9-12-2014

Criteria for Evolution of Successful Proteins: Fold Fitness and Domain Dynamics Explored

Tyrel Bryan

Tyrel Bryan

Candidate

Chemistry and Biological Chemistry

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Debra Dunaway-Mariano, Chairperson

Karen N. Allen

Patrick S. Mariano

Charles E. Melançon III

CRITERIA FOR EVOLUTION OF SUCCESSFUL PROTEINS: FOLD FITNESS AND DOMAIN DYNAMIC EXPLORED

by

TYREL BRYAN

B.S. Chemistry, University of New Mexico

DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy
Chemistry

The University of New Mexico

Albuquerque, New Mexico

July 2014

Acknowledgment

This collection of work is the product of hard work and continuous support of esteemed colleagues, good friends, and devoted family. I would like to take this moment to thank those people.

First and foremost I would like to give thanks to my family. Without their continuous support and love I would not be in this situation. My father fostered my curiosity for science and nature, and always supported my interests. A special thanks to my mother for her sacrifices that enable my growth as a person and scientist. To my sister her continuous support was always appreciated. Thank you for everything from the bottom of my heart.

During my time as a graduate student I have had the privilege to interact with many talented people and lucky enough to call many of them friends. To John and Luke: thank you for your insight and discussions pertaining to science, but more important the time we spent outside work from the links to the taps. To lab members past and present including Chunliang, Sarah, and Jenny thank you for your friendship. In most labs the barrier between undergrads and graduate student is stiff, but luckily a cultivated friendship was created with Rachel, Kaila, and Andrea, thank you for making those workdays exciting. In addition, special thanks for all the hard work put forth by Andrea.

Science is a collaboration of thoughts and efforts. I am thankful for the connections create with collaborators near and far. Thank you to Dan Saltzberg, Javier Gonzalez, Zoe Fisher, and help from Thomas Weiss.

Thanks to Dr. Patrick Mariano for all your direction and advice that will allow me to grow as a scientist. I gratefully acknowledge Dr. Charles Melançon for your input and patience.

To Karen: I am grateful for all your help and knowledge. You have been an integral part in shaping me as a structural protein chemist. From allowing me to invade your laboratory to Skype conversation, you have taught me much the whole way. I am grateful for your help and patients.

To DDM: I would like to thank you for every opportunity you have given me especially the opportunity to follow my own interests. You allowed to me express my interests and grow them into meaningful research. Your guidance and insight has allowed me to become a better scientist and is invaluable to me. Thank you for everything you have offered through my time as an undergrad and graduate student. You will always be a source of guidance in my career as a scientist.

And Last: my fiancée Hanna, you have been the continuous source of energy and persistence in my journey. Thank you for your patients and love especially through the long nights of prepping and many trips away.

CRITERIA FOR EVOLUTION OF SUCCESSFUL PROTEINS: FOLD FITNESS AND DOMAIN DYNAMICS EXPLORED

Tyrel Bryan

B.S., Chemistry, University of New Mexico, 2009

Ph.D., Chemistry, University of New Mexico 2014

ABSTRACT

The Haloacid Dehalogenase Superfamily (HADSF) is a ubiquitous family of enzyme with more than 32,000 members. A variety of reactions are catalyzed by HAD members, but a majority of HAD members are phosphatases. Dephosphorlyation of organophosphate metabolites are carried out in the conserved Rossmann-like core domain. The Rossman-like core domain houses the conserved active site residues necessary for catalysis and also supports domain inserts associated with substrate binding.

This dissertation will focus primarily on investigating fold fitness and domain dynamics of HAD proteins more specifically three C2 capped proteins. The first part will focus on whether or not the prevailing tenet that "*Catalytic Function Comes at the Expense of Protein Fold Stability*" is applicable to the HADSF phosphatase. To test function at the cost of stability site directed mutagenesis was used to replace catalytic residues and used an array of techniques (CD/Fluorescence/SAXS) to monitored denaturation to evaluate stability in the resulting mutants. From denature results we

conclude that although most catalytic residues play no role in stability, good or bad, one in particular Asp^2 of DXD squiggle shows to be necessary for not only activity but stability as well. To identify what about the residue is essential to stability the side chain and interacting residues were examined.

The Rossmann-like catalytic core domain is a robust domain, and to test the co-evolution of the cap and catalytic domains as a "coupled folding and catalytic units" will be examined. The structural co-dependency was tested by creating truncations of the catalytic and cap domains and identifying stability and activity of each domain and chimeric C2 capped proteins. Thermal denature of truncations and chimeric proteins reveal stability of each domain separately while the chimeric stability establish a robust core Rossmann-like catalytic domain.

As previously discussed the cap domain increases function by increasing recognition of substrates. To examine the cap domains role in substrate-induced-fit we examined changes in protein solution conformation using small angle X-ray scattering (SAXS). Because HAD proteins have evolved separate domains for catalysis and recognition SAXS will also be used to investigate changes in conformation change due to changes in domain-domain interfacing residues. Solution scattering of Bt4131 help demonstrate an induced-fit mechanism for HAD proteins. In the presence of VO_3 (phosphate analog) we see closure of cap domain whereas sugar moiety does not initiate closure of domain.

Table of Contents

CHAPTER ONE

Protein Structural Elucidation and Evolutionary Markers

1.1 Haloacid Dehalogenase Superfamily

The Haloacid Dehalogenase Superfamily (HADSF) is a ubiquitous family of enzymes (1,2). Although most family members are phosphotransferases, the family was named after the first characterized member, 2-haloacid dehalogenase (3,4). The HADSF enzymes are found in all forms of life and typically there are ten or more different HADSF enzymes per organism. The chemistry catalyzed by this enzyme includes phosphate and carbon group transfer processes (Figure 1.1) (1). The phosphatases, enzymes that catalyze phosphate ester hydrolysis, make up the majority of the HADSF. Phosphorylated biomolecules are as varied in structure and function as they are abundant in the cell. Therefore, phosphoryl group recycling is central in living systems, and the large size of the HADSF phosphatase subfamily can, in part, be attributed to the high demand for cellular phosphatases.

The fact that the HADSF phosphatases greatly outnumber the phosphatases from other superfamilies suggests that the HADSF is highly effective at evolving novel phosphatases. For instance, the phosphatases of *E. coli* derive from seven different enzyme superfamilies, yet the twenty-nine HADSF phosphatases present outnumber the combined sum of the phosphatases from the other six superfamilies. The dominance and success can be associated with the expanded library of substrates that support catalysis also known as "evolutionary exploration of substrate space" (5), but what special about

the HADSF has allowed this proliferation in living organism. Among a set of reactions catalyzed by the HADSF, the protein members have a conserved core domain which houses the active residues conserved in all HAD proteins.

With the vast amount of information available through X-ray crystal structures and bioinformatic networks a possible explanation for the evolutionary success of the HADSF can be proposed.

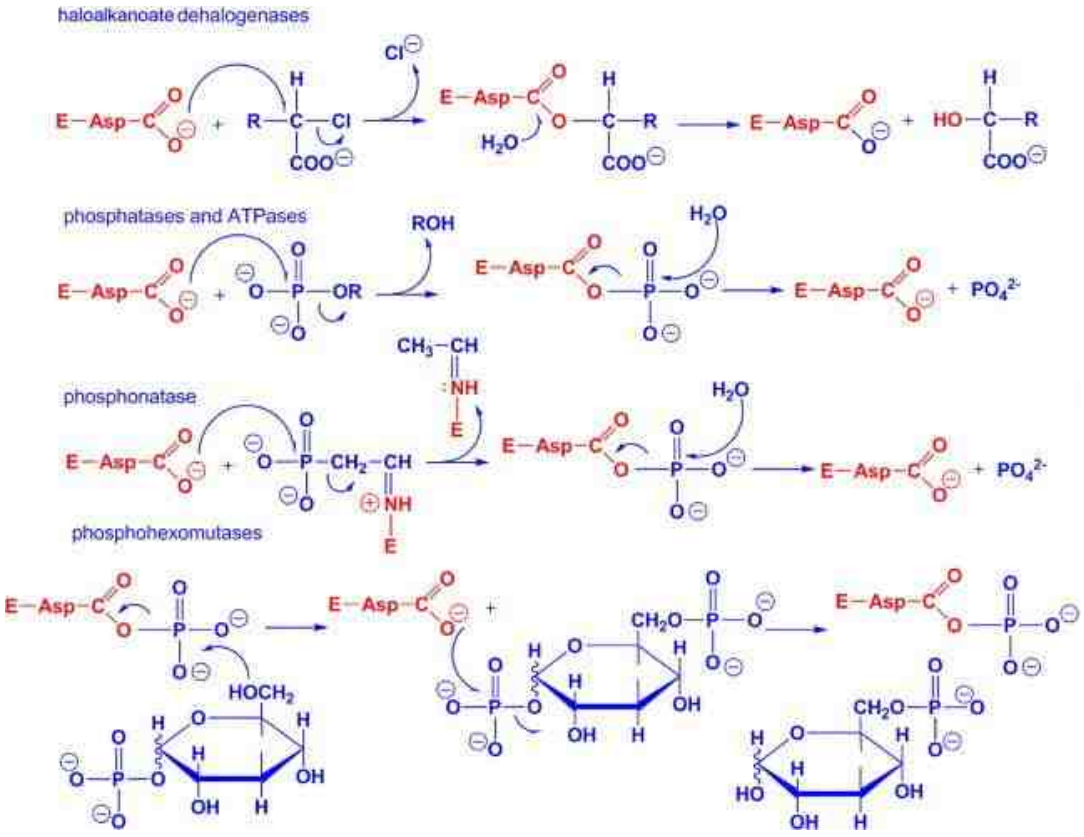

Figure 1.1: General reactions catalyzed by HAD proteins. RED: protein residues whereas BLUE represents substrates or coordinating water.

1.2 HADSF Rossman Core Domain

The core domains of the HADSF members consist of a central β-sheet in between α-helices, a topology that is reminiscent of Rossmannoid class of α/β folds. Unlike other

Rossmann-like folds, the HADSF fold has four conserved motifs that play a significant role in catalysis (Figure 1.2) (3, 6). The four conserved residues are associated with the active site of the core domain in all HADs. The active site of the core domain is the binding site for the phosphate moiety which is transferred from the phosphate ester substrate to in most cases a solvent water molecule.

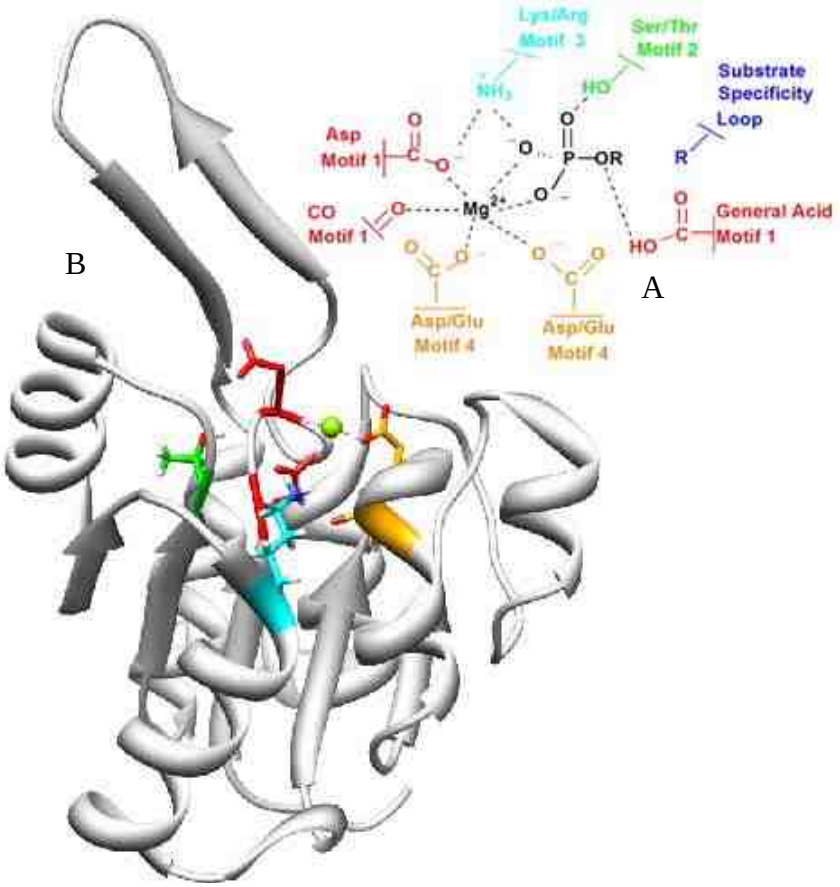

Figure 1.2: A. Cartoon representation of conserved HADSF motifs, each Motif is colored individually. **B.** represents monomer of KDN9P (PDB: 3E8M) colored residues correspond to motif coloring of **A.**

The conserved motifs of HADs, residing in the active site of the protein, consist of Motif I with a conserved Asp with the sequence DxD, Motif II with a loop 2 a conserved Ser/Thr, Motif III with a loop 3 conserved Arg/Lys, and Motif IV with a loop

4 conserved Asp. Because each motif includes conserved residues that are responsible for specific binding functions within the active site, sequence alignment allows for identification of HADSF members.

The role of the conserved aspartates (DxD) in Motif I, involving binding of the metal cofactor Mg^{2+}, is the cornerstone to catalysis. Asp^{1} acts as a nucleophile and creates an aspartyl-phosphoryl intermediate and Asp^{2} act as a general acid base to regenerate enzyme and the alcohol leaving group (7-9). Motifs II and III include conserved Ser/Thr and Lys/Arg respectively, which assist in stabilizing high energy intermediates. Loop 4 associated with Motif IV consists of two Asp residues, which partner with the aspartates in Motif I to bind Mg^{2+} (3, 4, 7, 10-12). Although the catalytic core comprised of the four motifs described above motifs is conserved in HADs, diversity in substrate recognition is directed by inserts into the Rossmann-like core domain. These insert or "caps" aid in active site shielding and substrate recognition.

1.3 Incorporation of Cap Domain and Expansion of Substrate Recognition

The caps found in a majority of HADSF proteins are categorized into subfamilies that are classified by size and insert point. The three subfamilies are: C0 comprise HAD proteins that include small loops or no cap, C1 and C2 capped proteins are distinguished by insertion point of cap domain (between loops 1 & 2 or between loop 2 & 3 respectively) Figure 1.3. As shown in Figure 1.2 substrate specificity is usually dominated by interactions with cap domains or extended loops (11-13).

Two well-studied C0 HAD proteins include 2-keto-3-deoxy-D-glycero-D-galacto-9-phosphatase (KDN9P) from the 8KDOP subfamily and D-glycero-D-manno-Heptose-1,7-bisphosphate phosphatase (GmhB) from the HisB subfamily (11, 14). The X-ray structure shows that KDN9P contains a β-loop-β insert at the same C1 insert site in the core domain. This extended loop functions in oligermerizing KDN9P monomers into a functional tetramer (14) GmhB of the C0 subfamily belonging to the HisB clade does not possess a true cap domain, but instead contains three ornate loops which create a rounded barrier for active site and bound substrate (11). In both of these enzymes, the inserted loops act to increase substrate affinity and to create a more active conformation.

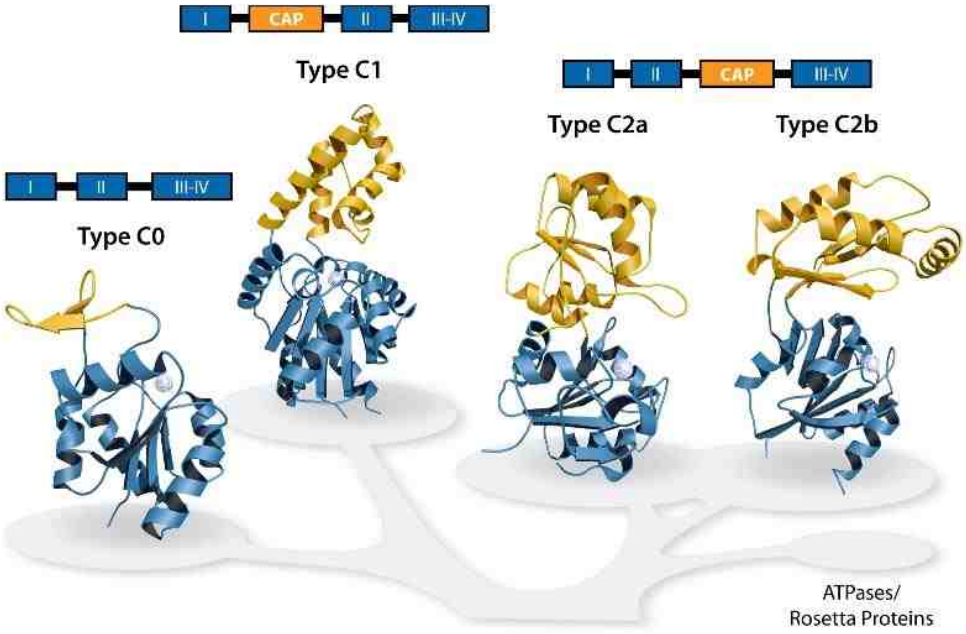

Figure 1.3: Representation of the HADSF subfamilies based on cap inserts and insert site. BLUE represents the Rossmann-like core domain and ORANGE represents the inserted cap domain.

The C1 subfamily proteins are distinguished by possessing one of two types of cap inserts, including an α-helical cap or a small α+β cap. One of the more studied C1 (α-helical) capped proteins is β-phosphoglucomutase (βPGM) from Lactococcus *lactis*.

βPGM catalyzes the interconversion of β-D-glucose-1-phosphate to D-glucose-6-phohspate. Important to substrate conversion is the cap domains and conserved residues that bind the substrate (10, 15).

The C2 subfamily is divided into 2 cap types, the NagD (type A) and Cof-type phosphatases (type B) (1). Both caps incorporate an α+β domain consisting of at least three β-strands. Two crystallized C2B proteins in these subfamilies include Bacteroides *thetaiotaomicron* Bt4131 and Escherichia *coli* YbiV. Both proteins have elaborate cap inserts with C2B topologies (12, 17, and 18). The results of substrate screening show that both proteins have affinities for sugar phosphates. The determination of classes and categorizing of HAD proteins has been simplified as a result of X-ray crystallographic studies which reveal secondary and tertiary structural features. However, in cases where X-ray structural information is lacking other methods are needed for elucidating subfamilies to which HADs belong.

1.4 Benefits of Neutron Crystallography

From the time of the first structure elucidations by John Kendrew and Max Perutz to today where numerous structures have been deposited into the Protein Data Bank, X-ray crystallography has found numerous applications to enzymes (19, 20). The advancements in X-ray energy sources and computational programs have allowed an exponential growth of the known enzyme structures (21-26). Although X-ray crystallography is a powerful technique it has some important limitations. In particular, X-ray diffraction does not reveal the positions of hydrogen atoms owing to their small

electron density. Although positions of the majority of hydrogen atoms in proteins can be assumed owing to the connectivity to amino acid residues, those that are in active site residues involved in catalysis exist in modified protonation states. Thus, the ability to locate these types of hydrogen can provide overwhelming insight into modes of catalysis (27-30). It is fortunate that, in contrast to the X-ray method, neutron crystallographic analysis of appropriate deuterated analogs provides data that can be used to identify protonation states of enzymes (32).

The ability to resolve the positions of hydrogen atoms within the structures of HADs is invaluable (Figure 1.4). As a result, incorporation of neutron crystallography into the study of members of this superfamily will enable a better understanding of the mechanism for the phosphoryl transfer reactions they catalyze. Because most HAD proteins contain two domains that interact through solvent linkers, their activities for the most part (C1 and C2 subfamilies) involve desolvation of the active site and substrate from bulk solvent. Although X-ray crystallography along with other techniques have already been used to observe specific conformations in which the HADs exist (10, 16, 33), results coming from Neutron crystallographic studies will serve as a third source of information about the domain dynamics and solution conformations of these enzymes.

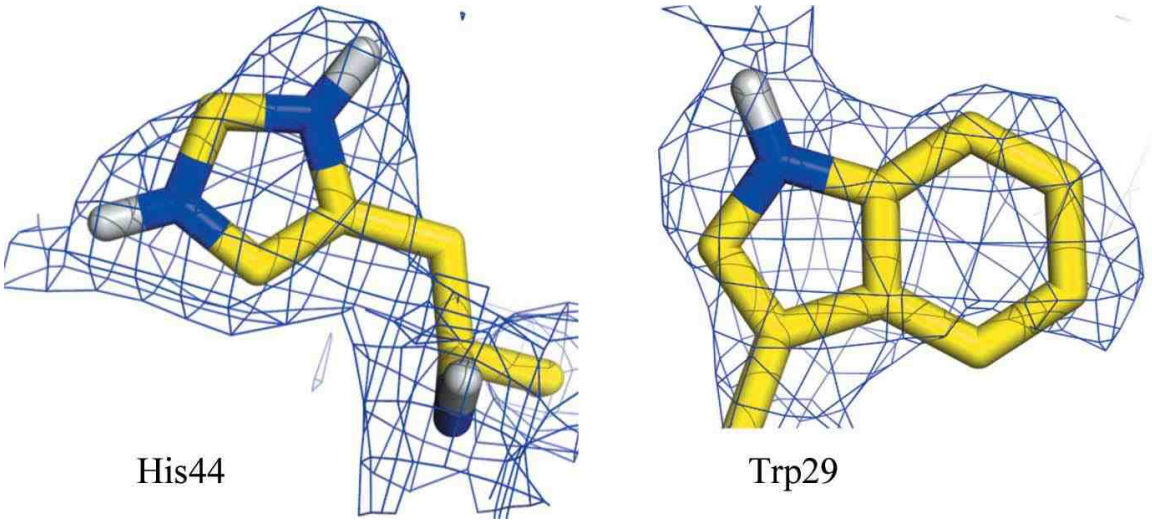

Figure 1.4: $2F_o$-F_c neutron density map (BLUE) contoured 1σ of His44 and Trp29 of KDN9P. The exchanged deuteriums are shown in white

1.5 Small Angle Xray Scattering

Small angle X-ray scattering (SAXS) is an elastic scattering technique that has been used to investigate the structures of biological macromolecules in solution. Although it has a lower resolution than do X-ray and Neutron crystallography, SAXS can be used to investigate almost biomolecules under near physiological conditions as demonstrated by its applications to exploring features of flexible protein systems, protein complexes, and systems as large as viral particles (36-38)

X-ray scattering data collected on samples in solution is a collection of scattering intensities as a function of scattering angle. The difference between electron densities of solvent and solute can be used to gain structural information about the solute. Because samples in solution lack crystal lattices and adopt random orientations, the resulting X-

ray intensities can be fit to a one dimensional curve (I(q)) averaged over all orientations Eq. 1.1 (39). Isotropic scattering is dependent on the modulus of the momentum transfer q ($q = 4\pi\sin(\theta)/\lambda$) where **2θ** is the scattering angle and **λ** incident wavelength. Data is collected for both the solvent (exact match buffer) and solute (buffer + sample(s)), and the subtracted curve is the radially average used for further processing.

$$I(q) = \langle I(q) \rangle_{\Omega} = \langle A(q) A^{*}(q) \rangle_{\Omega} \qquad \textbf{Eq. 1.1}$$

Eq. 1.1: I(s) Describes the intensity averaged over all orientations (**Ω)** where **A(s)** is a Fourier Transformation of the excess scattering length distance.

include (MM) molecular mass, (R_g) radius of gyration, (V_p) hydrated particle volume, and (d_{max}) maximum diameter of particle. The R_g value can be determined two ways including using the Guinier approximation and the Pair Distribution function Eq. 1.2-3. The R_g value is essentially the root mean square distance from the center of gravity (42,43).

$$I(s) = I(0) e^{\left(-\frac{1}{3} R_g^2 q^2\right)} \qquad \textbf{Eq. 1.2}$$

$$p(r) = \frac{r^2}{2\pi^2} \int_0^{\infty} q^2 I(q) \frac{\sin qr}{qr} dq \qquad \textbf{Eq. 1.3}$$

Eq. 1.2-3: Eq. 1.2 describes definition of Guinier equation. Equation 1.3 describes the Fourier transformation of the scattering intensity to yield the distance distribution function *p*(r)

Molecular mass can be calculated based a knowing the intensities arising from at least two known concentrations of a well-studied sample with known particle size (for example, Bovine Serum Albumin, BSA or lysozyme) (40). Possible determination of MM can also be made using the known scattering of water (41). In some instances

determination of MM is less useful than is the hydrated particle volume (V_p) owing to the dependence on concentrations of the sample. V_p is an alternative estimation of MM estimated by the electron density of the particle estimated on Parod's equation Eq. 1.4 (44).

$$V_p = \frac{2\pi^2 I(0)}{Q}. \qquad Q = \int_0^\infty q^2 I(q) \cdot dq \qquad \textbf{Eq. 1.4}$$

Eq. 1.4: Parod's equation

Once initial interpretation of scattering data has been completed, advancements in low resolution modeling makes it possible to create *ab intio* models using an array of techniques including dummy atom modeling such as DAMMIN (45) and sequence based modeling such as GASBOR (46). Implementation of other programs enable a comparison of the structure with models created by using high resolution structure data (X-ray, NMR Structures) using CRYSOL (47). Together the advancements in Synchrotron X-ray beam intensities and programs dedicated to understanding scattering profiles have allowed the SAXS technique to be utilized to gain structural information about biomolecules under near physiological conditions.

1.6 The Multi-Domain Protein Pyruvate Phosphate Dikinase

Pyruvate phosphate dikinase (PPDK) is a multi-domain protein that catalyzes the conversion of ATP, phosphate, and pyruvate into AMP, pyrophosphate, and phosphoenolpyruvate by a pathway involving two partial reactions. PPDK consist of

three separate domains including a N-terminal nucleotide binding domain reminiscent of an ATP grasp domain, C-terminal α/β barrel PEP/pyruvate binding domain, and a central domain. The two partial reactions involving ATP promoted N-phosphorylation of a histidine residue in the central domain followed by central domain swiveling and phosphoryl transfer to pyruvate bound in the C-terminal domain (Figure 1.5).

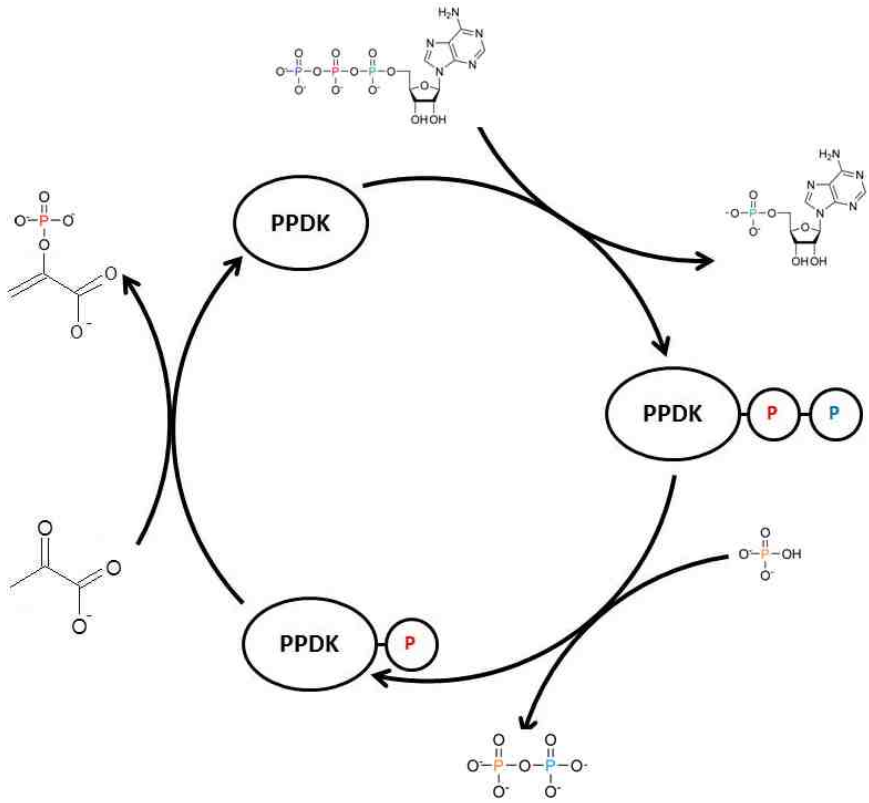

Figure 1.5: Displays the general reaction coordinates for PPDK. Each partial reaction is reversible in the presence of necessary substrates.

PPDK, whose crystal structure was solved earlier (48), is a 96 kDa homodimer created by interactions between α/β barrel domains (48). One of the two protein conformations (PDB: 1DIK) contains the central domain docked to ATP-grasp domain

(48). To crystallize the second conformation where the central domain is located near the PEP binding domain, a triple mutant of residues near the nucleotide binding domain were created in order to negate binding interactions between grasp-domain and central domain. (35).

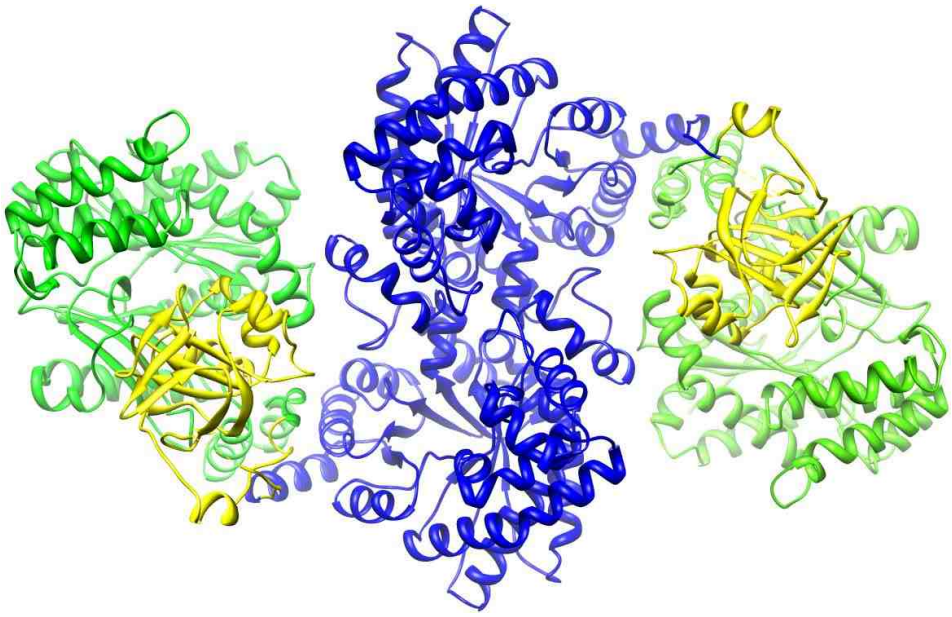

Figure 1.X: Crystal representation of PPDK dimer (PDB: 1DIK). GREEN: Nucleotide binding domain, YELLOW: Central domain and BLUE: α/β barrel PEP binding domain.

1.7. Summary

Although HADSF proteins are mechanistically unlike the colossal 96 kDa PPDK, both proteins have novel features that can be probed using techniques other than X-ray crystallography. The HADSF proteins have been shown to have evolved a very large

range of substrates for catalysis owing to the presence of a prominent and core domain robust fold. PPDKs ability to create an efficient and unique catalytic pathway through connection of domains to carry out partial reactions guided by domain swiveling is highly intriguing and warrants further investigation. An important common feature of both proteins is their use of multiple domains and ability to take advantage of the strengths of protein fold and structure. Research serving as the foundation of this dissertation focuses on structural adaptations of proteins in the HADSF and PPDK. A reoccurring theme throughout the chapters is protein stability, domain dynamics and structure.

In discussions of enzymes, a majority of attention is often given to mechanism and efficiency and not to protein fold and stability. It is known that the HADSF proteins play an important role as phosphatases in most organisms. However, little attention has been given to why these proteins in this family are so prevalent in nature. In the initial chapter of this dissertation, attention is given to studies probing the stability of the core domain as an evolutionary marker of the success of the HADs. To investigate this stability, different subfamily proteins and commonalities between all HADSF (common fold domains, and conserved residues) were explored with attention being given to determining the domain dynamic associated with catalysis.

Both PPDK and HAD proteins utilize domain movement as part of their catalytic mechanisms. In the HADSF small angle scattering was used to investigate key interactions between domains and the factors that initiate movement of the domains and how domain movement is connected to catalysis. Small angle scattering was used to investigate the catalytically intriguing domain movement of PPDK in the solution phase. Finally, neutron scattering structures of representative HADSF proteins were determined

and used to locate hydrogen interactions that lead to stability and catalytic activity. The prevailing goal of this thesis work was to investigate using modern structural techniques key evolutionary commonalities among members of protein super families that enable them to have a unique prevalence.

References

1. Maxwell Burroughs, Karen N. Allen, Debra Dunaway-Mariano, L. Aravind, Evolutionary Genomics of the HAD Superfamily: Understanding the Structural Adaptations and Catalytic Diversity in a Superfamily of Phosphoesterases and Allied Enzymes, Journal of Molecular Biology, Volume 361, Issue 5, 1 September 2006, Pages 1003-1034, ISSN 0022-2836

2. Eugene V. Koonin, Roman L. Tatusov, Computer Analysis of Bacterial Haloacid Dehalogenases Defines a Large Superfamily of Hydrolases with Diverse Specificity: Application of an Iterative Approach to Database Search, Journal of Molecular Biology, Volume 244, Issue 1, 17 November 1994, Pages 125-132, ISSN 0022-2836

3. Ridder, I. S., Rozeboom, H. J., Kalk, K. H., Janssen, D. B., and Dijkstra, B. W. (1997) Three-dimensional structure of L-2-haloacid dehalogenase from *Xanthobacter autotrophicus* GJ10 complexed with the substrate-analogue formate, *J. Biol. Chem*. *272*, 33015- 33022.

4. Hisano, T., Hata, Y., Fujii, T., Liu, J. Q., Kurihara, T., Esaki, N., and Soda, K. (1996) Crystal structure of L-2-haloacid dehalogenase from *Pseudomonas* sp. YL. An R/*â* hydrolase structure that is different from the R/*â* hydrolase fold, *J. Biol. Chem. 271*, 20322-20330.

5. Vivek Anantharaman, L Aravind, Eugene V Koonin, Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins, Current Opinion in Chemical Biology, Volume 7, Issue 1, February 2003, Pages 12-20, ISSN 1367-5931.

6. L Aravind, Michael Y Galperin, Eugene V Koonin, The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold, Trends in Biochemical Sciences, Volume 23, Issue 4, 1 April 1998, Pages 127-129, ISSN 0968-0004

7. Lahiri, S. D., G. Zhang, et al. (2002). "Caught in the Act: The Structure of Phosphorylated β-Phosphoglucomutase from Lactococcus lactis†,‡." Biochemistry **41**(26): 8351-8359.

8. Jean-François Collet, Isabelle Gerin, Mark H Rider, Maria Veiga-da-Cunha, Emile Van Schaftingen, Human l-3-phosphoserine phosphatase: sequence, expression and evidence for a phosphoenzyme intermediate, FEBS Letters, Volume 408, Issue 3, 26 May 1997, Pages 281-284

9. Seal, S. N. and Z. B. Rose (1987). "Characterization of a phosphoenzyme intermediate in the reaction of phosphoglycolate phosphatase." Journal of Biological Chemistry **262**(28): 13496-13500.

10. Lahiri, S. D., G. Zhang, et al. (2003). "The Pentacovalent Phosphorus Intermediate of a Phosphoryl Transfer Reaction." Science **299**(5615): 2067-2071.

11. Nguyen, H. H., L. Wang, et al. (2010). "Structural Determinants of Substrate Recognition in the HAD Superfamily Member d-glycero-d-manno-Heptose-1,7-bisphosphate Phosphatase (GmhB)." Biochemistry **49**(6): 1082-1092.

12. Lu, Z., D. Dunaway-Mariano, et al. (2005). "HAD Superfamily Phosphotransferase Substrate Diversification: Structure and Function Analysis of HAD Subclass IIB Sugar Phosphatase BT4131†,‡." Biochemistry **44**(24): 8684-8696.

13. Roberts, A., S.-Y. Lee, et al. (2005). "Ybiv from Escherichia coli K12 is a HAD phosphatase." Proteins: Structure, Function, and Bioinformatics **58**(4): 790-801.

14. Lu, Z., L. Wang, et al. (2009). "Structure-Function Analysis of 2-Keto-3-deoxy-D-glycero-D-galactononononate-9-phosphate Phosphatase Defines Specificity Elements in Type C0 Haloalkanoate Dehalogenase Family Members."

15. Lahiri, S. D., G. Zhang, et al. (2004). "Analysis of the Substrate Specificity Loop of the HAD Superfamily Cap Domain†,‡." Biochemistry **43**(10): 2812-2820.

16. Allen, K. N., and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: evolution of a catalytic scaffold, *Trends Biochem Sci 29*, 495-503.

17. Roberts, A., S.-Y. Lee, et al. (2005). "Ybiv from Escherichia coli K12 is a HAD phosphatase." Proteins: Structure, Function, and Bioinformatics **58**(4): 790-801.

18. Lu, Z., D. Dunaway-Mariano, et al. (2008). "The catalytic scaffold of the haloalkanoic acid dehalogenase enzyme superfamily acts as a mold for the trigonal bipyramidal transition state." Proceedings of the National Academy of Sciences **105**(15): 5687-5692.

19. Kendrew, J. C.,Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., Phillips, D. C. (1958). " A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis." Nature 181(4610): 662-666.

20. Perutz, M. F. (1963). "X-ray Analysis of Hemoglobin: The results suggest that a marked structural change accompanies the reaction of hemoglobin with oxygen." Science **140**(3569): 863-869.

21. Abola, E. Kuhn, P. Earnest, T. Stevens , R C. (2000). "Automation of X ray Crystallography." Nature *Structural Biology* **7**: 973- 977.

22. J.R. Helliwell, Protein crystallography with synchrotron radiation, Journal of Molecular Structure, Volume 130, Issues 1–2, August 1985, Pages 63-91, ISSN 0022-2860,

23. Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, *60*(12), 2126-2132

24. Sheldrick, G. M. (1990). Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallographica Section A: Foundations of Crystallography*, *46*(6), 467-473.

25. Collaborative, C. P. (1994). The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography*,*50*(Pt 5), 760.

26. Adams, P. D., P. V. Afonine, et al. (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallographica Section D **66**(2): 213-221

27. Phillips, S E. V. Schoenborn, B P. (1981) "Neutron diffraction reveals oxygen-histidine hydrogen bond in oxymyoglobin" Nature 292(5818): 81-82

28. Norvell, J., A. Nunes, et al. (1975). "Neutron diffraction analysis of myoglobin: structure of the carbon monoxide derivative." Science **190**(4214): 568-570.

29. Kossiakoff, A. A., and Spencer, S. A. (1981) Direct determination of the protonation states of aspartic acid-102 and histidine-57 in the tetrahedral intermediate of the serine proteases: neutron structure of trypsin, *Biochemistry* 20, 6462-6474.

30. Shu, F., V. Ramakrishnan, et al. (2000). "Enhanced visibility of hydrogen atoms by neutron crystallography on fully deuterated myoglobin." Proceedings of the National Academy of Sciences **97**(8): 3872-3877.

31. Niimura, N., Minezaki, Y., Nonaka, T., Castagna, J. C., Cipriani, F., Høghøj, P., ... & Wilkinson, C. (1997). Neutron Laue diffractometry with an imaging plate provides an effective data collection regime for neutron protein crystallography.*Nature Structural & Molecular Biology*, *4*(11), 909-914.

32. Bryan, T., J. M. Gonzalez, et al. (2013). "Neutron diffraction studies towards deciphering the protonation state of catalytic residues in the bacterial KDN9P phosphatase." Acta Crystallographica Section F **69**(9): 1015-1019.

33. Tremblay, L., Zhang, G., Dai, J., et al. (2005) Chemical confirmation of the oxyphosphorane-*â*-phosphoglucomutase complex, *J. Am. Chem. Soc. 127*, 5298-5299.

34. Dmitri I. Svergun, Maxim V. Petoukhov, Michel H.J. Koch, Determination of Domain Structure of Proteins from X-Ray Solution Scattering, Biophysical Journal, Volume 80, Issue 6, June 2001, Pages 2946-2953.

35. Lim, K. Read, R. Chen, C C H. Tempczyk, A. Wei, M. Ye, D. Wu, C. Dunaway-Mariano, D. Herzburg, O. (2007) "Swiveling Domain Mechanism in Pyruvate Phosphate Dikinase" Biochemistry, Volume 45, Issue 51, 14845-14853.

36. Bernadó, P., E. Mylonas, et al. (2007). "Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering." Journal of the American Chemical Society **129**(17): 5656-5664

37. Lee, K. K., H. Tsuruta, et al. (2005). "Cooperative Reorganization of a 420 Subunit Virus Capsid." Journal of Molecular Biology **352**(3): 723-735.

38. Walker, R. G., X. Deng, et al. (2014). "The Structure of Human Apolipoprotein A-IV as Revealed by Stable Isotope-assisted Cross-linking, Molecular Dynamics, and Small Angle X-ray Scattering." Journal of Biological Chemistry **289**(9): 5596-5608.

39. Mertens, H. D. T. and D. I. Svergun (2010). "Structural characterization of proteins and complexes using small-angle X-ray solution scattering." Journal of Structural Biology **172**(1): 128-141.

40. Mylonas, E., & Svergun, D. I. (2007). Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal of applied crystallography*, *40*, s245-s249.

41. Orthaber, D., Bergmann, A., & Glatter, O. (2000). SAXS experiments on absolute scale with Kratky systems using water as a secondary standard.*Journal of Applied Crystallography*, *33*(2), 218-225.

42. Guinier, A. (1939). La diffraction des rayons X aux tres petits angles: applications a l'etude de phenomenes ultramicroscopiques.

43. Glatter, O. (1977). "A new method for the evaluation of small-angle scattering data." Journal of Applied Crystallography 10(5): 415-421.

44. Porod, G., 1982. General theory. In: Glatter, O., Kratky, O. (Eds.), Small-angle X-ray Scattering. Academic Press, London, pp. 17–51.

45. D. I. Svergun (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J.* 2879-2886.

46. D. I. Svergun, M. V. Petoukhov, and M. H. Koch, _Determination of domain structure of proteins from x-ray solution scattering,_ Biophysical journal, vol. 80, no. 6, pp. 2946_2953, 2001.

47. Svergun, D., Barberato, C., & Koch, M. H. J. (1995). CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography*, *28*(6), 768-773

48. Herzberg, O., Chen, C. C., Kapadia, G., McGuire, M., Carroll, L. J., Noh, S. J., & Dunaway-Mariano, D. (1996). Swiveling-domain mechanism for enzymatic phosphotransfer between remote reaction sites. *Proceedings of the National Academy of Sciences*, *93*(7), 2652-2657.

CHAPTER TWO

Case Study: HAD Subclass II Phosphatase, Mesorhizobium loti D-glycero-D-manno-Heptose-1,7-bisphosphate Phosphatase, Stability and Structural Studies

2.1 Introduction

The HAD superfamily is one of the largest enzyme superfamilies, and it is the major supplier of organophosphate metabolite phosphohydrolases ("phosphatases") (4, 5). As described in Chapter 1, the focus of my doctoral research was to investigate underlying reasons for the robustness and prevalence of the HAD superfamily. The highly conserved catalytic domain (sometimes referred to as the "core domain") possesses a "Rossmann-like" fold (hereafter referred to as the Rossmann-fold). Based on the findings from bioinformatic analysis it has been shown that the vast majority of the family members also possess a cap domain which, as succinctly stated in the article by Pandya et al (1) "leads to a natural classification of the superfamily into different structural classes – C0 (no or minimal cap insert), C1 (α helical cap insert), C2 (α helical and β strand cap insert, further sub-divided into C2a and C2b depending on topology) and, C1+C2 (inserts in both positions), based on topology and location of the insert.".

My goal was to examine the impact of domain insertion into the Rossmann-fold core domain. As detailed in Chapter 1, the Rossmann-fold is one among only a handful of known superfolds. We posited that the intrinsic stability and plasticity of the Rossmann-fold made it possible for HAD superfamily core domain to accommodate the sequence

inserts, which ultimately evolved to the cap domains of the C1 and C2 HAD family members, which in turn lead to the explosive expansion of the HAD superfamily. The fact that the cap domain insertions occur within the catalytic platform is remarkable and at the same time, to be expected; "remarkable" because of the high potential for disruption of catalysis, and "expected" because active sites are commonly formed by loops which project from the structural framework of the enzyme (14).

In Burroughs et al (14-16) the authors divide the HAD superfamily into the type C0 proteins that have a core five-stranded β-sheet, and the rest of the superfamily which conserves the six-stranded β-sheet. As detailed in Chapter 1, in order to carry out a side-by-side analysis of the core domain stability as a function of sequence insert size (loop vs cap domain) we selected for study phosphatases from the six-stranded β-sheet Rossmann-fold core-domain division. The protein picked to represent the core domain having a "extended-loop insert" at the C2 position, is D-glycero-D-manno-heptose-1,7-bisphosphate phosphatase (GmhB). Determination of the structure-stability relationship in GmhB is the focus of this chapter.

GmhB is a monomeric protein (2,3), which is ideal for structure-stability analysis. Moreover, there are three GmhB orthologs (from *E. coli*, *B. bronchiseptica* and *M. loti*) represented in the Protein Structure Data Base (PDB accession codes: 3L8H, 3L8E, 3L8F, 3L8G and 202X). In Figure 2.1, the structures of the three orthologs are compared. Each possesses the family "squiggle motif" (colored purple), which positions the Asp nucleophile and the Asp acid/base, and at the C2 insertion position an extended loop (colored blue), which in *E. coli* and *B. bronchiseptica* binds a Zn^{2+} ion. An overlay of three ortholog structures is shown in Figure 2.2, highlighting the three active site three

loops (the Zn^{2+}-binding loop-blue, loop 2-red and loop 1-green) which function in place of a cap domain to bind the substrate-leaving group.

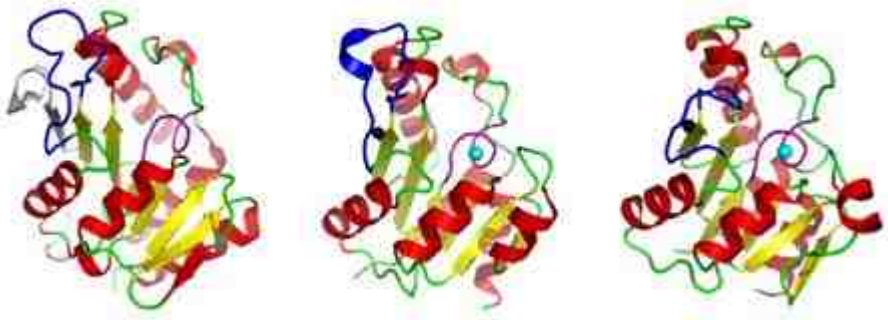

Figure 2.1: The GmhB orthologs from M. loti (left), E. coli (center), and B. bronchiseptia (right) represented in the Protein Structure Data Base (PBD accession codes: 3L8H, 3L8E, 3L8F, 3L8G AND 2O2X).

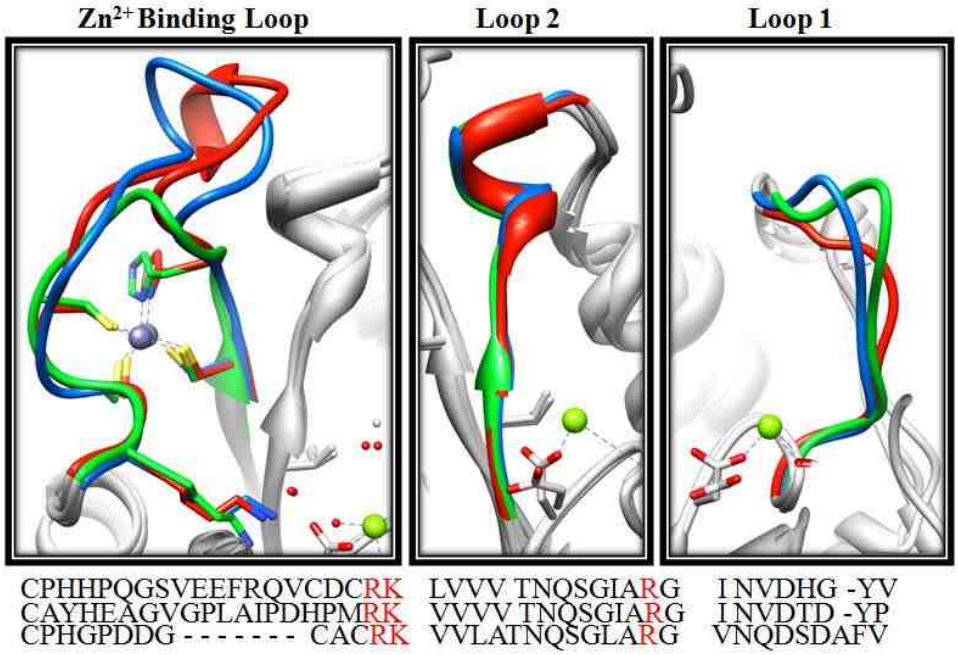

Figure 2.2: (A) An overlay of three structures (PDB accession codes, 3L8G-red, 2O2X-blue, and 3L8H-green) is shown in Figure 2.2, highlighting the collaboration between three loops (the Zn^{2+} binding loop, loop2 and loop 1) in a productive binding and in desolvation of the catalytic site. (B) shows represented loop as surface modeled on PDB accession 2O2X.

2.1.1 GmhB Function

GmhB catalysis is required in the S-layer glycoprotein and lipid A biosynthesis. The GmhB substrate, the C(1) α- or β-anomer of D-*glycero*-D-*manno*-heptose-1,7-bisphosphate, is formed as an intermediate in the D-*glycero*-D-*manno*-heptose-1α-GDP and L-*glycero*-D-*manno*-heptose-1β-ADP pathways, respectively (see Figure 2.1). The GmhB catalyzes the hydrolysis of the C(7)phosphate group of the D-*glycero*-D-*manno*-heptose-1,7-biphosphate C(1) α or β-anomer to form D-*glycero*-D-*manno*-heptose-1α-phosphate or D-*glycero*-D-*manno*-heptose-1−β-phosphate, respectively (see Figure 2.3) (2). The D-*glycero*-D-*manno*-heptose-1-phosphate anomers are activated by nucleotidation for incorporation of the *glycero*-D-*manno*-heptose unit into the glycan of the S-layer glycoprotein or lipid A.

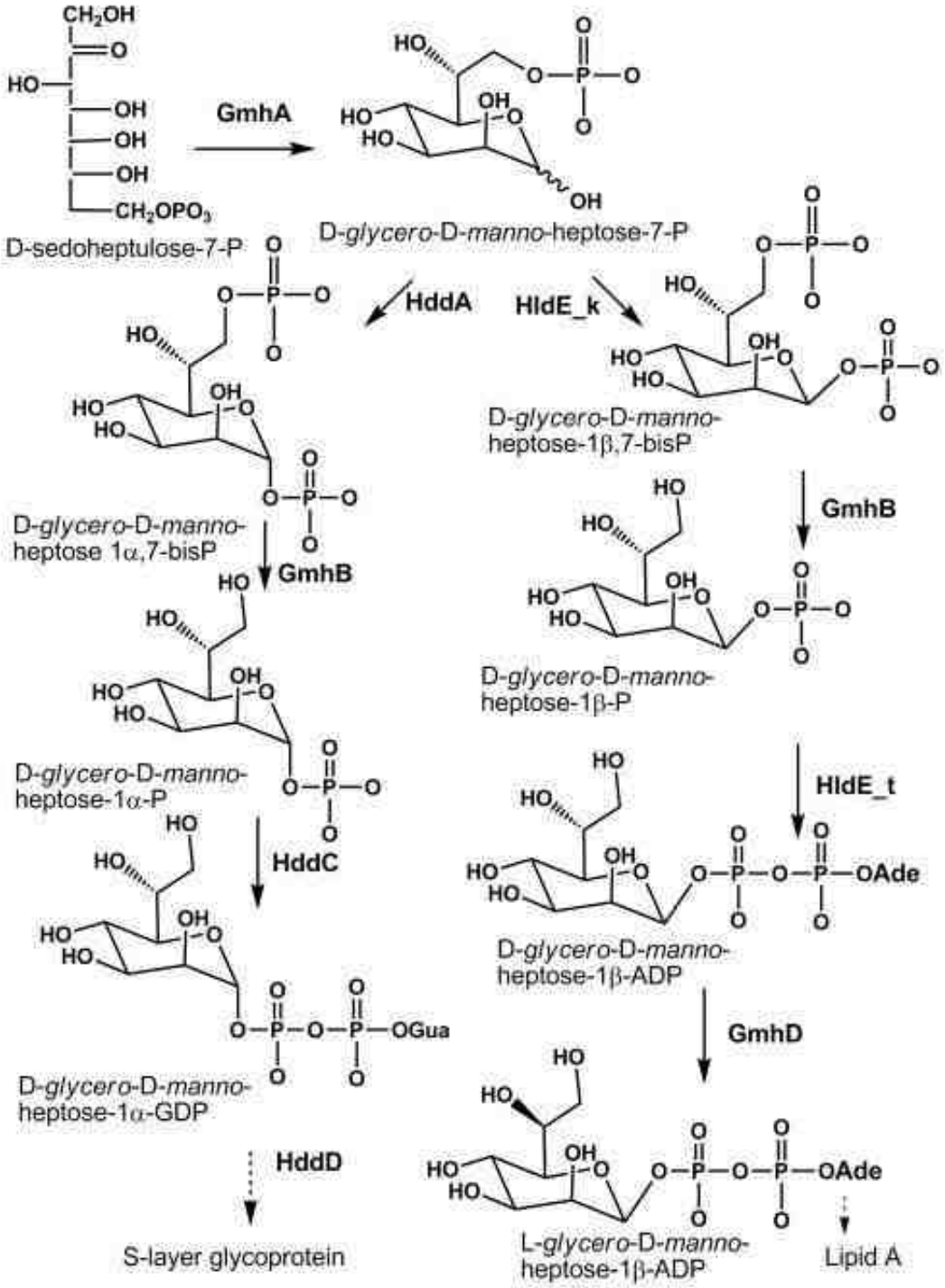

Figure 2.3: Chemical steps and enzyme catalysts of the A. thermoaerophilu D-glycero-D-manno-heptose-1α-GDP pathway and the E. coli L-glycero-D-manno-heptose-1β-ADP pathway.

2.1.2 GmhB Active Site.

The active site of *E. coli* GmhB bound with the Mg^{2+} cofactor and the D-*glycero*-D-*manno*-heptose-1β,7-bisphosphate substrate is shown in Figure 2.4. The subject of the structure-stability studies is *M. loti* GmhB, the X-ray structure of which was determined for the apo enzyme (Figure 2.5). The residues that were mutated during the course of the studies reported in this chapter are highlighted in the figure.

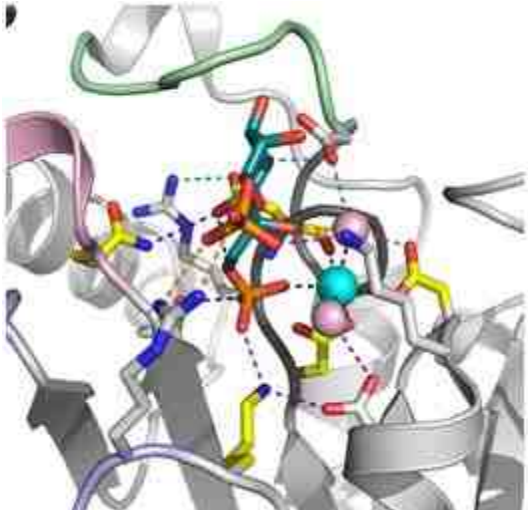

Figure 2.4: The active site of E. coli GmhB with Mg2+ cofactor (cyan sphere; water ligand pink sphere) and the substrate D-glycero-D-manno-heptose-1b,7-bisphosphate bound (stick with teal carbon atoms). Oxygen atoms are red, nitrogen atoms are blue and phosphorus atoms are orange. Side chains are shown in stick with gray are yellow carbon atoms.

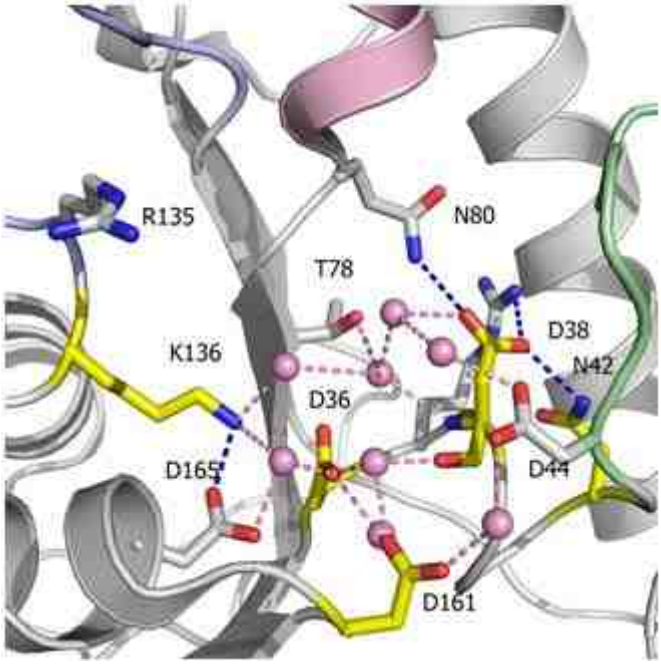

Figure 2.5: The active site of *M. loti* GmhB (*apo*). Pink spheres represent water molecules and dashed line represents hydrogen bonds. Residues in which the carbon atoms are colored yellow were replaced by site-directed mutagenesis

2.2 Material and Methods

2.2.1 Materials

All restriction enzymes and T4 ligase were purchased from Invitrogen. P*fu* Turbo polymerase was purchased from Strategene. Oligonucleotides were custom synthesized by Invitrogen. Components and buffer additives were all purchased from Sigma Aldrich if not specified. DNA sequencing was performed by MC labs and MW mass spectrometry determinations were carried out by UNM Mass Spec Facilities.

2.2.2 Cloning and Purification of M loti. GhmB

The wild-type M. *loti* GmhB clone was obtained from PSI Materials Repository (Clone ID: M1CD00090897). For purification, cell stock containing the plasmid gmhB/pSPEED in BL21 (DE3) competent cells was used. M. *loti* GmhB was purified using a modification of the procedure described by Wang et al. (2). GmhB/pSPEED/BL21 (DE3) cells were grown aerobically at 37 ºC in LB broth containing 50 µg/mL kanamycin to an $O.D_{600}$ of 0.5-0.7 and induced for production of protein by adding 4 mM L-arabinose. Following a 12 h induction period at 18 ºC , cells were harvested by centrifugation at 6,500 rpm for 10 min. Harvested cells were resuspended in 10 x weight lysis buffer (50 mM HEPES, 200 mM NaCl, 50 mM imidazole pH 7.5) and lysed using a Amico French press at 1,200 psi. the lysate was centrifuged at 20,000 rpm for 20 min. Supernatant from centrifugation was loaded onto a NTA-Ni HisTrap FF 5 mL column (GE Life Sciences) and the column was washed with lysis buffer to remove unbound protein and the protein was eluted using elution buffer (50 mM HEPES, 200 mM NaCl, 500 mM imidazole pH 7.5). Column fractions were analyzed for protein using SDS-page gel electrophoresis. Fractions containing protein were combined and dialyzed at 4 ºC against 3 changes 2L/2L/4L of 50 mM HEPES, 50 mM NaCl, and 5mM $MgCl_2$ pH 7.2. Protein yield was *ca.* 20 mg/g wet cells.

2.2.3 Site-Directed Mutagenesis

Site-directed mutagenesis (SDM) was carried out using the Quick Change PCR Strategy (Stratagene). Plasmids of WT/mutants were used as templates (mutant plasmid as template in the event multiple mutations are desired) in PCR amplifications along with

oligionucleotides, including mutated codon(s) as primer, and *Pfu* Turbo polymerase (Strategene). Recombinant mutant plasmids transformed into BL21 (DE3) were expressed initially on 10 mL culture scales. Cultures showing expression of protein were verified using either gene sequencing or mass spectrometric analysis. All mutated proteins were cultured and purified according to native protein protocol.

2.2.4 Determination of Steady-State Kinetic Constants

Phosphatase activities were determined using the EnzChek Assay (Invitrogen). Reactions were monitored at 360 nm using a Shimadzu UV1800 UV/Vis spectrometer. Reactions were carried out at 25 ºC in total volumes of 0.5 mL in 25 µL 20X buffer (1.0 M Tris-HCl, 20 mM $MgCl_2$, pH 7.5), containing 100µL 2-amino-6-mercapto-7-methyl-purine riboside (MESG), 5 µL purine nucleoside phosphorylase (PNP), and varying concentration (0.5-5.0X K_m) of phosphorylated substrates. Initial velocities, measured as a function of substrate concentration, were analyzed using Enzyme Kinetics v1.4 and equation 2.1, where V is the initial velocity, V_{max} is maximum velocity, [S] is substrate concentration, and K_m is the Michaelis constant. The k_{cat} values were calculated from V_{max}/[E] where [E] is the total enzyme concentration determined by uing the Bradford method.

$$V = \frac{V_{max}[S]}{[S]} + K_m \qquad \text{eq. 2.1}$$

2.2.5 Thermal Denaturation Monitored by Using Circular Dichroism

GmhB was thermally denatured while using Circular Dichorism (CD) at wavelength 222 nm to detect changes in secondary structure (12). Each assay solution had a total volume of 0.6 mL and contained 0.1-0.2 mg/mL protein in buffer (10 mM HEPES, 2.5 mM $MgCl_2$ pH 7.1). The CD signal was recorded at 1 ºC step intervals from 20 ºC - 90 ºC (equilibration of temperature in between each step was made by incubation for 30 s before). To determine the changes associated with complete denaturation, a full wavelength CD scan between 260 and 215 nm was conducted before and after thermal melting.

CD data were fitted to both a 2-state or 3-state denaturation model Table 2.1 (9). SigmaPlot 11.0 was used to fit and normalize the data to the equation shown in eq. 2.2, where **y** is the measured molar ellipticity (f_U for normalized graph), $\mathbf{y_o}$ is the smallest molar ellipticity (f_U for normalized graph), **a** is Δy, **x** is Temperature, $\mathbf{x_o}$ **(**$\mathbf{T_m}$**)** is the temperature at one-half **a**, and **b** is the slope associated with the transition from folded to unfolded forms.

$$y = y_o + \frac{a}{1+e^{-\left(\frac{x-x_o}{b}\right)}} \qquad \text{eq. 2.2}$$

Description	2-state model	3-state model
	$N \overset{K_{eq}}{\longrightarrow} U$	$N \overset{K_1}{\leftrightarrow} I \overset{K_2}{\leftrightarrow} U$
Equilibrium constant and total protein concentration	$K = \frac{[U]}{[N]}$ $P_T = [N] + [U]$	$K_1 = \frac{[I]}{[N]}\ K_2 = \frac{[U]}{[I]}$ $P_T = [N] + [I] + [U]$
Definition of molar fraction f_N	$\frac{1}{1+K}$	$\frac{1}{1+K_1+K_1K_2}$
Definition of molar fraction f_I	---	$\frac{K_1}{1+K_1+K_1K_2}$
Definition of molar fraction f_U	$\frac{K}{1+K}$	$\frac{K_1K_2}{1+K_1+K_1K_2}$
Fitting equation	$Y = Y_N f_N + Y_U f_U$	$Y = Y_N f_N + Y_I f_I + Y_U f_U$

Table 2.1: Equilibrium-folding models for monomeric proteins. **N** represents native folded protein, **U** represents unfolded proteins, **I** represents intermediate folded protein, **[N]** represents concentration of native protein, **[I]** represents concentration of intermediate, **[U]** represents concentration of unfolded protein, $\mathbf{P_T}$ is total protein concentration, **K** represents equilibrium for each transition, **Y** represents fitting equation for respective model.

2.2.6 Chemical Denaturation Monitored by Using Circular Dichroism

Chemical denaturation of proteins was carried out while being monitored with an Aviv model 420 Circular Dichroism at wavelengths 217, 222 nm. All samples consisted of protein (*ca.* 0.1-0.15 mg/mL), buffer (10 mM HEPES, 2.5 mM $MgCl_2$ pH 7.1), and urea (concentrations between 0.0 -7.0 M) at a total volume of 0.6 mL. Each sample was incubated at specified denaturant concentration for 10 min at room temperature to ensure equilibrium is attained. CD data were collected at specified wavelengths in triplicate with

a 5 second averaging time. To ensure complete unfolding occurred, the protein was tested for catalytic activity and full scans between 215 - 260 nm were recorded on samples treated with 0.0 and 8.0 M denaturant. The data were fit to a 2-state or 3-state model as described above and used to calculate Gibbs Free Energy (ΔG) using the extrapolation method (10,11).

2.2.7 Chemical Denaturation Monitored by Using Small Angle Scattering

Tryptophan fluorescence and Circular Dichroism probe the respective changes taking place in the amino acid environment and secondary structure of proteins. In order explore changes occurring in the tertiary structures of proteins, denaturation was monitored by using Small Angle X ray Scattering (SAXS) and the Stanford Synchrotron Radiation Lightsource on the 4-2 bio-SAXS beamline. Protein samples ranged from 3-10 mg/mL in 50 mM HEPES, 5 mM $MgCl_2$, 75 mM NaCl, pH 7.2, containing varying amounts of denaturant (urea, 0.0 - 7.0 M). The R_g values associated with SAXS for proteins were used as an indicator of changes in conformation caused by changes in the denaturant concentration. To determine R_g, data for the protein samples and corresponding buffer samples were collected and processed using the ATSAS program package (PRIMUS) (13). Values used to interpret the results (ΔG, m, d_{eq}) were calculated in the same manner as described handling CD or fluorescence data arising from chemical denaturation.

2.3 Results and Discussion

2.3.1 Chemical Denature of Native M. *loti* GmhB Unfolds According to a 3-State Model

M. *loti* GmhB was denatured using urea and monitored using three different techniques that investigate different aspects of the protein structure. Fluorescence was used to track the tryptophan environment and changes to its surroundings (emission 285 nm, excitation ~340 nm). To investigate the secondary structure of the protein, CD was used (molar ellipticity was monitored at 222 nm). Lastly to track the tertiary structure of the protein small angle X ray scattering (SAXS) was used (monitored the change in d_{max}). Table 2.1 tabulates the denaturant constants from denaturant experiments using urea. Figure 2.6 visually describes the difference between d_{eq} of the three denaturant curves.

The calculated Gibbs Free Energy of GmhB was ~-4824.00 ± 239.00 in dependent of detection technique. This amount in modest compared to some similar structures tested (6). This value tells us that the Rossmann-like HAD core domain is stable,

Observing the protein denature using different techniques allow further interpretation into the denaturing model. From Figure 2.6 we see that the denature curves show different transition regions (d_{eq}); this signifies the protein follows a more complex model that a simple 2-state model. A 3-state model was used to interpret the data, Figure 2.7. Interestingly the tertiary structure is modified, whereas the secondary structure is intact until the second transition. From this data we can imagine the protein losses the compact tertiary structure near ~2.0 M urea and the rest of the fold is lost near 3.5 M urea.

M. *loti* GmhB is a reasonable starting point to investigate the fold stability of the HAD core domain. GmhB is a capless monomer protein that unfolds through a three-state model such that the protein tertiary structure is lost and an intermediate structure is short lived before all structure is lost. This could be explained with the tertiary structure of the core domain; the topology map, Figure 2.8, shows a possible scenario in which the core domain consist of two sub-domains that are held together using intramolecular interactions. The loss of tertiary structure but not secondary structure can be explained through the dis-association of the sub-domains.

Technique	**Gibbs Free Energy ($cal*mol^{-1}$)**	**m ($cal*mol^{-1}*M^{-1}$)**	**d_{eq} (M)**
Fluorescence	-4594.9	1315.00	3.49
Circular Dichroism	-4804.7	1432.20	3.28
Small angle scattering	-5072.8	2445.10	2.07

Table 2.1: Table includes calculated constants from chemical denature. All reactions were carried out in buffer containing (20 mM HEPES, 2.5 mM $MgCl_2$ pH 7.2). Urea concentration was carried out every 0.25 M from 0.0 – 7.0.

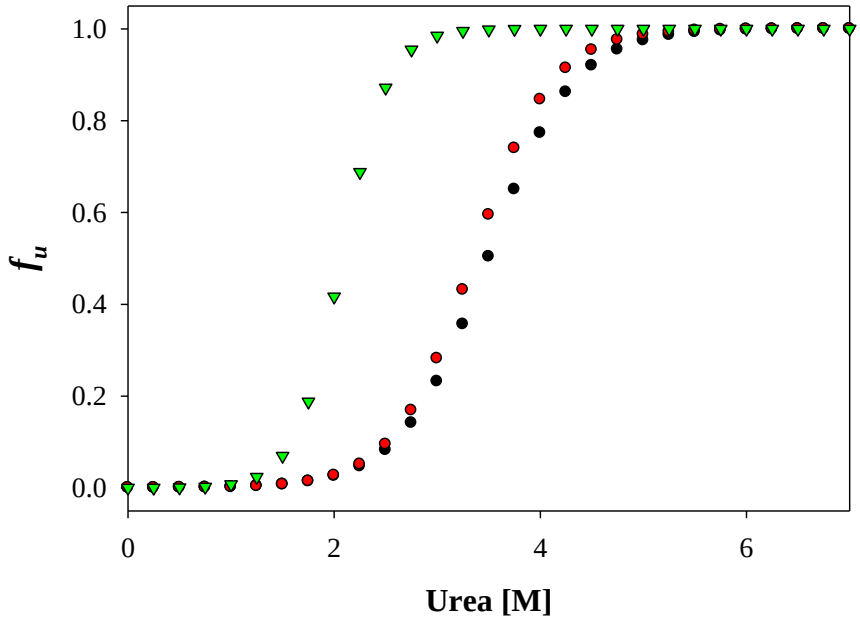

Figure 2.6: Normalized denature plot of native M. loti GmhB. Chemical denature was performed using urea as chaotrope. BLACK: native GmhB observed using Fluorescence, RED: native GmhB observed using Circular Dichroism, and GREEN: native GmhB observed using SAXS.

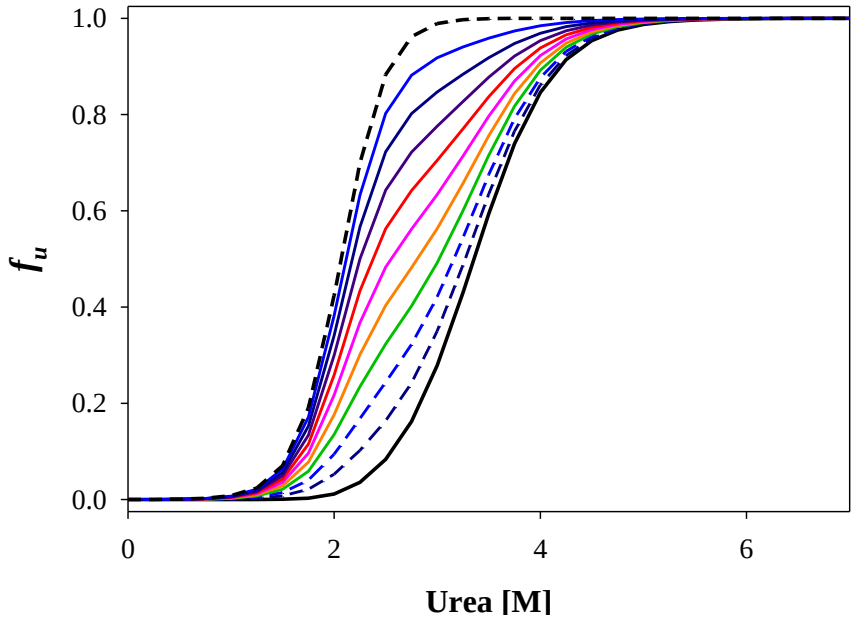

Figure 2.7: Chemical denature plots were fit separately to a 2-state model, but because the denature data observed with different techniques do not overlap suggests a more complex model. BLACK lines represent the two chemical denature plots fit to a 2-state model whereas colored lines represent fitting to a 3-state model with variation in amount of intermediate structure, each line is an increment of 10% intermediate (molar fraction).

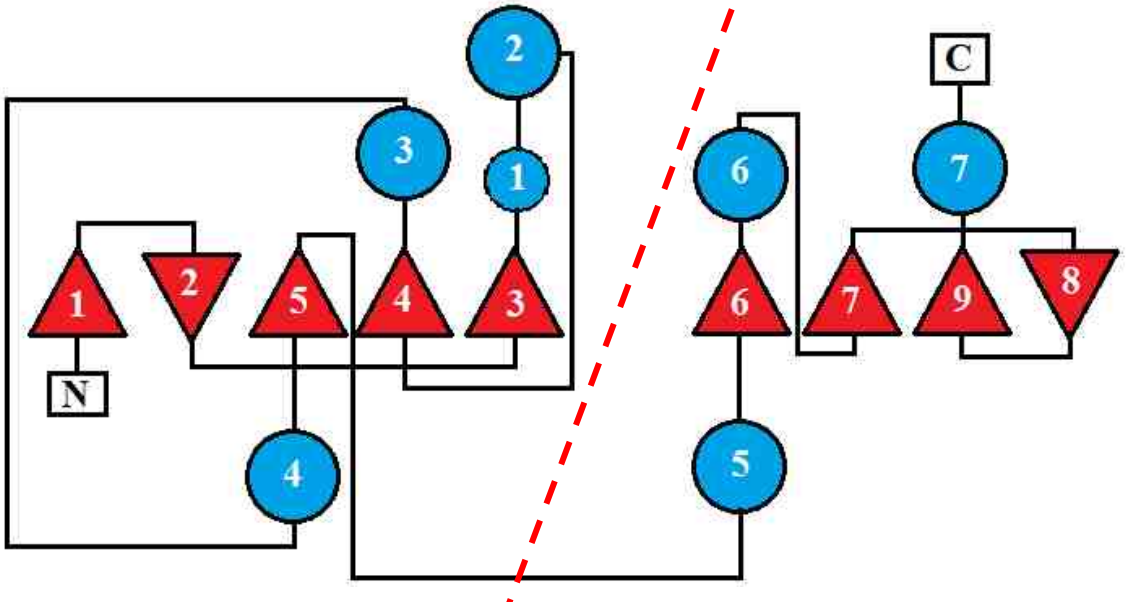

Figure 2.8: Protein topology map of M. *loti* GmhB. RED dashed line suggests interacting surfaces between sub-domains in GmhB. The loop from residue 118-136 is the only connecting sections between the two sub-domains.

2.3.2 Mg^{2+} Cofactor Show Minimal Effect on Protein Stability

In HAD proteins the active site consists of 5 conserved motif throughout all sub-classes. Also conserved is the use of Mg^{2+} cofactor which is bound by catalytic aspartates and interaction with a backbone carbonyl group. Previous studies have shown that activity is abolished without Mg^{2+} (7), experiments on activity of GmhB in the absence of cofactor support previous studies (data not shown), but what affect does the cofactor have on structure and stability? GmhB without Mg^{2+} was tested for changes in stability similar to native experiments with the exception of no Mg^{2+} in buffer.

Data collected from thermal and chemical denature experiments show similar thermal stability, but a decrease in Gibbs Free Energy from chemical denature calculations. Table 2.2 displays changes in biophysical calculated constants. As shown

the thermal stability based on T_m show very similar values 49.83 ± 0.06 and 50.80 ± 0.10 for $+Mg^{2+}$ and ΔMg^{2+} respectively. Chemical denature of GmhB presents the protein as less stable when cofactor is not present (Table 2.2 and Figure 2.9). Although the d_{eq} for the two samples are essentially the same the deciding factor to the minimized stability is the slope associated with the samples. The slope, m, has been correlated to the accessible surface are to solvent during unfolding (9). Chemical denature of the native protein has illustrated a possible scenario in which the protein consists of two sub-domains held together through intramolecular interactions. This scenario can be supported by the chemical denature data collected on the protein without Mg^{2+} cofactor. From crystal structures we can see that the cofactor is bound near the interacting area of the two "sub-domains", and when the cofactor is not present we see larger accessible surface area due to an increase in m-value which would express a loss of possible favorable interactions that keep the sub-domains together.

Protein	**T_m (ºC)**	**Gibbs Free Energy (cal^{-1} mol^{-1})**	**m ($cal*mol^{-1}*M^{-1}$)**	**d_{eq} (M)**
WT with Mg^{2+}	49.83 ± 0.06	-4804.7	-1432.2	3.35
WT Δ Mg^{2+}	50.80 ± 0.10	-3088.0	-942.55	3.28

Table 2.2: Table includes calculated constants from chemical denature. Reactions including Mg^{2+} were carried out in buffer containing (10 mM HEPES, 2.5 mM $MgCl_2$ pH 7.1). Reactions without Mg^{2+} were carried out in buffer including (10 mM HEPES, 5 mM NaCl pH 7.1). Urea concentration was carried out every 0.25 M from 0.0 – 7.0.

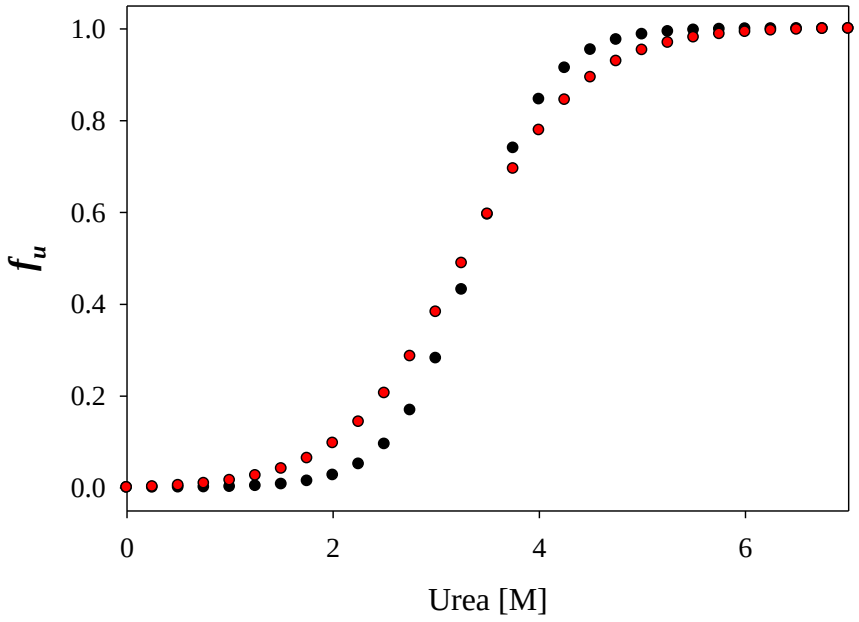

Figure 2.9: Normalized denature plot of native M. loti GmhB with/without Mg2+. Data was observed using Fluorescence. BLACK: M. *loti* GmhB with Mg^{2+}, and RED: M. *loti* GmhB without Mg^{2+}.

2.3.3Active Site Residues Play Unique Roles in Protein Stability

Engineered point mutations in protein chemistry have been widely used to experimentally investigate protein evolution. Protein activity is closely tied to protein structure and due to the intricate interactions that make a functional protein (tertiary structure). In many cases charged residues are positioned within the active site, this phenomenon gave rise to the theory of fold/function trade-off. This suggests that because the necessity of charge residues near substrate for activity the stability is compromised due unfavorable interactions between similarly charged residues. To test the idea that active site activity is traded for fold stability we created point mutations of conserved active site residues. Conserved motifs were mutated to alanine and tested for activity and stability.

Phosphatase activity is driven by nucleophilic catalysis including motif 1 (DXD). The first sets of point mutations were D36A and D38A (DXD). In both cases the protein is inactive. Table 2.3 display calculated values from chemical and thermal denature. Results show distinct changes in stability between the two aspartate mutants.

The D36A mutation shows a small decrease in stability when compared to native values, but not drastic when compared to changes due to the D38A mutation. The D38A mutation shows a substantial change in stability, both thermal and chemical stability. Although both residues are necessary for catalysis they affect protein stability differently. For completeness the third coordinating (interaction with cofactor Mg^{2+}) aspartate was mutated to alanine. From stability tests the protein showed to be as stable as the native protein suggesting no effect on fold stability.

It could be possible that changes in the active site (i.e., point mutations) can affect and possibly drive changes in surrounding residues. To test the possible additive or deleterious effects on protein stability due to single mutations of active site residues, both a double and triple mutant were created (D36A/D38A and D36A/D38A/D161A respectively).

The constructed multi-point mutation proteins performed similar to the D38A GmhB, both chemically and thermally the two proteins were destabilized. The possible rescue of stability by creating the multi-mutants was not observed, even though the decrease in charge repulsion due to multiple aspartates was created. Although the stability was not rescued from the D38A mutation, there was also no decrease in stability due to the additional point mutations.

HAD activity is derived from the conserved DXD motif. The two aspartates play different roles in the catalysis of substrates, but nucleophilic attack is promoted by the first aspartate of the conserve sequence, followed by the acid/base regeneration of the protein. From our results we see that although Asp36 is necessary for catalysis it has no effect on stability when changed. On the other hand the Asp38 shows effect on both activity but most intriguing is the effect on stability. Asp38 shows to destabilize the protein when absent which suggests an evolved necessity for not only activity but stability which could possibly refute the fold/ function trade-off.

Protein	**T_m (ºC)**	**Gibbs Free Energy (cal^{-1} mol^{-1})**	**m ($cal*mol^{-1}*M^{-1}$)**	**d_{eq} (M)**
WT	49.83 ± 0.06	-4804.7	-1432.2	3.35
D36A	48.99 ± 0.05	-4181.1	-1328.9	3.15
D38A	38.04 ± 0.14	-1603.0	-839.2	1.91
D161A	49.38 ± 0.09	-3937.7	-1203.6	3.27
D36A/D38A	40.12 ± 0.18	-1609.4	-874.3	1.84
D36A/D38A/D161A	37.35 ± 0.11	-1284.8	-792.4	1.62

Table 2.3: Table includes calculated constants from chemical denature. All reactions (both chemical and thermal denature) are carried out in following reaction buffer (20 mM HEPES, 2.5 mM $MgCl_2$ pH 7.2).

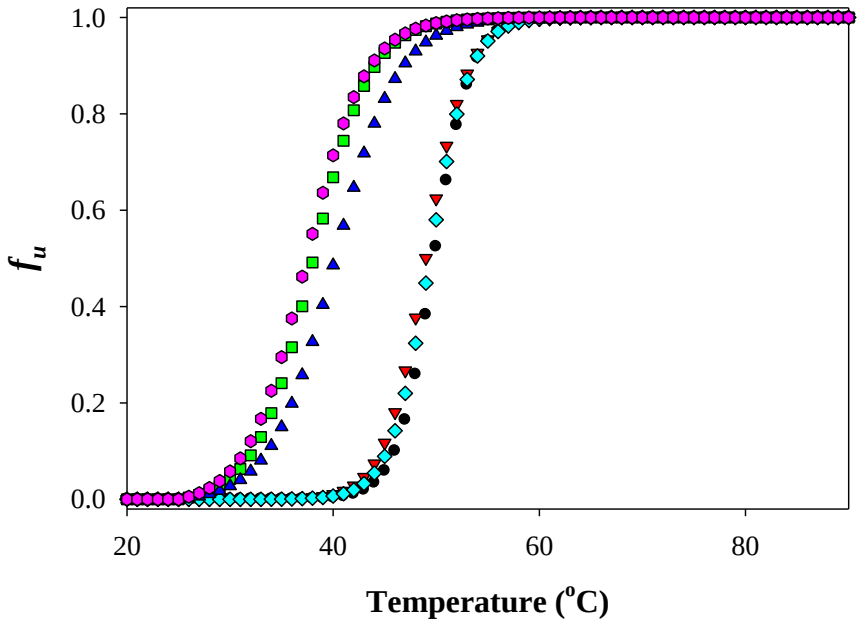

Figure 2.10: Thermal Denature of *M. loti* GmhB DxD mutants. MAGENTA: D36A/D38A/D161A GmhB, GREEN: D38A GmhB, BLUE; D36A/D38A GmhB, TEAL: D161A GmhB, RED: D36A GmhB, and BLACK: native GmhB

2.3.4 Conserved HAD Motifs and Pin Residues

As in every fold family there are many residues conserved throughout members of each family. In the HAD superfamily there are also other motifs conserved along with the acitve site DXD squiggle. For completeness, the other conserved motifs were tested for stabilty when mutated to alanine. The other conserved motifs include motif II: a conserved theronine/serine, motif III is centered around a conserved lysine. Motif IV is associated with a series of amino acids exhibiting one of three signatures: DD, GDxxxD, or GDxxxxD. The acidic residues are required for coordinating the Mg^{2+} cofactor.

Protein	T_m (ºC)	Gibbs Free Energy (cal* mol^{-1})	m (cal*mol^{-1}*M^{-1})	d_{eq} (M)
WT	49.83 ± 0.06	-4804.7	-1432.2	3.35
T78A	41.86 ± 0.11	-1851.0	-889.42	2.08
K136A	49.04 ± 0.10	-	-	
N42A	---	---	---	
N42A/K136A		-1848.7	-930.24	1.99

Table 2.4: Calculated stability variable for conserved HAD residues. All samples contained 20 mM HEPES, 2.5 $MgCl_2$ pH 7.2.

Motif IV residue Asp161 has been evaluated previously,in Table 2.4. The calculatated value for the other conserved motifs are presented. From collected data we see that the Thr78 plays a part in protein stability with a decrease in stability based on T_m and ΔG of 41.86 ± 0.11 ºC and -1851.0 cal*mol^{-1} respectively. Motif III Lys136 showed no particular change in stability (T_m) when compared to native GmhB.

Thr78 from crystal structure (PDB: 2O2X) is within H-bonding distance to amine group of Asn80 side chain, but also in vicinity of the Asn80 is active site residue Asp38 (from previous studies is showed to be key in stability). From the H-binding network created by solvent and residue side chains it could be possible that the T78A mutant disrupts the bonding network connected to Asp38 in which causes a destabilization similar to the D38A mutant.

Lys136 has been conserved and shown to bind and stabilize the phosphoryl group on the bound substrate. The lysine shows thermal stability (T_m) similar to native protein, ~0.8 °C which suggests a small to no effect on the stability.

Asparagine 42 has been shown to interact through H-bonding with Asp 38, Figure 2.11. To test if the stability of the structure is altered by the loss of bonding between the two residues the mutant was attempted. Unfortunately the point mutation Asn42 could not be produced, although the double mutant N42A/K136A was created. As previously observed the K136A mutant showed no effect on stability whereas the double mutant shows a decrease in stability (ΔG). From the collective of data between the single mutant and double mutant we can suspect that the decrease in stability is due to the Asn42 rather than the Lys136 although we cannot overlook the possibility that the decrease is an additive effect of the two point mutations.

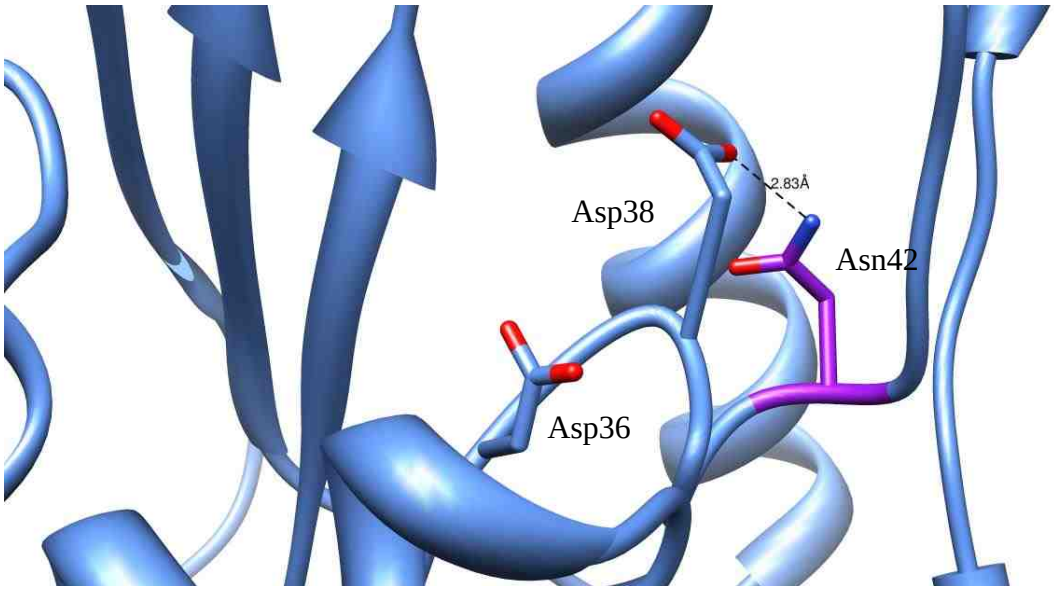

Figure 2.11: Ribbon diagram from M .loti GmhB (PDB 2O2X) displaying the interaction of Asp38 and pin residue Asn42.

The HAD superfamily has shown to be a prominent staple in phosphatase activity. The construct of the structure has allowed the protein to separate catalysis and substrate recognition through separating phosphoryl binding and metabolite binding. The

conserved residues are notably associated with the phosphate binding and the active site. From our stability studies we have observed some conserved residues that play a role in not only activity but stability (Asp38, Thr78), whereas other conserved motifs play small to no role in stability (Lys136, Asp36, and Asp161). The pin residues when mutated in accompany with K136A shows a noticeable decrease in stability similar to D38A. This decrease cannot be tracked to the Asn mutation without knowing the affect due to the single point mutation, but when compared to other residues that interact with D38A (T78A) we do in fact see a decrease in stability which supports the idea that Asn42 plays a role in the stability of the protein.

2.3.5 Effect of Side Chain on Active Site Stability

It is customary to mutate active site residues to homologous amino acids for X-ray crystallography to mimic ligand binding without proceeding through reduction of substrate to product. For this reason three mutants were created to investigate the effects on stability due to changes inside chain residues at the second aspartate in the DXD motif. The aspartic acid was mutated to glutamic acid, asparagine and previously mentioned alanine. Table 2.5 and Figure 2.12 show that changes in stability constants when mutations were tested. When changed to alanine the interaction of D38A is lost between it and the pin residue

Protein	T_m (°C)	Gibbs Free Energy (cal^{-1} mol^{-1})	m ($M*cal*mol^{-1}$)	d_{eq} (M)
WT	49.8 ± 0.1	-4804.7	1432.2	3.35
D38A	38.0 ± 0.1	---	---	---
D38E	41.7 ± 0.2	---	---	---
D38N	46.6 ± 0.4	---	---	---

Table 2.5: Calculated stability variables for point mutation variations at residue Asp38. Thermal denature buffer was 20 mM HEPES, 2.5 $MgCl_2$, pH 7.2. Sample concentration ~0.1 mg/mL.

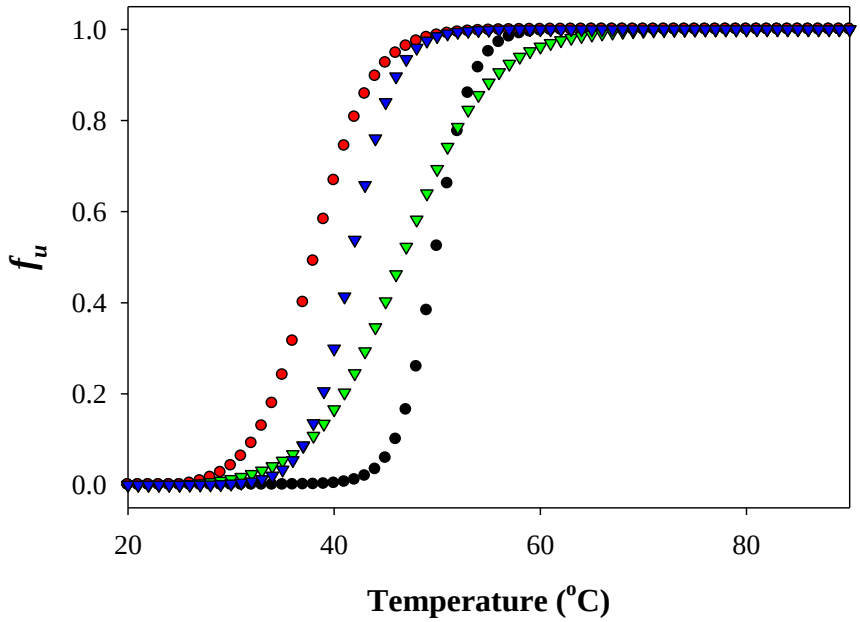

Figure 2.12: Thermal Denature of GmhB with different amino acids at Asp38. Black: Native, RED: D38A, BLUE: D38N, and GREEN: D38G

2.4 Conclusion

Mesorhizobium *loti* D-glycero-D-manno-Heptose-1,7-bisphosphate Phosphatase, GmhB is a subclass 0 HAD protein. The protein catalyzes the dephosphorylation of D-*glycero*-D-*manno*-heptose 1,7-biphosphate. Because of the prior knowledge about GmhB

(bioinformatics, structure, etc.) (1-3) it is natural to use GmhB as a starting point in investigating the stability of the HAD proteins. The overarching goal is to identify possible reasons for the prominences and reliability of HAD members used as a majority of phosphatases in organisms.

Results collected from GmhB studies have identified possible trends in stability based on the structure of the protein especially the active site residues. We have shown that GmhB unfolds through a 3-state unfolding model with a complete loss of tertiary structure before the loss of secondary structure. The relative stability of the protein is modest when compared to similar sized proteins (6). Conserved residues shows either no change to structural stability whereas a select few decreased the stability base on both T_m and ΔG; residues associated with Asp38 (bonding networks or H-bonding directly) show similar decrease as the D38A point mutation.

A majority of HAD proteins incorporate large sub-domains to help increase binding to native substrate. To better understand the structural stability of the HAD family a model of stability for the conserved active site domain (core domain) needs to be established. GmhB has been shown through bioinformatics to be a capable model for the standard core domain.

References

1. Pandya, C., S. Brown, et al. (2013). "Consequences of domain insertion on sequence-structure divergence in a superfold." Proceedings of the National Academy of Sciences **110**(36): E3381-E3387.

2. Wang, L., H. Huang, et al. (2010). "Divergence of Biochemical Function in the HAD Superfamily: d-glycero-d-manno-Heptose-1,7-bisphosphate Phosphatase (GmhB)." Biochemistry **49**(6): 1072-1081.

3. Nguyen, H. H., L. Wang, et al. (2010). "Structural Determinants of Substrate Recognition in the HAD Superfamily Member d-glycero-d-manno-Heptose-1,7-bisphosphate Phosphatase (GmhB)." Biochemistry **49**(6): 1082-1092.

4. Collet, J. F., van Schaftingen, E., and Stroobant, V. (1998) A new family of phosphotransferases related to P-type ATPases. Trends Biochem. Sci. 23, 284.

5. Koonin, E. V., and Tatusov, R. L. (1994) Computer Analysis of Bacterial Haloacid Dehalogenases Defines a Large Superfamily of Hydrolases with Diverse Specificity: Application of an Iterative

6. Filimonov, V. V., J. Prieto, et al. (1993). "Thermodynamic analysis of the chemotactic protein from Escherichia coli, CheY." Biochemistry **32**(47): 12906-12921.

7. Lu, Z., D. Dunaway-Mariano, et al. (2005). "HAD Superfamily Phosphotransferase Substrate Diversification: Structure and Function Analysis of HAD Subclass IIB Sugar Phosphatase BT4131†,‡." Biochemistry **44**(24): 8684-8696

8. Myers, J. K., C. Nick Pace, et al. (1995). "Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding." Protein Science **4**(10): 2138-2148.

9. Jad Walters, Sara L. Milam, A. Clay Clark, Chapter 1 Practical Approaches to Protein Folding and Assembly: Spectroscopic Strategies in Thermodynamics and Kinetics, In: Michael L. Johnson, Jo M. Holt and Gary K. Ackers, Editor(s), Methods in Enzymology, Academic Press, 2009, Volume 455, Pages 1-39,

10. Pace CN. Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods Enzymol.1986;131:266–280.

11. Greene RFJ, Pace CN. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. J Biol Chem. 1974;249:5388–5393.

12. Provencher, S. W., & Gloeckner, J. (1981). Estimation of globular protein secondary structure from circular dichroism. *Biochemistry*, *20*(1), 33-37.

13. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H., & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, *36*(5), 1277-1282.

14. Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D., & Aravind, L. (2006). Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *Journal of molecular biology*, *361*(5), 1003-1034.

15. Allen, K. N., & Dunaway-Mariano, D. (2004). Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends in biochemical sciences*, *29*(9), 495-503.

16. Allen, K. N. and D. Dunaway-Mariano (2009). "Markers of fitness in a successful enzyme superfamily." Current Opinion in Structural Biology **19**(6): 658-665.

CHAPTER THREE

Case Study: Stability of HAD Subtype II Phosphatase YbiV

3.1 Introduction

Escherichia *coli* K12 YbiV is a sugar phosphatase belonging to the Haloacid Dehalogenase Superfamily (HAD). The function in E. *coli* has not yet been discovered. From substrate screening and bioinformatics the protein is regarded as a sugar phosphatase that prefers ribose-5-phosphate and slightly less glucose-6-phopshate (1). It is hypothesized that the protein could be associated with regulation and recycling of molecules in the cell or possibly included in stress induced production of metabolites similar to trehalose production (2).

YbiV belongs to the subclass II of the HAD family. Characteristic of HAD proteins is a core Rossmann-like domain housing conserved active site residues directing phosphate binding. The subclass is defined by a α/β domain insert between motif II and III. Because the core domain binds the phosphate group for dephosphorylation, the superfamily has evolved inserts to increase binding affinity to substrates. The large α/β domain insert of the subclass II family has been shown to help bind alcohol leaving group of substrate (1, 3-5).

In previous chapters we looked at the fold-fitness of GmhB, a C2 protein, is a "capless" HAD protein that will be used as a model for stability of the core domain of HAD proteins. In the HAD family, the majority of HAD proteins include a cap insert responsible for substrate binding. How do these large insert affect the stability of the core

domain? To explore the effect of domain inserts on the stability of the HAD proteins we will characterize YbiV in terms of fold fitness and robustness. Because previous stability studies have uncovered possible trends in the HAD family we will look into the stability of the protein in terms of Mg^{2+} binding, conserved active site residues and active site bonding networks Figure 3.1 . In order to determine the most complete picture possible a multitude of techniques will be applied.

Because proteins fold in complex structures through secondary interactions as well and complex systems; denature studies will include 3 techniques (fluorescence, circular dichroism, and small angle X ray scattering) to investigate the change in conformation of different ordered structures (1º, 2º, and 3º respectively). All denature data will be fit to either a 2-state or 3-state model.

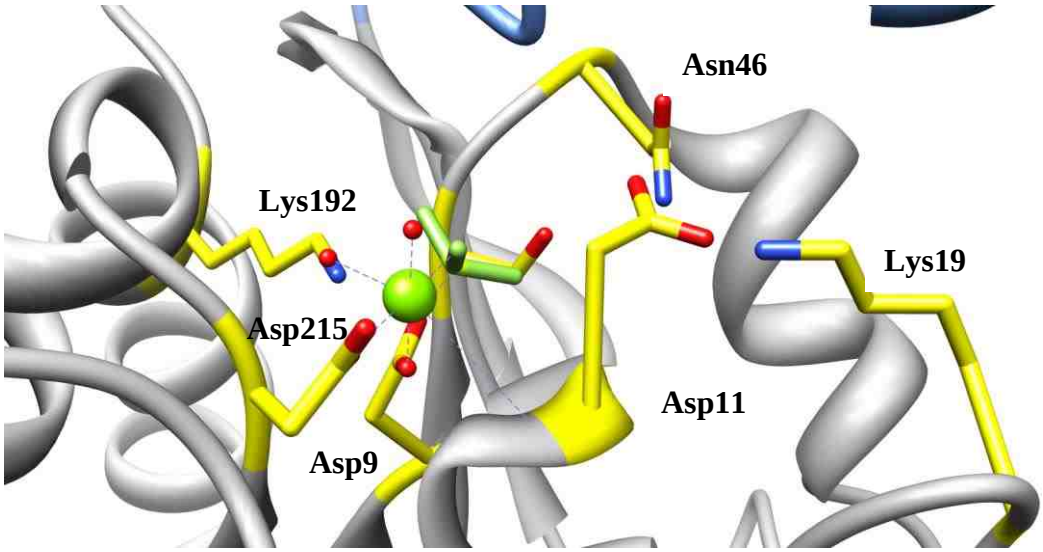

Figure 3.1: X-ray crystal of YbiV (PDB: 1RLO) with residues colored in yellow in which were mutated to investigate change in stability.

3.2 Material and Methods

3.2.1 Materials

All restriction enzymes and T4 ligase were purchased from Invitrogen. *Pfu* Turbo polymerase was purchased from Stratagene. Oligonucleotides were custom-synthesized from Invitrogen. Components and buffer additives were all purchased from Sigma Aldrich if not specified. To verify correct sequence, DNA sequencing was performed by MC labs and MW mass spec determination was carried out by UNM Mass Spec Facilities.

3.2.2 Cloning and Purification of E. *coli* YbiV

The wild-type YbiV protein was cloned using PCR amplification from K-12 strain E. *coli* (ATCC). PCR experiment used K-12 E. *coli* as a template with custom oligonucleotides incorporating restriction sites for future ligation into vector, *Pfu* Turbo polymerase (Stratagene) was used for amplification. The forward primer included a NheI restriction site whereas the reverse primer included a XhoI restriction site. The amplified PCR product was digested with previously mentioned restriction enzymes and ligated into digested vector pET-28a (Novagen) using T4 ligase (NEB). The cloned genes were transformed into BL21* (DE3) competent cells (Invitrogen). Colonies were tested for expression and verified by gene sequencing upon expression.

YbiV/pET-28a/BL21 (DE3) cells were grown aerobically at 37 ºC containing 50 µg/mL kanamycin in LB broth to an $O.D_{600}$ 0.6-0.9 and induced for production of protein by adding 0.4 mM isopropyl-β-D-galactopyranoside (IPTG). Following a 12 hour

induction period at 18 ºC, cells were harvested by centrifugation at 6,500 rpm for 10 minutes. Harvested cells were resuspended in 10x w/v lysis buffer (50 mM HEPES, 200 mM NaCl, 50 mM imidazole pH 7.5). Resuspended cells were lysed using a Amico French press at 1,200 psi, lysate was centrifuged at 20,000 rpm for 20 minutes. Supernatant from centrifugation was loaded onto a NTA-Ni HisTrap FF 5 mL column (GE Life Sciences). Column was washed with lysis buffer to clean any un-bound protein and bound protein was eluted using elution buffer (50 mM HEPES, 200 mM NaCl, 500 mM imidazole pH 7.5). Column fractions were tested for clean protein using SDS-page gel; all fractions containing clean protein samples were combined and dialyzed at 4 ºC against 3 rounds ,2L/2L/4L, of 50 mM HEPES, 50 mM NaCl, 5mM $MgCl_2$ pH 7.2. Protein yield was ~15 mg/g wet cells.

3.2.3 Site-Directed Mutagenesis

Site directed Mutagenesis was carried out using the Quick Change PCR Strategy (Stratagene). WT/mutants plasmid were used as template (mutant plasmid as template in the event multiple mutations are desired) in PCR amplification, oligionucleotides including mutated codon as primer, and *Pfu* Turbo polymerase (Stratagene). Recombinant mutant plasmids were transformed into BL21 (DE3) and tested in small scale cultures for expression. Cultures showing expression of protein were verified using either gene sequencing or mass Spec. All mutated proteins were cultured and purified according to wild-type (WT) protocols.

3.2.4 Mg^{2+} Chelatating Protocol

To investigate the changes in stability due to the absence of Mg^{2+} cofactor the protein needs to be purified and cofactor chelated from sample. M. *loti* GmhB was purified similar to native protocol with the following exceptions: After elution of protein from NTA-Ni column, sample is equilibrated with 15 mM EDTA. Protein sample is incubated for 30 minutes on ice with occasional disruption. Protein/EDTA solution was dialyzed against buffer excluding Mg^{2+}, buffer was chelexed against BioRad Chelex ® 100 resin, (25 mM HEPES, 50 mM NaCl, pH 7.2. Three rounds of dialysis was carried out to ensure absolute loss of Mg^{2+}.

3.2.5 Determination of Steady-State Kinetic Constants

The K_{cat} and K_m values for native and mutant YbiV and substrate ribose-5-phosphate were determined as described in Chapter 2.

3.2.6 Thermal Denature Using Circular Dichroism

Circular dichroism was used to observe the secondary structure of YbiV as a function of temperature. Procedures and fitting models are described in Chapter 2.

3.2.7 Chemical Denature of Ordered Protein Structure

Three techniques were used to investigate the ordered structure of YbiV. Tryptophan fluorescence was used to investigate the aromatic (mainly tryptophan)

surroundings and local polarity. To observe changes in the secondary structure of the protein circular dichroism was used at wavelengths 222 and 217 nm. For the tertiary structure small angle scattering was implemented; SAXS was used to generate a calculated solution volume in which was used to identify the general shape of the solution conformation of the protein. Procedures and fitting models for chemical denature experiments are discussed previously in Chapter 2.

3.3 Results and Discussions

3.3.1 Native E. *coli* YbiV Stability

Fold-fitness of the native YbiV protein was experimentally observed using chemical and thermal denature techniques. Thermal denature monitored only changes in secondary structures whereas chemical denature occurred using three detection techniques to ensure the correct unfolding model was used. Table 3.1 displays variables used to rate protein stability.

Results from denature experiments describe the protein to be very stable with an average ΔG of 10.839 kcal*mol^{-1} and T_m ~55.1 ºC. The stability when compared to M. *loti* GmhB is significantly higher; this result suggests that the protein is more stable when the cap domain is present. Similar to the M. *loti* GmhB, According to data collected from tryptophan fluorescence and circular dichroism YbiV folds through a 2-state model, Figures 3.2-3.8. Because the normalized data overlay between the two techniques we assume the protein unfolds from native to denatured without the presence of an

intermediate structure. Data was collected with a third detection technique to observe the loss of tertiary structure.

Technique	ΔG (cal*mol^{-1})	m (cal*mol^{-1}*M^{-1})	d_{eq} (M)
Fluorescence	-8207.7	-1808.8	4.53
Circular Dichroism (CD)	-5685.4	-1327.8	4.28
Small angle X ray scattering (SAXS)	-18626	-5162.8	3.60

Table 3.1: Comparison of stability variable of native E. *coli* YbiV using different detection techniques. Buffer in all experiments was 20 mM HEPES, 2.5 mM $MgCl^{2+}$ pH 7.20 with increasing urea concentration. protein concentration was ~10 µM in all experiments.

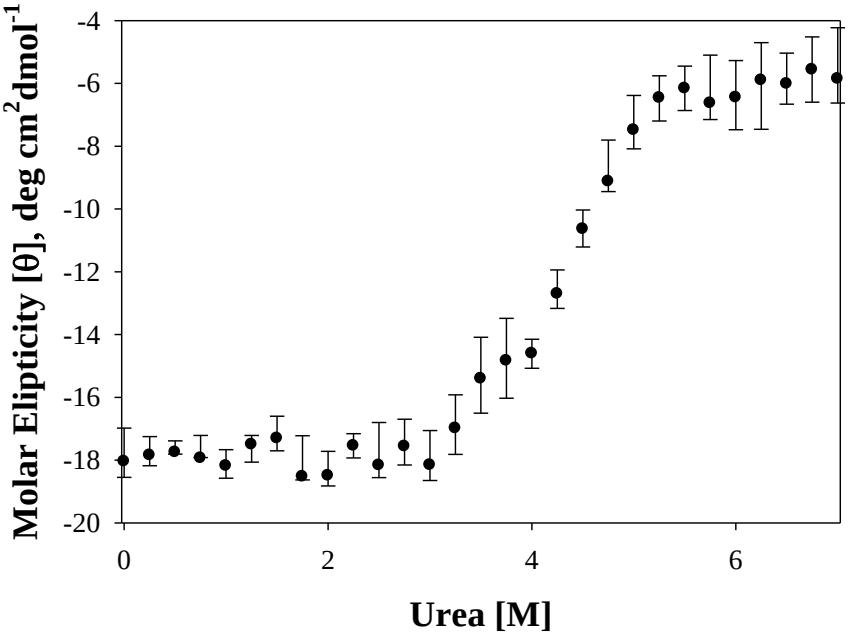

Figure 3.2: Four chemical denature sets were collected for YbiV using CD. Standard deviation from average is shown for CD detection at 222 nm. Buffer: 20 mM HEPES, 2.5 mM $MgCl^{2+}$ pH 7.20 with increasing urea concentration. Protein concentration was 0.1 mg/mL.

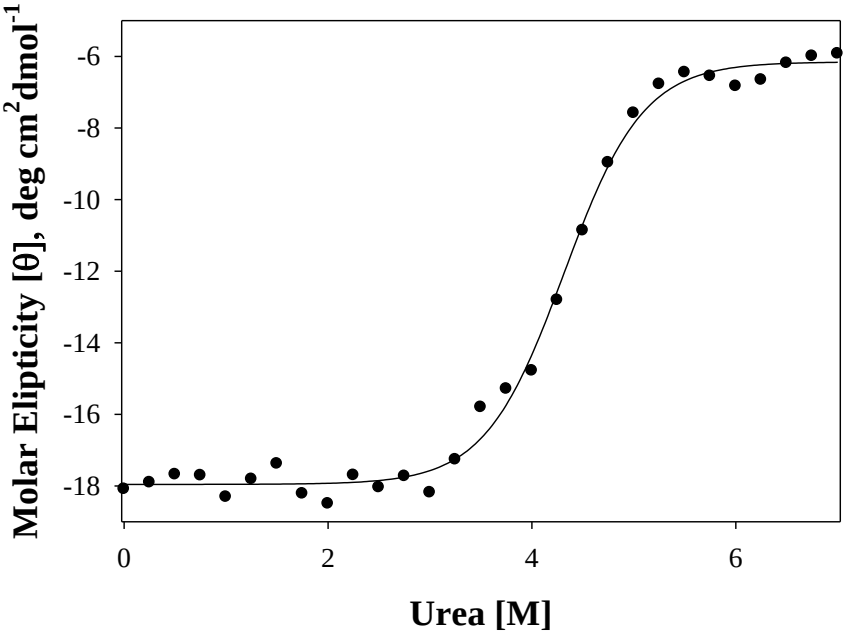

Figure 3.3: Fit line with averaged CD data for chemical denature of YbiV. Fit is a 4-parameter sigmoidal fit with variables: **a**= 11.8014, **b**= 0.3879, $\mathbf{x_0}$= 4.3157, and $\mathbf{y_o}$= -17.9573

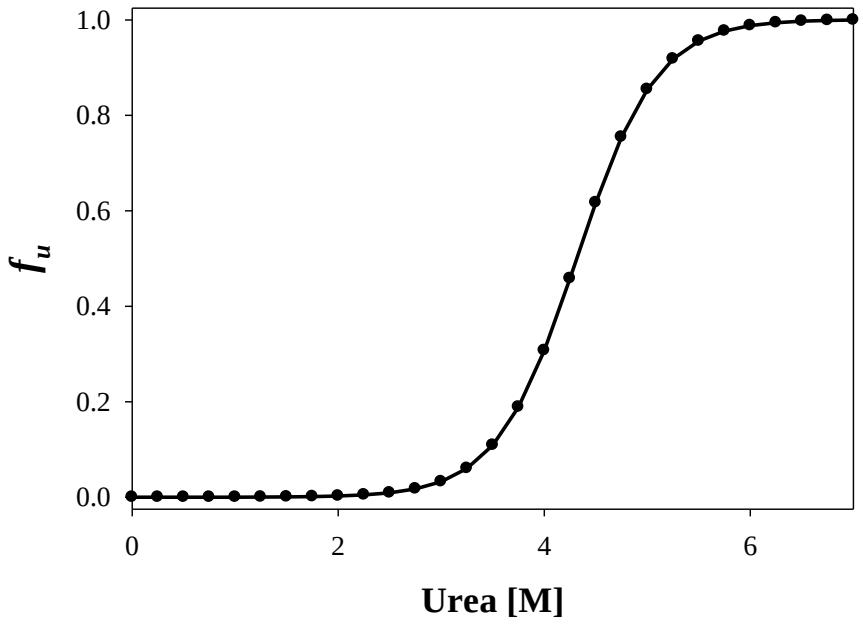

Figure 3.4: Normalized fit of chemical denature data using CD.

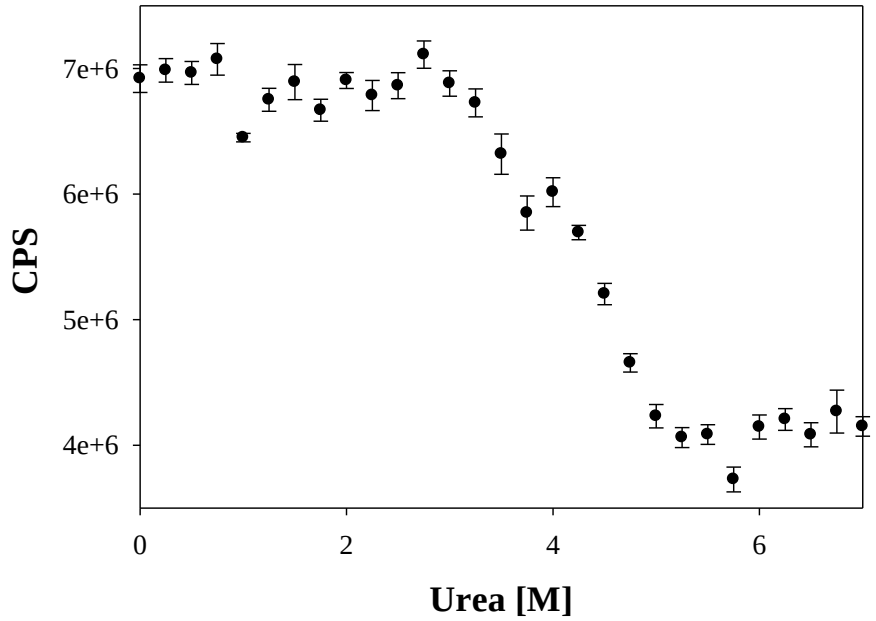

Figure 3.5: Four chemical denature sets were collected for YbiV using tryptophan fluorescence. Standard deviation from average is shown for fluorescence detection at excitations: 280 and 285 nm, emission: 340 and 352 nm. Buffer: 20 mM HEPES, 2.5 mM $MgCl^{2+}$ pH 7.20 with increasing urea concentration. Protein concentration was ~2-5 µM.

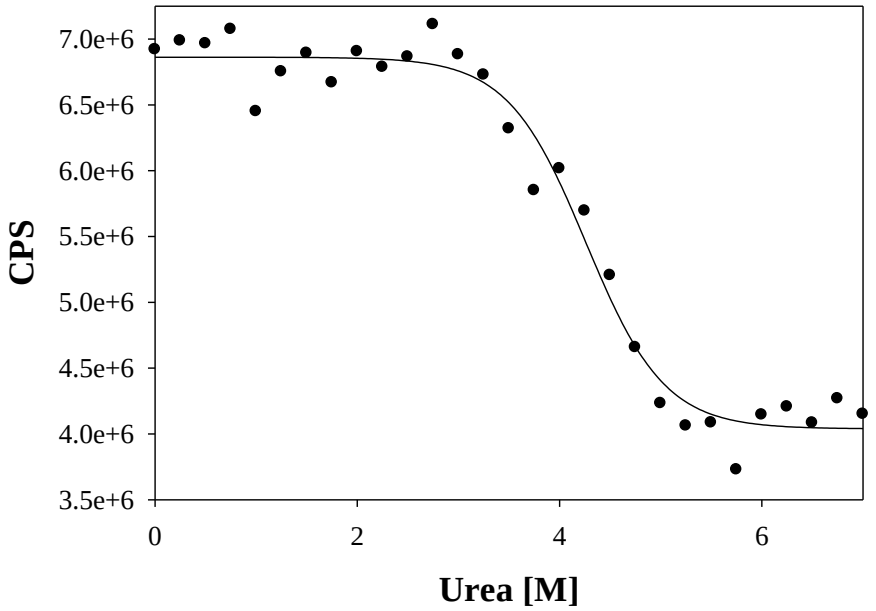

Figure 3.6: Fit line with averaged tryptophan fluorescence data for chemical denature of YbiV. Fit is a 4-parameter sigmoidal fit with variables: **a**= 2823502, **b**= -0.3876, $\mathbf{x_0}$= 4.2624, and $\mathbf{y_o}$= 4038595.

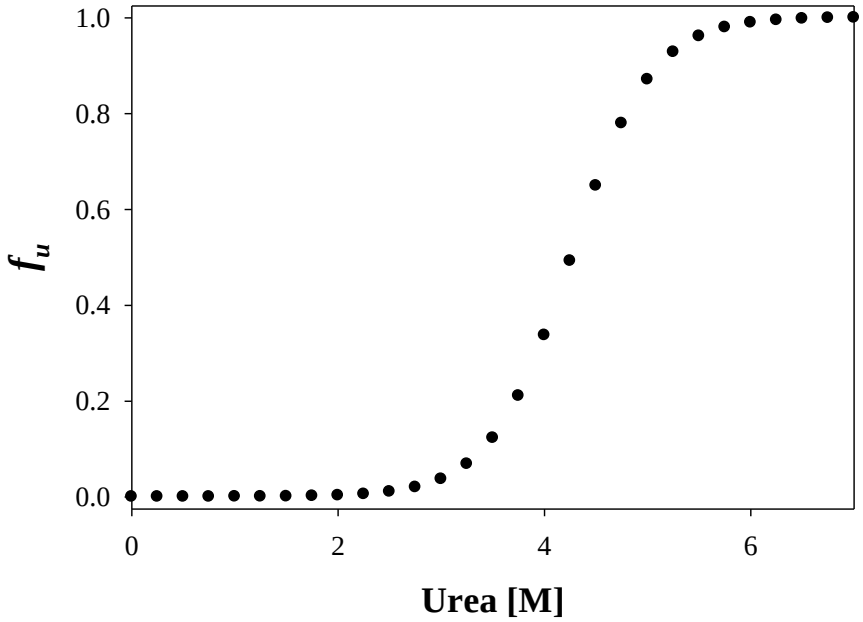

Figure 3.7: Normalized fit according to a 4-parameter fit of chemical denature of YbiV using tryptophan fluorescence. Excitation: 280, 285 nm and emission: 340 and 352 nm.

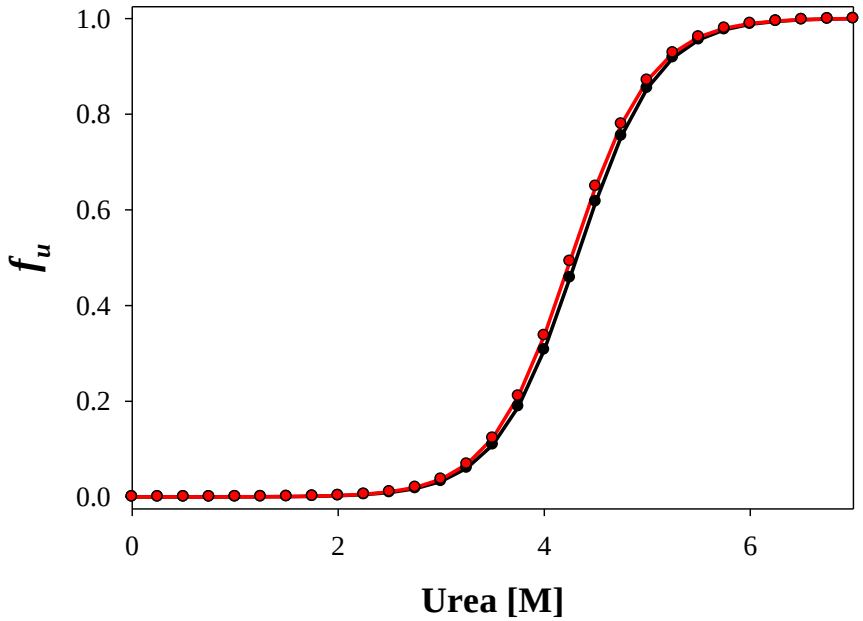

Figure 3.8: Chemical denature of YbiV using 2 different techniques to investigate correct fitting model of denature data. BLACK: circular Dichroism and RED: Fluorescence.

Small angle X-ray scattering data was collected to observe changes in tertiary structure as a function of urea. From the previous chemical denature data sets we can see that the data overlap which suggest a 2-state unfolding model, but with the incorporation of SAXS data Figure 3.9-3.15 we can see YbiV possibly unfolds through a 3-state model (due to absence of overlap of titration data from the three techniques.

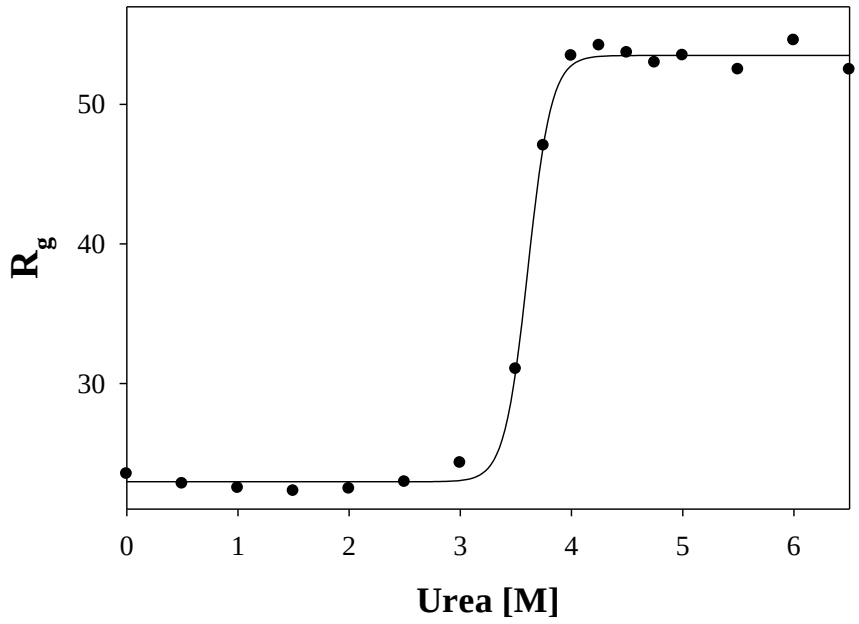

Figure 3.9: Fit line with small angle X ray scattering data for chemical denature of YbiV. Fit is a 4-parameter sigmoidal fit with variables: **a**= 30.5510, **b**= 0.1050, $\mathbf{x_0}$**=** 3.6083, and $\mathbf{y_o}$**=** 22.9548. R_g was calculated for each [urea] using PRIMUS.

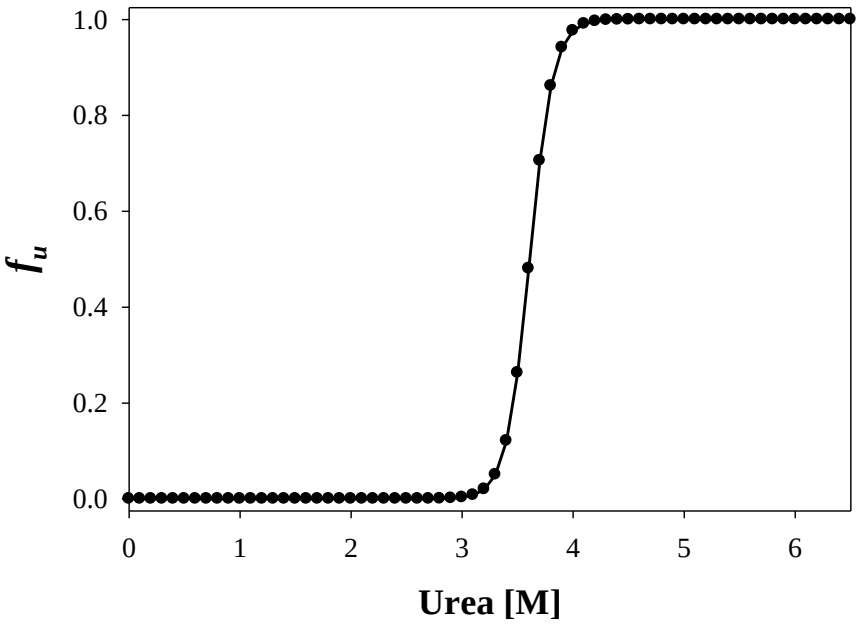

Figure 3.10: Normalized fit according to a 4-parameter fit of chemical denature of YbiV using SAXS.

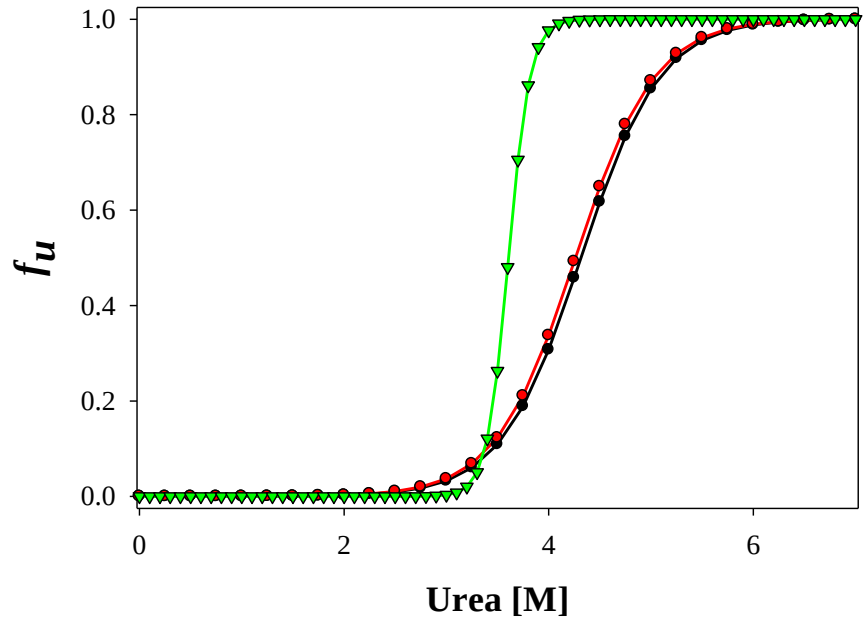

Figure 3.11: Chemical denature of YbiV using three detection techniques. GREEN: SAXS, RED: fluorescence, and BLACK: circular dichroism.

From Figure 3.11 we can see the SAXS data does not overlap with the other two techniques suggesting the protein unfolds according to a 3-state model. With further investigation into collected denature data from both fluorescence and CD (particularly fluorescence) Figure 3.2 and 3.5; we can see a possible accumulation of intermediate near 3.75 M urea. This possible accumulation of intermediate is near 0.2 molar fraction unfolded protein. Figure 3.12 represents a universal fit to SAXS data and overlaid CD/fluorescence and describes possible intermediate fraction 0.2 (6). Data collected favor a stable 3-state model of unfolding for YbiV.

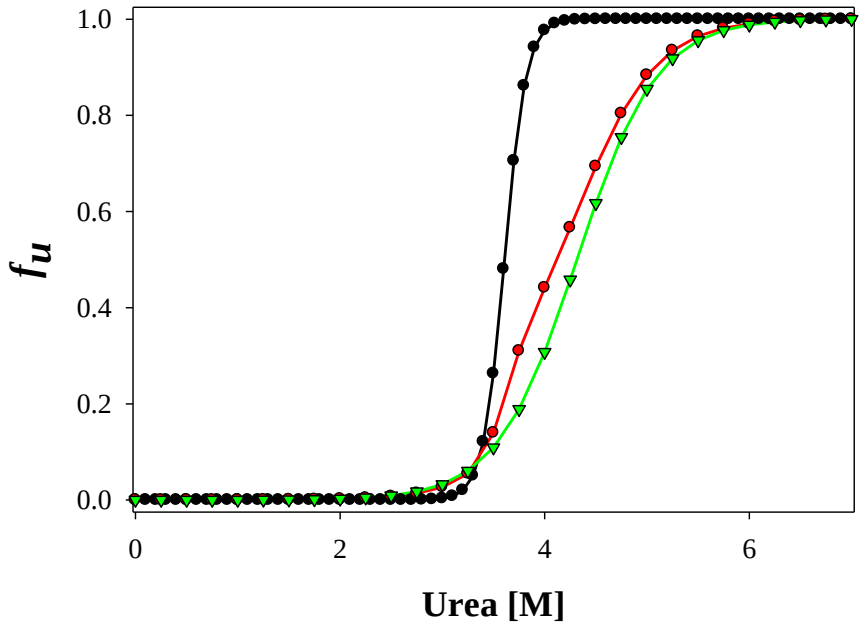

Figure 3.12: Normalized fit of chemical denature of YbiV compared to calculated 3-state model. BLACK: SAXS detection technique, GREEN: CD detection technique, and RED: calculated denature plot with a 0.2 molar fraction intermediate.

3.3.2 Effect of $Mg2^{+}$ Cofactor on Stability

Conserved among HAD proteins is the Mg^{2+} cofactor, it is well known that it is necessary for catalysis. How does it affect the structure strength of the subclass II HAD protein? Cofactor was chelated according to previously mentioned methods, and tested similar to native protein for structure stability.

Thermal and chemical stability results Table 3.2 we can see that Mg^{2+} plays an import roll in strength of the structure. From the collected data, results show two distinct denature paths, Figures 3.2-3.4, 3.13-3.16 display the two chemical denature data sets. From denature of native YbiV we know that it unfolds according to a 3-state model although the intermediate is not present in the sample without Mg^{2+} does not necessarily mean it does not exist. Stability is decrease in the absence of cofactor suggesting a role in structure stability. Evan though thermal denature alone cannot determine if an intermediate is sampled during the unfolding process, it is a good marker for stability changes among same protein. Thermal denature of YbiV with and without Mg^{2+} display similar trends in stability (T_m) as ΔG between the two proteins Figure 3.17-3.21.

Protein	**T_m (ºC)**	**ΔG (cal* mol^{-1})**	**m (cal*mol^{-1}*M^{-1})**	**d_{eq} (M)**
WT with Mg^{2+}	55.4 ± 0.1	-5685.4	-1327.8	4.28
WT without Mg^{2+}	50.2 ± 0.1	-5033.2	-1535.3	3.28

Table 3.2: Comparison of stability variables of WT E. *coli* YbiV with and without cofactor, Mg^{2+}. Protein buffers were made to have similar charge stabilization effects from added ions without the addition of Mg^{2+}. Protein buffer for native included 20 mM HEPES, 2.5 mM $MgCl^{2+}$, pH 7.20, whereas protein buffer for ΔMg^{2+} included 20 mM HEPES, 5 mM NaCl, pH 7.20.

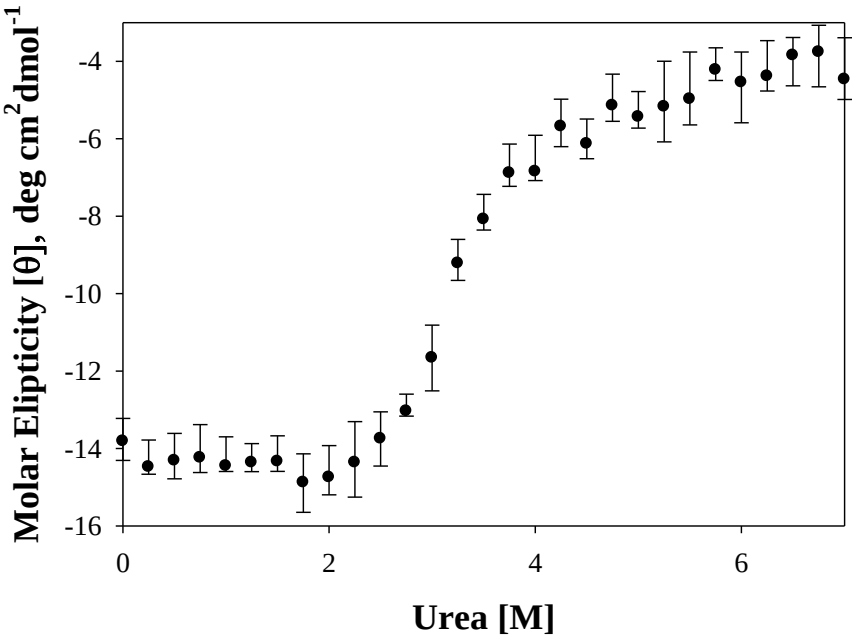

Figure 3.13: Four chemical denature sets were collected for YbiV ΔMg^{2+} using CD. Standard deviation from average is shown for CD at 222 nm. Buffer: 20 mM HEPES, 5.0 mM NaCl pH 7.20 with increasing urea concentration. Protein concentration was ~0.1 mg/mL.

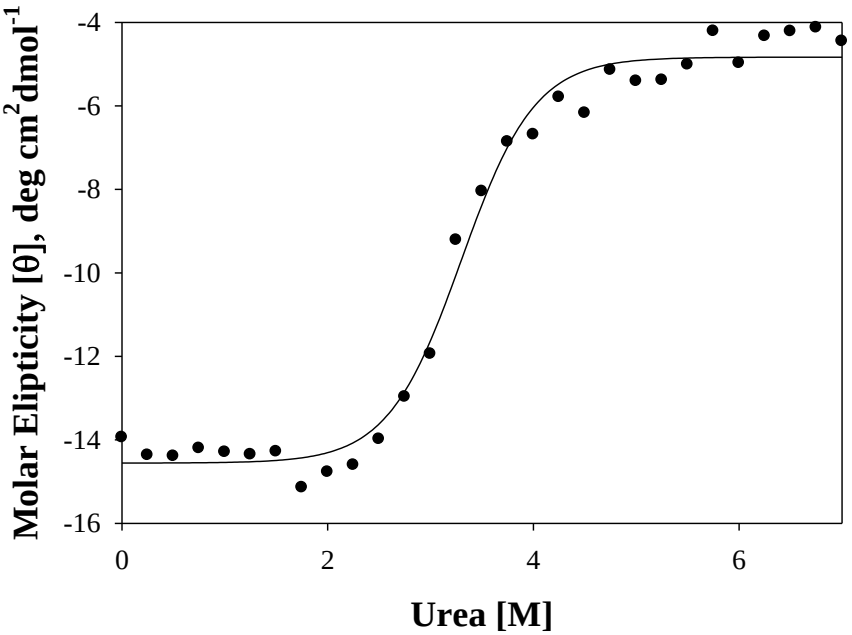

Figure 3.14: Fit line with averaged CD data for chemical denature of YbiV ΔMg^{2+}. Fit is a 4-parameter sigmoidal fit with variables: **a**= 9.7226, **b**= 0.3574, $\mathbf{x_0}$= 3.2997, and $\mathbf{y_o}$= -14.5561.

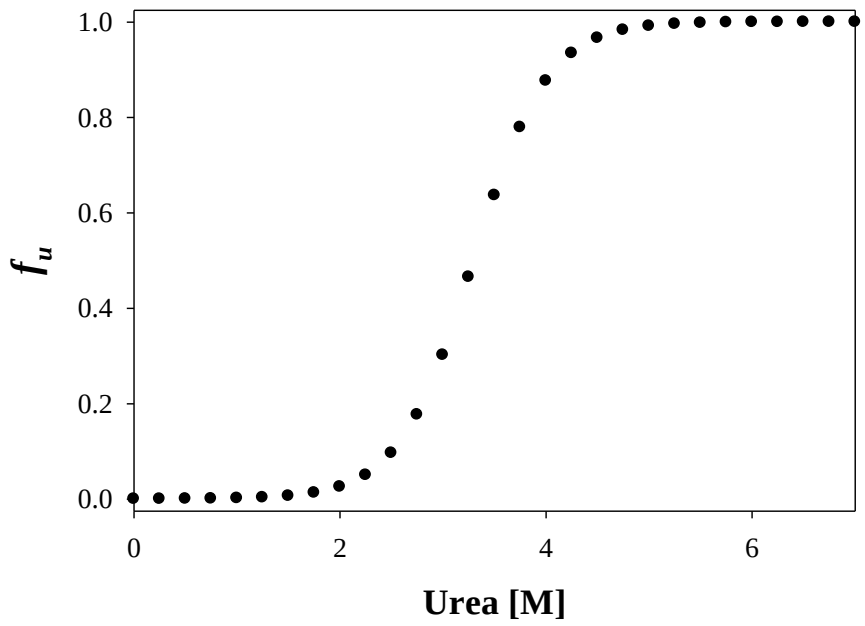

Figure 3.15: Normalized fit according to a 4-parameter fit of chemical denature of YbiV ΔMg^{2+} using circular Dichroism at 222 nm.

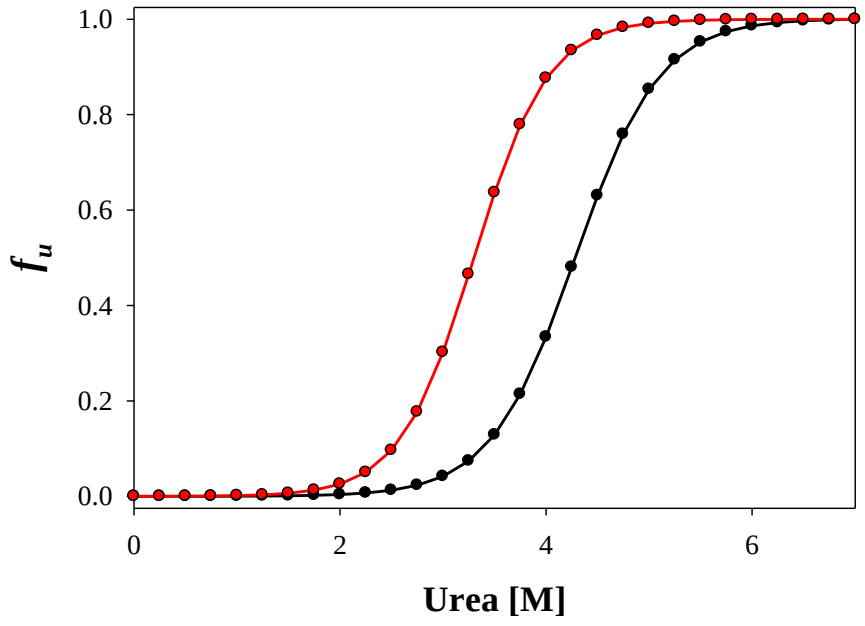

Figure 3.16: Chemical denature plot (CD- 222nm) of YbiV +/ ΔMg^{2+} (BLACK, RED respectively). Chemical denature clearly displays a destabilization in the absence Mg^{2+}.

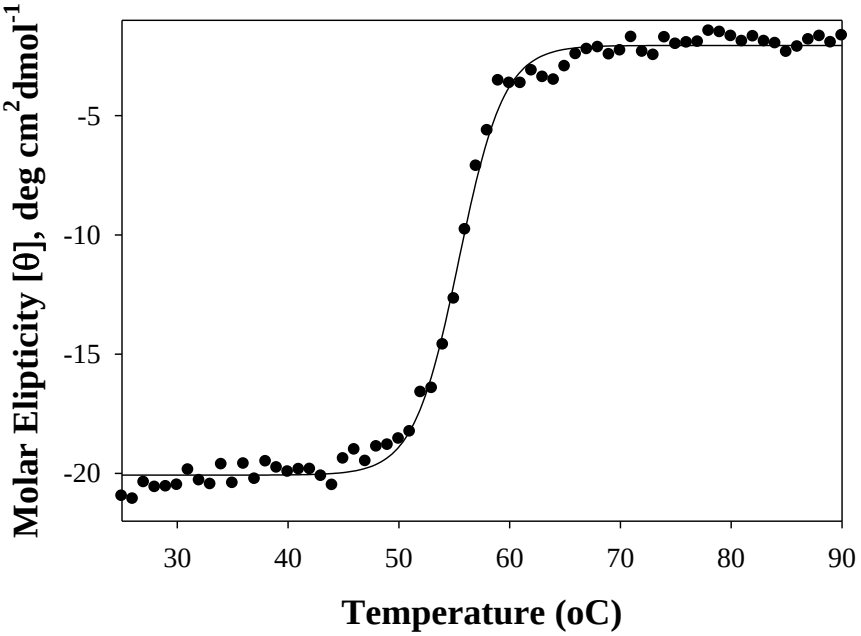

Figure 3.17: Fit line with thermal denature data of YbiV. Fit is a 4-parameter sigmoidal fit with variables: **a**= 18.0148, **b**= 2.0370, $\mathbf{x_0}$= 55.4105, and $\mathbf{y_o}$=- 20.07231.

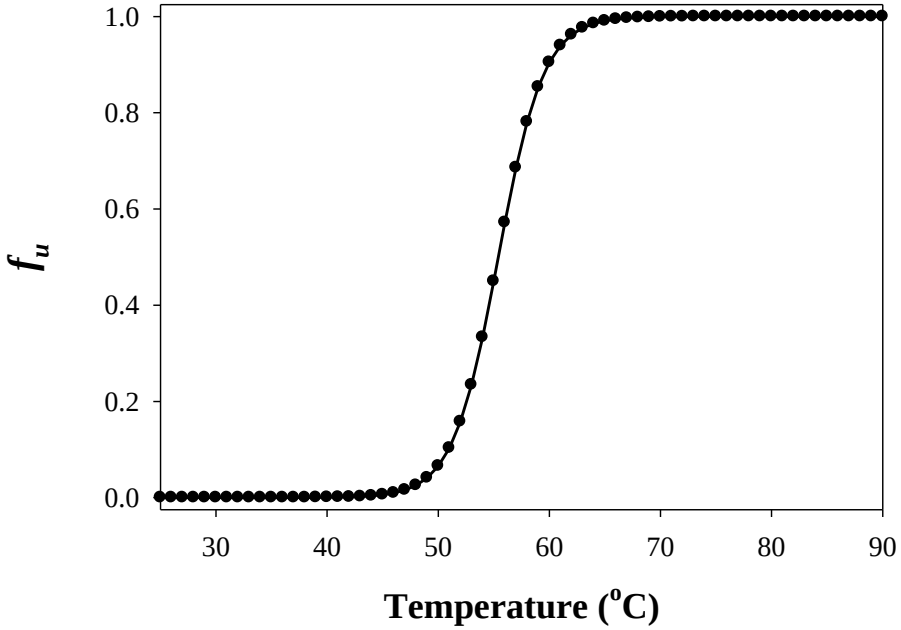

Figure 3.18: Normalized fit according to a 4-parameter fit of thermal denature of YbiV using circular Dichroism at 222 nm.

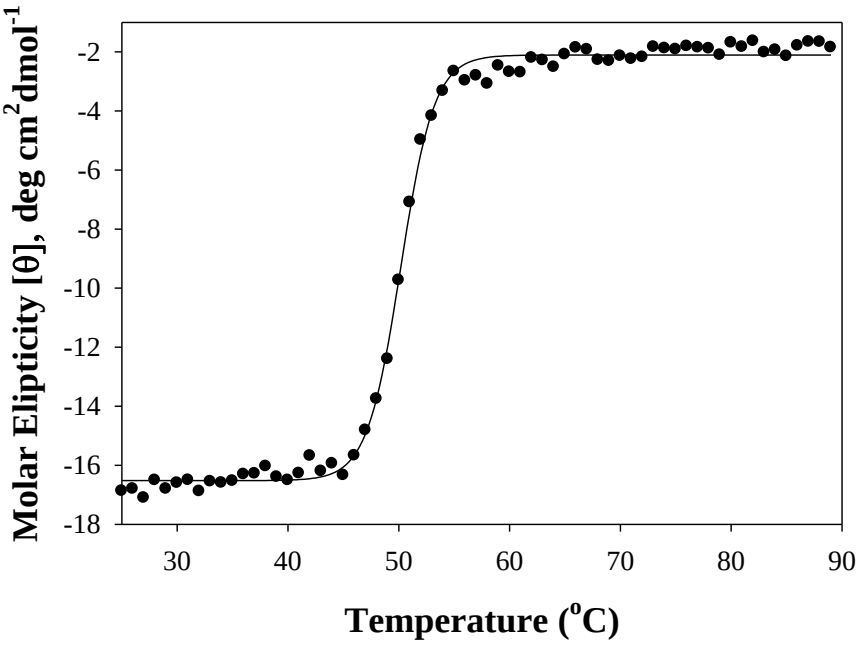

Figure 3.19: Fit line with thermal denature data of YbiV ΔMg^{2+}. Fit is a 4-parameter sigmoidal fit with variables: **a**= 14.4131, **b**= 1.5176, $\mathbf{x_0}$= 50.1690, and $\mathbf{y_o}$= -16.5185.

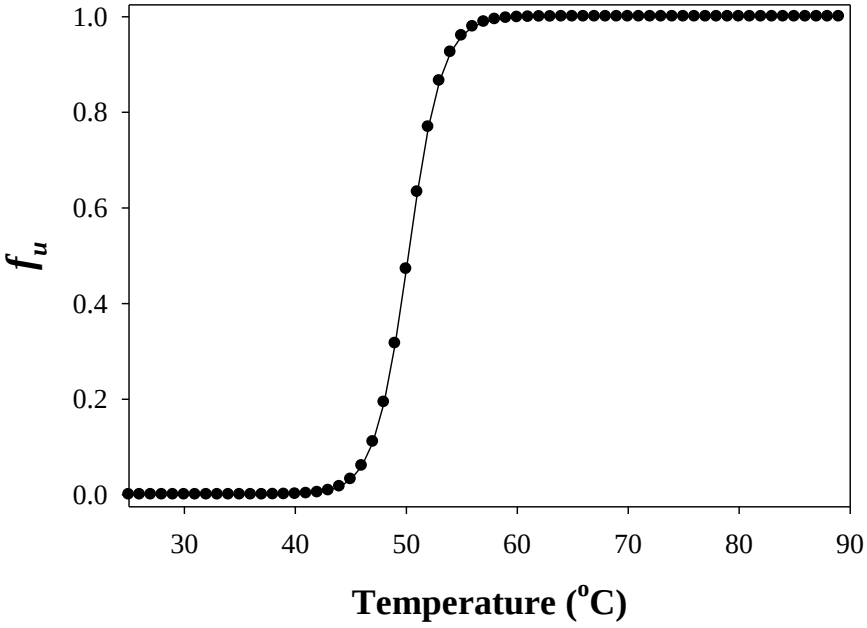

Figure 3.20: Normalized fit according to a 4-parameter fit of thermal denature of YbiV ΔMg^{2+} using circular Dichroism at 222 nm.

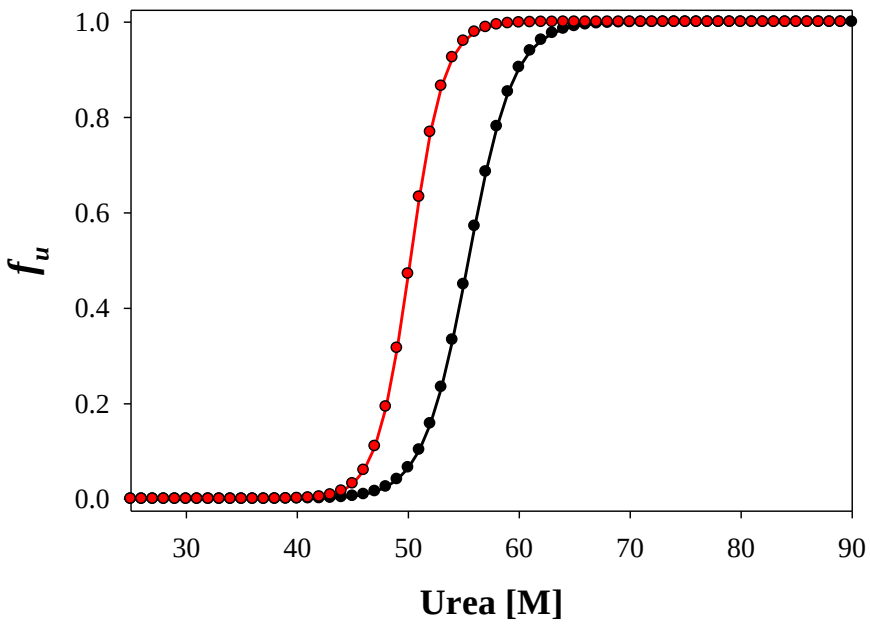

Figure 3.21: Thermal denature plot (CD- 222nm) of YbiV +/ ΔMg^{2+} (BLACK, RED respectively). Thermal denature clearly displays a destabilization in the absence Mg^{2+}.

3.3.3 Motif I Point Mutations Show Distinct Changes in Structural Stability

M. *loti* GmhB results on point mutations of motif I conserved residues, DxD squiggle, displayed unique changes in stability. Asp36 when mutated to alanine exhibited no change in stability whereas Asp 38 revealed a dramatic decrease in stability. These changes were displayed in the GmhB protein. Can these changes be extended to other HADs or are they specific to M. *loti* GmhB? To probe the generality of changes in stability of DxD point mutants; the Asp 9 and Asp 11 to alanine mutants were created. Point mutants were tested similar other HADs tested.

Similarly to M. *loti* GmhB the point mutations of YbiV's active site DxD motif stability trends changed due to each Asp. Data presented in Table 3.3 display values that

correlate to previous experiment D9A is as stable or more and D11A shows a destabilization in the fold of YbiV. As seen previously, the values are consistent, Asp^2 (second Asp in DxD motif) is decreased while Asp^1 tends to stabilize YbiV and play little role in M. *loti* GmhB stability.

Native YbiV as well as M. *loti* GmhB unfold according to a 3-state model, but in each case the elucidation of 3-state unfolding is greatly decreased when DxD mutants are created. YbiV chemical denatures have a distinct intermediate structure seen in all three detection techniques whereas in chemical denature of D9A and D11A, Figures 3.22-3.27, D11A losses transition and D9A retains 3-state transition. The D9A mutant from chemical and thermal denature proves to be a modestly more stable than the native YbiV protein and displays a slightly longer intermediate lifespan (based in chemical denature plots). On the other end of the spectra the D11A mutation decreased the stability of the protein as seen in both denature techniques. From the chemical denature using fluorescence as a detection technique the protein did not unfold according to a 3-state model (native and D9A had distinct intermediates), to verify this loss of structure the CD denature was obtained and showed similar but not overlaying spectra. The addition of the two spectra also suggests a 3-state model, but the longevity of the intermediate is far less than the native and D9A counterparts.

In M. *loti.* GmhB the double DxD mutant was created to test the dual-effect on the protein stability as well as probe the possible canceling/amplifying effects when compared to the single point mutations. Because YbiV has the included cap domain when compared to GmhB; the double mutant D9A/D11A was created to investigate modification to a sub-class II HAD protein.

The double mutant was cloned and express in high yields, but when purified the protein was unstable in native purification conditions. Protein began precipitating after elution from Ni-NTA 6xHIS column fractions were centrifuged and transferred for dialysis although continuation of precipitation occurred. Because stability is the target of our experiments, the ability to stabilize the protein through additive would jeopardize our results. From observations of purification we can assume the protein was destabilized by the double mutation. When interpreted with point mutation data we can conclude that the double mutation has a negative effect on the structure and stability of the protein and the mutations create an additive effect of the destabilizing effects of the D11A point mutation.

Protein	**T_m (ºC)**	**Gibbs Free Energy (cal mol^{-1})**	**M (cal mol^{-1} M^{-1})**	**D_{eq} (M)**
WT	55.41 ± 0.10	-5685.4	-1327.8	4.28
D9A	58.65 ± 0.09	-7265.1	-1502.4	4.84
D11A	48.64 ± 0.11	-4029.3*,	-1340.5*,	3.01*
		-5693.7	-1766.8	3.22
D9A/D11A	NA	NA	NA	NA

Table 3.3: Comparison of stability variables of native E. coli ybiV and DXD squiggle site-directed mutations. Protein buffer in all experiments was 10 mM HEPES, 2.5 mM $MgCl^{2+}$ pH 7.20 with increasing urea concentration. protein concentration was ~10 μM in all experiments. * denotes chemical denature detected by fluorescence, whereas all other chemical denature variable are detected using CD.

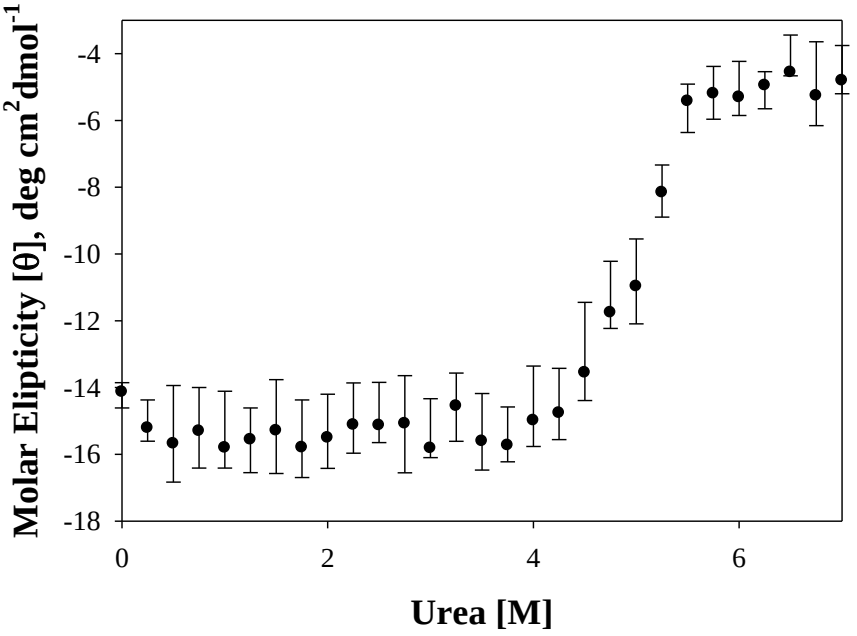

Figure 3.22: Four chemical denature sets were collected for D9A YbiV using CD. Standard deviation from average is shown for CD detection at 222 nm. Buffer: 20 mM HEPES, 2.5 mM $MgCl^{2+}$ pH 7.20 with increasing urea concentration. Protein concentration was 0.1 mg/mL.

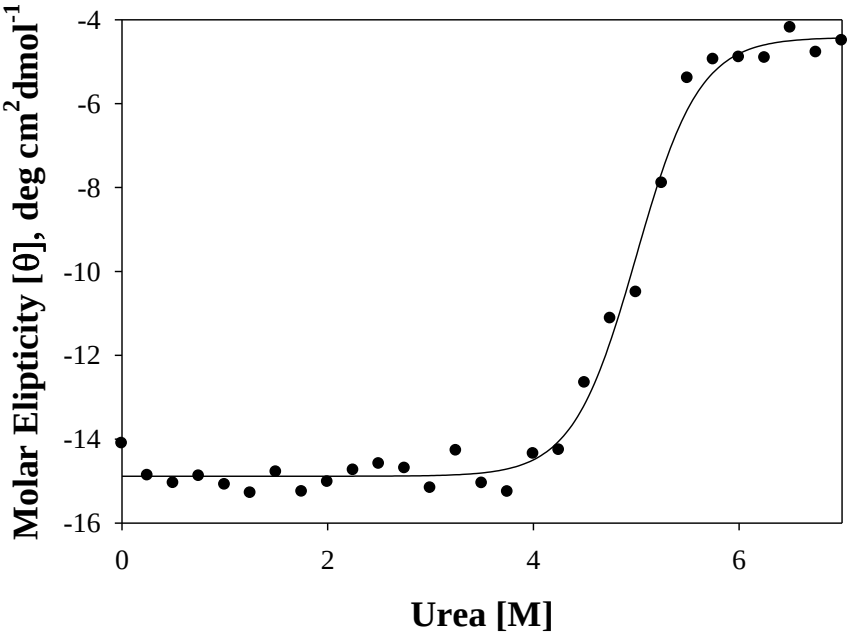

Figure 3.23: Fit line with averaged CD data for chemical denature of D9A YbiV. Fit is a 4-parameter sigmoidal fit with variables: **a**= 10.4623, **b**= 0.3073, $\mathbf{x_0}$= 4.9981, and $\mathbf{y_o}$= -14.8837.

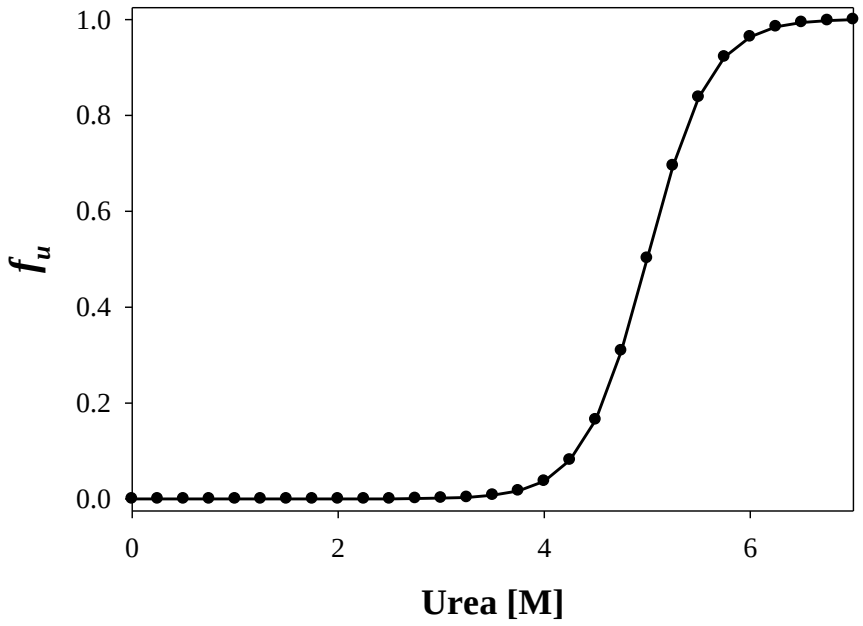

Figure 3.24: Normalized fit of chemical denature data using CD for D9A YbiV.

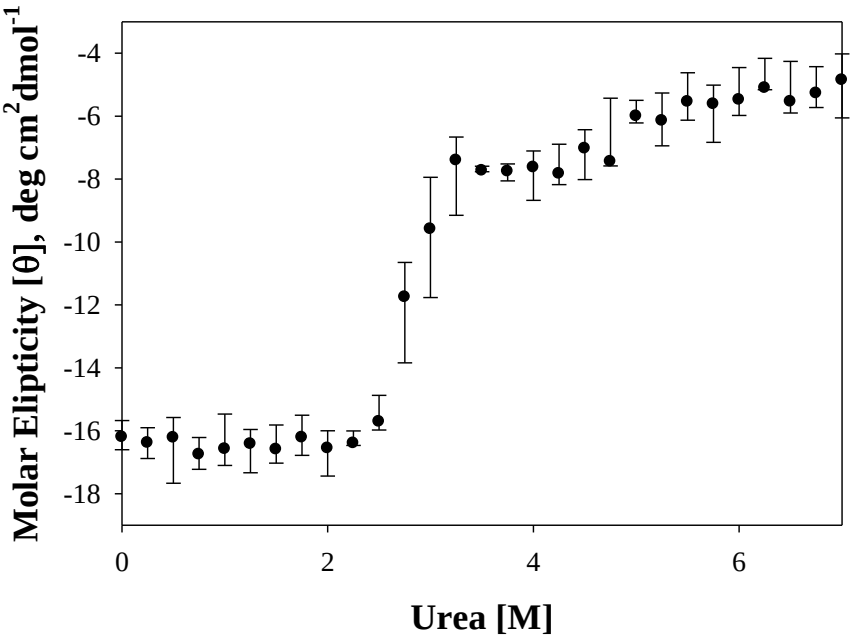

Figure 3.25: Four chemical denature sets were collected for D11A YbiV using CD. Standard deviation from average is shown for CD detection at 222 nm. Buffer: 20 mM HEPES, 2.5 mM $MgCl^{2+}$ pH 7.20 with increasing urea concentration. Protein concentration was 0.1 mg/mL.

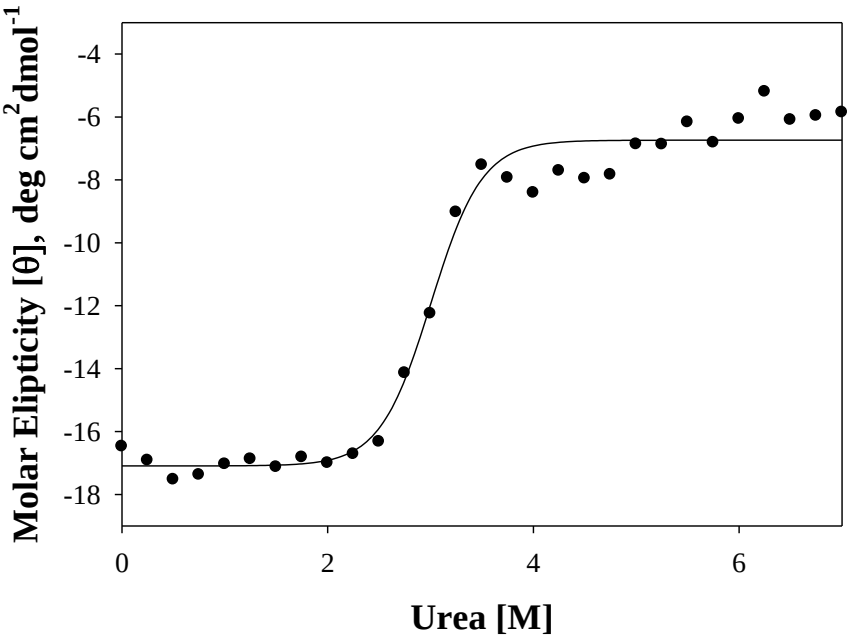

Figure 3.26: Fit line with averaged CD data for chemical denature of D11A YbiV. Fit is a 4-parameter sigmoidal fit with variables: **a**= 10.3538, **b**= 0.2482, $\mathbf{x_0}$= 3.0045, and $\mathbf{y_o}$= -17.0919.

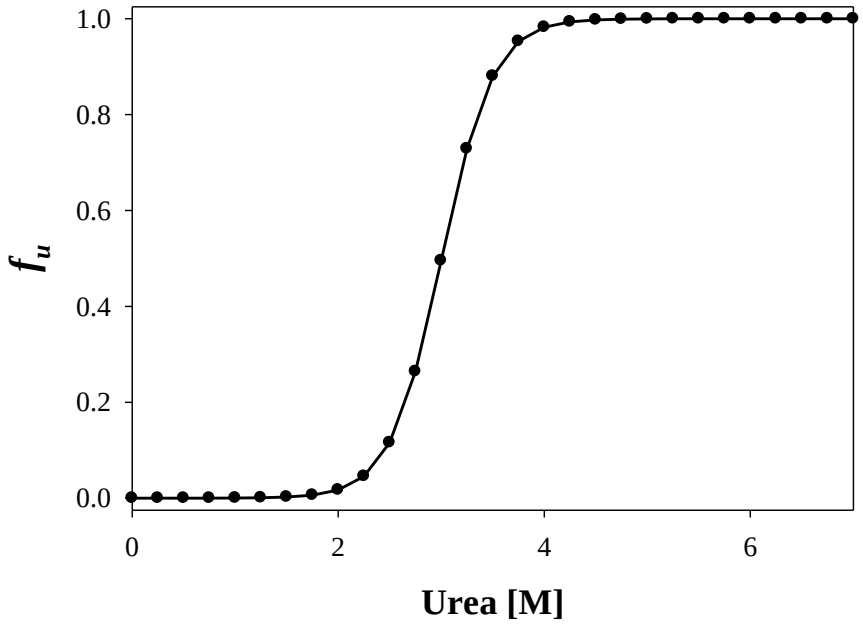

Figure 3.27: Normalized fit of chemical denature data using CD for D11A YbiV.

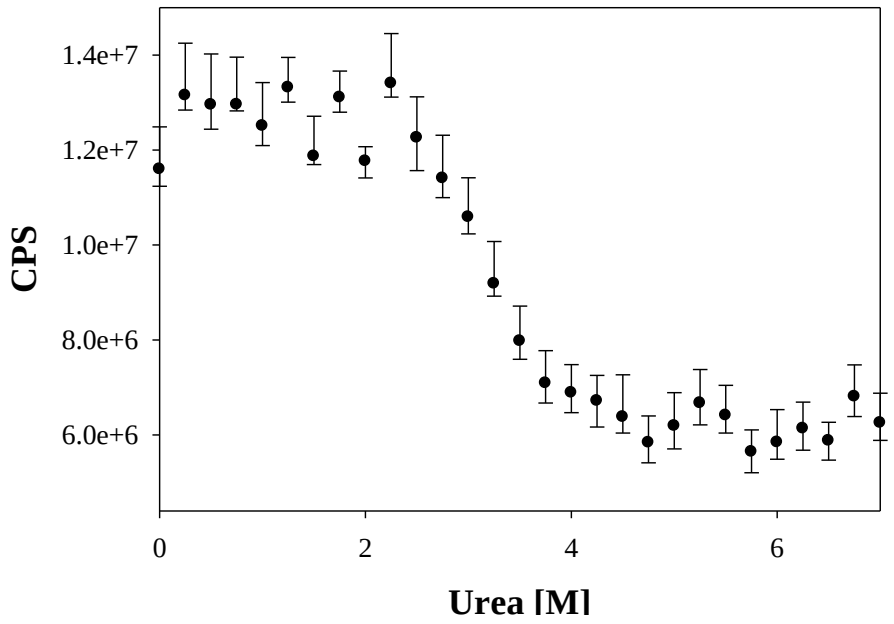

Figure 3.28: Four chemical denature sets were collected for D11A YbiV using fluorescence. Standard deviation from average is shown for tryptophan florescence detection at excitation: 280 and 285nm, and emission: 340 and 352 nm. Buffer: 20 mM HEPES, 2.5 mM $MgCl^{2+}$ pH 7.20 with increasing urea concentration. protein concentration was 0.1 mg/mL.

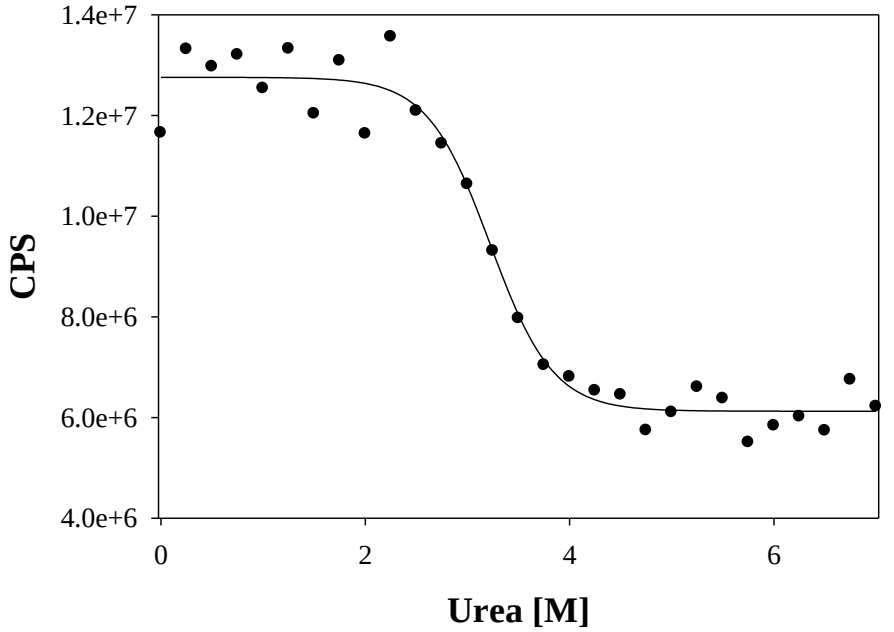

Figure 3.28: Fit line with averaged tryptophan fluorescence data of D11A YbiV. Fit is a 4-parameter sigmoidal fit with variables: **a**= 6632028, **b**= -0.3071, $\mathbf{x_0}$**=** 3.2227, and $\mathbf{y_o}$**=** 6125087

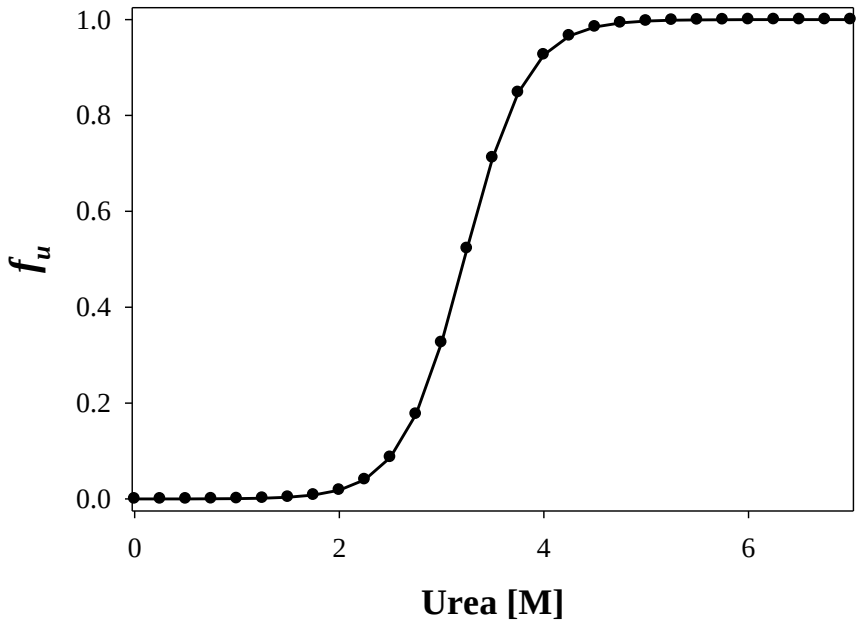

Figure 3.29: Normalized fit of chemical denature data using tryptophan fluorescence of D11A YbiV

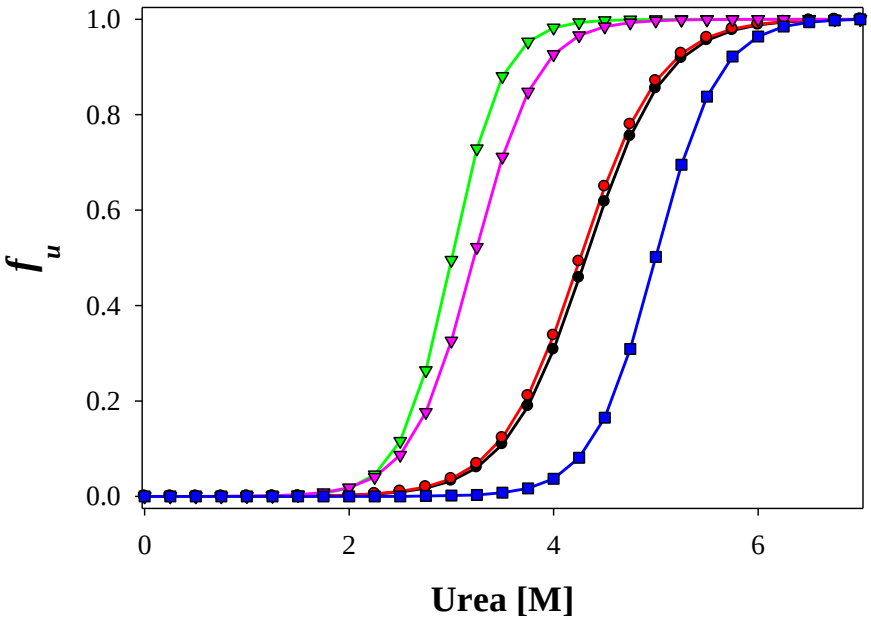

Figure 3.30: Graph includes chemical denature of native YbiV and DxD point mutations: D9A and D11A. GREEN: D11A (CD), MAGENTA: D11A (fluorescence), RED: YbiV (fluorescence), BLACK: YbiV (CD), and BLUE: D9A (CD). From the graph we can see clearly that D11A is less stable and D9A is more stable than native YbiV.

3.3.4 Asp11 Interacting Pin Residues: Role Played In Protein Structure Stability

Previous experiments on Asp11 show a destabilization in structure stability when mutated to alanine. Does the decrease in stability result from the loss of residue when mutated or does bonding networks play a role in destabilization? To test the underlying reason for decreases the interacting pin residues were tested for stability when mutated to alanine. The two residues interacting with Asp11 are Asn46 and Lys19, Figure 3.6. From the published YbiV crystal structure (PDB: 1RLM) we see interaction between Asp11-Asn46 and Asp11-Lys19, more specifically interaction between carboxylic functional group of Asp11 and the amide of Asn46, and Asp11 carboxylic and Lys19 amine group.

Asparagine 46 was mutated to alanine to investigate effects of protein behavior. The N46A mutant decreased activity, but did not affect structure stability. From thermal and chemical denature changes in biophysical variables were not affected (same T_m and +0.9 kcal*mol^{-1} ΔG).

Lysine 19 point mutation was attempted and successfully cloned although the protein was ultimately found in the pellet after lyses using French press technique. Because the K19A protein was tested due to stability issues; the double mutation K19A/N46A protein was attempted. Similarly to the point mutation K19A the double mutant was unstable. The protein purified using native Ni-NTA procedures, but once eluted protein quickly precipitated in collection tube.

Interaction between Asp11 and pin residues Asn46 and Lys19 propose the idea that without the intramolecular interactions between Asp11 and pin residue Lys19 the protein significantly losses stability. This connection identifies key residues in HAD

stability. To further investigate the connection between the residues, Asp11 side change mutations were created to evaluate the specific interactions that play a key role in protein stability.

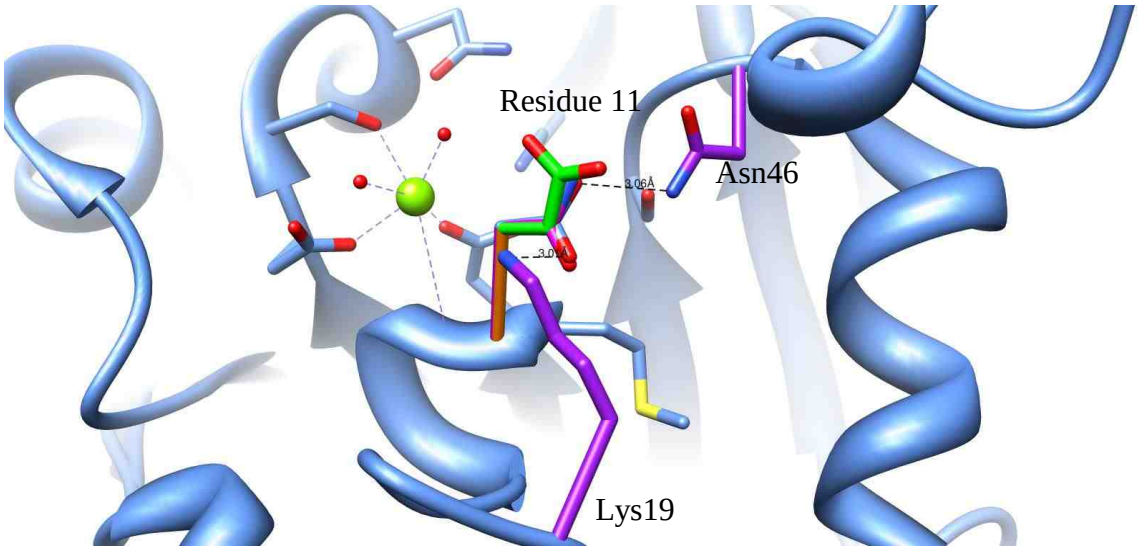

Table 3.31: Crystal structure **(**PDB: 1RLM) displaying point mutations at Asp11 (ORANGE: D11A, BLUE: Asp11, MAGENTA: D11N, and GREEN: D11E). PURPLE: pin residues Lys189 and Asn46.

3.3.5 Side Chain Mutations of Asp11 Demonstrate Key Interactions Necessary for Stability

As seen previously core stability is stringently connected to Asp11 and its environment. To pin down specific interactions created by Asp11, point mutations were created to monitor changes in protein stability. Among the discovery of stability factors the point mutation created give an insight into possible effects created by creating non-active proteins for ligand binding in X ray crystallography. It is a common practice in X ray crystallography to elucidate ligand interaction in proteins to do that the ligand must bind in native-like fashion either by substrate analogs or deactivation of protein. It is

common in HAD proteins to deactivate the protein by mutating Asp11 to non-functional side chains such as alanine, asparagine, and some cases glutamic acid.

A set of Asp11 point mutations were created to evaluate the change in stability based on side chain interactions, the mutations included D11A, D11E, and D11N. To evaluate the effect of side chain interactions thermal denature was executed and changes in T_m observed, Table 3.5

Protein	T_m (ºC)
Native	55.41 ± 0.10
Alanine	48.64 ± 0.11
Glutamic Acid	55.77 ± 0.16
Asparagine	48.35 ± 0.12

Table 3.4: Thermal T_m values for native YbiV and Asp11 point mutants. Experiments were carried out in 10 mM HEPES, 2.5 mM $MgCl^{2+}$ pH 7.2.

Interestingly, the changes in stability as observed by T_m proposed similar results as the pin mutation experiments. From Table 3.5 and Figure 3.7 we can see that the carboxylic functional group is absolutely necessary for stability of the protein. In both cases where the carboxylic acid functional group is replaced the stability is decrease (~ 7 ºC) in a similar manner compared to native T_m. The third point mutation, D11E, was observed to be as stable as the native protein and is the only one of the point mutations that retains the carboxylic acid side chains although it is one carbon length longer than Asp.

From variation in side chain, it is apparent the interaction of the carboxylic acid side chain of aspartate and glutamate are required regardless of carbon length. Although

most interesting is the effect asparagine has on the stability of the protein. Asparagine is the same length as aspartate but includes a amide functional group. From previous results we have seen the interaction between Asn46 and Asp11 does not affect stability whereas Lys19-Asp11 does, so why does the D11N affect stability? An explanation for this distinguishment could be accredited to distinct interactions between residues. The Asp11 could be locked into a single rotomer which creates distinguishment between the two results.

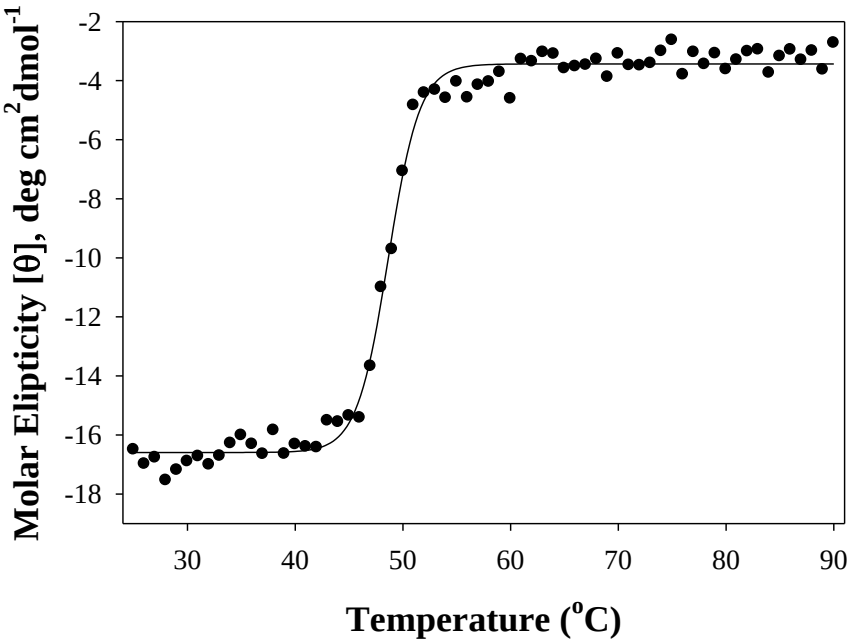

Table 3.31: Fit line with thermal denature data of D11A YbiV. Fit is a 4-parameter sigmoidal fit with variables: **a**= 13.5179, **b**= 1.4598, $\mathbf{x_0}$= 48.6474, and $\mathbf{y_o}$= -16.5922.

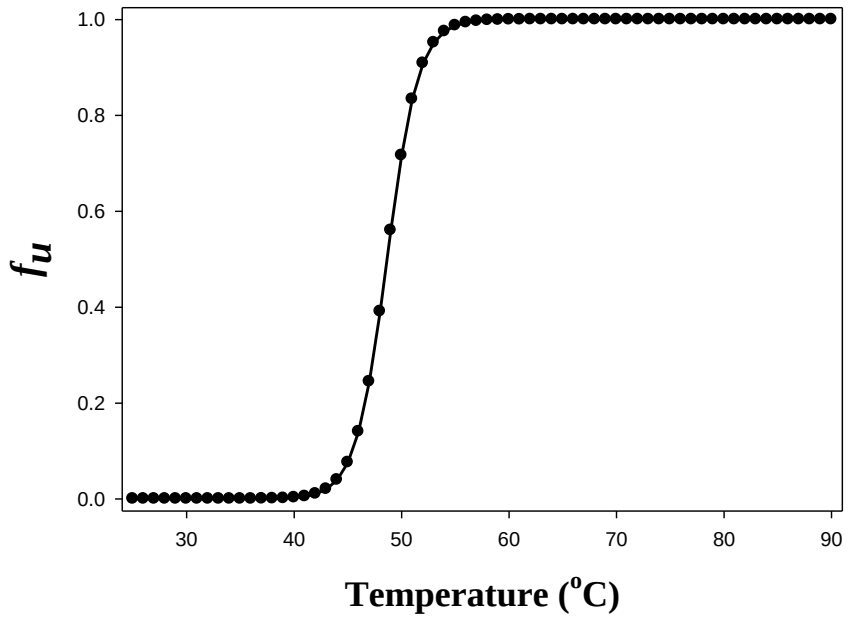

Table 3.31: Normalized fit for thermal denature of D11A YbiV.

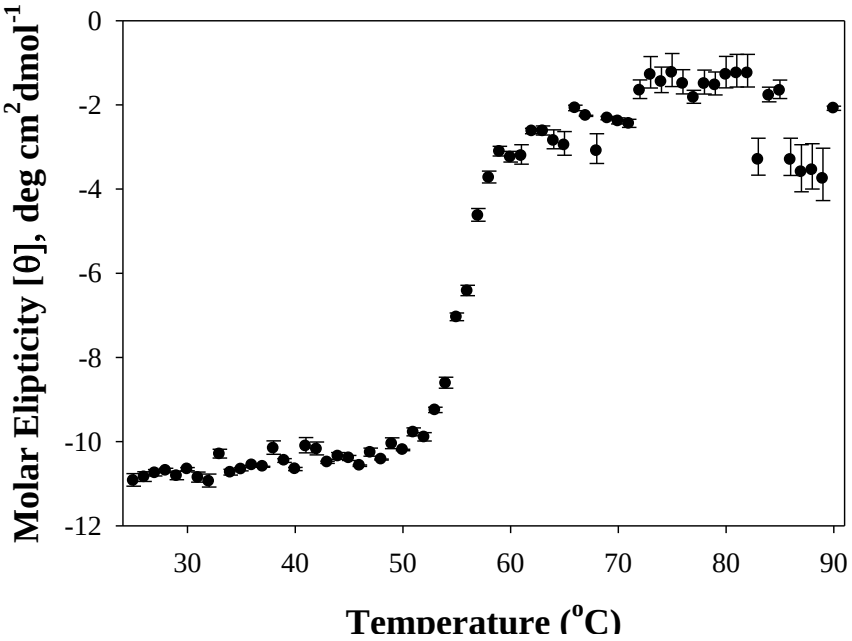

Table 3.31: Standard deviation from average is shown for thermal denature of D11E YbiV. Buffer: 20 mM HEPES, 2.5 mM $MgCl^{2+}$ pH 7.20. Protein concentration was ~.01mg/mL.

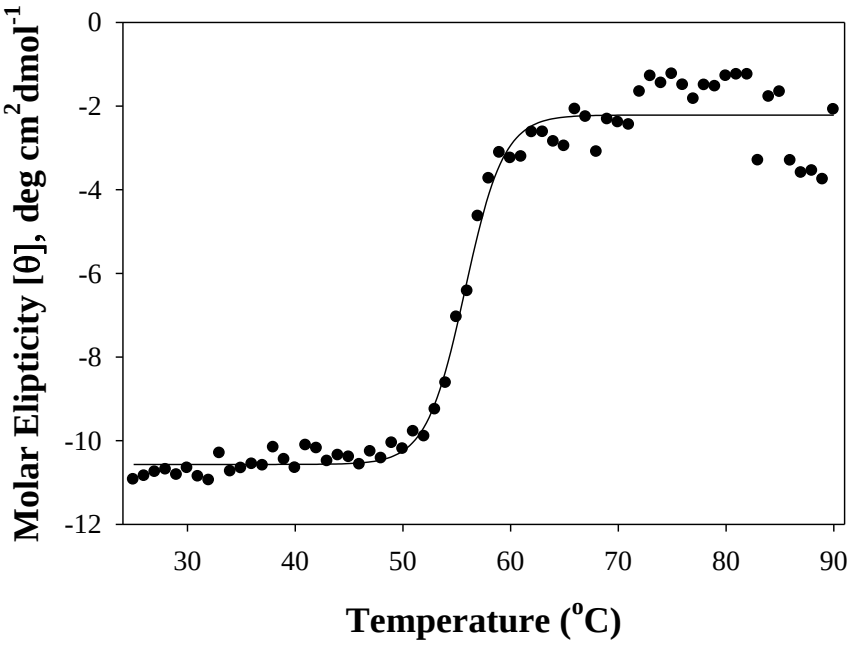

Table 3.31: Fit line with thermal denature data of D11E YbiV. Fit is a 4-parameter sigmoidal fit with variables: **a**= 8.3524, **b**= 1.7955, $\mathbf{x_0}$= 55.7723, and $\mathbf{y_o}$= -10.5674.

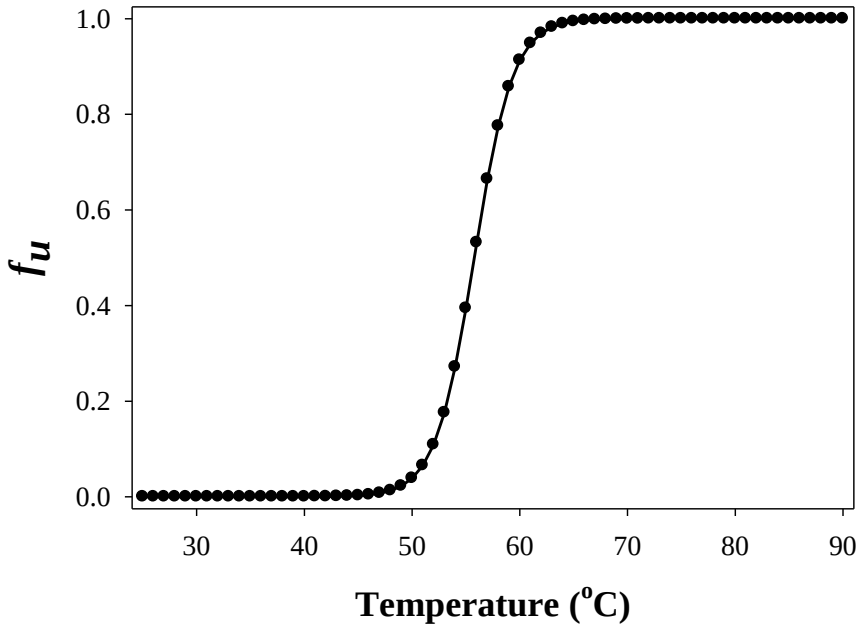

Table 3.31: Normalized fit for thermal denature of D11E YbiV.

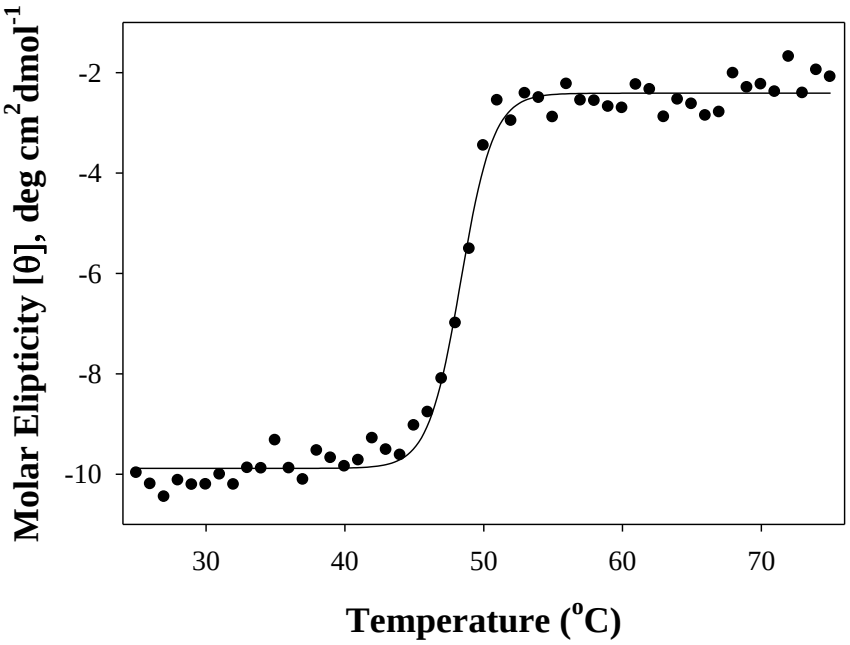

Table 3.31: Fit line with thermal denature data of D11N YbiV. Fit is a 4-parameter sigmoidal fit with variables: **a**= 7.4732, **b**= 1.1727, $\mathbf{x_0}$= 48.3544, and $\mathbf{y_o}$= -9.8842.

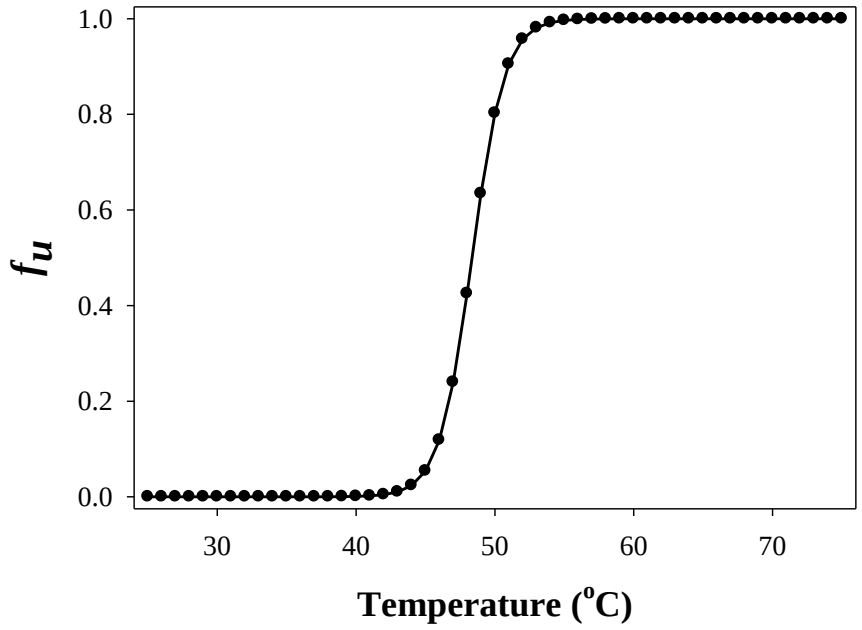

Table 3.31: Normalized fit for thermal denature of D11N YbiV.

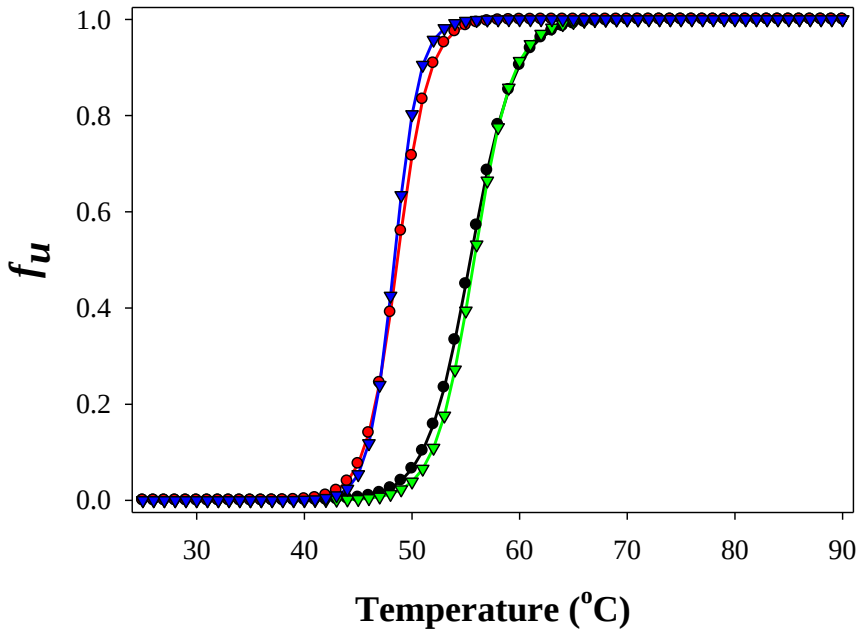

Table 3.31: Thermal overlay of native, D11A, D11E, and D11N YbiV (BLACK, RED, GREEN, and BLUE respectively). Data clearly shows destabilization of D11A and D11N whereas D11E is as stable as native YbiV.

3.4 Conclusions

E. *coli* YbiV is a sub-type II HAD protein. It is identified by a large cap insert between motif II and III. This elaborate cap domain has been associated with expanded substrate binding in HAD protein. In comparison M. *loti* GmhB the protein is overall more stable. This can be observed by both chemical and thermal stability values.

Albeit E. *coli* YbiV is more stable than its "capless" counterpart stability trends including changes in the active site are still present. Similar to GmhB, YbiV shows a positive effect on stability with the Asp^1 mutation and a negative effect on stability when residues interacting with or include Asp^2 mutations. Combined results of experiments on these proteins advocate trends among HAD proteins and the intricate arrangement

between structural stability, activity, and robustness of the protein superfamily. The underlying examination of the HAD superfamily entails the reason for the prominence and success of the fold family among organism. Because the structural stability is linked to the activity and success of the protein through key family trends the next logical step is to investigate the relationship between increased activity and fold.

References

1. Roberts, A., S.-Y. Lee, et al. (2005). "Ybiv from Escherichia coli K12 is a HAD phosphatase." Proteins: Structure, Function, and Bioinformatics **58**(4): 790-801.

2. Kandror, O., A. DeLeon, et al. (2002). "Trehalose synthesis is induced upon exposure of Escherichia coli to cold and is essential for viability at low temperatures." Proceedings of the National Academy of Sciences **99**(15): 9727-9732

3. Lu, Z., D. Dunaway-Mariano, et al. (2005). "HAD Superfamily Phosphotransferase Substrate Diversification: Structure and Function Analysis of HAD Subclass IIB Sugar Phosphatase BT4131†,‡." Biochemistry **44**(24): 8684-8696

4. Peisach, E., J. D. Selengut, et al. (2004). "X-ray Crystal Structure of the Hypothetical Phosphotyrosine Phosphatase MDP-1 of the Haloacid Dehalogenase Superfamily†,‡." Biochemistry **43**(40): 12770-12779.

5. Karen N. Allen, Debra Dunaway-Mariano, Phosphoryl group transfer: evolution of a catalytic scaffold, Trends in Biochemical Sciences, Volume 29, Issue 9, September 2004, Pages 495-503

6. Harder, M. E., Deinzer, M. L., Leid, M. E., & Schimerlik, M. I. (2004). Global analysis of three-state protein unfolding data. *Protein science*, *13*(8), 2207-2222.

CHAPTER FOUR

CRITERIA FOR EVOLUTION OF A SUCCESSFUL ENZYME SUPERFAMILY: FOLD FITNESS AND DOMAIN DYNAMICS EXPLORED.

4.1 Introduction

Bt4131 from Bacteriodes *thetaiotaomicron* is a HAD protein. Although the native physiological role of Bt4131 is not known, it has been tentatively assigned as a phosphatase in chitin metabolism (2). Gene context of Bt4131 suggests a role in recycling nitrogen and carbon from the cell wall polymer chitin. Neighboring enzymes consist of Bt4132 and Bt4127 which are (chitobiase and glucosamine-6-phosphate isomerase respectively) (1,2). Chitobiase is an enzyme that cleaves oligomers of N-acetyl-glucosimine (NAG) into monomers, whereas Bt4127 recycles "NH_3" though conversion of glucosamine-6-phopshate to fructose-6-phosphate. The gene context and broad-range sugar phosphatase activity suggests Bt4131 might function as a fructose-6-phosphate phosphatase in chitin catabolism (1).

Bt4131 is a HAD subfamily II class B protein. The subfamily II class B is designated by the topology of the cap insert. The cap domain is a slightly smaller α/β domain; more specifically $\alpha\beta\beta(\alpha\beta\alpha)\alpha\beta\beta$ type, Figure 4.1. The cap domain in HAD proteins acts to interact with alcohol group of substrate for tighter binding during catalysis. This can be observed due to accumulation of hydrophobic residues on the cap interacting with the sugar ring of the substrate(s) (1). The many interactions between

substrate and protein are created with binding of the phosphate group to the core domain and hydrophobic interactions between the cap domain and metabolite, but binding is not enough for catalysis; the protein undergoes a conformation change, wherein the active site is sequestered from bulk solvent. Closure of the cap domain to desolvate the active site is a major barrier in the Bt4131 reaction, Figure 4.2. Previous studies suggest an induced fit mechanism for catalysis; this suggests that the protein must undergo a conformation change to progress through mechanism (4). In our cases we will look at the conformation changes due to binding ligands as well as the affect domain interacting residues have on transitions between open and closed conformation.

For completeness Bt4131 was tested for stability similar to other HAD proteins mentioned previously. In this chapter we look into the possible compromise between activity and stability through investigation into point mutations in the active site and throughout the domain - domain interface. Point mutations were created in the active site to evaluate stability trends seen in previous proteins. Domain - domain interacting point mutations were created to analyze effect on activity through modification of conformation changes observed with small angle X ray scattering (SAXS). Small angle X-ray scattering is a technique that employs high energy X ray energy to evaluate solution conformations of proteins. As previously mentioned the subfamily II class B HAD proteins include a cap domain that closes onto the active site initiate catalysis, therefore we used SAXS to investigate the change in conformation due to modification of protein or binding of substrate analogs.

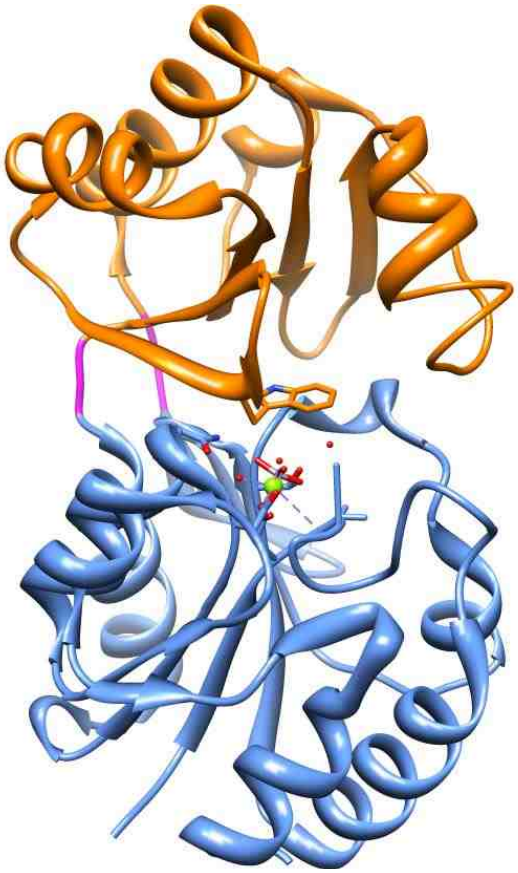

Figure 4.1: X ray crystal structure representing Bt4131 (PDB: 2RAR). Structure represents the subfamily II class B HAD protein: ORANGE: represents cap domain in Bt4131, BLUE: represents the core domain, and MAGENTA: represents linkers between the two domains.

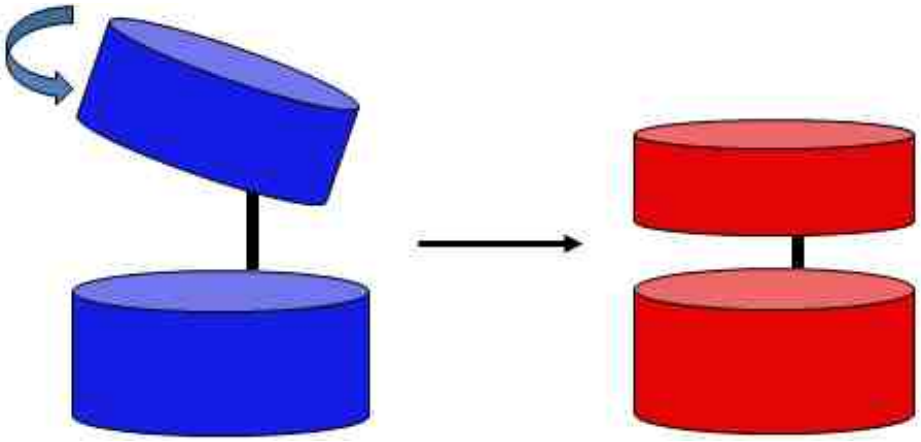

Figure 4.2: Cartoon graphic describes the process of closing the active site for catalysis. BLUE represents open structure and RED suggests closed structure.

4.2 Material and Methods

4.2.1 Materials

Restriction enzymes and T4 ligase were purchased from Invitrogen. Pfu Turbo polymerase was purchased from Stratagene. Oligonucleotides were custom synthesized by Invitrogen. Components and buffer additives were purchased from Sigma Aldrich if not otherwise specified. DNA sequencing was performed at MC labs and by using MW mass spectrometric determinations carried out by UNM Mass Spec Facilities.

4.2.2 Cloning and Purification of Bt4131

Wild-type Bt4131 was cloned previously (1). Cell stock containing the Bt4131 plasmid in pLyseS competent cells was used for purification. Bt4131/pET3A/pLysesS cells were grown aerobically at 37 ºC in TB broth containing 50 µg/mL ampicillin, and 4 mL glycerol to an $O.D_{600}$ 0.7 - 0.9 and induced for production of protein by addition of 0.1 mM isopropyl-β-D-galactopyranoside (IPTG). Following a 12 h induction period at 18 ºC, cells were harvested by centrifugation at 6,500 rpm for 10 min, resuspended in 10x weight Buffer A (50 mM HEPES pH 7.5), and lysed using French press at 1,200 psi. Lysate was centrifuged at 20,000 rpm and 4 ºC for 20 min., giving a supernatant that was fractioned using ammonium sulfate induced precipitation. The 50-80% (w/v) ammonium sulfate fraction was collected by using centrifugation at 20,000 rpm, 4 ºC for 20 min. The collected pellet was resuspended in 30 mL Buffer B (50 mM HEPES, 12% $(NH_4)_2SO_4$ pH 7.5, and loaded onto a 10 x 30 mm (HR10 column) BS Sepharose column, pre equilibrated with Buffer B at 4 ºC. The column was washed with 5 CV Buffer B and eluted with a linear gradient (10 CV) 100 - 0% Buffer B (12% -0% $(NH_4)_2SO_4$). Fractions

were analyzed by using SDS-PAGE, and those containing the desired protein were combined and concentrated using 10K Amicon Ultra Centrifugal filter.

4.2.3 Site-Directed Mutagenesis

Site-directed mutagenesis (SDM) was carried out using the Quick Change PCR Strategy (Stratagene). Plasmid was used as template with custom oligionucleotides expressing mutated condon (s). P*f*u Turbo polymerase was used to copy mutated PCR product. Recombinant mutants plasmids were transformed into pLyseS competent cells and used in 10 mL scale cultures for expression. Cultures showing expression of protein were subjected to either gene sequencing or mass spectrometric analysis. All mutants were cultured and purified using the procedure employed for the wild-type (WT) protein.

4.2.4 Determination of SteadyState Kinetic Constants

The k_{cat} and K_m values for native and mutant Bt4131 against glucose-6-phosphate were determined using the method described in Chapter 2.

4.2.5 Thermal Denature Monitored by Using Circular Dichroism

Circular dichroism was used to observe changes in the secondary structure of Bt4131 as a function of temperature. Procedures and fitting models used to analyze the data are described in Chapter 2.

4.2.6 Chemical Denature Monitored by Using Circular Dichroism, Fluorescence and SAXS

Three techniques were used to probe changes in the structure of Bt4131 promoted by addition of urea. Tryptophan fluorescence was used to monitor changes in the surroundings and local polarity of aromatic (mainly tryptophan) residues. To observe changes in the secondary structure of the protein, circular dichroism was used at wavelengths 222 and 217 nm. Finally, to monitor changes in the tertiary structure, small angle scattering was employed. SAXS data were used to generate a calculated solution volume, which is used to determine the general shape of the solution conformation of the protein. Procedures and fitting models for handling the chemical denaturation data are discussed in Chapter 2.

4.2.7 Preparation of SAXS Samples

Recombinant Bt4131 in stock solution was concentrated to 30 mg/mL. Individual stock solutions (200 mM) of sodium metavanadate (VO_3), 2-deoxy-d-glucose, and d-ribose were prepared in 0.22 μM filtered 18MΩ deionized water (typically stock ligands would be in dialysis buffer, but multiple samples were tested using stock solutions). Metavanadate (VO_3) was used to mimic phosphate mimic and both d-ribose and 2-deoxy-d-glucose were used to mimic substrate. Prior to sample collection, the solution containing the protein was centrifuged at 14,000 rpm 4ºC for fifteen min to ensure possible aggregates are pelleted. Protein and substrate solutions were diluted with dialysis

buffer to proper concentrations (protein range 0.5-30 mg/mL, substrate analogs 10x K_I) and placed in PCR strips for SAXS analysis.

4.2.8 Collection of SAXS Data

SAXS data were collected at beamline 4-2 at the Stanford Synchrotron Radiation Lightsource (SSRL). Samples subjected to this technique were prepared by mixing protein and matched buffer each ligand concentration. During collection, samples were held at 10 ºC using an aluminum block. Then, 30µL of a sample, drawn into a capillary using an automated fluid handling system, was subjected to ten-seventeen 1 s exposures giving data that were collected using a Pilatus 300K photon detector at a distance of 0. 5m covering a range of $0.013 < q^5 < 1.23$. Scattering profiles were processed using the BL4-2 SasTool utility program (5).

4.2.9 Theoretical Interpretation of Protein Conformation

To interpret data arising from SAXS experiments, the absolute scattering profiles of open and closed conformations of HAD proteins, with previously determined X ray crystal structures, were used in the to create theoretical scattering profiles Table: 4.1. Because the open conformation Bt4131 does not crystallize, Phyre[2] was used to thread the Bt4131 amino acid sequence through a known open conformation HAD protein ,TM0651, from Thermotoga *martima* (9). To create scattering profiles FoXS (7,8) and CRYSOL (6) were used

PDB ID	Ligand	Resolution (Å)	Conf. State
1YMQ	Mg, SO_4	1.90	D10A closed
2RBK	Mg, VO_3	1.00	Closed
1YMQ*	Mg, SO_4	1.90	Open

Table 4.1: Crystal structures used to investigate different conformations of Bt4131. 1YMQ* signifies open conformation of 1YMQ determined using Phyre[2] modeling of Bt4131 sequence with PDB: 1RLO.

4.3 Results and Discussion

4.3.1 Stability of Bt4131 and Effects of Active Site Mutation

Recombinant Bt4131 is a sub-type II HAD protein similar to E. *coli* YbiV. We have observed trends that follow through both capped CII and capless C0 HAD proteins. In this chapter we want to investigate the affect and mutation on catalysis observed through changes in solution conformation. To inspect the generality of stability trends amongst HAD proteins stability experiments were tested on Bt4131 Table 4.2 and Figures 4.3-4.9.

Akin to the previous two proteins tested, Bt4131 showed to unfold according to a 3-state model Figure 4.8. Although the transitions do not include a notable intermediate we can infer the extra state due to the separate transitions among different detection limits.

Technique	ΔG (cal*mol^{-1})	m (cal*mol^{-1}*M^{-1})	d_{eq} (M)
Fluorescence	-	-	
Circular Dichroism (CD)	-3646.7	-1466.4	2.49
Small angle X ray scattering (SAXS)	-4606.6	-2590.0	1.78

Table 4.2: Bt4131 chemical denature using urea with different observation techniques. All reactions were carried out in 20 mM HEPES, 2.5 $MgCl_2$. pH 7.2 with increasing urea concentration. Protein concentration was ~ 10 µM.

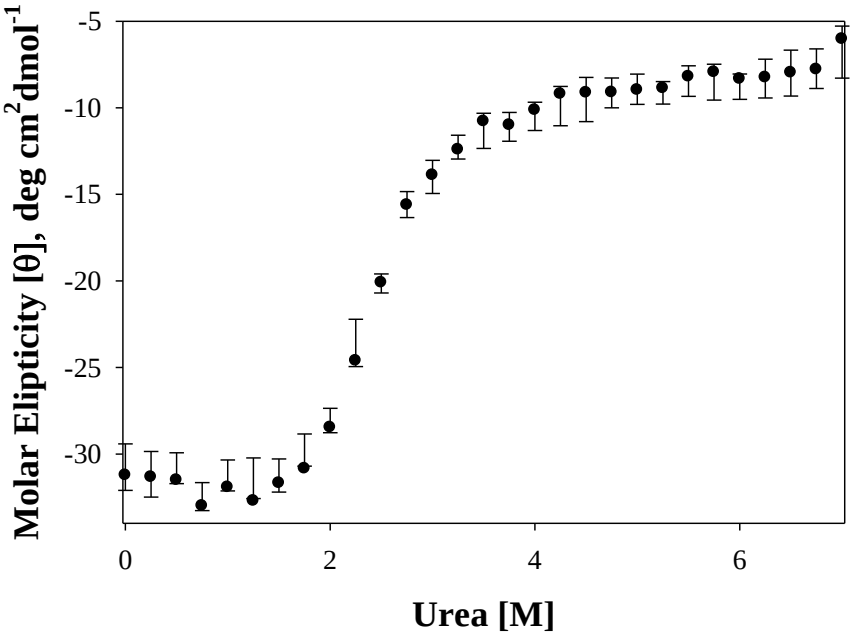

Figure 4.3: Raw chemical data from chemical denature of Bt4131 using CD. Error bars are standard deviation from average of 4 data sets. To ensure protein unfolds equally both 217 and 222 nm wavelength was collected and verified to overlay each other.

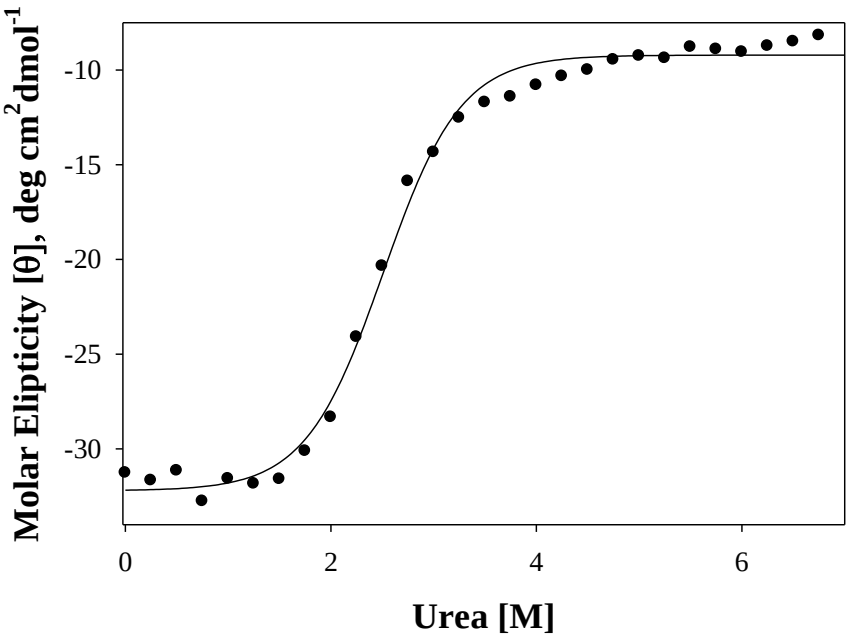

Figure 4.4: Fit line with averaged CD data for chemical denature of Bt4131. Fit is a 4-parameter sigmoidal fit with variables: a= 22.9957, b= .3773, x_0= 2.5088 and y_o= -32.2096

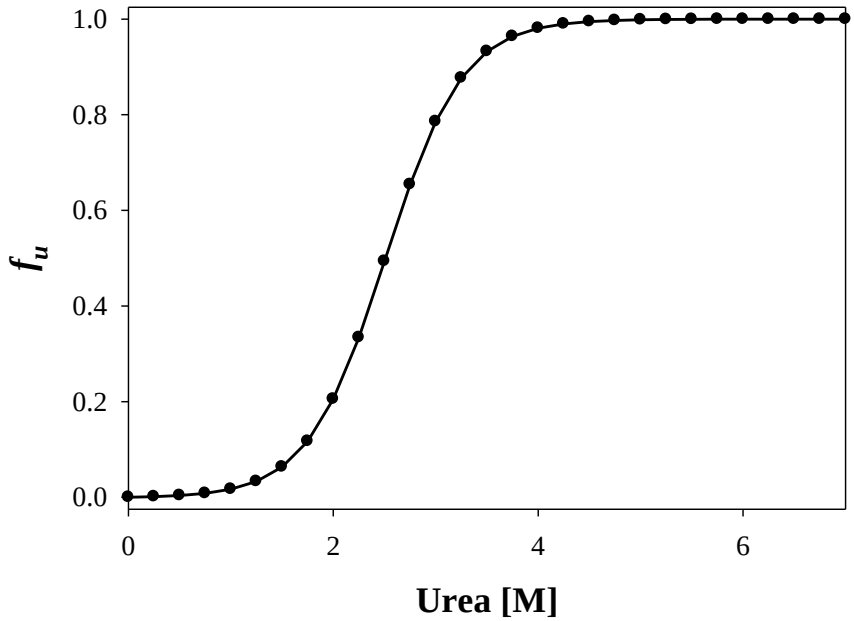

Figure 4.5: Normalized fit of chemical denature data using CD for Bt4131.

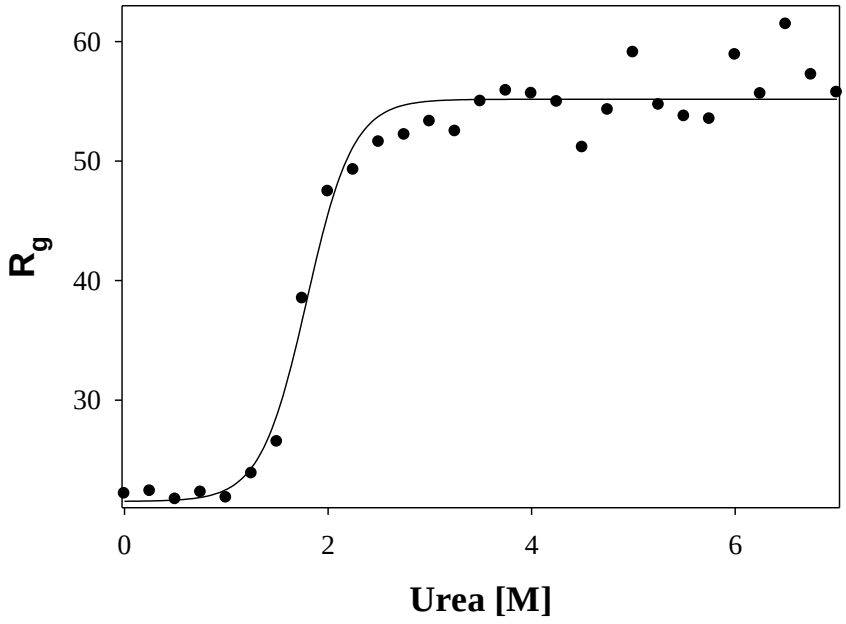

Figure 4.6: Fit line with small angle X ray scattering data for chemical denature of Bt4131. Fit is a 4-parameter sigmoidal fit with variables: **a**= 36.6441, **b**= 0.2265, $\mathbf{x_0}$= 1.7909, and $\mathbf{y_o}$= 21.5256. R_g was calculated for each [urea] using PRIMUS.

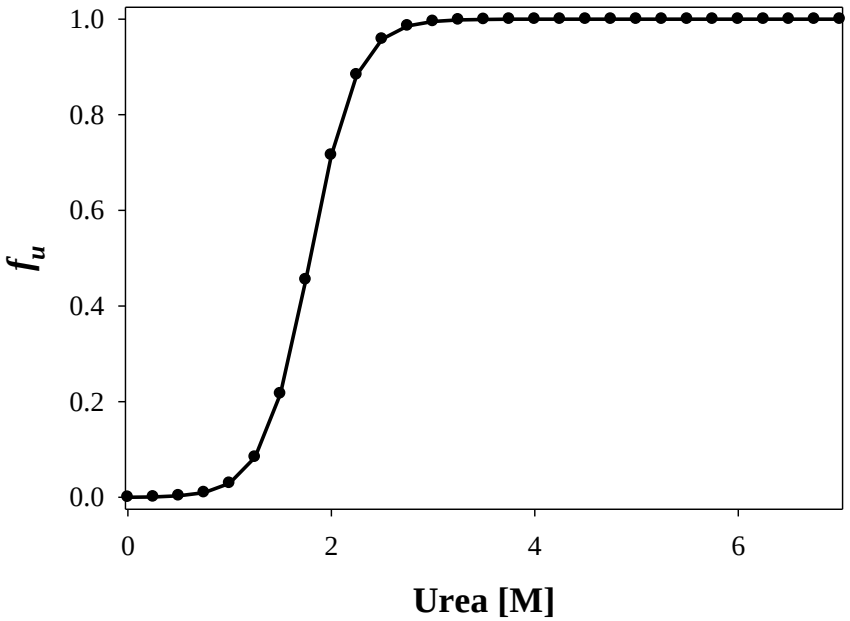

Figure 4.7: Normalized fit according to a 4-parameter fit of chemical denature of Bt4131 using SAXS.

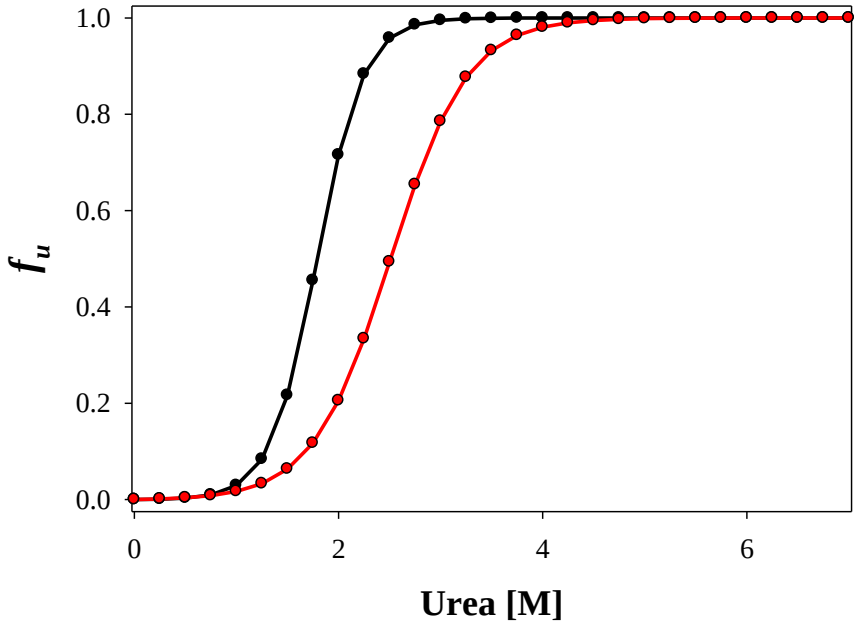

Figure 4.8: Overlay of chemical denature of Bt4131 with different observational techniques. Because denature plots do not overlay suggests the protein unfolds through a 3-state model.

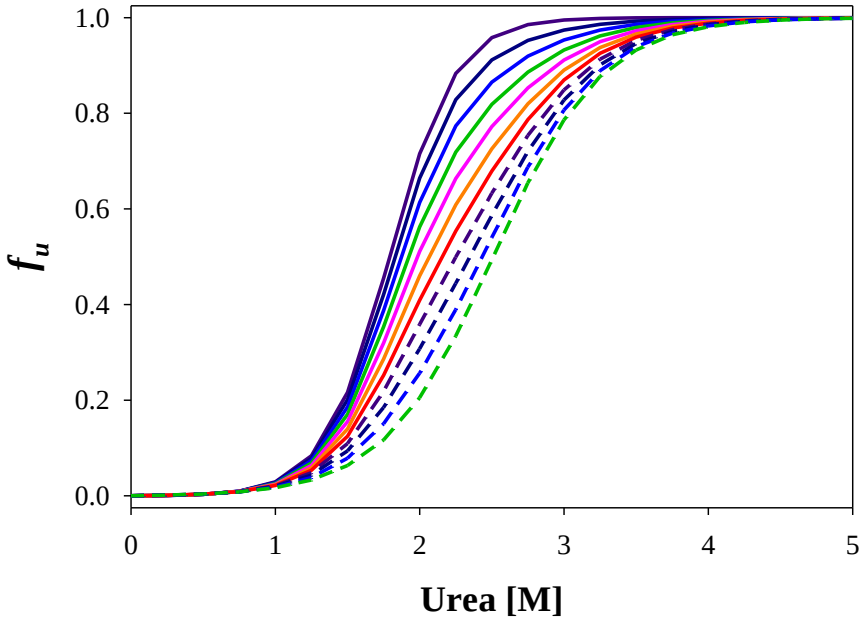

Figure 4.9: Calculation of combined theoretical 3-state denature curves at different molar ratio of intermediate conformation. From left to right, individual colors represent different molar fraction of intermediate from 0.0 to 1.0 by increments of 0.1. Solid DARK PURPLE: chemical denature using SAXS, GREEN SHORT DASH: chemical denature using CD.

4.3.2 Effect of Mg^{2+} Cofactor on Structural Stability

We have already recognized the importance of binding the Mg^{2+} cofactor in HAD proteins; results from E. *coli* YbiV display a decrease in stability and loss of activity when cofactor was not present. Bt4131 displays similar trend to YbiV. This decrease in stability can be attributed to changes in charge of the active site, i.e. a decrease in lysine in active and increase in aspartate.

Thermal denature of Bt4131 with and without cofactor resulted in a loss of stability when Mg^{2+} was not present 53.84 ± 0.09 and 50.05 ± 0.05 respectively, Figures 4.13-4.18. Common amongst all thermal denature experiments is the loss of 3-state unfolding and in this case we see no intermediate which suggests a 2-state unfolding model. Unlike the GmhB the loss of cofactor destabilized both capped C2 HAD proteins. The similarity between C2 proteins and difference from GmhB suggests a key role albeit small in stability of specific protein.

Previous explanation for the intermediate structure suggests a loss of intramolecular interactions between two possible sub-units of the core domain. The deficient Mg^{2+} experiments support this theory that the accumulation of the cap domain in the C2 proteins create a greater strain on the interaction between sub-units. Interactions between Mg^{2+} and Asp residues reside on the interacting surfaces of the two sub-units suggesting a role in increased affinity between sub-domains. A second theory could be attributed to the changes in charge of the active site Figure 4.10-4.12. In GmhB there are 2 lysines and 3 aspartates interacting near Mg^{2+} whereas in the capped C2 proteins an additional Asp is present and loss of Lys.

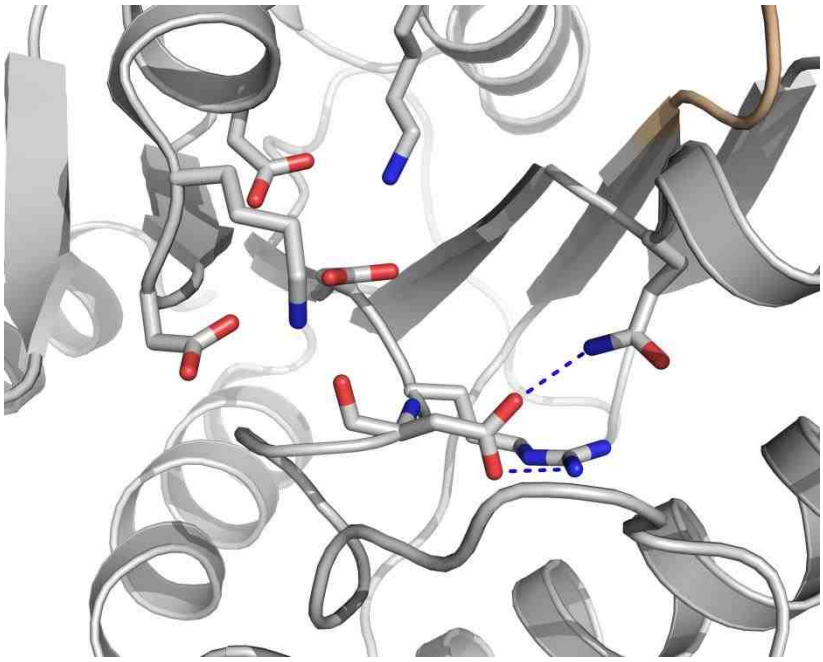

Figure 4.10: Displays the active site of M. *loti* GmhB. From the figure you can see three aspartates and 2 lysines near active site which suggest a neutralization of charge in the absence of Mg^{2+}.

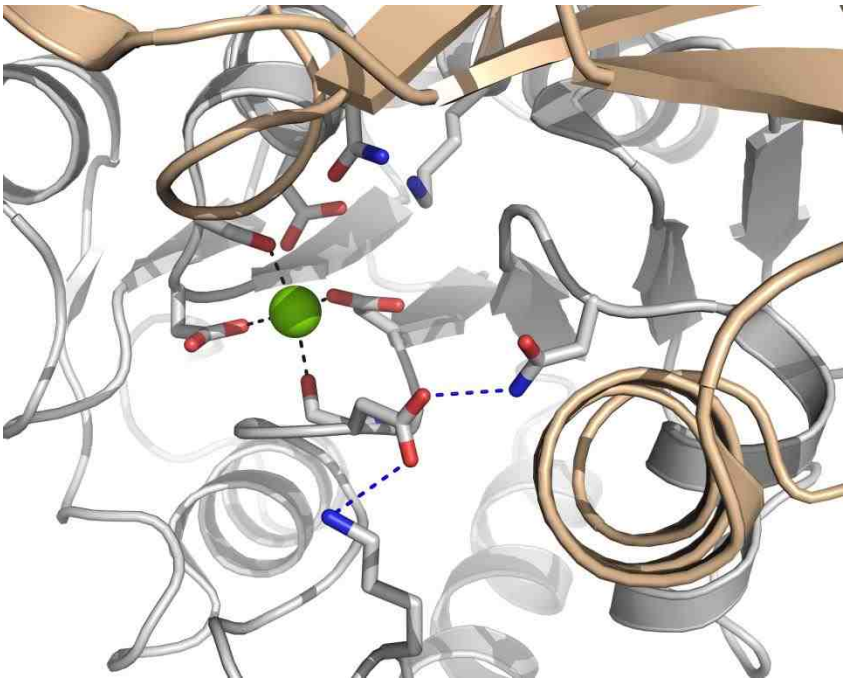

Figure 4.11: Displays the active site of E. *coli* YbiV. From the figure we can see that the loss of lysine and addition of aspartate could destabilize the charge equilibrium when Mg^{2+} is not present.

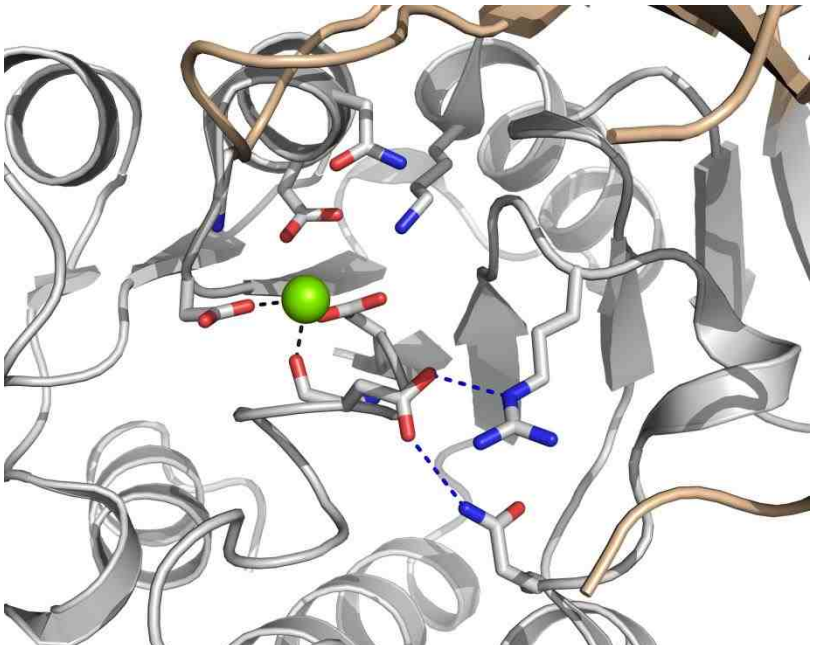

Figure 4.12: Displays the active site of Bt4131. From the figure we can see that the loss of lysine and addition of aspartate could destabilize the charge equilibrium when Mg^{2+} is not present.

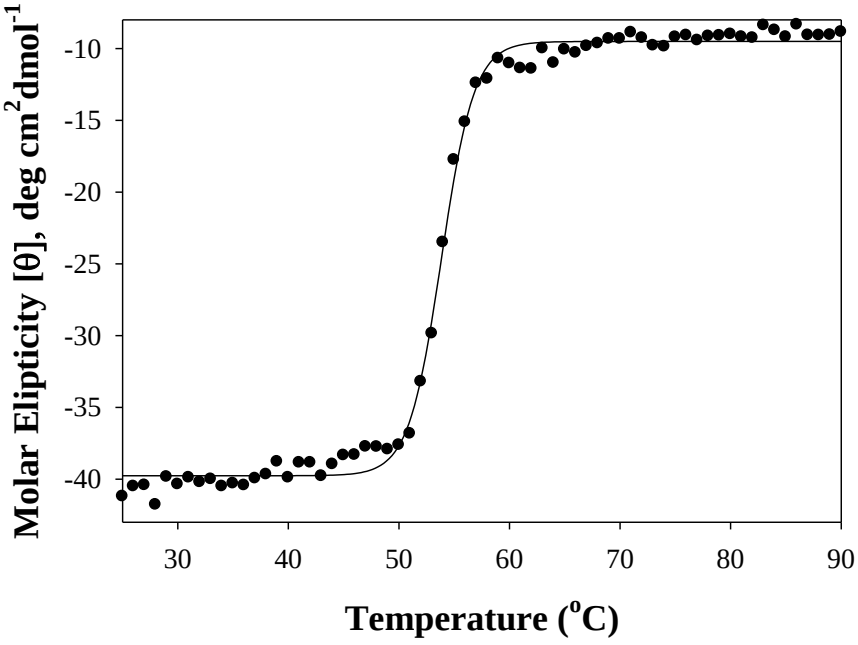

Figure 4.13: Fit line with CD data for thermal denature of Bt4131 +Mg^{2+}. Fit is a 4-parameter sigmoidal fit with variables: **a**= 30.2542, **b**= 1.4810, $\mathbf{x_0}$= 53.8440, and $\mathbf{y_o}$= -39.7638.

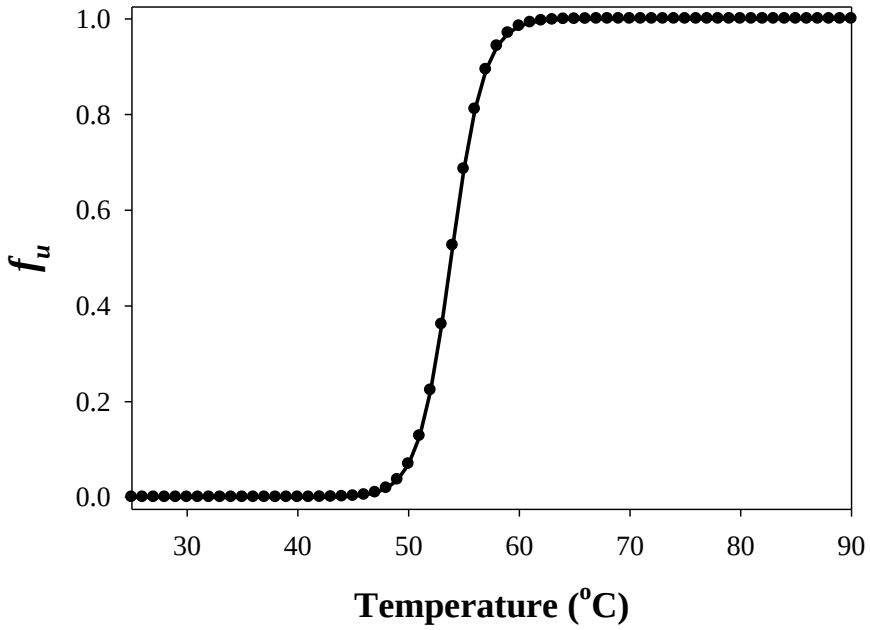

Figure 4.14: Normalized fit according to a 4-parameter fit of thermal denature of Bt4131 +Mg^{2+} using CD

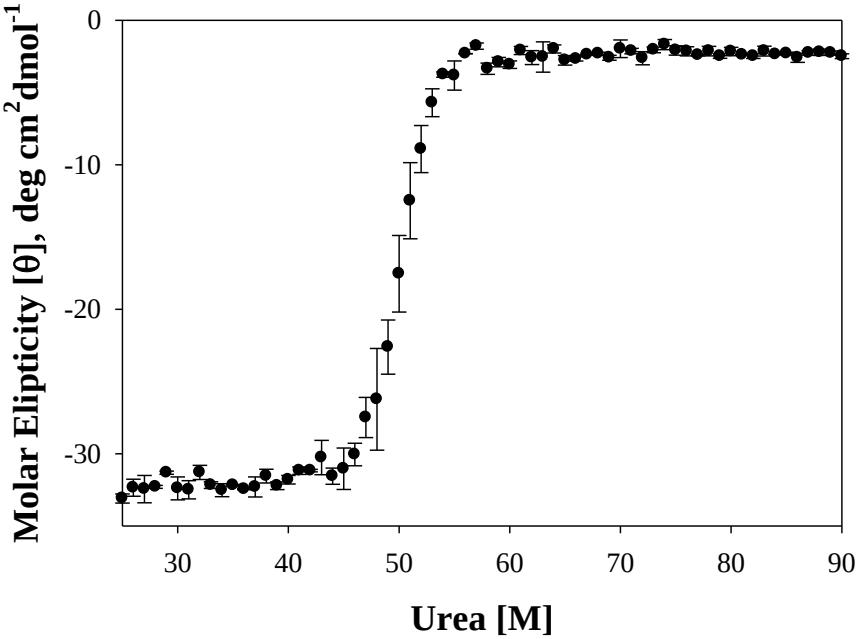

Figure 4.15: Three thermal denature sets were collected for Bt4131 ΔMg^{2+} using CD. Standard deviation from average is shown for 222 nm detection. Buffer: 20 mM HEPES, 5.0 mM NaCl pH 7.20. Protein concentration was 0.1 mg/mL

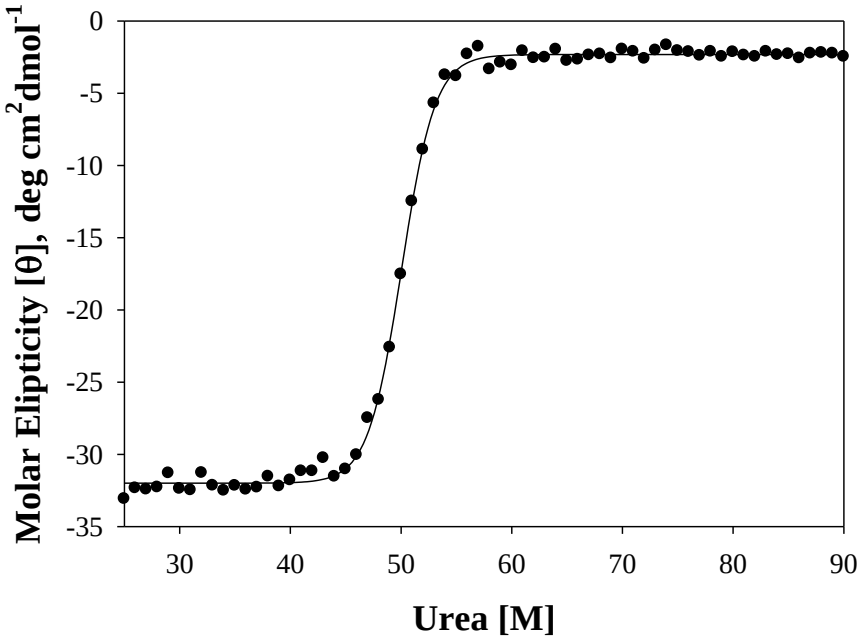

Figure 4.16: Fit line with averaged thermal CD data for Bt4131 ΔMg^{2+}. Fit is a 4-parameter sigmoidal fit with variables: **a**= 29.6701, **b**= 1.5126, $\mathbf{x_0}$= 50.0505, and $\mathbf{y_o}$= -31.9956.

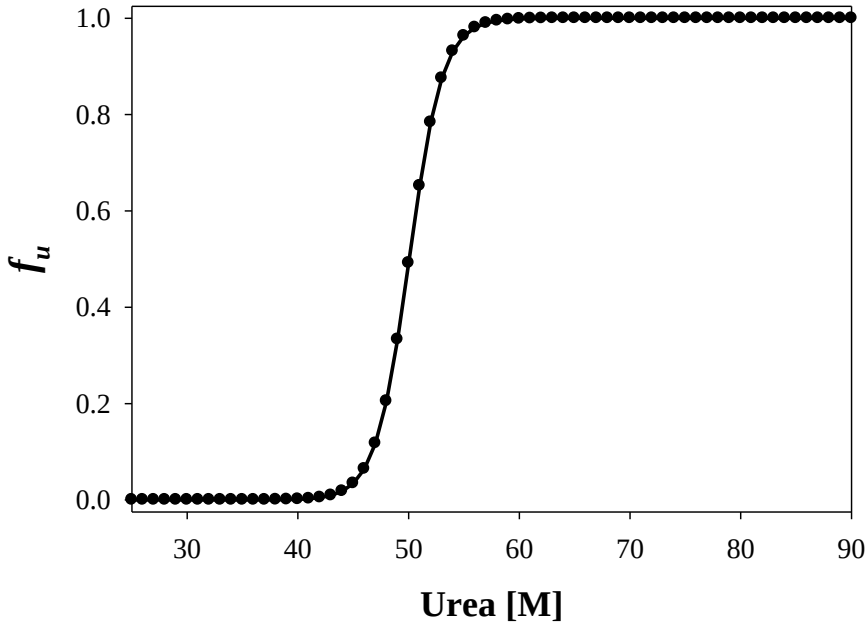

Figure 4.17: Normalized thermal denature fit line for Bt4131 ΔMg^{2+}.

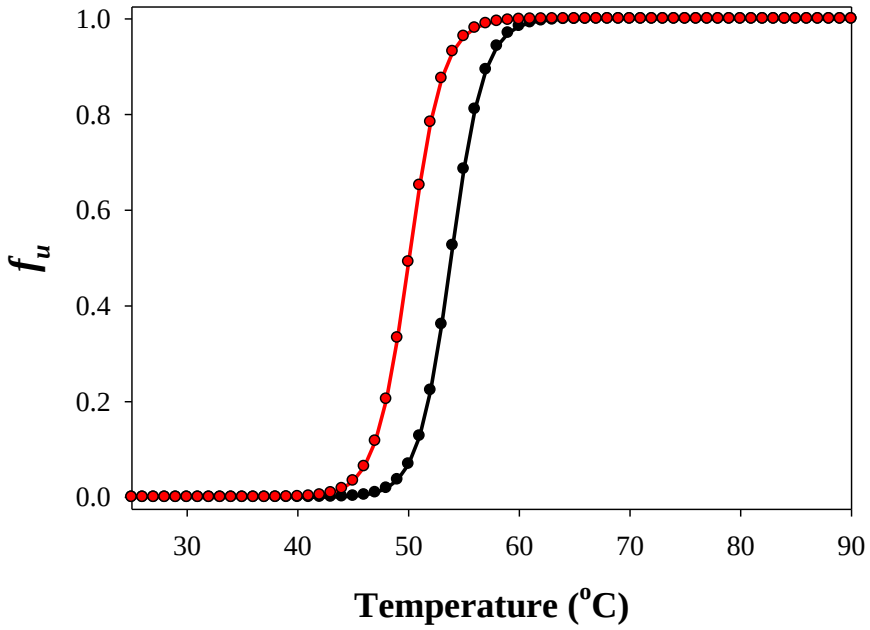

Figure 4.18: Overlay of thermal denature of Bt4131 with/without Mg^{2+} (BLACK, RED respectively). Plot clearly shows a destabilization when cofactor is absent.

4.3.3 Active site residues and the effects on stability when mutated

In previous chapters we have looked into the stability of the protein with active site mutations and the two that make the most impact are the two aspartates involved in the dephosphorylation of the substrate. In Bt4131 the two residues are D8 and D10. From the thermal denature we can see similar trends in the two mutations D8A and D10A with T_m of 52.17 ± 0.008 and 43.6 ± 0.5 respectively. Again we see the same trend of the first aspartate in the DXD motif is either stabilizing or have essentially no effect on the stability of the protein; whereas the D10A mutation decreases the stability of the protein Table 4.3.

Protein	T_m (ºC)
WT	53.8 ± 0.1
D8A	52.2 ± 0.1
D10A	43.6 ± 0.5

Table 4.3: Comparison of stability variables of native Bt4131 and DXD squiggle site-directed mutation. Protein buffer in all experiments was 20 mM HEPES, 2.5 mM $MgCl^{2+}$ pH 7.20. Chemical denature done with increasing urea concentration. protein concentration was ~10 μM in all experiments.

4.3.4 Activity and Stability of Domain Interfacing Point Mutations

To explore ligand induced conformation changes in Bt4131 both activity and stability of each point mutation was obtained Table 4.4. From T_m comparisons between domain-domain interacting point mutations we can see they play little to no part in structure stability with all mutations (except D10A) varying by ± 1.8 ºC from native Bt4131 T_m. Drastic variation in activity base on domain interfacing point mutation suggest connection between activity and interactions between cap and core.

Protein	T_m (°C)	k_{cat}	k_m	k_{cat}/k_m
WT	53.84 ± 0.09	3.6 ± 0.2	11 ± 2	
D8A	52.17 ± 0.08	---		
D10A	43.6 ± 0.5	[1]2 x 10^{-4}	[1]14	
R45A	53.96 ± 0.09	[1]ND	[1]ND	
R45K	53.92 ± 0.06	[1]5 x 10^{-3}	[1]5 x 10^{-3}	
K47A	55.77 ± 0.09	---		
N51A	54.74 ± 0.11	---		

Table 4.4: Includes the changes in stability and activity when compared to native Bt4131. All thermal denatures were carried out in same buffer, 20 mM HEPES, 2.5 mM $MgCl_2$ and pH 7.2. Activity was tested using Enzchek assay. [1]corresponds to published kinetic values for specific mutations.

4.3.3 Crystal Structure Generated Scattering Profiles Display Distinct Conformations

Scattering profiles display very distinct conformations depending on apo or liganded, Figure 4.19. From the CRYSOL generated scattering profiles, Figure 4.20, we can see that SAXS can distinguish conformational difference between the open and closed structures. Although SAXS cannot distinguish small changes; the techniques allows us to investigate near native solution conditions and how the protein reacts and shifts especially during the progression of reaction. From the theoretical scattering plots generated by CRYSOL we can look distinct modifications in low and mid q-range intensity.

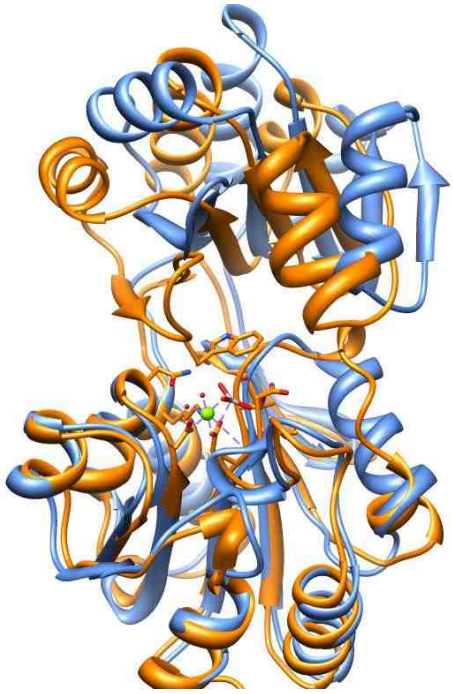

Figure 4.19: Overlay of Bt4131, PDB: 2RBK and Phyre generated structure of Bt4131. ORANGE displays the close conformation of Bt4131 whereas BLUE displays the generated open conformation of Bt4131.

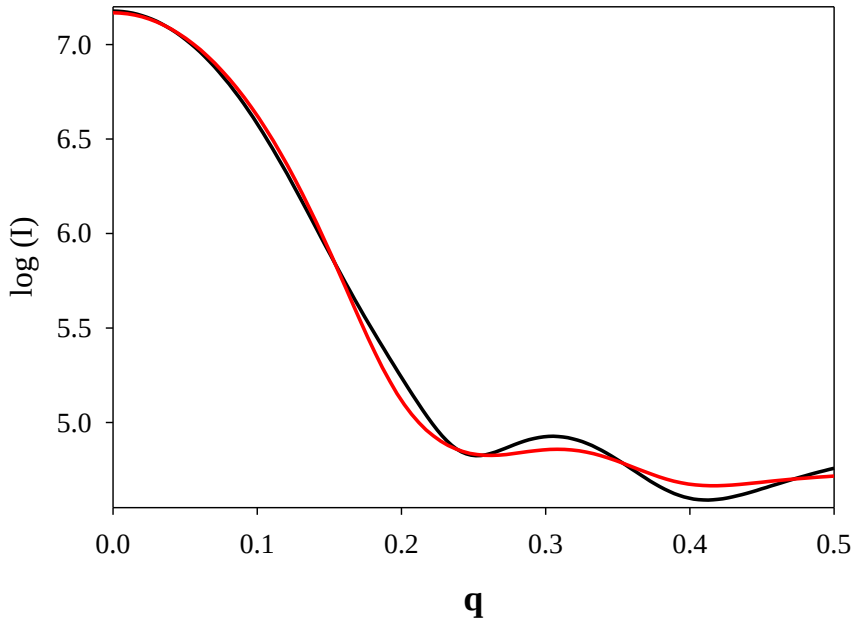

Figure 4.20: Theoretical scattering profiles for Bt4131 in open and closed conformation (RED, BLACK respectively). Profiles were generated from PDB structures (1YMQ* and 1YMQ) using CRYSOL.

4.3.5 Ligand Binding Induces Conformation Changes in Bt4131

Recombinant Bt4131 samples were exposed in the absence and presence of substrate analogs. Substrate analogs were used to bind similar to substrate without reacting; ligands were tested with protein at varying concentration (0-50 mM). Ligands included sodium metavanadate (VO_3), d-ribose and 2-dexoy-d-glucose Table 4.5.

From collected SAXS we can see distinct changes in conformation Figures 4.21-4.23 whereas predominant changes in structure were observed in samples that contained VO_3. Sodium metavanadate acts as a phosphate analog, which suggests binding of the phosphate group to active site of the core domain initiates a conformation change although changes in conformation due to sugar moiety binding was not near as extreme some samples induced a change in structure. Combination of both VO_3 and sugar moiety induced change similar to conformation change with VO3 independently.

Scattering data support an induced-fit mechanism of catalysis. For catalysis to occur a conformational change must occur to sequester bound substrate and active site from bulk solvent. Results suggest binding of substrate analogs initiates catalysis with binding of VO_3 inducing the largest change followed by VO_3 with d-ribose. Observed results show minimal change when binding sugar moiety but in the presence of vanadate change is substantially increased. Changes made by bound VO_3 were not seen in any other samples which could be accredited to the increased volume in the active when bound with VO_3 and sugar moiety. Off all the sugar moieties bound 2-doexy-d-glucose was the only ligand than induced a change in conformation, which is interesting because

2-deoxy-d-glucose is the suggested native substrate. This result may suggest a evolved interaction between native sugar moiety and cap domain which allows initiation of catalysis. From a collective view of substrate analog binding experiments we can conclude that Bt4131 progresses through an induced-fit mechanism tailored to activation through binding phosphate group and native substrate analog.

Substrate Present	**Guinier R_g**	**P (r) distribution Rg**	**d_{max}**
Empty	19.55	19.39	68.43
2-deoxy-d-glucose	19.26	19.03	63.09
d-riboose	19.18	19.26	67.14
l-arabinose	20.59	20.73	72.09
Pyridoxal	20.56	20.78	72.02
VO_3	18.79	18.62	57.31
2-deoxy-d-glucose + VO_3	18.88	18.73	63.24
d-riboose + VO_3	18.75	18.70	58.60
l-arabinose + VO_3	20.49	20.16	71.70
Pyridoxal + VO_3	20.18	20.43	70.61

Table 4.5: shows the calculated SAXS variable R_g (from Guinier and P(r) Distribution) as well as the d_{max} (calculated from P(r) Distribution). All experiments were runs at multiple concentrations with ligand in excess in 50 mM HEPES, 5mM $MgCl^2$, 75 mM NaCl, and 1 mM DTT, pH 7.0

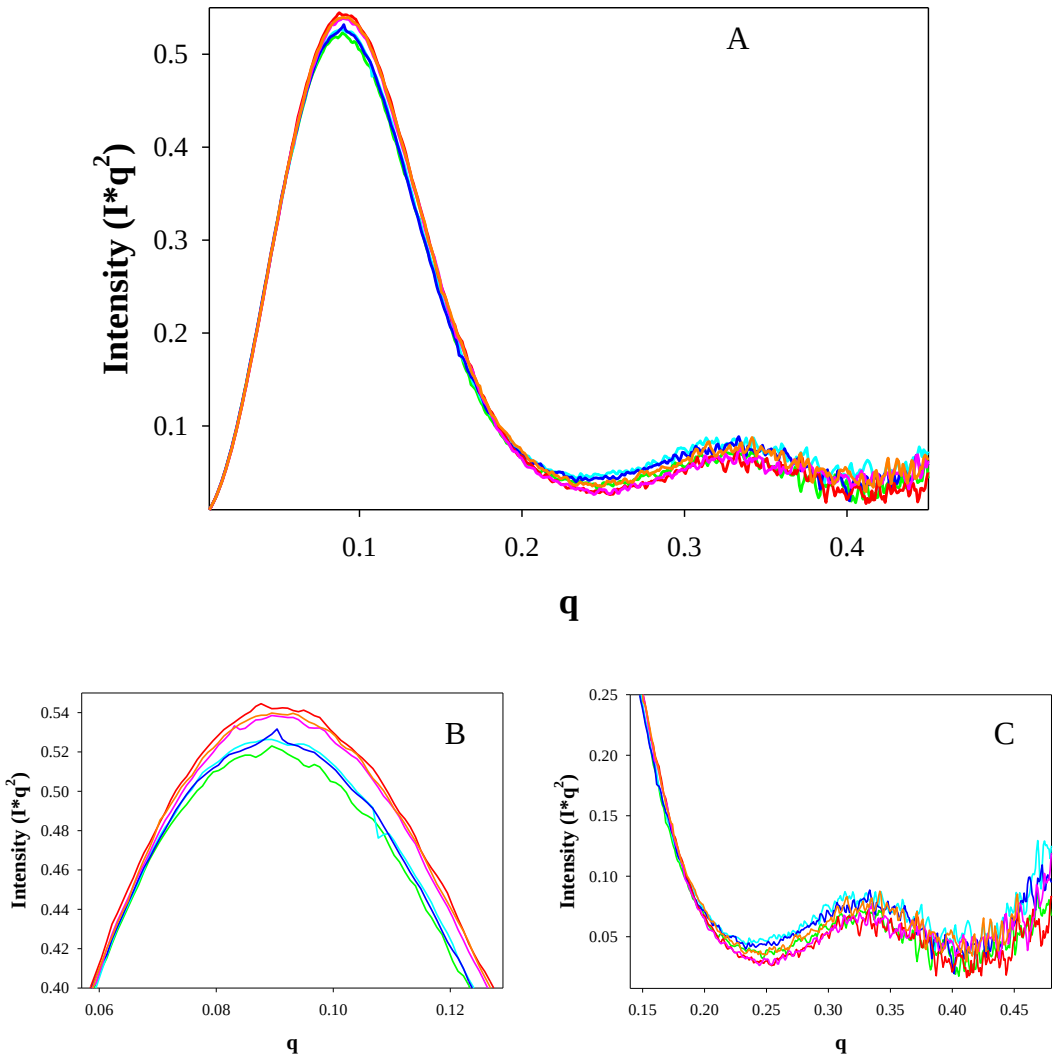

Figure 4.21: Kratky plot of Bt4131 w/wo substrate analogs. 6 experimental plots are represented although 2 distinct shapes are seen. Cool colors teal, green, blue) represent Bt4131 (empty, d-ribose, or 2-deoxy-d-glucose).Warms colors (red, orange, and pink) represent Bt4131+ VO_3 ± (d-ribose, or 2-deoxy-d-glucose). B and C are magnifications of key regions in scattering plots.

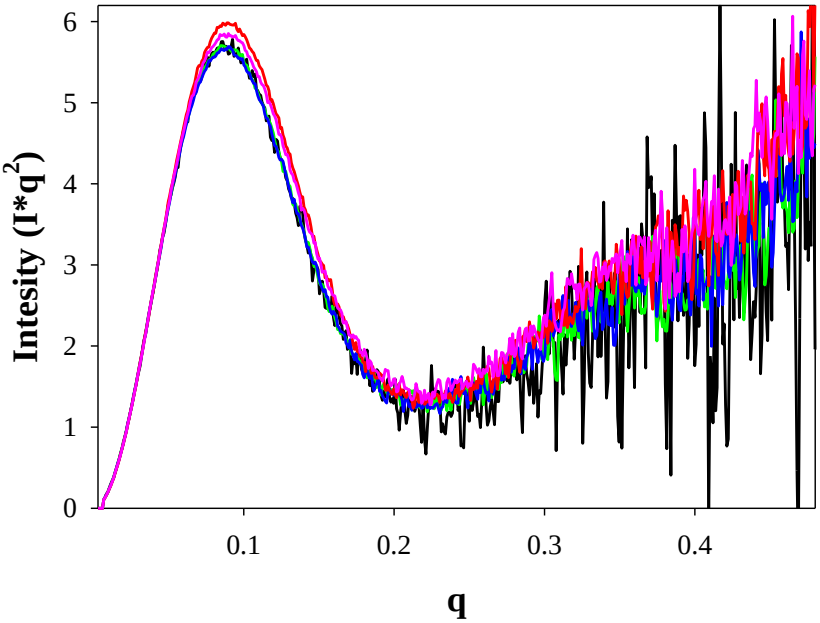

Figure 4.22: Kratky plot of Bt4131 w/wo substrate analogs. 5 experimental plots are represented although 2 distinct shapes are seen. Cool colors (black, green, blue) represent Bt4131 (empty, l-arabinose, or pyridoxal).Warms colors (red and pink) represent Bt4131+ VO_3 + (l-arabinose, or pyridoxal).

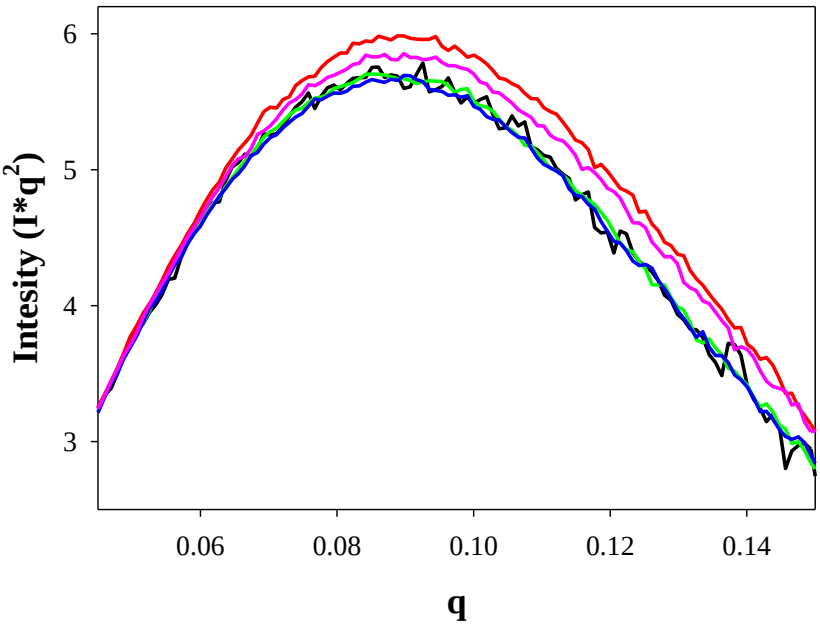

Figure 4.23: magnified Kratky plot (Figure 4.19) of Bt4131 w/wo substrate analogs. 5 experimental plots are represented although 2 distinct shapes are seen. Cool colors (black, green, blue) represent Bt4131 (empty, l-arabinose, or pyridoxal).Warms colors (red and pink) represent Bt4131+ VO_3 + (l-arabinose, or pyridoxal).

4.3.6 Domain-Domain interacting mutations and Solution Conformational changes induced.

After identifying the ligand responsible for induced closure of Bt4131, the next step is to identify key residues responsible for stabilization of a cap-closed conformation; most notable are domain-domain interacting residues. The table below lists the mutations tested and the values obtained Table 4.6. As we can see the mutations have a variety of effects on the empty solution conformations Figure 4.24. This set of results without substrate analogs identifies that the changes in these residues indeed effects the conformation and thus our activity, the table below Table 4.6 displays the effects on R_g and d_{max} when protein is with/without substrate analogs 2-deoxy-d-glucose and vanadate (the substrate used to investigate closure of cap were determine based on changes observed in the native protein). From previously proposed theories on induced conformational changes in proteins is necessary for catalysis (Johnson) we can see factors into activity in this C2 HAD protein through cap closure induced by phosphate binding.
In a specific case, R45A, had no activity and from our scattering experiments we can see that there is no changes in conformation when binding vanadate in combination with sugar metabolites (substrate analogs) Figure 4.25. This case helps demonstrate the necessity of cap movement in activity. Other mutants that do not move are R45K, and D8A.

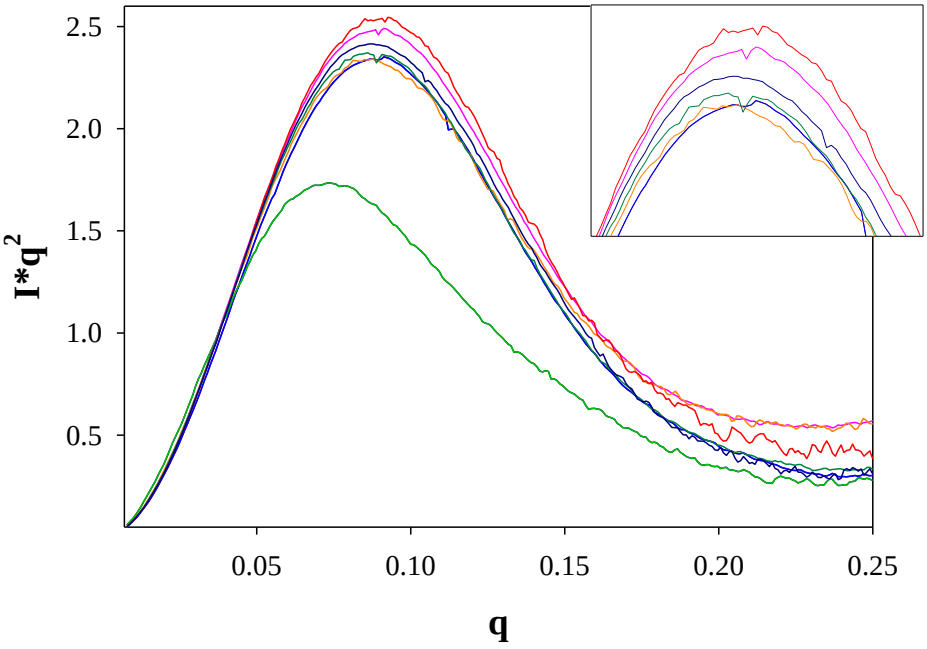

Figure 4.24: Kratky plot of SDMs of Bt4131 (Apo). From graph we can see that the proteins occupy many equilibrium conformations without ligand binding.

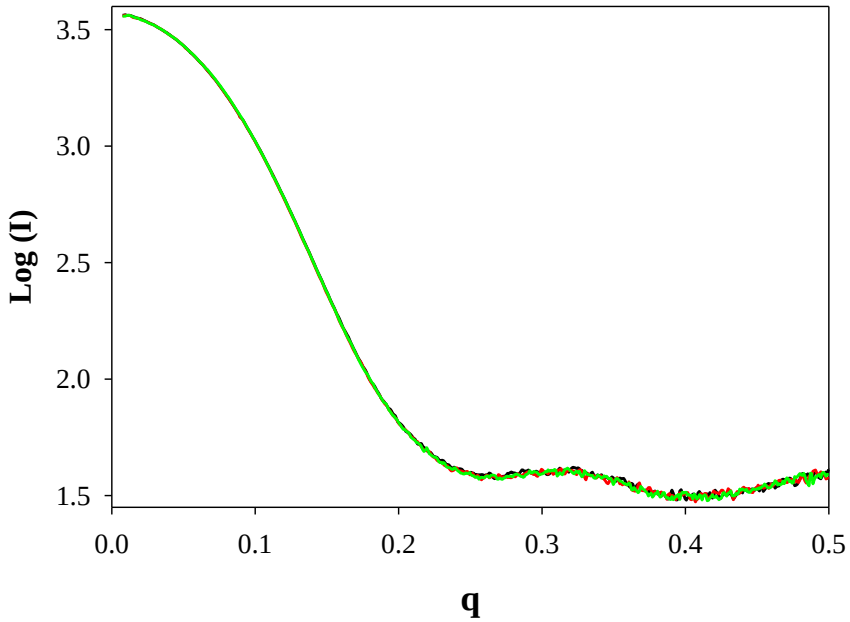

Figure 4.23: Scattering profile of R45A with and without substrate. As you can see the protein does not undergo any conformation change due to ligand binding (2-deoxy-d-glucose and VO_3).

Bt4131	$^{13}R_g$	$^{24}R_g$	$^{13}R_g$	$^{24}R_g$	$^{1}d_{max}$	$^{2}d_{max}$
WT	19.529 ± 0.315	19.087 ± 0.400	19.33	18.82	68.35	64.85
D8A	19.132 ± 0.353	17.036 ± 0.930	19.43	19.21	62.66	55.80
D10A	19.304 ± 0.558	18.449 ± 0.159	19.30	18.44	58.88	54.89
F16A	27.323 ± 1.247	28.789 ± 2.778	25.79	25.73	87.43	87.81
R45A	19.346 ± 0.178	19.330 ± 0.093	19.26	19.29	60.46	58.96
R45K	20.019 ± 1.043	19.339 ± 1.281	20.12	19.69	70.07	67.69
K47A	18.723 ± 0.200	17.798 ± 0.221	18.87	18.33	57.11	54.28
N51A	19.809 ± 0.197	NA	19.80	NA	67.84	NA
E77A	19.162 ± 0.257	17.727 ± 0.368	19.51	18.69	61.32	57.74

Table 4.6: Table includes R_g and d_{max} values for SD point mutation with and without substrate (vanadate and 2-deoxy-d-glucose). [1] represents protein without ligands, [2] denoted protein with ligands.

4.4 Conclusion

Once again we have seen similar stability effects due to changes in active site residues. Similar to other proteins examined, Bt4131 shows no change in stability due to D8 mutation, but a decrease in stability due to D10 mutation. We have seen the same trends through many different proteins, but do the changes correspond to a change in structure?

Small angle X ray scattering was used to investigate how the mutations effect the solution conformations and if so how does it affect activity. The DXD mutants showed similar structures and scattering values; the major difference between them and native protein was the difference between open and closed conformations.

As we have seen the stability of Bt4131 is similar to the other C2 HAD proteins tested. In all tested proteins the DXD plays a large role in protein stability. The stability is affected similar in the C2 capped proteins when Mg^{2+} is absent. In this chapter we used SAXS to connect fold and activity. From collected data we see that both interaction between residues at the interface and interaction between parts of substrate modify conformations. Results displayed two distinct conformations with the closed structure being stabilized with the binding of substrate analogs VO_3 and VO_3 + sugar moiety. Tested domain-domain interfacing residues modified structure as well, with the most interesting being R45A, R45K, D8A, and F16A. The Arg45 mutants and Asp8 mutants did not induce changes in structure whereas F16A displays a completely unique structure compared the rest. SAXS data helps establish a connection between solution structure and activity.

References

1. Lu, Z., D. Dunaway-Mariano, et al. (2005). "HAD Superfamily Phosphotransferase Substrate Diversification: Structure and Function Analysis of HAD Subclass IIB Sugar Phosphatase BT4131†,‡." Biochemistry **44**(24): 8684-8696

2. Xu, J., M. K. Bjursell, et al. (2003). "A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis." Science **299**(5615): 2074-2076

3. Midelfort, C. F. and I. A. Rose (1977). "Studies on the mechanism of Escherichia coli glucosamine-6-phosphate isomerase." Biochemistry **16**(8): 1590-1596.

4. Johnson, K. A. (2008). "Role of Induced Fit in Enzyme Specificity: A Molecular Forward/Reverse Switch." Journal of Biological Chemistry **283**(39): 26297-26301

5. L. Smolsky, P. Liu, M. Niebuhr, K. Ito, T. M. Weiss, and H. Tsuruta, Biological small-angle x-ray scattering facility at the stanford synchrotron radiation laboratory, Applied Crystallography. 2007

6. Svergun D.I., Barberato C. & Koch M.H.J. (1995) *J. Appl. Cryst.* **28**, 768-773.

7. D. Schneidman-Duhovny, M. Hammel, JA. Tainer, and A. Sali. Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophysical Journal 2013.

8. D. Schneidman-Duhovny, M. Hammel, and A. Sali. FoXS: A Web server for Rapid Computation and Fitting of SAXS Profiles. NAR 2010.38 Suppl:W540-4

CHAPTER FIVE

PROTEIN STABILITY AS A FUNCTION OF DOMAIN

5.1 Introduction

Previous studies on multi-domain proteins suggest that the stability of the 'host domain" (core domain) increases when a "guest domain" (cap domain) is absent (1). The stabilities of two sub-types in the HAD superfamily have been described previously. Proteins in each subfamily display similar stability traits associated with their active site residues and cofactor affinities, but members of each group have the ability to discriminate substrate binding through the presence of unique adornments such as loops and larger cap domains. Although the core domains are similar in these subgroups, the cap portions vary, which leads to differences in overall stabilities.

It is widely believed that the acquisition of a new activity by a protein compromises stability (1-3). A new function of HAD members would be introduced not only by a change in the mechanism of the catalyzed reaction but also by a change in cap residues involved in substrate binding. In this context, we asked the following important question. How does insertion of the cap domain affect the stability of the core domain of a HAD member. To gain information about this issue, proteins having domains present in two representative HADs, YbiV and Bt4131, were prepared and their stabilities were assessed. The six protein generated by using this strategy included the separate core and cap domains of YbiV and Bt4131 and, two chimera proteins, in which the cap domains of

YbiV and Bt4131 are swapped (Figure 5.1). The activities and structures of these two chimera proteins have been evaluated in a previous study (2). Stability, activity, and structure based investigations with the six truncated proteins were carried out in order to elucidate the complete story about why the HAD protein family is so robust and abundant.

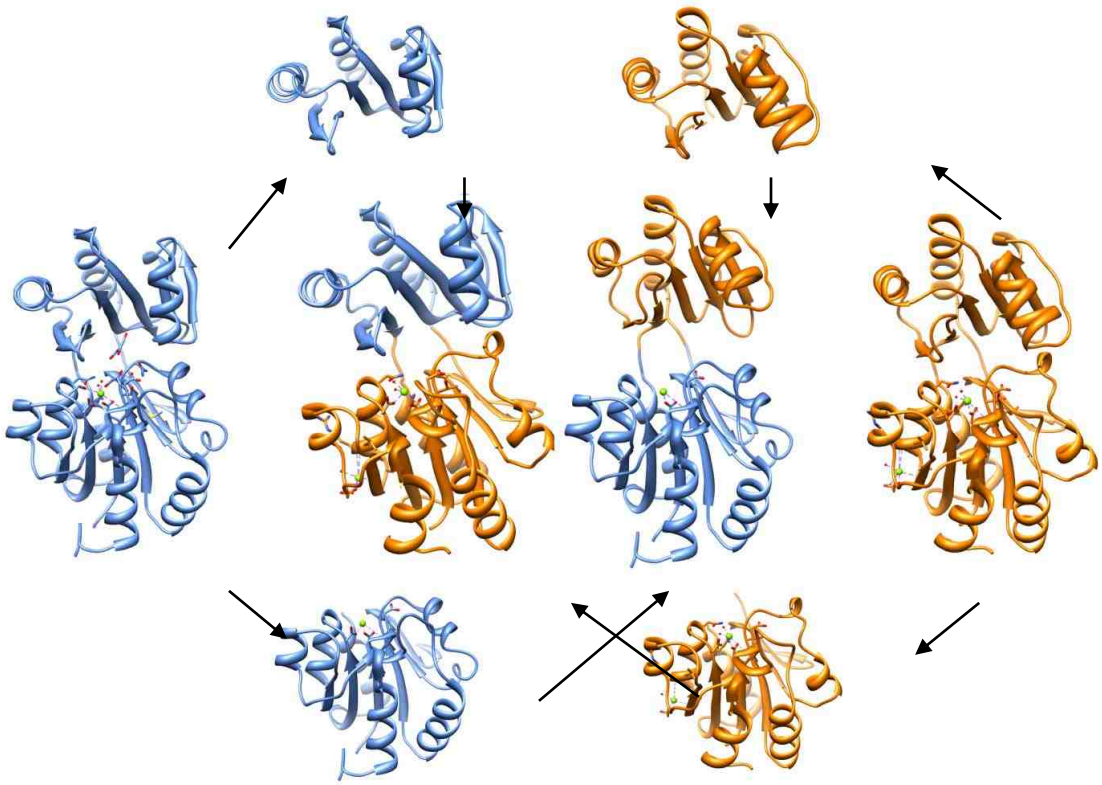

Figure 5.1: Cartoon representation of engineered truncations of YbiV and Bt4131. BLUE: YbiV and ORANGE: Bt4131.

5.2 Methods and Material

5.2.1 Synthesis of Hybrid Proteins

The constructed proteins were synthesized using Genetech Technologies. Proteins were codon optimized and transformed in both pUC57 and pET 28a vectors.

5.2.2 Purification of Capless ybiV

The capless YbiV protein plasmid capless YbiV/pET28a was transformed into Bl21 Iq competent cells using a given protocol. The capless YbiV/pET28a/Bl21 Iq cells were grown aerobically at 37 ºC containing 50 µg/mL kanamycin in LB broth to an $O.D_{600}$ 0.6-0.9 and induced for production of protein by adding 0.4 mM isopropyl-β-D-galactopyranoside (IPTG). Following a 12 h induction period at 18 ºC , cells were harvested by centrifugation at 6,500 rpm for 10 min at 4 ºC. Harvested cells were resuspended in 10 x weight lysis buffer (40 mM HEPES, 200 mM NaCl, 50 mM imidazole pH 7.5). Resuspended cells were lysed using a Amico French press at 1,200 psi, lysate was centrifuged at 20,000 rpm for 20 min 4 ºC. Supernatant from the centrifugation was loaded onto a NTA-Ni HisTrap FF 5 mL column (GE Life Sciences). The column was washed with lysis buffer and eluted using elution buffer (40 mM HEPES, 200 mM NaCl, 500 mM imidazole pH 7.5). Column fractions were tested for protein using SDS-page gel. All fractions containing pure protein samples were combined and dialyzed at 4 ºC against 3 changes 2L/2L/4L of 50 mM HEPES, 50 mM NaCl, 5mM $MgCl_2$ pH 7.3. The protein yield was *ca.* 15 mg/g wet cells.

5.2.3 Purification of Capless Bt4131

The Capless Bt4131 plasmid, **C**Bt4131/pET28a, was transformed into Bl21 Iq competent cells using a directed protocol. The **C**Bt4131/pET28a/Bl21 Iq cells were grown in a manner that is similar to that of the Capless YbiV with the exception of the final dialysis, which consisted of 3 rounds at 4 ºC 2L/2L/2L of 50 mM HEPES, 50 mM NaCl and 5 mM $MgCl_2$ pH 6.8. The protein yield was *ca.* 10 mg/g wet cells.

5.2.4 Purification of Bt413-Cap

The Bt4131 Cap plasmid, Bt4131**C**/pET28a, was transformed into Bl21 Iq competent cells using a directed protocol. The Bt4131**C**/pET28a/Bl21 Iq cells were grown in a manner that is similar to that of the protocol described above with the exception of the final dialysis, which consisted of 3 rounds at 4 ºC 2L/2L/2L of 50 mM HEPES, 50 mM NaCl and 5 mM $MgCl_2$ pH 6.8. The protein yield was *ca.* 12 mg/g wet cells.

5.2.5 Purification of ybiV-Cap

The YbiV Cap plasmid, YbiV-(cap)/pET28a, was transformed into Bl21 Iq competent cells using a directed protocol. The YbiV-(cap)/pET28a/Bl21 Iq cells were grown in a similar manner (see above) with the exception of the final dialysis; which consisted of 3 rounds at 4 ºC 2L/2L/2L of 50 mM HEPES, 50 mM NaCl and 5 mM $MgCl_2$ pH 7.0. The protein yield was *ca.* 15 mg/g wet cells.

5.2.6 Purification of Chimera ybiV(cap)/Bt4131 protein

The chimera YbiV(cap)/Bt4131 protein plasmid ybiV-Bt4131/pET28a was transformed into Bl21 Iq competent cells using the given protocol. The ybiV-Bt4131/pET28a/Bl21 Iq cells were grown in a similar manner (see above) with the following exceptions. This protein in less stable than the previous constructs and undergoes precipitation during elution and dialysis. When precipitation occurred, the protein sample was centrifuged for 5 min at 6500 rpm at 4 ºC. The supernatant was dialyzed with 3 buffer exchanges 2L/2L/2L of dialysis buffer (50 mM HEPES, 150 mM NaCl, and 5 mM $MgCl_2$ pH 7.0). The protein yield was *ca.* 10mg/g wet cells.

5.2.7 Purification of Chimera ybiV-core/Bt4131 protein

The chimera YbiV(core)/Bt4131 protein plasmid ybiV-Bt4131/pET28a was transformed into Bl21 Iq competent cells using a given protocol. The ybiV-Bt4131/pET28a/Bl21 Iq cells were grown in a similar manner (see above) with the exception of the final dialysis, which consisted of 3 rounds at 4 ºC 2L/2L/2L of 50 mM HEPES, 50 mM NaCl and 5 mM $MgCl_2$ pH 7.0. The protein yield was *ca.* 12 mg/g wet cells.

5.2.8 Site-Directed Mutagenesis

Site directed Mutagenesis was carried out using the Quick Change PCR Strategy (Stratagene). WT plasmids were used as templates and mutant plasmids were employed as templates when multiple mutations are generated for PCR amplification, oligionucleotides including mutated codon as primer and *Pfu* Turbo polymerase

(Strategene). Recombinant mutant plasmids were transformed into BL21 (DE3) Iq using small scale cultures for expression. Cultures showing expression of protein were verified using either gene sequencing or mass spectrometric analysis All mutated proteins were cultured and purified employing the protocol used for wild-type (WT) proteins.

5.2.9 Determination of Steady-State Kinetic Constants

K_{cat} and K_m values were determined using Enzchek phosphate assay kit, when applicable. This assay procedure is described in Chapter2.

5.2.10 Assessing Thermal Denaturation Using Circular Dichroism

Changes in the circular dichroism spectra as a function of temperature were used to detect changes in the secondary structures of the proteins. The procedures and fitting models employed for this purpose are described in Chapter 2.

5.2.11 Chemical Denaturation of Ordered Protein Structure

Three techniques were used to investigate the ordered structure of the truncated and chimeric proteins. Tryptophan fluorescence was used to investigate changes in the environment of aromatic amino acid residues (mainly tryptophan). To observe changes in the secondary structure of the protein, circular dichroism measurement were made at 222 and 217 nm. In order to evaluate changes in the tertiary structure, small angle scattering was employed. SAXS was used to generate a calculated solution volume, which was then

employed to identify the general shape of the protein in solution. The procedures and fitting models utilized in the chemical denaturation experiments are described in Chapter 2.

5.3 Results and Discussion

5.3.1 Constructed Capless Proteins (Capless YbiV, Bt4131) Exhibit Stabilities that Are Similar to Those of the Corresponding Native Proteins

To activities and stabilities of the truncated capless proteins were investigated to assess the possible destabilizing effects of cap domain inserts. Table 5.1 displays stability related values arising from the chemical and thermal denaturation studies. Importantly, the data show that both capless proteins are as or more stable than the corresponding native proteins Bt4131 and YbiV. In addition, the chemical denaturation data, and in particular differences in changes detected by using fluorescence and CD, demonstrate that capless YbiV unfolds by using a 3-state process. In previous studies with GmhB, YbiV and Bt4131, we observed that the fluorescence and CD denaturation profiles are similar and differ from those arising from SAXS measurements. Furthermore, although it is likely that capless Bt4131 also folds through a 3-state pathway, the shortage of data arising from a secondary observational technique prevents an unambiguous assignment of this route.

The data arising from studies with GmhB, YbiV and Bt4131, and the two capless proteins derived from Bt4131 and YbiV enables the reasonable conclusion that the core Rossmann-like domain in these proteins folds through a 3-state pathway Most notable is

the fact that capless proteins do not display a large increases in their stabilities, a result that suggests that the stability of the core domain is not greatly affected by changes associated with the addition of cap domains.

Protein	T_M	ΔG (cal*mol^{-1})	m (M*cal*mol^{-1})	d_{eq} (M)
Bt4131	53.84 ± 0.09	-3646.7[1]	-1466.4[1]	2.49[3]
Bt4131 Capless	51.43 ± 0.15	-3757.5[1]	-1321.0[1]	2.84[3]
YbiV	55.41 ± 0.10	-5685.4[1]	-1327.8[1]	4.28[3]
YbiV Capless	57.00 ± 0.05	-4497.1[1]	-1292.1[1]	3.48[3]
		9718.8[2]	2748.7[2]	3.53[3]

Table 5.1: Table shows the denature values, thermal and chemical denature, for both the WT and Capless proteins tested.

5.3.2 Structure Determination of Capless Proteins

In a preliminary effort, crystal trays (hanging drop method) were set up to test possible crystallization conditions for both capless Bt4131 and capless YbiV. Because the capless proteins are comprised of cores only, the starting conditions used were those employed for crystallization of M. *loti* GmhB (Structural Genomics Center). Under the tested conditions some crystals grew (Figure 5.2-5.3). For example, capless Bt4131 grew long thin rod-like crystals in 25% PEG 3350, 0.1 mM MES, 5mM $CaCl_2$ pH 6.5, whereas capless YbiV grew rectangular crystals in 0.1 M Bis-Tris, PEG 3350 20% pH 5.5, and 0.1 M MES, 5mM $CaCl_2$, PEG 3350 25% and 0.2 M $(NH_4)_2SO_4$, pH 6.5. Although the sizes

of the crystal are too small to be useful for crystallographic studies, this effort ha created a starting point for optimization of conditions to allow larger crystal growth.

Figure 5.2: Crystals grown for capless Bt4131. Crystal conditions: 25% PEG 3350, 0.1 mM MES, 5mM $CaCl_2$ pH 6.5. Crystal are the same conditions two separate wells

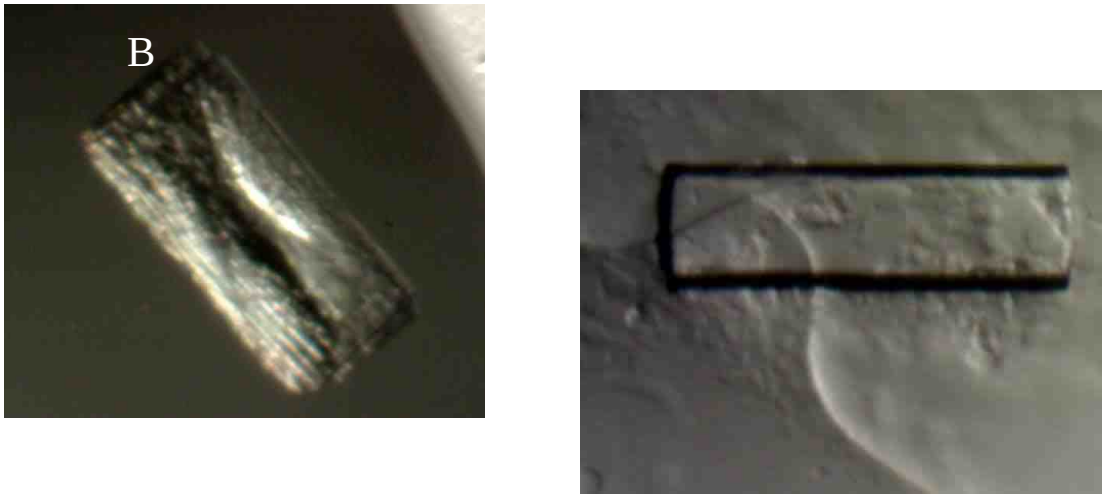

Figure 5.3: Crystals grown from capless YbiV protein. Crystal conditions A: 0.1 M Bis-Tris, PEG 3350 20% pH 5.5, and B: and 0.1 M MES, 5mM $CaCl_2$, PEG 3350 25% and 0.2 M $(NH_4)_2SO_4$, pH 6.5

In order to determine if the capless proteins are folded in solution, the small angle scattering techniques was employed. Inspection of the scattering profiles of both capless Bt4131 and capless YbiV, displayed as Kratky plots in Figure 5.4, shows that each protein has a defined tertiary structure.

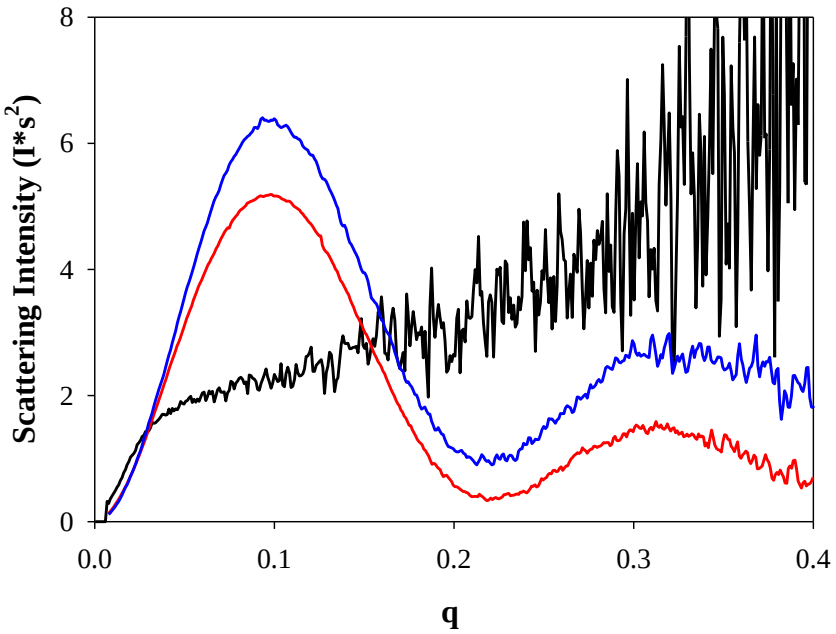

Figure 5.4: Kratky plot of capless YbiV (BLUE) and capless Bt4131 (RED) overlaid with scattering profile of protein without structure (BLACK).

5.3.3 Cap Proteins are stable in Absence of Core Domains

The results show that the capless HAD proteins are as stable as their native counterparts. Our studies were next directed at determining the stabilities of the two cap domains. In attempts to purify both proteins in large quantities, precipitation occurred during dialysis although soluble protein was collected. Both proteins were thermally denatures to investigate the stability. From thermal denature of the cap-only proteins we can see that both caps are stable (T_m 48.0 ± 1.5, and 52.5 ± 0.5 for YbiV-(cap) and

Bt4131-cap respectively). From the collected cap-protein thermal we can conclude that the cap only proteins are as stable as the core domains separately. With a stable core and cap domain separately we can infer stability when the two domains are together seen in the native protein structures

5.3.4 HAD Active Site Point Mutation Trends Observed in Capless Proteins

In previous chapters, we described the effects of active site point mutations on the stabilities of HAD proteins. Observations made show that effects of specific point mutations display trends. Because the truncated capless YbiV and capless Bt4131 proteins have been shown to fold and are stable, active site point mutations were created to see if the stability trends are also followed by the mutant truncated proteins. The results of studies with these proteins show that the effects of mutations follow similar trends in both the native and capless proteins. The D9A (capless YbiV) and D8A (capless Bt4131) have similar stabilities to their respective native capless proteins (Table 5.2). As predicted, the D10A and D11A mutants have substantially lower T_m values (*ca.* 22 ºC for capless YbiV and precipitated capless Bt4131) corresponding to lower stabilities. Point mutations at third aspartic acid moiety (Asp122) of capless YbiV that interacts with Mg^{2+} in the native structure do not alter the stability of the protein as reflected by both the T_m and ΔG values. One noticeable difference between the native protein and capless Bt4131 is that D8A/D10A mutation of the capless protein leads to recovered stability corresponding to an increase in T_m. The changes in stability brought about by point

mutations can be attributed to changes in the of interactions between the side chains of the cap and core amino acids and switches to interactions with solvent.

Protein	T_m (°C)	ΔG (cal*mol^{-1})	m (M*cal*mol^{-1})	d_{eq} [M]
Capless Bt4131	51.43 ± 0.15	-3841.6	-1366.8	2.81
D8A	59.53 ± 0.07	-4791.4	-1065.7	4.50
D10A	ND*	ND*	ND*	ND*
D8A/D10A	49.09 ± 0.15			
Capless YbiV	57.00 ± 0.05	-4497.1	-1292.1	3.48
D9A	58.44 ± 0.08			
D11A	35.35 ± 0.11			
D122A	61.79 ± 0.08	-4825.3	-1079.9	4.47

Table 5.2: Table shows the denature values, thermal and chemical denature, for both core-only proteins (Capless Bt4131 and Capless YbiV) and corresponding aspartate mutants (catalytic and cofactor binding (Mg2+) mutants.

5.3.5 Folding of the Bt4131-core/YbiV-cap protein.

The recombinant chimeric protein Bt4131-core/YbiV-cap, constructed and purified on a large scale, was found to be instable (Table 5.3). Circular dichroism studies showed that this protein has a preserved secondary structure at 20 ºC. Kinetic analysis demonstrated that this chimeric protein does not catalyze hydrolysis of ribose-5-phosphate, glucose-6-phosphate, or para-nitrophenyl phosphate (pNPP). Thus results suggest that destabilization is caused by fusing the these cap and core domains but that

this effect does not alter the ability of the protein to fold, indicating that folding of the core domain is able to endure the destabilizing incorporation of the non-natural cap.

Protein	T_M	ΔG (cal*mol^{-1})	m (M*cal*mol^{-1})	d_{eq} (M)
Bt4131	53.84 ± 0.09	-3646.7	-1466.4	2.49
YbiV	55.41 ± 0.10	-5685.4	-1327.8	4.28
Chimera (Bt4131/YbiV)	33.04 ± 0.14	---	---	---
Chimera (YbiV/Bt4131)		-2714.8	-1037.2	2.62

Table 5.3: Table shows the denature values, thermal and chemical denature, for both the WT and constructed chimera proteins (Bt4131/YbiV and YbiV/Bt4131).

5.3.6 The YbiV(core)/Bt4131(cap) Protein Has Native-Like Characteristics

A chimeric protein consisting of the core domain of YbiV and cap domain of Bt4131 was prepared and subjected to studies probing its activity, folding ability and stability. The results of chemical denaturation experiments indicate that this chimeric protein is moderately stable. Although displaying no catalytic activity toward ribose-5-phosphate, glucose-6-phosphate, or pNPP, the protein is marginally more stable than the chimeric protein, Bt4131 (core)/YbiV(cap) as demonstrated by its handling characteristics. While the difference between these two chimeric proteins could merely reside in linker connections, it is possible that the difference is a consequence of the fact that Bt4131 is less stable than YbiV (see above), and that the relative stabilities of the chimeric proteins is a consequence of the characteristics of the core domains.

5.4 Conclusion

The HAD proteins have evolved by incorporation of cap domains, which assists substrate binding and creates diversity (3-7). In the effort described in this Chapter, truncated and chimeric HAD proteins were prepared in order to examine the stabilities of individual domains in HAD members. As seen by inspecting the results of investigations of other proteins with multiple domains, in most cases the truncated proteins are more stable (1). In contrast, the current studies show that core domains of the HADs have stabilities that are similar to those of the native proteins. This observations demonstrates that the Rossmann-like core domains of the HADs are robust, a possible reason why this family tolerates cap domain changes that bring about substrate diversification.

Importantly, the stabilities of the truncated core proteins and their native HAD counterparts respond to point mutations of active site residues in the same manner. In the absence of the cap domain, capless YbiV core domain still undergoes folding through the 3-state pathway. The truncated cap domains of both Bt4131 and YbiV show similar stability to the core domain truncations.

The chimeric proteins, in which the cap domains are switched with minimal changes of linker residues, retain the ability to fold into ordered. This phenomenon structures leads us to believe the HAD family is sufficiently robust to accommodate large changes (for example electrostatics, domain interfacing surface interactions) in the structures of cap domains. However, the YbiV-core/Bt4131-cap chimeric protein does not retain the catalytic activities of either of its domain components. Thus, the stability of

the structure of this chimeric protein serves supports a protein evolutionary model that involves incorporation of recognition domains into catalytically active cores domains. For this model to be evolutionarily applicable the core protein domain, like that in the HADs, must be robust so that alterations of the recognition domains do not create unacceptable levels of instability.

References

1. Gabriel Zoldák, Linn Carstensen, Christian Scholz, Franz X. Schmid, Consequences of Domain Insertion on the Stability and Folding Mechanism of a Protein, Journal of Molecular Biology, Volume 386, Issue 4, 6 March 2009, Pages 1138-1152, ISSN 0022-2836, http://dx.doi.org/10.1016/j.jmb.2008.12.052.

2. Seifried, A., G. Knobloch, et al. (2014). "Evolutionary and Structural Analyses of Mammalian Haloacid Dehalogenase-type Phosphatases AUM and Chronophin Provide Insight into the Basis of Their Different Substrate Specificities." Journal of Biological Chemistry **289**(6): 3416-3431.

3. Maxwell Burroughs, Karen N. Allen, Debra Dunaway-Mariano, L. Aravind, Evolutionary Genomics of the HAD Superfamily: Understanding the Structural Adaptations and Catalytic Diversity in a Superfamily of Phosphoesterases and Allied Enzymes, Journal of Molecular Biology, Volume 361, Issue 5, 1 September 2006, Pages 1003-1034, ISSN 0022-2836.

4. Eugene V. Koonin, Roman L. Tatusov, Computer Analysis of Bacterial Haloacid Dehalogenases Defines a Large Superfamily of Hydrolases with Diverse Specificity: Application of an Iterative Approach to Database Search, Journal of Molecular Biology, Volume 244, Issue 1, 17 November 1994, Pages 125-132, ISSN 0022-2836.

5. Vivek Anantharaman, L Aravind, Eugene V Koonin, Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins, Current Opinion in Chemical Biology, Volume 7, Issue 1, February 2003, Pages 12-20, ISSN 1367-5931.

6. Nguyen, H. H., L. Wang, et al. (2010). "Structural Determinants of Substrate Recognition in the HAD Superfamily Member d-glycero-d-manno-Heptose-1,7-bisphosphate Phosphatase (GmhB)." Biochemistry **49**(6): 1082-1092.

7. Lu, Z., D. Dunaway-Mariano, et al. (2005). "HAD Superfamily Phosphotransferase Substrate Diversification: Structure and Function Analysis of HAD Subclass IIB Sugar Phosphatase BT4131†,‡." Biochemistry **44**(24): 8684-8696.

CHAPTER SIX

STRUCTURAL INSGIHT: AN INVESTIGATION OF SOLUTION CONFORMATION OF PYRUVATE PHOPSHATE DIKINASE

6.1 Introduction

Pyruvate phosphate dikinase (PPDK) is a multi-domain protein that reversible catalyzes the conversion of ATP, inorganic phosphate, and pyruvate to AMP, pyrophosphate and phospho*enol*pyruvate (Figure 6.1). The protein in solution is a monodimer (monomer MW ~96,000 Da) consisting of three separate domains: N-terminal nucleotide binding domain, central domain phosphocarrier, and C-terminal PEP/pyruvate binding domain.

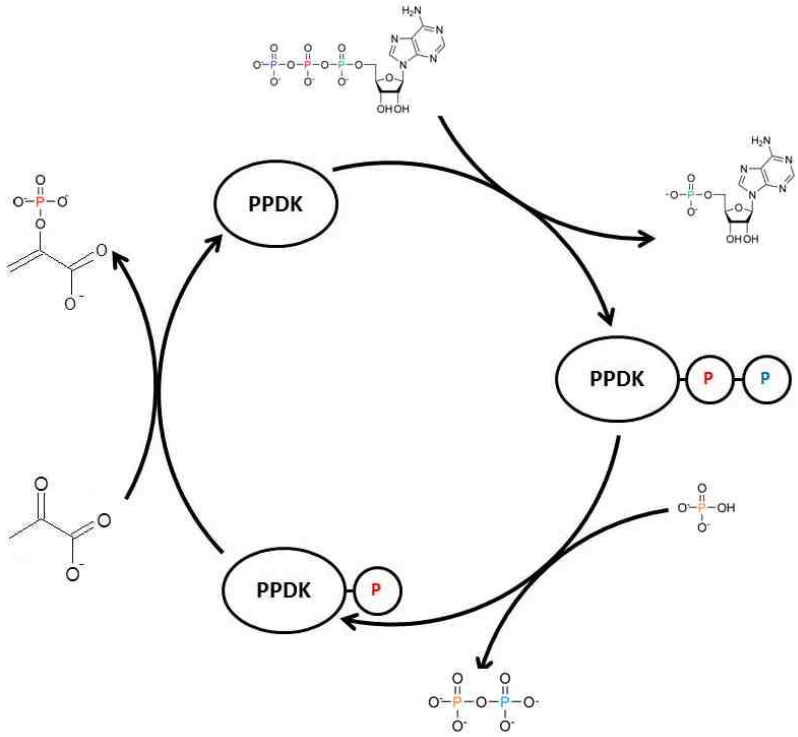

Figure 6.1: schematic represents the cycle through catalysis of PPDK.

PPDK has been extensively characterized. The two predominant catalytic conformations have been crystallized (PDB: 1DIK and 2R82) Figure 6.2. These two structures differ by which domain is interacting with the phosphocarrier central domain. To obtain the phosphocarrier central domain interacting with the PEP/pyruvate domain a triple mutant was created to 'trap' the conformation (1).

Recent studies have shown that X-ray crystal structures do not always represent accurate solution conformations (3-4). In this chapter we investigate modest differences between X-ray structures and what is in solution. Experiments using small angle X-ray scattering were used to observe solution conformations of PPDK. PPDK is an ideal candidate for SAXS due to the size and unique conformations cycled through due to catalysis.

In this chapter we will look at the differences in conformation due to detection technique (X-ray crystallography and X-ray scattering) as well as investigate the equilibrium state of the central domain in solution. We have X-ray structures showing interaction between central domain and the two other domains, but what is the equilibrium state of the central domain in solution? Does the central domain interact closely to the rest of the protein structure or does the central domain 'float' in solution until conformation changes summon interactions between domains.

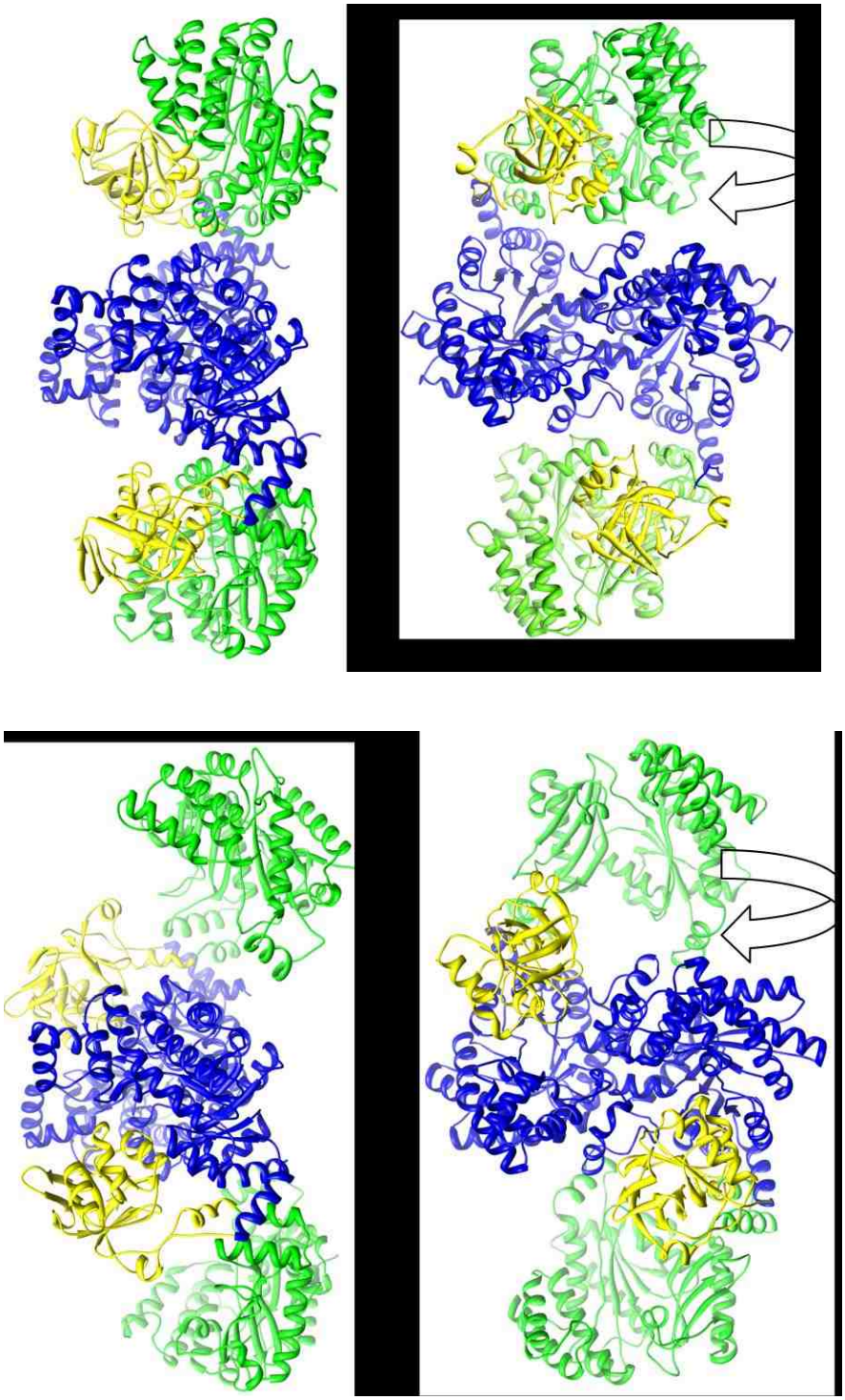

Figure 6.2: X-ray crystal structures of PPDK in both conformations (central domain interacting with N-terminal or C-terminal domains. **A.** PDB: 1DIK**,** left: side view and right: 90º turn (central domain interacting with N-terminal) **B.** PDB: 2R82 left: side view and right: 90º turn (central domain interacting with C-terminal)

6.2 Material and Methods

6.2.1 Materials

All restriction enzymes and T4 ligase were purchased from Invitrogen. Pfu Turbo polymerase was purchased from Stratagene. Oligonucleotides were custom-synthesized from Invitrogen. To verify correct sequence DNA sequencing was performed by MC labs and MW mass spec determination was carried out by UNM Mass Spec Facilities.

6.2.2 Cloning and Purification of WT C. *symbiosum* PPDK

WT PPDK was grown from plasmid encoding the gene pACYC184-D12 according to 2. Cells were grown aerobically at 37 ºC (with agitation at 250 rpm) containing 10 µg/mL tetracyclin in TB broth supplemented with 4 mL glycerol to an $O.D_{600}$ 0.5-1.0 and induced for production of protein by adding 1mM IPTG (final concentration). Following a 12 hour induction period at 17 ºC, cells were harvested by centrifugation at 6,500 rpm for 10 minutes. Harvested cells were resuspended in 10x w/v buffer A (50 mM HEPES, pH 7.5), and were lysed using a Amico French press at 1,200 psi, lysate was centrifuged at 20,000 rpm for 20 minutes. Protein was fractioned by ammonium sulfate induced precipitation. The 50-90% ammonium sulfate fraction was collected and centrifuged at 20,000 rpm for 20 minutes, the pellet was collected and re-suspended in buffer B (50 mM HEPES 16% ammonium sulfate pH 7.5). The protein solution was loaded onto a 10 x 300 mm (HR10 column) BS sepharose column, pre equilibrated with Buffer B at 4º C. The column was washed with 10 COLUMN VOLUME buffer B and protein was eluted with a linear gradient (10 CV) 100-0 % buffer

B (16-0% $(NH_4)_2SO_4$). The fractions containing the protein were collected and concentrated using 50K Amicon Ultra centrifugal filters. Protein solution was dialyzed with buffer A (3 rounds X 2L X2 hours each) to rid the solution of ammonium sulfate. The dialyzed supernatant was loaded onto a DEAE sepharose column (10 X 300 mm HR10 column) equilibrated with buffer A. the column was washed with 5 CV buffer A and protein was eluted with a linear gradient from 0 - 1M KCl. Protein fractions were collected and concentrated to load onto Size exclusion column (SEC) for buffer exchange and final clean-up of sample. Final fractions were collected and concentrated. Protein samples were stored at -80 ^oC. The protein yield was 10 mg protein/ 1 g wet cells.

6.2.3 Cloning and Purification of Δ340 PPDK

Δ340 PPDK protein was cloned using PCR amplification from the plasmid pACYC184-D12 containing WT PPDK. PCR using pACYC184-D12 as a template and Pfu Turbo polymerase. Custom oligonucleotides (5`-AAGAGAACAGCTAGCGCTGCTCTTCAGATT, and 3` -TTTGCTCGAGAACTCTACGGAAGAAGGAT), containing the restriction endonuclease cleavage sites NheI and XhoI, were used in the PCR reactions. The pET28a vector was cut by the restriction enzymes NheI and XhoI and ligated to the PCR product that has been isolated and digested with the same restriction enzymes. The ligation product (pET28a-Δ340 PPDK) was used to transform E. *coli* Iq competent cells.

The transformed cells were grown at 37 ^oC (with agitation at 250 rpm) in Terrific broth (TB) containing 4 mL glycerol and 50 µg/mL kanamycin until OD_{600} reached

approximately 0.6-1.0. The cells were induced overnight at 17^{o} C with 1 mM IPTG. Cells were harvested by centrifugation at 6,500 rpm for 10 minutes. Harvested cells were resuspended in 10x w/v buffer A (50 mM HEPES, 200 mM NaCl, and 50 mM imidazole, pH 7.5). Suspended cells were lysed using a Amico French press at 1,200 psi, cell lysate was centrifuged at 4 ^{o}C for 20 minute at 20,000 rpm. Supernatant from centrifugation was loaded onto a NTA-Ni HisTrap FF 5 mL column (GE Life Sciences). Column was washed with buffer A to clean any un-bound protein and bound protein was eluted using buffer B (50 mM HEPES, 200 mM NaCl, 500 mM imidazole pH 7.5). Column fractions were tested for clean protein using SDS-page gel; all fractions containing clean protein samples were combined and dialyzed at 4 ^{o}C against 3 changes 2L/2L/4L of 50 mM HEPES, 50 mM NaCl, 5mM $MgCl_2$ pH 7.2. Protein yield was ~10 mg/g wet cells.

6.2.4 Purification of CentI PPDK

Truncated protein named CentI was created to express only the central domain of the WT PPDK protein which included residues 382-503. Cells were grown from cell stock encoding the His-tagged CentI protein. Cells were grown aerobically at 37 ^{o}C (with agitation at 250 rpm) in Terrific broth (TB) containing 4 mL glycerol and 50 µg/mL ampicillin to an $O.D_{600}$ 0.5-1.0 and induced for production of protein by adding 1 mM IPTG (final concentration). Following a 12 hour induction period at 17 ^{o}C, cells were harvested by centrifugation at 4^{o}C for 10 minutes at 20,000 rpm. Cell pellet was suspended in 10X w/v buffer A (50 mM HEPES, 200 mM NaCl, 50 mM imidazole pH 7.5). Supernatant was lysed using an Amico French press at 1,200 psi. Cell lysate was

centrifuged at 4 ºC for 20 minute at 20,000 rpm. Supernatant from centrifugation was loaded onto a NTA-Ni HisTrap FF 5 mL column (GE Life Sciences). Column was washed with buffer A to clean any un-bound protein and bound protein was eluted using buffer B (50 mM HEPES, 200 mM NaCl, 500 mM imidazole pH 7.5). Column fractions were tested for clean protein using SDS-page gel; all fractions containing clean protein samples were combined and dialyzed at 4 ºC against 3 changes 2L/2L/4L of 50 mM HEPES, 50 mM NaCl, 5mM $MgCl_2$ pH 7.2. Protein yield was ~5 mg/g wet cells.

6.2.5 Purification of (R219E/E271R/S262D) PPDK

(R219E/E271R/S262D) PPDK was grown from cell stock encoding the gene according to (1). Cells were grown aerobically at 37 ºC (with agitation at 250 rpm) containing 10 µg/mL tetracyclin in TB broth supplemented with 4 mL glycerol to an $O.D_{600}$ 0.5-1.0 and induced for production of protein by adding 1mM IPTG (final concentration). Following a 12 hour induction period at 17 ºC, cells were harvested by centrifugation at 6,500 rpm for 10 minutes. Harvested cells were suspended in 10x w/v buffer A (50 mM HEPES, pH 7.5) and lysed using an Amico French press at 1,200 psi. Lysate was centrifuged at 4 ºC for 20 minutes at 20,000 rpm. Protein was fractioned by ammonium sulfate induced precipitation. The 50-90% ammonium sulfate fraction was collected and centrifuged at 4 ºC for 20 minutes at 20,000 rpm, the pellet was collected and re-suspended in buffer B (50 mM HEPES 16% ammonium sulfate pH 7.5). The protein solution was loaded onto a 10 x 300 mm (HR10 column) BS sepharose column, pre equilibrated with Buffer B at 4º C. The column was washed with 10 CV buffer B and

protein was eluted with a linear gradient (10 CV) 100-0 % buffer B (16-0% $(NH_4)_2SO_4$). The fractions containing the protein were collected and concentrated using 50K Amicon Ultra centrifugal filters. Protein solution was dialyzed with buffer A (3 rounds X 2L X2 hours each) to rid the solution of ammonium sulfate. The dialyzed supernatant was loaded onto a DEAE sepharose column (10 X 300 mm HR10 column) equilibrated with buffer A. the column was washed with 5 CV buffer A and protein was eluted with a linear gradient from 0 - 1M KCl. Protein fractions were collected and concentrated to load onto Size exclusion column (SEC) for buffer exchange and final clean-up of sample. Final fractions were collected and concentrated. Protein samples were stored at -80 ^oC. The protein yield was 12 mg protein/ 1 g wet cells.

6.2.6 Purification of Tem540 PPDK

Truncated protein Tem533 was created to express the N-terminal and central domains of the WT PPDK protein which included residues 1-533. Cells were grown from cell stock encoding TEm533 protein. Cells were grown aerobically at 37 ^oC (with agitation at 250 rpm) in Terrific broth (TB) containing 4 mL glycerol and 50 µg/mL ampicillin to an $O.D_{600}$ 0.5-1.0 and induced for production of protein by adding 1 mM IPTG (final concentration). Cells were harvested by centrifugation at 4 ^oC for 10 minutes at 6,500 rpm. Harvested cells were suspended in 10x w/v buffer A (50 mM HEPES, pH 7.5) and lysed using an Amico French press at 1,200 psi. Lysate was centrifuged at 4 ^oC for 20 minutes at 20,000 rpm. Collected supernatant was loaded onto a DEAE sepharose column (10 X 300 mm HR10 column) equilibrated with buffer A. Column was washed

with 10 CV buffer A and protein was eluted using a linear gradient 0-100% buffer B (50 mM HEPES, 1 M KCl pH 7.5). Protein fractions were collected equilibrated with 16% ammonium sulfate and loaded onto a BS sepharose column (10 X 300 mm HR10 column) equilibrated with buffer A (50 mM HEPES, 16% $(NH_4)_2SO_4$ pH 7.5). Column was washed with 5 CV buffer A and protein was eluted with linear gradient 0-100% buffer B (50 mM HEPES pH7.5). Protein fractions were collected and concentrated using a 30K Amicon Ultra Centrifugal filter. The final protein sample was loaded onto a size exclusion column to buffer exchange and clean any residual contaminants. Purification yielded 10 mg/g wet cells.

6.2.7 Purification of Tem340 PPDK

Truncated protein Tem340 PPDK was created to express the N-terminal domain of the WT PPDK protein which included residues 1-340. Cells were grown from cell stock encoding Tem340 protein. Cells were grown aerobically at 37 ºC (with agitation at 250 rpm) containing 50 µg/mL ampicillin in TB broth supplemented with 4 mL glycerol to an $O.D_{600}$ 0.5-1.0 and induced for production of protein by adding 1mM IPTG (final concentration). Following a 12 hour induction period at 17 ºC, cells were harvested by centrifugation at 6,500 rpm for 10 minutes. Harvested cells were resuspended in 10x w/v buffer A (50 mM HEPES, pH 7.5) were lysed using an Amico French press at 1,200 psi, lysate was centrifuged at 20,000 rpm for 20 minutes. Protein was fractioned by ammonium sulfate induced precipitation. The 40-90% ammonium sulfate fraction was collected and centrifuged at 20,000 rpm for 20 minutes, the pellet was collected and re-

suspended in buffer B (50 mM HEPES 16% ammonium sulfate pH 7.5). The protein solution was loaded onto a 10 x 300 mm (HR10 column) BS sepharose column, pre equilibrated with Buffer B at 4^{o} C. The column was washed with 10 CV buffer B and protein was eluted with a linear gradient (10 CV) 100-0 % buffer B (16-0% $(NH_4)_2SO_4$). The fractions containing the protein were collected and concentrated using 50K Amicon Ultra centrifugal filters. Protein solution was dialyzed with buffer A (3 rounds X 2L X2 hours each) to rid the solution of ammonium sulfate. The dialyzed supernatant was loaded onto a DEAE sepharose column (10 X 300 mm HR10 column) equilibrated with buffer A. the column was washed with 5 CV buffer A and protein was eluted with a linear gradient from 0 - 1M KCl. Protein fractions were collected and concentrated to load onto Size exclusion column (SEC) for buffer exchange and final clean-up of sample. Final fractions were collected and concentrated. Protein samples were stored at -80 ^{o}C. The protein yield was 5 mg protein/ 1 g wet cells.

6.2.8 Phosphorylation of PPDK and Δ340 PPDK

To investigate the solution conformation of the phosphorylated protein we need to phosphorylate the catalytic histidine (H455). For the WT PPDK the 1st half reaction was carried out in the absence of pyruvate. This was carried out according to the paper (5). Mass Spec shows conversion from WT to phosphorylated enzyme Figure 6.3 and 6.4. Reaction was checked periodically to determine complete phosphorylation.

A similar process was carried out to phosphorylate Δ340 PPDK. Modification of the above reaction included manipulating the 2nd partial reaction; PEP was added to protein sample. Similar method was used to observe reaction over time.

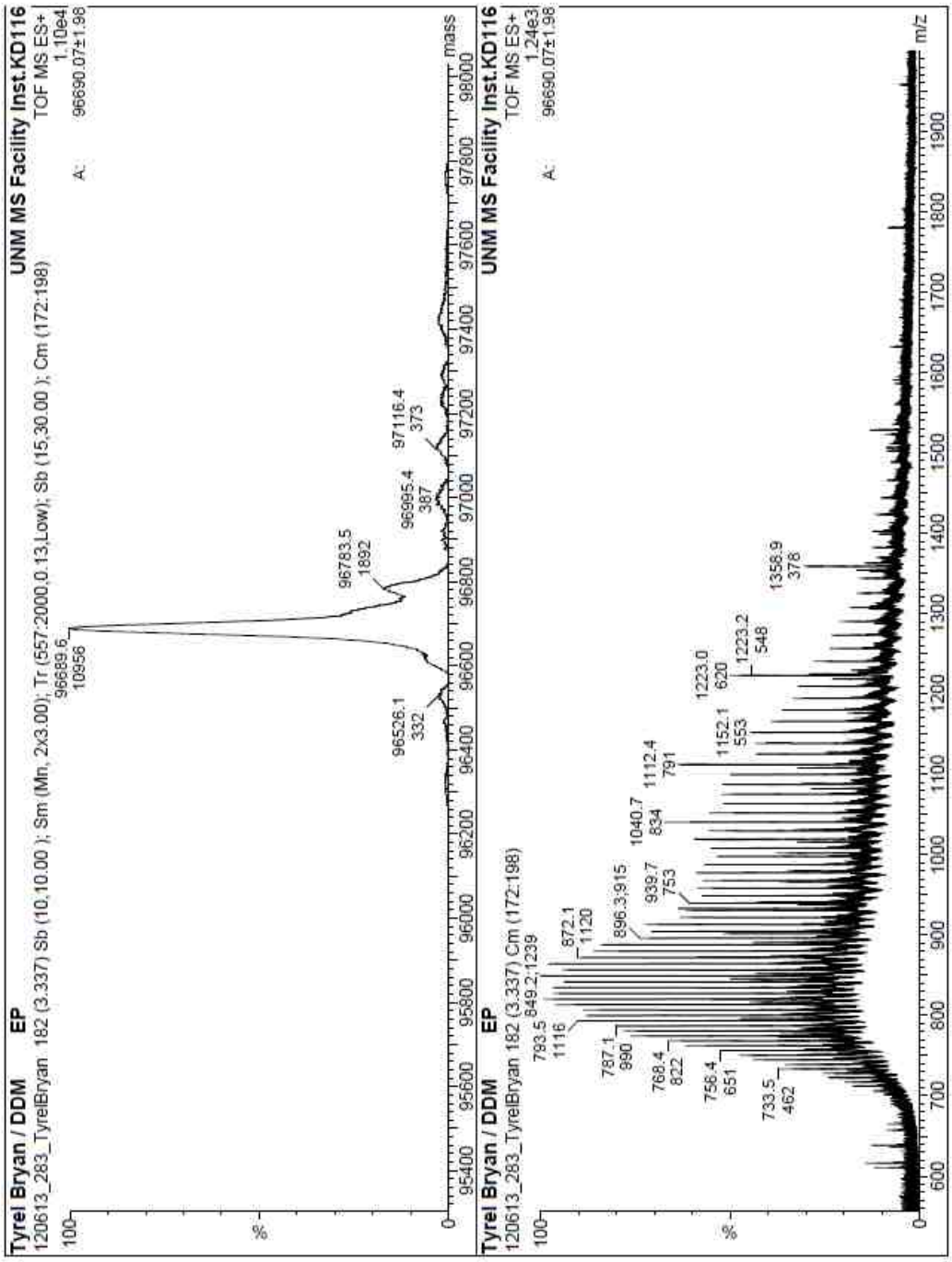

Figure 6.3: Mass Spec profile for phosphorylated PPDK. Secondary peak is the result of cleavage of N-terminal Methionine

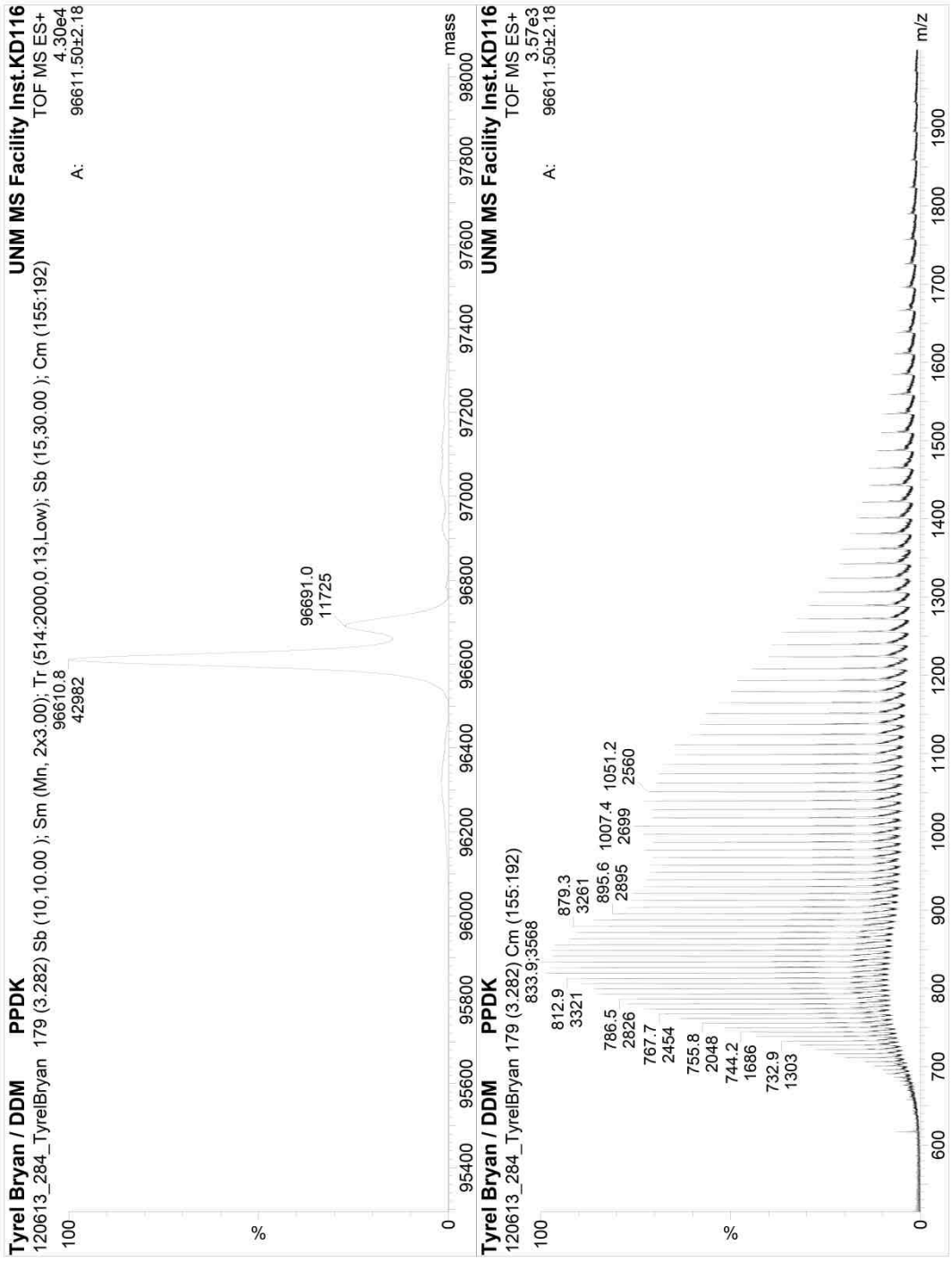

Figure 6.4: Mass Spec profile for native PPDK. Secondary peak is the result of cleavage of N-terminal Methionine

6.2.9 Small angle X ray Analysis of Solution Conformations of PPDK

Synchrotron X-ray scattering data was collected at the SSRL 4-2 BioSAXS beamline (SLAC, Menlo Park CA). The scattering patterns were tested at a range of concentration to ensure a monodisperse solution. Data was collected with a MarCCD225 detector at wavelength of 1.3 A, at a distance of 1.7 m. Data processing and evaluation was carried out using programs in the ATSAS package. Primary evaluation of parameters was carried out using standard procedures in the PRIMUS package.

6.3 Results and Discussion

6.3.1 SAXS observed data shows modified quaternary structure compared to X-ray crystal.

In cases involving proteins with large domain movement it has been normal to see changes in solution conformation compared to X-ray crystal counterparts. In the event of PPDK we know that the central domain swings from one domain to the other to complete 2 independent partial reactions. X-ray structure has been solved for both partial reactions (central domain docked at both domains). The solution structure has been identified for both conformations through site directed mutagenesis and substrate analog binding.

Theoretical scattering plots were created using CRYSOL (6). CRYSOL is used to reverse calculate scattering curves from X-ray crystal structures. Figure 6.5 shows the two scattering curves for PDB structures 1DIK and 2R82. From the scattering curves we can see distinct differences between the two conformations.

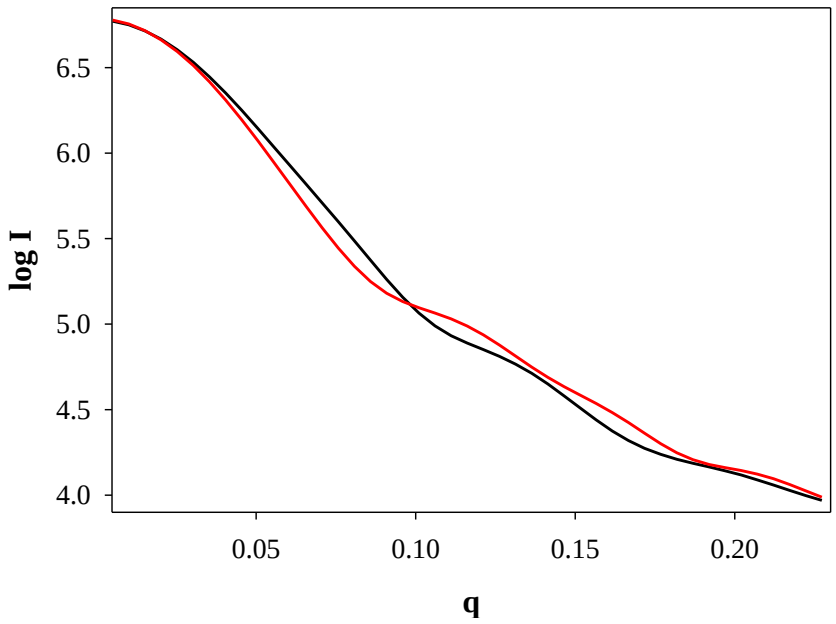

Figure 6.5: Theoretical scattering plot based on known PDB structures. BLACK: 1DIK (central domain /C-terminal; RED: 2R82 (central domain /N-Terminal). Scattering curves were generated using CRYSOL.

6.3.2 PPDK shows distinct solution conformations in native state upon binding 1st partial reaction transition state intermediate: AMPPNP

The first partial reaction is the phosphorylation of the His455. The nucleotide binding domain binds ATP to phosphorylate histidine. AMPPNP has previously been shown to be a good inhibitor toward the 1st partial reaction. Binding AMPPNP (10x k_i) reveals a different solution conformation than the native PPDK (Figure 6.6-6.10)

Although results show that the native protein without any ligand shows a structure similar to theoretical volume of central domain interacting with C-terminal domain (2R82); when the native protein is tested with substrate analog AMPPNP the structure

shifts to look more like the native X ray crystal structure (1DIK) in which the central domain is interacting with the N-terminal domain.

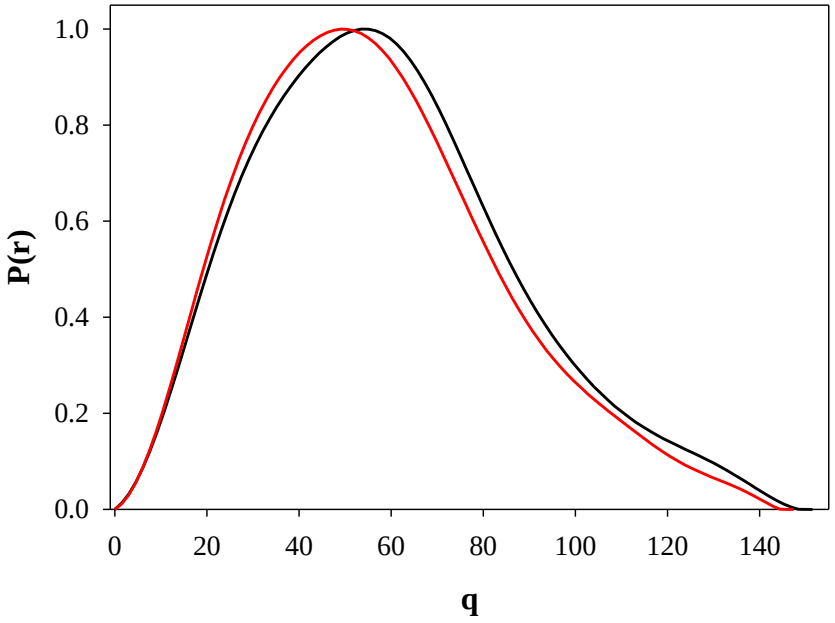

Figure 6.6: Native PPDK with and with Substrate inhibitor AMPPNP. The protein shows a shift in conformational structure when in the presence of ligand. BLACK: native protein without AMPPNP, RED: native protein with AMPPNP.

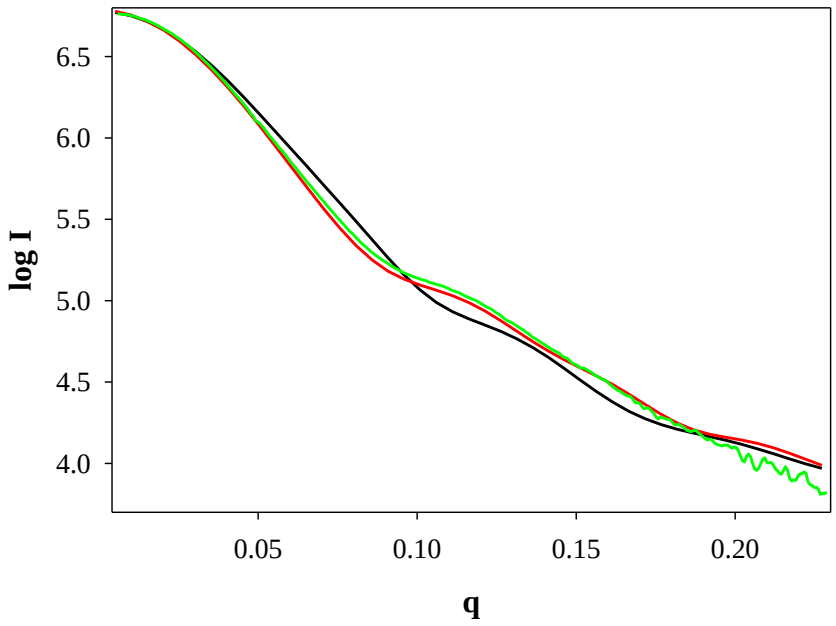

Figure 6.7: Experimental scattering curve of native PPDK without substrate, AMPPNP compared to PDB generated scattering curves. BLACK: 1DIK (central domain /C-terminal; RED: 2R82 (central domain /N-Terminal); and GREEN: Experimental data.

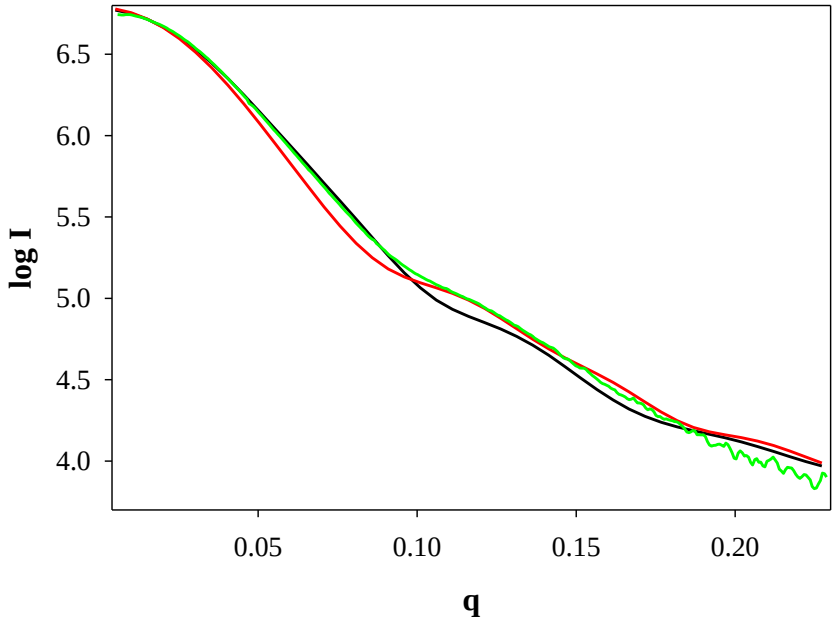

Figure 6.8: Experimental scattering curve of native PPDK with substrate, AMPPNP (10X K_i) compared to PDB generated scattering curves. BLACK: 1DIK (central domain /C-terminal; RED: 2R82 (central domain /N-Terminal); and GREEN: Experimental data.

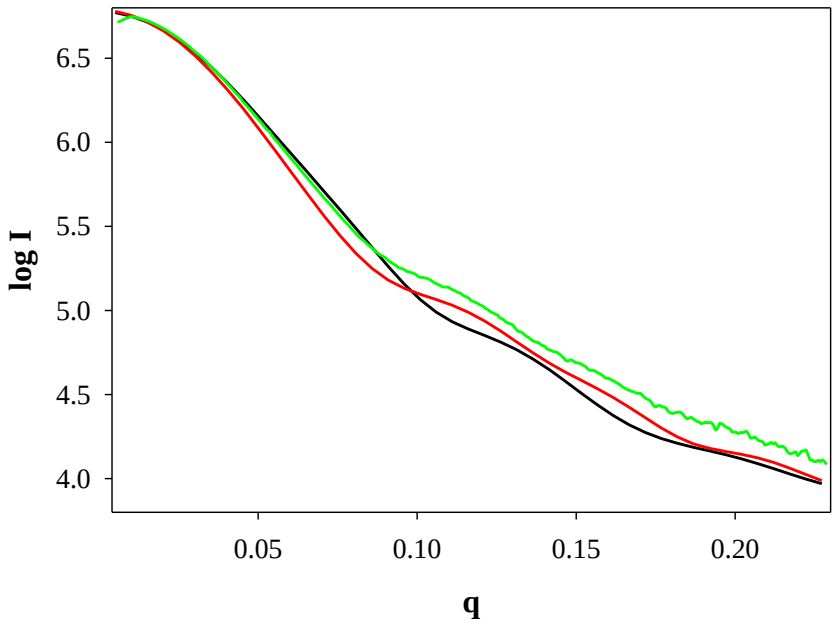

Figure 6.9: Experimental scattering curve of phosphorylated PPDK without substrate, AMPPNP compared to PDB generated scattering curves. BLACK: 1DIK (central domain /C-terminal; RED: 2R82 (central domain /N-Terminal); and GREEN: Experimental data.

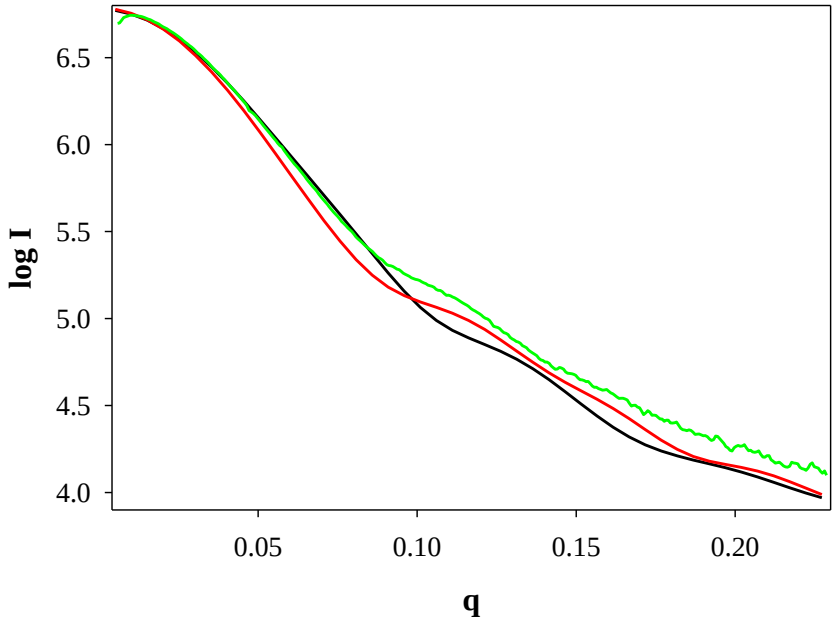

Figure 6.10: Experimental scattering curve of phosphorylated PPDK with substrate, AMPPNP (10X K_i) compared to PDB generated scattering curves. BLACK: 1DIK (central domain /C-terminal; RED: 2R82 (central domain /N-Terminal); and GREEN: Experimental data.

6.3.3 PPDK shows two distinct conformations according to phosphorylation of His-455

Solution conformations were observed using SAXS of both the native PPDK and the phosphorylated structures. From known X ray structures (PDB: 1DIK and 2R82) we see PPDK in two distinct conformations: N-terminal/central domain and C-terminal/central domain.

Both samples were tested at a variety of concentrations (0.5 -30 mg/mL) to ensure no concentration dependence. There are four possible conformational changes: native: empty, native: substrate, phosphorylated: empty, and phosphorylated: substrate. From the observed scattering we see two predominant structures (Table 6.1, Figure 6.11). The structures seen are from the native PPDK with substrate analog AMPPNP and the three other combinations showing the other scattering volumes.

From the scattering experiments we can see that that the native PPDK shows conformation similar to the phosphorylated PPDK structures. While the native structure shows two separate conformations, the phosphorylated central domain have only one conformation regardless of empty or bound substrate analog (oxelate).

Protein	Ligand	$^{1}R_g$	$^{2}R_g$	$^{2}d_{max}$
WT	EMPTY	43.804 ± 0.413	45.96	153.32
	AMPNP	41.415 ± 0.599	43.76	144.95
Phosphorylated PPDK	EMPTY	41.449 ± 1.226	43.41	145.07
	Oxelate/Mn^{2+}	42.785 ± 0.471	44.48	149.75
Tem340 PPDK	EMPTY	22.851 ± 0.150	22.81	79.98
	AMP	21.795 ± 0.191	21.74	75.78
	AMPPNP	22.253 ± 0.158	22.10	72.88
Tem540 PPDK	EMPTY	32.109 ± 0.353	32.58	112.38
	AMP	30.163 ± 0.857	30.44	105.57
	AMPNP	29.471 ± 0.856	29.92	103.15
Cent PPDK	EMPTY	16.907 ± 1.599	17.72	63.17

Table 6.1: Displayed are all the scattering values for the PPDK constructs

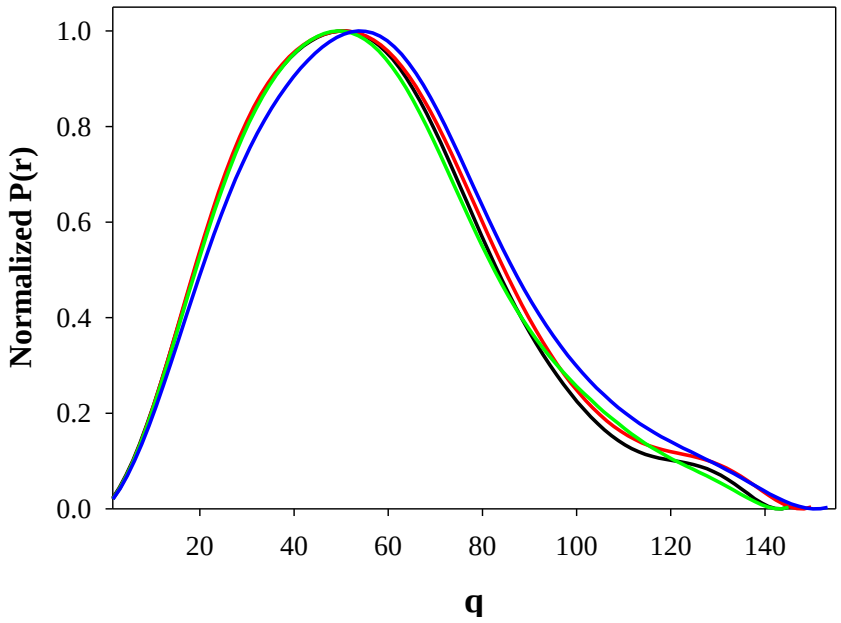

Figure 6.11: P(r) distributions of native and phosphorylated PPDK with/without substrate analogs. GREEN: native PPDK without substrate analog (AMPPNP), BLUE: native PPDK with substrate analog (AMPPNP), BLACK: phosphorylated PPDK without substrate analog (oxelate), and RED: phosphorylated PPDK with substrate analog (oxelate).

6.3.4 Nucleotide binding domain shows closure upon introduction of substrate and analogs

The N-terminal domain consisting of residues 1-340 fold into a nucleotide binding domain where the 1st partial reaction occurs; transfer of β-phosphoryl of ATP to hisitne (455) forming a pyrophosphorylated enzyme intermediate. Scattering data shows conformational change in nucleotide binding domain upon binding substrate analog AMP and transition state analog AMPPNP. SAXS studies show 3 distinct conformation (empty, AMP, AMPNP) suggesting conformational change due to activity. The nucleotide binding domain goes through a closure upon binding substrate this can be observes from the decrease in size (as seen through R_g and volume max, d_{max}) as

substrates bind from empty>AMP>AMPNP, Figure 6.12 and Table 6.1 shows the modification in P (r) distribution according to binding.

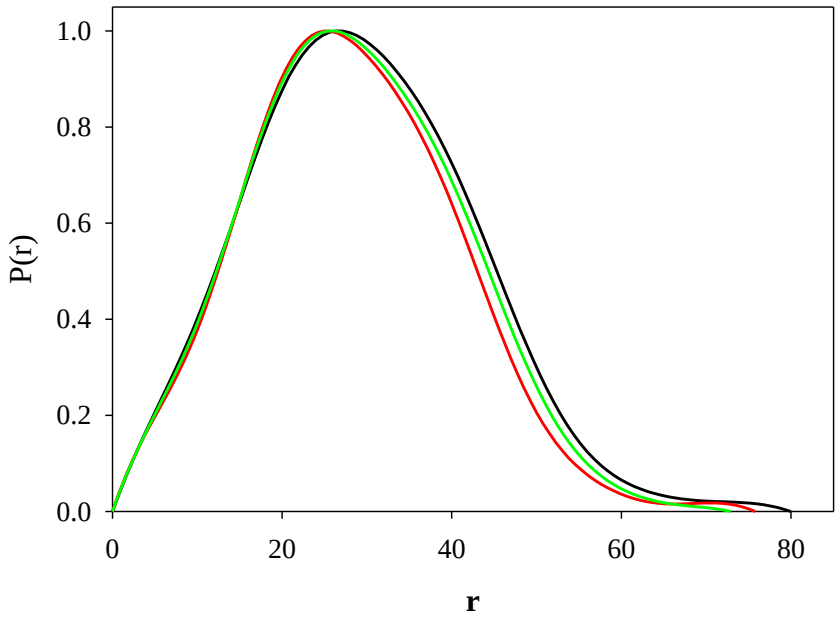

Figure 6.12: Truncated PPDK, Tem340, w/wo substrate analogs. BLACK: Tem340 empty, RED: Tem340 with substrate analog (AMP), and GREEN: Tem340 with substrate analog (AMPPNP).

6.3.5 Truncated PPDK (Tem540) shows modification in volume due to ligand binding.

Tem 540 is a truncated mutant of PPDK in which the C-terminal domain in cleaved. SAXS was collected on the truncated protein in the absence and presence of AMP and AMPNP to identify change in the conformation of the protein; more specifically does central domain interact with the N-terminal or is it in solution. SAXS data shows a decrease in d_{max} upon ligand introduction (empty>AMP=AMPPNP) with empty having the largest d_{max}.

When looking at solution conformations of proteins with drastic domain shifts, it is important to identify equilibrium positions. This truncation was used to investigate if the central domain would interact with the N-terminal domain without the presence of the C-terminal domain. We have already seen that the central domain possibly sets close to the C-terminal domain at equilibrium but as ligands are introduced the conformation shifts to the central/N-terminal conformation.

From the collected scattering data we can see the protein created a larger volume when compared to the N-terminal alone, d_{max} of 79.98 to 112.38 nm for TEM340 and TEM540 PPDK respectively. From creating theoretical distances, longest distances from end-to-end from PDB structures, of the central domain interacting w/wo N-terminal domain we have concluded that the largest distance is from ~94 nm for 2R82 in which central domain is interacting with C-terminal, and ~75 nm for 1DIK (central + N-terminal) (Figure 6.13).

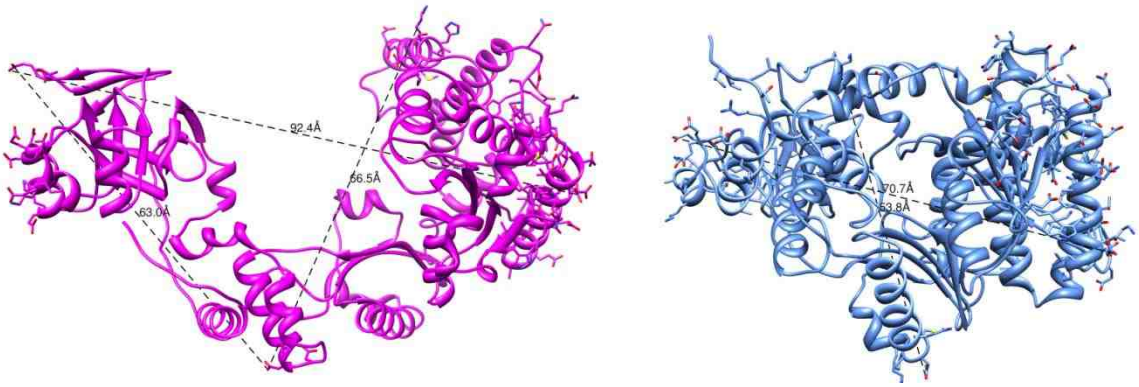

Figure 6.13: Theoretical Tem540 truncations created by deleting N-terminal domain of known PDBs (MAGENTA: 2R82 and BLUE: 1DIK). Distances were created to create estimate of d_{max} in PDB structures. 2R82 longest distance ~92 nm and 1DIK longest distance ~70 nm.

From the d_{max} derived from the calculated protein volumes we see that the central domain indeed does not interact in the same way as our theoretical models. This holds

both good news and bad news, as our models are created from using truncated X ray structures which does not accurately portray how the protein acts without the N-terminal domain, although our lengths are larger than the theoretical, it does tell us that the central domain in solution does not interact as closely as thought (Figure 6.14)

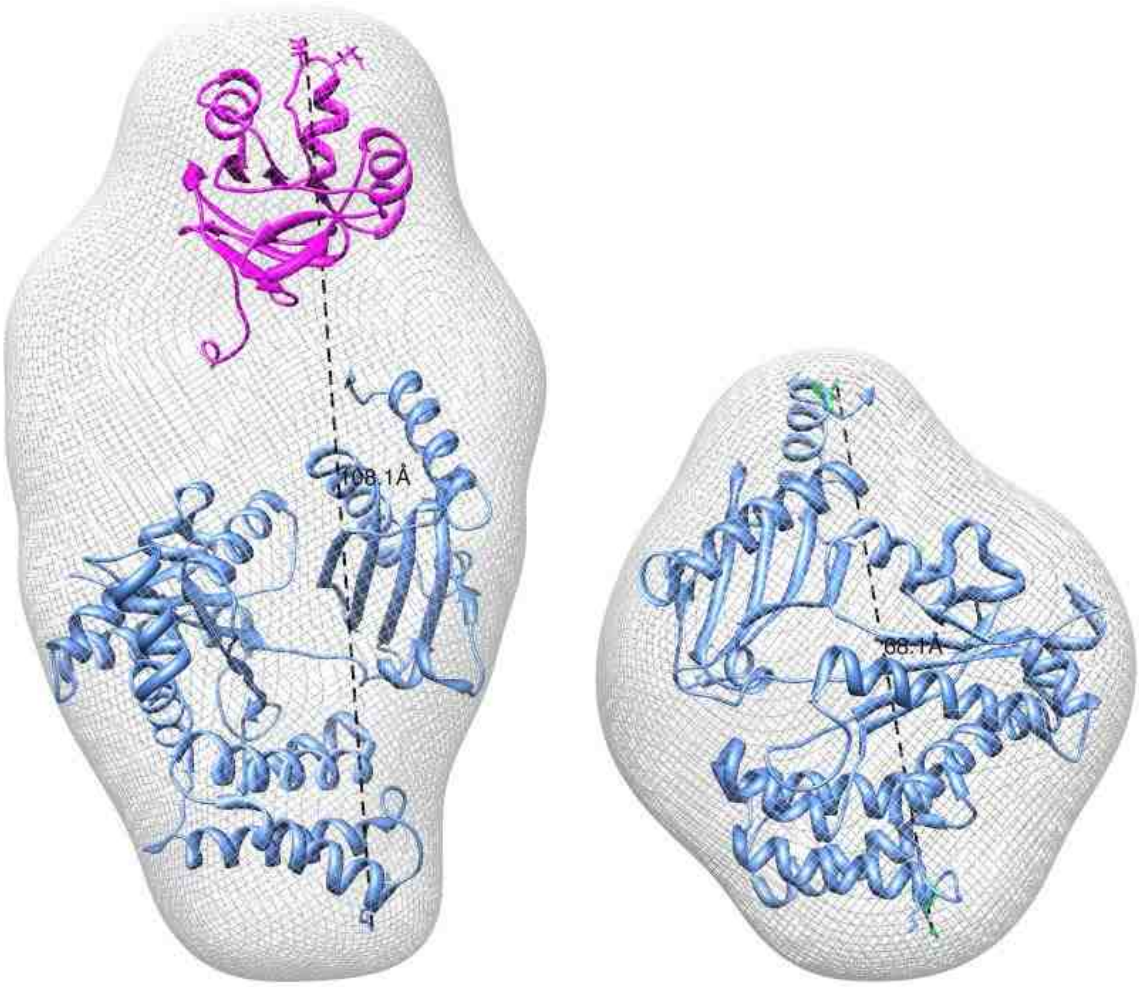

Figure 6.14: Displayed volumes with truncated PDB structures fit to volumes. BLUE X ray crystal structures are the N-terminal domain from PDB: 2R82 and MAGENTA: X ray crystal structure of central domain (PDB: 2FM4. Volumes (left to right) are TEM540 with substrate analog AMPPNP and TEM340 with substrate analog AMPPNP

solution of the Tem540 PPDK truncation does not interact closely we do see 'closure' of the d_{max} length upon ligand binding.

6.3.6 Truncated central domain folds completely

The central domain is responsible for connecting the two partial reactions. It has previously been characterized and crystallized separately (PDB: 2FM4). For completeness and to investigate the structure in solution SAXS data was collected on CentI PPDK domain Figure 6.15.

Scattering data supports a folded central domain. Figure 6.16 shows PDB 2FM4 fit into solution volume created from central domain scattering data. Protein was used to investigate interaction between it and TEM340 PPDK truncation.

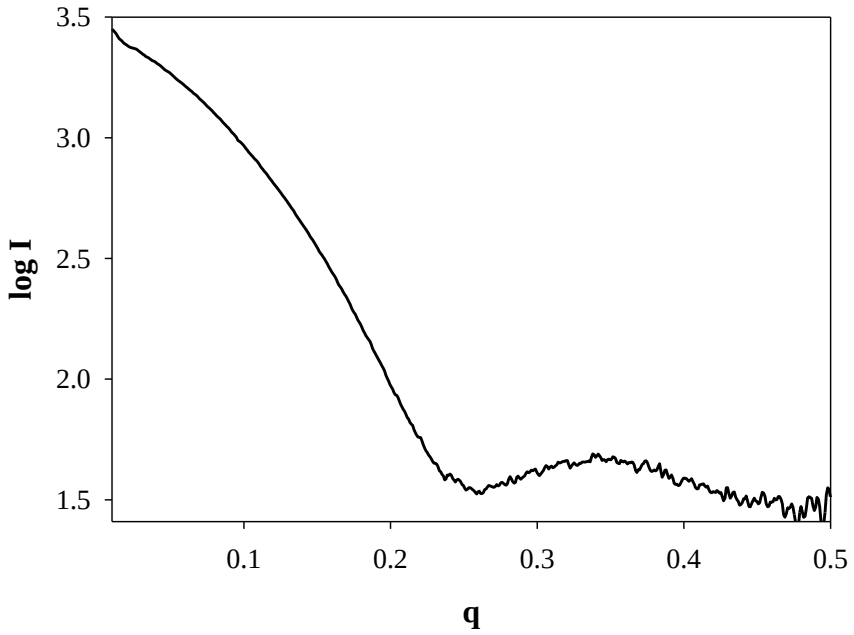

Figure 6.15: Scattering plot of CentI.

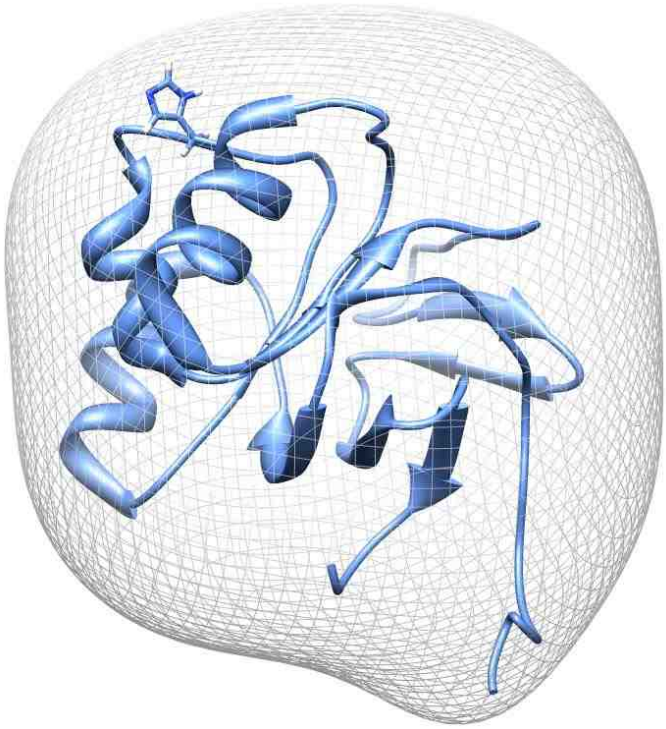

Figure 6.16: Central domain of PPDK (PDB: 2FM4) fit within calculated volume from scattering data.

6.4 Conclusion

PPDK has been thoroughly investigated mechanistically, but some questions have yet to be answered. The central domain swings between the two other domains and connects two partial reactions, but how does the domain interact and move between domains? There are two possibilities in which the central domain can move: 1. move through solution between the two reaction centers or 2. 'roll' between reaction centers while continuously making contact with protein. To investigate the conformational changes in PPDK SAXS was used.

From the collected scattering data we can see that the PPDK's solution conformations are not as cut and dry as the two published structures (PDB 1DIK and 2R82). We can see that the protein in the native form has a conformation similar to

PDB:2R82 rather than 1DIK. From our data we see that the native PPDK conformation equilibrium favors the PDB: 2R82 structure when substrate analog is not present, but when present the other conformation is favored (PDB: 1DIK), after phosphorylation of His455 the 2R82 structure is favored again.

From the truncated domains we see folded functional forms of both the N-terminal grasp domain and central domains. From the N-terminal domain separate we can identify change in conformation due to binding. The more unique truncation of the bundle is the N-terminal-central domain truncation (Tem540 PPDK). From the SAXS data collected on the Tem540 PPDK construct we can see an elongated structure when compared to the N-terminal only which suggests minimal interactions among the two domains. Although we cannot rule out interaction the SAXS data allows us to infer the central domain may transverse through solvent between active sites rather along the protein surface. For a complete picture deletion mutations of linkers between central domain and rest of protein will be addressed to investigate solution conformation and position of central domain compared to native and phosphorylated PPDK.

References

1. Lim, K. Read, R. Chen, C C H. Tempczyk, A. Wei, M. Ye, D. Wu, C. Dunaway-Mariano, D. Herzburg, O. (2007) “Swiveling Domain Mechanism in Pyruvate Phosphate Dikinase” Biochemistry, Volume 45, Issue 51, 14845-14853.

2. Herzberg, O., Chen, C. C., Kapadia, G., McGuire, M., Carroll, L. J., Noh, S. J., & Dunaway-Mariano, D. (1996). Swiveling-domain mechanism for enzymatic phosphotransfer between remote reaction sites. *Proceedings of the National Academy of Sciences*, *93*(7), 2652-2657.

3. Walker, R. G., X. Deng, et al. (2014). "The Structure of Human Apolipoprotein A-IV as Revealed by Stable Isotope-assisted Cross-linking, Molecular Dynamics, and Small Angle X-ray Scattering." Journal of Biological Chemistry **289**(9): 5596-5608.

4. Mertens, H. D. T. and D. I. Svergun (2010). "Structural characterization of proteins and complexes using small-angle X-ray solution scattering." Journal of Structural Biology **172**(1): 128-141.

5. Thrall, S. H. and D. Dunaway-Mariano (1994). "Kinetic Evidence for Separate Site Catalysis by Pyruvate Phosphate Dikinase." Biochemistry **33**(5): 1103-1107.

6. Svergun, D., Barberato, C., & Koch, M. H. J. (1995). CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography*, *28*(6), 768-773
7. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H., & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, *36*(5), 1277-1282.