STRUCTURAL AND FUNCTIONAL STUDIES OF PROLINE CATABOLIC ENZYMES AND HUMAN ALDEHYDE DEHYDROGENASES

A Dissertation

Presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

MIN LUO

Dr. John J. Tanner, Dissertation Supervisor

MAY 2015

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

STRUCTURAL AND FUNCTIONAL STUDIES OF PROLINE CATABOLIC ENZYMES AND HUMAN ALDEHYDE DEHYDROGENASES

presented by Min Luo,

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 Jason W. Cooley

Professor Kent S. Gates

For my parents, parents in law and wife.

献给我的父母，岳父母和爱人

ACKNOWLEDGEMENTS

This dissertation is a summary of the research I have performed in the University of Missouri-Columbia. On my way toward a PhD degree, many people contributed tirelessly to drive me to succeed. Here I acknowledge those who helped me successfully accomplish my PhD study.

The person I appreciate most is my advisor, Dr. Jack Tanner, who contributed largely to realizing my "Dr. Luo" dream. He is always patiently mentoring my learning, tolerantly correcting my mistakes, tirelessly giving me new ideas, and never stopping being encouraging. The biggest thing I appreciate is the flexibility he provided in my PhD, I am a person who has lots of weird ideas, curious to try some crazy experiments, and he is always there supporting me, smartly keeping me on the right track. Besides, his passion in science and diligence putting in research inspired me a lot, I never understand how much energy is inside him to keep working until the early morning every day, which became my biggest pressure, but also motivation to work hard in my PhD. Another important lesson I learned from him is the accurate assessment of scientific data. For any experiment, he always asks me to test from different experiments, check from different aspects, and finally draws a conclusion carefully. I cannot forget his description of the three stages for graduate students: 1. Cannot generate data, 2. Can generate data, and 3. Can generate and also analyze data. He is the person making me finish the transition from a fresh graduate student that cannot generate data until now a successful PhD that can not

only generate data but also analyze data well.

I would also like to thank our cooperator: Professor Donald Becker from University of Nebraska. Dr. Becker is such an intellectual and modest biochemist, from who I learned lots of functional studies on protein. I benefit a lot from the discussions in the weekly Skype meetings, and his meaningful suggestion contributes a lot in successfully finishing my PhD projects.

At the same time, I sincerely appreciate the following professors: Dr. Kent Gates, Dr. Lesa Beamer, Dr. Jason Cooley and Dr. Michael Henzl. I learned organic chemistry from Dr. Gates, a great chemist who can explain a complicated reaction mechanism with easy language; Dr. Beamer gave me lots of professional opinions on protein chemistry and crystallography, and always selflessly let me use her lab facilities; Two of my projects were finished directly related to Dr. Cooley, he gave me hints on crystallizing a mono-functional proline dehydrogenase, and generously provided me the *Rhodobacter capsulus* genomic DNA, without his help, I cannot finish my PhD on time; Dr. Henzl is such a good scientist who has great love in research, and I would never forget his joke after I submit several samples to him for ITC " I really regret that I get involved into this project ... " All of the above professors impacted and shaped me becoming a better scientist.

As a group I would like to extend my acknowledgement to the former and current members of the Tanner lab, who together created a good research environment for me. The people I would like to name first here are John Larson, Vivek Furtado and Dale Karr,

they taught me most of the basic techniques in Tanner lab. Specific thanks to my best friends Ranjan Singh and Travis Pemberton, it is my fortune to have them in the same lab that I can work and learn with. I also appreciate Dhiraj Srivastava, Harkewal Singh and Richa Dhatwalia for giving several good introductive talks in our group meeting.

No words of gratitude can address the great deal of thanks I owe to my parents and wife. As a son, I can only say “Happy birthday” and “Happy new year” to my parents on the phone; when they get old and sometimes get sick, what I can do is just pray for them on the other side of the world. As a husband, I did not give enough time and love to my wife, she has to put up with my long working hours in the lab. Without their sacrifice and support behind me, I would never succeed in completing my PhD degree.

Finally, thank University of Missouri and the Department of Chemistry for giving me the chance to fulfill my PhD here.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Oxidation of amino acids, like proline catabolism, is a central part of energy metabolism. Proline is oxidized to glutamate by two enzymes: proline dehydrogenase (PRODH) and Δ^1-pyrroline-5-carboxylate dehydrogenase (P5CDH). PRODH catalyzes the first reaction of proline to 1-pyrroline-5-carboxylate (P5C). P5C undergoes a non-enzymatic hydrolysis to glutamate semialdehyde (GSA), which is oxidized to glutamate by a NAD^+-dependent enzyme P5CDH. PRODH and P5CDH are mono-functional enzymes in eukaryotes and Gram-positive bacteria; while in Gram-negative bacteria, the two enzymes are fused into one protein as two domains, known as proline utilization A (PutA). This dissertation work involved structural and functional studies of PRODH, P5CDH, PutA, and human aldehyde dehydrogenases (ALDHs). The results illuminated the substrate recognition for mono-functional PRODH and hot spot oligomerization mechanism for mono-functional P5CDH, also, demonstrated that diethylamino-benzaldehyde (DEAB) is a mechanism based inactivator for aldehyde dehydrogenase 7A1. Furthermore, the C-terminal domain found in PutAs, the only domain without any structural and functional information has been structurally and biochemically characterized. Put all together, this dissertation work successfully resolved several main questions in the world of proline catabolic enzymes, and largely broadened the understanding for proline catabolism.

CHAPTER 1

1. INTRODUCTION

1.1 PROLINE CATABOLISM

L-proline is the only natural occurred amino acid that occupies a bulky backbone, which gives its unique role to make sharp turns in protein structure. Besides the role as a unique building block in protein biosynthesis, its catabolism is an important part in amino acid metabolism, which has been gradually recognized playing critical role in bioenergetics, cellular redox homeostasis, cancer and apoptosis. Because of its important biological relevance, the study of enzymes involved in proline catabolism is important.

The proline catabolism pathway involves the oxidation of L-proline to glutamate, which utilizes two enzymes: proline dehydrogenase (PRODH) and Δ^1-pyrroline -5-carboxylate dehydrogenase (P5CDH), and this pathway is widely conserved in both bacteria and eukaryotes (1). Figure 1.1 shows the reaction steps of proline catabolism. First the substrate proline is oxidized by PRODH into pyrolline-5-carboxylate (P5C), in which, two electrons was transferred as a hydride to the co-factor Flavin adenine dinucleotide (FAD) of PRODH. Then P5C undergoes a non-enzymatic hydrolysis to form glutamate semialdehyde (GSA). GSA is then oxidized by P5CDH to the final product glutamate, in which the co-factor nicotinamide adenine dinucleotide (NAD^+) accepts a hydride to form NADH.

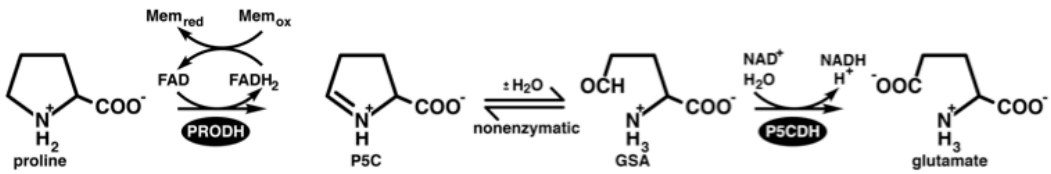


Figure 1-1 The proline catabolic pathway. This figure was taken from Luo et al. Biochemistry 2012.

1.2 BIOLOGICAL RELEVANCE OF PROLINE CATABOLIC ENZYMES

In eukaryotes, PRODH is localized to the inner mitochondrial membrane and P5CDH exists in mitochondrial matrix. Proline catabolism plays important role in establishing the mitochondrial redox homeostasis (2, 3).

PRODH is important in apoptosis, tumor cell growth, and schizophrenia. Previous study of Phang's group has characterized that human PRODH is a tumor suppressor protein (3−9). Tumor suppressor p53 has been shown capable of inducing the expression of PRODH, and PRODH has been observed to activate intrinsic and extrinsic apoptotic pathways (6). The function that PRODH reacts as a tumor suppressor relies on its ability to generate superoxide (6, 10). Also, inborn errors in the gene encoding PRODH could cause type I hyperprolinemia, which is a risk factor for schizophrenia (11, 12).

Defects in P5CDH are pathogenic, which cause the type II hyperprolinemia (15-19). Certain missense and frame shift mutations in the gene of P5CDH could impair enzyme function, which causes elevated levels of proline and P5C in plasma, and cerebrospinal fluid (20). Type II hyperprolinemia could induce neurologic manifestations, such as increased incidence of seizures and intellectual disabilities

(10), but why the deficiency of P5C dehydrogenase activity would cause these diseases is still unknown – the high level concentration of proline may be a possible reason, considering it has been observed as a neurotransmitter, and oxidative stress (22-26). P5CDHs has been shown important for the pathogenicity of some microorganisms. Previous work on fungal pathogen *Cryptococcus neoformans* revealed that P5CDH is essential for optimal production of the virulence factors, and a dysfunctional mutation in P5CDH disabled the infection for mouse models by *C. neoformans* (27).

1.3 DIVERSITY OF PROLINE CATABOLIC ENZYMES

PRODH and P5CDH have large diversity in different organisms. In eukaryotes and Gram-positive bacteria, PRODH and P5CDH are distinct enzymes (1); while in Gram-negative bacteria and *Corynebacterium*, these two enzymes are fused in a single polypeptide chain, which is known as proline utilization A (PutA) (28).

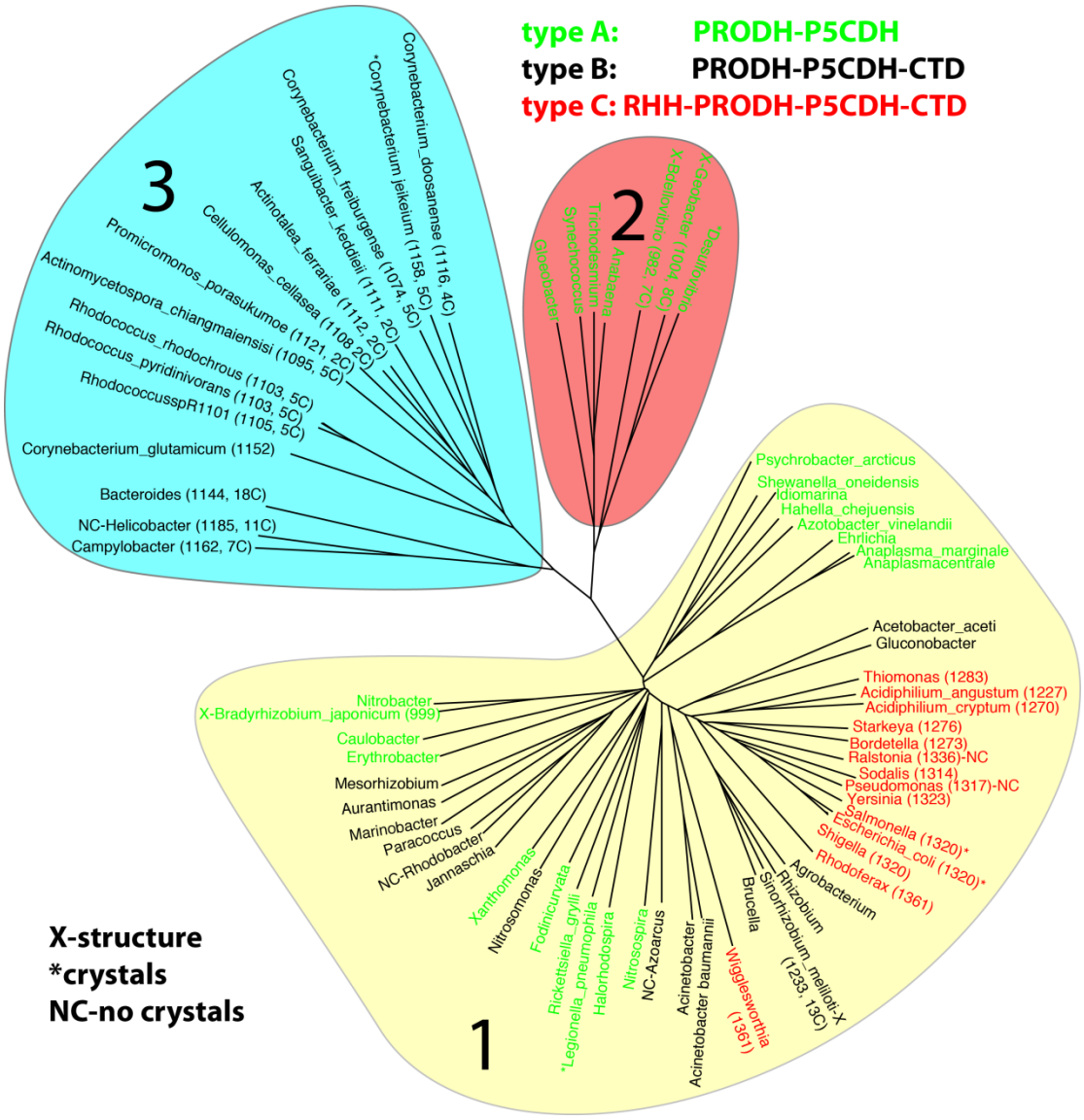


Figure 1-2 PutA classification scheme.

There are three main PutA branches based on multiple sequence alignments, the sequence identity between them are normally less than 30%. Based on their domain architecture difference, PutAs can be divided into three types of A, B, and C (Fig. 1.2) (29). Type A PutAs only contain the minimal catalytic core, which includes N-terminal PRODH and C-terminal P5CDH domains. *Bradyrhizobium japonicum* PutA (BjPutA) from Branch 1 and *Geobacter sulfurreducens* PutA (GsPutA) from Branch 2

belong to type A PutAs (30, 31). Types B PutAs, compared to type A enzymes, have an additional C-terminal domain (CTD), which normally contains 100-200 residues, *Rhodobacter capsulatus* PutA (RcPutA) from Branch 1 and *Helicobacter Pylori* PutA (HpPutA) from Branch 3 are two examples of type B PutAs (28). Type C PutAs, besides the CTD, have an extra N-terminal DNA binding domain, which give it another function as a transcription relator (32). DNA binding domain only exist in Branch 1 PutAs, PutA from *Escherichia coli* (EcPutA) is an example of type C PutA (32).

1.4 SUMMARY OF THESIS RESEARCH

An interesting aspect for PRODH and P5CDH is their diversity in different organism. In my PhD, I have worked on both mono-functional PRODH and P5CDH, also figured out several main questions in PutA.

For the mono-functional PRODH study, by using X-ray diffraction, I have determined the structures of proline dehydrogenase from *Deinococcus radiodurans* in the oxidized state complexed with the proline analogue l-tetrahydrofuroic acid and in the reduced state with the proline site vacant. The PRODH complex structure is the first structure of a mono-functional PRODH complexed with a proline analogue and the highest-resolution structure of any PRODH or PutA. These results provide new insight into substrate recognition by mono-functional PRODHs and the conformational changes that accompany flavin reduction and product release.

In mono-functional P5CDH, using a combination of small-angle X-ray scattering, dynamic light scattering, and X-ray crystallography, my colleague Ranjon Singh and I

have determined the oligomeric states and quaternary structures of four bacterial P5CDHs from *Thermus thermophilus* (TtP5CDH*), Deinococcus radiodurans* (DrP5CDH), *Bacillus halodurans* (BhP5CDH), and *Bacillus licheniformis* (BlP5CDH). The hot spot oligomerization mechanism of ALDH4 family has been determined, and which can be used to predict the oligomeric state of the other P5CDHs.

The C-terminal domain of the type B and type C PutA is the only domain that is without structure and function information. Combining small-angle X-ray scattering (SAXS), kinetics, and mutagenesis, the location and function of a 100-200 residue C-terminal domain (CTDUF) in RcPutA was determined. SAXS-guided rigid body model has been built on CTDUF, the kinetic and mutagenesis studies suggest that CTDUF contributes to P5CDH activity and substrate channeling. More specifically, I resolved the full-length structure of type B PutA from *Sinorhizobium meliloti* that contains the CTDUF, which further confirmed CTDUF plays a major role in facilitating substrate channeling by helping to prevent diffusion of the intermediate into the bulk medium.

Finally, besides the study on proline catabolic enzymes, inactivation of human aldehyde dehydrogenase 7A1 (ALDH7A1) by 4-diethylaminobenzaldehyde (DEAB) was studied. DEAB is widely used as a competitive inhibitor for Human aldehyde dehydrogenase 1, which was known as a novel cancer stem cell marker. In a commercially available ALDHFLUOR assay kit, DEAB is used as a specific inhibitor of ALDH to control for background fluorescence. Although DEAB has been commonly used for decades in stem cell biology, its specificity and inhibition mechanism has never been clearly determined. With this purpose, I performed DEAB

inhibition studies on human ALDH7A1. Several high-resolution ALDH7A1-DEAB complex structures were determined, a clear covalent bond was found between DEAB and catalytic Cys302. Further kinetic and mass spectrometry studies confirmed DEAB as a mechanism based inactivator instead of a competitive inhibitor as previously thought.

1.5 REFERENCES

1. Tanner, J. J. (2008) Structural biology of proline catabolism, Amino Acids 35, 719-730.

2. Phang, J. M. (1985) The regulatory functions of proline and pyrroline-5-carboxylic acid, Curr. Top. Cell. Reg. 25, 92-132.

3. Phang, J. M., Donald, S. P., Pandhare, J., and Liu, Y. (2008) The metabolism of proline, a stress substrate, modulates carcinogenic pathways, Amino Acids 35, 681-690.

4. Donald, S. P., Sun, X. Y., Hu, C. A., Yu, J., Mei, J. M., Valle, D., and Phang, J. M. (2001) Proline oxidase, encoded by p53-induced gene-6, catalyzes the generation of proline-dependent reactive oxygen species, Cancer Res. 61, 1810-1815.

5. Hu, C. A., Donald, S. P., Yu, J., Lin, W. W., Liu, Z., Steel, G., Obie, C., Valle, D., and Phang, J. M. (2007) Overexpression of proline oxidase induces proline-dependent and mitochondria-mediated apoptosis, Mol. Cell. Biochem. 295, 85-92.

6. Liu, Y., Borchert, G. L., Surazynski, A., Hu, C. A., and Phang, J. M. (2006) Proline oxidase activates both intrinsic and extrinsic pathways for apoptosis: the role of ROS/superoxides, NFAT and MEK/ERK signaling, Oncogene 25, 5640-5647.

7. Pandhare, J., Cooper, S. K., and Phang, J. M. (2006) Proline oxidase, a proapoptotic gene, is induced by troglitazone: evidence for both peroxisome proliferator-activated receptor gamma-dependent and -independent mechanisms, J Biol Chem 281, 2044-2052.

10. Phang, J. M., Pandhare, J., Zabirnyk, O., and Liu, Y. (2008) PPARgamma and Proline Oxidase in Cancer, PPAR Res 2008, 542694.

11. Liu, Y., Borchert, G. L., Donald, S. P., Diwan, B. A., Anver, M., and Phang, J. M. (2009) Proline oxidase functions as a mitochondrial tumor suppressor in human cancers, Cancer Res. 69, 6414-6422.

12. Liu, Y., Borchert, G. L., Donald, S. P., Surazynski, A., Hu, C. A., Weydert, C. J., Oberley, L. W., and Phang, J. M. (2005) MnSOD inhibits proline oxidase-induced apoptosis in colorectal cancer cells, Carcinogenesis 26, 1335-1342.

13. Phang, J. M., Hu, C. A., and Valle, D. (2001) Disorders of proline and

hydroxyproline metabolism, In Metabolic and molecular basis of inherited disease (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., Eds.), pp 1821-1838, McGraw Hill, New York.

14. Willis, A., Bender, H. U., Steel, G., and Valle, D. (2008) PRODH variants and risk for schizophrenia, Amino Acids 35, 673-679.

15. Efron, M. L. (1965). 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. 272, 1243-54.

16. Baumgartner, M. R., Rabier, D., Nassogne, M. C., Dufier, J. L., Padovani, J. P., Kamoun, P., Valle, D. & Saudubray, J. M. (2005). 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. 164, 31-6.

17. Geraghty, M. T., Vaughn, D., Nicholson, A. J., Lin, W. W., Jimenez-Sanchez, G., Obie, C., Flynn, M. P., Valle, D. & Hu, C. A. (1998). Mutations in the Delta1-pyrroline 5-carboxylate dehydrogenase gene cause type II hyperprolinemia. Hum. Mol. Genet. 7, 1411-5.

18. Scriver, C. R., Sly, W. S., Childs, B., Beaudet, A. L., Valle, D., Kinzler, K. W. & Vogelstein, B., Eds. (2001). The Metabolic and Molecular Bases of Inherited Disease. 8th edit. New York: McGraw-Hill.

19. Valle, D., Goodman, S. I., Applegarth, D. A., Shih, V. E. & Phang, J. M. (1976). Type II hyperprolinemia. Delta1-pyrroline-5-carboxylic acid dehydrogenase deficiency in cultured skin fibroblasts and circulating lymphocytes. J. Clin. Invest. 58, 598-603.

20. Wyse, A. T. & Netto, C. A. (2011). Behavioral and neurochemical effects of proline. Metab. Brain Dis. 26, 159-72.

21. Phang, J. M., Hu, C. A. & Valle, D. (2001). Disorders of proline and hydroxyproline metabolism. In Metabolic and molecular basis of inherited disease (Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D., eds.), pp. 1821-1838. McGraw Hill, New York.

22. Fremeau, R. T., Jr., Caron, M. G. & Blakely, R. D. (1992). Molecular cloning and expression of a high affinity L-proline transporter expressed in putative glutamatergic pathways of rat brain. Neuron 8, 915-26.

23. Felix, D. & Kunzle, H. (1976). The role of proline in nervous transmission. Adv. Biochem. Psychopharmacol. 15, 165-73.

24. Takemoto, Y. & Semba, R. (2006). Immunohistochemical evidence for the localization of neurons containing the putative transmitter L-proline in rat brain. Brain Res. 1073-1074, 311-5.

25. Gogos, J. A., Santha, M., Takacs, Z., Beck, K. D., Luine, V., Lucas, L. R., Nadler, J. V. & Karayiorgou, M. (1999). The gene encoding proline dehydrogenase modulates sensorimotor gating in mice. Nat. Genet. 21, 434-9.

26. He, F. & DiMario, P. J. (2011). 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 11, 397-404.

27. Lee, I. R., Lui, E. Y., Chow, E. W., Arras, S. D., Morrow, C. A. & Fraser, J. A. (2013). Reactive Oxygen Species Homeostasis and Virulence of the Fungal Pathogen Cryptococcus neoformans Requires an Intact Proline Catabolism Pathway. Genetics, doi: 10.1534/genetics.113.150326.

28. Tanner, J. J., and Becker, D. F. (2013) PutA and proline metabolism. In Handbook of Flavoproteins (Hille, R., Miller, S. M., and Palfey, B., Eds.) pp 31−56, Walter de Gruyter, Boston.

29. Luo M, Christgen, S., Sanyal, N, Arentson BW, Becker DF, Tanner JJ. (2014) Evidence that the C-terminal Domain of a Type B PutA Protein Contributes to Aldehyde Dehydrogenase Activity and Substrate Channeling. Biochemistry. 53(35):5661-73

30. Srivastava, D., Schuermann, J. P., White, T. A., Krishnan, N., Sanyal, N., Hura, G. L. Tan, A., Henzl, M. T., Becker, D. F., and Tanner, J. J. (2010) Crystal structure of the bifunctional proline utilization A flavoenzyme from Bradyrhizobium japonicum, Proceedings of the National Academy of Sciences of the United States of America 107, 2878-2883.

31. Singh, H., Arentson, B. W., Becker, D. F., and Tanner, J. J. (2014) Structures of the PutA peripheral membrane flavoenzyme reveal a dynamic substrate-channeling tunnel and the quinone-binding site, Proceedings of the National Academy of Sciences of the United States of America 111, 3389-3394.

32. Ranjan K. Singh, John D. Larson, Weidong Zhu, Robert P. Rambo, Greg L. Hura, Donald F. Becker, and John J. Tanner. (2011) Small-Angle X-ray Scattering Studies of the Oligomeric State and Quaternary Structure of the Trifunctional Proline Utilization A (PutA) Flavoprotein from Escherichia coli. J. Biol. Chem. 286 (50) 43144-53.

CHAPTER 2

2. CRYSTAL STRUCTURES AND KINETICS OF MONOFUNCTIONAL PROLINE DEHYDROGENASE PROVIDE INSIGHT INTO SUBSTRATE RECOGNITION AND CONFORMATIONAL CHANGES ASSOCIATED WITH FLAVIN REDUCTION AND PRODUCT RELEASE

Luo M, et al. Crystal Structures and Kinetics of Monofunctional Proline Dehydrogenase Provide Insight into Substrate Recognition and Conformational Changes Associated With Flavin Reduction and Product Release. Biochemistry (2012) 51(50):10099-108).

ABBREVIATIONS

PRODH, proline dehydrogenase; P5C, Δ^1-pyrroline-5-carboxylate; GSA, glutamate semialdehyde; P5CDH, 1-pyrroline-5-carboxylate dehydrogenase; POX, human proline oxidase; TtPRODH, proline dehydrogenase from *Thermus thermophilus*; PutA, proline utilization A; DrPRODH, proline dehydrogenase from *Deinococcus radiodurans* R1; $DrPRODH_o$, oxidized proline dehydrogenase from *Deinococcus radiodurans* R1; $DrPRODH_r$, reduced proline dehydrogenase from *Deinococcus radiodurans* R1; THFA, L-tetrahydrofuroic acid; CoQ1, Coenzyme Q1; TEVP, Tobacco Etch Virus protease; PutA86-669, PRODH construct of *E. coli* PutA containing residues 86-669.

2.1 ABSTRACT

Proline dehydrogenase catalyzes the FAD-dependent oxidation of proline to D^1-pyrroline-5-carboxylate, which is the first step of proline catabolism. Here, we report the structures of proline dehydrogenase from *Deinococcus radiodurans* in the oxidized state complexed with the proline analog L-tetrahydrofuroic acid and in the reduced state with the proline site vacant. The analog binds against the *si* face of the FAD isoalloxazine and is protected from bulk solvent by the a8 helix and the b1-a1 loop. The FAD ribityl chain adopts two conformations in the E-S complex, which is unprecedented for flavoenzymes. One of the conformations is novel for the PRODH superfamily and may contribute to the low substrate affinity of *Deinococcus* PRODH. Reduction of the crystalline enzyme-inhibitor complex causes profound structural changes, including 20° butterfly bending of the isoalloxazine, crankshaft rotation of the ribityl, shifting of a8 by 1.7 Å, reconfiguration of the b1-a1 loop, and rupture of the Arg291-Glu64 ion pair. These changes dramatically open the active site to facilitate product release and allow electron acceptors access to the reduced flavin. The structures suggest that the ion pair, which is conserved in the PRODH superfamily, functions as the active site gate. Mutagenesis of Glu64 to Ala decreases catalytic efficiency 27-fold, which demonstrates the importance of the gate. Mutation of Gly63 decreases efficiency 140-fold, which suggests that flexibility of the b1-a1 loop is essential for optimal catalysis. The large conformational changes that are required to form the E-S complex suggest that conformational selection plays a role in substrate recognition.

2.2 INTRODUCTION

Proline dehydrogenase (PRODH) catalyzes the first reaction of proline

catabolism (Figure 2.1A).[1] PRODH is an FAD-dependent enzyme that catalyzes the oxidation of L-proline to Δ^1-pyrroline-5-carboxylate (P5C). The electrons stored in the reduced flavin are subsequently transferred to the electron transport chain for eventual ATP production. P5C forms an equilibrium with its hydrolysis product glutamate γ-semialdehyde (GSA), which is oxidized to glutamate by the second enzyme of proline catabolism, NAD^+-dependent P5C dehydrogenase (P5CDH). PRODH and P5CDH are distinct enzymes in eukaryotes and Gram-positive bacteria, whereas the two enzymatic activities are combined into a single polypeptide chain, known as proline utilization A (PutA), in Gram-negative bacteria.[2, 3]

PRODH is of interest because of its roles in apoptosis, cancer, and schizophrenia. In eukaryotes, PRODH and P5CDH are localized to the inner mitochondrial membrane and mitochondrial matrix respectively, and proline catabolism is important for establishing the mitochondrial redox status.[4, 5] Seminal work from Phang's group has established that human PRODH (aka proline oxidase or POX) is a tumor suppressor protein.[5-13] POX expression is induced by the tumor suppressor p53, and POX itself activates intrinsic and extrinsic apoptotic pathways.[8] Crucial to the role of POX as a tumor suppressor is its ability to generate superoxide.[5, 8, 14] Also, certain mutations in the gene encoding POX cause type I hyperprolinemia,[15] which is a risk factor for schizophrenia.[16]

Crystal structures of PRODHs suggest that conformational changes are essential for catalysis, but the details of these conformational changes have not been elucidated. We previously determined structures of the PRODH domain of *Escherichia coli* PutA (PutA86-669) complexed with proline and proline analogs,[17, 18, 19] as well as a structure of *Bradyrhizobium japonicum* PutA with a sulfate ion bound in the proline

site.[20] In all of those structures, proline (or the analog) is completely buried, implying that the active site opens to allow product release and closes again in response to substrate binding. Since the structure of the empty PutA PRODH active site is unknown, it has not been possible to deduce these conformational changes for PutA. We also reported a structure of the monofunctional PRODH from *Thermus thermophilus* (TtPRODH) in a substrate-free conformation.[21] However, the corresponding structure of the closed, substrate-bound active site is unknown. Thus, our understanding of how conformational changes facilitate catalysis has been limited by difficulties in crystallizing the same PRODH in both the open (substrate-free) and closed (substrate-bound) conformations. Although some information has been gained by comparing the structures of PutA669-proline and ligand-free TtPRODH, this comparison is problematic because it is unknown whether PutAs and monofunctional PRODHs bind proline similarly. Another factor that complicates such a comparison is that the PRODH domain of PutA makes tertiary structural contacts that are absent in monofunctional PRODHs, and thus is unlikely that PutA PRODH domains and monofunctional PRODHs exhibit similar degrees of flexibility.

We therefore surveyed several monofunctional PRODHs in search of one that crystallizes in the presence of the proline analog L-tetrahydrofuroic acid (THFA). This survey uncovered *Deinococcus radiodurans* R1 PRODH (DrPRODH) as a suitable enzyme for determining the structural changes that occur in monofunctional PRODHs during the reductive half-reaction (Figure 2.1B). Herein we report high resolution crystal structures and kinetic studies of DrPRODH, which provide new insight into substrate recognition by monofunctional PRODHs and the conformational changes that accompany flavin reduction and product release.

2.3 EXPERIMENTAL PROCEDURES

2.3.1 Protein Expression and Purification.

The gene encoding DrPRODH (NCBI RefSeq code NP_294538.1, 310 residues) in the vector pMH1F was obtained from the Joint Center for Structural Genomics. The gene was amplified by PCR and ligated into plasmid pKA8H using *NdeI* and *BamHI* sites. The expressed protein includes an N-terminal His_8 tag and Tobacco Etch Virus protease (TEVP) site. The G63A mutant of DrPRODH was created with the QuickChange II site-directed mutagenesis kit (Agilent) using a forward primer of 5'-GAACCTTGACCTGCTCGCCGAATTTATCGACAGCCCG-3', and a reverse primer of 5'-CGGGCTGTCGATAAATTCCGCGAGCAGGTCAAGGTTC-3'. The E64A mutant of DrPRODH was prepared similarly using a forward primer of 5'-CTTGACCTGCTCGGCGCCTTTATCGACAGCCCGGCC-3', and a reverse primer of 5'- GGCCGGGCTGTCGATAAAGGCGCCGAGCAGGTCAAG-3'. The mutations were confirmed using DNA sequencing.

DrPRODH and mutants G63A and E64A were expressed in BL21(DE3)pLysS cells as follows. Starter cultures of 10 mL were grown in LB media overnight and used to inoculate 3 L of LB broth. After the culture reached an optical density of OD_{600} = 0.8, 0.5 mM IPTG was added to induce protein expression for 5 hours at 22 °C. The cells were collected by centrifugation, resuspended in 50 mM Tris, 100 mM NaCl, 10 mM imidazole, and 5% glycerol at pH 7.5, and frozen at -80 °C.

DrPRODH and mutants G63A and E64A were purified as follows. The frozen cells were thawed at 4 °C in the presence of protease inhibitors (0.1 mM TPCK, 0.05

mM AEBSF, 0.1 μM Pepstatin, 0.01 mM Leupeptin, 5μM E-64) and broken using sonication. The mixture was centrifuged at 16500 rpm in an SS34 rotor for 1 hour at 4 °C, filtered through a 0.45 μm filter (Millipore) and loaded to HisTrap HP column (5 mL) that had been equilibrated in 50 mM Tris, 300 mM NaCl, 10 mM imidazole, and 5% glycerol at pH 7.5. Washing steps with the loading buffer supplemented with 10 mM imidazole followed by 30 mM imidazole were performed. The protein was eluted with 300 mM imidazole. The histidine tag was removed by incubating the protein with 0.2 mg/mL TEVP for 1 hour at 28 °C followed by dialysis at 4 °C against 50 mM Tris, 50 mM NaCl, and 5% glycerol at pH 7.5. The mixture was applied to the HisTrap HP column to separate the cleaved protein, which appeared in the flow-through, from the tag and TEVP. The cleaved protein was dialyzed overnight in the dark at 4 °C into 50 mM Tris, 0.5 mM EDTA, 0.5 mM DTT, and 5% glycerol at pH 7.8 in preparation for anion exchange chromatography (HiTrap Q). The sample was loaded onto the column using a buffer of 50 mM Tris and 5% glycerol at pH 7.8, and a linear NaCl gradient was applied. DrPRODH eluted in the range 360 - 420 mM NaCl. The protein concentration was estimated using the bicinchoninic acid method (Pierce kit) with bovine serum albumin as the standard. The flavin contents of the purified mutants G63A and E64A were similar to that of DrPRODH.

2.3.2 Crystallization and Crystal Soaking Experiments.

All crystallization experiments were performed at 22 °C using the sitting-drop method of vapor diffusion with drops formed by mixing equal volumes of the reservoir and the protein stock solutions. The latter solution consisted of 2.8 mg/mL DrPRODH in 50 mM Tris, 50 mM NaCl, 0.5 mM THP, 0.5 mM EDTA, 5% glycerol, and 400 mM THFA at pH 7.5. Initial conditions were identified using commercially

available crystal screens (Hampton Research). Optimized crystals were grown using a reservoir solution of 0.2 M $MgCl_2$, 25% (w/v) PEG 3350, and 0.1 mM Bis-tris at pH 5.8. The crystals were cryoprotected with the reservoir solution supplemented with 25% (v/v) PEG 200, picked up with Hampton loops, and flash-cooled in liquid nitrogen. The space group is $P2_12_12_1$ with unit cell dimensions of a = 44 Å, b = 95 Å, and c = 136 Å. The asymmetric unit includes two enzyme-THFA complexes and 40% solvent.

Crystals of DrPRODH with the FAD reduced ($DrPRODH_r$) were obtained by soaking the aforementioned crystals for ~17 minutes in 50 mM sodium dithionite, 0.2 M $MgCl_2$, 25% PEG 3350, 0.1 mM Bis-Tris pH 5.5, and 25% PEG 200. The crystal color changed from deep yellow to colorless during soaking, which is consistent with reduction of the FAD. The crystals were flash cooled in liquid nitrogen to trap the reduced enzyme conformation.

2.3.3 X-ray Diffraction Data Collection, Processing, and Refinement.

Diffraction data were collected on beamlines 4.2.2 of the Advanced Light Source and 24-ID-C of the Advanced Photon Source. The reflections were integrated with XDS[22] and scaled with SCALA[23] (Table 2.1). The phase problem was solved using molecular replacement as implemented in MOLREP.[24] The search model was the $(\beta\alpha)_8$ barrel of *Thermus thermophilus* PRODH (residues 37 - 279 of chain A of PDB code 2G37, 47 % identical to DrPRODH). The correct solution had a correlation coefficient of 0.36 with two molecules in the $P2_12_12_1$ asymmetric unit. For reference, molecular replacement calculations that assumed other primitive orthorhombic space groups yielded correlation coefficients of 0.26 – 0.31. The model from molecular

replacement was used as the starting point for several rounds of model building with COOT[25] and refinement with PHENIX.[26] Table 2.1 lists refinement statistics.

The oxidized flavin exhibits two conformations in the THFA complex. The two conformations have occupancies of 0.56 (conformation A) and 0.44 (conformation B) and differ in the orientations of the 2′ and 3′ hydroxyl groups of the ribityl chain. Justification for building two conformations was obtained by refining the structure with single conformations of the FAD. For example, the F_o-F_c map calculated after a refinement that included only conformation A at occupancy of 1.0 showed strong features for the 2'-OH and 3'-OH ribityl groups of conformation. Conversely, the map calculated after a refinement that included just conformation B at occupancy of 1.0 showed strong features for the 2'-OH and 3'-OH ribityl groups of conformation A.

2.3.4 Enzyme Activity Assays.

All chemicals used during kinetic characterization were purchased from Fischer Scientific or Sigma-Aldrich. All steady-state assays were conducted at 23 °C in 50 mM potassium phosphate, 25 mM NaCl at pH 7.5. For all assays, Coenzyme Q1 (CoQ1) was used as the electron acceptor with CoQ1 reduction monitored by the decrease in absorbance at 278 nm using a molar extinction coefficient of 14.5 mM^{-1} cm^{-1}.[27] The K_m for proline and k_{cat} for wild-type DrPRODH (0.25 μM) were determined by varying the concentration of proline (0 - 500 mM) while keeping CoQ1 constant (200 μM). Inhibition of wild-type DrPRODH (0.25 μM) by THFA was analyzed by varying proline (10 - 500 mM) and THFA (0 - 200 mM), while keeping CoQ1 constant (200 μM). These assays were performed in a total volume of 200 μl per assay using a Powerwave XS microplate spectrophotometer (Bio-Tek). The K_m

for CoQ1 and k_{cat} for wild-type DrPRODH (0.25 μM) were determined by varying CoQ1 (0 - 450 μM) and keeping proline constant (500 mM). For the DrPRODH mutants G63A (11 μM) and E64A (6.3 μM), the K_m for proline and k_{cat} were determined by varying proline (0 - 1000 mM) and holding CoQ1 constant (200 μM). The K_m for CoQ1 and k_{cat} for the DrPRODH mutants were determined by varying CoQ1 (10 - 450 μM) and holding proline constant (500 mM). These assays were performed in a total volume of 150 μl by mixing enzyme and substrate solutions using a Hi-Tech Scientific SF-61DX2 stopped flow instrument equipped with a 0.15 cm path length cell. Steady-state parameters were calculated by fitting initial rate data to the Michaelis-Menten equation and inhibition data were globally fit to a competitive inhibition model using Enzyme Kinetic Wizard (SigmaPlot 12.0).

2.4 RESULTS

2.4.1 Structure of DrPRODH$_o$-THFA.

The crystal structure of oxidized DrPRODH (DrPRODH$_o$) complexed with the proline analog THFA was determined at 1.36 Å resolution (Table 2.1). It is the first structure of a monofunctional PRODH complexed with a proline analog and the highest resolution structure of any PRODH or PutA.

The DrPRODH$_o$-THFA complex exhibits a distorted $(\beta\alpha)_8$ barrel fold in which the last helix, denoted as α8 (residues 285-295), is located above the C-terminal ends of the strands rather than alongside β8 as in the classic triosephosphate isomerase barrel (Figure 2.2A). This fold is also observed in the structures of TtPRODH (Figure 2.2B) and PutA and is considered to be a defining characteristic of the PRODH

superfamily, which includes both monofunctional PRODHs and PutAs. The root mean square deviation between $DrPRODH_o$-THFA and TtPRODH is 1.2 Å; the variation from PutA PRODH domains is slightly higher at 1.5 - 1.9 Å.

THFA binds between the *si* face of the FAD and α8 (Figure 2.2A). Arg291, Arg292, and Lys98 form ion pairs with the carboxylate of THFA, while Leu257, Tyr278, Tyr288 provide hydrophobic contacts with the pyrrolidine ring (Figure 2.3A). Arg291 and Arg292 are stabilized by ion pairs with Glu64 and Glu295, respectively. The plethora of residues clustered around THFA make the inhibitor and the *si* face of the flavin inaccessible to solvent.

The oxidized flavin exhibits two conformations, which is unprecedented for PRODHs and PutAs (Figure 2.4A). Furthermore, we are unaware of another flavin oxidoreductase that exhibits multiple flavin conformations in the E-S complex. The two conformations have occupancies of 0.56 (conformation A) and 0.44 (conformation B). Interestingly, the two FAD conformations do not correspond to THFA-bound and THFA-free states because the occupancy of THFA is clearly 1.0 (*B*-factor = 17.5 $Å^2$). Rather, the enzyme is able to bind THFA using either of two different FAD conformations.

The two FAD conformations differ in the orientations of the 2′ and 3′ hydroxyl groups of the ribityl chain (Figures 2.4A). In conformation A (yellow in Figure 2.4A), the 2′-OH points toward the proline binding site, while the 3′-OH is oriented 180° away and interacts with the carbonyl of Gly191. Conformation B is related to conformation A by a crankshaft rotation around the C2′-C3′ bond. In conformation B (pink in Figure 2.4A), the 2′-OH is tucked under the pyrimidine ring where it forms

hydrogen bonds with the flavin N1 atom and N-H group of Gly191, while the 3′-OH interacts with Glu295 of α8.

Ribityl conformation A appears to be unique among the PRODH superfamily. Although it resembles the ribityl of PutA86-669-THFA in that 2′-OH points into the proline site, the other two hydroxyl groups of conformation A are rotated 180° from the corresponding hydroxyls of PutA86-669-THFA (Figure 2.5A). Although the 4′-OH of DrPRODH overlaps the 3′-OH of PutA86-669-THFA (Figure 2.5A), the two hydroxyls are nonequivalent because they form different interactions. The 3′-OH of PutA86-669-THFA forms a hydrogen bond with conserved Glu559 of α8 (Figure 2.3B), which corresponds to Glu295 of DrPRODH. Glu295, however, adopts a different conformation and is thus unable to form an analogous hydrogen bond with the 4′-OH (Figure 2.3B).

Ribityl B is more familiar. It is similar to the ribityl of ligand-free TtPRODH (Figure 2.5B). In both structures, the 2′- and 3′-hydroxyls are below the pyrimidine ring, while the 4′-OH is below the dimethylbenzene ring. One notable difference is that the 4′-OH of DrPRODH ribityl B is located on the *si* face of the FAD, whereas that of TtPRODH is torsioned over to the *re* side. The *re* side location allows the hydroxyl to engage a structurally conserved water molecule that bridges the main chains of β5 and β6 in all PRODH and PutA structures (Figure 2.5B).

The flavins of DrPRODH and PutA also differ in the position of the adenosine group (Figure 2.5A). Because of a rotation around the pyrophosphate, the adenine ring positions of the two structures differ by 13 Å. This difference was also noted in TtPRODH and is due to the presence of an extra α-helix in PutAs that is absent in

monofunctional PRODHs.[21]

2.4.2 Structure of DrPRODH$_r$.

A 1.75 Å resolution structure of DrPRODH$_r$ was determined from a crystal of the DrPRODH$_o$-THFA complex that was soaked in sodium dithionite. Reduction causes large conformational changes in the flavin, which are observed in both molecules in the asymmetric unit. Large protein conformational changes are also observed in chain A of the asymmetric unit, and the electron density map is consistent with release of THFA. In the other chain, crystal contacts prevent large protein conformational changes, and the electron density map suggests that an acetate ion replaces THFA. We note that PutA86-669 crystallized in PEG 3350 also has an acetate ion bound in the active site.[28] The discussion below focuses on chain A, since it provides information about conformational changes in both the flavin and the protein.

Reduction by dithionite induces substantial conformational changes in the flavin. The reduced flavin exhibits a 20° butterfly bend in the isoalloxazine (Figure 2.4C), which contrasts the planar isoalloxazine of the oxidized enzyme. Severe butterfly bending (20°-30°) has also been observed in the dithionite-reduced flavin of PutA86-669[28] and the covalently modified, reduced flavins of N-propargylglycine-inactivated PRODH[29] and PutA.[30] The bending of the flavin pushes the N5-N10 axis toward the proline binding site, which results in the *si* face being convex. The reduced flavin of DrPRODH$_r$ exhibits just one ribityl conformation (Figure 2.4B), which is nearly identical to the ribityl chains of the oxidized FAD in TtPRODH and the reduced flavin in the N-propargylglycine-inactivated *E. coli* PutA PRODH domain[30] (Figure 2.5C).

Flavin reduction also dramatically changes the protein conformation. Upon flavin reduction, α8 shifts away from the isoalloxazine by 1.7 Å, and the β1-α1 loop (residues 62 - 69) withdraws from the active site by over 5 Å (Figure 2.6A). One major consequence of these movements is the rupture of the Arg291-Glu64 ion pair, which links α8 and the β1-α1 loop in the THFA complex. The electron density for the ion pair in the THFA complex is exceptional, indicating that this interaction is quite strong (Figure 2.3A). Its rupture is thus a significant event. Breaking of the ion pair causes Glu64 and Arg291 to separate from each other. Glu64 is driven to the protein surface along with the rest of the β1-α1 loop (Figures 2.6A and 2.2B), although we note that electron density for the β1-α1 loop is relatively weak in $DrPRODH_r$, implying high mobility. In fact, density for the side chain of Glu64 in $DrPRODH_r$ is absent, indicating that the side chain is disordered. Electron density for the side chain of Arg291 is likewise weak, but suggests that the guanidinium group turns away from the proline binding site (Figure 2.6A). Flavin reduction also induces disorder in another key side chain of α8, Glu295, which hydrogen bonds with the FAD adenine and ribityl groups in the THFA complex.

The large movements of α8 and the β1-α1 loop dismantle the proline binding site and profoundly change the solvent accessibility of the flavin. In the THFA complex, the isoalloxazine and THFA are buried beneath the Arg291-Glu64 ion pair (Figure 6B), and the interaction between Glu295 and the ribityl seals the bottom of the active site (Figure 2.6B, inset). In the reduced enzyme, the separation of Arg291 from Glu64 creates a large, solvent exposed cavity in the upper part of the active site, which contains the isoalloxazine (Figure 2.6C). Furthermore, the combination of disorder in Glu295 and change in the ribityl conformation creates a hole in the bottom

of the active site of the reduced enzyme (Figure 2.6C, inset). Thus, active site of the reduced, ligand-free enzyme is open and the flavin is highly solvent exposed.

2.4.3 Kinetic Parameters of DrPRODH and DrPRODH β1-α1 Loop Mutants.

The observation of large conformational changes in the β1-α1 loop induced by flavin reduction and release of THFA prompted investigation of this loop using site-directed mutagenesis. As described by White, *et al*., the β1-α1 loop harbors one of 9 conserved sequence motifs that define the PRODH superfamily.[21] This motif includes a Gly-Glu pair that is present in all monofunctional PRODHs and PutAs and corresponds to Gly63-Glu64 of DrPRODH. These residues were therefore investigated by creating the site-directed mutant enzymes G63A and E64A.

The kinetic parameters for wild-type DrPRODH were first determined (Table 2.2). The values of k_{cat} and K_m using proline as the variable substrate with fixed CoQ1 are 8.7 s^{-1} and 290 mM, respectively, resulting in a k_{cat}/K_m of 30 $s^{-1}M^{-1}$. For comparison, the corresponding parameters of the closely related TtPRODH (47 % identical to DrPRODH) are k_{cat} = 13 s^{-1}, K_m = 27 mM.[21] Those of the *E. coli* PutA are k_{cat} = 5.2 s^{-1}, K_m = 42 mM.[27] Thus, the K_m for proline of DrPRODH is higher than expected. The kinetic parameters were also determined using CoQ1 as the variable substrate at fixed proline concentration. These values are k_{cat} = 14 s^{-1}, K_m = 155 μM. The corresponding values for TtPRODH are not available. Those of the *E. coli* PutA are k_{cat} = 3.4 s^{-1}, K_m = 110 μM.[27] Finally, THFA was found to inhibit DrPRODH competitively with proline. The estimated K_I value of 38 mM is over 10 times higher than those of TtPRODH (1 mM[21]) and PutA (1.6 mM[27]). These results suggest that the affinity of DrPRODH for proline is atypically low.

Catalytic activity is severely impaired in the mutants G63A and E64A (Table 2). Using proline as the varying substrate, k_{cat} values of 0.08 s^{-1} and 0.055 s^{-1} were obtained for G63A and E64A, respectively, which are over 100 times lower than that of DrPRODH. The catalytic efficiencies of G63A and E64A for proline are 140-fold and 27-fold lower than that of DrPRODH. Using CoQ1 as the varying substrate, the kcat values are 0.043 s-1 and 0.046 s-1 for G63A and E64A, respectively, which are 300-fold lower than that of DrPRODH. The efficiencies of the mutants for CoQ1 are 60-fold lower relative to DrPRODH. In summary, the kinetic analysis of G63A and E64A confirms the significance of the conserved sequence motif and is consistent with the structures, which suggest that Gly63 is important for the flexibility of the b1-a1 loop, and Glu64 is important for stabilizing the closed active site.

2.5 DISCUSSION

The DrPRODH structures provide new information about the conformations populated by monofunctional PRODHs during the reductive half-reaction (Figure 2.1B). In particular, the $DrPRODH_o$-THFA structure represents the E-S complex (state II, Figure 2.1B). The active site of the E-S complex is closed, and the substrate is buried. The tight packing within the complex is also observed in PutA86-669 (Figure 2.3B) and presumably helps align proline with the flavin for hydride transfer. A water molecule is present in the E-S complex, which bridges conserved Tyr193 and the THFA heteroatom (N atom of proline). An analogous water molecule is also present in the PutA86-669-THFA complex (Figure 2.3B), suggesting that it is conserved in the PRODH superfamily. It is tempting to speculate that this water molecule hydrolyzes P5C to GSA. If so, the hydrolysis reaction occurs within the

active site rather than in solution as implied in Figure 1A and most Biochemistry textbooks and literature articles on proline catabolism.

The E-S complexes of DrPRODH and PutA86-669 are very similar (Figure 2.3B). In both cases, the inhibitor is completely buried between the *si* face of the FAD and α8. The enzyme-inhibitor interactions and ion pair gate conformation are virtually identical in the two structures. Thus, substrate recognition is highly conserved in the PRODH superfamily. The only notable difference involves the conserved glutamate of α8 (Glu295 of DrPRODH and Glu559 of PutA). The different orientations of this side chain seem to be related to the different conformations of the FAD adenosine groups. In DrPRODH, the adenine ring packs against α8 and forms a hydrogen bond with Glu295 (Figure 2.3B). This interaction is not possible in PutA because of the markedly different location of the adenosine (Figure 2.5A), which is due to the presence of an extra α-helix in PutAs that is absent in monofunctional PRODHs.[21]

Given the sequence and structural conservation of the proline binding site, it is surprising that DrPRODH has markedly lower substrate affinity than PutA and TtPRODH. One possibility is that the FAD ribityl plays a role in tuning substrate affinity. DrPRODH is unique in that two conformations of the ribityl are observed in the E-S complex, neither of which is corresponds to that of the high affinity PutA86-669-THFA complex, although conformation A is reminiscent of the flavin in PutA86-669-THFA. It is possible that conformation B represents an inactive state and that a conformational change is required to transform it to conformation A before hydride transfer occurs. If so, Glu295 may play a role in this transformation since it hydrogen bonds with the 3′-OH of conformation B (Figure 2.4A). Additional structures of

PRODH-THFA complexes would shed light on this issue.

Comparison of the $DrPRODH_o$-THFA and $DrPRODH_r$ structures provides insight into conformational changes that occur upon flavin reduction and product release (state II to state IV in Figure 2.1B). These include severe bending of the isoalloxazine, dihedral rotation of the ribityl chain, shifting of α8, reconfiguration of the β1-α1 loop, rupture of the conserved Arg291-Glu64 ion pair, and disordering of Glu295.

The structures suggest a scenario for how the enzyme moves from state II to state IV of the reductive half-reaction. The butterfly bending of the isoalloxazine appears to initiate a cascade of events that leads to product release. Bending of the ring system pushes the N5-N10 axis of the flavin toward the newly formed product, creating steric clash in the highly crowded active site. The shifting of α8 and rupture of the Arg291-Glu64 ion pair alleviate this clash, creating a large cavity that presumably allows facile product dissociation.

The Arg291-Glu64 ion pair is a central player in the conformational changes observed in DrPRODH. Because it is formed in the closed E-S complex and broken in the ligand-free state, the ion pair appears to function as the active site gate. Since the ion pairing residues are identically conserved in monofunctional PRODHs and PutAs, we suggest that the gate is a universal aspect of PRODH catalysis.

Whereas the ion pair gate opens to the bulk medium in DrPRODH, the gate of PutA separates the proline binding site from a large, internal cavity that traverses 41 Å to the P5CDH active site.[20] This cavity serves as a conduit through which P5C/GSA is channeled to the P5CDH active site. In the structure of full-length

Bradyrhizobium japonicum PutA, the gate is closed and a sulfate ion is bound in the PRODH active site. A structure of this enzyme with the proline site empty is not available, but based on the DrPRODH structures we predict that reduction of the FAD in PutA opens the gate allowing P5C/GSA to enter the substrate-channeling cavity. Curiously, dithionite-reduction of crystalline PutA86-669-THFA does not break the ion pair, although the flavin conformation is altered as in $DrPRODH_r$.[28] It is possible that crystal contacts inhibit rupture of the gate in crystalline PutA86-669. Clearly, new structures of full-length PutAs are needed to understand how the ion pair gate functions in PutA.

The structural changes observed in DrPRODH perhaps also provide insight into functional switching of trifunctional PutAs. Trifunctional PutA is a type of flavin switch protein that serves as a transcriptional repressor of the *put* regulon in addition to having PRODH and P5CDH catalytic activities.[31] How trifunctional PutA switches from being a DNA-bound transcriptional repressor to membrane-associated bifunctional enzyme (aka functional switching) is a major question in PutA biochemistry. Recent rapid reaction kinetic measurements of the paradigmatic trifunctional PutA from *E. coli* revealed an isomerization step that occurs after reduction of FAD, and this step was proposed to report on flavin-dependent, global conformational changes that drive functional switching.[32] Since rupture of the ion pair gate and movement of α8 occur after flavin reduction in DrPRODH, it is possible that analogous changes are also part of functional switching. Because α8 and the β1-α1 loop contact other structural elements in PutA that are absent in monofunctional PRODHs, rupture of the ion pair and shifting of α8 in trifunctional PutA could help initiate a cascade of events that eventually leads to a larger, more global isomerization

of the protein that unveils a high affinity membrane-association domain.

Finally, although it was not possible to crystallize DrPRODH in the absence of THFA, presumably the active site of the oxidized enzyme is open prior to substrate binding (state I in Figure 2.1B). Indeed, this is true for TtPRODH. The structure of oxidized, ligand-free TtPRODH resembles $DrPRODH_r$ in that the ion pair gate is open and α8 is shifted away from the isoalloxazine (Figure 2.2B). In fact, the shift of α8 in TtPRODH is 1.3. Å larger than that of $DrPRODH_r$ (Figure 2.2B). Whether substrate binding to the open enzyme induces closure of the gate (induced fit) or the substrate binds to closed-like conformations that arise from thermal fluctuations (conformational selection) remains to be determined. However, the large conformational changes that are required to form the closed E-S complex seem to favor conformational selection as a plausible mechanism for substrate binding.[33]

2.6 REFERENCES

1. Adams, E., and Frank, L. (1980) Metabolism of proline and the hydroxyprolines, *Annu. Rev. Biochem. 49*, 1005-1061.

2. Tanner, J. J. (2008) Structural biology of proline catabolism, *Amino Acids 35*, 719-730.

3. Singh, R. K., and Tanner, J. J. (2012) Unique Structural Features and Sequence Motifs of Proline Utilization A (PutA), *Front. Biosci. 17*, 556-568.

4. Phang, J. M. (1985) The regulatory functions of proline and pyrroline-5-carboxylic acid, *Curr. Top. Cell. Reg. 25*, 92-132.

5. Phang, J. M., Donald, S. P., Pandhare, J., and Liu, Y. (2008) The metabolism of proline, a stress substrate, modulates carcinogenic pathways, *Amino Acids 35*, 681-690.

6. Donald, S. P., Sun, X. Y., Hu, C. A., Yu, J., Mei, J. M., Valle, D., and Phang, J. M. (2001) Proline oxidase, encoded by p53-induced gene-6, catalyzes the generation of proline-dependent reactive oxygen species, *Cancer Res. 61*, 1810-1815.

7. Hu, C. A., Donald, S. P., Yu, J., Lin, W. W., Liu, Z., Steel, G., Obie, C., Valle, D., and Phang, J. M. (2007) Overexpression of proline oxidase induces proline-dependent

and mitochondria-mediated apoptosis, *Mol. Cell. Biochem. 295*, 85-92.

8. Liu, Y., Borchert, G. L., Surazynski, A., Hu, C. A., and Phang, J. M. (2006) Proline oxidase activates both intrinsic and extrinsic pathways for apoptosis: the role of ROS/superoxides, NFAT and MEK/ERK signaling, *Oncogene 25*, 5640-5647.

9. Pandhare, J., Cooper, S. K., and Phang, J. M. (2006) Proline oxidase, a proapoptotic gene, is induced by troglitazone: evidence for both peroxisome proliferator-activated receptor gamma-dependent and -independent mechanisms, *J Biol Chem 281*, 2044-2052.

10. Cooper, S. K., Pandhare, J., Donald, S. P., and Phang, J. M. (2008) A novel function for hydroxyproline oxidase in apoptosis through generation of reactive oxygen species, *J Biol Chem 283*, 10485-10492.

11. Phang, J. M., Pandhare, J., and Liu, Y. (2008) The metabolism of proline as microenvironmental stress substrate, *J Nutr 138*, 2008S-2015S.

12. Phang, J. M., Pandhare, J., Zabirnyk, O., and Liu, Y. (2008) PPARgamma and Proline Oxidase in Cancer, *PPAR Res 2008*, 542694.

13. Liu, Y., Borchert, G. L., Donald, S. P., Diwan, B. A., Anver, M., and Phang, J. M. (2009) Proline oxidase functions as a mitochondrial tumor suppressor in human cancers, *Cancer Res. 69*, 6414-6422.

14. Liu, Y., Borchert, G. L., Donald, S. P., Surazynski, A., Hu, C. A., Weydert, C. J., Oberley, L. W., and Phang, J. M. (2005) MnSOD inhibits proline oxidase-induced apoptosis in colorectal cancer cells, *Carcinogenesis 26*, 1335-1342.

15. Phang, J. M., Hu, C. A., and Valle, D. (2001) Disorders of proline and hydroxyproline metabolism, In *Metabolic and molecular basis of inherited disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., Eds.), pp 1821-1838, McGraw Hill, New York.

16. Willis, A., Bender, H. U., Steel, G., and Valle, D. (2008) PRODH variants and risk for schizophrenia, *Amino Acids 35*, 673-679.

17. Lee, Y. H., Nadaraia, S., Gu, D., Becker, D. F., and Tanner, J. J. (2003) Structure of the proline dehydrogenase domain of the multifunctional PutA flavoprotein, *Nat. Struct. Biol. 10*, 109-114.

18. Zhang, M., White, T. A., Schuermann, J. P., Baban, B. A., Becker, D. F., and Tanner, J. J. (2004) Structures of the Escherichia coli PutA proline dehydrogenase domain in complex with competitive inhibitors, *Biochemistry 43*, 12539-12548.

19. Ostrander, E. L., Larson, J. D., Schuermann, J. P., and Tanner, J. J. (2009) A conserved active site tyrosine residue of proline dehydrogenase helps enforce the preference for proline over hydroxyproline as the substrate, *Biochemistry 48*, 951-959.

20. Srivastava, D., Schuermann, J. P., White, T. A., Krishnan, N., Sanyal, N., Hura, G. L., Tan, A., Henzl, M. T., Becker, D. F., and Tanner, J. J. (2010) Crystal structure of the bifunctional proline utilization A flavoenzyme from Bradyrhizobium japonicum, *Proc. Natl. Acad. Sci. U. S. A. 107*, 2878-2883.

21. White, T. A., Krishnan, N., Becker, D. F., and Tanner, J. J. (2007) Structure and kinetics of monofunctional proline dehydrogenase from Thermus thermophilus, *J*

Biol Chem 282, 14316-14327.

22. Kabsch, W. (2010) XDS, *Acta Crystallogr. D Biol. Crystallogr. 66*, 125-132.

23. Evans, P. (2006) Scaling and assessment of data quality, *Acta Cryst. D62*, 72-82.

24. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement, *J. Appl. Cryst. 30*, 1022-1025.

25. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Cryst. D60*, 2126-2132.

26. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr., Sect. D 66*, 213-221.

27. Moxley, M. A., Tanner, J. J., and Becker, D. F. (2011) Steady-state kinetic mechanism of the proline:ubiquinone oxidoreductase activity of proline utilization A (PutA) from Escherichia coli, *Arch Biochem Biophys 516*, 113-120.

28. Zhang, W., Zhang, M., Zhu, W., Zhou, Y., Wanduragala, S., Rewinkel, D., Tanner, J. J., and Becker, D. F. (2007) Redox-induced changes in flavin structure and roles of flavin N(5) and the ribityl 2'-OH group in regulating PutA-membrane binding, *Biochemistry 46*, 483-491.

29. White, T. A., Johnson, W. H., Jr., Whitman, C. P., and Tanner, J. J. (2008) Structural basis for the inactivation of Thermus thermophilus proline dehydrogenase by N-propargylglycine, *Biochemistry 47*, 5573-5580.

30. Srivastava, D., Zhu, W., Johnson, W. H., Jr., Whitman, C. P., Becker, D. F., and Tanner, J. J. (2010) 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 49*, 560-569.

31. Becker, D. F., Zhu, W., and Moxley, M. A. (2011) Flavin redox switching of protein functions, *Antioxid Redox Signal 14*, 1079-1091.

32. Moxley, M. A., and Becker, D. F. (2012) Rapid reaction kinetics of proline dehydrogenase in the multifunctional proline utilization A protein, *Biochemistry 51*, 511-520.

33. Grant, B. J., Gorfe, A. A., and McCammon, J. A. (2010) Large conformational changes in proteins: signaling and other functions, *Curr. Opin. Struct. Bio. 20*, 142-147.

34. Engh, R. A., and Huber, R. (1991) Accurate bond and angle parameters for x-ray protein structure refinement, *Acta Cryst. A47*, 392-400.

35. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation, *Proteins 50*, 437-450.

36. DeLano, W. L. (2002) *The PyMOL User's Manual*, DeLano Scientific, Palo Alto, CA, USA.

Table 2-1 X-ray Diffraction Data Collection and Refinement[a]

	$DrPRODH_o$-THFA	$DrPRODH_r$
FAD redox state	oxidized	reduced
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters (Å)	a = 44.5, b = 95.5, c = 136.4	a = 43.4, b = 95.7, c = 136.0
Wavelength	0.979	1.00
Resolution (Å)	95.5 – 1.36 (1.43 - 1.36)	47.8 – 1.75 (1.84 - 1.75)
Observations	450935	415890
Unique reflections	124272	58089
$R_{merge}(I)$	0.032 (0.431)	0.065 (0.837)
$R_{meas}(I)$	0.043 (0.576)	0.071 (0.905)
$R_{pim}(I)$	0.021 (0.294)	0.026 (0.341)
Mean I/σ	17.1 (2.4)	22.1 (2.5)
Completeness (%)	98.9 (98.5)	100.0 (100.0)
Multiplicity	3.6 (3.7)	7.2 (7.0)
No. of protein residues	562	550
No. of atoms	5091	4557
No. of FAD atoms	212	106
No. of THFA atoms	16	0
No. of water molecules	475	297
R_{cryst}	0.148 (0.239)	0.182 (0.276)
R_{free}[b]	0.181 (0.279)	0.213 (0.323)
rmsd bond lengths (Å)[c]	0.015	0.013
rmsd bond angles (°)	1.75	1.45
Ramachandran plot[d]		
Favored (%)	99.3	99.3
Allowed (%)	0.7	0.7
Outliers (%)	0.0	0.0
Average B (Å^2)		
Protein	21.4	28.6
FAD	17.7	23.7
THFA	17.5	n/a
Acetate	n/a	26.2
Water	31.5	31.4
Coordinate error (Å)[e]	0.14	0.19
PDB code	4H6Q	4H6R

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

[b]A common test set (5 %) was used for refinement of both structures.

[c]Compared to the parameters of Engh and Huber.[34]

[d]The Ramachandran plot was generated with RAMPAGE.[35]

[e]Maximum likelihood-based coordinate error estimate.

Table 2-2 Kinetic Parameters for DrPRODH and DrPRODH Mutant Enzymes Using Proline and CoQ1 as the Substrates

	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} M^{-1})	Variational Effect (k_{cat}/K_m of mutant)/ (k_{cat}/K_m of DrPRODH)
		Proline as the variable substrate[a]		
DrPRODH	8.7 ± 0.58	290 ± 39	30 ± 4.5	
G63A	0.080 ± 0.006	384 ± 68	0.21 ± 0.04	0.007 ± 0.002
E64A	0.055 ± 0.001	50 ± 6	1.1 ± 0.13	0.037 ± 0.007
		CoQ1 as the variable substrate[b]		
DrPRODH	14 ± 1	0.155 ± 0.04	90323 ± 24185	
G63A	0.043 ± 0.003	0.028 ± 0.01	1535 ± 559	0.017 ± 0.008
E64A	0.046 ± 0.003	0.032 ± 0.01	1438 ± 458	0.016 ± 0.007

[a][CoQ1] was fixed at 200 μM.

[b][Proline] was fixed at 500 mM.

FIGURE LEGENDS

Figure 2-1. (A) Reactions catalyzed by PRODH and P5CDH. (B) Scheme showing the enzyme states that occur during the PRODH reductive half-reaction

Figure 2-2. Overall fold of PRODH. (A) Two views of the DrPRODH$_o$-THFA structure. The backbone is colored according to a rainbow scheme, with blue at the N-terminus and red at the C-terminus. FAD is colored yellow. THFA is colored pink. Residue numbers of the N- and C-termini are noted. (B) Superposition of DrPRODH$_o$-THFA (yellow), DrPRODH$_r$ (gray), and TtPRODH (cyan, PDB code 2g37). The R291-E64 ion pair is observed only in the DrPRODH$_o$-THFA structure. This figure and others were prepared with PyMOL.[36]

Figure 2-3. The active site of DrPRODH$_o$-THFA (relaxed stereographic views). (A) Electron density map for DrPRODH$_o$-THFA. The cage represents a simulated annealing σ_A-weighted F_o - F_c omit map (3.0 σ). (B) Superposition of DrPRODH$_o$-THFA (yellow) and PutA86-669-THFA (gray, PDB code 1tiw), highlighting differences in the orientation of the conserved glutamate of α8 (Glu295 in DrPRODH, Glu559 in PutA). Yellow and black dashes denote the unique electrostatic interactions of DrPRODH and PutA86-669, respectively. Conformations A and B of the FAD in DrPRODH$_o$-THFA are colored yellow and pink, respectively.

Figure 2-4. Conformations of the flavin in DrPRODH (relaxed stereographic views). (A) The dual FAD conformations of DrPRODH$_o$-THFA. Conformations A (q = 0.56) and B (q = 0.44) are colored yellow and pink, respectively. Yellow and red dashes denote the unique hydrogen bonds of conformations A and B, respectively. Note that the two conformations differ mainly in the orientations of the 2′-OH and 3′-

OH groups of the ribityl. (B and C) Two views of the dithionite-reduced flavin of $DrPRODH_r$. In all three panels, the cage represents a simulated annealing σ_A-weighted F_o - F_c omit map (3.0 σ).

Figure 2-5. Comparison of flavin conformations in PRODHs and PutAs. (A) Superposition FAD conformation A of $DrPRODH_o$-THFA (yellow) and the FAD of PutA86-669-THFA (salmon, PDB code 1tiw). (B) Superposition of FAD conformation B of $DrPRODH_o$-THFA (pink) and the oxidized flavin of TtPRODH (cyan, PDB code 2g37). (C) Superposition of the reduced flavin of $DrPRODH_r$ (gray), the oxidized FAD of TtPRODH (cyan), and the inactivated flavin of the PutA PRODH domain (magenta, PDB code 3ITG).

Figure 2-6. Conformational changes induced by flavin reduction and product release. (A) Superposition of the active sites of $DrPRODH_o$-THFA (yellow) and $DrPRODH_r$ (gray) highlighting the shift of α8, reconfiguration of the β1-α1 loop, and breaking of the Arg291-Glu64 ion pair. Black dashes indicate electrostatic interactions in $DrPRODH_o$-THFA. (B) Surface representation of $DrPRODH_o$-THFA. FAD and THFA are drawn as spheres in yellow and pink, respectively. Glu64 and Arg291 are colored red and blue, respectively. Glu295 is colored green. The inset shows a view from the bottom of the active site. (C) Surface representation of $DrPRODH_r$. The coloring is the same as in panel B. The dashes outline the openings to the active site.

Figure 2-1 (A) Reactions catalyzed by PRODH and P5CDH. (B) Scheme showing the enzyme states that occur during the PRODH reductive half-reaction

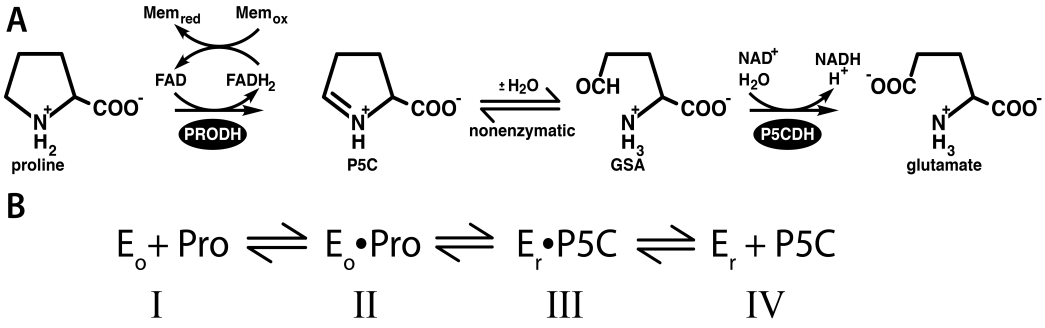


Figure 2-2 Overall fold of PRODH.

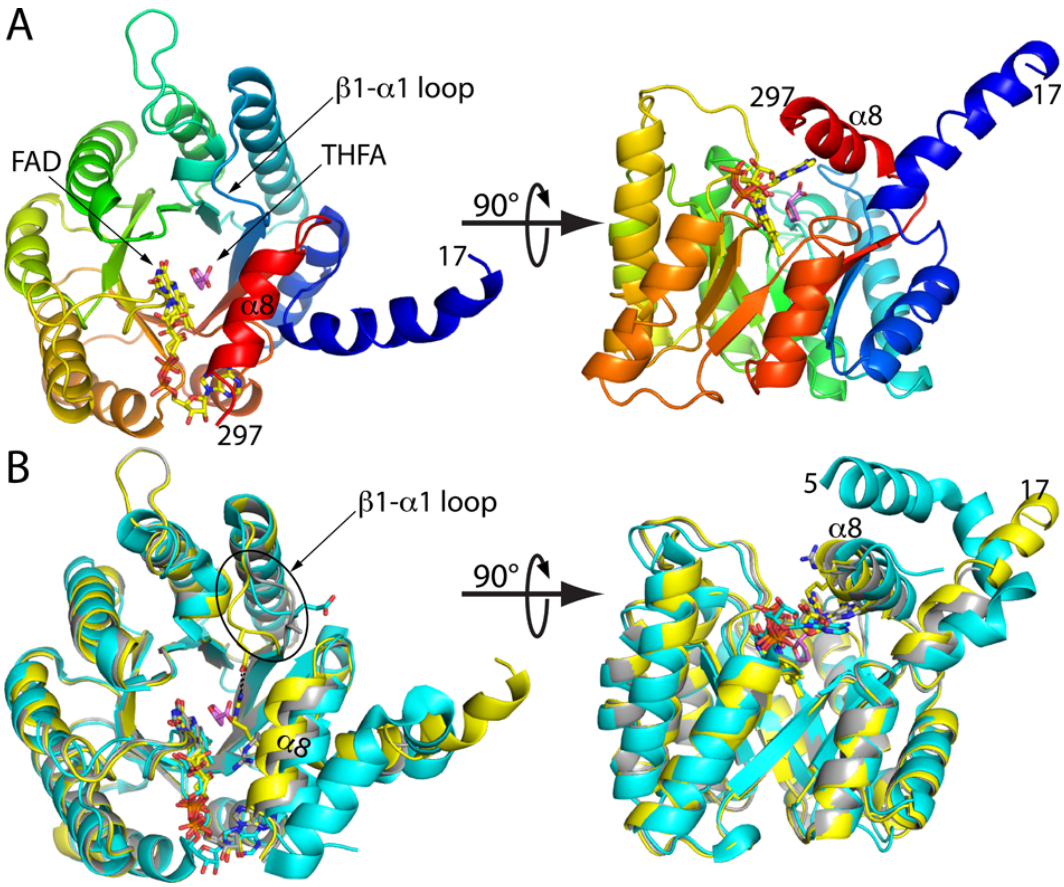


Figure 2-3 The active site of DrPRODHo-THFA (relaxed stereographic views).

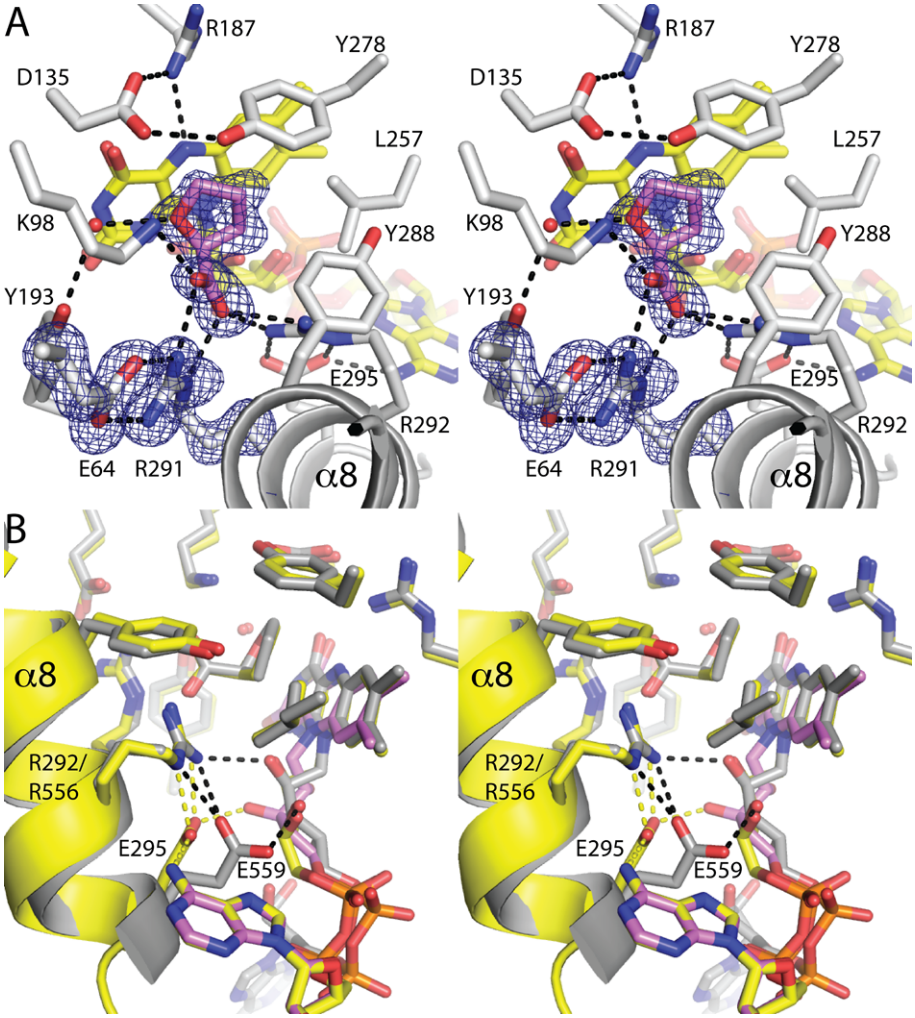


Figure 2-4 Conformations of the flavin in DrPRODH (relaxed stereographic views).

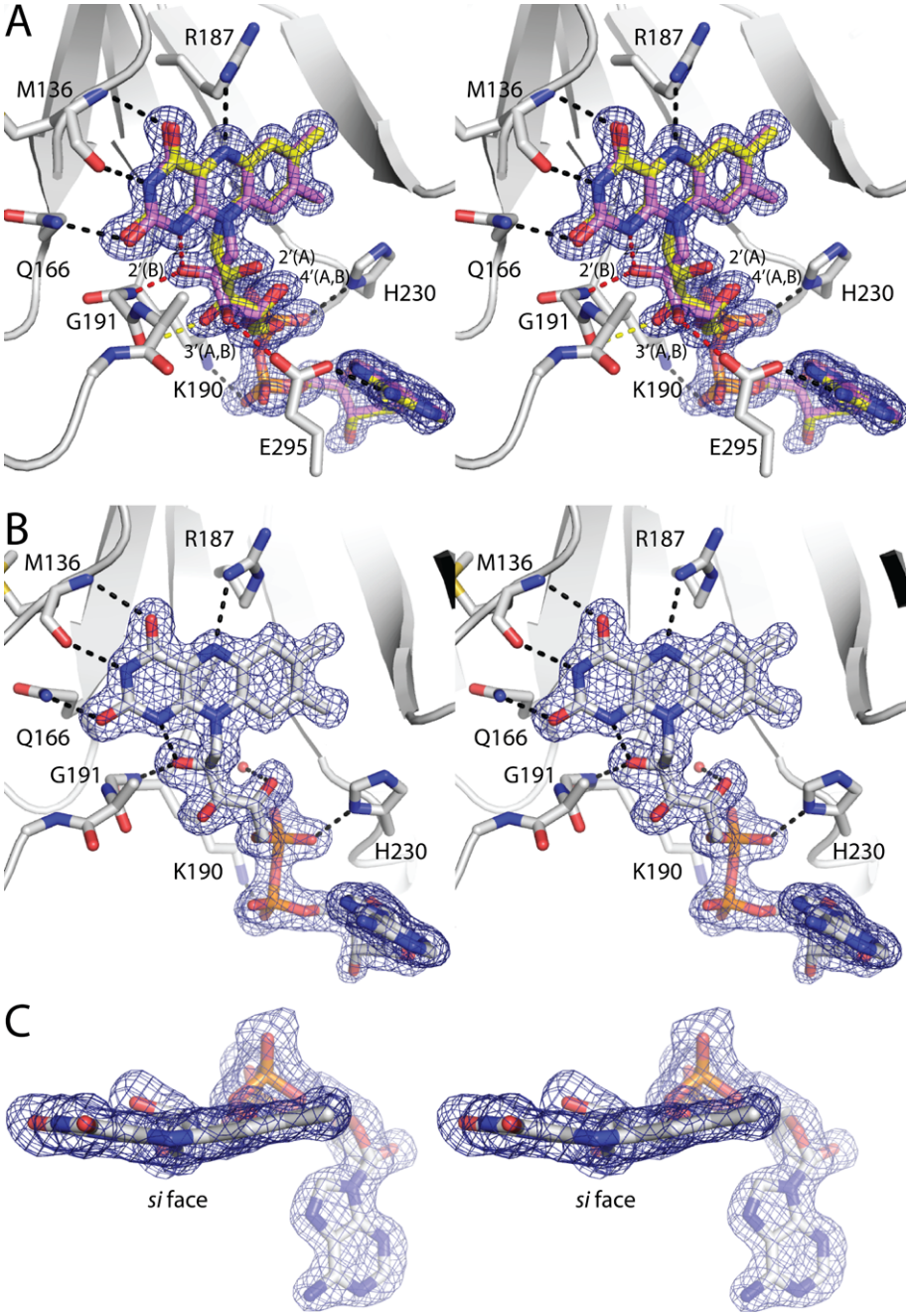


Figure 2-5 Comparison of flavin conformations in PRODHs and PutAs.

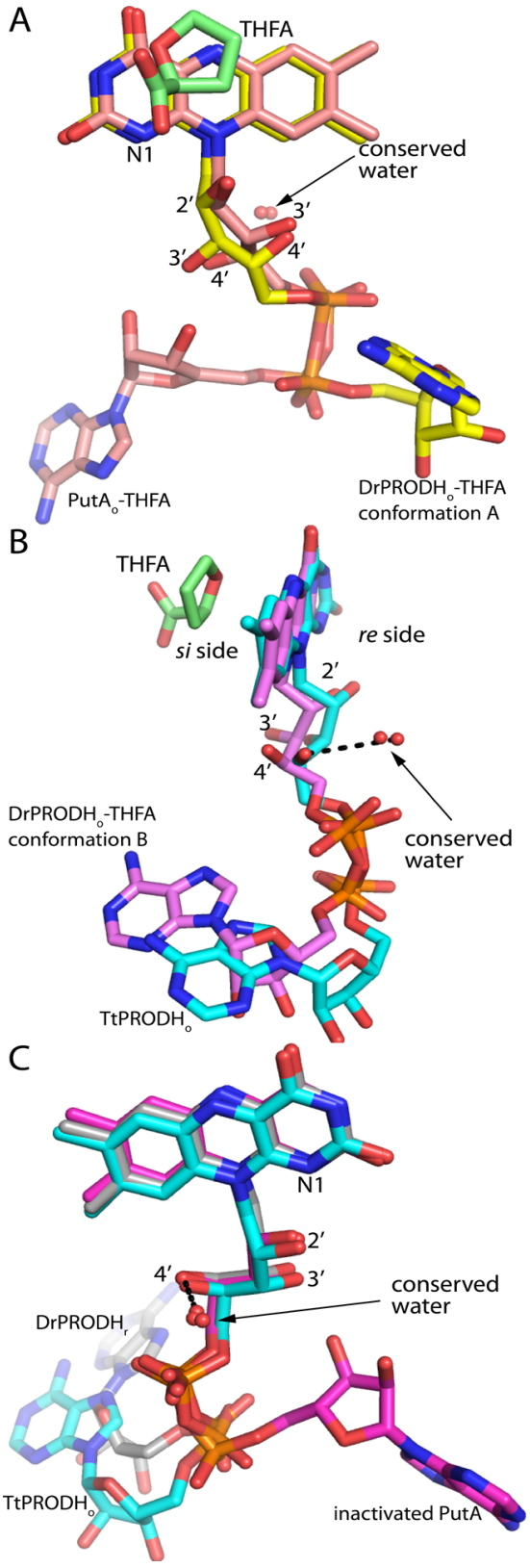
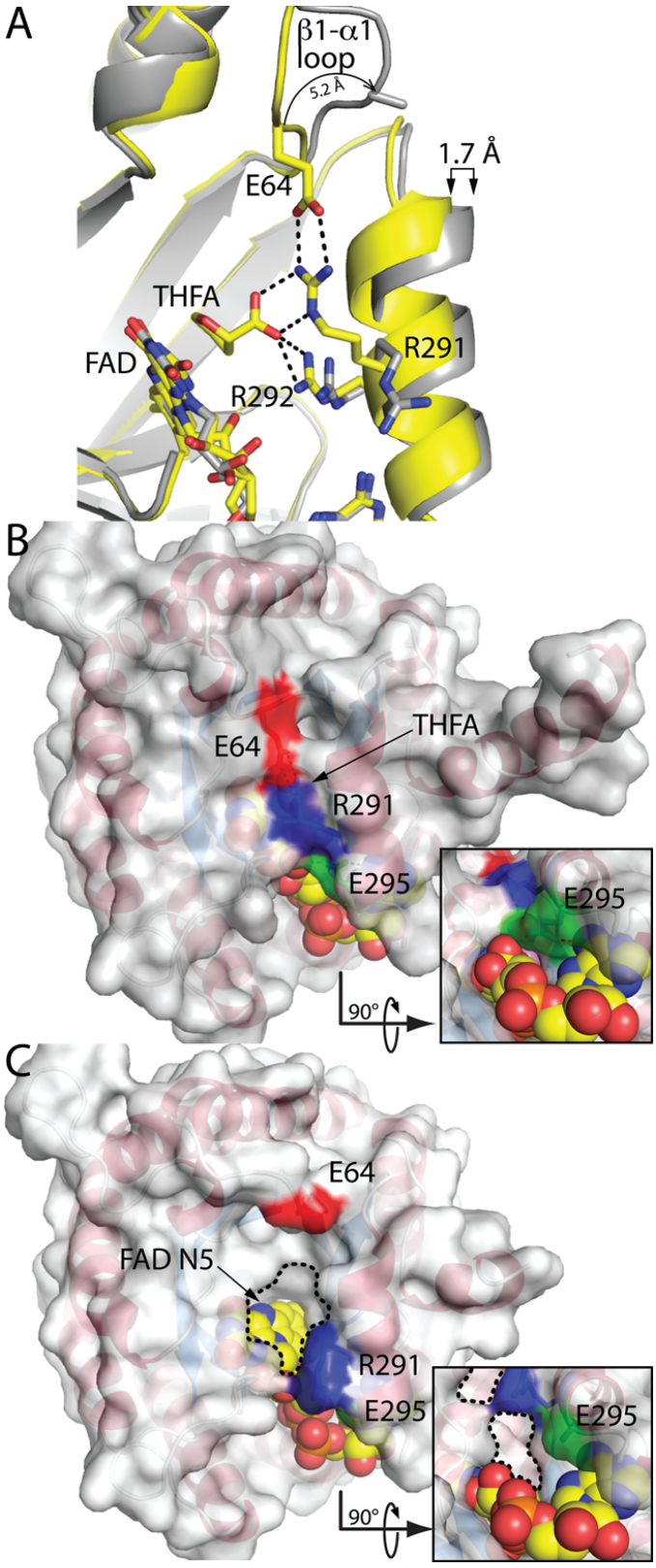


Figure 2-6 Conformational changes induced by flavin reduction and product release.



CHAPTER 3

3. STRUCTURAL DETERMINANTS OF OLIGOMERIZATION OF 1-PYRROLINE-5-CARBOXYLATE DEHYDROGENASE: IDENTIFICATION OF A HEXAMERIZATION HOT SPOT

Min Luo, et al. Structural Determinants of Oligomerization of 1-Pyrroline-5-Carboxylate Dehydrogenase: Identification of a Hexamerization Hot Spot. J. Mol. Biol. (2013) 425(17):3106-3120.

3.1 ABSTRACT

The aldehyde dehydrogenase (ALDH) superfamily member Δ^1-pyrroline-5-carboxylate dehydrogenase (P5CDH) catalyzes the NAD^+-dependent oxidation of glutamate semialdehyde to glutamate, which is the final step of proline catabolism. Defects in P5CDH activity lead to the metabolic disorder type II hyperprolinemia, P5CDH is essential for virulence of the fungal pathogen *Cryptococcus neoformans*, and bacterial P5CDHs have been targeted for vaccine development. Although the enzyme oligomeric state is known to be important for ALDH function, the oligomerization of P5CDH has remained relatively unstudied. Here we determine the oligomeric states and quaternary structures of four bacterial P5CDHs using a combination of small-angle X-ray scattering, X-ray crystallography, and dynamic light scattering. The P5CDHs from *Thermus thermophilus* and *Deinococcus radiodurans* form trimer-of-dimers hexamers in solution, which is the first observation of a hexameric ALDH in solution. In contrast, two *Bacillus* P5CDHs form dimers in solution but do not assemble into a higher order oligomer. Site-directed mutagenesis was used to identify a hexamerization hot spot that is centered on an arginine residue in the NAD^+-binding domain. Mutation of this critical Arg residue to Ala in either of the hexameric enzymes prevents hexamer formation in solution. Paradoxically, the dimeric Arg-to-Ala *T. thermophilus* mutant enzyme packs as a hexamer in the crystal state, which illustrates the challenges associated with predicting the biological assembly in solution from crystal structures. The observation of different oligomeric states among P5CDHs suggests potential differences in cooperativity and protein-protein interactions.

ABBREVIATIONS

PRODH, proline dehydrogenase; P5C, Δ^1-pyrroline-5-carboxylate; P5CDH, Δ^1-pyrroline-5-carboxylate dehydrogenase; PutA, proline utilization A; ALDH, aldehyde dehydrogenase; TtP5CDH, Δ^1-pyrroline-5-carboxylate dehydrogenase from *Thermus thermophilus*; DrP5CDH, Δ^1-pyrroline-5-carboxylate dehydrogenase from *Deinococcus radiodurans*; BhP5CDH, Δ^1-pyrroline-5-carboxylate dehydrogenase from *Bacillus halodurans*; BlP5CDH, Δ^1-pyrroline-5-carboxylate dehydrogenase from *Bacillus licheniformis*; SAXS, small-angle X-ray scattering; DLS, dynamic light scattering; TEVP, Tobacco Etch Virus protease; SEC, size exclusion chromatography.

3.2 INTRODUCTION

The enzyme Δ^1-pyrroline-5-carboxylate (P5C) dehydrogenase (P5CDH, EC 1.5.1.12) catalyzes the NAD^+-dependent oxidation of glutamate semialdehyde to glutamate, which is the final step of proline catabolism (Fig. 3.1).[1] In humans, P5CDH also catalyzes the final step of hydroxyproline catabolism, which is the conversion of 4-hydroxyglutamate semialdehyde to 4-erythro-hydroxy-L-glutamate. The enzyme is widely distributed in eukaryotes and bacteria. In the former organisms, P5CDH is localized to the mitochondrial matrix. In some bacteria, mostly Gram-negative bacteria, P5CDH is combined with the first enzyme of proline catabolism, proline dehydrogenase (PRODH), into the bifunctional enzyme proline utilization A (PutA).[2,3] However, in Gram-positive bacteria, PRODH and P5CDH are separate monofunctional enzymes encoded by distinct genes. Monofunctional P5CDH is the subject of this research.

Defects in P5CDH underlie the autosomal recessive disorder type II hyperprolinemia.[4-8] Certain missense and frameshift mutations in the gene encoding P5CDH abrogate enzyme function, resulting in elevated levels of P5C and proline in plasma, urine, and cerebrospinal fluid.[9] Type II hyperprolinemia is causally linked to neurologic manifestations, such as increased incidence of seizures and intellectual and developmental disabilities,[10] although exactly how the enzyme deficiency contributes to these conditions is unclear. Possible mechanisms involve the role of proline as a neurotransmitter,[11-14] oxidative stress,[9] and mitochondrial dysfunction.[15]

P5CDHs from microorganisms have been gaining attention as well. A recent

study of proline catabolism genetics in the fungal pathogen *Cryptococcus neoformans* showed that P5CDH is required for optimal production the major cryptococcal virulence factors.[16] This work also showed that a mutant strain of *C. neoformans* in which P5CDH was disabled is avirulent in a mouse model of infection. Recent immunological studies have shown that P5CDH is associated with the outer surface of two important bacterial pathogens, *Streptococcus pneumoniae* and *Staphylococcus aureus*.[17] Although the secretion mechanism and nature of the physical interaction that anchors the enzyme to the surface are unknown, P5CDH appears to be a new member of an emerging class of anchorless surface proteins found in Gram-positive bacteria that are thought to be potential virulence factors.[18] Furthermore, *S. pneumoniae* and *S. aureus* P5CDH are immunogenic and have been proposed as components of vaccines against *S. aureus*, a bacterial pathogen that causes significant morbidity, mortality, and healthcare costs worldwide.[17] Three-dimensional structural studies of P5CDHs should aid efforts to design inhibitors of fungal P5CDHs and new vaccines based on bacterial P5CDHs.

P5CDH is a member of the vast aldehyde dehydrogenase (ALDH) superfamily and is known as ALDH4A1. Like other ALDHs, P5CDH exhibits a 3-domain fold, as exemplified by *Thermus thermophilus* P5CDH, the first P5CDH to be structurally characterized (Fig. 3.2a).[19] The catalytic domain provides the essential Cys nucleophile that attacks the C atom of the substrate aldehyde group. The $NAD(P)^+$-binding domain exhibits the Rossmann fold. The oligomerization domain is a β substructure that protrudes from the NAD^+-binding domain and consists of a β-hairpin and the final β-strand of the polypeptide.

Oligomerization is an important aspect of the ALDH structure-function paradigm. All ALDHs form a domain-swapped dimer in which the oligomerization domain of one protomer engages the catalytic domain of the other protomer (Fig. 3.2b). In some ALDHs, such as ALDH1 and ALDH2, two of these dimers assemble into a dimer-of-dimer homotetramer.[20] Oligomerization appears to be important for ALDH function and stability.[20] For example, the inactive form of ALDH2 found in 40% of the East Asian population has a glutamate to lysine substitution in an oligomerization interface.[21] Also, the tetrameric assembly of ALDH protomers appears to be essential for the half-of-the-sites reactivity exhibited by ALDH1A1 and ALDH2.[22]

In contrast to other ALDHs, relatively little is known about the oligomeric states and quaternary structures of P5CDHs. The oligomeric state in solution has been determined only for human P5CDH, which forms the classic ALDH homodimer, but does not assemble into a higher order oligomer.[23] To address this knowledge gap, we report an analysis of the oligomeric states and quaternary structures of the bacterial P5CDHs from *Thermus thermophilus* (TtP5CDH), *Deinococcus radiodurans* (DrP5CDH), *Bacillus halodurans* (BhP5CDH), and *Bacillus licheniformis* (BlP5CDH).

3.3 EXPERIMENTAL PROCEDURES

3.3.1 Expression and purification of TtP5CDH

The gene encoding P5CDH from *Thermus thermophilus* HB27 (NCBI Reference Sequence YP_005182.1, 516 residues) was cloned from genomic DNA into pKA8H using *NdeI* and *BamHI* sites by the University of Missouri DNA Core. The expressed

protein includes an N-terminal His_8 tag and Tobacco Etch Virus Protease (TEVP) cleavage site. Treatment with TEVP results in Gly-His followed by the TtP5CDH polypeptide.

TtP5CDH was expressed in *Escherichia coli* BL21(DE3)pLysS. A 10 mL overnight culture was used to inoculate 1 L of LB media. The cells were grown at 37 °C at 250 rpm and induced with 0.5 mM IPTG at OD_{600} = 0.6 with an induction temperature of 22 °C at 200 rpm for 18 hours. Cells were harvested by centrifugation, resuspended in 50 mM HEPES, 300 mM NaCl, 10 mM imidazole, and 5% glycerol at pH 8.0, and flash-frozen in liquid nitrogen. The frozen cells were thawed with protease inhibitors (AEBSF, TPCK, E64, Pepstatin, Leupeptin) and broken by sonication. Cell debris and unbroken cells were separated by centrifugation at 16,500 rpm for one hour in a SS34 rotor. The supernatant was applied to a His-Trap HP column charged with Ni^{2+} (GE Healthcare) equilibrated with 20 mM HEPES, 300 mM NaCl, and 5% glycerol at pH 8.0. TtP5CDH was eluted with equilibration buffer containing 300 mM imidazole. Fractions were pooled, and TEVP, 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 °C for 2 hours and dialyzed overnight at 4 °C before injection onto the His-Trap HP column. The tag-free 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 anion exchange chromatography (HiTrap Q). The protein was eluted with a linear 0 - 1 M NaCl gradient. Size exclusion chromatography (SEC) (Superdex 200, 25 mL) was used as the final step of purification.

3.3.2 Expression and purification of DrP5CDH

A gene encoding DrP5CDH (NCBI Reference Sequence NP_294537, 523 residues) with codons optimized for expression in *E. coli* was synthesized (Genscript) and subcloned into pKA8H using *NdeI* and *BamHI* sites. The expressed protein includes an N-terminal His_8 tag and TEVP site. Cleavage with TEVP produces Gly-His followed by the DrP5CDH polypeptide.

DrP5CDH was expressed in BL21(DE3)pLysS (induction at OD_{600} = 0.8 with 0.5 mM IPTG for 5 hours at 22 °C). The cells were collected by centrifugation, resuspended in 50 mM Tris, 100 mM NaCl, 10 mM imidazole, 5% glycerol, at pH 7.5, and frozen at -80 °C.

The frozen cells were thawed at 4 °C in the presence of protease inhibitors (0.1 mM TPCK, 0.05 mM AEBSF, 0.1 μM Pepstatin, 0.01 mM Leupeptin, 5μM E-64) and broken using sonication. The mixture was centrifuged at 16500 rpm in an SS34 rotor for 1 hour at 4°C, filtered through a 0.45 μm filter (Millipore) and loaded on a HisTrap HP column (5 mL) that had been charged with $NiCl_2$ and equilibrated in 50 mM Tris, 300 mM NaCl, 10 mM imidazole, and 5% glycerol at pH 7.5. Washing steps were performed using the loading buffer supplemented with 10 mM imidazole followed by 30 mM imidazole. The protein was eluted with 300 mM imidazole. The histidine tag was removed by incubating the protein with 0.2 mg/ml TEVP for 1 hour at 28°C followed by dialysis at 4 °C against 50 mM Tris, 50 mM NaCl, and 5% glycerol at pH 7.5. The mixture was applied to the HisTrap HP column to separate the cleaved protein, which appeared in the flow-through, from the tag and TEVP. The cleaved protein was dialyzed overnight at 4 °C into 50 mM Tris, 0.5 mM EDTA, 0.5 mM DTT, and 5% glycerol at pH 7.8 in preparation for anion exchange

chromatography (HiTrap Q). The sample was loaded onto the column using a buffer of 50 mM Tris and 5% glycerol at pH 7.8, and a linear NaCl gradient was applied. DrP5CDH eluted at 280 - 340 mM NaCl. The protein concentration was estimated using the bicinchoninic acid method (Pierce kit) with bovine serum albumin as the standard.

3.3.3 Expression and purification of BhP5CDH and BlP5CDH

Expression constructs for BhP5CDH (NCBI Reference Sequence NP_243603.1, 515 residues) and BlP5CDH (NCBI Reference Sequence YP_077616.1, 516 residues) were obtained from the New York Structural Genomics Research Consortium. The expressed enzymes have C-terminal His_8-tags and a TEVP cleavage site. TEVP cleavage produces the enzyme polypeptide followed by AENLYFQ. The enzymes were expressed and purified using the protocols described for TtP5CDH.

3.3.4 Site-directed mutagenesis

Site-directed mutants of TtP5CDH and DrP5CDH were generated using the QuikChange II site-directed mutagenesis kit (Agilent). The mutations were confirmed with sequencing performed by the University of Missouri DNA core. The mutant enzymes were purified as described above for the native enzymes.

3.3.5 SAXS

SAXS experiments were performed at beamline 12.3.1 of the Advanced Light Source via the mail-in program.[37,38] Prior to data collection, all protein samples were subjected to SEC using a Superdex 200 column. The column buffer was typically 50 mM Tris, 5% glycerol, 0.5 mM THP, and 50 mM NaCl at pH 7.5. In some cases (e.g., DrP5CDH), the SEC fractions were pooled, concentrated to ~12 mg/mL, and

dialyzed at 4 °C for 24 hours against 50 mM Tris, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM THP, and 5% glycerol at pH 7.8. For each protein, scattering intensities were measured at three nominal protein concentrations (1 - 10 mg/mL). For each protein concentration, exposure times of 0.5, 1.0, 3.0, and 6.0 sec were used. Scattering curves collected from the protein samples were corrected for background scattering using intensity data collected from the SEC effluent or dialysis buffer.

The SAXS data were analyzed as follows. A composite scattering curve for each sample was generated with PRIMUS[39] by scaling and merging the high q region from one of the longer time exposures with the low q region from a shorter time exposure. The scattering curves were multiplied with a concentration factor and overlaid on each other to check for concentration dependent variation of the profile. No substantial concentration effects were observed for any of the samples. PRIMUS was also used to perform Guinier analysis. FoXS was used to calculate theoretical scattering profiles from atomic models.[40] FoXS was also used to perform minimum ensemble calculations.[26] GNOM was used to calculate pair distribution functions.[41] MOLEMAN was used to calculate R_g from atomic coordinates.[42] The SASTBX server[43] was used for shape reconstruction calculations.

3.3.6 Estimation of molecular weight from SAXS volume of correlation

The molecular weight was estimated from the volume of correlation (V_c) as described recently by Rambo and Tainer.[27] Briefly, V_c is a new SAXS invariant defined as the ratio of the zero angle scattering intensity, $I(0)$, to the total scattered intensity. The latter quantity is equal to the integral $\int qI(q)\mathrm{d}q$ performed over the entire range of the scattering data. V_c thus has units of Å^2. Rambo and Tainer showed that V_c is independent of solute concentration and the aforementioned integral

converges for both folded-compact and unfolded-flexible particles. For proteins, they also demonstrated that the molecular weight (M) in units of Da can be estimated from a single SAXS curve by the relationship, $M = V_c^2 R_G^{-1}/0.1231$. An analogous relationship was provided for RNA.

V_c was calculated as follows. First, $I(0)$ and R_G were estimated from Guinier analysis using PRIMUS. These values were used to extrapolate the experimental SAXS curves to $q = 0$ using the equation, $I(q) = I(0) \exp(-q^2 R_G^2/3)$. Extrapolation to $q = 0$ is needed for accurate calculation of the total scattered intensity. The extrapolated region consisted of 18 – 19 points with q spacing of 0.000610 Å^{-1}, which matches the q spacing of the experimental data. The area under the curve of $qI(q)$ versus q was calculated using the Polygon Area utility of Origin 9 software, and V_c was calculated as the ratio of $I(0)$ to the area.

3.3.7 Crystallization of TtP5CDHR100A

Crystals of TtP5CDHR100A were grown in sitting drops at room temperature using the protocol for TtP5CDH.[44] Briefly, screening using commercially available kits (Hampton Research) yielded positive results in several conditions, which led to either the twinned form described earlier[44] or the more desired *H*3 form that was used for structure determination. Space group *H*3 crystals of TtP5CDHR100A were grown using sitting drops formed by mixing 1 μL of the protein stock solution (7 mg/mL protein in 50 mM Tris-HCl at pH 7.5, 100 mM NaCl, 5 % (v/v) glycerol, 0.5 mM THP, and 0.5 mM EDTA) and 1 μL of the reservoir containing 45 % 2-methyl-2,4-pentanediol and 0.05 M sodium citrate buffer at pH 5.2. The reservoir was used as the cryoprotectant. The space group is *H*3 with unit cell dimensions of a = 173 Å and

c = 279 Å and two protein molecules in the asymmetric unit.

3.3.8 Crystallization of TtP5CDHR100A/K104A/R111A

Triclinic crystals of the triple mutant TtP5CDHR100A/K104A/R111A were grown at 295 K with the sitting-drop method of vapor diffusion. Initial conditions were identified using commercially available crystal screens (Hampton Research). TtP5CDHR100A/K104A/R111A was crystallized using a reservoir of 0.2 M $MgCl_2$, 0.1 M Tris-HCl pH 8.5, and 30% w/v PEG 4000. The protein stock solution contained 5 mg/mL TtP5CDH triple mutant in the buffer of 50 mM Tris, 100 mM NaCl, 0.5 mM THP, 0.5 mM EDTA, and 5% glycerol at pH 7.8. Crystals were cryoprotected with 0.2 M $MgCl_2$, 0.1 M Tris-HCl at pH 8.5, 30% w/v PEG 4000, and 28% PEG 200. The space group is $P1$ with unit cell dimensions a = 65.1 Å, b = 102.7 Å, c = 160.6 Å, α = 86.3°, β =87.5°, and γ = 79.4°. The asymmetric unit includes 8 protein molecules (4 dimers), which implies 48 % solvent and V_M of 2.35 $Å^3$/Da.

3.3.9 X-ray diffraction data collection, phasing, and refinement

X-ray diffraction data from crystals of TtP5CDHR100A in space group $H3$ were collected at beamline 4.2.2 of Advanced Light Source. The 1.54 Å resolution data set used for refinement consisted of 360 frames with an oscillation width of 0.5^o per image, detector distance 110 mm, and exposure time 2 s/image. The data were processed with XDS[45] and SCALA[46] via CCP4i.[47] Refinement using PHENIX[48] commenced from the coordinates of TtP5CDH with Arg100 truncated to Ala. COOT[49] was used for model building. Data collection and refinement statistics are listed in Table 3.2.

X-ray diffraction data from crystals of the triple mutant

TtP5CDHR100A/K104A/R111A were collected on an in-house Rigaku rotating anode generator coupled to an R-AXIS IV++ detector. Two data sets were recorded and merged. The first one consisted of 360 frames collected with an oscillation width of 0.5°, detector distance of 270 mm, and exposure time of 7 minutes per frame. The second set consisted of 360 frames collected with an oscillation width of 0.5°, detector distance of 200 mm, and exposure time of 5.2 minutes per frame. The data were integrated with MOSFLM[50] and merged to 2.42 Å resolution with SCALA. The arrangement of the dimers in the asymmetric unit was determined using molecular replacement as implemented in PHASER[51] with a search model derived from a dimer of TtP5CDHR100A with Lys104 and Arg111 truncated to Ala. A clear solution with 4 dimers in the unit cell and log-likelihood gain of 31444 was obtained. Rigid body refinement yielded $R_{work} = R_{free} = 0.290$ for all data to 2.42 Å resolution.

3.3.10 DLS

These experiments were performed on a Protein Solutions DynaPro 99 Molecular Sizing Instrument (Wyatt Technology) at 20 °C with a wavelength of 836.3 nm and scattering angle of 90°. Protein concentrations were in the range 0.5 - 2 mg/ml in 50 mM Tris buffer at pH 7.5. Prior to DLS, each sample was centrifuged for 10 min at 13,000 × g at 4°C and passed through a 0.22 μm Millipore filter (Whatman). The data were collected with acquisition time of 10 s with at least 18 acquisitions. The DLS data were analyzed by the program DYNAMICS v.5.26.38 by performing regularization fit using the regularization algorithm on the measured autocorrelation functions.

3.3.11 Steady-state kinetics

The P5CDH activity of TtP5CDH and TtP5CDH mutant enzymes were measured at 20 °C by monitoring NADH production at 340 nm as described previously.[23] The final assay mixture (1 mL) contained 6 μg/ml P5CDH (0.1 μM), 1 mM NAD^+, and various concentrations of P5C in 0.1 M potassium phosphate buffer at pH 7.5. The pH of the P5C stock solution was adjusted to 7.5 before adding to the reaction mixture. Kinetic constants were estimated by fitting the initial rate data to the Michaelis-Menten equation using Origin 9.0.

3.4 RESULTS

3.4.1 TtP5CDH and DrP5CDH form hexamers in solution

TtP5CDH and DrP5CDH were analyzed with small-angle X-ray scattering (SAXS), which can define solution conformation and assembly states in combination with crystallography.[24,25] The SAXS curve for TtP5CDH is shown in Fig. 3.3a. The Guinier plot exhibits good linearity and yields radius of gyration (R_g) of 43.4 ± 0.3 Å (Table 3.1). Calculations of the pair distribution function (P(r)) suggest R_g of 43.0 – 43.2 Å for maximum particle dimension (D_{max}) of 120 - 125 Å. The P(r) function has one maximum centered at r = 60 Å. Similar results were obtained for DrP5CDH, suggesting that the two enzymes adopt the same oligomeric state in solution (Fig. 3.3c). The Guinier R_g for DrP5CDH is 43.1 ± 0.1 Å. The R_g and D_{max} estimates from P(r) calculations are 43.6 Å and 120 - 125 Å, respectively. The P(r) function for DrP5CDH is nearly identical to that of TtP5CDH.

The crystal structure of TtP5CDH was used to help determine the quaternary structures of TrP5CDH and DrP5CDH. The crystal structure of TtP5CDH was reported by Inagaki and coworkers in 2006.[19] The space group is *H*3 with one classic

ALDH dimer (Fig. 3.2b) in the asymmetric unit. The R_g of the dimer is only 30 Å, which suggests that TtP5CDH assembles into a higher order oligomer in solution. Furthermore, the SAXS curve calculated from the dimer deviates substantially from the experimental curves (Figs. 3.3a, 3.3c). The *H*3 crystal lattice was inspected to identify a higher order assembly that is consistent with the SAXS data. As described previously,[19] application of the crystallographic 3-fold rotation to the asymmetric unit generates a trimer-of-dimers hexamer (Fig. 3.2c). The hexamer has R_g of 42.6 Å, which agrees with the experimental R_g of 43 Å. The theoretical SAXS curve calculated from the hexamer exhibits good agreement with the experimental ones for both TtP5CDH (Fig. 3.3a) and DrP5CDH (Fig. 3.3c). Ensembles containing both the hexamer and dimer models were also considered using the minimal ensemble search method.[26] Slightly better fits to the experimental profiles are obtained with ensembles consisting of 94 % hexamer and 6% dimer for TtP5CDH (Fig. 3.3a), and 87 % hexamer and 13 % dimer for DrP5CDH (Fig. 3.3C). These calculations suggest that the hexamer is the predominant oligomer in solution. Moreover, the *ab initio* SAXS envelope matches the size and shape of the hexamer (Fig. 3.3b). It is concluded that TtP5CDH and DrP5CDH exist primarily as a trimer-of-dimers hexamer in solution. This is the first observation of a hexameric ALDH in solution.

The hexamer oligomeric state was confirmed using dynamic light scattering (DLS) (Table 3.1) and the SAXS volume of correlation (V_c).[27] DLS data indicate a hydrodynamic radius (R_H) of 7.2 nm for TtP5CDH, which implies a molecular weight (M) of 351 kDa. The latter value is within 3 % of the predicted M of 342 kDa for a hexamer. Similarly, the estimated R_H and M of DrP5CDH are 7.3 nm and 361 kDa. The V_c values of TtP5CDH and DrP5CDH are 1294 Å^2 and 1295 Å^2, respectively

(Table 3.1). The corresponding *M* values are 314 kDa for TtP5CDH and 316 kDa for DrP5CDH (Table 3.1), which are within 8 % of the *M* of the hexamer. The DLS and V_c results confirm that TtP5CDH and DrP5CDH exist primarily as hexamers in solution.

3.4.2 BhP5CDH and BlP5CDH are dimeric in solution

The SAXS profile of BhP5CDH is profoundly different from those of TtP5CDH and DrP5CDH (Fig. 3.4a). In particular, the pronounced valley at q = 0.083 Å^{-1} and peak at q = 0.102 Å^{-1}, which are characteristic of hexameric P5CDHs, are absent in the SAXS profile for BhP5CDH. The R_g of BhP5CDH from Guinier analysis is 31.3 ± 0.1 Å, while the real space R_g from calculations of P(r) is 31.8 – 31. 9 Å for D_{max} = 95 – 105 Å. These values are much smaller than the R_g of 43 Å of hexameric P5CDHs, suggesting that BhP5CDH does not form a hexamer. Furthermore, the P(r) has a maximum at r = 35 Å, whereas the P(r) for hexameric P5CDH has a maximum at 60 Å. The R_g calculated from a P5CDH dimer is 30 Å, implying that BhP5CDH forms the classic ALDH dimer but does not assemble into higher order oligomers. *M* estimated from SAXS V_c is 98 kDa (Table 3.1), which is suggestive of the dimer (114 kDa) and certainly eliminates tetramers (228 kDa) and hexamers (342 kDa) from consideration. Also, DLS data are consistent with BhP5CDH forming a dimer (R_H = 4.4 nm, *M* = 106 kDa).

The hypothesis that BhP5CDH forms a dimer in solution was tested using crystallographic data. The structure of BhP5CDH was determined by the New York Structural Genomics Research Consortium. The enzyme crystallizes in space group *C*2 with three molecules in the asymmetric unit (PDB code 3QAN, unpublished). Two of the molecules form the classic ALDH dimer, while the crystallographic 2-fold

rotation generates the other half of the dimer for the third molecule. Analysis of the crystal lattice with PDBePISA[28] indicates that the dimer is the most probable assembly in solution. Inspection of crystal packing with COOT confirmed that the hexamer is absent. The SAXS profile calculated from the BhP5CDH dimer agrees well with the experimental profile (Fig. 3.4a), and the SAXS reconstruction exhibits good agreement with the dimer (Fig. 3.4b). It is concluded that BhP5CDH is dimeric in solution.

The oligomeric state of BlP5CDH (75 % identical to BhP5CDH) was determined using DLS and analysis of crystal packing. The R_H is 4.6 nm, which corresponds to M = 119 kDa, consistent with a dimer. The crystal structure of BlP5CDH was also determined by the New York Structural Genomics Research Consortium. The enzyme crystallizes in space group $P2_1$ with eight molecules in the asymmetric unit. As deposited, dimers are not evident in the asymmetric unit. However, application of crystallographic symmetry allows a different choice of asymmetric unit that contains four dimers. Analysis of the BlP5CDH crystal lattice with PDBePISA indicates that the dimer is the most probable assembly in solution. It is concluded that BlP5CDH is also dimeric in solution.

3.4.3 Site-directed mutagenesis rationale

Site-directed mutagenesis (to Ala) was used to identify residues important for hexamerization. TtP5CDH was chosen as the model for this study because high resolution crystal structures are available (e.g., PDB 2BHQ). The major interface between dimers in the hexamer is formed by a helix from the NAD^+-binding domain (α3, residues 96-111) and the oligomerization domain (red in Fig. 3.2a). These

structural elements form a symmetric interface lining the inside surface of the tunnel that surrounds the 3-fold axis (red in Fig. 3.5a). This interface buries 1100 $Å^2$ of surface area. For comparison, the interfacial area of the domain-swapped dimer interface is 2900 $Å^2$. A smaller dimer-dimer contact surface (600 $Å^2$) is located on the outside of the hexamer (blue in Fig. 3.5b). The α14 helix of the catalytic domain (residues 454-463) and α3 are prominent in this interface (Fig. 3.2a).

The α3 helix was targeted for site-directed mutagenesis because it participates in both dimer-dimer interfaces (Fig. 3.5c). Specifically, three positively charged residues that form dimer-dimer electrostatic interactions were individually mutated to Ala: Arg100, Lys104, and Arg111. Arg100 is unique in that it is the only residue in the protein that interacts with two protomers outside of its own dimer (3.2 Å cutoff). These dimer-dimer interactions include ion pairs with Asp166, Glu168, and Glu458, and a hydrogen bond with Tyr154 (Fig. 3.5c). Arg100 also forms an intermolecular stacking interaction with Arg461. The other two residues targeted for mutagenesis, Lys104 and Arg111, form dimer-dimer ion pairs with the carboxyl-terminus of the polypeptide chain (Fig. 3. 5c).

Arg153 of the oligomerization domain was also mutated to Ala. This residue is interesting because it is next to the two-fold axis of the major dimer-dimer interface, and thus its guanidinium group stacks in parallel with that of the symmetry-related related Arg153 (Fig.3. 5c).

3.4.4 Steady-state kinetic measurements

The kinetic constants for the native and mutant enzymes were estimated using P5C as the variable substrate and NAD^+ fixed at 1 mM in order to assess whether

mutation of the hexamer interface causes any gross change in enzyme activity. The catalytic efficiencies (k_{cat}/K_m) of the TtP5CDH mutant enzymes are within 20 % of that of the native enzyme, indicating that these particular mutations do not have an obvious, substantial effect on activity.

3.4.5 Hexamerization hot spot

SAXS analysis of TtP5CDHR100A clearly shows that mutation of Arg100 to Ala is sufficient to disrupt the hexamer. The SAXS curve for TtP5CDHR100A lacks the valley and peak features that are diagnostic of the hexamer (Fig. 3.6a). The R_g from Guinier analysis is 32.1 ± 0.2 Å, which is close to the value of 30 Å calculated from the dimer and substantially smaller than that of the hexamer (43 Å). Furthermore, the SAXS curve calculated from the dimer exhibits excellent agreement with the experimental one (Fig. 3.6a), and the SAXS envelope matches the dimer (Fig. 6b). Also, the R_H estimated from DLS is 4.7 nm, which is substantially smaller than the R_H of 7.2 for the TtP5CDH hexamer. The corresponding *M* from DLS is 123 kDa, which is similar to value of 114 kDa expected for the dimer. *M* estimated from SAXS V_c is 100 kDa (Table 3.1), which is also consistent with the dimer. It is concluded that TtP5CDHR100A exists in solution primarily as a dimer.

The analogous mutation was generated for DrP5CDH (DrP5CDHR102A). The SAXS profile of DrP5CDHR102A is likewise indicative of a dimer (Fig. 3.6a). The R_g is 32.6 ± 0.1 Å, and the SAXS curve calculated from the dimer agrees well with the experimental curve. The SAXS envelope exhibits good agreement with the dimer (Fig. 3.6b). Also, the *M* of DrP5CDHR102A estimated from DLS is 126 kDa, which is similar to the value of 114 kDa expected for a dimer. *M* estimated from SAXS V_c is 111 kDa (Table 3.1), which is also consistent with the dimer. These data show that

DrP5CDHR102A likewise exists primarily as a dimer in solution.

In contrast, mutation of Arg111, Lys104, or Arg153 individually to Ala does not disrupt the hexamer. The SAXS curve for TtP5CDHR111A clearly exhibits the valley and peak features that are diagnostic of the hexamer (Fig. 3.6a). The R_g from Guinier analysis is 46.1 ± 0.2 Å, which is consistent with a hexamer. Furthermore, the profile calculated from the hexamer matches the experimental profile (Fig. 3.6a), and the shape reconstruction resembles the hexamer (Fig. 3.6b). Minimum ensemble search calculations also suggest that the hexamer is the predominant (88%) oligomer in solution (Fig. 3.6a). The M value estimated from SAXS V_c is 319 kDa (Table 3.1), which is just 7 % lower than M of the hexamer. TtP5CDHR111A, TtP5CDHK104A, and TtP5CDHR153A were analyzed with DLS. The hydrodynamic radii for these proteins span the range 6.8 – 7.2 nm, implying M = 304 - 337 kDa, which is consistent with the hexamer being the major species in solution. These results suggest that TtP5CDHR111A, TtP5CDHK104A, and TtP5CDHR153A exist primarily as hexamers in solution.

3.4.6 Crystallization of dimeric TtP5CDHs

TtP5CDHR100A was targeted for crystallization to better understand why mutation of Arg100 to Ala disrupts hexamerization. Surprisingly, the mutant enzyme crystallized in the same *H*3 lattice as TtP5CDH (Table 3.2), and thus the trimer-of-dimers hexamer is evident in crystalline TtP5CDHR100A despite loss of the numerous interactions formed by Arg100.

Electron density maps indicated that the Arg100Ala mutation causes a cascade of conformational changes in the vicinity of residue 100 (Fig. 3.7a). Glu168 rotates into

the void created by removal of the Arg100 side chain, while Tyr154 moves into the space vacated by Glu168 (Fig. 3.7b). Also, electron density for the side chains of Arg153 and Glu458 is weak and diffuse, implying disorder.

Formation of the *H*3 lattice by TtP5CDHR100A prompted the generation of double and triple mutants designed to induce a new crystal form that is devoid of the hexamer. This goal was achieved with the triple mutant R100A/K104A/R111A, which crystallizes in space group *P*1. A moderate resolution diffraction data set was obtained (Table 3.2), which was sufficient to determine the arrangement of dimers in the crystal lattice. The calculations show that the triclinic cell contains 4 dimers (Fig. 3.8). Analysis of protein-protein interfaces using manual inspection and PDBePISA shows that the hexamer is not present in the lattice.

3.5 DISCUSSION

Global sequence identity apparently is not a good predictor of the oligomeric state of P5CDHs. We previously showed using analytical ultracentrifugation that human P5CDH is a dimer in solution. The human enzyme has 30% sequence identity to the bacterial enzymes studied here. The *Bacillus* enzymes are also dimeric despite having 50% identity to the hexameric enzymes TtP5CDH and DrP5CDH. Thus, prediction of the oligomeric state of P5CDH from sequence requires a more careful examination of both sequence and structure. This conclusion is consistent with the hot spot theory of protein-protein interaction, which posits that a few critical residues in the interface account for most of the binding energy.[29,30]

We used structure-guided alanine scanning mutagenesis to identify a

hexamerization hot spot for bacterial P5CDHs. The hot spot is centered on Arg100, which is consistent with early work showing that protein-protein interaction hot spots tend to be enriched in arginine, surrounded by less important residues.[29] Arg100 is unique in that it is the only residue that interacts with two other protomers of the hexamer (Fig. 3.5c). This key residue forms dimer-dimer interactions with five residues: Tyr154, Asp166, Glu168, Glu458, and Arg461. Arg100 is also present in DrP5CDH but is substituted by Ala/Asn in the *Bacillus* enzymes and Gln in human P5CDH, consistent with it being essential for hexamerization. Arg461 is also found in DrP5CDH, while Asp166 is conservatively substituted with Glu. On the other hand, Arg461 and Asp166 are not conserved in the three dimeric P5CDHs. Also, a clear trend is not evident for Tyr154 or Glu168. This analysis suggests that the triplet of Arg100, Asp/Glu166, and Arg461 is a hexamerization hot spot, which could be used to predict the oligomeric states of P5CDHs. For example, *S. aureus* P5CDH, which has been proposed as a vaccine reagent, is predicted to be dimeric since it has Glu in place of Arg100.

The prediction of stable oligomeric species in solution from crystal structures is an important area of research. PDBePISA is used widely for this purpose and has been adopted by the PDB for predicting the biological assembly of deposited structures (remark 350 of PDB entries). TtP5CDHR100A is interesting in this regard because it forms a hexamer in the crystal yet is predominately dimeric in solution under the conditions used for SAXS and DLS (1-5 mg/mL). Analysis of the TtP5CDHR100A *H*3 lattice with PDBePISA indicates that the hexamer is the most probable assembly in solution, which contradicts the SAXS data. However, the PDBePISA results do suggest that the R100A hexamer is less stable than the native

one. For example, considering only the protein component and omitting solvent, the mutation causes 5 % and 7 % decreases in the surface area buried in the major and minor hexamer interfaces, respectively. Also, the complexation significance score of the major hexamer interface decreases from 1.0 for TtP5CDH (maximum possible value) to only 0.2 for TtP5CDHR100A. Thus, the program was able to discern a difference in the interfaces of the two crystal structures. Our results suggest that the prediction of the solution oligomeric state from crystal packing remains challenging in some cases.

Consideration of dynamic self-association equilibrium provides plausible explanations for the discrepancy in the oligomeric state of TtP5CDHR100A in the aqueous and solid states. The SAXS data show that the dimer-hexamer equilibrium of TtP5CDHR100A lies far on the dimer side under conditions of 1-10 mg/mL protein in Tris/NaCl buffers at pH 7.5. It is possible that the conditions used for crystallization shift the equilibrium to the hexamer, enabling crystal growth. However, DLS performed on TtP5CDHR100A under solution conditions similar to those of the initial crystallization drop yielded R_H of 4.3 nm and M of 101 kDa. Thus, the combination of low pH and high concentration of 2-methyl-2,4-pentanediol does not dramatically shift the equilibrium toward the hexamer. Another explanation is that the $H3$ crystal form may have sufficiently favorable free energy to allow crystal growth even at very low hexamer concentration. As hexamers join the growing crystal, the dimer-hexamer equilibrium shifts toward the hexamer according to Le Chatelier's principle, sustaining crystal growth.

The role of oligomerization in P5CDH function is unknown, but possibilities include cooperativity and substrate channeling. Positive cooperatively in cofactor

binding and half-of-the-sites reactivity are exhibited by some tetrameric ALDHs.[22,31] Analogous kinetic studies have not been performed for P5CDHs, but the observation of a higher order oligomer for TtP5CDH and DrP5CDH suggests that such experiments are worth pursuing. Oligomerization of P5CDH could also be connected with substrate channeling.[32,33] The potential for substrate channeling between monofunctional PRODH and P5CDH is suggested by the observation that the two enzymes are combined into a single polypeptide chain (PutA) in some bacteria. Eisenberg's group first proposed that such fused proteins could be used to predict protein-protein interactions (Rosetta Stone hypothesis of protein-protein interactions).[34] Kinetic studies have indeed shown that PutAs exhibit substrate channeling.[35,36] Thus, it is possible that monofunctional PRODH and P5CDH physically interact and engage in intermolecular substrate channeling. If so, it seems likely that the oligomeric states and quaternary structures of the interacting proteins will play an important role in dictating PRODH-P5CDH associations. This idea also remains to be tested.

Accession numbers

Coordinates and structure factor amplitudes for TtP5CDHR100A have been deposited in the PDB under the accession number 4K57.

Acknowledgements

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health via Grant GM065546. We thank Dr. Tommi White for purifying *T. thermophilus* genomic DNA and Dr.

Mingyi Zhou for cloning the TtP5CDH gene. We thank Prof. Steven Almo and the New York Structural Genomics Research Consortium for providing the BhP5CDH and BlP5CDH clones, and Prof. Donald Becker for providing P5C. We thank Kevin Dyer of the SIBYLS Mail-In SAXS Program for collecting the SAXS data. We thank Dr. Jay Nix for help with diffraction data collection at beamline 4.2.2. X-ray scattering and diffraction technologies and their applications to the determination of macromolecular shapes and conformations at the SIBYLS beamline at the Advanced Light Source, Lawrence Berkeley National Laboratory, are supported in part by the DOE program Integrated Diffraction Analysis Technologies (IDAT) under Contract Number DE-AC02-05CH11231 with the U.S. Department of Energy. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

3.6 REFERENCES

1. Phang, J. M. (1985). The regulatory functions of proline and pyrroline-5-carboxylic acid. *Curr. Top. Cell. Reg.* 25, 92-132.

2. Tanner, J. J. (2008). Structural biology of proline catabolism. *Amino Acids* 35, 719-30.

3. Singh, R. K. & Tanner, J. J. (2012). Unique Structural Features and Sequence Motifs of Proline Utilization A (PutA). *Front. Biosci.* 17, 556-568.

4. Efron, M. L. (1965). 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.* 272, 1243-54.

5. Baumgartner, M. R., Rabier, D., Nassogne, M. C., Dufier, J. L., Padovani, J. P., Kamoun, P., Valle, D. & Saudubray, J. M. (2005). 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.* 164, 31-6.

6. Geraghty, M. T., Vaughn, D., Nicholson, A. J., Lin, W. W., Jimenez-Sanchez, G., Obie, C., Flynn, M. P., Valle, D. & Hu, C. A. (1998). Mutations in the Delta1-pyrroline 5-carboxylate dehydrogenase gene cause type II hyperprolinemia. *Hum. Mol. Genet.* 7, 1411-5.

7. Scriver, C. R., Sly, W. S., Childs, B., Beaudet, A. L., Valle, D., Kinzler, K. W. & Vogelstein, B., Eds. (2001). The Metabolic and Molecular Bases of Inherited Disease. 8th edit. New York: McGraw-Hill.

8. Valle, D., Goodman, S. I., Applegarth, D. A., Shih, V. E. & Phang, J. M. (1976). Type II hyperprolinemia. Delta1-pyrroline-5-carboxylic acid dehydrogenase deficiency in cultured skin fibroblasts and circulating lymphocytes. *J. Clin. Invest.* 58, 598-603.

9. Wyse, A. T. & Netto, C. A. (2011). Behavioral and neurochemical effects of proline. *Metab. Brain Dis.* 26, 159-72.

10. Phang, J. M., Hu, C. A. & Valle, D. (2001). Disorders of proline and hydroxyproline metabolism. In *Metabolic and molecular basis of inherited disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D., eds.), pp. 1821-1838. McGraw Hill, New York.

11. Gogos, J. A., Santha, M., Takacs, Z., Beck, K. D., Luine, V., Lucas, L. R., Nadler, J. V. & Karayiorgou, M. (1999). The gene encoding proline dehydrogenase modulates sensorimotor gating in mice. *Nat. Genet.* 21, 434-9.

12. Felix, D. & Kunzle, H. (1976). The role of proline in nervous transmission. *Adv. Biochem. Psychopharmacol.* 15, 165-73.

13. Takemoto, Y. & Semba, R. (2006). Immunohistochemical evidence for the localization of neurons containing the putative transmitter L-proline in rat brain. *Brain Res.* 1073-1074, 311-5.

14. Fremeau, R. T., Jr., Caron, M. G. & Blakely, R. D. (1992). Molecular cloning and expression of a high affinity L-proline transporter expressed in putative glutamatergic pathways of rat brain. *Neuron* 8, 915-26.

15. He, F. & DiMario, P. J. (2011). 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* 11, 397-404.

16. Lee, I. R., Lui, E. Y., Chow, E. W., Arras, S. D., Morrow, C. A. & Fraser, J. A. (2013). Reactive Oxygen Species Homeostasis and Virulence of the Fungal Pathogen Cryptococcus neoformans Requires an Intact Proline Catabolism Pathway. *Genetics*, doi: 10.1534/genetics.113.150326.

17. Lijek, R. S., Luque, S. L., Liu, Q., Parker, D., Bae, T. & Weiser, J. N. (2012). Protection from the acquisition of Staphylococcus aureus nasal carriage by cross-reactive antibody to a pneumococcal dehydrogenase. *Proc. Natl. Acad. Sci. USA* 109, 13823-8.

18. Chhatwal, G. S. (2002). Anchorless adhesins and invasins of Gram-positive bacteria: a new class of virulence factors. *Trends Microbiol.* 10, 205-8.

19. Inagaki, E., Ohshima, N., Takahashi, H., Kuroishi, C., Yokoyama, S. & Tahirov, T.

H. (2006). Crystal structure of Thermus thermophilus Delta1-pyrroline-5-carboxylate dehydrogenase. *J. Mol. Biol.* 362, 490-501.

20. Rodriguez-Zavala, J. S. & Weiner, H. (2002). Structural aspects of aldehyde dehydrogenase that influence dimer-tetramer formation. *Biochemistry* 41, 8229-37.

21. Larson, H. N., Weiner, H. & Hurley, T. D. (2005). Disruption of the coenzyme binding site and dimer interface revealed in the crystal structure of mitochondrial aldehyde dehydrogenase "Asian" variant. *J. Biol. Chem.* 280, 30550-6.

22. Belem, Y. S., Pablo, P. J. & Jose, R. Z. (2013). New insights into the half-of-the-sites reactivity of human aldehyde dehydrogenase 1A1. *Proteins*.

23. Srivastava, D., Singh, R. K., Moxley, M. A., Henzl, M. T., Becker, D. F. & Tanner, J. J. (2012). The Three-Dimensional Structural Basis of Type II Hyperprolinemia *J. Mol. Biol.* 420, 176-189.

24. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* 40, 191-285.

25. Rambo, R. P. & Tainer, J. A. (2010). Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Curr. Opin. Struct. Biol.* 20, 128-37.

26. Pelikan, M., Hura, G. L. & Hammel, M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.* 28, 174-89.

27. Rambo, R. P. & Tainer, J. A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496, 477-81.

28. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774-97.

29. Bogan, A. A. & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* 280, 1-9.

30. Moreira, I. S., Fernandes, P. A. & Ramos, M. J. (2007). Hot spots--a review of the protein-protein interface determinant amino-acid residues. *Proteins* 68, 803-12.

31. Wei, B. & Weiner, H. (2001). Making an Oriental equivalent of the yeast cytosolic aldehyde dehydrogenase as well as making one with positive cooperativity in coenzyme binding by mutations of glutamate 492 and arginine 480. *Chemico-biological interactions* 130-132, 173-9.

32. Arentson, B. W., Sanyal, N. & Becker, D. F. (2012). Substrate channeling in proline metabolism. *Front. Biosci.* 17, 375-88.

33. Anderson, K. S. (1999). Fundamental mechanisms of substrate channeling. *Methods Enzymol.* 308, 111-45.

34. Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O. & Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science* 285, 751-3.

35. Srivastava, D., Schuermann, J. P., White, T. A., Krishnan, N., Sanyal, N., Hura, G. L., Tan, A., Henzl, M. T., Becker, D. F. & Tanner, J. J. (2010). Crystal structure of the bifunctional proline utilization A flavoenzyme from Bradyrhizobium japonicum. *Proc. Natl. Acad. Sci. USA* 107, 2878-83.

36. Surber, M. W. & Maloy, S. (1998). The PutA protein of Salmonella typhimurium catalyzes the two steps of proline degradation via a leaky channel. *Arch. Biochem. Biophys.* 354, 281-287.

37. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., 2nd, Tsutakawa, S. E., Jenney, F. E., Jr., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S. J., Scott, J. W., Dillard, B. D., Adams, M. W. & Tainer, J. A. (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* 6, 606-612.

38. Classen, S., Hura, G. L., Holton, J. M., Rambo, R. P., Rodic, I., McGuire, P. J., Dyer, K., Hammel, M., Meigs, G., Frankel, K. A. & Tainer, J. A. (2013). Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J. Appl. Crystallogr.* 46, 1-13.

39. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* 36, 1277-1282.

40. Schneidman-Duhovny, D., Hammel, M. & Sali, A. (2010). FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* 38, W540-4.

41. Svergun, D. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* 25, 495-503.

42. Kleywegt, G. J. (1997). Validation of protein models from Calpha coordinates alone. *J. Mol. Biol.* 273, 371-6.

43. Liu, H., Hexemer, A. & Zwart, P. H. (2012). The Small Angle Scattering ToolBox (SASTBX): an open-source software for biomolecular small-angle scattering. *J. Appl. Crystallogr.* 45, 587-593.

44. Inagaki, E., Takahashi, H., Kuroishi, C. & Tahirov, T. H. (2005). Crystallization and avoiding the problem of hemihedral twinning in crystals of Delta1-pyrroline-5-carboxylate dehydrogenase from Thermus thermophilus. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 61, 609-11.

45. Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125-32.

46. Evans, P. (2006). Scaling and assessment of data quality. *Acta Cryst.* D62, 72-82.

47. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Cryst.* D59, 1131-7.

48. Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C. & Terwilliger, T. C. (2004). Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Rad.* 11, 53-5.

49. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics.

Acta Cryst. D60, 2126-32.

50. Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Cryst.* D62, 48-57.

51. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658-674.

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

53. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* D66, 12-21.

Table 3-1 Parameters derived from SAXS and DLS

	Guinier RG (Å)	Real space RG (Å)	Vc (Å2)	SAXS M (kDa)[a]	DLS M (kDa)[b]	Oligomeric state
TtP5CDH	43	43	1294	314	351	Hexamer
TtP5CDHR100A	32	33	628	100	123	Dimer
TtP5CDHK104A	ND[c]	ND[c]	ND[c]	ND[c]	337	Hexamer
TtP5CDHR111A	46	44	1346	319	314	Hexamer
TtP5CDHR153A	ND[c]	ND[c]	ND[c]	ND[c]	304	Hexamer
TtP5CDHR100A/K104A/R111A	ND[c]	ND[c]	ND[c]	ND[c]	120	Dimer
DrP5CDH	43	44	1295	316	361	Hexamer
DrP5CDHR102A	33	33	666	111	126	Dimer
BhP5CDH	31	32	616	98	106	Dimer
BlP5CDH	ND[c]	ND[c]	ND[c]	ND[c]	119	Dimer

[a]The molecular weight estimated from the volume of correlation as M = Vc2RG-1/0.1231.
[b]The molecular weight estimated from dynamic light scattering.
[c]Not determined.

Table 3-2 X-ray diffraction data collection and refinement statistics

	TtP5CDH R100A	TtP5CDH R100A/K104A/R111A
Wavelength (Å)	1.000	1.542
Space group	H3	P1
Unit cell parameters (Å, °)	a = 102.7, c = 279.5	a = 65.1, b = 102.7 c = 160.6, α = 86.3, β = 87.5, γ = 79.4
Resolution (Å)	47.3 - 1.54 (1.62 - 1.54)	83.2-2.42 (2.55-2.42)
Total observations	835877	287864
Unique reflections	158507	114331
Multiplicity	5.3 (2.8)	2.5 (1.8)
Rmerge[b]	0.045 (0.461)	0.038 (0.110)
Rmeas[b]	0.050 (0.549)	0.047 (0.156)
Rpim[b]	0.021(0.291)	0.028 (0.110)
$\langle I/\sigma(I)\rangle$	20.7 (2.3)	15.4 (5.2)
Completeness (%)	96.8 (78.5)	73.5 (64.4)
Predicted oligomeric state[c]	hexamer	dimer
Rwork	0.168 (0.282)	
Rfree[d]	0.185 (0.291)	
Number of atoms	8608	
Protein residues	1032	
Water molecules	623	
RMSD bond lengths (Å)	0.006	
RMSD bond angles (°)	1.04	
Ramachandran plot[e]		
Favored (%)	98.64	
Outliers (%)	0	
MolProbity score (percentile)	99	
Average B-factors		
Protein (Å2)	22.5	
Water (Å2)	29.7	
Coordinate error (Å)[f]	0.13	
PDB code	4K57	

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]Definitions of Rmerge, Rmeas, and Rpim can be found in Weiss.[52]
[c]Oligomeric state predicted from crystal packing using PDBePISA.
[d]5% random test set.
[e]The Ramachandran plot was generated with MolProbity.[53]
[f]Maximum - likelihood based coordinate error from PHENIX.

FIGURE LEGENDS

Figure 3-1. The reactions of proline catabolism.

Figure 3-2. Structure of TtP5CDH (PDB code 2BHQ). (A) The protomer is shown with the catalytic domain in green, NAD^+-binding domain in light blue, and oligomerization domain in yellow. Red patches indicate residues that form the major dimer-dimer interface of hexametric P5CDHs. Dark blue denotes residues involved in the minor dimer-dimer interface of hexametric P5CDHs. (B) Structure of the classic ALDH dimer. (C) Two views of the TtP5CDH hexamer deduced from SAXS and X-ray crystallography.

Figure 3-3. SAXS analyses of TtP5CDH and DrP5CDH. (A) Experimental and calculated SAXS curves for TtP5CDH. The inset shows a Guinier plot spanning the range of qR_g from 0.489 to 1.28. The linear fit of the Guinier plot has R^2 of 0.996. (B) Superposition of the TtP5CDH SAXS shape reconstruction and the hexamer generated from crystallographic symmetry. The SAXS envelope was calculated using the SASTBX server. Two orthogonal views are shown. (C) Experimental and calculated SAXS curves for DrP5CDH. The inset shows a Guinier plot spanning the range of qR_g from 0.458 to 1.30 ($R^2 = 0.999$).

Figure 3-4. SAXS analysis of BhP5CDH. (A) Experimental and calculated SAXS curves. The inset shows a Guinier plot spanning the qR_g range of 0.352 – 1.30 ($R^2 = 0.998$). (B) Superposition of the SAXS shape reconstruction and the dimer.

Figure 3-5. Dimer-dimer interfaces within the P5CDH hexamer. (A and B) Orthogonal views of the hexamer with the A-B dimer colored cyan-green, C-D dimer in gray-yellow, and E-F dimer in brown-magenta. Red and blue denote residues in

the major and minor dimer-dimer interfaces, respectively. (C) Dimer-dimer interactions formed by the residues targeted for mutagenesis (Arg100, Lys104, Arg111, and Arg153). As in panels A and B, protomer A is colored cyan, and dimer C-D is colored gray-yellow. The oval denotes the approximate location of one of the 2-fold axes of the hexamer.

Figure 3-6. SAXS analysis of P5CDH mutant enzymes. (A) Experimental and theoretical SAXS curves. Arbitrary offsets were applied to the curves for clarity. The theoretical SAXS curves for TtP5CDHR100A (red) and DrP5CDHR102A (green) were calculated from a TtP5CDH dimer. The theoretical curve in cyan for TtP5CDHR111A was calculated from a TtP5CDH hexamer, while the curve in orange was calculated from an ensemble consisting of 88% hexamer and 12 % dimer. The inset shows Guinier plots spanning the qR_g ranges of 0.361-1.30 for TtP5CDHR100A (R^2 = 0.9992), 0.347-1.30 for DrP5CDHR102A (R^2 =0.9996), and 0.519-1.31 for TtP5CDHR111A (R^2 = 0.9997). (B) Superpositions of the SAXS shape reconstructions and oligomer models for TtP5CDHR100A, DrP5CDHR102A, and TtP5CDHR111A.

Figure 3-7. Hexamer interface of TtP5CDHR100A. (A) Electron density for residues of TtP5CDHR100A that change conformation. The cage represents a simulated annealing σ_A-weighted F_o - F_c omit map (3σ). The orientation and coloring scheme are identical to those of Fig. 5c. (B) Comparison of the hexamer interfaces of TtP5CDH (gray) and TtP5CDHR100A (pink).

Figure 3-8. Triclinic unit cell of TtP5CDHR100A/K104A/R111A. The red spheres denote the locations of the mutated residues, Ala100, Ala104, and Ala111.

Figure 3-1 The reactions of proline catabolism.

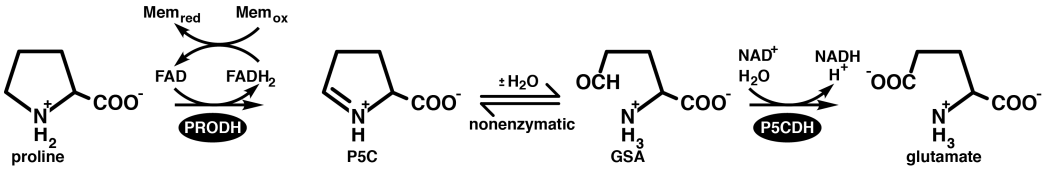


Figure 3-2 Structure of TtP5CDH (PDB code 2BHQ).

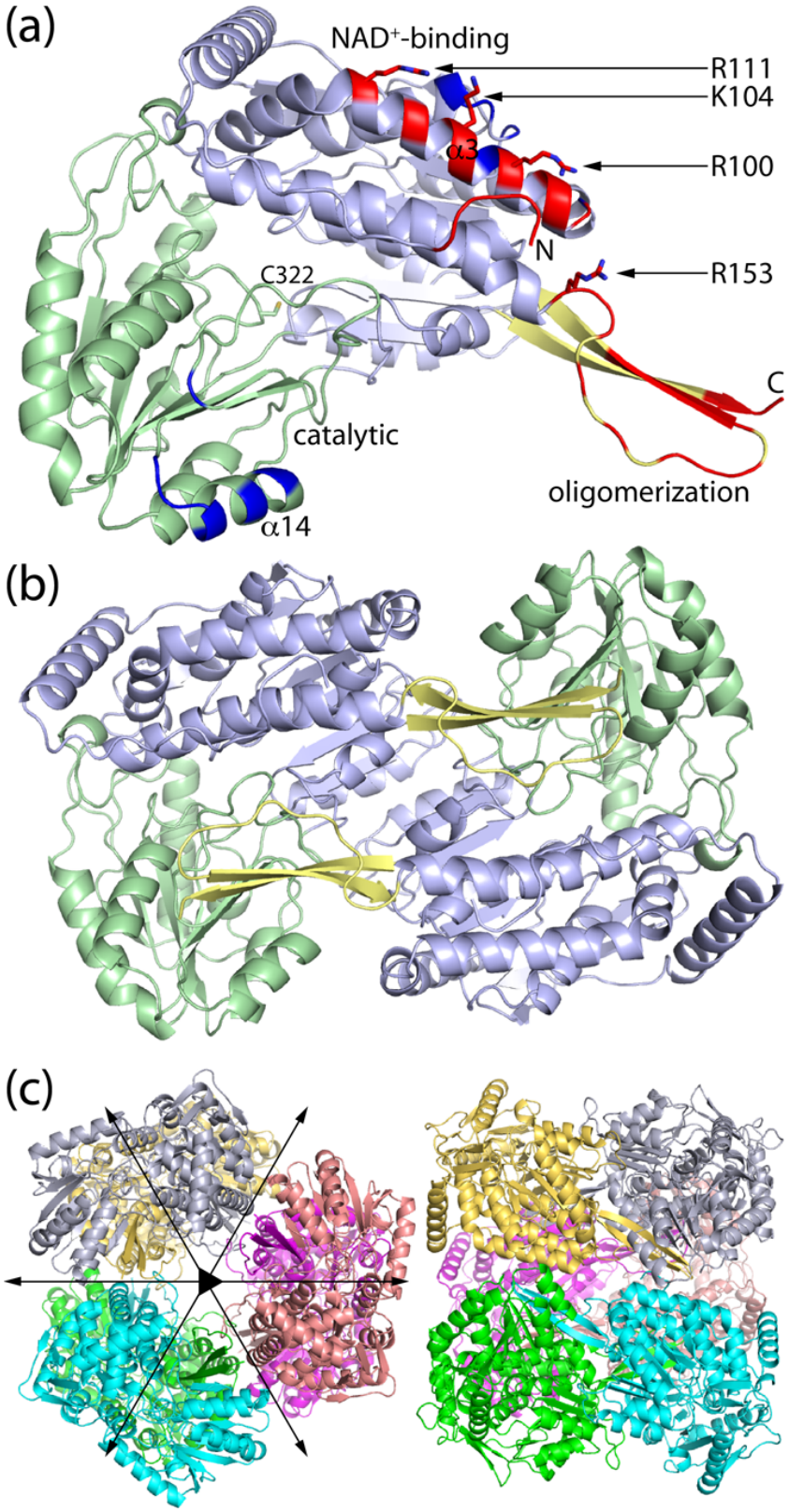


Figure 3-3 SAXS analyses of TtP5CDH and DrP5CDH.

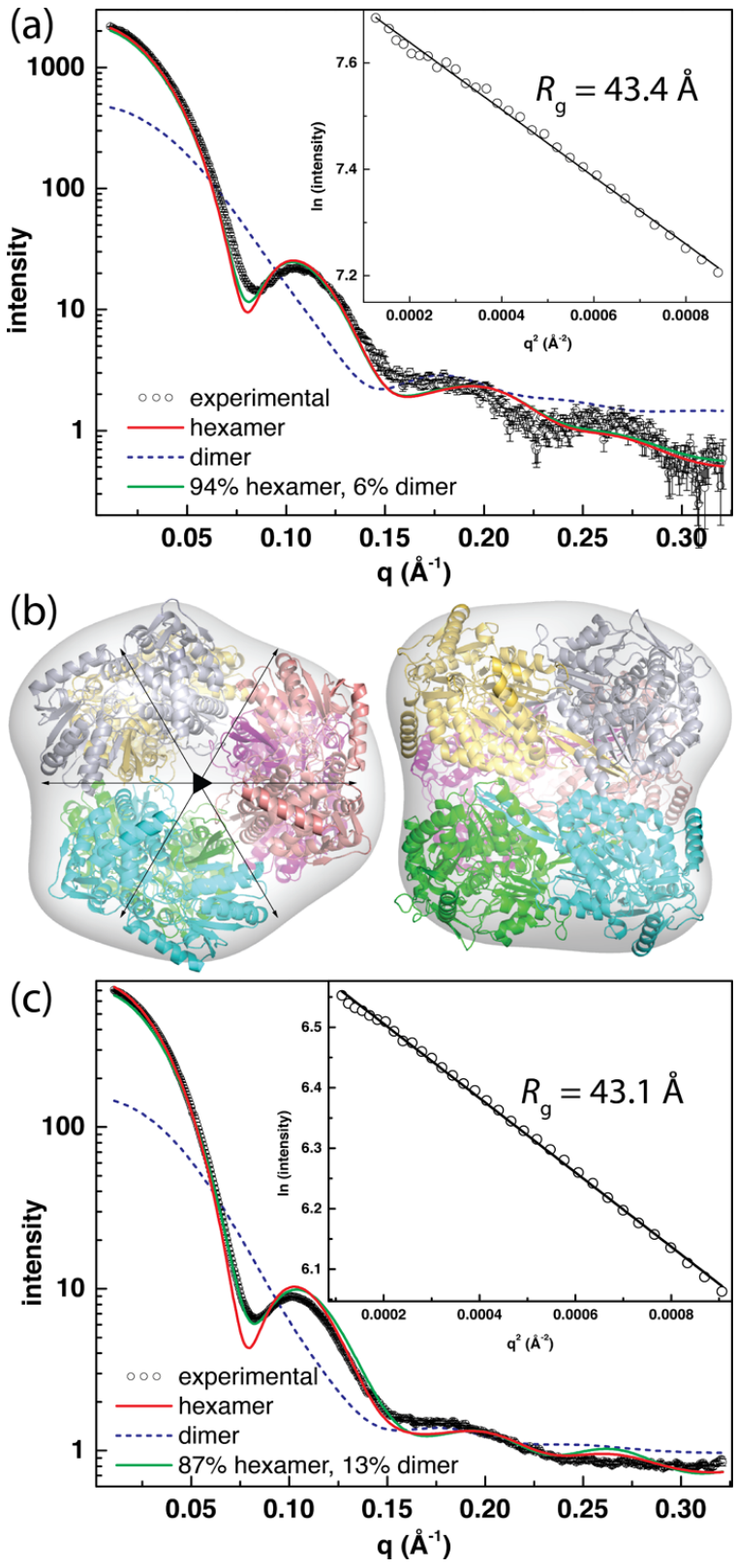


Figure 3-4 SAXS analysis of BhP5CDH.

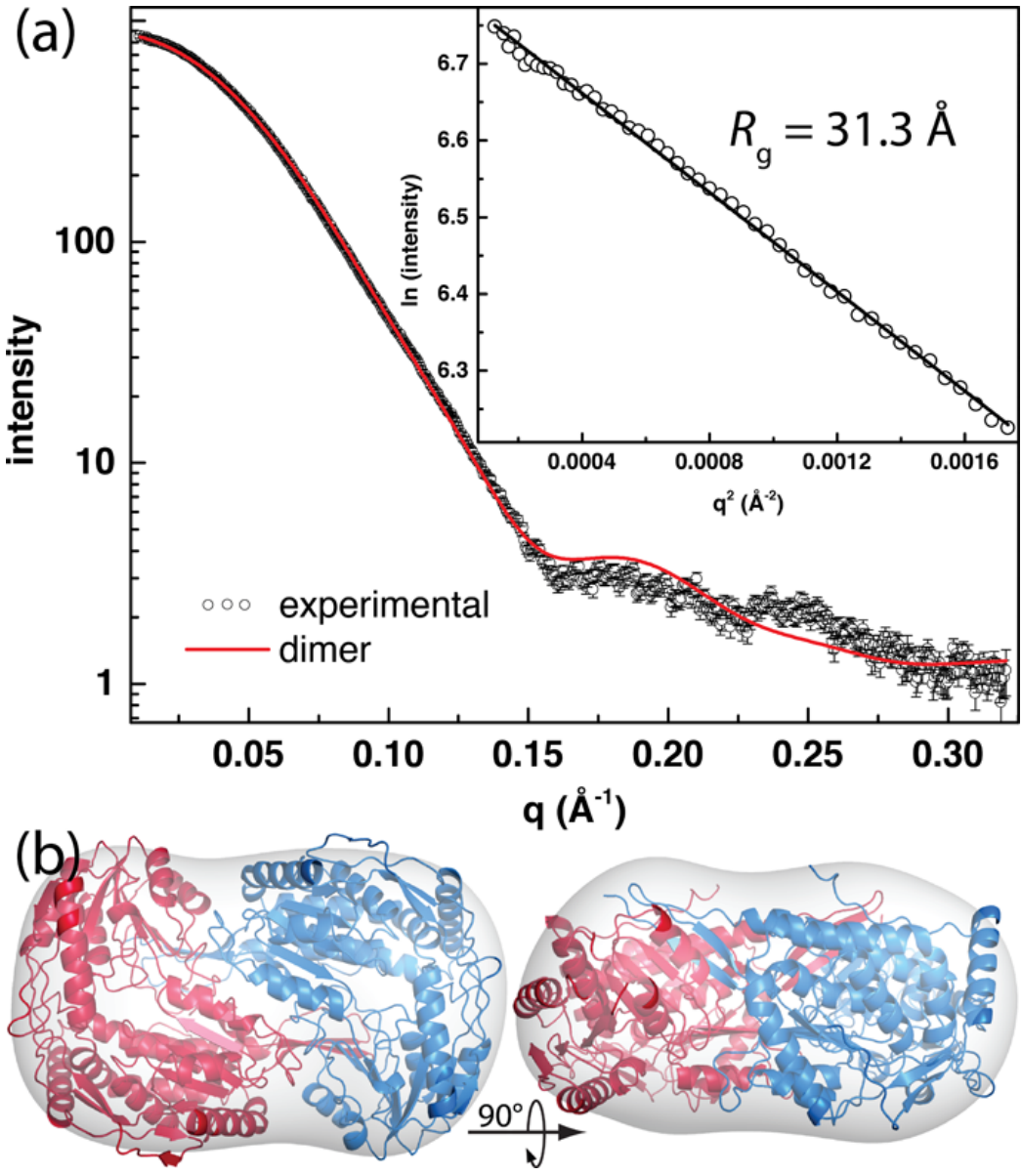


Figure 3-5 Dimer-dimer interfaces within the P5CDH hexamer.

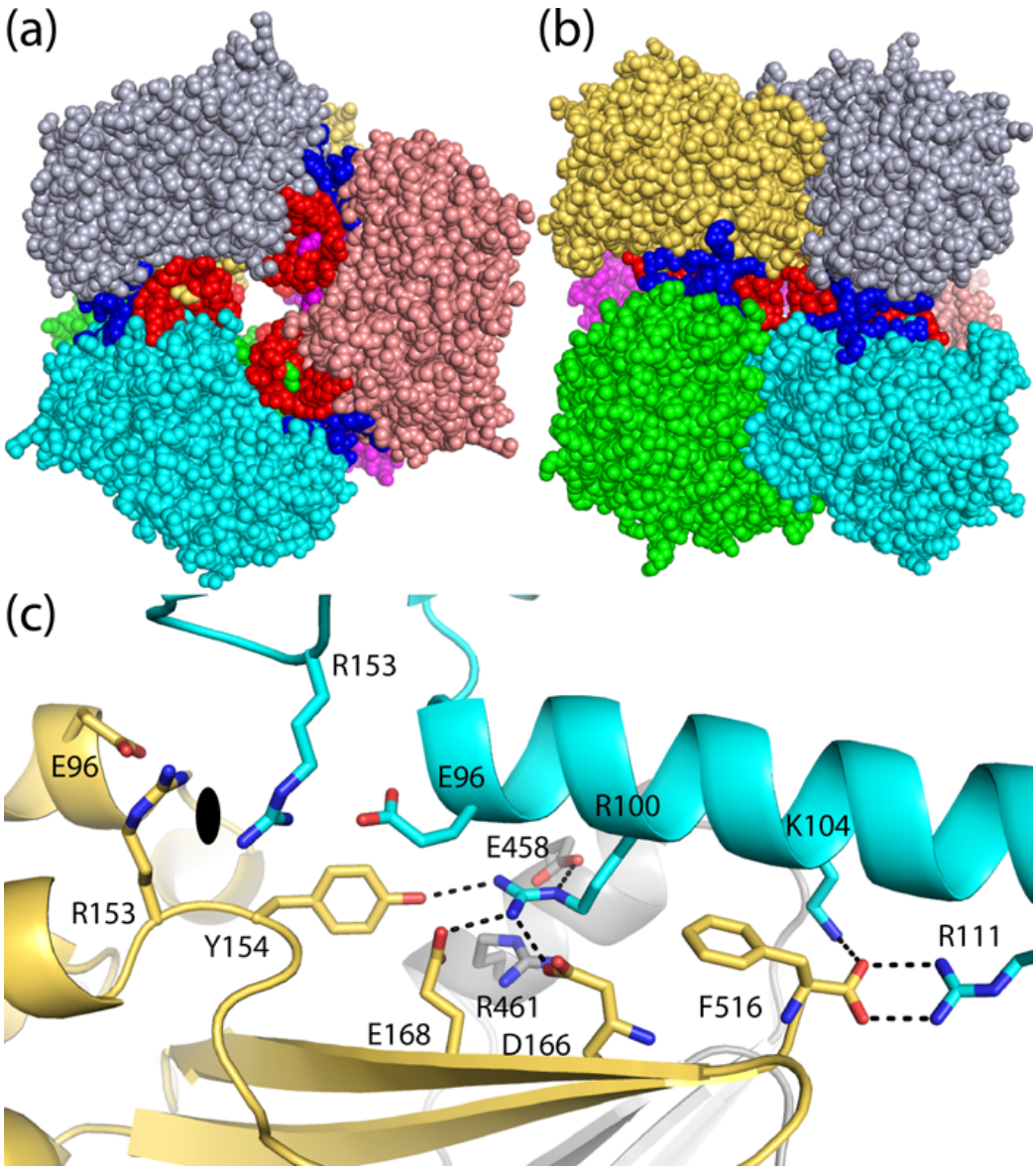


Figure 3-6 SAXS analysis of P5CDH mutant enzymes.

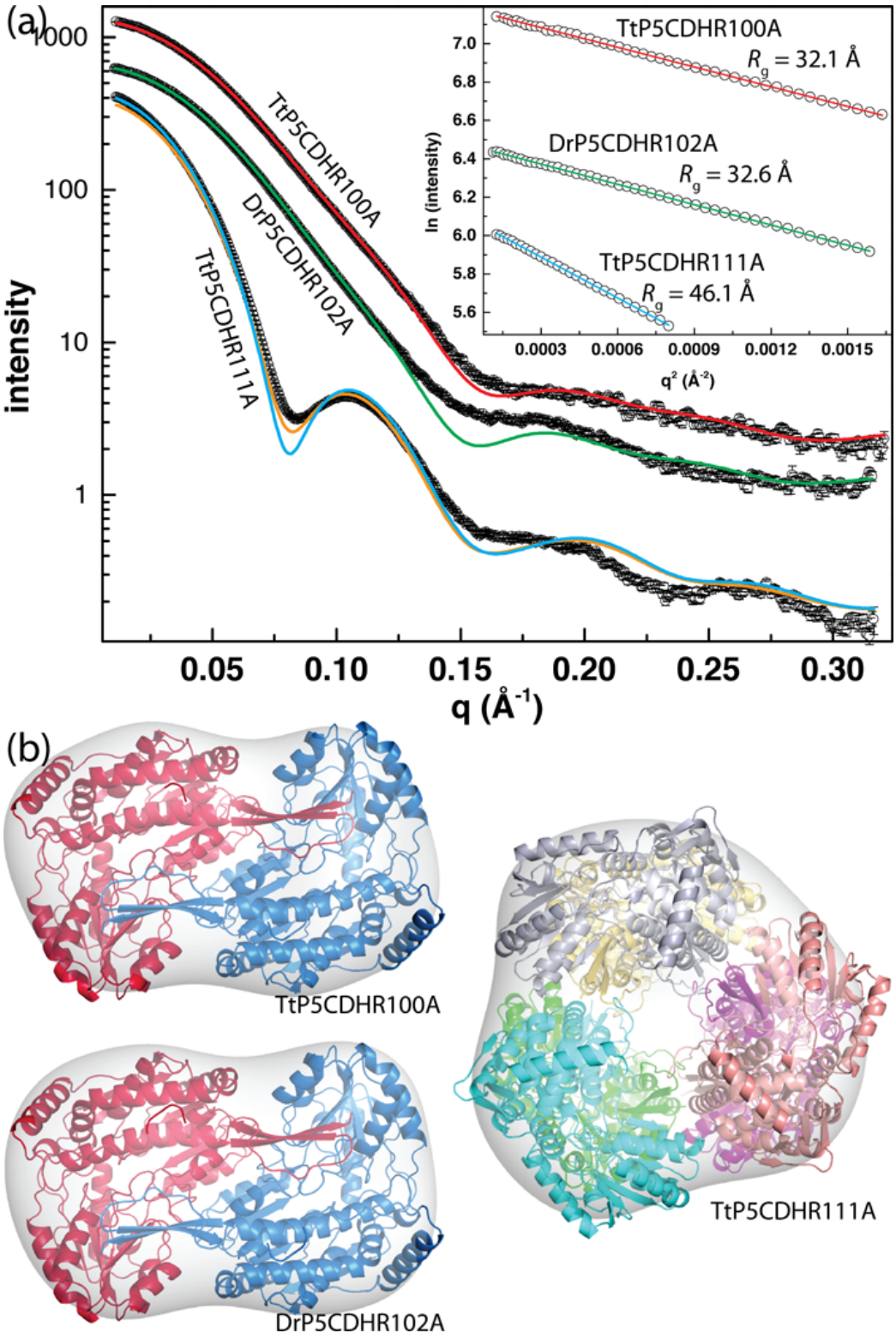


Figure 3-7 Hexamer interface of TtP5CDHR100A.

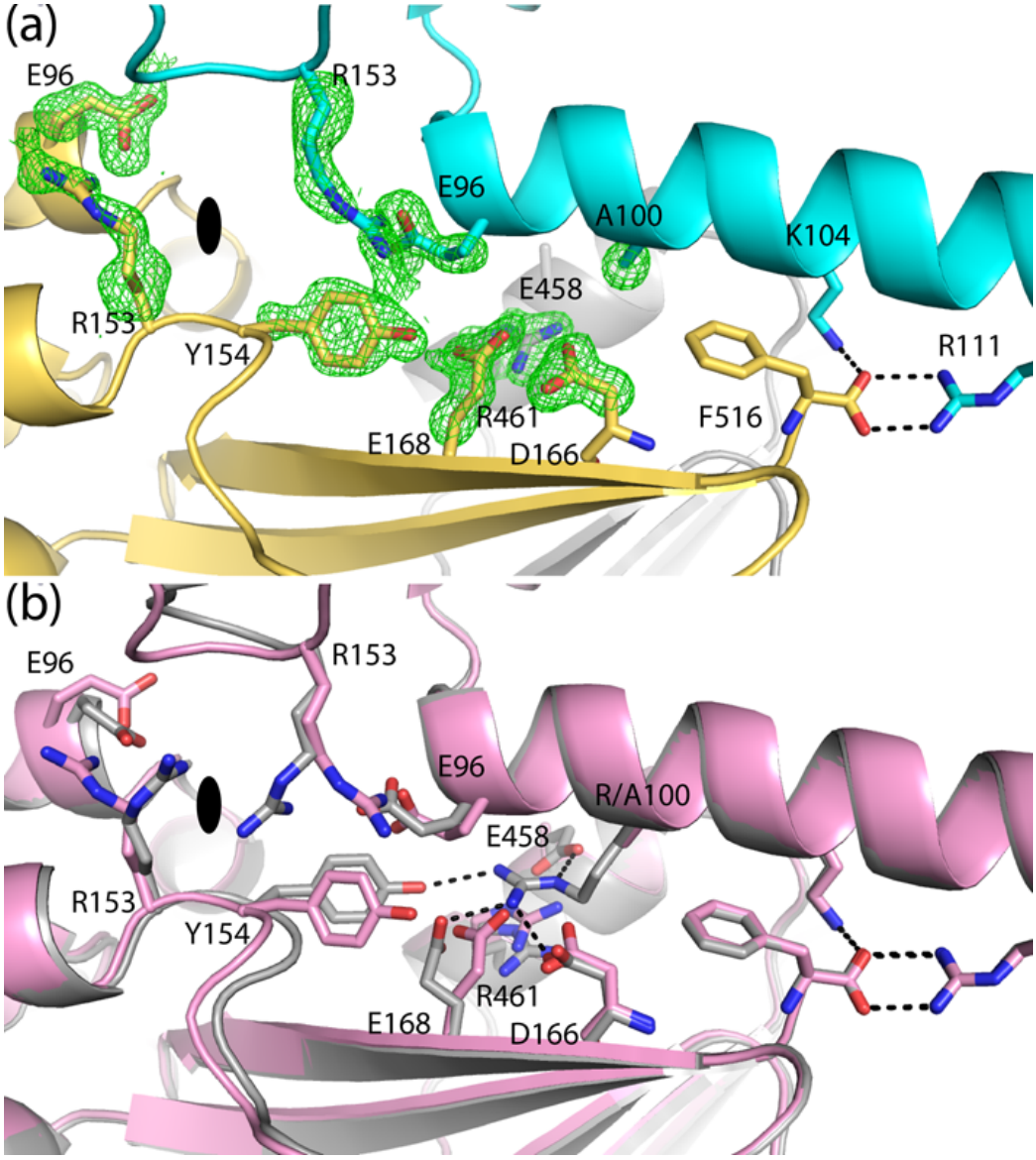
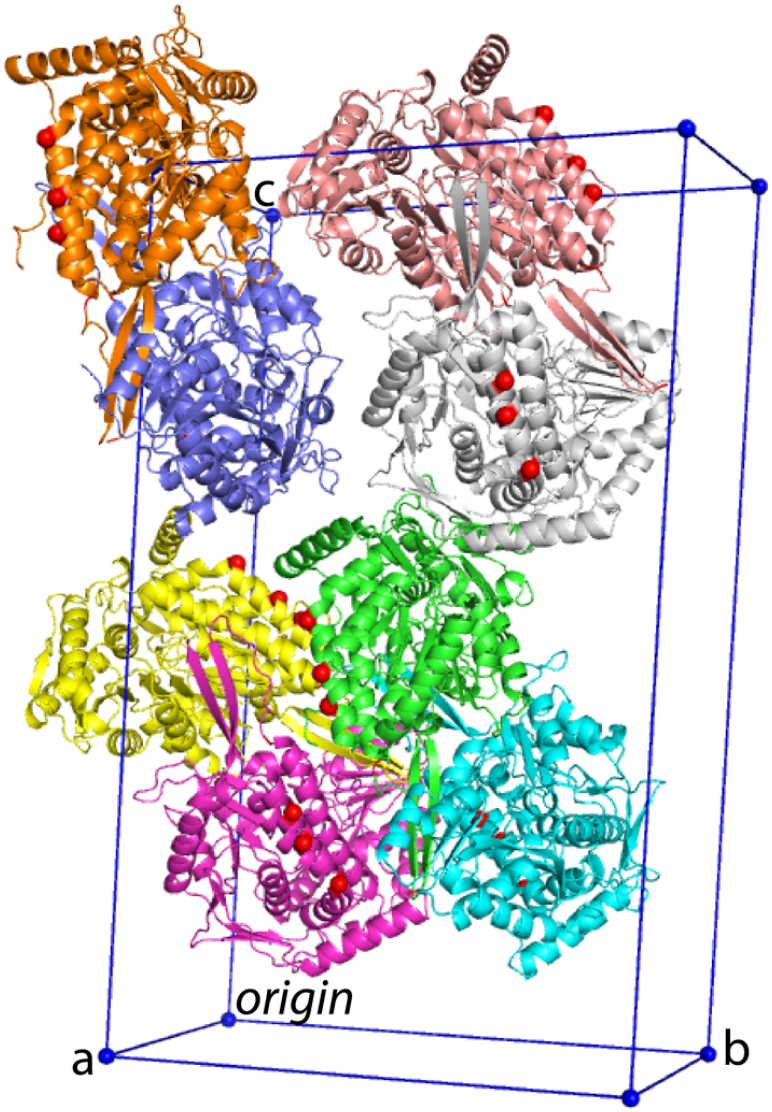


Figure 3-8 Triclinic unit cell of TtP5CDHR100A/K104A/R111A.



CHAPTER 4

4. EVIDENCE THAT THE C-TERMINAL DOMAIN OF A TYPE B PUTA PROTEIN CONTRIBUTES TO ALDEHYDE DEHYDROGENASE ACTIVITY AND SUBSTRATE CHANNELING

Luo M., et al. Evidence that the C-terminal Domain of a Type B PutA Protein Contributes to Aldehyde Dehydrogenase Activity and Substrate Channeling.Biochemistry (2014) 53(35):5661-73.

ABBREVIATIONS

PRODH, proline dehydrogenase; P5C, Δ^1-pyrroline-5-carboxylate; P5CDH, Δ^1-pyrroline-5-carboxylate dehydrogenase; FAD, flavin adenine dinucleotide; GSA, L-glutamate-γ-semialdehyde; PutA, proline utilization A; BjPutA, *Bradyrhizobium japonicum* proline utilization A; GsPutA, *Geobacter sulfurreducens* proline utilization A; RcPutA, *Rhodobacter capsulatus* proline utilization A; EcPutA, *Escherichia coli* proline utilization A; CTDUF, C-terminal domain of unknown function; ALDH, aldehyde dehydrogenase; RMSD, root mean square deviation; PDB, Protein Data Bank; SAXS, small-angle X-ray scattering; CoQ_1, ubiquinone-1; WT, wild-type; PCR, polymerase chain reaction; THP, Tris(3-hydroxypropyl)phosphine; TEVP, Tobacco Etch Virus protease; DCPIP, dichlorophenolindophenol; MALS, multi-angle light scattering; BjPutAG638P, site-directed mutant of BjPutA in which Gly638 is replaced with Pro; BjPutAE639A, site-directed mutant of BjPutA in which Glu639A is replaced with Ala; BjPutAN641A, site-directed mutant of BjPutA in which Asn641 is replaced with Ala; BjPutAΔ635-639, loop deletion mutant of BjPutA in which residues 635-639 are deleted; BjPutA1-986, C-terminal deletion mutant of BjPutA containing residues 1-986; RcPutAG1011P, site-directed mutant of RcPutA in which Gly1011 is replaced with Pro; RcPutAE1012A, site-directed mutant of RcPutA in which Glu1012A is replaced with Ala; RcPutAN1014A, site-directed mutant of RcPutA in which Asn1014 is replaced with Ala; RcPutAΔ1008-1012, loop deletion mutant of RcPutA in which residues 1008-1012 are deleted; RcPutA1-1116, C-terminal deletion mutant of RcPutA containing residues 1-1116.

4.1 ABSTRACT

Proline utilization A (PutA) is a bifunctional enzyme that catalyzes the oxidation of proline to glutamate. Structures of type A PutAs have revealed the catalytic core consisting of proline dehydrogenase (PRODH) and Δ^1-pyrroline-5-carboxylate dehydrogenase (P5CDH) modules connected by a substrate-channeling tunnel. Type B PutAs also have a C-terminal domain of unknown function (CTDUF) that is absent in type A PutAs. Small-angle X-ray scattering (SAXS), mutagenesis, and kinetics are used to determine the contributions of this domain to PutA structure and function. The 1127-residue *Rhodobacter capsulatus* PutA (RcPutA) is used as a representative CTDUF-containing, type B PutA. The reaction progress curve for the coupled PRODH-P5CDH activity of RcPutA does not exhibit a time lag, implying a substrate channeling mechanism. RcPutA is monomeric in solution, which is unprecedented for PutAs. SAXS rigid body modeling with target-decoy validation is used to build a model of RcPutA. Based on homology to aldehyde dehydrogenases (ALDHs), the CTDUF is predicted to consist of a β-hairpin fused to a non-catalytic Rossmann fold domain. The predicted tertiary structural interactions of the CTDUF resemble the quaternary structural interactions in the type A PutA dimer interface. The model is tested by mutagenesis of the dimerization hairpin of a type A PutA and the CTDUF hairpin of RcPutA. Similar functional phenotypes are observed in the two sets of variants, supporting the hypothesis that the CTDUF mimics the type A PutA dimer interface. These results suggest annotation of the CTDUF as an ALDH superfamily domain that facilitates P5CDH activity and substrate channeling by stabilizing the aldehyde-binding site and sealing the inter-active site tunnel from the bulk medium.

4.2 INTRODUCTION

Proline utilization A (PutA) proteins are bifunctional enzymes that catalyze the oxidation of L-proline to L-glutamate in Gram-negative bacteria (Figure 4.1A).[1-3] PutAs contain spatially separated proline dehydrogenase (PRODH) and Δ^1-pyrroline-5-carboxylate (P5C) dehydrogenase (P5CDH) active sites that catalyze this 4-electron oxidation process. The PRODH active site catalyzes the oxidation of L-proline to P5C with concomitant reduction of the enzyme-bound FAD. The P5CDH active site catalyzes the NAD^+-dependent oxidation of L-glutamate-γ-semialdehyde (GSA, the hydrolysis product of P5C) to L-glutamate. In addition to these enzymatic activities, some PutAs, known as trifunctional PutAs, also function as transcriptional regulators of the *put* regulon.

PutAs are large, multidomain proteins, and sequence analysis suggests three types of PutA based on domain architecture (Figure 4.1B).[1, 2] Type A PutAs contain just the catalytic core, which consists of N-terminal PRODH and C-terminal P5CDH modules separated by about 50 residues. The 999-residue PutA from *Bradyrhizobium japonicum* (BjPutA) is the best-characterized type A PutA.[4-6] Type B PutAs have an additional 100-200 residue C-terminal domain of unknown function (CTDUF) following the P5CDH module. The main subject of this study, PutA from *Rhodobacter capsulatus* (RcPutA), is a type B PutA. Type C PutAs have an N-terminal DNA-binding domain in addition to the catalytic core and CTDUF. These PutAs are transcriptional repressors of the *put* regulon and are thus trifunctional.[7-9] PutA from *Escherichia coli* PutA (EcPutA) is the archetypal trifunctional PutA.

The crystal structures of the type A PutAs BjPutA[5] and *Geobacter sulfurreducens* PutA (GsPutA)[10] revealed the structure of the PutA catalytic core (Figure 4.2). The PRODH active site is located in a distorted $(\beta\alpha)_8$ barrel. The P5CDH module adopts the classic aldehyde dehydrogenase (ALDH) superfamily fold, which consists of three domains: Rossmann NAD^+-binding, catalytic, and oligomerization. The two active sites are separated by a linear distance of 42 Å and connected by a buried tunnel that serves as a protected conduit for channeling the intermediate P5C/GSA from the PRODH site to the P5CDH site. The structures also revealed ancillary domains, denoted arm, α, and linker (Figure 4.2A), which help maintain the orientations of the two catalytic domains and form the walls of the substrate-channeling tunnel. Although BjPutA and GsPutA share only 27 % sequence identity, the two structures are very similar (2.0 Å root mean square deviation (RMSD)), implying high structural conservation of the catalytic core throughout the PutA family. Thus, the domain architecture, spatial disposition of the two active sites, and the intervening tunnel observed in BjPutA and GsPutA are likely present in all PutAs.

The DNA-binding domain of trifunctional PutA is also well characterized. It can be expressed as a ~50-residue fragment that retains DNA-binding activity,[7] which has enabled structural and biophysical analysis. X-ray crystal and solution NMR structures have shown that the DNA-binding domain adopts the ribbon-helix-helix fold and revealed the details of protein-DNA interactions.[7, 8, 11] Also, domain deletion analysis of EcPutA showed that the DNA-binding domain mediates dimerization of type C PutAs.[12]

In contrast to the catalytic core and DNA-binding domain, less is known about the CTDUF. Study of the CTDUF is hampered by the fact that it is not amenable to

soluble expression as an isolated fragment (D.F. Becker and J.J. Tanner, unpublished results). Nevertheless, a few clues about function have been obtained. Domain deletion analysis of EcPutA ruled out a role for the CTDUF in dimerization of type C PutAs.[12] The fact that type A PutAs exhibit both PRODH and P5CDH activities implies that the CTDUF is not directly involved in catalysis. Deletion of the CTDUF from EcPutA results in loss of P5CDH activity, which suggested that the CTDUF might be situated near the P5CDH active site.[12] By process of elimination, we speculated that the CTDUF could play a role in facilitating substrate channeling by providing tertiary structural interactions that help prevent escape of the intermediate P5C into the bulk medium.[1, 12]

Here, we use small-angle X-ray scattering (SAXS), steady-state kinetics, and site-directed mutagenesis to evaluate a homology model of the CTDUF and elucidate the contributions of this domain to PutA function. Our results suggest that the CTDUF is homologous to the dinucleotide binding domain of ALDHs and consists of a β-hairpin fused to a non-catalytic Rossmann fold domain. The β-hairpin facilitates P5CDH activity and substrate channeling by stabilizing the GSA anchor loop in the P5CDH active site and sealing the substrate-channeling tunnel from the bulk medium.

4.3 EXPERIMENTAL PROCEDURES

4.3.1 Materials.

P5C was synthesized by the method of Williams and Frank[13] and stored in 1 M HCl at 4 ^{o}C. On the day of experiments, the DL-P5C (50:50) was neutralized on ice by titrating with 6 M NaOH. All experiments were conducted in

Nanopure water.

4.3.2 Cloning of the RcPutA Gene.

The gene encoding full-length RcPutA was PCR-amplified from *R. capsulatus* SB 1003 genomic DNA (generously provided by Dr. Jason Cooley) using oligonucleotide primers with the restriction recognition sites NdeI and XhoI located at 5′ and 3′ ends. PCR was conducted using 50-µl volumes containing 5.0 µL 10X buffer, 2 µl of *R. capsulatus* genomic DNA, 2.5 µL of each PCR forward and reverse primer (1 µM), 1 µL 10 mM dNTPs, and 1 µL Pfu DNA polymerase (Agilent); distilled water was added to a final volume of 50 µL. PCR was run under the following cycling conditions: 95°C for 5 min, followed by 30 cycles of 95°C for 1 min, T_m at 56°C for 30 s, 72°C for 6 min, and a final extension of 10 min at 72°C. PCR products were detected by 1% agarose gel electrophoresis, and extracted using a gel extraction kit (Qiagen). The amplified full-length RcPutA sequence was digested with Nde1 and Xho1, and then cloned into pET-28a. The gene sequence was confirmed by sequencing at the DNA core of the University of Missouri-Columbia.

4.3.3 Site-Directed Mutagenesis.

Site-directed mutants and deletion constructs of RcPutA and BjPutA were generated using the QuikChange II site-directed mutagenesis kit (Agilent). The mutations were confirmed with DNA sequencing. The mutant enzymes were purified as described for the wild-type (WT) enzymes, except where noted.

4.3.4 Protein Expression and Purification.

BjPutA and BjPutA mutant enzymes were expressed and purified using immobilized metal affinity chromatography (Ni^{2+}-charged His-Trap HP, GE

Healthcare) as described previously.[14] The N-terminal His tag was removed as described.[14] The purified proteins were dialyzed into 50 mM Tris-HCl, 0.5 mM THP, 5% glycerol, 0.5 mM EDTA, at pH 8.0 and quick frozen in liquid nitrogen and stored at -80°C.

RcPutA and RcPutA mutants were expressed in *E. coli* BL21(DE3)pLysS and purified as follows. Starter cultures of 10 mL were grown in LB media overnight and used to inoculate 4 L of LB broth supplemented with 50 μg/mL kanamycin. After the culture reached an optical density of OD_{600} = 0.7, 0.3 mM isopropyl β-D-1-thiogalactopyranoside was added to induce protein expression for 12 hours at 18 °C. The cells were collected by centrifugation, resuspended in 50 mM Tris, 300 mM NaCl, 10 mM imidazole, 5% glycerol, at pH 7.5, and frozen at -80 °C. The frozen cells were thawed at 4 °C in the presence of protease inhibitors (0.1 mM tosyl phenylalanyl chloromethyl ketone, 0.05 mM 4-(2-Aminoethyl)benzenesulfonyl fluoride, 0.1 μM Pepstatin, 0.01 mM Leupeptin, 5μM E-64) and broken using sonication. The mixture was centrifuged at 16500 rpm in an SS34 rotor for 1 hour at 4^{o}C, filtered through a 0.45 μm filter, and loaded on a HisTrap HP column (5 mL) that had been charged with $NiCl_2$ and equilibrated in 50 mM Tris, 300 mM NaCl, 10 mM imidazole, and 5% glycerol at pH 7.5. Washing steps were performed using the loading buffer supplemented with 10 mM imidazole followed by 30 mM imidazole. The target protein was eluted with 200 mM imidazole and then dialyzed overnight in the dark at 4 °C into 50 mM Tris, 0.5 mM EDTA, 0.5 mM THP, and 5% glycerol at pH 7.8 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. The final purified proteins were supplemented with 0.05 mM FAD and NAD^+ and then dialyzed against 50 mM Tris-HCl, 0.5 mM THP, 5% glycerol, 0.5 mM EDTA, pH 8.0 at 4 °C overnight. This procedure was modified slightly for RcPutA mutant RcPutAE1012A. Because of poor expression and low protein yields for RcPutAE1012A, 12 L of LB broth was used for expression, 0.2% Triton-X-100 was included in the sonication buffer, and the ion exchange step was omitted. The concentrations of the PutA proteins were determined from the amount of FAD bound to normalize for differences in FAD content.

4.3.5 Steady-State Assays.

All steady-state assays were performed at 23°C. Two PRODH assays were used. First, the PRODH kinetic constants for wild-type RcPutA were determined using proline and CoQ_1 as the substrates as described previously.[15] Second, the PRODH activities of wild-type RcPutA, RcPutA variants, and BjPutA variants were measured using dichlorophenolindophenol (DCPIP) as the terminal electron acceptor and phenazine methosulfate as the mediator (proline/DCPIP oxidoreductase assay) as previously described.[16] K_m and k_{cat} for proline were determined using PutA (0.085-0.09 μM) and varying proline (0-300 mM) while fixing DCPIP at 75 μM. The above assays were conducted in 20 mM Tris buffer (pH 8.0, 10% glycerol). P5CDH activity was measured in 50 mM potassium phosphate (pH 7.5, 600 mM NaCl) as previously described[17] using PutA enzyme (0.17-0.19 μM) and varying DL-P5C (0 - 5.5 mM) with NAD^+ fixed at 200 μM. Reaction progress was monitored by following NADH formation at 340 nm ($\varepsilon_{340\ nm} = 6.2\ mM^{-1}\ cm^{-1}$). Data were collected using a 0.15 cm path length on a Hi-Tech Scientific SF-61DX2 stopped-flow instrument. Assays were performed in triplicate and values for K_m and k_{cat} were estimated by fitting initial

velocities to the Michaelis-Menten equation (SigmaPlot 12.0).

The PRODH-P5CDH coupled activity, in which proline is converted to glutamate, was measured as described previously.[5] Briefly, PutA enzyme (0.17-0.19 μM) was mixed with 200 μM CoQ_1, 40 mM proline, and 200 μM NAD^+ in 50 mM potassium phosphate (pH 7.5, 600 mM NaCl). The progress of the reaction was followed by NAD^+ reduction at 340 nm ($\varepsilon = 6200\ M^{-1}\ cm^{-1}$). The transient time was estimated by fitting the linear portion of the product concentration progress curve to a line and extrapolating to the x-axis.

4.3.6 Homology Modeling.

Homology models of RcPutA domains were used in SAXS rigid body modeling. The models were obtained from the following servers using default settings: the SWISS-MODEL Workspace server,[18] MODELLER[19] via the HHPred server[20] of MPI Bioinformatics Toolkit,[21] the Phyre2 server,[22] and the I-TASSER server.[23]

All four servers identified BjPutA (PDB code 3HAZ) as the best template for modeling the catalytic core (RcPutA residues 1-972). The sequence identity between BjPutA and RcPutA in this region is 52 %. The four models are very similar; the pairwise RMSDs for Cα atoms span the range of 0.5 - 0.7 Å.

The four servers were also used to calculate models of the CTDUF (residues 994-1097). All four servers identified BjPutA residues 622-756 as the template (27% identity). This region of BjPutA corresponds to the oligomerization hairpin and Rossmann dinucleotide-binding domain (Figure 4.2A). The pairwise RMSDs of the four models span the range 0.7 - 2.3 Å.

A model of the conserved C-terminal motif of RcPutA (residues 1108-1119) was built using MODELLER. The other servers did not produce models because of the short sequence length. The template structure was BjPutA residues 978-989. This region of BjPutA corresponds to a β-strand followed by a turn of α-helix (Figure 4.2A). The sequence identity of the modeled region is 5/12 residues.

4.3.7 Small-Angle X-ray Scattering (SAXS).

Prior to SAXS data collection, purified RcPutA was subjected to size exclusion chromatography using a Superdex 200 column. The column buffer was 50 mM Tris, 5% glycerol, 0.5 mM THP, and 50 mM NaCl at pH 7.8. The fractions were pooled, concentrated to ~8.5 mg/mL, and dialyzed at 4 °C for 24 hours against 50 mM Tris, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM THP, and 5% glycerol at pH 7.8. The dialysate was reserved for use as the SAXS reference.

SAXS experiments were performed at SIBYLS beamline 12.3.1 of the Advanced Light Source through the mail-in program.[24, 25] For each sample, scattering intensities were measured at three nominal protein concentrations. Data were collected for each protein concentration at exposure times of 0.5 s, 1.0 s, 3.0 s, and 6.0 s. The scattering curves collected from the protein samples were corrected for background scattering using intensity data collected from the reference buffer.

The SAXS data were analyzed as follows. Composite scattering curves were generated with PRIMUS[26] by scaling and merging the background-corrected low q region data from the 0.5 s or 1.0 s exposure with the high q region data from the 3.0 s exposure. PRIMUS was also used to perform Guinier analysis. GNOM was used to calculate pair distribution functions.[27] GASBOR[28] was used to calculate shape

reconstructions. Fifty independent models were generated with GASBOR using a maximum particle dimension (D_{max}) of 107 Å and no symmetry constraints. DAMAVER[29] was used to average and filter the GASBOR models. The Situs module pdb2vol was used to convert the averaged, filtered models into volumetric maps.[30] SUPCOMB was used to superimpose dummy atom models.[31] CRYSOL was used to calculate theoretical SAXS curves from atomic models.[32] The molecular mass in solution was determined from SAXS data using the volume of correlation invariant[33] as implemented previously[34] and Porod-Debye analysis.[35]

4.3.8 SAXS Rigid Body Modeling.

The program CORAL (COmplexes with RAndom Loops) of the ATSAS package[36] was used to determine the structural relationship between the catalytic core and the CTDUF of RcPutA. The default settings of CORAL were used for all calculations. Three sets of rigid body calculations, denoted as CORAL set 1, CORAL set 2, and CORAL decoy set, were performed as follows.

For CORAL set 1, the catalytic core model (residues 1-972) was held fixed, residues 973-993 were modeled as string of dummy residues, and the CTDUF model (residues 994-1097) was treated as a movable, rigid body. Each of the four models of the catalytic core was combined with each of the four models of the CTDUF for a total of 16 pairs. For every pair, ten independent simulated annealing optimization calculations were performed, each starting from different random number seed. Thus, a total of 160 CORAL poses were generated.

A second set of calculations (CORAL set 2) was performed in which the model of the conserved C-terminal motif (residues 1108-1119) was combined with the CTDUF

models using structural similarity to the oligomerization flap of BjPutA. The resulting composite model was considered to be a single, moveable rigid body in these calculations. Four such composite models were made by adding the conserved C-terminal motif model to each of the four CTDUF models. These four composite models were paired with the four catalytic core models, and thirty CORAL calculations were performed for each pair to generate a total of 480 poses. Two different starting configurations were used for these calculations to ensure that the initial arrangement of domains did not bias the results.

The CORAL decoy set was generated to validate the results of rigid body modeling. These calculations were performed using decoy structures in place of the CTDUF model. Four decoy structures were used: profilin IB (PDB code 1ACF, 125 residues), ketosteroid isomerase (3SED, 126 residues), a VH domain (1T2J, 116 residues), and human bromodomain (3HMF, 118 residues). Each decoy domain was paired with each of the four models of the RcPutA catalytic core, and twenty CORAL calculations were performed for each pair for a total of 320 decoy poses.

4.3.9 Multi-Angle Light Scattering.

The molecular mass of RcPutA in solution was estimated using a multi-angle light scattering (MALS) detector (Wyatt Technology) coupled to a G5000PWXL size exclusion chromatography column (Tosoh Bioscience, Montgomeryville, PA). The column buffer was 50 mM Tris-HCl, 50 mM NaCl, 0.5 mM EDTA, 0.5 mM THP, and 5% glycerol at pH 8.0. The flow rate was 0.75 ml/min. The data were analyzed using ASTRA software (Wyatt Technology).

4.4 RESULTS

4.4.1 Kinetic Characterization of RcPutA.

The RcPutA gene was cloned from genomic DNA to create a recombinant expression system. Recombinant RcPutA purifies as soluble, yellow protein. Analysis of the purified protein with SDS-PAGE is consistent with the predicted molecular mass of 117 kDa. The UV-visible spectrum of RcPutA exhibits maxima at 380 nm and 451 nm, consistent with a flavoprotein. The flavin content is estimated to be 97%.

The PRODH steady-state kinetic parameters of RcPutA were determined using proline and CoQ_1 as substrates. With proline as the variable substrate at fixed CoQ_1 concentration of 300 μM, K_m is 5.6 ± 0.8 mM and k_{cat} is 1.0 ± 0.1 s^{-1}, which corresponds to a catalytic efficiency of 180 ± 30 M^{-1} s^{-1}, a value that is similar to that previously reported for EcPutA (k_{cat}/K_m = 124 M^{-1} s^{-1}).[17] With CoQ_1 as the variable substrate at fixed proline concentration of 200 mM, the kinetic parameters are K_m = 94 ± 19 μM, k_{cat} = 2.0 ± 0.1 s^{-1}, and k_{cat}/K_m = 21000 ± 4000 M^{-1} s^{-1}.

The P5CDH activity of RcPutA using P5C as the variable substrate is characterized by K_m = 1530 ± 160 μM and k_{cat} = 7.3 ± 0.6 s^{-1}. The efficiency of RcPutA (k_{cat}/K_m = 4800 ± 1000 M^{-1} s^{-1}) appears to be nearly two-fold higher than that of EcPutA (2580 M^{-1} s^{-1}).[17]

The PRODH-P5CDH coupled activity of RcPutA was analyzed by monitoring NADH production from a reaction mixture containing proline, CoQ_1, and NAD^+. The PRODH-P5CDH coupled assay for wild-type RcPutA shows no apparent lag phase in the reaction progress curve (Figure 4.3A), suggesting that the intermediate L-P5C/GSA is channeled between the PRODH and P5CDH active sites.[5] For a

nonchanneling control, an equimolar mixture of the monofunctional RcPutA variants R454M and C791A was assayed. The RcPutA mutants R454M and C791A are devoid of PRODH and P5CDH activity, respectively. As described previously for other PutAs, a mixture containing both PutA monofunctional variants is able to oxidize proline to glutamate, but channeling is not possible.[5, 17] The reaction progress curve with the mixed variants shows a considerable lag time of ~ 8 min before attaining steady-state formation of NADH (Figure 4.3A). This result with RcPutA is a diagnostic feature of substrate channeling in PutAs and is consistent with results reported previously for EcPutA[17] and BjPutA,[5] and repeated here for BjPutA (Figure 4.3B).

4.4.2 SAXS Analysis and Oligomeric State of RcPutA.

The solution structural properties of RcPutA were studied using SAXS (Figure 4.4). Guinier analysis of data collected at three protein concentrations yields a radius of gyration (R_g) of 32.3 ± 0.4 Å. Calculations of the pair distribution function (P(r)) suggest R_g of 32.4 – 32.8 Å and maximum particle dimension (D_{max}) of 100 - 110 Å. The P(r) function exhibits a single maximum (r = 37 Å), which is indicative of a particle having a single lobe (Figure 4.4B).

The SAXS data suggest that RcPutA is compact and folded. The Kratky plot exhibits the classic inverted parabola shape that is characteristic of a compact protein that lacks unfolded domains (Figure 4.4C). Likewise, the Porod-Debye plot has a distinct plateau, indicating a sharp boundary between the scattering particle and the solvent, which is a signature of a well-folded protein (Figure 4.4C, inset). These results suggest that the domains of PutA are folded and in close contact with each

other.

The molecular mass (*M*) of RcPutA in solution was estimated from the SAXS data using the volume of correlation invariant.[33] The volume of correlation of RcPutA is 670 Å^2, which corresponds to *M* of 108 kDa. This value is within 8 % of the theoretical *M* of an RcPutA monomer of 117 kDa, implying that RcPutA is monomeric in solution.

Porod-Debye analysis[35] was used to confirm the oligomeric state of RcPutA. The average excluded particle volume determined from 50 independent shape reconstruction calculations performed using DAMMIF[37] is 195600 Å^3. This value agrees well with the volume of 188931 Å^3 from DATPOROD.[36] The density of a monomeric RcPutA protein having a volume of 189000-196000 would be 1.0 g/mL, which is within the expected range for proteins of 0.8 - 1.6 g/mL.[35] In contrast, the assumption of higher order oligomers of RcPutA having volume of 189000-196000 Å^3 implies densities ≥ 2.0 g/mL, which is well outside of the expected range. This analysis is also consistent with RcPutA forming a monomer in solution.

The oligomeric state of RcPutA was confirmed with MALS (Figure 4.5). The data suggest that the purified protein is monodisperse with an apparent *M* of 124 ± 2 kDa. This value is within 6 % of the predicted *M* of an RcPutA monomer. Thus, SAXS and MALS indicate that RcPutA is primarily monomeric in solution, which is unprecedented for PutAs.

Shape reconstruction calculations were performed with GASBOR assuming a monomer (Figure 4.6A). The normalized spatial discrepancy of the ensemble is 1.02 ± 0.2 and only two of the 50 models were discarded during averaging (2 σ cutoff),

which indicates a reliable reconstruction. RcPutA resembles an ellipsoid with dimensions of 50 x 70 x 100 Å. The shape exhibits a distinct curvature when viewed down the 70-Å axis, which results in a pronounced convex surface (Figure 4.6A).

Conservation of the β-hairpin in the RcPutA CTDUF. Remote homology detection predicts that the CTDUF is homologous to the N-terminal half of ALDHs.[1, 12] In particular, the CTDUF of RcPutA is 27 % identical to residues 622-756 of BjPutA (Figure 4.7). This region of BjPutA corresponds to the β-hairpin of the oligomerization domain and the Rossmann fold NAD^+-binding domain (Figure 4.2A).

Accordingly, homology models of the RcPutA CTDUF calculated from the BjPutA structure using online servers feature a β-hairpin followed by a 4-stranded parallel β-sheet flanked by three α-helices (Figure 4.7). Although the predicted β-hairpin is structurally homologous to the oligomerization β-hairpin of BjPutA and other ALDHs, it is not involved in dimerization of RcPutA, which is a monomeric protein. The predicted β-α-β motifs and strand order of the parallel β-sheet (2134) are diagnostic of the Rossmann dinucleotide binding fold.[38] The predicted Rossmann domain of the CTDUF is abbreviated in that it contains fewer than the standard 6 β-strands.

The predicted β-hairpin of RcPutA was tested by mutagenesis of conserved residues in the β-hairpin of BjPutA and RcPutA. Site-directed mutations were made to three conserved residues in the dimerization hairpin of BjPutA (BjPutAG638P, BjPutAE639A, BjPutAN641A) and the corresponding sites in the predicted hairpin of RcPutA (RcPutAG1011P, RcPutAE1012A, RcPutAN1014A) (Figure 4.2D). The β-hairpins were also more aggressively perturbed by deleting the conserved GPTGE

sequence at residues 635-639 of BjPutA (BjPutAΔ635-639) and 1008-1012 of RcPutA (RcPutAΔ1008-1012).

The steady-state kinetic parameters for the mutants of RcPutA and BjPutA are reported in Table 4.1. In general, the PRODH activity was not substantially decreased in any of the mutants. The most significant decrease in PRODH activity was observed for mutants RcPutA RcPutAN1014A and BjPutAΔ635-639, which had ~ 3.5-fold lower k_{cat}/K_m relative to the wild-type enzyme. Interestingly, some of the RcPutA mutants showed increased PRODH activity with the RcPutA mutant RcPutAG1011P having a 5-fold increased k_{cat}/K_m.

In contrast to PRODH activity, the P5CDH activity was eliminated in all of the hairpin mutants. These results are consistent with the hairpin being close to the P5CDH active site. In particular, the crystal structure of BjPutA shows that the hairpin forms electrostatic interactions that stabilize the loop that binds GSA in the P5CDH active site (Figure 2.2D). Disruption of these interactions in BjPutA apparently abrogates P5CDH activity. The fact that the same phenotype is seen in the analogous RcPutA variants suggests that the hairpin of RcPutA likewise helps stabilize the GSA binding loop, thus validating the homology model of the RcPutA CTDUF.

SAXS Rigid Body Modeling of RcPutA: CORAL Set 1. Theoretical SAXS curves were calculated from models of the catalytic core to provide a reference for evaluating SAXS rigid body modeling calculations. The catalytic core represents 85% of RcPutA and is 52 % identical in sequence to BjPutA. SAXS curves calculated from four homology models of the core obtained from different servers yield χ values

of 3.5 - 4.5. The calculated curve having the lowest χ is compared to the experimental curve in Figure 4.8A (orange curve). Note that the agreement is very good for $q < 0.08$ Å^{-1} but deviation is observed for $0.08 < q < 0.16$ Å^{-1} (Figure 4.8A inset).

SAXS-constrained rigid body modeling was used to generate three-dimensional models of RcPutA. These calculations were performed with the program CORAL using homology models of the RcPutA catalytic core (residues 1-972) and CTDUF (residues 994-1097, Figure 4.7). The catalytic core was held fixed, the CTDUF was allowed to move, and the intervening peptide (residues 973-993) was modeled as a string of dummy residues. In the starting configuration, the catalytic core and CTDUF were far apart, with residue 972 of the core and residue 994 of the CTDUF separated by 68 Å.

The results of these calculations are shown in Figure 4.9A as a scatter plot of the clash penalty versus χ for 160 independent poses. The clash penalty reports on steric interference between domains, while χ expresses the agreement between the experimental scattering curve and the theoretical curve calculated from the model. Low values of both parameters are desirable. The scatter plot shows a tight cluster of 78 poses centered at ($\chi = 1.9$, clash penalty = 1.1). All the poses in this cluster have $\chi < 2.15$ and penalty < 1.5, and represent 49 % of the 160 poses generated.

The poses in the cluster of 78 are structurally similar (Figure 4.9C). All 78 poses feature the CTDUF interacting mainly with the P5CDH part of the catalytic core. The β-hairpin of the CTDUF extends into the open region between the two active sites, and the abbreviated Rossmann fold of the CTDUF packs against the NAD^+-

binding domain of the catalytic core. This arrangement is highly reminiscent of the dimer interface of BjPutA (Figure 4.2C) but in the context of a monomeric protein.

The structural similarity between the CORAL poses and the BjPutA dimer interface was quantitated by calculating the RMSD between the β-hairpin of each CORAL pose and the β-hairpin as positioned in the BjPutA dimer interface. For this calculation, the BjPutA dimer was first superimposed onto RcPutA using the residues of the catalytic core, then the RMSD for the Cα atoms of the hairpin was calculated. The scatter plot of RMSD versus χ is shown in Figure 4.9B. The poses from the cluster of 78 centered at (χ = 1.9, clash penalty = 1.1) have relatively low RMSD compared to the other poses (Figure 4.9B), which is consistent with the idea that the CTDUF in RcPutA mimics the dimer interface of BjPutA.

SAXS Rigid Body Modeling: Decoy Set. CORAL calculations were also performed using decoy structures in place of the CTDUF model in order to identify the high confidence regions of target function space. Four decoy structures were used: profilin IB, ketosteroid isomerase, a VH domain, and human bromodomain. These structures are suitable decoys because they have approximately the same number of residues as the CTDUF model but are functionally unrelated to PutAs. The decoys exhibit diversity in protein fold and secondary structure content. The VH domain is primarily β-strand, whereas the bromodomain is mostly α-helical. Profilin IB and ketosteroid isomerase have mixed α/β folds. Each decoy domain was paired with each of the four models of the RcPutA catalytic core, and twenty CORAL calculations were performed for each pair for a total of 320 decoy poses.

The results of the decoy calculations are shown in Figure 4.9A. The χ distribution

of the decoy poses ranges from 2.3 to 5.9 with an average of 3.9 ± 0.8, while the clash penalty is 1.1 - 10.4 with an average of 4.2 ± 1.8. We suggest that the decoy results correspond to the low confidence region of parameter space where the false positive rate is high. Note that the aforementioned cluster of 78 poses from CORAL set 1 is well separated from the noise region defined by the decoy calculations, suggesting that these poses are meaningful (Figure 4.9A).

SAXS Rigid Body Modeling of RcPutA: CORAL Set 2. The results from CORAL set 1 motivated a second, more extensive set of calculations designed to test the idea that the CTDUF mimics the type A PutA dimer interface (CORAL set 2). The BjPutA dimerization flap consists of two elements: the β-hairpin and the C-terminal motif (Figure 4.2C). The latter element was incorporated into the SAXS modeling by first docking a model of the C-terminal motif to the CTDUF models using the BjPutA oligomerization domain as the superposition template, and then treating the resulting composite model as a rigid body in CORAL.

The CORAL set 2 calculations produced a high confidence cluster of 253 poses centered at (χ = 1.6, penalty = 1.2) that is clearly separated from the decoy region (Figure 4.10A). The average χ (1.6) and minimum χ (1.3) of this cluster are lower than the corresponding values of the high confidence cluster from CORAL set 1, which indicates that inclusion of the C-terminal motif improved the model. All but one of the 253 poses has RMSD < 10 Å, indicating that they mimic the BjPutA dimer interface (Figure 4.10B), a result that is confirmed by visual inspection (Figure 4.10C). Thus, both sets of rigid body calculations suggest that the CTDUF packs against the catalytic core in the same way that the oligomerization domain mediates

dimerization in type A PutAs.

Finally, the models in the high confidence clusters have good agreement with the experimental scattering profile and shape reconstruction. For example, Figure 4.8A shows the curves calculated from three poses from CORAL set 2 corresponding to the (1) best combination of low χ and low RMSD, (2) lowest RMSD regardless of χ, and (3) lowest χ regardless of RMSD. In all three cases, the fit to the experimental curve is noticeably better than that provided by the catalytic core alone. In particular, the rigid body models significantly improve the fit in the intermediate region of q = 0.08 - 0.16 $Å^{-1}$ (Figure 4.8A, inset). The rigid body models are also consistent with the shape reconstruction (Figure 4.6B).

Function of the C-terminal Motif in Substrate Channeling. The function of the conserved C-terminal motif was tested by engineering C-terminal truncations BjPutA1-986 and RcPutA1-1116. The C-terminal truncations BjPutA1-986 and RcPutA1-1116 exhibit PRODH and P5CDH activity similarly to the respective wild-type PutA (Table 4.1) indicating that the C-terminal motif is not critical for the individual activity of the catalytic domains. Kinetic measurements of the coupled PRODH-P5CDH reaction, however, indicate a role for the C-terminal motif in substrate channeling. Reaction progress curves for the coupled PRODH-P5CDH activity of RcPutA1-1116 (Figure 4.3A) and BjPutA1-986 (Figure 4.3B) show a lag prior to steady-state formation of NADH. Linear extrapolation yields estimated transient times of ~4.6 min and 1.3 min for RcPutA1-1116 and BjPutA1-986, respectively. Thus, C-terminal truncations disrupt substrate channeling in both PutAs with a more significant effect observed in RcPutA. These results are consistent with the C-terminus forming part of the lid that seals the substrate channeling tunnel from

the bulk medium (Figure 4.8B).

4.5 DISCUSSION

RcPutA is the first example of a monomeric PutA. This result is consistent with sequence analysis. For example, an alignment of BjPutA and RcPutA shows a gap in RcPutA in the region of the dimerization β-hairpin of BjPutA. Accordingly, homology models of the RcPutA catalytic core lack a β-hairpin protruding from the 640s loop. The discovery of a monomeric PutA suggests that the PutA family is diverse in terms of oligomeric states and quaternary structures. Elucidating the full extent of this diversity is a subject of current study in our labs.

We used SAXS to determine the tertiary structural interactions of a domain that represents only 15 % of the protein, which is somewhat challenging. The challenge can be appreciated by observing that the SAXS profile calculated from a model lacking the CTDUF shows good agreement with the experimental curve, particularly for $q < 0.08$ Å^{-1} (Figure 4.8A, orange curve). This indicates that methods based solely on the low q part of the SAXS data, such as Guinier analysis, have limited utility in this case. Data in the intermediate range of $0.08 < q < 0.16$ Å^{-1} contain structural information about the CTDUF, and deviations between the experimental and calculated profiles in this region drive rigid body modeling. Since these deviations are rather subtle (Figure 4.8A, orange), we developed a rigid body modeling strategy involving the generation of hundreds of models using both the predicted domain model and decoy models. We note that target decoy analysis is used in computational ligand docking[39] but has not been used in SAXS modeling to our knowledge. The

approach we used has two main benefits. First, the decoy calculations allow delineation of the low and high confidence regions of parameter space. Second, the generation of many rigid body models allows identification of clusters containing models sharing a common theme. Densely populated clusters in the high confidence zone represent valid solutions. We suggest that this approach has general applicability in SAXS modeling.

Our working model of RcPutA (Figure 4.8B) implies that the CTDUF is a Rossmann dinucleotide binding domain that does not bind NAD^+. Dinucleotides bind at the C-termini of the strands of Rossmann domains. The potential NAD^+ surface of the CTDUF is on the perimeter of RcPutA (Figure 4.8C), and NAD^+ bound to the CTDUF would be 40 Å away from the catalytic Cys, which is incompatible with the known mechanism of ALDHs in which a hydride is transferred from the hemithioacetal intermediate to NAD^+. Furthermore, homology modeling and sequence alignments suggest that the Rossmann fold domain in the catalytic core of RcPutA (residues 510-761, 955-975) is fully functional. Thus, we suggest that the CTDUF is a non-catalytic Rossmann domain.

Non-catalytic Rossmann domains are rare. The γ subunit of ATP synthase contains a Rossmann mononucleotide motif that plays a structural role.[40] Also, the DNA-processing protein DprA contains a Rossmann domain that binds single-stranded DNA.[41, 42] PutA is unique in that it contains both catalytic and non-catalytic Rossmann domains in the same protein.

Although it is unlikely that the CTDUF binds NAD^+, it nevertheless contributes to P5CDH activity. Mutation of residues in the CTDUF hairpin or the BjPutA

dimerization hairpin abrogates P5CDH activity. This result is consistent with the crystal structure of BjPutA, which shows conserved residues of the oligomerization hairpin forming intermolecular electrostatic interactions with conserved residues of the loop that anchors GSA in the P5CDH site (Figure 4.2D). Our SAXS model implies that analogous intramolecular interactions are present in RcPutA. We conclude that the CTDUF makes an essential, albeit indirect, contribution to the P5CDH activity of RcPutA.

The CTDUF likely also plays a role in substrate channeling. The reaction progress curve for the coupled PRODH-P5CDH activity of RcPutA does not exhibit a perceptible time lag, whereas the nonchanneling control displays a lag of 8 min (Figure 4.3A). These results are consistent with a substrate channeling mechanism, implying that the pathway between the two active sites is protected from the bulk medium. The SAXS rigid body model suggests that the hairpin and C-terminal motif form a flap that spans the region between two active sites, serving as a lid that seals the tunnel from the bulk medium (Figure 4.8B). The model is consistent with the observation that truncation of the flap in RcPutA1-1116 results in a 5-minute lag in the coupled PRODH-P5CDH assay (Figure 4.3A). Interestingly, truncation of the C-terminus in BjPutA1-986 results in just a 1-minute lag (Figure 4.3B). We suggest that proximity effects[43] due to the tetrameric structure of BjPutA lowers the observed transient time. This effect is not possible for monomeric RcPutA.

In summary, our results suggest that the CTDUF resembles the N-terminal half of ALDH superfamily enzymes and consists of a non-catalytic Rossmann fold domain connected to a β-hairpin. Furthermore, the hairpin facilitates P5CDH activity and substrate channeling by stabilizing the aldehyde substrate binding loop and sealing the

inter-active site tunnel from the bulk medium.

ACKNOWLEDGEMENTS

We thank Prof. Jason Cooley for providing genomic DNA and Kevin Dyer for collecting the SAXS data through the SIBYLS mail-in program. We thank Prof. Krishna K. Sharma for the use of the SEC-MALS instrument, and Drs. Santhoshkumar Puttur and Ranjan Singh for help with SEC-MALS data collection and analysis. Part of this work was conducted at the Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional support comes from the National Institute of Health project MINOS (R01GM105404).

4.6 REFERENCES

1. Singh, R. K., and Tanner, J. J. (2012) Unique Structural Features and Sequence Motifs of Proline Utilization A (PutA), *Front. Biosci. 17*, 556-568.

2. Tanner, J. J., and Becker, D. F. (2013) PutA and proline metabolism, In *Handbook of Flavoproteins. Volume 1 Oxidases, Dehydrogenases and Related Systems* (Hille, R., Miller, S. M., and Palfey, B. A., Eds.), pp 31-56, De Gruyter, Berlin.

3. Tanner, J. J. (2008) Structural biology of proline catabolism, *Amino Acids 35*, 719-730.

4. Krishnan, N., and Becker, D. F. (2005) 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 44*, 9130-9139.

5. Srivastava, D., Schuermann, J. P., White, T. A., Krishnan, N., Sanyal, N., Hura, G. L., Tan, A., Henzl, M. T., Becker, D. F., and Tanner, J. J. (2010) Crystal structure of the bifunctional proline utilization A flavoenzyme from Bradyrhizobium japonicum, *Proc. Natl. Acad. Sci. USA 107*, 2878-2883.

6. Zhang, W., Krishnan, N., and Becker, D. F. (2006) Kinetic and thermodynamic analysis of Bradyrhizobium japonicum PutA-membrane associations, *Arch Biochem Biophys 445*, 174-183.

7. Gu, D., Zhou, Y., Kallhoff, V., Baban, B., Tanner, J. J., and Becker, D. F. (2004) Identification and characterization of the DNA-binding domain of the multifunctional PutA flavoenzyme, *J Biol Chem 279*, 31171-31176.

8. Zhou, Y., Larson, J. D., Bottoms, C. A., Arturo, E. C., Henzl, M. T., Jenkins, J. L., Nix, J. C., Becker, D. F., and Tanner, J. J. (2008) Structural basis of the transcriptional regulation of the proline utilization regulon by multifunctional PutA, *J. Mol. Biol. 381*, 174-188.

9. Brown, E. D., and Wood, J. M. (1992) Redesigned purification yields a fully functional PutA protein dimer from Escherichia coli, *J Biol Chem 267*, 13086-13092.

10. Singh, H., Arentson, B. W., Becker, D. F., and Tanner, J. J. (2014) Structures of the PutA peripheral membrane flavoenzyme reveal a dynamic substrate-channeling tunnel and the quinone-binding site, *Proc. Nat. Acad. Sci. USA 111*, 3389-3394.

11. Halouska, S., Zhou, Y., Becker, D. F., and Powers, R. (2009) Solution structure of the Pseudomonas putida protein PpPutA45 and its DNA complex, *Proteins 75*, 12-27.

12. Singh, R. K., Larson, J. D., Zhu, W., Rambo, R. P., Hura, G. L., Becker, D. F., and Tanner, J. J. (2011) Small-angle X-ray Scattering Studies of the Oligomeric State and Quaternary Structure of the Trifunctional Proline Utilization A (PutA) Flavoprotein from Escherichia coli, *J Biol Chem 286*, 43144-43153.

13. Williams, I., and Frank, L. (1975) Improved chemical synthesis and enzymatic assay of delta-1-pyrroline-5-carboxylic acid, *Anal Biochem 64*, 85-97.

14. Schuermann, J. P., White, T. A., Srivastava, D., Karr, D. B., and Tanner, J. J. (2008) Three crystal forms of the bifunctional enzyme proline utilization A (PutA) from Bradyrhizobium japonicum, *Acta Cryst. F64*, 949-953.

15. Moxley, M. A., Tanner, J. J., and Becker, D. F. (2011) Steady-state kinetic mechanism of the proline:ubiquinone oxidoreductase activity of proline utilization A (PutA) from Escherichia coli, *Arch Biochem Biophys 516*, 113-120.

16. Zhang, W., Zhang, M., Zhu, W., Zhou, Y., Wanduragala, S., Rewinkel, D., Tanner, J. J., and Becker, D. F. (2007) Redox-induced changes in flavin structure and roles of flavin N(5) and the ribityl 2'-OH group in regulating PutA-membrane binding, *Biochemistry 46*, 483-491.

17. Moxley, M. A., Sanyal, N., Krishnan, N., Tanner, J. J., and Becker, D. F. (2014) Evidence for Hysteretic Substrate Channeling in the Proline Dehydrogenase and Delta1-Pyrroline-5-carboxylate Dehydrogenase Coupled Reaction of Proline Utilization A (PutA), *J Biol Chem 289*, 3639-3651.

18. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling,

Bioinformatics 22, 195-201.

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

20. Soding, J. (2005) Protein homology detection by HMM-HMM comparison, *Bioinformatics 21*, 951-960.

21. Biegert, A., Mayer, C., Remmert, M., Soding, J., and Lupas, A. N. (2006) The MPI Bioinformatics Toolkit for protein sequence analysis, *Nucleic Acids Res. 34*, W335-339.

22. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web: a case study using the Phyre server, *Nat. Protoc. 4*, 363-371.

23. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction, *BMC Bioinformatics 9*, 40.

24. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., 2nd, Tsutakawa, S. E., Jenney, F. E., Jr., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S. J., Scott, J. W., Dillard, B. D., Adams, M. W., and Tainer, J. A. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS), *Nat. Methods 6*, 606-612.

25. Classen, S., Hura, G. L., Holton, J. M., Rambo, R. P., Rodic, I., McGuire, P. J., Dyer, K., Hammel, M., Meigs, G., Frankel, K. A., and Tainer, J. A. (2013) Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source, *J. Appl. Crystallogr. 46*, 1-13.

26. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis, *J. Appl. Crystallogr. 36*, 1277-1282.

27. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *J. Appl. Crystallogr. 25*, 495-503.

28. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) Determination of domain structure of proteins from X-ray solution scattering, *Biophys. J. 80*, 2946-2953.

29. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering, *J. Appl. Crystallogr. 36*, 860-864.

30. Wriggers, W. (2010) Using Situs for the integration of multi-resolution structures, *Biophys. Rev. 2*, 21-27.

31. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models, *J. Appl. Crystallogr. 34*, 33-41.

32. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates, *J. Appl. Crystallogr. 28*, 768-773.

33. Rambo, R. P., and Tainer, J. A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering, *Nature 496*, 477-481.

34. Luo, M., Singh, R. K., and Tanner, J. J. (2013) Structural determinants of

oligomerization of delta(1)-pyrroline-5-carboxylate dehydrogenase: identification of a hexamerization hot spot, *J. Mol. Biol. 425*, 3106-3120.

35. Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and instrinsically unstructured biological macromolecules by SAS using the porod-debye law, *Biopolymers 95*, 559-571.

36. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis, *J. Appl. Cryst. 45*, 342-350.

37. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering, *J. Appl. Cryst. 42*, 342-346.

38. Brändén, C., and Tooze, J. (1991) *Introduction to protein structure*, Garland Publishing, Inc., New York.

39. Huang, N., Shoichet, B. K., and Irwin, J. J. (2006) Benchmarking sets for molecular docking, *J. Med. Chem. 49*, 6789-6801.

40. Gibbons, C., Montgomery, M. G., Leslie, A. G., and Walker, J. E. (2000) The structure of the central stalk in bovine F(1)-ATPase at 2.4 A resolution, *Nat. Struct. Biol. 7*, 1055-1061.

41. Wang, W., Ding, J., Zhang, Y., Hu, Y., and Wang, D. C. (2014) Structural insights into the unique single-stranded DNA-binding mode of Helicobacter pylori DprA, *Nucleic Acids Res. 42*, 3478-3491.

42. Quevillon-Cheruel, S., Campo, N., Mirouze, N., Mortier-Barriere, I., Brooks, M. A., Boudes, M., Durand, D., Soulet, A. L., Lisboa, J., Noirot, P., Martin, B., van Tilbeurgh, H., Noirot-Gros, M. F., Claverys, J. P., and Polard, P. (2012) Structure-function analysis of pneumococcal DprA protein reveals that dimerization is crucial for loading RecA recombinase onto DNA during transformation, *Proc. Nat. Acad. Sci. USA 109*, E2466-2475.

43. Bauler, P., Huber, G., Leyh, T., and McCammon, J. A. (2010) Channeling by Proximity: The Catalytic Advantages of Active Site Colocalization Using Brownian Dynamics, *J. Phys. Chem. Lett. 1*, 1332-1335.

44. Berka, K., Hanak, O., Sehnal, D., Banas, P., Navratilova, V., Jaiswal, D., Ionescu, C. M., Svobodova Varekova, R., Koca, J., and Otyepka, M. (2012) MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels, *Nucleic Acids Res. 40*, W222-227.

45. Srivastava, D., Singh, R. K., Moxley, M. A., Henzl, M. T., Becker, D. F., and Tanner, J. J. (2012) The Three-Dimensional Structural Basis of Type II Hyperprolinemia *J. Mol. Biol. 420*, 176-189.

Table 4-1 Steady-state Kinetic Parameters for Wild-Type and Variants of RcPutA and BjPutA

	PRODH[a]			P5CDH[b]		
	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
RcPutA (wild-type)	20.7 ± 0.4	54.3 ± 2.8	383 ± 4	7.3 ± 0.6	1530 ± 160	4800 ± 1000
RcPutAG1011P	22.4 ± 1.4	10.1 ± 3.2	2200 ± 840	-	-	-
RcPutAE1012A	19.8 ± 0.7	15.3 ± 2.6	1300 ± 270	-	-	-
RcPutAN1014A	4.8 ± 0.2	44.5 ± 6.7	110 ± 21	-	-	-
RcPutA Δ1008-1012	8.9 ± 0.2	17.8 ± 1.8	490 ± 62	-	-	-
RcPutA1-1116	26.9 ± 0.9	36.2 ± 4.5	740 ± 120	4.7 ± 0.2	1560 ± 86	3010 ± 290
BjPutA (wild-type)[c]	5.6 ± 0.7	150 ± 11	37 ± 7.4	2.2 ± 0.7	200 ± 6	11000 ± 3830
BjPutAG638P	6.1 ± 1.6	147 ± 82	42 ± 34	-	-	-
BjPutAE639A	1.95 ± 0.24	85 ± 29	23 ± 11	-	-	-
BjPutAN641A	2.5 ± 0.4	100 ± 37	29 ± 15	-	-	-
BjPutA Δ635-639	0.5 ± 0.1	43 ± 14	11 ± 5	-	-	-
BjPutA1-986	10.7 ± 1.2	151 ± 38	71 ± 26	4.4 ± 0.1	275 ± 22	16000 ± 1800

[a]DCPIP assay with proline as the variable substrate and fixed [DCPIP] = 75 μM.
[b]P5C is the variable substrate with fixed [NAD^+] = 200 μM.
[c]Values were determined previously.[4, 5]

FIGURE LEGENDS

Figure 4-1. PutA reactions and domain architectures. (A) The reactions catalyzed by PutA. (B) Domain diagrams representing the three types of PutA domain architectures.

Figure 4-2. Structure of BjPutA. (A) Protomer structure emphasizing the arrangement of domains. The surface represents the substrate-channeling tunnel calculated using Mole.[44] (B) BjPutA dimer. The two protomers are colored according to the legend in panel A. The surfaces represent the substrate channeling tunnels. (C) Close-up view of the dimer interface, emphasizing how the dimerization flap covers the substrate-channeling tunnel. (D) Interactions between the oligomerization domain (orange) and the GSA anchor loop of the P5CDH catalytic domain (cyan) in BjPutA. The locations of NAD^+ and glutamate in this figure were inferred from the structures of BjPutA-NAD^+ (PDB code 3HAZ[5]) and P5CDH-glutamate (PDB code 3V9K[45]) complexes. Residue numbers of RcPutA are listed in parenthesis. The surface represents the substrate-channeling tunnel.

Figure 4-3. Kinetic analysis of the coupled PRODH-P5CDH reaction of RcPutA and BjPutA. (A) Steady-state progress curves of NADH production from proline by wild-type RcPutA (black circles), RcPutA1-1116 (red circles), and an equimolar mixture of RcPutA monofunctional variants R454M and C791A (green circles). (B) Steady-state progress curves of NADH production from proline by wild-type BjPutA (black circles), BjPutA1-986 (red circles), and an equimolar mixture of BjPutA monofunctional variants R456M and C792A (green circles). The dashed lines show the linear extrapolation used to estimate the lag time for each reaction.

Figure 4-4. SAXS analysis of RcPutA. (A) Scattering curve and Guinier analysis (inset). The Guinier plot spans the range of $qR_g = 0.364 - 1.29$. (B) Pair distribution function. (C) Kratky plot and Porod-Debye plot (inset).

Figure 4-5. Determination of the molecular mass of RcPutA using 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 mass.

Figure 4-6. Shape reconstruction of RcPutA. (A) Shape reconstruction from GASBOR. The surface represents the averaged and filtered volume from 50 independent GASBOR calculations. The normalized spatial discrepancy is 1.02 ± 0.2. (B) Superposition of the GASBOR shape with a representative model from SAXS rigid body modeling (χ = 1.5, RMSD = 1.0; corresponds to the green curve in Figure 8A). The catalytic core is colored red. The CTDUF with C-terminal peptide composite model is colored blue. The spheres represent dummy residues linking the catalytic core and CTDUF.

Figure 4-7. Local sequence alignment of BjPutA residues 622-756 and the CTDUF of RcPutA. A homology model of the RcPutA CTDUF based on this alignment from the PHYRE2 server is shown, with identical residues in red.

Figure 4-8. Comparison of the experimental and theoretical SAXS curves. (A) Comparison of the experimental SAXS curve with theoretical curves calculated from the catalytic core model (orange) and three representative models of RcPutA from rigid body modeling. (B) The rigid body model of RcPutA that was used to calculate

the green curve in panel A (χ = 1.5, RMSD = 1.0).

Figure 4-9. SAXS rigid-body modeling results from CORAL set 1. (A) Scatter plot of the clash penalty versus χ for the 160 poses of CORAL set 1 (red circles) and the 320 decoy poses (blue squares). The green oval encloses the high confidence cluster of 78 poses, which is separated from the decoy poses. (B) Scatter plot of RMSD from the BjPutA dimer interface versus χ. The green circles represent the high confidence cluster of poses from panel A. (C) The 78 poses of the high confidence cluster. The catalytic core is colored red. The β-hairpin and abbreviated Rossmann fold of the CTDUF are colored orange and blue, respectively.

Figure 4-10. SAXS rigid-body modeling results from CORAL set 2. (A) Scatter plot of the clash penalty versus χ for the 480 poses of CORAL set 2 (red circles) and the 320 decoy poses (blue squares). The green oval encloses the high confidence cluster of 253 poses, which is separated from the decoy poses. (B) Scatter plot of RMSD from the BjPutA dimer interface versus χ. The green circles represent the high confidence cluster of poses from panel A. (C) The 253 poses of the high confidence cluster. The catalytic core is colored red, the β-hairpin orange, the Rossmann fold blue, and the conserved C-terminal motif green.

Figure 4-1 PutA reactions and domain architectures.

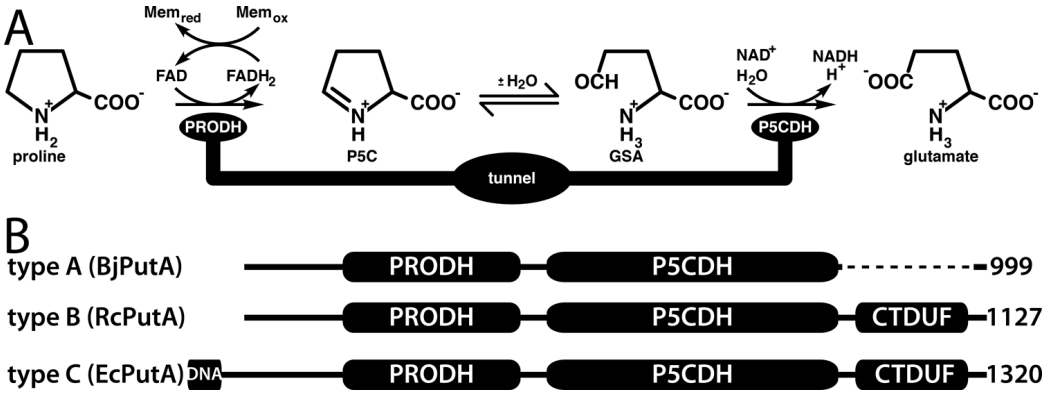


Figure 4-2 Structure of BjPutA.

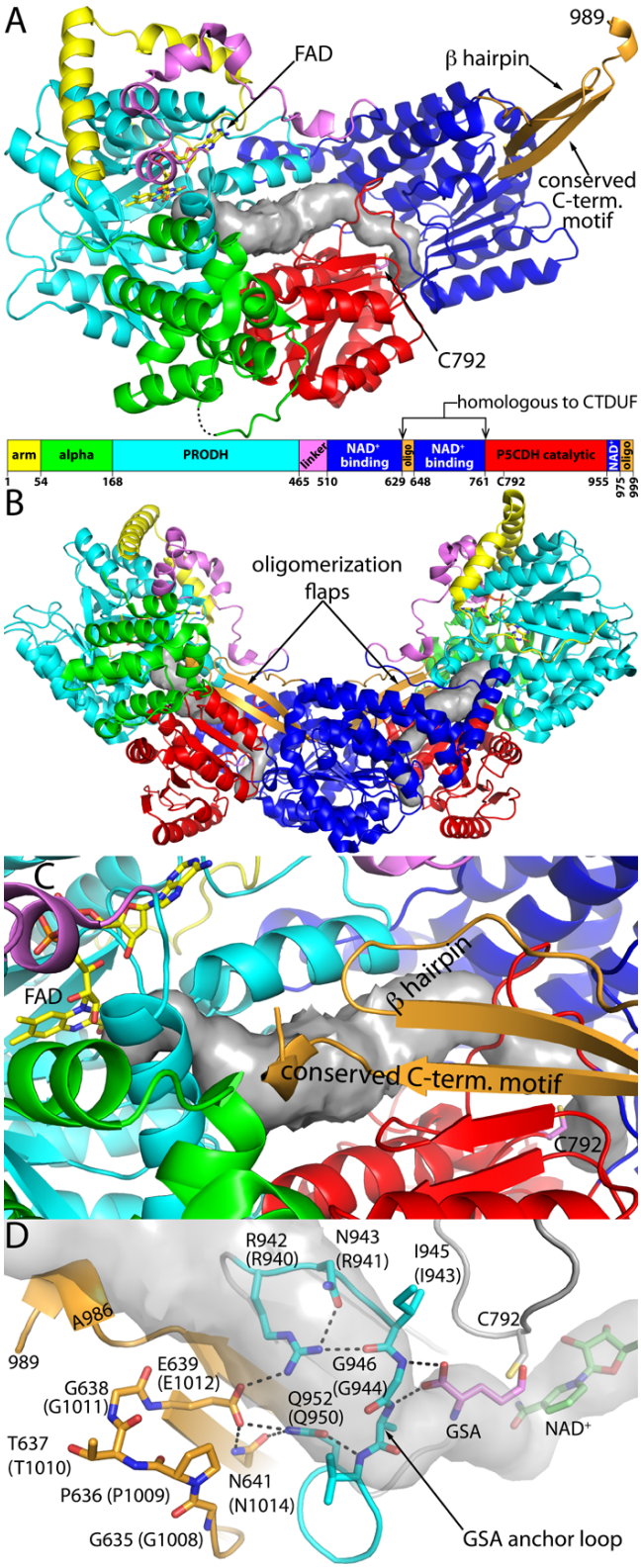


Figure 4-3 Kinetic analysis of the coupled PRODH-P5CDH reaction of RcPutA and BjPutA.

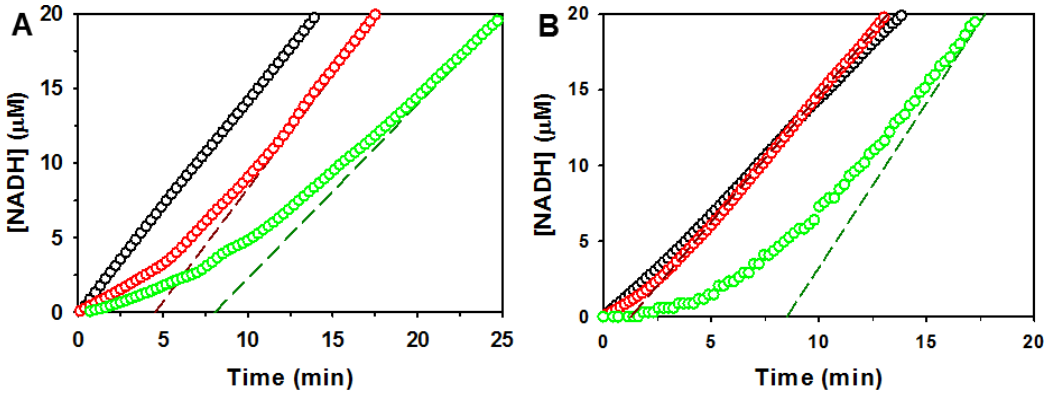


Figure 4-4 SAXS analysis of RcPutA.

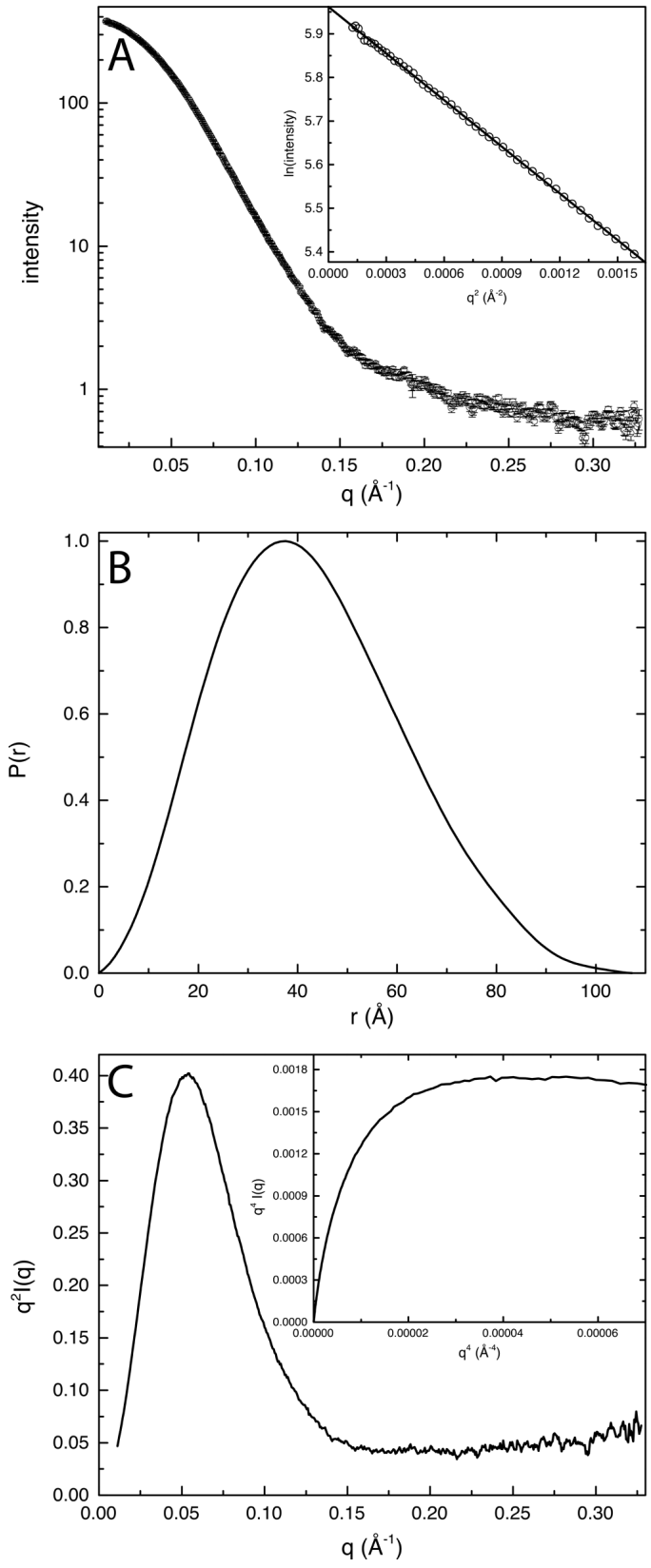


Figure 4-5 Determination of the molecular mass of RcPutA using MALS.

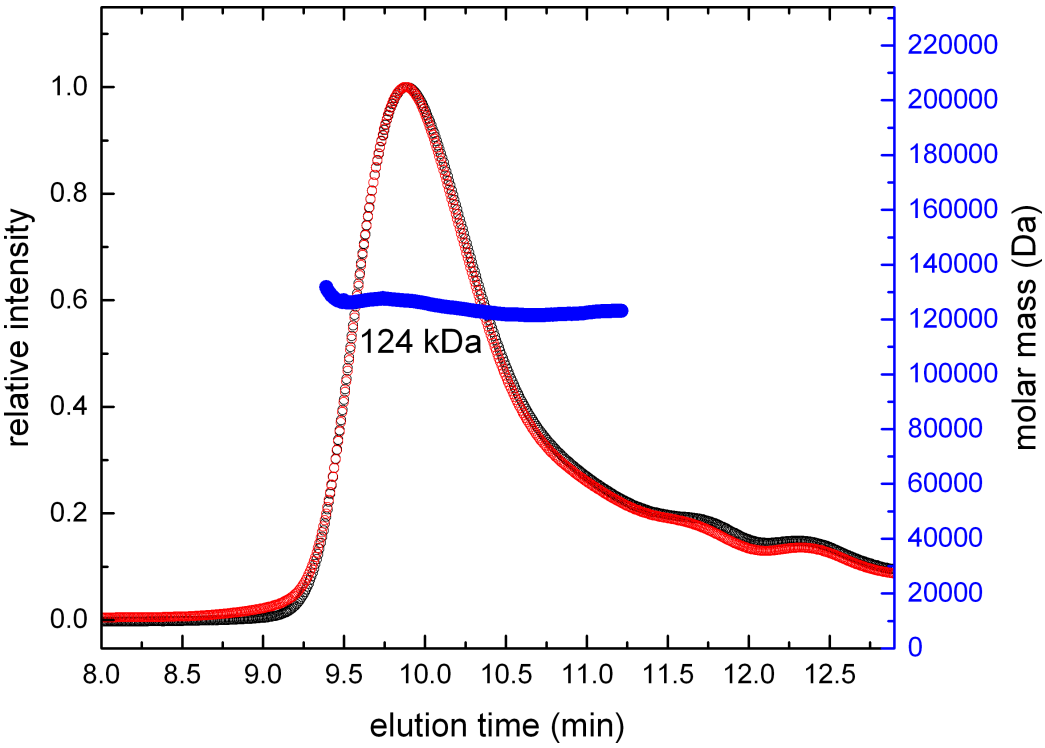


Figure 4-6 Shape reconstruction of RcPutA.

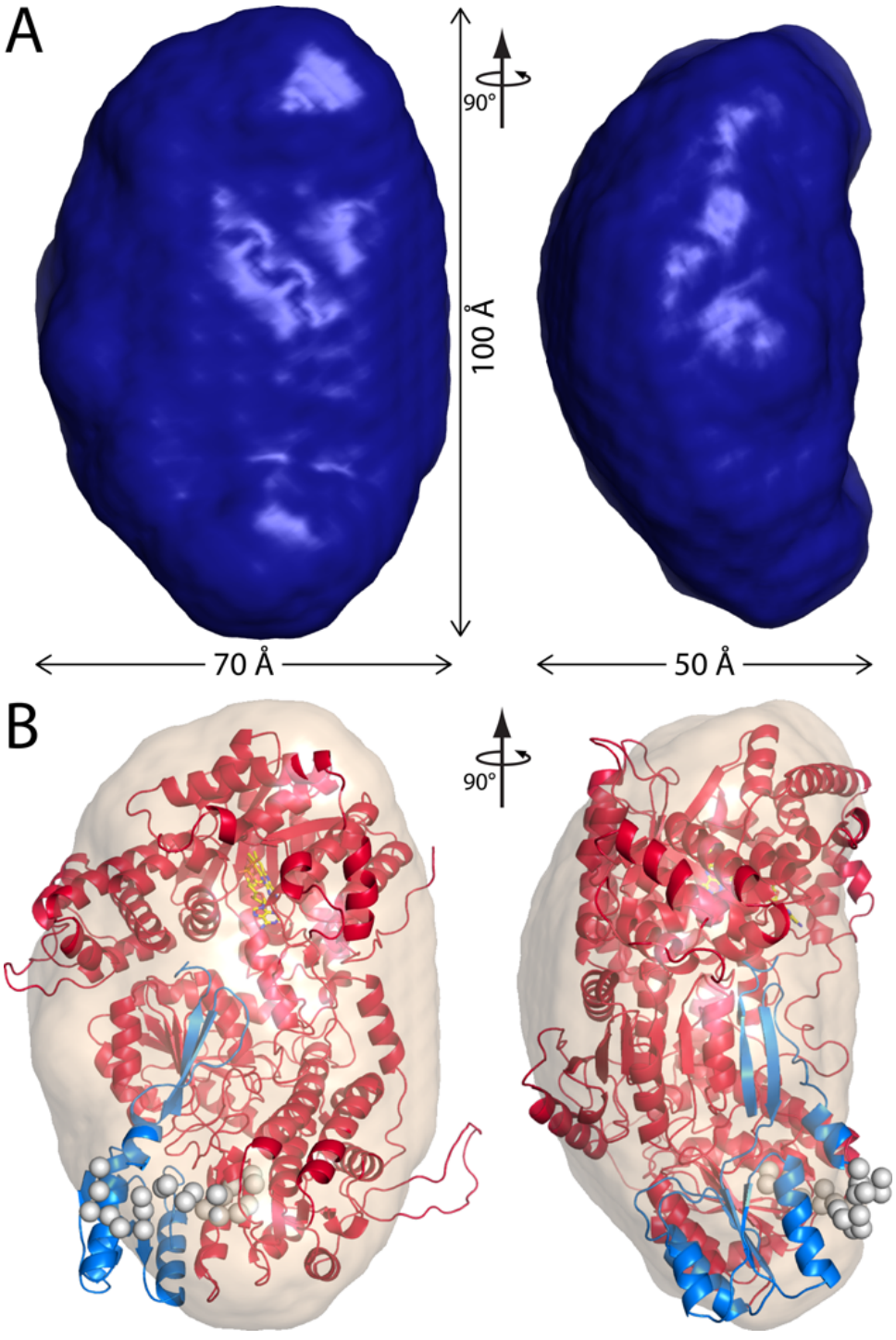


Figure 4-7 Local sequence alignment of BjPutA residues 622-756 and the CTDUF of RcPutA.

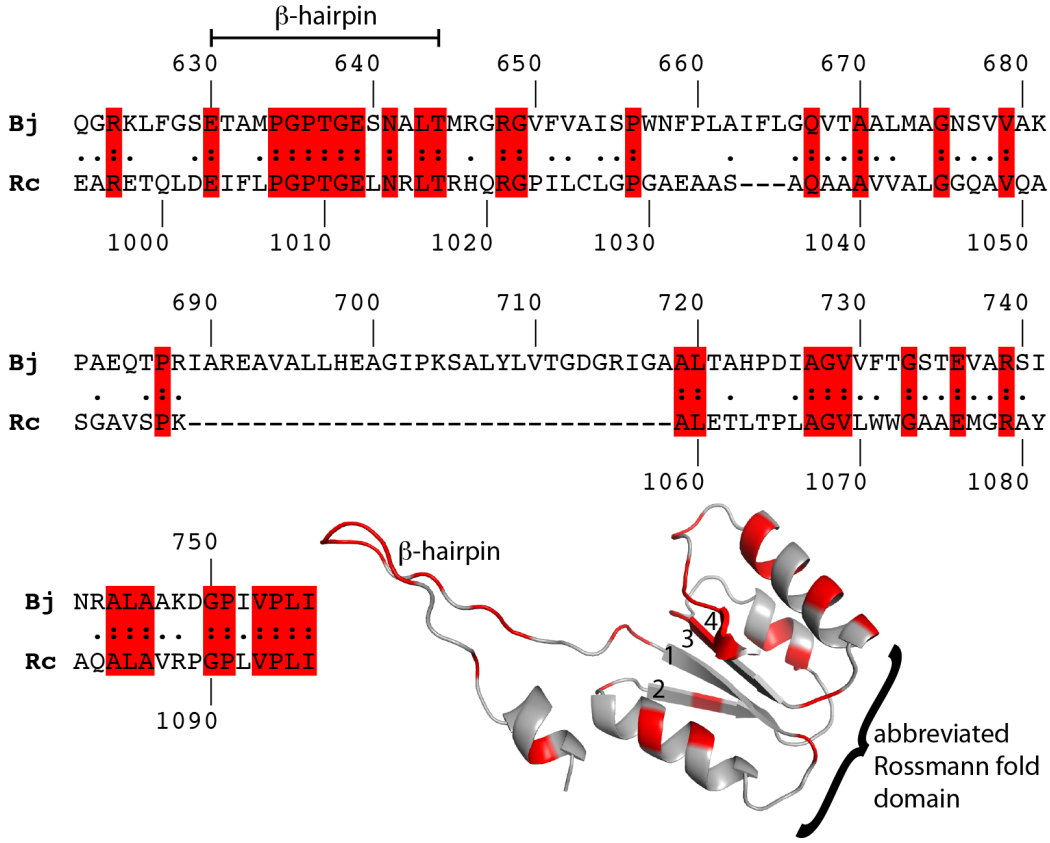


Figure 4-8 Comparison of the experimental and theoretical SAXS curves.

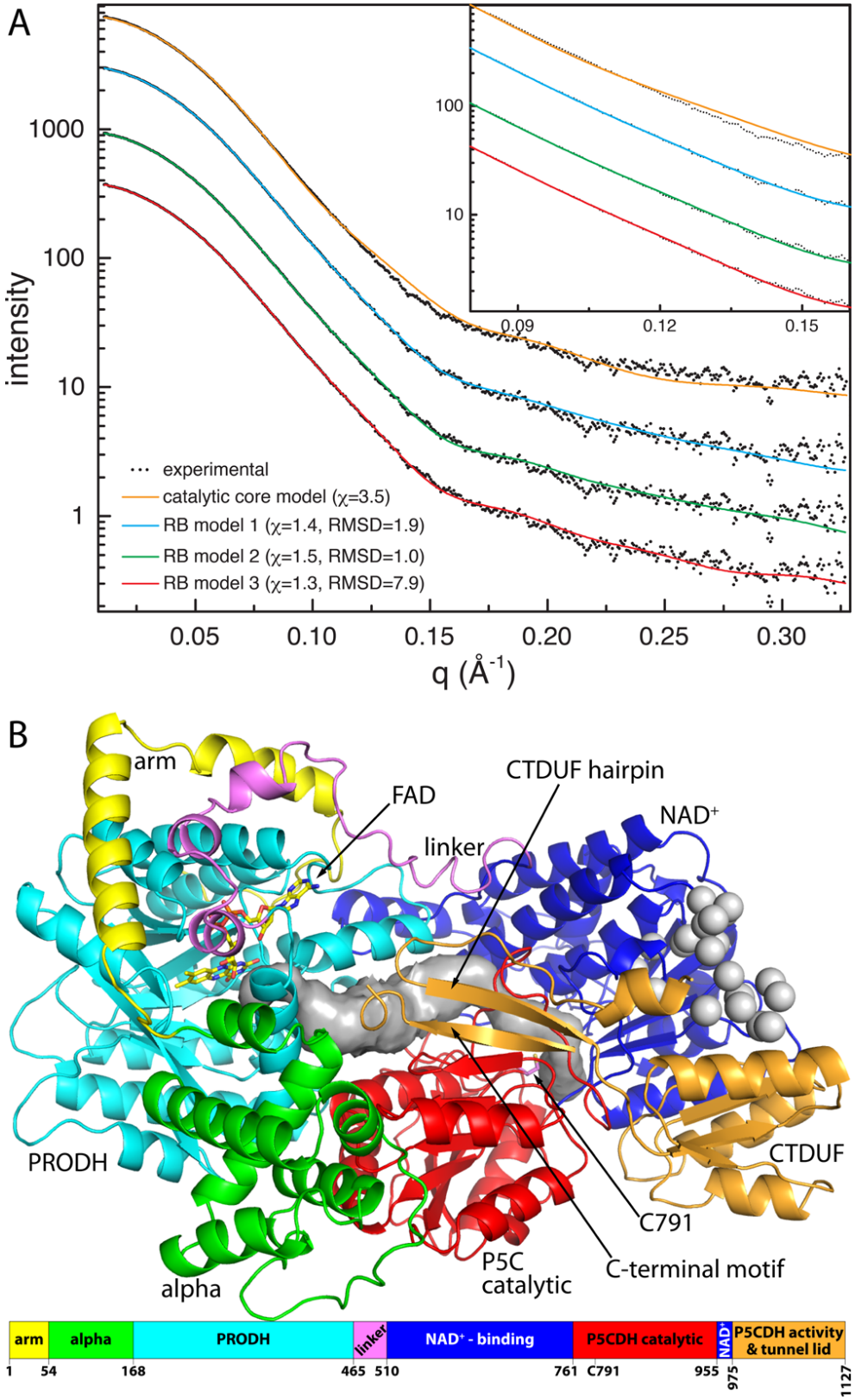


Figure 4-9 SAXS rigid-body modeling results from CORAL set 1.

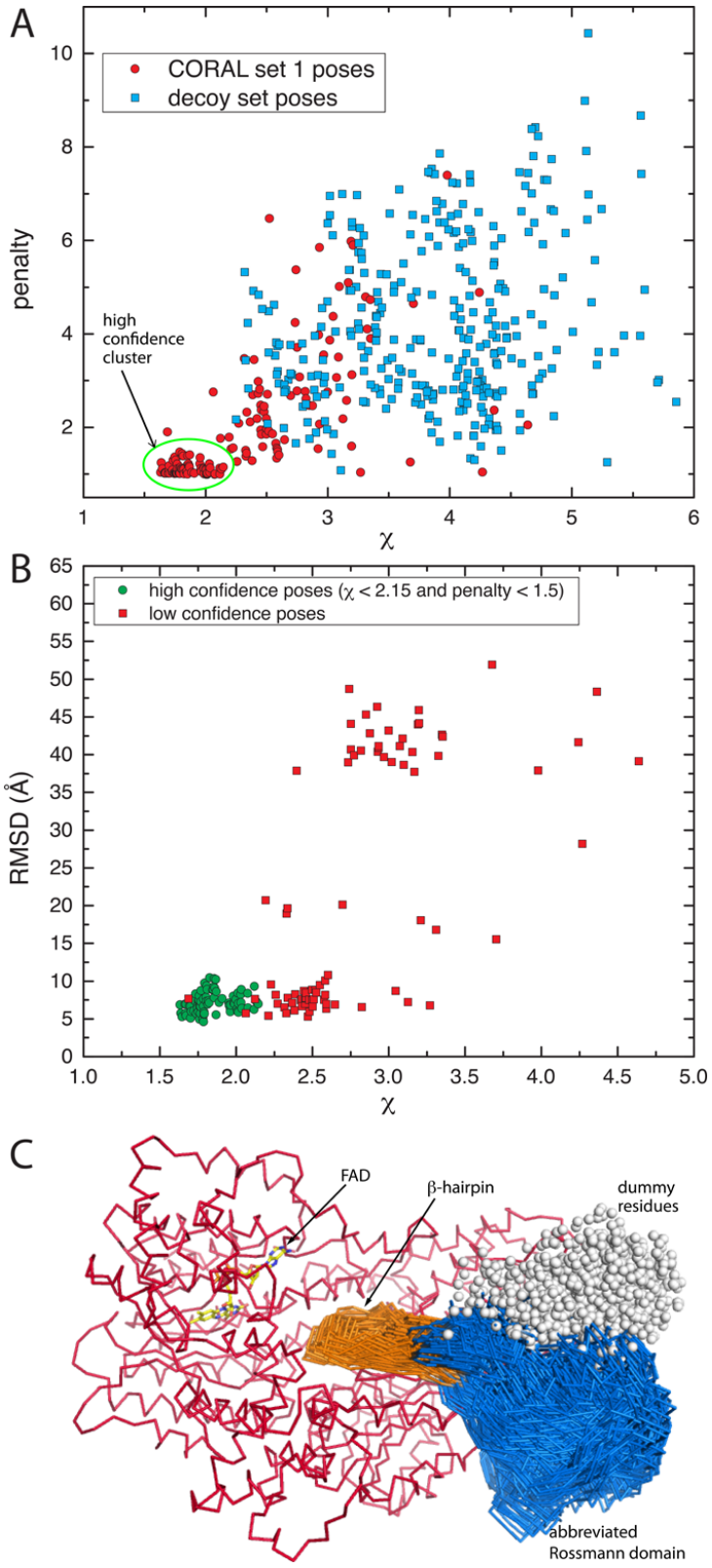
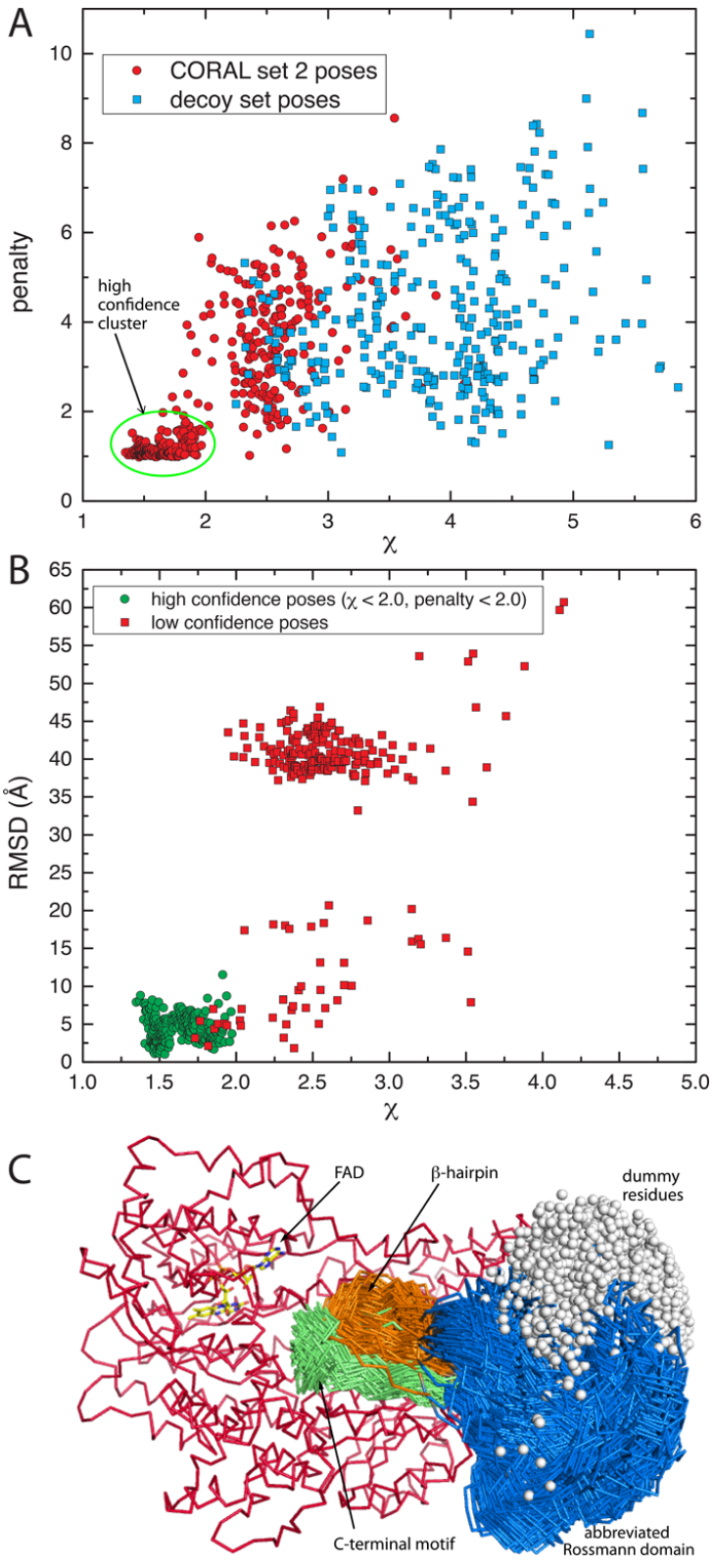


Figure 4-10 SAXS rigid-body modeling results from CORAL set 2.



CHAPTER 5

5. NEW INSIGHTS INTO PRODH DOMAIN FUNCTION IN PROLINE DEHYDROGENASE AND SUBSTRATE CHANNELING OF TYPE 1A PUTA FROM *B. JAPONICUM*

Min Luo, Travis A. Pemberton

ABBREVIATIONS

PutA, proline utilization A; BjPutA, proline utilization A from *Bradyrhizobium japonicum*; PRODH, proline dehydrogenase; P5C, 1-pyrroline-5-carboxylate; GSA, glutamate semialdehyde; P5CDH, 1-pyrroline-5-carboxylate dehydrogenase; GsPutA, proline utilization A from *Geobacter sulfurreducens*; DrPRODH, proline dehydrogenase from *Deinococcus radiodurans R1*; BjPutAo, oxidized proline utilization A from *Bradyrhizobium japonicum*; BjPutAr, reduced proline utilization A from *Bradyrhizobium japonicum*; LAC, L-lactate; CoQ1, Coenzyme Q1.

5.1 ABSTRACT

Proline utilization A from gram-negative bacteria *Bradyrhizobium japonicum* (BjPutA), a Type 1A peripheral membrane flavoenzyme, oxidizes L-proline to glutamate in proline metabolic pathway. PutAs play important role in pathogenicity and energy metabolism in many bacteria. Previous structural study and kinetic characterization suggest it utilizes a substrate channel to transfer the intermediates from the active site of proline dehydrogenase domain (PRODH) to the second active site of Δ^1-pyrroline-5-carboxylate dehydrogenase domain (P5CDH). However the only Type 1A PutA structure of BjPutA available so far, cannot provide enough information on the potential dynamic substrate channel pathway, leaving the question where and how the PRODH domain transfer the product to the downstream channel remains enigmatic. In this study, by investigating the conformational changes happening within the reductive half-reaction of PRODH from four BjPutA structures in different reaction states, together with performing kinetic and structural characterization of 4 active-sites mutants, we determined a previously unidentified complex and dynamic substrate-channeling pathway in BjPutA, and different flexibility profile of the PRODH active site was revealed. Our data provides new insights of how PRODH domain functions in proline dehydrogenase activity and substrate channeling in Type 1A PutA.

5.2 INTRODUCTION

The proline catabolism pathway catalyzes the two-step oxidation of proline to glutamate (Fig. 5.1A). Proline dehydrogenase (PRODH) catalyzes the first oxidation by transferring 2 electrons, in the form of hydride, from proline to the FAD cofactor – these steps happen in the reductive half-reaction (Fig. 5.1B), followed by the oxidation of the FAD, in which 2 electrons abstracted from proline flow into the electron transport chain. The product of PRODH, Δ^1-pyrroline-5-carboxylate (P5C), is translated into glutamate-γ-semialdehyde (GSA) via a non-enzymatic hydrolysis reaction. GSA is the substrate for the NAD^+-dependent P5C dehydrogenase (P5CDH), which catalyzes the final oxidation to glutamate. Proline catabolism is important for the pathogenicity and energy metabolism of bacteria, like *Helicobacter pylori*, which can harness energy exclusively from proline in human gut.

PRODH and P5CDH are fused in a single polypeptide chain known as proline utilization A (PutA) in Gram-negative bacteria and *Corynebacterium*. Based on multiple sequence alignments, there are two main PutA branches: Branch 1 and Branch 2. The amino acid sequence identity between these two branches are normally less than 30%, which suggests significant differences between Branches 1 and 2 PutAs. Branch 1 PutAs mainly come from *α*-, *β*-, and *γ*-proteobacteria; Branch 2 PutAs normally exist in Gram-negative cyanobacteria, δ- and ε-proteobacteria, and *Corynebacterium*. According to the domain architecture difference, PutAs can be further sub-classified into Type A, B, and C (Fir. 2). Type A PutAs contain the minimal catalytic core, which consists of N-terminal PRODH and C-terminal P5CDH domains separated by about 50 residues, *Bradyrhizobium japonicum* PutA (BjPutA)

from Branch 1 and *Geobacter sulfurreducens* PutA (GsPutA) from Branch 2 are well-characterized type A PutAs.[6,7] Types B PutAs contain an additional C-terminal domain (CTD) that is not present in type A PutAs, *Rhodobacter capsulatus* PutA (RcPutA)[5] from Branch 1 and *Helicobacter Pylori* PutA (HpPutA) from Branch 2 are two examples of type B PutAs. Type C PutAs have an N-terminal DNA binding domain in addition of the catalytic core and CTD,[8] which only exist in Branch 1 PutAs, PutA from *Escherichia coli* (EcPutA) is such a Type C PutA.

So far, there are only two full-length PutA structures available, one is BjPutA from Type 1A PutA, another one is GsPutA, which is a Type 2A PutA, the amino acid sequence between them is only 27% identical. In both BjPutA and GsPutA structures, a large internal tunnel was observed to connect the PRODH and P5CDH active sites, which implies a potential substrate channeling for the intermediates GSA/P5C transfer. Consistent with the potential substrate channel found from structural basis, kinetic studies have solidly proved substrate-channeling character in PutAs. Early studies from *Salmonella typhimurium* PutA showed exogenous P5C does not efficiently compete over endogenous ^{14}C-labeled P5C produced by PRODH for P5CDH activity, which is a direct evidence for a channeling mechanism. More recent steady-state and rapid-reaction kinetic measurements from BjPutA, GsPutA and RcPutA clearly showed the production of NADH from proline is without a perceptible time lag, which further confirmed the intermediates is directly transported from PRODH to P5CDH active site rather than released into the bulk solution, the substrate channeling.

Based on the BjPutA and GsPutA structures, the whole substrate-channeling pathway within the protein can be divided into three parts: the two active sites of PRO

and P5C site, and the middle section between them (Fig. 5.3). The middle section also refers to hydrolysis cavity, because it locates right between the two active sites and is formed by lots of hydrophilic residues, which provides an appropriate environment for the P5C hydrolysis reaction. The hydrolysis cavity occupies over 90% of the total volume of the entire channeling system. The hydrolysis cavity in BjPutA is 24 Å by 14 Å in its two largest dimensions, with the third dimension of 3-7 Å, and its volume calculated from VOIDOO is about 1325 $Å^3$ with the whole substrate channel estimated to be 1400 $Å^3$.[6]

The structures of BjPutA and GsPutA have characterized the structural basis of substrate channeling, but still farness than complete - the route between PRODH active site and the hydrolysis cavity where P5C get transferred and how does PRODH active site function to promote P5C getting channeled to the hydrolysis cavity have never been clearly determined. The available BjPutA structures, in which a sulfate ion is always bound in PRO site and an iron pair is forming between Glu197 and Arg456, depicted a circuitous route (route A) (Fig. 5.3), which goes down the ribityl chain of the FAD, then passes through a potential tunnel to the bulk solution, and finally wanders under the side chain of Arg456 until connecting with the hydrolysis cavity. A more direct route (route B) was found that directly pass through the corresponding iron pair Arg421-Glu149 to the hydrolysis cavity based on the more recent ligand-free GsPutA structure. GsPutA structure of both ligand free and bound states has been determined, which revealed large dynamic flexibility of PRODH active sites. Assuming the PRO site of BjPutA has similar dynamic profile by cutting the side chains of the ion pair of Arg456-Glu197, which corresponds the Arg421-Glu149 from GsPutA, the route B is also observed from computational analysis (Fig.5. 3).

BjPutA is the only Type 1A PutA whose full-length structure is available, but its structure has been locked in one conformation with a sulfate ion bound in the active site of PRODH. Without the structural information from different enzymatic states, the substrate recognition mechanism, conformational changes upon flavin reduction, and dynamic aspect of the PRODH active site is literally unknown. And different routes between PRO site and hydrolysis cavity have been found in Type 1A and Type 2A PutAs, which raised the question: which one is the correct tunnel that function in substrate channeling? Also, large flexibility of the β1−α1 loop, α8 and the Arg421-Glu149 iron pair in GsPutA has been determined, and their dynamic movements control the route B open and close, which raises another question - that whether the PRODH active site of Type 1A PutA also has large flexibility to make the route B forming.

We therefore used BjPutA as a model of type 1A PutAs. By determining the enzyme structure of all the four states in PRODH reductive half-reduction (Fig. 5.1B), we revealed a previously unidentified dynamic tunnel system in PRO site of BjPutA, and verified where and how the P5C get transferred from PRODH to the hydrolysis cavity. Also, different dynamic flexibility profile of the PRODH active site in the type 1A BjPutA was observed compared to type 2A GsPutA. Furthermore, we report four active site mutant structures of BjPutA and their kinetic parameters of PRODH activity, which provide new insights into PRODH function and substrate channeling that, accompany substrate binding and flavin reduction.

5.3 MATERIALS AND METHODS

5.3.1 Site-Directed Mutagenesis.

Site-directed mutants of BjPutA were generated using the QuikChange II site-directed mutagenesis kit (Agilent). The mutations were confirmed by DNA sequencing. The mutant enzymes were purified as described for the wild-type (WT) enzymes, except where noted.

5.3.2 Protein Expression and Purification.

BjPutA and BjPutA mutant enzymes were expressed and purified using immobilized metal affinity chromatography (Ni^{2+}-charged His-Trap HP, GE Healthcare) as described previously. The N-terminal His tag was removed as described previously. The purified proteins were dialyzed into 50 mM Tris-HCl, 0.5 mM THP, 5% glycerol, and 0.5 mM EDTA (pH 8.0), quick frozen in liquid nitrogen, and stored at −80 °C.

5.3.3 Crystallization of wild type BjPutA and BjPutA mutants G196A, G196P, E197A and E460A in the space group of *C*2.

Centered monoclinic crystals were grown similarly as described for BjPutA previously, except that *C*2 crystals got from original crystal screen were used as microseeds for further optimization. Crystals with good diffraction quality were grown in sitting drops at room temperature in the presence of 1.7-2.1 M ammonium sulfate and Tris buffer between pH 7.5-9.0, picked up with Hampton nylon loops, and plunged into liquid nitrogen.

5.3.4 Crystallization of Proline-reduced BjPutA.

Wild-type crystals were soaked and cryoprotected with reservoir supplemented with 2 M proline, which was achieved through 0.5 M incremental steps at a starting concentration of 0.5 M proline. The cryoprotection in proline took approximately 10-

15 minutes to reach the desired concentration of 2 M. BjPutA crystals have a distinct yellow color as a result of using FAD as a cofactor. As proline concentrations increased, the crystals visibly turned colorless from the edges inward, finally becoming completely colorless between 1 and 1.5 M proline. The colorless crystals were then looped and plunged and stored in liquid nitrogen. If many crystals were being pulled from the same drop, the cryobuffer being used was replaced with fresh solution after every 2-3 crystals.

5.3.5 Crystallization of BjPutA-lactate complex in both oxidized and reduced states.

The complex of BjPutAo-lactate in the oxidized state was obtained by soaking wild-type BjPutA crystals with cryobuffer including 1M ammonium sulfate, 0.1 M Tris pH 7.8, 25% (w/v) PEG3350, 0.6 M L-lactate and 25% glycerol for 0.5 minutes, then picked up with Hampton nylon loops, and plunged into liquid nitrogen. The BjPutA-lactate complex in its reduced state was obtained by soaking with the same cryobuffer for BjPutAo-lactate supplemented with 0.1 M reducing agent sodium dithionite for 0.5 minutes, picked up with Hampton nylon loops, and plunged into liquid nitrogen.

5.3.6 X-ray Data Collection.

The space group is *C*2 with a BjPutA dimer in the asymmetric unit. X-ray diffraction data sets were collected at beamline 4.2.2 of the Advanced Light Source (ALS) using a NOIR-1 detector. The data were integrated with MOSFLM and scaled with SCALA. The mutants G196A, G196P, E460A and E197A were collected using the ALS 4.2.2 beamline on their new Pilatus detector.

5.3.7 Refinement and validation.

XDS was used for integration and SCALA was used for scaling. Refinements in PHENIX were initiated from models derived from the structure of wild-type BjPutA (Protein Data Bank (PDB) entry 3HAZ). COOT was used for model building. The structures were validated with MolProbity and the PDB validation server. Data collection and refinement statistics are listed in Table 5. 2.

5.3.8 Computational analysis.

The substrate-channeling tunnel system was analyzed and visualized with MOLE, which finds tunnels that connect cavities to the bulk medium. In order to find the tunnel that connects the active sites of PRODH and P5CDH, the cofactor NAD^+ from the P5CDH domain was moved, and the heteroatoms except FAD were deleted too. MOLE was run with default options and using sulfate ion from 3HAZ as the starting point, because it right bound at the proline binding site. The main substrate-channeling pathway between the two active sites was discussed in this paper.

5.3.9 Steady-State Kinetics.

All steady-state assays were conducted at 23 °C in 50 mM potassium phosphate and 25 mM NaCl (pH 7.5). For WT BjPutA, G196A and G196P, coenzyme Q1 (CoQ1) was used as the electron acceptor with CoQ1 reduction monitored by the decrease in absorbance at 278 nm using a molar extinction coefficient of 14.5 mM-1 cm-1. The Km for proline and the kcat for the enzymes were determined by varying the concentration of proline (0–300 mM) while keeping the CoQ1 concentration constant (0.1 mM).

5.4 RESULTS

5.4.1 Structure of oxidized BjPutA complexed with L-lactate.

Structure of oxidized BjPutA complexed with the proline analog L-lactate was determined (Table 5.1). This structure is the first Type 1A PutA complexed with a proline analog and represents state II of E_O-Pro complex (Fig. 5.1B), which mimics the state of the enzyme before P5C get transferred from PRO site to hydrolysis cavity.

The overall structure of BjPutAo-LAC is the same as 3HAZ;[6] their root-mean-square deviation calculated from the backbone is 0.18 Å.[11] Besides the two main catalytic domain of PRODH and P5CDH, there are three smaller domains which are named arm, alpha domain, and linker. The *C*2 crystal form has two molecules of protein in one asymmetric unit, which forms a U-shaped dimer. The architecture of BjPutA protomer is present in Figure 5.4. The N-terminal 51 residues wraps around the PRODH domain like an arm (yellow). The arm is followed by the alpha domain (green), which tightly packed against PRODH domain, and also contact with P5CDH domain. A linker (pink) of about 83 residues connect PRODH and P5CDH domain. The overall folding for the PRODH domain is a distorted $(\alpha\beta)_8$ barrel with the distortion part comes from the $\alpha_{8,}$ which sits on top of the barrel instead of packing along with the other helix. P5CDH is formed by the 3 classic domains as the other aldehyde dehydrogenases: the catalytic domain (red), NAD^+-binding domain (blue) and the oligomerization domain (brown).[12] The oligomerization domain in Type 1A PutA has been noticed that not only function in dimerization, but also plays important role in substrate channeling by serving as a lid that covers the hydrolysis cavity.[6]

This structure first time revealed the substrate recognition mechanism for type 1A

PutA, which is similar to that of type 2A PutA.[7] L-lactate (LAC) binds between the α_8 and the *si* face of the FAD isoalloxazine (Fig. 5.5A). The carboxylate of LAC forms ion pairs with Arg456, Arg457, and Lys237. The hydroxyl oxygen of LAC, that mimics the N4 of proline which is next to the hydride donor atom C5, is positioned 3.3 Å away from the hydride acceptor of flavin N5 atom, this arrangement suggests the direct hydride transfer mechanism, which was observed in the other PRODH family members.[13,14] Upon lactate binding, two ion pairs form: one is Arg456-Glu197 between β1−α1 loop and α8; another one is Arg457-Glu460 with both residues coming from α8. Furthermore, the 2' and 3'-OH group of the FAD ribityl chain hydrogen bonds with the ion pair of Arg457-Glu460.

5.4.2 Structure of Proline-reduced BjPutA.

Structure of proline-reduced BjPutA was determined from soaking the crystals of wild-type enzyme with 2 M proline. This is the first reduced type 1A PutA structure, also the first PutA structure reduced by using the natural substrate proline. This structure represents the state IV of E_r (Fig. 5.1B), which provides insight into the state of enzyme right after P5C get transferred from PRO site to hydrolysis cavity.

This structure provides the first structural evidence for conformational changes happening in type 1A PutA upon flavin reduction. Reduction triggers large conformational changes in the flavin (Fig. 5.5A and 5B, Fig. 5.6). A 22° butterfly bend is found in the isoalloxazine, which differs from the planar isoalloxazine from 3HAZ, this bend is caused by the moving of N5-N10 axis toward the substrate binding site. Also, the three hydroxyl groups of the ribityl chain rotates substantially from its oxidized form, 2'- and 3'-OH shifts to the pyrimidine side of the isoalloxazine and the 4'-OH points to the dimethylbenzene side. Similar

conformational changes of flavin were observed in GsPutA.[7]

Reduction of flavin also induces conformational changes of the enzyme (Fig. 5.5B). Between the si face of flavin and α8, no density was observed, which indicates the active site is empty. Compared to the structure of BjPutAo-LAC, the ion pair of Arg456-Glu197 is ruptured with the side chains 6.8 Å away from each other, in this reduced structure, guanidinium group of R456 points away from the active site and hydrogen bonds with the backbone carbonyl of alanine 344, the carboxyl group of E197 bends away from the active site toward the internal cavity and hydrogen bonds with arginine 200 as well as water mediated hydrogen bonds with arginine 284. Another ion pair Arg457-Glu460 does not exist anymore, the side chain of Glu460 gets disordered and the side chain of Arg457 flip away from proline binding site, thus the hydrogen bonds between flavin and Arg457 and Glu460 are disrupted. Simply speaking, the reduction of flavin disrupts both interactions between α8 with flavin and β1−α1 loop.

Surprisingly, reduction of flavin does not induce any dramatic shift on the backbone of the active site. This is quite different from GsPutA,[7] in which large shifts of β1−α1 loop and α8 have been observed accompany flavin reduction. This indicates different dynamic profile of the active sites from BjPutA, a type 1A PutA.

5.4.3 Structure of dithionite-reduced BjPutA complexed with L-lactate.

A 2.3 Å resolution structure of reduced BjPutAr-LAC complex was determined from a crystal of the wild-type enzyme that was soaked with L-lactate and sodium dithionite. This is the first PutA structure with proline analog bound in the reduced state. This structure represents state III of E_r-P5C, and it captures a snapshot that

PRODH active site begins to transfer P5C to the hydrolysis cavity, which serves as the structural basis of how PRODH functions to transfer P5C in substrate channeling.

The reduction induces substantial conformation changes of the flavin, and they are similar with what we have observed in the proline-reduced BjPutA, which includes bending of the isoalloxyzine and crankshaft rotation of the ribityl chain.

Accompany the reduction of flavin, the side chain of Arg457 and Glu460 get disordered (Fig. 5.5C), which renders the ion pair between lactate and Arg457, and also disrupts the interaction between FAD and α8. Without these interactions, the L-lactate still locates in the substrate pocket. Tight salt bridge still forms between the carboxylate of lactate and Arg456, and even tighter ion pair is forming between Lys237 and lactate. Considering that Lys237 is located near the Arg456-Glu197 ion pair, the movement of lactate, that pass though this ion pair upon its rupture, toward the hydrolysis cavity is easily forecasted.

5.4.4 Structure of R456A.

The structure of R456M is investigated to illuminate the structural conformation of the resting enzyme with the substrate pocket unoccupied. The lone BjPutA structure (3HAZ) has a sulfate ion from the crystallization solution bond in the PRODH active site,[6] which makes it problematic to reveal the substrate-channeling tunnel in its resting state. And because Arg456 is the main residue interacting with the sulfate, a mutant of R456M mutant was proposed to mimic the resting enzyme. The structure of R456M is determined in the same *C*2 space group as 3HAZ, which represents state I of E_O.

The substrate pocket is confirmed to be unoccupied since no density was observed

between the si face of flavin and α8, which confirms this mutant R456A a good mimic of the resting enzyme (Fig. 5.5D). Mutation of Arg456 to Met does not result in any noticeable conformational change on the backbone of the active site, but large flip has been observed on the side chain of the residues of Glu197 and the Met456 (Fig.5.5D). Similar to the proline-reduced BjPutA structure, the side chain of these two residues flips away from the active site toward hydrolysis cavity with a distance larger than 7 Å between Met456 and Glu197.

5.4.5 The dynamic profile of PRODH domain within its reductive half-reaction.

The backbone of the PRODH active site does not undergo large movements (Fig. 5.7), which is quite different from what was observed in GsPutA.[7] During the reductive half-reaction of PRODH, no obvious shift is exhibited in the β1−α1 loop of BjPutA, while 5.7Å relocation were seen in GsPutA. Movement is shown on the backbone of α8, which moves < 1.2 Å toward proline binding site in the residue region of 457-462 upon substrate binding. This region contains the ion pair of Arg457-Glu460 in E-S complex, and high B-factor of this region suggests high mobility too (Fig. 5.8). While compared to the 3.5 Å movement of α8 from GsPutA, the flexibility in BjPutA is small. These results revealed a much lower dynamic profile for the PRODH active sites of BjPutA.

Unlike the backbone, large dynamic aspect was revealed involving the ion pair of Arg456-Glu197 and Arg457-Glu460 (Fig. 5.8, and Fig. 5.9). These two ion pairs were formed in the closed E_o-S complex, the ion pair of Arg457-Glu460 also hydrogen bonds with FAD, which becomes broken upon flavin reduction. The ion pair Arg456-Glu197 formed between the β1−α1 loop and α8, which was disrupted only with the substrate pocket unoccupied. Breaking of this ion pair would open the Rout B, which

provides a quick, controlled and straightforward tunnel for P5C transfer to the hydrolysis cavity (Fig. 5.8A and 5.8C).

5.4.6 The tunnel system change of PRODH domain within its reductive half-reaction.

With the conformational changes determined within the reductive half-reaction of PRODH domain, computational analysis on the tunnel system by MOLE[15] was performed on the structures of R456M (a mimic of resting enzyme), the BjPutAo-LAC complex (E_o-S) and the proline-reduced BjPutA with its proline binding site empty (E_r), aiming to resolve the puzzle that where and how PRODH functions to transfer P5C to the following substrate channel in BjPutA. Mole calculation was performed by using the default parameters and origin from the position of the sulfate ion in 3HAZ, since it right located in the substrate-binding site of PRODH domain.

An unverified tunnel of Route B in Type 1A PutA was found in the states of E_o and E_r (Fig. 5.8A and 5.8C), this tunnel has been previously shown in the Type 2A PutA of GsPutA.[7] A tunnel of Route A was found in E-S state (Fig. 5.8B), which is similar to the one already found in 3HAZ structure (Fig. 5.3).[6] Route B utilizes an exit toward the hydrolysis cavity much more straightforward than Route A. It origins from the proline binding site, pass through the broken ion pair of Arg456-Glu197, and then directly merge into the hydrolysis cavity. The ion pair of Arg456-Glu197 serves as an active site gate. Accompany with proline binding and P5C release, it forms and disrupts to close and open the Route B, functioning in control traffic between PRO site and the hydrolysis cavity. The 6.8 Å hole formed between the broken ion pair is just big enough for P5C to pass through, which is estimated to have a sphere diameter

of approximately 5.8 Å.

The dithionite-reduced structure of BjPutAr-LAC captured a snapshot that P5C begins to get transferred by PRODH, which provides information on the transfer mechanism of P5C to the hydrolysis cavity by PRODH (Fig. 5.9A). After reduction of flavin, the ion pair of Arg457-Glu460 becomes broken, at the time, the salt bridge between Arg457 and lactate (mimic of product P5C)) is disrupted. Without travelling through the widely open tunnel of Route A (Fig. 5.9B), lactate binds more tightly onto the Route B side with one more hydrogen bond forming between Lys237 and shorter salt bridge with Arg456. Accompany with the undergoing rupture of the Arg456-Glu197 ion gate, a movement of lactate along with Arg456 is not hard to be predicated. This suggests that Arg456 not only serves as a substrate-binding residue, but also plays the role of transferring the product toward the hydrolysis cavity through the ion gate, where the Route B locates.

5.4.7 Rationale for generating the Gly and Glu mutants.

The Gly mutants of G196A and G196P were generated to confirm the unexpected low dynamic profile of the β1−α1 loop (Fig. 5.7). Large flexibility of the β1−α1 loop has been determined in GsPutA and a mono-functional PRODH from *Deinococcus radiodurans* (DrPRODH), in these two enzymes; β1−α1 loop withdraws from the active site of 5.7 and 5.2 Å respectively. The Gly196 is conserved among PRODH family, and in DrPRODH, this loop has shown to be essential for optimal catalysis: the Gly to Ala mutant severely impaired the catalytic activity to 140 times lower than the wild type enzyme.[14] If the rigid β1−α1 loop we observed from BjPutA were true, the Gly mutant would not have dramatic effect on the PRODH activity.

Two Glu mutants of E197A and E460A were made to check the importance of the two ion pairs on the PRODH function, and E460 is located right on α8 region where high B factors were observed (Fig. 5.8), its mutation could confirm the mobile character through the structural determination. Substrate recognition seems to be conserved in PRODH family,[7,14,16] the key active site components in PRODH are FAD, the β1−α1 loop and α8. Upon proline binding, two ion pairs forms. The Arg456-Glu197 ion pair bridges the β1−α1 loop and α8, and the other ion pair connects FAD and α8. BjPutA structure analysis in this paper imply the importance of these two ion pairs in PRODH function again, but so far none of their mutagenesis and kinetic study have been performed in any PutAs, in order to fill this gap, we generated E197A and E460A mutants for biochemical and structural analysis.

5.4.8 Kinetic studies of the mutant enzymes.

The kinetic parameters of the wild-type BjPutA were determined first (Table 5.2), and which were similar with the kinetic data reported by Srivastava *et al*..[6] The values of k_{cat} and K_m using proline as the variable substrates with fixed concentration of CoQ1 as the electron acceptor are 2.2 s^{-1} and 30 mM, respectively, which gives a k_{cat}/K_m of 71.7 s^{-1} M^{-1}. For compares, the kinetic efficiency of the Type 2A GsPutA is listed here: 13 s^{-1} M^{-1} using menadione bisulfite as electron acceptor, and 7 s^{-1} M^{-1} using menadione as electron acceptor.[7]

Catalytic activity does not change much in mutants G196A and G196P (Table 5.2). Using proline as the varying substrate, k_{cat} value for G196A is 1.9 s^{-1}, which is similar to the wild-type enzyme; and the catalytic efficiency is 16.6 s^{-1} M^{-1}, which is only about 4-fold lower than that of wild-type enzyme. For G196P, a k_{cat} value of 0.14 s^{-1} was obtained, and the catalytic efficiency is estimated at 8.3 s^{-1} M^{-1}, which is 10-

fold lower than that of wild-type enzyme. While compared to the pervious DrPRODH study, the activity change of these two Gly mutants is pretty small. The catalytic efficiency for the corresponding Gly to Ala mutation is 140-fold lower than that of wild-type enzyme in DrPRODH.[14] These data suggest the flexibility of the β1−α1 loop does contribute to the PRODH function, but quite limited, which is consistent with the low flexibility profile observed from the structural data.

Proline dehydrogenase activity was severely impaired in mutants E197A and E460A (Table 5.2). Using proline as the variable substrate while keeping another substrate CoQ1 constant, the activity of both mutants is below detection limit. By mutating these two Glu that involved in the two ion pairs of Arg456-Glu197 and Arg457-Glu460, the enzyme's activity was largely affected; this confirms the two ion pair's importance in the function of PRODH.

5.4.9 Structures of the mutant enzymes.

The structures of the mutants G196A, G196P, E197A and E460A were determined to investigate the structural basis of their function in proline dehydrogenase and substrate channeling. They were crystallized in the same *C*2 space group as 3HAZ.[6]

No obvious conformational changes in the PRODH active site were observed in the two Gly mutants of G196A and G196P (Fig. 5.10C). The bulky density for the Gly196 site confirms the mutations are successfully (Fig 5.10A and 5.10B), but the mutations do not induce much conformational change of β1−α1 loop, which includes residues 195-199. The Arg456-E197 ion pair still exists as a sulfate ion is bound in the proline-binding site; and the two proline-binding residues of Arg456 and Arg457

occupy similar conformation as that in 3HAZ.[6] Only subtle difference is observed in the side chain of Glu460, but this may not be resulted from the Gly mutation, by taking its high flexibility into account. Very identical PRODH active site conformation to the wild-type enzyme were uncovered from the Gly196 mutants, this is consistent with the kinetic data that only small activity change are observed, and further confirmed the low dynamic aspect of the β1−α1 loop, where Gly196 locates.

Dramatic reconfiguration of the PRODH active site was noticed in the two Glu mutants of E197A and E460A. Both mutations break the helix structure of residues 458-462 from α8, while keeping the rest of the active site well folded as the wild-type enzyme. In E197A mutant, residues 459-462 shifts toward the FAD with Glu460 reside near the proline-binding site, and Arg457 occupies two conformations with one flipping away from the active site and hydrogen bond with the newly ordered Glu59. For E460A, residues 458-462 of α8 are disordered with very little density. These unfavored conformational changes are consistent with the structural data that residues 459-462 region on α8 is of high mobility. Also, this suggests that the two glutamates 197 and 460 play essential role in keeping a fit electrostatic environment for the flexible α8 region in its functional conformation. Furthermore, the disruption of the α8 helix structure is consistent with the kinetic data that these two Glu mutants play essential role in the function of proline dehydrogenase although they are not directly involved in substrate binding and catalysis.

5.5 DISCUSSION

The structures of BjPutAo-LAC, the first Type 1A PutA-substrate analog complex structure, revealed a similar substrate recognition mechanism as GsPutA[7] and mono-functional DrPRODH[14]. Upon substrate binding, two ion pairs of Arg456-Glu197 and

Arg457-Glu460 formed, together with the hydrogen bond between Arg457-ion pair with FAD, tightly bury the substrate between the si face of FAD and the α8. Also, these residues are conserved in PutA and PRODH sequence,[17,18] which suggests the proline-binding mechanism among PRODH family members, is highly conserved.

With the structures of all four states in the PRODH half-reduction getting resolved, different dynamic profile of the PRODH active site from the Type 1A BjPutA was observed. Unlike the large flexibility determined from GsPutA, smaller mobility from the backbone of α8 and quite low dynamic aspect from the β1−α1 loop were observed. The residues of the active site are highly conserved in all PutAs,[19] and the structures of BjPutA and GsPutA, which has only 27% sequence identity, further confirmed a highly conserved active site of PRODH domain. This raised an interesting question, that what is the functional variance resulted from the different dynamic profile in the proline dehydrogenase active sites - specifically the β1−α1 loop. A possible and reasonable hypothesis is, the different flexibility aspect of the β1−α1 loop would impact on the rate of P5C release from PRODH active site. A previous comprehensive kinetic mechanism study of *E. coli* PutA demonstrated the channeling step of PRODH domain has the slowest rate constant, and different transient time for the intermediates have been determined in PutAs. Our finding that the active sites of PRODH domain that with different flexibility, would provide a possible structural basis for the different substrate channeling rates among PutAs.

Unlike the rigid backbones, large dynamic aspect was observed from the side chains of the active site residues in BjPutA PRODH domain, especially from the ion pair of Arg456-Glu197. Based on the conformational changes determined from the PRODH reductive half-reaction, the iron pair of Glu197 and Arg456 constitutes an

active site gate that connects PRODH domain active site and the hydrolysis cavity, which leads to P5CDH active site. Upon substrate binding, this gate is closed and help providing a steric environment for the first step proline oxidation; FAD reduction triggers the dissociation of the iron pair, which opens the gate and allows the product P5C/GSA entering the hydrolysis cavity and channels to P5CDH active site.

The conformational changes in the PRODH half-reduction of BjPutA, for the first time, showed a complex and dynamic substrate-channeling pathway in Type 1A PutAs. No obvious change was observed in the parts of hydrolysis cavity and P5CDH active site, while multiple routes between PRODH active site and the hydrolysis cavity were determined. Route A appears upon the substrate pocket was occupied and the ion pair Arg456-Glu197 forms, Route B is found in both the resting and natural substrate proline-reduced enzyme.

As the dynamic tunnel system getting revealed from different states of the PRODH half-reduction, the real tunnel, which plays the role in transferring P5C from PRODH domain to the hydrolysis cavity, is specified. First consider the advantages of the ion pair of Arg456-Glu197: its dramatic dynamic aspect, gives its ability to open and close the route B; the high conservation among PutAs[17] implies its importance and universal function in the PutAs as a substrate-channeling gate; and the perfect location and tight interaction of Arg456 to proline via a salt bridge upon substrate binding, makes the ion pair a perfect candidate in transferring the product. Next the structure of BjPutAr-LAC provides a snapshot that, after flavin reduction, the tight salt bridge between Arg456 and product does not break, but the interactions on the side of route A do not exist any more, thus a motion of product with the flip of the side chain from Arg456 toward the hydrolysis cavity is quite reasonable and

convincing. Furthermore, our structures showed that the tunnel of route B opens while route A getting closed in the resting and the proline-reduced enzyme. Taking them together, we conclude that the route B is the real substrate tunnel where P5C get transferred from the active site of PRODH to the hydrolysis cavity. Even lower dynamic profile are observed on the backbone of the PRODH active site, the large dynamic aspect from ion gate of Arg456-Glu197 in BjPutA is enough to open and close the route B, and it resembles a identical substrate-channeling pathway as that of GsPutA, which suggests a universally common substrate-channeling mechanism in PutAs: gating.

Finally, a recent kinetic analysis on EcPutA showed the activity of PRODH domain increases after several turnovers,[20] establishing the hysteresis character of PutA - while the structural evidence has never been found, leaving the structural mechanism a main question in PutA. Our results in this paper, first time showed that the backbone of α8, specifically the region including residues 458-462, is the only active site component of PRODH domain that with high B factor from BjPutA. And the single Glu mutation exclusively induces the region 458-462 of α8 becoming disordered while keeping the rest of the active site well folded. This obviously higher dynamic aspect of α8 could be the potential key to explain the hysteresis phenomenon in PutAs: it could be disordered in a non-functional state; after several turnovers of substrate-binding and FAD reduction, the ordered level of α8 would go up, which resulted in a phenomenon that the PRODH activity goes higher, the hysteresis. However, to test this hypothesis, more structural and biochemical studies are required on a good model of hysteretic PutA.

5.6 REFERENCES

1. Adams, E., and Frank, L. (1980) Metabolism of proline and the hydroxyprolines, *Annu. Rev. Biochem. 49*, 1005-1061.

2. Krishnan, N., Doster, A. R., Duhamel, G. E., and Becker, D. F. (2008) Characterization of a Helicobacter hepaticus putA mutant strain in host colonization and oxidative stress. Infect. Immun. 76, 3037− 3044.

3. Bringaud, F., Riviere, L., and Coustou, V. (2006) Energy metabolism of trypanosomatids: Adaptation to available carbon sources. Mol. Biochem. Parasitol. 149, 1−9.

4. Tanner, J. J., and Becker, D. F. (2013) PutA and proline metabolism. In Handbook of Flavoproteins (Hille, R., Miller, S. M., and Palfey, B., Eds.) pp 31−56, Walter de Gruyter, Boston.

5. Luo M, Christgen, S., Sanyal, N, Arentson BW, Becker DF, Tanner JJ. (2014) Evidence that the C-terminal Domain of a Type B PutA Protein Contributes to Aldehyde Dehydrogenase Activity and Substrate Channeling. Biochemistry. 53(35):5661-73

6. Srivastava, D., Schuermann, J. P., White, T. A., Krishnan, N., Sanyal, N., Hura, G. L., Tan, A., Henzl, M. T., Becker, D. F., and Tanner, J. J. (2010) Crystal structure of the bifunctional proline utilization A flavoenzyme from Bradyrhizobium japonicum, Proceedings of the National Academy of Sciences of the United States of America 107, 2878-2883.

7. Singh, H., Arentson, B. W., Becker, D. F., and Tanner, J. J. (2014) Structures of the PutA peripheral membrane flavoenzyme reveal a dynamic substrate-channeling tunnel and the quinone-binding site, Proceedings of the National Academy of Sciences of the United States of America 111, 3389-3394.

8. Tanner, J. J. (2008) Structural biology of proline catabolism, Amino acids 35, 719-730.

9. Arentson BW, Luo M, Pemberton TA, Tanner JJ, Becker DF. (2014) Kinetic and Structural Characterization of Tunnel-Perturbing Mutants in Bradyrhizobium japonicum Proline Utilization A (PutA). Biochemistry 12; 53(31):5150-61.

10. Surber, M. W., and Maloy, S. (1998) The PutA protein of Salmonella typhimurium catalyzes the two steps of proline degradation via a leaky channel. Arch. Biochem. Biophys. 354, 281−287.

11. Rajarshi Maiti, Gary H. Van Domselaar, Haiyan Zhang, and David S. Wishart. (2014) "SuperPose: a simple server for sophisticated structural superposition" Nucleic Acids Res. 32 (Web Server issue): W590W594.

12. Inagaki, E., Ohshima, N., Takahashi, H., Kuroishi, C., Yokoyama, S. & Tahirov, T. H. (2006). Crystal structure of Thermus thermophilus Delta1-pyrroli- ne-5-carboxylate dehydrogenase. J. Mol. Biol. 362, 490–501.

13. Moxley MA, Becker DF (2012) Rapid reaction kinetics of proline dehydrogenase in

the multifunctional proline utilization A protein. Biochemistry 51(1):511–520.

14. Luo, M., Arentson, B. W., Srivastava, D., Becker, D. F., and Tanner, J. J. (2012) Crystal structures and kinetics of monofunctional proline dehydrogenase provide insight into substrate recognition and conformational changes associated with flavin reduction and product release, Biochemistry 51, 10099-10108.

15. Petrek M, Kosinova P, Koca J, Otyepka M (2007) MOLE: A Voronoi diagram-based explorer of molecular channels, pores, and tunnels. Structure, 15(11):1357–1363.

16. M. Zhang, T.A. White, J.P. Schuermann, B.A. Baban, D.F. Becker, and J. J.Tanner. (2004) Structures of the Escherichia coli PutA proline dehydrogenase domain in complex with competitive inhibitors. Biochemistry. 43(39) 12539-48.

17. T.A. White, N. Krishnan, D.F. Becker and J.J. Tanner. (2007) Structure and Kinetics of Monofunctional Proline Dehydrogenase from Thermus thermophilus. J. Biol. Chem. 282(19):14316-14327.

18. Krishnan, N., and Becker, D. F. (2005) 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 44, 9130-9139.

19. R. K. Singh and J.J. Tanner. (2012) Unique structural features and sequence motifs of proline utilization A (PutA). Frontiers in Bioscience 17, 556-568.

20. Moxley, M. A., Sanyal, N., Krishnan, N., Tanner, J. J., and Becker, D. F. (2014) Evidence for hysteretic substrate channeling in the proline dehydrogenase and Delta1-pyrroline-5-carboxylate dehydrogenase coupled reaction of proline utilization A (PutA), The Journal of biological chemistry 289, 3639-3651.

21. Engh, R. A., and Huber, R. (1991) Accurate bond and angle parameters for x-ray protein structure refinement, *Acta Cryst. A47*, 392-400.

22. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation, *Proteins 50*, 437-450.

Table 5-1 BjPutA X-ray Diffraction Data Collection and Refinement[a]

	BjPutAo-Lactate	BjPutAr-Lactate	BjPutAr	BjPutA R456M
Beamline	ALS 4.2.2	ALS4.2.2	ALS4.2.2	ALS4.2.2
Space group	*C*2	*C*2	*C*2	*C*2
Unit cell parameters (Å)	α = 167.52	α = 166.92	*a* = 167.8	*a* = 167.2
	b = 195.80	*b* = 195.10	*b* = 196.1	*b* = 195.2
	c = 108.52	*c* = 108.54	*c* = 109.0	*c* = 109.2
	β = 121.75	β = 121.43	β = 121.7	β = 121.5
Wavelength	1.00000	1.00000	1.0000	1.0000
Resolution (Å)	31.83-2.30	33.76-2.30	37.84 – 2.20	29.96 – 2.03
Observations	489561	477185	461836	708040
Unique reflections	131364	130820	138636	190232
$R_{merge}(I)$	0.123(0.444)	0.177(0.363)	0.091 (0.480)	0.124 (0.589)
$R_{meas}(I)$	0.143(0.522)	0.208(0.439)	0.109 (0.648)	0.146 (0.723)
$R_{pim}(I)$	0.074(0.272)	0.109(0.243)	0.058 (0.430)	0.076 (0.410)
Mean I/σ	7.4(2.8)	6.5(4.6)	8.4 (2.0)	6.5 (2.5)
Completeness (%)	99.9(99.9)	99.9(99.5)	91.7 (63.0)	99.1 (93.9)
Multiplicity	3.7(3.6)	3.6(3.2)	3.3 (2.1)	3.7 (3.2)
No. of protein residues			1938	1941
No. of protein atoms			14631	14681
No. of FAD atoms			106	106
No. of ligand atoms			80	96
No. of water molecules			706	980
R_{cryst}			0.2300 (0.3559)	0.2217 (0.3856)
R_{free}[b]			0.2681 (0.3693)	0.2634 (0.4165)
rmsd bond lengths, Å[c]			0.010	0.007
rmsd bond angles, °			1.392	1.046
Ramachandran plot[d]				
Favored (no. residues)			1883	1891
Outliers (no. residues)			6	12
Average B-factor (Å^2)				
Protein			28.14	26.72
FAD			23.92	18.65
Water			37.75	45.71
Coordinate error (Å)[e]			0.35	28.86

Table 5-2 BjPutA Mutants X-ray Diffraction Data Collection and Refinement[a]

	G196P	G196A	E197A	E460A
Beamline	ALS	ALS	ALS	ALS
Space group	*C*2	*C*2	*C*2	*C*2
Unit cell parameters (Å)	α = 166.9	*a* = 167.4	*a* = 167.3	*a* = 166.5
	b = 195.7	*b* = 196.1	*b* = 196.0	*b* = 195.8
	c = 108.8	*c* = 109.2	*c* = 109.0	*c* =108.7
	β = 121.7	β = 121.5	β = 121.4	β = 121.4
Wavelength	1.00000	1.00000	1.0000	1.0000
Resolution (Å)	62.05-2.04	62.21-2.10	62.15 – 2.10	61.97 – 2.30
Observations	696584	650417	642493	486982
Unique reflections	186324	173918	171782	130904
$R_{merge}(I)$	0.075(0.321)	0.137(0.496)	0.086 (0.611)	0.072 (0.509)
$R_{meas}(I)$	0.088(0.390)	0.160(0.586)	0.119 (0.839)	0.099 (0.698)
$R_{pim}(I)$	0.045(0.218)	0.082(0.308)	0.082 (0.572)	0.068 (0.476)
Mean I/σ	11.2(3.2)	5.9(2.1)	9.4 (1.8)	11.6 (2.1)
Completeness (%)	99.1(90.1)	99.9(99.1)	98.8 (98.3)	99.7 (99.8)
Multiplicity	3.7(3.1)	3.7(3.5)	3.7 (3.5)	3.7 (3.5)
No. of protein residues				
No. of protein atoms			14372	14285
No. of FAD atoms			106	106
No. of ligand atoms			0	0
No. of water molecules			692	
R_{cryst}			0.1926	0.2238
R_{free}[b]			0.2239	0.2618
rmsd bond lengths, Å[c]			0.009	0.009
rmsd bond angles, °			1.080	1.145
Ramachandran plot[d]				
Favored (no. residues)				
Outliers (no. residues)				
Average B-factor (Å^2)				
Protein			35.15	43.315
FAD			26.47	37.017
Water			36.77	-

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]A common test set (5 %) was used for refinement of both structures.
[c]Compared to the parameters of Engh and Huber.[21]
[d]The Ramachandran plot was generated with RAMPAGE.[22]
[e]Maximum likelihood-based coordinate error estimate.

Table 5-3 Kinetic Parameters for BjPutA and BjPutA Mutant Enzymes Using Proline and CoQ1 as the Substrates

	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} M^{-1})	Variational Effect (k_{cat}/K_m of mutant)/ (k_{cat}/K_m of BjPutA)
		Proline as the variable substrate[a]		
BjPutA	2.18 ± 0.12	30.33 ± 5.88	71.73 ± 14.48	
G196A	1.92 ± 0.12	115.72 ± 15.30	16.55 ± 2.43	0.231 ± 0.058
G196P	0.14 ± 0.004	16.45 ± 1.96	8.25 ± 1.04	0.115 ± 0.027
E197A	BD[b]	BD	BD	BD
E460A	BD	BD	BD	BD

[a][CoQ1] was fixed at 100 μM.
[b]Below detection limit of assay.

FIGURE LEGENDS

Figure 5-1. (A) Reactions catalyzed by PRODH and P5CDH. (B) Scheme showing the enzyme states that occur during the PRODH reductive half-reaction.

Figure 5-2. Domain diagrams representing the PutA classification.

Figure 5-3. Substrate channeling pathway calculated from MOLE 2.0. The three parts of the pathway is named PRO site, hydrolysis cavity and P5C site. Route A found in 3HAZ is represented in hot pink surface; Route B calculated by cutting the side chain of Arg456-Glu197 ion pair was shown with blue surface. The β1−α1 loop and α8 from PRODH are drawn with cartoon in cyan, FAD is shown with yellow stick; and the catalytic Cys792 of P5CDH is drawn in green stick.

Figure 5-4. Ribbon drawing of the protomer. The domains are colored as follows: N-terminal arm, yellow; alpha domain, green; PRODH domain, cyan; linker, magenta; and the three domains of P5CDH: catalytic domain, NAD+-binding domain and oligomerization domain are colored in red, blue and orange respectively. FAD is represented in yellow sphere, and the catalytic residue Cys792 of P5CDH is drawn as sphere in green. The hot pink and blue surface represents the internal cavity shown in Figure 5.3 that connects the two active sites.

Figure 5-5. PRODH active site of the four enzyme states during the PRODH reductive half-reaction. The cage represents a simulated annealing σ_A-weighted Fo − Fc omit map (2.5σ). Lactate is represented with stick in magenta, and the interactions between the enzyme and lactate are shown with black doted line; The proline binding site is roughly drawn with red circle with pro labeled; FAD in its oxidized state is drawn with yellow stick; FAD in its reduced state is shown with grey stick. (A)

BjPutAo-LAC (Eo-PRO); (B) BjPutAr-LAC complex (Er-P5C); (C) BjPutAr (Er); (D) BjPutA mutant of R456M (Eo), which is a mimic of the resting enzyme.

Figure 5-6. Electron density map for reduced flavin. The gray cage represents a simulated annealing σ_A-weighted Fo − Fc omit map (2.5σ); Reduced flavin is shown with gray sticks, for compares, oxidized flavin is represented with yellow sticks; The bending of the isoalloxazine ring is shown in the right.

Figure 5-7. Superposition of the PRODH active site from different enzyme states in its reductive half-reduction: BjPutA R456M (Eo, grey), the BjPutAo-LAC complex (Eo-PRO, magenta), the BjPutAr-LAC complex (Er-P5C, green) and proline-reduced BjPutA (Er, yellow). β1−α1 loop and α8 are presented with cartoon, the rest part of the enzyme is shown in ribbon; FAD is drawn with sticks, and colored the same to their corresponding enzyme; Lactate in the BjPutAo-LAC complex is shown with sticks in magenta, lactate from the BjPutAr-LAC complex is represented with yellow sticks. No movement was observed from β1−α1 loop, and the largest shift of α8 is observed in its C-terminus end, which is about 1.2 Å.

Figure 5-8. Surface rendering of the tunnel routes that connects PRO site and hydrolysis cavity, the trace of the main pathway is shown with red doted line. FAD is represented with sticks, the oxidized and reduced form is colored in yellow and grey, respectively; β1−α1 loop and α8 are drawn with cartoon, the size of cartoon represents B factor (mobility), more bulky means higher B factor. (A) R456M of BjPutA (Eo); (B) The BjPutAo-LAC (Eo-PRO); (C) Proline-reduced BjPutA (Er). Note that the ion pair is broken in A and C, and LAC of A and C are not present in their structures, which is modeled there to indicate the proline-binding site.

Figure 5-9. P5C transfer mechanism of PRODH domain. (A) Superposition of BjPutAr-LAC (a mimic of Er-P5C, colored in yellow) and proline-reduced BjPutA (Er, colored in green). The route B represented in blue surface is the hydrolysis cavity calculated from proline-reduced BjPutA (Er); LAC shown in magenta sticks is from the complex of Eo-LAC; LAC (yellow sticks) is from the complex of Er-P5C, and its interactions with enzyme is labeled. Note that upon the reduction of flavin, LAC binds more tightly to the route B side, while all the interactions on the side of route A are broken; (B) Surface rendering of the tunnel that connects PRO site and hydrolysis cavity in the structure of BjPutAr-LAC. Note the route A is found in this state.

Figure 5-10. PRODH active site view of the two mutants of G196A and G196P. (A) G196A mutant, the cage represents a simulated annealing σ_A-weighted Fo − Fc omit map (2.5σ); (B) G196P mutant, the cage represents a simulated annealing σ_A-weighted Fo − Fc omit map (2.5σ); (C) Superposition of structures of 3HAZ (grey), G196A (magenta), and G196P (green). Note that no significant conformation changes were observed from the two Gly mutants.

Figure 5-1 (A) Reactions catalyzed by PRODH and P5CDH. (B) Scheme showing the enzyme states that occur during the PRODH reductive half-reaction.

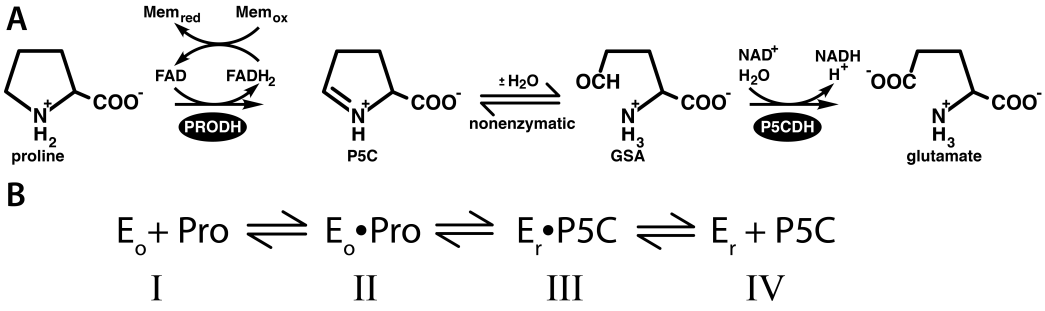


Figure 5-2 Domain diagrams representing the PutA classification.

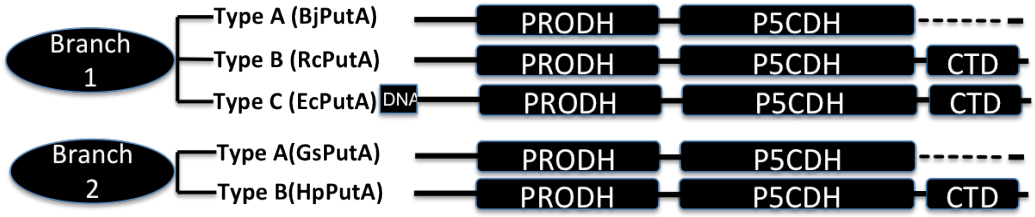


Figure 5-3 Substrate channeling pathway calculated from MOLE 2.0.

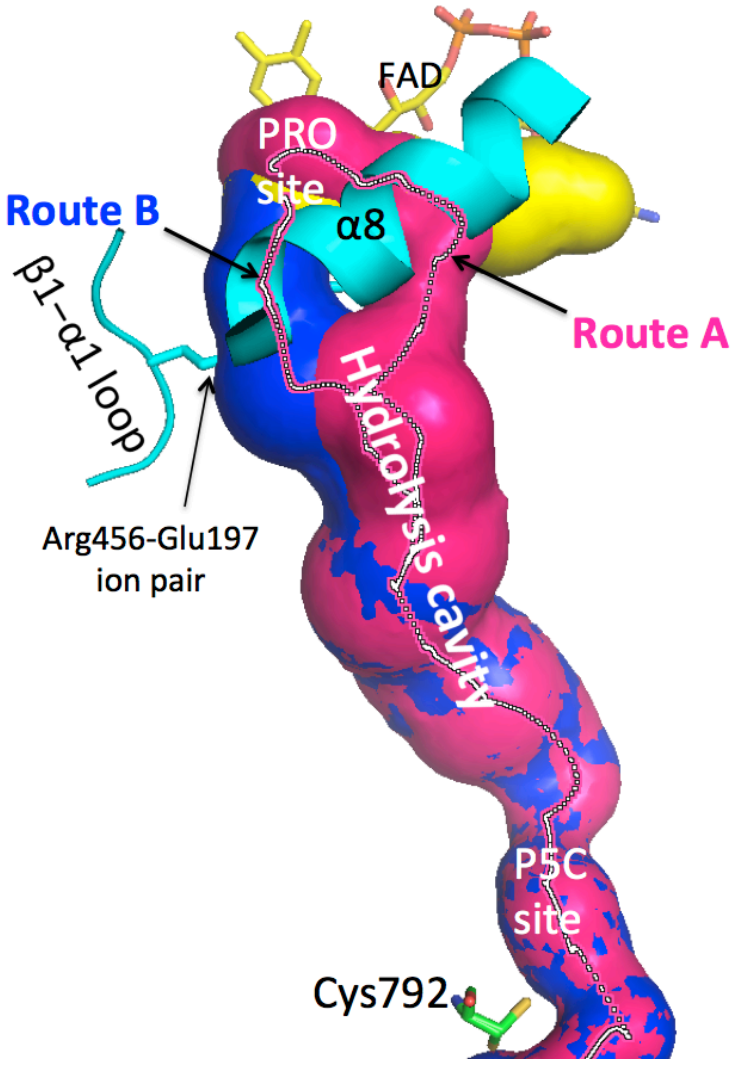


Figure 5-4 Ribbon drawing of the protomer.

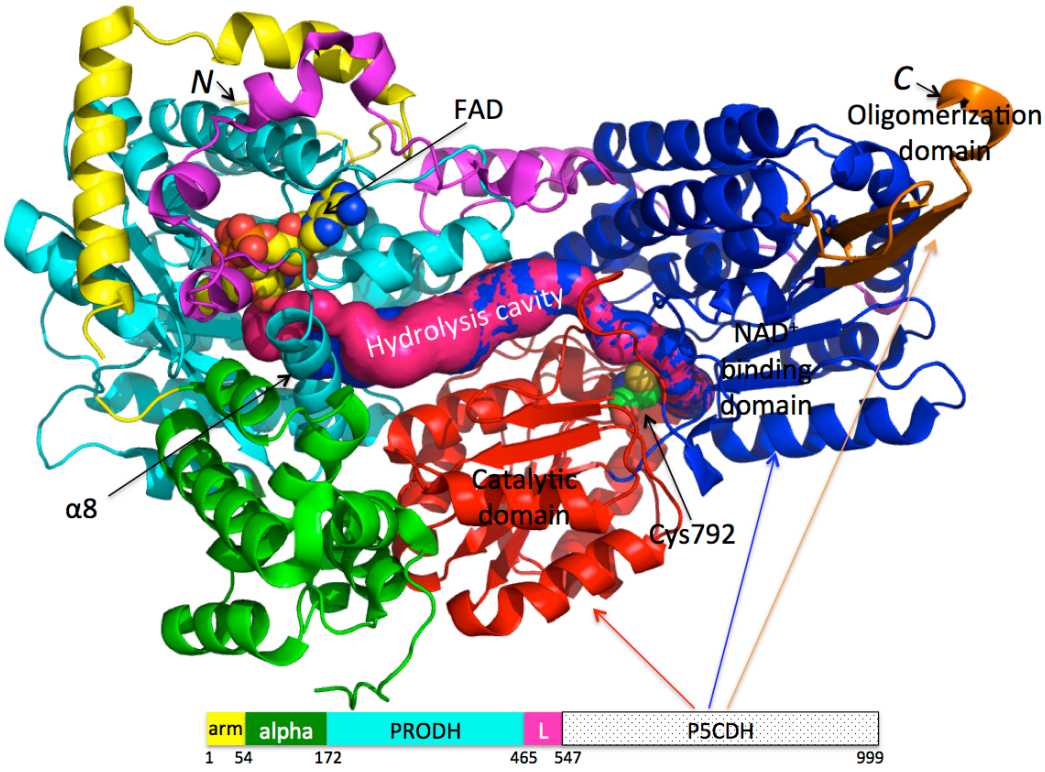


Figure 5-5 PRODH active site of the four enzyme states during the PRODH reductive half-reaction.

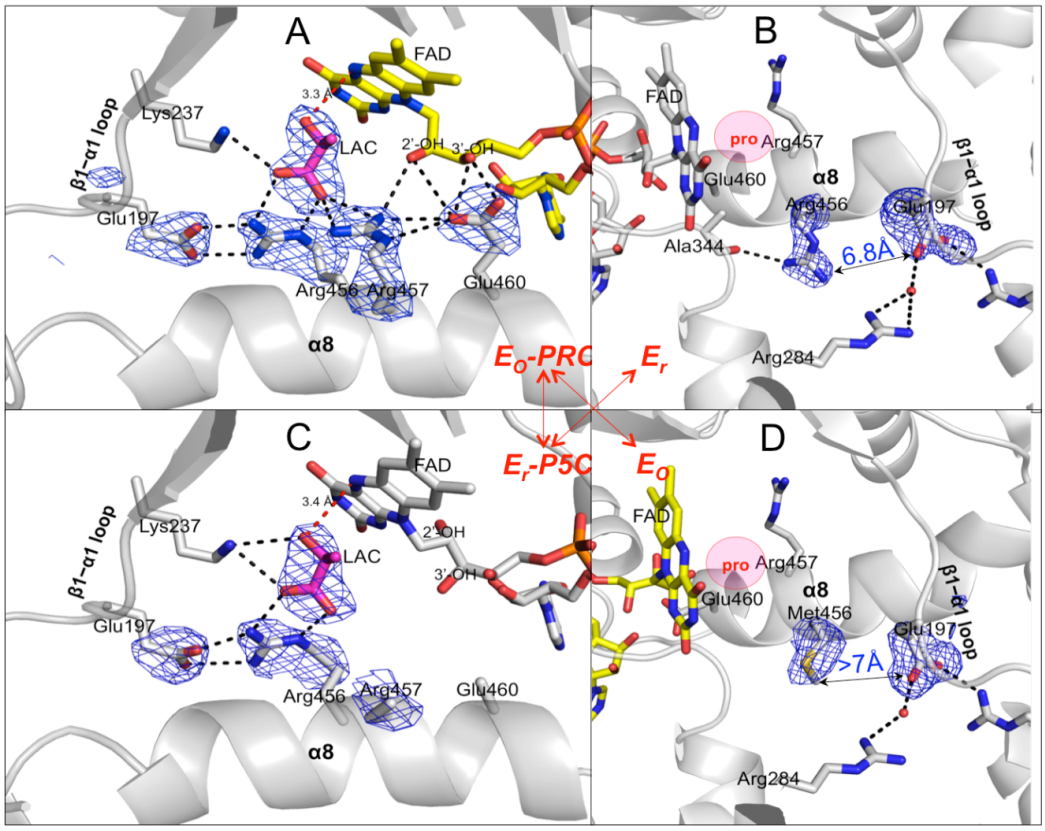


Figure 5-6 Electron density map for reduced flavin.

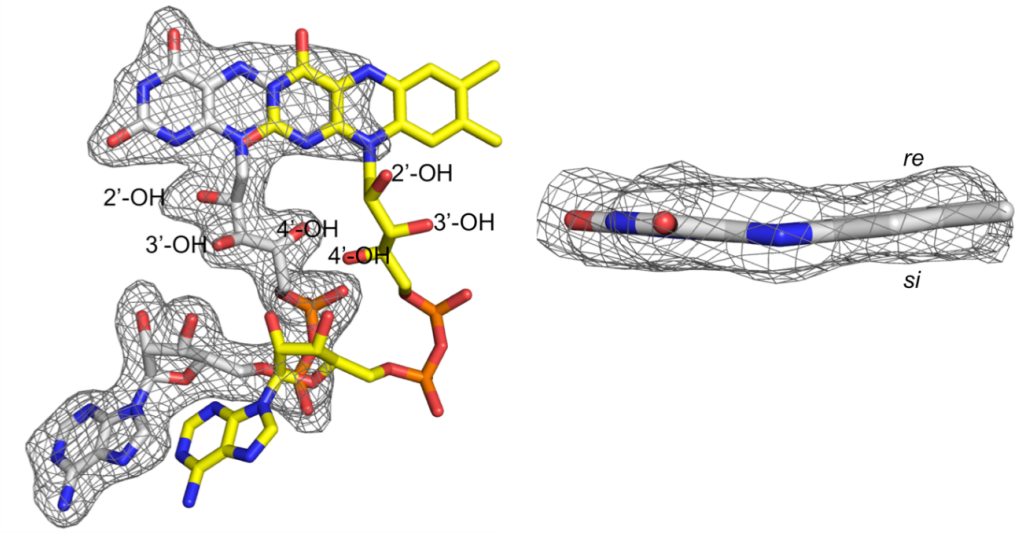


Figure 5-7 Superposition of the PRODH active site from different enzyme states in its reductive half-reduction

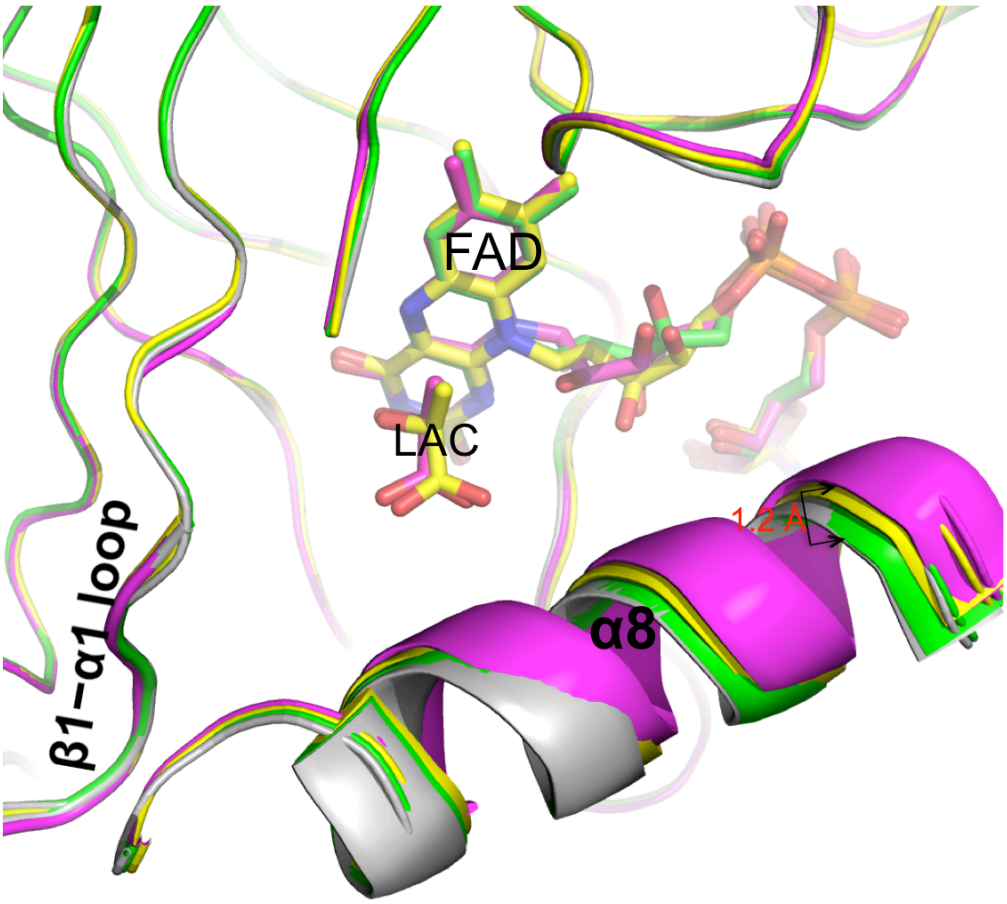


Figure 5-8 Surface rendering of the tunnel routes that connects PRO site and hydrolysis cavity

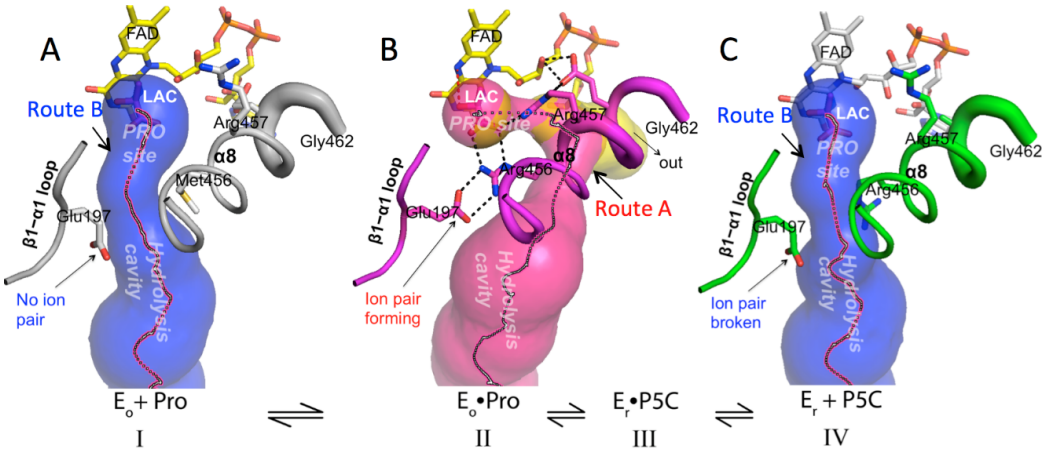


Figure 5-9 P5C transfer mechanism of PRODH domain.

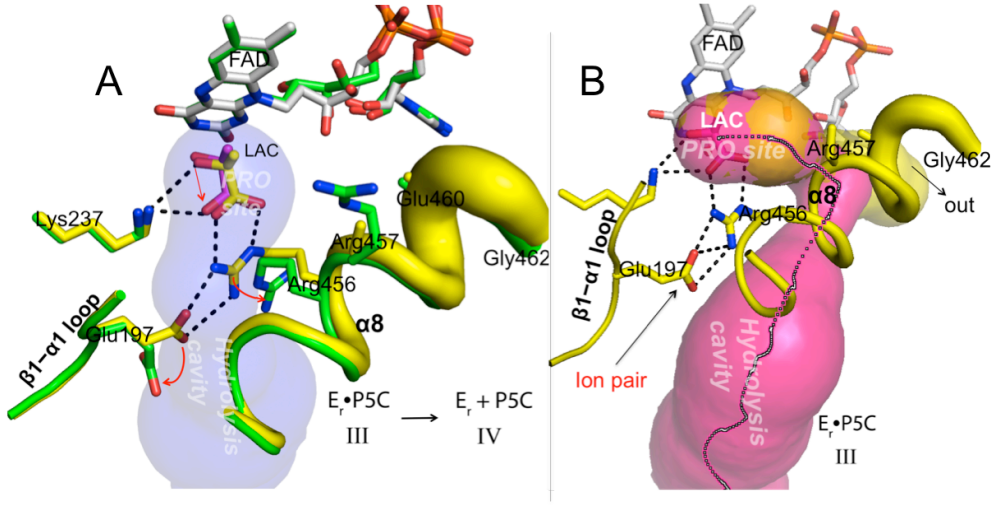
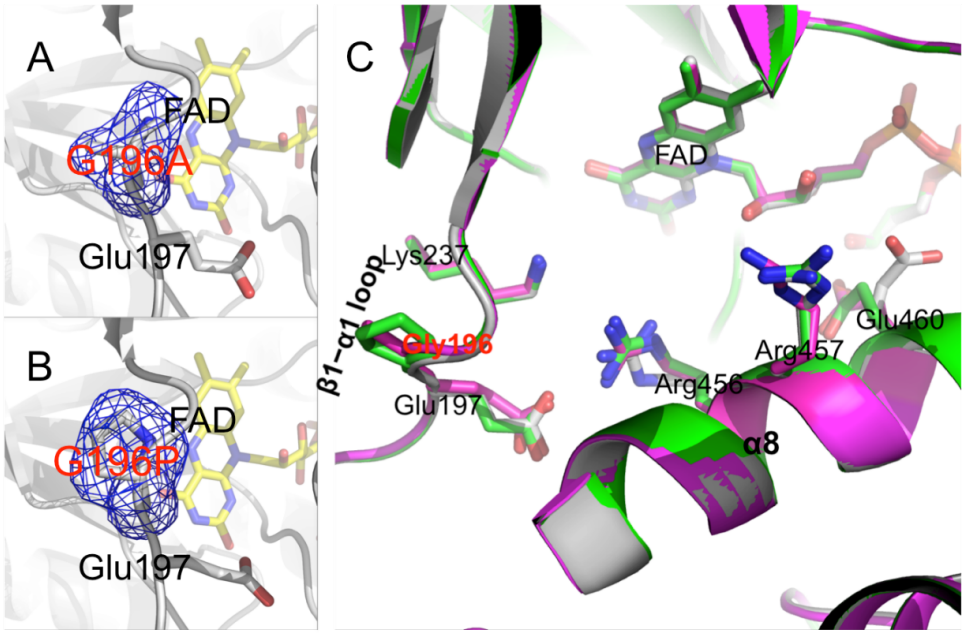


Figure 5-10 PRODH active site view of the two mutants of G196A and G196P.



CHAPTER 6

6. DIETHYLAMINOBENZALDEHYDE IS A COVALENT, IRREVERSIBLE INACTIVATOR OF ALDH7A1

Luo M., et al. Diethylaminobenzaldehyde is a Covalent, Irreversible Inactivator of ALDH7A1. ACS Chem. Biol. (2015) Jan. 2. [Epub ahead of print]

6.1 ABSTRACT

There is growing interest in aldehyde dehydrogenases (ALDHs) because of their overexpression in cancer stem cells and ability to mediate resistance to cancer drugs. Here we report the first crystal structure of an aldehyde dehydrogenase complexed with the inhibitor 4-diethylaminobenzaldehyde (DEAB). Contrary to the widely held belief that DEAB is a reversible inhibitor of ALDHs, we show that DEAB irreversibly inactivates ALDH7A1 via formation of a stable, covalent acyl-enzyme species.

6.2 INTRODUCTION

Aldehyde dehydrogenases catalyze the $NAD(P)^+$-dependent oxidation of aldehydes to carboxylic acids. In humans, the aldehyde dehydrogenase (ALDH) superfamily comprises 19 enzymes, which process a diverse array of substrates, including small aldehydes such as acetaldehyde, amino acid derivatives including aminoadipate semialdehyde and glutamate semialdehyde, and large lipids like retinaldehyde and fatty aldehydes such as octanal. ALDHs are important in detoxification of reactive aldehydes, amino acid metabolism, embryogenesis and development, neurotransmission, oxidative stress, and cancer.[1] Consistent with the diversity of substrates and widespread tissue distribution of ALDHs, mutations in ALDH genes are associated with numerous inherited metabolic disorders.[1]

ALDHs have also emerged as important biomarkers of cancer stem cells (CSCs) and mediators of cancer cell resistance to chemotherapeutic agents.[2-5] First discovered in acute myeloid leukemia,[6] CSCs constitute a subpopulation of cells within a heterogeneous tumor and have enhanced tumor-initiating potential because of their ability to regenerate and initiate metastasis. According to the CSC model, conventional cancer therapies reduce tumor mass by killing non-tumorigenic cells, but leave behind CSCs, resulting in eventual recurrence. The most common method of identifying CSCs is flow cytometry detection of cells expressing high levels of ALDH activity (so-called ALDH bright cells) using fluorescent aldehyde substrates, such as the ALDEFLUOR reagent.[7, 8] Overexpression of ALDH1A1, ALDH1A3, ALDH2, ALDH4A1, and ALDH7A1 isoforms has been found in several types of cancers.[9] Because of the functional involvement of ALDHs in CSCs, the correlation between

ALDH bright cells and poor clinical outcomes, and the role of ALDHs in cancer drug resistance mechanisms, there is a need for isoform-specific ALDH inhibitors that can be used to probe the various roles of ALDHs in CSCs and as leads in drug development.[5, 10]

4-diethylaminobenzaldehyde (DEAB) is a well-known, but poorly characterized, ALDH inhibitor. DEAB was first identified as an ALDH inhibitor over 25 years ago when the inhibition of ALDH-mediated drug metabolism by DEAB was shown to sensitize cancer cells to the cytotoxic action of cyclophosphamide.[11] More recently, DEAB has been shown to increase the sensitivity of ALDH bright human breast cancer cells to paclitaxel and doxorubicin.[12] DEAB and the related compound 4-dipropylaminobenzaldehyde have been described as reversible competitive inhibitors of ALDH1 (competitive with the aldehyde substrate) with K_i values of 0.04 μM and 0.01 μM, respectively.[11, 13] Although early work hinted at specificity toward ALDH1A1, more recent studies suggest that DEAB may be a broad inhibitor of ALDHs.[14] Understanding the mechanism and specificity of DEAB is important because DEAB has become an important reagent in cancer research. For example, DEAB is used extensively in cancer drug resistance studies.[5, 12, 14-16] Furthermore, DEAB is employed as an allegedly ALDH1A1-specific inhibitor in the widely used ALDEFLUOR CSC detection kit.[7] Despite its longstanding and widespread use in medicinal chemistry and cancer biology, the structural basis for the inhibition of ALDH by DEAB has remained unknown.

We herein describe the structural basis of the inhibition of ALDH7A1 by DEAB. ALDH7A1 plays a role in lysine catabolism by catalyzing the NAD^+-dependent

oxidation of α-aminoadipate semialdehyde to α-aminoadipate. High expression of ALDH7A1 has been found in prostate cancer cell lines, prostate cancer tissue, and matched bone metastasis samples, suggesting that ALDH7A1 plays a functional role in prostate cancer bone metastasis.[17] In addition, patients with ALDH7A1-expressing non-small-cell lung carcinoma tumors have a significantly increased incidence of lung cancer recurrence,[18] and ALDH7A1 expression is increased in ovarian tumors.[19]

6.3 METHODS

6.3.1 Subcloning, expression, and purification of ALDH7A1.

A plasmid containing the cDNA of human cytosolic ALDH7A1 isoform 2 (NCBI RefSeq NP_001188306.1) was purchased from Thermo-Fisher (catalog number MHS6278-202758773, Mammalian Gene Collection CloneId 5518212). The cDNA was subcloned into pKA8H between *NdeI* and *BamHI* sites, and the sequence was verified by the University of Missouri DNA Core. The expressed protein includes an N-terminal His_8 tag and Tobacco Etch Virus Protease (TEVP) cleavage site. Treatment with TEVP results in Gly-His followed by the ALDH7A1 polypeptide.

ALDH7A1 was expressed in *Escherichia coli* BL21(DE3). A 10 mL overnight culture was used to inoculate 1000 mL of LB media. The culture was then grown at 37 °C at 250 rpm and induced with 0.2 mM IPTG when the OD_{600} reached 0.5. Induction was performed for 12 hours at 18 °C at 200 rpm. Cells were harvested and frozen in 50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, and 5% (v/v) glycerol.

The enzyme was purified using affinity and ion exchange chromatography as follows. The frozen cells were thawed in the presence of 0.4 mM protease inhibitor

phenylmethylsulfonyl fluoride and 0.5% (v/v) Triton X-100 detergent and broken by sonication. Cell debris and unbroken cells were separated by centrifugation at 16,500 rpm for one hour in an SS34 rotor. The supernatant was applied to a Ni^{2+}-charged His-Trap HP column (GE Healthcare) equilibrated with 50 mM Hepes, 300 mM NaCl, and 5% (v/v) glycerol at pH 8.0. ALDH7A1 was eluted with equilibration buffer containing 250 mM imidazole. The fractions were pooled, and TEVP and 20X TEV buffer (1 M Hepes, 10 mM EDTA, pH 8.0) were added so that the final solution contained 1 mg of TEVP per 10 mg of protein in 50 mM Hepes and 0.5 mM EDTA at pH 8.0. The sample was incubated at 28 °C for 2 hours, then dialyzed overnight at 4 °C, and finally injected onto the His-Trap HP column. The untagged ALDH7A1 was collected at 10 mM imidazole and dialyzed into 50 mM Hepes, 5% (v/v) glycerol, 0.5 mM EDTA, at 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. The column buffer for enzyme to be used in kinetics experiments contained 50 mM Hepes and 100 mM NaCl at pH 7.8. The pooled fractions were supplemented with NAD^+ to a final concentration of 1 mM and stored at 4 °C. The column buffer used for enzyme destined for crystallization trials was 50 mM Hepes pH 7.8, 100 mM NaCl, 1 mM dithiothreitol, and 5% (v/v) glycerol. The pooled fractions were supplemented with NAD^+ to a final concentration of 1 mM, dialyzed against pre-crystallization buffer containing 50 mM Tris pH 7.8, 0.5 mM Tris(hydroxypropyl)phosphine, 50 mM NaCl, and 5% (v/v) glycerol, and stored at –

80 °C.

6.3.2 Crystallization of apo ALDH7A1 in space group *C*2.

Centered monoclinic crystals of apo ALDH7A1 were grown for use as microseeds in other crystallization experiments. These crystals were grown in sitting drops (Crystal*EX* 96 well plates from Hampton Research) at 22 °C using drops formed by mixing 3 μL each of the protein and reservoir solutions. The protein stock solution contained ALDH7A1 at 3 mg mL^{-1} in the pre-crystallization buffer. The reservoir contained 0.2 M ammonium sulfate, 20% (w/v) polyethylene glycol (PEG) 3350, and 0.1 mM Bis-Tris pH 6.5, which was obtained by optimizing a condition identified in crystal screening trials using commercially available kits (Hampton Research). The space group is *C2* with unit cell dimensions a = 154 Å, b = 163 Å, c = 159 Å, β = 95°. We note that this is the same crystal form that was reported by Brocker *et al.* for ALDH7A1 complexed with NADH.

6.3.3 Crystallization of ALDH7A1 inactivated by 4-diethylaminobenzaldehyde in space group *F*222.

4-Diethylaminobenzaldehyde (DEAB) was purchased from Sigma-Aldrich (catalogue number D86256). Soaking of apo ALDH7A1 crystals with DEAB was not successful, so cocrystallization experiments were performed as follows. A stock solution of 100 mM DEAB in DMSO was prepared. DEAB was added to the enzyme solution (3 mg mL^{-1} in pre-crystallization buffer) to a final concentration ~200 μM DEAB. The enzyme - DEAB mixture was incubated overnight and then used in matrix microseeding crystal screening trials using commercially available kits (Hampton Research). The microseed stock was prepared by crushing crystals of apo

ALDH7A1 in a buffer of 0.2 M ammonium sulfate, 20% (w/v) PEG 3350, and 0.1 mM Bis-Tris pH 6.5. Seeds were added to the drops by streaking with a horse tail hair. A hit was obtained with a reservoir solution containing NaBr and 30% (w/v) PEG monomethyl ether 2000. These crystal clusters were crushed and used as microseeds for another round of crystal screening, which resulted in several hits. Optimization of these hits resulted in diffraction quality crystals grown in sitting drops using a reservoir solution of 0.2 M $MgCl_2$, 25% (w/v) PEG 3350, 0.1 M Tris at pH 8.2, 1% (v/v) DMSO, and 200 μM DEAB. The crystals were cryoprotected in the reservoir solution supplemented with 15 % (v/v) ethylene glycol, picked up with Hampton nylon loops, and plunged into liquid nitrogen. The space group is *F*222 with unit cell dimensions of a = 161 Å, b = 219 Å, and c = 233 Å. The asymmetric unit includes four protomers arranged as two dimers. Each dimer is half of a tetramer that is built from crystallographic symmetry. The estimated solvent content[2] and V_M are 47 % and 2.3 $Å^3$/Da, respectively.

6.3.4 Crystallization of DEAB-inactivated ALDH7A1 complexed with NAD^+ in space group *I*422.

ALDH7A1 was cocrystallized with DEAB and NAD^+. The enzyme-DEAB mixture was made as described in the preceding paragraph, except that NAD^+ was included at 5 mM. Crystal screens using the ALDH7A1-DEAB-NAD^+ mixture were performed as described for ALDH7A1-DEAB crystallization, except that the microseed stock included both the *C*2 form apo enzyme crystals and the *F*222 enzyme-DEAB crystals. Several hits were found during matrix microseeding, and the best diffraction quality crystals were grown in the reservoir solution of 0.1 M magnesium formate dihydrate, 15% (w/v) PEG 3350, and 0.1 M Hepes pH 7.5. The

crystals were cryoprotected in the reservoir solution supplemented with 22 % (v/v) ethylene glycol, picked up with Hampton nylon loops, and plunged into liquid nitrogen. The space group is *I*422 with unit cell dimensions of $a = b = 160.3$ Å, and $c = 320.5$ Å. The asymmetric unit contains one tetramer (Figure 6.1b), which implies 47 % solvent and V_M of 2.3 $Å^3$/Da.

6.3.5 Mass spectrometry.

A solution of ALDH7A1 at 2 mg mL^{-1} was first dialyzed versus 0.1 M ammonium acetate at pH 7.2, and then DEAB (100 μM) and NAD^+ (2 mM) were added to an aliquot of the dialyzed protein. The samples were then analyzed at the MU Proteomics core using Nano-LC Nanospray QTOF MS (Agilent 6520A QTOF MS).

6.3.6 Kinetics of the time-dependent inactivation of ALDH7A1 by DEAB.

ALDH activity was measured by monitoring NADH production at 340 nm using hexanal as the substrate. Hexanal was purchased from Sigma-Aldrich and dissolved in 0.5% (v/v) DMSO at a concentration of 40 mM. The activity assay volume was 1 mL; the assay buffer was 100 mM sodium pyrophosphate buffer at pH 8. This assay was used to determine a K_m for hexanal of 35 μM.

A fresh 0.1 mM stock solution of DEAB was prepared in 100 mM sodium pyrophosphate buffer (pH 8) and 1% (v/v) DMSO. ALDH7A1 (4 μM) was incubated with NAD^+ (2.5 mM) and DEAB (20, 30, 40, and 50 μM) on ice. At intervals of 4 minutes, a 50 μL aliquot was diluted into a 1 mL assay mixture containing hexanal (0.8 mM) and NAD^+ (2.5 mM), and the activity was immediately measured. The activity data were fit globally using Origin 9 to a pseudo-first order inactivation model as described previously (White T. A. et al.) using Equation 1. In Equation 1,

A(t) is the percent activity remaining after incubating the enzyme with the inactivator at concentration I for time t. The parameters A and y_0 are fitting parameters, which are determined for each inactivator concentration simultaneously. K_I and k_{inact} are the desired inactivation kinetics parameters.

$$A(t) = A\exp(-\frac{k_{inact}It}{K_I+I}) + y_0$$

6.3.7 Persistence of inactivation following removal of excess DEAB.

Three samples of ALDH7A1 (~10 μM) were incubated with 400 μM DEAB and 2.5 mM NAD^+ in 100 mM sodium pyrophosphate buffer at pH 8 containing 1% (v/v) DMSO at room temperature for 2 hours. As a control, another three samples of ALDH7A1 were incubated with 2.5 mM NAD^+ in 100 mM sodium pyrophosphate buffer at pH 8 containing 1% (v/v) DMSO at room temperature for 2 hours. After incubation, these samples were applied to PD-10 desalting columns (GE healthcare) using a running buffer of 100 mM sodium pyrophosphate at pH 8.

ALDH activity was measured after desalting by monitoring NADH production at 340 nm using hexanal as the substrate. The assay volume was 1 mL; the assay buffer was 100 mM sodium pyrophosphate at pH 8. The final assay includes 1 μM of the desalted enzyme, 2 mM hexanal and 3 mM NAD^+. All reactions were initiated by adding hexanal and NAD^+.

6.4 RESULTS

Two crystal structures of ALDH7A1 with DEAB bound were determined (Table 6.1). A 2.4 Å resolution structure having space group $I422$ was determined following

co-crystallization of ALDH7A1 with DEAB and a large excess of NAD^+ (5 mM). Electron density for NAD^+ is strong in the *I*422 form, consistent with the high concentration of NAD^+ used in crystallization. A second structure was determined following co-crystallization with DEAB without adding excess NAD^+ to the crystallization setup. In this case, 1 mM NAD^+ was added to the enzyme during purification, and the enzyme was dialyzed against a buffer devoid of NAD^+ prior to crystallization. This protocol produced an *F*222 crystal form that diffracts to 1.95 Å resolution. Electron density for NAD^+ is absent in the *F*222 form, which suggests that the final dialysis step reduced the NAD^+ concentration to a level that was insufficient to yield high occupancy of the cofactor in this particular crystal form.

ALDH7A1 exhibits the expected ALDH superfamily fold consisting of NAD^+-binding, catalytic, and oligomerization domains (Figure 6.1a). The catalytic domain contains the nucleophilic Cys302 residue that attacks the carbonyl C atom of the aldehyde substrate. The NAD^+-binding domain adopts the Rossmann fold. The oligomerization domain mediates domain-swapped dimerization. As with certain other ALDHs, two dimers of ALDH7A1 assemble into a tetramer (Figure 6.1b). We note that this tetramer is present in both crystal lattices reported here as well as the monoclinic lattice reported previously for ALDH7A1.[20] The active site is located in the crevice between the NAD^+-binding and catalytic domains, underneath the interdomain linker peptide.

Electron density representing DEAB is evident near the catalytic cysteine, Cys302, in both crystal forms (Figures 6.1c, 6.1d). The maps are consistent with a covalent bond between the S atom of Cys302 and the carbonyl C atom of DEAB. A protruding electron density feature near the S atom indicates the orientation of the C-

O bond of DEAB. This feature is more obvious in the higher resolution *F*222 form (Figure 6.1d). The remaining volume of the density feature is sufficient to accommodate the diethylaminobenzene moiety. This is the first evidence that DEAB forms a stable covalent adduct with an ALDH.

The covalently attached DEAB forms several interactions with the enzyme. We first focus on the *I*422 form (Figure 6.1c). The O atom of DEAB occupies the oxyanion hole and accepts hydrogen bonds from the main-chain amide of Cys302 and the side chain of Asn167. Whereas one face of the aromatic ring of DEAB packs against the NAD^+ nicotinamide and Val172, the other face is highly solvent-accessible. The phenyl ring of DEAB forms edge-to-face π-π interactions (3.3 Å) with the phenyl rings of Phe168 and Phe468, while the diethylamino moiety sits atop Trp175. These three residues form an aromatic box, which is a conserved feature of ALDH substrate recognition.[21]

Although both structures clearly indicate covalent attachment of DEAB to Cys302, they differ with respect to the local polypeptide conformations and interactions with DEAB. The major difference is that the active site in the *F*222 form is much more open due to conformational disorder of the interdomain linker and the C-terminal 13 residues of the polypeptide chain. Because of this disorder, the interaction of DEAB with aromatic box residue Phe468 is lost in the *F*222 lattice (Figure 6.1d). Also, Trp175 adopts a different rotamer, which results in less contact with the diethyl groups of DEAB (Figure 6.1d). Finally, the carbonyl O atom of DEAB forms only one hydrogen bond (with the backbone of Cys302) in the *F*222 form, since the interaction distance with Asn167 is too long for a hydrogen bond (3.7 Å). This difference may be due to the absence of the NAD^+ nicotinamide bracing one

face of DEAB.

Mass spectrometry was used to confirm the covalent modification seen in the structures. ALDH7A1 was incubated with DEAB and NAD^+ prior to analysis with Nano-LC Nanospray QTOF-MS. The spectrum shows a species that has molecular mass of the native enzyme plus 175 Da (Figure 6.2a). Given a molecular mass of 177 Da for DEAB, the mass spectrum is consistent with a covalent modification having the stoichiometry of one DEAB per protein chain, as observed in the crystal structures.

ALDH activity assays were used to provide further evidence that DEAB is an irreversible inactivator of ALDH7A1. ALDH7A1 was incubated with excess DEAB and NAD^+, and the residual ALDH activity was measured as a function of incubation time using hexanal as the substrate. As shown in Figure 6.2b, DEAB inactivates ALDH7A1 in a time- and concentration-dependent manner. The residual enzyme activity measured as a function of time at five DEAB concentrations was globally fit to a pseudo-first order inactivation model (Equation 1 of Supporting Information) yielding apparent kinetic parameters of $k_{inact} = 0.4 \pm 0.1$ min^{-1} and $K_I = 100 \pm 36$ μM. The apparent second-order rate constant for enzyme inactivation by DEAB is $k_{inact}/K_I = 70 \pm 30$ $M^{-1}s^{-1}$. These results are consistent with DEAB being a covalent inactivator of ALDH7A1.

A hallmark of irreversible inactivation is the persistence of inhibition following removal of unbound inhibitor.[22] To investigate this issue, the enzyme (10 μM) was inactivated by treatment with DEAB (400 μM) and NAD^+ (2.5 mM) and then passed through a size exclusion column to remove excess DEAB. As a control, a sample of

ALDH7A1 was treated similarly except that DEAB was omitted. After desalting, the DEAB-treated sample exhibited less than 1 % of the activity compared to the control (Figure 6.2c). These results are consistent with covalent, irreversible inactivation of ALDH7A1 by DEAB.

Our results contradict the conventional wisdom that DEAB is a noncovalent, reversible inhibitor of ALDHs.[11, 13] The notion that DEAB is a reversible inhibitor of ALDHs has always seemed curious, given that DEAB is an aldehyde, and that the related compound, benzaldehyde, is a known substrate for many ALDHs.

Our data indicate that DEAB forms a stable acyl-enzyme species that is a "stalled" intermediate in the enzyme's normal catalytic cycle (Figure 6.3). In the accepted ALDH mechanism,[10] nucleophilic attack by the catalytic cysteine on the aldehyde (Figure 6.3a) produces a hemithioacetal intermediate (Figure 6.3b). Hydride transfer to NAD^+ generates NADH and the acyl-enzyme intermediate (Figure 6.3c). Hydrolysis of the acyl-enzyme intermediate yields the carboxylic acid product (Figure 6.3d) and regenerates the native enzyme. The acyl-enzyme intermediate derived from DEAB is apparently stalled. Mass spectral data support the proposed inactivation mechanism. Formation of a stable acyl-enzyme species would be accompanied by loss of two protons – loss of one to solvent during the activation of Cys302 and transfer of a second, as a hydride ion, from the hemithioacetal to NAD^+. Thus, the expected mass-shift for formation of the DEAB acyl-enzyme would be 175 Da, in perfect agreement with the observed mass-shift, measured by mass spectrometry.

The mechanism-based inactivation of SARS main proteinase by benzotriazole esters provides a precedent for the proposed mechanism of inactivation of ALDH7A1

by DEAB.[23, 24] In particular, reaction of the proteinase with 1-(4-dimethylaminobenzoyloxy)-benzotriazole results in acylation of the catalytic Cys with 1-(4-dimethylaminobenzoyloxy) (PDB code 2V6N), which is very similar to the covalent adduct that we observe. Donation of electrons from the amino substituent toward the dimethylaminobenzoyloxy thioester is thought to render the adduct resistant to hydrolysis.[23, 24] Analogous electronic effects may be at work in our case, and may help explain why DEAB is an inactivator of ALDH7A1, whereas benzaldehyde is a substrate.[20]

Alternatively, or at the same time, interactions of DEAB with active site residues may serve to disable the enzyme's hydrolytic machinery. For example, various interactions of DEAB may force the carbonyl of the acyl-enzyme intermediate out of the ideal stereoelectronic alignment required for reaction with the activated water molecule bound to Glu268 at the enzyme active site.[25] Also, Glu268 itself may not be positioned correctly for activating the attacking water molecule. In the normal catalytic cycle, following hydride transfer, the NMN group of NADH rotates out of the active site, allowing Glu268 to rotate into the nicotinamide site to activate the attacking water molecule. In the *F*222 structure, which does not have NAD^+ bound, the side chain of Glu268 is disordered in two chains and in the "out" rotamer in the other two chains.

In summary, the inhibition mechanism of DEAB is much different than previously thought. Whether DEAB similarly inactivates other ALDHs remains to be determined, but such investigations should be pursued, given the widespread use of DEAB in the ALDH and CSC research communities. The discovery that DEAB can be an irreversible, covalent inactivator suggests a new strategy for designing ALDH-

targeted drugs and probes.

Accession codes

Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 4X0T (*I*422 form) and 4X0U (*F*222 form).

6.5 REFERENCES

1. Vasiliou, V., Thompson, D. C., Smith, C., Fujita, M., and Chen, Y. (2013) Aldehyde dehydrogenases: from eye crystallins to metabolic disease and cancer stem cells, *Chem. Biol. Interact. 202*, 2-10.

2. Ma, I., and Allan, A. L. (2011) The role of human aldehyde dehydrogenase in normal and cancer stem cells, *Stem. Cell. Rev. 7*, 292-306.

3. Muzio, G., Maggiora, M., Paiuzzi, E., Oraldi, M., and Canuto, R. A. (2012) Aldehyde dehydrogenases and cell proliferation, *Free Radic. Biol. Med. 52*, 735-746.

4. Abdullah, L. N., and Chow, E. K. (2013) Mechanisms of chemoresistance in cancer stem cells, *Clin. Transl. Med. 2*, 3.

5. Januchowski, R., Wojtowicz, K., and Zabel, M. (2013) The role of aldehyde dehydrogenase (ALDH) in cancer drug resistance, *Biomed. Pharmacother. 67*, 669-680.

6. Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M. A., and Dick, J. E. (1994) A cell initiating human acute myeloid leukaemia after transplantation into SCID mice, *Nature 367*, 645-648.

7. Balber, A. E. (2011) Concise review: aldehyde dehydrogenase bright stem and progenitor cell populations from normal tissues: characteristics, activities, and emerging uses in regenerative medicine, *Stem Cells 29*, 570-575.

8. Minn, I., Wang, H., Mease, R. C., Byun, Y., Yang, X., Wang, J., Leach, S. D., and Pomper, M. G. (2014) A red-shifted fluorescent substrate for aldehyde dehydrogenase, *Nat. Commun. 5*, 3662.

9. Marcato, P., Dean, C. A., Giacomantonio, C. A., and Lee, P. W. (2011) Aldehyde dehydrogenase: its role as a cancer stem cell marker comes down to the specific isoform, *Cell Cycle 10*, 1378-1384.

10. Koppaka, V., Thompson, D. C., Chen, Y., Ellermann, M., Nicolaou, K. C., Juvonen, R. O., Petersen, D., Deitrich, R. A., Hurley, T. D., and Vasiliou, V. (2012) Aldehyde

dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application, *Pharmacol. Rev. 64*, 520-539.

11. Russo, J. E., Hauguitz, D., and Hilton, J. (1988) Inhibition of mouse cytosolic aldehyde dehydrogenase by 4-(diethylamino)benzaldehyde, *Biochem. Pharmacol. 37*, 1639-1642.

12. Croker, A. K., and Allan, A. L. (2012) Inhibition of aldehyde dehydrogenase (ALDH) activity reduces chemotherapy and radiation resistance of stem-like ALDHhiCD44(+) human breast cancer cells, *Breast Cancer Res. Treat. 133*, 75-87.

13. Russo, J., Chung, S., Contreras, K., Lian, B., Lorenz, J., Stevens, D., and Trousdell, W. (1995) Identification of 4-(N,N-dipropylamino)benzaldehyde as a potent, reversible inhibitor of mouse and human class I aldehyde dehydrogenase, *Biochem. Pharmacol. 50*, 399-406.

14. Moreb, J. S., Ucar, D., Han, S., Amory, J. K., Goldstein, A. S., Ostmark, B., and Chang, L. J. (2012) The enzymatic activity of human aldehyde dehydrogenases 1A2 and 2 (ALDH1A2 and ALDH2) is detected by Aldefluor, inhibited by diethylaminobenzaldehyde and has significant effects on cell proliferation and drug resistance, *Chem. Biol. Interact. 195*, 52-60.

15. Moreb, J. S., Maccow, C., Schweder, M., and Hecomovich, J. (2000) Expression of antisense RNA to aldehyde dehydrogenase class-1 sensitizes tumor cells to 4-hydroperoxycyclophosphamide in vitro, *J. Pharmacol. Exp. Ther. 293*, 390-396.

16. Pappa, A., Brown, D., Koutalos, Y., DeGregori, J., White, C., and Vasiliou, V. (2005) Human aldehyde dehydrogenase 3A1 inhibits proliferation and promotes survival of human corneal epithelial cells, *J Biol Chem 280*, 27998-28006.

17. van den Hoogen, C., van der Horst, G., Cheung, H., Buijs, J. T., Pelger, R. C., and van der Pluijm, G. (2011) The aldehyde dehydrogenase enzyme 7A1 is functionally involved in prostate cancer bone metastasis, *Clin. Exp. Metastasis. 28*, 615-625.

18. Giacalone, N. J., Den, R. B., Eisenberg, R., Chen, H., Olson, S. J., Massion, P. P., Carbone, D. P., and Lu, B. (2013) ALDH7A1 expression is associated with recurrence in patients with surgically resected non-small-cell lung carcinoma, *Future Oncol. 9*, 737-745.

19. Saw, Y. T., Yang, J., Ng, S. K., Liu, S., Singh, S., Singh, M., Welch, W. R., Tsuda, H., Fong, W. P., Thompson, D., Vasiliou, V., Berkowitz, R. S., and Ng, S. W. (2012) Characterization of aldehyde dehydrogenase isozymes in ovarian cancer tissues and sphere cultures, *BMC cancer 12*, 329.

20. Brocker, C., Lassen, N., Estey, T., Pappa, A., Cantore, M., Orlova, V. V., Chavakis, T., Kavanagh, K. L., Oppermann, U., and Vasiliou, V. (2010) Aldehyde dehydrogenase 7A1 (ALDH7A1) is a novel enzyme involved in cellular defense against hyperosmotic stress, *J Biol Chem 285*, 18452-18463.

21. Riveros-Rosas, H., Gonzalez-Segura, L., Julian-Sanchez, A., Diaz-Sanchez, A. G., and Munoz-Clares, R. A. (2013) Structural determinants of substrate specificity in aldehyde dehydrogenases, *Chem. Biol. Interact. 202*, 51-61.

22. Silverman, R. B. (1995) Mechanism-based enzyme inactivators, *Methods Enzymol 249*, 240-283.

23. Verschueren, K. H., Pumpor, K., Anemuller, S., Chen, S., Mesters, J. R., and Hilgenfeld, R. (2008) A structural view of the inactivation of the SARS coronavirus main proteinase by benzotriazole esters, *Chem. Biol. 15*, 597-606.

24. Wu, C. Y., King, K. Y., Kuo, C. J., Fang, J. M., Wu, Y. T., Ho, M. Y., Liao, C. L., Shie, J. J., Liang, P. H., and Wong, C. H. (2006) Stable benzotriazole esters as mechanism-based inactivators of the severe acute respiratory syndrome 3CL protease, *Chem. Biol. 13*, 261-268.

25. Bürgi, H. B., Dunitz, J. D., Lehn, J. M., and Wipff, G. (1974) Stereochemistry of reaction paths at carbonyl centres, *Tetrahedron 30*, 1563–1572.

26. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation, *Proteins 50*, 437-450.

27. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. D Biol. Crystallogr. 66*, 213-221.

Table 6-1 X-ray Diffraction Data Collection and Refinement[a]

	DEAB-NAD^+	DEAB
Space group	*I*422	*F*222
Unit cell parameters (Å)	$a = b = 160.3$ $c = 320.5$	$a = 161.5$ $b = 219.4$ $c = 233.2$
Wavelength	1.000	1.000
Resolution (Å)	59.52 – 2.40 (2.45 – 2.40)	51.49 – 1.95 (1.98 – 1.95)
Observations	606813	780650
Unique reflections	81067	148899
$R_{merge}(I)$	0.144 (0.996)	0.054 (0.615)
$R_{meas}(I)$	0.166 (1.146)	0.068 (0.803)
$R_{pim}(I)$	0.079 (0.545)	0.041 (0.510)
Mean I/σ	12.0 (1.9)	22.1 (1.8)
Completeness (%)	99.7 (95.4)	99.9 (98.3)
Multiplicity	7.5 (7.5)	5.2 (3.6)
No. of protein chains	4	4
No. of protein residues	2035	1933
No. of protein atoms	15109	14441
No. of DEAB atoms	39	39
No. of water molecules	314	516
No. of NAD^+ atoms	176	n/a
R_{cryst}	0.1803 (0.2532)	0.1879 (0.2375)
R_{free}[b]	0.2348 (0.3057)	0.2270 (0.2857)
rmsd bond lengths (Å)	0.003	0.007
rmsd bond angles (°)	0.682	0.969
Ramachandran plot[c]		
Favored (%)	96.65	98.06
Outliers (%)	0.10	0.10
Average B-factor ($Å^2$)		
Protein	38.2	33.7
DEAB	39.6	42.3
Water	29.9	29.8
NAD^+	32.3	n/a
Coordinate error (Å)[d]	0.29	0.20
PDB code	4X0T	4X0U

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

[b]5 % random test set.

[c]The Ramachandran plot was generated with RAMPAGE.[26]

[d]Maximum likelihood-based coordinate error estimate reported by PHENIX.[27]

Figure Legends

Figure 6-1. Structure of DEAB-inactivated ALDH7A1. (a) Protomer structure, highlighting the domain architecture and DEAB binding site. The three major domains are colored red (NAD+-binding), blue (catalytic), and green (oligomerization). The polypeptide section that links the catalytic and NAD+-binding domains is colored gold. Modified Cys302 and NAD+ are colored pink and cyan, respectively. The inset shows a close-up view of the active site. (b) Two views of the ALDH7A1 tetramer, which is present in both the I422 (pictured here) and F222 crystal lattices. The four protomers have different colors. Modified Cys302 and NAD+ are colored pink and cyan, respectively. (c) Electron density for covalently-modified Cys302 (pink) in the I422 crystal form. The mesh represents a simulated annealing Fo-Fc omit map contoured at 2.5σ. NAD+ is shown in cyan. (d) Electron density for covalently-modified Cys302 (pink) in the F222 crystal form. The mesh represents a simulated annealing Fo-Fc omit map contoured at 2.5σ.

Figure 6-2. Confirmation of covalent inactivation of ALDH7A1 by DEAB. (a) Nano-LC Nanospray QTOF mass spectra of native ALDH7A1 (top) and DEAB-inactivated ALDH7A1 (bottom). (b) Residual ALDH activity measured after incubating ALDH7A1 with 20 – 50 μM DEAB and 2.5 mM NAD^+. (c) ALDH activity measured after gel filtration of native ALDH7A1 (left) and DEAB-inactivated ALDH7A1 (right). The activities are normalized to that of the native enzyme. The error bars were calculated from three trials.

Figure 6-3. Diagram summarizing the accepted mechanism of ALDHs. (a) The aldehyde substrate binds to the active site of the enzyme-NAD^+ complex to form a

ternary complex. Catalytic Cys302 attacks the carbonyl C atom of the aldehyde substrate. (b) Nucleophilic attack by the catalytic cysteine on the aldehyde produces a hemithioacetal intermediate. Hydrogen bond donors in the oxyanion hole stabilize the negative charge of the hemithioacetal intermediate. (c) Hydride transfer to NAD^+ generates NADH and the acyl-enzyme intermediate. Conserved Glu268 activates the water molecule that attacks the acyl-enzyme intermediate. (d) Hydrolysis of the acyl-enzyme intermediate yields the carboxylic acid product, which is released from the enzyme. NADH dissociates and NAD^+ binds to regenerate the native enzyme.

Figure 6-1 Structure of DEAB-inactivated ALDH7A1.

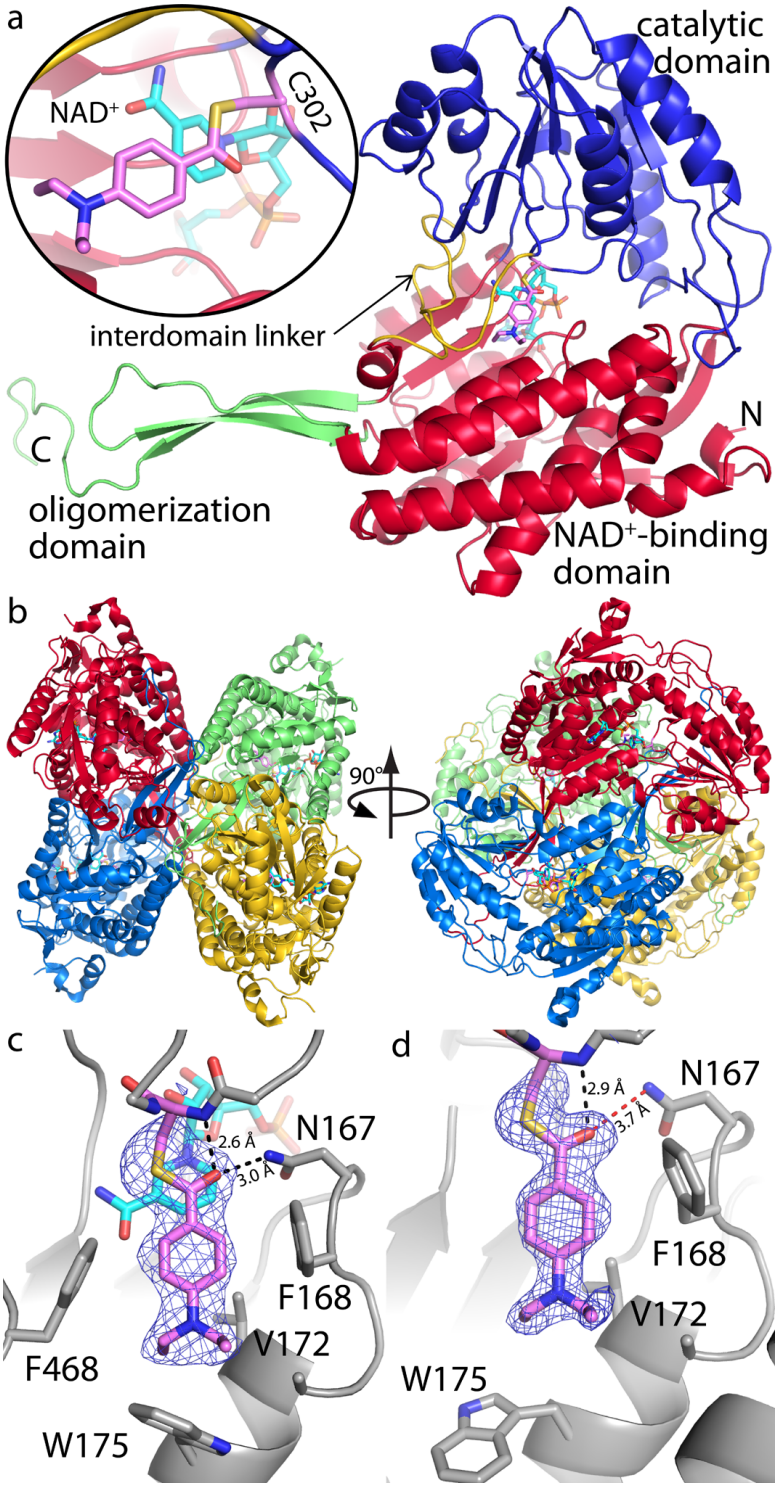


Figure 6-2 Confirmation of covalent inactivation of ALDH7A1 by DEAB

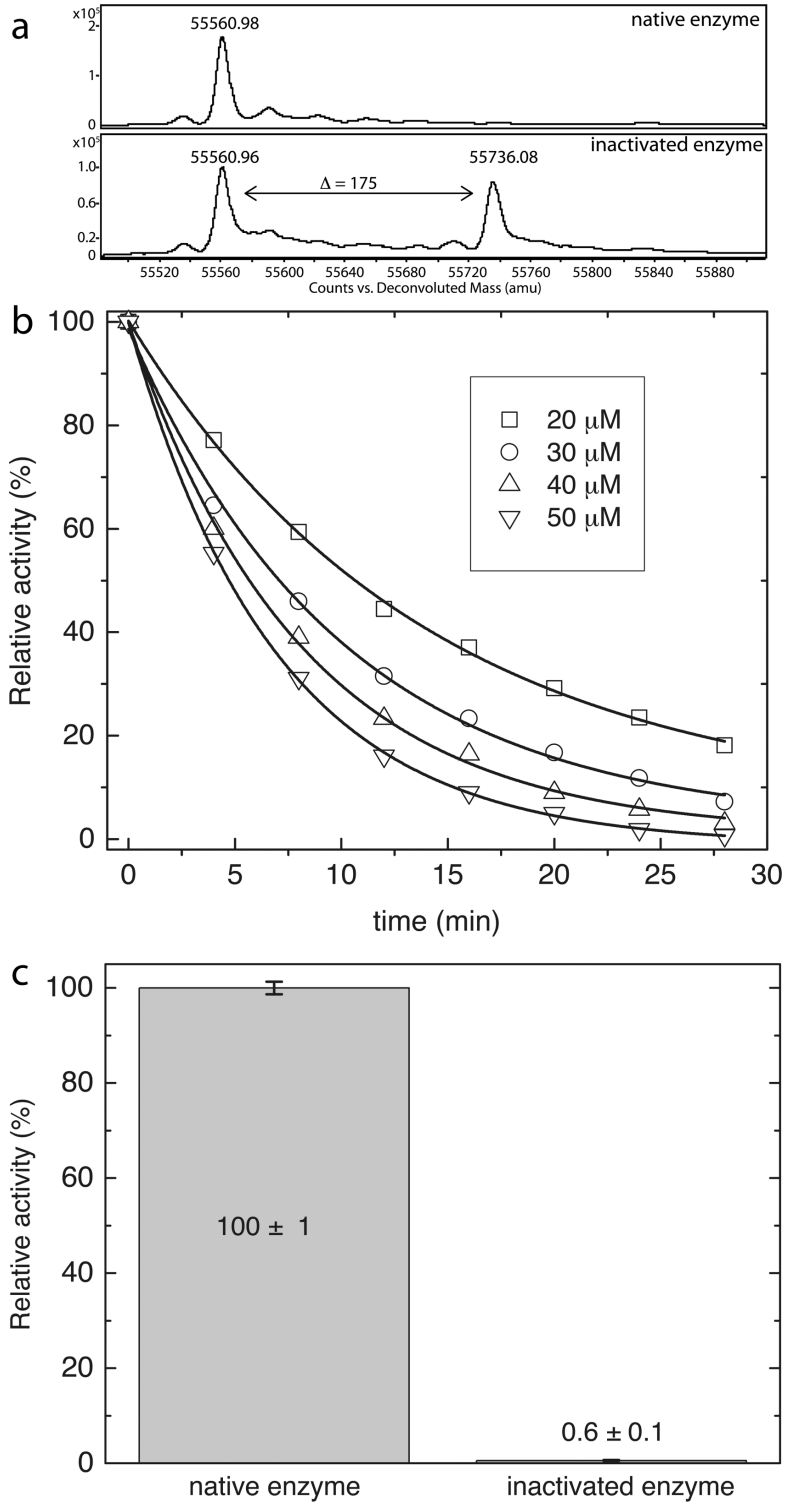
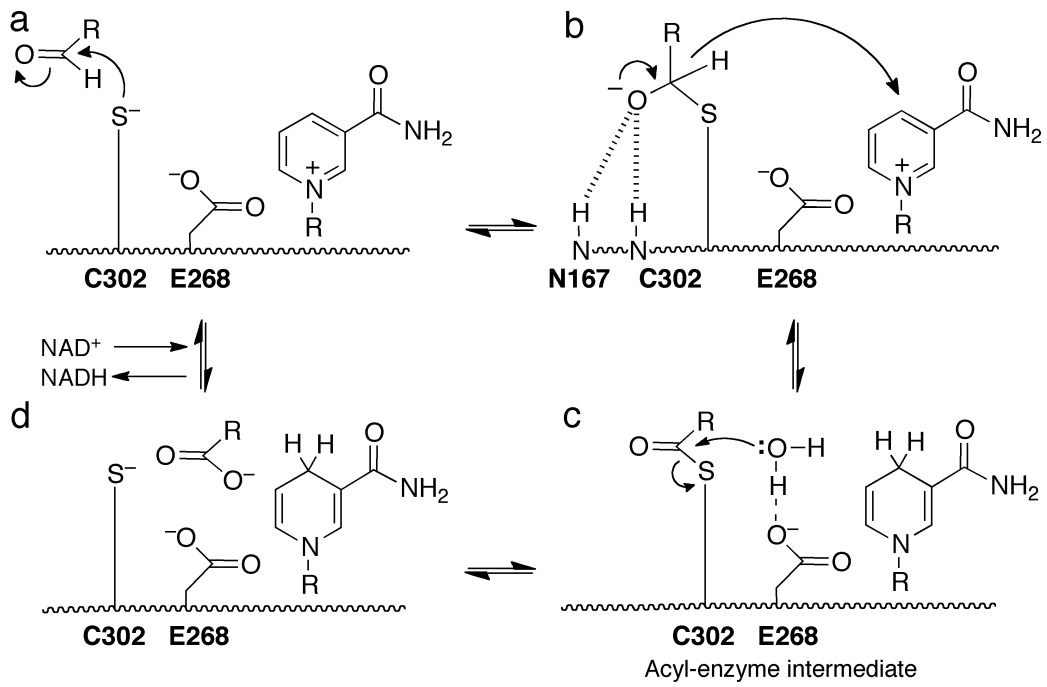


Figure 6-3 Diagram summarizing the accepted mechanism of ALDHs



7. VITA

Min Luo was born in Hunan province of China. He first attended Shandong University, as an undergraduate student majored in Mechanic engineering. But after one year study in engineering, he found his real interest is in life science. In 2006, the second year of his undergraduate study, he followed his heart and changed his major to biology and got his Bachelors in 2009. After that, he came to America to start his doctoral career in Department of Chemistry, University of Missouri - Columbia. In the spring of 2010, he joined Professor John J. Tanner Group to start his research on the structural and functional study of important enzymes. In February 2015, Min will start his postdoctoral career at Harvard University.