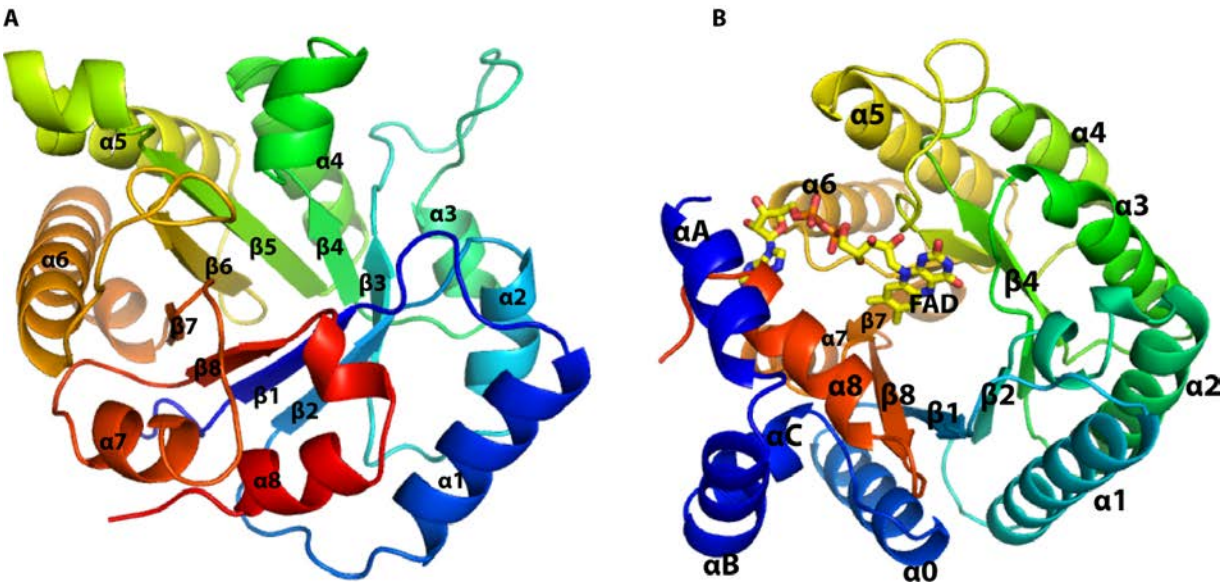


Figure 1.3. Ribbon drawing of the A) classic TIM barrel (pdb code 1TIM), B) TtPRODH protomer. The protein chains are colored in the rainbow scheme, with dark blue at the N-terminus and red at the C-terminus. Selected α-helices and β-strands are labeled. The FAD of TtPRODH is drawn as a stick model in yellow.

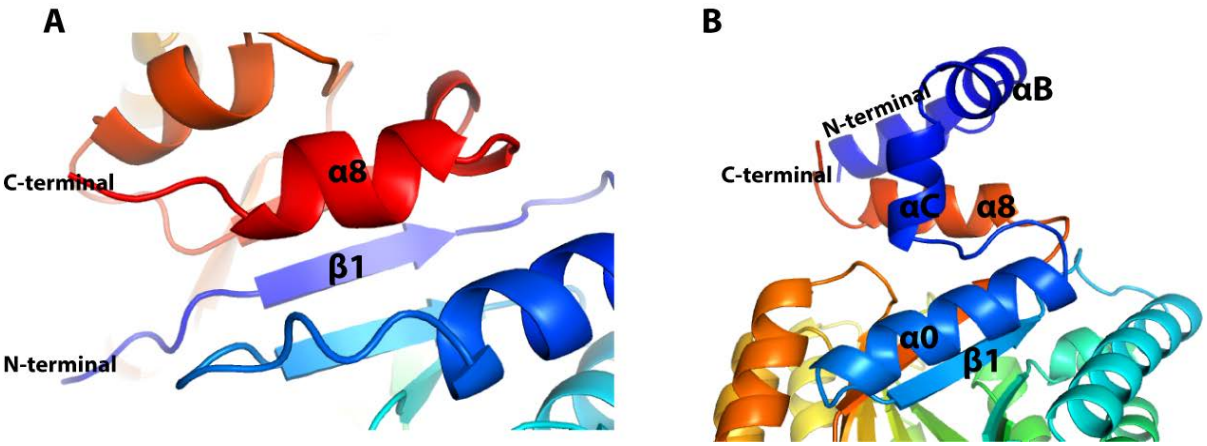


Figure 1.4.Rainbow drawing to show the difference in location of the α8 helices of A) the classical TIM barrel and B) TtPRODH. Note that α0 of TtPRODH corresponds to α8 of the TIM barrel.

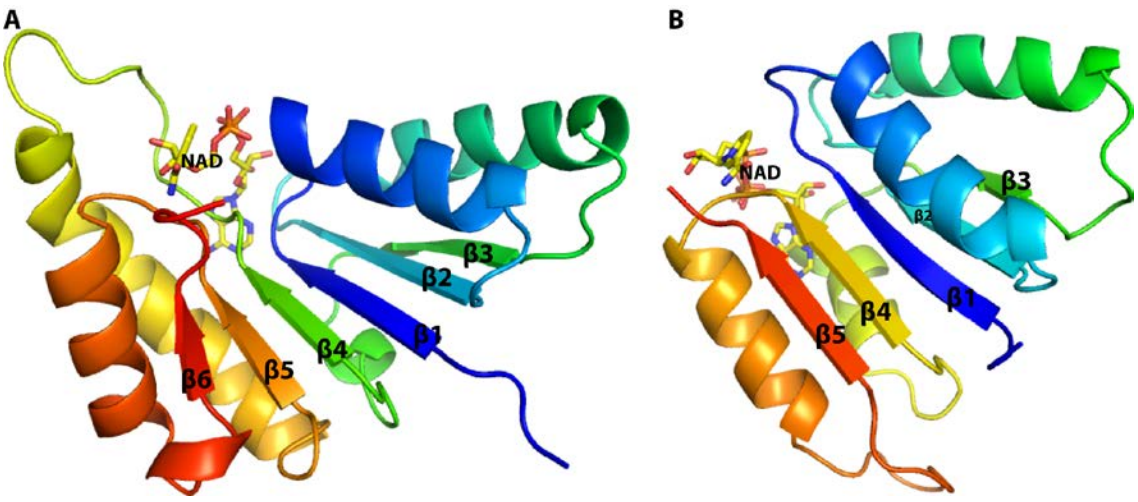


Figure 1.5 Cartoon diagram for (A) classical Rossamann fold domain (from PDB code 2FM3) and (B) non-classical Rossmann fold from TtP5CDH. The protein chains are colored in rainbow scheme. NAD^+ is represented in sticks in yellow. The domains are colored in a rainbow scheme, with blue at the N-terminus and red at the C-terminus.

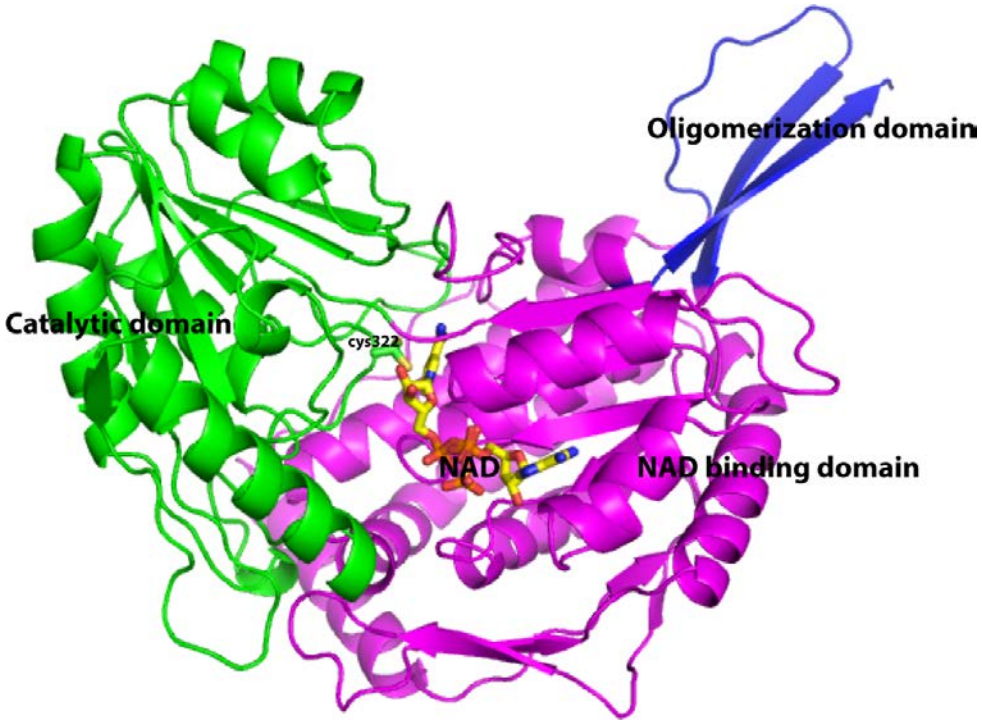


Figure 1.6 Cartoon representation of a protomer of TtP5CDH. Catalytic cysteine and NAD^+ is shown in yellow sticks.

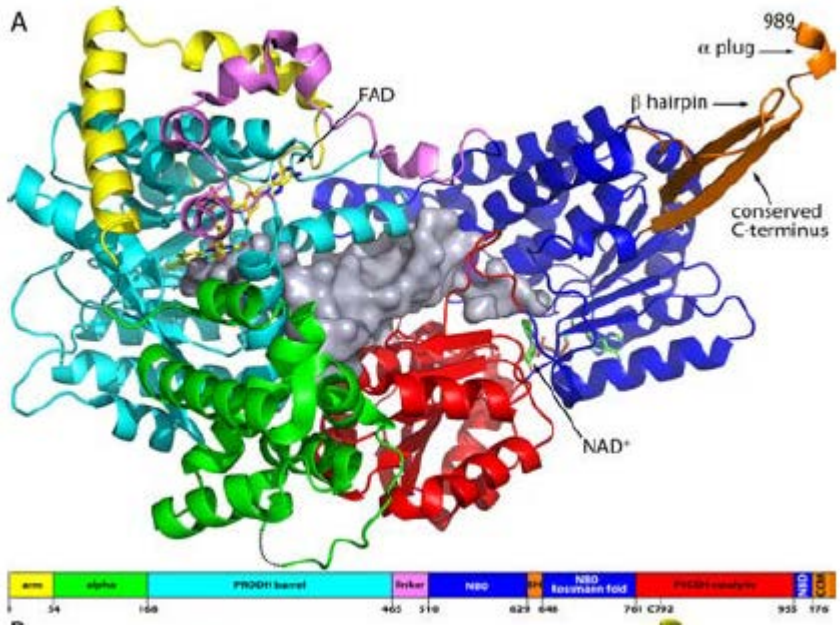


Figure 1.7 Protomer structure of BjPutA with the domains colored according to the domain diagram.The silver surface represents the substrate-channeling cavity. FAD and NAD^+ are represented as yellow and green sticks, respectively. Abbreviations used in the domain diagram: NBD, NAD^+-binding domain; BH, beta-hairpin; CCM, conserved C-terminal motif. This figure was taken from Singh *et al* FBS 2012 (48).

1.5.2. Trifunctional PutA: A moonlighting protein

The trifunctional PutA comes under the class of protein called moonlighting protein. Before going into the brief aspect of the structural studies of trifunctional PutA, the term moonlighting protein and its properties are discussed in the following paragraph.

1.5.2.a. Protein moonlighting

The term moonlighting proteins was coined by Prof. Constance Jeffery in 1999 to illuminate the original discovery of Joram Piatigorsky about multitasking proteins with his finding on crystallin (49-51). He discovered an exciting feature of crystallin (an eye lens protein), which functions as an enzyme when expressed at low levels in tissues but

when expressed in high levels in eye tissue, become densely packed thus forming the lens. He originally termed it as gene sharing proteins which was later termed "moonlighting protein". We can compare the moonlighting proteins with a part time graduate student who is working in lab as a researcher in the day time and in evening he works as customer care representative in call-center. (Such a student would be immediately dismissed from graduate school.) It means that these proteins, along with their original function of catalysis, are also involved in other important cellular functions like signal transduction, transcriptional regulation, apoptosis etc. There are many kinds of moonlighting proteins. Some examples includes moonlighting due to change in cellular location (52, 53), change in oligomeric state (54), and substrate concentration (55). Trifunctional PutA can perform different functions on the basis of different subcellular location. It is an auto-regulatory protein that switches its location for its different activities. Thus, trifunctional PutA regulates its own transcription depending on the concentration of proline in the environment.

1.5.2.b. *Escherichia coli* PutA

One of the well-studied examples of trifunctional PutA is PutA from *Escherichia coli* (EcPutA). EcPutA is a 1320 amino acid residue long PutA which along with catalyzing proline to glutamate oxidation, also acts as a transcriptional repressor (55). Thus, EcPutA act as a sensor for cellular (proline) metabolism. This protein has two sub-cellular locations 1) peripheral membrane enzyme and 2) DNA-bound transcriptional repressor. Similar to BjPutA, the PRODH and P5CDH domains are responsible for the enzymatic function. An N-terminal DNA binding domain is responsible for the

autotranscriptional function. Therefore, the two different functions are performed by entirely separated and specific domains. The protein not only binds to the DNA but it also is able to show enough conformational flexibility to bind to the membrane when the proline level in cellular environment is high. The proline concentration dependent redox change in the flavin within the PRODH domain is thought to be responsible for the conformational change in the overall protein (55-60) that underlies functional switching. It makes sense when we think in terms of evolution; it gives the cell the leisure of using same protein as a tool for performing two exclusively different functions.

The *put* regulon contains two genes encoding PutA and the Na^+ -proline transporter PutP, which are transcribed divergently (39). PutA functions as a transcription repressor by binding in between in region separating these two genes. PutA act as a DNA binding protein when it locates itself in the cytoplasm. The DNA binding activity of the PutA stems from the ribbon-helix-helix domain (RHH) located at the N-terminus of the protein. The secondary structure elements of the RHH domain are comprised of a beta strand followed by two alpha helices and is widespread in prokaryotes (61). The crystal structure of DNA binding domain of EcPutA with and without DNA bound has been determined (38, 39).

1.5.2.b.1. DNA binding domain (PutA1-52)

Sequence analysis and molecular dissection study showed that the first 47 residues of EcPutA is the RHH domain. The above observation was confirmed by solving the crystal structure of the polypeptide corresponding to the first 52 residues of EcPutA (38). The structure showed that the RHH domain exists as dimer, like other RHH domains

(Figure 1.8A). The RHH fold consists of one β strand followed by two α helices. The dimeric assembly is formed by intermolecular β-sheet formation and has two-fold symmetry. There are 9 main chain hydrogen bonds between the two strands. In total, there are 21 hydrogen bonds and 10 salt bridges along with many hydrophobic interactions between the two subunits. In addition to hydrogen bonds and salt bridges, many non-polar side chains contribute to the hydrophobic interaction. The total buried surface area for the dimer interface is 1712 which is 49 % of the total surface area. The aforementioned analysis was calculated from PDBePISA (62).

1.5.2.b.2. DNA binding domain complexed with DNA

The crystal structure of PutA52 with a DNA fragment containing one of the operator sites, which contains the consensus sequence motif GTTGCA, was determined (39). It showed a typical way of interaction of RHH domain and DNA where the intermolecular beta sheet inserts into the DNA major groove. Thr5, Gly7 and Lys9 within the β-strands are involved in base specific hydrogen bonds with the consensus DNA motif. Thr28, Pro29 and His30 form interactions with the DNA backbone (Figure 1.8B). The interaction of the dimeric PutA52 with DNA suggests that the full length PutA would interact in a similar manner, which would result in dimeric EcPutA.

1.5.2.b.3. Proline catabolic activity

When proline levels are high, trifunctional EcPutA acts as proline catabolic enzyme. It binds to the membrane to perform its proline catabolic activity. Amino acid sequence analysis clearly shows that the PRODH domain is located within residue 228-572 and the

P5CDH domain are located near 680-1110.The structure of isolated PRODH domain of EcPutA with inhibitor bound has been determined (40, 41). The structure of P5CDH domain from *E. coli* has not been determined. The pairwise sequence identity of the P5CDH domain of EcPutA with TtP5CDH and P5CDH domain of BjPutA is ~ 38 and 52% respectively. The residues ~ 660-1110 of EcPutA corresponds to the P5CDH domain. EcPutA also have an extra C-terminal domain not found in bifunctional BjPutA. The other unique features of EcPutA are described in details in chapter 2. Following paragraph briefly discusses how the PRODH domain interacts with the substrate. This result was obtained by solving crystal structure of L-THFA bound structure of isolated PRODH domain (PutA86-669).

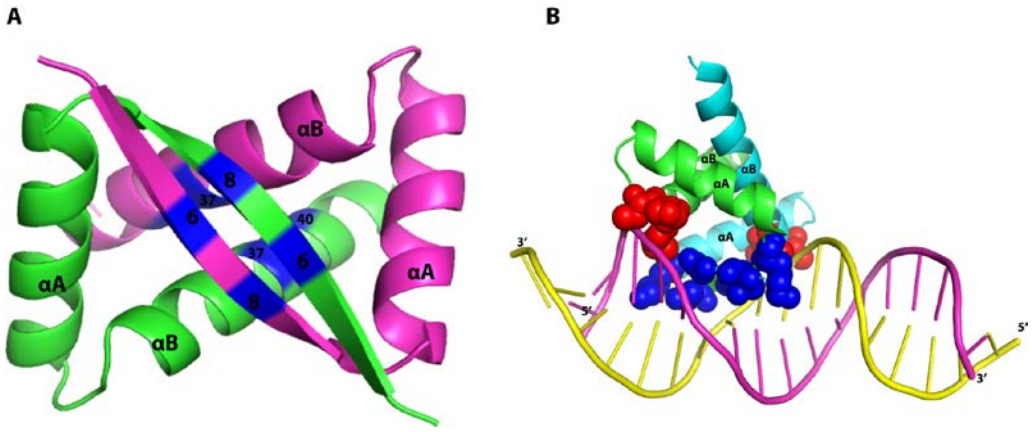


Figure 1.8. A) PutA1-52 dimer. The patches in blue cover most of the surface area buried in the dimer interface. (B) PutA52 bound to DNA. The red spheres are the residues electrostatically interacting with the sugar phosphate backbone. The blue spheres are the residues interacting with the DNA bases.

1.5.2.b.4. PRODH domain (PutA 86-669) complexed with L-THFA

Structure of the EcPutA PRODH domain was solved with the inhibitor bound to the

active site (41) (Figures. 1.9, 1.10). The best represented structure of inhibitor bound PRODH is PutA86-669 bound to THFA. THFA is a very close analog of the substrate proline and therefore the THFA bound structure is best way to know how proline interacts with the active site. Similar to monofunctional TtPRODH and the BjPutA PRODH domain, the PRODH domain of EcPutA is also a distorted $(\beta\alpha)_8$ barrel. In addition, the FAD binds to the C-terminal end of the strands of the barrel in similar fashion. The conformation of FAD is similar to the FAD conformation in BjPutA. The region 188-241 is disordered in the crystal structure of PutA86-669. The purified protein appeared to be a monomer. The inhibitor bound structure suggests that residues of alpha-helix 8 are critical for substrate recognition as it contains a conserved Arg555-Arg556 motif that is shown to interact the carboxylate group of L-THFA. The inhibitor is completely buried within the enzyme, suggesting some flexibility in the protein is required to allow the inhibitor to enter the active site. Furthermore, the structure of inhibitor free TtPRODH discussed earlier is an open and highly solvent exposed (position of helix 8 in the structure is significantly responsible for open structure of TtPRODH) structure.

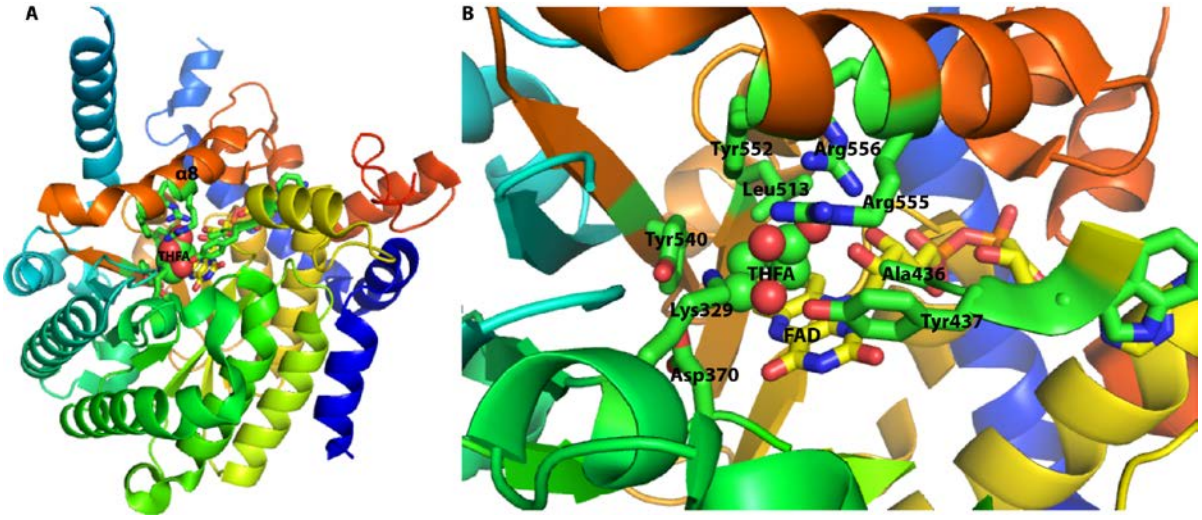


Figure 1.9: The overall fold of PutA86-669 complexed with THFA. Some parts of the complete crystal structure are omitted for clarity. The buried THFA is shown in spheres and the protein is shown in rainbow coloring. The FAD is shown in sticks and colored yellow.

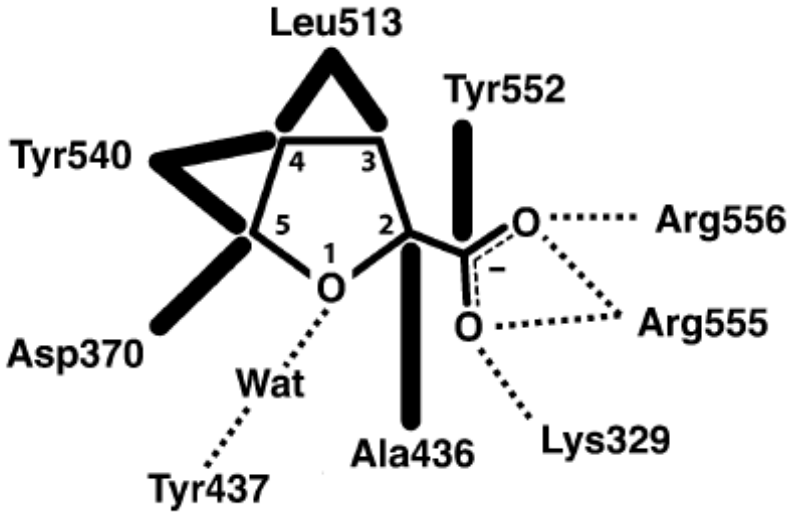


Figure 1.10 Interaction of L-THFA with PutA86-669. Dotted lines represent hydrogen bonds and ion pairs. Thick solid lines indicate van der Waals interactions. This figure was taken from Zhang *et. al*. 2004 (41).

1.6. Summary of Thesis Research

Although we know that PutA switches its function from DNA binding domain to membrane binding domain depending on the level of the proline, a major outstanding challenge is to understand what kind of conformational changes trigger the protein to

switch its function. The lack of a crystal structure of full-length trifunctional PutA impedes this area of research. However, the crystal structure of the N-propargylglycine (NPPG) inactivated PRODH domain of EcPutA has shed some light on the functional switching event of the trifunctional PutA (43). The structure of the NPPG-inactivated PRODH domain identified important interactions between PutA residues and the FAD that appear to be critical for reductive activation of membrane binding. Figure 1.11 shows the probable mechanism of irreversible reduction of FAD with NPPG. A major challenge is to understand how the structural changes originated from FAD reduction in the PRODH domain are transmitted to the overall protein, which affects the membrane binding activity. This question could be answered if we have the structural information of full length EcPutA in oxidized and reduced form. The associated conformational change would be visible when comparing the two structures. The research described in this thesis is a major step toward solving the aforementioned problem. Innumerable attempts to crystallize full-length trifunctional PutAs have been done, but failed to produce crystals that are suitable for structure determination. We therefore utilized small angle X-ray scattering (SAXS) to obtain the overall shape of the protein. In addition, rigid body modeling of full length PutA has been done with the help of SAXS data and crystal structures of DNA-binding and PRODH domains of EcPutA, and a homology model from BjPutA structure.

Unique structural features of PutA have also been explored through multiple sequence alignments and homology modeling using the webservers like ClustalW, Espript, Phyre, and Swiss Model. The results obtained from the sequence alignment study were interesting. It led us to work on oligomeric state of other bifunctional PutAs from

branchs 1 and 2. One of them from branch 1 (RpPutA) is 70% identical to BjPutA in amino acid sequence while the other two are from branch 2 and have very low sequence similarity to BjPutA. The oligomeric states of these proteins have been determined with multi-angle light scattering (MALS) and SAXS. In addition, the oligomeric states of monofunctional P5CDHs have been studied with MALS and SAXS.

Finally, the structural basis of HPII disease was determined through X-ray crystallographic studies of human P5CDH (HsP5CDH). In particular, the structure of the disease-associated S352L variant was determined.

Altogether, the research described in this dissertation is a stepping stone in the field of proline catabolism in various organisms that will surely lead to advancement in understanding many aspects of proline catabolism.

1.7. References

1. Phang JM. The regulatory functions of proline and pyrroline-5-carboxylic acid. Curr Top Cell Reg. 1985;25:92-132.

2. Donald SP, Sun XY, Hu CA, Yu J, Mei JM, Valle D, et al. Proline oxidase, encoded by p53-induced gene-6, catalyzes the generation of proline-dependent reactive oxygen species. Cancer Res. 2001;61(5):1810-5.

3. Phang JM, Hu CA, Valle D. Disorders of proline and hydroxyproline metabolism. In: Scriver CR, Beaudet AL, Sly WS, Valle D, editors. Metabolic and molecular basis of inherited disease. New York: McGraw Hill; 2001. p. 1821-38.

4. Pandhare J, Cooper SK, Phang JM. Proline oxidase, a proapoptotic gene, is induced by troglitazone: evidence for both peroxisome proliferator-activated receptor gamma-dependent and -independent mechanisms. J Biol Chem. 2006;281(4):2044-52.

5. Singh RK, Larson JD, Zhu W, Rambo RP, Hura GL, Becker DF, et al. Small-angle X-ray Scattering Studies of the Oligomeric State and Quaternary Structure of the Trifunctional Proline Utilization A (PutA) Flavoprotein from Escherichia coli. Journal of Biological Chemistry. 2011;286(50):43144-53.

6. Mollica F, Pavone L. HYPERPROLINAEMIA: A DISEASE WHICH DOES NOT NEED TREATMENT? Acta Pædiatrica. 1976;65(2):206-8.

7. Mitsubuchi H, Nakamura K, Matsumoto S, Endo F. Inborn Errors of Proline Metabolism. The Journal of Nutrition. 2008;138(10):2016S-20S.

8. Jacquet H, Berthelot J, Bonnemains C, Simard G, Saugier-Veber P, Raux G, et al. The severe form of type I hyperprolinaemia results from homozygous inactivation of the PRODH gene. Journal of Medical Genetics. 2003;40(1):e7.

9. Jacquet H, Demily C, Houy E, Hecketsweiler B, Bou J, Raux G, et al. Hyperprolinemia is a risk factor for schizoaffective disorder. Mol Psychiatry. 2005;10(5):479-85. Epub 2004/10/21.

10. Woody NC, Snyder CH, Harris JA. HYPERPROLINEMIA: CLINICAL AND BIOCHEMICAL FAMILY STUDY. Pediatrics. 1969;44(4):554-63.

11. Campbell HD, Webb GC, Young IG. A human homologue of the <i>Drosophila melanogaster sluggish-A (proline oxidase) gene maps to 22q11.2, and is a candidate gene for type-I hyperprolinaemia. Human Genetics. 1997;101(1):69-74.

12. Geraghty MT, Vaughn D, Nicholson AJ, Lin WW, Jimenez-Sanchez G, Obie C, et al. Mutations in the Delta1-pyrroline 5-carboxylate dehydrogenase gene cause type II hyperprolinemia. Hum Mol Genet. 1998;7(9):1411-5.

13. Valle D, Goodman SI, Applegarth DA, Shih VE, Phang JM. Type II hyperprolinemia. Delta1-pyrroline-5-carboxylic acid dehydrogenase deficiency in cultured skin fibroblasts and circulating lymphocytes. J Clin Invest. 1976;58(3):598-603.

14. Ratzkin B, Roth J. Cluster of genes controlling proline degradation in Salmonella typhimurium. J Bacteriol. 1978;133(2):744-54.

15. Tanner JJ. Structural biology of proline catabolism. Amino Acids. 2008;35(4):719-30.

16. Downing SJ, Phang JM, Kowaloff EM, Valle D, Smith RJ. Proline oxidase in cultured mammalian cells. J Cell Physiol. 1977;91(3):369-76.

17. Kowaloff EM, Phang JM, Granger AS, Downing SJ. Regulation of proline oxidase activity by lactate. Proc Natl Acad Sci U S A. 1977;74(12):5368-71.

18. Hu CA, Donald SP, Yu J, Lin WW, Liu Z, Steel G, et al. Overexpression of proline oxidase induces proline-dependent and mitochondria-mediated apoptosis. Mol Cell Biochem. 2006.

19. Liu Y, Borchert GL, Surazynski A, Hu CA, Phang JM. Proline oxidase activates both intrinsic and extrinsic pathways for apoptosis: the role of ROS/superoxides, NFAT and MEK/ERK signaling. Oncogene. 2006;25(41):5640-7.

20. Cooper SK, Pandhare J, Donald SP, Phang JM. A novel function for hydroxyproline oxidase in apoptosis through generation of reactive oxygen species. J Biol Chem. 2008;283(16):10485-92.

21. Phang JM, Donald SP, Pandhare J, Liu Y. The metabolism of proline, a stress substrate, modulates carcinogenic pathways. Amino Acids. 2008;35(4):681-90. Epub 2008/04/11.

22. Phang JM, Pandhare J, Liu Y. The metabolism of proline as microenvironmental stress substrate. J Nutr. 2008;138(10):2008S-15S. Epub 2008/09/23.

23. Yoshida A, Rzhetsky A, Hsu LC, Chang C. Human aldehyde dehydrogenase gene family. Eur J Biochem. 1998;251(3):549-57. Epub 1998/03/07.

24. Yoon KA, Nakamura Y, Arakawa H. Identification of ALDH4 as a p53-inducible gene and its protective role in cellular stresses. J Hum Genet. 2004;49(3):134-40. Epub 2004/02/27.

25. Forte-McRobbie C, Pietruszko R. Human glutamic-gamma-semialdehyde dehydrogenase. Kinetic mechanism. Biochem J. 1989;261(3):935-43. Epub 1989/08/01.

26. Forte-McRobbie CM, Pietruszko R. Purification and characterization of human liver "high Km" aldehyde dehydrogenase and its identification as glutamic gamma-semialdehyde dehydrogenase. J Biol Chem. 1986;261(5):2154-63.

27. White TA, Krishnan N, Becker DF, Tanner JJ. Structure and kinetics of monofunctional proline dehydrogenase from Thermus thermophilus. J Biol Chem. 2007;282(19):14316-27.

28. Krishnan N, Becker DF. Characterization of a bifunctional PutA homologue from Bradyrhizobium japonicum and identification of an active site residue that modulates proline reduction of the flavin adenine dinucleotide cofactor. Biochemistry. 2005;44(25):9130-9.

29. Schuermann JP, White TA, Srivastava D, Karr DB, Tanner JJ. Three crystal forms of the bifunctional enzyme proline utilization A (PutA) from Bradyrhizobium japonicum. Acta Cryst. 2008;F64(Pt 10):949-53.

30. Straub PF, Reynolds PH, Althomsons S, Mett V, Zhu Y, Shearer G, et al. Isolation, DNA sequence analysis, and mutagenesis of a proline dehydrogenase gene (putA) from Bradyrhizobium japonicum. Appl Environ Microbiol. 1996;62(1):221-9.

31. Krishnan N, Doster AR, Duhamel GE, Becker DF. Characterization of a Helicobacter hepaticus putA mutant strain in host colonization and oxidative stress. Infect Immun. 2008;76(7):3037-44.

32. Krishnan N, Becker DF. Oxygen Reactivity of PutA from Helicobacter Species and Proline-Linked Oxidative Stress. J Bacteriol. 2006;188(4):1227-35.

33. Nakajima K, Inatsu S, Mizote T, Nagata Y, Aoyama K, Fukuda Y, et al. Possible involvement of put A gene in Helicobacter pylori colonization in the stomach and motility. Biomed Res. 2008;29(1):9-18. Epub 2008/03/18.

34. White TA, Tanner JJ. Cloning, purification and crystallization of Thermus thermophilus proline dehydrogenase. Acta Cryst. 2005;F61(Pt 8):737-9.

35. White TA, Johnson WH, Jr., Whitman CP, Tanner JJ. Structural basis for the inactivation of Thermus thermophilus proline dehydrogenase by N-propargylglycine. Biochemistry. 2008;47(20):5573-80. Epub 2008/04/23.

36. Inagaki E, Ohshima N, Takahashi H, Kuroishi C, Yokoyama S, Tahirov TH. Crystal structure of Thermus thermophilus Delta1-pyrroline-5-carboxylate dehydrogenase. J Mol Biol. 2006;362(3):490-501.

37. Inagaki E, Ohshima N, Sakamoto K, Babayeva ND, Kato H, Yokoyama S, et al. New insights into the binding mode of coenzymes: structure of Thermus thermophilus [Delta]1-pyrroline-5-carboxylate dehydrogenase complexed with NADP+. Acta Cryst. 2007;F63(6 %R doi:10.1107/S1744309107021422):462-5.

38. Larson JD, Jenkins JL, Schuermann JP, Zhou Y, Becker DF, Tanner JJ. Crystal structures of the DNA-binding domain of Escherichia coli proline utilization A flavoprotein and analysis of the role of Lys9 in DNA recognition. Protein Sci. 2006;15:1-12.

39. Zhou Y, Larson JD, Bottoms CA, Arturo EC, Henzl MT, Jenkins JL, et al. Structural basis of the transcriptional regulation of the proline utilization regulon by multifunctional PutA. J Mol Biol. 2008;381(1):174-88.

40. Lee YH, Nadaraia S, Gu D, Becker DF, Tanner JJ. Structure of the proline dehydrogenase domain of the multifunctional PutA flavoprotein. Nat Struct Biol. 2003;10(2):109-14.

41. Zhang M, White TA, Schuermann JP, Baban BA, Becker DF, Tanner JJ. Structures of the Escherichia coli PutA proline dehydrogenase domain in complex with competitive inhibitors. Biochemistry. 2004;43(39):12539-48.

42. Ostrander EL, Larson JD, Schuermann JP, Tanner JJ. A conserved active site tyrosine residue of proline dehydrogenase helps enforce the preference for proline over hydroxyproline as the substrate. Biochemistry. 2009;48(5):951-9. Epub 2009/01/15.

43. Srivastava D, Zhu W, Johnson WH, Jr., Whitman CP, Becker DF, Tanner JJ. The structure of the proline utilization a proline dehydrogenase domain inactivated by N-propargylglycine provides insight into conformational changes induced by substrate binding and flavin reduction. Biochemistry. 2010;49(3):560-9. Epub 2009/12/10.

44. Srivastava D, Schuermann JP, White TA, Krishnan N, Sanyal N, Hura GL, et al. Crystal structure of the bifunctional proline utilization A flavoenzyme from Bradyrhizobium japonicum. Proc Natl Acad Sci U S A. 2010;107(7):2878-83. Epub 2010/02/06.

45. Wierenga RK. The TIM-barrel fold: a versatile framework for efficient enzymes. FEBS Letters. 2001;492(3):193-8.

46. Banner DW, Bloomer AC, Petsko GA, Phillips DC, Wilson IA. Atomic coordinates for triose phosphate isomerase from chicken muscle. Biochemical and Biophysical Research Communications. 1976;72(1):146-55.

47. Zhang W, Zhou Y, Becker DF. Regulation of PutA-membrane associations by flavin adenine dinucleotide reduction. Biochemistry. 2004;43(41):13165-74.

48. Singh RK, Tanner JJ. Unique structural features and sequence motifs of proline utilization A (PutA). Frontiers in bioscience : a journal and virtual library. 2012;17:556-68. Epub 2011/12/29.

49. Jeffery CJ. Moonlighting proteins. Trends in Biochemical Sciences. 1999;24(1):8-11.

50. Piatigorsky J, O'Brien WE, Norman BL, Kalumuck K, Wistow GJ, Borras T, et al. Gene sharing by delta-crystallin and argininosuccinate lyase. Proceedings of the National Academy of Sciences. 1988;85(10):3479-83.

51. Jeffery CJ. Moonlighting proteins: old proteins learning new tricks. Trends in Genetics. 2003;19(8):415-7.

52. Jeffery CJ. Mass spectrometry and the search for moonlighting proteins. Mass Spectrometry Reviews. 2005;24(6):772-82.

53. Muro-Pastor AM, Ostrovsky P, Maloy S. Regulation of Gene Expression by Repressor Localization: Biochemical Evidence that Membrane and DNA Binding by the PutA Protein are Mutually Exclusive. J Bacteriol. 1997;179(8):2788-91.

54. Meyer-Siegler K, Mauro DJ, Seal G, Wurzer J, deRiel JK, Sirover MA. A human nuclear uracil DNA glycosylase is the 37-kDa subunit of glyceraldehyde-3-phosphate dehydrogenase. Proceedings of the National Academy of Sciences. 1991;88(19):8460-4.

55. Ostrovsky de Spicer P, Maloy S. PutA protein, a membrane-associated flavin dehydrogenase, acts as a redox-dependent transcriptional regulator. Proc Natl Acad Sci USA. 1993;90(9):4295-8.

56. Zhu W, Becker DF. Flavin redox state triggers conformational changes in the PutA protein from Escherichia coli. Biochemistry. 2003;42(18):5469-77.

57. Brown ED, Wood JM. Conformational change and membrane association of the PutA protein are coincident with reduction of its FAD cofactor by proline. J Biol Chem. 1993;268(12):8972-9.

58. Wood JM. Membrane association of proline dehydrogenase in Escherichia coli is redox dependent. Proc Natl Acad Sci USA. 1987;84(2):373-7. Epub 1987/01/01.

59. Becker DF, Thomas EA. Redox properties of the PutA protein from Escherichia coli and the influence of the flavin redox state on PutA-DNA interactions. Biochemistry. 2001;40(15):4714-21.

60. Zhang W, Zhang M, Zhu W, Zhou Y, Wanduragala S, Rewinkel D, et al. Redox-induced changes in flavin structure and roles of flavin N(5) and the ribityl 2'-OH group in regulating PutA--membrane binding. Biochemistry. 2007;46(2):483-91.

61. Schreiter ER, Drennan CL. Ribbon-helix-helix transcription factors: variations on a theme. Nat Rev Microbiol. 2007;5(9):710-20.

62. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. Journal of Molecular Biology. 2007;372(3):774-97.

CHAPTER 2

Unique structural features and sequence motifs of proline utilization A

This chapter has been adapted from
Ranjan K. Singh and John J. Tanner
Front. in Biosci. (2012) 17 556-568.

My Contribution to the paper

Sequence alignments of branch 1 and branch 2 PutAs

Abstract

Proline utilization A proteins (PutAs) are bifunctional enzymes that catalyze the oxidation of proline to glutamate using spatially separated proline dehydrogenase and pyrroline-5-carboxylate dehydrogenase active sites. Here we use the 2.1 Å crystal structure of the minimalist PutA from *Bradyrhizobium japonicum* (BjPutA) along with amino acid sequence analysis to identify unique structural features of PutAs. This analysis shows that PutAs are not simply fusions of the related monofunctional PRODH and P5CDH enzymes, but rather have secondary structural elements and domains not found in the separate enzymes. Some of these extra PutA-specific features are predicted to be necessary for substrate channeling in PutA. Conserved elements of PutAs are also described. Multiple sequence alignment analysis shows that some PutAs have a 17-residue conserved motif located in the C-terminal 20-30 residues of the polypeptide chain. The BjPutA structure shows that this motif participates in domain-swapped dimerization and helps to seal the internal substrate-channeling cavity from the bulk medium. Finally, it is shown that some PutAs have a 100-200 residue domain of unknown function in the C-terminus that is not found in minimalist PutAs. Remote homology detection suggests that this domain is homologous to the oligomerization beta-hairpin and Rossmann fold domain of BjPutA. This result implies the novel hypothesis that some PutAs have a second NAD^+-binding domain. Whether this domain actually binds NAD^+ or is a pseudo-domain that plays a structural role remains to be determined.

2.1. Introduction

The oxidation of proline to glutamate, i.e. proline catabolism, is catalyzed by two enzymes, proline dehydrogenase (PRODH) and pyrroline-5-carboxylate (P5C) dehydrogenase (P5CDH) (Figure 2.1). The former catalyzes the oxidation of L-proline to P5C with concomitant reduction of an enzyme-bound FAD cofactor. The latter enzyme catalyzes the oxidation of L-glutamate semi-aldehyde (GSA) to L-glutamate using NAD^+ as the electron acceptor. Note that the product of the PRODH reaction is not the substrate for P5CDH. Instead, the two reactions are coupled by the hydrolysis of P5C, which has traditionally been thought of as a nonenzymatic process. As noted by Phang over two decades ago, researchers typically refer to P5C and GSA interchangeably, since the two species are not distinguished in most experiments (1). Phang's observation remains valid today.

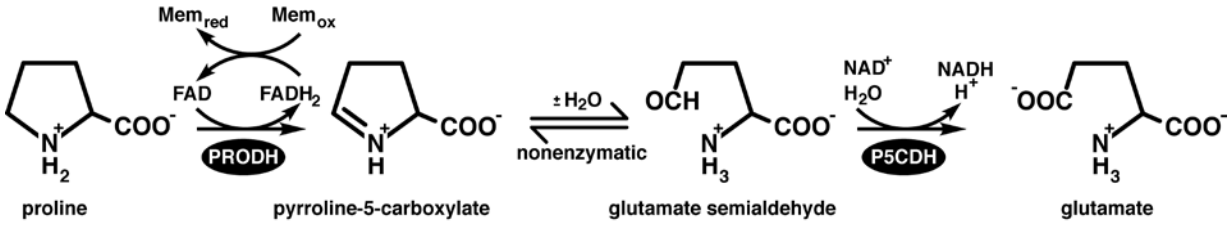


Figure 2.1. The reactions of proline catabolism.

One interesting feature of proline catabolism is that PRODH and P5CDH are combined into a single polypeptide in some organisms (Figure 2.2). The combined enzymes are known as proline utilization A (PutA) and were discovered by Roth's group in the late 1970s during their studies of proline utilization in *Salmonella typhimurium* (2).

Analysis of genome sequence data suggests that PutAs are limited to Gram-negative bacteria (Figure 2.2, branches 1, 2), whereas PRODH and P5CDH are separate enzymes encoded by distinct genes in Gram-positive bacteria (branch 3B) (3). In eukaryotes, PRODH and P5CDH are also separate enzymes and are localized to mitochondria (branch 3A). Human PRODH is a p53-induced tumor suppressor protein localized to the inner mitochondrial membrane and is often referred to as POX to emphasize its role as a superoxide-generating oxidase (4-12). Human P5CDH (ALDH4 (13)) is also induced by p53 (14) and is located in the mitochondrial matrix. ALDH4 has been characterized biochemically, including elucidation of the oligomeric state in solution (dimer) and kinetic mechanism (15, 16).

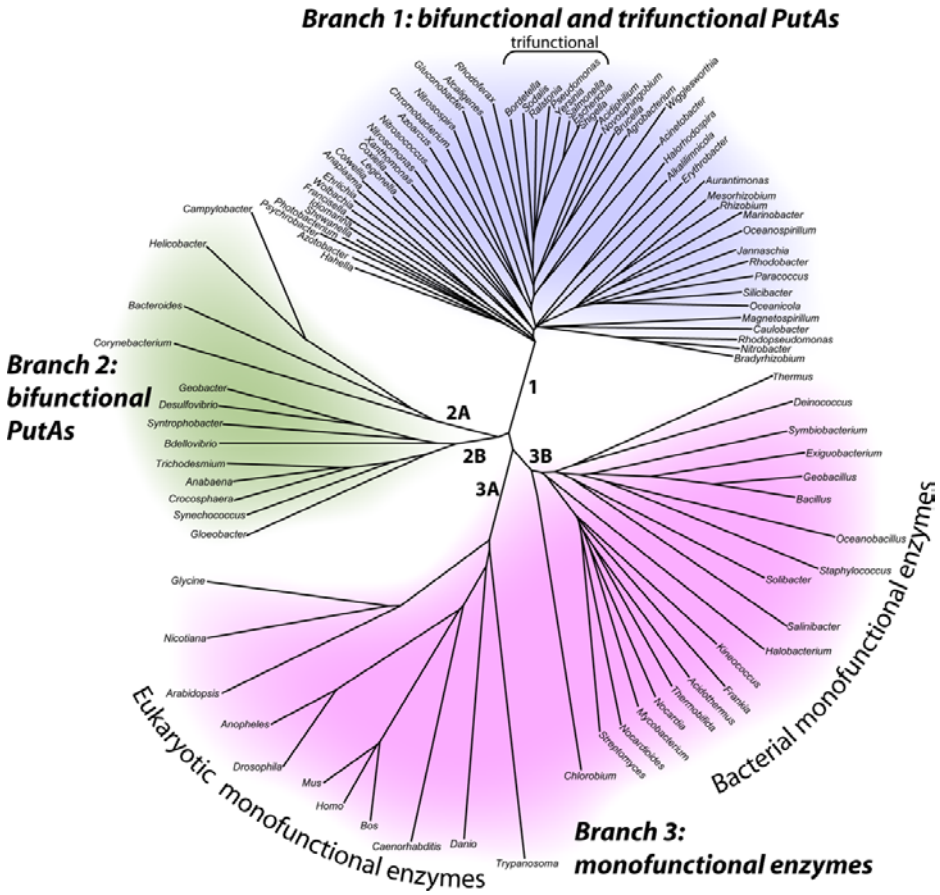


Figure 2.2. Phylogenetic tree representing the organization of proline catabolic enzymes in bacteria and eukaryotes. PutAs are found in branches 1 and 2. Monofunctional PRODH and P5CDH enzymes are found in branch 3. A cluster of trifunctional PutAs is indicated.

The PutA part of the PutA/PRODH/P5CDH family tree has two branches (3, 17). Branch 1 primarily consists of PutAs from alpha-, beta-, and gamma-proteobacteria. Branch 2 includes PutAs from delta- and epsilon-proteobacteria as well as cyanobacteria. The PutAs in branch 1 have chain lengths from 999 to almost 1400 residues, and the pairwise sequence identities are greater than 38 %. The polypeptide length for branch 2 PutAs ranges from around 980 to almost 1300 residues, and the pairwise sequence identity range can be as low as 23 %. Thus, branch 2 PutAs appear to be a more diverse group than branch 1 PutAs. Between branches 1 and 2, the pairwise sequence identities are typically less than 30 %. Nevertheless, the residues in the PRODH and P5CDH active sites are highly conserved, indicating that the three-dimensional structures of the catalytic domains are conserved by PutAs. Whether the three-dimensional arrangement of the catalytic and other domains is likewise conserved remains to be determined.

PutAs are further classified as bifunctional or trifunctional. Bifunctional PutAs exhibit only PRODH and P5CDH catalytic activities, have polypeptide chain lengths in the range of ~980 residues to over 1300 residues, and are found in both PutA branches. Bifunctional PutAs from *Bradyrhizobium japonicum* (BjPutA, (18-20)) and *Helicobacter* species (21-23) have been studied. Trifunctional PutAs constitute a subset of branch 1 PutAs and are distinguished by the presence of a DNA-binding domain (a ribbon-helix-helix domain) in the first ~50 residues of the polypeptide chain. The polypeptide chain length of trifunctional PutAs are in the range of ~1270-1361. In addition to functioning as dual PRODH/P5CDH enzymes, trifunctional PutAs have a third function of repressing transcription of the *put* regulon, which contains the genes encoding PutA and the proline transporter PutP, when proline levels are low (24-27).

High levels of proline in the bacterium's environment cause PutA to disengage from the *put* control region thus activating transcription of *putA* and *putP*. Thus, trifunctional PutAs are remarkable proteins that link transcription and metabolism in response to an environmental cue (proline level). Trifunctional PutAs from *S. typhimurium* (25, 26, 28-31) and *Escherichia coli* (EcPutA) (24, 32-46) have been studied. PutA from *E. coli* is the most studied trifunctional PutA and is considered to be the archetypal trifunctional PutA.

The observation that enzymes catalyzing successive reactions in a metabolic pathway are combined into a single polypeptide chain as in PutA has intriguing implications. First, the covalent linking of the two active sites may allow the transfer of the reaction product of one enzyme to the next without equilibrating with the bulk medium. Substrate channeling is the term used for such kinetic mechanisms, and Arentson *et al*. 2011 provide a review of substrate channeling in proline metabolism (47). Two limiting channeling mechanisms are possible: direct transfer and proximity. In the former, the intermediate moves through an internal cavity or tunnel connecting the two active sites without leaving the confines of the protein. Proximity refers to a spectrum of cases in which the reaction product dissociates from the first enzyme but encounters a locally high concentration of the second enzyme and thus does not truly equilibrate with the bulk medium. There are potential advantages of substrate channeling, including decreased transit time between enzymes, protection of labile intermediates, and isolation of intermediates from competing enzymatic reactions (48-50). The latter point is relevant for proline metabolism, because P5C/GSA is common to proline catabolism, proline biosynthesis, and arginine biosynthesis. Another implication of fused enzymes was

recognized by Eisenberg's group in 1999 and concerns the prediction of protein-protein interactions (51). The basic idea is that the observation that two proteins are separate in some organisms and fused in others implies that the two separate proteins form a functional association. Thus, protein fusion (or gene fusion) provides the 'Rosetta stone' for identifying potential protein–protein interactions. In proline catabolism, for example, the Rosetta Stone hypothesis suggests the possibility that monofunctional PRODH and P5CDH interact and engage in intermolecular substrate channeling.

Three-dimensional structural studies have contributed to our understanding of proline catabolic proteins. Crystal structures have been solved for the monofunctional PRODH and P5CDH enzymes from *Thermus thermophilus* (TtPRODH (17, 52, 53), TtP5CDH (54, 55)), the DNA-binding domains of two trifunctional PutAs (56-58), a PRODH domain construct of EcPutA (EcPutA86-630, (46, 59-62)), and full-length BjPutA (63). Here, we use the BjPutA structure as a platform for identifying the unique features that distinguish PutAs from their monofunctional relatives and to gain insight into the structure and function of the C-terminal domains of PutAs.

2.2. Research Background

2.2.1. Comparision of a bifunctional PutA with monofunctional PRODH and P5CDH

2.2.1.a. Structure of a minimalist PutA

PutA from *Bradyrhizobium japonicum* (BjPutA) is one of the simplest of bifunctional PutAs. At 999 residues in length, it is the shortest branch 1 PutA known and is thus considered to be a minimalist PutA. We recently reported the crystal structure of

BjPutA (PDB code 3haz), which is the first structure of a full-length PutA. The structure shows that PutAs are more than simply a fusion of two catalytic domains. The protomer comprises seven domains: arm, alpha, PRODH barrel, linker, NAD^+-binding, P5CDH catalytic, and oligomerization domain (Figure 2.3). The PRODH active site is located in a distorted $(beta\text{-}alpha)_8$ barrel. The barrel structure is very similar to that of EcPutA (59). The P5CDH active site is located in the interface between the NAD^+-binding domain and the P5CDH catalytic domain. The two active sites are separated by 41 Å and connected by a large, irregularly shaped internal cavity (silver surface in Figure 2.3A). This cavity most likely functions in substrate channeling, as described in Arentson *et al* (47).

Oligomerization is essential for substrate channeling in BjPutA. The enzyme forms a U-shaped dimer (Figure 2.3B), and two of these dimers assemble into a ring-shaped dimer-of-dimers tetramer (see reference (63)). The dimer shown in Figure 2.3B is the relevant oligomeric state for understanding substrate channeling, so we will describe it in detail here. Dimerization is mediated by an oligomerization domain (orange in Figure 2.3A) that protrudes from the NAD^+-binding domain. The oligomerization domain consists of two elements that are far apart in sequence, a beta-hairpin formed by residues 629-647 and a beta-strand followed by a short helix at the C-terminus of polypeptide chain (residues 976-989). The C-terminal strand of the oligomerization domain of one protomer forms main chain hydrogen bonds with the beta-sheet of the P5CDH catalytic domain of the other protomer to form a large, twisted intermolecular beta-sheet (Figures 2.3B and 2.3C). This type of oligomerization is an example of domain swapping (64). Both the bipartite oligomerization domain and the domain-swapped dimer are conserved

by aldehyde dehydrogenases. In BjPutA, the oligomerization domain not only mediates dimerization but also covers the cavity of the other protomer (Figure 2.3C). Without the lid provided by the oligomerization domain, the cavity would be open to the bulk medium. Thus dimerization appears to be essential for formation of the substrate channeling cavity. The sealed, internal cavity is consistent with direct transfer rather than proximity as the predominant mechanism of substrate channeling in BjPutA.

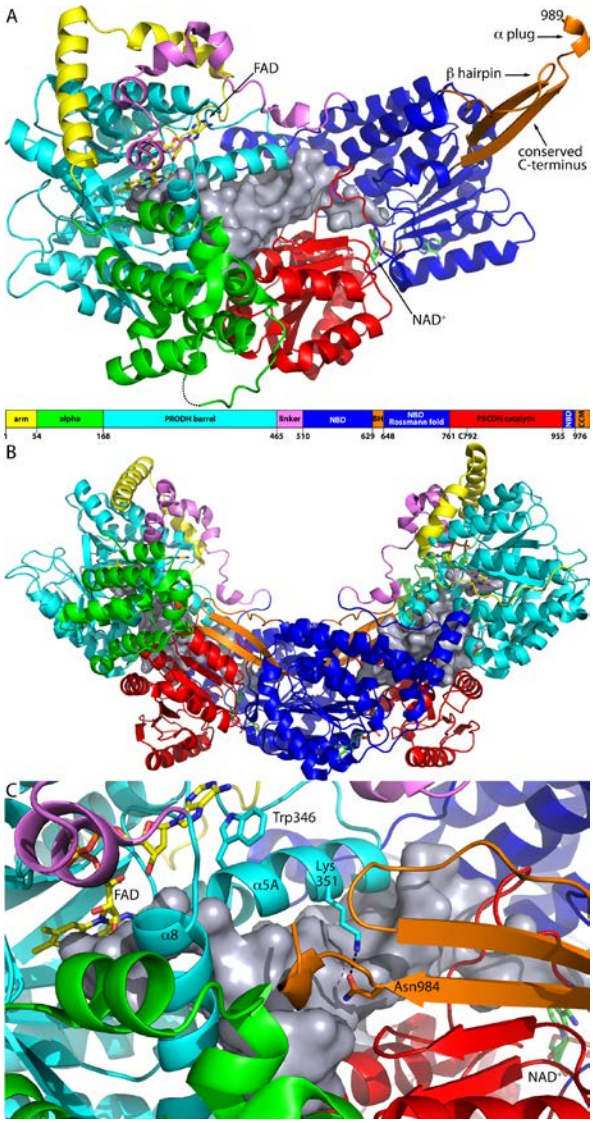


Figure 2.3. Structure of BjPutA. (A) Protomer structure with the domains colored according to the domain diagram. The silver surface represents the substrate-channeling cavity. FAD and NAD^+ are represented as yellow and green sticks, respectively. Abbreviations used in the domain diagram are as follows: NBD, NAD^+-binding domain; BH, beta-hairpin; CCM, conserved C-terminal motif. (B) The domain-swapped dimer of BjPutA. The domains are colored according to the domain diagram in panel A. The silver surfaces represent the substrate-channeling cavities of the two protomers. (C) Close-up view of the oligomerization domain covering the cavity of the other protomer. This figure and others were prepared with PyMol (65).

2.2.1.b. The PutA PRODH barrel has an extra helix involved in FAD binding and substrate channeling

Both the PutA PRODH domain and the monofunctional enzyme TtPRODH exhibit the distorted (beta-alpha)$_8$ barrel that is characteristic of the PRODH/PutA family. As shown in Figure 2.4, the two folds are nearly identical; the RMSD between the two proteins is 2.0 Å for 256 aligned residues. In both structures, the FAD is bound at the C-terminal ends of the strands of the barrel. The fold is unusual in that the final helix, alpha8, sits atop the barrel rather than alongside beta8 as in the classical triosephosphate isomerase barrel. Helix 8 contains residues that are critical for substrate recognition, including a conserved Arg-Arg motif that has been shown to bind the substrate carboxylate group (60). Both PutA and the monofunctional PRODH exhibit this distortion from the classical (beta-alpha)$_8$ barrel fold.

Despite the similarities in overall fold and amino acid sequence (28 % identity), there is an important difference between monofunctional PRODH and PutA. The PutA PRODH barrel has an extra helix (alpha5a) inserted between beta5 and alpha5 (Figure 2.4). This additional secondary structural element is also present in EcPutA86-630, suggesting that it conserved in branch 1 PutAs. Helix 5a contains a conserved tryptophan residue (Trp346 in BjPutA, Trp438 in EcPutA) that stacks against the FAD adenine (Figures 2.3C and 2.4), and this interaction presumably contributes to the different FAD conformations in PutA and TtPRODH, as described previously (17). The BjPutA structure reveals a new function for helix 5a. The helix forms a large section of the wall of the internal substrate-channeling cavity (Figure 2.3C), and its absence would leave a large hole to the bulk medium. Thus, helix 5a seems to be essential for

channeling in PutA.

Finally, we note that TtPRODH has three additional helices that precede the start of the barrel (alphaA, alphaB, alphaC in Figure 2.4). These helices are located in the vicinity of the alpha domain of BjPutA, but they do not superimpose well with any of the helices of the alpha domain.

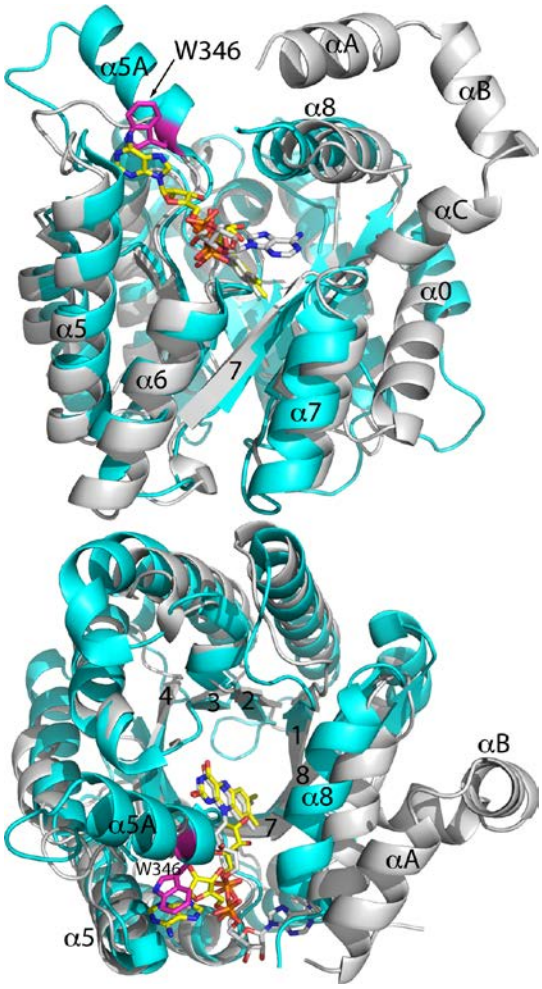


Figure 2.4. Comparison of the monofunctional enzyme TtPRODH (white) and the PRODH barrel of BjPutA (cyan). Strands of the barrel are labeled 1-8 using the standard convention for (beta-alpha)$_8$ barrels. The extra helix of the PutA barrel is denoted alpha5a. Trp346 of alpha5a is colored magenta. The three helices of TtPRODH that precede the barrel are labeled alphaA, alphaB, and alphaC. Two orthogonal views are shown.

2.2.1.c. An extra helix at the C-terminus of PutA plugs the substrate-channeling cavity

The structure of the monofunctional enzyme TtP5CDH is very similar to that of the P5CDH half of BjPutA (Figure 2.5). The sequence identity of TtP5CDH to PutA is about 38 %, implying substantial structural similarity. Indeed, the RMSD between the two proteins is 1.4 Å for 463 residues. Both enzymes exhibit the characteristic aldehyde dehydrogenase fold (66), which consists of three domains, NAD^+-binding, catalytic, and oligomerization. (The oligomerization is also referred to as the bridging domain (66) and beta-flap (63)). The NAD^+-binding domain is structurally contiguous but actually consists of three separate sections of the polypeptide chain, as indicated in the domain diagram in Figure 2.3A. Sections 1 and 2 are separated in primary structure by the beta-hairpin of the oligomerization domain, while sections 2 and 3 are separated by the P5CDH catalytic domain. Section 2 exhibits a variation of the Rossmann dinucleotide binding fold. The classical Rossmann fold consists of two beta-alpha-beta-alpha-beta motifs that form a 6-stranded parallel beta sheet with relative strand order 321456 (67). The Rossmann domain of aldehyde dehydrogenase lacks the final strand and helix and is thus referred to as a non-classical Rossmann fold domain. The catalytic domains of TtP5CDH and BjPutA are also quite similar, as are the details of the P5CDH active sites (63).

Closer inspection of BjPutA and TtP5CDH, however, reveals an important difference at the C-terminus (Figure 2.5). Structure-based alignment of BjPutA and TtP5CDH shows that the BjPutA chain extends 14 residues past the final residue of TtP5CDH (Phe516). Some of these extra residues were resolved in the BjPutA structure, and they

form a turn of helix that plugs a hole in the cavity wall (Figure 2.3C). Without this plug, there would be a significant hole in the cavity leading to the bulk medium. Interestingly, the extra residues of BjPutA are incompatible with the observed oligomeric state of TtP5CDH. TtP5CDH forms a hexamer, which can be thought of as a trimer of domain-swapped dimers. If the TtP5CDH chain were longer as in BjPutA, the C-terminus would clash with another dimer of the hexamer.

2.2.1.d. BjPutA has domains not found in the monofunctional enzymes

PutAs also have extra domains not found in the monofunctional enzymes, and the BjPutA structure exhibits three of them: arm, alpha domain, and linker (Figure 2.3A). The alpha-helical arm at the N-terminus (yellow in Figure 3A) wraps around the PRODH barrel and sits below the linker.

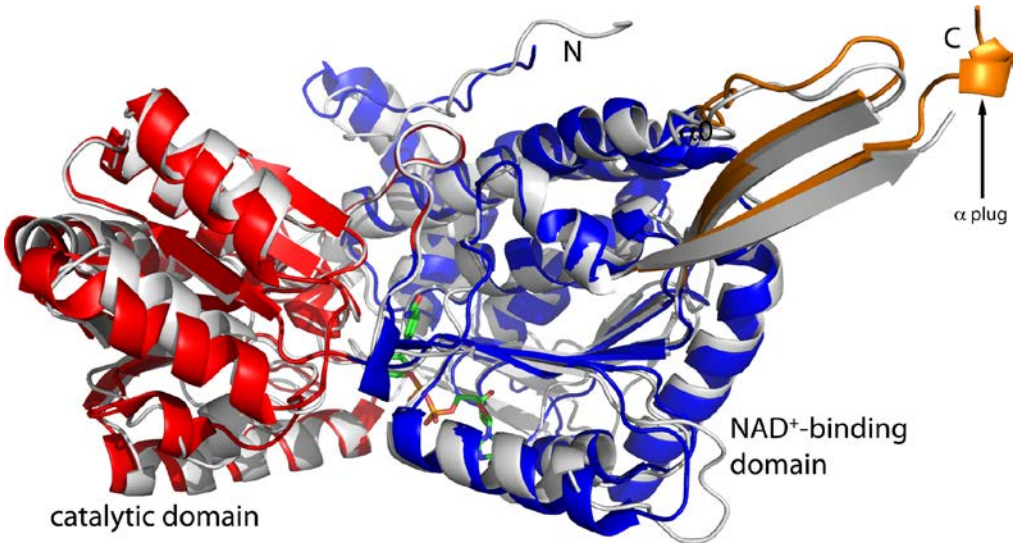


Figure 2.5. Comparison of the monofunctional enzyme TtP5CDH (white) and the P5CDH half of BjPutA. BjPutA is colored according to the legend in Figure 3A. NAD^+ is drawn in green sticks.

The arm domain is observed in both BjPutA and EcPutA86-630, indicating that it is conserved by branch 1 PutAs. The arm connects to the alpha domain, which is a globular domain consisting of 6 helices (green in Figure 2.3A). The alpha domain contacts both the PRODH and P5CDH domains, as well as the oligomerization domain of the other protomer of the domain-swapped dimer (Figures 2.3B and 2.3C). Its strategic location at the confluence of three domains suggests that it is critical for properly orienting the two active sites for channeling and for formation of the internal cavity. We note that the alpha domain was disordered in the structure of EcPutA86-630, indicating that contacts with the P5CDH half of the enzyme are required for proper folding. As noted above, the alphaA, alphaB, and alphaC helices of TtPRODH are in the same general location as the alpha domain of BjPutA. Whether these helices play a similar role as the PutA alpha domain is unknown.

The polypeptide that links the PRODH barrel to the NAD^+-binding domain is also unique to PutA. The linker is not simply a flexible tether that keeps the two catalytic domains in close proximity, but rather has a well-defined structure that appears to be essential for maintaining the tertiary and quaternary structure of the enzyme. The linker (residues 465-509, violet in Figure 2.3) joins the C-terminus of helix 8 of the PRODH barrel to the N-terminus of the NAD^+-binding domain. The 45-residue linker consists of 5 short helical segments that form a meandering U-turn, effectively redirecting the chain toward the P5CDH domain. We note that a nearly identical meandering U-turn is also found in the structure of EcPutA86-630, suggesting that the linker structure is conserved by branch 1 PutAs. Because of its wide, curved path, the linker traverses 100 Å, although the two residues it connects are separated by only 30 Å. The linker forms

extensive interactions with other domains, which likely are essential for maintaining its three-dimensional structure. The majority of these interactions are with the PRODH domain (2000 $Å^2$ of inter-domain buried surface area) and the arm (1300 $Å^2$). The large contact area with the arm reflects the fact that the linker sits atop the arm, essentially tracking its curved path around the barrel. In fact, it is possible that the primary role of the arm is to help stabilize the conformation of the linker. In summary, the linker is a structural element that interacts with disparate parts of the polypeptide chain via noncovalent interactions, and as a result is important for properly orienting the two active sites and creating the substrate channeling cavity.

2.3. The conserved C-terminal motif of branch 1 PutAs

The second element of the oligomerization domain contains a conserved sequence motif, which to our knowledge has not been described previously. Multiple sequence alignment (MSA) analysis reveals a conserved stretch of 17-residues located in the C-terminal 20-30 residues of branch 1 PutAs (Figure 2.6). The motif is found in both bifunctional and trifunctional branch 1 PutAs, but it does not appear to be present in branch 2 PutAs and monofunctional P5CDHs. Based on the sequences analyzed, the consensus motif is Exxxxv[N or D]t[T or A]AaGGnaxL, where upper case denotes identity, lowercase denotes presence in over half of the sequences, and x denotes no significant conservation.

The BjPutA structure provides the three-dimensional context of the conserved motif. The first 12 residues of the motif were resolved in the structure. The motif begins at the N-terminus of the final strand of the enzyme and extends beyond the helical plug. The

identically conserved Glu at position 1 forms an intersubunit hydrogen bond to the backbone of Lys965. The Asn at position 7, which is Asp in some sequences, forms an intersubunit hydrogen bond with Lys351 (Figure 2.3C). Lys351 is highly conserved in branch 1 PutAs (Arg in some sequences) and is part of helix 5a of the PRODH barrel. As described above, helix 5A is not found in monofunctional PRODHs. Positions 9-11 form the turn of helix that helps plugs a hole in the cavity wall. This analysis suggests that that the C-terminal motif is important for formation of the substrate-channeling cavity.

2.4. Beyond the minimalist PutA

Many PutAs have chain lengths that are much longer than that of the 999-residue BjPutA, so it is natural to ask how these extra residues beyond the minimalist PutA are incorporated into the polypeptide chain. MSAs provide information about the domains that are shared among PutAs and the locations of the extra domains of long PutAs. We will focus here on branch 1 PutAs, because the sequence identity is high within this group and the resulting trends are obvious. Although we have analyzed many branch 1 PutAs using MSAs, we present an alignment of just three due to space limitations. The results presented here are valid for the larger group. Figure 2.7 shows an MSA of BjPutA with a representative long bifunctional branch 1 PutA (*Azoarcus* PutA) and the archetypal trifunctional PutA, EcPutA. The alignment exhibits a long region of high identity (48 - 59 % pairwise) corresponding to BjPutA residues 5-974. This region corresponds to the arm, alpha, PRODH, linker, NAD^+-binding, and P5CDH catalytic domains of BjPutA. Note that there is a gap in the long PutAs corresponding to the

beta-hairpin of the oligomerization domain (Figure 2.7, beta11). The second region of high identity corresponds to the conserved C-terminal motif. These results suggest that all the domains described here for BjPutA are also present in the long PutAs except for the beta-hairpin, which appears to be abbreviated or absent. The MSA also indicates an extra domain at the N-terminus of trifunctional PutAs and another domain immediately preceding the conserved C-terminal motif in long branch 1 PutAs. The N-terminal domain is the ribbon-helix-helix DNA-binding domain. The function of the extra C-terminal domain is unknown.

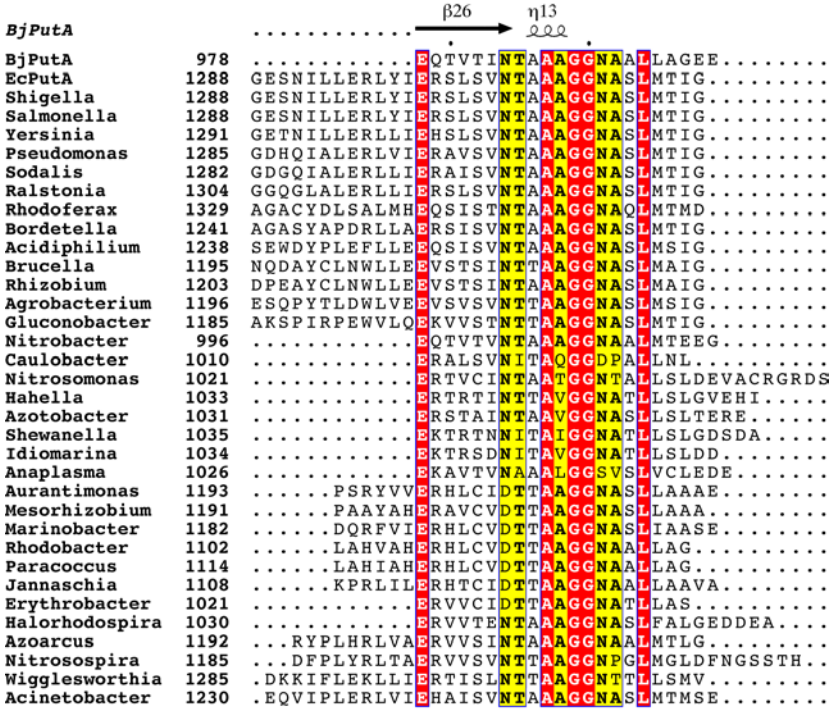


Figure 2.6. Section of an MSA of branch 1 PutAs showing the conserved motif at the C-terminus. The trifunctional PutAs are EcPutA through *Acidiphilium*, plus *Wigglesworthia*. The other PutAs are bifunctional. The secondary structure elements are from the BjPutA structure (PDB code 3haz). This figure and others were prepared with ClustalW2 (68) and ESPript (69).

2.4.1. The DNA-binding domain of trifunctional PutA

The DNA-binding domains EcPutA and *Pseudomonas putida* PutA have been extensively characterized using X-ray crystallography, NMR, and an array of biophysical and biochemical techniques (56-58, 70). The PutA DNA-binding domain has the ribbon-helix-helix (RHH) fold, which identifies PutA as a member of a large superfamily of transcription factors that includes Arc, MetJ, CopG, and NikR, among others. Schreiter and Drennan have written an excellent review on the RHH superfamily (71). The DNA-binding domain connects to the arm domain via a ~35-residue polypeptide (Figure 2.7). No structural information is available for these residues, thus it is not known whether this polypeptide is a flexible tether or has a well-defined three-dimensional structure. The sequence identity of the linker is relatively low (results not shown), which perhaps argues in favor of a flexible tether. Furthermore, analysis of the sequences of several trifunctional PutAs using the Disopred server suggests a high probability of disorder within residues 50-80 (72).

2.4.2. The C-terminal domain of unknown function

MSAs indicate that trifunctional PutAs have a ~200-residue domain near the C-terminus (Figure 2.7, EcPutA). Bifunctional PutAs of branch 1 that are longer than about 1100 residues also have this domain; the PutA from *Azoarcus* sp. BH72 is one example (Figure 2.7, Az). Interestingly, long bifunctional PutAs of branch 2 also have an extra domain in the C-terminus, but it is not clear whether it is related to the C-terminal domain of branch 1 PutAs. Thus, we will restrict our discussion here to the conserved C-terminal domain of branch 1 PutAs, which we denote by CTD. Using

BjPutA as a reference, the CTD is inserted between the third section of the NAD^+-binding domain and the conserved C-terminal motif of the oligomerization domain. The size of the CTD ranges from about 130 residues to 220 residues.

Although the function of the CTD is unknown, amino acid sequence analysis provides some intriguing ideas to test. Analysis of over 2 dozen branch 1 PutA CTDs using the remote homology detection algorithm HHSearch (73, 74) suggests that the CTD is homologous to the beta-hairpin and Rossmann fold regions of aldehyde dehydrogenases. For example, using the CTD of EcPutA as the query, HHSearch returned 33 aldehyde dehydrogenases with probability scores ranging from 99.1 % to 99.9 %. The top match was BjPutA with a probability score of 99.9 and E-value of 7.7E-26. The alignment shows that the CTD of EcPutA is 25 % identical to residues 551-761 of BjPutA. In the other cases tested, the BjPutA beta-hairpin/Rossmann domain was also identified as the closest homolog, and the probability score was in the range 99.7-100.0 %. These results are indicative of meaningful homology.

The homology is evident in an MSA of BjPutA with the CTDs of several branch 1 PutAs (Figure 2.8A). The core region of homology corresponds to the beta-hairpin of the oligomerization domain and the Rossmann fold domain of BjPutA. The structure of this region is highlighted in green, yellow, and red in Figure 2.8B. The highest similarity is found in the beta-hairpin, which contains the conserved motif lpGPtGExN. This result is significant, because multiple sequence alignments show that long branch 1 PutAs have a gap corresponding to the beta-hairpin of BjPutA (Figure 2.7, beta11). Thus, it appears that the beta-hairpin of the oligomerization domain has been shifted to the CTD in long PutAs.

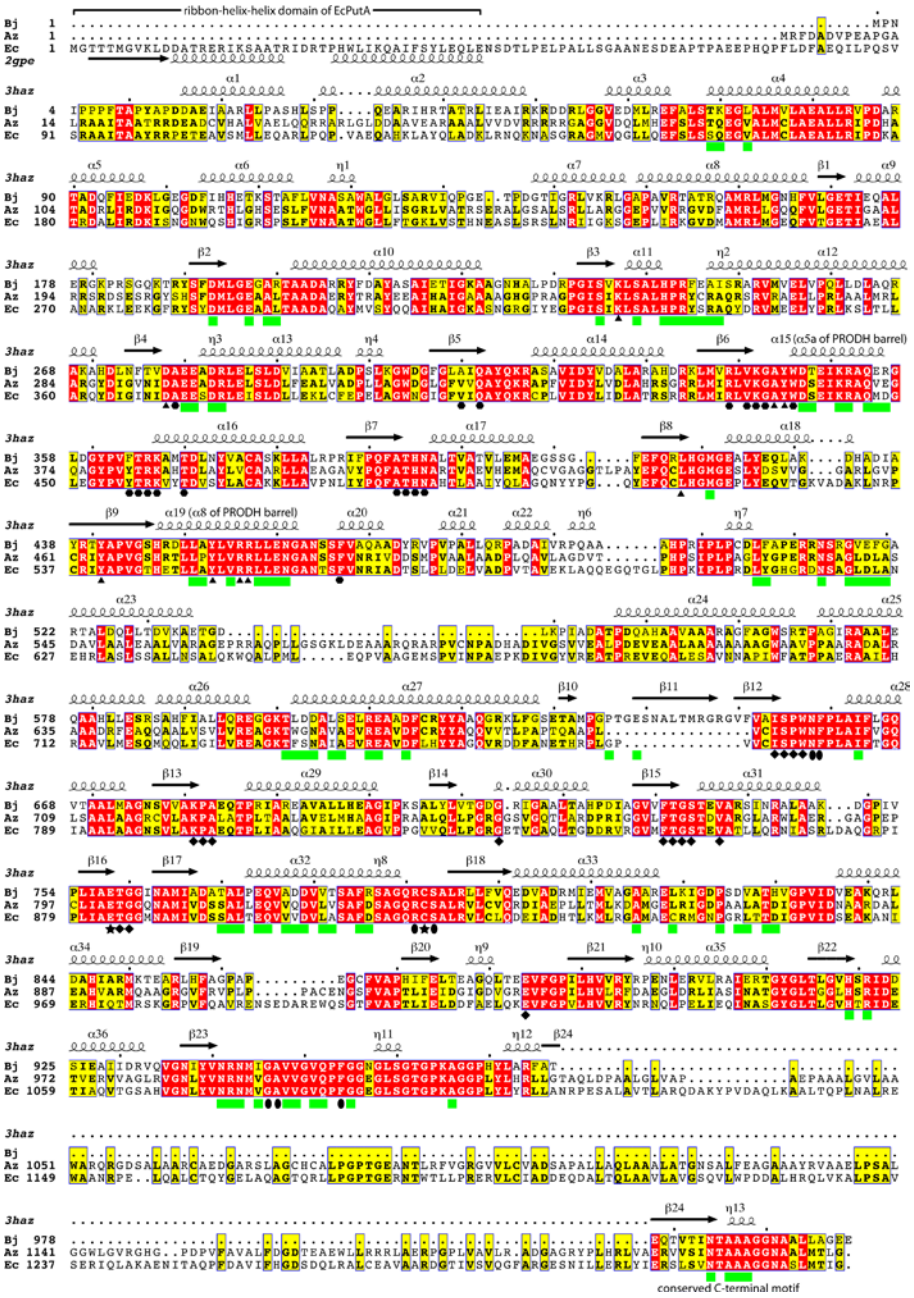


Figure 2.7. MSA of three branch 1 PutAs: BjPutA (Bj, GenBank BAC52526.1), *Azoarcus* sp. BH72 PutA (Az, GenBank CAL96369.1), and EcPutA (Ec, GenBank AAB59985.1). BjPutA is a minimalist PutA. Az is a long bifunctional PutA. Ec is a trifunctional PutA. The secondary structure elements above the sequence blocks are from the BjPutA structure (PDB code 3haz). The secondary structure elements for the N-terminal ribbon-helix-helix domain are from a structure of the EcPutA DNA-binding domain (PDB code 2gpe). Symbols below the sequence blocks denote the following: green squares, substrate-channeling cavity; triangles, proline binding site; hexagons, FAD binding site; diamonds, NAD^+ binding site; ovals, GSA binding site; stars, catalytic Cys of the P5CDH catalytic domain and the Glu that is predicted to assist in hydrolysis of the thioacylenzyme.

In summary, remote homology detection suggests that the CTD includes a beta-hairpin that is normally found in the oligomerization domain of aldehyde dehydrogenases and a Rossmann fold domain. Thus, long branch 1 PutAs are predicted to have two Rossmann fold domains, one that has high identity (~50 %) to the Rossmann fold domain of BjPutA and a second one in the CTD that has lower identity (14 - 33 %).

2.5. Summary

Comparison of the PutA structures to those of the monofunctional enzymes is useful for thinking about the possibility of protein-protein interactions predicted by the Rosetta Stone hypothesis. Several unique features of PutAs are absent in the monofunctional enzymes, including helix alpha5a, the arm domain, the alpha domain, and the linker domain. These components appear to be important for orienting the catalytic domains of the PutA protomer so that the two active sites face each other and for sealing the substrate-channeling cavity from the bulk medium. The absence of these PutA-specific structural features in the monofunctional enzymes, at least those from branch 3B, perhaps argues against the formation of an efficient PRODH-P5CDH channeling complex. However, kinetic measurements of substrate channeling and biophysical measurements of protein-protein association are still needed to test the Rosetta Stone hypothesis for monofunctional proline catabolic enzymes.

The CTD is the only PutA domain that has not been structurally characterized. Remote homology detection suggests the tantalizing hypothesis that the CTD of branch 1 PutAs contains a beta-hairpin like the one in the oligomerization domain of aldehyde dehydrogenases and a Rossmann fold domain. Interestingly, remote homology

detection did not find meaningful homology for the C-terminal domains of branch 2 enzymes. Thus, the C-terminal domains of branch 1 and 2 enzymes may differ in structure. These results raise several intriguing questions.

Does the CTD bind NAD^+ or is it a pseudo-NAD^+-binding domain that primarily plays a structural role? We note that the oxidation of GSA to Glu requires only one equivalent of NAD^+ and thus only one functional NAD^+-binding domain is expected. Furthermore, the CTD is missing the NFP motif found in BjPutA residues Asn658-Phe659-Pro660 (Figure 2.8A). This motif is highly conserved by P5CDHs (including the first Rossmann domain of all PutAs), and the Asn residue is thought to help anchor the substrate in the oxyanion hole by forming a hydrogen bond with the epsilon O atom of GSA (54). The absence of this critical residue argues against the CTD participating in catalysis. If the CTD plays a structural role, does the predicted beta-hairpin interact with the C-terminal strand and participate in domain-swapped dimerization? And, does it help cover the substrate-channeling cavity as in BjPutA? New biochemical and structural studies on PutAs designed to answer these questions represent an exciting next phase of research in proline catabolism.

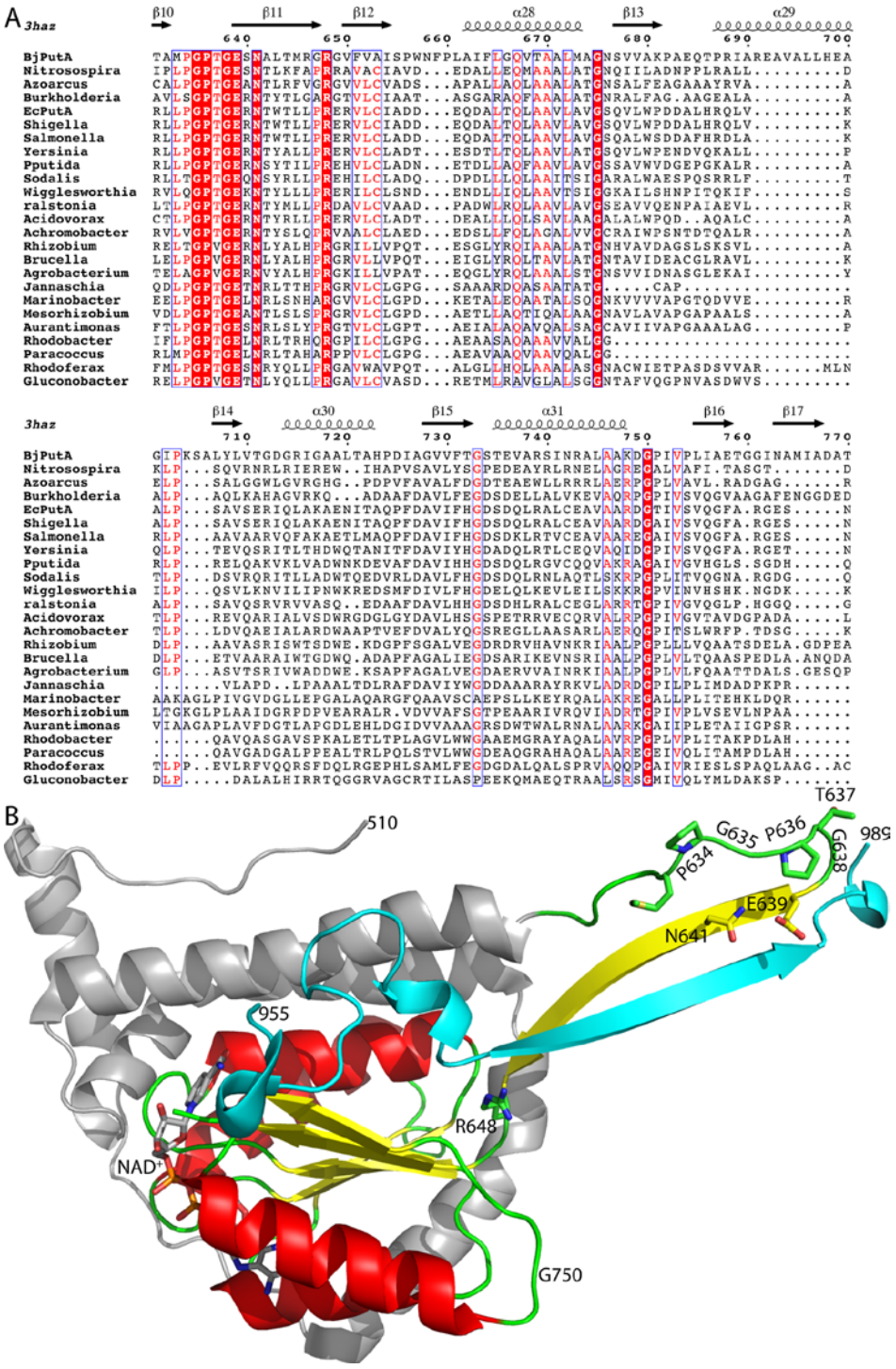


Figure 2.8. Homology of the CTD to the BjPutA beta-hairpin and Rossmann fold domain. (A) MSA of BjPutA with the CTDs of several branch 1 PutAs. (B) The structure of the NAD^+-binding and oligomerization domains of BjPutA. The core region of homology with the CTD (residues 629-760) is colored according to secondary structure, with alpha helices in red, beta strands in yellow, and loops in green. Conserved residues of the CTD are indicated. Residues 510-628 are colored silver. Residues 955-989 are colored cyan.

2.6. References

1. Phang JM. The regulatory functions of proline and pyrroline-5-carboxylic acid. Curr Top Cell Reg. 1985;25:92-132.

2. Ratzkin B, Roth J. Cluster of genes controlling proline degradation in Salmonella typhimurium. J Bacteriol. 1978;133(2):744-54.

3. Tanner JJ. Structural biology of proline catabolism. Amino Acids. 2008;35(4):719-30.

4. Downing SJ, Phang JM, Kowaloff EM, Valle D, Smith RJ. Proline oxidase in cultured mammalian cells. J Cell Physiol. 1977;91(3):369-76.

5. Kowaloff EM, Phang JM, Granger AS, Downing SJ. Regulation of proline oxidase activity by lactate. Proc Natl Acad Sci U S A. 1977;74(12):5368-71.

6. Donald SP, Sun XY, Hu CA, Yu J, Mei JM, Valle D, et al. Proline oxidase, encoded by p53-induced gene-6, catalyzes the generation of proline-dependent reactive oxygen species. Cancer Res. 2001;61(5):1810-5.

7. Hu CA, Donald SP, Yu J, Lin WW, Liu Z, Steel G, et al. Overexpression of proline oxidase induces proline-dependent and mitochondria-mediated apoptosis. Mol Cell Biochem. 2006.

8. Liu Y, Borchert GL, Surazynski A, Hu CA, Phang JM. Proline oxidase activates both intrinsic and extrinsic pathways for apoptosis: the role of ROS/superoxides, NFAT and MEK/ERK signaling. Oncogene. 2006;25(41):5640-7.

9. Pandhare J, Cooper SK, Phang JM. Proline oxidase, a proapoptotic gene, is induced by troglitazone: evidence for both peroxisome proliferator-activated receptor gamma-dependent and -independent mechanisms. J Biol Chem. 2006;281(4):2044-52.

10. Cooper SK, Pandhare J, Donald SP, Phang JM. A novel function for hydroxyproline oxidase in apoptosis through generation of reactive oxygen species. J Biol Chem. 2008;283(16):10485-92.

11. Phang JM, Donald SP, Pandhare J, Liu Y. The metabolism of proline, a stress substrate, modulates carcinogenic pathways. Amino Acids. 2008;35(4):681-90. Epub 2008/04/11.

12. Phang JM, Pandhare J, Liu Y. The metabolism of proline as microenvironmental stress substrate. J Nutr. 2008;138(10):2008S-15S. Epub 2008/09/23.

13. Yoshida A, Rzhetsky A, Hsu LC, Chang C. Human aldehyde dehydrogenase gene family. Eur J Biochem. 1998;251(3):549-57. Epub 1998/03/07.

14. Yoon KA, Nakamura Y, Arakawa H. Identification of ALDH4 as a p53-inducible gene and its protective role in cellular stresses. J Hum Genet. 2004;49(3):134-40. Epub 2004/02/27.

15. Forte-McRobbie CM, Pietruszko R. Purification and characterization of human liver "high Km" aldehyde dehydrogenase and its identification as glutamic gamma-semialdehyde dehydrogenase. J Biol Chem. 1986;261(5):2154-63.

16. Forte-McRobbie C, Pietruszko R. Human glutamic-gamma-semialdehyde dehydrogenase. Kinetic mechanism. Biochem J. 1989;261(3):935-43. Epub 1989/08/01.

17. White TA, Krishnan N, Becker DF, Tanner JJ. Structure and kinetics of monofunctional proline dehydrogenase from Thermus thermophilus. J Biol Chem. 2007;282(19):14316-27.

18. Krishnan N, Becker DF. Characterization of a bifunctional PutA homologue from Bradyrhizobium japonicum and identification of an active site residue that modulates proline reduction of the flavin adenine dinucleotide cofactor. Biochemistry. 2005;44(25):9130-9.

19. Schuermann JP, White TA, Srivastava D, Karr DB, Tanner JJ. Three crystal forms of the bifunctional enzyme proline utilization A (PutA) from Bradyrhizobium japonicum. Acta Cryst. 2008;F64(Pt 10):949-53.

20. Straub PF, Reynolds PH, Althomsons S, Mett V, Zhu Y, Shearer G, et al. Isolation, DNA sequence analysis, and mutagenesis of a proline dehydrogenase gene (putA) from Bradyrhizobium japonicum. Appl Environ Microbiol. 1996;62(1):221-9.

21. Krishnan N, Becker DF. Oxygen Reactivity of PutA from Helicobacter Species and Proline-Linked Oxidative Stress. J Bacteriol. 2006;188(4):1227-35.

22. Krishnan N, Doster AR, Duhamel GE, Becker DF. Characterization of a Helicobacter hepaticus putA mutant strain in host colonization and oxidative stress. Infect Immun. 2008;76(7):3037-44.

23. Nakajima K, Inatsu S, Mizote T, Nagata Y, Aoyama K, Fukuda Y, et al. Possible involvement of put A gene in Helicobacter pylori colonization in the stomach and motility. Biomed Res. 2008;29(1):9-18. Epub 2008/03/18.

24. Brown ED, Wood JM. Redesigned purification yields a fully functional PutA protein dimer from Escherichia coli. J Biol Chem. 1992;267(18):13086-92.

25. Menzel R, Roth J. Regulation of genes for Proline Utilization in *Salmonella typhimurium*: Autogenous Repression by the *putA* gene Product. J Mol Biol. 1981;148:21-44.

26. Ostrovsky de Spicer P, O'Brien K, Maloy S. Regulation of proline utilization in Salmonella typhimurium: a membrane-associated dehydrogenase binds DNA in vitro. J Bacteriol. 1991;173(1):211-9.

27. Vilchez S, Manzanera M, Ramos JL. Control of expression of divergent Pseudomonas putida put promoters for proline catabolism. Appl Environ Microbiol. 2000;66(12):5221-5.

28. Menzel R, Roth J. Enzymatic properties of the purified putA protein from Salmonella typhimurium. J Biol Chem. 1981;256(18):9762-6.

29. Ostrovsky de Spicer P, Maloy S. PutA protein, a membrane-associated flavin dehydrogenase, acts as a redox-dependent transcriptional regulator. Proc Natl Acad Sci USA. 1993;90(9):4295-8.

30. Surber MW, Maloy S. The PutA protein of Salmonella typhimurium catalyzes the two steps of proline degradation via a leaky channel. Arch Biochem Biophys. 1998;354(2):281-7.

31. Surber MW, Maloy S. Regulation of Flavin Dehydrogenase Compartmentalization: Requirements for PutA-Membrane Association in *Salmonella typhimurium*. Biochim Biophys Acta. 1999;1421:5-18.

32. Wood JM, Zadworny D. Amplification of the *put* Genes and Identification of the *put* Gene Products in *Escherichia coli* K12. Can J Biochem. 1980;58:787-96.

33. Wood JM. Genetics of L-proline Utilization in Escherichia coli. J Bacteriol. 1981;146:895-901.

34. Abrahamson JL, Baker LG, Stephenson JT, Wood JM. Proline dehydrogenase from Escherichia coli K12. Properties of the membrane-associated enzyme. Eur J Biochem. 1983;134(1):77-82.

35. Graham S, Stephenson JT, Wood JM. Proline Dehydrogenase from *Escherichia coli* K12, Reconstitution of Functional Membrane Association. J Biol Chem. 1984;259(4):2656-61.

36. Wood JM. Membrane association of proline dehydrogenase in Escherichia coli is redox dependent. Proc Natl Acad Sci USA. 1987;84(2):373-7. Epub 1987/01/01.

37. Brown ED, Wood JM. Conformational change and membrane association of the PutA protein are coincident with reduction of its FAD cofactor by proline. J Biol Chem. 1993;268(12):8972-9.

38. Ling M, Allen SW, Wood JM. Sequence analysis identifies the proline dehydrogenase and delta 1-pyrroline-5-carboxylate dehydrogenase domains of the multifunctional Escherichia coli PutA protein. J Mol Biol. 1994;243(5):950-6. Epub 1994/11/11.

39. Becker DF, Thomas EA. Redox properties of the PutA protein from Escherichia coli and the influence of the flavin redox state on PutA-DNA interactions. Biochemistry. 2001;40(15):4714-21.

40. Vinod MP, Bellur P, Becker DF. Electrochemical and functional characterization of the proline dehydrogenase domain of the PutA flavoprotein from *Escherichia coli*. Biochemistry. 2002;41:6525-32.

41. Zhu W, Gincherman Y, Docherty P, Spilling CD, Becker DF. Effects of proline analog binding on the spectroscopic and redox properties of PutA. Arch Biochem Biophys. 2002;408(1):131-6.

42. Zhu W, Becker DF. Flavin redox state triggers conformational changes in the PutA protein from Escherichia coli. Biochemistry. 2003;42(18):5469-77.

43. Baban BA, Vinod MP, Tanner JJ, Becker DF. Probing a hydrogen bond pair and the FAD redox properties in the proline dehydrogenase domain of Escherichia coli PutA. Biochim Biophys Acta. 2004;1701(1-2):49-59.

44. Zhang W, Zhou Y, Becker DF. Regulation of PutA-membrane associations by flavin adenine dinucleotide reduction. Biochemistry. 2004;43(41):13165-74.

45. Zhu W, Becker DF. Exploring the proline-dependent conformational change in the multifunctional PutA flavoprotein by tryptophan fluorescence spectroscopy. Biochemistry. 2005;44(37):12297-306. Epub 2005/09/15.

46. Zhang W, Zhang M, Zhu W, Zhou Y, Wanduragala S, Rewinkel D, et al. Redox-induced changes in flavin structure and roles of flavin N(5) and the ribityl 2'-OH group in regulating PutA--membrane binding. Biochemistry. 2007;46(2):483-91.

47. Arentson BW, Sanyal N, Becker DF. Substrate channeling in proline metabolism. Front Biosci. 2011.

48. Anderson KS. Fundamental mechanisms of substrate channeling. Methods Enzymol. 1999;308:111-45.

49. Miles EW, Rhee S, Davies DR. The molecular basis of substrate channeling. J Biol Chem. 1999;274(18):12193-6.

50. Huang X, Holden HM, Raushel FM. Channeling of substrates and intermediates in enzyme-catalyzed reactions. Annu Rev Biochem. 2001;70:149-80.

51. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences. Science. 1999;285(5428):751-3.

52. White TA, Tanner JJ. Cloning, purification and crystallization of Thermus thermophilus proline dehydrogenase. Acta Cryst. 2005;F61(Pt 8):737-9.

53. White TA, Johnson WH, Jr., Whitman CP, Tanner JJ. Structural basis for the inactivation of Thermus thermophilus proline dehydrogenase by N-propargylglycine. Biochemistry. 2008;47(20):5573-80. Epub 2008/04/23.

54. Inagaki E, Ohshima N, Takahashi H, Kuroishi C, Yokoyama S, Tahirov TH. Crystal structure of Thermus thermophilus Delta1-pyrroline-5-carboxylate dehydrogenase. J Mol Biol. 2006;362(3):490-501.

55. Inagaki E, Ohshima N, Sakamoto K, Babayeva ND, Kato H, Yokoyama S, et al. New insights into the binding mode of coenzymes: structure of Thermus thermophilus [Delta]1-pyrroline-5-carboxylate dehydrogenase complexed with NADP+. Acta Cryst. 2007;F63(6 %R doi:10.1107/S1744309107021422):462-5.

56. Larson JD, Jenkins JL, Schuermann JP, Zhou Y, Becker DF, Tanner JJ. Crystal structures of the DNA-binding domain of Escherichia coli proline utilization A flavoprotein and analysis of the role of Lys9 in DNA recognition. Protein Sci. 2006;15:1-12.

57. Zhou Y, Larson JD, Bottoms CA, Arturo EC, Henzl MT, Jenkins JL, et al. Structural basis of the transcriptional regulation of the proline utilization regulon by multifunctional PutA. J Mol Biol. 2008;381(1):174-88.

58. Halouska S, Zhou Y, Becker DF, Powers R. Solution structure of the Pseudomonas putida protein PpPutA45 and its DNA complex. Proteins. 2009;75(1):12-27. Epub 2008/09/04.

59. Lee YH, Nadaraia S, Gu D, Becker DF, Tanner JJ. Structure of the proline dehydrogenase domain of the multifunctional PutA flavoprotein. Nat Struct Biol. 2003;10(2):109-14.

60. Zhang M, White TA, Schuermann JP, Baban BA, Becker DF, Tanner JJ. Structures of the Escherichia coli PutA proline dehydrogenase domain in complex with competitive inhibitors. Biochemistry. 2004;43(39):12539-48.

61. Ostrander EL, Larson JD, Schuermann JP, Tanner JJ. A conserved active site tyrosine residue of proline dehydrogenase helps enforce the preference for proline over hydroxyproline as the substrate. Biochemistry. 2009;48(5):951-9. Epub 2009/01/15.

62. Srivastava D, Zhu W, Johnson WH, Jr., Whitman CP, Becker DF, Tanner JJ. The structure of the proline utilization a proline dehydrogenase domain inactivated by N-propargylglycine provides insight into conformational changes induced by substrate binding and flavin reduction. Biochemistry. 2010;49(3):560-9. Epub 2009/12/10.

63. Srivastava D, Schuermann JP, White TA, Krishnan N, Sanyal N, Hura GL, et al. Crystal structure of the bifunctional proline utilization A flavoenzyme from Bradyrhizobium japonicum. Proc Natl Acad Sci U S A. 2010;107(7):2878-83. Epub 2010/02/06.

64. Bennett MJ, Schlunegger MP, Eisenberg D. 3D domain swapping: a mechanism for oligomer assembly. Protein Sci. 1995;4(12):2455-68. Epub 1995/12/01.

65. The PyMOL user's manual [database on the Internet]. DeLano Scientific, Palo Alto, CA, USA. 2002.

66. Liu ZJ, Sun YJ, Rose J, Chung YJ, Hsiao CD, Chang WR, et al. The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. Nat Struct Biol. 1997;4(4):317-26.

67. Bottoms CA, Smith PE, Tanner JJ. A structurally conserved water molecule in Rossmann dinucleotide-binding domains. Protein Sci. 2002;11(9):2125-37.

68. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, et al. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 2003;31(13):3497-500. Epub 2003/06/26.

69. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics. 1999;15(4):305-8. Epub 1999/05/13.

70. Gu D, Zhou Y, Kallhoff V, Baban B, Tanner JJ, Becker DF. Identification and characterization of the DNA-binding domain of the multifunctional PutA flavoenzyme. J Biol Chem. 2004;279(30):31171-6.

71. Schreiter ER, Drennan CL. Ribbon-helix-helix transcription factors: variations on a theme. Nat Rev Microbiol. 2007;5(9):710-20.

72. Watanabe H, Hastings JW. Specificities and properties of three reduced pyridine nucleotide-flavin mononucleotide reductases coupling to bacterial luciferase. Mol Cell Biochem. 1982; 44:181-7.

73. Soding J. Protein homology detection by HMM-HMM comparison. Bioinformatics. 2005;21(7):951-60. Epub 2004/11/09.

74. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006;22(2):195-201. Epub 2005/11/23.

CHAPTER 3

Small Angle X-ray Scattering Studies of the Oligomeric State and Quaternary Structure of the trifunctional Proline Utlization A (PutA) Flavoprotein from *Escherichia coli.*

This chapter has been adapted from

Ranjan K. Singh, John D. Larson, Weidong Zhu, Robert P. Rambo, Greg L. Hura, Donald F. Becker, and John J. Tanner

J. Biol. Chem. (2011) 286(50), 43144-53

My contribution to the paper

Purified proteins for kinetics and gel shift assays, cloned PutA86-1320 and analyzed the SAXS data.

3.1. Introduction

Proline catabolism in Gram-negative bacteria is catalyzed by the bifunctional enzyme proline utilization A (PutA) (1, 2). The catalytic apparatus of PutAs consists of an FAD-dependent proline dehydrogenase (PRODH) active site that catalyzes the oxidation of proline to Δ^1-pyrroline-5-carboxylate (P5C), and an NAD -dependent P5C dehydrogenase (P5CDH) active site that catalyzes the oxidation of glutamate semialdehyde to glutamate (Figure 3.1). These two reactions are linked by a hydrolysis step that converts P5C into glutamate semialdehyde.

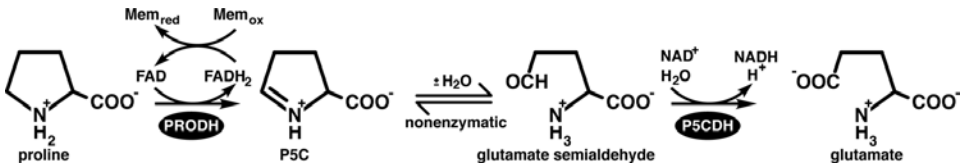


Figure 3.1. Reactions catalyzed by PutA.

In addition, some PutAs provide genetic regulation by acting as transcriptional repressors, thereby establishing a class of trifunctional PutAs (3-9). The best studied trifunctional PutA is EcPutA from *Escherichia coli* (5, 10-17). EcPutA controls the expression of the put regulon, which contains the genes encoding EcPutA and the proline transporter PutP. When proline levels are low, EcPutA blocks transcription by binding to operator sites located between the two divergently transcribed genes. Increased proline levels cause EcPutA to dissociate from DNA and bind the inner membrane, which activates gene transcription, uptake of proline, and proline catabolic enzymatic activity. The mechanism by which proline stimulates proline utilization, e.g. functional switching,

involves global conformational changes induced by proline reduction of the FAD, which enhance membrane binding affinity and shift the equilibrium of PutA from DNA-bound to membrane-associated. Thus, trifunctional PutAs are dynamic proteins that function both as a sensor that monitors the level of environmental proline and a transducer that converts proline into usable energy for the organism.

The crystal structure of a bifunctional PutA is known(18) (Figure 3.2). At 999 residues, PutA from *Bradyrhizobium japonicum* (BjPutA) is among the smallest of PutAs and is therefore considered to be a minimalist PutA. The PRODH active site is located in a $(\beta\alpha)_8$ barrel, whereas the P5CDH site resides in a crevice between the Rossmann-like NAD -binding domain and the catalytic domain that furnishes the essential reactive Cys. The two active sites are separated by 41 Å and connected by a large, irregularly shaped internal cavity (silver surface in Figure 3.2A). It has been hypothesized that the cavity functions in substrate channeling by serving as an internal reaction vessel for the hydrolysis of P5C to glutamate semialdehyde and a protected conduit for the diffusion of the semialdehyde to the P5CDH active site. In addition to the catalytic domains, the structure includes four ancillary domains (arm, α, linker, and oligomerization) that not only help to create the aforementioned cavity, but also provide the sites for oligomerization. Of particular note for the current work is the oligomerization domain.

The oligomerization domain of BjPutA is a bipartite flap consisting of a β-hairpin and the C-terminal 20 residues of the chain (Figure 3.2A, orange). The latter part of the oligomerization domain forms a β -strand followed by a turn of -helix and contains the conserved sequence motif EXXXXv(N or D)t(T or A)AaGGnaXL, where uppercase, lowercase, and X indicate identical, highly conserved, and any residue, respectively. The

sequence appears to be in all branch 1 PutAs, a group that includes both BjPutA and trifunctional PutAs (1). BjPutA forms a domain-swapped dimer in which the flap of one protomer forms main chain hydrogen bonds with the β -sheet of the catalytic domain of the other protomer (Figure 3.2B). As a consequence of dimerization, the flap of one protomer seals the cavity of the other protomer from the bulk medium preventing loss of the intermediate (Figure 3.2C).

Trifunctional PutAs have two additional domains not found in the minimalist PutA (Figure 3.1). The DNA-binding domain (residues 1–49) has a ribbon helix helix (RHH) fold and forms the canonical RHH dimer in solution (9, 19). The other extra domain (CTD, C-terminal domain) has 200 residues and is inserted between the NAD -binding domain and the predicted conserved C-terminal motif. The function of the CTD is unknown.

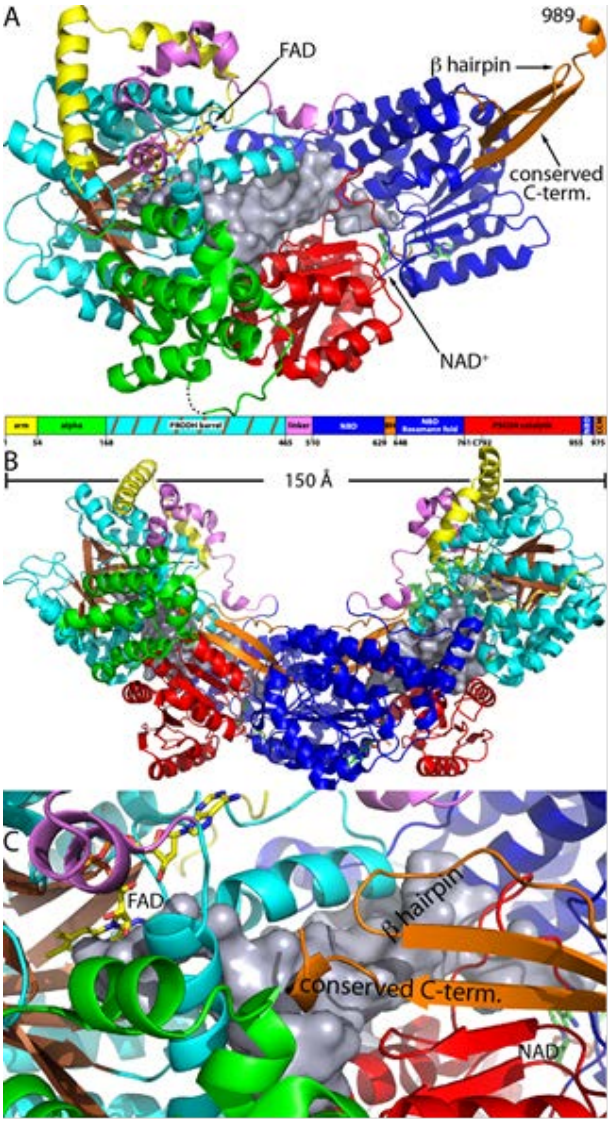


Figure 3.2. Structure of the minimalist PutA, BjPutA. (A) Structure of the protomer with the domains colored according to the domain diagram. FAD and NAD^+ are drawn as sticks in yellow and green, respectively. Abbreviations used in the domain diagram: NBD, NAD^+-binding domain; BH, β-hairpin; CCM, conserved C-terminal motif. (B) Structure of the domain-swapped dimer. (C) Close-up view of the dimer interface highlighting how the β-flap (orange) of one protomer seals the substrate-channeling cavity of the other protomer of the dimer.

Although structures of the DNA-binding (9, 19) and PRODH(14, 16, 20-22) domains of EcPutA are known, the three-dimensional structure of a full-length

trifunctional PutA has remained elusive since the discovery of PutAs in the late 1970s (23). Here, we report small-angle x-ray scattering (SAXS) studies of EcPutA, which provide the first view of the three-dimensional architecture of a trifunctional PutA.

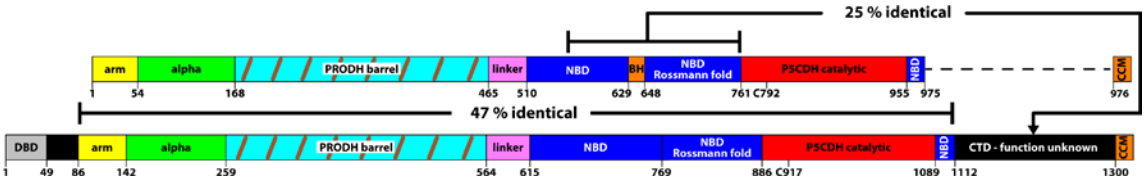


Figure 3.3. Schematic diagram depicting a multiple sequence alignment of minimalist PutAs, represented by BjPutA (top), and trifunctional PutAs, represented by EcPutA (bottom). BjPutA and EcPutA have 999 and 1320 residues, respectively. Abbreviations used: DBD, DNA-binding domain; NBD, NAD^+-binding domain; BH, β-hairpin; CTD, C-terminal domain; CCM, conserved C-terminal motif.

3.2. Materials and Methods

3.2.1. Expression and Purification of EcPutA

The plasmid used to express EcPutA (pKA8H-EcPutA) was created by subcloning the putA gene from a previously described pET-3a vector (11) into pKA8H (kindly provided by Dr. Christopher Hill) using NdeI and BamHI restriction sites. The expressed protein includes the 1320-residue EcPutA with an N-terminal His_8 tag and intervening tobacco etch virus protease (TEVP) cleavage site. Treatment with TEVP results in the native polypeptide preceded by Gly-His. EcPutA was expressed in E. coli using standard methods and purified using immobilized metal affinity chromatography (His-Trap Ni^{2+} -Sepharose HP, GE Healthcare). Fractions eluted from the Ni^{2+}-Sepharose column that contained EcPutA were pooled, and TEVP, 1 M DTT, and 20 TEV buffer (1 M Tris-HCl, 10 mM EDTA, pH 8.0) were added so that the resulting solution contained 3mg of TEVP per 50mg of EcPutA in 50 mM Tris-HCl, 0.5 mM DTT, and 0.5 mM EDTA, pH 8.0. The

sample was incubated for 3 h at 30 °C, dialyzed overnight at 4 °C, and injected onto the Ni^{2+} -Sepharose column. The flowthrough was collected, dialyzed into 50 mM Tris-HCl, 50 mM NaCl, 0.5mM EDTA, 0.5mM DTT, and 5% glycerol, pH 8.0, and concentrated to 10–25 mg/ml using a centrifugal concentrator. The protein concentration was measured with the bicinchoninicacid method (Pierce kit). Size exclusion chromatography (SEC) was used as the final step of purification. As described by Brown and Wood (24), EcPutA exhibits two apparent species in SEC (Figure 3.4A) the major species is the functional dimeric protein (Figure 3.4B), and the minor species appears to be the isolated subunit. The dimer was isolated using either a Superdex 200 SEC column or a Shodex KW-803 SEC column.

3.2.2. Subcloning and Purification of Domain Deletion Constructs.

A domain deletion construct having EcPutA residues 1–1085(PutA1–1085) was created. The coding sequence for residues 1–1085 was amplified by PCR from a pET-23b plasmid harboring the putA gene (25) and subcloned into pET-23b using NdeI and EcoRI restriction sites. PutA1–1085 was purified as described above for EcPutA, except the C-terminal His tag was retained. Another domain deletion construct having EcPutA residues 86–1320 (PutA86–1320) was also created. The coding sequence for these residues was amplified by PCR from the plasmid pKA8H-EcPutA and subcloned into pKA8H using NdeI and BamHI restriction sites. PutA86–1320 was purified using immobilized metal affinity chromatography followed by cleavage of the affinity tag, passage through the affinity column to remove the tag and uncleaved protein, and finally anion exchange chromatography (HiTrap Q-Sepharose, GE Healthcare). For the latter

step, the loading buffer was 50mM Tris, 0.5.mM EDTA, and 5% glycerol, pH 8.0, and the protein was eluted with a 0–0.5 M linear NaCl gradient.

3.2.3. Small-angle X-ray Scattering.

SAXS experiments were performed at the SIBYLS beamline (12.3.1) of the ALS (26). For each sample, scattering intensities (I) were measured at three protein concentrations to ensure concentration-independent scattering. Exposures of 0.5, 1.0, and 5.0 s were used to check for radiation damage. The scattering curves collected from the protein sample were corrected for background scattering using intensity data collected from the dialysis buffer, SEC effluent, or flow-through from a centrifugal concentrator. Composite scattering curves were generated with PRIMUS (27) by scaling and merging the background-corrected high q region data from the 5.0-s exposure with the low q region data from a shorter exposure (0.5 or 1.0 s).

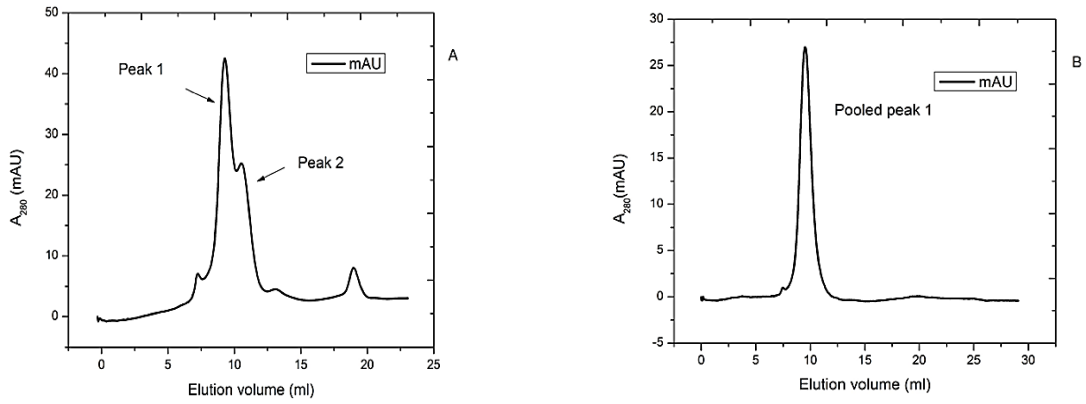


Figure 3.4. Size exclusion chromatogram of EcPutA. (A) Two species in chromatogram represented by peak 1 and peak 2 respectively. (B) pooled peak1 (functional dimeric protein).

For one of the EcPutA samples, the scattering curves (3.3, 6.7, and 10.0 mg/ml) were extrapolated to zero concentration, and composite scattering curves were generated by

scaling and merging the background-corrected high q region data from the 10 mg/ml sample with the low q region zero extrapolated data. Scattering curves were subjected to indirect Fourier transform using GNOM (28) to yield the pair distribution function (P(r)), from which the radius of gyration (R_g) and the maximum particle dimension (Dmax) were estimated. PRIMUS was used to calculate Porod volumes. The molecular weight in solution was estimated from SAXS data using the relationship, M I(0)/Kc, where M is the molecular weight, I(0) is the intensity extrapolated to zero scattering angle, c is the protein concentration in mg/ml, and K is a constant determined from beamline calibration measurements using glucose isomerase as a standard (29). GASBOR (30) was used to calculate shape reconstructions, and DAMAVER (31) was used to average and filter the resulting dummy atom models. The Situs module pdb2vol was used to convert the averaged, filtered models into volumetric maps (32). SUPCOMB was used to superimpose dummy atom models (33).

3.2.4. Rigid Body Modeling Using SAXS Data.

Modeling of EcPutA was performed using two rigid bodies: the 1.9-Å resolution crystal structure of the DNA-binding domain dimer (PDB 2GPE)(19) and a hybrid x-ray/homology model of residues 87–1113. The model of residues 87–1113 was built by combining the crystal structure of an EcPutA PRODH domain construct (PutA86–630, PDB code 1TIW) (16) with homology models based on the BjPutA structure (PDB code 3HAZ) generated with I-TASSER (34) and SWISS-MODEL (35). The BjPutA structure is a good template for modeling this part of EcPutA because the two enzymes are 47% identical (62% similar) in this region (Figure 3.2, Figure 3.3). Furthermore, PutA86–630

exhibits the arm, PRODH barrel, and linker domains also found in BjPutA (1.4-Å root mean square deviation). Neither the CTD nor the conserved C-terminal motif was included in rigid body modeling.

The strategy for rigid body modeling was based on the assumption that the DNA-binding domain is essential for dimerization, which follows from the domain deletion studies (see below). This assumption implies that the DNA-binding domain dimer resides in the connector region of the SAXS envelope, whereas residues 87–1113 are located in the large, spatially separated lobes. The EcPutA DNA-binding domain dimer (PDB code 2GPE) was manually fitted into the connector region of the consensus SAXS volumetric map with its 2-fold axis coincident with the 2-fold axis of the SAXS map. The structure fit equally well in two orientations corresponding to the DNA-binding surface facing the concave or convex surfaces of the envelope.

COLORES (36) was used to dock the x-ray/homology model of residues 87–1113 into the lobes. The volumetric maps used for these calculations were created as follows. First, the dummy atoms of the averaged, filtered consensus reconstruction model that overlapped the docked DNA-binding domain dimer were identified manually with PyMol and deleted. The remaining dummy atoms formed two clusters corresponding to the two lobes. These clusters of atoms were saved as two separate coordinate files and converted to volumetric maps, which were used for two COLORES docking calculations.

The poses of the x-ray/homology model of residues 87–1113 from COLORES were combined with the two poses of the docked DNA-binding domain dimer to generate several models, which were then ranked according to the agreement with the experimental scattering profiles using the FoXS parameter (37). The linker between

residues 47 and 87 was modeled using rapper (38) via ccp4i (39). The SAXS volumetric map was input to rapper to constrain the modeled peptide inside the SAXS envelope.

3.2.5. Multi Angle Light Scattering.

The molecular weight of EcPutA in solution was estimated using a multiangle light scattering (MALS) detector coupled to a Shodex KW-803 SEC column. The MALS analysis was performed in-line to SEC separation using an 18-angle DAWN HELEOS detector (WyattTechnology) with detector 12 replaced with a DynaPro quasielasti light scattering detector. Protein concentrations were simultaneously monitored with an Optilab refractive index detector (Wyatt Technology). System calibrations were per-formed with glucose isomerase (Hampton Research) dissolved in 50 mM Tris-HCl, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM DTT, and 5% glycerol, pH 8.0. The flow rate was 0.9 ml/min. The molecular weight of PutA86–1320 was estimated similarly using a MALS detector coupled to a G5000PWXL SEC column (Tosoh Bioscience, Montgomeryville, PA). The column buffer was 50mM Tris-HCl, 50mM NaCl, 0.5mM EDTA, 0.5mM THP, and 5% glycerol, pH 8.0. The flow rate was 0.75 ml/min.

3.2.6. Biochemical Assays.

Kinetic parameters for PRODH activity were determined at 25 °C in 100 mM MOPS buffer, pH 8.0, using proline as the substrate (0–500 mM) as previously described (17). P5CDH activity was measured using P5C as the substrate (0–1.5 mM, L-P5C) and 0.2 mM NAD as previously described (24). NADH formation was monitored at 340 nm, and the extinction coefficient of 6220 M^{-1} cm^{-1} for NADH was used to calculate the kinetic

parameters. P5C is not commercially available, therefore DL-P5C was chemically synthesized as described (40). DL-P5C was stored in acid at 4 °C. Immediately before performing kinetics experiments DL-P5C was neutralized to pH 6.5–8.0, quantified with O-aminobenzaldehyde (forms complex absorbing at 443 nm with €= 2590 M^{-1} cm^{-1}), and diluted with ice-cold assay buffer to make a stock solution. Due to the limited availability of P5C, the saturation region of the Michaelis-Menten curve was not accessible, which prevented determination of k_{cat} and K_m. Therefore, the ratio of k_{cat} to K_m in the limit of [S] << K_m was estimated from the slope of the linear region of the Michaelis-Menten curve at low substrate concentration.

Tryptophan fluorescence quenching was used to study the binding of NAD to PutA following an approach described previously (41). PutA (0.5–1.0 µM) was excited at 295 nm, and the maximal emission at 335 nm was measured at increasing concentrations of NAD (0–60 µM). A control assay without protein was performed similarly and used to correct for any inner filter effect. The dissociation constant (K_d) was estimated by fitting the corrected fluorescence quenching data to a single site binding isotherm. The DNA binding activities of PutA and PutA1–1085 were studied using gel mobility shift assays as previously described (9).

3.3. Results.

3.3.1. Oligomeric state of EcPutA from SEC-MALS.

The oligomeric state of EcPutA in solution was studied using SEC-MALS (Figure 3.5A). The data suggest that the purified protein is monodisperse with an apparent molecular mass of 274. 3 kDa. This value is within 6% of the predicted molecular mass

of 290 kDa for the dimer, in agreement with previous studies (24).

3.3.2. SAXS Analysis of EcPutA

SAXS data from five EcPutA samples differing in the reference buffer used for subtraction or the protein batch are shown in Figure 3.6A, and parameters derived from SAXS are listed in Table 3.1. The scattering curves exhibit a perceptible depression near q ~ 0.045 $Å^{-1}$ and bump near q ~ 0.065 $Å^{-1}$ (arrows in Figure 3.6A). The Guinier plots show good linearity, with R^2 scores greater than 0.994 (Figure 3.6A, inset). The P(r) curves exhibit a major peak at 44 Å and a prominent shoulder peak at 110 Å (Figure 3.6B), which is consistent with a particle having two spatially separated lobes. The maximum particle dimension (D_{max}) is in the range 200–210 Å. The real space radius of gyration (R_g) derived from P(r) calculations spans 62.4–63.3 Å with an average of 62.8 ± 0.4 Å. The I(0) value obtained from a scattering curve that was collected on a calibrated beamline yielded a molecular mass of 285 kDa, consistent with a dimeric protein.

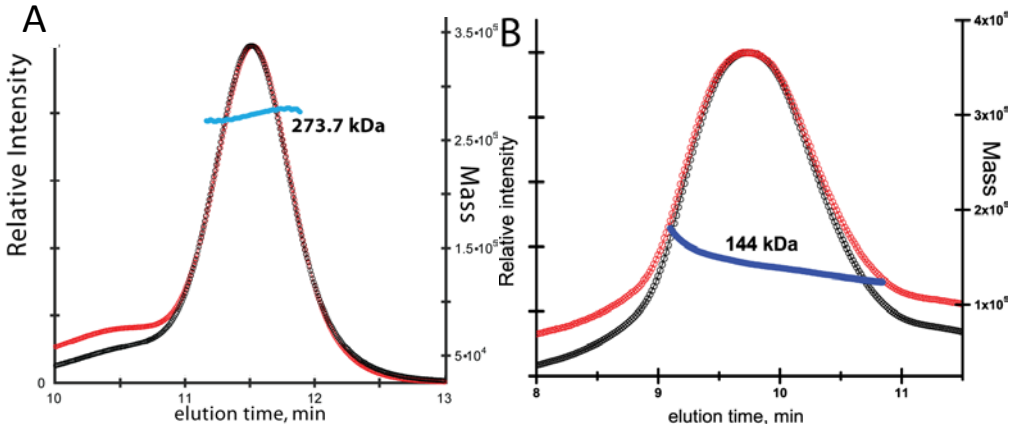


Figure 3.5. Determination of the molecular weight of (A) EcPutA and (B) PutA86-1320 using SEC-MALS. The red curve represents the light scattering response measured at 90°. The black curve represents the response of the refractive index detector. The blue curve shows the derived molecular weight.

Porod-Debye analysis was performed to assess the flexibility of EcPutA (Figure 3.6C). For well folded proteins, a plot of $q^4I(q)$ versus q4 restricted to low q reaches an asymptotic value, and the absence of a Porod-Debye plateau suggests that the protein contains substantial regions of unstructured polypeptide (42). Thus, the Porod-Debye plateau is a diagnostic indicator of foldedness akin to Kratky analysis. The Porod-Debye plots for EcPutA exhibit an obvious plateau, suggesting that the protein is well folded and does not have large regions of unstructured polypeptide (Figure 3.6C).

Porod-Debye analysis also provides confirmation of the oligomeric state. The Porod volume estimated from the five data sets is 396,000 ±16,000 $Å^3$. The assumption of a dimeric protein leads to a value for the protein density of 1.22 ± 0.05 g/ml. This value is well within the range of 0.9–1.5 g/ml obtained in a recent analysis of SAXS data from 31 different proteins (42). In contrast, the assumption of a monomeric or trimeric protein results in density values of 0.6 or 1.8 g/ml, respectively, which are unrealistic for a compact folded protein. The MALS and SAXS data support the hypothesis that EcPutA is a stable, monodisperse dimer in solution.

3.3.3. Shape Reconstructions of EcPutA.

The low resolution shape of EcPutA was derived from the SAXS data using the shape reconstruction program GASBOR (Figure 3.7). Shape reconstructions were performed for each of the 5 data sets, and a consensus shape was obtained by averaging 10 independent models from each data set. The mean normalized spatial discrepancy of the 50-model reconstruction performed without enforcing symmetry (P1) is 1.60 ± 0.07. The normalized spatial discrepancy for the P2 consensus model is 1.51±0.08, thus neither the

mean normalized spatial discrepancy nor the variation increased substantially when enforcing 2-fold symmetry. This result is consistent with the fact that the P1 shape exhibits approximate 2-fold symmetry (Figure 3. 7A).

Table 3.1. Parameters derived from SAXS experiments

Sample	R_g **(Å)**[a]	D_{max} **(Å)**[b]	V_{Porod} **(Å^3)**[c]	M **(kDa)**[d]
EcPutA	62.8 ± 0.4	200 - 210	396,000 ± 16,000	285 ± 29
PutA1-1085	59.9 ± 0.1	200	345,000	nd[e]
PutA86-1320	43.2 ± 0.1	165	186,000	140 ± 14

[a]The real space radius of gyration was estimated from calculations of P(r) using GNOM. The uncertainty for EcPutA is the standard deviation from five replicate samples. The uncertainties for PutA1-1085 and PutA86-1320 were obtained from single P(r) calculations.
[b]The maximum particle dimension was estimated from calculations of P(r) using GNOM. The range listed for EcPutA is based on five replicate samples.
[c]The Porod volume was calculated using PRIMUS. The uncertainty for EcPutA is the standard deviation from five replicate samples.
[d]The molecular weight is based on I(0), which was estimated using GNOM. The protein concentration was 3.3 mg/mL for EcPutA and 9 mg/mL for PutA86-1320. The quoted uncertainty of 10 % is from Mylonas and Svergun (29).
[e]Not determined.

The reconstructions suggest that EcPutA is a symmetric, V-shaped dimer having dimensions of 205 X 85 X 55 Å (Figure 3.7B). The particle has two large lobes that connect via a short cylindrical section with a diameter of 30 Å. The molecular shape resembles a curved and slightly twisted dumbbell. The molecular 2-fold axis passes through a connecting cylinder and is perpendicular to the longest axis of the dimer and

parallel to the 85-Å axis. Thus, the two protein chains presumably meet in the connecting cylinder to form a symmetric dimer interface.

3.3.4. Domain Deletion Analysis.

Domain deletion analysis was used to identify the domain(s) involved in dimerization. Two domain deletion mutants were created: PutA1–1085 and PutA86–1320. PutA1–1085 lacks the CTD and the conserved C-terminal motif, whereas PutA86–1320 contains all domains except the DNA-binding domain and the polypeptide that connects the DNA-binding domain to the arm (see Figure 3.3).

Enzyme and DNA-binding activities were measured to examine the impacts of the deletions on EcPutA function. Both deletion mutants exhibit PRODH activity similar to that of EcPutA (Table 3.2). The PRODH kinetic parameters for PutA86–1320 were k_{cat} of 7.5 ± 0.1 s^{-1} and K_m of 67 ± 4 mM (k_{cat}/K_m = 112 $M^{-1}s^{-1}$). Those of PutA1–1085 were k_{cat} of 10.6 ± 0.2 s^{-1} and K_m of 122 ± 5 mM (k_{cat}/K_m = 87 $M^{-1}s^{-1}$). For reference, the PRODH kinetic constants for EcPutA were 7.5 s^{-1} and 100 (k_{cat}/K_m = 75 M^{-1} s^{-1}) (12). The P5CDH activity of PutA86–1320 (k_{cat}/K_m =783 M^{-1} s^{-1}) is near that of EcPutA (1409 M^{-1} s^{-1}), but the P5CDH activity of PutA1–1085 was below detection (Table 2). The latter result likely reflects that fact that PutA1–1085 lacks a conserved 20-residue section of the P5CDH domain (1086–1108, Figure 3.8).

To determine whether the loss of P5CDH activity in PutA1– 1085 was due to diminished NAD binding, tryptophan fluorescence quenching experiments were performed. The K_d value for NAD binding to PutA1–1085 was 5.5 ± 2 µM, which was similar to that of EcPutA (2.6±0.2 µM), demonstrating that the NAD -binding domain

was functional in PutA1–1085. Finally, the DNA-binding activity of PutA1–1085 was comparable with that of EcPutA, as demonstrated by gel mobility shift assays (Fig 3.9). In summary, the domain deletion mutants exhibit the expected activities.

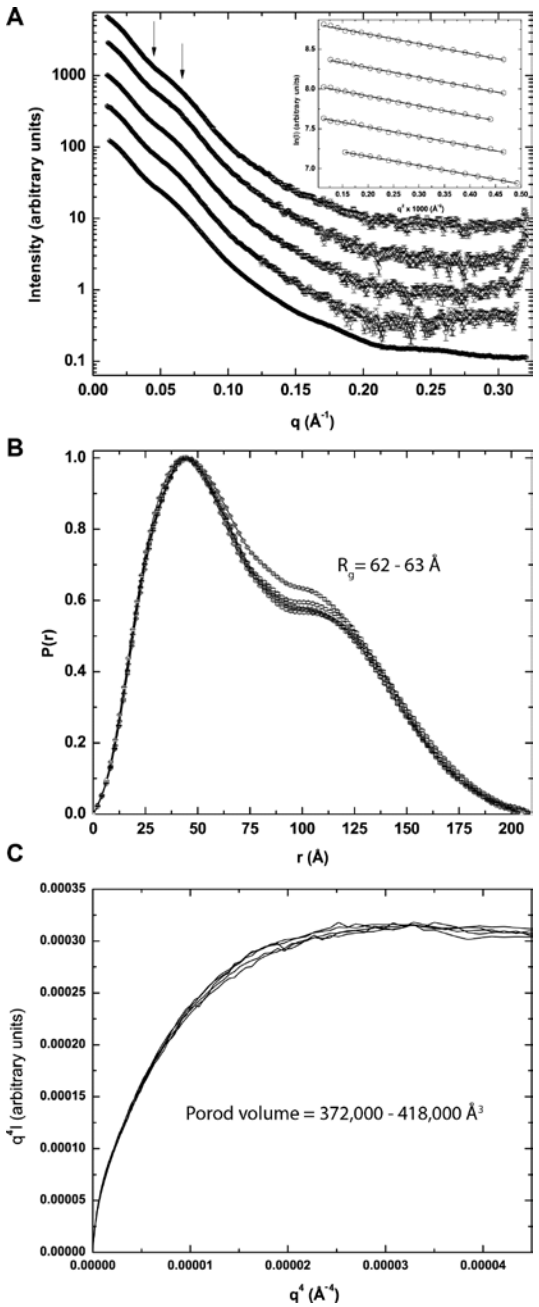


Figure 3.6. SAXS data for five EcPutA samples. (A) Composite scattering curves and Guinier plots (restricted to $qR_g \leq 1.3$). The arrows mark the bump and depression features at q = 0.045 Å^{-1} and q = 0.065 Å^{-1}, respectively, which are characteristic of full-length dimeric EcPutA. (B) P(r) curves. (C) Porod-Debye plots

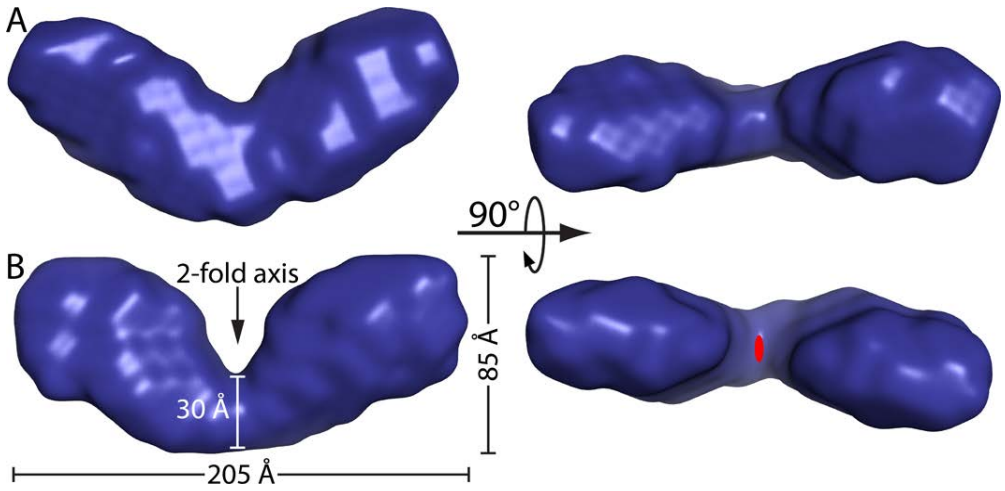


Figure 3.7. Consensus shape reconstructions for EcPutA calculated using GASBOR assuming *P*1 (A) and *P*2 (B) symmetries. Each surface represents the averaged, filtered volume based on 10 independent models from each of five samples (50 models total). Two orthogonal views of each shape are shown. The oval in panel B (right) represents the two-fold axis of the envelope.

The SAXS curve for PutA1–1085 is very similar to that of EcPutA (Figure 3.10A). In particular, the curve exhibits the depression and bump at q~0.045 $Å^{-1}$ and q ~ 0.065 $Å^{-1}$, respectively, which are characteristic of the full-length, dimeric protein. The P(r) curve for PutA1–1085 is strikingly similar to that of EcPutA (Figure 3.10B). The Rg value was 60 Å, which is comparable with the value of 63 Å for EcPutA. Porod-Debye plots for PutA1–1085 exhibit a well-defined plateau resulting in a Porod volume of 345,000 $Å^3$ (Figure 3.10C), which is just 13% smaller than that of EcPutA. These results suggest that PutA1–1085 is dimeric in solution.

The scattering curve for PutA86–1320 is noticeably different from that of EcPutA (Figure 3.10A). In particular, the characteristic features observed in EcPutA SAXS curves at q~0.045 $Å^{-1}$ and q ~0.065 $Å^{-1}$ are absent. The P(r) function shows a more profound difference, exhibiting just a single maximum near 41Å (Figure 3.10B). The Rg

of PutA86–1320 is 43 Å, which is 20 Å smaller than that of EcPutA. Also, the maximum particle dimension is 165 Å, which is about 40 Å shorter than that of EcPutA. The estimated Porod volume for PutA86–1320 is 186,000 Å3, which is about one-half of the volume of EcPutA. The molecular mass estimated from I(0) is 140 ± 14 kDa, which is within 4% of expected monomer molecular mass of 134 kDa (Table 3.1). Furthermore, analysis of PutA86–1320 using SEC-MALS suggests a molecular mass in solution of 144 kDa (Figure 3.5B). These data are consistent with PutA86–1320 being monomeric.

In summary, domain deletion analysis suggests that residues 1086–1320 are not essential for dimerization, and that an essential dimerization domain is located within residues 1–85. We suggest that the DNA-binding domain (residues 1–47) is the essential dimerization domain, because RHH domains bind DNA as obligate dimers (43). Furthermore, the DNA-binding domain of EcPutA has been expressed as an isolated protein (PutA52) and shown to form the classic RHH dimer in solution (9, 19).

Table 3.2

Kinetic parameters for EcPutA and domain deletion mutants

	PRODH activity			P5CDH activity
	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/K_m ($M^{-1}s^{-1}$)
EcPutA[a]	7.5[a]	100[a]	75[a]	1410 ± 31
PutA1-1085	10.6 ± 0.2	122 ± 5	87 ± 4	bd[b]
PutA86-1320	7.5 ± 0.1	67 ± 4	112 ± 7	783 ± 57

[a]From Vinod *et al*. (12).
[b]Below the detection limit of 0.03 mM NADH/min.

3.3.5. Rigid Body Modeling.

Rigid body modeling was performed to generate hypotheses about the spatial arrangement of domains in the EcPutA dimer. Two rigid bodies were used: the 1.9-Å resolution crystal structure of the DNA-binding domain dimer (PDB 2GPE, Figure 3.11A, left) and a hybrid x-ray/homology model of residues 87–1113 (Figure 3.11A, right). As described under "Experimental Procedures," the strategy used for modeling was based on the assumption that the DNA-binding domain is located in the connector section of the SAXS envelope with its 2-fold axis coincident with that of the envelope. This assumption is consistent with the domain deletion results, which show that the DNA-binding domain is essential for dimerization. Consequently, the catalytic units (residues 87–1113) correspond to the two large lobes of the envelope.

The best model (model 1), as judged by the lowest value from FoXS, is shown in Figure 3.11B. Model 1 shows good agreement with the experimental scattering profiles (Figure 3.11C). Note that the model curve exhibits the characteristic depression near q = 0.045 $Å^{-1}$ and bump near q = 0.065 $Å^{-1}$. The values calculated from FoXS for the fits to the five experimental curves are in the range of 2.0–5.2 over the entire q range. The fit is remarkably good, considering that 208 of the 1320 residues are missing in the homology model.

Several features of the model are notable. The catalytic units are oriented with the PRODH half of the polypeptide chain near the DNA-binding domain and the P5CDH half in the outermost part of the lobe. Residues 47 and 87 are located on the same face of the envelope and separated by 44 Å, which is close enough to be connected by 39 residues. The domains line the trench between the two lobes and face each other at a distance of 40 Å. A consequence of the large separation between the two catalytic units is that the putative substrate channeling cavity is open to the bulk medium (dashed oval in Figure 3.11B).

Other models generated by the COLORES docking calculations were examined to test the reliability of model 1. The top two of these alternative models, as judged by the fit to the experimental scattering profiles, are shown in Figure 3.12. In model 2 (χ = 3.2–10.2 for the five replicate EcPutA SAXS curves) COLORES positioned the catalytic units such that the domains face the convex surface of the envelope. With the catalytic units in this orientation, the DNA-binding surface is constrained to also face the convex side of the envelope for residues 47 and 87 to be connected by the intervening 39 residues. Thus, the DNA-binding domain dimer was manually rotated by 180° from that

of model 1. Model 3 (χ = 4.0–9.1) is different from models 1 and 2 in that the P5CDH domains are near the DNA binding domains, whereas the PRODH domains are in the distal ends of the lobes. Although models 2 and 3 have satisfactory agreement with the consensus SAXS shape, the fits to the scattering profiles are substantially worse than that of model 1. Thus, the experimental SAXS data are sufficiently sensitive to rule out models having domain orientations that are substantially different from those of model 1.

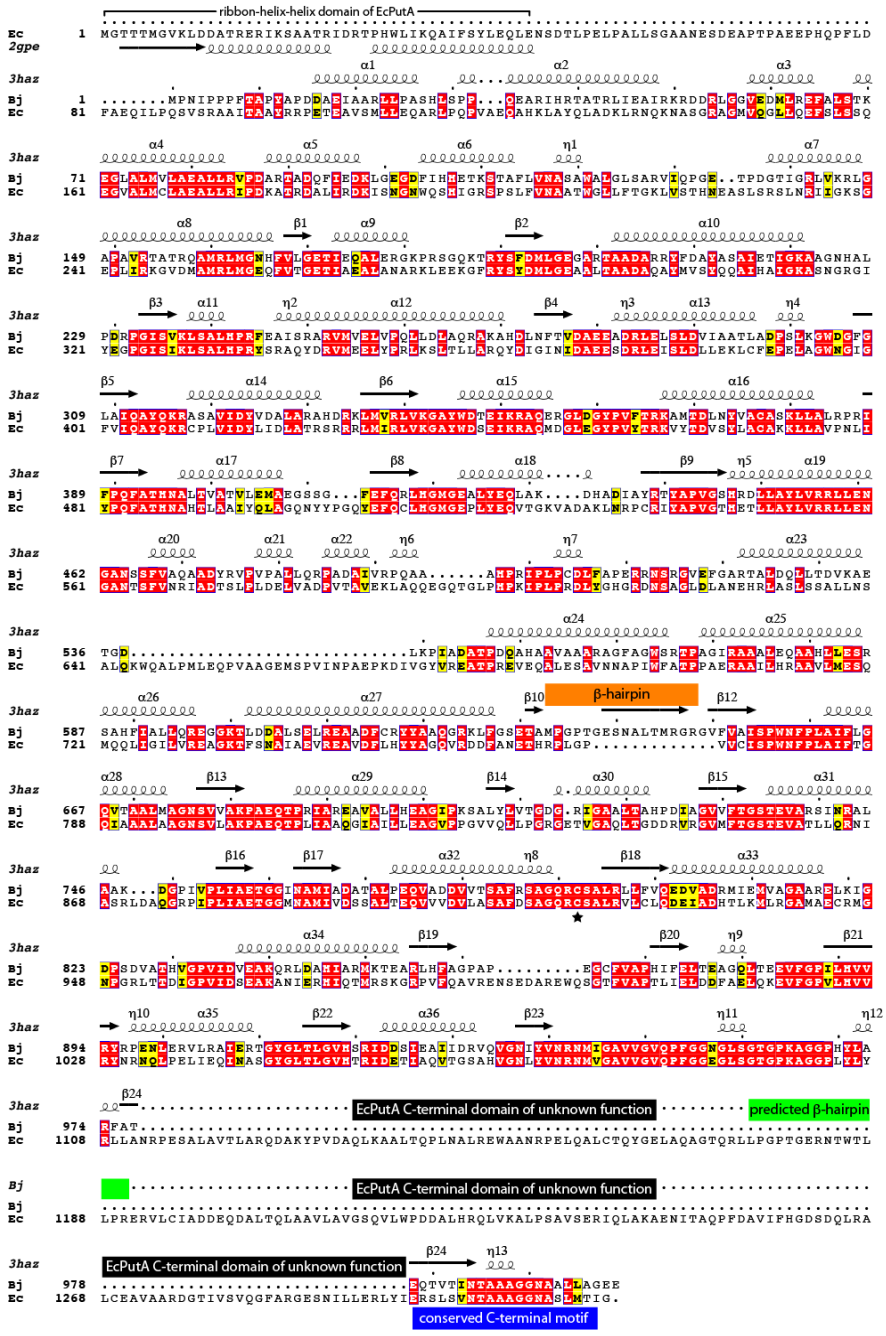


Figure 3.8. Global sequence alignment of BjPutA (Bj, GenBank BAC52526.1) and EcPutA (Ec, GenBank AAB59985.1). The secondary structure elements above the sequence are from the BjPutA structure (PDB code 3haz). The secondary structure elements below the sequence for the N-terminal ribbon-helix-helix domain are from a structure of the EcPutA DNA-binding domain (PDB code 2GPE). The star denotes the catalytic Cys of the P5CDH catalytic domain. The orange box denotes the β-hairpin of BjPutA, which is abbreviated in EcPutA. The green box denotes the β-hairpin of EcPutA predicted by remote homology detection. The blue box denotes the conserved C-terminal motif shared by minimalist and trifunctional PutAs.

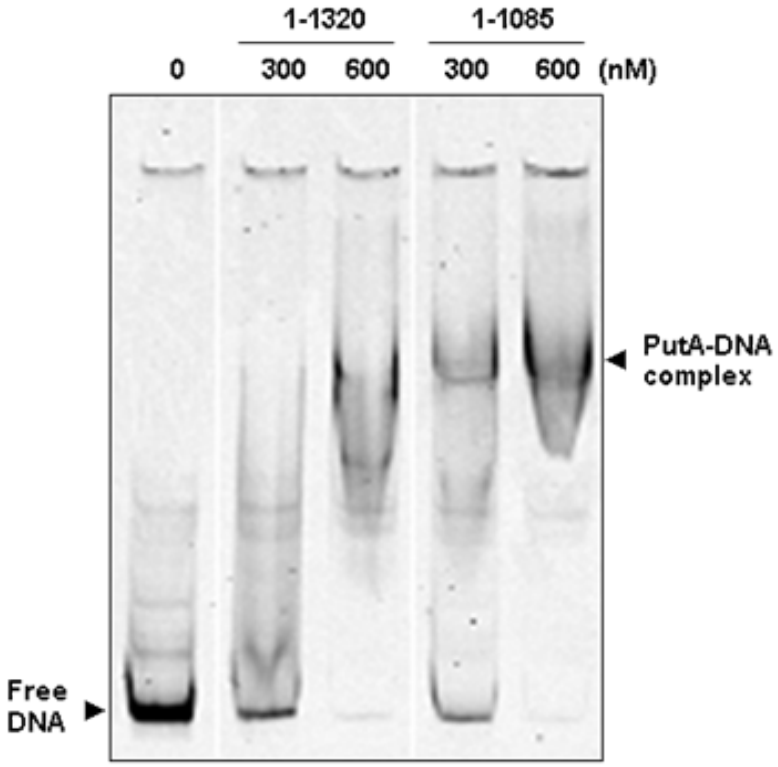


Figure 3.9. Gel-mobility shift assay of EcPutA and PutA1-1085. Two different concentrations of EcPutA and PutA1-1085 were added to binding mixtures containing IRdye-700 labeled *E. coli put* intergenic DNA (2 nM) and 100 µg/mL of nonspecific calf thymus DNA at 23°C.

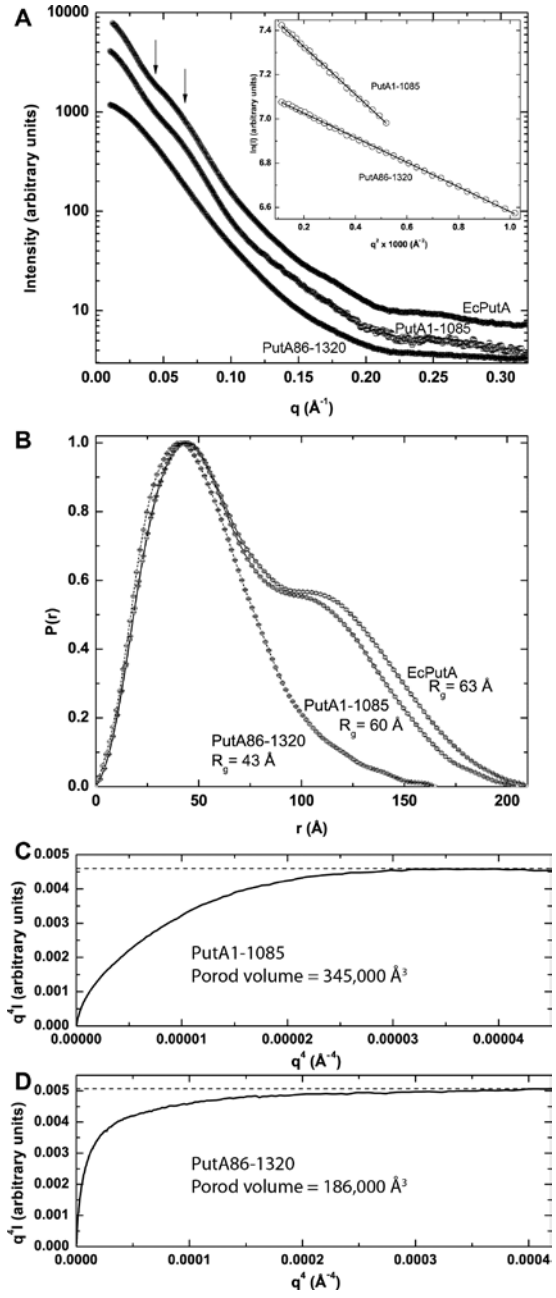


Figure 3.10. SAXS analysis of EcPutA domain deletion mutants PutA1-1085 and PutA86-1320. (A) Composite scattering curves and Guinier plots (restricted to $qR_g \leq 1.3$). The arrows mark q = 0.045 Å^{-1} and q = 0.065 Å^{-1}. (B) P(r) curves for the domain deletion mutants and EcPutA. (C) Porod-Debye plot for PutA1-1085. (D) Porod-Debye plot for PutA86-1320.

3.4. Conclusions

The oligomeric states and quaternary structures of PutAs are not conserved. To date, the oligomeric states of just two PutAs have been determined using rigorous biophysical methods. Almost 20 years ago, Brown and Wood (24) used sedimentation and light scattering to show that EcPutA forms a dimer in solution. We confirmed this result using MALS and SAXS. Last year, we reported SAXS and equilibrium sedimentation data showing that BjPutA, a minimalist bifunctional PutA, forms a ring-shaped, dimer-of-dimers tetramer having 222 symmetry (18). Thus, despite nearly 50% amino acid sequence identity, these two PutAs have different oligomeric states.

Considering the high identity between EcPutA and BjPutA, one might hypothesize that the EcPutA dimer resembles one of the three 2-body assemblies of the BjPutA tetramer. However, the R_g values of those assemblies are 44.8, 44.3, and 47.5 Å, which are much smaller than the R_g of 63 Å for EcPutA. Furthermore, the P(r) curves calculated from the BjPutA assemblies are distinctly different from that of EcPutA (Fig 3.13). In particular, the distributions of vectors in the BjPutA dimers lack the prominent shoulder at 110 Å, suggesting that the catalytic units are farther apart in EcPutA than in BjPutA. In fact, if two BjPutA protomers are separated as in rigid body model 1, the resulting P(r) exhibits the characteristic bimodal shape of EcPutA (Fig 3.13, *dashed curve*). We thus conclude that neither the oligomeric state nor the quaternary structure are conserved in the PutA family.

This lack of conservation perhaps makes sense considering the additional function of EcPutA as a transcriptional repressor, which requires dimerization of the DNA-binding domain. Indeed, the DNA-binding domain was found here to be essential for dimerization

of EcPutA. The connecting cylinder of the SAXS envelope accommodates the RHH dimer, but is not large enough to fit additional domains, suggesting that the DNA binding domain is the sole dimerization domain. Thus, it appears that function trumps homology in determining the oligomeric state and quaternary structure of PutA.

DNA is predicted to bind in the trench on the concave surface of the protein (Figure 3.14). The model is consistent with structural data on EcPutA-DNA association. In particular, side chains of the RHH domain that are known to contact DNA (9) are solvent exposed in the SAXS model. These critical residues are located in the β-strand (residues 5, 7, and 9) and N- terminus of the second helix of the RHH-fold (residues 28–30). The model suggests the hypothesis that elements outside of the RHH domain, such as residues in domain and PRODH barrel, may influence DNA binding. The SAXS model also provides new insights into the redox-dependent transcriptional regulation of the *putA* and *putP* genes by EcPutA. The 419-bp *put* control DNA region was previously shown to have five operator sites, with two of the operators (sites 3 and 4) separated by only one nucleotide (9). The mode of DNA binding predicted in our model (Figure 3.14) suggests that PutA binding to sites 3 or 4 preclude binding at the neighboring operator sequence. Thus, PutA most likely binds only four operator sites at one time to repress transcription of the *putA* and *putP* genes.

Consideration of the SAXS data in the context of existing biochemical and biophysical data on EcPutA provides a new model for gene regulation by EcPutA. Previous studies showed that reduction of the FAD causes just a 2-fold increase in the dissociation constant of PutA with *put* control DNA (11), indicating that the FAD redox state has little influence on the intrinsic affinity of EcPutA for DNA. On the other hand,

reduction of FAD increases the binding constant for membrane association by several orders of magnitude (15). Essentially, oxidized EcPutA has negligible affinity for the membrane, whereas the reduced protein exhibits nanomolar affinity. Two studies have shown that membrane binding and DNA binding are mutually exclusive (15, 44) . Finally, limited proteolysis and Trp fluorescence studies of EcPutA showed that FAD reduction induces a conformational change in the α domain, implying that reduction of the flavin triggers a global conformational change involving, in part, the α domain, that causes EcPutA to switch from being a transcriptional repressor to membrane bound enzyme (13, 17).

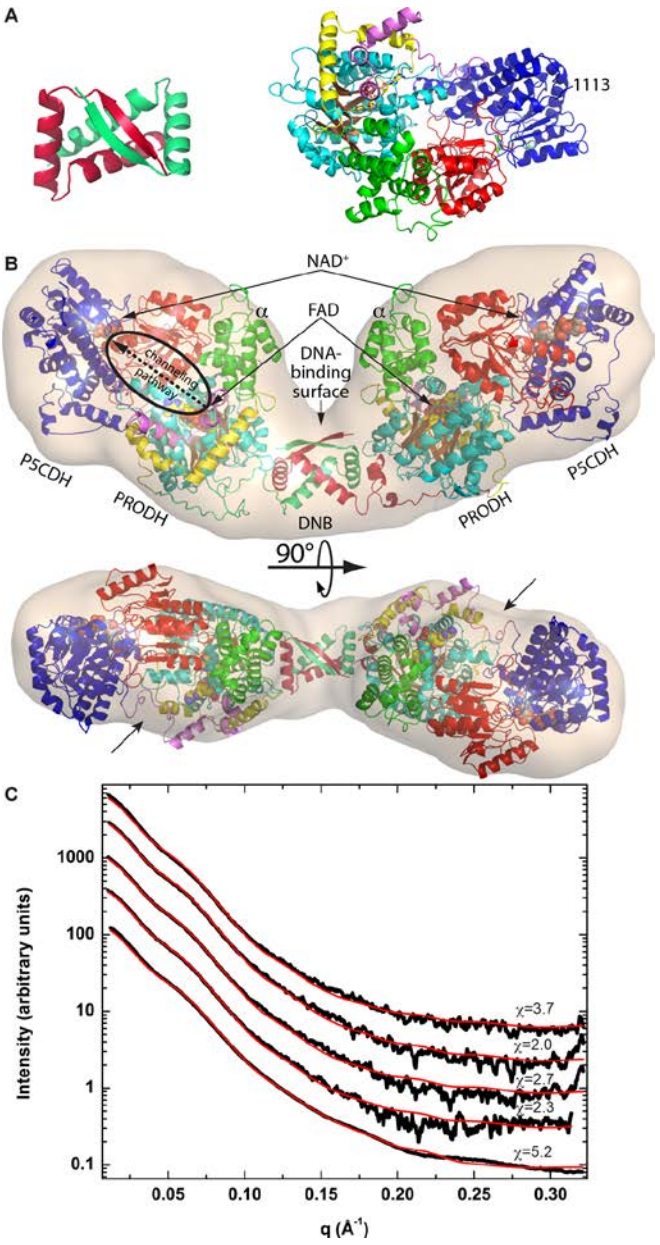


Figure 3.11. Rigid body model 1 of EcPutA. (A) Crystal structure of the DNA-binding domain dimer (left, PDB code 2GPE) and a hybrid X-ray/homology model for residues 87-1113 (right). The model of 87-1113 has the same orientation as the BjPutA protomer in Figure S1A. Note that this model is essentially identical to the BjPutA protomer, except the oligomerization flap is absent. The domains are colored according to the domain diagrams in Figgure 3.2. (B) Two views of the current working model of EcPutA. The dashed oval shows the location of the putative substrate-channeling cavity. As modeled, the cavity is open to bulk solvent. The arrows in the lower panel indicate the possible locations of the CTD. (C) Comparison of the experimental SAXS curves (black) and curves calculated from the model using FoXS (red).

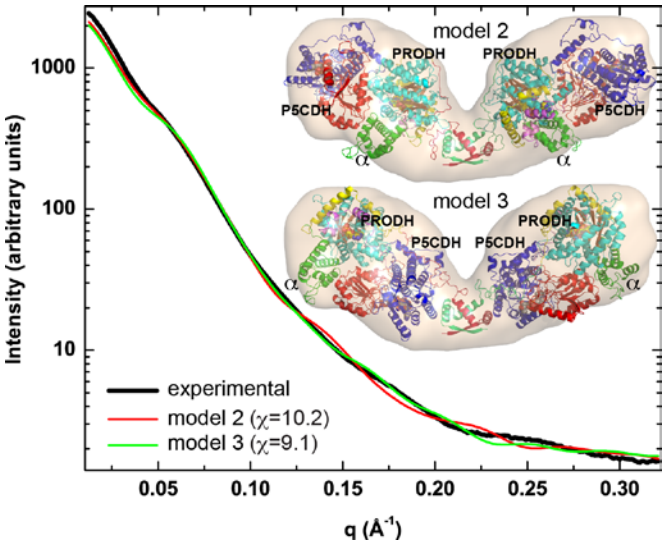


Figure 3.12. Two alternative models of EcPutA.

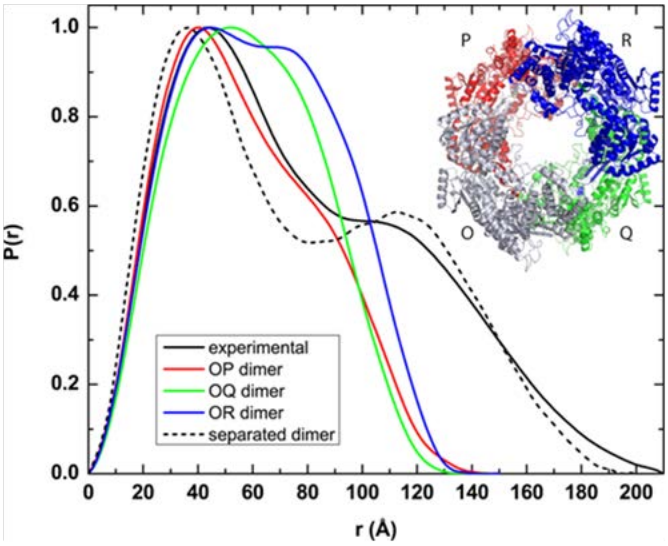


Figure 3.13. Comparison of an experimental P(r) curve for EcPutA (solid black) with theoretical curves calculated from the three two-body assemblies of the BjPutA tetramer (red, green, and blue). The chains of the tetramer are labeled O, P, Q, and R, and there are three unique dimeric assemblies: OP, OQ, and OR. The OP dimer corresponds to the one shown in Figure 3.2B. The dashed curve was calculated from a model in which two BjPutA protomers were superimposed onto SAXS model 1 (Figure 3.7B). The theoretical P(r) curves were calculated using GNOM from theoretical scattering data calculated using FoXS

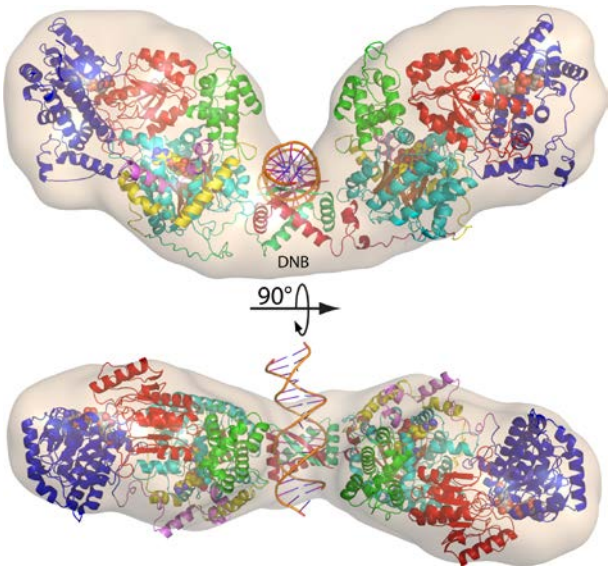


FIGURE 3.14. Model of DNA bound to SAXS model 1. This model was created by superimposing the crystal structure of the EcPutA DNA binding domain complexed with DNA (PDB code 2RBF) onto the DNA binding domain of the SAXS model 1.

A new model of gene regulation that is consistent with these observations and the SAXS model is that the DNA-binding and membrane-association interfaces are located on the same face of the protein, and that in the oxidized state, the former interface is exposed, whereas the latter is concealed. Reduction of FAD induces a conformational change that exposes the high affinity membrane-binding interface *without* disrupting the DNA-binding interface. The unveiling of the membrane-binding interface drives EcPutA to the membrane surface, which hides the DNA-binding interface from the *put* regulon thereby activating gene transcription. In SAXS model 1, the DNA-binding interface and both α domains are located on the same face of the protein, *i.e.* the concave face (Figure 3.11B). We thus hypothesize that the concave face of EcPutA supports both DNA- and membrane- binding, enabling a cloaking mechanism of gene regulation.

The SAXS model also provides insight into the nature of the global conformational changes associated with functional switching. The V-shape of oxidized EcPutA suggests hinge bending analogous to the ν_2 normal mode of water as a natural degree of freedom. Bending in one direction (closing) brings the α domains closer together and narrows the DNA-binding trench, resulting in a less elongated particle. Hinge bending in the opposite direction (opening) extends and flattens the dimer. Whether reduction of the FAD closes or opens the hinge is difficult to predict, but it may be possible to distinguish between these two general models for redox-linked conformational change using SAXS of reduced EcPutA and fluorescence resonance energy transfer of labeled EcPutA.

Spatial separation of the P5CDH domains, as in our rigid body models, is unprecedented in the aldehyde dehydrogenase (ALDH) gene superfamily. The superfamily includes $NAD(P)^+$-dependent enzymes catalyzing the oxidation of a variety of aldehyde substrates to their corresponding carboxylic acids(45, 46). P5CDHs belong to the ALDH4 family. Crystal structures of several ALDHs are known (47-49). All feature a 3-domain tertiary structure consisting of a Rossmann-like cofactor-binding domain (also called the N domain), a catalytic domain that furnishes the reactive Cys (also called the C domain), and an oligomerization flap domain formed by a β-hairpin and C-terminal β-strand. BjPutA and the monofunctional P5CDH from *Thermus thermophilus* (50) exhibit this defining architecture. The known structures also show that ALDHs form a domain-swapped dimer in which the oligomerization domain of one protomer interacts with the catalytic domain of the other protomer as in Figure 3.2C. In some ALDHs, the domain-swapped dimers assemble to form tetramers (48, 49) or hexamers (50).

The domain deletion and SAXS data suggest that dimerization of EcPutA is *not* mediated by traditional ALDH oligomerization flap domains. Although unprecedented in the ALDH superfamily, this result is consistent with multiple sequence alignments. In particular, the alignments clearly show that the β-hairpin of the minimalist PutA (residues 633–648) is truncated in trifunctional PutAs (Figure 3.8, *orange box*). Consequently, we hypothesize that the ALDH domains of trifunctional PutAs do not have the traditional dimerization flap seen in other ALDHs. Rather, our results suggest that trifunctional PutAs are unique members of the ALDH superfamily because of their mode of dimerization.

Finally, our results provide insight into the function of the CTD. The observation that PutA1–1085 is dimeric suggests that the CTD is not involved in dimerization. Remote homology detection analysis suggests that the CTD is homologous to the Rossmann domain of ALDHs. For example, HHSearch (35) identifies 33 ALDHs with probability scores of 99.1–99.9%, with the top match being BjPutA (probability = 99.9%, *E*-value = 7.7*E*-26). The alignment shows that the CTD of EcPutA is 25% identical to residues 551–761 of BjPutA (Figure 3.3). Of particular note is the prediction that CTD residues 1175–1190 (Figure 3.8, *green box*) form an ALDH β-hairpin homologous to the one found in BjPutA residues 633–648. As noted above, the substrate channeling cavity is open to the bulk medium in the SAXS models, which is inconsistent with the observation that PutAs exhibit several kinetic signatures of substrate channeling (18, 51) (Dr. Don Becker group, unpublished results) These results suggest the hypothesis that the CTD occupies the vacant space in the SAXS envelope between the two active sites (*arrows in lower part* of Figure 3.11B), thereby forming an intramolecular lid analogous to the intermolecular lid

of BjPutA. Future studies will be needed to test this, and other, hypotheses raised by the model of EcPutA proposed here.

3.5. References.

1. Tanner JJ. Structural biology of proline catabolism. Amino Acids. 2008;35(4):719-30.

2. Zhou Y, Zhu W, Bellur PS, Rewinkel D, Becker DF. Direct linking of metabolism and gene expression in the proline utilization A protein from Escherichia coli. Amino Acids. 2008;35(4):711-8. Epub 2008/03/08.

3. Menzel R, Roth J. Regulation of genes for Proline Utilization in *Salmonella typhimurium*: Autogenous Repression by the *putA* gene Product. J Mol Biol. 1981;148:21-44.

4. Maloy S, Roth JR. Regulation of Proline Utilization in *Salmonella typhimurium*: Characterization of *put*:Mu *L*(Ap, *lac*) Operon Fusions. J Bacteriol. 1983;154:561-8.

5. Wood JM. Membrane association of proline dehydrogenase in Escherichia coli is redox dependent. Proc Natl Acad Sci USA. 1987;84(2):373-7. Epub 1987/01/01.

6. Ostrovsky de Spicer P, Maloy S. PutA protein, a membrane-associated flavin dehydrogenase, acts as a redox-dependent transcriptional regulator. Proc Natl Acad Sci USA. 1993;90(9):4295-8.

7. Muro-Pastor AM, Maloy S. Proline Dehydrogenase Activity of the Transcriptional Repressor PutA is Required for Induction of the *put* Operon by Proline. J Biol Chem. 1995;270(17):9819-27.

8. Vilchez S, Manzanera M, Ramos JL. Control of expression of divergent Pseudomonas putida put promoters for proline catabolism. Appl Environ Microbiol. 2000;66(12):5221-5.

9. Zhou Y, Larson JD, Bottoms CA, Arturo EC, Henzl MT, Jenkins JL, et al. Structural basis of the transcriptional regulation of the proline utilization regulon by multifunctional PutA. J Mol Biol. 2008;381(1):174-88.

10. Brown ED, Wood JM. Conformational change and membrane association of the PutA protein are coincident with reduction of its FAD cofactor by proline. J Biol Chem. 1993;268(12):8972-9.

11. Becker DF, Thomas EA. Redox properties of the PutA protein from Escherichia coli and the influence of the flavin redox state on PutA-DNA interactions. Biochemistry. 2001;40(15):4714-21.

12. Vinod MP, Bellur P, Becker DF. Electrochemical and functional characterization of the proline dehydrogenase domain of the PutA flavoprotein from *Escherichia coli*. Biochemistry. 2002;41:6525-32.

13. Zhu W, Becker DF. Flavin redox state triggers conformational changes in the PutA protein from Escherichia coli. Biochemistry. 2003;42(18):5469-77.

14. Lee YH, Nadaraia S, Gu D, Becker DF, Tanner JJ. Structure of the proline dehydrogenase domain of the multifunctional PutA flavoprotein. Nat Struct Biol. 2003;10(2):109-14.

15. Zhang W, Zhou Y, Becker DF. Regulation of PutA-membrane associations by flavin adenine dinucleotide reduction. Biochemistry. 2004;43(41):13165-74.

16. Zhang M, White TA, Schuermann JP, Baban BA, Becker DF, Tanner JJ. Structures of the Escherichia coli PutA proline dehydrogenase domain in complex with competitive inhibitors. Biochemistry. 2004;43(39):12539-48.

17. Zhu W, Becker DF. Exploring the proline-dependent conformational change in the multifunctional PutA flavoprotein by tryptophan fluorescence spectroscopy. Biochemistry. 2005;44(37):12297-306. Epub 2005/09/15.

18. Srivastava D, Schuermann JP, White TA, Krishnan N, Sanyal N, Hura GL, et al. Crystal structure of the bifunctional proline utilization A flavoenzyme from Bradyrhizobium japonicum. Proc Natl Acad Sci U S A. 2010;107(7):2878-83. Epub 2010/02/06.

19. Larson JD, Jenkins JL, Schuermann JP, Zhou Y, Becker DF, Tanner JJ. Crystal structures of the DNA-binding domain of Escherichia coli proline utilization A flavoprotein and analysis of the role of Lys9 in DNA recognition. Protein Sci. 2006;15:1-12.

20. Zhang W, Zhang M, Zhu W, Zhou Y, Wanduragala S, Rewinkel D, et al. Redox-induced changes in flavin structure and roles of flavin N(5) and the ribityl 2'-OH group in regulating PutA--membrane binding. Biochemistry. 2007;46(2):483-91.

21. Ostrander EL, Larson JD, Schuermann JP, Tanner JJ. A conserved active site tyrosine residue of proline dehydrogenase helps enforce the preference for proline over hydroxyproline as the substrate. Biochemistry. 2009;48(5):951-9. Epub 2009/01/15.

22. Srivastava D, Zhu W, Johnson WH, Jr., Whitman CP, Becker DF, Tanner JJ. The structure of the proline utilization a proline dehydrogenase domain inactivated by N-propargylglycine provides insight into conformational changes induced by substrate binding and flavin reduction. Biochemistry. 2010;49(3):560-9. Epub 2009/12/10.

23. Ratzkin B, Roth J. Cluster of genes controlling proline degradation in Salmonella typhimurium. J Bacteriol. 1978;133(2):744-54.

24. Brown ED, Wood JM. Redesigned purification yields a fully functional PutA protein dimer from Escherichia coli. J Biol Chem. 1992;267(18):13086-92.

25. Zhu W, Gincherman Y, Docherty P, Spilling CD, Becker DF. Effects of proline analog binding on the spectroscopic and redox properties of PutA. Arch Biochem Biophys. 2002;408(1):131-6.

26. Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL, 2nd, Tsutakawa SE, et al. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods. 2009;6(8):606-12. Epub 2009/07/22.

27. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr. 2003;36(5):1277-82.

28. Svergun D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr. 1992;25(4):495-503.

29. Mylonas E, Svergun DI. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. J Appl Cryst. 2007;40:s245-s9.

30. Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. Biophys J. 2001;80(6):2946-53. Epub 2001/05/24.

31. Volkov VV, Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. J Appl Crystallogr. 2003;36(3 Part 1):860-4.

32. Wriggers W. Using Situs for the integration of multi-resolution structures. Biophys Rev. 2010;2(1):21-7. Epub 2010/02/23.

33. Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. J Appl Crystallogr. 2001;34(1):33-41.

34. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9:40. Epub 2008/01/25.

35. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006;22(2):195-201. Epub 2005/11/23.

36. Chacon P, Wriggers W. Multi-resolution contour-based fitting of macromolecular structures. J Mol Biol. 2002;317(3):375-84. Epub 2002/04/02.

37. Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res. 2010;38(Web Server issue):W540-4. Epub 2010/05/29.

38. de Bakker PI, DePristo MA, Burke DF, Blundell TL. Ab initio construction of polypeptide fragments: Accuracy of loop decoy discrimination by an all-atom statistical potential and the AMBER force field with the Generalized Born solvation model. Proteins. 2003;51(1):21-40. Epub 2003/02/22.

39. Potterton E, Briggs P, Turkenburg M, Dodson E. A graphical user interface to the CCP4 program suite. Acta Cryst. 2003;D59(Pt 7):1131-7.

40. Williams I, Frank L. Improved chemical synthesis and enzymatic assay of delta-1-pyrroline-5-carboxylic acid. Anal Biochem. 1975;64(1):85-97. Epub 1975/03/01.

41. Easley KE, Sommer BJ, Boanca G, Barycki JJ, Simpson MA. Characterization of human UDP-glucose dehydrogenase reveals critical catalytic roles for lysine 220 and aspartate 280. Biochemistry. 2007;46(2):369-78. Epub 2007/01/11.

42. Rambo RP, Tainer JA. Characterizing flexible and instrinsically unstructured biological macromolecules by SAS using the porod-debye law. Biopolymers. 2011;95(8):559-71. Epub 2011/04/22.

43. Schreiter ER, Drennan CL. Ribbon-helix-helix transcription factors: variations on a theme. Nat Rev Microbiol. 2007;5(9):710-20.

44. Muro-Pastor AM, Ostrovsky P, Maloy S. Regulation of Gene Expression by Repressor Localization: Biochemical Evidence that Membrane and DNA Binding by the PutA Protein are Mutually Exclusive. J Bacteriol. 1997;179(8):2788-91.

45. Sophos NA, Vasiliou V. Aldehyde dehydrogenase gene superfamily: the 2002 update. Chem Biol Interact. 2003;143-144:5-22.

46. Yoshida A, Rzhetsky A, Hsu LC, Chang C. Human aldehyde dehydrogenase gene family. Eur J Biochem. 1998;251(3):549-57. Epub 1998/03/07.

47. Liu ZJ, Sun YJ, Rose J, Chung YJ, Hsiao CD, Chang WR, et al. The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. Nat Struct Biol. 1997;4(4):317-26.

48. Moore SA, Baker HM, Blythe TJ, Kitson KE, Kitson TM, Baker EN. Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. Structure. 1998;6(12):1541-51. Epub 1998/12/24.

49. Steinmetz CG, Xie P, Weiner H, Hurley TD. Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. Structure. 1997;5(5):701-11. Epub 1997/05/15.

50. Inagaki E, Ohshima N, Takahashi H, Kuroishi C, Yokoyama S, Tahirov TH. Crystal structure of Thermus thermophilus Delta1-pyrroline-5-carboxylate dehydrogenase. J Mol Biol. 2006;362(3):490-501.

51. Surber MW, Maloy S. The PutA protein of Salmonella typhimurium catalyzes the two steps of proline degradation via a leaky channel. Arch Biochem Biophys. 1998;354(2):281-7.

CHAPTER 4

Structural Diversity of the Proline Utlization A (PutA) Family Revealed by Small Angle X-ray Scattering.

4.1. Introduction

Proline utilization A (PutA) catalyzes the oxidation of proline to glutamate (Figure 4.1). PutAs are bifunctional enzymes and have two catalytic domains. The first catalytic domain corresponds to an FAD dependent proline dehydrogenase (PRODH) active site and is accountable for catalyzing the oxidation of proline to Δ^1-pyrroline-5-carboxylate (P5C), whilst the second domain accounts for a NAD^+ dependent P5C dehydrogenase (P5CDH) active site which catalyzes the oxidation of glutamate semialdehyde (GSA) to glutamate. These two steps are connected with a non-enzymatic step, which hydrolyses P5C to GSA. Interestingly, some PutAs have an additional transcriptional repressor function and are thus trifunctional. These PutAs have a third functional domain, which is a DNA binding domain (DBD) and accounts for the auto-transcriptional regulation activity of the enzyme (1-4).

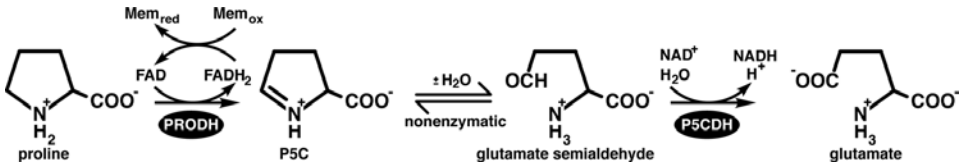


Figure 4.1. Reactions catalyzed by PutA. Figure taken from Singh *et.al* 2011.

The phylogenetic tree in Figure 2.2 of Chapter 2 shows that PutAs are divided into two distinct branches, simply called branch 1 and branch 2. Sequence alignment of branch 1 PutAs was the subject of investigation in Singh *et al.* FBS 2012. Here is the brief summary of finding of that work. The pairwise sequence identity within branch 1 varies from 39% to as high as 99%. Figure 4.2 and Appendix I show the sequence alignment of branch 1 PutAs. Important regions identified by the alignment include the PRODH and P5CDH catalytic domains and a C-terminal motif (CCM), which are highly

conserved in branch 1 PutAs. However, the sequence alignment also shows some striking differences within the branch 1. Those differences divide branch 1 PutA into three different kinds, as described next.

Short bifunctional PutAs. have polypeptide chain lengths varying from 999 (*Bradyrhizobium japonicum*, BjPutA) to 1080 (*Psychrobacter arcticus*). This class of bifunctional PutA is termed as "short branch 1 bifunctional" or "minimalist" PutA. The crystal structure of BjPutA has been determined (5). In addition, the Becker group reported biochemical studies of BjPutA (6, 7).

Long bifunctional PutAs. Have chain length varies from 1127 (*Rhodobacter capsulatus,* RcPutA) to 1262 (*Acinetobacter sp.*). This class of bifunctinal PutA called "long branch 1 bifunctional PutA". Our lab has determined the oligomeric state of RcPutA. Otherwise, long bifunctional PutAs have not been investigated extensively.

Trifunctional PutAs. have chain lengths from 1270 (*Acidiphilium cryptum)* to 1361 (*Rhodoferax ferrireducens*). The trifunctional PutAs from *Escherichia coli* (EcPutA) and Salmonella typhimurium (StPutA) have been extensively studied (2, 8-16).

Based on the known crystal structure of minimalist BjPutA and the current sequence alignment, two important attributes differentiate long and short branch 1 bifunctional PutAs: the β-hairpin dimerization domain and C-terminal domain (CTD). Short branch 1 PutAs have the β-hairpin dimerization domain and lack the CTD. On the other hand, long branch 1 PutAs have the CTD but lack the β-hairpin dimerization domain. The domains are discussed in more detail next.

The β-hairpin is an essential component of the domain swapped dimerization and substrate channeling phenomenon in short PutAs, such as BjPutA (residues 635-646). It

is also important to mention that this β-hairpin is homologous to oligomerization domain of a protein of the aldehyde dehydrogenase (ALDH) superfamily, of which PutAs are a member.

Long bifunctional PutAs and trifunctional PutAs have an extra ~200-residue domain that is not found in the minimalist PutAs, which is called the C-terminal domain (CTD). The CTD is an insertion between the conserved 17-residue C-terminal motif (CCM) and the NAD^+-binding domain of the P5CDH active site. A sequence alignment of trifunctinal PutA with these long bifunctional PutAs clearly shows the sequence similarity in the C-terminal end (Figure 4.3). The pairwise sequence identity between the CTD of trifunctional and the long bifunctional PutAs varies from 14 to 35%. The role of these extra residues is uncertain. However, recent work suggests that the CTD adopts a Rossmann fold (17).

A β-hairpin corresponding to residues 635-646 of minimalist BjPutA appears to be absent at the analogous position of the longer bifunctional and trifunctional PutAs. The region is shown in Figure 4.2. The absence of the β-hairpin at the analogous position in long bifunctional and trifunctional PutAs raises the question about the mechanism of oligomerization in these PutAs. Contrary to a typical ALDH member, our recent work on the EcPutA explains that the N-terminal DNA binding domain is responsible for dimerization of trifunctional PutAs (18). The mechanism of oligomerization is uncertain in long bifunctional PutAs. However, careful inspection of the sequence analysis suggests that a region in the CTD is similar to the sequence of the β-hairpin of BjPutA. It is not clear how this possible shift of the β-hairpin from the ALDH region to the CTD affects the oligomerization of long bifunctional PutAs. However, we have developed a working

hypothesis that the apparent β-hairpin in the CTD is not used for dimerization and that branch 1 long bifunctional PutAs are monomeric.

Branch 2 PutAs consist of only bifunctional PutAs (Table 4.2). The pairwise sequence identity is as low as 23 to 65 % (Table 4.3). The low sequence identity suggests that the branch 2 PutAs are more diverse than branch 1 PutAs. The pairwise sequence identity of branch 2 PutA with the minimalist BjPutA varies from 19 to 29 %. Based on the length of the amino acid sequence, branch 2 PutAs can also be divided into two kinds: short bifunctional branch 2 PutAs and long bifunctional branch 2 PutAs. Unlike branch 1 PutAs, the sequence alignment does not suggest the presence of a conserved C-terminal motif in branch 2 PutAs. In addition, it is not clear whether the β-hairpin is present in the ALDH region of the longer bifunctional PutAs. The sequence analysis of branch 2 PutAs suggests that the extra C-terminal region is more like an appendage at the C-terminal end rather than an insertion as in longer branch 1 PutAs.

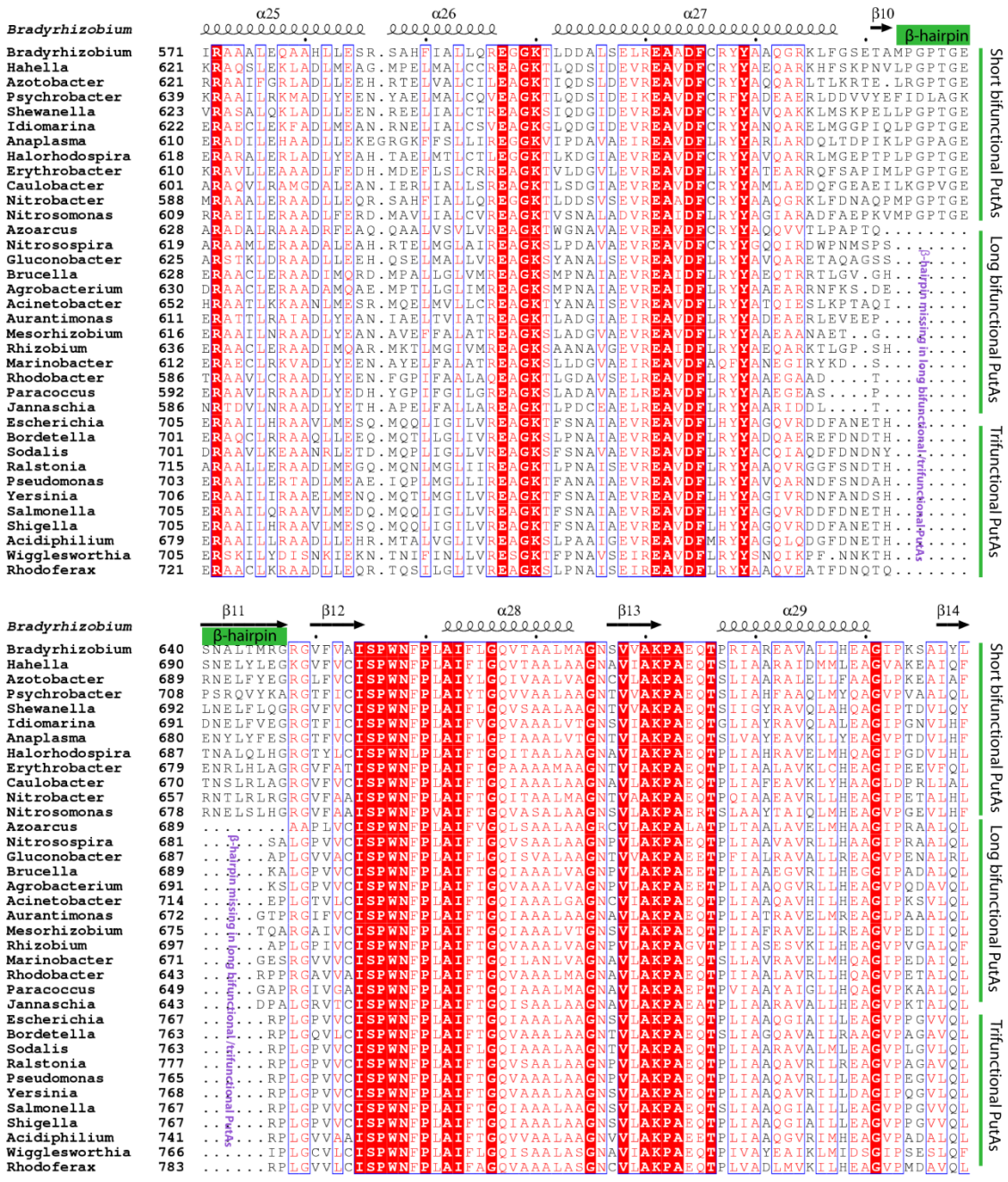


Figure 4.2. A region of the sequence alignment of branch 1 PutAs shows a missing β-hairpin region in long bifunctional and trifunctional branch 1 PutAs. The secondary structure alignment shown on the top corresponds to the structure of BjPutA (first sequence). The alignment was generated from ClustalW and ESPript (19, 20).

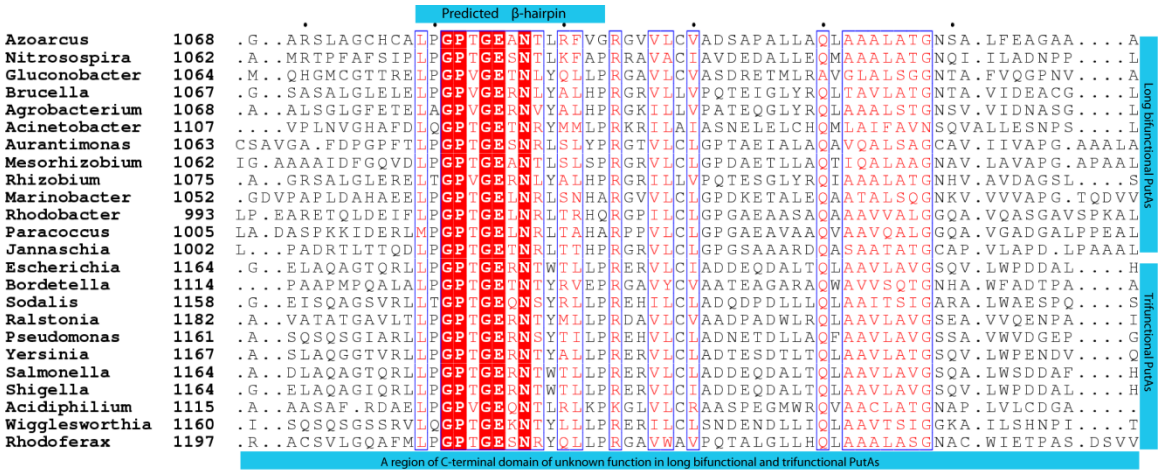


Figure 4.3. Alignment of the CTDs of long bifunctional and trifunctional PutAs of branch 1. It also shows a conserved region predicted to be the β-hairpin in these kinds of PutAs. The alignment is generated from ClustalW and ESPript (19, 20).

One of the well-studied examples of a short bifinctional PutA is from *Bradyrhizobium japonicum* (BjPutA). This PutA has 999 amino acid residues and a known crystal structure (5). Based on the number of amino acid residues and the known crystal structure, BjPutA is considered to be a minimalist PutA (17). The crystal structure and the biochemical analysis suggest that BjPutA is a dimer-of-dimers, a tetramer (5). The dimerization interface is formed by a domain-swapping event. The crystal structure of BjPutA has shown that the β-hairpin and conserved C-terminal motif are involved in making a domain swapped dimer interface with the catalytic domain of another subunit. Insertion or appendage of extra residues near the C-terminal, or absence of β-hairpin in branch 1 longer bifunctional PutAs raises the question about the oligomeric states of these classes of PutAs.

We do not have any biochemical and structural information regarding long branch 1 or branch 2 bifunctional PutAs. BjPutA is the only short bifunctional PutA which has

been studied comprehensively, although a structure of the short branch 2 PutA from *Geobacter sulfurreducens* PCA (GsPutA) has been determine been determined (PDB code 4f9i). The minimalist BjPutA is a tetramer, GsPutA is a traditional ALDH domain-swapped dimer, and the trifunctional EcPutA is an N-terminal DNA binding domain mediated dimer. The study of short and long bifunctional PutAs is required to understand the structural diversity within the PutA family. Here we have chosen various bifunctional PutAs from branch 1 and branch 2 to understand the diversity in the oligomeric states of PutAs within these two branches. Small angle X-ray scattering (SAXS) was used to study the oligomeric states of 5 bifunctional PutAs. Table 4.1 is the list of PutAs whose oligomeric states are discussed in this chapter.

Table 4.1. List of PutAs under investigation

Organism	Abbreviation	Type
Rhodopseudomonas palustris	RpPutA	Branch 1 Short bifunctional
Rhodobacter capsulatus	RcPutA*	Branch 1 Long bifunctional
Desulfovibrio vulgaris	DvPutA	Branch 2 Short bifunctional
Corynebacterium jeikeium	CjPutA	Branch 2 Long bifunctional
Helicobacter pylori	HpPutA*	Branch 2 Long bifunctional

*The work related to RcPutA and HpPutA was performed by Min Luo and only the final oligomeric state of these proteins is mentioned in conclusion section.

Table 4.2: Pair wise sequence identity of Branch 2 PutAs calculated from ClustalW2.

Organism	residues	Abbreviation	Identity with BjPutA
Campylobacter jejuni subsp. Jejuni	1162	Cje	19
Helicobacter pylori	1185	Hp	19
Bacteroides sp.	1144	Bs	22
Corynebacterium jeikeium	1158	Cj	24
Geobacter bemidjiensis Bem	1004	Gb	26
Desulfovibrio vulgarisstr. Hildenborough	1006	Dv	26
Bdellovibrio bacteriovorus	982	Bb	28
Trichodesmium erythraeum	993	Te	28
Anabaena variabilis	993	Av	29
Synechococcus sp.	1007	Ss	29
Gloeobacter violaceus	996	Gv	29

Table 4.3: Pairwise sequence alignment of Branch 2 PutAs from ClustalW2.

	Cje	Hp	Bs	Cj	Gb	Dv	Bb	Te	Av	Ss	Gv
Cje	100	65	37	30	23	24	25	24	24	23	23
Hp		100	37	31	24	24	25	25	25	24	24
Bs			100	33	23	25	27	26	27	26	26
Cj				100	26	26	27	27	26	26	28
Gb					100	63	47	42	44	43	41
Dv						100	47	44	44	44	43
Bb							100	44	44	45	45
Te								100	73	67	51
Av									100	69	54
Ss										100	51
Gv											100

4.2. Materials and Methods

Three PutAs have been expressed and purified: *Corynebacterium jeikeium* Put A (CjPutA), *Desulfovibrio vulgaris* PutA (DvPutA), *Rhodopseudomona palustris* (RpPutA). The purification methods for RpPutA were performed by Min Luo, will not be discussed here.

4.2.1. Protein expression and purification

The gene of CjPutA was synthesized from Bio basic. CjPutA gene was cloned in vector pKA8H by Min Luo between Nde1 and BamH1 restriction sites. The vector has an N-terminal His_8-tag and an intervening Tobacco Etch Virus Protease enzyme (TEVP) cleavage site. Treatment with TEVP results in the native polypeptide preceded by Gly-His. The New York Structural Genomic Research Consortium generously provided the clone of DvPutA in PNIC-28-BsaI vector The expressed enzyme has a N-terminal His_8-tag and an intervening TEVP cleavage site.

All the proteins mentioned above were purified as follows. The protein was expressed using an *E.coli* strain BL21(DE3)pLysS. A 10 mL overnight culture for each 1000 mL LB media was used to inoculate. The cells were grown at 37^{o}C at 250 rpm and induced with 0.5 mM IPTG at OD_{600} = 0.6 with an induction temperature of 22 ^{o}C at 200 rpm for 18 hours. Cells were harvested and frozen in 50 mM Hepes pH 8.0, 300 mM NaCl, 10 mM Imidazole and 5% glycerol. The frozen cells were thawed before purification, with the addition of protease inhibitors (AEBSF, TPCK, E64, Pepstatin, Leupeptin) and cells were broken by sonication. Cell debris and unbroken cells were separated by centrifugation at 16,500 rpm for one hour in a SS34rotor. The supernatant was applied to

a His-Trap HP column (His-Trap Ni^{2+}-Sepharose HP, GE Healthcare) equilibrated with 20 mM Hepes, 300 mM NaCl, and 5% glycerol at pH 8.0. The protein of interest was eluted with equilibration buffer congaing 300 mM imidazole. The fractions were pooled, and TEVP, 1M THP, and 20X TEV buffer (1 M Tris-HCl, 10 mM EDTA, pH 8.0) were added so that the final solution contained 5 mg of TEVP per 30 mg of protein in 50 mM Tris–HCl, 0.5 mM THP, and 0.5 mM EDTA, pH 8.0. The sample was incubated at 30^{o}C for 2 hrs, and dialyzed overnight at 4 ^{o}C and injected onto the His-Trap HP column. The untagged protein was collected at 30 mM imidazole and dialyzed in 50 mM Tris-HCl, 0.5 mM THP, 5% glycerol, 0.5 mM EDTA, pH 8.0 in preparation for further purification using anion exchange chromatography (HiTrap Q). The protein was bound to the HiTrap Q anion exchange column equilibrated with a buffer similar to the dialysis buffer and was eluted with a linear 0-1 M NaCl gradient. Size exclusion chromatography (Superdex 200, 25 mL) was used as the final step of purification.

4.2.2. Small Angle X-Ray Scattering (SAXS)

SAXS experiments were performed at beamline 12.3.1 of the Advanced Light Source via the mail-in program (1). Prior to analysis, all protein samples were subjected to size exclusion chromatography (SEC) using a Superdex 200 column, and the protein from the major peak was collected and dialyzed overnight with three buffer changes. The final dialysis buffer was reserved for the reference for the SAXS experiment. Scattering intensities (I) were measured at three nominal protein concentrations using exposure times of 0.5, 1.0, 3.0, and 6.0 sec. For DvPutA, the protein concentrations were 1.2, 3.0, and 4.1 mg / mL, based on the BCA (Pierce) assay, whereas CjPutA was used at

concentrations of 1.0, 2.2, and 3.6 mg/mL (BCA). The protein concentration of RpPutA was not measured prior to the SAXS experiment, although the three different dilutions were in the ratio of 3:1.5:1.

The SAXS data were analyzed as follows. The scattering curves collected from the protein sample were corrected for background scattering using intensity data collected from the dialysis buffer. A composite scattering curve for each sample was generated with PRIMUS (21) by scaling and merging the high q region from one of the longer time exposures with the low q region from a lower time exposure. Table XX shows the summary of how composite data sets were generated for each concentration for the different proteins under investigation. The scattering curves were multiplied with a concentration factor and overlaid on each other to check concentration dependent variation of the profile. A plot of ln (I) versus q^2 was plotted at the lower q region ($qR_g<1.3$). This region is called Guinier region, and the Guinier plot for a well behaved, non-aggregated protein that is free of interparticle interference should be linear in this region. The Guinier regions used for the various samples are listed as footnotes to Tables 4.1, 4.2 and 4.3). The AutoRg utility of PRIMUS was also used to validate the Guinier region obtained manually. AutoR_g finds the q region that yields the best linear fit and outputs a quality index of the fit expressed as a percentage. One-hundred percent is the best score; factors that affect the score include the number of points at low q that are omitted and the residual from linear regression analysis. GNOM was used to calculate the pair distribution function (P(r)) in order to estimate the radius of gyration (R_g) and the maximum particle dimension (D_{max}) (22). MOLEMAN was used to calculate the R_g from atomic coordinates.

Shape reconstruction calculations were performed using the *ab initio* based program GASBOR (23). The D_{max} values used for the shape reconstructions are listed in Tables 4.3, 4.4 and 4.5. Each reconstruction consisted of 18-20 independent GASBOR calculations, and the models were averaged and filtered using damaver. The point group symmetry used during the GASBOR calculations was *P*1or *P*2. The resulting damfilt pdb from damaver was converted into a volumetric map using pdb2vol. The FoXS server was used to calculate SAXS curve from atomic coordinates (24).

4.3. Results

4.3.1. DvPutA

The composite scattering profiles and the corresponding Guinier plot are shown in Figure 4.4. The Guinier plot exhibits good linearity for the three different concentrations and suggests and R_g value in the range of 43.2-43.7Å (Table 4.4 and Figure 4.4c). The Figure 4.4b shows scattering curves multiplied by the dilution factor of the sample. The q range shown in the figure is $0 < q < 0.15$ $Å^{-1}$. The figure clearly shows that the scaled curves are essentially identical, which indicates that I(0)/C does not vary significantly with concentration. These results confirm that the particle is free from concentration dependent behavior. The AutoR_g software gives an R_g estimate of 45.0 - 45.9 Å. The slight discrepancy in R_g is due to the fact that AutoR_g looks for the best linear fit in the low q region, and for that it sacrifices some of the initial data points (among lowest q). Moreover, the pair distribution analysis gives an R_g value in the range of 46.4 to 47.4Å. The R_g range calculated from pair distribution function (46.8 – 47.4 Å) is more reliable as it is not as sensitive to the data points at low q region. Calculation of pair distribution

function also gives a maximum particle dimension in the range of 160 - 164Å.

For DvPutA, the medium concentration sample was used for the shape reconstruction. Figure 4.5 shows volumetric map of the damfilt model of the GASBOR reconstruction assuming *P1 and P2* symmetry. The volumetric map obtained from GASBOR constrained in P1 and P2 shape reconstruction suggests that the particle has two interconnected lobes and have twisted V shape. The overall of shape of the protein is reminiscent of one of the dimers in the BjPutA tetramer. Furthermore, the FoXS calculation suggests that the DvPutA experimental SAXS profile matches very well with the scattering profile of theoretical BjPutA dimer calculated from its atomic co-ordinates (Figure 4.6). Also, the R_g of the BjPutA dimer of 45 Å is very close to the SAXS estimate.

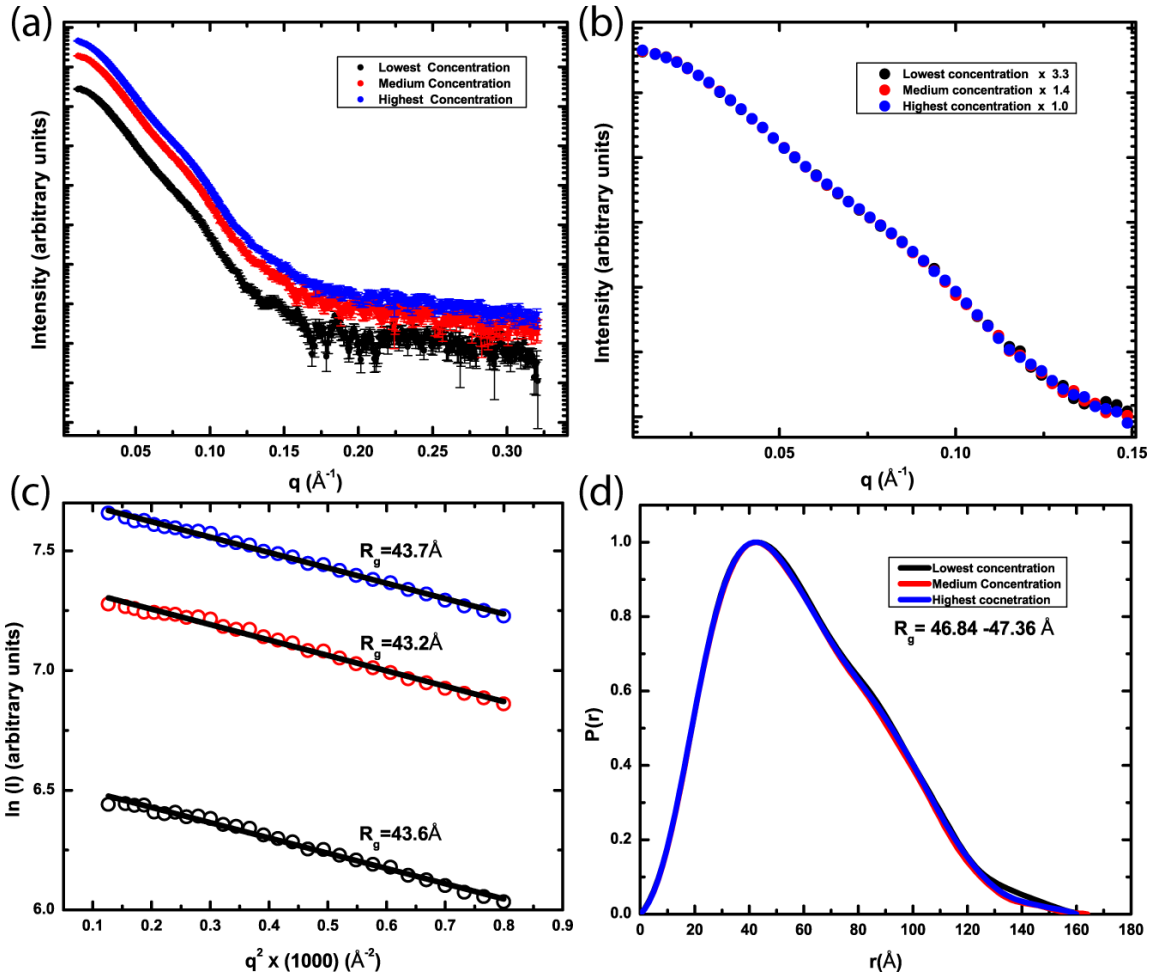


Figure 4.4. SAXS analysis of DvPutA (a) scattering profile at three different concentrations. (b) Overlaid scattering profiles at three concentrations in the range of 0 < q < 0.15Å. (c) Guinier plots and the corresponding radii of gyration for three different concentrations. (d) Normalized pair distribution functions for three different concentrations.

Table 4.4. Statistics of Guinier and Pair distribution analysis of DvPutA

Sample	Conc (mg/mL)	Guinier Analysis			P(r) Analysis			
		R_g	I(0)	I(0)/C	R_g	I(0)	I(0)/C	D_{max}
1[a]	1.24	43.6 (±0.4)	702.3 (±4.2)	566.4	47.4 (±0.2)	725 (3)	585.7	164
2[b]	2.92	43.2 (±0.4)	1591 (±7)	544.9	46.4 (±0.2)	1631 (5)	558.6	164
3[c]	4.1	43.7 (±0.4)	2314 (±11)	564.4	46.8 (±0.1)	2367 (7)	577.3	160

[a] $0.49 < qR_g < 1.2$ (points 1-27); AutoRg (90%) $R_g = 45.6 \pm 0.2$ (points 9-27) $0.76<qR_g<1.3$

[b] $0.49 < qR_g < 1.2$ (points 1-27); AutoRg (91%) $R_g = 45.93 \pm 0.05$ (points 9-28) $0.77<qR_g<1.3$

[c] $0.49 < qR_g < 1.2$ (points 1-27); AutoRg (90%) $R_g = 45.0 \pm 0.2$ (points 9-29) $0.75<qR_g<1.3$

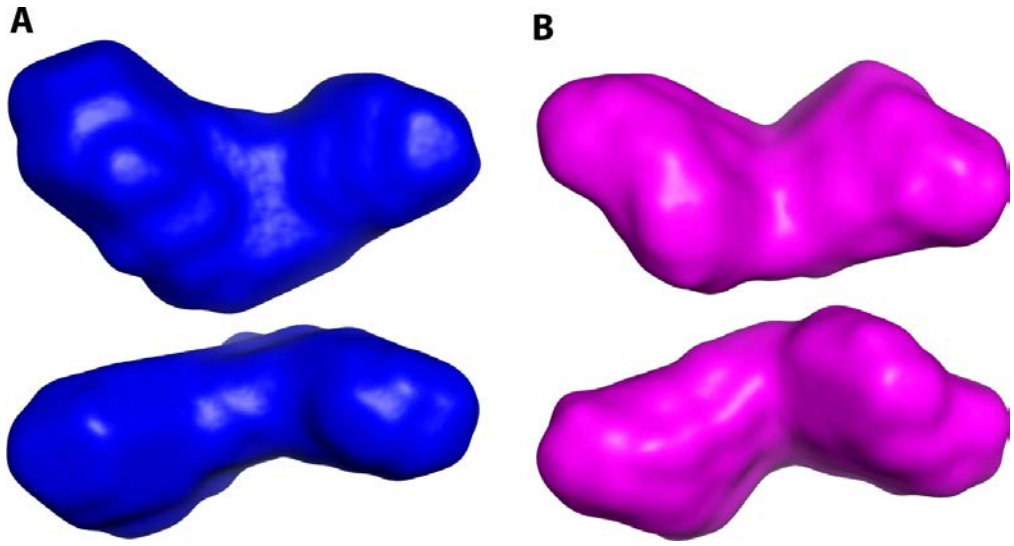


Figure 4.5. (A) Shape reconstruction for DvPutA constrained in P1 symmetry. The envelope corresponds to the average, filtered model calculated from independent models (NSD = 1.3 ± 0.0). (B) GASBOR shape reconstruction for DvPutA constrained in P2 symmetry. The filtered model is the average of 20 independent GABOR jobs with NSD = 1.3 ± 0.1).

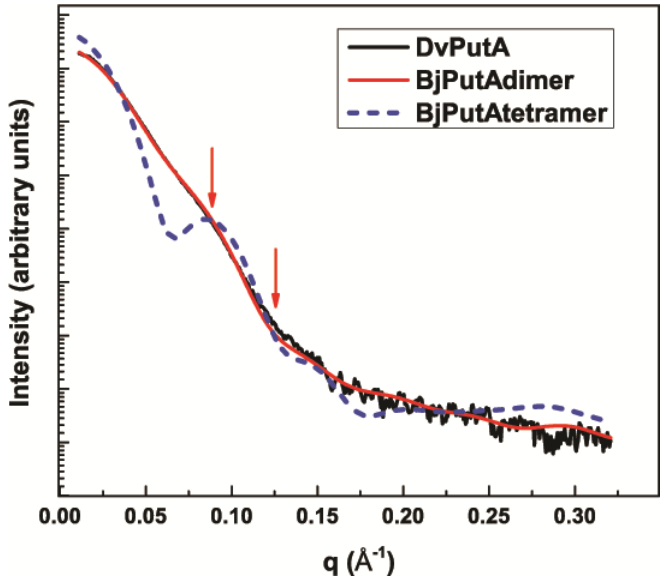


Figure 4.6 Comparison of the experimental profile of DvPutA with the theoretical profiles calculated from the BjPutA crystal structure (3HAZ) using FoXS.

4.3.2. CjPutA

The composite scattering profiles and the corresponding Guinier plots are shown in Figure 4.7a. The Guinier plots exhibit good linearity for the three different concentrations and suggests an R_g of 46.4 – 48.0Å (mention below respective Guinier plot in the figure 4.7c). The scattering curves were further analyzed by superimposing the curve on each other by multiplying with their concentration factor. Figure 4.7b shows scattering curves multiplied by their concentration factor. All the three concentration curves match very well in the q value of $0<q<0.15Å^{-1}$. The results related to the preliminary analysis of the data are summarized in Table 4.5. The AutoR_g software gives a R_g estimate of 47.1 -48.8 Å. The slight discrepancy in R_g is due to the fact that AutoR_g looks for the best linear fit in the low q region and for that it omits some of the lower q data points. Moreover, the pair distribution analysis gives an R_g value in the range of 49.7 to 50.9Å. The value of I(0)/C from Guinier and Gnom analysis does not vary significantly. All these analysis suggests that particle does not have significant concentration dependence variation. The R_g value calculated from pair distribution function is more reliable as it is not sensitive to the data points at low q region. Calculation of pair distribution function also gives a maximum particle dimension in the range of 169 - 178Å. For CjPutA, maximum concentration sample was used for the shape reconstruction. GASBOR shape reconstructions constrained to *P1* and *P2* symmetry were performed. Figure 4.8 shows volumetric map of of the average filtered *ab initio* GASBOR model. The volumetric map of shape reconstruction constrained to *P2* point group symmetry suggests that the protein does not the same twist as is obtained in DvPutA shape reconstruction. The CjPutA volumetric map seems more elongated and lacks any twist in between two lobes. The

FoXS calculation suggests that CjPutA agrees better with the curve calculated from the BjPutA dimer that that from the the tetramer (Figure 4.9). However, the theoretical BjPutA dimer profile does do not match as good as it does with DvPutA. Particularly, the difference in the scattering profile can be found at $0.087Å^{-1}$ and $0.125Å^{-1}$. The BjPutA dimer curve has a bump and depression at above mentioned q values respectively but the experimental CjPutA profiles lacks that signature. It matches poorly with biological BjPutA tetramer.

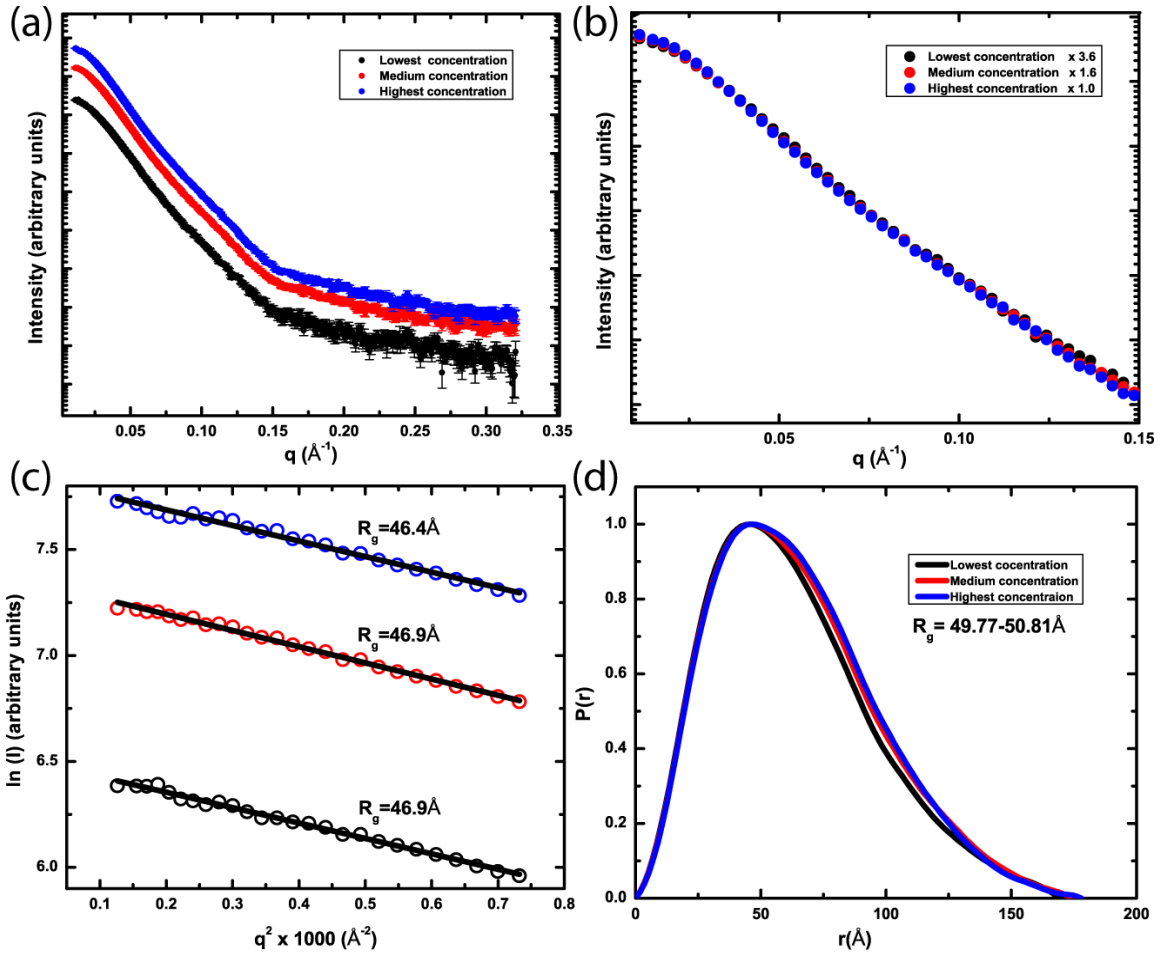


Figure 4.7: SAXS analysis of CjPutA. **(**a) Scattering profiles at three different protein concentrations. (b) Scattering profiles after scaling for concentration in the range of 0 < q< 0.15 Å. (c) Guinier plot and the corresponding radii of gyration for three different concentrations. (d) Normalized pair distribution functions for three different concentrations.

Table 4.5. Statistics of Guinier and pair distribution analysis of CjPutA

Sample	Conc. from BCA (mg/mL)	Guinier Analysis			P(r) Analysis			
		R_g	I(0)	I(0)/C	R_g	I(0)	I(0)/C	D_{max}
1[a]	1.00	46.4 (±0.4)	668 (±4)	668.4	49.7 (±0.2)	679 (2)	678.6	169
2[b]	2.25	48.0 (±0.4)	1554 (±9)	691.1	50.9 (±0.2)	1579 (5)	701.8	174
3[c]	3.6	46.9 (±0.4)	2502 (±13)	695.0	50.8 (±0.2)	2579 (10)	719.2	178

[a] $0.53<qR_g<1.3$ (points 1-26); AutoRg (87%) R_g = 47.1±0.2 (points 10-27) 0.82<qRg<1.3

[b] $0.54<qR_g<1.3$ (points 1-26); AutoRg (91%) R_g = 49.0 ±0.0 (points 7-25) 0.77<qRg<1.3

[c] $0.49<qR_g<1.2$ (points 1-26); AutoRg (89%) R_g = 48.8±1.4 (points 7-25) 0.76<qRg<1.3

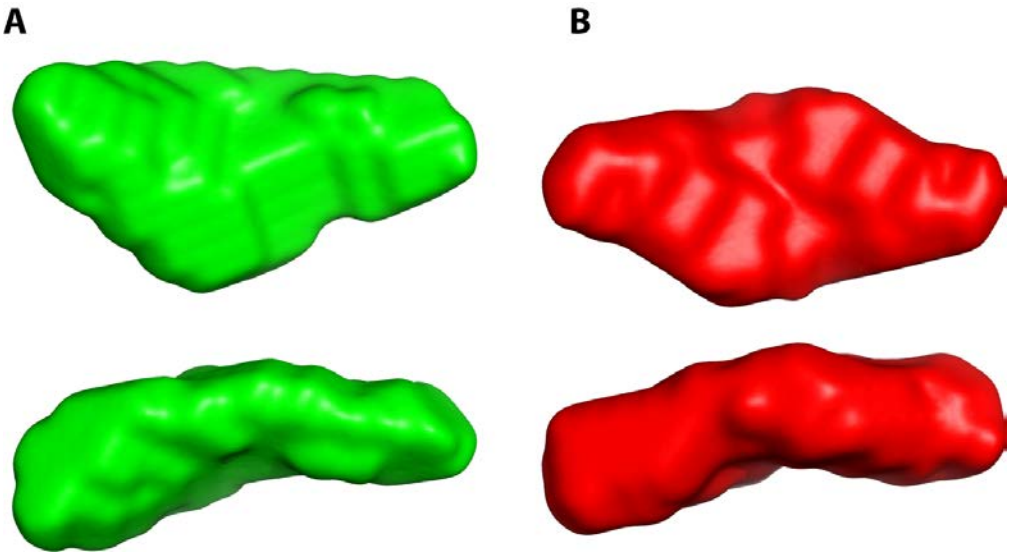


Figure 4.8. A) Damfilt volumetric map of the GASBOR shape reconstructions for CjPutA constrained in P1 symmetry. The filtered model is the average of 20 jobs with NSD=1.7 ± 0.1. B) Damfilt volumetric map of the GASBOR shape reconstruction for DvPutA constrained in P2 symmtery. The filtered model is the average of 19 independent GABOR jobs with NSD =1.7 ± 0.2.

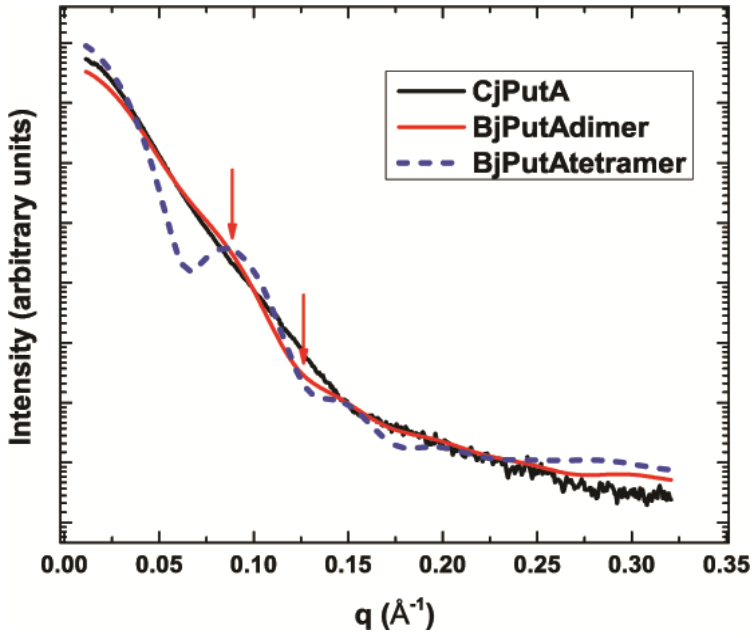


Figure 4.9. Comparison of the experimental profile of CjPutA with the theoretical profiles calculated from the BjPutA crystal structure (3HAZ) using FoXS.

4.3.3. RpPutA

The composite scattering profiles and the corresponding Guinier plots for RpPutA are shown in Figure 4.10a. The Guinier plots exhibit good linearity for the three different concentrations and suggests R_g value of 45.4-45.8Å. The scattering curves were further analyzed by superimposing the curves on each other by multiplying with their concentration factor. The figure 4.10b shows scattering curves multiplied by their concentration factor. All the three concentration curves match very well in the lower q value. The results related to the preliminary analysis of the data are summarized in Table 4.6. The AutoRg software gives a R_g estimate of 45.0 -46.2Å. Moreover, the pair distribution analysis gives an R_g value in the range of 46.12 to 46.15Å. This analysis suggests that the particle is independent of concentration dependence variation of the radius of gyration. Furthermore, the R_g estimation from Guinier and pair distribution analysis matches very well and suggesting high quality dataset. The R_g value calculated from the pair distribution function is more reliable as it is not sensitive to the data points at low q region. Calculation of pair distribution function also gives a maximum particle dimension in the range of 155 - 158Å. For RpPutA, the medium concentration sample was used for the shape reconstruction. Figure 4.11 shows volumetric map of the damfilt model of the GASBOR reconstruction assuming *P1 and P2* symmetry. The volumetric map obtained from GASBOR constrained in *P1 and P2* shape reconstruction suggests that similar to DvPutA, the particle has two interconnected lobes and have twisted V shape. In addition, similar to DvPutA, the FoXS calculation suggests RpPutA matches very well with the theoretical scattering curve consistent with the BjPutA dimer (Figure 4.12). Similar to DvPutA and CjPutA it also matches poorly with biological BjPutA

tetramer. This was unexpected as the sequence identity of RpPutA with BjPutA is around 70%. Therefore, a similar oligomeric state was expected.

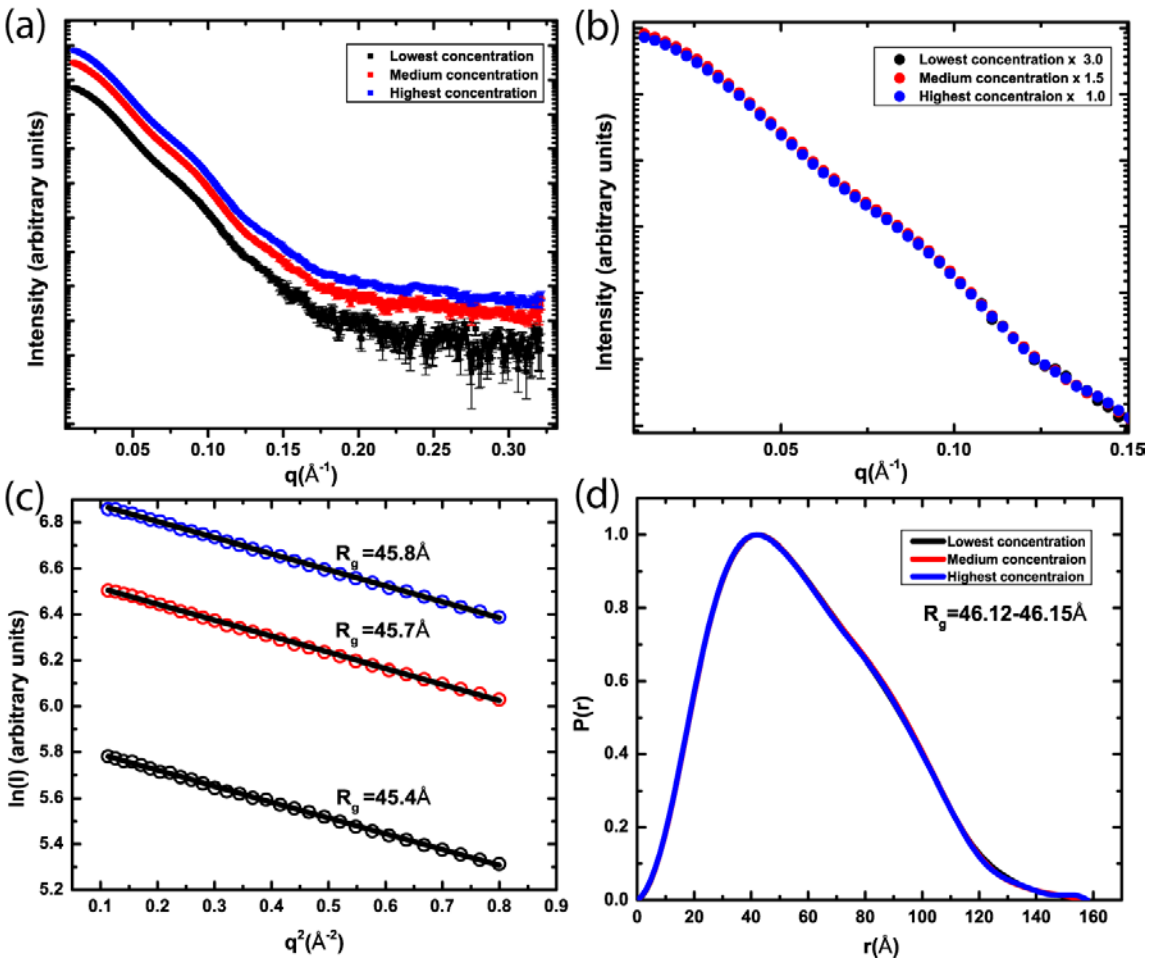


Figure 4.10. SAXS analysis of RpPutA **(**a) Scattering profiles at three protein concentrations. (b) Overlaid scaled scattering profiles at three concentration in the range of 0<q<0.15Å. (c) Guinier plots and the corresponding R_g for different concentrations. (d) Normalized pair distribution functions for three different concentrations.

Table 4.6. Statistics of Guinier and pair distribution analysis of RpPutA

Sample	Cocn (mg/mL) BCA	Gunier Analysis			P(r) Analysis			
		Rg	I(0)	I(0)/C	Rg	I(0)	I(0)/C	D_{max}
1[a]	3X	45.4 (±0.1)	350.1 (±0.8)	1050X	46.15 (±0.05)	351.0 (±0.4)	1053	155
2[b]	1.5X	45.70 (±0.1)	722.5 (±1.1)	1083X	46.13 (±0.04)	721.6 (±0.6)	1081	155
3[c]	X	45.80 (±0.1)	1036.4 (±1.5)	1036.4X	46.12 (±0.04)	1034.0 (±0.8)	1034	158

[a]0.483<qRg<1.28 (1-30 point) AutoRg (94%) Rg 45.4(±0.0) (6-28points) 0.62<qRg<1.2

[b]0.486<qRg<1.29(1-30 point) AutoRg (87%) Rg 44.9 (±0.1) (13-51points) 0.81<qRg<1.3

[c]0.487<qRg<1.29(1-30 point) AutoRg (97%) Rg 46.2(±0.0) (2-23points) 0.52<qRg<1.1

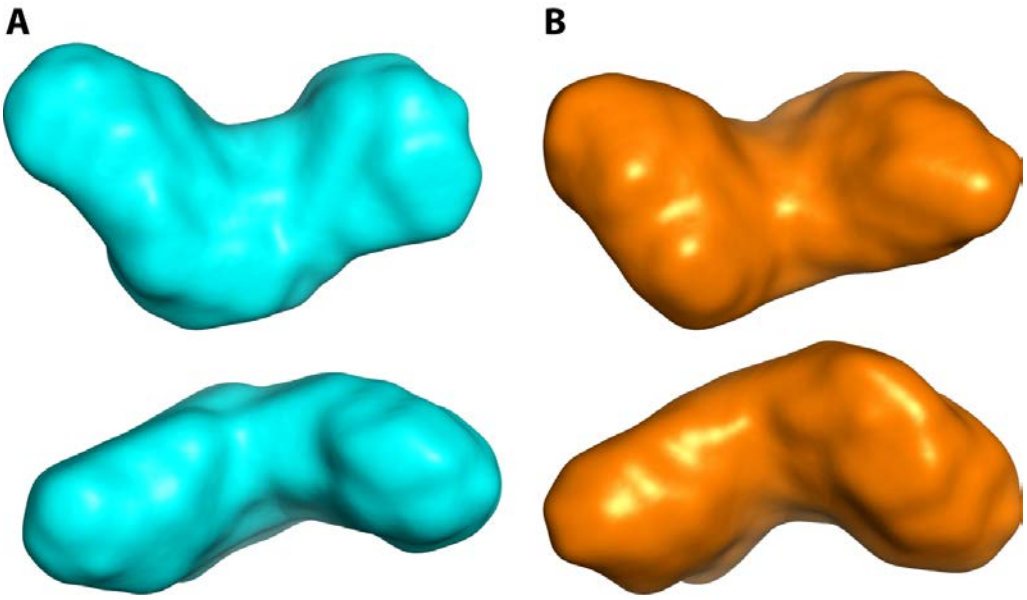


Figure 4.11. (A) Damfilt volumetric map of the GASBOR shape reconstructions for RpPutA constrained in P1 symmetry. The filtered model is the average of 18 jobs with NSD=1.3 ± 0.1. (B) Damfilt volumetric map of the GASBOR shape reconstruction for DvPutA constrained in P2 symmtery. The filtered model is the average of 20 independent GABOR jobs with NSD =1.4 ± 0.1.

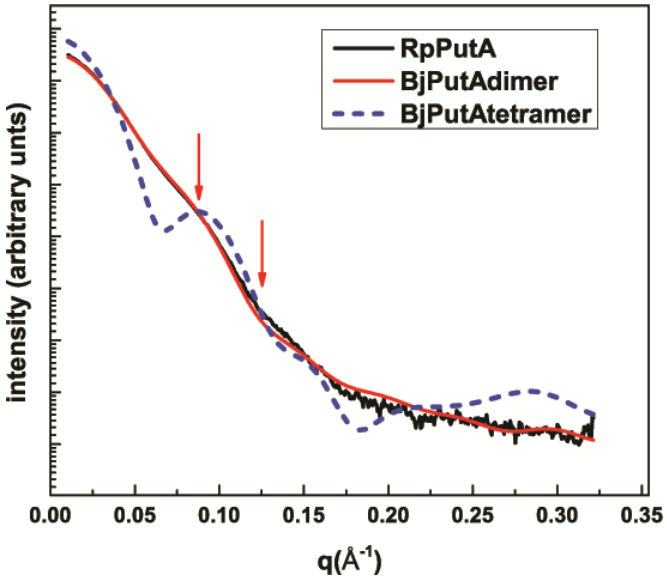


Figure 4.12. Comparison of the experimental profile of RpPutA with the theoretical profiles calculated from the BjPutA crystal structure (3HAZ) using FoXS.

4.4. Summary and Conclusion

Oligomeric states of three bifunctional PutAs have been determined using SAXS. The short bifunctional branch 2 DvPutA and long bifunctional branch 2 CjPutA was found to be dimers. The results obtained clearly suggest presence of dimeric PutA in both branches. Our lab has also determined that RcPutA which belongs to the long branch 1 bifunctional PutA is a monomer. The difference in oligomeric states of long branch 1 and branh 2 PutA is intetesting as we. note that, the long bifunctional branch 1 PutAs has a clear conserve C-terminal motif (CCM). In addition the absence of β-hairpin essential for dimerization is also obvious. These two features are not obvious in long bifunctional branch 2 PutAs. Though there is no structural or biochemical explanation to understand the basis of oligomerization in long branch 1 and branch 2 bifunctioanl. Earlier work on BjPutA showed that it's a dimer of dimers tetramer. Surprisingly,the 70% identical RpPutA is a dimer. Though sequence alignment does not show any remarkable difference that could be responsible for higher order oligomerization of BjPutA. Study of more branch 1 and branch 2 PutA is needed to understand the diversity of oligomeric states in PutAs.

4.5. References.

1. Tanner JJ. Structural biology of proline catabolism. Amino Acids. 2008;35(4):719-30.

2. Zhou Y, Zhu W, Bellur PS, Rewinkel D, Becker DF. Direct linking of metabolism and gene expression in the proline utilization A protein from Escherichia coli. Amino Acids. 2008;35(4):711-8. Epub 2008/03/08.

3. Ostrovsky de Spicer P, Maloy S. PutA protein, a membrane-associated flavin dehydrogenase, acts as a redox-dependent transcriptional regulator. Proc Natl Acad Sci USA. 1993;90(9):4295-8.

4. Lee YH, Nadaraia S, Gu D, Becker DF, Tanner JJ. Structure of the proline dehydrogenase domain of the multifunctional PutA flavoprotein. Nat Struct Biol. 2003;10(2):109-14.

5. Srivastava D, Schuermann JP, White TA, Krishnan N, Sanyal N, Hura GL, et al. Crystal structure of the bifunctional proline utilization A flavoenzyme from Bradyrhizobium japonicum. Proc Natl Acad Sci U S A. 2010;107(7):2878-83. Epub 2010/02/06.

6. Krishnan N, Becker DF. Characterization of a bifunctional PutA homologue from Bradyrhizobium japonicum and identification of an active site residue that modulates proline reduction of the flavin adenine dinucleotide cofactor. Biochemistry. 2005;44(25):9130-9. Epub 2005/06/22.

7. Zhang W, Krishnan N, Becker DF. Kinetic and thermodynamic analysis of Bradyrhizobium japonicum PutA-membrane associations. Archives of biochemistry and biophysics. 2006;445(1):174-83. Epub 2005/11/29.

8. Brown ED, Wood JM. Redesigned purification yields a fully functional PutA protein dimer from Escherichia coli. J Biol Chem. 1992;267(18):13086-92.

9. Wood JM. Membrane association of proline dehydrogenase in Escherichia coli is redox dependent. Proc Natl Acad Sci USA. 1987;84(2):373-7. Epub 1987/01/01.

10. Brown ED, Wood JM. Conformational change and membrane association of the PutA protein are coincident with reduction of its FAD cofactor by proline. J Biol Chem. 1993;268(12):8972-9.

11. Surber MW, Maloy S. The PutA protein of Salmonella typhimurium catalyzes the two steps of proline degradation via a leaky channel. Arch Biochem Biophys. 1998;354(2):281-7.

12. Maloy S, Roth JR. Regulation of Proline Utilization in *Salmonella typhimurium*: Characterization of *put*:Mu *L*(Ap, *lac*) Operon Fusions. J Bacteriol. 1983;154:561-8.

13. Muro-Pastor AM, Maloy S. Proline Dehydrogenase Activity of the Transcriptional Repressor PutA is Required for Induction of the *put* Operon by Proline. J Biol Chem. 1995;270(17):9819-27.

14. Zhu W, Gincherman Y, Docherty P, Spilling CD, Becker DF. Effects of proline analog binding on the spectroscopic and redox properties of PutA. Arch Biochem Biophys. 2002;408(1):131-6.

15. Vinod MP, Bellur P, Becker DF. Electrochemical and functional characterization of the proline dehydrogenase domain of the PutA flavoprotein from *Escherichia coli*. Biochemistry. 2002;41:6525-32.

16. Zhang W, Zhou Y, Becker DF. Regulation of PutA-membrane associations by flavin adenine dinucleotide reduction. Biochemistry. 2004;43(41):13165-74.

17. Singh RK, Tanner JJ. Unique structural features and sequence motifs of proline utilization A (PutA). Frontiers in bioscience : a journal and virtual library. 2012;17:556-68. Epub 2011/12/29.

18. Singh RK, Larson JD, Zhu W, Rambo RP, Hura GL, Becker DF, et al. Small-angle X-ray Scattering Studies of the Oligomeric State and Quaternary Structure of the Trifunctional Proline Utilization A (PutA) Flavoprotein from Escherichia coli. Journal of Biological Chemistry. 2011;286(50):43144-53.

19. Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology. 2000;302(1):205-17.

20. Gouet P, Courcelle E, Stuart DI, M√©toz F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics. 1999;15(4):305-8.

21. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr. 2003;36(5):1277-82.

22. Svergun D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr. 1992;25(4):495-503.

23. Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. Biophysical journal. 2001;80(6):2946-53. Epub 2001/05/24.

24. Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res. 2010;38(Web Server issue):W540-4. Epub 2010/05/29.

CHAPTER 5

Investigation of two oligomeric states of pyrroline -5-carboxylate dehydrogenase by small angle X-ray scattering, X-ray crystallography and light scattering.

5.1. Introduction

5.1.1. Protein oligomerization

Proteins have evolved as oligomers in all kingdoms of living organisms. The oligomerization of proteins is related to evolution. Proteins exist as oligomers to satisfy the need for their diverse cellular functions. These cellular functions include allosteric regulation, active site formation at the interface, DNA binding, gene expression, stability etc. To fulfill its cellular functions, proteins exist as monomers, dimers, tetramers or even higher order oligomers. Protein oligomerization has also been rationalized on the basis of economy scale of protein synthesis and small genome size. In addition, the burial of hydrophobic interfaces to avoid the solubility problems has been mentioned as another possible justification for the oligomerization. It has also been suggested that higher order oligomers have more extensive internal interactions, which would enhance stability relative to lower order oligomers (1, 2).

5.1.2. Characteristics of protein-protein interface

Understanding protein – protein interfaces has been investigated for more than three decades (3-8). Various experimental and computational studies have been reported (9-13) to understand the characteristic features of residues participating in interface formation and as well as their secondary structures. Interfaces may differ from the rest of the protein in terms of amino acid composition, number of hydrophobic residues, amount of hydration, and number of electrostatic interactions between the charged residues. These features have been investigated comprehensively using experimental and computational methods. At the protein- protein interface, one finds a combination of hydrogen bonds,

salt bridges, and non-polar interactions. Some of these interactions could be more important to maintain the stability of the interface than the others. Andrew *et al.* suggested that the energetic contribution of the individual residues across each interface is not even. His work on heterodimer interfaces showed that there are certain residues across the interface that are responsible for most of the binding energy (14). These regions were termed “hot spots” of the protein interface (15). The hot spots are enriched in residues like tryptophan, tyrosine, and arginine and are surrounded by energetically less important residues. Andrew *et al* also observed that the residues that make up the hot spots tend to cluster at the center of the interface rather than being at the periphery of the interface.

5.1.3. Protein engineering at the interface

Site directed mutagenesis has been used to alter protein quaternary structure. These alterations were meant to increase the thermostability (16, 17), creating medically important (18) oligomers or just to understand the functional properties of the proteins (Table 5.1). Some examples of single or double mutations changing the quaternary structure are described next.

5.1.3.a. Tyrosyl-tRNA synthetase

Tyrosyl-tRNA synthetase (TyrTS) is a homodimeric protein. One of the short construct of ΔTyrTS was also found to be homodimeric. A pair of homodimeric mutants from TyrTS and ΔTyrTS was created. Phe164 of the two protomers interacts across the dyad axis in the symmetrical dimer in both TyrTS and ΔTyrTS. In TyrTS, Phe164 was

mutated to Asp. On the other hand, in ΔTyrTS, the interface residue Phe164 was mutated to Lys. These mutations resulted in the pH-dependent monomeric forms of the enzyme enzymes on the basis of identical charge repulsion at the interface. They observed that at pH above 6.0 TyrTSF164D is predominantly monomeric as the D164 ionizes (19). In a similar manner, ΔTyrTSF164K becomes predominantly monomeric by lowering the pH, as, K164 becomes protonated. The authors generated the heterodimers by mixing equivalent amount the TyrTSF164D and ΔTyrTSF164K at pH 7.8. The pH dependent heterodimers were generated on the basis of attraction of positive charged Lys and negative charged Asp residue at the interface (3, 20).

5.1.3.b. Insulin

Human insulin largely exists as a trimer-of-dimers homohexamer. The medically important monomeric variants of insulin were created using single and double mutations at the dimer interface. The monomeric variants of the protein successfully retained the *in vivo* potency of the hormone (4, 18). Interestingly, they successfully adopted the idea of identical charge repulsion at the interface from Jones *et al.*(19). For example, one of the mutants B12Val to B12Glu was essentially monomeric at the concentration of 10^{-3}M. One double mutant B9ser+B27Thr to B9Asp +B27Glu was also found to be monomeric at the similar concentrations. The oligomeric state of the monomeric variants was investigated by osmotic pressure measurement and circular dichroism (CD).

5.1.3.c. *Serratia* endonuclease

In this paper the authors observed that most endonucleases are dimers and a few are monomers. One example of a monomeric endonuclease is *Anabaena* nuclease. The authors performed site-directed mutagenesis in the interface of the dimeric endonuclease from *Serratia* to create a monomeric variant. The mutations were designed on the crystal structure of the *Serratia* endonuclease dimer (21). They chose a histidine residue at the interface as a candidate for mutation. This residue interacts with a loop in another subunit containing residues Asn, Ala, Val and Pro. The idea was to disrupt the electrostatic or steric complementarity by mutating the histidine. H184A, H184N, H184T and H184R were created. These variants were observed to be monomers in analytical ultracentrifugation and gel filtration experiments (22, 23). In addition they also observed that the stability towards the chemical denaturants and the activity is same as the wild type enzyme.

5.1.4. Pyrroline-5-carboxylate dehydrogenase

Δ^1-Pyrroline-5-carboxylate dehydrogenase (P5CDH, EC 1.5.1.12) is involved in the final step of proline catabolism (24). It catalyzes the NAD^+-dependent oxidation of L-glutamate-semialdehyde (GSA) to L-glutamate and exists in all kingdoms of living organisms. P5CDHs exist either as monofunctional enzymes or as a component of a multifunctional enzyme (25).

Recently, our lab reported the crystal structure and analytical ultracentrifugation analysis of human P5CDH (26). The protein forms a domain-swapped dimer in which the oligomerization domain of one protomer interacts with the catalytic domain of the other

promoter. This kind of dimerization is common to members of the aldehyde dehydrogenase superfamily.

Besides human P5CDH, crystal structures of three more P5CDHs have been deposited in the PDB by other labs (Table 5.2). The crystal packing of these enzymes also suggests a domain swapped dimer. In addition the crystal structure of TtP5CDH suggests one more distinct interface where three domain swapped dimer packs around a 3-fold axis to make a trimer-of-dimers hexamer (27). Whether the hexamer is formed in solution is unknown.

Here we report small angle X-ray scattering (SAXS) and multi angle light scattering (MALS) studies aimed at determining the oligomeric states of P5CDHs in solution (Table 5.2). In addition, the structural determinants of P5CDH hexamer formation is investigated using site directed mutagenesis of TtP5CDH.

Table 5.1. Summary of published quaternary structural engineering studies. This table is taken from chapter 5 of "Protein Dimerization and Oligomerization in Biology"(2).

Protein	Natural oligomeric form	Engineered form
Bacillus stearothermophilus tyrosyl t-RNA synthetase	Homodimer	Monomer (3, 20, 28)
Human insulin	Homohexamer (trimer of dimers)	Monomer (4, 29)
Human interleukin-5	Homodimer (domain swapped)	Monomer (30, 31)
Serratia marcessens endonuclease	Homodimer	Monomer (22, 23)
Streptavidin/avidin	Homotetramer	Dimer and monomer (32-34)
E.coli inorganic pyrophosphatase	Homohexamer (dimer of trimers)	Trimer, dimer and monomer (35, 36)
EcoRI endonuclease	Homodimers	Monomer (37)
Bse634I endonuclease	Homotetramer	Dimer (38)
λ Cro	Homodimer	Monomer (39)
Trypansomal triosephosphatase isomerase	Homodimer	Monomer (40, 41)
Human triosephosphatase isomerase	Homodimer	Monomer (42)
Human superoxide dismutase	Homodimer	Monomer (43, 44)
Rat prostatic acid phosphatase	Homodimer	Monomer (45)
E. coli Malate dehydrogeanse	Homodimer	Monomer (46)
Rabbit fructose-1-6-bisphosphate aldolase	Homotetramer	Dimer and monomer (6, 7)
Bacillus stearothermophilus Lactate dehydrogenase	Homotetramer	Dimer (47)
E. coli Succinyl CoA Synthetase	Heterotetramer (αβ-dimer)	Dimer (αβ) (48)

Table 5.2. P5CDHs under investigation

Organism	Abbreviation	PDB Code	amino acid residues
Thermus thermophilus	TtP5CDH	2BHP	516
Bacillus halodurans	BhP5CDH	3RJL	515
Bacillus licheniformis	BlP5CDH	3QAN	515
Bacillus subtilis	BsP5CDH	ND*	515
Deinococcus radiodurans	DrP5CDH	ND*	523

*Crystal structure has not been determined.

Table 5.3. Pairwise sequence identity of the P5CDHs under investigation

	Tt	Dr	Bl	Bs	Bh
Tt	100	52	50	50	49
Dr		100	51	51	49
Bl			100	90	74
Bs				100	75
Bh					100

5.2. Materials and Methods

Four P5CDHs have been expressed and purified: *Bacillus halodurans* P5CDH (BhP5CDH), *Thermus thermophilus* P5CDH (TtP5CDH), *Bacillus subtilis* P5CDH (BsP5CDH), *Deinococcus radiodurans* P5CDH (DrP5CDH). Also, a site-directed mutant of TtP5CDH (TtP5CDHR100A) has been studied. The methods for DrPRODH and BsP5CDH which were performed by Min Luo, will not be discussed here.

5.2.1. Protein expression and purification

TtP5CDH was expressed from vector pKA8H with an N-terminal His_8-tag and an intervening Tobacco Etch Virus Protease enzyme (TEVP) cleavage site. Treatment with TEVP results in the native polypeptide preceded by Gly-His. The New York Structural Genomic Research Consortium generously provided the clone of BhP5CDH in pET30CHS. The expressed enzyme has a C-terminal His_8-tag and an intervening TEVP cleavage site. The sequence remains after TEVP cleavage is AENLYFQ at the C-terminal end.

All the proteins mentioned above were purified as follows. The protein was expressed using an *E.coli* strain BL21(DE3)pLysS. A 10 mL overnight culture for each 1000 mL LB media was used to inoculate. The cells were grown at 37^{o}C at 250 rpm and induced with 0.5 mM IPTG at OD_{600} = 0.6 with an induction temperature of 22 ^{o}C at 200 rpm for 18 hours. Cells were harvested and frozen in 50 mM Hepes pH 8.0, 300 mM NaCl, 10 mM Imidazole and 5% glycerol. The frozen cells were thawed before purification, with the addition of protease inhibitors (AEBSF, TPCK, E64, Pepstatin, Leupeptin) and cells were broken by

sonication. Cell debris and unbroken cells were separated by centrifugation at 16,500 rpm for one hour in a SS34rotor. The supernatant was applied to a His-Trap HP column (His-Trap Ni^{2+}-Sepharose HP, GE Healthcare) equilibrated with 20 mM Hepes, 300 mM NaCl, and 5% glycerol at pH 8.0. The protein of interest was eluted with equilibration buffer congaing 300 mM imidazole. The fractions were pooled, and TEVP, 1M THP, and 20X TEV buffer (1 M Tris-HCl, 10 mM EDTA, pH 8.0) were added so that the final solution contained 5 mg of TEVP per 30 mg of protein in 50 mM Tris–HCl, 0.5 mM THP, and 0.5 mM EDTA, pH 8.0. The sample was incubated at 30^{o}C for 2 hrs, and dialyzed overnight at 4 ^{o}C and injected onto the His-Trap HP column. The untagged protein was collected at 30 mM imidazole and dialyzed in 50 mM Tris-HCl, 0.5 mM THP, 5% glycerol, 0.5 mM EDTA, pH 8.0 in preparation for further purification using anion exchange chromatography (HiTrap Q). The protein was bound to the HiTrap Q anion exchange column equilibrated with a buffer similar to the dialysis buffer and was eluted with a linear 0-1 M NaCl gradient. Size exclusion chromatography (Superdex 200, 25 mL) was used as the final step of purification.

5.2.2 TtP5CDHR100A mutation

The quick change mutagenesis kit (Stratagene) was used to create the TtP5CDH mutation R100A. The following primers were used

Forward 5' GGAGGACCGGAGCGCCCTCCTCCTCAAGGC 3'

Reverse 5' GCCTTGAGGAGGAGGGCGCTCCGGTCCTCC 3'

The mutations was confirmed with sequencing performed by the University of Missouri DNA core. TtP5CDHR100A mutant was purified as described above.

5.2.3. Small Angle X-ray Scattering (SAXS)

SAXS experiments were performed at beamline 12.3.1 of the Advanced Light Source via the mail-in program (49). Prior to analysis, all protein samples were subjected to size exclusion chromatography (SEC) using a Superdex 200 column. Scattering intensities (I) were measured at three nominal protein concentrations using exposure times of 0.5, 1.0, 3.0, and 6.0 sec. Scattering curves collected from the protein samples were corrected for background intensity using intensity data collected from dialysis buffer (for BhP5CDH and TtP5CDHR100A) or SEC effluent (for TtP5CDH).

The SAXS data are analyzed as follows. A composite scattering curve for each sample was generated with PRIMUS (50) by scaling and merging high q region from one of the longer time exposures with the lower q region from a lower time exposure. The scattering curves were multiplied with a concentration factor and overlaid on each other to check concentration dependent variation of the profile. A plot of $\ln(I)$ versus q^2 was plotted at the lower q region ($qR_g<1.3$). This region is called the Guinier region, and the Guinier plot for a well behaved, non-aggregated protein that is free from interparticle interference should be linear in this region. The Guinier regions used for various samples are listed as footnotes to Tables 5.1, 5.2 and 5.3. The AutoR_g utility of PRIMUS was also used to validate the Guinier region obtained manually. AutoR_g finds the q region that yields the best linier fit and outputs a quality index of the fit expressed as a percentage. One hundred percent is the best score; factors that affect the score include the number of points at low q that are omitted and the residual from linear regression analysis. GNOM was used to calculate the pair distribution function (P(r)) in order to estimate the radius of gyration (R_g) and the maximum particle dimension (D_{max}) (51) . MOLEMAN was used to

calculate the R_g from atomic coordinates.

Shape reconstruction calculations were performed using the *ab initio* based program GASBOR (52). The D_{max} values used for the shape reconstructions are listed in Tables 5.4, 5.5 and 5.6. Each reconstruction consisted of 18-20 independent GASBOR calculations, and the models were averaged and filtered using damaver. The point group symmetry used during the GASBOR calculations was *P*1, *P*2 *or P*32. The averaged, filtered dummy atom model from damaver (53) (damfilt.pdb) was converted into a volumetric map using pdb2vol. The TtP5CDH hexamer and BhP5CDH dimer were created using the crystal symmetry. The dimer and the hexamer P5CDH models were superimposed on the respective GASBOR model using supcomb. The web-server FoXS was used to calculate SAXS curves from atomic coordinates (54)**.**

5.2.4. Multi Angle Light Scattering

The molecular mass of BsP5CDH, BhP5CDH, TtP5CDH and DrP5CDH were determined from multi-angle light scattering (MALS) detector coupled to G5000PWXLSEC column (Tosoh Bioscience, Montgomeryville, PA) connected to a HPLC system. The MALS analysis was performed in line to SEC separation using multi-angle light scattering and quasi-angle light scattering detectors (Wyatt Technology Corp., Santa Barbara, CA). The molar mass, hydrodynamic radius and the polydispersity of the sample were estimated using ASTRA software as described previously (55). The column buffer was 50 mM Tris-HCl, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM THP, and 5% glycerol, pH 8.0. The flow rate was 0.75 ml/min.

5.2.5. Crystallization and data collection of TtP5CDHR100A

Crystals of TtP5CDH R100A mutant were grown similar to the protocol for TtP5CDH described earlier (56). X-ray diffraction data were collected at beamline 4.2.2 of Advanced Light Source. The data set consisted of 360 frames with an oscillation width of 0.5^{o} per image, detector distance 110 mm, exposure time 2s/image. The data were processed using XDS(57) and SCALA (58)via CCP4i(59). The data collection and refinement statistics is shown in Table 5.11.

5.3. Results

5.3.1. Small Angle X-ray Scattering

The composite scattering profiles and the corresponding Guinier plots of BhP5CDH are shown in Figure 5.1. The Guinier plots exhibit good linearity for the three different concentration and suggests an R_g value in the range of 29.8-30.3Å (Table 5.4 and Figure 5.1c). Figure 5.1b shows scattering curves multiplied by the dilution factor of the sample. The q range shown in this figure is $0 < q < 0.15$ $Å^{-1}$. The figure clearly shows that the scaled curves are essentially identical, which indicates that I(0)/C does not vary significantly with concentration. These results confirm that the particle is free from concentration dependent behavior. The AutoR_g utility in PRIMUS (50) gives an R_g estimate of 31.4 -31.8 Å. Moreover, the pair distribution analysis gives an R_g value in the range of 31.5 to 32.0 Å. The difference in R_g calculated from the two methods is less than 1%, which shows that R_g is well defined by the data. I note that the R_g range (31.5-32.0 Å) calculated from the pair distribution function is more reliable as it is not as sensitive to the data points at low q region. Calculation of pair distribution function also gives a

maximum particle dimension in the range of 105-110Å.

For BhP5CDH, the medium concentration sample was used for the shape reconstruction. Figure 5.2b shows the volumetric representation of the damfilt model from the GASBOR reconstruction assuming *P*2 symmetry. The normalized spatial discrepancy (NSD) for the set of 19 GASBOR models was 0.97 ± 0.05, which indicates a trustworthy shape reconstruction. The volumetric map obtained from GASBOR constrained in *P*2 is consistent with the domain-swapped dimeric protein predicted from PDBePISA (Figure 5.2b). Furthermore, the FoXS calculation suggests that the BhP5CDH experimental SAXS profile matches very well with the theoretical scattering profile calculated from the domain-swapped dimer atomic coordinates (Figure 5.2a). Also, the R_g of the BhP5CDH dimer of 30.7 Å is very close to the SAXS estimate (31 – 32 Å). These results show that the dimer predicted by PDBePISA is also formed in solution.

The composite scattering profiles and the corresponding Guinier plots of TtP5CDH are shown in Figure 5.3. The Guinier plots exhibit good linearity for the three different concentrations and suggest an R_g of 42.1 - 43.0 Å (Table 5.5 and Figure 5.3c). Figure 5.3b shows scattering curves scaled to the dilution factor of the sample. The q range shown in the figure is $0 < q < 0.15$ $Å^{-1}$. The figure clearly suggests all the three concentration-scaled curves are overlaid on top of each other. In addition I(0)/C does not vary significantly with concentration. These results confirm that the particle is free from concentration dependent behavior. The AutoR_g software gives an R_g estimate of 44.2 -45.4 Å. The slight discrepancy in R_g is due to the fact that AutoR_g looks for the best linear fit in the low q region and for that it omits some of the lower q data points.

Moreover, the pair distribution analysis gives an R_g value in the range of 43.2 to 44.0Å. The difference in R_g calculated from the two methods is less than 1%, which shows that the determination of R_g from the SAXS data is robust. Though, the R_g value calculated from pair distribution function is more reliable as it is not sensitive to the data points at low q region. Calculation of pair distribution function also gives a maximum particle dimension in the range of 130-135Å.

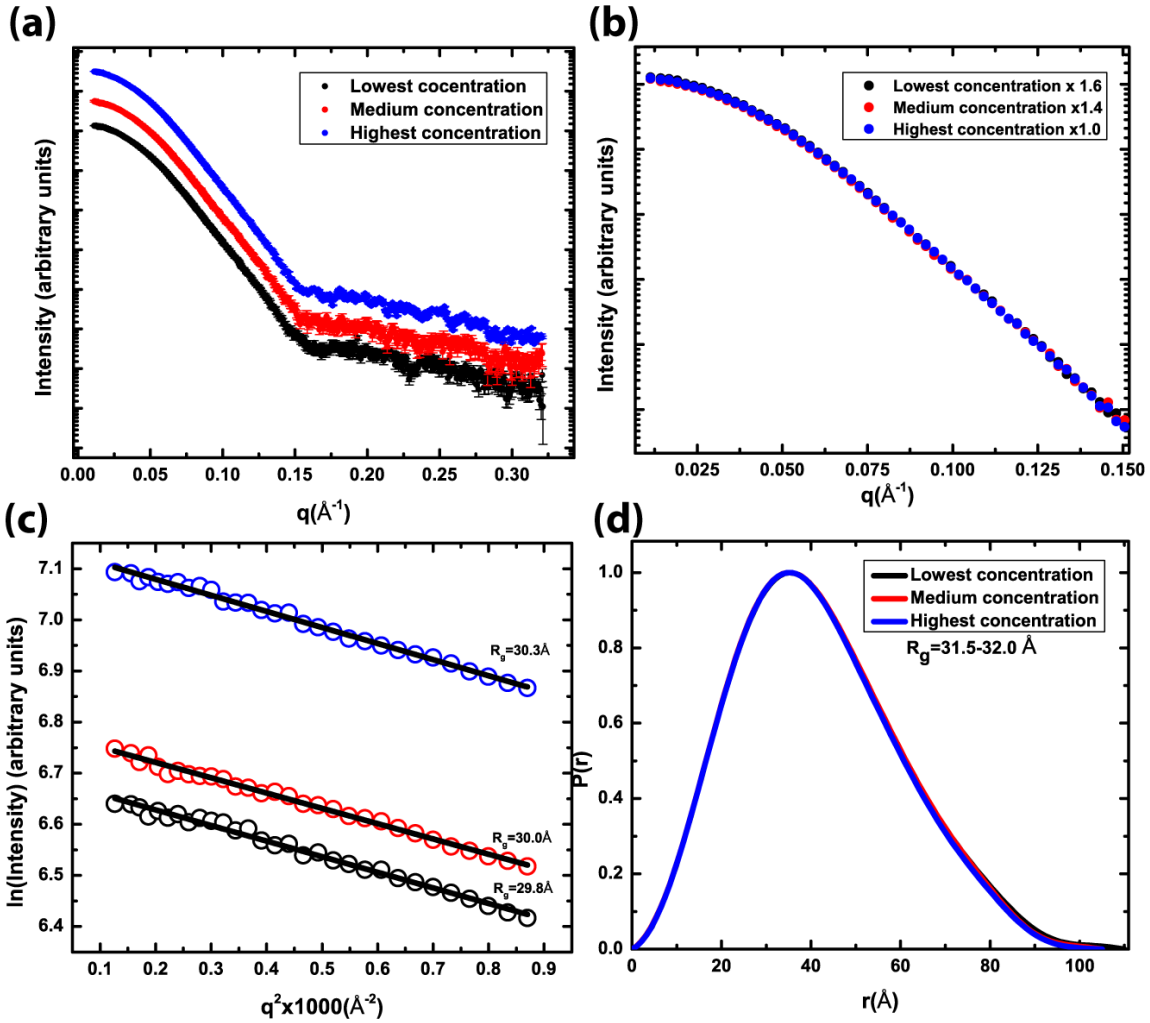


Figure 5.1. SAXS of BhP5CDH (a) Scattering profiles at three different concentrations. Data are offset for clear visibility (b) Scattering profiles after scaling for concentration in the range of 0< q< 0.15Å. (c) Guinier plots and the corresponding radii of gyration for three different concentrations. (d) Normalized pair distribution functions for three different concentrations.

Table 5.4. Statistics of Guinier and pair distribution analysis of BhP5CDH

Sample	Conc (mg/mL)	Guinier Analysis			P (r) Analysis			
		R_g(Å)	I(0)	I(0)/C	R_g(Å)	I(0)	I(0)/C	D_{max} (Å)
1[a]	2.7	29.8 (0.2)	799 (1)	296	31.5 (0.0)	812 (10)	300	110
2[b]	3.2	30.0 (0.3)	881 (3)	276	32.0 (0.1)	902 (1)	281	105
3[c]	4.4	30.3 (0.2)	1258 (3)	285	31.8 (0.0)	1276 (1)	290	105

[a]$0.33<qR_g<0.88$ (points 1-30) AutoR_g (94%) R_g =31.8 ± 0.1 (points 9-47) $0.53<qR_g<1.3$

[b]$0.34<qR_g<0.88$ (points 1-30) AutoR_g (87%) R_g =31.7 ± 0.3 (points 18-49) $0.70<qR_g<1.3$

[c]$0.34<qR_g<0.89$ (points 1-30) AutoR_g (93%) R_g =31.4 ± 0.0 (points 12-49) $0.58<qR_g<1.3$

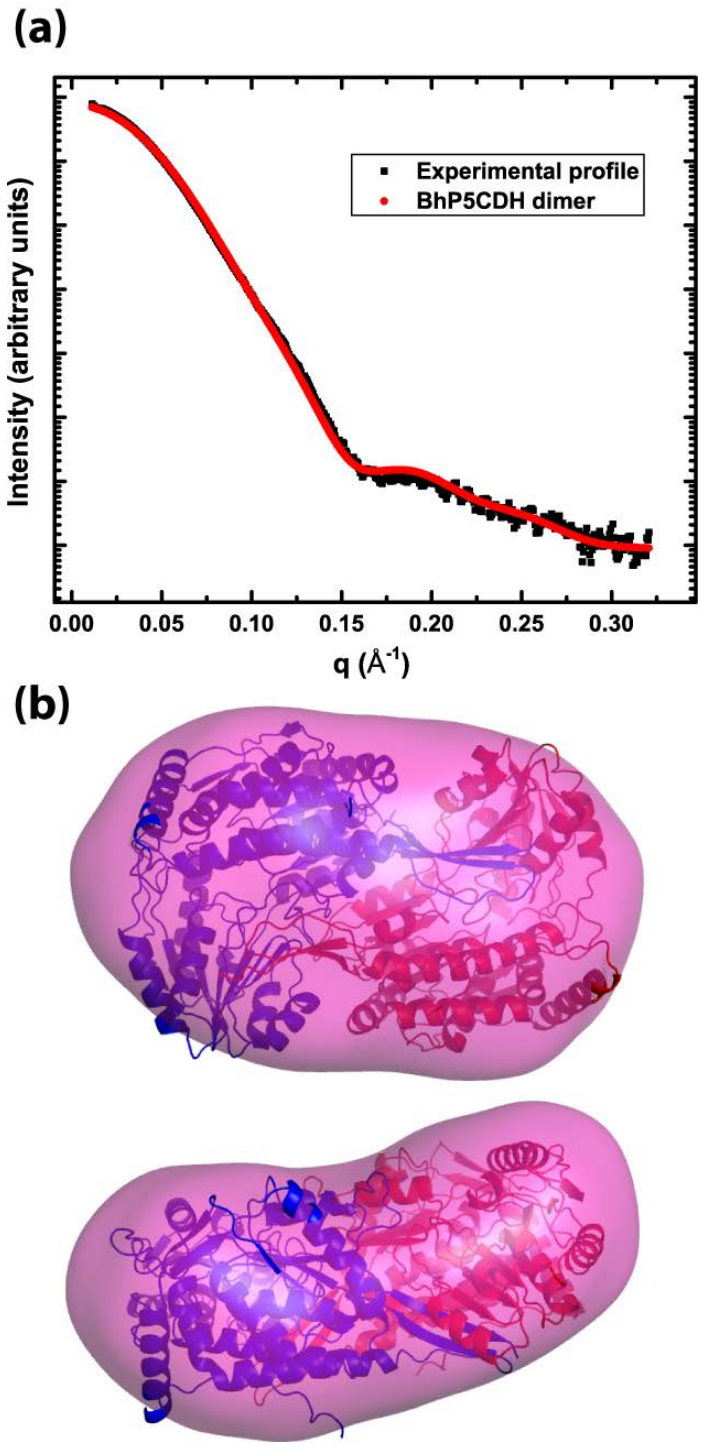


Figure 5.2. SAXS of BhP5CDH. (a) The black curve represents the composite scattering curve and the red curve is the SAXS curve calculated from the domain swapped dimer predicted from the PDBePISA. The calculation has been done using FoXS webserver. (b) Two orthogonal views of the GASBOR P2 envelope (0.5 sigma) with superimposed dimer.

For TtP5CDH, the medium concentration sample was used for the shape reconstruction. Figure 5.4b shows the volumetric map resulting from a GASBOR reconstruction assuming *P*32 symmetry. The NSD for the set of 19 GASBOR models was 1.4 ± 0.1, which indicates a high quality reconstruction. The volumetric map is consistent with the trimer-of-dimers hexamer predicted from PDBePISA (Figure 5.4b). Furthermore, the FoXS calculation suggests that the TtP5CDH experimental SAXS profile is in good agreement with the theoretical TtP5CDH hexamer calculated from its atomic coordinates (Figure 5.4a). On the other hand**,** the theoretical SAXS profile of the domain-swapped dimer deviates substantially from the experimental curve (Figure 5.4a). Figure 5.4b shows the hexamer superimposed onto the GASBOR shape. Note that the envelope matches the size and shape of the hexamer. Furthermore, the R_g of the TtP5CDH hexamer of 43 Å is close to the SAXS estimate (43 – 44 Å). For reference, R_g of the TtP5CDH domain-swapped dimer is only 31 Å. These results suggest that the hexamer predicted from the PDBePISA is stable in solution and represents the oligomeric state of TtP5CDH.

TtP5CDHR100A. The influence of Arg100 on hexamer formation was tested by creating the R100A mutant of TtP5CDH (TtP5CDHR100A). Arg100 forms hydrogen bonds and ion pairs that appear to stabilize the TtP5CDH hexamer (Figure 5.9a, 5.9b, 5.10).

The composite scattering profiles and the corresponding Guinier plots for TtP5CDHR100A are shown in Figure 5.6. The Guinier plots exhibit good linearity for the three different concentrations and suggest an R_g of 32.1-34.3Å (Table 5.6 and Figure 5.6).

Figure 5.6b shows scattering curves scaled to the dilution factor of the sample. The q range shown in the figure is $0 < q < 0.15$ Å^{-1}. The figure suggests that the scattering profile for the lowest concentration sample (black curve) does not match very well with the other two higher concentration sample. Though, the I(0)/C value among the three sample does not vary significantly. The deviation in the curve could be due to the bad buffer subtraction, and may not represent the true profile for the protein sample. Therefore, the lowest concentration was discarded during further investigation. The AutoR_g software gives an R_g estimate of 32 -35 Å. The slight discrepancy in R_g is due to the fact that AutoR_g looks for the best linear fit in the low q region, and for that it omits some of the lower q data points. Pair distribution analysis gives an R_g value in the range of 32.7 to 35.1 Å. The difference in R_g calculated from the two methods is less than 1%, which assures its correct value. Though, the R_g range (32.7-35.1 Å) calculated from pair distribution function is more reliable as it is not as sensitive to the data points at low q region. Note that the R_g of TtP5CDHR100A (33 – 35 Å, Table 5.6) is substantially smaller than that of TtP5CDH (43 – 44 Å, Table 5.5). Calculation of pair distribution function also gives a maximum particle dimension in the range of 110-120Å. Recall that D_{max} of TtP5CDH is 130 – 135 Å (Table 5.5).

For TtP5CDHR100A, the medium concentration sample was used for the shape reconstruction. Figure 5.7b shows volumetric representation of the averaged, filtered GASBOR dummy atom model calculated with P2 symmetry enforced. The NSD for the set of 20 GASBOR models was 1.05 ± 0.05, which indicates a high quality shape reconstruction. The size and shape of GASBOR envelope is consistent with the domain swapped dimer of TtP5CDH and does not resemble the hexamer (Figure 5.7b).

Furthermore, the FoXS calculation suggests that the TtP5CDHR100A experimental SAXS profile is in good agreement with the theoretical TtP5CDH domain swapped dimer rather than a hexamer (Figure 5.7a). This result shows that the R100A mutant of TtP5CDH is a dimer in solution, which implies that Arg100 is essential for hexamer formation and not important for formation of the classic ALDH dimer.

5.3.2. Multi Angle Light Scattering

Multi angle light scattering reveals two different elution profiles for P5CDHs. The first elution time corresponds to DrP5CDH and TtP5CDH which is at 10.2-10.70 minutes (Figure 5.5). These enzymes form hexamers in solution. The second elution profile corresponds to BhP5CDH and BsP5CDH which is at 11.5-11.7 minutes. Recall that these enzymes form dimers in solution. The RI detector gives an average molecular weight of 347, 314, 113, and 114 kDa for TtP5CDH, DrP5CDH, BhP5CDH and BsP5CDH respectively. For reference, the expected molecular weights for dimers and hexamers of P5CDH are 114 kDa and 342 kDa, respectively.

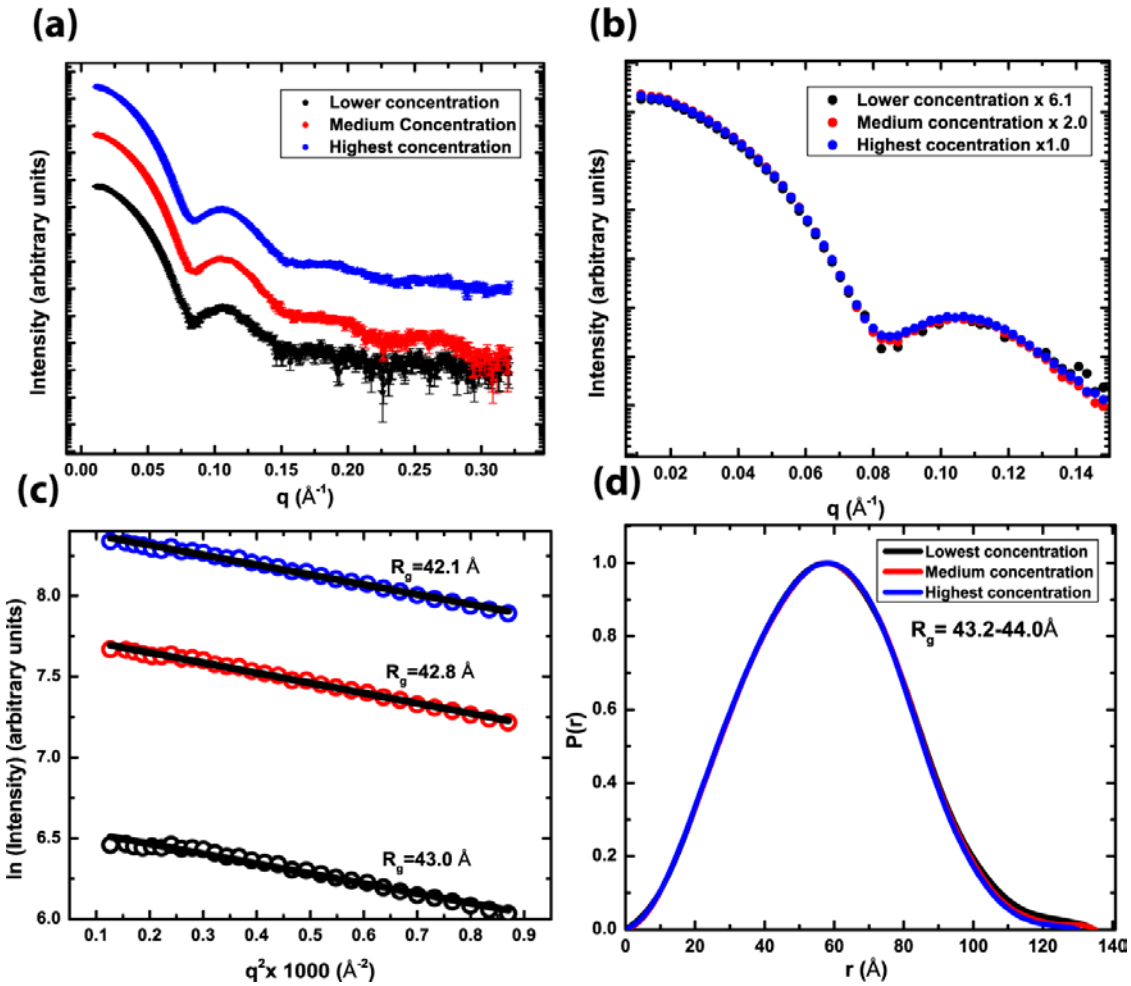


Figure 5.3. SAXS analysis of TtP5CDH (a) Scattering profiles at three different concentrations. Data are offset for clear visibility (b) Scattering profiles after scaling for concentration in the range of 0< q< 0.15Å. (c) Guinier plots and the corresponding radii of gyration for three different concentrations. (d) Normalized pair distribution functions for three different concentrations.

Table 5.5. Statistics of Guinier and pair distribution analysis of TtP5CDH

Sample	Conc (mg/mL)	Guinier Analysis			P (r) Analysis			
		R_g (Å)	I(0)	I(0)/C (relative)	R_g (Å)	I(0)	I(0) /C (relative)	D_{max} (Å)
1[a]	6.3X	43.0 (±0.3)	729 (±3)	4593X	44.0 (±0.1)	732 (±3)	4611X	135
2[b]	2X	42.8 (±0.3)	2359 (±9)	4717X	43.6 (±0.6)	2354 (±5)	4708X	135
3[c]	X	42.1 (±0.3)	4563 (±17)	4563X	43.2 (±0.0)	4574 (±9)	4574X	130

[a] $0.48<qR_g<1.3$ (points 1-30); AutoR_g (88%) R_g =45.4±0.1 (points 9 to 28) $0.76<qR_g<1.3$

[b] $0.48<qR_g<1.3$ (points 1-30); AutoR_g (89%) R_g =44.7±0.2 (points 9 to 29) $0.75<qR_g<1.3$

[c] $0.47<qR_g<1.2$ (points 1-30); AutoR_g (89%) R_g =44.2±0.1 (points 9 to 29) $0.74<qR_g<1.3$

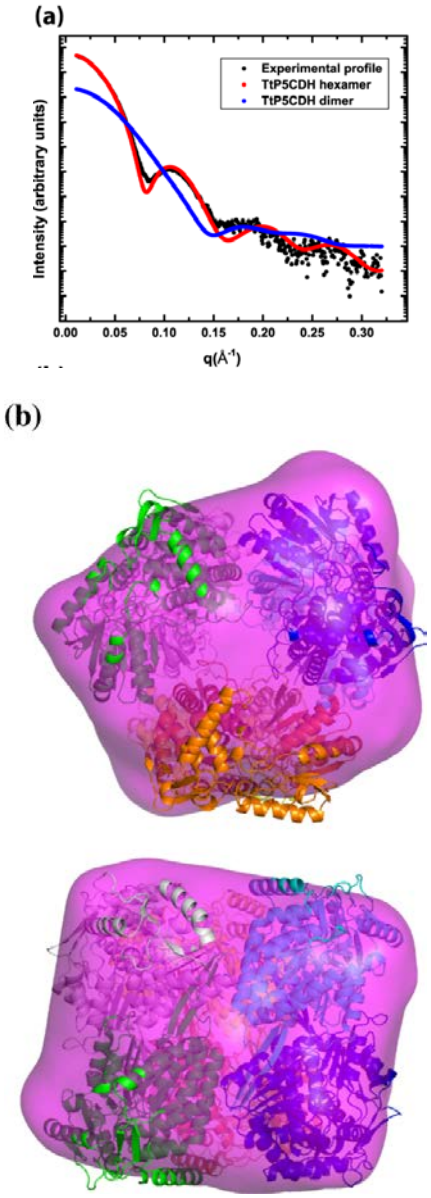


Figure 5.4. SAXS analysis of TtP5CDH. (a) The black curve represents the composite scattering curve and the red curve is the SAXS profile calculated from TtP5CDH hexamer (trimer-of-dimers) predicted from PDBePISA. The blue curve was calculated from the domain-swapped dimer predicted from PISA. The calculation has been done using the FoXS webserver. (b) Two orthogonal views of the GASBOR shape reconstruction in *P*32 symmetry (0.5 sigma). The structure of the predicted hexamer was superimposed onto the surface map using supcomb13. The top structure is viewed down three fold symmetry axis.

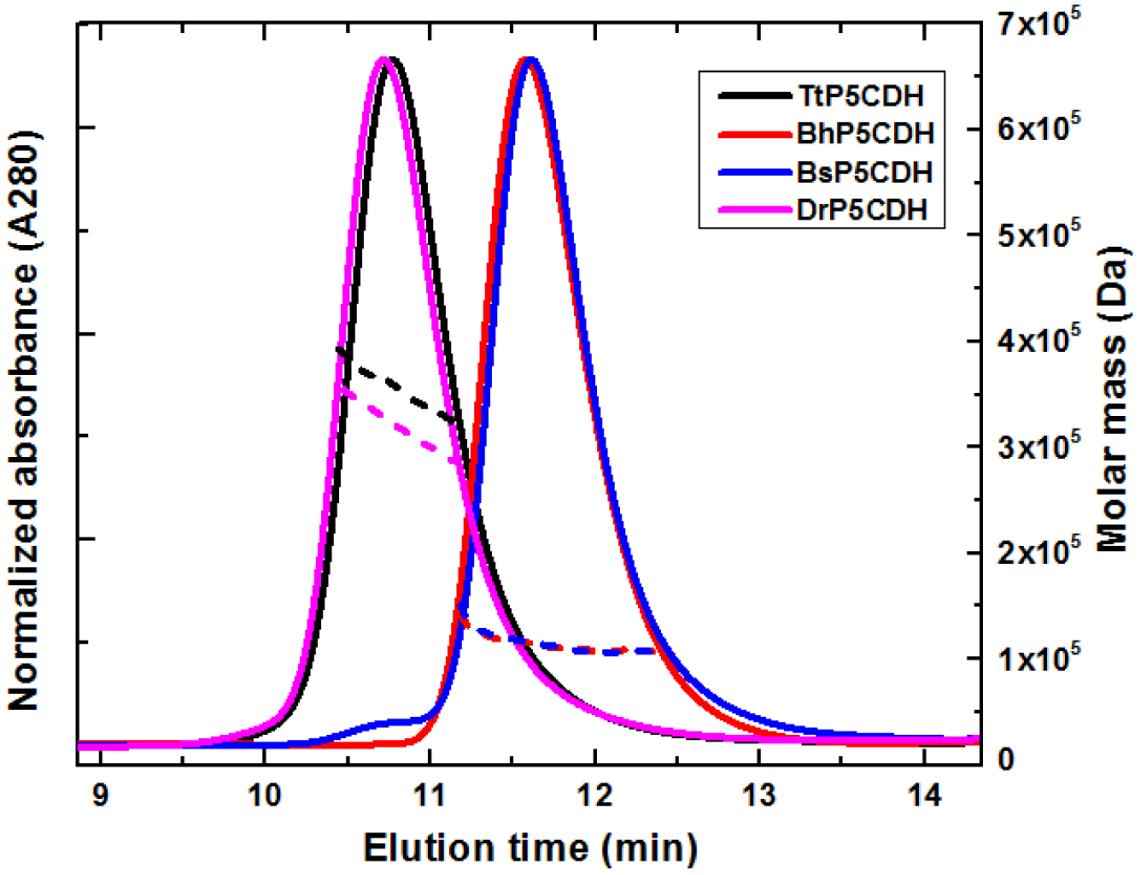


Figure 5.5. Molar mass distribution and elution profile of hexameric and dimeric P5CDHs. The short dashed curve is the molar mass distribution and the smooth curve is the elution profile. Curves for each protein are colored differently.

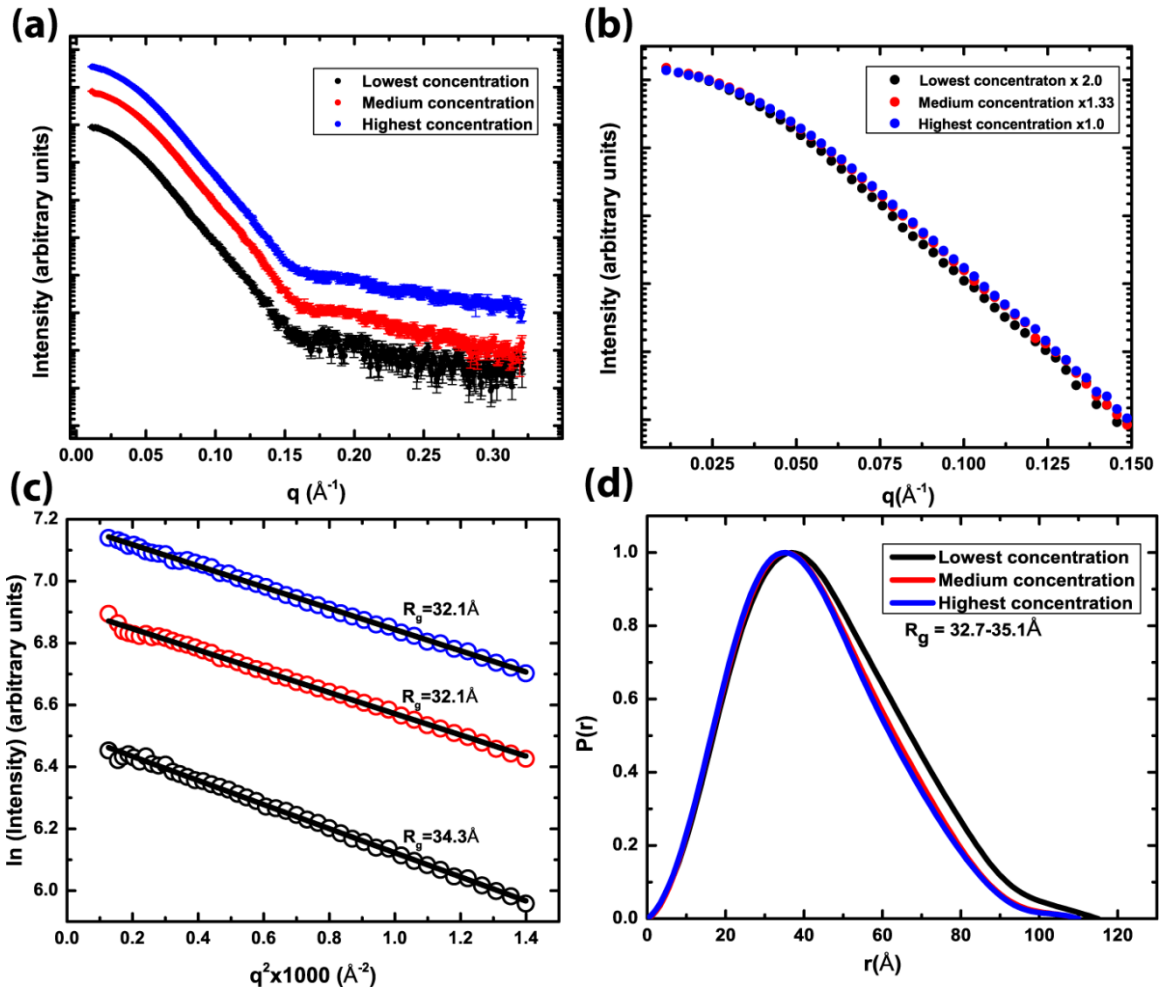


Figure 5.6. SAXS analysis of TtP5CDHR100A (a) Scattering profiles at three different concentrations. Data are offset for clear visibility (b) Scattering profile at three concentration in the range of 0< q< 0.15Å. (c) Guinier plots and the corresponding radii of gyration for three different concentrations. (d) Normalized pair distribution functions for three different concentrations.

Table 5.6. Statistics of Guinier and pair distribution analysis of TtP5CDHR100A

Sample	Conc (mg/mL)	Guinier Analysis			P (r) Analysis			
		R_g (Å)	I(0)	I(0)/C	R_g(Å)	I(0)	I(0)/C	D_{max} (Å)
1[a]	2.2	34.3 (±0.2)	675 (±2)	307	35.1 (±0.1)	680 (1)	309	120
2[b]	3.4	32.1 (±0.1)	1007 (±2)	296	33.0 (±0.0)	1017	299	110
3[c]	4.5	32.1 (±0.2)	1323 (±3)	294	32.7 (±0.0)	1330	295	110

[a]$0.39<qR_g<1.3$(points 1-43); AutoR_g (90%) R_g =34.6±0.0 (points 15 to 43) $0.70<qR_g<1.3$

[b]$0.36<qR_g<1.2$ (points 1-43); AutoR_g (93%) R_g =32.2±0.1 (points 9 to 39) $0.54<qR_g<1.1$

[c]$0.36<qR_g<1.2$(points 1-43); AutoR_g (84%) R_g =31.9±0.1 (points 23 to 48) $0.81<qR_g<1.3$

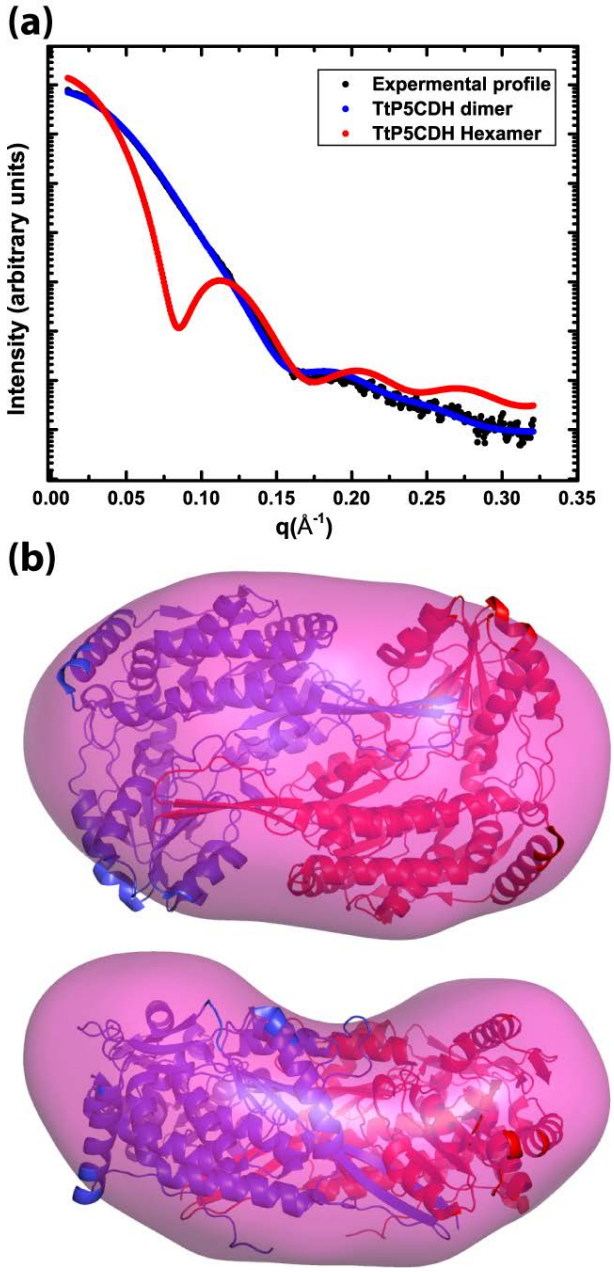


Figure 5.7. SAXS analysis of TtP5CDHR100A. (a) The black curve represents the composite scattering curve and the red curve is the SAXS profile calculated from the TtP5CDH hexamer (trimer-of-dimers) predicted from PDBePISA. The blue curve is SAXS profile calculated from the domain swapped dimer predicted from PISA. The calculation was done using FoXS. (b) Two orthogonal views of the GASBOR shape reconstruction in *P*2 symmetry (0.5 sigma) and superimposed dimer.

The protein-protein interfaces of TtP5CDH were analyzed to understand the structural basis of hexamer formation. PDBePISA predicts three symmetric and one asymmetric interface in the TtP5CDH crystal with space group H3. Only the symmetric interfaces of TtP5CDH will be considered here.

The first symmetric interface covers an interface area of 2900Å^2. More than 16% of the residues of the two subunits participate in the interface formation. These interfacial residues engage themselves in multiple inter-subunit hydrogen bonding, salt-bridge and non-polar interactions and correspond to a well-known domain-swapped dimer. This kind of dimerization is common to a protein of the aldehyde dehydrogenase (ALDH) superfamily. This interface was indeed expected as P5CDH is an ALDH.

The second and third interfaces predicted by PISA account for a higher oligomeric state of TtP5CDH. Figure 5.8 and Tables 5.7 to 5.10 shows the important residues involved in the second and the third interface. The interaction involved in the third predominant interface is also shown in Figures 5.9 and 5.10. The two chains (chain A and chain D) shown in the figure are each subunits from the domain-swapped dimers involved in making the hexamer interface. These interactions create an interface amidst the two domain-swapped dimers, resulting in a *trimer-of-dimers* hexamer. The second interface has a smaller interfacial area of 307-320Å^2. Further, only 1.4 % of the residues are involved in the interface formation. On the other hand, the third interface has a buried surface area of 1087Å^2 and 5.6% of the residues from each participating subunits are in the interfacing region. Furthermore, the inter-subunit hydrogen bonding, salt bridges and the non-polar interactions are much more extensive in the third interface than in the second interface. Because of the larger buried surface area and extensive hydrogen

bonding, salt bridge and non-polar interactions, the third interface was assumed to be most important for the hexamer formation and has been considered for further investigation.

The interactions in the major (third) hexamer interface is symmetric and can be divided into three different regions: (1) interaction of Arg111 and Lys104 of one subunit with the C-terminal end of the other subunit; (2) interaction of Arg100 of one subunit with Asp166, Glu168 and Tyr154 of the other subunit; and (3) interaction involving Arg153 from each subunit (Table 5.7). The interactions involving Arg100 is particularly interesting, as it contributes to 33% of hydrogen bonding and 100% of salt bridge formation. This Arg residue is also present in hexameric DrP5CDH, though is absent in the dimeric P5CDHs. To understand the hexamer enabling and disabling regions in different P5CDHs under investigation, theoretical hexamers of BhP5CDH and BlP5CDH have been created (Figure 5.11, Figure 5.12). In order to make artificial interfaces, the individual subunits of dimeric P5CDHs were taken apart and superimposed individually on hexamer interfaces of TtP5CDH. The surface representations firmly establish that the dimeric P5CDHs lack the predominant hexamer interface (Figure 5.12). Note in particular that the artificial interface has gaps between the protomers (Figure 5.12), whereas the interface of the true hexamer is tightly sealed (Figure 5.10).

Sequence alignments were performed to further understand why some P5CDHs form hexamers (TtP5CDH, DrP5CDH) and other form dimers (BhP5CDH, BlP5CDH, BsP5CDH). The pairwise sequence identities for the 5 proteins are listed in table 5.3, while the sequence alignment is shown in Figure 5.13. The regions highlighted as blue

correspond to the hexamer interface of TtP5CDH. The possible enabling or disabling regions are not evident from the sequence alignment results. An explicit reasoning for the hexamerization event was not revealed, as there were no specific regions conserved in hexameric P5CDH in the hexamer interface region which is not found dimeric P5CDH. Though it is interesting to observe that the Arg100 is also present in DrP5CDh and not present in the dimeric P5CDH. This residue is replaced by Ala or Asn in the dimeric P5CDHs.

The TtP5CDHR100A results suggest that positively charged residues in interface 2 are important for hexamer formation. The TtP5CDH structure shows that Arg111 and Arg153 also engage in intersubunit contacts. It would be interesting create the R111A and R153A mutants of TtP5CDH.

5.3.3. Crystal structure of TtP5CDHR100A mutant

R100 is located at the predominant hexamer interface. This residue interacts with D166, E168 and Y154 at the major (third) hexamer interface (Figure 5.14, Figure 15A). Interestingly, this residue is also interacts with E458 in the second (smaller) interface. Mutation of R100 to A disrupts these interactions. In addition, the loop containing Y154 moved by 3.7 Å (Figure 15A). This shift resulted change in the conformation of R153 (Figure 5.15B) that disrupted the inter-subunit stacking interaction of R153 at the third interface.

5.4 Summary and Conclusion.

Two oligomeric states of P5CDH has been determined. SAXS and MALS study showed TtP5CDH and DrP5CDH is a hexamer, On the other hand BhP5CDH and BlP5CDH is a dimer. Hexamer interface of TtP5CDH has been investigated using PDBePISA and site

directed mutagenesis. Based on the TtP5CDH R100A solution and crystal structure, possible hot spot for the hexamer formation is located near the center of the third hexamer interface.

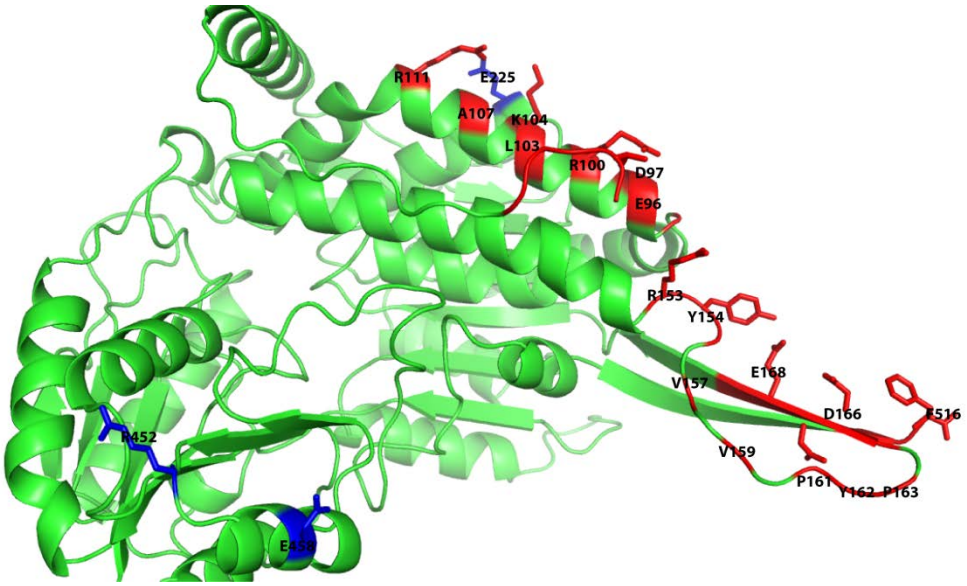


Figure 5.8. Protomer of TtP5CDH. The residues shown in red and blue are the interfacial residue involved in making the hexamer interfaces. Blue color residues represent smaller hexamer interface while the red colored residues represent the bigger hexamer interface. Residues involved in hydrogen bonds or salt bridges are shown in sticks.

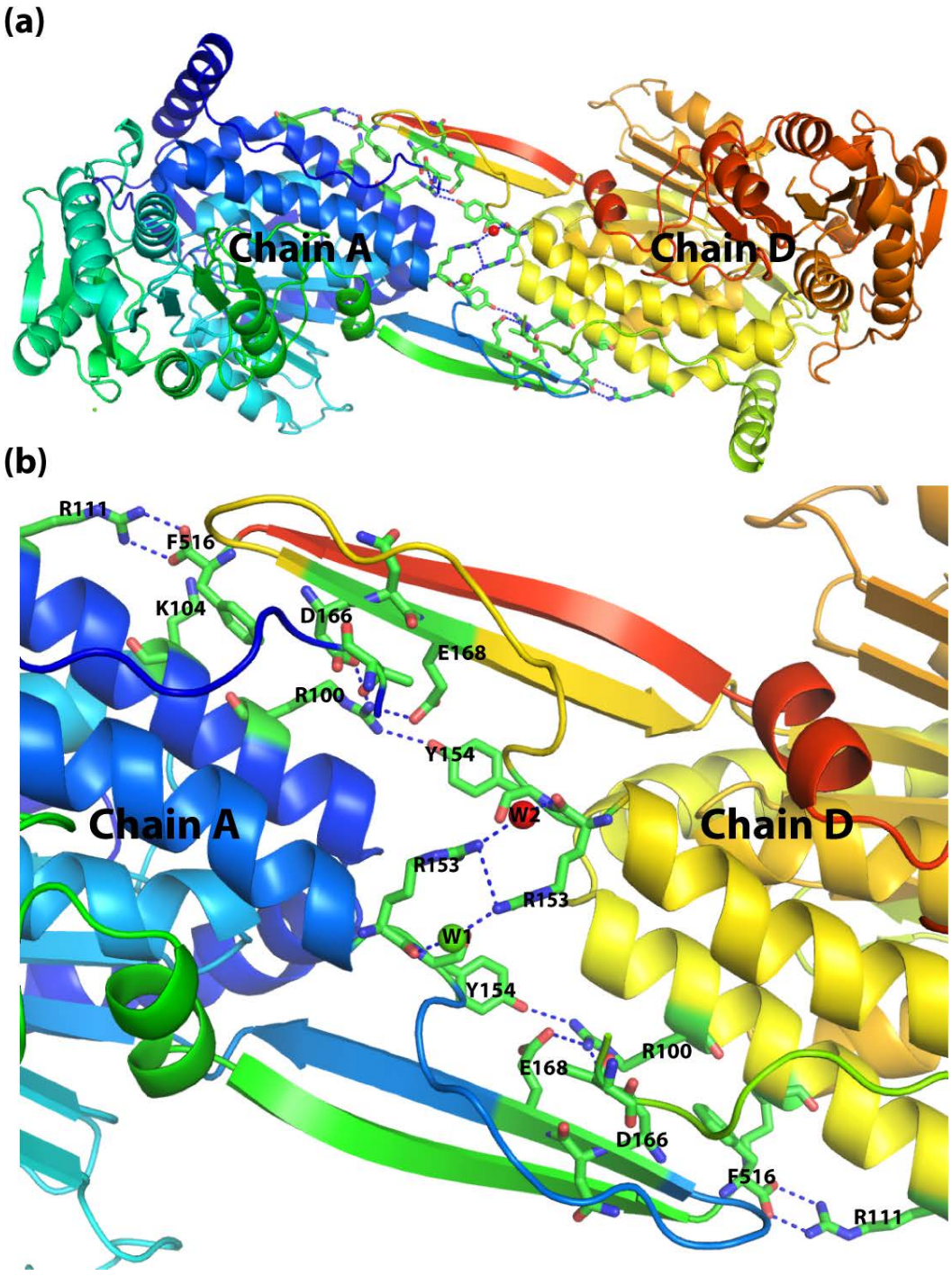


Figure 5.9. Predominant hexamer interface TtP5CDH. (a) Chains A and D are colored in a rainbow scheme with blue at the N-terminus and red at the C-terminus. Interfacial residues are shown in sticks. Two water molecules in the interface are shown in spheres. (b) Residues involved in hydrogen bonding and salt bridges between chains A and D of the hexamer interface are shown in stick.

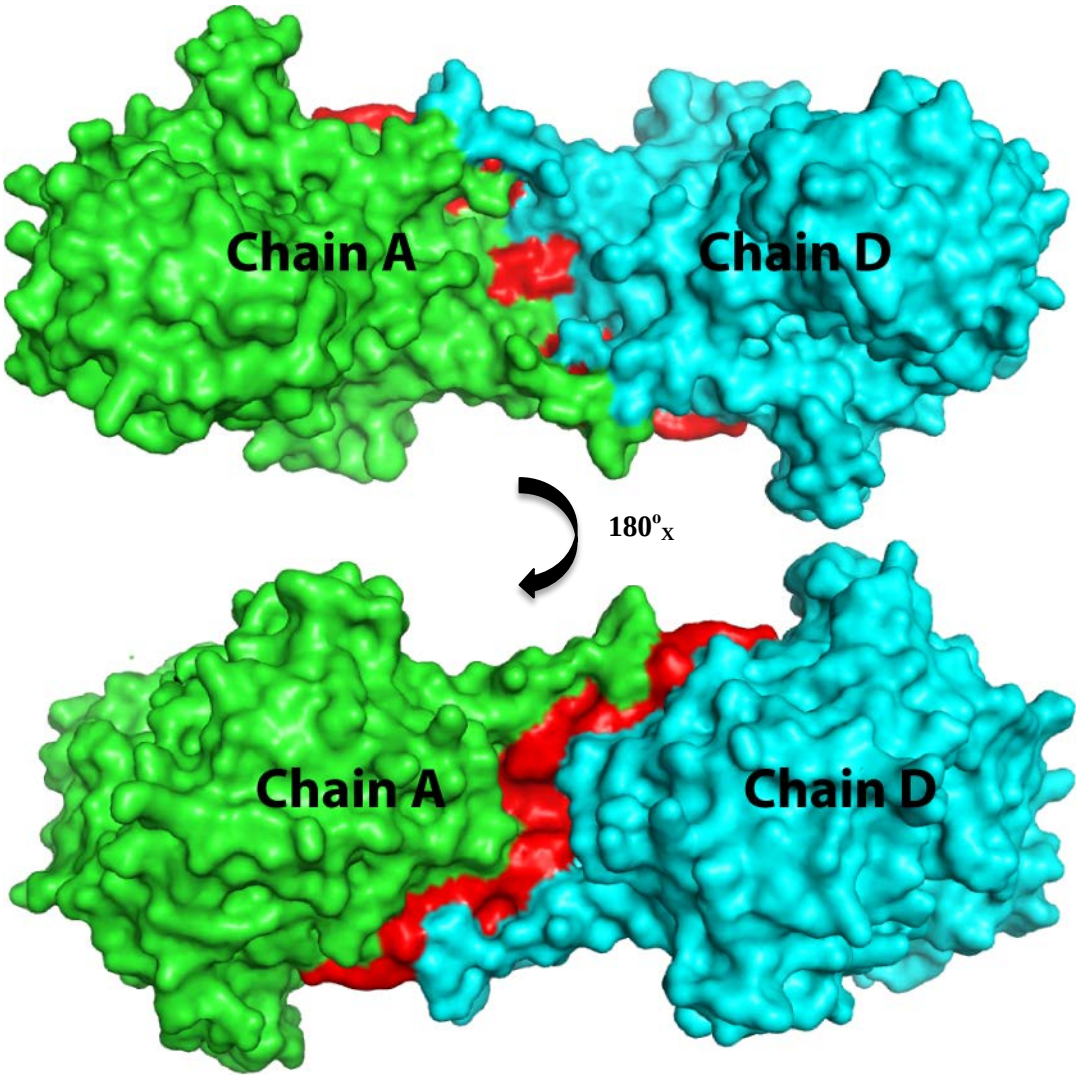


Figure 5.10. Surface representation of the hexamer interface of TtP5CDH. The two chains are colored in green and cyan, while the interface residues involved in hydrogen bonds and salt bridges are colored red.

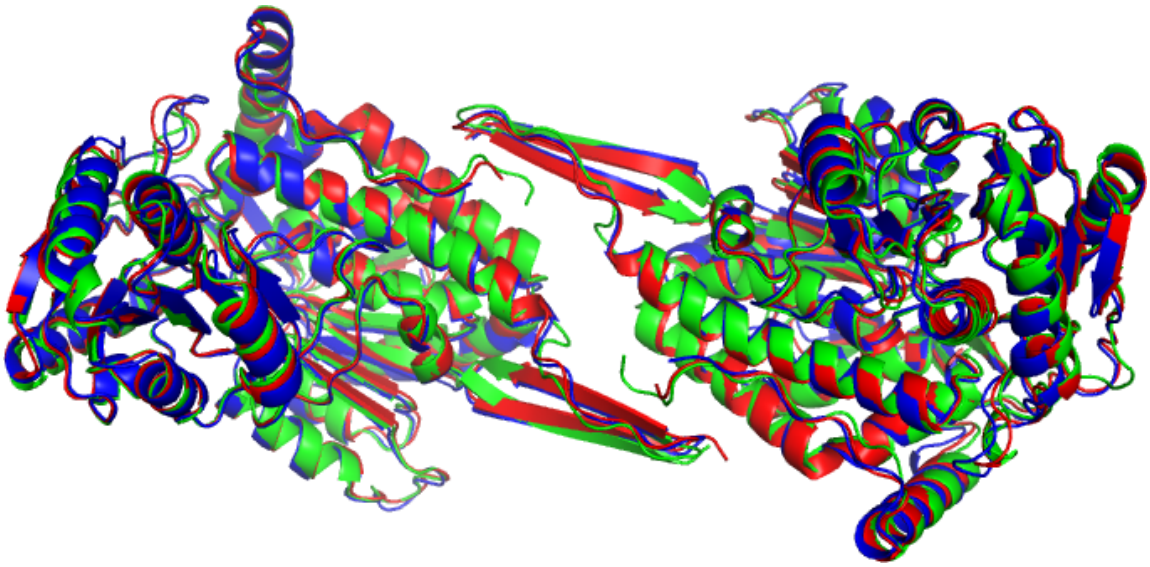


Figure 5.11. Superposition of hexamer interface of TtP5CDH (green) with the theoretical hexamer interface of BhP5CDH (blue) and BlP5CDH (red). The theoretical hexamer was created by superposition of individual subunits of natural domain swapped dimer onto the respective subunits of the TtP5CDH hexamer.

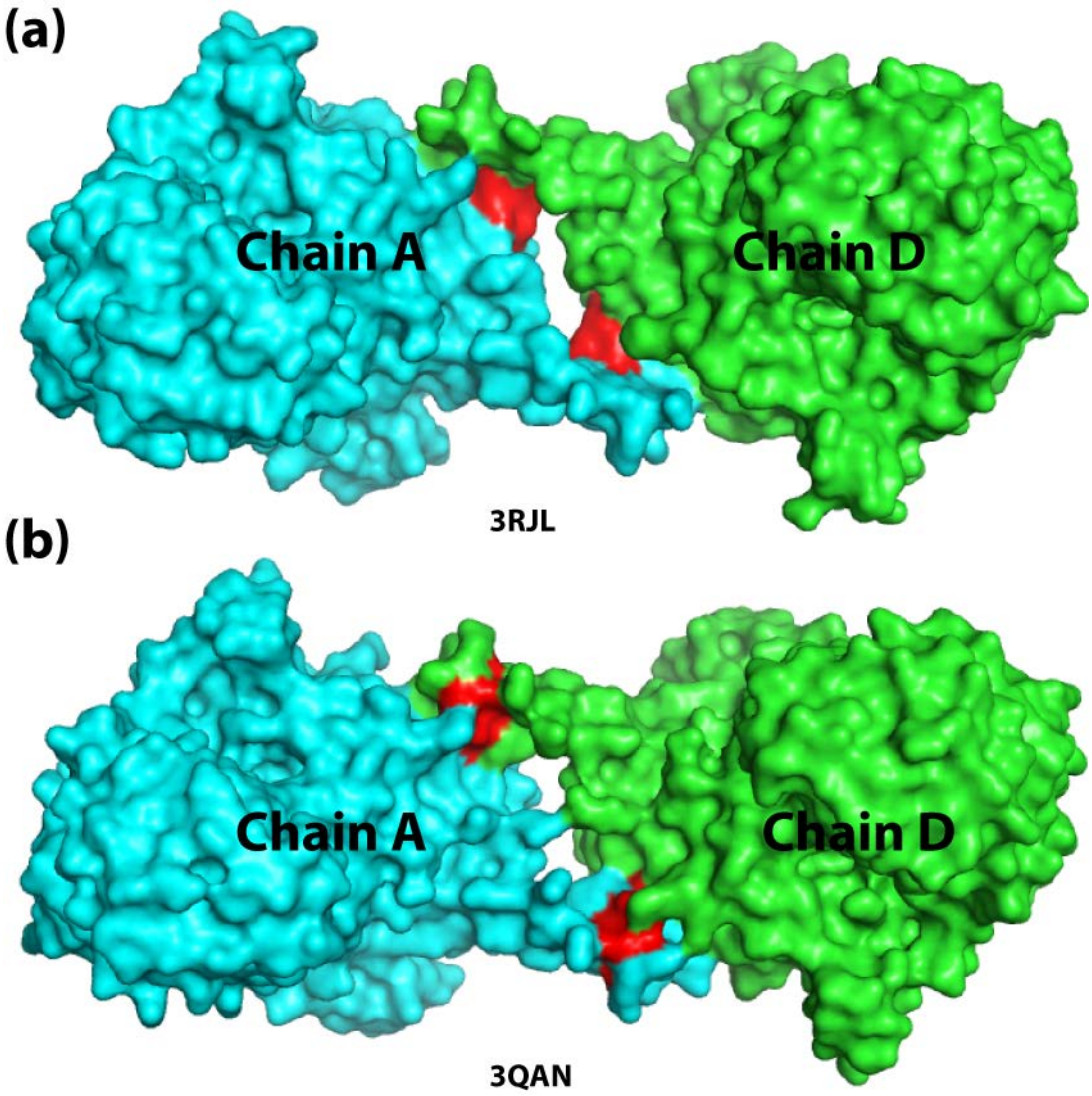


Figure 5.12. Artificial hexamer interfaces of BhP5CDH and BlP5CDH. These were created by superposition of individual chains of natural domain-swapped dimers (chain A and chain B) of BhPhP5CDH and BlP5CDH onto the TtP5CDH hexamer.

TABLE 5.7. Predominant hexamer interface hydrogen bond interactions.

Interaction	Residue 1	Distance (Å)	Residue 2
1	A:THR 2 [OG1]	2.4	B:ASP 166 [OD2]
2	A:ARG 100 [NH1]	3.1	B:ASP 166 [OD1]
3	A:ARG 100 [NH1]	2.9	B:GLU 168 [OE2]
4	A:ARG 100 [NH2]	3.2	B:TYR 154 [OH]
5	A:LYS104 [NZ]	2.8	B:PHE516[O]
6	A:ARG111[NH1]	2.9	B:PHE516[O]
7	A: ARG111[NH2]	2.6	B:PHE516{OXT]
8	A:ARG153[NH1]	3.1	B:TYR154[O]
9	A:ASN167[N]	2.8	B:THR2[O]
10	A:ASP166[OD2]	2.5	B:THR2[OG1]
11	A:GLU168[OE2]	2.9	B:ARG100[NH1]
12	A:ASP166[OD1]	3.1	B:ARG100[NH1]
13	A:TYR154[OH]	3.2	B:ARG100[NH2]
14	A:PHE516[O]	2.8	B:LYS104[NZ]
15	A:PHE516[O]	2.9	B:ARG111[NH1]
16	A:PHE516[OXT]	2.6	B:ARG111[NH2]
17	A:TYR154[O]	2.9	B:ARG153[NH1]
18	A:THR2[O]	2.8	B:ASN167[N]

Table 5.8. Predominant hexamer Interface salt bridge interactions

Interaction	Residue 1	Distance (Å)	Residue 2
1	A:ARG100 [NH1]	3.1	B:ASP1666 [OD1]
2	A:ARG100[NH1]	3.6	B:GLU168[OE1]
3	A:ARG100[NH1]	2.9	B:GLU168[OE2]
4	A:ARG100[NH2]	3.9	B:GLU168[OE1]
5	A:ARG100[NH2]	3.9	B:GLU168[OE1]
6	A:GLU168[OE1]	3.6	B:ARG100[OE2]
7	A:GLU168[OE2]	2.9	B:ARG100[NH1]
8	A:ASP 166[OD1]	3.1	B:ARG100[NH1]
9	A:GLU168[OE1]	3.9	B:ARG100[NH2]
10	A:GLU168[OE2]	3.9	B:ARG100[NH2]

Table 5.9. Small hexamer interface hydrogen bond interaction.

Interaction	Residue 1	Distance (Å)	Residue 2
1	B:ARG 452[NE]	3.4	B:GLU225[OE1]
2	B:ARG452[NH1]	3.8	B:GLU225[OE1]
3	B:GLU458[OE1]	3.0	B:ARG100[NE]
4	B:GLU458[OE2]	3.5	B:ARG100[NH2]

Table 5.10. Small hexamer interface salt bridge interaction.

Interaction	Residue 1	Distance (Å)	Residue 2
1	B:ARG452[NE]	3.4	B:GLU225[OE1]
2	B:ARG452[NH1]	3.8	B:GLU225[OE1]
3	B:GLU458[OE1]	3.0	B:ARG100[NE]
4	B:GLU458[OE1]	3.8	B:ARG100[NH2]
5	B:GLU458[OE2]	3.8	B:ARG100[NE]
6	B:GLU458[OE2]	3.5	B:ARG100[NH2]

Table 5.11. Data collection and refinement statistics of TtP5CDHR100A

	R100AMPD
Wavelength (Å)	1.000
Detector distance (mm)	110
Space group	H3
Cell dimensions *a*, *c* (Å)	102.7, 279.5
Resolution (Å)	47.3 - 1.54 (1.62 - 1.54)
I / (σI)	20.7 (2.3)
Completeness (%)	96.8 (78.5)
Redundancy	5.3 (2.8)
Total observations	835877
Unique reflections	158507
R_{meas}#	0.051(0.549)
R_{pim}#	0.021(0.291)
R_{merge}#	0.041(0.527)
Protein	
Water molecules	532
B-factors	
Protein	22.1
Water	27.4
Rwork/Rfree	0.172/0.190
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.01
Ramachandran plot*	
Favored	1016
Allowed	10
Outliers	2

Values of the outer resolution shell of data are given in parenthesis.
#Definition of R_{merge},R_{meas} and R_{pim} can be found in Weiss *et al.* 2001(60).
*Ramachandran plot was generated from RAMPAGE (61).

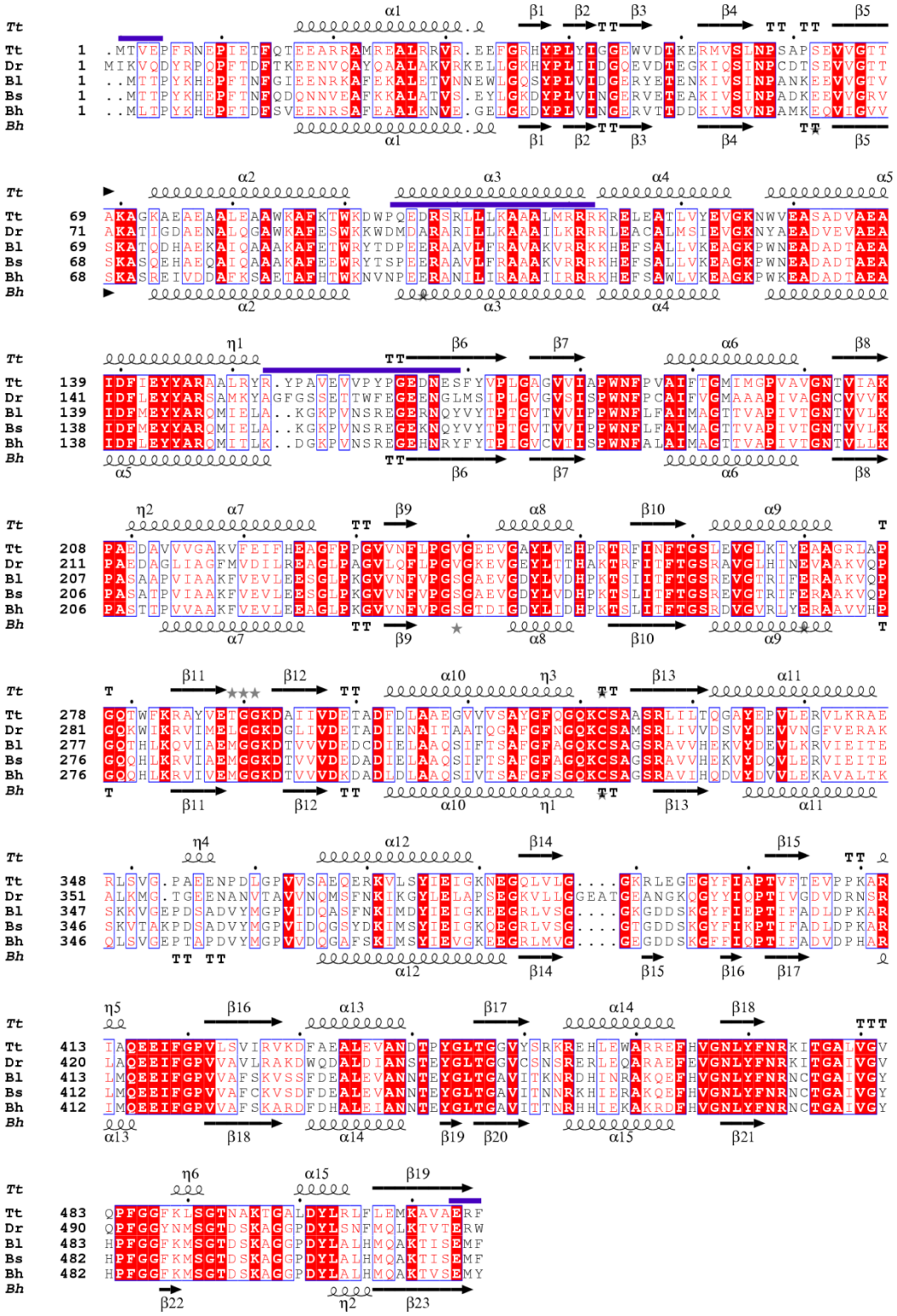


Figure 5.13. Multiple sequence alignment of two hexameric P5CDHs (TtP5CDH, DrP5CDH) and three dimeric P5CDHs (BlP5CDH, BsP5CDH, BhP5CDH).

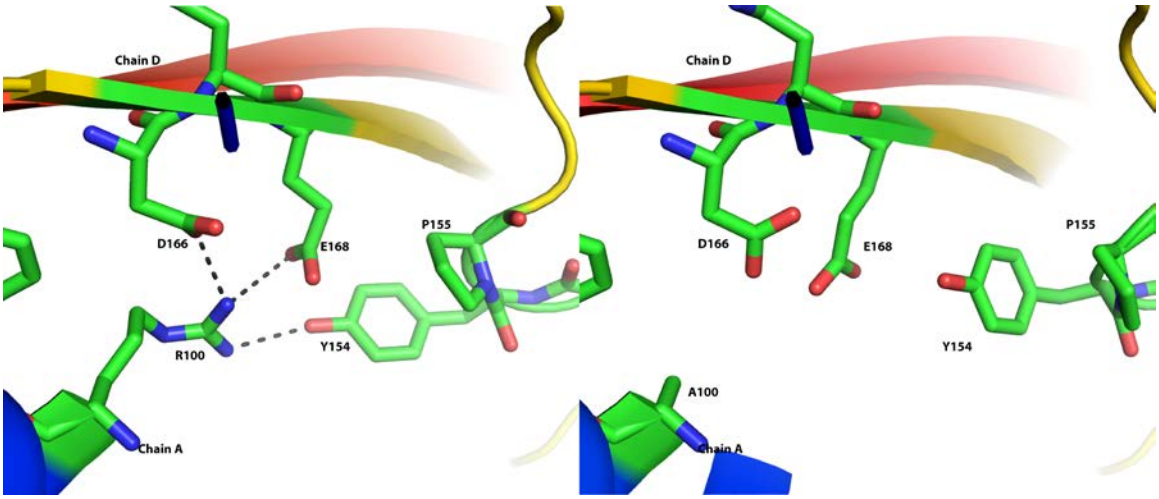


Figure 5.14. A) Interactions of Arg100 of chain A with D166,E168 and Y154 of chain D at the third hexamer interface of TtP5CDH (left) B) These interactions are missing at TtP5CDHR100A (right). In addition, the interaction partners of R100 moved away from their original position.

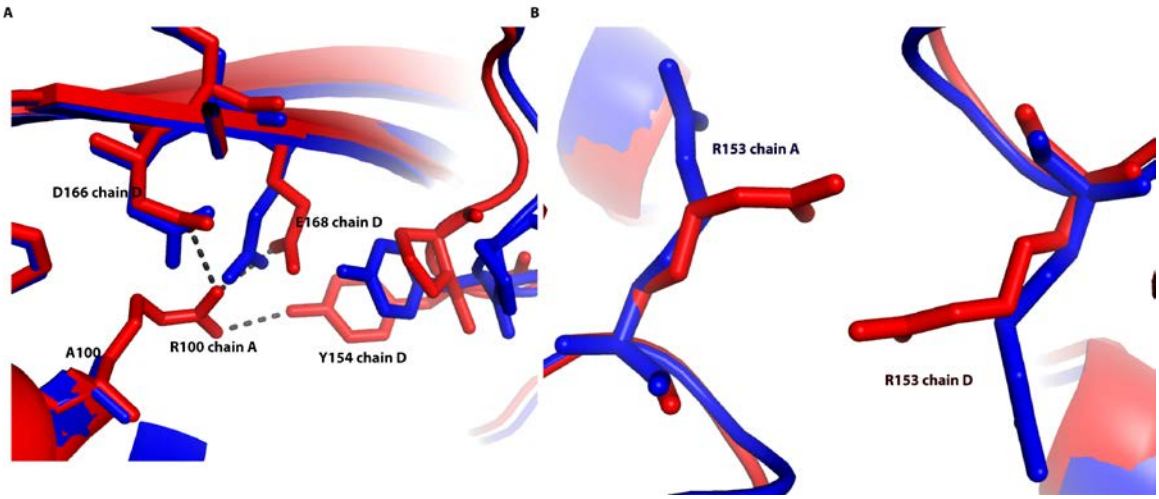


Figure 5.15. Superposition of TtP5CDH structure on TtP5CDHR100A structure.(A) disruption of R100 interaction at the hexamer interface B) Disruption of R153 interaction at the hexamer interface

5.5. References

1. Goodsell DS, Olson AJ. Structural symmetry and protein function. Annual review of biophysics and biomolecular structure. 2000;29:105-53. Epub 2000/08/15.

2. Matthews jM. Protein Dimerization and Oligomerization in Biology: Springer Science 2012. 166 p.

3. Ward WH, Jones DH, Fersht AR. Protein engineering of homodimeric tyrosyl-tRNA synthetase to produce active heterodimers. The Journal of biological chemistry. 1986;261(21):9576-8. Epub 1986/07/25.

4. Brange J, Ribel U, Hansen JF, Dodson G, Hansen MT, Havelund S, et al. Monomeric insulins obtained by protein engineering and their medical implications. Nature. 1988;333(6174):679-82. Epub 1988/06/16.

5. Perham RN. Self-assembly of biological macromolecules. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 1975;272(915):123-36. Epub 1975/11/06.

6. Beernink PT, Tolan DR. Subunit interface mutants of rabbit muscle aldolase form active dimers. Protein science : a publication of the Protein Society. 1994;3(9):1383-91. Epub 1994/09/01.

7. Beernink PT, Tolan DR. Disruption of the aldolase A tetramer into catalytically active monomers. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(11):5374-9. Epub 1996/05/28.

8. Salwinski L, Eisenberg D. Computational methods of analysis of protein-protein interactions. Current opinion in structural biology. 2003;13(3):377-82. Epub 2003/07/02.

9. Jones S, Thornton JM. Principles of protein-protein interactions. Proceedings of the National Academy of Sciences. 1996;93(1):13-20.

10. Ansari S, Helms V. Statistical analysis of predominantly transient protein–protein interfaces. Proteins: Structure, Function, and Bioinformatics. 2005;61(2):344-55.

11. Ofran Y, Rost B. Analysing six types of protein-protein interfaces. Journal of molecular biology. 2003;325(2):377-87. Epub 2002/12/19.

12. Bahadur RP, Chakrabarti P, Rodier F, Janin J. A dissection of specific and non-specific protein-protein interfaces. Journal of molecular biology. 2004;336(4):943-55. Epub 2004/04/21.

13. Miller S. The structure of interfaces between subunits of dimeric and tetrameric proteins. Protein engineering. 1989;3(2):77-83. Epub 1989/11/01.

14. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. Journal of molecular biology. 1998;280(1):1-9. Epub 1998/07/07.

15. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. Science. 1995;267(5196):383-6. Epub 1995/01/20.

16. Gerk LP, Leven O, Müller-Hill B. Strengthening the dimerisation interface of lac repressor increases its thermostability by 40 deg. C. Journal of molecular biology. 2000;299(3):805-12.

17. Vetriani C, Maeder DL, Tolliday N, Yip KS-P, Stillman TJ, Britton KL, et al. Protein thermostability above 100°C: A key role for ionic interactions. Proceedings of the National Academy of Sciences. 1998;95(21):12300-5.

18. DeFelippis MR, Chance RE, Frank BH. Insulin self-association and the relationship to pharmacokinetics and pharmacodynamics. Critical reviews in therapeutic drug carrier systems. 2001;18(2):201-64. Epub 2001/04/28.

19. Jones DH, McMillan AJ, Fersht AR, Winter G. Reversible dissociation of dimeric tyrosyl-tRNA synthetase by mutagenesis at the subunit interface. Biochemistry. 1985;24(21):5852-7. Epub 1985/10/08.

20. Ward WH, Jones DH, Fersht AR. Effects of engineering complementary charged residues into the hydrophobic subunit interface of tyrosyl-tRNA synthetase. Appendix: Kinetic analysis of dimeric enzymes that reversibly dissociate into inactive subunits. Biochemistry. 1987;26(13):4131-8. Epub 1987/06/30.

21. Miller MD, Krause KL. Identification of the Serratia endonuclease dimer: Structural basis and implications for catalysis. Protein Science. 1996;5(1):24-33.

22. Franke I, Meiss G, Pingoud A. On the advantage of being a dimer, a case study using the dimeric Serratia nuclease and the monomeric nuclease from Anabaena sp. strain PCC 7120. The Journal of biological chemistry. 1999;274(2):825-32. Epub 1999/01/05.

23. Franke I, Meiss G, Blecher D, Gimadutdinow O, Urbanke C, Pingoud A. Genetic engineering, production and characterisation of monomeric variants of the dimeric Serratia marcescens endonuclease. FEBS letters. 1998;425(3):517-22. Epub 1998/05/01.

24. Adams E, Frank L. Metabolism of proline and the hydroxyprolines. Annual review of biochemistry. 1980;49:1005-61. Epub 1980/01/01.

25. Tanner JJ. Structural biology of proline catabolism. Amino Acids. 2008;35(4):719-30.

26. Srivastava D, Singh RK, Moxley MA, Henzl MT, Becker DF, Tanner JJ. The three-dimensional structural basis of type II hyperprolinemia. Journal of molecular biology. 2012;420(3):176-89. Epub 2012/04/21.

27. Inagaki E, Ohshima N, Takahashi H, Kuroishi C, Yokoyama S, Tahirov TH. Crystal structure of Thermus thermophilus Delta1-pyrroline-5-carboxylate dehydrogenase. Journal of molecular biology. 2006;362(3):490-501. Epub 2006/08/29.

28. Carter P, Bedouelle H, Winter G. Construction of heterodimer tyrosyl-tRNA synthetase shows tRNATyr interacts with both subunits. Proceedings of the National Academy of Sciences of the United States of America. 1986;83(5):1189-92. Epub 1986/03/01.

29. Vora JP, Owens DR, Dolben J, Atiea JA, Dean JD, Kang S, et al. Recombinant DNA derived monomeric insulin analogue: comparison with soluble human insulin in normal subjects. Bmj. 1988;297(6658):1236-9. Epub 1988/11/12.

30. McKenzie AN, Ely B, Sanderson CJ. Mutated interleukin-5 monomers are biologically inactive. Molecular immunology. 1991;28(1-2):155-8. Epub 1991/01/01.

31. Dickason RR, English JD, Huston DP. Engineering of a functional interleukin-5 monomer: a paradigm for redesigning helical bundle cytokines with therapeutic potential in allergy and asthma. Journal of molecular medicine. 1996;74(9):535-46. Epub 1996/09/01.

32. Sano T, Vajda S, Smith CL, Cantor CR. Engineering subunit association of multisubunit proteins: a dimeric streptavidin. Proceedings of the National Academy of Sciences of the United States of America. 1997;94(12):6153-8. Epub 1997/06/10.

33. Pazy Y, Eisenberg-Domovich Y, Laitinen OH, Kulomaa MS, Bayer EA, Wilchek M, et al. Dimer-tetramer transition between solution and crystalline states of streptavidin and avidin mutants. Journal of bacteriology. 2003;185(14):4050-6. Epub 2003/07/03.

34. Laitinen OH, Nordlund HR, Hytonen VP, Uotila ST, Marttila AT, Savolainen J, et al. Rational design of an active avidin monomer. The Journal of biological chemistry. 2003;278(6):4010-4. Epub 2002/11/30.

35. Velichko IS, Mikalahti K, Kasho VN, Dudarenkov VY, Hyytia T, Goldman A, et al. Trimeric inorganic pyrophosphatase of Escherichia coli obtained by directed mutagenesis. Biochemistry. 1998;37(2):734-40. Epub 1998/02/21.

36. Salminen A, Efimova IS, Parfenyev AN, Magretova NN, Mikalahti K, Goldman A, et al. Reciprocal effects of substitutions at the subunit interfaces in hexameric pyrophosphatase of Escherichia coli. Dimeric and monomeric forms of the enzyme. The Journal of biological chemistry. 1999;274(48):33898-904. Epub 1999/11/24.

37. Fritsche P, Alves J. A monomeric mutant of restriction endonuclease EcoRI nicks DNA without sequence specificity. Biological chemistry. 2004;385(10):975-85. Epub 2004/11/24.

38. Zaremba M, Sasnauskas G, Urbanke C, Siksnys V. Conversion of the tetrameric restriction endonuclease Bse634I into a dimer: oligomeric structure-stability-function correlations. Journal of molecular biology. 2005;348(2):459-78. Epub 2005/04/07.

39. Mossing MC, Sauer RT. Stable, monomeric variants of lambda Cro obtained by insertion of a designed beta-hairpin sequence. Science. 1990;250(4988):1712-5. Epub 1990/12/21.

40. Borchert TV, Abagyan R, Jaenicke R, Wierenga RK. Design, creation, and characterization of a stable, monomeric triosephosphate isomerase. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(4):1515-8. Epub 1994/02/15.

41. Schliebs W, Thanki N, Jaenicke R, Wierenga RK. A double mutation at the tip of the dimer interface loop of triosephosphate isomerase generates active monomers with reduced stability. Biochemistry. 1997;36(32):9655-62. Epub 1997/08/12.

42. Mainfroid V, Terpstra P, Beauregard M, Frere JM, Mande SC, Hol WG, et al. Three hTIM mutants that provide new insights on why TIM is a dimer. Journal of molecular biology. 1996;257(2):441-56. Epub 1996/03/29.

43. Banci L, Benedetto M, Bertini I, Del Conte R, Piccioli M, Viezzoli MS. Solution structure of reduced monomeric Q133M2 copper, zinc superoxide dismutase (SOD). Why is SOD a dimeric enzyme? Biochemistry. 1998;37(34):11780-91. Epub 1998/08/26.

44. Banci L, Bertini I, Chiu CY, Mullenbach GT, Viezzoli MS. Synthesis and characterization of a monomeric mutant Cu/Zn superoxide dismutase with partially reconstituted enzymic activity. European journal of biochemistry / FEBS. 1995;234(3):855-60. Epub 1995/12/15.

45. Porvari KS, Herrala AM, Kurkela RM, Taavitsainen PA, Lindqvist Y, Schneider G, et al. Site-directed mutagenesis of prostatic acid phosphatase. Catalytically important aspartic acid 258, substrate specificity, and oligomerization. The Journal of biological chemistry. 1994;269(36):22642-6. Epub 1994/09/09.

46. Breiter DR, Resnik E, Banaszak LJ. Engineering the quaternary structure of an enzyme: construction and analysis of a monomeric form of malate dehydrogenase from Escherichia coli. Protein science : a publication of the Protein Society. 1994;3(11):2023-32. Epub 1994/11/01.

47. Jackson RM, Gelpi JL, Cortes A, Emery DC, Wilks HM, Moreton KM, et al. Construction of a stable dimer of Bacillus stearothermophilus lactate dehydrogenase. Biochemistry. 1992;31(35):8307-14. Epub 1992/09/08.

48. Bailey DL, Fraser ME, Bridger WA, James MN, Wolodko WT. A dimeric form of Escherichia coli succinyl-CoA synthetase produced by site-directed mutagenesis. Journal of molecular biology. 1999;285(4):1655-66. Epub 1999/01/26.

49. Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL, 2nd, Tsutakawa SE, et al. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods. 2009;6(8):606-12. Epub 2009/07/22.

50. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr. 2003;36(5):1277-82.

51. Svergun D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr. 1992;25(4):495-503.

52. Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. Biophysical journal. 2001;80(6):2946-53. Epub 2001/05/24.

53. V.V. Volkov DIS. Uniqueness of *ab initio* shape determination in small -angle scattering. J Appl Crystallogr. 2003;36:860-4.

54. Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res. 2010;38(Web Server issue):W540-4. Epub 2010/05/29.

55. Wyatt P. Light scattering and absolute characterization of macromolecules. Anal Chim Acta. 1993;272:1-40.

56. Inagaki E, Takahashi H, Kuroishi C, Tahirov TH. Crystallization and avoiding the problem of hemihedral twinning in crystals of Delta1-pyrroline-5-carboxylate dehydrogenase from Thermus thermophilus. Acta crystallographica Section F, Structural biology and crystallization communications. 2005;61(Pt 6):609-11. Epub 2006/03/03.

57. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):125-32. Epub 2010/02/04.

58. Evans P. Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr. 2006;62(Pt 1):72-82.

59. Potterton E, Briggs P, Turkenburg M, Dodson E. A graphical user interface to the CCP4 program suite. Acta Crystallogr D Biol Crystallogr. 2003;59(Pt 7):1131-7.

60. Weiss M. Global indicators of X-ray data quality. J Appl Cryst. 2001;34(2):130-5.

61. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins. 2003;50(3):437-50.

CHAPTER 6

The Three–Dimensional Structural Basis of Type II Hyperprolinemia.

This chapter is adopted from

D. Srivastava, Ranjan K. Singh*, M.A. Moxely, M.T. Henzl, Donald F. Becker, and John J tanner.

J. Mol. Biol (2012) 420 (3):176-89

My Contribution to the Paper

***R.K. Singh has purified all the proteins for kinetics experiments and solved the structure of S352A and S352L.**

6.1. Introduction

The enzyme Δ^1-pyrroline-5-carboxylate (P5C) dehydrogenase (P5CDH, EC 1.5.1.12) is a nuclear-encoded mitochondrial matrix protein that catalyzes the final step of proline catabolism (Figure 6.1, upper reactions) (Adams and Frank (1)). Despite its name, P5CDH is an aldehyde dehydrogenase (ALDH); it catalyzes the NAD^+-dependent oxidation of L-glutamate-γ-semialdehyde (GSA) to L-glutamate. GSA is the hydrolysis product of Δ^1-pyrroline-5-carboxylate (P5C), which is generated from proline by proline dehydrogenase (PRODH). In humans, PRODH is called proline oxidase (POX) in recognition of its function as a superoxide-generating tumor suppressor protein (2). GSA is also produced from ornithine by ornithine aminotransferase.

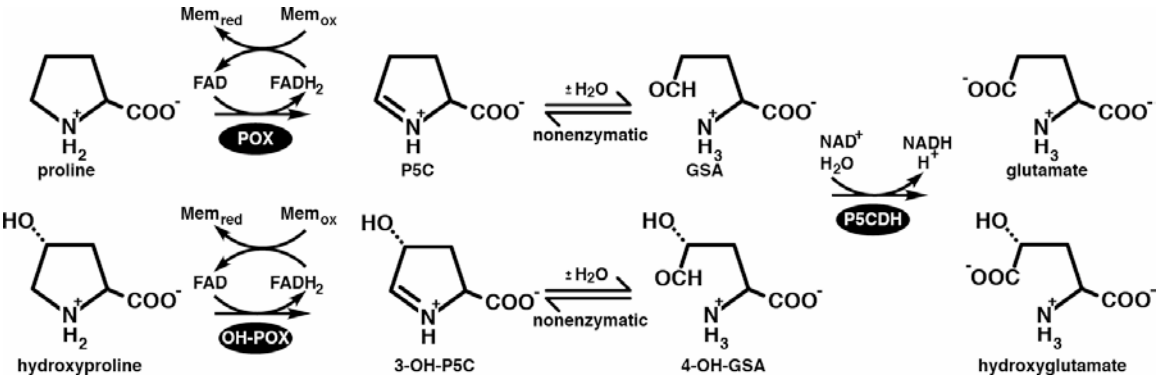


Figure 6.1. The reactions of proline (upper) and hydroxyproline (lower) catabolism in humans.

P5CDH is also the second enzyme of hydroxyproline catabolism in humans (Figure 6.1, lower reactions) (3). Analogous to proline, catabolism of hydroxyproline begins with the oxidation of hydroxyproline (trans-4-hydroxy-L-proline) to Δ^1-pyrroline-3-hydroxy-5-carboxylate (3OH-P5C) by hydroxyproline oxidase (OH-POX, 45 % identical to POX). The nonenzymatic hydrolysis of 3OH-P5C produces

4-hydroxyglutamate semialdehyde (OH-GSA), which is oxidized to 4-erythro-hydroxy-L-glutamate (OH-Glu) by P5CDH. The structural basis for the dual substrate specificity of human P5CDH (HsP5CDH) is unknown.

P5CDH belongs to the ALDH superfamily, a vast group of divergently evolved enzymes that share a common three-dimensional fold and catalytic strategy for oxidizing aldehydes to carboxylic acids. As of a 2002 census,(4) the superfamily comprises several hundred enzymes from all three kingdoms of life, which have been classified into 20 families based on sequence identity. P5CDH belongs to the ALDH4 family and is encoded by the *ALDH4A1* gene. The generally accepted catalytic mechanism of ALDHs begins with nucleophilic attack by a universally conserved Cys residue at the C atom of the aldehyde group of the substrate to form a hemithioacetal (5, 6). Hydride transfer to $NAD(P)^+$ yields a thioacyl intermediate and NAD(P)H. Lastly, hydrolysis of the thioacyl generates the carboxylic acid product. Crystal structures of P5CDH from *Thermus thermophilus* suggest that this mechanism holds for P5CDH (6).

Certain mutations in the *ALDH4A1* gene cause type II hyperprolinemia (HPII), an inherited autosomal recessive disorder characterized by a deficiency in P5CDH activity (7-11). Individuals with HPII have elevated levels of P5C and proline in plasma, urine, and cerebrospinal fluid (12). Elevation of P5C is consistent with the block in the conversion of P5C to glutamate. The increase in proline reflects the conversion of some of the accumulated P5C pool to proline by P5C reductase (11).

HPII is causally linked to neurologic manifestations and is associated with an increased incidence of seizures and possibly mental retardation (13). Hyperprolinemia is also a risk factor for schizophrenia (14, 15). Precisely how a deficiency of P5CDH contributes to

these disorders is uncertain but may reflect the role of proline as a neurotransmitter (16-19). Animal studies have established a connection between high proline levels and neurological dysfunction and suggest that excessive proline disrupts energy metabolism in the brain, leading to mitochondrial dysfunction and oxidative stress (12). Recent studies of a fly model of HPII show that loss of P5CDH activity results in swollen mitochondria and suggest that proper P5CDH activity is necessary for normal mitochondrial homeostasis (20).

Human genetics studies have provided insight into the molecular basis of the HPII. Several nonsynonymous single-nucleotide polymorphisms (SNPs) in the P5CDH gene are associated with the disease (9). Known pathological SNPs include the missense mutation of Ser352 to Leu (S352L) and frameshift mutations that cause premature termination of translation.

Here we provide the first structural analysis of HsP5CDH and investigate the impact of the S352L mutation on the structure and catalytic properties of HsP5CDH. Several crystal structures of HsP5CDH and the 93%-identical mouse P5CDH (MmP5CDH) are reported, including complexes of the latter enzyme with NAD^+ and glutamate. The effect of mutating Ser352 to Leu or Ala is probed using steady-state kinetics assays, NAD^+ binding measurements, and X-ray crystallography. The data provide insight into three-dimensional structural basis of HPII as well as the dual substrate recognition of GSA and OH-GSA.

6.2. Materials and Methods

6.2.1. Subcloning

The genes encoding HsP5CDH (NCBI RefSeq NP_003739) and MmP5CDH (NCBI RefSeq NP_780647.3) were obtained from ATCC. The coding sequences for HsP5CDH residues 18 - 563 and MmP5CDH residues 21 - 562, which omit the mitochondrial leader sequences, were subcloned into pET28a between *NdeI* and *XhoI* restriction sites. The site-directed mutants S352L and S352A of HsP5CDH were created with the QuikChange XL II mutagenesis kit (Agilent Technologies).

6.2.2. Expression and purification of HsP5CDH

HsP5CDH and site-directed mutants of HsP5CDH were expressed in Bl21(DE3)pLysS as follows. Two liters of LB media, supplemented with 50 μg/L kanamycin, were inoculated with 1 % of a starter culture and grown at 37 °C with constant aeration at 250 rpm until the OD_{600} reached 0.6. IPTG was added and protein expression proceeded for about 20 hours at 22 °C and 200 rpm. The cells were harvested by centrifugation at 3500 rpm for 30 minutes in a Sorvall SLC 6000 rotor at 4 °C. The cells were resuspended into 20 mM HEPES, 60 mM NaCl, and 5 % glycerol at pH 8.0, quick frozen in liquid nitrogen, and stored at -80 °C until further purification.

Frozen cells were thawed and broken by sonication. Cell debris and unbroken cells were separated by centrifugation at 16500 rpm for 1 hour in a SS34 rotor. The supernatant was applied to a HisTrap HP column that had been charged with $NiCl_2$ and equilibrated with buffer A (20 mM HEPES, 300 mM NaCl, 5 % glycerol, pH 8.0). The column was washed with buffer A followed by a second wash with buffer A supplemented with 30 mM

imidazole. The protein was eluted with 300 mM imidazole in buffer A. The sample was dialyzed into 50 mM Tris, 0.5 mM EDTA, 0.5 mM THP, and 5 % glycerol at pH 8.0 in preparation for further purification using anion exchange chromatography (HiTrap Q). His-tagged HsP5CDH was collected in the flow-through, while contaminating proteins were retained in the column. The purified protein was dialyzed into pre-crystallization buffer (50 mM Tris, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM THP, 5 % glycerol, pH 8.0). The His-tag was not removed.

6.2.3. Expression and purification of MmP5CDH

MmP5CDH was expressed as described above for HsP5CDH ([IPTG] = 0.5 mM, overnight induction at 22 °C and 200 rpm). The protein was purified using Ni^{2+}- affinity and ion exchange chromatography. Unlike His-tagged HsP5CDH, His-tagged MmP5CDH bound to the anion exchange column when loaded in a buffer of 50 mM Tris, 0.5 mM EDTA, 0.5 mM THP, and 5 % glycerol at pH 7.5, and was eluted with a linear 0 – 1.0 M NaCl gradient. The purified protein was dialyzed into 50 mM Tris, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM THP and 5 % glycerol at pH 7.5 in preparation for crystallization trials. The His-tag was not removed.

6.2.4. Preparation of Se-Met HsP5CDH

The Se-Met derivative of HsP5CDH was prepared using the method of metabolic inhibition (21). An overnight starter culture was grown in LB media at 37 °C. Cells from the starter culture were pelleted and resuspended in minimal media supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 0.4 % glucose as the carbon source. Cells were grown

until the OD_{600} reached 0.5, at which time, the Met biosynthetic pathway was inhibited by adding 1 mg each of lysine, threonine, and phenylalanine, and 0.5 mg each of leucine, isoleucine, valine, and L-selenomethionine per liter of media. The temperature was lowered to 22 °C and the rotation rate was set to 200 rpm. After 30 minutes, IPTG was added (50 μM). The cells were harvested after ~20 hours. Se-Met HsP5CDH was purified as describe above for the native protein.

6.2.5. Crystallization

Se-Met HsP5CDH was crystallized at 4 °C in sitting drops using reservoirs containing 20 - 25 % (w/v) PEG 3350, 0.2 M $(NH_4)_2SO_4$, and 0.1 M HEPES at pH 7.0 - 8.0. The drops were formed by mixing equal volumes of the reservoir and the protein stock solution (6.0 mg/mL protein in 50 mM Tris, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM THP, 5 % glycerol, pH 8.0). The crystals were cryoprotected in 25 % PEG 3350, 0.2 M $(NH_4)_2SO_4$, 0.1 M HEPES, 25 % glycerol at pH 7.5. The space group is $P6_5$ with unit cell parameters of a = 150.7 Å and c = 192.0 Å. The asymmetric unit includes two dimers, which implies 53 % solvent and V_M of 2.6 $Å^3$/Da (22).

Crystals of S352A and S352L were grown at 20 °C using the recipe described above for HsP5CDH except that microseeding was used. Sitting drops were formed by mixing equal volumes of the protein stock (7.0 mg/ml for S352A, 2.0 mg/ml for S352L) and reservoir solutions, and the drops were streak seeded 12 hours later using a horse hair. The seed stock was prepared by diluting crushed S352A crystals by a factor of 5000-10000 with the reservoir. Crystallization of S352L was unsuccessful without microseeding.

MmP5CDH was crystallized in sitting drops (1 μl of protein, 1 μl of reservoir) at 20 °C using reservoir solutions containing 20 - 25 % (w/v) PEG 3350, 0.2 M Li_2SO_4, 0.1 M Bis-Tris, and pH of 6.0 - 7.0. The protein stock solution contained 6 mg/mL His-tagged MmP5CDH in 50 mM Tris, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM THP and 5 % glycerol at pH 7.5. The crystals were cryoprotected with 25 % PEG 3350, 0.2 M Li_2SO_4, 0.1 M Bis-Tris, 25 % glycerol, and pH 6.25. The space group is $P2_12_12_1$ with unit cell lengths of a = 85.2 Å, b = 94.0 Å, and c = 132.4 Å. The asymmetric unit contains one dimer, implying 44 % solvent and V_M of 2.2 $Å^3$/Da. The glutamate complex was formed by soaking crystals for 6 – 8 hours in the cryobuffer without Li_2SO_4 and supplemented with 300 mM sodium glutamate. The NAD^+ complex was obtained similarly using 20 mM NAD^+.

6.2.6. X-ray diffraction data collection

Crystals of Se-Met HsP5CDH were analyzed at NE-CAT beamline 24-ID-C at the Advanced Photon Source using a Quantum 315 detector (Table 5.1). The data set that was used for single-wavelength anomalous diffraction (SAD) phasing was collected at the energy corresponding to the experimentally-determined peak of f'' (λ = 0.979181 Å). This data set consisted of 180 frames collected with an oscillation width of 1.0° per image, detector distance of 375 mm, exposure time of 1.0 s/image, and transmission of 2.5 %. The data were processed to 2.85 Å resolution using HKL2000,(23) with I^+ and I^- treated as nonequivalent reflections during scaling. The crystal was translated, and a second data set was collected at higher transmission for purpose of phase extension and refinement. Denoted as Se-Met-2 in Table 5.1, this data set consisted of 60 frames

collected with an oscillation width of 1.0° per image, detector distance of 350 mm, exposure time of 1.0 s/image, and transmission of 10.0 %. The data were processed using HKL2000 to 2.5 Å resolution, with I^+ and I^- treated as equivalent reflections during scaling.

Data from crystals of S352A and S352L were collected at 24-ID-C using continuous vector scanning, in which the crystal is translated after each image to reduce radiation damage (Table 5.1). The S352A data set consisted of 68 frames with an oscillation width of 1.0° per image, detector distance of 275 mm, exposure time of 1.0 s/image, and a transmission of 50%. The S352L data set comprised 50 frames with an oscillation width of 1.0° per image, detector distance of 320 mm, exposure time of 1.0 s/image, and a transmission of 100.0 %. The data were integrated with XDS(24) and scaled with SCALA(25) via CCP4i (26).

A 1.3 Å resolution data set for the MmP5CDH-sulfate complex was collected at 24-ID-C (Table 5.2). The data set consisted of 120 frames collected with an oscillation width of 1° and detector distance of 150 mm. The data were processed with HKL2000. Data for the MmP5CDH-Glu and MmP5CDH-NAD^+ complexes were collected at beamline 4.2.2 of the Advanced Light Source (Table 5.2). Each data set consisted of 360 frames with an oscillation width of 0.5°, detector distance of 100 mm, and exposure time of 1 s. These data sets were processed with XDS and SCALA.

Table 6.1. Data collection and refinement statistics for HsP5CDH[a]

	Se-Met-1	Se-Met-2	S352A	S352L
Space group	$P6_5$	$P6_5$	$P6_5$	$P6_5$
Unit cell parameters (Å)	a = 150.7	a = 150.7	a = 149.1	a = 150.4
	c = 191.6	c = 192.0	c = 190.7	c = 192.5
Resolution (Å)	50.0 - 2.85	50.0 - 2.50	48.8 - 2.40	48.8 - 2.85
	(2.95 - 2.85)	(2.59 - 2.50)	(2.53 - 2.40)	(3.00 - 2.85)
Total observations	661829	321938	342969	177351
Unique reflections	113595	84960	93350	55880
Multiplicity	5.8 (5.8)	3.8 (3.8)	3.7 (3.2)	3.2 (3.1)
R_{merge}[b]	0.100 (0.524)	0.075	0.072	0.101
		(0.539)	(0.450)	(0.626)
R_{meas}[b]			0.083	0.119
			(0.539)	(0.742)
R_{pim}[b]			0.041	0.062
			(0.291)	(0.390)
$\langle I/\sigma(I)\rangle$	22.9 (4.4)	17.0 (2.5)	10.7 (2.3)	8.6 (2.2)
Completeness (%)	100.0 (100.0)	99.7 (100.0)	99.7 (99.5)	97.5 (99.0)
R_{work} / R_{free}[c]		0.199 /	0.210 /	0.209 /
		0.237	0.251	0.264
Number of atoms		16032	15844	13924
Protein residues		2162	2151	1982
Water molecules		0	167	0
RMSD bond lengths (Å)		0.009	0.008	0.011
RMSD bond angles (°)		1.18	1.16	1.23
Ramachandran plot[d]				
Favored (residues)		2099	2059	1846
Allowed (residues)		55	73	72
Outliers (residues)		0	1	6
Average *B*-factors				
Protein (Å^2)		62	49	56
Water (Å^2)			34	
Coordinate error (Å)[e]		0.70	0.72	0.90
PDB code		3V9G	3V9H	3V9I

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]Definitions of R_{merge}, R_{meas}, and R_{pim} can be found in Weiss.(27)
[c]A common test set of reflections was used for all refinements (2.4 %).
[d]The Ramachandran plot was generated with RAMPAGE.(28)
[e]Maximum - likelihood based coordinate error from PHENIX.

Table 6.2. Data collection and refinement statistics for MmP5CDH[a]

	Sulfate	Glutamate	NAD^+
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters (Å)	a = 85.2	a = 84.8	a = 84.9
	b = 94.0	b = 93.9	b = 94.0
	c = 132.4	c = 132.2	c = 132.4
Resolution (Å)	50.0 - 1.30 (1.32 - 1.30)	47.0 - 1.50 (1.58 - 1.50)	47.0 - 1.50 (1.58 - 1.50)
Total observations	914888	1222700	1221816
Unique reflections	247337	168017	168338
Multiplicity	3.7 (3.6)	7.3 (6.7)	7.3 (6.8)
R_{merge}[b]	0.055 (0.502)	0.066 (0.418)	0.061 (0.399)
R_{meas}[b]		0.071 (0.454)	0.065 (0.432)
R_{pim}[b]		0.026 (0.174)	0.024 (0.164)
$\langle I/\sigma(I)\rangle$	26.6 (2.4)	25.3 (4.8)	25.3 (4.9)
Completeness (%)	94.9 (86.0)	99.5 (96.5)	99.6 (97.6)
R_{work} / R_{free}[c]	0.132 / 0.160	0.151 / 0.169	0.158 / 0.176
Number of atoms	9479	9233	9246
Protein residues	1091	1081	1081
Water molecules	1035	896	872
Active site ligand atoms	10	20	46
RMSD bond lengths (Å)	0.005	0.006	0.006
RMSD bond angles (°)	1.05	1.09	1.11
Ramachandran plot[d]			
Favored (%)	98.3	98.2	98.0
Allowed (%)	1.7	1.8	2.0
Outliers (%)	0.0	0.0	0.0
Average *B*-factors			
Protein (Å^2)	12	10	11
Water (Å^2)	25	20	21
Active site ligand (Å^2)	17	11	14
Coordinate error (Å)[e]	0.30	0.29	0.31
PDB code	3V9J	3V9K	3V9L

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]Definitions of R_{merge}, R_{meas}, and R_{pim} can be found in Weiss.(27)
[c]A common test set of reflections was used for all refinements (5.0 %).
[d]The Ramachandran plot was generated with RAMPAGE.(28)
[e]Maximum - likelihood based coordinate error from PHENIX.

6.2.7. Phasing and refinement

The structure of HsP5CDH was determined using SAD phasing based on data collected from crystals of Se-Met HsP5CDH. The phasing potential of each data set was analyzed with the HKL2MAP(29) interface to the SHELXC/D/E programs (30-32). Promising data sets were input to PHENIX AutoSol for SAD phasing, density modification, and automated building calculations (33). For the best data set, 30 of the expected 36 Se sites (i.e., four HsP5CDH molecules per asymmetric unit) were identified, which resulted in a figure of merit of 0.39 for reflections to 2.85 Å resolution. Density modification, which included 4-fold non-crystallographic symmetry (NCS) averaging, increased the figure of merit to 0.69. The model from automated building included 1790 residues (10194 atoms) and had an *R-factor* of 0.48 and map-model correlation coefficient of 0.59. PHENIX AutoBuild(33) was used for phase extension to 2.5 Å resolution and additional model building. The resulting model included 1751 protein residues and 505 water molecules, with R_{cryst} of 0.32, R_{free} of 0.37, and map-model correlation coefficient of 0.72.

The protein part of the model from automated building was used as the starting point for several rounds of manual building in COOT(34) and simulated annealing refinement against the 2.5 Å resolution data set in PHENIX (35). The B-factor model used during refinement consisted of an isotropic B-factor for each non-hydrogen atom plus one TLS group per chain. Four-fold NCS restraints were used during refinement. The final structure served as the starting point for NCS-restrained refinement of the S352A and S352L structures. Refinement statistics are listed in Table 5.1.

The 1.3-Å structure of MmP5CDH complexed with sulfate was determined using molecular replacement as implemented in the programs MOLREP (36) and PHASER (37) with the HsP5CDH dimer serving as the search model. Several rounds of model building in COOT and refinement in PHENIX were performed. Initially, the *B*-factor model consisted of an isotropic *B*-factor for each non-hydrogen atom and one TLS group per protein chain, but anisotropic *B*-factors were used during the final few rounds of refinement. The refined MmP5CDH-sulfate structure was the starting point for refinements of the 1.5 Å resolution glutamate and NAD^+ complexes. For these two structures, the *B*-factor model consisted of an isotropic *B*-factor for each non-hydrogen atom and one TLS group per protein chain. Refinement statistics are listed in Table 5.2.

6.2.8. Analytical ultracentrifugation

The quaternary structure of HsP5CDH in solution was analyzed using equilibrium analytical ultracentrifugation. Data were acquired at 20 °C using a Beckman XL-I Optima analytical ultracentrifuge equipped with an An50Ti rotor. Prior to centrifugation, a protein sample having concentration of 0.8 mg/mL was dialyzed into a buffer containing 50 mM Tris, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM THP, and 5 % glycerol at pH 8.0. Absorbance data (λ = 278 nm) were collected at three protein concentrations (0.2 mg/mL, 0.4 mg/mL, 0.8 mg/mL) and three rotor speeds (4000 rpm, 8000 rpm, 12000 rpm). The nine sets of data were fit globally to a single-species model (Eq. 9 of Lebowitz *et al.*(38)). A solvent density of 1.02 g/cm^3 and partial specific volume of 0.74 cm^3/g were used in the calculations. The 68 % confidence interval for the molar mass was estimated by analyzing the χ^2 surface using the F-statistic.

6.2.9. Isothermal titration calorimetry

The binding of NAD^+ to HsP5CDH, S352A, and S352L was studied at 25°C using a VP-ITC calorimeter (MicroCal, LLC). Prior to the titration, the protein sample was dialyzed in 50 mM Tris, 50 mM NaCl, 0.5 mM THP, 0.5 mM EDTA, and 5 % glycerol at pH 8.0. The protein concentration was estimated using absorbance with a theoretical extinction coefficient of 1.1 A_{280}/mg. NAD^+ was dissolved in the dialysate, and the pH was adjusted to that of the dialyzed protein sample. The concentration of NAD^+ was estimated by absorbance using an extinction coefficient of 16.9 A_{259}/mM. The protein and NAD^+ solutions were loaded into the sample cell and buret, respectively. Following thermal equilibration, 20 μl aliquots of the titrant were injected into the sample cell at 240-s intervals with a stirring speed of 250 rpm. A pre-injection of 4 μl was used and discarded during fitting. Fitting of the data to a single-site binding model was performed using Origin software.

6.2.10. Kinetic characterization

P5CDH activity was measured at 20 °C by monitoring the production of NADH at 340 nm. The assay buffer contained 0.1 M sodium phosphate (pH 7.0) and 1 mM EDTA. The enzyme concentration was 11 μg/ml (0.16 μM). Specific activity is expressed as μmoles of NADH produced per minute per mg of enzyme. Initial velocity data were collected using NAD^+ concentrations from 1 to 1500 μM at different fixed concentrations of a 50/50 DL-P5C mixture with L-P5C concentrations from 10 to 500 μM. Data were globally fitted to a kinetic scheme for the Theorell-Chance mechanism as shown in Supplemental Information. The fitted rate constants and kinetic parameters for HsP5CDH are reported in

Table 5.3.

Table 5.3. Kinetic constants for HsP5CDH determined from global fitting

Parameter	Best fit value	Lower Bound	Upper Bound
k_1 (k_{cat}/K_m, NAD)	98.7 $mM^{-1}s^{-1}$	74.7 $mM^{-1}s^{-1}$	144 $mM^{-1}s^{-1}$
k_{-1}	0.473 s^{-1}	0.294 s^{-1}	0.813 s^{-1}
k_2 (k_{cat}/K_m, P5C)	316 $mM^{-1}s^{-1}$	253 $mM^{-1}s^{-1}$	406 $mM^{-1}s^{-1}$
k_{-2}[a]	$\leq$ 300 s^{-1}	-	-
k_3[b]	$\geq$ 500 s^{-1}	-	-
k_{-3}[c]	0	-	-
k_4[b]	$\geq$ 500 s^{-1}	-	-
k_{-4}[d]	0	-	-
k_5 (k_{cat})	10 s^{-1}	9.1 s^{-1}	11 s^{-1}
k_{-5}[e]	0	-	-
k_6[f]	81.7 $mM^{-1}s^{-1}$	-	-
k_{-6}	9.2 s^{-1}	5.1	14.8
$K_I = k_{-6}/k_6$	112 µM	49[g]	181

[a] k_{-2} has little effect on the initial velocity progress curves and was held fixed to a value of 22 s^{-1} but could be $\leq$ 300 s^{-1}.
[b] k_3 and k_4 have little effect on the initial velocity progress curves beyond 500 s^{-1} and were fixed at 3290 s^{-1} and 569s^{-1} so that these steps were non rate limiting according to the Theorell-Chance mechanism (39).
[c] k_{-3} was fixed to zero based on the previous observation of no observable turnover in the reverse direction.
[d] k_{-4} was fixed to zero according to the Theorell-Chance mechanism.
[e] k_{-5} was fixed to zero based on the observation of a $K_d \geq 1$ mM (data not shown). Once k_{-5} is 10-fold below k_5 it has no effect on the progress curves.
[f] k_6 was held fixed to obtain an error on the K_I; this step is at equilibrium and thus the fixed rate constant shown here is only the best fit value and should not be considered as a constrained estimate of this rate constant.
[g] The percent error from the bounds of k_{-6} was used to estimate a rigorous boundary for the K_I.

6.2.11. Tryptophan fluorescence quenching

An estimate of NAD^+ binding to HsP5CDH was performed to better constrain the k_{-1}/k_1 ratio in the global fitting analysis shown in Figure 6.2. HsP5CDH (1 µM) in 0.1 M phosphate buffer (pH 7.0, 1 mM EDTA) was excited at 295 nm and the fluorescence emission maximum at 330 nm was recorded. Quenching of tryptophan fluorescence by NAD^+ (0 - 30 µM) was monitored at 330 nm and the emission was normalized to represent fractional quenching. Increasing amounts of NAD^+ were added to the HsP5CDH solution

allowing for a 2 min equilibration time after each addition of NAD^+. The tryptophan fluorescence quenching data from the NAD^+ titration was then simulated with Kinetic Global explorer software and fitted globally along with the kinetic data to better constrain the K_d for NAD^+ binding to HsP5CDH (Figure 6.2).

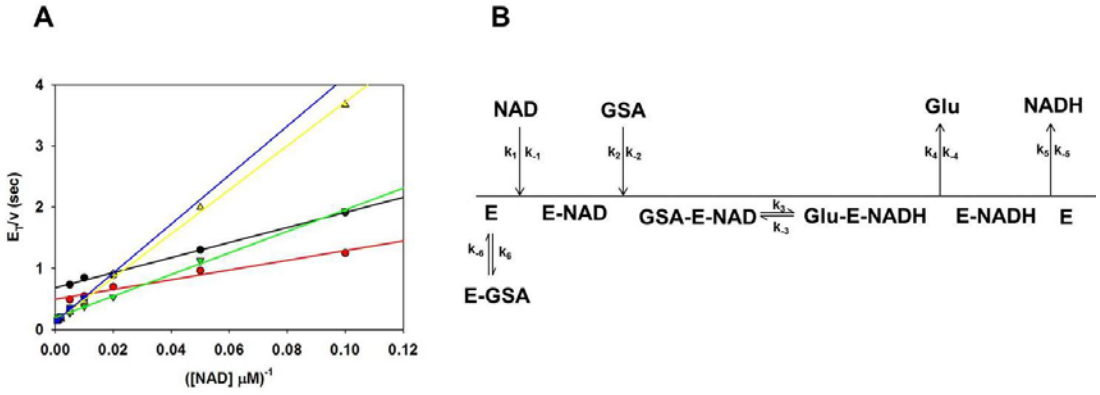


Figure 6.2. (A) Lineweaver-Burk analysis of initial velocity data for HsP5CDH collected with NAD^+ concentrations in the range 10 - 1500 µM in different fixed concentrations of a 50/50 mixture of DL-P5C with L-P5C concentrations as follows: 10 (black), 15 (red), 50 (green), 150 (yellow), 500 µM (blue). Solid lines are the best-fit line to the individual data sets with the corresponding colored data points. (B) Kinetic scheme suggested by previously published data (40) but with an additional step for substrate inhibition as observed in panel A at higher P5C concentrations (150 and 500 µM). The original kinetic mechanism established for HsP5CDH was a Theorell-Chance mechanism, which is a limiting case of the scheme shown above where k_3 and k_4 are much faster than k_5.

6.3. Results

6.3.1. Overall fold

The structure of HsP5CDH was determined at 2.5 Å resolution using single-wavelength anomalous diffraction (SAD) phasing (Table 6.1). HsP5CDH exhibits the classic ALDH fold (Figure 6.3a), which consists of an N-terminal

NAD^+-binding domain (residues 23-182, 199-315, 525-542), C-terminal catalytic domain (residues 316-524), and oligomerization domain (residues 183-198, 543-562). At the core of the NAD^+-binding domain is an open α/β substructure that resembles the Rossmann dinucleotide-binding fold (residues 203 - 314). As first described for ALDH3, the Rossmann fold of ALDH differs from the classic one in that the final helix and strand of the classic Rossmann fold are absent in ALDHs.(41) The catalytic domain also exhibits an open α/β fold and features a twisted 7-stranded β-sheet with all but one strand in parallel. This domain furnishes the essential cysteine nucleophile (Cys348) and several residues that bind GSA, including Ser349. The oligomerization domain is a bipartite structure consisting of β-hairpin protruding from the NAD^+-binding domain and the C-terminal ~20 residues of the polypeptide chain. The latter part forms a β-strand followed by an extended section. The hairpin and C-terminal strand combine to form a two-stranded antiparallel β-sheet that resembles a flap.

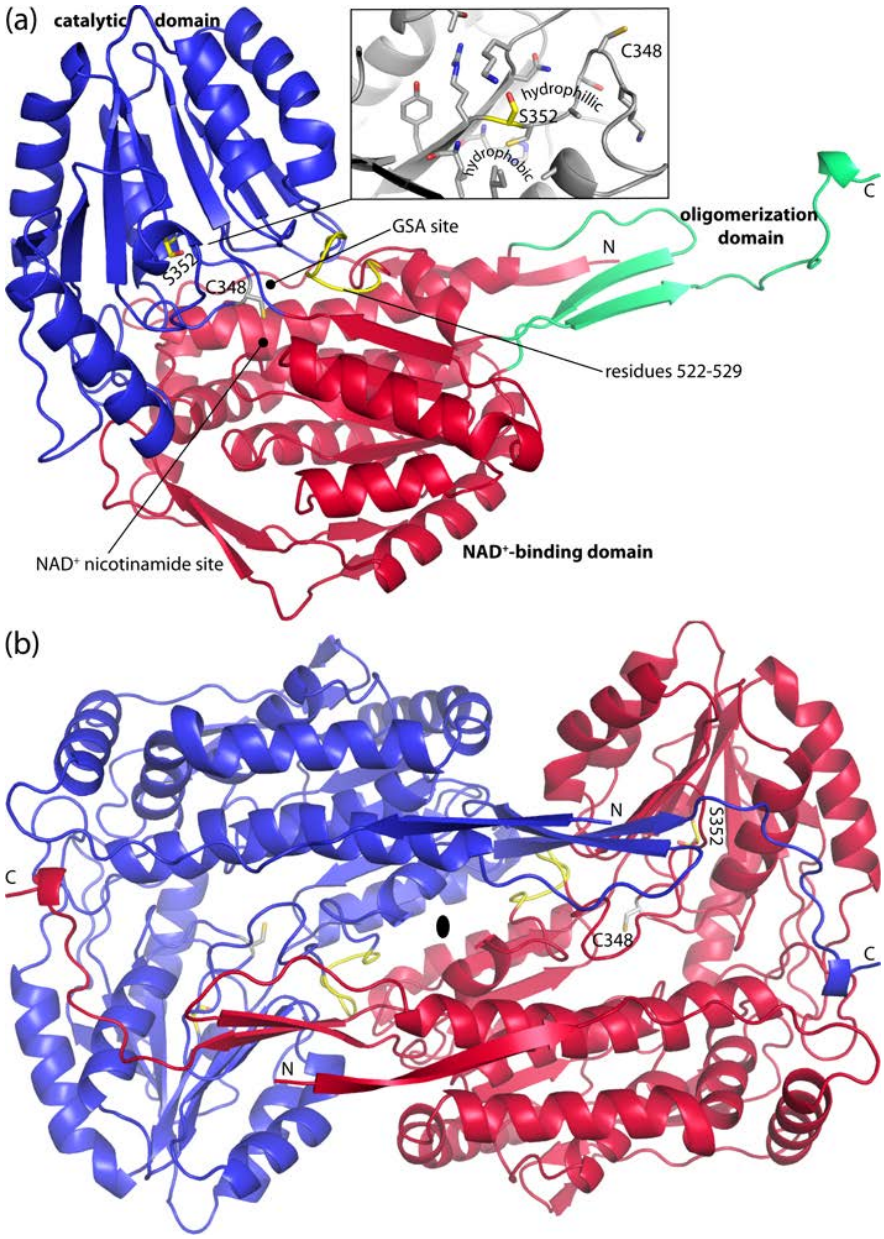


Fig 6.3. Protomer and dimer structure of HsP5CDH. (a) Ribbon drawing of the protomer. The NAD^+-binding, catalytic, and oligomerization domains are colored red, blue, and green, respectively. Ser352 and residues 522-529 are colored yellow. The side chain of catalytic Cys348 is shown. Inset: close-up view of the environment around Ser352. (b) The HsP5CDH dimer. The two chains are colored red and blue. This figure and others were created with PyMOL.(42)

6.3.2. Oligomeric state and quaternary structure

The oligomeric state of HsP5CDH in solution was determined using equilibrium analytical ultracentrifugation (Figure 6.4). The data from nine sedimentation experiments corresponding to three loading concentrations and three rotor speeds were fitted globally to a single-species model. The molar mass is estimated to be 122 kDa (120 – 125 kDa), which is within 2 % of the value predicted for the dimer (124 kDa).

Inspection of the crystal lattice revealed a two-body assembly that represents the dimer formed in solution (Figure 6.3b). This dimeric assembly is also found in the crystal lattices of MmP5CDH and other ALDHs (41, 43, 44). The dimer is a domain-swapped assembly in which the oligomerization domain of one protomer engages the other protomer. The domain swapping results in the formation of an intermolecular β-sheet involving the C-terminal strand of the oligomerization domain of one protomer and the final strand of the catalytic domain of the other protomer. Also, the extended region at the C-terminus of one protomer wraps over the catalytic domain of the other protomer forming several hydrogen bonds. Another major part of the dimer is located on the face opposite to that of the intermolecular β-sheet and consists of extensive nonpolar and electrostatic interactions involving the 290s helices of the two protomers (residues 288-304). In total, the dimer interface buries 4300 $Å^2$ of surface area.

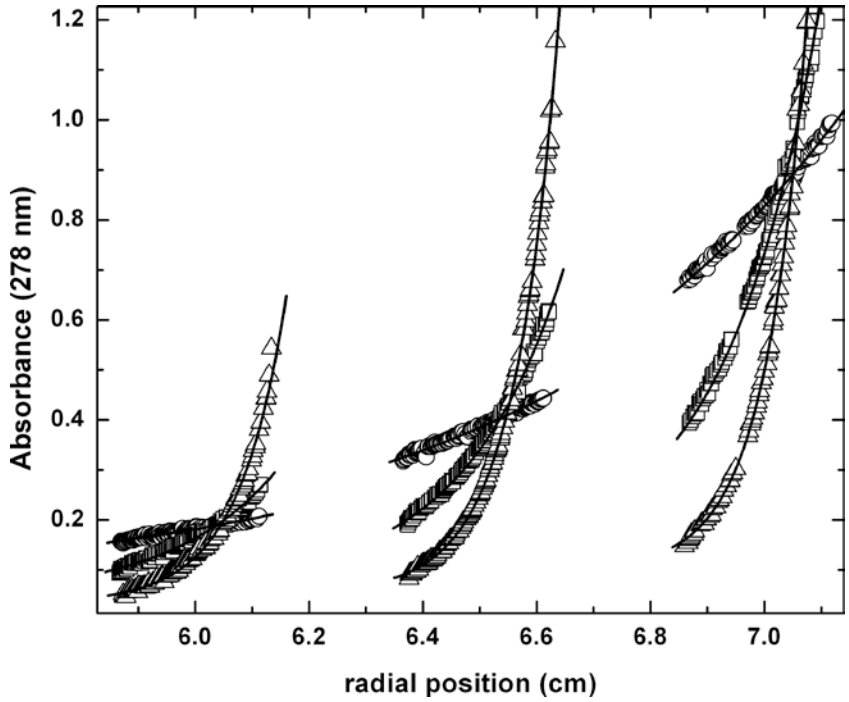


Figure 6.4. Analytical ultracentrifugation data for HsP5CDH. Data were acquired at 4000 (circles), 8000 (squares), and 12000 (triangles) rpm at three nominal loading concentrations. The curves represent the best least-squares fit to an ideal single-species model. The molar mass obtained from global fitting is 122 kDa (120 – 125 kDa), which is within 2 % of the value predicted for the dimer (124 kDa).

6.3.3. Binding of ligands to the active site.

MmP5CDH was used for structure determination of enzyme-ligand complexes because of the exceptional crystallographic resolution (Table 5.2). The overall conformations of MmP5CDH and HsP5CDH are essentially identical except for a deviation at the N-terminus, which is due to different crystal lattice interactions (Figure 6.5). The root mean square deviation between the two structures is 0.3 Å for 533 residues. The sequence identity is 100 % for active site residues. Moreover, the conformations of active site residues in the two enzymes are essentially identical. Thus MmP5CDH is an excellent surrogate for HsP5CDH.

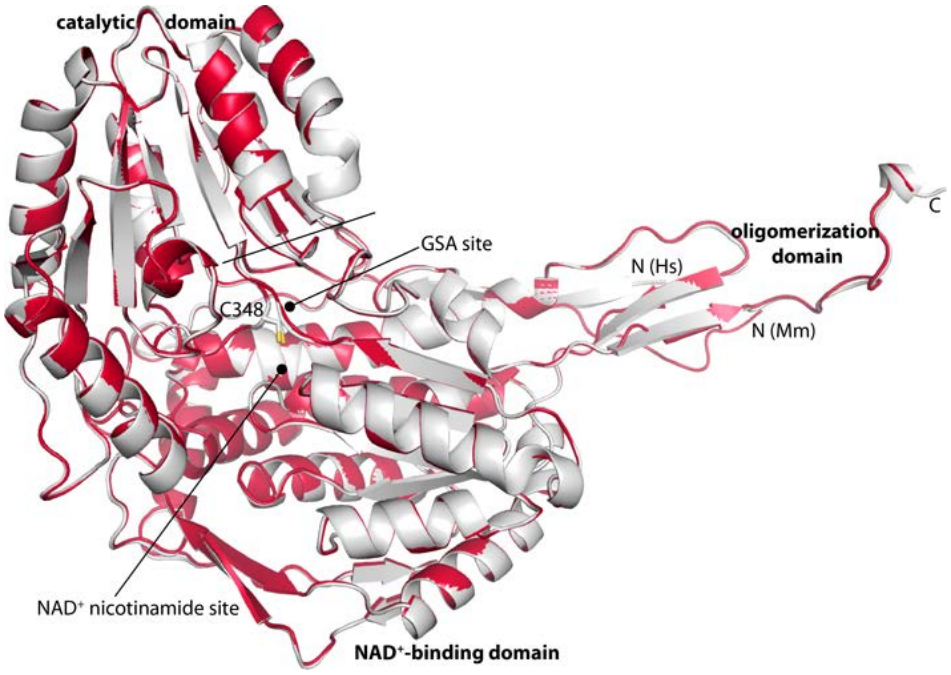


Figure 6.5. Superposition of HsP5CDH (white) and MmP5CDH (red).

The 1.3 Å resolution structure of MmP5CDH was determined from crystals grown in the presence of 0.2 M Li_2SO_4. Electron density maps clearly indicate a sulfate ion bound in the active site (Figure 6.6). The ion interacts with Gly512, Ser513, and Ser349. We note that the sulfate site corresponds to the binding site for the phosphate group of glyceraldehyde-3-phosphate of non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase, which also belongs to the ALDH superfamily (45).

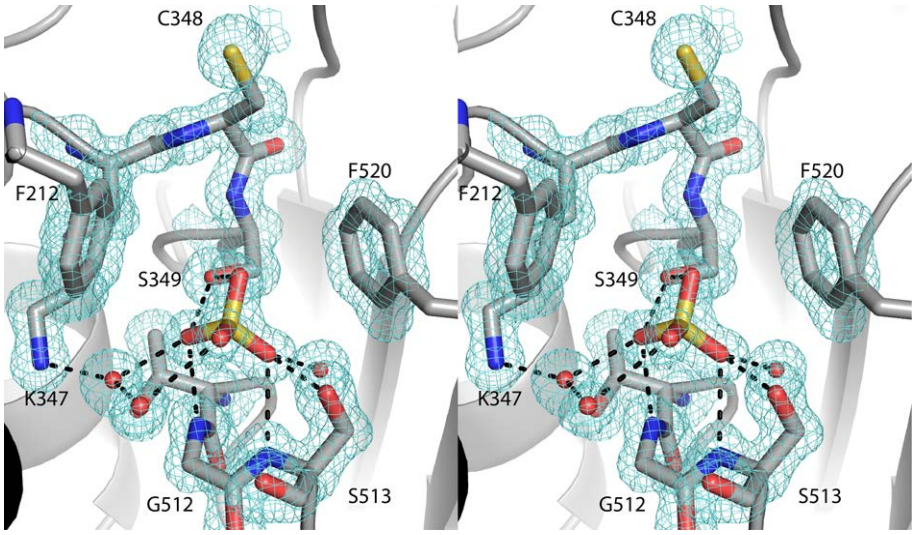


Figure 6.6. Electron density for the active site of the MmP5CDH-sulfate complex (stereographic view). The cage represents a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 3.0 σ.

The structure of MmP5CDH complexed with the product glutamate was solved at 1.5 Å resolution using data from a crystal soaked in ~300 mM glutamate. Electron density maps clearly showed that glutamate had displaced the bound sulfate ion (Figure 6.7a). Glutamate binds with its α-carboxylate in the sulfate site and its side chain extending toward catalytic Cys348. The amino and α-carboxylate groups of glutamate form hydrogen bonds with Ser349, Gly512, and Ser513. The aliphatic part of the glutamate side chain packs between Phe512 and Phe520. One of the oxygen atoms of the γ-carboxylate interacts with Asn211 and the main chain of Cys348. This atom likely represents the O atom of the aldehyde group of GSA. The other O atom of the carboxylate points into a solvent-filled cavity and represents the O atom derived from the attack of water on the thioacyl intermediate. This mode of product binding is similar to that observed for *Thermus thermophilus* (TtP5CDH, 30 % identical to HsP5CDH) (6).

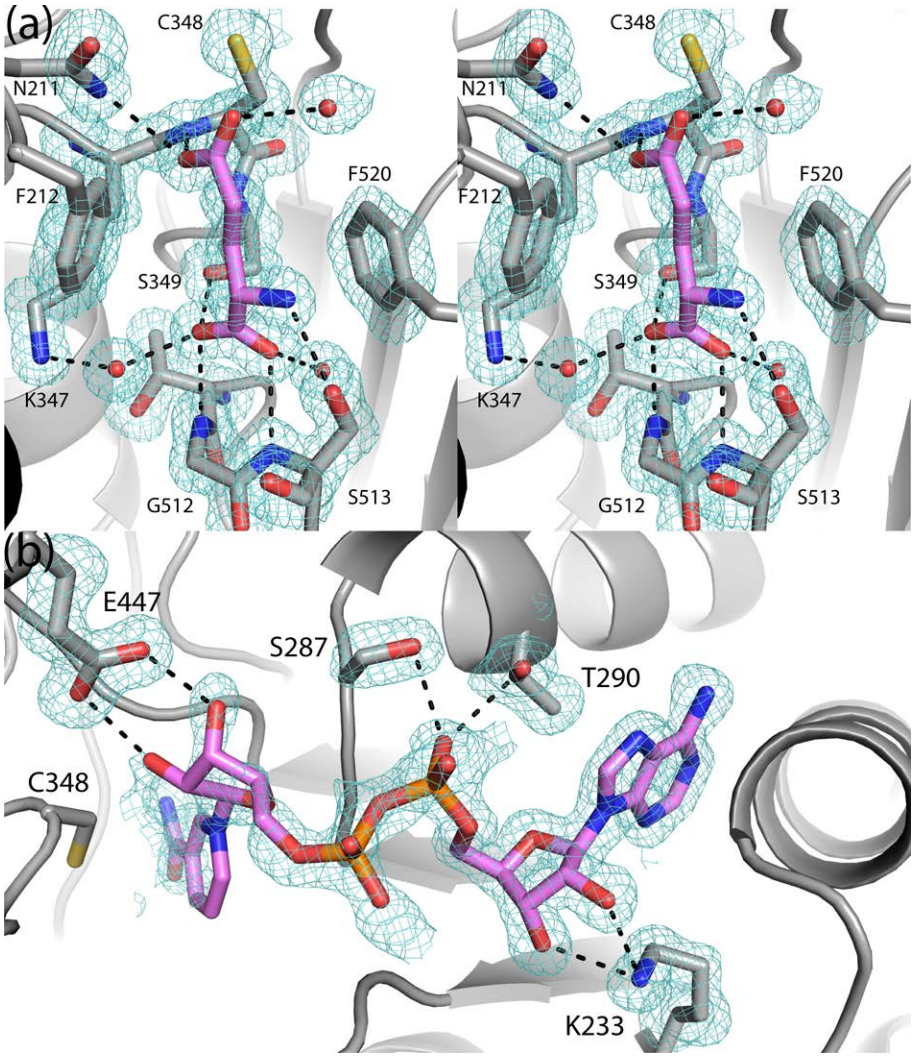


Figure 6.7. Electron density and interactions for (a) glutamate (stereographic view) and (b) NAD^+ bound to MmP5CDH. The cages represent simulated annealing σ_A-weighted F_o - F_c omit maps contoured at 3.0 σ. The NMN part of NAD^+ in panel b was modeled based on other ALDH structures and has zero occupancy in the structure deposited in the PDB.

A 1.5 Å resolution structure of MmP5CDH complexed with NAD^+ was determined. Electron density for the AMP moiety is strong, while the density for NMN is weaker (Figure 6.7b). Nevertheless, the weak density is consistent with the hydride transfer NAD^+ conformation that is seen in other ALDHs, including TtP5CDH. We note that conformational disorder is common for ALDHs (5, 6). Based on these observations, the

NMN and AMP moieties were modeled with occupancy values of 0.0 and 1.0, respectively. The adenine ring is wedged between the last two helices of the Rossmann dinucleotide-binding fold. Lys233 engages the adenine ribose, while Ser287 and Thr290 interact with the pyrophosphate. The interactions involving Ser287 and Lys233 are also seen in TtP5CDH. The NMN ribose is predicted to form hydrogen bonds with Glu447. This interaction is universally conserved in ALDHs and is presumably essential for binding NAD^+.

6.3.4. The structural context of G521fs(+1)

The frameshift mutation G521fs(+1) consists of an insertion of T following nucleotide 1563 in exon 16 of the *ALDH4A1* gene (9). This mutation alters the sequence of residues 522-529 (yellow in Figure 6.3a) from GARASGTN to WGPSLWNQ and introduces a stop codon that truncates translation after residue 529. Residue 529 is located in the inter-domain linker that connects the end of the catalytic domain to the strand of the oligomerization domain. The truncated protein thus lacks the final β-strand and C-terminal extension. These elements form a major part of the dimer interface, suggesting that they are essential for proper dimer formation. Since there are no examples of functional monomeric ALDHs, the truncated protein is probably nonfunctional. The lack of proper dimer formation most likely explains why P5CDH-deficient yeast expressing G521fs(+1) failed to grow on proline and had no detectable P5CDH activity (9).

6.3.5. The structural context of S352L

Ser352 is located at one end of the catalytic loop (Figure 6.3a inset). This loop contains catalytic Cys348 and Ser349, a residue that interacts with the aldehyde substrate (Figure 6.7a). The MmP5CDH structure provides a high resolution view of the three-dimensional context of Ser352 (Figure 6.8). We note that the conformations of the catalytic loops in HsP5CDH and MmP5CDH are almost identical (Figure 6.9a).

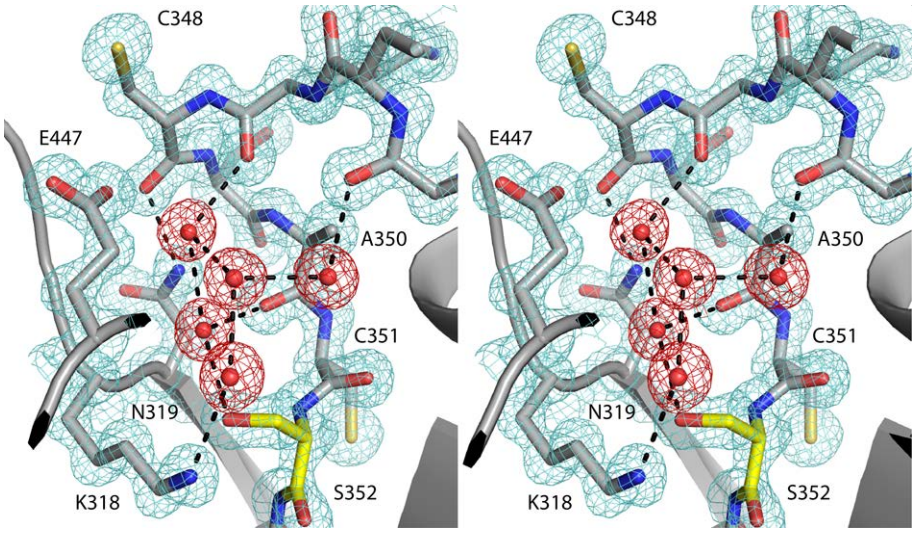


Figure 6.8. High resolution (1.3 Å) view of the catalytic loop of MmP5CDH (stereographic view). The cage represents a $2F_o$ - F_c map contoured at 1.5 σ. Protein density is colored aquamarine. Water density is colored red. Ser352 is colored yellow.

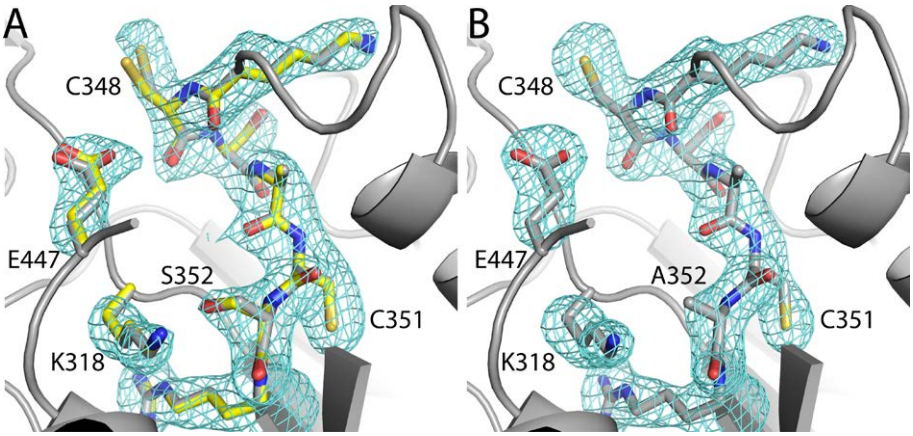


Figure 6.9. Electron density for the catalytic loops of (a) HsP5CDH and (b) S352A. (a) Superposition of HsP5CDH (gray) and MmP5CDH (yellow). The cage represents a simulated annealing σ_A-weighted F_o - F_c omit map for HsP5CDH contoured at 3.0 σ. Note that the two enzymes have almost identical active site conformations. (b) Catalytic loop of S352A shown in the same orientation as in panel a. The cage represents a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 3.0 σ. Note that the conformation of the active site of S352A is nearly identical to those of HsP5CDH and MmP5CDH.

Ser352 sits at the bottom of a water-filled pocket (Figure 6.8). The adjacent residue, Cys351, is turned away from this hydrophilic space, and its side chain is tucked into a hydrophobic region. Catalytic Cys348 and Glu447 sit at the top of the water-filled pocket where it flows into the crevice between the catalytic and NAD^+-binding domains. Ser352, Lys318, Asn319 and backbone carbonyls of the catalytic loop project into the pocket and form hydrogen bonds with water molecules, thus creating an electrostatic network that appears to stabilize the conformation of the catalytic loop (Figure 6.8). This region appears to be conserved among ALDHs. For example, the constellation of water molecules in the pocket is also found in human mitochondrial aldehyde dehydrogenase (PDB code 1O04) (5).

6.3.6. Characterization of the S352L and S352A mutants of HsP5CDH

The mutation of Ser352 to Leu in HsP5CDH abolishes catalytic activity. The mutant enzyme was inactive under all conditions tested. For example, the activity of S352L was below detection in an assay containing 100 μg/mL of enzyme, 355 μM NAD^+, and 200 μM P5C (Figure 6.10). For reference, an identical assay performed with HsP5CDH present at 10-fold lower concentration exhibited substrate depletion within 1 minute (Figure 6.10). The absence of measurable catalytic activity for recombinant S352L is in agreement with previous studies showing that P5CDH-deficient yeast expressing S352L failed to grow on proline and had no P5CDH activity (9).

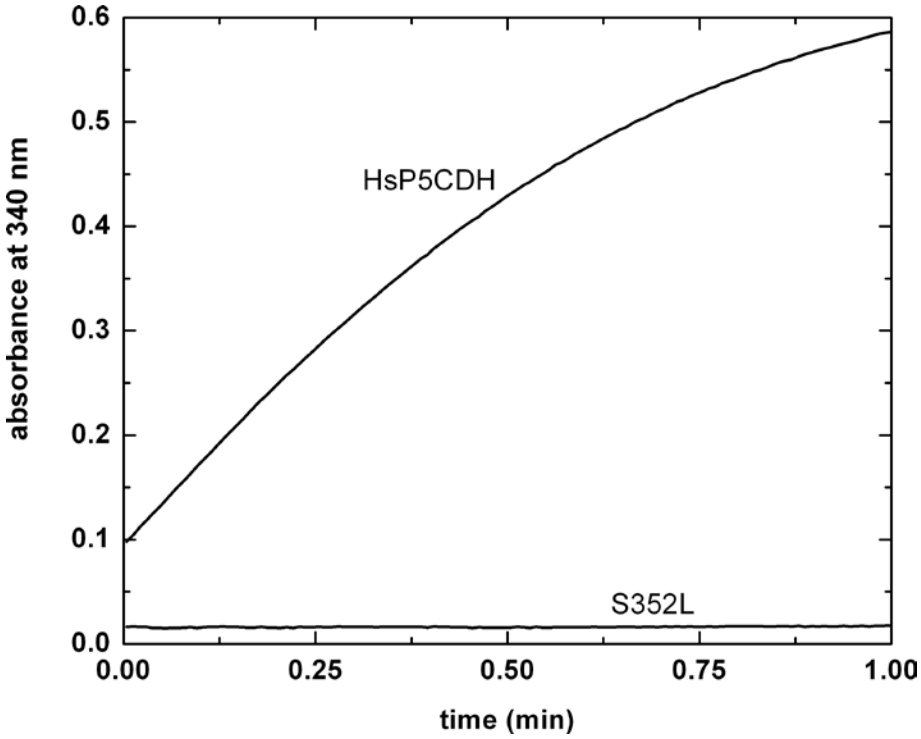


Figure 6.10. Mutation of Ser352 to Leu abolishes catalytic activity. Progress curves for HsP5CDH and S352L. The NAD^+ and P5C concentrations are 350 μM and 200 μM, respectively. The concentration of S352L is 10-fold higher than that of HsP5CDH.

The functional consequences of mutating Ser352 was probed further using mutagenesis, kinetics, and isothermal titration calorimetry. Mutation of Ser to Leu has the dual effect of removing a hydroxyl group and adding a branched aliphatic side chain. The mutant enzyme S352A was created to examine the contribution of the former effect.

In contrast to S352L, the mutant enzyme S352A exhibits catalytic activity that is commensurate with that of HsP5CDH. The kinetic properties of HsP5CDH and S352A were thus more closely compared. Competitive type inhibition by L-P5C at concentrations > 50 µM was observed for HsP5CDH and S352A in assays varying NAD^+ as previously reported but not characterized (40) (Figure 6.2). In order to obtain kinetic constants and a substrate inhibition constant for L-P5C (K_I), reaction progress curves at various NAD^+ concentrations using different fixed P5C concentrations were globally fitted to a Theorell-Chance mechanism (Figure 6.11). Global fitting using KinTek Global Kinetic Explorer (46) yielded k_{cat} values of 10 and 5 s^{-1} for HsP5CDH and S352A, respectively, with the NADH dissociation step being rate limiting for the overall reaction (Table 5.3). The k_{cat}/K_m for HsP5CDH and S352A are 98.7 and 102 $mM^{-1}s^{-1}$, respectively. The K_m values for NAD^+ and L-P5C are thus 100 and 32 µM for HsP5CDH, and 49 and 51 µM for S352A respectively. The similar kinetic parameters for HsP5CDH and S352A suggest that the catalytic defect of S352L is due primarily to the addition of the nonpolar Leu side chain rather than the removal of the hydroxyl group of Ser352.

Isothermal titration calorimetry (ITC) was used to study the binding of NAD^+ to HsP5CDH, S352A, and S352L (Figure 6.12). The association of NAD^+ with HsP5CDH is exothermic (Figure 6.6a). Global fitting of data from three replicate titrations (Figure 6.6b) yielded an apparent enthalpy change of -18 ± 1 kcal/mol and dissociation constant

(K_d) of 15 ± 1 μM (Table 5.4). The binding of NAD^+ to S352A is likewise exothermic (Figure 6.12c), and the binding parameters are similar to those of HsP5CDH (ΔH = -18 ± 2 kcal/mol, K_d, = 10 ± 2 μM). In contrast, the injection heats for S352L are small and nearly constant during the titration (Figure 6.12e, 6.12f), implying that S352L has negligible affinity for NAD^+. This result implies that the catalytic defect of S352L is due, in part, to a lack of NAD^+ binding. Furthermore, the ITC results show that the defect in NAD^+ binding of S352L results primarily from the introduction of the nonpolar Leu side chain rather than removal of the hydroxyl group of Ser352.

The structure of S352A was determined at 2.4 Å resolution (Table 6.1). The structure is essentially identical to that of HsP5CDH. In particular, the catalytic loop and constellation of residues that bind GSA and NAD^+ have the same conformations as in the native enzyme (Figure 6.9b). The S352A structure, in agreement with the ITC and kinetic data, shows that the catalytic defect of S352L is due primarily to the introduction of the nonpolar Leu side chain.

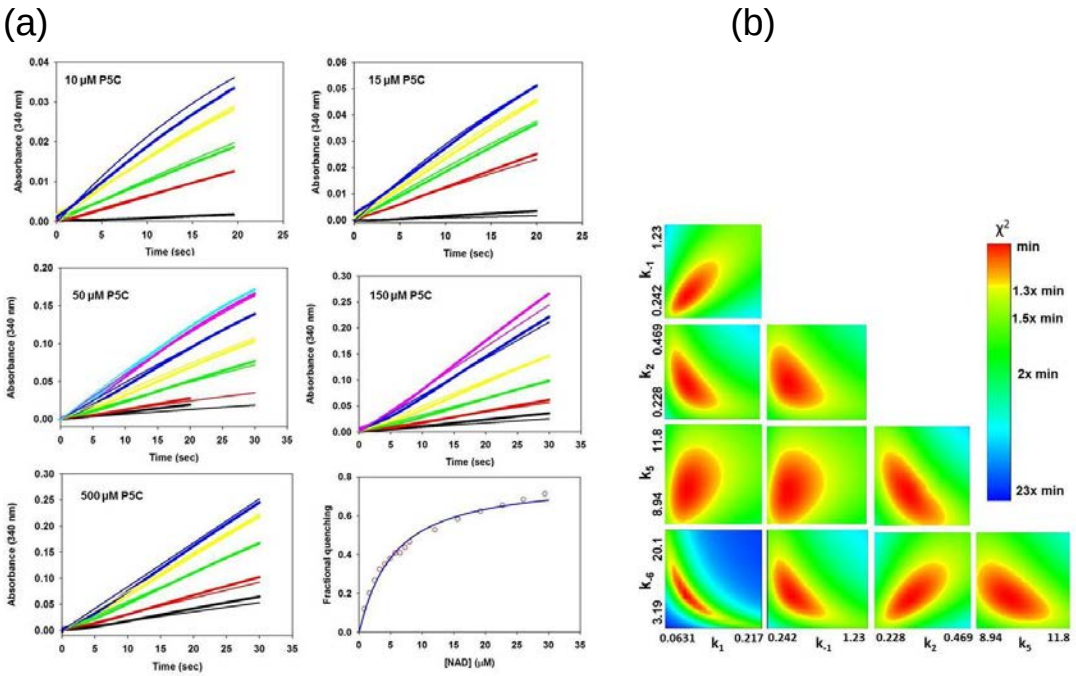


Figure 6.11 (a) Initial velocity progress curves for HsP5CDH at various NAD^+ concentrations (1-1500 µM) and different fixed L-P5C concentrations followed at 340 nm. Data were globally fitted to the simulated mechanism shown in Supplemental Figure 6B using KinTek Global Kinetic Explorer (46). The bottom right graph shows a titration of HsP5CDH with NAD^+ monitored by tryptophan fluorescence quenching, which was also included in the global fitting analysis to help constrain k_1 and k_{-1}. The rate constants for the chemical step (k_3) and glutamate dissociation step (k_4) were held fixed at values well above the NADH dissociation step (k_5) in accordance with the Theorell-Chance mechanism (39). (b) FitSpace (47) contour plots of the global fitting showing how variation in the fitted parameters affects the χ^2 value.

6.3.7. Crystal structure of the S352L mutant of HsP5CDH

The structure of S352L was determined at 2.85 Å resolution (Table 6.1). The mutant enzyme exhibits the same overall fold and dimeric structure as HsP5CDH. The active site of S352L, however, is highly perturbed. In particular, the catalytic loop adopts a non-native conformation (Figure 6.13a).

Mutation of Ser352 to Leu profoundly changes the structure of the catalytic loop (Figure 6.13b). Leu352 packs into a hydrophobic pocket that is occupied by Cys351 in the

native enzyme. Consequently, Cys351 and Ala350 have moved into the hydrophilic pocket of the native enzyme, forcing Lys318 out of the pocket. Ser349 moves by 8Å into the space that is reserved for Glu447 in the native enzyme, ejecting Glu447 from the active site (see arrows in Figure 6.13b). These changes are significant because Ser349 and Glu447 interact with the aldehyde substrate and the nicotinamide ribose of NAD^+, respectively (Figure 6.7). The conformational change of Glu447 may be particularly significant, since mutagenesis of this residue in ALDH2 (Glu399) substantially lowers k_{cat} and increases K_m for NAD^+ (48). Finally, catalytic Cys348 shifts by 3 Å, and its side chain is turned into the enzyme rather than pointing into the open space between the catalytic and NAD^+-binding domains.

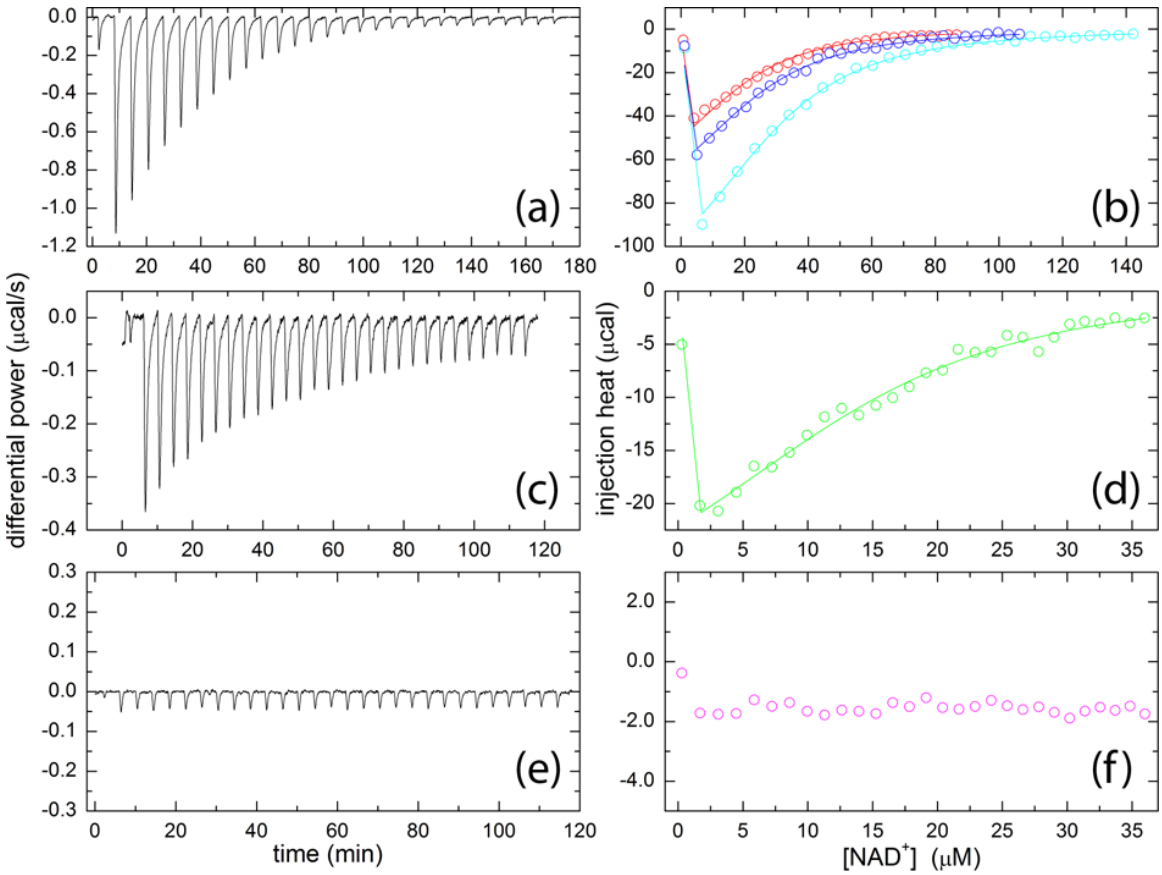


Figure 6.12. ITC analysis of NAD^+ binding to (a-b) HsP5CDH, (c-d) S352A, and (e-f) S352L. (a) Raw data for the titration of 47 µM HsP5CDH with 0.77 mM NAD^+. (b) Integrated data for titrations of 30, 34, and 47 µM HsP5CDH with NAD^+. The solid lines represent the optimal global fit of the three data sets to a single-site binding model. (c) Raw data for the titration of 18 µM S352A with 0.20 mM NAD^+. (d) Integrated data for the experiment displayed in panel c, along with the optimum least-squares fit to a single-site model. (e) Raw data for titration of 18 µM S352L with 0.20 mM NAD^+. (f) Integrated data for the experiment displayed in panel e.

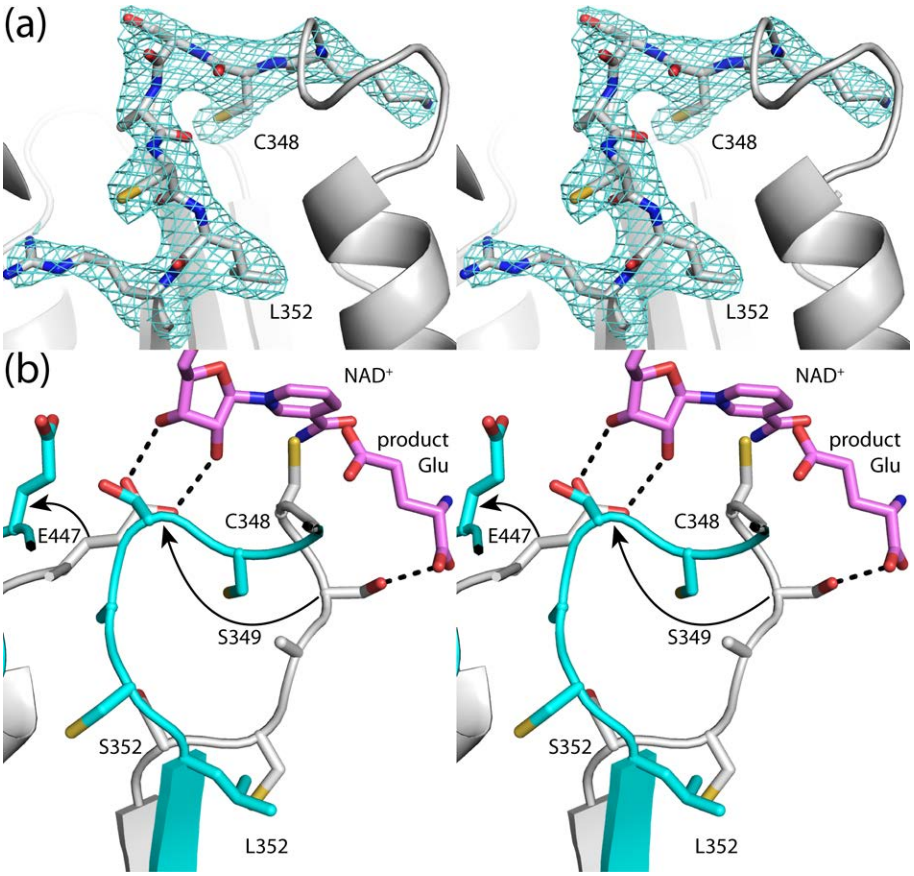


Figure 6.13. The active site of S352L. (a) Electron density map for the catalytic loop of S352L (stereographic view). The cage represents a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 2.5 σ. (b) Superposition of HsP5CDH (white) and S352L (cyan) (stereographic view). The arrows denote the movements of Ser349 and Glu447 that are induced by the mutation of Ser352 to Leu. For reference, NAD^+ and the product glutamate from the MmP5CDH structures are shown in pink. The dashed lines indicated hydrogen bonds to NAD^+ and the product glutamate in the MmP5CDH structures.

6.4. Discussion

Human and bacterial P5CDHs share a common fold. Structures of P5CDH from *Thermus thermophiles* (6, 49), *Bacillus licheniformis* (PDB code 3RJL), and *Bacillus halodurans* (3QAN) have been determined. The root mean square deviation between HsP5CDH and the bacterial enzymes is 1.5 Å over about 500 residues, indicating a high level of structural similarity at the fold level (Figure 6.14). A notable difference is that the

C-terminal extension is not found in bacterial P5CDHs. The extension is involved in dimerization in HsP5CDH. It remains to be determined whether bacterial and mammalian P5CDHs share a common oligomeric state in solution.

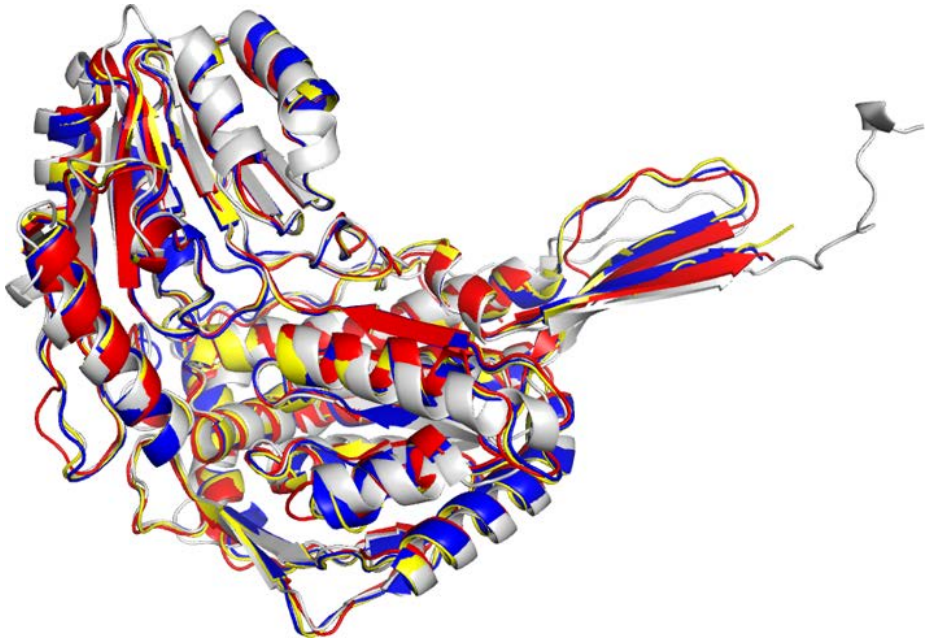


Figure 6.14. Superposition of HsP5CDH with bacterial P5CDHs from *Thermus thermophilus* (red, PDB code 2EIW), *Bacillus licheniformis* (blue, PDB code 3RJL), and *Bacillus halodurans* (yellow, PDB code 3QAN).

The MmP5CDH structures provide insight into the dual substrate specificity of HsP5CDH for GSA and OH-GSA. A model of the ternary enzyme-OH-GSA-NAD^+ complex was created by superimposing the MmP5CDH-Glu and MmP5CDH-NAD^+ structures and replacing the glutamate ligand with a model of OH-GSA (Figure 6.15). The hydroxyl group occupies a solvent-filled cavity and is bounded by Phe520 and the nicotinamide of NAD^+. The shortest contact distances with the O atom of the OH-GSA hydroxyl are 3.2 Å with the phenyl group of Phe520 and 3.1 Å with the nicotinamide

carbonyl. Thus, there are no severe steric clashes in the model. The absence of a protein group that clashes with hydroxyl of OH-GSA and the potential for hydrogen bonding with water molecules and the carboxamide of NAD^+ provide an explanation for the dual recognition of GSA and OH-GSA.

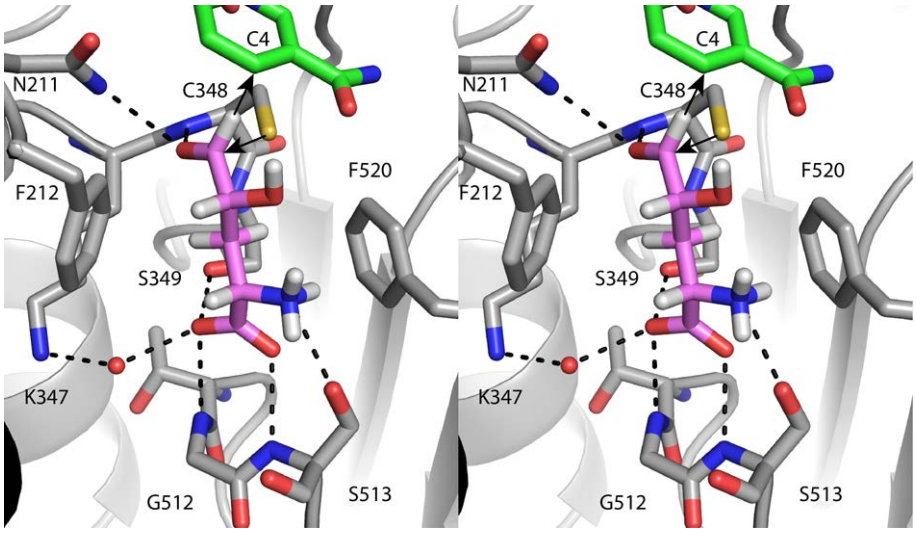


Figure 6.15. A model of OH-GSA bound to MmP5CDH (stereographic view). OH-GSA is shown in pink. NAD^+ is colored green. The arrows depict the directions of nucleophilic attack by Cys348 and hydride transfer to NAD^+.

The S352L mutant is reminiscent of the E487K mutant of ALDH2. As with S352L, E487K is associated with a human health disorder, impaired ethanol metabolism. In both cases the mutated residue does not directly contact the substrate or NAD^+, yet the mutant enzymes are catalytically deficient. Mutation of Glu487 to Lys in ALDH2 results in a 200-fold increase in K_m for NAD^+ (50), and structural studies have shown that NAD^+ induces partial reordering of the E487K active site (51). In contrast, we see no evidence for association of NAD^+ with S352L and hence no reordering of the active site.

Finally, our data provide a satisfying explanation of the role of the S352L mutation in HPII.

It was previously suggested, based on crystal structures and molecular dynamics simulations, that the main consequence of this mutation is to disrupt a water network that connects Ser352 to the catalytic Cys (6, 52). This idea was based, in part, on a static model derived from a TtP5CDH structure and assumed that the Leu side chain projects into the solvent-filled pocket that connects Ser352 to Cys348 (6). In fact, the S352L structure shows that Leu352 avoids the solvent-filled pocket and instead seeks a hydrophobic pocket (Figure 6.13b). The structure further shows that the S352L mutation dramatically alters the constellation of residues responsible for aldehyde binding, NAD^+ binding, and covalent bond formation. In particular, Ser349 is not positioned properly to interact with the carboxylate of the aldehyde substrate. Conserved Glu447 is no longer poised to bind the nicotinamide ribose of NAD^+, an interaction that is common to all ALDHs. And Cys348 is facing the wrong direction for nucleophilic attack. These structural features are consistent with the observed absence of catalytic activity and NAD^+ binding in S352L. In conclusion, the mutation of Ser352 to Leu disrupts three major aspects of catalysis - aldehyde recognition, NAD^+ binding, and nucleophilic attack - rendering the enzyme nonfunctional.

6.5. References.

1. Adams E, Frank L. Metabolism of proline and the hydroxyprolines. Annu Rev Biochem. 1980;49:1005-61.

2. Liu Y, Borchert GL, Donald SP, Diwan BA, Anver M, Phang JM. Proline oxidase functions as a mitochondrial tumor suppressor in human cancers. Cancer Res. 2009;69(16):6414-22. Epub 2009/08/06.

3. Valle D, Goodman SI, Harris SC, Phang JM. Genetic evidence for a common enzyme catalyzing the second step in the degradation of proline and hydroxyproline. J Clin Invest. 1979;64(5):1365-70. Epub 1979/11/01.

4. Sophos NA, Vasiliou V. Aldehyde dehydrogenase gene superfamily: the 2002 update. Chem Biol Interact. 2003;143-144:5-22.

5. Perez-Miller SJ, Hurley TD. Coenzyme isomerization is integral to catalysis in aldehyde dehydrogenase. Biochemistry. 2003;42(23):7100-9.

6. Inagaki E, Ohshima N, Takahashi H, Kuroishi C, Yokoyama S, Tahirov TH. Crystal structure of Thermus thermophilus Delta1-pyrroline-5-carboxylate dehydrogenase. J Mol Biol. 2006;362(3):490-501.

7. Efron ML. Familial Hyperprolinemia. Report Of A Second Case, Associated With Congenital Renal Malformations, Hereditary Hematuria And Mild Mental Retardation, With Demonstration Of An Enzyme Defect. N Engl J Med. 1965;272:1243-54.

8. Baumgartner MR, Rabier D, Nassogne MC, Dufier JL, Padovani JP, Kamoun P, et al. Delta1-pyrroline-5-carboxylate synthase deficiency: neurodegeneration, cataracts and connective tissue manifestations combined with hyperammonaemia and reduced ornithine, citrulline, arginine and proline. Eur J Pediatr. 2005;164(1):31-6.

9. Geraghty MT, Vaughn D, Nicholson AJ, Lin WW, Jimenez-Sanchez G, Obie C, et al. Mutations in the Delta1-pyrroline 5-carboxylate dehydrogenase gene cause type II hyperprolinemia. Hum Mol Genet. 1998;7(9):1411-5.

10. Scriver CR, Sly WS, Childs B, Beaudet AL, Valle D, Kinzler KW, et al., editors. The Metabolic and Molecular Bases of Inherited Disease. 8th ed. New York: McGraw-Hill; 2001.

11. Valle D, Goodman SI, Applegarth DA, Shih VE, Phang JM. Type II hyperprolinemia. Delta1-pyrroline-5-carboxylic acid dehydrogenase deficiency in cultured skin fibroblasts and circulating lymphocytes. J Clin Invest. 1976;58(3):598-603.

12. Wyse AT, Netto CA. Behavioral and neurochemical effects of proline. Metab Brain Dis. 2011;26(3):159-72. Epub 2011/06/07.

13. Phang JM, Hu CA, Valle D. Disorders of proline and hydroxyproline metabolism. In: Scriver CR, Beaudet AL, Sly WS, Valle D, editors. Metabolic and molecular basis of inherited disease. New York: McGraw Hill; 2001. p. 1821-38.

14. Jacquet H, Demily C, Houy E, Hecketsweiler B, Bou J, Raux G, et al. Hyperprolinemia is a risk factor for schizoaffective disorder. Mol Psychiatry. 2005;10(5):479-85. Epub 2004/10/21.

15. Clelland CL, Read LL, Baraldi AN, Bart CP, Pappas CA, Panek LJ, et al. Evidence for association of hyperprolinemia with schizophrenia and a measure of clinical outcome. Schizophr Res. 2011;131(1-3):139-45. Epub 2011/06/08.

16. Gogos JA, Santha M, Takacs Z, Beck KD, Luine V, Lucas LR, et al. The gene encoding proline dehydrogenase modulates sensorimotor gating in mice. Nat Genet. 1999;21(4):434-9.

17. Felix D, Kunzle H. The role of proline in nervous transmission. Adv Biochem Psychopharmacol. 1976;15:165-73. Epub 1976/01/01.

18. Takemoto Y, Semba R. Immunohistochemical evidence for the localization of neurons containing the putative transmitter L-proline in rat brain. Brain Res. 2006;1073-1074:311-5. Epub 2006/02/07.

19. Fremeau RT, Jr., Caron MG, Blakely RD. Molecular cloning and expression of a high affinity L-proline transporter expressed in putative glutamatergic pathways of rat brain. Neuron. 1992;8(5):915-26. Epub 1992/05/11.

20. He F, DiMario PJ. Drosophila delta-1-pyrroline-5-carboxylate dehydrogenase (P5CDh) is required for proline breakdown and mitochondrial integrity-Establishing a fly model for human type II hyperprolinemia. Mitochondrion. 2011;11(3):397-404. Epub 2010/12/21.

21. Doublie S. Preparation of selenomethionyl proteins for phase determinations. Methods Enzymol. 1997;276:523-30.

22. Matthews BW. Solvent content of protein crystals. J Mol Biol. 1968;33:491-7.

23. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997;276:307-26.

24. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):125-32. Epub 2010/02/04.

25. Evans P. Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr. 2006;62(Pt 1):72-82.

26. Potterton E, Briggs P, Turkenburg M, Dodson E. A graphical user interface to the CCP4 program suite. Acta Crystallogr D Biol Crystallogr. 2003;59(Pt 7):1131-7.

27. Weiss M. Global indicators of X-ray data quality. J Appl Cryst. 2001;34(2):130-5.

28. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins. 2003;50(3):437-50.

29. Pape T, Schneider TR. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. J Appl Crystallogr. 2004;37(5):843-4.

30. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr. 2002;58(Pt 10 Pt 2):1772-9. Epub 2002/09/28.

31. Sheldrick GM. Macromolecular phasing with SHELXE. Z Kristallogr. 2002;217:644-50.

32. Sheldrick GM. A short history of SHELX. Acta Crystallogr A. 2008;64(Pt 1):112-22. Epub 2007/12/25.

33. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr. 2002;58(Pt 11):1948-54.

34. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004;60(Pt 12 Pt 1):2126-32.

35. Adams PD, Gopal K, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, et al. Recent developments in the PHENIX software for automated crystallographic structure determination. J Synchrotron Rad. 2004;11(Pt 1):53-5.

36. Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. J Appl Cryst. 1997;30:1022-5.

37. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40(Pt 4):658-74. Epub 2007/08/01.

38. Lebowitz J, Lewis MS, Schuck P. Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Sci. 2002;11(9):2067-79. Epub 2002/08/23.

39. Cook PF, Cleland WW. Enzyme kinetics and mechanism. New York: Garland Science; 2007.

40. Forte-McRobbie C, Pietruszko R. Human glutamic-gamma-semialdehyde dehydrogenase. Kinetic mechanism. Biochem J. 1989;261(3):935-43. Epub 1989/08/01.

41. Liu ZJ, Sun YJ, Rose J, Chung YJ, Hsiao CD, Chang WR, et al. The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. Nat Struct Biol. 1997;4(4):317-26.

42. DeLano WL. The PyMOL User's Manual. Palo Alto, CA, USA: DeLano Scientific; 2002.

43. Steinmetz CG, Xie P, Weiner H, Hurley TD. Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. Structure. 1997;5(5):701-11. Epub 1997/05/15.

44. Moore SA, Baker HM, Blythe TJ, Kitson KE, Kitson TM, Baker EN. Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. Structure. 1998;6(12):1541-51. Epub 1998/12/24.

45. Lorentzen E, Hensel R, Knura T, Ahmed H, Pohl E. Structural Basis of allosteric regulation and substrate specificity of the non-phosphorylating glyceraldehyde 3-Phosphate dehydrogenase from Thermoproteus tenax. J Mol Biol. 2004;341(3):815-28. Epub 2004/08/04.

46. Johnson KA, Simpson ZB, Blom T. Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. Anal Biochem. 2009;387(1):20-9. Epub 2009/01/22.

47. Johnson KA, Simpson ZB, Blom T. FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. Anal Biochem. 2009;387(1):30-41. Epub 2009/01/27.

48. Ni L, Sheikh S, Weiner H. Involvement of glutamate 399 and lysine 192 in the mechanism of human liver mitochondrial aldehyde dehydrogenase. J Biol Chem. 1997;272(30):18823-6. Epub 1997/07/25.

49. Inagaki E, Ohshima N, Sakamoto K, Babayeva ND, Kato H, Yokoyama S, et al. New insights into the binding mode of coenzymes: structure of Thermus thermophilus [Delta]1-pyrroline-5-carboxylate dehydrogenase complexed with NADP+. Acta Cryst. 2007;F63(6 %R doi:10.1107/S1744309107021422):462-5.

50. Farres J, Wang X, Takahashi K, Cunningham SJ, Wang TT, Weiner H. Effects of changing glutamate 487 to lysine in rat and human liver mitochondrial aldehyde

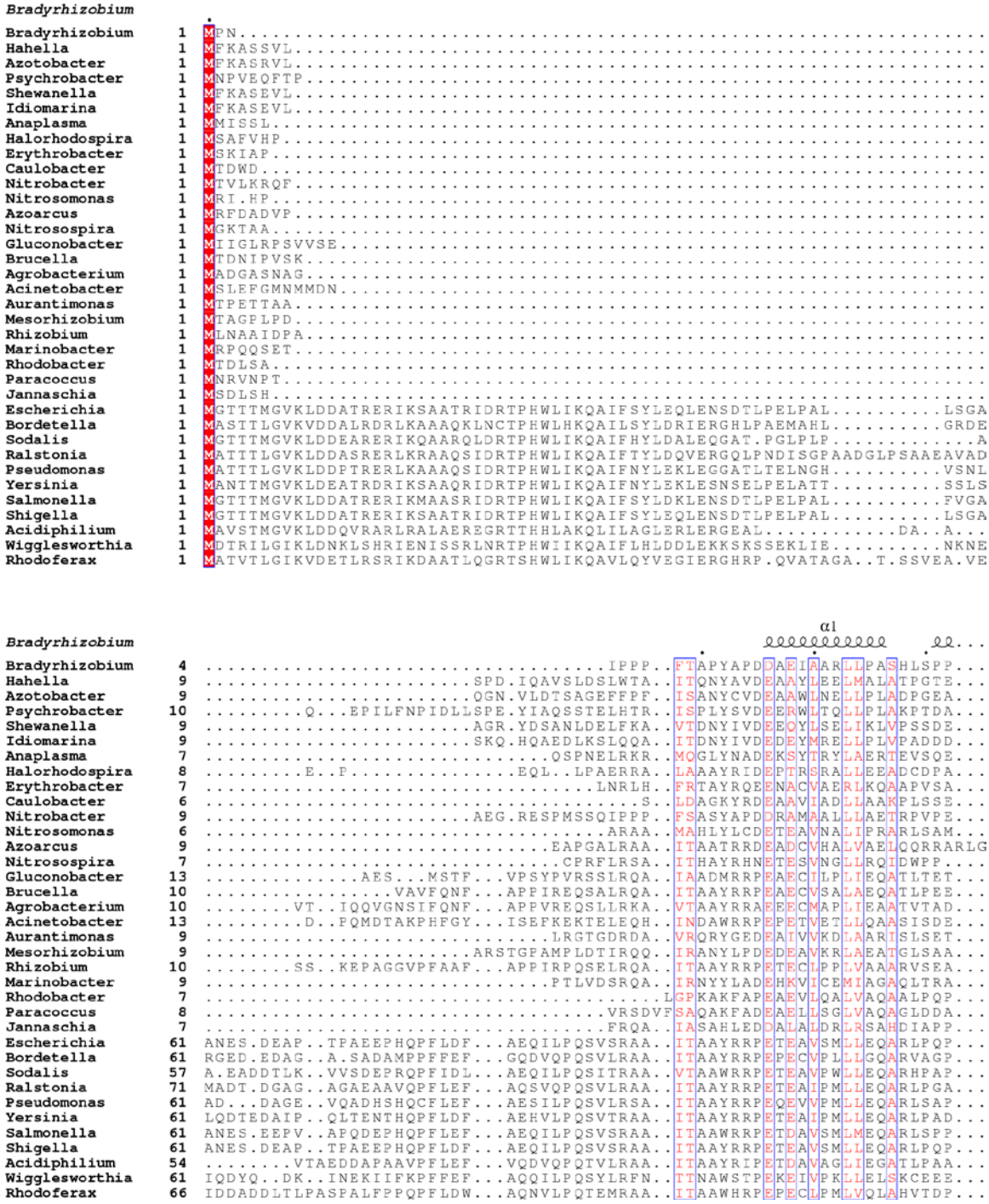
dehydrogenase. A model to study human (Oriental type) class 2 aldehyde dehydrogenase. J Biol Chem. 1994;269(19):13854-60. Epub 1994/05/13.

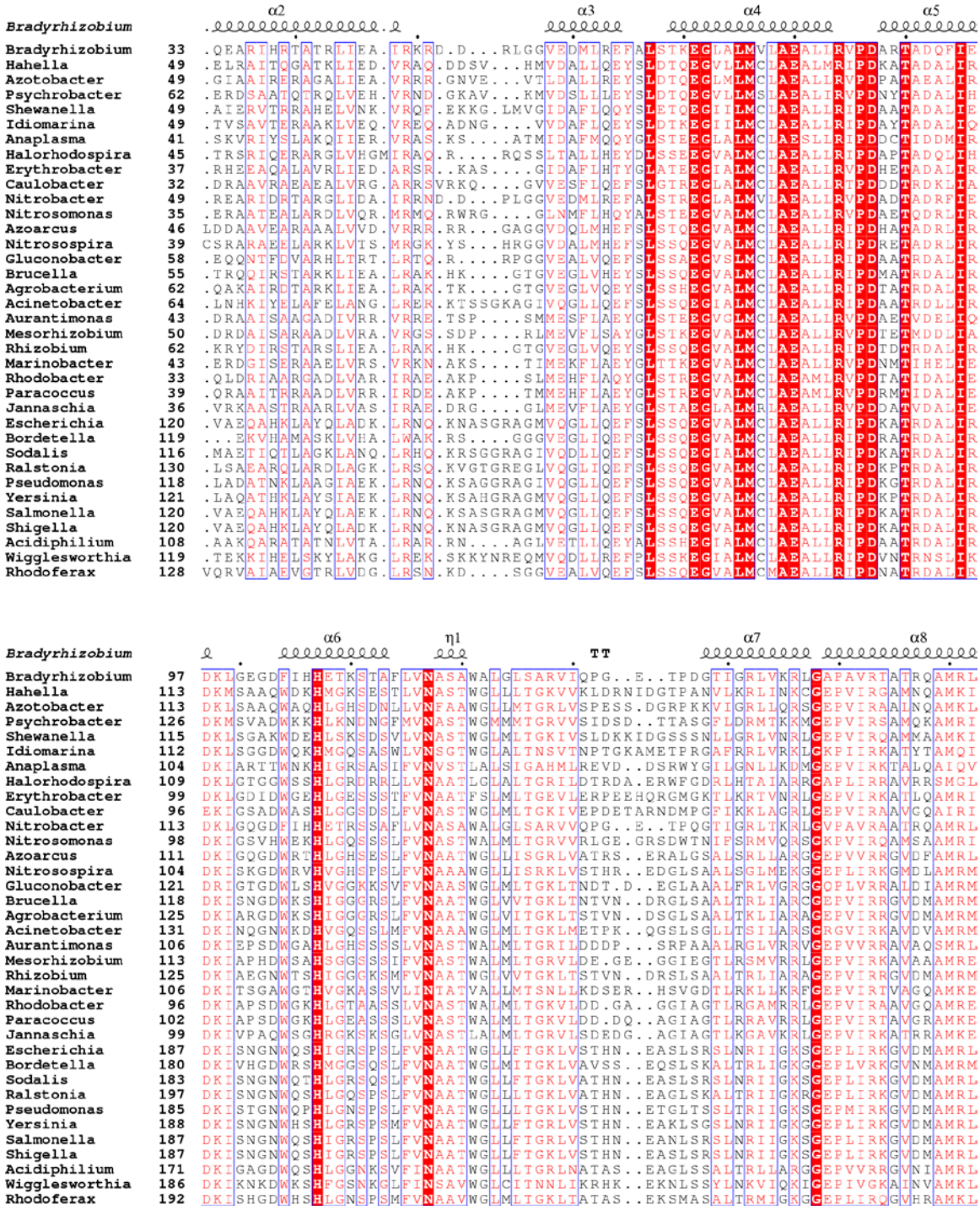
51. Larson HN, Zhou J, Chen Z, Stamler JS, Weiner H, Hurley TD. Structural and functional consequences of coenzyme binding to the inactive asian variant of mitochondrial aldehyde dehydrogenase: roles of residues 475 and 487. J Biol Chem. 2007;282(17):12940-50. Epub 2007/03/01.

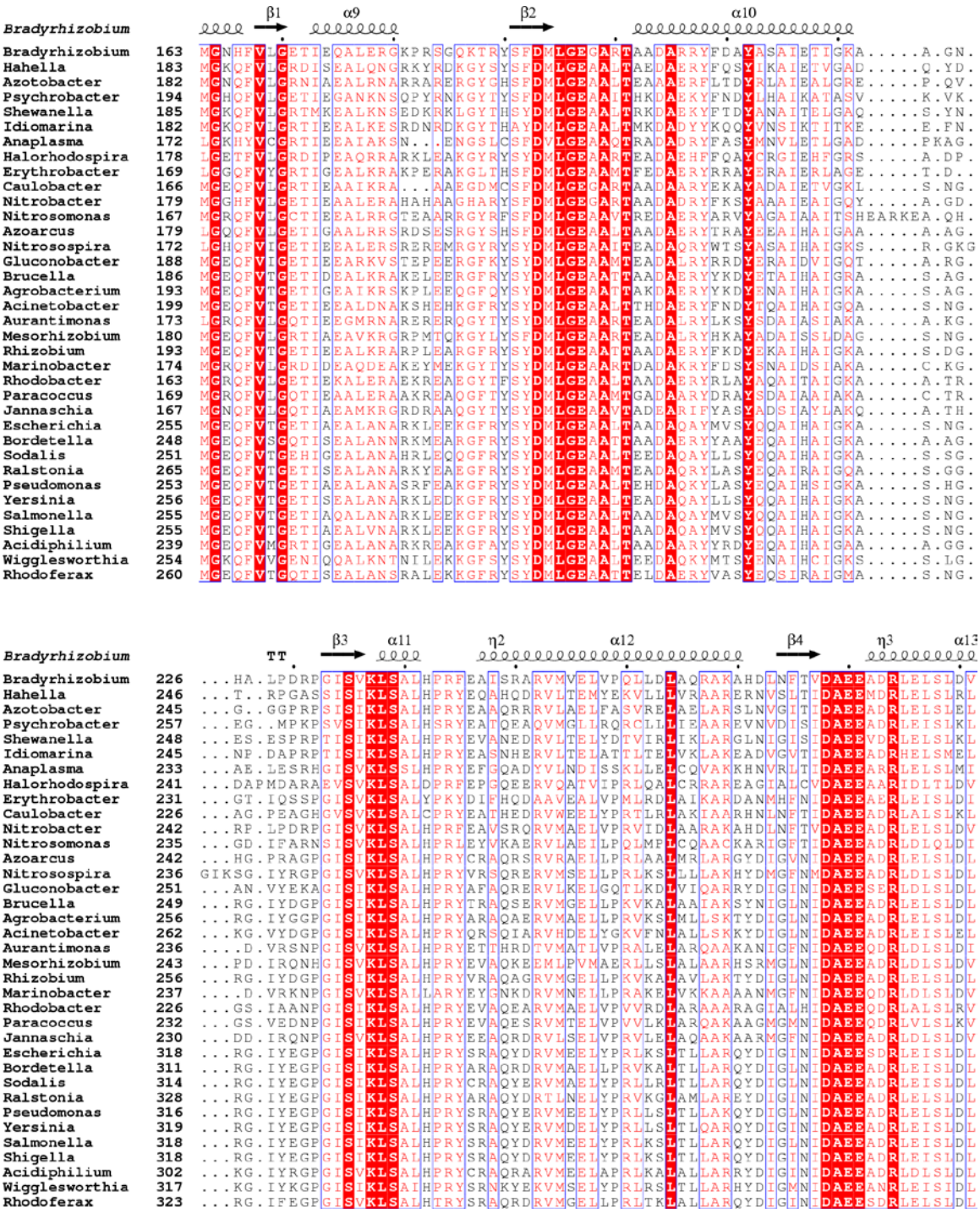
52. Hempel J, Kraut A, Wymore T. Gamma glutamyl semialdehyde dehydrogenase: simulations on native and mutant forms support the importance of outer shell lysines. Chem Biol Interact. 2009;178(1-3):75-8. Epub 2008/11/13.

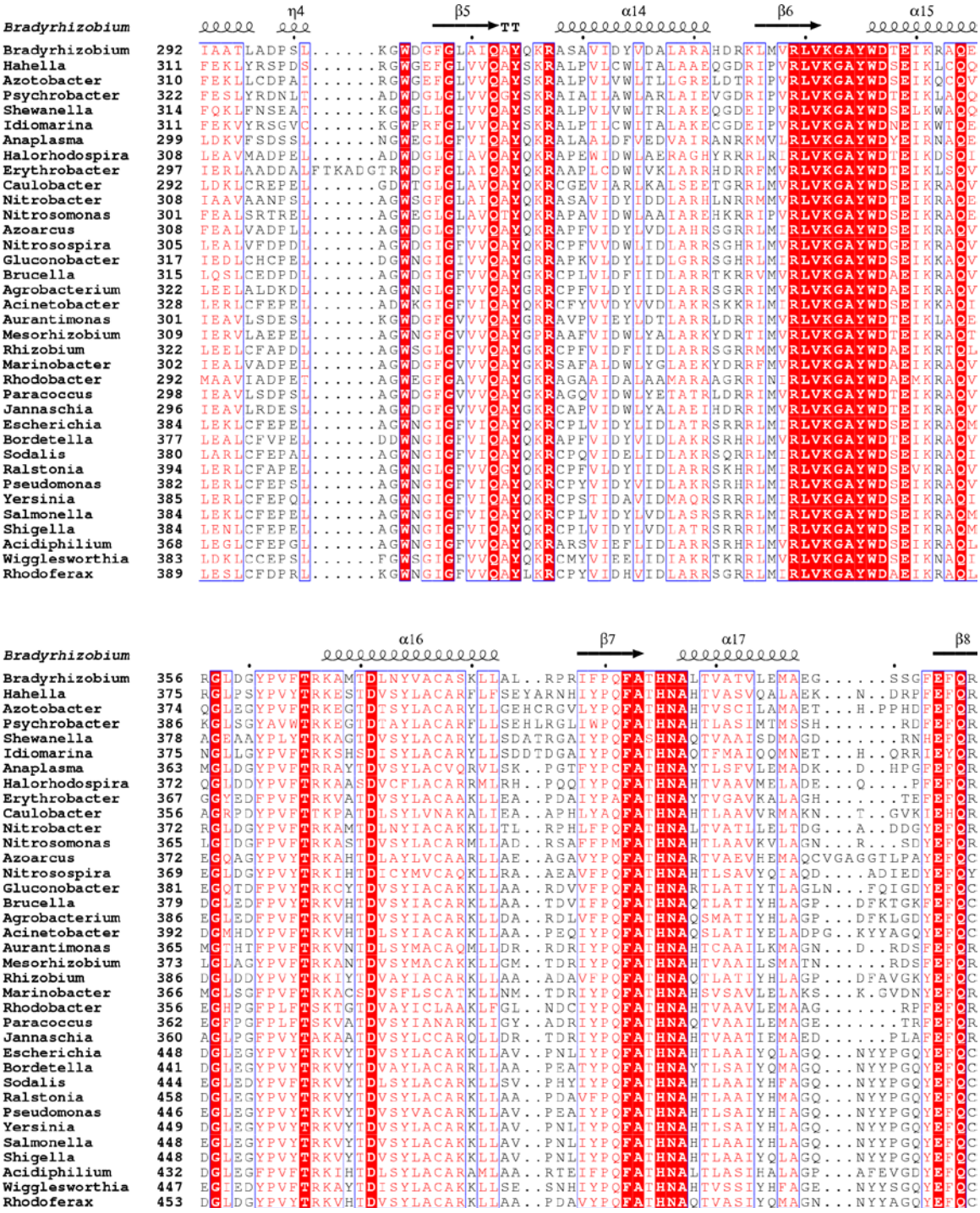
Appendix I

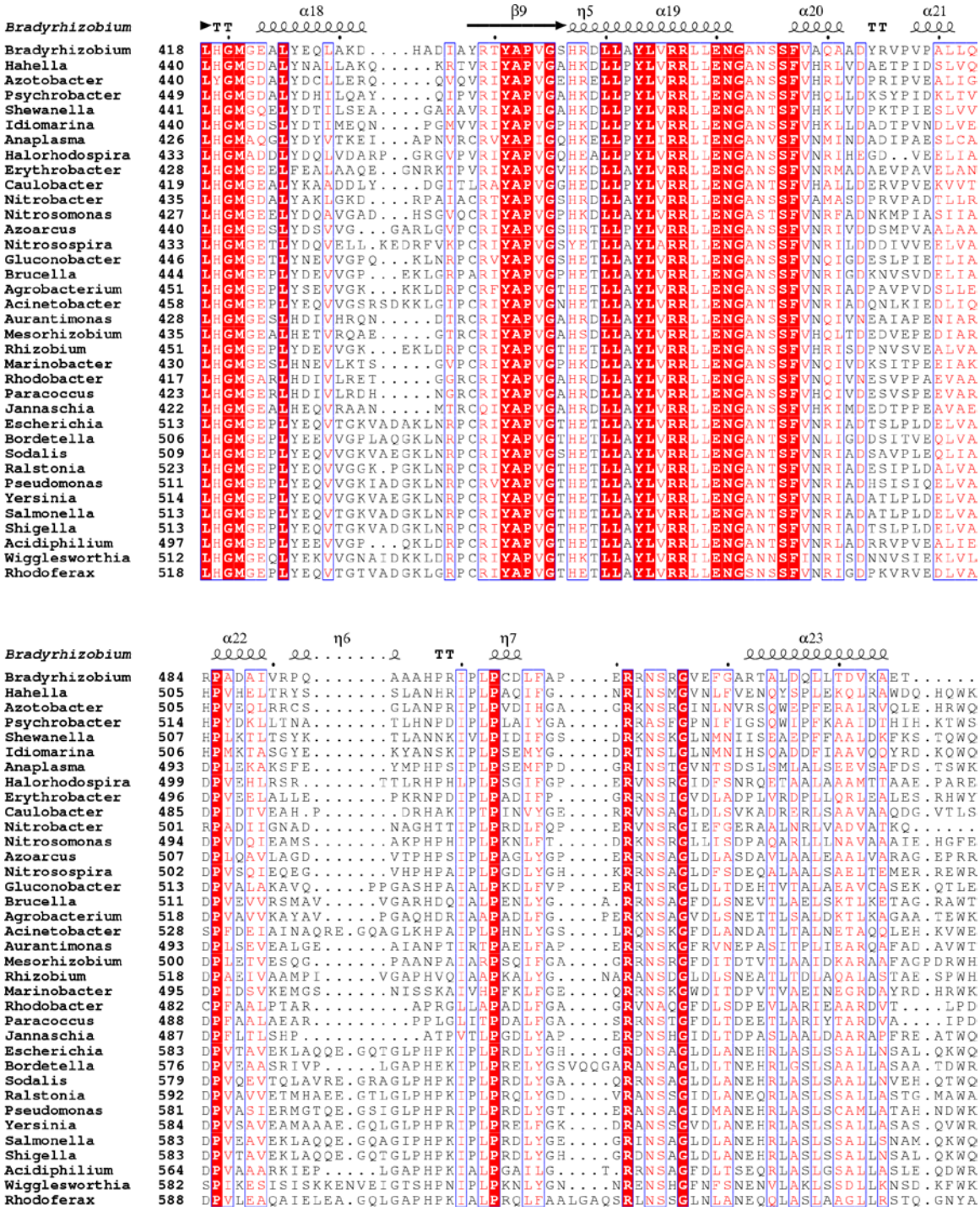
Sequence alignments of branch 1 PutAs

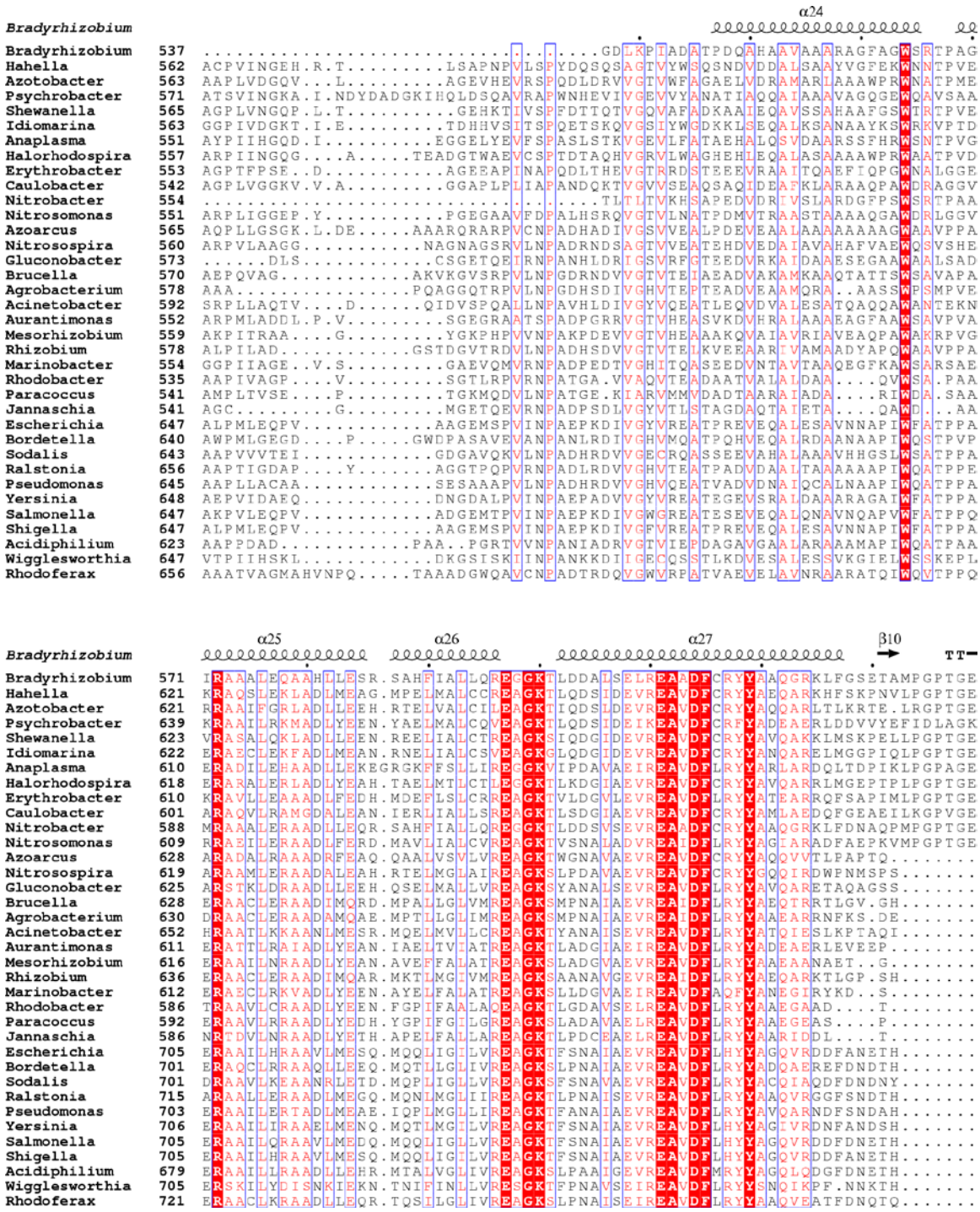

Bradyrhizobium
Bradyrhizobium 1 MPN
Hahella 1 MFKASSVL
Azotobacter 1 MFKASRVL
Psychrobacter 1 MNPVEQFTP
Shewanella 1 MFKASEVL
Idiomarina 1 MFKASEVL
Anaplasma 1 MMISSL
Halorhodospira 1 MSAFVHP
Erythrobacter 1 MSKIAP
Caulobacter 1 MTDWD
Nitrobacter 1 MTVLKRQF
Nitrosomonas 1 MRI.HP
Azoarcus 1 MRFDADVP
Nitrosospira 1 MGKTAA
Gluconobacter 1 MIIGLRPSVVSE
Brucella 1 MTDNIPVSK
Agrobacterium 1 MADGASNAG
Acinetobacter 1 MSLEFGMNMMDN
Aurantimonas 1 MTPETTAA
Mesorhizobium 1 MTAGPLPD
Rhizobium 1 MLNAAIDPA
Marinobacter 1 MRPQQSET
Rhodobacter 1 MTDLSA
Paracoccus 1 MNRVNPT
Jannaschia 1 MSDLSH
Escherichia 1 MGTTTMGVKLDDATRERIKSAATRIDRTPHWLIKQAIFSYLEQLENSDTLPELPAL LSGA
Bordetella 1 MASTTLGVKVDDALRDRLKAAAQKLNCTPHWLHKQAILSYLDRIERGHLPAEMAHL GRDE
Sodalis 1 MGTTTMGVKLDDEARERIKQAARQLDRTPHWLIKQAIFHYLDALEQGAT.PGLPLP A
Ralstonia 1 MATTTLGVKLDDASRERLKRAAQSIDRTPHWLIKQAIFTYLDQVERGQLPNDISGPAADGLPSAAEAVAD
Pseudomonas 1 MATTTLGVKLDDPTRERLKAAAQSIDRTPHWLIKQAIFNYLEKLEGGATLTELNGH VSNL
Yersinia 1 MANTTMGVKLDEATRDRIKSAAQRIDRTPHWLIKQAIFNYLEKLESNSELPELATT SSLS
Salmonella 1 MGTTTMGVKLDDATRERIKMAASRIDRTPHWLIKQAIFSYLDKLENSDTLPELPAL FVGA
Shigella 1 MGTTTMGVKLDDATRERIKSAATRIDRTPHWLIKQAIFSYLEQLENSDTLPELPAL LSGA
Acidiphilium 1 MAVSTMGVKLDDQVRARLRALAEREGRTTHHLAKQLILAGLERLERGEAL DA..A
Wigglesworthia 1 MDTRILGIKLDNKLSHRIENISSRLNRTPHWIIKQAIFLHLDDLEKKSKSSEKLIE NKNE
Rhodoferax 1 MATVTLGIKVDETLRSRIKDAATLQGRTSHWLIKQAVLQYVEGIERGHRP.QVATAGA..T.SSVEA.VE
α1
Bradyrhizobium
Bradyrhizobium 4 IPPP..FTAPYAPDDAEIAARLLPASHLSPP
Hahella 9 SPD.IQAVSLDSLWTA..ITQNYAVDEAAYLEELMALATPGTE
Azotobacter 9 OGN.VLDTSAGEFFPF..ISANYCVDEAAWLNELLPLADPGEA
Psychrobacter 10 Q...EPILFNPIDLLSPE.YIAQSSTELHTR..ISPLYSVDEERWLTQLLPLAKPTDA
Shewanella 9 AGR.YDSANLDELFKA..VTDNYIVDEEQYLSELIKLVPSSDE
Idiomarina 9 SKQ.HQAEDLKSLQQA..ITDNYIVDEDEYMRELLPLVPADDD
Anaplasma 7 QSPNELRKR..MQGLYNADEKSYTRYLAERTEVSQE
Halorhodospira 8 E..P EQL..LPAERRA..LAAAYRIDEPTRSRALLEEADCDPA
Erythrobacter 7 LNRLH..FRTAYRQEENACVAERLKQAAPVSA
Caulobacter 6 S..LDAGKYRDEAAVIADLLAAKPLSSE
Nitrobacter 9 AEG.RESPMSSQIPPP..FSASYAPDDRAMAALLAETRPVPE
Nitrosomonas 6 ARAA..MAHLYLCDETEAVNALIPRARLSAM
Azoarcus 9 EAPGALRAA..ITAATRRDEADCVHALVAELQQRRARLG
Nitrosospira 7 CPRFLRSA..ITHAYRHNETESVNGLLRQIDWPP
Gluconobacter 13 AES...MSTF...VPSYPVRSSLRQA..IAADMRRPEAECILPLIEQATLTET
Brucella 10 VAVFQNF...APPIREQSALRQA..ITAAYRRPEAECVSALAEQATLPEE
Agrobacterium 10 VT..IQQVGNSIFQNF...APPVREQSLLRKA..VTAAYRRAFEFCMAPLIEAATVTAD
Acinetobacter 13 D..PQMDTAKPHFGY...ISEFKEKTELEQH..INDAWRRPEPETVETLLQAASISDE
Aurantimonas 9 LRGTGDRDA..VRQRYGEDEAIVVKDLAARISLSET
Mesorhizobium 9 ARSTGPAMPLDTIRQQ..IRANYLPDEDEAVKRLAEATGLSAA
Rhizobium 10 SS..KEPAGGVPFAAF...APPIRPQSELRQA..ITAAYRRPETECLPPLVAAARVSEA
Marinobacter 9 PTLVDSRQA..IRNYYLADEHKVICEMIAGAQLTRA
Rhodobacter 7 LGPKAKFAPEAEVLQALVAQAALPQP
Paracoccus 8 VRSDVFSAQAKFADEAELLSGLVAQAGLDDA
Jannaschia 7 FRQA..IASAHLEDDALALDRLRSAHDIAPP
Escherichia 61 ANES.DEAP..TPAEEPHQPFLDF...AEQILPQSVSRAA..ITAAYRRPETEAVSMLLEQARLPQP
Bordetella 61 RGED.EDAG..A.SADAMPPFFEF...GQDVQPQSVLRAA..ITAAYRRPEPECVPLLLGQARVAGP
Sodalis 57 A.EADDTLK..VVSDEPRQPFIDL...AEQILPQSITRAA..VTAAWRRPETEAVPWLLEQARHPAP
Ralstonia 71 MADT.DGAG..AGAEAAVQPFLEF...AQSVQPQSVLRAA..ITAAYRRPETEAIPMLLEQARLPGA
Pseudomonas 61 AD...DAGE..VQADHSHQCFLEF...AESILPQSVLRSA..ITAAYRRPEQEVVPMLLEQARLSAP
Yersinia 61 LQDTEDAIP..QLTENTHQPFLDF...AEHVLPQSVTRAA..ITAAYRRPETEAIPMLLEQARLPAD
Salmonella 61 ANES.EEPV..AFQDEPHQPFLEF...AEQILPQSVSRAA..ITAAWRRPETDAVSMLMEQARLSPP
Shigella 61 ANES.DEAP..TPAEEPHQPFLDF...AEQILPQSVSRAA..ITAAYRRPETEAVSMLLEQARLPQP
Acidiphilium 54 VTAEDDAPAAVPFLEF...VQDVQPQTVLRAA..ITAAYRIPETDAVAGLIEGATLPAA
Wigglesworthia 61 IQDYQ..DK..INEKIIFKPFFEF...AQQILPQSYLRFN..ITNAWSTPEKEIVPKLLELSKCEEE
Rhodoferax 66 IDDADDLTLPASPALFPPQPFLDW...AQNVLPQTEMRAA..ITAAWHRPEPECLPMLVQLAHVTDP

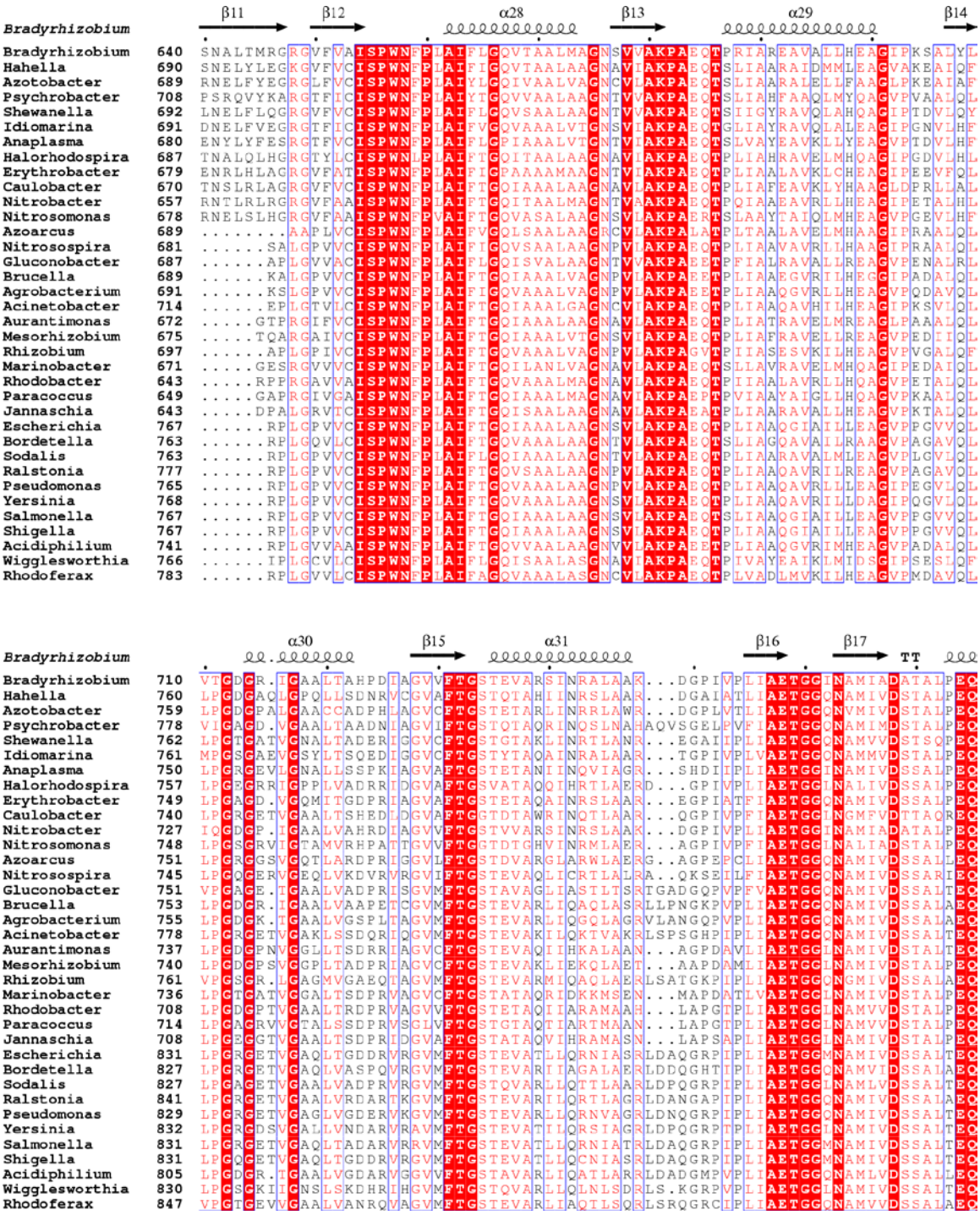

α2
α3
α4
α5
α6
η1
α7
α8
Bradyrhizobium
Bradyrhizobium 33
Hahella 49
Azotobacter 49
Psychrobacter 62
Shewanella 49
Idiomarina 49
Anaplasma 41
Halorhodospira 45
Erythrobacter 37
Caulobacter 32
Nitrobacter 49
Nitrosomonas 35
Azoarcus 46
Nitrosospira 39
Gluconobacter 58
Brucella 55
Agrobacterium 62
Acinetobacter 64
Aurantimonas 43
Mesorhizobium 50
Rhizobium 62
Marinobacter 43
Rhodobacter 33
Paracoccus 39
Jannaschia 36
Escherichia 120
Bordetella 119
Sodalis 116
Ralstonia 130
Pseudomonas 118
Yersinia 121
Salmonella 120
Shigella 120
Acidiphilium 108
Wigglesworthia 119
Rhodoferax 128
Bradyrhizobium 97
Hahella 113
Azotobacter 113
Psychrobacter 126
Shewanella 115
Idiomarina 112
Anaplasma 104
Halorhodospira 109
Erythrobacter 99
Caulobacter 96
Nitrobacter 113
Nitrosomonas 98
Azoarcus 111
Nitrosospira 104
Gluconobacter 121
Brucella 118
Agrobacterium 125
Acinetobacter 131
Aurantimonas 106
Mesorhizobium 113
Rhizobium 125
Marinobacter 106
Rhodobacter 96
Paracoccus 102
Jannaschia 99
Escherichia 187
Bordetella 180
Sodalis 183
Ralstonia 197
Pseudomonas 185
Yersinia 188
Salmonella 187
Shigella 187
Acidiphilium 171
Wigglesworthia 186
Rhodoferax 192

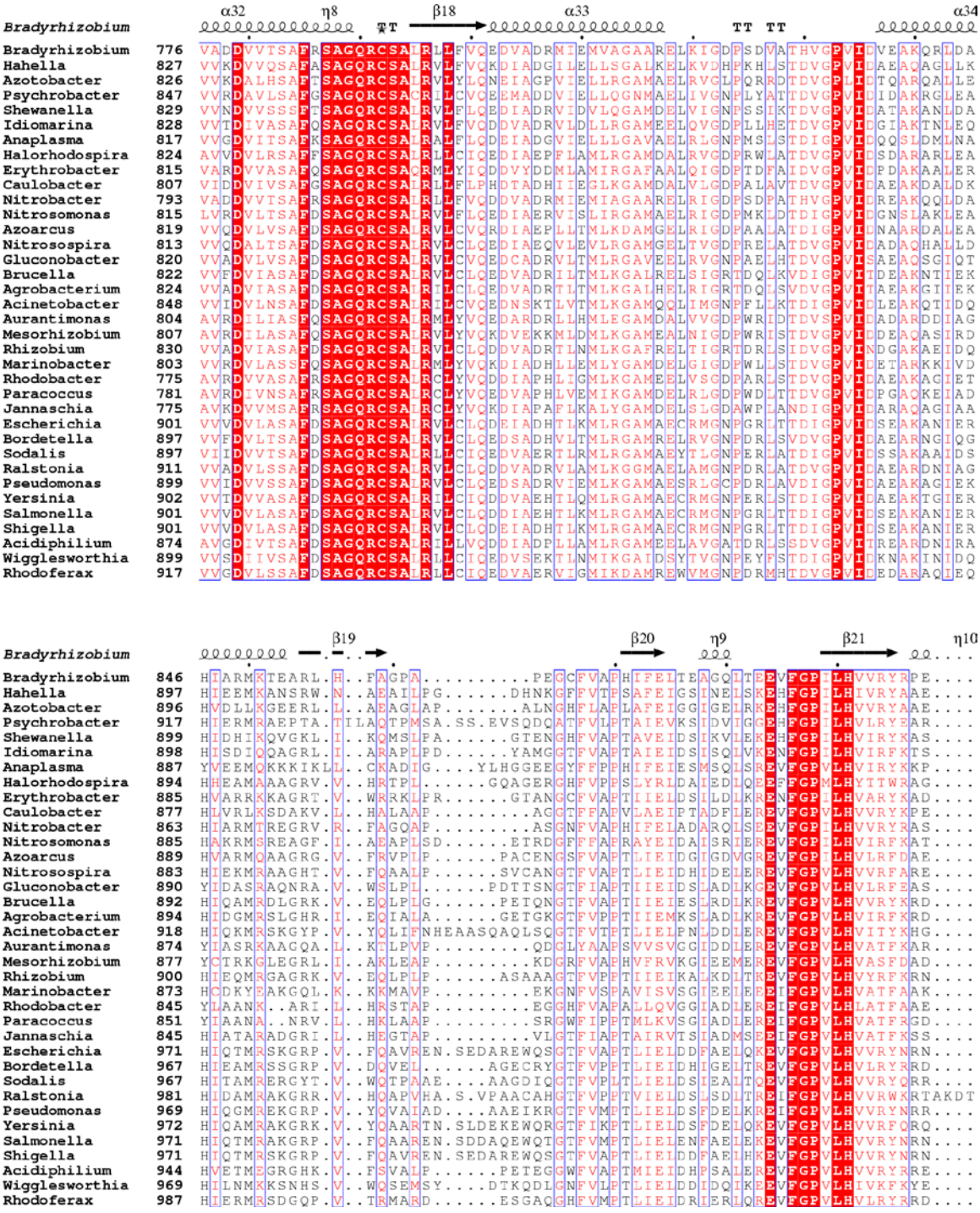

Bradyrhizobium
β1
α9
β2
α10
Bradyrhizobium 163 MGNHFVLGETIEQALERGKPRSGQKTRYSFDMLGEGARTAADARRYFDAYASAIETIGKA.....A.GN.
Hahella 183 MGKQFVLGRDISEALQNGRKYRDKGYSYSFDMLGEAALTAEDAERYFQSYIKAIETVGAD.....Q.YD.
Azotobacter 182 MGNQFVLGRNIAEALRNARRARERGYGHSFDMLGEAALTEAAAERFLTDYRLAIEALGRE.....P.QV.
Psychrobacter 194 MGHQFVLGETIEGANKNSQPYRNKGYIYSFDMLGEAAITHKDAEKYFNDYLHAIKATASV.....K.VK.
Shewanella 185 MGKQFVLGRTMKEALKNSEDKRKLGYTHSYDMLGEAALTRKDAEKYFTDYANAITELGAQ.....S.YN.
Idiomarina 182 MGKQFVLGRTIEEALKESRDNRDKGYTHAYDMLGEAALTMKDADYYKQQYVNSIKTITKE.....E.FN.
Anaplasma 172 LGKHYVCGRTIEEAIAKSN...ENGSLCSFDVLGEAAQTRADADRYFASYMNVLETLGAD.....PKAG.
Halorhodospira 178 LGETFVLGRDIPEAQRRARKLEAKGYRYSYDMLGEAARTEADAEHFFQAYCRGIEHFGRS.....A.DP.
Erythrobacter 169 LGGQFVYGRTIGEALKRAKPERAKGLTHSFDMLGEAAMTFEDAERYRRAYERAIERLAGE.....T.D..
Caulobacter 166 MGEQFVLGRTIEAAIKRA...AAEGDMCSFDMLGEGARTAADAARYEKAYADAIETVGKL.....S.NG.
Nitrobacter 179 MGGHFVLGETIEAALERAHAHAAGHARYSFDMLGEGARTAADADRYFKSYAAAIEAIGQY.....A.GD.
Nitrosomonas 167 MGRQFVLGCTIEEALRRGTEAARRGYRFSFDMLGEAAVTREDAERYARVYAGAIAAITSHEARKEA.QH.
Azoarcus 179 LGQQFVLGETIGAALRRSRDSESRGYSHSFDMLGEAALTAALAERYTRAYEEAIHAIGAA.....A.AG.
Nitrosospira 172 LGHQFVIGETIEEALERSRERERMRGYRYSYDMLGEAAITEADAQRYWTSYASAIHAIGKS.....R.GKG
Gluconobacter 188 MGEQFVIGETIEEARKVSTEPEERGFKYSYDMLGEAAMTEADALRYRRDYERAIDVIGQT.....A.RG.
Brucella 186 MGEQFVTGETIDEALKRAKELEERGFRYSYDMLGEAATTAADAERYYKDYETAIHAIGRA.....S.AG.
Agrobacterium 193 MGEQFVTGETIGEAIKRSKPLEEQGFQYSYDMLGEAATTAKDAERYYKDYENAIHAIGKA.....S.AG.
Acinetobacter 199 MGEQFVTGETIEEALDNAKSHEHKGFRYSYDMLGEAALTTHDAERYFNDYTQAIHAIGQA.....S.NG.
Aurantimonas 173 LGRQFVLGQTIEEGMRNARERERQGYTYSYDMLGEAARTEADALRYLKSYSDAIASIAKA.....A.KG.
Mesorhizobium 180 MGEQFVLGRTIAEAVKRGRPMTQKGYLYSFDMLGEAARTEADALRYHKAYADAISSLDAG.....S.NG.
Rhizobium 193 MGEQFVTGETIEEALKRARPLEAARGFRYSYDMLGEAATTAADAERYFKDYEKAIHAIGKA.....S.DG.
Marinobacter 174 MGRQFVLGRDIDEAQDEAKEYMEKGYIYSYDMLGEAARTDADAKRYFDSYSNAIDSIAKA.....C.KG.
Rhodobacter 163 MGRQFVLGETIEKALERAEKREAEGYTFSYDMLGEAALTAADAERYRLAYAQAITAIGKA.....A.TR.
Paracoccus 169 MGRQFVLGQTIEAALERAKREAQGFTYSYDMLGEAAMTGADAARYDRAYSDAIAAIAKA.....C.TR.
Jannaschia 167 MGNQFVLGQTIAEAMKRGRDRAAQGYTYSYDMLGEAAVTADEARIFYASYADSIAYLAKQ.....A.TH.
Escherichia 255 MGEQFVTGETIAEALANARKLEEKGFRYSYDMLGEAALTAADAQAYMVSYQQAIHAIGKA.....S.NG.
Bordetella 248 MGEQFVSGQTISEALANNRKMEARGFRYSYDMLGEAATTAADAERYYAAYEQAIHAIGKA.....A.AG.
Sodalis 251 MGEQFVTGEHIGEALANAHRLEQQGFRYSYDMLGEAALTEEDAQAYLLSYQQAIHAIGKA.....S.SG.
Ralstonia 265 MGEQFVTGETISEALANARKYEAEGFRYSYDMLGEAAMTEADAQRYLASYEQAIRAIGQA.....S.GG.
Pseudomonas 253 MGEQFVTGETIAEALANASRFEAKGFRYSYDMLGEAALTEHDAQKYLASYEQAIHSIGKA.....S.HG.
Yersinia 256 MGEQFVTGETISEALANARKLEDKGFRYSYDMLGEAALTEADAQAYLLSYQQAIHAIGKA.....S.NG.
Salmonella 255 MGEQFVTGETIAQALANARKLEEKGFRYSYDMLGEAALTAADAQAYMVSYQQAIHAIGKA.....S.NG.
Shigella 255 MGEQFVTGETIAEALVNARKLEEKGFRYSYDMLGEAALTAADAQAYMVSYQQAIHAIGKA.....S.NG.
Acidiphilium 239 MGEQFVMGRTIGEALANARKREAKGFAYSYDMLGEAALTAADAARYYRDYEQAIHAIGAA.....A.GG.
Wigglesworthia 254 MGKQFVVGENIQQALKNTNILEKKGFSYSYDMLGEAALTEEDAQKYMTSYENAIHCIGKS.....S.LG.
Rhodoferax 260 MGEQFVTGQTISEALANSRALEKKGFRYSYDMLGEAATTELDAERYVASYEQSIRAIGMA.....S.NG.
Bradyrhizobium
TT
β3
α11
η2
α12
β4
η3
α13
Bradyrhizobium 226 ...HA.LPDRPGISVKLSALHPRFEAISRARVMVELVPQLLDLAQRAKAHDLNFTVDAEEADRLELSLDV
Hahella 246 ...T..RPGASSISIKLSALHPRYEQAHQDRVLTEMYEKVLLLVRAARERNVSLTIDAEEMDRLELSLRL
Azotobacter 245 ...G..GGPRPSISIKLSALHPRYEAAQRRRVLAELFASVRELAELARSLNVGITIDAEEADRLELSLEL
Psychrobacter 257 ...EG..MPKPSVSIKLSALHSRYEATQFAQVMGLLRQRCLLLIEAARFVNVDISIDAEEADRLEISLKL
Shewanella 248 ...ES.ESPRPTISIKLSALHPRYEVANEDRVLTELYDTVIRLIKLARGLNIGISIDAEEVDRLELSLKL
Idiomarina 245 ...NP.DAPRPTISIKLSALHPRYEASNHERVLTELATTLTELVKLAKEADVGVTIDAEEADRHELSMEL
Anaplasma 233 ...AE.LESRHGISVKLSSLHPRYEFGQADYVLNDISSKLLELCQVAKKHNVRLTIDAEEARRLELSLMI
Halorhodospira 241 ...DAPMDARAEVSVKLSALDPRFEPGQEERVQATVIPRLQALCRRAREAGIALCVDAEEAARIDLTLDV
Erythrobacter 231 ...GT.IQSSPGISVKLSALYPKYDIFHQDAAVEALVPMLRDLAIKARDANMHFNIDAEEAERLEISLDI
Caulobacter 226 ...AG.PEAGHGVSVKLSALCPRYEATHEDRVWEELYPRTLRLAKIAARHNLNFTIDAEEADRLALSLKL
Nitrobacter 242 ...RP.LPDRPGISVKLSALHPRFEAVSRQRVMAELVPRVIDLAARAKAHDLNFTVDAEEADRLELSLDV
Nitrosomonas 235 ...GD.IFARNSISVKLSALHPRLEYVKAERVLAEILPQLMPLCQAACKARIGFTIDAEEADRLDLQLDI
Azoarcus 242 ...HG.PRAGPGISIKLSALHPRYCRAQRSRVRAELLPRLAALMRLARGYDIGVNIDAEEADRLELSLDL
Nitrosospira 236 GIKSG.IYRGPGISVKLSALHPRYVRSQRERVMSELLPRLKSLLLLAKHYDMGFNMDAEEADRLDLSLDL
Gluconobacter 251 ...AN.VYEKAGISIKLSALHPRYAFAQRERVLKELGQTLKDLVIQARRYDIGINIDAEESERLDLSLDL
Brucella 249 ...RG.IYDGPGISIKLSALHPRYTRAQSERVMGELLPKVKALAAIAKSYNIGLNIDAEEADRLELSLDL
Agrobacterium 256 ...RG.IYGGPGISIKLSALHPRYARAQAERVMAELLPRVKSLMLLSKTYDIGLNIDAEEADRLELSLDL
Acinetobacter 262 ...KG.VYDGPGISIKLSALHPRYQRSQIARVHDELYGKVFNLALLSKKYDIGLNIDAEEADRLEISLEL
Aurantimonas 236 ....D.VRSNPGISVKLSALHPRYETIHRDTVMATLVPRALELARQAAKANIGFNIDAEEQDRLDLSLDI
Mesorhizobium 243 ...PD.IRQNHGISVKLSALHPRYEVAQKEMLPVMAERLLSLALAARHSRMGLNIDAEEADRLDLSLDV
Rhizobium 256 ...RG.IYDGPGISIKLSALHPRYVRAQAGRVMGELLPKVKALAVLAKTYDIGLNIDAEEADRLELSLDL
Marinobacter 237 ....D.VRKNPGISVKLSALLARYEYGNKDRVMNELLPRAKELVKKAAANMGFNIDAEEQDRLDLSLDV
Rhodobacter 226 ...GS.IAANPGISIKLSALHPRYEVAQEARVMAELVPVVRDLARAAAAAGIALHIDAEEQDRLALSLRV
Paracoccus 232 ...GS.VEDNPGISIKLSALHPRYEVAQESRVMTELVPVVLKLARQAKAAGMGMNIDAEEQDRLVLSLKV
Jannaschia 230 ...DD.IRQNPGISVKLSALHPRYEEAQRDRVLSELVPRVLELAQAAKAARMGFNIDAEEADRLDLSLDV
Escherichia 318 ...RG.IYEGPGISIKLSALHPRYSRAQYDRVMEELYPRLKSLTLLARQYDIGINIDAEESDRLEISLDL
Bordetella 311 ...RG.IYEGPGISIKLSALHPRYARAQRDRVMAELLPRVKALTLLARQYDIGLNIDAEEADRLEISLDL
Sodalis 314 ...RG.IYEGPGISIKLSALHPRYCRAQYERVMAELYPRLLRLTLLARQYDIGLNIDAEEADRLEISLDL
Ralstonia 328 ...RG.IYEGPGISIKLSALHPRYARAQYDRTLNELYPRVKGLAMLARKYDIGINIDAEEADRLELSLDL
Pseudomonas 316 ...RG.IYEGPGISIKLSALHPRYSRAQYERVMEELYPRLLSLTLLAKQYDIGLNIDAEEADRLELSLDL
Yersinia 319 ...RG.IYEGPGISIKLSALHPRYSRAQYERVMDELYPRLLSLTLQARQYDIGINIDAEEADRLEISLDL
Salmonella 318 ...RG.IYEGPGISIKLSALHPRYSRAQYDRVMEELYPRLKSLTLLARQYDIGLNIDAEEADRLEISLDL
Shigella 318 ...RG.IYEGPGISIKLSALHPRYSRAQYDRVMEELYPRLKSLTLLARQYDIGINIDAEEADRLEISLDL
Acidiphilium 302 ...KG.IYRGPGISVKLSALHPRYCRAQRARVMEELAPRLKALALLARRYDIGINIDAEEADRLDLSLDL
Wigglesworthia 317 ...KG.IYKGPGISIKLSAIHPRYSRNKYEKVMSELYPRLRSLTLLARHYDIGMNIDAEEANRLEISIDL
Rhodoferax 323 ...RG.IFEGPGISVKLSALHTRYSRAQRDRVMGELLPRLTKLALLARQYDIGINIDAEESDRLELSLDL

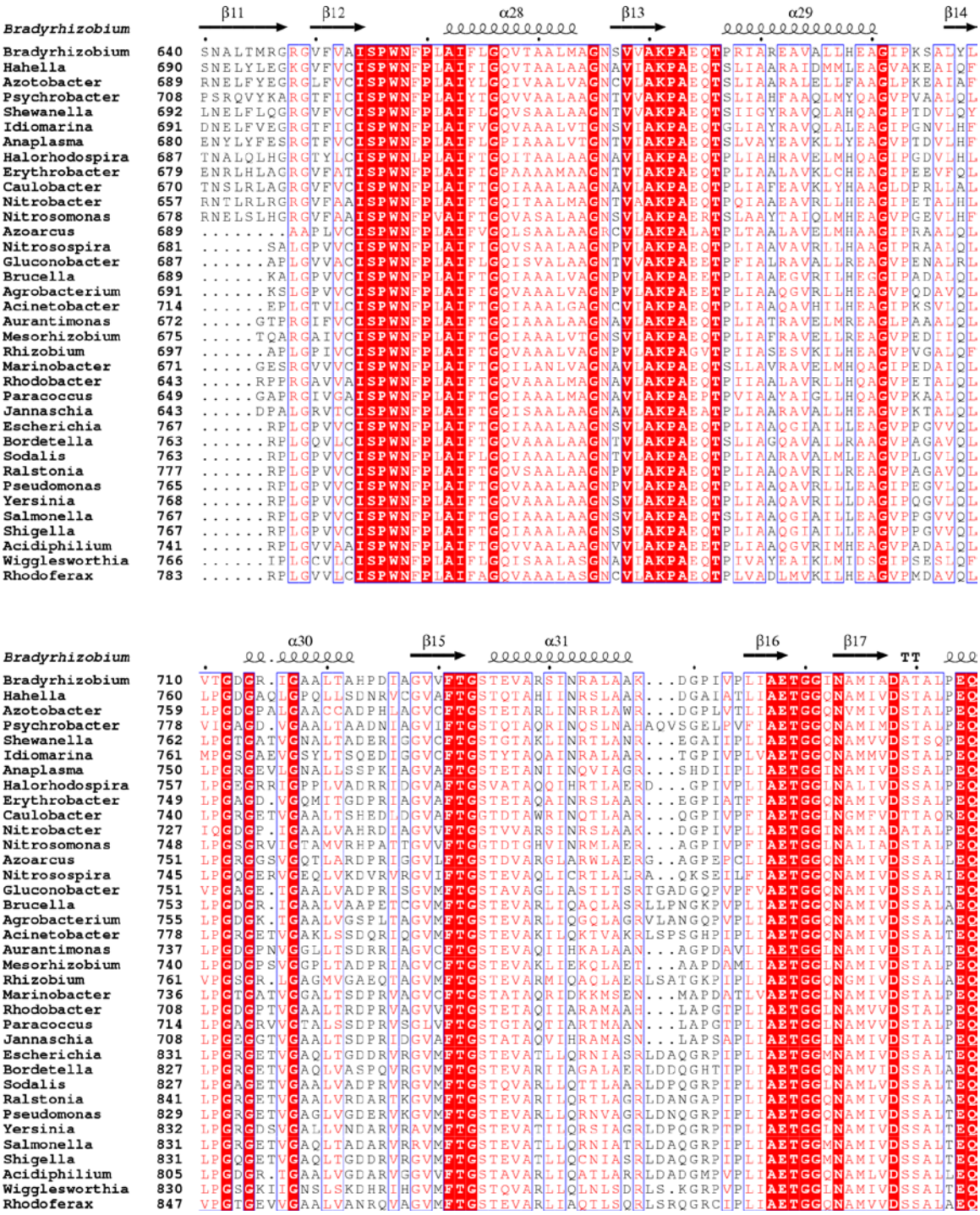

Bradyrhizobium
η4
β5
TT
α14
β6
α15
Bradyrhizobium 292 IAATLADPSL.....KGWDGFGLAIQAYQKRASAVIDYVDALARAHDRKLMVRLVKGAYWDTEIKRAQE
Hahella 311 FEKLYRSPDS.....RGWGEFGLVVQAYSKRALPVLCWLTALAAEQGDRIPVRLVKGAYWDSEIKLCQQ
Azotobacter 310 FEKLLCDPAI.....RGWGEFGLVVQAYSKRALPVLVWLTLLGREQDTRIPVRLVKGAYWDSEIKQCQV
Psychrobacter 322 FESLYRDNLT.....ADWDGLGLVVQGYSKRAIAILAWLARLAIEVGDRIPVRLVKGAYWDTEIKLAQQ
Shewanella 314 FQKLFNSEAT.....KGWGLLGIVVQAYSKRALPVLVWLTRLAKEQGDEIPVRLVKGAYWDSELKWAQQ
Idiomarina 311 FEKVYRSGVC.....KGWPRFGLVVQAYSKRALPTLCWITALAKECGDEIPVRLVKGAYWDNEIKWTQE
Anaplasma 299 LDKVFSDSSL.....NGWEGLGFVIQAYQKRALAALDFVEDVAIRANRKMVLRLVKGAYWDYEIRNAQE
Halorhodospira 308 LEAVMADPEL.....ADWDGLGIAVQAYQKRAPEWIDWLAERAGHYRRRLRIRLVKGAYWDIEIKDSQI
Erythrobacter 297 IERLAADDALFTKADGTRWDGFGLAIQAYQKRAAPLCDWIVKLARRHDRRFMVRLVKGAYWDTEIKLSQV
Caulobacter 292 LDKLCREPEL.....GDWTGLGLAVQAYQKRCGEVIARLKALSEETGRRLMVRLVKGAYWDSEIKRAQV
Nitrobacter 308 IAAVAANPSL.....AGWSGFGLAIQAYQKRASAVIDYIDDLARHLNRRMMVRLVKGAYWDTEIKRAQE
Nitrosomonas 301 FEALSRTREL.....AGWEGLGLAVQIYQKRAPAVIDWLAAIAREHKRRIPVRLVKGAYWDSEIKRAQE
Azoarcus 308 FEALVADPLL.....AGWDGLGFVVQAYQKRAPFVIDYLVDLAHRSGRRLMIRLVKGAYWDSEIKRAQV
Nitrosospira 305 LEALVFDPDL.....AGWDGIGFVVQAYQKRCPFVVDWLIDLARRSGHRLMVRLVKGAYWDGEIKRAQV
Gluconobacter 317 IEDLCHCPEL.....DGWNGIGIVVQAYGRRAPKVLDYLIDLGRRSGHRLMIRLVKGAYWDSEIKKAQV
Brucella 315 LQSLCEDPDL.....AGWDGIGFVVQAYGKRCPLVLDFIIDLARRTKRRVMVRLVKGAYWDAEIKRAQV
Agrobacterium 322 LEELALDKDL.....AGWNGLGFVVQAYGRRCPFVLDYIIDLARRSGRRIMVRLVKGAYWDAEIKRAQV
Acinetobacter 328 LERLCFEPEL.....ADWKGIGFVIQAYQKRCFYVVDYVVDLAKRSKKRLMIRLVKGAYWDSEIKKAQI
Aurantimonas 301 IEAVLSDESL.....KGWDGFGVVVQAYGRRAVPVIEYLDTLARRLDRRIMVRLVKGAYWDTEIKLAQE
Mesorhizobium 309 IERVLAEPEL.....AGWNGFGVVVQAYGFRAAFVIDWLYALARKYDRTIMVRLVKGAYWDTEIKRAQT
Rhizobium 322 LEELCFAPDL.....AGWSGLGFVVQAYGKRCPFVIDFIIDLARRSGRRMMVRLVKGAYWDAEIKRTQL
Marinobacter 302 IEALVADPEL.....AGWDGFGVVVQAYGKRSAFALDWLYGLAEKYDRRFMVRLVKGAYWDAEIKRAQV
Rhodobacter 292 MAAVIADPET.....AGWEGFGAVVQAYGKRAGAAIDALAAMARAAGRRINIRLVKGAYWDAEMKRAQV
Paracoccus 298 IEAVLSDPSL.....AGWDGFGVVVQAYGKRAGQVIDWLYETATRLDRRIMVRLVKGAYWDTEIKRAQV
Jannaschia 296 IEAVLRDESL.....AGWDGFGVVVQAYGKRCAPVIDWLYALAEIHDRRIMVRLVKGAYWDTEIKRAQV
Escherichia 384 LEKLCFEPEL.....AGWNGIGFVIQAYQKRCPLVIDYLIDLATRSRRRLMIRLVKGAYWDSEIKRAQM
Bordetella 377 LEALCFVPEL.....DDWNGIGFVIQAYQKRAPFVIDYVIDLAKRSRHRLMVRLVKGAYWDIEIKRAQV
Sodalis 380 LARLCFEPAL.....AGWNGIGFVIQAYQKRCPQVIDELIDLAKRSQRRLMIRLVKGAYWDSEIKRAQI
Ralstonia 394 LERLCFAPEL.....AGWNGLGFVVQGYQKRCPFVLDYIIDLARRSKHRLMIRLVKGAYWDSEVKRAQV
Pseudomonas 382 LERLCFEPSL.....AGWNGIGFVIQAYQKRCPYVIDYVIDLAKRSRHRLMIRLVKGAYWDSEIKRAQV
Yersinia 385 LERLCFEPQL.....AGWNGIGFVIQAYQKRCPSTIDAVIDMAQRSRRRLMIRLVKGAYWDSEIKRAQI
Salmonella 384 LEKLCFEPEL.....AGWNGIGFVIQAYQKRCPLVIDYLVDLASRSRRRLMIRLVKGAYWDSEIKRAQM
Shigella 384 LENLCFEPEL.....AGWNGIGFVIQAYQKRCPLVIDYLVDLATRSRRRLMIRLVKGAYWDSEIKRAQM
Acidiphilium 368 LEGLCFEPGL.....AGWNGIGFVVQAYQKRARSVIEFLIDLARRSRHRLMVRLVKGAYWDSEIKRAQL
Wigglesworthia 383 LDKLCCEPSL.....FGWSGIGFVVQAYQKRCMYVIEELIDIAKRTKRRLMVRLVKGAYWDSEIKNAQI
Rhodoferax 389 LESLCFDPRL.....KGWNGIGFVVQAYLKRCPYVIDHVIDLARRSGRRLMIRLVKGAYWDAEIKRAQL
α16
β7
α17
β8
Bradyrhizobium 356 RGLDGYPVFTRKAMTDLNYVACASKLLALRPRIFPQFATHNALTVATVLEMAEGSSGFEFQR
Hahella 375 RGLPSYPVYTRKESTDVSYLACARFLFSEYARNHIYPQFATHNAHTVASVQALAEKNDRPFEFQR
Azotobacter 374 QGLEGYPVFTRKEGTDTSYLACARYLLGEHCRGVLYPQFATHNAHTVSCILAMAETHPPHDFEFQR
Psychrobacter 386 KGLSGYAVWTRKEGTDTAYLACARFLLSEHLRGLIWPQFATHNAHTLASIMTMSSHRDFEFQR
Shewanella 378 AGEAAYPLYTRKAGTDVSYLACARYLLSDATRGAIYPQFASHNAQTVAAISDMAGDRNHEFQR
Idiomarina 375 NGLLGYPVFTRKSHSDISYLACARYLLSDDTDGAIYPQFATHNAQTFMAIQQMNETHQRRIEYQR
Anaplasma 363 MGLDGYPVFTRRAYTDVSYLACVQRVLSKPGTFYPCFATHNAYTLSFVLEMADKHPGFEFQR
Halorhodospira 372 QGLDDYPVFTRKAASDVCFLACARRMLRHPQQIYPQFATHNAHTVAAVMELADEQPFEFQR
Erythrobacter 367 GGYEDFPVFTRKVATDVSYLACAAKLLEAPDAIYPAFATHNAYTVGAVKALAGHTEFEFQR
Caulobacter 356 AGRPDYPVFTTKPATDLSYLVNAKALIEAAPHLYAQFATHNAHTLAAVVRMAKNTVGVKIEHQR
Nitrobacter 372 RGLDGYPVFTRKAMTDLNYIACAKKLLTLRPHLFPQFATHNALTVATLLELTDGADDGYEFQR
Nitrosomonas 365 LGIDGFPVFTRKASTDVSYLACARMLLADRSAFFPMFATHNAHTLAAVKVLAGNRSDYEFQR
Azoarcus 372 EGQAGYPVYTRKAHTDLAYLVCAARLLAEAGAVYPQFATHNARTVAEVHEMAQCVGAGGTLPAYEFQC
Nitrosospira 369 EGLDGYPVYTRKIHTDICYMVCAQKLLRAAEAVFPQFATHNAHTLSAVYQIAQDADIEDYEFQY
Gluconobacter 381 EGQTDFPVYTRKCYTDVSYIACAKKLLAARDVVFPQFATHNARTLATIYTLAGLNFQIGDYEFQC
Brucella 379 DGLEDFPVYTRKVHTDVSYIACARKLLAATDVIFPQFATHNAQTLATIYHLAGPDFKTGKFEFQC
Agrobacterium 386 EGLEDFPVFTRKVHTDVSYIACARKLLDARDLVFPQFATHNAQSMATIYHLAGPDFKLGDYEFQC
Acinetobacter 392 DGMHDYPVFTRKVHTDLSYIACAKKLLAAPEQIYPQFATHNAQSLATIYELADPGKYYAGQYEFQC
Aurantimonas 365 MGTHTFPVFTRKVNTDLSYMACAQMLLDRRDRIYPQFATHNAHTCAAILKMAGNDRDSFEFQR
Mesorhizobium 373 LGLAGYPVFTRKANTDVSYMACAKKLLGMTDRIYPQFATHNAHTVAAILSMATNRDSFEFQR
Rhizobium 386 DGLDDYPVYTRKIYTDVAYIACARKLLAAADAVFPQFATHNAQTLATIYHLAGPDFAVGKYEFQC
Marinobacter 366 MGLSGFPVFTRKACSDVSFLSCATKLLNMTDRIYPQFATHNAHSVSAVLELAKSKGVDNYEFQR
Rhodobacter 356 EGHPGFPLFTSKTGTDVAYICLAAKLFGLNDCIYPQFATHNAHTVAAVLEMAAGRPFEFQR
Paracoccus 362 EGFPGFPLFTSKVATDVSYIANARKLIGYADRIYPQFATHNAQTVAAILEMAGETRFEFQR
Jannaschia 360 AGLPGFPVYTAKAATDVSYLCCARQLLDRTDRIYPQFATHNAHTVAATIEMAEDPLAFEFQR
Escherichia 448 DGLEGYPVYTRKVYTDVSYLACAKKLLAVPNLIYPQFATHNAHTLAAIYQLAGQNYYPGQYEFQC
Bordetella 441 DGLEGYPVYTRKVYTDVAYLACARRLLAAPEATYPQFATHNAYTLSAIYQMAGQNYYPGQYEFQC
Sodalis 444 EGLEDYPVFTRKVYTDLSYLACARKLLSVPHYIYPQFATHNAHTLAAIYHFAGQNYYPGQYEFQC
Ralstonia 458 DGLEGYPVYTRKVYTDVSYLACARKLLAAPDAVFPQFATHNAHTLAAIYHMAGQNYYPGQYEFQC
Pseudomonas 446 EGLEGYPVYTRKVYTDVSYVACARKLLAVPEAIYPQFATHNAHTLSAIYHIAGQNYYPGQYEFQC
Yersinia 449 DGLEGYPVYTRKVYTDVSYLACARKLLAVPNLIYPQFATHNAHTLSAIYHLAGQNYYPGQYEFQC
Salmonella 448 EGLEGYPVYTRKVYTDVSYLACAKKLLAVPNLIYPQFATHNAHTLAAIYHLAGQNYYPGQYEFQC
Shigella 448 DGLEGYPVYTRKVYTDVSYLACAKKLLAVPNLIYPQFATHNAHTLAAIYQLAGQNYYPGQYEFQC
Acidiphilium 432 DGLEGFPVFTRKIHTDLSYLACARAMLAARTEIFPQFATHNALTLASIHALAGPAFEVGDYEFQC
Wigglesworthia 447 EGIEDYPVFTRKAYTDISYLACAKKLLSESNHIYPQFATHNAHTVSSIYHFAGENYYSGQYEFQC
Rhodoferax 453 DGLDGYPVYTRKVHTDVSYLACARKLLAAPDAVYPQFATHNAQTVASIYQMAGNNYYAGQYEFQC

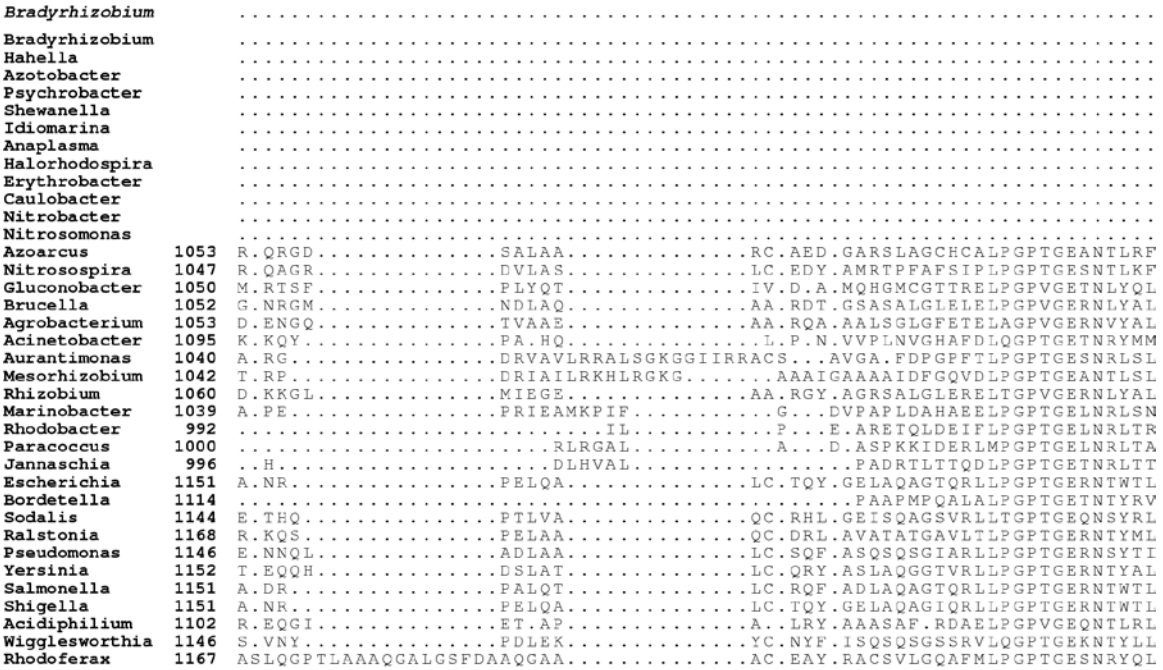

Bradyrhizobium
α18
β9
η5
α19
α20
α21
α22
η6
η7
α23

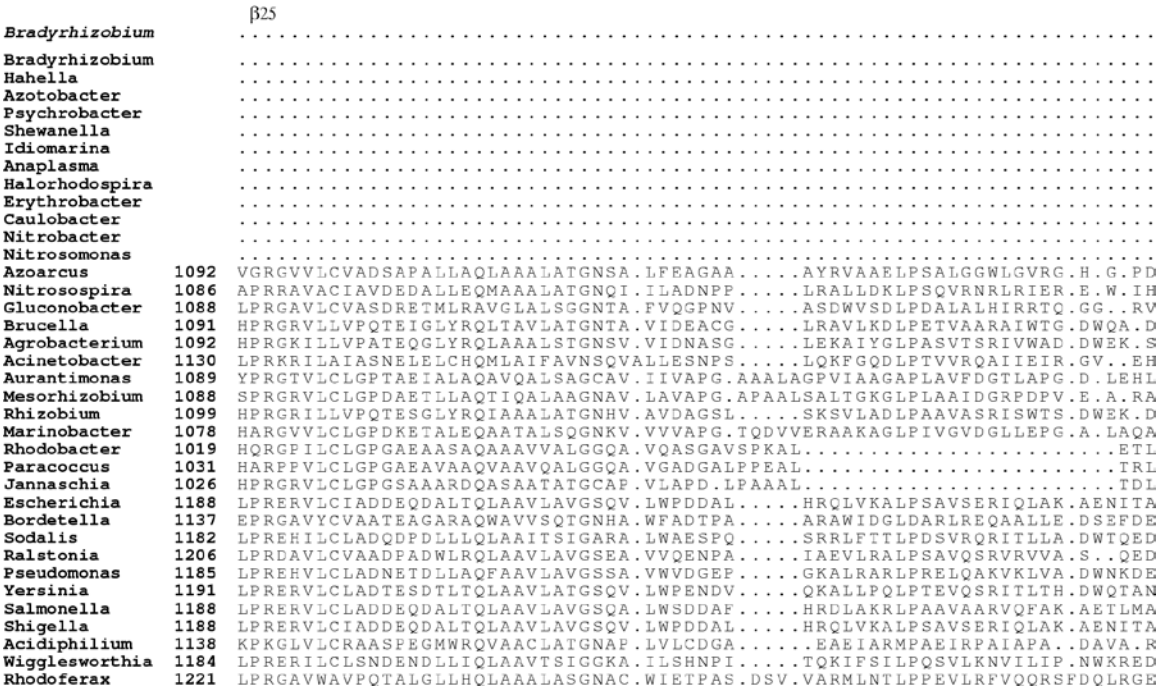

α24
α25
α26
α27
β10
Bradyrhizobium
Bradyrhizobium 537
Hahella 562
Azotobacter 563
Psychrobacter 571
Shewanella 565
Idiomarina 563
Anaplasma 551
Halorhodospira 557
Erythrobacter 553
Caulobacter 542
Nitrobacter 554
Nitrosomonas 551
Azoarcus 565
Nitrosospira 560
Gluconobacter 573
Brucella 570
Agrobacterium 578
Acinetobacter 592
Aurantimonas 552
Mesorhizobium 559
Rhizobium 578
Marinobacter 554
Rhodobacter 535
Paracoccus 541
Jannaschia 541
Escherichia 647
Bordetella 640
Sodalis 643
Ralstonia 656
Pseudomonas 645
Yersinia 648
Salmonella 647
Shigella 647
Acidiphilium 623
Wigglesworthia 647
Rhodoferax 656
Bradyrhizobium 571
Hahella 621
Azotobacter 621
Psychrobacter 639
Shewanella 623
Idiomarina 622
Anaplasma 610
Halorhodospira 618
Erythrobacter 610
Caulobacter 601
Nitrobacter 588
Nitrosomonas 609
Azoarcus 628
Nitrosospira 619
Gluconobacter 625
Brucella 628
Agrobacterium 630
Acinetobacter 652
Aurantimonas 611
Mesorhizobium 616
Rhizobium 636
Marinobacter 612
Rhodobacter 586
Paracoccus 592
Jannaschia 586
Escherichia 705
Bordetella 701
Sodalis 701
Ralstonia 715
Pseudomonas 703
Yersinia 706
Salmonella 705
Shigella 705
Acidiphilium 679
Wigglesworthia 705
Rhodoferax 721

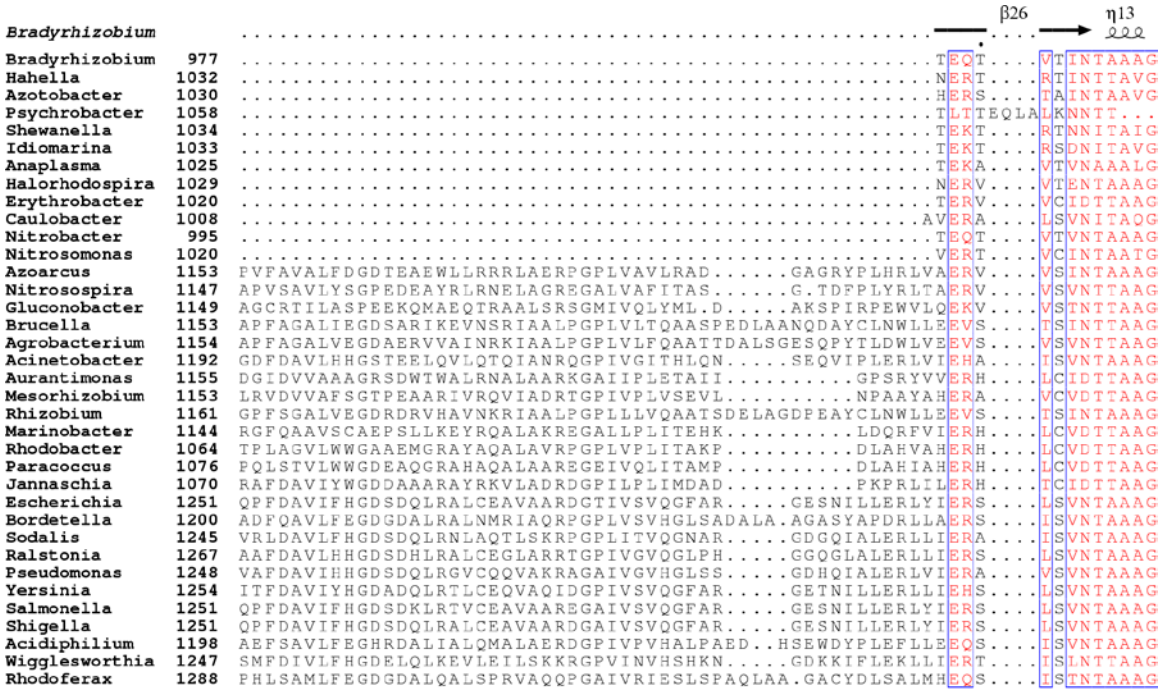

β11 β12 α28 β13 α29 β14
Bradyrhizobium
Bradyrhizobium 640 SNALTMRGRGVFVAISPWNFPLAIFLGQVTAALMAGNSVVAKPAEQTPRIARAEAVALLHEAGIPKSALYL
Hahella 690 SNELYLEGKGVFVCISPWNFPLAIFIGQVTAALAAGNAVIAKPAEQTSLIAARAIDMMLEAGVAKEAIQF
Azotobacter 689 RNELFYEGRGLFVCISPWNFPLAIYLGQIVAALVAGNCVLAKPAEQTSLIAARALELLFAAGLPKEAIAF
Psychrobacter 708 PSRQVYKARGTFICISPWNFPLAIYTGQVVAALAAGNTVVAKPAEQTSLIAHFAAQLMYQAGVPVAALQL
Shewanella 692 LNELFLQGRGVFVCISPWNFPLAIFLGQVSAALAAGNTVVAKPAEQTSIIGYRAVQLAHQAGIPTDVLQY
Idiomarina 691 DNELFVEGRGTFICISPWNFPLAIFVGQVAAALVTGNSVIAKPAEQTGLIAYRAVQLALEAGIPGNVLHF
Anaplasma 680 ENYLYFESRGTFVCISPWNFPLAIFLGPIAAALVTGNTVIAKPAEQTSLVAYEAVKLLYEAGVPTDVLHF
Halorhodospira 687 TNALQLHGRGTYLCISPWNLPLAIFTGQITAALAAGNAVIAKPAEQTPLIAHRAVELMHQAGIPGDVLHL
Erythrobacter 679 ENRLHLAGRGVFATISPWNFPLAIFIGPAAAAMAAGNTVIAKPAEQTPLIAALAVKLCHEAGIPEEVFQL
Caulobacter 670 TNSLRLAGRGVFVCISPWNFPLAIFTGQIAAALAAGNAVLAKPAEQTPLIAFEAVKLYHAAGLDPRLLAL
Nitrobacter 657 RNTLRLRGRGVFAAISPWNFPLAIFTGQITAALMAGNTVAAKPAEQTPQIAAEAVRLLHEAGIPETALHL
Nitrosomonas 678 RNELSLHGRGVFAAISPWNFPVAIFTGQVASALAAGNSVLAKPAERTSLAAYTAIQLMHEAGVPGEVLHF
Azoarcus 689 .....AAPLVCISPWNFPLAIFVGQLSAALAAGRCVLAKPALATPLTAALAVELMHAAGIPRAALQL
Nitrosospira 681 .....SALGPVVCISPWNFPLAIFIGQVSAALAAGNPVLAKPAEQTPLIAAVAVRLLHAAGIPRAALQL
Gluconobacter 687 .....APLGVVACISPWNFPLAIFLGQISVALAAGNTVVAKPAEETPFIALRAVALLREAGVPENALRL
Brucella 689 .....KALGPVVCISPWNFPLAIFTGQIAAALVAGNPVLAKPAEETPLIAAEGVRILHEGGIPADALQL
Agrobacterium 691 .....KSLGPVVCISPWNFPLAIFIGQVAAALVAGNPVLAKPAEETPLIAAQGVRLLHEAGVPQDAVQL
Acinetobacter 714 .....EPLGTVLCISPWNFPLAIFTGQIAAALGAGNCVIAKPAEQTPLIAAQAVHILHEAGIPKSVLQL
Aurantimonas 672 .....GTPRGIFVCISPWNFPLAIFTGQIAAALAAGNAVLAKPAEQTPLIATRAVELMREAGLPAAALQL
Mesorhizobium 675 .....TQARGAIVCISPWNFPLAIFTGQIAAALVTGNSVIAKPAEQTPLIAFRAVELLREAGVPEDIIQL
Rhizobium 697 .....APLGPIVCISPWNFPLAIFTGQVAAALVAGNPVLAKPAGVTPIIASESVKILHEAGVPVGALQF
Marinobacter 671 .....GESRGVVVCISPWNFPLAIFTGQILANLVAGNAVLAKPAEQTSLLAVRAVELMHQAGIPEDAIQL
Rhodobacter 643 .....RPPRGAVVAISPWNFPLAIFTGQVAAALMAGNAVLAKPAEQTPIIAALAVRLLHQAGVPETALQL
Paracoccus 649 .....GAPRGIVGAISPWNFPLAIFTGQIAAALMAGNAVIAKPAEPTPVIAAYAIGLLHQAGVPKAALQL
Jannaschia 643 .....DPALGRVTCISPWNFPLAIFTGQISAALAAGNAVLAKPAEATPLIAARAVALLHEAGVPKTALQL
Escherichia 767 ......RPLGPVVCISPWNFPLAIFTGQIAAALAAGNSVLAKPAEQTPLIAAQGIALLLEACVPPGVVQL
Bordetella 763 ......RPLGQVLCISPWNFPLAIFTGQVAAALAAGNTVLAKPAEQTSLIAGQAVAILRAAGVPAGAVQL
Sodalis 763 ......RPLGPVVCISPWNFPLAIFLGQIAAALAAGNTVLAKPAEQTPLIAARAVALMLEAGVPLGVLQL
Ralstonia 777 ......RPLGPVVCISPWNFPLAIFTGQVSAALAAGNPVLAKPAEQTPLIAAQAVRILRRAGVPAGAVQL
Pseudomonas 765 ......RPLGPVVCISPWNFPLAIFTGQVAAALAAGNPVLAKPAEQTPLIAAQAVRLLLEAGIPEGVLQL
Yersinia 768 ......RPLGPVVCISPWNFPLAIFTGQVAAALAAGNSVLAKPAEQTPLIAAQAVRILLDAGIPQGVLQL
Salmonella 767 ......RPLGPVVCISPWNFPLAIFTGQIAAALAAGNSVLAKPAEQTSLIAAQGIALLLEAGVPFPGVVQL
Shigella 767 ......RPLGPVVCISPWNFPLAIFTGQIAAALAAGNSVLAKPAEQTPLIAAQGIAILLEAGVPFPGVVQL
Acidiphilium 741 ......RPLGVVAAISPWNFPLAIFTGQVVAALAAGNVVLAKPAEETPLIAAQGVRIMHEAGVPADALQL
Wigglesworthia 766 ......IPLGCVLCISPWNFPLAIFLGQISAALASGNAVIAKPAEQTPIVAYEAIKLMIDSGIPSESLQF
Rhodoferax 783 ......RPLGVVLCISPWNFPLAIFAGQVAAALASGNCVLAKPAEQTPLVADLMVKLLHEAGVPMDAVQL
α30 β15 α31 β16 β17 TT
Bradyrhizobium
Bradyrhizobium 710 VTGDGR.IGAALTAHPDIAGVVFTGSTEVARSINRALAAK...DGPIVPLIAETGGINAMIADATALPEQ
Hahella 760 LPGDGAQLGPQLLSDNRVCGVAFTGSTQTAHIINRSLAAR...DGAIATLIAETGGQNAMIVDSTALPEQ
Azotobacter 759 LPGDGPALGAACCADPHLAGVCFTGSTETARLINRRLAWR...DGPLVTLIAETGGQNVMIVDSTALPEQ
Psychrobacter 778 VIGAGD.VGAALTAADNIAGVIFTGSTQTAQRINQSLNAHAQVSGELPVFIAETGGQNAMIMDSTALPEQ
Shewanella 762 LPGTGATVGNALTADERIGGVCFTGSTGTAKLINRTLANR...EGAIIPLIAETGGQNAMVVDSTSQPEQ
Idiomarina 761 MPGSGAEVGSYLTSQEDIGGVCFTGSTYTAQAINRALAAR...TGPIVPLVAETGGQNAMMVDSTALPEQ
Anaplasma 750 LPGREVLGNALLSSPKIAGVAFTGSTETANIINQVIAGR...SHDIIPLIAETGGINAMIVDSSALPEQ
Halorhodospira 757 LPGEGRRIGPPLVADRRIDGVAFTGSVATAQQIHRTLAERD...GPIVPLIAETGGINALIVDSSALPEQ
Erythrobacter 749 LPGAGD.VGQMITGDPRIAGVAFTGSTETAQAINRSLAAR...EGPIATFIAETGGQNAMIVDSSALPEQ
Caulobacter 740 LPGRGETVGAALTSHEDLDGVAFTGGTDTAWRINQTLAAR...QGPIVPFIAETGGINGMFVDTTAQREQ
Nitrobacter 727 IQGDGP.IGAALVAHRDIAGVVFTGSTVVARSINRSLAAK...DGPIVPLIAETGGINAMIADATALPEQ
Nitrosomonas 748 LPGSGRVIGTAMVRHPATTGVVFTGGTDTGHVINRMLAER...AGPIVPFIAETGGINALIADSTALPEQ
Azoarcus 751 LPGRGGSVGQTLARDPRIGGVLFTGSTDVARGLARWLAERG..AGPEPCLIAETGGQNAMIVDSSALLEQ
Nitrosospira 745 LPGQGERVGEQLVKDVRVRGVIFTGSTEVAQLICRTLALRA..QKSEILFIAETGGQNAMIVDSSARIEQ
Gluconobacter 751 VPGAGE.IGAALVADPRISGVMFTGSTAVAGLIASTLTSRTGADGQPVPFVAETGGQNAMIVDSSALTEQ
Brucella 753 LPGDGR.IGAALVAAPETCGVMFTGSTEVARLIQAQLASRLLPNGKPVPLIAETGGQNAMIVDSSALAEQ
Agrobacterium 755 LPGDGK.TGAALVGSPLTAGVMFTGSTEVARLIQGQLAGRVLANGQPVPLIAETGGQNAMIVDSSALAEQ
Acinetobacter 778 LPGRGETVGAKLSSDQRIQGVMFTGSTEVAKILQKTVAKRLSPSGHPIPLIAETGGQNAMIVDSSALTEQ
Aurantimonas 737 LPGDGPNVGGLLTSDRRIAGVCFTGSTEVAQIIHKALAANN...AGPDAVLIAETGGLNAMIVDSTALTEQ
Mesorhizobium 740 LPGDGPSVGGPLTADPRIAGVCFTGSTEVAKLIEKQLAET...AAPDAMLIAETGGLNAMIVDSTALPEQ
Rhizobium 761 VPGSGR.LGAGMVGAEQTAGVMFTGSTEVARMIQAQLAERLSATGKPIPLIAETGGQNGMIVDSSALAEQ
Marinobacter 736 LPGTGATVGGALTSDPRVAGVCFTGSTATAQRIDKKMSEN...MAPDATLVAETGGLNAMIVDSTALPEQ
Rhodobacter 708 LPGDGPTVGAALTRDPRVAGVVFTGSTETAQIIARAMAAH...LAPGTPLIAETGGLNAMVVDSTALPEQ
Paracoccus 714 LPGAGRVVGTALSSDPRVSGLVFTGSTGTAQTIARTMAAN...LAPGTPLIAETGGLNAMVVDSTALPEQ
Jannaschia 708 LPGEGGTVGAALTSDPRIDGVAFTGSTATAQVIHRAMASN...LAPSAPLIAETGGLNAMIVDSTALPEQ
Escherichia 831 LPGRGETVGAQLTGDDRVRGVMFTGSTEVATLLQRNIASRLDAQGRPIPLIAETGGMNAMIVDSSALTEQ
Bordetella 827 LPGRGETVGAQLVASPQVRGVMFTGSTEVARIIAGALAERLDDQGHTIPLIAETGGQNAMVIDSSALAEQ
Sodalis 827 LPGAGETVGAALVADPRVRGVMFTGSTQVARLLQTTLAARLDPQGRPIPLIAETGGLNAMLVDSSALTEQ
Ralstonia 841 LPGRGETVGAALVRDARTKGVMFTGSTEVARILQRTLAGRLDANGAPIPLIAETGGQNAMIVDSSALAEQ
Pseudomonas 829 LPGRGETVGAGLVGDERVKGVMFTGSTEVARLLQRNVAGRLDNQGRPIPLIAETGGQNAMIVDSSALTEQ
Yersinia 832 LPGRGDSVGALLVNDARVRAVMFTGSTEVATILQRSIAGRLDPQGRPTPLIAETGGINAMIVDSSALTEQ
Salmonella 831 LPGRGETVGAQLTADARVRRVMFTGSTEVATLLQRNIATRLDAQGRPIPLIAETGGMNAMIVDSSALTEQ
Shigella 831 LPGQGETVGAQLTGDDRVRGVMFTGSTEVATLLQCNIASRLDAQGRPIPLIAETGGMNAMIVDSSALTEQ
Acidiphilium 805 LPGDGR.IGAALVGDARVGGVVFTGSTAVARLIQATLARRLDADGMPVPLIAETGGQNAMLVDSSALTEQ
Wigglesworthia 830 LPGSGKIIGNSLSKDHRIHGVMFTGSTQVARLLQLNLSDRLS.KGRPVPLIAETGGLNAMIVDSSALTEQ
Rhodoferax 847 VPGTGEVVGAALVANRQVAGVMFTGSTEVARLIAQTLSQRLSRQGRCIPLIAETGGQNAMVVDSSALAEQ


α32 η8 β18 α33 α34
Bradyrhizobium
Bradyrhizobium 776 VADDVVTSAFRSAGQRCSALRLLFVQEDVADRMIEMVAGAARELKIGDPSDVATHVGPVIDVEAKQRLDA
Hahella 827 VVKDVVQSAFASAGQRCSALRVLFVQKDIADGILELLSGALKELKVDHPKHLSTDVGPVIDAEAQAGLLK
Azotobacter 826 VVKDALHSAFTSAGQRCSALRVLYLQNEIAGPVIELLRGAMAELRVGLPQRRDTDVGPLIDTQARQALLE
Psychrobacter 847 VVRDAVLSAFGSAGQRCSACRILCVQEEMADDVIELLQGNMAELIVGNPLYATTDVGPVIDIDAKRGLEA
Shewanella 829 VVNDVVSSSFTSAGQRCSALRVLFLQEDIADRVIDVLQGAMDELVIGNPSSIKTDVGPVIDATAKANLDA
Idiomarina 828 VVTDIVASAFQSAGQRCSALRVLFVQDDVADRVLDLLRGAMEELQVGDPLLHETDVGPVIDGIAKTNLEQ
Anaplasma 817 VVGDVITSAFKSAGQRCSALRALFLQDEIADGVIELLLGAVAELRLGNPMSLSTDIGPVIDQQSLDMLNA
Halorhodospira 824 AVVDVLRSAFTSAGQRCSALRLLCIQEDIAEPFLAMLRGAMDALRVGDPRWLATDVGPVIDSDARARLEA
Erythrobacter 815 VARDVVASAFQSAGQRCSAQRMLYIQDDVYDDMLAMIRGAFAALQIGDPTDFAIDVGPVIDPDAKAALER
Caulobacter 807 VIDDVIVSAFGSAGQRCSALRLLFLPHDTADHIIEGLKGAMDALVLGDPALAVTDVGPVIDAEAKDALDK
Nitrobacter 793 VADDVVTSAFRSAGQRCSALRLLFVQDDVADRMIEMIAGAARELRIGDPSDPATHVGPVIDREAKQQLDA
Nitrosomonas 815 LVRDVLTSAFDSAGQRCSALRVLFLQEDIAERVISLIRGAMAELRIGDPMKLDTDIGPVIDGNSLAKLEA
Azoarcus 819 VVQDVLVSAFDSAGQRCSALRVLCVQRDIAEPLLTMLKDAMGELRIGDPAALATDIGPVIDNAARDALEA
Nitrosospira 813 VVQDALTSAFDSAGQRCSALRVLCLQEDIAEQVLEVLRGAMGELTVGDPRELATDVGPVIDADAQHALLD
Gluconobacter 820 VVADVLVSAFDSAGQRCSALRVLCVQEDCADRVLTMLRGAVEELRVGNPAELHTDVGPVISAEAQSGIQT
Brucella 822 VVFDVIASAFDSAGQRCSALRVLCLQEDVADRILTMLKGALRELSIGRTDQLKVDIGPVITDEAKNTIEK
Agrobacterium 824 VVADVIASAFDSAGQRCSALRILCLQEDVADRTLTMLKGALHELRIGRTDQLSVDVGPVITAEAKGIIEK
Acinetobacter 848 VVIDVLNSAFDSAGQRCSALRILCVQEDNSKTLVTMLKGAMQQLIMGNPFLLKTDIGPVIDLEAKQTIDQ
Aurantimonas 804 AVRDILIASFQSAGQRCSALRMLYVQEDARDRLLHMLEGAMDALVVGDPWRIDTDVSPVIDADARDDIAG
Mesorhizobium 807 AVRDILASAFQSAGQRCSALRVLYVQKDVEKKMLDMLKGAMEALNIGDPWRISTDVGPVIDDEAQASIRD
Rhizobium 830 VVADVIASAFDSAGQRCSALRVLCLQDDVADRTLNMLKGAFRELTIGRTDRLSIDVGPVINDGAKAEIDQ
Marinobacter 803 VVRDVLASSFQSAGQRCSALRMLYVQKDIADHLLELMLYGAMDELGIGDPWLLSTDVGPVIDETARKKIVD
Rhodobacter 775 AVRDVVASAFRSAGQRCSALRCLYVQDDIAPHLIGMLKGAMEELVSGDPARLSTDVGPVIDAEAKAGIET
Paracoccus 781 AVRDIVNSAFRSAGQRCSALRCLYVQEDVAPHLVEMIKGAMDELRLGDPWDLTTDVGPVIDPGAQKEIAD
Jannaschia 775 AVKDVVMSAFRSAGQRCSALRCLYVQKDIAPAFLKALYGAMDELSLGDAWPLANDIGPVIDARAQAGIAA
Escherichia 901 VVVDVLASAFDSAGQRCSALRVLCLQDEIADHTLKMLRGAMAECRMGNPGRLTTDIGPVIDSEAKANIER
Bordetella 897 VVFDVLTSAFDSAGQRCSALRVLCLQEDSADHVLTMLRGAMRELRVGNPDRLSVDVGPVIDAEARNGIQG
Sodalis 897 VIIDVVTSAFDSAGQRCSALRLLCIQEDVAERTLRMLRGAMAEYTLGNPERLATDIGPVIDSSAKAAIDS
Ralstonia 911 VVADVLSSAFDSAGQRCSALRVLCLQDDVADRVLAMLKGGMAELAMGNPDRLATDVGPVIDAEARDNIAG
Pseudomonas 899 VVIDVVSSAFDSAGQRCSALRVLCLQEDSADRVIEMLKGAMAESRLGCPDRLAVDIGPVIDAEAKAGIEK
Yersinia 902 VVTDVVASAFDSAGQRCSALRILCICDDVAEHTLQMLRGAMAECRMGNPERLSTDIGPVIDAEAKTGIER
Salmonella 901 VVVDVLASAFDSAGQRCSALRVLCLQDDIAEHTLKMLRGAMAECRMGNPGRLTTDIGPVIDSEAKANIER
Shigella 901 VVVDVLASAFDSAGQRCSALRVLCLQDEIADHTLKMLRGAMAECRMGNPGRLTTDIGPVIDSEAKANIER
Acidiphilium 874 AVGDIVTSAFDSAGQRCSALRILLVQDDIADPLLAMLRGAMEELAVGATDRLSSDVGPVITREARDNIRA
Wigglesworthia 899 VVSDIIVSAFDSAGQRCSALRLLCIQEDVSEKTLNMIKGAIDSYTVGNPEYFSTDIGPVIDKNAKINIDQ
Rhodoferax 917 VVGDVLSSAFDSAGQRCSALRLLCIQEDVAERVIGMIKDAMREWVMGNPDRMHTDVGPVIDEDARAQIEQ
β19 β20 η9 β21 η10
Bradyrhizobium
Bradyrhizobium 846 HIARMKTEARL.H..FAGPA.....PEGCFVAPHIFELTEAGQLTEEVFGPILHVVRYRPE....
Hahella 897 HIEEMKANSRW.N..AEAILPG.....DHNKGFFVTPSAFEIGSINELSKEHFGPILHIVRYAAE....
Azotobacter 896 HVDLLKGEERL.L..AEAGLAP.....ALNGHFLAPLAFEIGGIGELRKEHFGPILHVVRYAAE....
Psychrobacter 917 HIERMRAEPTA.TILAQTPMSA.SS.EVSQDQATFVLPTAIEVKSIDVIGGEHFGPILHVLRYEAR....
Shewanella 899 HIDHIKQVGKL.I..KQMSLPA.....GTENGHFVAPTAVEIDSIKVLEKEHFGPILHVIRYKAS....
Idiomarina 898 HISDIQQAGRL.I..ARAPLPD.....YAMGGTFVAPTAIEIDSISQLVKENFGPILHVIRFKTS....
Anaplasma 887 YVEEMQKKKIKLL..CKADIG....YLHGGEEGYFFPPHIFEIESMSQLSREVFGPVLHVIRYKPA....
Halorhodospira 894 HHEAMAAAGRV.V..HRTPL.....GQAGERGHFVPPSLYRLDAIEDLQEEFFGPMLHYTTWRAG....
Erythrobacter 885 HVARRKKAGRT.V..WRRKLPR.....GTANGCFVAPTIIELDSILDLKRENFGPILHVARYKAD....
Caulobacter 877 HLVRLKSDAKV.L..HALAAP.....AGGTFFAPVLAEIPTADFLERDEVFGPVLHVVRYKPE....
Nitrobacter 863 HIARMTREGRV.R..FAGQAP.....ASGNFVAPHIFELADARQLSEEVFGPILHVVRYRAS....
Nitrosomonas 885 HAKRMSREAGF.I..AEAPLSD.....ETRDGFFFAPRAYEIDAISRIERDVFGPILHVIRFKAT....
Azoarcus 889 HVARMQAAGRG.V..FRVPLP.....PACENGSFVAPTLIEIDGIGDVGREVFGPILHVLRFDAE....
Nitrosospira 883 HIEKMRAAGHT.V..FQAALP.....SVCANGTFVAPTLIEIDHIDELERDEVFGPVLHVVRFARE....
Gluconobacter 890 YIDASRAQNRA.V..WSLPL.....PDTTSNGTFIAPTIIEIDSLADLKGEVFGPVLHVLRFEAS....
Brucella 892 HIQAMRDLGRK.V..EQLPLG.....PETQNGTFVAPTIIEIESLRDLKREVFGPVLHVVRYKRD....
Agrobacterium 894 HIDGMRSLGHR.I..EQIALA.....GETGKGTFVPPTIIEMKSLADLKREVFGPVLHVIRFKRD....
Acinetobacter 918 HIQKMRSKGYP.V..YQLIFNHEAASQAQLSQGTFVTPTLIELPNLDDLERDEVFGPVLHVITYKHG....
Aurantimonas 874 YIASRKAAGQA.L..KTLPVP.....QDGLYAAPSVVSVGGIDDLERDEVFGPVLHVATFKAR....
Mesorhizobium 877 YCTRKGLEGRL.I..AKLEAP.....KDGRFVAPHVFRVKGIEEMERDEVFGPVLHVASFDAD....
Rhizobium 900 HIEQMRGAGRK.V..EQLPLP.....ASAAAGTFVPPTIIEIKALKDLTKEVFGPVLHVYRFKRN....
Marinobacter 873 HCDKYEAKGQL.K..KKMAVP.....EKGNFVSPAVISVSGIEELEEEIFGPVLHVATFAAK....
Rhodobacter 845 YLAANK..ARI.L..HRSTAP.....EGGHFVAPALLQVGGIADLERDEIFGPVLHLATFAAE....
Paracoccus 851 YIAANA..NRV.L..HKLAAP.....SRGWFIPPTMLKVSGIADLERDEIFGPVLHVATFRGD....
Jannaschia 845 HIATARADGRI.L..HEGTAP.....VLGTFIAPTAIRVTSIADMSEEIFGPVLHVATFKSS....
Escherichia 971 HIQTMRSKGRP.V..FQAVRDN.SEDAREWQSGTFVAPTLIELDDFAELQKEVFGPVLHVVRYRN....
Bordetella 967 HIEAMRSSGRP.V..DQVEL.....AGECRYGTFVPPTLIEIDHIGELTREVFGPVLHVLRYRRD....
Sodalis 967 HITAMRERGYT.V..WQTPAAE....AAGDIQGTFVPPLTLIELDSIEALTQEVFGPVLHVVRYQRR....
Ralstonia 981 HIDAMRAKGRR.V..HQAPVHA.S.VPAACAHGTFVPPTVIELDSLSDLTREIFGPVLHVVRWKRTAKDT
Pseudomonas 969 HIQGMREKGRP.V..YQVAIAD....AAEIKRGTFVMPTLIELDSFDELKREIFGPVLHVVRYNRR....
Yersinia 972 HIQAMRAKGRK.V..YQAARTN.SLDEKEWQRGTFIKPTLIELDSFDELQKEVFGPVLHVVRFQRQ....
Salmonella 971 HIQTMRAKGRP.V..FQAARDN.SDDAQEWQTGTFVMPTLIELENFAELEKEVFGPVLHVVRYNRN....
Shigella 971 HIQTMRSKGRP.V..FQAVRDN.SEDAREWQSGTFVAPTLIELDDFAEIHKEVFGPVLHVVRYNRN....
Acidiphilium 944 HVETMEGRGHK.V..FSVALP.....PETEGGWFVAPTMIEIDHPSALEREVFGPVLHVVRYRRE....
Wigglesworthia 969 HILNMKKSNHS.V..WQSEMSY....DTKQDLGNFVLPTLIELDDIDQIKEEVFGPVLHIVKFKYE....
Rhodoferax 987 HIERMRSDGQP.V..TRMARD.....ESGAQGHFVMPTLIEIDRIERLQREVFGPVLHVLRYRRD....


β11 β12 α28 β13 α29 β14
Bradyrhizobium 640 SNALTMRGRGVFVAISPWNFPLAIFLGQVTAALMAGNSVVAKPAEQTPRIARAEAVALLHEAGIPKSALYL
Hahella 690 SNELYLEGKGVFVCISPWNFPLAIFIGQVTAALAAGNAVIAKPAEQTSLIAARAIDMMLEAGVAKEAIQF
Azotobacter 689 RNELFYEGRGLFVCISPWNFPLAIYLGQIVAALVAGNCVLAKPAEQTSLIAARALELLFAAGLPKEAIAF
Psychrobacter 708 PSRQVYKARGTFICISPWNFPLAIYTGQVVAALAAGNTVVAKPAEQTSLIAHFAAQLMYQAGVPVAALQL
Shewanella 692 LNELFLQGRGVFVCISPWNFPLAIFLGQVSAALAAGNTVVAKPAEQTSIIGYRAVQLAHQAGIPTDVLQY
Idiomarina 691 DNELFVEGRGTFICISPWNFPLAIFVGQVAAALVTGNSVIAKPAEQTGLIAYRAVQLALEAGIPGNVLHF
Anaplasma 680 ENYLYFESRGTFVCISPWNFPLAIFLGPIAAALVTGNTVIAKPAEQTSLVAYEAVKLLYEAGVPTDVLHF
Halorhodospira 687 TNALQLHGRGTYLCISPWNLPLAIFTGQITAALAAGNAVIAKPAEQTPLIAHRAVELMHQAGIPGDVLHL
Erythrobacter 679 ENRLHLAGRGVFATISPWNFPLAIFIGPAAAAMAAGNTVIAKPAEQTPLIAALAVKLCHEAGIPEEVFQL
Caulobacter 670 TNSLRLAGRGVFVCISPWNFPLAIFTGQIAAALAAGNAVLAKPAEQTPLIAFEAVKLYHAAGLDPRLLAL
Nitrobacter 657 RNTLRLRGRGVFAAISPWNFPLAIFTGQITAALMAGNTVAAKPAEQTPQIAAEAVRLLHEAGIPETALHL
Nitrosomonas 678 RNELSLHGRGVFAAISPWNFPVAIFTGQVASALAAGNSVLAKPAERTSLAAYTAIQLMHEAGVPGEVLHF
Azoarcus 689 .....AAPLVCISPWNFPLAIFVGQLSAALAAGRCVLAKPALATPLTAALAVELMHAAGIPRAALQL
Nitrosospira 681 .....SALGPVVCISPWNFPLAIFIGQVSAALAAGNPVLAKPAEQTPLIAAVAVRLLHAAGIPRAALQL
Gluconobacter 687 .....APLGVVACISPWNFPLAIFLGQISVALAAGNTVVAKPAEETPFIALRAVALLREAGVPENALRL
Brucella 689 .....KALGPVVCISPWNFPLAIFTGQIAAALVAGNPVLAKPAEETPLIAAEGVRILHEGGIPADALQL
Agrobacterium 691 .....KSLGPVVCISPWNFPLAIFIGQVAAALVAGNPVLAKPAEETPLIAAQGVRLLHEAGVPQDAVQL
Acinetobacter 714 .....EPLGTVLCISPWNFPLAIFTGQIAAALGAGNCVIAKPAEQTPLIAAQAVHILHEAGIPKSVLQL
Aurantimonas 672 .....GTPRGIFVCISPWNFPLAIFTGQIAAALAAGNAVLAKPAEQTPLIATRAVELMREAGLPAAALQL
Mesorhizobium 675 .....TQARGAIVCISPWNFPLAIFTGQIAAALVTGNSVIAKPAEQTPLIAFRAVELLREAGVPEDIIQL
Rhizobium 697 .....APLGPIVCISPWNFPLAIFTGQVAAALVAGNPVLAKPAGVTPIIASESVKILHEAGVPVGALQF
Marinobacter 671 .....GESRGVVVCISPWNFPLAIFTGQILANLVAGNAVLAKPAEQTSLLAVRAVELMHQAGIPEDAIQL
Rhodobacter 643 .....RPPRGAVVAISPWNFPLAIFTGQVAAALMAGNAVLAKPAEQTPIIAALAVRLLHQAGVPETALQL
Paracoccus 649 .....GAPRGIVGAISPWNFPLAIFTGQIAAALMAGNAVIAKPAEPTPVIAAYAIGLLHQAGVPKAALQL
Jannaschia 643 .....DPALGRVTCISPWNFPLAIFTGQISAALAAGNAVLAKPAEATPLIAARAVALLHEAGVPKTALQL
Escherichia 767 ......RPLGPVVCISPWNFPLAIFTGQIAAALAAGNSVLAKPAEQTPLIAAQGIALLLEACVPPGVVQL
Bordetella 763 ......RPLGQVLCISPWNFPLAIFTGQVAAALAAGNTVLAKPAEQTSLIAGQAVAILRAAGVPAGAVQL
Sodalis 763 ......RPLGPVVCISPWNFPLAIFLGQIAAALAAGNTVLAKPAEQTPLIAARAVALMLEAGVPLGVLQL
Ralstonia 777 ......RPLGPVVCISPWNFPLAIFTGQVSAALAAGNPVLAKPAEQTPLIAAQAVRILRRAGVPAGAVQL
Pseudomonas 765 ......RPLGPVVCISPWNFPLAIFTGQVAAALAAGNPVLAKPAEQTPLIAAQAVRLLLEAGIPEGVLQL
Yersinia 768 ......RPLGPVVCISPWNFPLAIFTGQVAAALAAGNSVLAKPAEQTPLIAAQAVRILLDAGIPQGVLQL
Salmonella 767 ......RPLGPVVCISPWNFPLAIFTGQIAAALAAGNSVLAKPAEQTSLIAAQGIALLLEAGVPFPGVVQL
Shigella 767 ......RPLGPVVCISPWNFPLAIFTGQIAAALAAGNSVLAKPAEQTPLIAAQGIAILLEAGVPFPGVVQL
Acidiphilium 741 ......RPLGVVAAISPWNFPLAIFTGQVVAALAAGNVVLAKPAEETPLIAAQGVRIMHEAGVPADALQL
Wigglesworthia 766 ......IPLGCVLCISPWNFPLAIFLGQISAALASGNAVIAKPAEQTPIVAYEAIKLMIDSGIPSESLQF
Rhodoferax 783 ......RPLGVVLCISPWNFPLAIFAGQVAAALASGNCVLAKPAEQTPLVADLMVKLLHEAGVPMDAVQL
α30 β15 α31 β16 β17 TT
Bradyrhizobium 710 VTGDGR.IGAALTAHPDIAGVVFTGSTEVARSINRALAAK...DGPIVPLIAETGGINAMIADATALPEQ
Hahella 760 LPGDGAQLGPQLLSDNRVCGVAFTGSTQTAHIINRSLAAR...DGAIATLIAETGGQNAMIVDSTALPEQ
Azotobacter 759 LPGDGPALGAACCADPHLAGVCFTGSTETARLINRRLAWR...DGPLVTLIAETGGQNVMIVDSTALPEQ
Psychrobacter 778 VIGAGD.VGAALTAADNIAGVIFTGSTQTAQRINQSLNAHAQVSGELPVFIAETGGQNAMIMDSTALPEQ
Shewanella 762 LPGTGATVGNALTADERIGGVCFTGSTGTAKLINRTLANR...EGAIIPLIAETGGQNAMVVDSTSQPEQ
Idiomarina 761 MPGSGAEVGSYLTSQEDIGGVCFTGSTYTAQAINRALAAR...TGPIVPLVAETGGQNAMMVDSTALPEQ
Anaplasma 750 LPGREVLGNALLSSPKIAGVAFTGSTETANIINQVIAGR...SHDIIPLIAETGGINAMIVDSSALPEQ
Halorhodospira 757 LPGEGRRIGPPLVADRRIDGVAFTGSVATAQQIHRTLAERD...GPIVPLIAETGGINALIVDSSALPEQ
Erythrobacter 749 LPGAGD.VGQMITGDPRIAGVAFTGSTETAQAINRSLAAR...EGPIATFIAETGGQNAMIVDSSALPEQ
Caulobacter 740 LPGRGETVGAALTSHEDLDGVAFTGGTDTAWRINQTLAAR...QGPIVPFIAETGGINGMFVDTTAQREQ
Nitrobacter 727 IQGDGP.IGAALVAHRDIAGVVFTGSTVVARSINRSLAAK...DGPIVPLIAETGGINAMIADATALPEQ
Nitrosomonas 748 LPGSGRVIGTAMVRHPATTGVVFTGGTDTGHVINRMLAER...AGPIVPFIAETGGINALIADSTALPEQ
Azoarcus 751 LPGRGGSVGQTLARDPRIGGVLFTGSTDVARGLARWLAERG..AGPEPCLIAETGGQNAMIVDSSALLEQ
Nitrosospira 745 LPGQGERVGEQLVKDVRVRGVIFTGSTEVAQLICRTLALRA..QKSEILFIAETGGQNAMIVDSSARIEQ
Gluconobacter 751 VPGAGE.IGAALVADPRISGVMFTGSTAVAGLIASTLTSRTGADGQPVPFVAETGGQNAMIVDSSALTEQ
Brucella 753 LPGDGR.IGAALVAAPETCGVMFTGSTEVARLIQAQLASRLLPNGKPVPLIAETGGQNAMIVDSSALAEQ
Agrobacterium 755 LPGDGK.TGAALVGSPLTAGVMFTGSTEVARLIQGQLAGRVLANGQPVPLIAETGGQNAMIVDSSALAEQ
Acinetobacter 778 LPGRGETVGAKLSSDQRIQGVMFTGSTEVAKILQKTVAKRLSPSGHPIPLIAETGGQNAMIVDSSALTEQ
Aurantimonas 737 LPGDGPNVGGLLTSDRRIAGVCFTGSTEVAQIIHKALAANN...AGPDAVLIAETGGLNAMIVDSTALTEQ
Mesorhizobium 740 LPGDGPSVGGPLTADPRIAGVCFTGSTEVAKLIEKQLAET...AAPDAMLIAETGGLNAMIVDSTALPEQ
Rhizobium 761 VPGSGR.LGAGMVGAEQTAGVMFTGSTEVARMIQAQLAERLSATGKPIPLIAETGGQNGMIVDSSALAEQ
Marinobacter 736 LPGTGATVGGALTSDPRVAGVCFTGSTATAQRIDKKMSEN...MAPDATLVAETGGLNAMIVDSTALPEQ
Rhodobacter 708 LPGDGPTVGAALTRDPRVAGVVFTGSTETAQIIARAMAAH...LAPGTPLIAETGGLNAMVVDSTALPEQ
Paracoccus 714 LPGAGRVVGTALSSDPRVSGLVFTGSTGTAQTIARTMAAN...LAPGTPLIAETGGLNAMVVDSTALPEQ
Jannaschia 708 LPGEGGTVGAALTSDPRIDGVAFTGSTATAQVIHRAMASN...LAPSAPLIAETGGLNAMIVDSTALPEQ
Escherichia 831 LPGRGETVGAQLTGDDRVRGVMFTGSTEVATLLQRNIASRLDAQGRPIPLIAETGGMNAMIVDSSALTEQ
Bordetella 827 LPGRGETVGAQLVASPQVRGVMFTGSTEVARIIAGALAERLDDQGHTIPLIAETGGQNAMVIDSSALAEQ
Sodalis 827 LPGAGETVGAALVADPRVRGVMFTGSTQVARLLQTTLAARLDPQGRPIPLIAETGGLNAMLVDSSALTEQ
Ralstonia 841 LPGRGETVGAALVRDARTKGVMFTGSTEVARILQRTLAGRLDANGAPIPLIAETGGQNAMIVDSSALAEQ
Pseudomonas 829 LPGRGETVGAGLVGDERVKGVMFTGSTEVARLLQRNVAGRLDNQGRPIPLIAETGGQNAMIVDSSALTEQ
Yersinia 832 LPGRGDSVGALLVNDARVRAVMFTGSTEVATILQRSIAGRLDPQGRPTPLIAETGGINAMIVDSSALTEQ
Salmonella 831 LPGRGETVGAQLTADARVRRVMFTGSTEVATLLQRNIATRLDAQGRPIPLIAETGGMNAMIVDSSALTEQ
Shigella 831 LPGQGETVGAQLTGDDRVRGVMFTGSTEVATLLQCNIASRLDAQGRPIPLIAETGGMNAMIVDSSALTEQ
Acidiphilium 805 LPGDGR.IGAALVGDARVGGVVFTGSTAVARLIQATLARRLDADGMPVPLIAETGGQNAMLVDSSALLEQ
Wigglesworthia 830 LPGSGKIIGNSLSKDHRIHGVMFTGSTQVARLLQLNLSDRLS.KGRPVPLIAETGGLNAMIVDSSALTEQ
Rhodoferax 847 VPGTGEVVGAALVANRQVAGVMFTGSTEVARLIAQTLSQRLSRQGRCIPLIAETGGQNAMVVDSSALAEQ


Bradyrhizobium
Bradyrhizobium
Hahella
Azotobacter
Psychrobacter
Shewanella
Idiomarina
Anaplasma
Halorhodospira
Erythrobacter
Caulobacter
Nitrobacter
Nitrosomonas
Azoarcus 1053 R.QRGD.....SALAA.....RC.AED.GARSLAGCHCALPGPTGEANTLRF
Nitrosospira 1047 R.QAGR.....DVLAS.....LC.EDY.AMRTPFAFSIPLPGPTGESNTLKF
Gluconobacter 1050 M.RTSF.....PLYQT.....IV.D.A.MQHGMCGTTRELPGPVGETNLYQL
Brucella 1052 G.NRGM.....NDLAQ.....AA.RDT.GSASALGLELELPGPVGERNLYAL
Agrobacterium 1053 D.ENGQ.....TVAAE.....AA.RQA.AALSGLGFETELAGPVGERNVYAL
Acinetobacter 1095 K.KQY.....PA.HQ.....L.P.N.VVPLNVGHAFDLQGPTGETNRYMM
Aurantimonas 1040 A.RG.....DRVAVLRRALSGKGGIIRRACS...AVGA.FDPGPFTLPGPTGESNRLSL
Mesorhizobium 1042 T.RP.....DRIAILRKHLRGKG.....AAAIGAAAAIDFGQVDLPGPTGEANTLSL
Rhizobium 1060 D.KKGL.....MIEGE.....AA.RGY.AGRSALGLERELTGPVGERNLYAL
Marinobacter 1039 A.PE.....PRIEAMKPIF.....G...DVPAPLDAHAEELPGPTGELNRLSN
Rhodobacter 992 .....IL.....P...E.ARETQLDEIFLPGPTGELNRLTR
Paracoccus 1000 .....RLRGAL.....A...D.ASPKKIDERLMPGPTGELNRLTA
Jannaschia 996 ..H.....DLHVAL.....PADRTLTTQDLPGPTGETNRLTT
Escherichia 1151 A.NR.....PELQA.....LC.TQY.GELAQAGTQRLLPGPTGERNTWTL
Bordetella 1114 .....PAAPMPQALALPGPTGETNTYRV
Sodalis 1144 E.THQ.....PTLVA.....QC.RHL.GEISQAGSVRLLTGPTGEQNSYRL
Ralstonia 1168 R.KQS.....PELAA.....QC.DRL.AVATATGAVLTLPGPTGERNTYML
Pseudomonas 1146 E.NNQL.....ADLAA.....LC.SQF.ASQSQSGIARLLPGPTGERNSYTI
Yersinia 1152 T.EQQH.....DSLAT.....LC.QRY.ASLAQGGTVRLLPGPTGERNTYAL
Salmonella 1151 A.DR.....PALQT.....LC.RQF.ADLAQAGTQRLLPGPTGERNTWTL
Shigella 1151 A.NR.....PELQA.....LC.TQY.GELAQAGIQRLLPGPTGERNTWTL
Acidiphilium 1102 R.EQGI.....ET.AP.....A..LRY.AAASAF.RDAELPGPVGEQNTLRL
Wigglesworthia 1146 S.VNY.....PDLEK.....YC.NYF.ISQSQSGSSRVLQGPTGEKNTYLL
Rhodoferax 1167 ASLQGPTLAAAQGALGSFDAAQGAA.....AC.EAY.RACSVLGQAFMLPGPTGESNRYQL


β25
Bradyrhizobium
Bradyrhizobium
Hahella
Azotobacter
Psychrobacter
Shewanella
Idiomarina
Anaplasma
Halorhodospira
Erythrobacter
Caulobacter
Nitrobacter
Nitrosomonas
Azoarcus 1092 VGRGVVLCVADSAPALLAQLAAALATGNSA.LFEAGAA.....AYRVAAELPSALGGWLGVRG.H.G.PD
Nitrosospira 1086 APRRAVACIAVDEDALLEQMAAALATGNQI.ILADNPP.....LRALLDKLPSQVRNRLRIER.E.W.IH
Gluconobacter 1088 LPRGAVLCVASDRETMLRAVGLALSGGNTA.FVQGPNV.....ASDWVSDLPDALALHIRRTQ.GG..RV
Brucella 1091 HPRGRVLLVPQTEIGLYRQLTAVLATGNTA.VIDEACG.....LRAVLKDLPETVAARAIWTG.DWQA.D
Agrobacterium 1092 HPRGKILLVPATEQGLYRQLAAALSTGNSV.VIDNASG.....LEKAIYGLPASVTSRIVWAD.DWEK.S
Acinetobacter 1130 LPRKRILAIASNELELCHQMLAIFAVNSQVALLESNPS.....LQKFGQDLPTVVRQAIIEIR.GV..EH
Aurantimonas 1089 YPRGTVLCLGPTAEIALAQAVQALSAGCAV.IIVAPG.AAALAGPVIAAGAPLAVFDGTLAPG.D.LEHL
Mesorhizobium 1088 SPRGRVLCLGPDAETLLAQTIQALAAGNAV.LAVAPG.APAALSALTGKGLPLAAIDGRPDPV.E.A.RA
Rhizobium 1099 HPRGRILLVPQTESGLYRQIAAALATGNHV.AVDAGSL.....SKSVLADLPAAVASRISWTS.DWEK.D
Marinobacter 1078 HARGVVLCLGPDKETALEQAATALSQGNKV.VVVAPG.TQDVVERAAKAGLPIVGVDGLLEPG.A.LAQA
Rhodobacter 1019 HQRGPILCLGPGAEAASAQAAAVVALGGQA.VQASGAVSPKAL.....ETL
Paracoccus 1031 HARPPVLCLGPGAEAVAAQVAAVQALGGQA.VGADGALPPEAL.....TRL
Jannaschia 1026 HPRGRVLCLGPGSAAARDQASAATATGCAF.VLAPD.LPAAAL.....TDL
Escherichia 1188 LPRERVLCIADDEQDALTQLAAVLAVGSQV.LWPDDAL.....HRQLVKALPSAVSERIQLAK.AENITA
Bordetella 1137 EPRGAVYCVAATEAGARAQWAVVSQTGNHA.WFADTPA.....ARAWIDGLDARLREQAALLE.DSEFDE
Sodalis 1182 LPREHILCLADQDPDLLLQLAAITSIGARA.LWAESPQ.....SRRLFTTLPDSVRQRITLLA.DWTQED
Ralstonia 1206 LPRDAVLCVAADPADWLRQLAAVLAVGSEA.VVQENPA.....IAEVLRALPSAVQSRVRVVA.S..QED
Pseudomonas 1185 LPREHVLCLADNETDLLAQFAAVLAVGSSA.VWVDGEP.....GKALRARLPRELQAKVKLVA.DWNKDE
Yersinia 1191 LPRERVLCLADTESDTLTQLAAVLATGSQV.LWPENDV.....QKALLPQLPTEVQSRITLTH.DWQTAN
Salmonella 1188 LPRERVLCLADDEQDALTQLAAVLAVGSQA.LWSDDAF.....HRDLAKRLPAAVAARVQFAK.AETLMA
Shigella 1188 LPRERVLCIADDEQDALTQLAAVLAVGSQV.LWPDDAL.....HRQLVKALPSAVSERIQLAK.AENITA
Acidiphilium 1138 KPKGLVLCRAASPEGMWRQVAACLATGNAP.LVLCDGA.....EAEIARMPAEIRPAIAPA..DAVA.R
Wigglesworthia 1184 LPRERILCLSNDENDLLIQLAAVTSIGGKA.ILSHNPI.....TQKIFSILPQSVLKNVILIP.NWKRED
Rhodoferax 1221 LPRGAVWAVPQTALGLLHQLAAALASGNAC.WIETPAS.DSV.VARMLNTLPPEVLRFVQQRSFDQLRGE


β26
η13
Bradyrhizobium
Bradyrhizobium 977 ....TEQT....VTINTAAAG
Hahella 1032 ....NERT....RTINTTAVG
Azotobacter 1030 ....HERS....TAINTAAVG
Psychrobacter 1058 ....TLTTEQLALKNNTT...
Shewanella 1034 ....TEKT....RTNNITAIG
Idiomarina 1033 ....TEKT....RSDNITAVG
Anaplasma 1025 ....TEKA....VTVNAAALG
Halorhodospira 1029 ....NERV....VTENTAAAG
Erythrobacter 1020 ....TERV....VCIDTTAAG
Caulobacter 1008 ....AVERA....LSVNITAQG
Nitrobacter 995 ....TEQT....VTVNTAAAG
Nitrosomonas 1020 ....VERT....VCINTAATG
Azoarcus 1153 PVFAVALFDGDTEAEWLLRRRLAERPGPLVAVLRAD.....GAGRYPLHRLVAERV....VSINTAAAG
Nitrosospira 1147 APVSAVLYSGPEDEAYRLRNELAGREGALVAFITAS.....G.TDFPLYRLTAERV....VSVNTTAAG
Gluconobacter 1149 AGCRTILASPEEKQMAEQTRAALSRSGMIVQLYML.D.....AKSPIRPEWVLQEKV....VSTNTTAAG
Brucella 1153 APFAGALIEGDSARIKEVNSRIAALPGPLVLTQAASPEDLAANQDAYCLNWLLEEVS....TSINTTAAG
Agrobacterium 1154 APFAGALVEGDAERVVAINRKIAALPGPLVLFQAATTDALSGESQPYTLDWLVEEVS....VSVNTTAAG
Acinetobacter 1192 GDFDAVLHHGSTEELQVLQTQIANRQGPIVGITHLQN.....SEQVIPLERLVIEHA....TSVNTAAAG
Aurantimonas 1155 DGIDVVAAAGRSDWTWALRNALAARKGAIIPLETAII.....GPSRYVVERH....LCIDTTAAG
Mesorhizobium 1153 LRVDVVAFSGTPEAARIVRQVIADRTGPIVPLVSEVL.....NPAAYAHERA....VCVDTTAAG
Rhizobium 1161 GPFSGALVEGDRDRVHAVNKRIAALPGPLLLVQAATSDELAGDPEAYCLNWLLEEVS....TSINTAAAG
Marinobacter 1144 RGFQAAVSCAEPSLLKEYRQALAKREGALLPLITEHK.....LDQRFVIERH....LCVDTTAAG
Rhodobacter 1064 TPLAGVLWWGAAEMGRAYAQALAVRPGPLVPLITAKP.....DLAHVAHERH....LCVDTTAAG
Paracoccus 1076 PQLSTVLWWGDEAQGRAHAQALAAREGEIVQLITAMP.....DLAHIAHERH....LCVDTTAAG
Jannaschia 1070 RAFDAVIYWGDDAAARAYRKVLADRDGPILPLIMDAD.....PKPRLILERH....TCIDTTAAG
Escherichia 1251 QPFDAVIFHGDSDQLRALCEAVAARDGTIVSVQGFAR.....GESNILLERLYIERS....LSVNTAAAG
Bordetella 1200 ADFQAVLFEGDGDALRALNMRIAQRPGPLVSVHGLSADALA.AGASYAPDRLLAERS....ISVNTAAAG
Sodalis 1245 VRLDAVLFHGDSDQLRNLAQTLSKRPGPLITVQGNAR.....GDGQIALERLLIERA....ISVNTAAAG
Ralstonia 1267 AAFDAVLHHGDSDHLRALCEGLARRTGPIVGVQGLPH.....GGQGLALERLLIERS....LSVNTAAAG
Pseudomonas 1248 VAFDAVIHHGDSDQLRGVCQQVAKRAGAIVGVHGLSS.....GDHQIALERLVIERA....VSVNTAAAG
Yersinia 1254 ITFDAVIYHGDADQLRTLCEQVAQIDGPIVSVQGFAR.....GETNILLERLLIEHS....LSVNTAAAG
Salmonella 1251 QPFDAVIFHGDSDKLRTVCEAVAAREGAIVSVQGFAR.....GESNILLERLYIERS....LSVNTAAAG
Shigella 1251 QPFDAVIFHGDSDQLRALCEAVAARDGAIVSVQGFAR.....GESNILLERLYIERS....LSVNTAAAG
Acidiphilium 1198 AEFSAVLFEGHRDALIALQMALAERDGPIVPVHALPAED..HSEWDYPLEFLLEEQS....ISVNTAAAG
Wigglesworthia 1247 SMFDIVLFHGDELQLKEVLEILSKKRGPVINVHSHKN.....GDKKIFLEKLLIERT....ISLNTTAAG
Rhodoferax 1288 PHLSAMLFEGDGDALQALSPRVAQQPGAIVRIESLSPAQLAA.GACYDLSALMHEQS....LSTNTAAAG
Bradyrhizobium
Bradyrhizobium 990 GNAALLAG.....EE
Hahella 1045 GNATLLSLGV.....EHI
Azotobacter 1043 GNASLLSLT.....ERE
Psychrobacter 1072 ENITQTTI.....A
Shewanella 1047 GNATLLSLGD.....SDA
Idiomarina 1046 GNATLLSL.....DD
Anaplasma 1038 GSVSLVCLE.....DE
Halorhodospira 1042 GNASLFALGE....DDEA
Erythrobacter 1033 GNATLLA.....S
Caulobacter 1022 GDPALLN.....L
Nitrobacter 1008 GNAALMTE.....EG
Nitrosomonas 1033 GNTALLSLDEVACRGRDS
Azoarcus 1213 GNAALMTL.....G
Nitrosospira 1206 GNPGLMGLDFN..GSSTH
Gluconobacter 1209 GNASLMTI.....G
Brucella 1219 GNASLMAI.....G
Agrobacterium 1220 GNASLMSI.....G
Acinetobacter 1253 GNASLMTM.....SE
Aurantimonas 1211 GNASLLAA.....AE
Mesorhizobium 1209 GNASLLA.....AA
Rhizobium 1227 GNASLMAI.....G
Marinobacter 1200 GNASLIAA.....SE
Rhodobacter 1120 GNAALLA.....G
Paracoccus 1132 GNAALLA.....G
Jannaschia 1126 GNAALLAA.....VA
Escherichia 1312 GNASLMTI.....G
Bordetella 1265 GNASLMTI.....G
Sodalis 1306 GNASLMTI.....G
Ralstonia 1328 GNASLMTI.....G
Pseudomonas 1309 GNASLMTI.....G
Yersinia 1315 GNASLMTI.....G
Salmonella 1312 GNASLMTI.....G
Shigella 1312 GNASLMTI.....G
Acidiphilium 1262 GNASLMSI.....G
Wigglesworthia 1308 GNTTLLSM.....V
Rhodoferax 1353 GNAQLMTM.....D

VITA

Ranjan Kumar Singh was born in a small town of Jharkhand state in India. He attended Banaras Hindu University, Varanasi as an undergraduate and got his Bachelors in Chemistry (Hons.) in 2005. Later on, he moved to the capital of India New Delhi and attended his Masters in Chemistry from Indian Institute of Technology, New Delhi. After getting his Master's degree in July 2008 he came to USA to start his doctoral career in Department of Chemistry, University of Missouri, Columbia. Subsequently, he joined Professor John J. Tanner Group in January 2009 to start his research. In November 2012, Ranjan will start his postdoctoral career at University of Texas, Health Science Center Houston. in Profesor Zheng Lei research group.